Sequence-based analysis of 31 Y-STRs using Massively Parallel Sequencing in the African Americans, Caucasians, Hispanics and Koreans

Mi Hyeon Moon[1,2], Kyoung-Jin Shin[1,2]

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

1. Introduction

Why? Y chromosome?

Applications

- Sexual assault cases
- Mixture deconvolution
- Biogeographic origin inference

1. Introduction

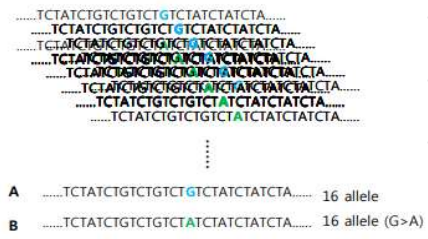
Y-STR analysis in forensic genetics

Capillary Electrophoresis (CE)

- Length-based analysis
- Limitations
  - Mixture DNA ?
  - Highly degraded DNA ?
  - Biogeographic origin ?

1. Introduction

Application of MPS to forensic genetics



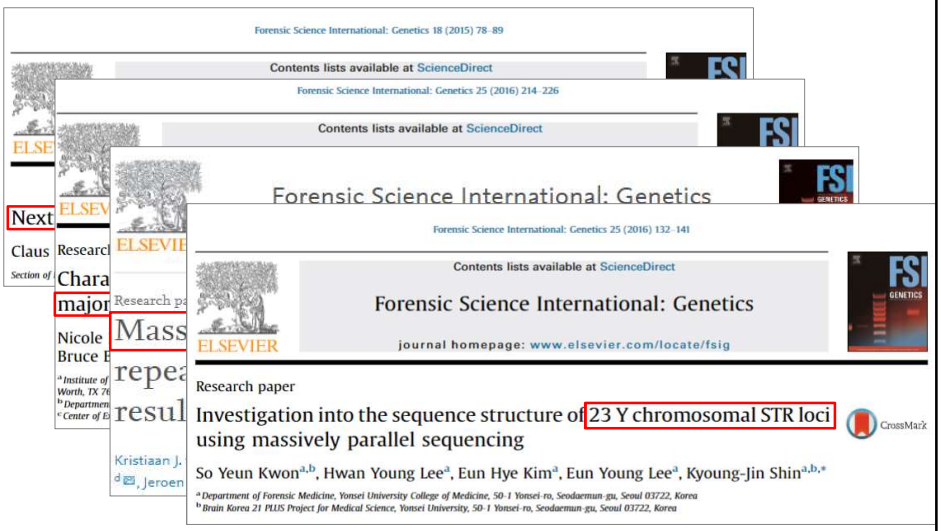
Massively Parallel Sequencing

- Sequence-based analysis
- Larger multiplexing

- Mixture DNA resolution
- Highly degraded DNA
- Biogeographic origin

1. Introduction

Recent studies



PowerPlex Y23 loci

1. Introduction

Objectives

✓ **Expand the MPS panel for Y-STRs**

- Compatible with CE-based panel

  (PowerPlex Y23, Yfiler Plus panel)

✓ **Compile the sequence variation and frequency data**

- For world-wide population

  (African American, Caucasians, Hispanics and Koreans)

✓ **Population-specific characteristics**

2. Materials and Methods

DNA samples

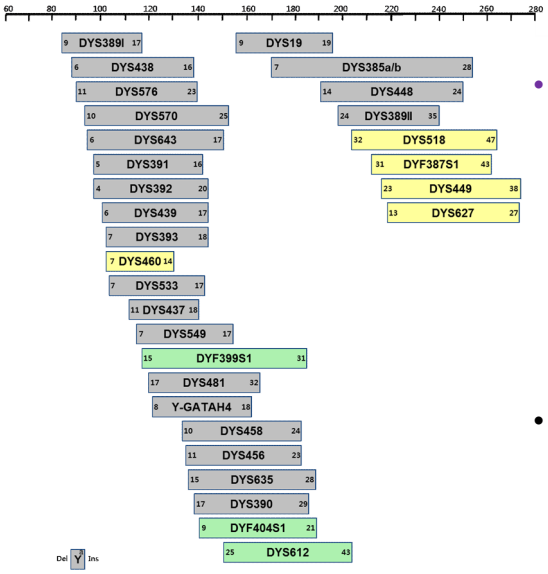
- 220 samples from 4 population groups

Population	No. of samples	Origin
African Americans (**AfAm**)	17	Cell line (Coriell)
Caucasians (**Cauc**)	50	Cell line (Coriell)
Hispanics (**Hisp**)	48	Cell line (Coriell)
Koreans (**Kor**)	105	Kwon et al. (2016)
Total	**220**	

