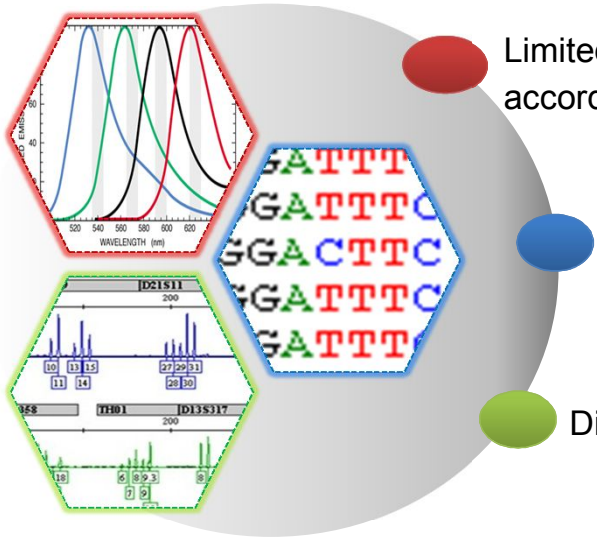
Massive Sequence Analysis of Forensic STR Loci using Next Generation Sequencing and Its Application to Mixture Analysis

Eun Hye Kim, In Seok Yang, Sang-Eun Jung, Hwan Young Lee, Woo Ick Yang, and Kyoung-Jin Shin

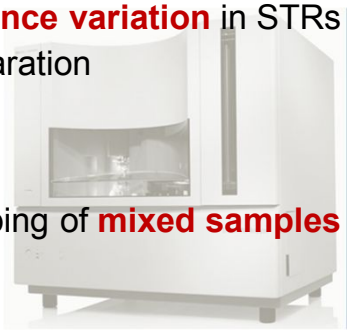
Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Current STR typing in forensic genetics

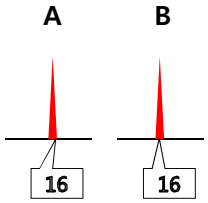


- Limited **the total of number** and **allelic size** of STRs according to available fluorescence dyes
- Can not identify **sequence variation** in STRs due to size based separation
- Difficulty in digital genotyping of **mixed samples**



YONSEI UNIVERSITY COLLEGE OF MEDICINE

Application of NGS to forensic STR typing

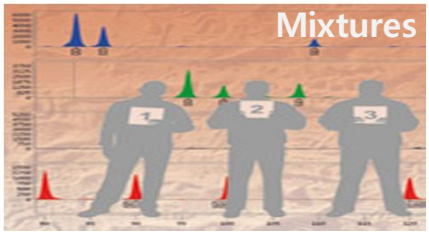


<CE method>

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

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

<NGS method>



YONSEI UNIVERSITY COLLEGE OF MEDICINE

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 monoplex 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 monoplex PCR
Fordyce *et al*. (In press)	Thermo Fisher Ion PGM	10 STRs (developing)	○	○	Custom multiplex PCR

Need for multiplex PCR system optimized for NGS with small amplicons

YONSEI UNIVERSITY COLLEGE OF MEDICINE

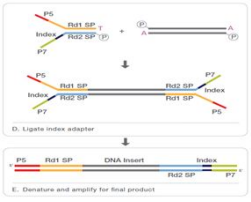
Outline

To analyze forensic STR data using next generation sequencing

- ✓ Construction of in-house multiplex PCR system for STR NGS analysis
- ✓ To validate the multiplex system, NGS data generated from two singles, mixtures with various ratio.
- ✓ Analysis of sequence variation in STR regions in 10 Koreans.

