Massively parallel sequencing of 33 Y-STRs including 11 rapidly mutating Y-STRs in three Asian populations

Mi Hyeon Moon[1] · Kyoung-Jin Shin[1,*]

[1]Department of Forensic Medicine, Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul 03722, Korea

Introduction

Y-chromosomal short tandem repeats (Y-STRs) are useful for male identification in sexual crimes and paternal lineage testing because of male-specific and non-recombinant characteristics. Recent studies have demonstrated that rapidly mutating (RM) Y-STRs have higher diversity and possibility of male individualization than conventional Y-STRs. Identifying sequence variations using massively parallel sequencing (MPS) can lead in much more increased diversity of RM Y-STRs than length-based analysis using capillary electrophoresis. As Korea becomes multicultural society, it is necessary to accumulate forensic genetic data on neighbouring Asian populations which have not been studied enough for Y-STRs. In this study, sequence variations of 33 Y-STRs including 11 RM Y-STRs (DYF387S1, DYF399S1, DYF404S1, DYS449, DYS518, DYS526a, DYS570, DYS576, DYS612, DYS626, and DYS627) were investigated using MPS for 261 males in Myanmar, Nepal and Pakistan populations. We will present sequence-based allele frequencies, distributions and gain in the number of alleles across Asian populations. This study is expected to show that identifying sequence variations of Y-STRs for Asians increases potential for male individualization in forensic practice.

Materials and Methods

Samples

- A total of 261 unrelated male samples across three Asian populations: Myanmar (MY, N=98), Nepal (NP, N=65), and Pakistan (Punjab;PJ, N=98) were obtained from Seoul National University Asian Sample Biobank.
- Approximately, 0.5 to 2 ng of DNA was used for library preparation.

MPS Library Preparation and Run

- The MPS library was generated using the developed in-house MPS panel which simultaneously amplified 33 Y-STRs including 11 RM Y-STRs by two-step PCR with DNA template.
- The MPS libraries were purified with 1.2 x Agencourt® AMPure® XP beads (Beckman Coulter) and quantified using KAPA library quantification kits (KAPA Biosystems) and 2100 Bioanalyzer (Agilent).
- The barcoded libraries were normalized to 10nM and then pooled in equal volumes. The pooled library was sequenced using MiSeq Reagent kit v3 (2 x 300 cycles, Illumina).

MPS Data Analysis

- FASTQ files were analyzed using the bioinformatic pipeline STRait Razor v3.0 and VisualSTR, in-ouse developed tool. We confirmed the genotype concordance between CE and MPS method and investigated sequence variation.

