creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

연세대학교
YONSEI UNIVERSITY

차세대 염기서열 분석법을 이용한 3개 아시아 인구집단의 Y 염색체 STR 염기서열 변이 분석

연세대학교 대학원
의 과 학 과
문 미 현

연세대학교
YONSEI UNIVERSITY

차세대 염기서열 분석법을 이용한 3개 아시아 인구집단의 Y 염색체 STR 염기서열 변이 분석

연세대학교 대학원
의 과 학 과
문 미 현

연세대학교
YONSEI UNIVERSITY

차세대 염기서열 분석법을 이용한 3개 아시아 인구집단의 Y 염색체 STR 염기서열 변이 분석

지도교수 신 경 진

이 논문을 석사 학위논문으로 제출함

2021 년 12 월

연세대학교 대학원

의 과 학 과

문 미 현

연세대학교
YONSEI UNIVERSITY

문미현의 석사 학위논문을 인준함

심사위원 이숭덕 (인)

심사위원 신경진 (인)

심사위원 김상우 (인)

연세대학교 대학원

2021 년 12 월

연세대학교
YONSEI UNIVERSITY

감사의 글

제 인생에 있어 크나큰 한 조각이 될 학위 과정을 많은 분들의 도움으로 마무리 짓게 되었습니다. 학위 과정 중에는 무거운 돌을 짊어진 것 같았지만 지나고 보니 그 돌이 거친 물살에 휩쓸리지 않게 해줬던 원천이었던 것 같습니다. 먼저, 세심한 지도로 부족한 저를 앞으로 나아갈 수 있게 도와주신 신경진 교수님께 감사 말씀드립니다. 좋은 말씀들로 항상 격려 해주시는 이승덕 교수님, 좋은 논문이 될 수 있도록 이끌어 주신 김상우 교수님께도 감사드립니다. 항상 먼저 챙겨주시는 이환영 교수님, 많은 관심과 격려 주신 양우익 교수님과 박종필 교수님께도 감사드립니다. 좋은 논문을 쓸 수 있도록 먼저 초석을 다져 주신 권소연, 이은혜 선배님께도 감사드립니다. 좋은 학위논문이 될 수 있도록 지원해주신 이지현, 조소희 선생님께도 감사드립니다. 내 일처럼 먼저 나서서 도와주시는 새롬언니 그리고 학위 과정 내내 함께 울고 웃었던 보민이와 예림언니, 나보다 학위 과정을 더 잘 해낼 수민이에게도 덕분에 견뎌낼 수 있었다는 감사의 말 전합니다. 무슨 일이 있어도 항상 내 편인 사랑하는 우리 가족, 엄마, 아빠, 동생에게 변함없는 믿음과 응원 보내줘서 감사하다는 말씀 글로 먼저 전합니다. 내 자존감 지킴이 설용오빠, 태성오빠, 소정이, 수빈오빠, 형안언니, 혁이오빠 그리고 생명과학과 동기들에게도 감사드립니다. 도움을 받고 나누며 모든 일은 혼자만으로는 되지 않는다는 것을 알게 된 소중한 시간이었습니다. 그동안 받은 사랑과 도움을 잊지 않고 필요로 하는 곳에 베풀 수 있는 사람이 되겠습니다. 정말 감사드립니다.

연세대학교
YONSEI UNIVERSITY

차 례

국문 요약····· 1

I. 서론 ····· 4
II. 재료 및 방법 ····· 7
1. 연구대상 ····· 7
2. DNA 추출 및 정량 ····· 8
3. MPS 분석을 위한 다중 중합효소연쇄반응 체계 구축·9
가. 연구 대상 Y 염색체 STR 표지자····· 9
나. 다중 중합효소연쇄반응 체계 구축····· 9
4. 자체 구축한 CE 기반 체계를 이용한 아시아 인구 집단의 유전자형 확보·····16
5. 세 아시아 인구 집단 시료의 Y 염색체 STR에 대한 MPS 수행·····26
가. 아시아 인구 집단 시료를 대상으로 한 Y 염색체 STR 라이브러리(library) 제작·····26
나. 아시아 인구 집단 시료를 대상으로 한 Y 염색체 STR 라이브러리(library) 정제 및 정도 관리 (Quality control) ·····27
다. MPS 장비를 이용한 Y 염색체 STR 염기서열 분석·····28
라. MPS 결과를 이용한 Y 염색체 STR 염기서열 변이 자료 분석·····29

연세대학교
YONSEI UNIVERSITY

결과 ·····33
1. 아시아 집단 시료의 Y 염색체 STR의 MPS 분석 결과·····33
2. CE와 MPS 방법을 통해 얻은 Y 염색체 STR 유전자형 비교 결과·····35
3. 아시아 집단 시료의 Y 염색체 STR의 반복 지역 및 주변 영역 염기서열 변이 분석 결과 및 대립 유전자 수의 증가·····36
4. 아시아 집단 시료 간 대립유전자 분포 양상·····94

IV. 고찰 ·····98
V. 결론 ·····113

참고 문헌 ·····115
ABSTRACT ·····120

연세대학교
YONSEI UNIVERSITY

그림 차례

그림 1. Electropherogram result of genotyping the 17 Y-chromosomal markers using in-house Euplex-Y17N PCR system for 1ng of 2800M control DNA·····17

그림 2. Electropherogram result of genotyping the 18 Y-chromosomal markers using in-house Kplex-Y18 PCR system for 1ng of 2800M control DNA··18

그림 3. Average depth of coverage (DoC) for the 34 Y-chromosomal markers used in this study for all 261 male DNA samples·····34

그림 4. Comparison of the number of observed alleles by CE and MPS of 33 Y-STRs·····87

그림 5. Difference in allele distribution patterns observed across Asian populations·····96

그림 6. The number of alleles observed by each length-, sequence-based analysis between conventional STR and RM Y-STR set·····99

그림 7-1. The number of increased alleles by sequence variations for each population in Moon et al.'s study·····102

연세대학교
YONSEI UNIVERSITY

그림 7-2. The number of increased alleles by sequence variations for each population in this study·····102

그림 8. The distributions of deletion and insertion in the Y-M175 marker for Asian populations on Moon et al. and this study·····109

그림 9-1. Example of CE genotyping results for DYS576 in the MY028 sample with genotype discordance·····111

그림 9-2. Flanking region SNP observed within the 3' MPS primer binding site leading to genotype discordance in DYS576 marker·····111

연세대학교
YONSEI UNIVERSITY

표 차례

표 1. Information of primers used for the multiplex PCR amplification of the 33 Y-STRs and a Y-SNP including allelic ranges and size range·····11

표 2. Primers' information used for the multiplex amplification of in-house Euplex-Y17N system···19

표 3. Primers' information used for the multiplex amplification of in-house Kplex-Y18 system·····22

표 4. Configuration file's information used in the STRait Razor v3.0 program for sequence variation analysis of 34 Y-chromosomal markers·····30

표 5. Sequence variations and frequencies for 34 Y-chromosomal markers·····37

표 6. Sequence variations observed in flanking region of Y-STRs·····91

연세대학교
YONSEI UNIVERSITY

국문요약

차세대 염기서열 분석법을 이용한 3개 아시아 인구집단의 Y 염색체 STR 염기서열 변이 분석

Y 염색체 STR (Y chromosomal STR, Y-STR)은 남성 특이적이며 세대 간 재조합 없이 전달되는 특성이 있어 부계 혈통을 추정하거나 남성·여성 혼합물 내 남성 프로필의 확인에 유용하게 사용된다. 그러나, 기존의 Y 염색체 STR은 동일 부계에 속한 남성은 동일한 Y-STR 하플로타입을 가지게 되어 남성 개인 식별에 제한이 있어 Rapidly mutating (RM) Y 염색체 STR을 도입하여 동일 부계 내에서 남성 개인을 식별하고자 하는 연구가 진행되고 있다. 한편, 차세대 염기서열 분석법 (Massively Parallel Sequencing, MPS)으로 Y 염색체 STR을 분석하게 되면 보다 많은 마커를 한 번에 분석할 수 있으며, 염기서열 변이도 동시에 분석할 수 있어 식별력을 증대시킬 수 있다.

본 연구에서는 법유전학적 연구가 많지 않은 미얀마, 네팔, 파키스탄 인구집단에 속하는 261명을 대상으로 11 RM Y 염색

연세대학교
YONSEI UNIVERSITY

체 STR 포함한 33개의 Y 염색체 STR을 MPS 방법으로 분석하여 STR 반복 영역 및 그 주변 영역의 염기서열 변이 및 분포를 조사하고, CE 분석법과 비교하여 관찰된 대립 유전자 수의 증가 정도를 제시하였으며, 기존 CE 방법과 MPS 방법 사이의 유전자형 일치정도를 조사하였다.

3개의 샘플이 DYS576 표지자에서 MPS 프라이머 결합 부위의 단일염기다형성 (Single nucleotide polymorphism, SNP) 돌연변이로 인해 모세관 전기영동법 (Capillary Electrophoresis, CE)과 MPS 방법 사이에 유전자형 불일치를 보였다. CE 기반 대립 유전자 수와 비교하였을 때, MPS 분석을 통해 240개의 대립 유전자 증가를 확인하였다 (287개에서 527개). 반복 영역의 변이로 인해 가장 많은 대립 유전자 수의 증가를 보인 표지자는 DYS449 였고 (4.17배 증가), 다음은 DYS518 이었다. 반복 영역의 주변 영역의 변이로 인해 가장 많은 대립 유전자 수의 증가를 보인 표지자는 DYF399S1 였고 (1.55배 증가), 다음은 DYS481 이었다. 분석된 33개 Y 염색체 STR 중 기존 22개 Y 염색체 STR보다 개수가 적은 11개의 RM Y 염색체 STR에서 훨씬 다양한 염기서열을 나타냈다. 특정 인구 집단에서 관찰된 대립 유전자의 상대적인 분

연세대학교
YONSEI UNIVERSITY

포 차이가 DYS393 및 DYS635 표지자에서 남아시아 (네팔 및 파키스탄) 그리고 동남아시아 (미얀마) 인구 집단으로 나뉘어 관찰되었다. 또한, 특정 인구 집단에서만 관찰된 반복 영역의 주변 영역의 다형성도 확인되었다.

본 연구를 통해 주변 아시아 인구집단을 대상으로 Y 염색체 STR의 염기서열 분석에 대한 효용성을 보여주었다. 다양성이 높은 RM Y 염색체 STR의 염기서열 분석으로 얻어진 대립유전자 수의 증가와 그 분포를 제시하여 남성 개인 식별과 생물지리학적 정보 추정 가능성을 제시하였다. 아시아 인구집단에서 관찰된 Y-STR 염기서열 변이 정보는 법과학 실무에서 주변 아시아 인구 집단 남성과 관련된 사건을 해결하는데 기초자료로 활용될 것으로 기대된다.

핵심되는 말: Y 염색체 STR, RM Y 염색체 STR, 차세대 염기서열 분석법, 염기서열 변이, 아시아 인구 집단

연세대학교
YONSEI UNIVERSITY

차세대 염기서열 분석법을 이용한 3개 아시아 인구집단의 Y 염색체 STR 염기서열 변이 분석

<지도교수 **신 경 진**>

연세대학교 대학원 의과학과

문 미 현

I. 서 론

Y 염색체 STR (Y-STR)은 부계 혈통 검사와 성범죄 현장에서 여성의 DNA가 남성의 것보다 상대적으로 과다하게 많은 혼합물이 발견되는 경우에 남성을 식별하는 데 유용하게 사용되고 있다.[1-3] 그러나 동일 부계의 경우 이론적으로 같은 하플로타입을 가지기 때문에 남성 친척 간의 구별이 어렵게 된다. 이에, 빠르게 변이하여 동일 부계 내에서도 하플로타입의 차이를 보여 남성 개인 식별의 가능성을 높여주는 rapidly mutating Y 염색체 STR (Rapidly mutating Y-STR; RM Y-STR)에 대한 연구가 진행되어 오고 있다.[4-7]

연세대학교
YONSEI UNIVERSITY

법유전학 영역에서는 모세관 전기영동법(Capillary electrophoresis, CE)을 사용하여 DNA 길이를 측정하여 STR의 대립유전자형을 결정해 왔으나 차세대 염기서열 분석법 (massively parallel sequencing, MPS)은 보다 많은 표지자 및 시료를 동시에 분석할 수 있고 리드 수 기반의 염기서열 변이를 정확하게 확인하고 구분할 수 있다는 여러 이점으로 인해 법유전학 영역에서 STR 분석에 적용하고자 하는 여러 연구들이 진행되어 오고 있다.[8-13]

현재까지 동남아시아 및 남아시아 인구 집단을 대상으로 Y 염색체 STR을 분석한 연구들이 진행되었으나,[14,15] 대부분 CE 방법에 기반한 하플로타입 빈도 분포 등을 확인하는데 그쳤다. 특히, Adnan 등[6] 과 Rakha 등[7]의 연구는 RM Y 염색체 STR을 분석하여 동족 내 결혼으로 인한 특정 가계 문화를 가져 기존 Y 염색체 STR의 다양성이 낮았던 남아시아 인구집단에서 식별력이 상당히 증가함을 보여 그 중요성을 보고하였다.

최근 상용화된 MPS 키트를 이용하여 대부분 대표적인 인구집단(아프리카계 아메리카인, 유럽인 등)을 대상으로 하고 있어[9, 10] 아시아 하위 인구 집단의 염기서열 기반의 자료는 부족한 실정이다. 한국인 인구집단을 대상으로 Y 염색체 STR 염기서열 변이를 분석한 연구가 있으나, [8, 11] 최근 다문화 사회로 이행하는 한국 사회에서 다양한 체류 외국인 집단에 대한 법유전학적 자료를 축적할 필요가 있다. 특히, 주변 아시아 인구 집단 중 법유전학 연구가 많지 않은

연세대학교
YONSEI UNIVERSITY

동남아시아 및 남아시아 인구 집단을 대상으로 RM Y 염색체 STR을 포함한 Y 염색체 STR 분석 자료가 필요한 실정이다.

본 연구에서는 MPS 방법을 사용하여 법유전학 연구가 많지 않은 네팔, 미얀마, 파키스탄 세 아시아 인구 집단을 대상으로 다수의 RM Y 염색체 STR을 포함한 Y 염색체 STR의 염기서열 변이를 조사하였다. 기존 CE 방법과 MPS 방법을 통해 얻은 유전자형의 일치 정도를 확인하였다. 또한, MPS 방법을 통해 식별한 대립 유전자 수의 증가 정도와 인구 집단 별로 염기 서열 분포를 파악해 보았다. 나아가 RM Y 염색체 STR을 포함하여 Y 염색체 STR의 서열 분석의 유용성을 제시하고, 확인된 염기서열 변이의 특징을 활용하여 남성 개인 식별 및 생물지리학적 기원을 추정하는데 필요한 기초자료를 제시하고자 하였다.

연세대학교
YONSEI UNIVERSITY

II. 재료 및 방법

1. 연구대상

본 연구에서는 MPS를 이용해 세 개의 아시아 인구 집단을 대상으로 Y 염색체 STR 염기서열을 분석하고, 이들의 빈도가 각 인구 집단에서 어떠한 분포 양상을 보이는지 조사하여 그에 따른 기초 통계량을 제시하고자 하였다. 이를 위해 서울대학교 아시아인 시료/정보 네트워크 시료 은행에 보관된 네팔 (Nepal; NP), 미얀마 (Myanmar, MY), 파키스탄 (Pakistan-Punjabi; PJ) 인구 집단 각 65, 98, 98명, 총 261명의 DNA 시료 혹은 FTA카드에 도말 건조된 혈액 시료를 지정하여 분양 받았다. DNA 시료 및 FTA카드에 도말 건조된 혈액은 1.5 ml의 보관 튜브에 넣고 DNA를 추출할 때까지 4℃ 냉장고에 보관하였다. 본 연구는 연세대학교 의료원 세브란스병원 연구심의위원회와 (승인번호: 4-2019-1266) 서울대학교 병원 연구심의위원회의 승인(승인번호: 1404-068-572)을 받아 수행되었다.

연세대학교
YONSEI UNIVERSITY

2. DNA 추출 및 정량

보관된 FTA카드에 도말 건조된 미얀마와 파키스탄 인구집단의 혈액 시료는 QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany)를 이용하여 제조사의 지시에 따라 DNA를 추출하였다. 분양 받은 DNA시료와 추출된 시료는 각각 Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) 그리고 Qubit 4 Fluorometer 와 Qubit dsDNA HS Assay Kit Qubit 4 (Thermo Fisher Scientific)를 이용하여 정량을 측정한 후 -20℃에 보관하였다. 1ng/µl 이상의 농도가 측정된 시료는 다음 단계 실험을 위하여 1ng/µl 로 희석되었다.

연세대학교
YONSEI UNIVERSITY

3. MPS 분석을 위한 다중 중합효소연쇄반응 체계 구축

가. 연구 대상 Y 염색체 STR 표지자

현재 법과학 영역에서 Y 염색체 STR 분석에 널리 사용되는 CE 기반의 상용 키트 중 PowerPlex® Y23 system (Promega, Madison, WI, USA)와 Yfiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific)에 포함된 표지자와 일부 RM Y 염색체 STR (DYF399S1, DYF404S1, DYS526a, DYS612, DYS626) 및 하나의 Y 염색체 단일 염기다형성 표지자(Y-M175)을 포함하여 총 34개의 Y 염색체 표지자 (DYF387S1, DYF399S1, DYF404S1, DYS19, DYS385ab, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS449, DYS456, DYS458, DYS460, DYS481, DYS518, DYS526a, DYS533, DYS549, DYS570, DYS576, DYS612, DYS626, DYS627, DYS635, DYS643, Y-GATAH4, Y-M175)를 분석 대상으로 하였다. 그 중 RM Y 염색체 STR은 총 11개(DYF387S1, DYF399S1, DYF404S1, DYS449, DYS518, DYS526a, DYS570, DYS576, DYS612, DYS626, DYS627)가 분석되었다.

나. 다중 중합효소연쇄반응 체계 구축

연세대학교
YONSEI UNIVERSITY

총 34개의 Y 염색체 표지자를 동시에 분석하기 위하여 다중 중합효소연쇄반응 체계를 구축하였다. 이 체계는 이전 연구에서[8] 구축된 24개의 Y 염색체 표지자를 동시에 분석할 수 있는 다중 중합효소연쇄반응 체계에 9개의 RM Y 염색체 STR (DYF387S1, DYF399S1, DYF404S1, DYS449, DYS518, DYS526a, DYS612, DYS626, DYS627)와 Yfiler Plus PCR Amplification Kit에 포함된 DYS460 표지자를 추가하여 구축하였다. 사용한 primer 염기서열과 농도는 표 1에 표시하였다.

