ICGSK 2019 *The Genetics Society of Korea*

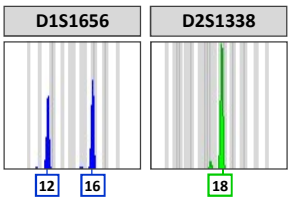
Sequence Variations of 25 Autosomal STRs Including **SE33** Analyzed by **Next Generation Sequencing** in the African Americans, Caucasians, Hispanics and Koreans

Ye-Lim Kwon[1,2] · Bo Min Kim[1,2] · Eun 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

1. Introduction

❖ Capillary electrophoresis (CE)



- Gold standard methodology
- **Length**-based analysis

➢ **Assays**
- PowerPlex® Fusion
- GlobalFiler™

etc...

❖ Next generation sequencing (NGS)

- **D1S1656**

Allele	Bracketed Repeat	Coverage
12	[TCTA]12	2296
16	CCTA [TCTA]15	1869

- **D2S1338**

Allele	Bracketed Repeat	Coverage
18 (a)	[GGAA]12 [GGCA]6	1722
18 (b)	[GGAA]11 [GGCA]7	1842

- **Sequence**-based analysis

➢ **Advantage**
- Mixture deconvolution
- Degraded DNA analysis

1

1. Introduction

❖ Next generation sequencing (NGS) for STRs

➢ Assays

- ForenSeq™ DNA Signature Prep Kit (Verogen)

- Precision ID GlobalFiler™ NGS STR Panel v2 (Thermo Fisher Scientific)

- PowerSeq™ 46GY System (Promega) - Prototype

➢ Recent studies

1. Introduction

❖ Our previous study..

Kim et al. (2017) FSIG

Forensic Science International: Genetics 30 (2017) 134–140

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Research paper

Sequence-based diversity of 23 autosomal STR loci in Koreans investigated using an in-house massively parallel sequencing panel

Eun Hye Kim[a], Hwan Young Lee[a], So Yeun Kwon[a,b], Eun Young Lee[a], Woo Ick Yang[a], Kyoung-Jin Shin[a,b,*]

[a] Department of Forensic Medicine, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
[b] Brain Korea 21 PLUS Project for Medical Science, Yonsei University, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

Analysis

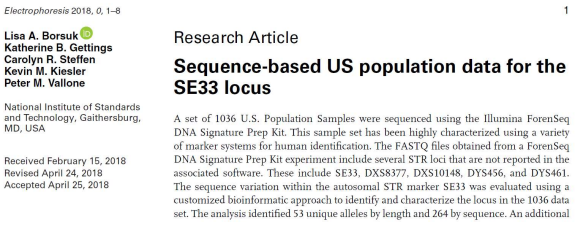
- 23 autosomal STR loci
- 250 Koreans
- In-house NGS panel

1. Introduction

❖ SE33 locus

- Highly polymorphic locus
- Few sequence-based data have been reported

Borsuk et al. (2018) Electrophoresis



Electrophoresis 2018, *0*, 1–8 1

Lisa A. Borsuk
Katherine B. Gettings
Carolyn R. Steffen
Kevin M. Kiesler
Peter M. Vallone

National Institute of Standards and Technology, Gaithersburg, MD, USA

Received February 15, 2018
Revised April 24, 2018
Accepted April 25, 2018

Research Article

Sequence-based US population data for the SE33 locus

A set of 1036 U.S. Population Samples were sequenced using the Illumina ForenSeq DNA Signature Prep Kit. This sample set has been highly characterized using a variety of marker systems for human identification. The FASTQ files obtained from a ForenSeq DNA Signature Prep Kit experiment include several STR loci that are not reported in the associated software. These include SE33, DXS8377, DXS10148, DYS456, and DYS461. The sequence variation within the autosomal STR marker SE33 was evaluated using a customized bioinformatic approach to identify and characterize the locus in the 1036 data set. The analysis identified 53 unique alleles by length and 264 by sequence. An additional

