Drilling Deeper Into the STR Allele: Enhanced Resolution and Statistical Power Through SNP Distributions Within the Short Tandem Repeats

John V. Planz[1], Bruce Budowle[2], Kristin Lowery[3], Arthur J. Eisenberg[1], Steven A. Hofstadler[3], and Thomas Hall[3]
[1] UNT Center for Human Identification, Laboratory for Molecular Identification
University of North Texas Health Science Center at Fort Worth, Ft. Worth, TX
[2] Federal Bureau of Investigation, Quantico VA
[3] Ibis Analytical Services Laboratory, Ibis Biosciences, Inc., Carlsbad, CA

The development, validation, and testing of short tandem repeats (STRs) has established a solid foundation for the evaluation of DNA samples from evidentiary items, convicted offenders, relationship testing analysis and numerous other genetic testing disciplines. The statistical power of the 13 or more loci commonly evaluated in forensic investigations is sufficient in most human identity testing comparisons dealing with unknown evidentiary materials and known exemplars. However, in evaluations that cross generational levels, such as in parentage testing or the identification of unidentified human remains, or in pairwise evaluations targeting the possible identification of a relative, the STR loci by themselves have distinct limitations. Most common of these is encountered during parentage evaluations, where the relatively high mutation rate of the STR loci increases the incidence of observing of single non-matching locus between a child and a true biological parent. These single inconsistencies pose a dilemma in that there are alternate explanations to explain the DNA results: non-parentage, true parentage given a mutational event, and existence of an alternate parent who is related to the tested putative parent and this relative is the true biological parent of the child in question. Coupled with this issue is the concept of "familial" association by virtue of a degree of shared STR alleles. A major assumption being made is that the alleles observed, either between a child and the non-excluded alleged parent or in two individuals associated in a database search, stem from common ancestry, i.e. are *Identical by Descent* (IBD). The Stepwise Model of evolution which is consistent with manner that STR alleles are generated, coupled with the rapid rate of mutation, makes it clear that many of the familial relationship assumptions are based only on STR allele associations which are Identical by State (IBS). The magnitude of this limited power becomes greater when evidentiary items do not produce a full complement of loci, as can often be the case with forensic evidence or unidentified human remains. However, there may be additional information within an STR allele which can aid in discerning whether the alleles shared between samples stem from common ancestry or are simply the result of commonality of a particular amplicon length at a locus within the population, i.e. IBS. Using the fully automated high throughput electrospray ionization mass spectrometry (ESI-MS) system developed by Ibis Biosciences Inc., population databases have been developed for the 13 CODIS core STR loci for African Americans, US Caucasians and Southwestern Hispanics which capture both the nominal repeat number of the allele, as well as SNP variation within the repeat motifs and flanking areas. Prevalence of SNPs among the STR loci has been observed to be as high as 40 percent in some loci. The presence of these internal polymorphisms at one or more loci now creates alleles that may contribute additional value in familial evaluations. The statistical effect of this expanded definition of an allele among the commonly used loci is discussed, and results from pedigree evaluations and parentage cases that contained single inconsistencies which have been re-evaluated in light of the additional data will be presented.