A New Multiplex PCR System for Forensic STR Profiling Using Next Generation Sequencing

KJ Shin, EH Kim, IS Yang, SE Jung, HY Lee

Department of Forensic Medicine
Yonsei University College of Medicine
Seoul, Korea

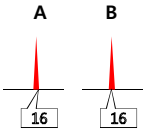
Current STR typing in forensic genetics

Short tandem repeat analysis using capillary electrophoresis

- Limited **a total of number** and **allelic size** of STRs according to available fluorescence dyes
- Can not identify **sequence variation** in STRs due to size based separation
- Digital genotyping of **mixed samples** is hard

Application of next generation sequencing to forensic STR typing

Next generation sequencing



<CE method>

A .....TCTATCTGTCTGTCTGTCTATCTATCTA..... 16 allele

B .....TCTATCTGTCTGTCTATCTATCTATCTA..... 16 allele (G>A)

<NGS method>

Previous studies for STR analysis using NGS

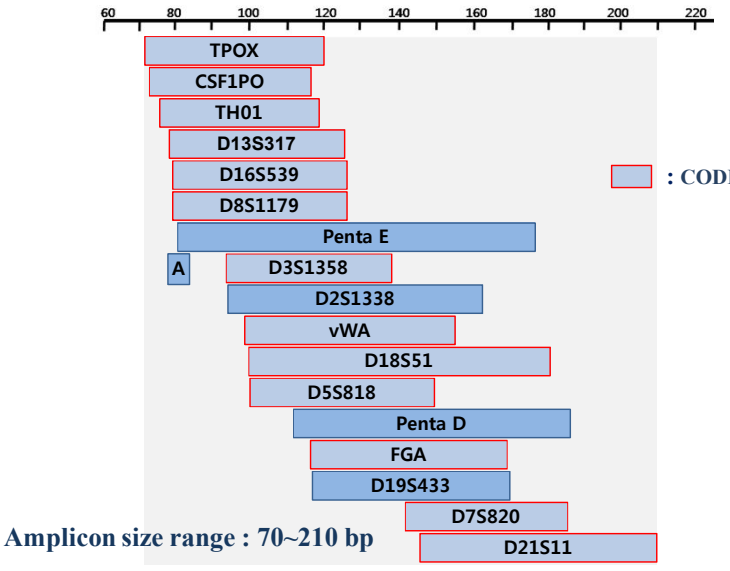
Publication	Platform	Target loci	Sample		Amplicon generation
			Single	Mixture	
Fordyce *et al.* (2011)	Roche 454 GS FLX	5 STRs	○		Custom Singleplex PCR
Van Neste *et al.* (2011)	Roche 454 GS FLX	9 STRs	○	○	Commercial Kit
Bornman *et al.* (2012])	Illumina GAIIx	13 STRs + Amelogenin	○	○	Custom designed long range PCR
Warshauerm *et al.* (2013)	Illumina GAIIx and MiSeq	22 STRs + 22 Y-STRs	○		Commercial Kits
Van Neste *et al.* (2013)	Illumina MiSeq	15 STRs + Amelogenin (developing)	○	○	Custom multiplex PCR
Dalsgaard *et al.* (2013)	Roche GS Junior	4 STRs	○		Commercial Kit
Rockenbauer *et al.* (2014)	Roche GS Junior	1 STR	○		Custom Singleplex PCR

➡ **Need for customized multiplex PCR system generating small amplicons**

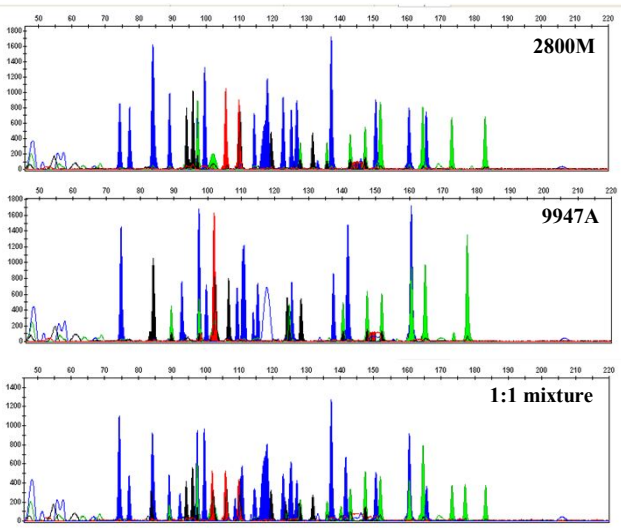
Criteria and resources for primer design