→ **SE33** is amplified by **ForenSeq DNA Signature Prep kit**, but **not reported** in the Universal Analysis Software (**UAS**)

4

1. Introduction

❖ Objectives

- **Upgrade the in-house NGS panel for autosomal STRs**
  - Add 2 autosomal STRs: **SE33** and **D4S2408**
- **Investigate the sequence structure and compile the sequence-based allele data**
  - 4 populations (African Americans, Caucasians, Hispanics and Koreans)
- **Discuss the coordinate issue across multiple NGS assays**

5

2. Materials and Methods

❖ DNA Samples

➢ **350 unrelated samples from 4 populations**

Population	# of samples
African Americans (AfAm)	83
Caucasians (Cauc)	82
Hispanics (Hisp)	82
Koreans (Kor)	103
Total	**350**

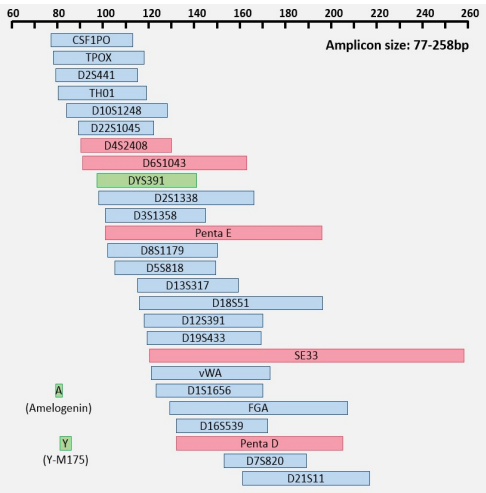
* Approved by the IRB of Severance Hospital, Yonsei University in Seoul, Korea

→ **All samples were genotyped by CE**

(using Kplex-23, Euplex-13 and PowerPlex® Fusion)

2. Materials and Methods

❖ Assay: In-house NGS panel v2.5



➢ **28 targeted loci**

- 20 expanded CODIS loci +
- 5 additional autosomal loci +
- 3 sex typing loci

➢ **5 additional autosomal loci**

- D4S2408 - Penta E - SE33
- D6S1043 - Penta D

➢ **3 sex typing loci**

- Amelogenin - DYS391
- Y Indel (M175)

2. Materials and Methods

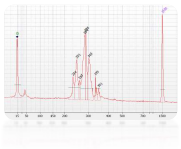
❖ Workflow for NGS on a MiSeq system

1 PCR-based NGS Library Preparation

Two-step PCR

> 1st PCR target-specific

> 2nd PCR platform-specific



Library validation

> Library quantification

> Bead purification

> Library pooling

2 Sequencing

> MiSeq Reagent kit v3

> MiSeq System

3 Data Analysis

> STRait Razor 3.0

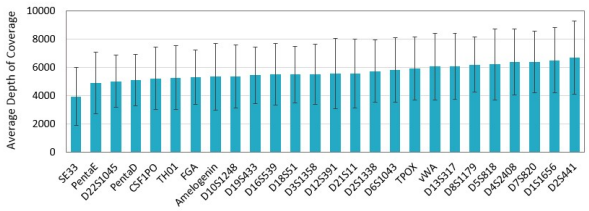
> Microsoft Excel

- Detailed protocol is available in http://forensic.yonsei.ac.kr/protocols.html and *Kim et al. (2017) FSIG*

3. Results and Discussions

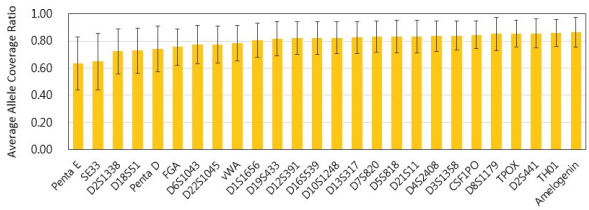
❖ NGS Quality

➢ **Average Depth of Coverage (DoC)**



- Average 5655 reads per marker
- Max/Min: less than 1.69-fold

➢ **Average Allele Coverage Ratio (ACR)**



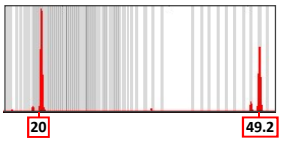
- Average 0.80 ACR per marker
- Min: 0.636 in Penta E
- Max: 0.865 in Amelogenin

