Investigation into the sequence structure of 23 autosomal STRs and 23 Y-STRs using massively parallel sequencing

So Yeun Kwon[1,2] · Eun Hye Kim[1,2] · Eun Young Lee[1] · Hwan Young Lee[1] · Kyoung-Jin Shin[1,2]

[1]Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea
[2]Brain Korea 21 PLUS Project for Medical Science, Yonsei University, Seoul, Korea

Introduction

Next-generation sequencing (NGS) can produce massively parallel sequencing (MPS) data for many targeted regions with a high depth of coverage, suggesting its successful application to the analysis of forensic genetic markers. In this regard, we evaluated the practical utility of MPS in autosomal short tandem repeat (STR) and Y-chromosomal STR (Y-STR) analysis using two in-house multiplex PCR systems. The first multiplex PCR system simultaneously amplified 23 autosomal STRs, DYS391 and amelogenin with the small-sized amplicons ranging from 77 to 217 bp. The second multiplex PCR system simultaneously amplified the PowerPlex® Y23 loci and the M175 with the small-sized amplicons ranging from 85 to 253 bp. We analyzed 250 unrelated Korean samples with MPS method using these two multiplex PCR system. Hence, we present the number of the observed sequence variations of the forensic markers through MPS analysis compared with the conventional capillary electrophoresis analysis.

Materials and Methods

1. DNA samples

DNA was extracted from buccal swab samples of 250 unrelated Korean samples using QIAamp DNA Mini Kit (Qiagen) and 1 ng/ul of DNA was prepared for test.

2. Procedure for MPS analysis

1) PCR-based library preparation
1 ng of the sample and appropriate concentration of target specific primers were used for the first PCR. In the second PCR, 1.0 µl of 100-fold diluted the first PCR products and Nextera XT Index Kit (Illumina) were used. The thermal cycling was conducted as following condition.

1st PCR (Target-specific PCR)	2nd PCR (Barcoding PCR)
Autosomal; 29 cycles Y chromosomal; 30 cycles	Autosomal; 15 cycles Y chromosomal; 17cycles

2) Following PCR cleanup with 1.2× Agencourt® AMPure® XP beads (Beckman Coulter), the libraries were quantified using KAPA library quantification kits (KAPA Biosystems) and Agilent 2100 Bioanalyzer.

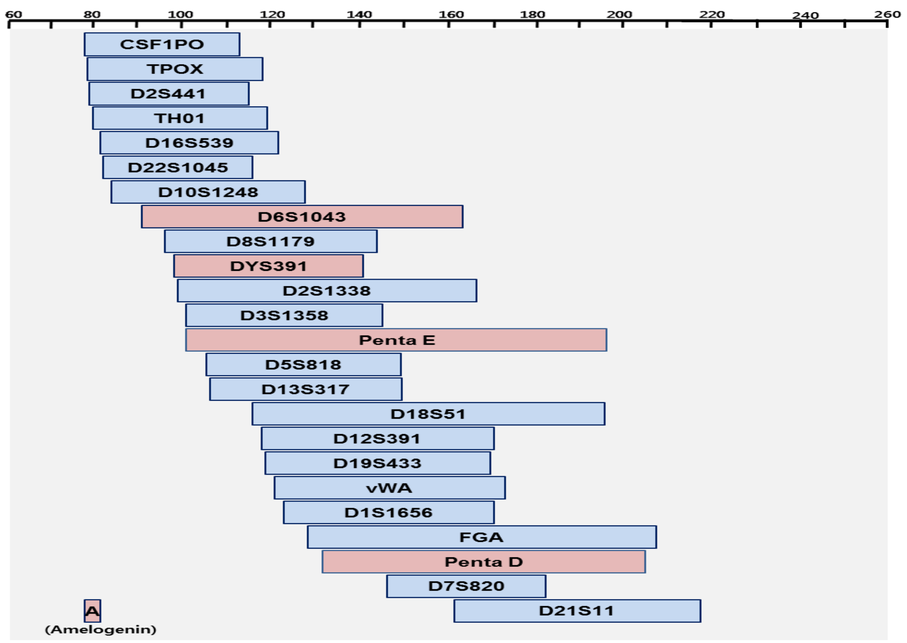
3) The barcoded libraries were normalized to 10nM and then pooled in equal volumes. Finally, the pooled library was sequenced on MiSeq™ (Illumina) using a MiSeq Reagent Kit (Illumina).