연세대학교
YONSEI UNIVERSITY

표 1. Information of primers used for the multiplex amplification of the 33 Y-STRs and a Y-SNP including allelic ranges and size range

Locus		Primer sequences (5'→3')*	Conc. (μM)	Allelic range	Size range (bp)
DYF387S1	F	NT1-ATT CCA TTT TAC CCC TAA CAA GA	0.45	31-43	213-263
	R	NT2-TGA GAA GTG TGA GAA GTG CTA CC	0.45		
DYF399S1	F	NT1-CAA AAG CCC AAC AGG AAA AA	0.90	15-31	117-184
	R	NT2-CCA GGA GAA AAG AAC CCA GA	0.90		
DYF404S1	F	NT1-GGC TTA AGA AAT TTC AAC GCA TA	0.25	9-21	140-188
	R	NT2-GAC AAA AAA ATG AGA CAC TGT CAA	0.25		
DYS19	F	NT1-CCA TCT GGG TTA AGG AGA GTG T	1.40	9-19	155-195
	R	NT2-TTC ACT ATG ACT ACT GAG TTT CTG TT	1.40		
DYS385ab	F	NT1-GGA AGG AGA AAG AAA GTA AAA AAG AA	2.00	7-28	169-253
	R	NT2-TTC CAA TTA CAT AGT CCT CCT TTC T	2.00		
DYS389I	F	NT1-CAA CTC TCA TCT GTA TTA TCT ATG TAT C	0.55	9-17	85-117
DYS389II	F	NT1-CAA CTC TCA TCT GTA TTA TCT ATG TGT G	1.80	24-35	197-241
	R*	NT2-GAT AGA TTG ATA GAG GGA GGG A	1.80		

연세대학교
YONSEI UNIVERSITY

Continued)

Locus		Primer sequences (5'→3')*	Conc. (μM)	Allelic range	Size range (bp)
DYS390	F	NT1-CAA TGT GTA TAC TCA GAA ACA AGG A	0.35	17-29	142-190
	R	NT2-CTG CAT TTT GGT ACC CCA TA	0.35		
DYS391	F	NT1-TTC AAT CAT ACA CCC ATA TCT GTC	0.38	5-16	97-141
	R	NT2-TGC AAG CAA TTG CCA TAG AG	0.38		
DYS392	F	NT1-TAA ACC TAC CAA TCC CAT TCC T	0.80	4-20	97-145
	R	NT2-TTT GTT ATT TAA AAG CCA AGA AGG	0.80		
DYS393	F	NT1-TGTGGTCTTCTACTTGTGTCAA	0.31	7-18	101-145
	R	NT2-AAA CTC AAG TCC AAA AAA TGA GG	0.31		
DYS437	F	NT1-GAC TAT GGG CGT GAG TGC AT	0.25	11-18	111-139
	R	NT2-GAT AAG TAG ATA GAC ATC ATT CAC AGA	0.25		
DYS438	F	NT1-TGG GGA ATA GTT GAA CGG TAA	0.42	6-16	88-138
	R	NT2-GCA ACA AGA GTG AAA CTC CAT T	0.42		
DYS439	F	NT1-ACA TAG GTG GAG ACA GAT AGA TGA	0.29	6-17	100-144
	R	NT2-GGC TTG GAA TTC TTT TAC CCA	0.29		

연세대학교
YONSEI UNIVERSITY

Continued)

Locus		Primer sequences (5'→3')*	Conc. (μM)	Allelic range	Size range (bp)
DYS448	F	NT1-AGA AAG GGA GAT AGA GAC ATG GA	0.80	14-24	190-250
	R	NT2-TGG CCG GTC TGG AAA TTT AT	0.80		
DYS449	F	NT1-AAG TTC CCC CAT GTT TTT TTC	2.00	23-38	214-274
	R	NT2-TGG ACA ACA AGA GTA AGA CAG AAA	2.00		
DYS456	F	NT1-CTG TTG TGG GAC CTT GTG ATA	0.48	11-23	134-182
	R	NT2-ACT CAG CCC AAA ACT TCT TAA A	0.48		
DYS458	F	NT1-GCA ACA GGA ATG AAA CTC CAA	0.42	11-24	130-182
	R	NT2-CCC AAA GTT CTG GCA TTA CAA	0.42		
DYS460	F	NT1-TCA TCT ATC CTC TGC CTA TCA TTT	0.33	7-14	101-129
	R	NT2-GAA TAC CAG AGG AAT CTG ACA CC	0.33		
DYS481	F	NT1-CTC ACC AGA AGG TTG CAA GAC	0.23	17-32	119-164
	R	NT2-GTC TGC TAA AAG GAA TGT GGC	0.23		
DYS518	F	NT1-GGC AAC ACA AGT GAA ACT GC	2.00	32-47	203-263
	R	NT2-TCC TTC CTT CCT TCT TTC TTT TC	2.00		

연세대학교
YONSEI UNIVERSITY

Continued)

Locus		Primer sequences (5'→3')*	Conc. (μM)	Allelic range	Size range (bp)
DYS526a	F	NT1-TGA TCC AAA CCT TTA CTT ACT TTC TAT	1.90	11-18	135-163
	R	NT2-GTT TGG GTT ACT TCG CCA GA	1.90		
DYS533	F	NT1-TCT TCT ACC TAT CAT CTT TCT AGC	0.55	7-17	102-142
	R	NT2-TCA GTT CTT AAC TCA ACC AAA CAA	0.55		
DYS549	F	NT1-GTC CCC TTT TCC ATT TGT GA	0.70	7-17	115-155
	R	NT2-GCA ATT AGG TAG GTA AAG AGG AAG A	0.70		
DYS570	F	NT1-CAA CCT AAG CTG AAA TGC AGA	0.23	10-25	100-160
	R	NT2-GCT GTG TCC TCC AAG TTC CT	0.23		
DYS576	F	NT1-AAT AAG CGT ATT TGT CTT GGC T	0.33	11-23	95-143
	R	NT2-CAT AGC AAG ACC TCA TCT CTG AA	0.33		
DYS612	F	NT1-AGA GCC AAA AAG GGA ACT GA	0.50	25-43	150-204
	R	NT2-TCA CAC AGG TTC AGA GGT TTG	0.50		
DYS626	F	NT1-GCA AGA CCC CAT AGC AAA AG	1.90	24-36	201-249
	R	NT2-CTT TCT TCC CTT TCT TTC TTT CC	1.90		

연세대학교
YONSEI UNIVERSITY

Continued)

Locus		Primer sequences (5'→3')*	Conc. (μM)	Allelic range	Size range (bp)
DYS627	F	NT1-GCC TTT CAT TCT CTC CTT CGT	0.80	13-27	217-273
	R	NT2-AAA TCA AAG AGA GAC ACA GGG A	0.80		
DYS635	F	NT1-TGG CTT CTC ACT TTG CAT AGA A	0.65	15-28	136-188
	R	NT2-GTG GAA CCA GCC CAA ATA TC	0.65		
DYS643	F	NT1-TGC CTG GTT AAA CTA CTG TGC	0.38	6-17	94-149
	R	NT2-TCC CCC CAA AAT TCT ACTGA	0.38		
YGATAH4	F	NT1-CTA TTC ATC CAT CTA ATC TAT CCA	1.80	8-18	121-161
	R	NT2-ATG CTG AGG AGA ATT TCC AA	1.80		
Y-M175	F	NT1-TGA TTT AAA CTC TCT GAA TCA GGC	0.27	1-2	87-92
	R	NT2-TGA TAC CTT TTT TTC TAC TGA TAC CTT T	0.27		

*NT1 and NT2 sequences are Nextera adapter sequences as follows

NT1 sequence: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

NT2 sequence: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

*DYS389Ⅱ reverse primer was used to amplify DYS389 I and DYS389Ⅱ commonly.

연세대학교
YONSEI UNIVERSITY

4. 자체 구축한 CE 기반 체계를 이용한 아시아 인구 집단의 유전자형 확보

MPS 수행을 통해 얻게 되는 유전자형과 비교, 확인하기 위하여 먼저 자체 구축한 CE 기반 다중 중합효소연쇄반응 체계 2개를 사용하여 총 261명의 아시아 인구 집단 시료를 대상으로 Y 염색체 STR 유전자형을 확보하였다. 해당 패널은 Euplex-Y17N과 Kplex-Y18로 명명하였으며, 각각 17개와 18개의 Y 염색체 표지자의 유전자형을 검출할 수 있게 설계하였다. 각 패널은 현재 법유전학 영역에서 주로 사용되는 상용화된 CE 패널의 좌위들과 호환되도록 구성되었다. Euplex-Y17N은 PowerPlex® Y23 system (Promega)과 이번 연구에서 분석한 RM Y-STR을 통합한 좌위 세트(그림 1), Kplex-Y18은 Yfiler™ PCR Amplification Kit (Thermo Fisher Scientific) 좌위 세트(그림 2)와 호환되도록 하였다. 4개의 형광 채널에서 (FAM, VIC, NED 및 PET) 증폭산물의 크기가 겹치지 않도록 하여 배치하였다. 각 형광 채널에 사용된 표지자의 프라이머 정보는 표 2와 3에 정리하였다.

연세대학교
YONSEI UNIVERSITY

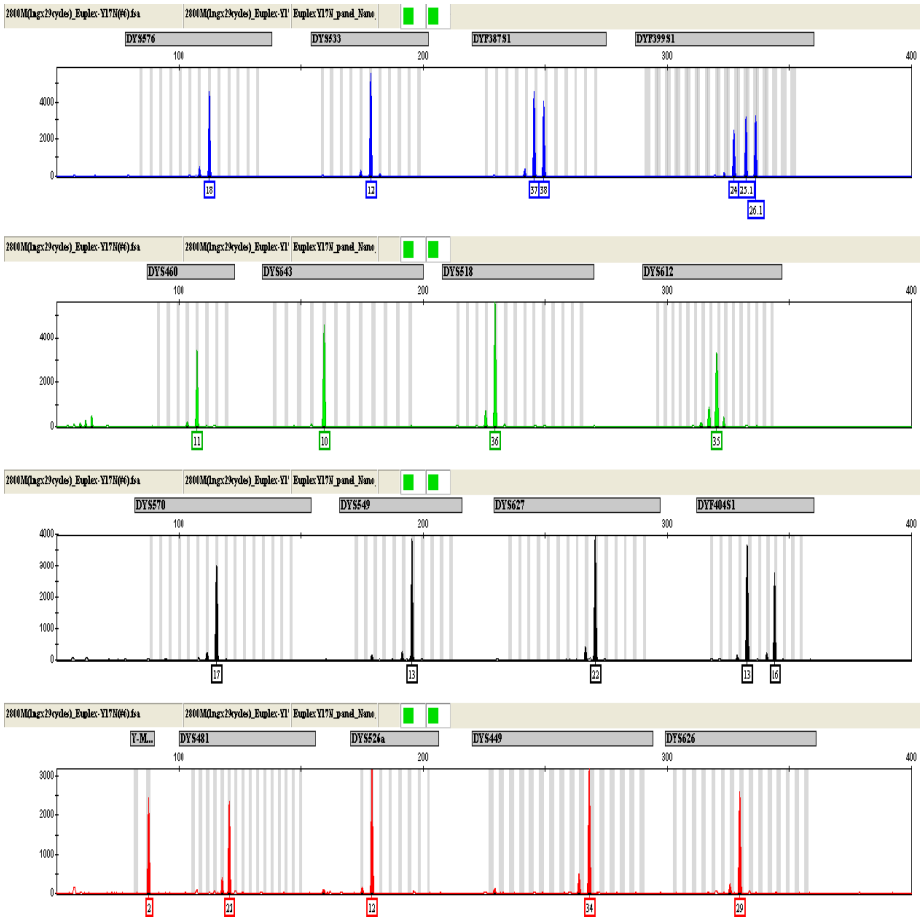

그림 1. Electropherogram result of genotyping the 17 Y-chromosomal markers using in-house Euplex-Y17N PCR system for 1ng of 2800M control DNA

연세대학교
YONSEI UNIVERSITY

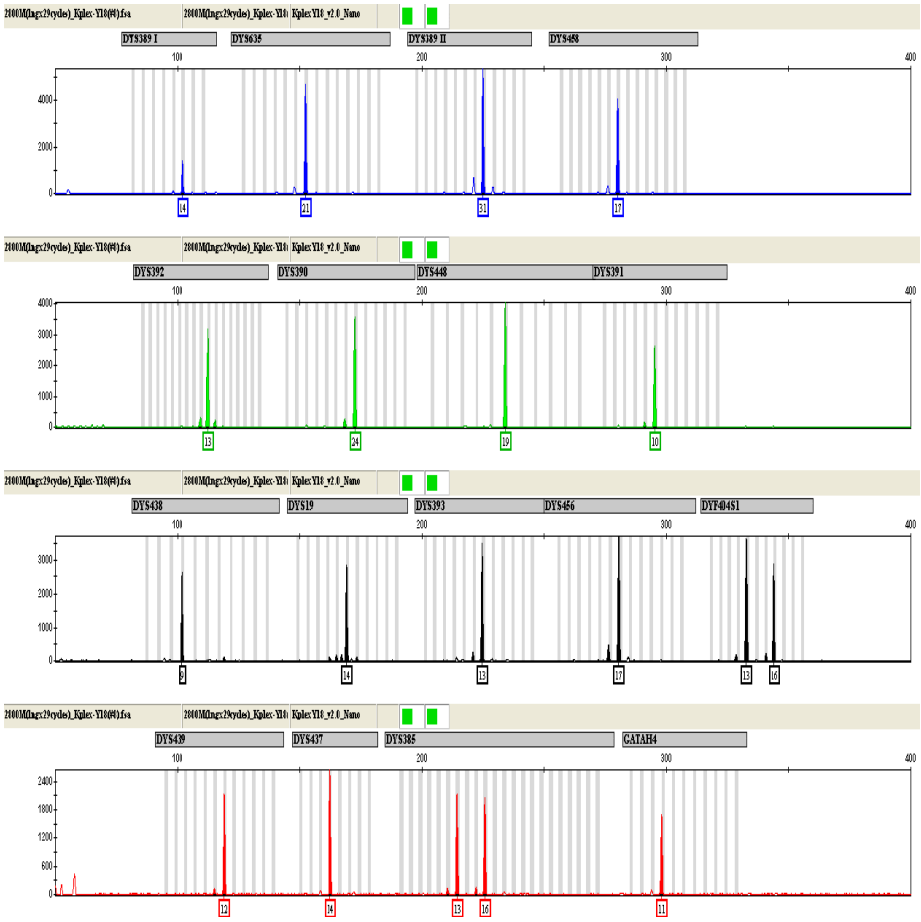

그림 2. Electropherogram result of genotyping the 18 Y-chromosomal markers using in-house Kplex-Y18 PCR system for 1ng of 2800M control DNA

연세대학교
YONSEI UNIVERSITY

표 2. Primers' information used for the multiplex amplification of in-house Euplex-Y17N system

Dye	STR loci		Primer Sequence (5'→3')	Conc. (μM)
FAM	DYS576	F	FAM-GCG TAT TTG TCT TGG CTY TTT C	2.25
		R	GCA AGA CCT CAT CTC TGA ATA AAA	2.25
	DYS533	F	FAM-TCT TCT ACC TAT CAT CTT TCT AGC	1.40
		R	ACG CTG TTA TTT CAT GGC TGT	1.40
	DYF387S1	F	GTC CAT TTT ACC CCT AAC AAG AAA	3.00
		R	FAM-GCC ACA GTG TGA GAA GTG TGA	3.00
	DYF399S1	F	FAM-GAA CCA AGG GAA ATG TGG AA	4.00
		R	GGG ACA TTC CTC TTC AAT GC	4.00
VIC	DYS460	F	GTC ATC TAT CCT CTG CCT ATC ATT T	0.90
		R	VIC-GAG GAA TCT GAC ACC TCT GAC A	0.90
	DYS643	F	GTG TGA TT T TTG CAG GTG TTC	1.50
		R	VIC-GTA ACC AAA CAC CAC CCA TTC	1.50

연세대학교
YONSEI UNIVERSITY

Continued)

Dye	STR loci		Primer Sequence (5'→3')	Conc. (μM)
VIC	DYS518	F	VIC-GGC AAC ACA AGT GAA ACT GC	4.25
		R	GCC TTC TTT TCC CTC CTT CCT	4.25
	DYS612	F	VIC-AGA GCC AAA AAG GGA ACT GA	5.00
		R	GTC AAC CAA CCA CTA ACC AAC A	5.00
NED	DYS570	F	NED-AAG CTG AAA TGC AGA TAT TCC C	2.50
		R	GCT GTG TCC TCC AAG TTC CT	2.50
	DYS549	F	NED-TGT GTG CAT AGA GGT GTT CAG A	3.00
		R	AGG TAG GTA AAG AGG AAG ATG A	3.00
	DYS627	F	GCC TTT CAT TCT CTC CTT CGT	4.50
		R	NED-GCG CAG GAT TCC ATC TAA AA	4.50
	DYF404S1	F	NED-CCT GGT AAT CTT CCA AAA TCG T	3.00
		R	GAA AGA TCA AAG GAG CCC AG	3.00

연세대학교
YONSEI UNIVERSITY

Continued)

Dye	STR loci		Primer Sequence (5'→3')	Conc. (μM)
PET	Y-M175	F	AAA CTC TCT GAA TCA GGC ACA	2.40
		R	PET-TGA TAC CTT TTT TTC TAC TGA TAC CTT T	2.40
	DYS481	F	PET-CTC ACC AGA AGG TTG CAA GAC	1.05
		R	AGG AAT GTG GCT AAC GCT GT	1.05
	DYS526a	F	CTT TCC TTC ATT TCT TCT TTC TTT C	3.50
		R	PET-GTT TGG GTT ACT TCG CCA GA	3.50
	DYS449	F	PET-AAG TTC CCC CAT GTT TTT TTC	2.50
		R	GTT GGA CAA CAA GAG TAA GAC AGA	2.50
	DYS626	F	PET-CTG GGT GAC AGA GTG CAA GA	5.00
		R	AAT GGA GAG ATC ATT ATT GAA GCA	5.00

연세대학교
YONSEI UNIVERSITY

표 3. Primers' information used for the multiplex amplification of in-house Kplex-Y18 system

Dye	STR loci		Primer Sequence (5'→3')	Conc. (μM)
FAM	DYS389I/II	F	FAM-CCA ACT CTC ATC TGT ATT ATC T	5.00
		R	GAT AGA TTG ATA GAG GGA GGG A	5.00
	DYS635	F	TGG CTT CTC ACT TTG CAT AGA A	0.95
		R	FAM-ACC AGC CCA AAT ATC CAT CA	0.95
	DYS458	F	FAM-GCA ACA GGA ATG AAA CTC CAA	2.50
		R	GAA GTA GCT GGG GCT AGA GGT T	2.50
VIC	DYS392	F	GTA AAC CTA CCA ATC CCA TTC CT	4.75
		R	VIC-AAA AGC CAA GAA GGA AAA CAA A	4.75
	DYS390	F	GCA ATG TGT ATA CTC AGA AAC AAG G	2.00
		R	VIC-CTG CAT TTT GGT ACC CCA TA	2.00
	DYS448	F	GCA GAA AGG GAG ATA GAG ACA TGG	1.25
		R	VIC-CCT CAT ATT TCT GGC CGG TCT	1.25

