DEVELOPMENT OF A COMPREHENSIVE PANEL OF SHORT TANDEM REPEAT AND SINGLE NUCLEOTIDE POLYMORPHISM MARKERS FOR HUMAN IDENTIFICATION USING MASSIVELY PARALLEL SEQUENCING TECHNOLOGY

Xiangpei Zeng[1], David H. Warshauer[2], Carey P. Davis[1,3], Jonathan L. King[1], Jennifer D. Churchill[1], Nicole Novroski[1], Frank R. Wendt[1], Ranajit Chakraborty[1], and Bruce Budowle[1,4]

[1]Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center
[2]Promega Corporation
[3]Illumina, Inc.
[4]Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University

Over the past two decades, capillary electrophoresis (CE) has been used as the gold standard method for short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs) typing in forensic genetics. However, CE technology has limitations, e.g., a limited number of markers can be multiplexed, different types of markers (STRs and SNPs) cannot be identified simultaneously, and sequence variants within STRs cannot be revealed. With the high throughput of massive parallel sequencing (MPS), it is possible to sequence a large number of markers in multiple samples. In addition, STRs and SNPs can be multiplexed in the same reaction, which allows greater discrimination power.

In this study, a comprehensive panel of STRs and SNPs was designed using Nextera™ Rapid Capture Custom Enrichment kit (Illumina). This panel consisted of 84 STRs (31 autosomal, 26 X-chromosomal, and 27 Y-chromosomal) and 275 SNPs (240 autosomal and 35 Y-chromosomal) markers. The panel was used to type 178 samples from four major US populations: African Americans, Asians, Caucasians, and Hispanic Americans. Sequencing was performed on the MiSeq™ instrument (Illumina). FASTQ files were analyzed using STRait Razor v2.0 to display STR alleles. SNP alleles were identified using Burrows-Wheeler Aligner (BWA), Sequence Alignment/Map Tools (SAMtools), and the Genome Analysis Toolkit (GATK). A concordance study was conducted to compare the genotype data generated by Nextera™ Rapid Capture Custom Enrichment kit with those of ForenSeq™ DNA Signature Prep Kit (Illumina). The genotype data were concordant with that of ForenSeq kit. For the STRs, relative locus performance (RLP) and allele coverage ratio (ACR) were assessed, and sequence variants were revealed. RLPs and ACRs of STRs were similar to those of the commercial PCR-based MPS kit. In addition to the RLPs and ACRs, statistical analyses also were performed to determine the SNP allele frequencies in four populations. Hardy-Weinberg equilibrium (HWE) and Linkage Disequilibrium (LD) were tested for the SNP markers. Y-STR haplotype diversity and haplogroups were predicted for each male individual based on the genotype data, and the results were consistent with the known ancestries. The results of this study indicate that a probe-based method can be used to construct a comprehensive panel of different types of markers and generate reliable data. Therefore, capture-based MPS technology can be considered as an alternative method to the PCR-based MPS kits, at least for analyzing reference samples.