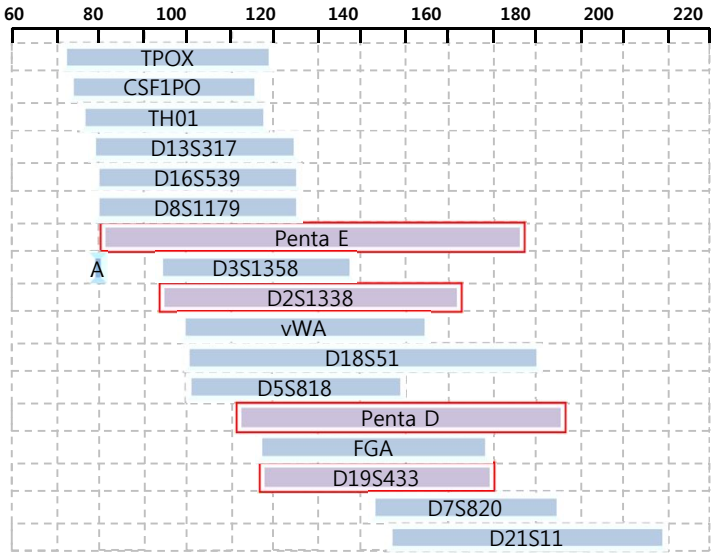
YONSEI UNIVERSITY COLLEGE OF MEDICINE

Experimental procedures

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) 10 Koreans • After PCR, primer digestion Using Exo-SAP IT. • Column purification using QIAquick column kit	• **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** ; Kapa Library Quantification Kit	• **Cluster gen and sequencing on MiSeq** ; 2 x 250 bp (Paired-end) 

YONSEI UNIVERSITY COLLEGE OF MEDICINE

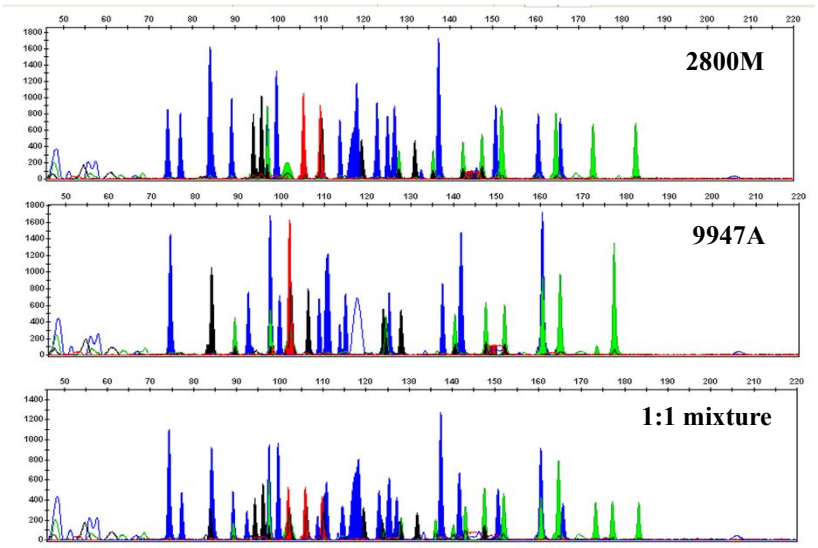
The in-house developed multiplex PCR system



- **Target markers (TOTAL 18 markers)**
  - CODIS STR 13 loci in **blue boxes**
  - Commonly used commercial kits **in red boxes**
  - Amelogenin
- **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/)