연세대학교
YONSEI UNIVERSITY

Continued)

Dye	STR loci		Primer Sequence (5'→3')	Conc. (μM)
VIC	DYS391	F	VIC-CAA GAC ACC CCA CCA CAG AT	3.00
		R	GGC AAG CAA TTG CCA TAG AGG	3.00
NED	DYS438	F	NED-TGG GGA ATA GTT GAA CGG TAA	1.45
		R	GCA ACA AGA GTG AAA CTC CAT T	1.45
	DYS19	F	NED-GGA GTC CAT CTG GGT TAA GGA	1.65
		R	CTA CTG AGT TTC TGT TAT AGT GTT TTT T	1.65
	DYS393	F	TGT CTT TAC TAG CAG CAT GAG AAC	5.00
		R	NED-AAA CTC AAG TCC AAA AAA TGA GG	5.00
	DYS456	F	TAC CCA GCC TAC ATC TTT CTC CA	4.00
		R	NED-CCA TCA ACT CAG CCC AAA AC	4.00
	DYF404S1	F	NED-CCT GGT AAT CTT CCA AAA TCG T	2.75
		R	GAA AGA TCA AAG GAG CCC AG	2.75

연세대학교
YONSEI UNIVERSITY

Continued)

Dye	STR loci		Primer Sequence (5'→3')	Conc. (μM)
PET	DYS439	F	GGT GGA GAC AGA TAG ATG ATA AA	1.05
		R	PET-GGC TTG GAA TTC TTT TAC CCA	1.05
	DYS437	F	PET-GAC TAT GGG CGT GAG TGC AT	1.50
		R	AGA TAG ATA GAT AAC CAC AGA TAA ATA T	1.50
	DYS385ab	F	PET-GAA ATG AAA TTC AGA AAG GAA GG	5.00
		R	TTC CAA TTA CAT AGT CCT CCT TTC T	5.00
	YGATAH4	F	TAA ACA ACT TAA CAG GAT AAA TCA CC	5.00
		R	PET-CAG AGT GGG TTC TGA AGA GCT A	5.00

연세대학교
YONSEI UNIVERSITY

전체 아시아 인구 집단 시료 0.5-2.0ng의 주형 DNA를 대상으로 자체 구축한 2개의 CE 다중 중합효소연쇄반응 체계를 사용하여 길이 기반 분석을 위한 증폭산물을 생성하였다. 중합효소연쇄반응을 위한 프로토콜은 http://forensic.yonsei.ac.kr/protocols.html 홈페이지에 공개된 것과 같다. 최종 증폭산물은 Applied Biosystems 3130 genetic analyzer (Thermo Fisher Scientific)에서 분리되었고 유전자형은 Applied Biosystems GeneMapper ID Software Version 3.2 (Thermo Fisher Scientific)을 사용하여 결정되었다. CE로부터 얻어진 유전자형은 MPS로부터 얻은 유전자형과 일치하는지 비교하였다.

연세대학교
YONSEI UNIVERSITY

5. 세 아시아 인구 집단 시료의 Y 염색체 STR에 대한 MPS 수행

가. 아시아 인구 집단 시료를 대상으로 한 Y 염색체 STR 라이브러리(library) 제작

전체 아시아 인구 집단 시료 261개를 대상으로 0.5ng에서 2ng 사이의 주형 DNA를 사용하여 이전 연구에서[8] 사용한 두 단계의 중합효소연쇄반응으로 라이브러리를 생성하였다. 첫 번째 단계의 중합효소연쇄반응에서는 표 1의 고안된 Nextera primer를 이용하여 Y 염색체 표지자 특이적인 증폭과 Nextera XT index kit v2 (Illumina, San Diego, CA, USA)와 호환이 되는 read 서열을 증폭하였다. 중합효소연쇄반응의 구성요소는 표 1의 최종 농도에 해당하는 primer와 0.5에서 2ng의 주형 DNA 및 6.0U AmpliTaq Gold DNA Polymerase (Thermo Fisher Scientific) 그리고 2.0µl의 Gold ST*R 10X Buffer (Promega)가 포함되고 최종 양은 20.0µl가 되도록 하여 잘 섞어 주었다. 혼합물은 Veriti® 96-Well Thermal Cycler (Thermo Fisher Scientific)에 장착하여, 95 ℃에서 11분간 변성시키고 94℃에서 20초, 60℃에서 1분, 72℃에서 45초의 조건으로 26-27회 중합반응 후 최종적으로 72 ℃에서 5분간 반응시킨 후 4 ℃에서 잠시 보관하였다. 다음으로 두 번째 단계의 중합효소연쇄반응

연세대학교
YONSEI UNIVERSITY

에서는 증폭 수행을 위하여 혼합물에 Nextera XT index kit v2 (Illumina)를 사용하여 다중색인 (Index) 1, 2를 각 1차 증폭산물 양쪽에 붙여 시료가 각각 개별적으로 분류될 수 있도록 어댑터를 부착한다. 중합효소연쇄반응의 구성요소는 1/10으로 희석한 1.0μl의 1차 증폭산물 및 10.0μl의 Platinum™ SuperFi II PCR Master Mix (Thermo Fisher Scientific) 그리고 각 시료마다 다른 조합으로 2.0μl의 Index 1, 2가 포함되고, 최종 양은 20.0μl가 되도록 하여 잘 섞어 주었다. 혼합물은 Veriti® 96-Well Thermal Cycler (Thermo Fisher Scientific)에 장착하여, 98℃에서 30초간 변성시키고 98℃에서 10초, 61℃에서 10초, 72℃에서 20초 혹은 30초의 조건으로 16회 중합반응 후 최종적으로 72 ℃에서 5분간 반응시킨 후 4 ℃에서 잠시 보관하였다. 두 단계의 중합효소연쇄반응을 통해 생성된 증폭산물을 확인하기 위하여 Agilent DNA 1000 Kit (Agilent Technologies, Santa Clara, CA, USA)를 사용하여 2100 Bioanalyzer (Agilent Technologies)를 통해 확인하였다. 200-400bp 크기의 증폭산물 존재 유무를 확인하고, 측정된 농도를 확인하였다. 각 시료의 증폭산물을 모두 10.0ng/μl의 농도의 같은 부피로 표준화시켜 준비하였다.

나. 아시아 인구 집단 시료를 대상으로 한 Y 염색체 STR 라이브러리(library) 정제 및 정도 관리 (Quality control)

연세대학교
YONSEI UNIVERSITY

증폭 산물의 크기를 고려하여 원하는 라이브러리만을 얻고, 중합효소연쇄반응을 통해 생성된 부산물을 제거하고자 Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, IN, USA)를 이용한 정제를 수행하였다. 증폭 산물 양의 1.2배의 AMPure Bead (Beckman Coulter)를 이용하여 제조사의 지시에 따라 정제하였다. 정제된 라이브러리는 다시 Agilent DNA 1000 Kit (Agilent Technologies)와 2100 Bioanalyzer (Agilent Technologies)를 사용하여 원하는 크기의 라이브러리로 정제되었는지 확인하였다. 제작된 라이브러리를 QuantStudio™ 3 Real-Time PCR System (Thermo Fisher Scientific)에서 KAPA Quantification Kit (KAPA Biosystems, Wilmington, MA, USA)를 사용하여 정량하고, 측정 값에 따라 10nM의 농도로 동량 섞어 하나의 튜브에 준비하였다.

다. MPS 장비를 이용한 Y 염색체 STR 염기서열 분석

10nM의 농도로 동량 섞인 Y 염색체 표지자 라이브러리를 대상으로 300bp씩 양 끝을 읽는 (Paired-end read) MiSeq Reagent Kit v3 kit (Illumina)를 이용하여 제조사의 지시에 따라 분석하였다.

연세대학교
YONSEI UNIVERSITY

라. MPS 결과를 이용한 Y 염색체 STR 염기서열 변이 자료 분석

얻어진 MPS 자료는 FASTQ 파일 형태로, 이의 분석은 법과학 영역에서 주로 사용되는 STR 대립유전자형 및 염기서열 확인을 위해 개발한 STRait Razor v3.0 프로그램과[16] 자체 구축한 VisualSTR 체계를 사용하여 MPS 결과를 확인하였다. STRiat Razor v3 프로그램은 대상 Y 염색체 표지자에 대하여 생성된 configuration 파일(표 4)을 사용하여 각 STR의 유전자형을 결정하고 각 유전자형에서 분석된 리드 수를 확인할 수 있었다. 유전자형 검출을 위한 최소 리드 수는 100으로 설정하였고, 일부 낮은 리드 수를 갖는 표지자에 대해서는 70을 적용하였다. 최소 리드 수 이상을 갖는 것으로 확인된 염기 서열을 바탕으로 반복단위와 그 주변 영역 서열을 확인하여 정리하였다. 반복 단위는 STRidER에서 제공하는 염기 서열을 통해 확인하였고 주변 영역에 대한 서열은 NCBI 1000 genomes 브라우저로부터 얻어진 서열을 통해 확인하였다. 반복영역 주변 영역에서 확인된 단일염기다형성(SNP)의 주석은 NCBI SNP build 155 정보를 기초로 하여 확인하였다.

연세대학교
YONSEI UNIVERSITY

표 4. Configuration file's information used in the STRait Razor v3.0 program for sequence variation analysis of 34 Y-chromosomal markers

#Marker	Type	5'Flank	3'Flank	Motif	Period	Offset
DYS19	Y-STR	TAAGGAGAGTGT	AACAGAAACTCA	TCTATCTATCTA	4	71
DYS389 I	Y-STR	TAGAGGGAGGGA	ACATAGATAATA	TAGATAGATAGA	4	1
DYS389 II	Y-STR	TAGAGGGAGGGA	ACATAGATAATA,2	TAGATAGATAGA	4	53
DYS390	Y-STR	CAGAAACAAGGA	TATGGGGTACCA	CAGACAGACAGA	4	29
DYS391	Y-STR	CCCATATCTGTC	CTCTATGGCAAT	TCTATCTATCTA	4	33
DYS392	Y-STR	AATCCCATTCCT	CCTTCTTGGCTT	ATAATAATAATA	3	39
DYS393	Y-STR	TACTTGTGTCAA	CCTCATTTTTTG	AGATAGATAGAT	4	28
DYS385ab	Y-STR	GTCCTCCTTTCT	TTCTTTTTTACT	TTTCTTTCTTTC	4	90
DYS437	Y-STR	GCGTGAGTGCAT	TCTGTGAATGAT	TCTATCTATCTA	4	20
DYS438	Y-STR	GTTGAACGGTAA	AATGGAGTTTCA	TTTTCTTTTC	5	15
DYS439	Y-STR	ACAGATAGATGA	TGGGTAAAAGAA	AGATAGATAGAT	4	31
DYS448	Y-STR	AGAGACATGGA	ATAAATTTCCAG	AGAGATAGAGAT	6	63

연세대학교
YONSEI UNIVERSITY

Continued)

#Marker	Type	5'Flank	3'Flank	Motif	Period	Offset
DYS456	Y-STR	GACCTTGTGATA	TTTAAGAAGTTT	AGATAGATAGAT	4	47
DYS458	Y-STR	ATGAAACTCCAA	TTGTAATGCCAG	GAAAGAAAGAAA	4	44
DYS635	Y-STR	AGCCCAAATATC	TTCTATGCAAAG	GATAGATAGATA	4	34
GATAH4	Y-STR	CTAATCTATCCA	TTGGAAATTCTC	TCTATCTATCTA	4	45
DYS570	Y-STR	CTCCAAGTTCCT	TCTGCATTTCAG	TTTCTTTCTTTC	4	19
DYS576	Y-STR	CTCATCTCTGAA	AGCCAAGACAAA	AAAGAAAGAAAG	4	6
DYS481	Y-STR	AAGGAATGTGGC	GTCTTGCAACCT	CTTCTTCTTCTT	3	26
DYS533	Y-STR	CATCTTTCTAGC	TTGTTTGGTTGA	ATCTATCTATCT	4	26
DYS549	Y-STR	TAAAGAGGAAGA	TCACAAATGGAA	GATAGATAGATA	4	42
DYS643	Y-STR	AAACTACTGTGC	TCAGTAGAATTT	CTTTTCTTTT	5	23
Y-M175	Y-STR	TCTGAATCAGGC	AAAGGTATCAGT	TTCTC	5	30
DYF387S1	Y-STR	TAACAAGAAAAA	AAACTGTGGTAG	GGAAGGAAGGAA	4	32
DYF399S1	Y-STR	CAACAGGAAAAA	TCTGGGTTCTTT	GAAAGAAAGAAA	4	37

연세대학교
YONSEI UNIVERSITY

Continued)

#Marker	Type	5'Flank	3'Flank	Motif	Period	Offset
DYF404S1	Y-STR	TTTCAACGCATA	TTGACAGTGTCT	TTTCTTTCTTTC	4	57
DYS449	Y-STR	TTCTTTTCTCTC	TTTGTTTCTGTC	TTTCTTTCTTTC	4	56
DYS460	Y-STR	TGCCTATCATTT	GGTGTCAGATTC	TCTATCTATCTA	4	26
DYS518	Y-STR	AAGTGAAACTG	GAGGGAAGAAGA	AAAGAAAGAAAG	4	23
DYS526a	Y-STR	CTTACTTTCTAT	TCTGGCGAAGTA	CCTTCCTTCCTT	4	40
DYS612	Y-STR	TTCAGAGGTTTG	TCAGTTCCCTTT	TCTTCTTCTTCT	3	34
DYS626	Y-STR	CCATAGCAAAAG	GGAAAGAAAGAAAGGG	AAAGAAAGAAAG	4	50
DYS627	Y-STR	ACACAGGGAGAA	ACGAAGGAGAGA	AAAGAAAGAAAG	4	119

연세대학교
YONSEI UNIVERSITY

III. 결 과

1. 아시아 집단 시료의 Y 염색체 STR의 MPS 분석 결과

261개 시료를 대상으로 생성된 라이브러리는 MiSeq system (Illumina)에서 여러 배치로 나누어 시퀀싱 되었다. 평균 샘플 리드 수는 201,043 이었으며 가장 많은 리드 수는 MY028 샘플에서 360,164, 가장 적은 리드 수는 PJ002 샘플에서 104,262로 관찰되었다. 34개의 Y 염색체 표지자 전체에 걸쳐 평균 커버리지 깊이 (Depth of Coverage; DoC)는 5,604 였으며, 가장 높은 DoC는 DYF404S1에서 9,488, 가장 낮은 DoC는 DYS626에서 827로 관찰되었다. (그림 3)

연세대학교
YONSEI UNIVERSITY

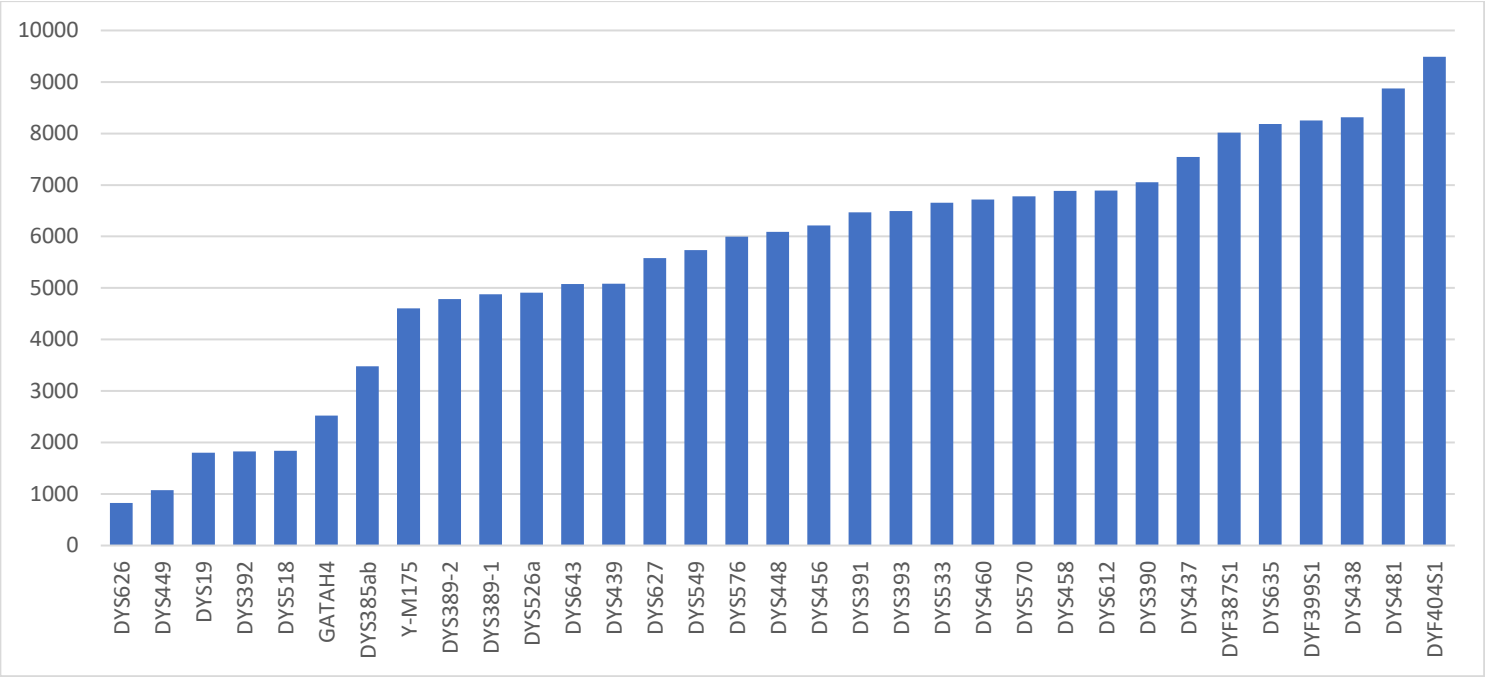

그림 3. Average depth of coverage (DoC) for the 34 Y-chromosomal markers used in this study for all 261 male DNA samples

연세대학교
YONSEI UNIVERSITY

2. CE와 MPS 방법을 통해 얻은 Y 염색체 STR 유전자형 비교 결과

CE와 MPS 방법 사이에 유전자형을 비교하기 위하여 전체 시료에 대한 33개의 Y 염색체 STR과 1개의 Y 단일염기다형성(SNP)의 유전자형을 자체 구축한 2개의 CE 기반 패널을 사용하여 얻었다. 얻어진 유전자형은 http://forensic.yonsei.ac.kr/presentations.html 홈페이지에 공개하였다. 확인된 하플로타입은 261개의 시료에 대하여 모두 고유했다. 한 샘플(PJ031)에서 Y 염색체 미세결손으로 인한 6개의 Y 염색체 STR (DYS449, DYS458, DYS481, DYS570, DYS576, 그리고 DYS627)에서 CE와 MPS에서 모두 유전자형이 검출되지 않았다 (allele 0). CE와 MPS로부터 얻어진 유전자형 사이의 일치율은 99.97% (8,871/8,874)로 확인되었다. 세 개의 샘플(MY028, MY041 및 PJ023)에서 유전자형 불일치가 관찰되었다. 모두 DYS576 표지자에서 불일치를 보였는데, CE에서는 각 샘플이 19, 13, 그리고 18의 유전자형을 보였으나 MPS에서는 모두 유전자형이 검출되지 않았다.