Results

Amplicon size and coverage

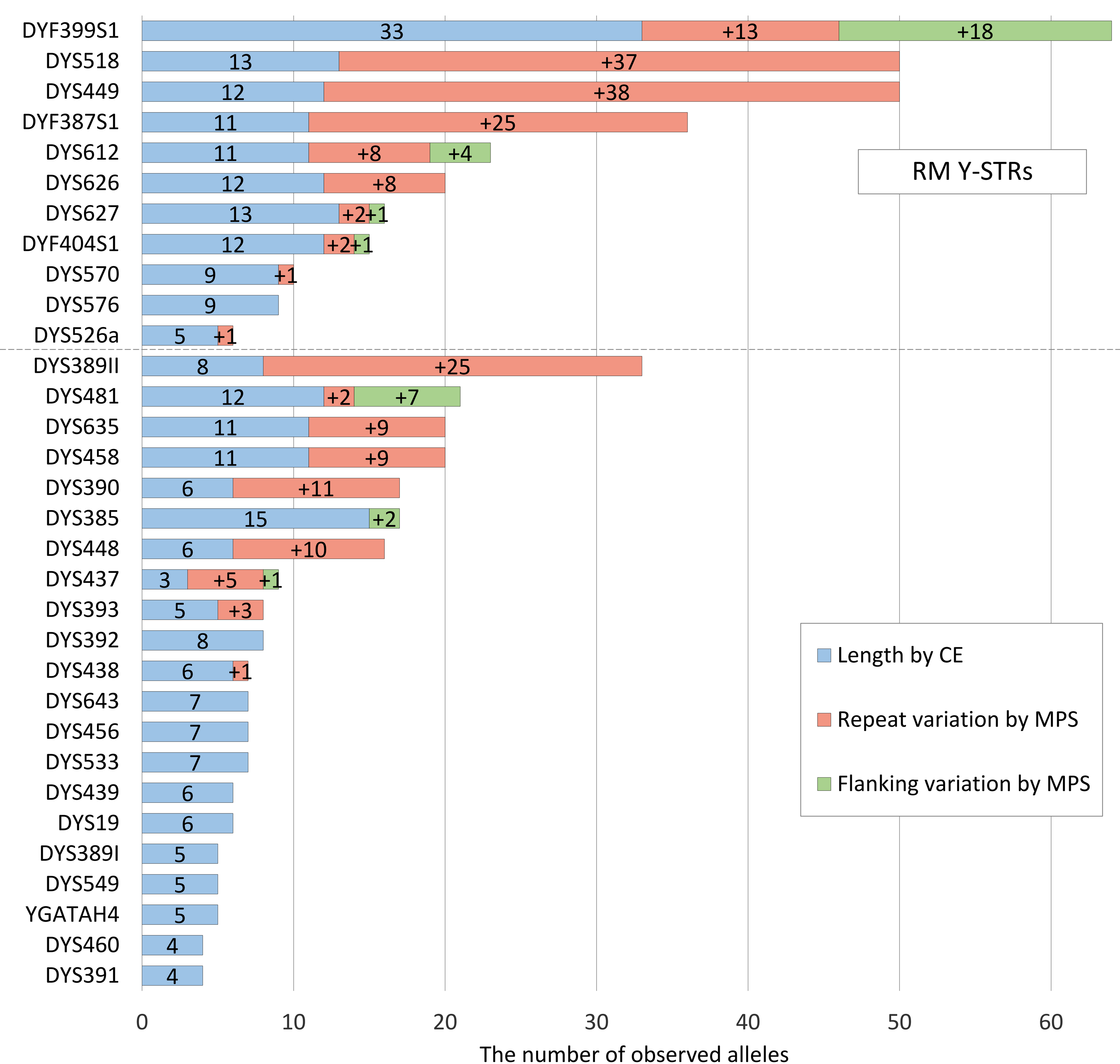
- The amplicon size of targeted markers ranged from 85 to 274bp.
- The depth of coverage (DoC) for each marker were calculated by read count obtained by STRait Razor. The average of minimum and maximum coverage was 827 in DYS626 and 9488 in DYF404S1, respectively.

Genotype concordance between CE and MPS method

- Three samples showed genotype discordance by SNP (rs757752030) of MPS primer binding site in DYS576 marker.
- In one Pakistan sample, null alleles were observed in six Y-STR markers by microdeletion.