*Approved by the Institutional Review Board of Severance Hospital, Yonsei University in Seoul, Korea.

2. Materials and Methods

In-house multiplex MPS panel



- **32** targeted markers
  - PowerPlex Y**23** loci + **1** Y-SNP (M175) (*Kwon et al*. FSIG; 2016)
  - **5** Yfiler Plus loci + **3** additional RM Y-STRs
- Small sized amplicons (85-274bp)

2. Materials and Methods

MPS workflow

STEP 1
PCR-based library prep. & validation

- 2100 Bioanalyzer
- KAPA quantification kit

STEP 2
Sequencing

- Miseq system
- Miseq reagent kit v3.0

STEP 3
Data analysis

- STRait Razor v 3.0
- Microsoft Excel

3. Result

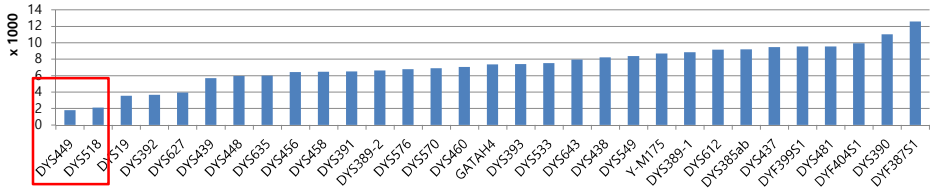
MPS coverage

✓ **Sample coverage**

- Average read counts : 256,089

✓ **Marker coverage**

- Average depth of coverage : 7,219

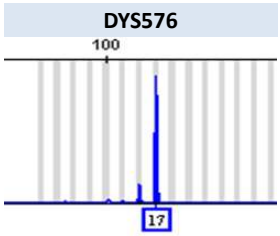


3. Result

Genotype discordance between CE and MPS

Marker: DYS576 (Allele drop-out) – 2 samples

1) Multiplex CE result (PowerPlex Y23)



Allele	Peak height
17	1195

2) MPS result

Allele	Read count	Sequence
0	37	SumbelowThreshold

3) Sanger sequencing result (MPS primer binding site mutation)

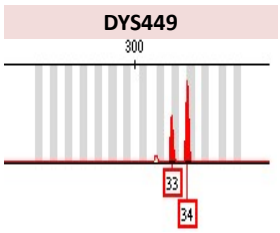
Reference	5’-GCGTATTTGTCTTGGCT**T**TTTC-3’
Sample	5’-GCATATTTGTCTTGGCT**C**TTTC-3’
dbSNP ID	T>C (rs754193694)

3. Result

Genotype discordance between CE and MPS

Marker: DYS449 (Allele drop-out) – 2 samples

1) Multiplex CE result (Euplex-Y17)

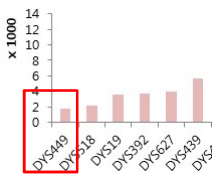


Allele	Peak height
33	1017(57%)
34	1786

2) MPS result

Allele	Read count	Sequence
34	206	[TTTC]17 N50 [TTTC]17

3) Relative low coverage (Minor allele drop-out)