연세대학교
YONSEI UNIVERSITY

3. 아시아 집단 시료의 Y 염색체 STR의 반복 지역 및 주변 영역 염기서열 변이 분석 결과 및 대립 유전자 수의 증가

세 개의 아시아 인구집단(네팔, 미얀마, 그리고 파키스탄)에 대하여 Y 염색체 STR의 모든 염기서열 구조와 그 빈도를 산출하였다(표 5).

연세대학교
YONSEI UNIVERSITY

표 5. Sequence variations and frequencies for 34 Y-chromosomal markers

Locus	CE allele	MPS allele*	Repeat Structure*	Flanking Region Variation	Frequency		
					NP	MY	PJ
DYF387S1	32		[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]9 [AAAG]10		-	0.005	-
DYF387S1	33		[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]8 [AAAG]12		-	0.005	-
DYF387S1	34	a	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]8 [AAAG]13		0.008	0.011	-
DYF387S1	34	b	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]9 [AAAG]12		0.016	0.005	-
DYF387S1	35	a	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]10 [AAAG]12		-	0.005	-
DYF387S1	35	b	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]7 [AAAG]15		-	0.005	-
DYF387S1	35	c	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]8 [AAAG]14		0.063	0.100	0.026
DYF387S1	35	d	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]9 [AAAG]13		0.016	0.011	0.005
DYF387S1	36	a	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]10 [AAAG]13		-	0.016	0.005
DYF387S1	36	b	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]11 [AAAG]12		0.008	0.026	0.005
DYF387S1	36	c	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]7 [AAAG]16		-	0.011	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat Structure	Flanking Region Variation	Frequency		
					NP	MY	PJ
DYF387S1	36	d	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]8 [AAAG]15		0.040	0.074	0.032
DYF387S1	36	e	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]9 [AAAG]14		0.111	0.079	0.126
DYF387S1	37	a	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]10 [AAAG]14		0.040	0.026	0.037
DYF387S1	37	b	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]11 [AAAG]13		0.032	0.063	0.047
DYF387S1	37	c	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]7 [AAAG]17		-	0.005	-
DYF387S1	37	d	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]8 [AAAG]16		0.024	0.011	0.016
DYF387S1	37	e	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]9 [AAAG]15		0.119	0.095	0.179
DYF387S1	38	a	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]10 [AAAG]15		0.024	0.042	0.063
DYF387S1	38	b	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]11 [AAAG]14		0.071	0.074	0.084
DYF387S1	38	c	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]12 [AAAG]13		0.016	-	0.011
DYF387S1	38	d	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]8 [AAAG]17		-	0.011	0.005

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat Structure	Flanking Region Variation	Frequency		
					NP	MY	PJ
DYF387S1	38	e	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]9 [AAAG]16		0.056	0.063	0.058
DYF387S1	38	f	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG AAAG [GAAG]10 [AAAG]16		-	0.005	-
DYF387S1	39	a	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]10 [AAAG]16		0.048	0.042	0.063
DYF387S1	39	b	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]11 [AAAG]15		0.159	0.132	0.142
DYF387S1	39	c	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]12 [AAAG]14		0.008	-	-
DYF387S1	39	d	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]9 [AAAG]17		0.016	0.011	0.011
DYF387S1	39	e	[AAAG]3 GTAG [GAAG]4 AAAG [GAAG]2 [AAAG]2 [GAAG]11 [AAAG]15		-	-	0.005
DYF387S1	40	a	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]10 [AAAG]17		0.016	0.026	0.005
DYF387S1	40	b	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]11 [AAAG]16		0.056	0.026	0.037
DYF387S1	40	c	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]12 [AAAG]15		0.016	0.011	0.005
DYF387S1	41	a	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]10 [AAAG]18		-	-	0.011

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat Structure	Flanking Region Variation	Frequency		
					NP	MY	PJ
DYF387S1	41	b	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]11 [AAAG]17		0.016	-	0.005
DYF387S1	41	c	[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]12 [AAAG]16		0.016	0.005	0.016
DYF387S1	42		[AAAG]3 GTAG [GAAG]4 [AAAG]2 GAAG [AAAG]2 [GAAG]11 [AAAG]18		0.008	-	-
DYF387S1					1.000	1.000	1.000
DYF399S1*	16		[GAAA]3 N7 [GAAA]11	rs4306075 (A>G)	-	0.003	-
DYF399S1	17		[GAAA]3 N7 [GAAA]12	rs4306075 (A>G)	0.005	-	0.007
DYF399S1	17.1		[GAAA]3 N8 [GAAA]12		-	-	0.003
DYF399S1	17.2		[GAAA]3 N7 [GAAA]13	rs2090175785 (delGA)	-	0.003	0.003
DYF399S1	18		[GAAA]3 N7 [GAAA]13	rs4306075 (A>G)	-	0.003	0.003
DYF399S1	18.1		[GAAA]3 N8 [GAAA]13		0.031	-	0.007
DYF399S1	18.2		[GAAA]3 N7 [GAAA]14	rs2090175785 (delGA)	-	-	0.010
DYF399S1	18.3		[GAAA]3 N7 [GAAA]14	rs4306075 (A>G)	-	0.003	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat Structure	Flanking Region Variation	Frequency		
					NP	MY	PJ
DYF399S1	19	a	[GAAA]3 N7 [GAAA]13 GAAC	rs4306075 (A>G)	-	-	0.003
DYF399S1	19	b	[GAAA]3 N7 [GAAA]14	rs4306075 (A>G)	0.005	0.024	0.014
DYF399S1	19	c	[GAAA]3 N7 [GAAA]14	rs878949651 (A>G)	-	0.003	0.003
DYF399S1	19.1		[GAAA]3 N8 [GAAA]14		0.031	0.003	0.034
DYF399S1	20	a	[GAAA]3 N7 [GAAA]14 GAAC	rs4306075 (A>G)	-	-	0.007
DYF399S1	20	b	[GAAA]3 N7 [GAAA]15	rs4306075 (A>G)	0.063	0.072	0.071
DYF399S1	20	c	[GAAA]3 N7 [GAAA]15		-	0.003	-
DYF399S1	20	d	[GAAA]3 N7 [GAAA]15	rs878949651 (A>G)	-	0.003	0.003
DYF399S1	20.1	a	[GAAA]3 N8 [GAAA]15		0.016	0.003	0.024
DYF399S1	20.1	b	[GAAA]3 N8 [GAAA]15	rs4306075 (A>G)	0.005	-	-
DYF399S1	21	a	[GAAA]3 N7 [GAAA]15 GAAC	rs4306075 (A>G)	-	-	0.003
DYF399S1	21	b	[GAAA]3 N7 [GAAA]16	rs878949651 (A>G)	0.005	0.010	0.020

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency		
					NP	MY	PJ
DYF399S1	21	c	[GAAA]3 N7 [GAAA]16	rs4306075 (A>G)	0.099	0.110	0.078
DYF399S1	21	d	[GAAA]3 N7 [GAAA]16		-	-	0.007
DYF399S1	21	e	[GAAA]3 N7 GGAA [GAAA]15		-	-	0.003
DYF399S1	21.1	a	[GAAA]3 N8 [GAAA]16		0.042	0.024	0.024
DYF399S1	21.1	b	[GAAA]3 N8 [GAAA]16	rs4306075 (A>G)	-	0.010	-
DYF399S1	22	a	[GAAA]3 N7 [GAAA]16 GAAC	rs4306075 (A>G)	-	-	0.010
DYF399S1	22	b	[GAAA]3 N7 [GAAA]17	rs878949651 (A>G)	0.037	0.024	0.048
DYF399S1	22	c	[GAAA]3 N7 [GAAA]17	rs4306075 (A>G)	0.052	0.062	0.031
DYF399S1	22	d	[GAAA]3 N7 [GAAA]17		-	0.003	0.010
DYF399S1	22	e	[GAAA]3 N7 GGAA [GAAA]16		-	0.007	0.003
DYF399S1	22.1		[GAAA]3 N8 [GAAA]17		0.079	0.031	0.020
DYF399S1	22.2		[GAAA]3 N7 [GAAA]16 AA GAAA	rs878949651 (A>G)	-	0.003	0.014

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency		
					NP	MY	PJ
DYF399S1	23	a	[GAAA]3 N7 [GAAA]17 GAAC	rs4306075 (A>G)	0.005	0.003	0.003
DYF399S1	23	b	[GAAA]3 N7 [GAAA]18		0.005	0.003	0.020
DYF399S1	23	c	[GAAA]3 N7 [GAAA]18	rs878949651 (A>G)	0.058	0.066	0.071
DYF399S1	23	d	[GAAA]3 N7 [GAAA]18	rs4306075 (A>G)	0.073	0.024	0.027
DYF399S1	23	e	[GAAA]3 N7 [GAAA]3 GAAG [GAAA]14	rs4306075 (A>G)	-	0.003	-
DYF399S1	23	f	[GAAA]3 N7 GGAA [GAAA]17		-	-	0.010
DYF399S1	23.1	a	[GAAA]3 N8 [GAAA]18		0.037	0.041	0.020
DYF399S1	23.1	b	[GAAA]3 N8 [GAAA]18	rs4306075 (A>G)	0.005	-	-
DYF399S1	23.2		[GAAA]3 N7 [GAAA]17 AA GAAA	rs878949651 (A>G)	-	0.007	0.014
DYF399S1	24	a	[GAAA]3 N7 [GAAA]18 GAAC	rs4306075 (A>G)	-	0.003	0.007
DYF399S1	24	b	[GAAA]3 N7 [GAAA]19	rs878949651 (A>G)	0.073	0.072	0.085
DYF399S1	24	c	[GAAA]3 N7 [GAAA]19	rs4306075 (A>G)	0.005	0.007	0.010

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYF399S1	24	d	[GAAA]3 N7 [GAAA]19		-	0.003	0.003
DYF399S1	24	e	[GAAA]3 N7 GGAA [GAAA]18		0.010	-	0.003
DYF399S1	24.1		[GAAA]3 N8 [GAAA]19		0.042	0.107	0.044
DYF399S1	24.2		[GAAA]3 N7 [GAAA]18 AA GAAA	rs878949651 (A>G)	0.005	0.007	0.003
DYF399S1	25	a	[GAAA]3 N7 [GAAA]20		0.005	0.003	0.010
DYF399S1	25	b	[GAAA]3 N7 [GAAA]20	rs878949651 (A>G)	0.058	0.048	0.048
DYF399S1	25	c	[GAAA]3 N7 GGAA [GAAA]19		-	0.003	-
DYF399S1	25.1		[GAAA]3 N8 [GAAA]20		0.047	0.052	0.068
DYF399S1	25.2		[GAAA]3 N7 [GAAA]19 AA GAAA	rs878949651 (A>G)	0.005	-	0.003
DYF399S1	26	a	[GAAA]3 N7 [GAAA]21	rs878949651 (A>G)	0.042	0.041	0.024
DYF399S1	26	b	[GAAA]3 N7 [GAAA]21		-	-	0.003
DYF399S1	26	c	[GAAA]3 N7 GGAA [GAAA]19		-	-	0.003

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYF399S1	26.1	a	[GAAA]3 N8 [GAAA]21		0.021	0.038	0.031
DYF399S1	26.1	b	[GAAA]3 N8 [GAAA]21	rs878949651 (A>G)	-	0.003	-
DYF399S1	26.2		[GAAA]3 N7 [GAAA]20 AA GAAA	rs878949651 (A>G)	-	-	0.003
DYF399S1	27		[GAAA]3 N7 [GAAA]22	rs878949651 (A>G)	0.021	0.014	-
DYF399S1	27.1		[GAAA]3 N8 [GAAA]22		0.005	0.017	0.010
DYF399S1	28	a	[GAAA]3 N7 [GAAA]23		0.005	-	-
DYF399S1	28	b	[GAAA]3 N7 [GAAA]23	rs878949651 (A>G)	-	0.003	-
DYF399S1	28.1		[GAAA]3 N8 [GAAA]23		-	0.014	0.007
DYF399S1					1.000	1.000	1.000
DYF404S1	11		[TTTC]11		0.010	-	-
DYF404S1	12		[TTTC]12		0.020	0.072	0.035
DYF404S1	13		[TTTC]13		0.158	0.355	0.112

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYF404S1	13.2		[TTTC]11 TT [TTTC]2		-	-	0.012
DYF404S1	14		[TTTC]14		0.248	0.296	0.329
DYF404S1	14.2		[TTTC]12 TT [TTTC]2		-	0.007	-
DYF404S1	15	a	[TTTC]16	rs34228393 (delTTCT)	-	-	0.006
DYF404S1	15	b	[TTTC]15		0.317	0.158	0.294
DYF404S1	15.2		[TTTC]13 TT [TTTC]2		-	-	0.006
DYF404S1	16	a	[TTTC]16		0.168	0.086	0.171
DYF404S1	16	b	[TTTC]5 TTTT [TTTC]10		0.010	-	-
DYF404S1	16.2		[TTTC]14 TT [TTTC]2		-	-	0.006
DYF404S1	17	a	[TTTC]17		0.069	0.020	0.018
DYF404S1	17	b	[TTTC]5 TTTT [TTTC]11		-	-	0.012
DYF404S1	18		[TTTC]18		-	0.007	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYF404S1					1.000	1.000	1.000
DYS19	13		[TCTA]10 ccta [TCTA]3		0.062	0.020	0.020
DYS19	14		[TCTA]11 ccta [TCTA]3		0.385	0.378	0.255
DYS19	15		[TCTA]12 ccta [TCTA]3		0.292	0.408	0.439
DYS19	16		[TCTA]13 ccta [TCTA]3		0.169	0.143	0.204
DYS19	17		[TCTA]14 ccta [TCTA]3		0.077	0.051	0.071
DYS19	18		[TCTA]15 ccta [TCTA]3		0.015	-	0.010
DYS19					1.000	1.000	1.000
DYS385ab	9		[TTTC]9		-	-	0.021
DYS385ab	10	a	[TTTC]10		-	-	0.005
DYS385ab	10	b	[TTTC]11	rs1248860842 (delCCTT)	0.008	-	-
DYS385ab	11		[TTTC]11		0.205	0.076	0.178

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS385ab	12		[TTTC]12		0.110	0.043	0.037
DYS385ab	13	a	[TTTC]13		0.126	0.211	0.110
DYS385ab	13	b	[TTTC]14	rs1248860842 (delCCTT)	-	0.005	-
DYS385ab	14		[TTTC]14		0.236	0.108	0.215
DYS385ab	14.2		[TTTC]5 TC [TTTC]9		-	0.011	-
DYS385ab	15		[TTTC]15		0.039	0.086	0.126
DYS385ab	16		[TTTC]16		0.047	0.059	0.089
DYS385ab	17		[TTTC]17		0.047	0.059	0.079
DYS385ab	18		[TTTC]18		0.071	0.103	0.073
DYS385ab	19		[TTTC]19		0.063	0.130	0.037
DYS385ab	20		[TTTC]20		0.031	0.081	0.016
DYS385ab	21		[TTTC]21		0.016	0.022	0.016

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS385ab	22		[TTTC]22		-	0.005	-
DYS385ab					1.000	1.000	1.000
DYS389 I	11		[TAGA]8 [CAGA]3		-	0.010	-
DYS389 I	12		[TAGA]9 [CAGA]3		0.262	0.439	0.214
DYS389 I	13		[TAGA]10 [CAGA]3		0.446	0.408	0.510
DYS389 I	14		[TAGA]11 [CAGA]3		0.277	0.133	0.265
DYS389 I	15		[TAGA]12 [CAGA]3		0.015	0.010	0.010
DYS389 I					1.000	1.000	1.000
DYS389 II*	26		[TAGA]9 [CAGA]3 N48 [TAGA]10 [CAGA]4		-	0.010	-
DYS389 II	27	a	[TAGA]8 [CAGA]3 N48 [TAGA]12 [CAGA]4		-	0.010	-
DYS389 II	27	b	[TAGA]9 [CAGA]3 N48 [TAGA]10 [CAGA]5		0.015	0.020	0.010
DYS389 II	27	c	[TAGA]9 [CAGA]3 N48 [TAGA]11 [CAGA]4		0.046	0.061	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS389Ⅱ	28	a	[TAGA]10 [CAGA]3 N48 [TAGA]10 [CAGA]5		-	0.010	0.010
DYS389Ⅱ	28	b	[TAGA]10 [CAGA]3 N48 [TAGA]11 [CAGA]4		-	0.020	0.010
DYS389Ⅱ	28	c	[TAGA]9 [CAGA]3 N48 [TAGA]11 [CAGA]5		0.046	0.051	0.102
DYS389Ⅱ	28	d	[TAGA]9 [CAGA]3 N48 [TAGA]12 [CAGA]4		0.092	0.204	0.010
DYS389Ⅱ	29	a	[TAGA]10 [CAGA]3 N48 [TAGA]11 [CAGA]5		0.031	0.082	0.092
DYS389Ⅱ	29	b	[TAGA]10 [CAGA]3 N48 [TAGA]12 [CAGA]4		0.046	0.153	0.041
DYS389Ⅱ	29	c	[TAGA]11 [CAGA]3 N48 [TAGA]11 [CAGA]4		0.015	0.010	-
DYS389Ⅱ	29	d	[TAGA]9 [CAGA]3 N48 [TAGA]12 [CAGA]5		0.046	0.010	0.041
DYS389Ⅱ	29	e	[TAGA]9 [CAGA]3 N48 [TAGA]13 [CAGA]4		0.015	0.051	-
DYS389Ⅱ	30	a	[TAGA]10 [CAGA]3 N48 [TAGA]11 [CAGA]6		-	-	0.031
DYS389Ⅱ	30	b	[TAGA]10 [CAGA]3 N48 [TAGA]12 [CAGA]5		0.154	0.041	0.184
DYS389Ⅱ	30	c	[TAGA]10 [CAGA]3 N48 [TAGA]13 [CAGA]4		0.031	0.071	0.020