- Criteria
  - Small amplicon size were adapted while primer is not overlapping with repeat region of STR
  - Size of the samllest amplicon of STR is greater than 70 bp
  - Avoid more than 1% mutation in primer binding area
- Resources
  - STRBase (http://www.cstl.nist.gov/ div831/strbase/)
  - GenBank (www.ncbi.nlm.nih.gov/genbank/)
  - Primer 3 v.0.4.0 (http://frodo.wi.mit.edu/primer3/)

18 markers of a new multiplex PCR system



Test new multiplex PCR system on CE



Library preparation and massive sequencing on MiSeq system

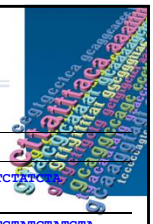
Step 1. PCR amplification	Step 2. Validate Amplicon	Step 3. Library preparation	Step 4. Validate Library	Step 5. Sequencing
• **Template DNA** ; 2800M, 9947A 1:1, 1:3, 1:6, 1:9 mixture (Male:Female)	• **Fluorometer** ; Quant-iT™ PicoGreen dsDNA assays (Invitrogen) • **Agilent BioAnalyzer**	• **TruSeq Nano DNA LT Sample preparation Kit** * Adjustment of beads ratio for size selection	• **Fluorometer** ; Quant-iT™ PicoGreen dsDNA assays (Invitrogen) • **Agilent BioAnalyzer** • **Library Quantification**	• **Cluster generation and sequencing on MiSeq** ; 2 x 250 bp (Paired-end)

STR genotyping in single source

STRs	2800M		9947A	
	CE	NGS	CE	NGS
D2S1338	22, 25	22, 25	19, 23	19, 23
D3S1358	17, 18	17, 18	14, 15	14, 15
D5S818	12	12	11	11
D7S820	8, 11	8, 11	10, 11	10, 11
D8S1179	14, 15	14, 15	13	13
D13S317	9, 11	9, 11	11	11
D16S539	9, 13	9, 13	11, 12	11, 12
D18S51	16, 18	16, 18	15, 19	15, 19
D19S433	13, 14	13, 14	14, 15	14, 15
D21S11	29, 31.2	29, 31.2	30	30
CSF1PO	12	12	10, 12	10, 12
FGA	20, 23	20, 23	23, 24	23, 24
Penta_D	12, 13	12, 13	12	12
Penta_E	7, 14	7, 14	12, 13	12, 13
TH01	6, 9.3	6, 9.3	8, 9.3	8, 9.3
TPOX	11	11	8	8
vWA	16, 19	16, 19	17, 18	17, 18

STRait Razor program (Warshauer *et al.*)

Sequence variations in target STR regions



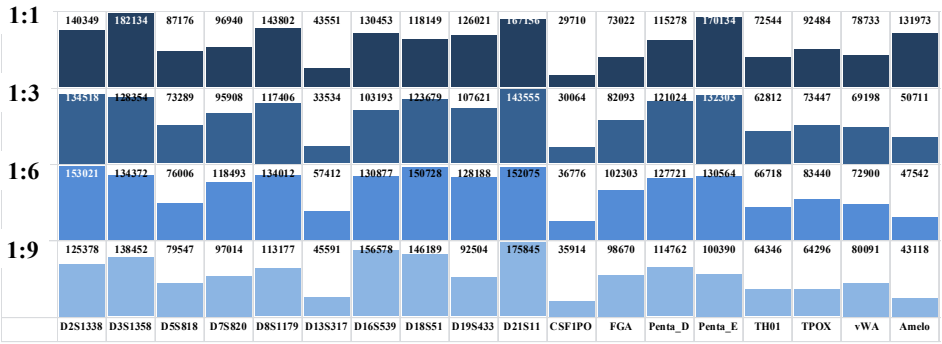
D8S1179

Samples	Alleles	Sequences
Reference (hg19)	13	TTTTTGTATTTCATGTGTACATTCGTATCTATCTGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA TTCCCCACAGTGAAAATAATCTAC
2800M	14	TTTTTGTATTTCATGTGTACATTCGTATCTATCTGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA TTCCCCACAGTGAAAATAATCTAC
	15	TTTTTGTATTTCATGTGTACATTCGTATCTATCTATCTGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA TTCCCCACAGTGAAAATAATCTAC
9947A	13	TTTTTGTATTTCATGTGTACATTCGTATCTATCTGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA TTCCCCACAGTGAAAATAATCTAC
	13'	TTTTTGTATTTCATGTGTACATTCGTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA TTCCCCACAGTGAAAATAATCTAC

STR loci	Sequence variation (SNPs)
D8S1179	G>A (rs13265375), A>G (rs111782616)
D3S1358	A>G (rs77577482), A>G (rs71325067)
D21S11	G>A (rs13049099), A>G (rs13050496), **Un-identified SNP**
vWA	G>A (rs78332633), G>A (rs216871)