3. MPS data analysis

FASTQ files generated through MPS with 250 unrelated Korean samples were analyzed by two bioinformatics methods, STRait Razor and in-house pipeline for MPS data processing. In-house pipeline included the alignment process using BWA software and a program coded with C# language. The genotyping concordance between MPS and the capillary electrophoresis method, as well as the sequence structure of the 23 autosomal STRs and 23 Y-STRs, were investigated.

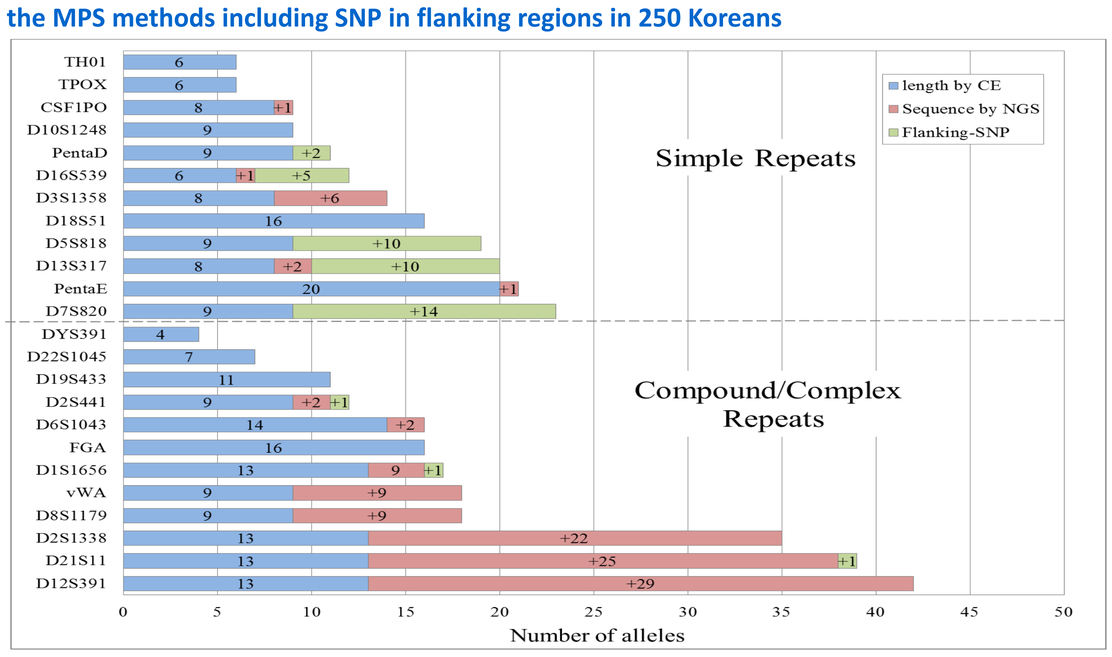
Results

Fig. 1a. Allelic size range of 23 autosomal STRs, DYS391 and amelogenin



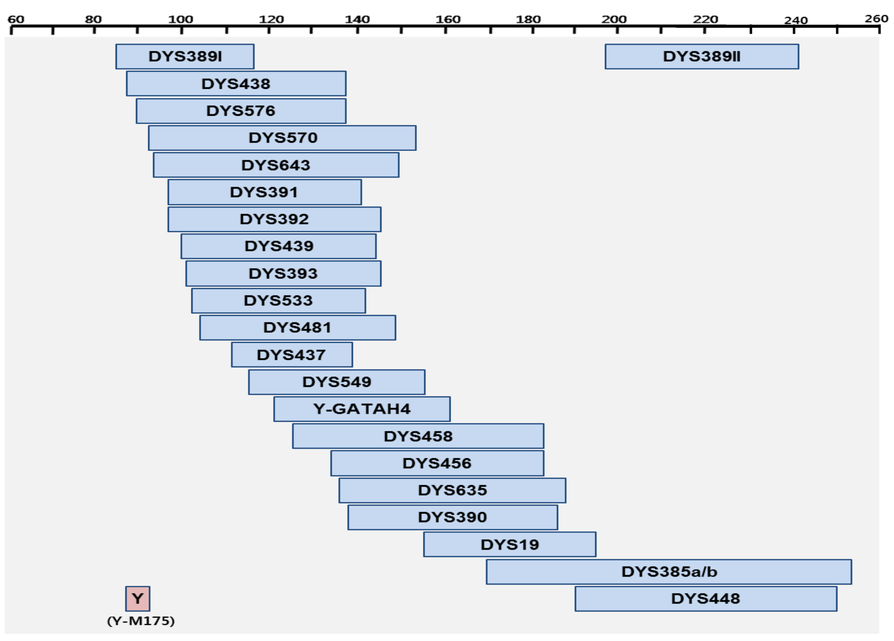
The 20 expanded CODIS core loci are marked in blue boxes and additional 4 loci and amelogein in red boxes. The amplicon sizes of all the targeted markers were ranging from 77 to 217 bp.

