DETECTION OF INTRA-ALLELIC SEQUENCE VARIANTS IN AUTOSOMAL AND X CHROMOSOME SHORT TANDEM REPEATS USING MASSIVELY PARALLEL SEQUENCING

Nicole M. Novroski, David H. Warshauer, Jonathan L. King, Xiangpei Zeng, Jennifer D. Churchill, Bruce Budowle

Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Boulevard, Fort Worth, TX 76107, USA

Forensic DNA typing routinely employs amplification using the polymerase chain reaction (PCR) followed by size separation by capillary electrophoresis (CE) to analyze and discriminate short tandem repeats (STRs). Although this method is robust and reliable, it is limited in both resolution and scalability. Massively parallel sequencing (MPS) is an alternative analytical methodology that allows for the generation of a large amount of data using minimal input DNA. Using this technology, genetic profiles of a large number of markers, including autosomal, X, and Y-STRs for multiple individuals (up to 96) can be generated in one sequencing run. The benefit is greater genetic information per individual sample pertaining to identification. A beneficial feature of using MPS is that it can identify and characterize sequence variation present within STR alleles – a capability not currently possible using traditional DNA typing by CE. This additional level of diversity could lead to better DNA mixture resolution and potentially increase the ability to develop investigative leads from forensic samples.

In this study, intra-allelic sequence variants that reside within STR repeat regions of 31 autosomal and 26 X chromosome markers were identified. Sequence variants were classified as either nominal allele variants or novel allele variants, where their classification was dependent on the existence of previously published sequence data for the particular alleles. In addition, variants were described as SNP-based or motif-based. Probes for the markers were generated using the Illumina® DesignStudio sequencing assay design tool, and the targets were captured with the Nextera™ Rapid Capture Custom Enrichment panel (Illumina, San Diego, CA). MPS was performed using the MiSeq™ Desktop Sequencer (Illumina) and analysis of FASTQ files was performed using STRait Razor v2.0 for a sample population of 190 unrelated individuals. These findings illustrate the potential of genetic variation that exists in the most commonly known STR markers and reinforces the need for further exploration of intra-allelic sequence variants for their use in forensic DNA typing.