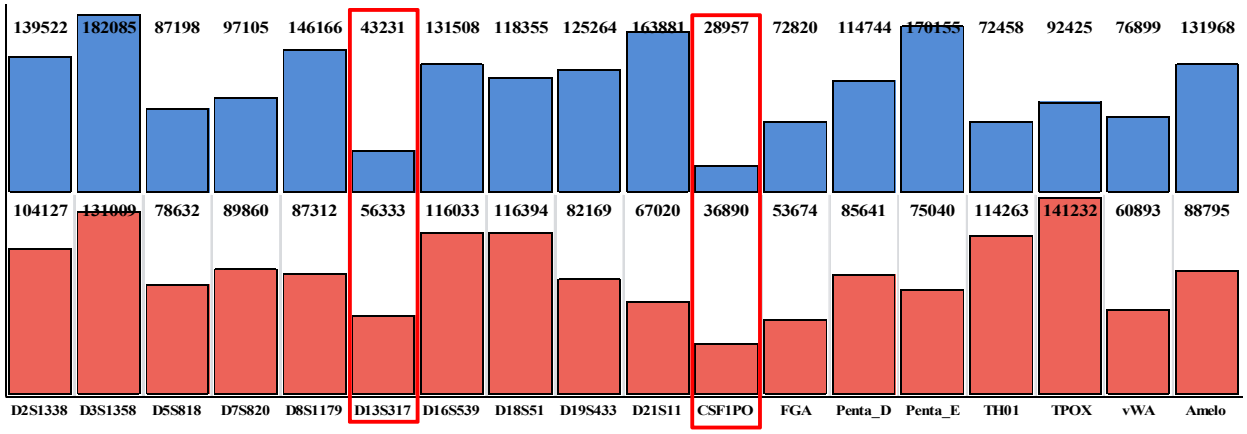
YONSEI UNIVERSITY COLLEGE OF MEDICINE

Test in-house multiplex PCR system on CE



YONSEI UNIVERSITY COLLEGE OF MEDICINE

NGS data from MiSeq



Improvement of coverage through the adjustment of primer concentration

Bowtie2 program (Langmead *et al.* Nat Methods; 2012)

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Results of STR genotyping in single-sources

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
AMEL	X, Y	X, Y	X	X

STRait Razor program (Warshauer *et al.* FSIG; 2013)

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Results of STR genotyping in mixtures on MiSeq

STRs	MiSeq STR data			
	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	13, 14, 15	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

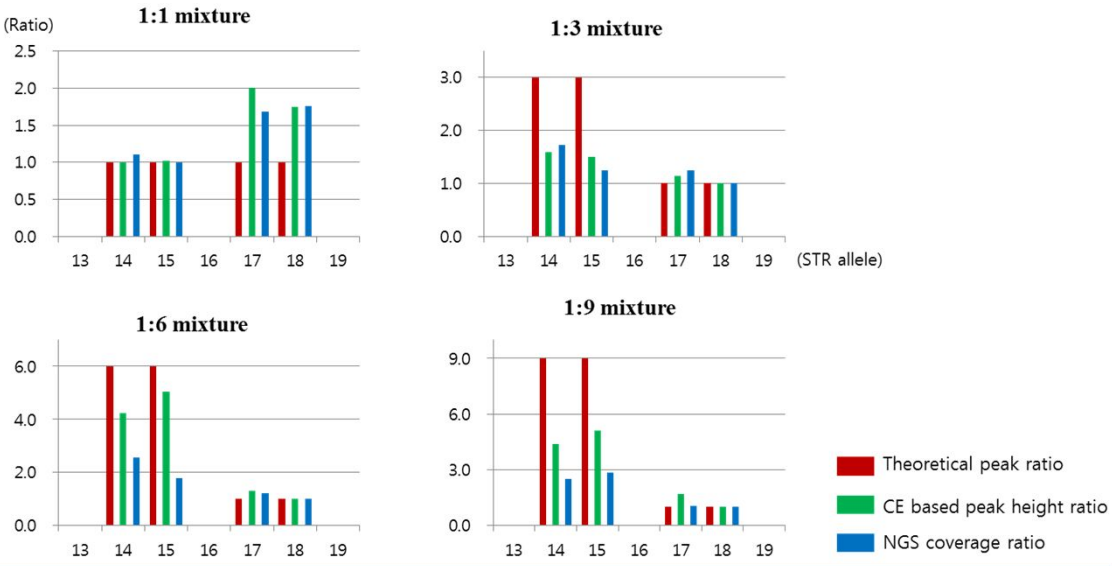
Blue color in parentheses - true allele less than coverage value of 10%

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

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Evaluation of mixture ratio

Example)
D3S1358



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

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Results of STR genotyping in 10 Koreans