Fig. 1b. Comparison of the number of observed alleles by size with CE and by sequence with the MPS methods including SNP in flanking regions in 250 Koreans



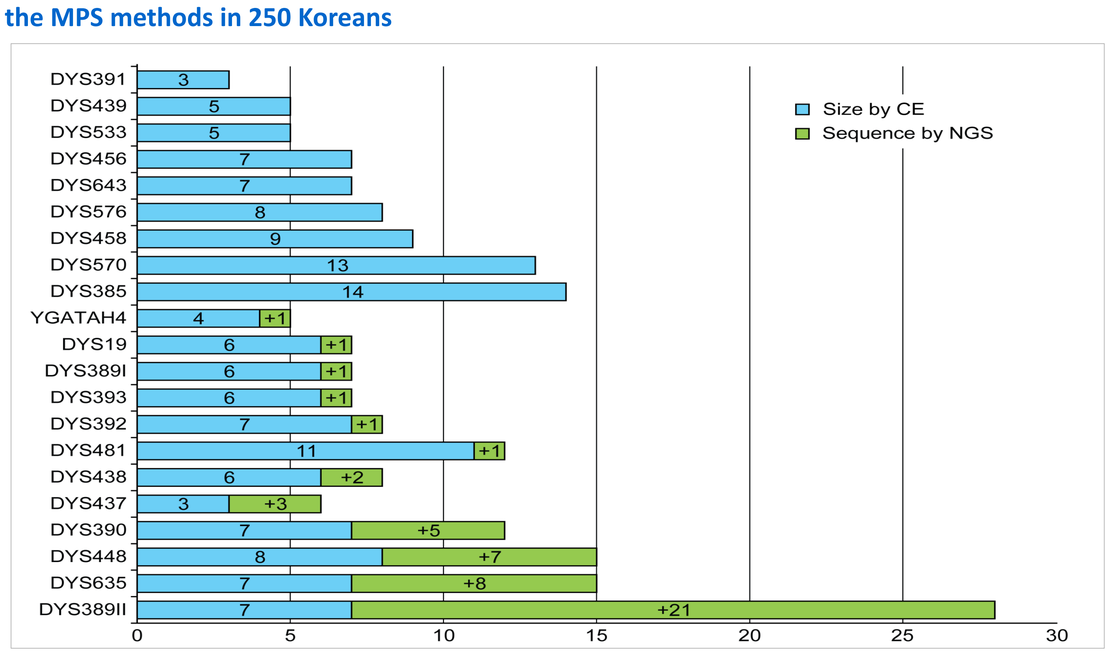
The increased number of alleles by MPS was more than doubled what was obtained by CE at D16S539, D5S818, D13S317, D7S820, vWA, D8S1179, D2S1338, D21S11 and D12S391. Another seven loci showed gains in the numbers of alleles obtained by MPS.

Fig. 2a. Allelic size range of 23 Y chromosomal STRs and Y-M175



The PowerPlex Y23 loci are marked in blue boxes and Y-M175 in a red box. The amplicon sizes of all the targeted markers were ranging from 85 to 253 bp.

Fig. 2b. Comparison of the number of observed alleles by size with CE and by sequence with the MPS methods in 250 Koreans



The increased number of alleles by MPS was more than doubled what was obtained by CE at DYS448, DYS635 and DYS389II. Another nine loci showed gains in the numbers of alleles obtained by MPS. (Forensic Sci Int Genet. 2016;25:132-141)

Conclusion

- We constructed two multiplex PCR system of 23 autosomal STRs and 23 Y chromosomal STRs optimized for MPS analysis with small-sized amplicon.
- These MPS system could discovered the sequence variations in repeat region as well as in flanking region of the forensic genetic markers.
- Therefore, the multiplex PCR systems of the present study will facilitate the MPS analysis of forensic genetic markers, and the observed sequence variations of global STRs through MPS analysis could be used to increase the discrimination power of STR loci in resolving of challenging caseworks.