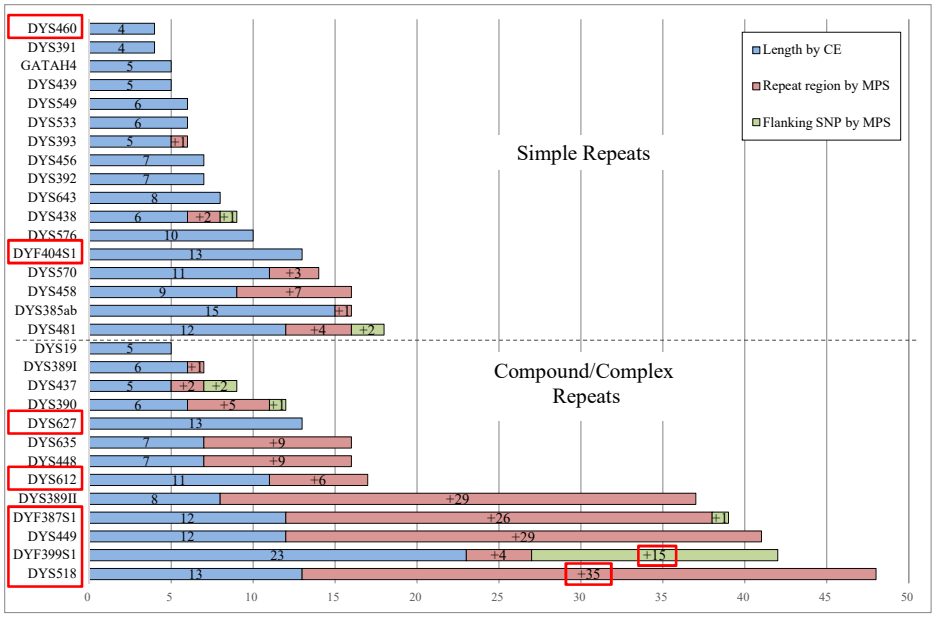


- Analytical Threshold (AT)

  = 100 reads

3. Results and discussion

Comparison of the observed alleles by CE and MPS method



3. Results and discussion

3.1. Gains from MPS analysis

- Examples of repeat region variations

· **Marker: DYS518**

13 (CE)	+35 (repeat region)

CE allele	MPS sub-allele	Repeat structure	Frequency AfAm	Cauc	Hisp	Kor
39	a	[AAAG]3 GAAG [AAAG]17 GGAG [AAAG]4 gaagag [AAAG]13	0.118	0.080	0.041	0.010
	b	[AAAG]3 GAAG [AAAG]15 GGAG [AAAG]4 gaagag [AAAG]14 **GAAG**	-	-	0.020	-
	c	[AAAG]3 GAAG [AAAG]16 GGAG [AAAG]4 gaagag [AAAG]12 **[GAAG]2**	-	0.020	-	-
	d	[AAAG]3 GAAG [AAAG]18 GGAG [AAAG]4 gaagag [AAAG]12	-	-	0.020	-
	e	[AAAG]3 GAAG [AAAG]16 GGAG [AAAG]4 gaagag [AAAG]14	-	0.080	0.122	0.038
	f	[AAAG]3 GAAG [AAAG]15 GGAG [AAAG]4 gaagag [AAAG]15	0.176	0.020	-	0.038
	g	[AAAG]3 GAAG [AAAG]13 GGAG [AAAG]4 gaagag [AAAG]17	-	-	-	0.010

3. Results and discussion

3.1. Gains from MPS analysis

- Examples of repeat region variations

· Marker: DYS449

12 (CE)	+29 (repeat region)

CE allele	MPS sub-allele	Repeat structure	Frequency AfAm	Cauc	Hisp	Kor
31	a	[TTTC]15 N50 [TTTC]16	0.176	0.020	0.042	**0.076**
	b	[TTTC]2 **TATC** [TTTC]12 N50 [TTTC]16	-	-	-	**0.010**
	c	**CTTC** [TTTC]15 N50 [TTTC]15	-	-	-	**0.010**
	d	[TTTC]16 N50 [TTTC]15	-	0.120	0.083	**0.038**
	e	[TTTC]14 N50 [TTTC]17	0.059	-	0.021	**0.029**
	f	[TTTC]17 N50 [TTTC]14	-	-	0.042	**0.010**
	g	[TTTC]13 N50 [TTTC]18	-	-	-	**0.010**

3. Results and discussion

3.1. Gains from MPS analysis

- Examples of repeat region variations

· Marker: DYF387S1

12 (CE)	+26 (repeat region)