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS389Ⅱ	30	d	[TAGA]11 [CAGA]3 N48 [TAGA]11 [CAGA]5		0.062	0.061	0.102
DYS389Ⅱ	30	e	[TAGA]11 [CAGA]3 N48 [TAGA]12 [CAGA]4		0.015	-	0.041
DYS389Ⅱ	30	f	[TAGA]12 [CAGA]3 N48 [TAGA]11 [CAGA]4		0.015	-	-
DYS389Ⅱ	30	g	[TAGA]9 [CAGA]3 N48 [TAGA]13 [CAGA]5		-	0.010	0.010
DYS389Ⅱ	30	h	[TAGA]9 [CAGA]3 N48 [TAGA]14 [CAGA]4		-	0.010	-
DYS389Ⅱ	31	a	[TAGA]10 [CAGA]3 N48 [TAGA]13 [CAGA]5		0.169	0.020	0.112
DYS389Ⅱ	31	b	[TAGA]10 [CAGA]3 N48 [TAGA]14 [CAGA]4		-	0.010	-
DYS389Ⅱ	31	c	[TAGA]11 [CAGA]3 N48 [TAGA]11 [CAGA]6		0.015	-	0.010
DYS389Ⅱ	31	d	[TAGA]11 [CAGA]3 N48 [TAGA]12 [CAGA]5		0.062	0.020	0.092
DYS389Ⅱ	31	e	[TAGA]11 [CAGA]3 N48 [TAGA]13 [CAGA]4		0.015	0.031	-
DYS389Ⅱ	31	f	[TAGA]12 [CAGA]3 N48 [TAGA]11 [CAGA]5		-	0.010	0.010
DYS389Ⅱ	31	g	[TAGA]9 [CAGA]3 N48 [TAGA]14 [CAGA]5		-	-	0.041

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS389Ⅱ	31	h	[TAGA]9 [CAGA]3 N48 [TAGA]15 [CAGA]4		-	0.010	-
DYS389Ⅱ	32	a	[TAGA]10 [CAGA]3 N48 [TAGA]14 [CAGA]5		0.015	-	0.010
DYS389Ⅱ	32	b	[TAGA]11 [CAGA]3 N48 [TAGA]13 [CAGA]5		0.077	0.010	-
DYS389Ⅱ	32	c	[TAGA]11 [CAGA]3 N48 [TAGA]14 [CAGA]4		-	-	0.010
DYS389Ⅱ	33		[TAGA]11 [CAGA]3 N48 [TAGA]14 [CAGA]5		0.015	-	0.010
DYS389Ⅱ					1.000	1.000	1.000
DYS390	21	a	[TAGA]4 CAGA [TAGA]8 [CAGA]8		-	-	0.031
DYS390	21	b	[TAGA]4 CAGA [TAGA]9 [CAGA]7		-	0.010	0.010
DYS390	21	c	[TAGA]4 CAGA [TAGA]11 [CAGA]5		-	-	0.010
DYS390	21	d	[TAGA]4 CAGA [TAGA]8 [CAGA]8		0.015	-	-
DYS390	22	a	[TAGA]4 CAGA [TAGA]9 [CAGA]8		0.092	0.061	0.194
DYS390	22	b	[TAGA]4 CAGA [TAGA]10 [CAGA]7		0.015	-	0.010

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS390	23	a	[TAGA]3 CAGA [TAGA]10 [CAGA]9		-	0.010	-
DYS390	23	b	[TAGA]4 CAGA [TAGA]9 [CAGA]9		-	0.020	0.010
DYS390	23	c	[TAGA]4 CAGA [TAGA]10 [CAGA]8		0.200	0.265	0.296
DYS390	23	d	[TAGA]4 CAGA [TAGA]11 [CAGA]7		0.015	-	-
DYS390	24	a	[TAGA]4 CAGA [TAGA]10 [CAGA]9		-	0.071	-
DYS390	24	b	[TAGA]4 CAGA [TAGA]11 [CAGA]8		0.308	0.286	0.163
DYS390	25	a	[TAGA]4 CAGA [TAGA]11 [CAGA]9		-	0.102	-
DYS390	25	b	[TAGA]4 CAGA [TAGA]12 [CAGA]8		0.338	0.122	0.245
DYS390	26	a	[TAGA]4 CAGA [TAGA]12 [CAGA]9		-	0.031	-
DYS390	26	b	[TAGA]4 CAGA [TAGA]13 [CAGA]8		0.015	0.010	0.031
DYS390	26	c	[TAGA]4 CATA [TAGA]12 [CAGA]9		-	0.010	-
DYS390					1.000	1.000	1.000

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS391	9		[TCTA]9		0.015	0.010	0.031
DYS391	10		[TCTA]10		0.615	0.714	0.643
DYS391	11		[TCTA]11		0.354	0.265	0.316
DYS391	12		[TCTA]12		0.015	0.010	0.010
DYS391					1.000	1.000	1.000
DYS392	10		[ATA]10		0.123	0.051	0.122
DYS392	11		[ATA]11		0.631	0.265	0.724
DYS392	12		[ATA]12		-	0.071	0.031
DYS392	13		[ATA]13		0.015	0.214	-
DYS392	14		[ATA]14		0.169	0.367	0.092
DYS392	15		[ATA]15		0.046	0.010	0.031
DYS392	15.2		[ATA]15 AT		-	0.010	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS392	16		[ATA]16		0.015	0.010	-
DYS392					1.000	1.000	1.000
DYS393	11		[AGAT]11		-	0.020	0.040
DYS393	12	a	[AGAT]12		0.277	0.459	0.273
DYS393	12	b	CGAT [AGAT]11		-	0.020	0.020
DYS393	13	a	[AGAT]13		0.231	0.194	0.192
DYS393	13	b	CGAT [AGAT]12		0.338	0.061	0.333
DYS393	14	a	[AGAT]14		0.062	0.235	0.091
DYS393	14	b	CGAT [AGAT]13		0.092	-	0.020
DYS393	15		[AGAT]15		-	0.010	0.030
DYS393					1.000	1.000	1.000
DYS437	14	a	[TCTA]9 TCTG [TCTA]4		-	0.051	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS437	14	b	[TCTA]8 [TCTG]2 [TCTA]4		0.662	0.480	0.633
DYS437	14	c	[TCTA]4 CCTA [TCTA]3 [TCTG]2 [TCTA]4		-	0.010	-
DYS437	14	d	[TCTA]8 [TCTG]2 [TCTA]4	rs9786886 (C>T)	-	-	0.010
DYS437	15	a	[TCTA]10 TCTG [TCTA]4		0.123	0.286	-
DYS437	15	b	[TCTA]9 [TCTG]2 [TCTA]4		0.062	0.153	0.163
DYS437	15	d	TCTG [TCTA]8 [TCTG]2 [TCTA]4		-	-	0.010
DYS437	16	a	[TCTA]11 TCTG [TCTA]4		0.015	0.010	-
DYS437	16	b	[TCTA]10 [TCTG]2 [TCTA]4		0.138	0.010	0.184
DYS437					1.000	1.000	1.000
DYS438	7		[TTTTC]7		-	-	0.010
DYS438	8		[TTTTC]8		-	0.010	0.020
DYS438	9		[TTTTC]9		0.123	0.122	0.184

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS438	10	a	[TTTTC]10		0.108	0.367	0.255
DYS438	10	b	TTTTC TTTTA [TTTTC]8		-	0.051	-
DYS438	11		[TTTTC]11		0.754	0.429	0.520
DYS438	12		[TTTTC]12		0.015	0.020	0.010
DYS438					1.000	1.000	1.000
DYS439	9		[GATA]9		0.015	-	-
DYS439	10		[GATA]10		0.369	0.102	0.418
DYS439	11		[GATA]11		0.308	0.276	0.306
DYS439	12		[GATA]12		0.231	0.388	0.235
DYS439	13		[GATA]13		0.046	0.194	0.041
DYS439	14		[GATA]14		0.031	0.041	-
DYS439					1.000	1.000	1.000

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS448*	16		[AGAGAT]8 N42 [AGAGAT]8		-	0.010	-
DYS448	17		[AGAGAT]10 N42 [AGAGAT]7		0.015	0.020	-
DYS448	18	a	[AGAGAT]11 N42 [AGAGAT]7		0.015	0.020	0.041
DYS448	18	b	[AGAGAT]10 N42 [AGAGAT]8		0.015	0.214	0.041
DYS448	18	c	[AGAGAT]9 N42 [AGAGAT]9		-	0.010	-
DYS448	19	a	[AGAGAT]13 N42 [AGAGAT]6		0.015	-	-
DYS448	19	b	[AGAGAT]12 N42 [AGAGAT]7		0.123	0.061	0.133
DYS448	19	c	[AGAGAT]11 N42 [AGAGAT]8		0.231	0.163	0.204
DYS448	19	d	[AGAGAT]10 N42 [AGAGAT]9		-	0.051	-
DYS448	20	a	[AGAGAT]14 N42 [AGAGAT]6		0.015	-	-
DYS448	20	b	[AGAGAT]13 N42 [AGAGAT]7		0.015	0.020	0.010
DYS448	20	c	[AGAGAT]12 N42 [AGAGAT]8		0.338	0.102	0.449

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS448	20	d	[AGAGAT]11 N42 [AGAGAT]9		0.169	0.214	0.010
DYS448	21	a	[AGAGAT]13 N42 [AGAGAT]8		0.015	0.010	0.082
DYS448	21	b	[AGAGAT]12 N42 [AGAGAT]9		0.031	0.092	0.031
DYS448	21	c	[AGAGAT]11 N42 [AGAGAT]10		-	0.010	-
DYS448					1.000	1.000	1.000
DYS449*	0		-		-	-	0.010
DYS449	25		[TTTC]12 N50 [TTTC]13		0.031	-	-
DYS449	26	a	[TTTC]15 N50 [TTTC]11		-	0.010	-
DYS449	26	b	[TTTC]13 N50 [TTTC]13		0.015	-	-
DYS449	26	c	[TTTC]12 N50 [TTTC]14		0.015	0.010	-
DYS449	26	d	[TTTC]11 N50 [TTTC]15		0.015	-	0.010
DYS449	27	a	[TTTC]16 N50 [TTTC]11		-	0.020	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS449	27	b	[TTTC]15 N50 [TTTC]12		0.015	-	0.010
DYS449	27	c	[TTTC]14 N50 [TTTC]13		-	-	0.010
DYS449	27	d	[TTTC]13 N50 [TTTC]14		-	-	0.010
DYS449	27	e	[TTTC]12 N50 [TTTC]15		-	-	0.020
DYS449	27	f	[TTTC]11 N50 [TTTC]16		-	0.010	0.031
DYS449	28	a	[TTTC]16 N50 [TTTC]12		-	0.010	0.020
DYS449	28	b	[TTTC]15 N50 [TTTC]13		0.015	0.082	-
DYS449	28	c	[TTTC]14 N50 [TTTC]14		-	-	0.020
DYS449	28	d	[TTTC]13 N50 [TTTC]15		-	0.010	0.010
DYS449	28	e	[TTTC]9 N50 [TTTC]19		-	-	0.010
DYS449	29	a	[TTTC]17 N50 [TTTC]12		0.015	-	0.010
DYS449	29	b	[TTTC]16 N50 [TTTC]13		0.015	0.082	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS449	29	c	[TTTC]15 N50 [TTTC]14		0.015	0.051	0.031
DYS449	29	d	[TTTC]14 N50 [TTTC]15		0.031	0.031	0.010
DYS449	29	e	[TTTC]13 N50 [TTTC]16		-	0.020	-
DYS449	29	f	[TTTC]11 N50 [TTTC]18		-	-	0.010
DYS449	30	a	[TTTC]17 N50 [TTTC]13		0.031	0.010	0.031
DYS449	30	b	[TTTC]16 N50 [TTTC]14		0.015	0.031	0.031
DYS449	30	c	[TTTC]15 N50 [TTTC]15		0.015	0.031	0.041
DYS449	30	d	[TTTC]14 N50 [TTTC]16		0.031	0.031	-
DYS449	30	e	[TTTC]13 N50 [TTTC]17		-	0.020	-
DYS449	30	f	[TTTC]12 N50 [TTTC]18		-	-	0.010
DYS449	31	a	[TTTC]18 N50 [TTTC]13		0.015	-	-
DYS449	31	b	[TTTC]17 N50 [TTTC]14		-	0.010	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS449	31	c	[TTTC]16 N50 [TTTC]15		0.062	0.031	0.082
DYS449	31	d	[TTTC]15 N50 [TTTC]16		0.046	0.102	0.051
DYS449	31	e	[TTTC]13 N50 [TTTC]18		0.015	-	-
DYS449	32	a	[TTTC]17 N50 [TTTC]15		0.092	0.041	0.041
DYS449	32	b	[TTTC]16 N50 [TTTC]16		0.231	0.112	0.163
DYS449	32	c	[TTTC]15 N50 [TTTC]17		0.046	-	0.071
DYS449	32	d	[TTTC]14 N50 [TTTC]18		-	0.010	-
DYS449	33	a	[TTTC]18 N50 [TTTC]15		0.015	-	0.010
DYS449	33	b	[TTTC]17 N50 [TTTC]16		0.062	0.122	0.051
DYS449	33	c	[TTTC]16 N50 [TTTC]17		0.062	0.041	0.122
DYS449	33	d	[TTTC]16 N50 [TTTC]5 CTTC [TTTC]11		-	-	0.010
DYS449	33	e	[TTTC]15 N50 [TTTC]18		-	-	0.020

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS449	33	f	[TTTC]14 N50 [TTTC]19		-	0.010	-
DYS449	34	a	[TTTC]18 N50 [TTTC]16		-	0.010	-
DYS449	34	b	[TTTC]17 N50 [TTTC]17		0.015	0.010	0.031
DYS449	34	c	[TTTC]16 N50 [TTTC]18		0.046	-	-
DYS449	34	d	[TTTC]15 N50 [TTTC]19		-	-	0.010
DYS449	35	a	[TTTC]18 N50 [TTTC]17		0.015	0.020	-
DYS449	35	b	[TTTC]17 N50 [TTTC]18		-	0.020	-
DYS449					1.000	1.000	1.000
DYS456	12		[AGAT]12		0.015	0.010	-
DYS456	13		[AGAT]13		0.046	0.041	0.031
DYS456	14		[AGAT]14		0.123	0.092	0.041
DYS456	15		[AGAT]15		0.646	0.582	0.622

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS456	16		[AGAT]16		0.169	0.173	0.255
DYS456	17		[AGAT]17		-	0.071	0.051
DYS456	18		[AGAT]18		-	0.031	-
DYS456					1.000	1.000	1.000
DYS458	0		-		-	-	0.010
DYS458	14		[GAAA]14		0.015	-	0.041
DYS458	15	a	[GAAA]15		0.185	0.184	0.235
DYS458	15	b	[GAAA]14 GGAA		0.015	0.031	0.020
DYS458	16	a	[GAAA]16		0.323	0.224	0.327
DYS458	16	b	[GAAA]15 GGAA		0.015	-	0.031
DYS458	17	a	[GAAA]17		0.215	0.204	0.163
DYS458	17	b	[GAAA]16 GGAA		-	-	0.020

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS458	18	a	[GAAA]18		0.154	0.184	0.082
DYS458	18	b	[GAAA]17 GGAA		-	-	0.020
DYS458	18	c	GAAG [GAAA]17		-	0.010	-
DYS458	19	a	[GAAA]19		0.046	0.102	-
DYS458	19	b	[GAAA]18		-	-	0.010
DYS458	19	c	[GAAA]17 GGAA		-	-	0.010
DYS458	20	a	[GAAA]20		-	0.041	-
DYS458	20	b	[GAAA]19		-	-	0.031
DYS458	21	a	[GAAA]21		0.015	-	-
DYS458	21	b	GAAG [GAAA]20		-	0.010	-
DYS458	22		[GAAA]22		0.015	-	-
DYS458	23		GAAG [GAAA]22		-	0.010	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS458					1.000	1.000	1.000
DYS460	9		[TCTA]9		0.077	0.306	0.010
DYS460	10		[TCTA]10		0.354	0.378	0.357
DYS460	11		[TCTA]11		0.431	0.296	0.469
DYS460	12		[TCTA]12		0.138	0.020	0.163
DYS460					1.000	1.000	1.000
DYS481	0		-		-	-	0.010
DYS481	18		[CTT]18		0.015	-	-
DYS481	19		[CTT]19		-	-	0.020
DYS481	20		[CTT]20		-	-	0.010
DYS481	21	a	CTG [CTT]20		-	0.010	-
DYS481	21	b	[CTT]21		0.015	0.020	0.041

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS481	22		[CTT]22		0.123	0.102	0.082
DYS481	23	a	[CTT]23	rs368663163 (G>A)	0.015	-	-
DYS481	23	b	[CTT]24	rs2015234517 (delGCT)	-	0.031	-
DYS481	23	c	CTG [CTT]22		-	-	0.020
DYS481	23	d	[CTT]23		0.492	0.480	0.408
DYS481	24	a	[CTT]24	rs368663163 (G>A)	-	0.010	0.031
DYS481	24	b	[CTT]25	rs2015234517 (delGCT)	-	0.010	-
DYS481	24	c	[CTT]24		0.108	0.194	0.214
DYS481	25	a	[CTT]25	rs368663163 (G>A)	0.077	-	0.071
DYS481	25	b	[CTT]25		0.031	0.082	0.061
DYS481	26	a	[CTT]26	rs368663163 (G>A)	0.015	-	0.010
DYS481	26	b	[CTT]26		0.031	0.010	0.010

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS481	27		[CTT]27		0.031	0.020	0.010
DYS481	28	a	[CTT]28	rs368663163 (G>A)	0.015	-	-
DYS481	28	b	[CTT]28		0.031	0.031	-
DYS481					1.000	1.000	1.000
DYS518	33	a	[AAAG]3 GAAG [AAAG]13 GGAG [AAAG]4 gaagag [AAAG]11		0.015	-	-
DYS518	33	b	[AAAG]3 GAAG [AAAG]12 GGAG [AAAG]4 gaagag [AAAG]12		-	0.010	-
DYS518	34	a	[AAAG]3 GAAG [AAAG]14 GGAG [AAAG]4 gaagag [AAAG]11		-	0.010	-
DYS518	34	b	[AAAG]3 GAAG [AAAG]13 GGAG [AAAG]4 gaagag [AAAG]12		-	0.010	-
DYS518	34	c	[AAAG]3 GAAG [AAAG]12 GGAG [AAAG]4 gaagag [AAAG]13		-	0.020	-
DYS518	34	d	[AAAG]3 GAAG [AAAG]11 GGAG [AAAG]4 gaagag [AAAG]14		-	0.010	-
DYS518	35	a	[AAAG]3 GAAG [AAAG]14 GGAG [AAAG]4 gaagag [AAAG]12		-	0.082	-
DYS518	35	b	[AAAG]3 GAAG [AAAG]13 GGAG [AAAG]4 gaagag [AAAG]13		0.031	0.071	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS518	35	c	[AAAG]3 GAAG [AAAG]12 GGAG [AAAG]4 gaagag [AAAG]14		-	0.031	-
DYS518	36	a	[AAAG]3 GAAG [AAAG]15 GGAG [AAAG]4 gaagag [AAAG]12		0.015	-	0.020
DYS518	36	b	[AAAG]3 GAAG [AAAG]14 GGAG [AAAG]4 gaagag [AAAG]13		0.062	0.051	0.041
DYS518	36	c	[AAAG]3 GAAG [AAAG]13 GGAG [AAAG]4 gaagag [AAAG]14		0.062	0.071	0.010
DYS518	37	a	[AAAG]3 GAAG [AAAG]17 GGAG [AAAG]4 gaagag [AAAG]11		0.015	-	-
DYS518	37	b	[AAAG]3 GAAG [AAAG]16 GGAG [AAAG]4 gaagag [AAAG]12		-	-	0.020
DYS518	37	c	[AAAG]3 GAAG [AAAG]15 GGAG [AAAG]4 gaagag [AAAG]13		0.015	0.041	0.071
DYS518	37	d	[AAAG]3 GAAG [AAAG]14 GGAG [AAAG]4 gaagag [AAAG]14		0.031	0.071	0.031
DYS518	37	e	[AAAG]3 GAAG [AAAG]13 GGAG [AAAG]4 gaagag [AAAG]15		-	0.020	0.010
DYS518	37	f	[AAAG]3 GAAG [AAAG]12 GGAG [AAAG]4 gaagag [AAAG]17		-	0.010	-
DYS518	38	a	[AAAG]3 GAAG [AAAG]16 GGAG [AAAG]4 gaagag [AAAG]13		0.031	0.020	0.031
DYS518	38	b	[AAAG]3 GAAG [AAAG]15 GGAG [AAAG]4 gaagag [AAAG]14		0.031	0.020	0.092

