## STR Sequence Diversity in Population Samples and Nomenclature Guidance for the "Next Generation"



| ABSTRACT <br> As STR loci were being identified in the 1990s, various nomenclature systems were developed for different loci, with the primary variation being whether or not to "count" nondeveloped for different loci, with the primary variation being whether or not to "count" non- repeat bases interspersed in the repeat motif. In 1997, the ISFG issued guidelines on STR nomenclature, in an attempt to provide a common currency for information exchange. Historical precedent already existed for some loci, and this was maintained to avoid confusion, resulting in several commonly used forensic loci having complicated and contradictory nomenclature systems. This has not been an issue within the forensic community, as the capillary electrophoresis (CE)-length analyses are kit-based, with corresponding computer programs that automatically count repeats in a standardized manner. Now, as the costs associated with next-generation sequencing (NGS) methods sequencing STR loci may provide. As a new generation of scientists begins interrogating these loci on a deeper level, an understanding of historical nomenclature is needed to achieve bioinformatic concordance with existing CE data. In the work presented here, NG results from population samples exemplify the sequence variation that exists in forensic STR loci (SNPs and InDels within and outside of STR allele regions and repeat motif changes) as well as the complexity and inconsistency of the current nomenclature. This experimental and provides examples of how sub-alleles can improve discrimination and mixture deconvolution in forensic casework. The different purposes of nomenclature-manal comparisons, forensic reports, database searching, court explanations-are discussed and examples of possible NGS-compatible nomenclature systems that may meet the needs of the forensic community are shown. |
| :---: |
|  |  |
|  |  |



Figur 1. Overiene of experimental desigr








 sequence within the repeat region were excluded from further analysis). Genotypes from both ExactID and STRaithazor were independently analyzed for concordance to $C$ b based
genotypes (generated previously with Powerplex Fusion (Promega)). Discordances were genotypes (generated previously with PowerPlex fusion (
evaluated further to determine the true genotype/sequence.