CE allele	MPS sub-allele	Repeat structure	Frequency AfAm	Cauc	Hisp	Kor
38	a	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]10 [AAAG]15	0.182	0.029	0.054	0.015
	b	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]11 [AAAG]14	0.030	0.058	0.011	0.111
	c	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]9 [AAAG]16	-	0.039	0.054	0.015
	d	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]8 [AAAG]17	-	0.010	-	0.015
	e	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]13 [AAAG]12	-	-	0.032	-
	f	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]12 [AAAG]13	-	0.010	-	-

3. Results and discussion

3.1. Gains from MPS analysis
- Flanking region SNP

STR Locus	rs number (dbSNP build 151)	Wild	Mutant	Position (GRCh38/hg38)	Frequency AfAm	Cauc	Hisp	Kor
DYF387S1	-	G	A	Chr Y: 23,785,347	-	0.010	-	-
DYF399S1	rs4306075	A	G	Chr Y: 22,950,382	0.333	0.333	0.248	0.365
	rs878949651	A	G	Chr Y: 22,950,401	0.137	0.367	0.404	0.278
DYS390	rs766823340	T	G	Chr Y: 15,163,167	-	0.040	-	-
DYS437	rs9786886	C	T	Chr Y: 12,346,264	0.588	-	-	-
DYS438	rs760613324	A	C	Chr Y: 12,825,955	-	-	0.020	-
DYS481	rs370750300	G	T	Chr Y: 8,558,336	-	-	-	0.038

3. Results and discussion

3.1. Gains from MPS analysis
- Examples of flanking region variation (SNP)

· Marker: DYF399S1

23 (CE)	+4 (repeat region)	+15 (flanking region)

CE allele	MPS sub-allele	Repeat structure	3' Flanking region	Frequency AfAm	Cauc	Hisp	Kor
21	a	[GAAA]3 N7 [GAAA]16	AAACTTTTACCCTTTTGACA	0.039	-	-	-
	b	[GAAA]3 N7 [GAAA]16	GAACTTTTACCCTTTTGACA	0.098	0.127	0.064	0.110
	c	[GAAA]3 N7 [GAAA]16	AAACTTTTACCCTTTTGACG	-	0.020	-	0.003
	d	[GAAA]3 N7 [GAAA]16	GAACTTTTACCCTTTTGACG	-	-	0.007	-
	e	[GAAA]3 N7 [GAAA]15 GAGA	GAACTTTTACCCTTTTGACA	-	-	-	0.003

rs4306075 rs878949651

3. Results and discussion

3.1. Gains from MPS analysis
- Examples of flanking region variation (SNP)

· **Marker: DYS437**

CE allele	MPS sub-allele	5' Flanking region	Repeat structure	Frequency AfAm	Cauc	Hisp	Kor
14	a	GCCCATC**C**GG	[TCTA]8 [TCTG]2 [TCTA]4	-	0.400	0.479	0.562
	b	GCCCATC**T**GG	[TCTA]8 [TCTG]2 [TCTA]4	**0.471**	-	-	-
	c	GCCCATC**C**GG	[TCTA]9 TCTG [TCTA]4	-	-	-	0.143

rs9786886

Conclusion

- ✓ We expanded the MPS panel for 31 Y-STRs by adding 7 RM Y-STRs and 1 Yfiler Plus loci to the PowerPlex Y23 loci.
- ✓ The markers with the increased number of alleles by repeat region variation was DYS518, DYF387S1, DYS389-Ⅱ, DYS449, and mainly RM Y-STRs.
- ✓ The increase in the number of alleles by flanking region SNP showed in DYF387S1, DYF399S1, DYS390, DYS437, DYS438 and DYS481, among which DYF399S1 was the most.
- ✓ The compilation of sequence-based data is necessary for giving statistics and applying in forensic practice.

Acknowledgement

National Research Foundation of Korea (NRF)
NRF-2014M3A9E1069989

BK21 PLUS
Brain Korea 21 PLUS Project for Medical Science, Yonsei University

KJSHIN@yuhs.ac
mhmoon@yuhs.ac
http://forensic.yonsei.ac.kr