3. Results and Discussions

❖ Genotype Concordance between CE and NGS - 99.88%

◆ **Genotype discordance** ① **FGA** ② vWA ③ SE33
(In 1 sample)

➢ **Multiplex CE result** (PowerPlex Fusion)



➢ **NGS result**

Allele	Bracketed repeat	DoC
20	[GGAA]2 GGAG [AAAG]12 AGAA AAAA [GAAA]3	1822

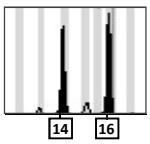
→ Allele 49.2 drop-out

3. Results and Discussions

❖ Genotype Concordance between CE and NGS - 99.88%

◆ **Genotype discordance** ① FGA ② **vWA** ③ SE33
(In 3 samples)

➢ **Multiplex CE result** (PowerPlex Fusion)



➢ **NGS result**

Allele	Bracketed repeat	DoC
16	[TAGA]11 [CAGA]4 TAGA	3749

→ Allele 14 drop-out

➢ **Sanger sequencing result** (SNP in the NGS primer binding site)

Reference	5′-TGATAAATACATAGGAT**G**GATGG-3′
Sample	5′-TGATAAATACATAGGAT**A**GATGG-3′

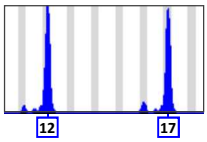
rs771794429 (G>A)

3. Results and Discussions

❖ Genotype Concordance between CE and NGS - 99.88%

◆ **Genotype discordance** ① FGA ② vWA ③ **SE33**
(In 8 samples)

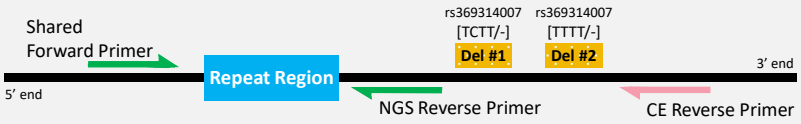
➢ **Monoplex CE result** (Euplex-13)



➢ **NGS result**

Allele	Bracketed repeat	DoC
13	[CTTT]13	1055
17	[CTTT]17	782

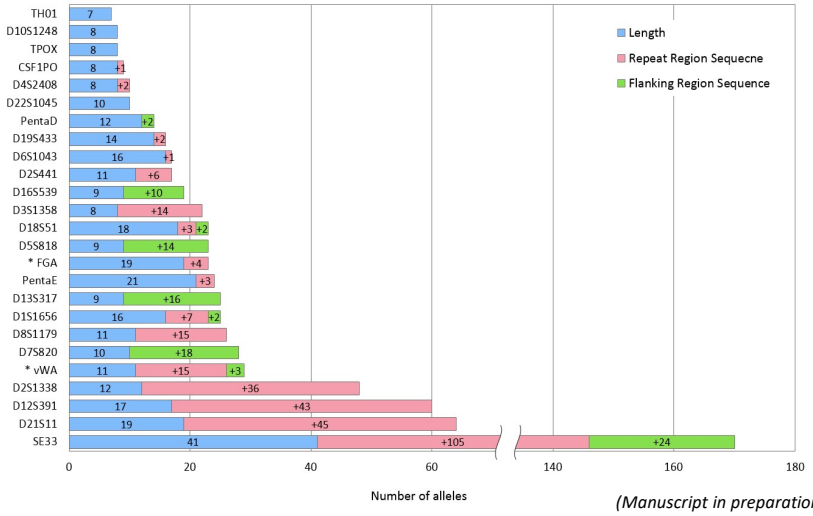
➢ **Sanger sequencing result** (4bp deletion between CE and NGS primer binding site)