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS518	38	c	[AAAG]3 GAAG [AAAG]15 GGAG [AAAG]3 gaagag [AAAG]15		-	-	0.020
DYS518	38	d	[AAAG]3 GAAG [AAAG]14 GGAG [AAAG]4 gaagag [AAAG]15		0.015	0.010	0.031
DYS518	38	e	[AAAG]3 GAAG [AAAG]13 GGAG [AAAG]4 gaagag [AAAG]16		0.015	0.020	-
DYS518	38	f	[AAAG]3 GAAG [AAAG]12 GGAG [AAAG]4 gaagag [AAAG]17		-	0.010	-
DYS518	39	a	[AAAG]3 GAAG [AAAG]18 GGAG [AAAG]4 gaagag [AAAG]12		-	0.010	-
DYS518	39	b	[AAAG]3 GAAG [AAAG]17 GGAG [AAAG]4 gaagag [AAAG]13		0.015	0.020	0.010
DYS518	39	c	[AAAG]3 GAAG [AAAG]16 GGAG [AAAG]4 gaagag [AAAG]14		0.077	0.051	0.082
DYS518	39	d	[AAAG]3 GAAG [AAAG]15 GGAG [AAAG]4 gaagag [AAAG]15		0.062	0.031	0.082
DYS518	39	e	[AAAG]3 GAAG [AAAG]14 GAAG GGAG [AAAG]4 gaagag [AAAG]15		-	-	0.010
DYS518	39	f	[AAAG]3 GAAG [AAAG]14 GGAG [AAAG]4 gaagag [AAAG]16		-	0.010	0.010
DYS518	40	a	[AAAG]3 GAAG [AAAG]18 GGAG [AAAG]4 gaagag [AAAG]13		-	0.010	-
DYS518	40	b	[AAAG]3 GAAG [AAAG]17 GGAG [AAAG]4 gaagag [AAAG]14		0.031	0.020	0.031

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency		
					NP	MY	PJ
DYS518	40	c	[AAAG]3 GAAG [AAAG]16 GGAG [AAAG]4 gaagag [AAAG]15		0.077	0.020	0.071
DYS518	40	d	[AAAG]3 GAAG [AAAG]15 GGAG [AAAG]4 gaagag [AAAG]16		0.046	0.031	0.092
DYS518	40	e	[AAAG]3 GAAG [AAAG]14 GGAG [AAAG]4 gaagag [AAAG]17		0.015	-	0.031
DYS518	41	a	[AAAG]3 GAAG [AAAG]17 GGAG [AAAG]4 gaagag [AAAG]15		0.015	0.020	0.020
DYS518	41	b	[AAAG]3 GAAG [AAAG]16 GGAG [AAAG]4 gaagag [AAAG]16		0.062	0.020	0.071
DYS518	41	c	[AAAG]3 GAAG [AAAG]15 GGAG [AAAG]4 gaagag [AAAG]17		0.062	-	0.010
DYS518	42	a	[AAAG]3 GAAG [AAAG]18 GGAG [AAAG]4 gaagag [AAAG]15		-	0.020	0.020
DYS518	42	b	[AAAG]3 GAAG [AAAG]17 GAAG GGAG [AAAG]4 gaagag [AAAG]15		-	-	0.010
DYS518	42	c	[AAAG]3 GAAG [AAAG]17 GGAG [AAAG]4 gaagag [AAAG]16		0.108	0.010	0.010
DYS518	42	d	[AAAG]3 GAAG [AAAG]16 GGAG [AAAG]4 gaagag [AAAG]17		0.031	0.010	0.031
DYS518	43	a	[AAAG]3 GAAG [AAAG]19 GGAG [AAAG]4 gaagag [AAAG]15		0.015	-	-
DYS518	43	b	[AAAG]3 GAAG [AAAG]18 GGAG [AAAG]4 gaagag [AAAG]16		-	0.051	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS518	43	c	[AAAG]3 GAAG [AAAG]17 GGAG [AAAG]4 gaagag [AAAG]17		0.015	0.010	0.010
DYS518	43	d	[AAAG]3 GAAG [AAAG]16 GGAG [AAAG]4 gaagag [AAAG]18		-	-	0.020
DYS518	44	a	[AAAG]3 GAAG [AAAG]19 GGAG [AAAG]4 gaagag [AAAG]16		-	0.031	-
DYS518	44	b	[AAAG]3 GAAG [AAAG]18 GGAG [AAAG]4 gaagag [AAAG]17		0.015	-	-
DYS518	45	a	[AAAG]3 GAAG [AAAG]20 GGAG [AAAG]4 gaagag [AAAG]16		-	0.031	-
DYS518	45	b	[AAAG]3 GAAG [AAAG]18 GGAG [AAAG]4 gaagag [AAAG]18		0.015	-	-
DYS518					1.000	1.000	1.000
DYS526a	12		CTCT [CCTT]11		0.123	0.082	0.071
DYS526a	13		CTCT [CCTT]12		0.169	0.306	0.276
DYS526a	14	a	CTCT [CCTT]13		0.108	0.133	0.122
DYS526a	14	b	CTCT [CCTT]12 CGTT		-	0.010	-
DYS526a	15		CTCT [CCTT]14		0.508	0.429	0.490

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS526a	16		CTCT [CCTT]15		0.092	0.041	0.041
DYS526a					1.000	1.000	1.000
DYS533	8		[TATC]8		-	-	0.010
DYS533	9		[TATC]9		-	-	0.010
DYS533	10		[TATC]10		0.046	0.265	0.041
DYS533	11		[TATC]11		0.354	0.388	0.286
DYS533	12		[TATC]12		0.523	0.306	0.602
DYS533	13		[TATC]13		0.062	0.031	0.051
DYS533	14		[TATC]14		0.015	0.010	-
DYS533					1.000	1.000	1.000
DYS549	10		[GATA]10		-	-	0.020
DYS549	11		[GATA]11		0.062	0.163	0.194

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS549	12		[GATA]12		0.662	0.612	0.531
DYS549	13		[GATA]13		0.231	0.224	0.204
DYS549	14		[GATA]14		0.046	-	0.051
DYS549					1.000	1.000	1.000
DYS570	0		-		-	-	0.010
DYS570	15		[TTTC]15		0.062	0.051	0.102
DYS570	16		[TTTC]16		0.062	0.245	0.082
DYS570	17		[TTTC]17		0.200	0.235	0.173
DYS570	18	a	[TTTC]18		0.308	0.255	0.235
DYS570	18	b	[TTTC]15 CTTC [TTTC]2		-	0.010	-
DYS570	19		[TTTC]19		0.277	0.122	0.204
DYS570	20		[TTTC]20		0.046	0.051	0.122

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS570	21		[TTTC]21		0.046	0.020	0.071
DYS570	23		[TTTC]23		-	0.010	-
DYS570					1.000	1.000	1.000
DYS576	0		-		-	0.020	0.020
DYS576	15		[AAAG]15		0.062	0.031	0.051
DYS576	16		[AAAG]16		0.092	0.051	0.082
DYS576	17		[AAAG]17		0.231	0.163	0.265
DYS576	18		[AAAG]18		0.385	0.398	0.296
DYS576	19		[AAAG]19		0.154	0.265	0.245
DYS576	20		[AAAG]20		0.062	0.061	0.031
DYS576	21		[AAAG]21		0.015	0.010	-
DYS576	22		[AAAG]22		-	-	0.010

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS576					1.000	1.000	1.000
DYS612	30	a	[CCT]5 CTT [TCT]4 CCT [TCT]19		-	-	0.010
DYS612	30	b	[CCT]5 CTT [TCT]3 TTT CCT [TCT]19		-	0.010	-
DYS612	31		[CCT]5 CTT [TCT]4 CCT [TCT]20		-	-	0.020
DYS612	32		[CCT]5 CTT [TCT]4 CCT [TCT]21		-	-	0.031
DYS612	33	a	[CCT]5 CTT [TCT]4 CCT [TCT]22		0.031	0.020	0.041
DYS612	33	b	[CCT]5 CTT [TCT]3 TTT CCT [TCT]22		-	0.010	-
DYS612	34	a	[CCT]5 CTT [TCT]4 CCT [TCT]23		0.062	0.051	0.041
DYS612	34	b	[CCT]5 CTT [TCT]3 TTT CCT [TCT]23		0.015	0.010	-
DYS612	35	a	[CCT]5 CTT [TCT]4 CCT [TCT]24	rs555095027 (T>C)	0.015	-	0.010
DYS612	35	b	[CCT]5 CTT [TCT]4 CCT [TCT]24		0.108	0.143	0.204
DYS612	36	a	[CCT]5 CTT [TCT]4 CCT [TCT]25	rs555095027 (T>C)	0.015	0.020	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency		
					NP	MY	PJ
DYS612	36	b	[CCT]5 CTT [TCT]4 CCT [TCT]25		0.246	0.184	0.235
DYS612	36	c	[CCT]5 CTT [TCT]3 TTT CCT [TCT]25		-	0.010	-
DYS612	37	a	[CCT]5 CTT [TCT]4 CCT [TCT]26	rs555095027 (T>C)	0.031	0.020	0.010
DYS612	37	b	[CCT]5 CTT [TCT]4 CCT [TCT]26		0.169	0.245	0.204
DYS612	37	c	[CCT]5 CTT [TCT]4 CCT [TCT]22 TAT [TCT]3		-	0.010	-
DYS612	37	d	[CCT]5 CTT [TCT]3 TTT CCT [TCT]26		0.015	0.010	-
DYS612	38	a	[CCT]5 CTT [TCT]4 CCT [TCT]27	rs555095027 (T>C)	0.015	-	-
DYS612	38	b	[CCT]5 CTT [TCT]4 CCT [TCT]27		0.185	0.163	0.122
DYS612	38	c	[CCT]5 CTT [TCT]3 TTT CCT [TCT]27		0.015	-	-
DYS612	39	a	[CCT]5 CTT [TCT]4 CCT [TCT]28		0.031	0.071	0.061
DYS612	39	b	[CCT]5 CTT [TCT]3 TTT CCT [TCT]27		0.015	-	-
DYS612	40		[CCT]5 CTT [TCT]4 CCT [TCT]29		0.031	0.020	0.010

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS612					1.000	1.000	1.000
DYS626	23		AGAA AGGA AGAA [AAAG]11 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.015	-	-
DYS626	25		AGAA AGGA AGAA [AAAG]13 [AGAA]2 AGAG [GAAG]3 [AAAG]3		-	0.010	0.010
DYS626	26		AGAA AGGA AGAA [AAAG]14 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.031	0.010	0.041
DYS626	27		AGAA AGGA AGAA [AAAG]15 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.015	0.080	0.010
DYS626	28		AGAA AGGA AGAA [AAAG]16 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.123	0.190	0.031
DYS626	29		AGAA AGGA AGAA [AAAG]17 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.108	0.200	0.122
DYS626	30		AGAA AGGA AGAA [AAAG]18 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.169	0.120	0.122
DYS626	31	a	AGAA AGGA AGAA [AAAG]19 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.169	0.110	0.204
DYS626	31	b	AGAA AGGA AGAA [AAAG]9 AAGG [AAAG]9 [AGAA]2 AGAG [GAAG]3 [AAAG]3		-	0.010	-
DYS626	32	a	AGAA AGGA AGAA [AAAG]20 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.169	0.110	0.163
DYS626	32	b	AGAA AGGA AGAA [AAAG]9 AAGG [AAAG]10 [AGAA]2 AGAG [GAAG]3 [AAAG]3		-	0.010	-

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS626	32	c	AGAA AGGA AGAA [AAAG]9 AAGG [AGAA]10 AGAG [GAAG]3 [AAAG]3		-	-	0.041
DYS626	33	a	AGAA AGGA AGAA [AAAG]21 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.092	0.090	0.112
DYS626	33	b	AGAA AGGA AGAA [AAAG]9 AAGG [AAAG]11 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.031	0.030	-
DYS626	33	c	AGAA AGGA AGAA [AAAG]9 AAGG [AGAA]11 AGAG [GAAG]3 [AAAG]3		-	-	0.041
DYS626	34	a	AGAA AGGA AGAA [AAAG]22 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.046	0.020	0.041
DYS626	34	b	AGAA AGGA AGAA [AAAG]18 AAGG [AAAG]2 [AGAA]2 AGAG [GAAG]3 [AAAG]3		-	0.010	-
DYS626	34	c	AGAA AGGA AGAA [AAAG]10 AAGG [AGAA]11 AGAG [GAAG]3 [AAAG]3		-	-	0.010
DYS626	34	d	AGAA AGGA AGAA [AAAG]9 AAGG [AGAA]12 AGAG [GAAG]3 [AAAG]3		-	-	0.020
DYS626	35		AGAA AGGA AGAA [AAAG]23 [AGAA]2 AGAG [GAAG]3 [AAAG]3		0.031	-	0.031
DYS626					1.000	1.000	1.000
DYS627	0		-		-	-	0.010
DYS627	13		[AGAG]3 [AAAG]10		-	-	0.010

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS627	15		[AGAG]3 [AAAG]12		0.015	0.010	-
DYS627	16		[AGAG]3 [AAAG]13		0.015	0.020	0.041
DYS627	17		[AGAG]3 [AAAG]14		0.246	0.071	0.255
DYS627	17.2		[AGAG]2 AG [AAAG]15		-	0.020	-
DYS627	18	a	[AGAG]2 [AAAG]16		0.015	-	-
DYS627	18	b	[AGAG]3 [AAAG]15		0.262	0.204	0.184
DYS627	18.2		[AGAG]3 [AAAG]8 AA [AAAG]7		-	-	0.010
DYS627	19	a	[AGAG]3 [AAAG]16		0.077	0.143	0.153
DYS627	19	b	[AGAG]3 [AAAG]16	rs571126660 (A>G)	-	0.010	-
DYS627	20	a	[AGAG]3 [AAAG]17		0.231	0.235	0.122
DYS627	20	b	[AGAG]3 [AAAG]13 AAAA AGAG [AAAG]2		0.015	-	-
DYS627	21		[AGAG]3 [AAAG]18		0.046	0.133	0.143

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS627	22		[AGAG]3 [AAAG]19		0.077	0.102	0.061
DYS627	23		[AGAG]3 [AAAG]20		-	0.051	0.010
DYS627					1.000	1.000	1.000
DYS635	17		[TAGA]11 [TACA]2 [TAGA]4		0.015	-	-
DYS635	18		[TAGA]8 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		-	0.010	0.010
DYS635	19	a	[TAGA]8 TACA TAGA TACA [TAGA]2 [TACA]2 [TAGA]4		-	0.010	-
DYS635	19	b	[TAGA]9 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.015	0.031	-
DYS635	20		[TAGA]10 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.138	0.255	0.031
DYS635	21	a	[TAGA]7 [TACA]2 [TAGA]2 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.031	-	-
DYS635	21	b	[TAGA]10 [TACA]3 [TAGA]2 [TACA]2 [TAGA]4		-	0.051	-
DYS635	21	c	[TAGA]11 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.108	0.357	0.194
DYS635	22	a	[TAGA]8 [TACA]2 [TAGA]2 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.031	-	0.020

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS635	22	b	[TAGA]12 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.046	0.133	0.082
DYS635	23	a	[TAGA]9 [TACA]2 [TAGA]2 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.292	0.061	0.378
DYS635	23	b	[TAGA]13 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		-	0.020	0.082
DYS635	24	a	[TAGA]10 [TACA]2 [TAGA]2 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.169	0.020	0.020
DYS635	24	b	[TAGA]14 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.062	0.020	0.071
DYS635	25	a	[TAGA]11 [TACA]2 [TAGA]2 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.077	0.010	0.082
DYS635	25	b	[TAGA]15 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		-	0.010	-
DYS635	26		[TAGA]12 [TACA]2 [TAGA]2 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		-	0.010	0.031
DYS635	28		[TAGA]14 [TACA]2 [TAGA]2 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4		0.015	-	-
DYS635					1.000	1.000	1.000
DYS643	8		[CTTTT]8		0.031	0.031	0.010
DYS643	9		[CTTTT]9		0.092	0.122	0.163

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
DYS643	10		[CTTTT]10		0.615	0.194	0.551
DYS643	11		[CTTTT]11		0.138	0.163	0.153
DYS643	12		[CTTTT]12		0.123	0.337	0.092
DYS643	13		[CTTTT]13		-	0.133	0.031
DYS643	14		[CTTTT]14		-	0.020	-
DYS643					1.000	1.000	1.000
GATAH4	10		[TCTA]10		0.015	0.010	0.041
GATAH4	11		[TCTA]11		0.215	0.408	0.235
GATAH4	12		[TCTA]12		0.385	0.449	0.490
GATAH4	13		[TCTA]13		0.338	0.122	0.235
GATAH4	14		[TCTA]14		0.046	0.010	-
GATAH4					1.000	1.000	1.000

연세대학교
YONSEI UNIVERSITY

Continued)

Locus	CE allele	MPS allele	Repeat structure	Flanking Region Variation	Frequency NP	MY	PJ
Y-M175	Del		TTCTC AC TTCTC		0.138	0.633	-
Y-M175	Ins		TTCTC AC [TTCTC]2		0.862	0.367	1.000
Y-M175					1.000	1.000	1.000

Abbreviations: CE, Capillary Electrophoresis; MPS, Massively Parallel Sequencing; NP, Nepal; MY, Myanmar; PJ, Pakistan (Punjab).

The repeat sequences of each allele were indicated with a square bracket and arranged with the forward strand direction.

*Alleles with the same size but different repeat structure in the MPS analysis were tentatively classified as sub-alleles and indicated with lowercase letters.

*Segments that were not counted for the allele designation were indicated with lowercase letters.

연세대학교
YONSEI UNIVERSITY

*The DYF399S1 marker contains the sequence written as N7: AAGAAAA or N8: AAGAAAAG. N7 sequence was defined as 2 repeats and N8 as 2.1 repeats.