STRs	Conformity of STR genotypes between CE based method and NGS analysis									
	Korean 01	Korean 02	Korean 03	Korean 04	Korean 05	Korean 06	Korean 07	Korean 08	Korean 09	Korean 10
D2S1338	O	O	O	O	O	O	O	O	O	O
D3S1358	O	O	O	O	O	O	O	O	O	O
D5S818	O	O	O	O	O	O	O	O	O	O
D7S820	O	O	O	O	O	O	O	O	O	O
D8S1179	O	O	O	O	O	O	O	O	O	O
D13S317	O	O	O	O	O	O	O	O	O	O
D16S539	O	O	O	O	O	O	O	O	O	O
D18S51	O	O	O	O	O	O	O	O	O	O
D19S433	O	O	O	O	O	O	O	O	O	O
D21S11	O	O	O	O	O	O	O	O	O	O
CSF1PO	O	O	O	O	O	O	O	O	O	O
FGA	O	O	O	O	O	O	O	O	O	O
Penta D	O	O	O	O	O	O	O	O	O	O
Penta E	O	O	O	O	O	O	O	O	O	O
TH01	O	O	O	O	O	O	O	O	O	O
TPOX	O	O	O	O	O	O	O	O	O	O
vWA	O	O	O	O	O	O	O	O	O	O
Amelogenin	O	O	O	O	O	O	O	O	O	O

Matched

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Determination of repeat structures in target STR regions

Example)

STR loci	Allele	Repeat structure
D2S1338	**Ref_23**	**$[TGCC]_7$ $[TTCC]_{13}$ GTCC $[TTCC]_2$**
	17	[TGCC]6 [TTCC]11
	18a	[TGCC]6 [TTCC]12
	18b	[TGCC]7 [TTCC]11
	19a	[TGCC]6 [TTCC]13
	19b	[TGCC]7 [TTCC]12
	20a	[TGCC]7 [TTCC]2 TTTC [TTCC]10
	20b	[TGCC]7 [TTCC]13
	21a	[TGCC]7 [TTCC]2 TTTC [TTCC]11
	21b	[TGCC]7 [TTCC]14
	22	[TGCC]7 [TTCC]12 GTCC [TTCC]2
	23	[TGCC]7 [TTCC]13 GTCC [TTCC]2
	24a	[TGCC]5 [TTCC]16 GTCC [TTCC]2
	24b	[TGCC]6 [TTCC]15 GTCC [TTCC]2
	25	[TGCC]7 [TTCC]15 GTCC [TTCC]2

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Determination of sequence variations in target STR regions

STR loci	Two Standard DNAs	10 Koreans	Separation of alleles CE/NGS
D2S1338	G>T (rs9678338)	G>T (rs62182233) G>T (rs6736805) G>T (rs9678338) non-identified SNP	9/14
D3S1358	A>G (rs77577482) A>G (rs71325067)	A>G (rs77577482) A>G (rs71325067)	4/6
D8S1179	G>A (rs13265375) A>G (rs111782616)	G>A (rs13265375) A>G (rs111782616)	5/9
D21S11	G>A (rs13049099) G>A (rs200026324) A>G (rs13050496)	G>A (rs13049099) G>A (rs200026324) A>G (rs13050496)	5/9
vWA	G>A (rs216871) A>G (rs112652289)	G>A (rs216871)	(5/5)

SNP info from NCBI – dbSNP Build 138

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Summary

- We constructed a in-house multiplex PCR system that is optimized for NGS analysis of 18 forensic markers.
- STR genotyping results obtained from NGS analysis were consistent with those from CE-based analyses both for single-source samples and mixed samples.
- Sequence variations which can help differentiation of alleles from different sources were also detected in some STR loci of two standard DNA and 10 Koreans.

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Further study

- Incorporation of extended STRs to support compatibility with recent CE based methods
- Fine adjustment of in-house multiplex PCR system for balancing coverage in inter-markers
- Validation study and application to casework samples

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Thank you for your attention !