12

3. Results and Discussions

❖ The number of alleles obtained by CE and NGS



(Manuscript in preparation)

13

3. Results and Discussions

❖ Characteristics of the SE33

➢ **Motif structure of SE33** ① rs9362477 ② rs536914220 ③ rs151261950

#	Motif structure	Flanking region polymorphisms	Freq.
BASIC	**[CTTT]n TT CT [CTTT]n**		
a	[CTTT]n	rs9362477	0.446
b	[CTTT]n TT [CTTT]n		0.420
c	[CTTT]n CT [CTTT]n		0.047
d	[CTTT]n		0.017
e	[CTTC]n [CTTT]n TT [CTTT]n		0.013
f	[CTTT]n TT [CTTT]n	rs151261950	0.009
g	CTTC [CTTT]n	rs9362477	0.009
h	[CTTT]n	rs536914220	0.007
i	CTTC [CTTT]n CT [CTTT]n		0.006
g	[CTTT]n TT [CTTT]n TT [CTTT]n		0.004
k	[CTTT]n TT [CTTT]n	rs536914220	0.004
	all other motifs (less than 1% for each population)		0.018

(Manuscript in preparation)

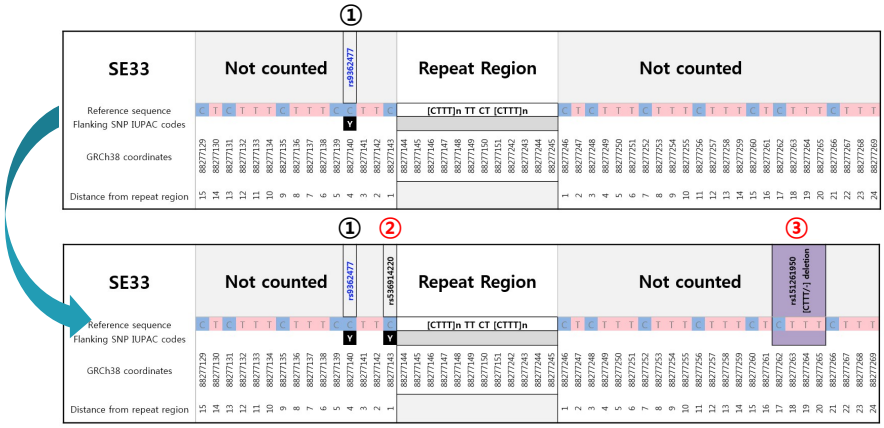
3. Results and Discussions

ICGSK 2019 *The Genetics Society of Korea*

❖ Characteristics of the SE33

➢ **Motif structure of SE33** ① rs9362477 ② rs536914220 ③ rs151261950

- Forensic STR Sequence Structure Guide v5.0 (https://strider.online/nomenclature)

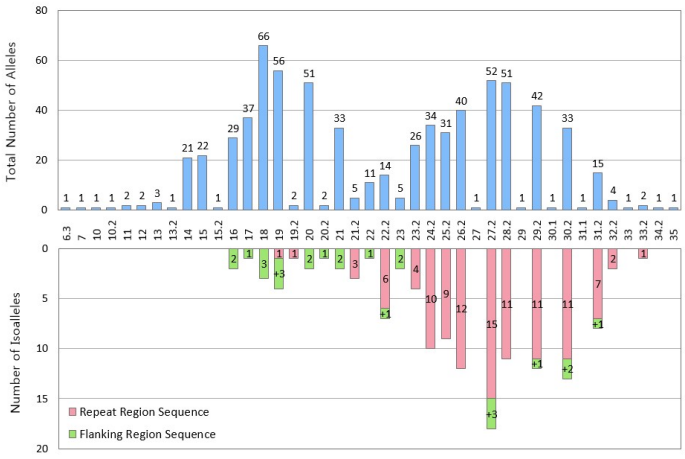


(Manuscript in preparation)