*The DYS389II marker contains the conservative sequence written as N48: TACATAGATAATACAGATGAGAGTTGGATACAGAAGTAGGTATAATGA.

*The DYS448 contains the conservative sequence written as N42: ATAGAGATAGAGAGATAGAGATAGAGATAGATAGATAGAGAA.

*The DYS449 contains the conservative sequence written as N50: TCTCTCTCCTCCTCTTTCTTTCCTTCTTTCTTTCTTTTCCTCTTTCCTTC.

연세대학교
YONSEI UNIVERSITY

염기 서열 분석으로 얻어진 대립유전자 변이의 수는 길이 분석만으로 얻어진 대립유전자 수에 비해 1.84배 증가하였다. 길이 기반 대립유전자의 수는 287개인데 반해, 염기서열 기반 대립유전자는 527개로, 염기서열 분석으로 240개의 대립유전자가 새로 확인되었다. 20개의 좌위에서 염기서열 변이를 보였는데, 19개의 좌위는 Y 염색체 STR의 반복 영역 변이를, 7개의 좌위는 주변 영역에서 변이를 보였다.

각 Y 염색체 표지자에 대하여 길이 및 염기서열 기반 대립유전자 수, 반복 영역 및 주변 영역 변이에 의해 얻어진 대립유전자 수의 증가 정도를 그림 4에 제시하였다. 특히, 이번 연구에서 사용한 MPS 패널에 포함한 11개의 RM Y-STR에서 더 많은 염기서열 변이가 확인되었다. 염기서열 분석으로 증가된 대립유전자 240개 중에 159개는 11개의 RM Y 염색체 STR 세트에서, 나머지 81개는 기존 22개 Y 염색체 STR에 의해 증가하였다.

연세대학교
YONSEI UNIVERSITY

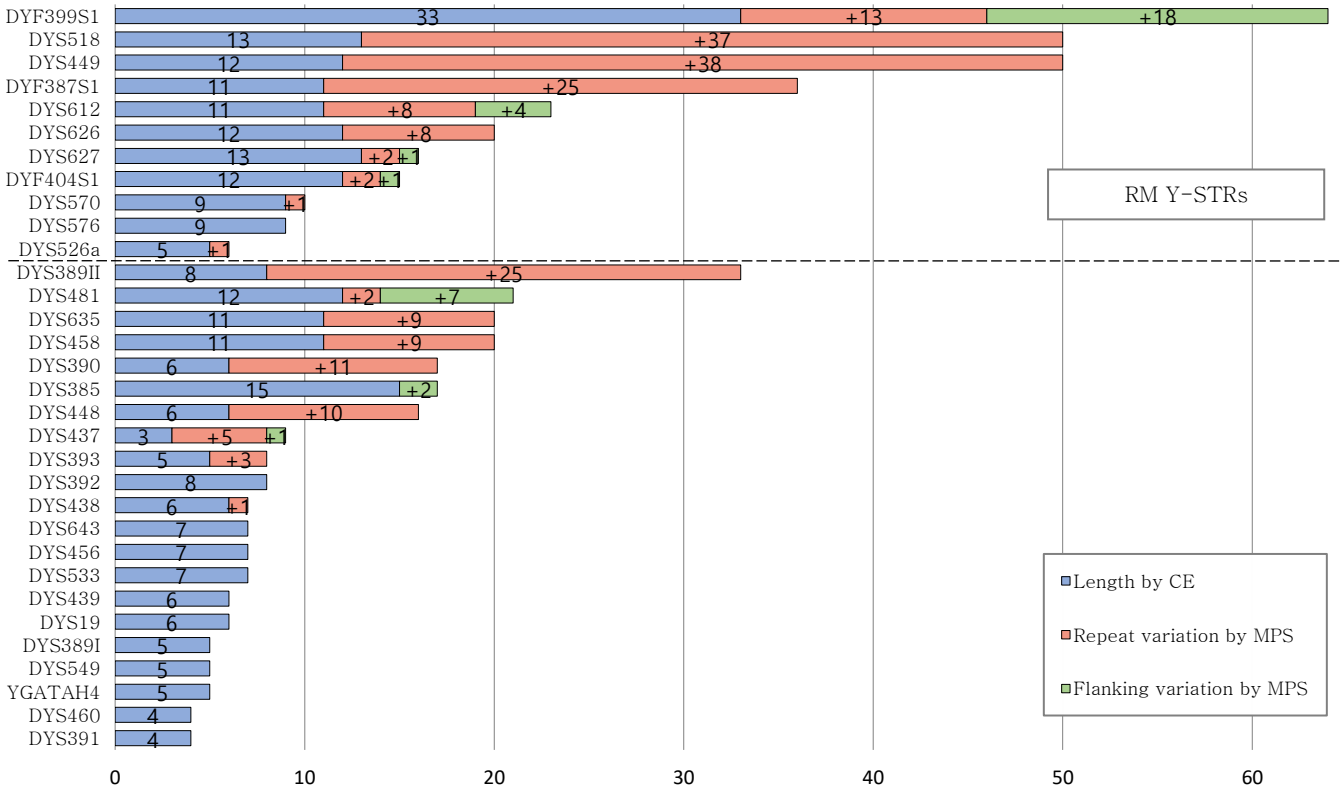

그림 4. Comparison of the number of observed alleles by CE and MPS of 33 Y-STRs

연세대학교
YONSEI UNIVERSITY

Abbreviations: CE, capillary electrophoresis; MPS, massively parallel sequencing; STR, short tandem repeat.

The blue bar presents the number of alleles by size-based method and the red bar presents the number of alleles observed by repeat region variations resulted from MPS method. The green bar presents the number of alleles observed by flanking region SNP resulted from MPS method.

연세대학교
YONSEI UNIVERSITY

염기서열 변이에는 Y 염색체 STR의 반복 영역의 변이 그리고 그 주변 영역의 변이가 있고, 반복 영역의 변이에는 반복 영역을 구성하는 반복 모티프의 다양한 수 조합과 반복 모티프 내 염기서열의 변이가 있다. DYS449 표지자는 반복 영역의 변이로 인해 대립유전자 수가 가장 많이 증가하였다. 길이 기반 12개의 대립 유전자에서 서열 기반 50개의 대립유전자로, 38개의 대립유전자가 서열 분석으로 인해 새로 나뉘었다 (약 4.17배 증가). 기본 구조는 [TTTC]n N50 [TTTC]m (n, m은 정수, N50은 반복 수 계산에 포함되지 않음)으로, 두 개의 반복 모티프에서 n과 m의 다양한 수의 조합이 관찰되었다. 예를 들어 대립 유전자 33 (allele 33)의 경우, 동일한 길이의 allele 33이 6개의 다른 염기서열을 갖도록 나뉘어졌다(33a-f). n과 m이 18과 15, 17과 16, 14와 19와 같이 다양하게 확인되었으며, 반복 모티프 내 염기서열 변이로 인하여 [TTTC]16 N50 [TTTC]5 CTTC [TTTC]11 (TTTC 모티프가 CTTC로)의 구조 또한 확인되었다. DYS449 다음으로 대립 유전자 수가 많이 증가한 표지자는 DYS518로, 길이 기반 13개 대립 유전자에서 서열 기반 50개의 대립유전자로 나뉘며, 37개의 대립유전자가 증가하였다 (약 3.85배 증가). 기본 구조는 [AAAG]3 GAAG [AAAG]n GGAG [AAAG]4 gaagag [AAAG]m (gaagag는 반복 수 계산에 포함되지 않음)으로, 두 개의 반복 모티프에서 n과 m의 다양한 수의 조합이 관찰되었다. 예를 들어 대립 유전자 39 (allele 39)의 경우, 동일한 길이지만 다

연세대학교
YONSEI UNIVERSITY

른 염기서열을 갖는 6개의 대립 유전자로 나뉘었다 (39a-f). n과 m이 18과 12, 17과 13, 16과 14 등 다양한 조합을 보였으며, 반복 모티프 내 염기서열 변이로 인하여 [AAAG]3 GAAG [AAAG]14 GAAG GGAG [AAAG]4 gaagag [AAAG]15 (AAAG 모티프가 GAAG로)와 같은 구조도 관찰하였다. 이외에도 DYF387S1와 DYS389II 등에서 서열 분석에 의해 각각 25개씩 대립유전자 수가 증가하였고, 그 다음은 DYS390, DYS448 표지자가 뒤를 이었다(11개, 10개).

Y 염색체 STR의 반복 영역의 주변 영역에서도 염기 서열 변이를 관찰하였다. 총 7개의 Y 염색체 표지자에서 관찰되었으며, 자세한 정보는 표 6에 제시하였다.

연세대학교
YONSEI UNIVERSITY

표 6. Sequence variations observed in flanking region of Y-STRs

Marker	Type	rs number (dbSNP build 155)	Wild	Mutant	Position on Chr Y (GRCh38.p13)	Frequency		
						NP	MY	PJ
DYF399S1	SNP	rs4306075	A	G	22,950,382	0.319	0.331	0.276
DYF399S1		rs878949651	A	G	22,950,401	0.304	0.307	0.340
DYS437		rs9786886	C	T	12,346,264	-	-	0.010
DYS481		rs368663163	G	A	8,558,321	0.123	0.010	0.112
DYS612		rs555095027	T	C	13,640,861	0.077	0.041	0.020
DYS627		rs571126660	A	G	8,782,059	-	0.010	-
DYF399S1	Indel	rs2090175785	GA	-	24,584,125-24,584,127	-	0.003	0.013
DYF404S1		rs34228393	TTCT	-	23,807,934-23,807,998	-	-	0.006
DYS385ab		rs1248860842	CCTT	-	18,639,757-18,639,780	0.008	0.005	-
DYS481		rs2015234517	GCT	-	8,558,332-8,558,338	-	0.041	-

Abbreviations: NP, Nepal; MY, Myanmar; PJ, Pakistan (Punjab).

연세대학교
YONSEI UNIVERSITY

주변 영역에서 단일염기다형성(SNP)와 염기의 삽입 또는 결실(Indel)이 관찰되었다. 관찰된 SNP에서 DYF399S1 표지자는 세 개의 인구 집단에서 모두 관찰되었으며, dbSNP의 명명은 rs4306075 (A에서 G로, 반복 영역으로부터 3’ 방향으로 한 뉴클레오티드(nt) 떨어져 위치)와 rs878949651 (A에서 G로, 반복 영역으로부터 3’ 방향 20nt 떨어져 위치)로 확인되었다. DYS437 표지자는 파키스탄 인구 집단에서만 관찰되었으며, dbSNP 명명 rs9786886 (C에서 T로, 반복 영역으로부터 5’ 방향 3nt 만큼 떨어져 위치)이었다. 또한, DYS481 표지자는 모든 인구 집단에서 dbSNP 명명 rs368663163 (G에서 A로, 반복 영역으로부터 5’ 방향으로 16nt 만큼) 이었고, DYS612 표지자는 모든 인구 집단에서 rs555095027 (T에서 C로, 3’ 방향으로 26nt 만큼)으로 확인되었다. DYS627 표지자는 미얀마 인구 집단에서 rs571126660 (A에서 G로, 3’ 방향으로 4nt 만큼) 변이를 확인하였다. 관찰된 Indel에서 DYF399S1 표지자는 미얀마, 파키스탄 인구 집단에서 관찰되었으며, rs2090175785 (GA의 결실, 반복 영역으로부터 3’ 방향)로 확인되었다. DYF404S1 표지자는 파키스탄 인구집단에서 rs34228393 (TTCT 결실, 반복 영역으로부터 3’ 방향)로 확인되었다. DYS385ab 표지자는 네팔과 미얀마 인구 집단에서 rs1248860842 (CCTT 결실, 반복 영역으로부터 3’방향), DYS481 표지자는 미얀마 인구집단에서 rs2015234517 (GCT 결실, 반복 영역으로부터 5’ 방향)이 관찰되었

연세대학교
YONSEI UNIVERSITY

다. 주변 영역 변이로 인해 가장 많은 대립유전자 수가 관찰된 표지자는 DYF399S1 이었다. 주변 영역 변이로 인해 18개의 대립유전자가 새로 나뉘었다. 기본 구조는 [GAAA]3 N7-8 [GAAA]n으로 (N7-8은 반복 수 계산에 포함되지 않는 염기서열), DYF399S1 표지자는 주변 영역 변이뿐만 아니라 반복 영역 변이에 의해서도 대립 유전자가 구분되었다. 대립 유전자 23의 경우, 6개의 다른 염기서열을 갖는 것을 확인하였다(23a-f). 반복 영역 구조는 같으나 DYF399S1에서 관찰된 두개의 주변영역 SNP (rs4306075와 rs878949651)로 두 개의 서열 구조로 나뉘었고, [GAAA]3 N7 [GAAA]17 GAAC (GAAA 모티프가 GAAC로)와 rs4306075 (A>G)가 함께 관찰된 서열 구조, [GAAA]3 N7 [GAAA]3 GAAG [GAAA]14 (GAAA 모티프가 GAAG로)와 rs4306075 (A>G)가 함께 관찰된 서열 구조, 그리고 [GAAA]3 N7 GGAA [GAAA]17 (GAAA 모티프가 GGAA로)와 같은 서열 구조로 총 6개로 나뉘었다. 다음으로 주변 영역 변이에 의해 대립 유전자 수가 두번째로 많이 증가한 표지자는 DYS481이었다 주변 영역 변이로 인해 7개의 대립 유전자가 새로 나뉘었으며 기본 구조는 [CTT]n이다. 대립 유전자 23의 경우, 주변 영역뿐만 아니라 반복 영역의 변이에 의해서 4개의 염기서열 구조를 가졌다. 기본 구조와 rs368663163의 SNP을 갖는 서열, rs2015234517의 deletion이 관찰된 서열, 그리고 CTG [CTT]22 (CTT 모티프가 CTG로)와 같은 서열 구조가 관찰되었다.

연세대학교
YONSEI UNIVERSITY

또한, 미세 변이(예: 22.2, 23.2 등의 대립유전자)를 가진 서열 구조도 확인되었다. 특히, DYF399S1 표지자에서 두 개의 유형으로 관찰되었다. 첫번째로는 반복 모티프 내에 2bp의 서열이 확인된 경우로, 예를 들어 대립 유전자 23.2를 보면 [GAAA]3 N7 [GAAA]17 AA GAAA와 같은 구조를 가지는 것을 확인하였다. 두번째로는 주변 영역의 2bp 결실(deletion)으로 인한 서열 구조이다. 예를 들어 17.2의 경우 [GAAA]3 N7 [GAAA]13 구조에 3' 쪽에 바로 인접한 2bp (GA)의 결실(rs2090175785)이 확인되었다.

4. 아시아 집단 시료 간 대립유전자 분포 양상

세 개의 아시아 인구집단 사이에 대립유전자가 다른 인구집단에 비해 상대적으로 높거나 낮은 대립유전자의 분포 양상을 확인하였다. 먼저 DYS635 표지자는 두 개의 서열 구조를 가지고 있는데, [TAGA]n [TACA]2 [TAGA]2 [TACA]2 [TAGA]2 [TACA]2 [TAGA]4 와, [TAGA]n [TACA]2 [TAGA]2 [TACA]2 [TAGA]4 구조이다. 후자의 서열 구조의 경우 미얀마 인구 집단에서 다른 두 인구집단에 비해 상대적으로 높은 빈도로 나타났다 (40%이상). 또한, DYS393 표지자의 경우, [AGAT]n과 CGAT [AGAT]n 두 개의 서열 구조 중 전자의 서열 구조가 미얀마 인구 집단에서 다른 두

연세대학교
YONSEI UNIVERSITY

인구집단에 비해 상대적으로 높은 빈도로 나타났다 (25%이상). (그림 5)

연세대학교
YONSEI UNIVERSITY

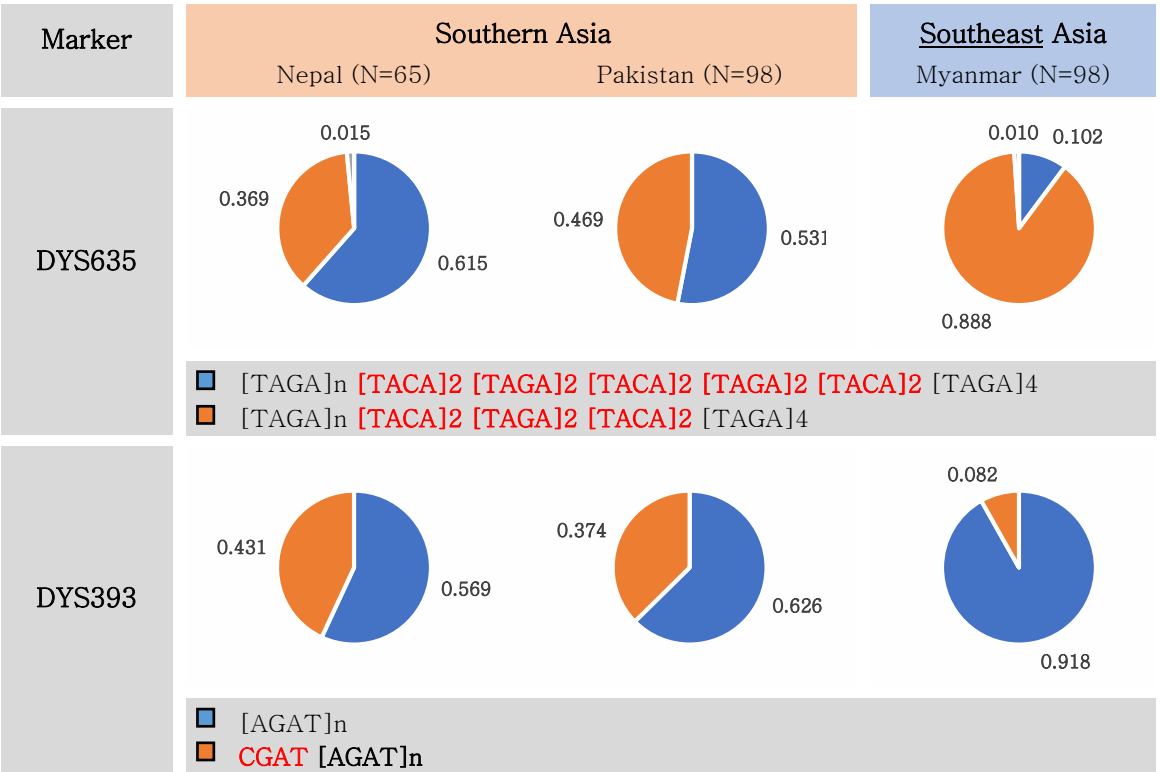

그림 5. Difference in allele distribution patterns observed across Asian population groups

연세대학교
YONSEI UNIVERSITY

The distribution of sequence structures was presented in circle graphs by classifying population groups into Southern Asia and Southeast Asia groups. The number next to the graph indicates the frequency of each sequence structure.