Read count from NGS data on MiSeq

Sample	Total reads
1:1 mixture	1,490,212
1:3 mixture	1,252,182
1:6 mixture	1,403,656
1:9 mixture	1,358,935



STR genotyping in mixed samples

STRs	Mixture sample			
	1:1	1:3	1:6	1:9
D2S1338	19, 22, 23, 25	19, 22, 23, 25	19, 22, 23, 25	19, 22, 23, 25
D3S1358	14, 15, 17, 18	14, 15, 17, 18	14, 15, 17, 18	14, 15, 17, 18
D5S818	11, 12	11, 12	11, 12	11, 12
D7S820	8, 10, 11	8, 10, 11	8, 10, 11	8, 10, 11
D8S1179	13, 14, 15	(12), 13, 14, 15	(12), 13, 14, 15	(12), 13, 14, 15
D13S317	9, 11	9, 11	9, 11	9, 11
D16S539	9, 11, 12, 13	9, 11, 12, 13	9, 11, 12, 13	9, (10), 11, 12, 13
D18S51	15, 16, 18, 19	15, 16, 18, 19	15, (16), 18, 19	(14), 15, 16, 18, 19
D19S433	13, 14, 15	13, 14, 15	13, 14, 15	13, 14, 15
D21S11	29, 30, 31.2	29, 30, 31.2	29, 30, 31.2	29, 30, 31.2
CSF1PO	10, 12	10, 12	10, 12	10, 12
FGA	20, 23, 24	20, 23, 24	20, 23, 24	20, 23, 24
Penta_D	12, 13	12, 13	12, 13	12, 13
Penta_E	7, 12, 13, 14	7, 12, 13, 14	7, 12, 13, 14	7, 12, 13, 14
TH01	6, 8, 9.3	6, 8, 9.3	6, 8, 9.3	6, 8, 9.3
TPOX	8, 11	8, 11	8, 11	8, 11
vWA	16, 17, 18, 19	16, 17, 18, 19	16, 17, 18, 19	16, 17, 18, 19

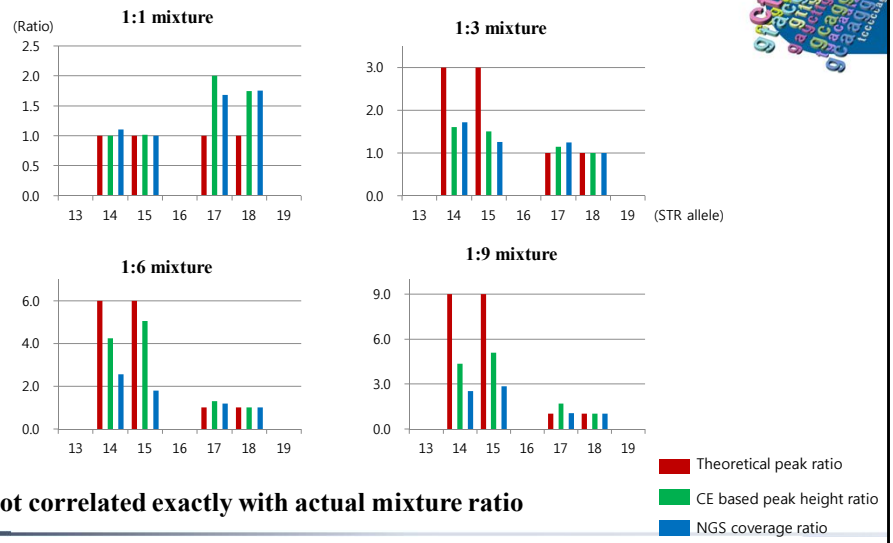
Blue color in parentheses - true allele less than coverage value of 10%

Red color in parentheses - stutter of true allele with coverage value between 5% and 10%

STRait Razor program (Warshauer et al.)

Evaluation of mixture ratio

Example) D3S1358



* **Not correlated exactly with actual mixture ratio**

Summary

- ✓ We constructed a singe tube new multiplex PCR system that is optimized for NGS analysis of forensic STR markers.
- ✓ Most STR alleles could be determined successfully in single-source DNA and even with mixed samples.
- ✓ Sequence variations could also be detected in targeted STR region.

Further study

- Fine adjustment of multiplex PCR system for read count balancing
- More tests on various mixed samples
- Application to degraded DNA samples

Acknowledgments

◆ MiSeq system (Illumina) supported by

BMS (주)비엠에스 Discover the Future bmskorea.co.kr

◆ This work was supported by the research project for practical use and advancement of forensic DNA analysis of Supreme Prosecutors' Office, Republic of Korea (1333-304-260, 2013).