3. Results and Discussions

❖ Characteristics of the SE33

➢ **Allele distribution pattern of SE33**

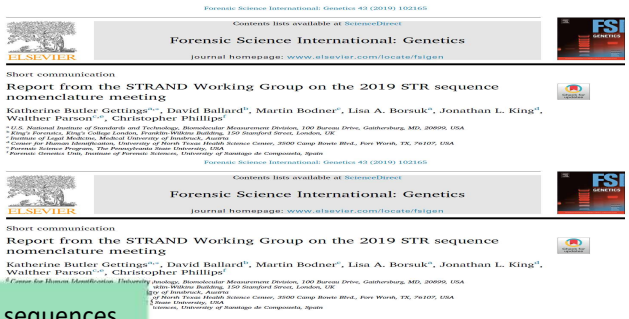


(Manuscript in preparation)

3. Results and Discussions

❖ Coordinate issue **(Topic to be discussed..)**

Gettings et al. (2019) FSIG


Forensic Science International: Genetics 43 (2019) 102165

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Short communication

Report from the STRAND Working Group on the 2019 STR sequence nomenclature meeting

Katherine Butler Gettings[a,*], David Ballard[b], Martin Bodner[c], Lisa A. Borsuk[a], Jonathan L. King[d], Walther Parson[c,e], Christopher Phillips[f]

1. Formats for STR sequences
2. **Defined coordinates** →
3. Forensic-specific reference

① **Assay specific**
② **Informative universal**
③ Unambiguous universal
④ Repeat region only

3. Results and Discussions

❖ Coordinate issue (Topic to be discussed..)

➢ **Assay specific coordinates** : Data compatibility ?

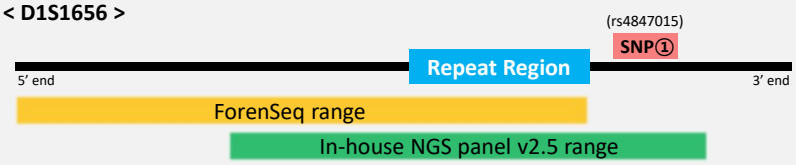
➢ **Informative universal coordinates**

: Assays include SNPs adjacent to the repeat region ?

◆ **Examples** **① rs4847015 in D1S1656** ② rs25768 in D5S818

+6bp, 28.34% (gnomAD)

< D1S1656 >



3. Results and Discussions

ICGSK 2019 The Genetics Society of Korea

❖ Coordinate issue (Topic to be discussed..)

➢ **Assay specific coordinates** : Data compatibility ?

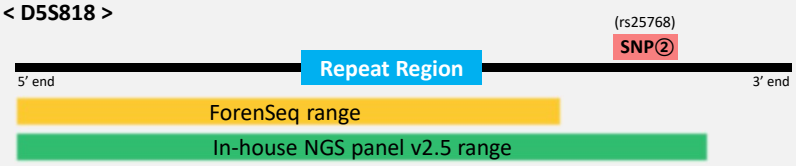
➢ **Informative universal coordinates**

: Assays include SNPs adjacent to the repeat region ?

◆ **Examples** ① rs4847015 in D1S1656 **② rs25768 in D5S818**

+13bp, 79.12% (gnomAD)

< D5S818 >



Conclusion

- We successfully analyzed **25 autosomal STRs** for 350 samples across **four representative populations** using the developed **in-house NGS panel v2.5**.
- We identified **sequence variation** which located in repeat and flanking region. The **SE33** is the most polymorphic locus and showed the largest increase of observed allele by sequence variations.
- Analyzing STRs using the **NGS method** provides additional sequence information and is useful by increasing the power of discrimination **in challenging caseworks** than CE method.
- **Coordinate issue** across the multiple assays should be further discussed in the international working group.

Acknowledgements

NRF National Research Foundation of Korea

Funded by the Ministry of Science and ICT
(NRF-2014M3A9E1069989)

BrainKorea21PLUS

Brain Korea 21 PLUS Project for
Medical Science, Yonsei University

Thank You

KJSHIN@yuhs.ac
yelim117@yuhs.ac
http://forensic.yonsei.ac.kr