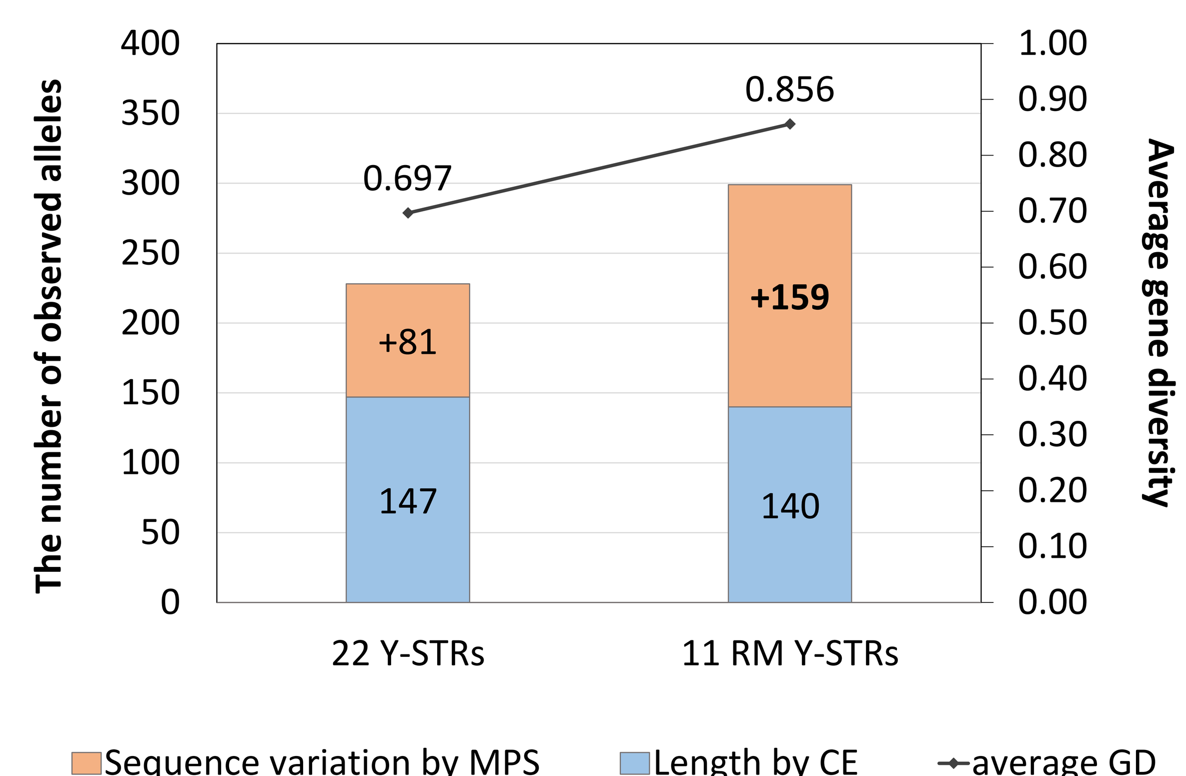
Sequence structure variation and allele gain

Fig. 1. The number of alleles obtained by length- and sequence-based method



- RM Y-STRs accounted for most of remarkable increase in the number of alleles by sequence variations. Repeat region variations are mainly observed in DYS449, DYS518, and DYF387S1. Most of flanking region variations are observed in DYF399S1 and DYS612.
- DYS449 marker showed the largest increase in the number of alleles due to repeat region variation (a 4.17-fold), followed by DYS518 marker (a 3.85-fold) and DYF387S1 marker (a 3.27-fold).

Fig. 2. The number of increased alleles and average gene diversity per Y-STR set



- A total of 527 alleles were identified by sequence analysis compared to 287 alleles by length-based analysis (a 1.84-fold). Of 33 Y-STRs, 11 RM Y-STRs accounted for 159 allele increase and 22 conventional Y-STRs accounted for 81 allele increase.
- Gene diversity of 11 RM Y-STRs was 0.856, which are higher than 22 conventional Y-STRs (0.697).

Table 1. Flanking region variations observed by sequence analysis

Marker	Flanking region variation	Wild	Mutant	Frequency		
				MY	NP	PJ
DYF399S1	rs4306075	A	G	0.331	0.319	0.276
DYF399S1	rs878949651	A	G	0.307	0.304	0.340
DYS437	rs9786886	C	T	-	-	0.010
DYS481	rs368663163	G	A	0.010	0.123	0.112
DYS612	rs555095027	T	C	0.041	0.077	0.020
DYS627	rs571126660	A	G	0.010	-	-
DYF399S1	rs2090175785	GA	-	0.003	-	0.013
DYF404S1	rs34228393	TTCT	-	-	-	0.006
DYS385ab	rs1248860842	CCTT	-	0.005	0.008	-
DYS481	rs2015234517	CTG	-	0.041	-	-

Conclusion

- We successfully analyzed sequence variations of 33 Y-STRs including 11 RM Y-STRs for 261 samples across three Asian populations using the developed in-house MPS panel.
- Significant increase from 287 (length) to 527 (sequence) in the number of alleles showed by sequence analysis.
- Part of RM Y-STRs showed more than three-fold increase in the number of sequence-based alleles compared to conventional Y-STRs.
- Sequence-based alleles were distinguished by flanking region variations and mainly observed in DYF399S1 marker.
- Applying MPS to Y-STRs analysis including RM Y-STRs improves the discrimination power and increases the possibility of distinguishing closely related males and especially in endogamous population which has low diversity of Y-STRs.

Acknowledgement

- This study was supported by a faculty research grant at Yonsei University College of Medicine (2020-2021)