연세대학교
YONSEI UNIVERSITY

IV. 고찰

본 연구에서는 11 개의 RM Y 염색체 STR 을 포함하여 34 개의 Y 염색체 표지자를 분석할 수 있는 자체 구축한 MPS 체계를 사용하여 길이 기반 287 개 대립유전자에서 서열 기반 527 개로 대립 유전자 수가 상당히 증가함을 확인하였다. 이 증가의 66% (159/240)는 11 개의 RM Y 염색체 STR 로부터 얻은 것으로, 나머지 기존 22 개 Y 염색체 STR 세트보다 적은 개수의 RM Y 염색체 STR 표지자로도 훨씬 더 많은 서열 다양성을 보이는 것을 확인하였다 (그림 6). 따라서, RM Y 염색체 STR 의 서열 구조를 보다 자세히 분석하고자 하였다.

연세대학교
YONSEI UNIVERSITY

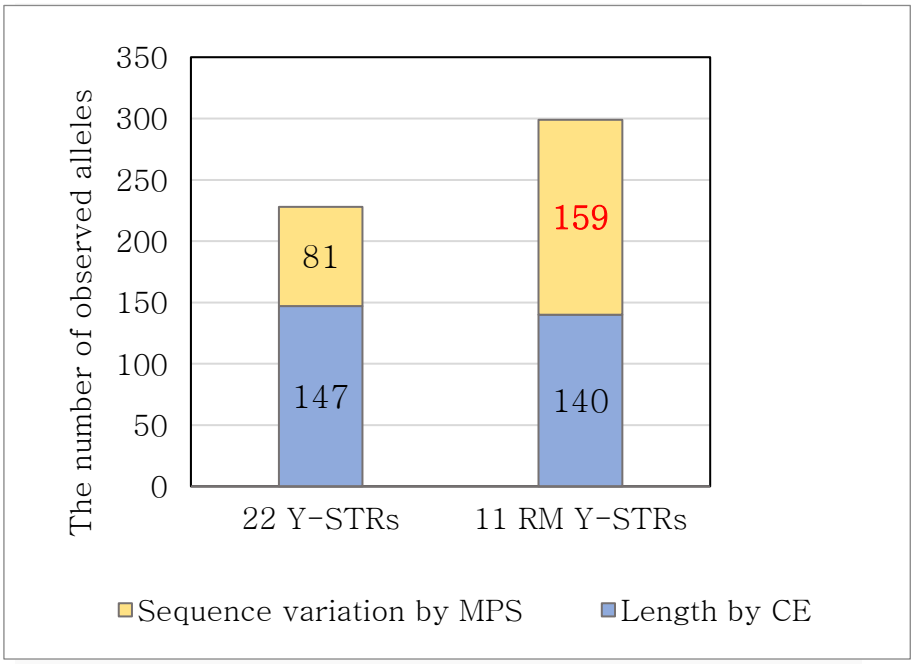

그림 6. The number of alleles observed by each length-, sequence-based analysis between conventional STR and RM Y-STR set

Abbreviations: STR, Short tandem repeat; CE, Capillary electrophoresis; MPS, Massively parallel sequencing.

The blue box below is the number of alleles obtained through the CE method and the yellow box is the number of alleles obtained by the MPS method.

연세대학교
YONSEI UNIVERSITY

많은 대립 유전자 수의 증가를 보인 대부분의 STR 은 여러 개의 반복 단위(repeat unit)로 구성된 복잡한 반복 구조(compound/complex repeat structure)를 갖는다.[8-13,17-19] 본 연구에서는 22 개의 기존 Y 염색체 STR 중 5 개, 11 개 RM Y 염색체 STR 중에는 6 개가 복잡한 반복 구조를 가지며 다양한 염기서열 구조로 나뉘었다. 따라서, 이런 서열 구조의 다양성은 길이기반 기법에서는 확인할 수 없었던 대립 유전자 내 변이들을 포함하며, 같은 길이지만 다른 염기서열을 가진 대립유전자를 구분해낼 수 있게 하여 Y 염색체 STR 의 식별력을 높여줄 것으로 사료된다. 따라서, 이런 고도로 다형화 된 좌위들의 염기서열 분석하는 것은 사람을 식별하는데 중요한 정보를 제공할 것으로 사료된다. [13, 17, 18]

염기서열 분석으로 관찰된 대립 유전자 수의 증가 정도를 3 개의 미국인 인구 집단(African American, U.S. Caucasians 및 U.S. Hispanics) 그리고 한국인 인구집단을 대상으로 32 개의 Y 염색체 표지자를 분석한 선행 연구와 비교하였을 때,[23] 대부분 미국인 집단은 DYF387S1 이나 DYS518 표지자에서 서열구조가 구분된 반면 한국인 집단은 DYS449 표지자에서 상대적으로 서열구조가 많이 구분되었다 (그림 7-1). 이번 연구에서 미얀마 인구 집단은 DYS518 및 DYF399S1 표지자에서, 파키스탄 인구 집단은 DYS449

연세대학교
YONSEI UNIVERSITY

및 DYS518 표지자에서 상대적으로 서열 구조가 다양하게 구분되었다 (그림 7-2).

연세대학교
YONSEI UNIVERSITY

AfAm (N=17)		Cauc(N=50)		Hisp(N=48)		Kor(N=105)	
Repeat region							
DYF387S1	6	DYF387S1	19	DYF387S1	13	DYS449	26
DYS518	5	DYS518	13	DYS518	13	DYF387S1	22
DYS449	5	DYS449	10	DYS449	11	DYS389 II	21
DYS389 II	4	DYS389 II	7	DYS389 II	10	DYS518	19
Flanking region							
DYF399S1	7	DYF399S1	1	DYF399S1	1	DYF399S1	3
DYS481						DYS481	1

그림 7-1. The number of increased alleles by sequence variations for each population in Moon et al.'s study[23]

Abbreviations: AfAm, African Americans; Cauc, Caucasians; Hisp, Hispanics; Kor, Koreans.

The top several markers with a large increase in the number of alleles by repeat region and flanking region variations were listed.

NP(N=65)		MY(N=98)		PJ(N=98)	
Repeat region					
DYS449	16	DYS518	24	DYS449	21
DYS518	16	DYF387S1	20	DYS518	21
DYF387S1	16	DYS449	19	DYF387S1	18
DYS389 II	14	DYS389 II	18	DYS389 II	15
Flanking region					
DYF399S1	8	DYF399S1	13	DYF399S1	11
DYS481	4	DYS481	3	DYS481	3

그림 7-2. The number of increased alleles by sequence variations for each population in this study

Abbreviations: NP, Nepal; MY, Myanmar; PJ, Pakistan (Punjab).

연세대학교
YONSEI UNIVERSITY

The top several markers with a large increase in the number of alleles by repeat region and flanking region variations were listed.

연세대학교
YONSEI UNIVERSITY

따라서, 법과학 실무에서 개인식별에 Y 염색체 STR 을 이용하고자 하는 경우에 대립유전자의 다양성이 요구되는데, 각 인구 집단에서 상대적으로 다양성이 높은 마커를 선정해 포함하여 분석한다면 인구집단 내에서 개인에 대한 변별력을 높일 수 있을 것으로 기대된다. 한국인을 대상으로 하여 상용화된 MPS 패널을 사용한 이전 연구와 비교하였을 때,[11] 이번 연구에서 분석한 표지자 중 동시에 포함하고 있는 DYF387S1 에서는 동일하게 가변 반복 모티프의 반복 수 조합으로만 서열 구조가 나뉘었다. DYS518 표지자도 마찬가지로 염기 서열 분석으로 대립 유전자의 수의 상당한 증가를 보였으나 반복 모티프 내의 뉴클레오티드 변이보다는 가변 반복 모티프의 다양한 수의 조합이 서열 구조의 주를 이루었다.

반복 영역뿐만 아니라 주변 영역에서도 변이가 다수 관찰되었다. DYS481 표지자에서 확인된 SNP (rs368663163)은 이번 연구에서 동남아시아에 속하는 미얀마 인구 집단은 0.010 의 빈도를 보인 반면 남아시아에 속하는 네팔과 파키스탄 인구 집단에서 약 0.123 과 0.112, 1000Genomes 데이터베이스에서는 남아시아 인구 집단에서 0.112 의 빈도로 보고되었다. 따라서, 해당 SNP 변이는 남아시아 인구 집단 특이적인 변이일 가능성이 높은 것으로 사료된다. 따라서, Y 염색체 STR 의 반복 영역뿐만 아니라 주변 영역에 대한 정보를 얻는 것은 반복 영역에 추가 정보를 제공하고,

연세대학교
YONSEI UNIVERSITY

나아가 특정 인구 집단을 추정할 수 있는 정보 또한 제시할 수 있을 것으로 기대된다. 반면, 주변 영역에서 반복 단위만큼(e.g., 3bp, 4bp …)의 결실은 분석 시에 대립 유전자가 동일한 반복이지만 한 반복 수 적게 호출되어 대립 유전자의 반복 수 예측에 혼란을 야기할 수 있다. 따라서, 이러한 정보의 축적은 CE 로부터 얻은 유전자형과의 역호환성을 유지하는데 중요할 것이다.

또한, 특히 DYF399S1 의 경우, 주변 영역 변이에 의해 많은 수의 대립 유전자로 나뉜 것을 확인하였다. 실제로 여러 인구 집단을 대상으로 Y 염색체 STR 을 분석한 선행연구에서 [23] 아프리카계 미국인 인구집단은 오직 주변 영역의 변이에 의해서만 서열 구조가 나뉜 것으로 나타났다. 주변 영역을 포함하여 반복영역의 다양성을 서열 분석하는 것은 반복영역만을 분석했을 때 보다 추가 정보를 제공하므로 식별력이 향상될 것으로 기대된다.[24] 이전 연구에서는 [23] 주변 영역의 2bp 결실로 DYF399S1.2 변이만을 보였으나 이번 연구에서는 해당 DYF399S1.2 변이뿐만 아니라 반복 모티프들 사이에 2bp 의 염기 서열이 확인되었다. 해당 구조를 갖는 서열은 모두 주변 영역에 rs878949651 (A>G)를 갖고 있었고, 이는 특정 SNP 과 특정 염기 서열 구조가 연관성을 가질 수 있다는 가능성을 보여주었다.[13] 이번 연구에서 확인한 RM Y 염색체 STR 의 서열 구조를 다른 연구와 비교하고자 하였으나, 이에 관한 연구가 적어 비교하기에 한계가 있었다. 더 많은 관련

연세대학교
YONSEI UNIVERSITY

염기 서열 구조 정보를 축적하게 된다면 연관성을 확인할 수 있을 것으로 사료된다.

다른 인구 집단과 비교했을 때 특정 인구 집단에서 상대적으로 많거나 적게 관찰된 염기 서열 구조를 확인하였다. 특히 DYS393 과 DYS635 표지자에서 관찰된 서열 구조는 남아시아 인구 집단인 네팔과 파키스탄, 동남아시아 인구 집단이 미얀마로 나뉘어 대립 유전자 분포 양상이 다르게 나타났다. DYS635 의 경우, 미얀마 인구집단에서 상대적으로 높은 빈도로 관찰된 서열 구조는 실제로 다수의 인구 집단을 대상으로 한 Novroski 등의 연구에서 [9] 동아시아 인구집단 내 0.884, 한국인 인구 집단을 대상으로 한 Kwon 등의 연구에서는 [8] 0.900 의 높은 빈도를 가져 결과가 일치하였다. 이를 통해 해당 서열 구조는 동남아시아 혹은 동아시아에서 주로 관찰된 다는 것을 유추해 볼 수 있었다. 더욱이 DYS393 표지자에서도 미얀마에서 상대적으로 높은 빈도로 관찰된 서열 구조는 한국인 집단 대상의 Kwon 등의 연구에서 [8] 0.996 의 높은 빈도로 확인되었다. 따라서, 이러한 결과들로 보아 특정 염기서열 구조가 물리적으로 가까운 두 나라 네팔과 파키스탄에서 양상이 비슷했던 것으로 추정해볼 수 있었다. 동시에 특정 인구집단에서 특정 서열 구조의 빈도 분포에 차이가 있다면 특정 집단 추정에 추가적인 정보가 될 수 있을 것으로 기대된다.

연세대학교
YONSEI UNIVERSITY

이번 연구에서는 이전에 연구된 패널 (PowerPlex Y23 system 과 호환)에 [8] 9 개의 RM Y 염색체 STR 과, Yfiler Plus system 에 포함된 하나의 표지자(DYS460)를 추가하여 새로 자체 구축하였다. 현재 상용화된 MPS 기반 패널은 Verogen 사의 ForenSeq DNA Signature Prep Kit 과 Promega 사의 PowerSeq 46GY system 이 있는데, 이들이 각각 4 개의 RM Y 염색체 STR (ForenSeq; DYF387S1, DYS570, DYS576, 그리고 DYS612)과 2 개의 RM Y 염색체 STR (DYS570 그리고 DYS576) 갖는데 반해 이번 연구에 사용된 자체 구축한 MPS 패널은 훨씬 많은 수의 RM Y 염색체 STR 을 분석할 수 있도록 하였다. 또한, ForenSeq kit 가 110-390bp[20], PowerSeq kit 가 140-300bp[21] 사이의 증폭산물 크기를 생성하는데 반해, 자체 구축한 MPS 패널은 85-274bp 의 상대적으로 훨씬 더 작은 크기의 증폭산물을 생성할 수 있게 설계되었다. 그러므로, 분해된 DNA 와 같이 법 과학 실무에서 해석하기에 어려운 시료의 해석에 있어 보다 많은 Y 염색체 STR 의 정보를 줄 수 있을 것으로 기대된다.[22]

자체 구축한 MPS 패널에 포함되는 하나의 Y-SNP 표지자인 Y-M175 좌위는 TTCTC 모티프의 결실(deletion)이 발생할 때, Y 하플로그룹 O 를 지정한다. Y 하플로그룹 O 는 주로 동아시아와 동남아시아 지역에서 가장 흔히 발견되는 것으로 알려져 있다. 실제로 여러 인구 집단을 대상으로 32 Y 염색체 표지자를 분석한

연세대학교
YONSEI UNIVERSITY

선행 연구에서도 해당 결실이 한국인 집단에서 79.0%의 높은 빈도로 나타났다.[23] 이번 연구 결과에서도 동남아시아에 속하는 미얀마 인구 집단에서 63.3%의 비교적 높은 빈도로 발견된 반면에 남아시아에 속하는 네팔 인구 집단에서는 13.8% 빈도로 나타났으며 파키스탄 인구 집단에서는 관찰되지 않았다 (그림 8). 즉, Y-M175 표지자의 결실이 Y 하플로그룹 O 를 지정하고 이는 동남아시아 및 동아시아에 많이 분포함으로 미루어 볼 때, 법과학 시료 분석에 있어서 Y-M175 표지자의 분석은 시료의 인구집단 추정에 추가적인 정보를 제공할 수 있을 것으로 기대된다.

연세대학교
YONSEI UNIVERSITY

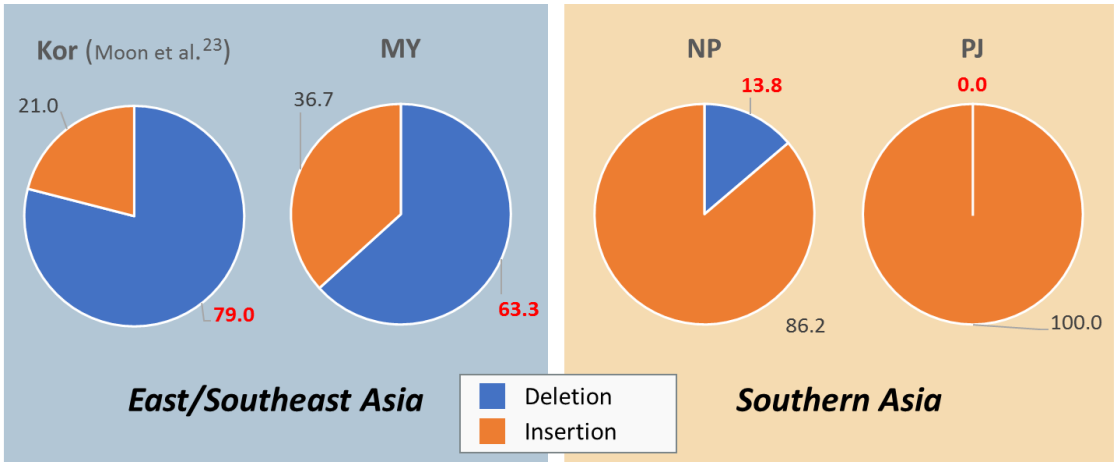

그림 8. The distributions of deletion and insertion in the Y-M175 marker for Asian populations on Moon et al.[23] and this study

Abbreviations: Kor, Koreans; NP, Nepal; MY, Myanmar; PJ, Pakistan (Punjab).

The blue portion indicates deletion and orange one indicates insertion, and each frequency is shown next to the circle graph.

연세대학교
YONSEI UNIVERSITY

이번 연구에서 자체 구축한 MPS 패널을 사용하여 분석된 리드는 염기서열 변이를 해석하기 위하여 최소 100 리드 수 이상을 갖도록 분석 한계치(threshold)를 설정하였다. ForenSeq DNA Signature Prep Kit 의 guide[20] 와 이를 사용한 연구들에서[9, 10] 낮은 표지자 커버리지를 가질 때 대립 유전자의 탈락과 증가된 n-1 stutter (반복 수보다 하나 적은 오류 증폭산물)의 발생의 가능성을 언급하며 DYS392 표지자의 해석에 어려움을 보고하였으나, 본 연구에서는 모두 100 리드 수 이상의 서열 구조를 가지며 대립 유전자 탈락은 관찰되지 않았다. 다만, 상대적으로 낮은 표지자 커버리지를 갖는 DYS449 에서 100 리드 수 이하를 가져 호출되지 않은 일부 서열의 경우, 분석 한계치를 70 으로 설정하여 서열을 재회수하였고 결과적으로 모든 표지자에 걸쳐 모든 염기 서열이 70 이상의 리드 수를 갖는 것으로 확인되었다. 모든 시료를 대상으로 CE 기법과 MPS 기법 사이에 유전자형의 일치 정도를 확인하였다. DYS576 에서 불일치를 보인 3 개의 시료는 Euplex-Y17N 의 결과에서 그림 9-1 의 예시와 같이 유전자형이 확인되었으나, MPS 결과에서는 유전자형이 검출되지 않았다. 3 개의 샘플에 대하여 MPS 프라이머 결합 부위를 포함하도록 하여 Sanger sequencing 을 수행한 결과, MPS 프라이머 결합 부위 내에서 공통된 하나의 SNP (rs757752030)을 확인하였다 (그림 9-2).

연세대학교
YONSEI UNIVERSITY

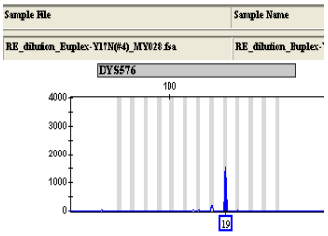

그림 9-1. Example of CE genotyping results for DYS576 in the MY028 sample with genotype discordance

The electropherogram obtained from Euplex-Y17N is shown.

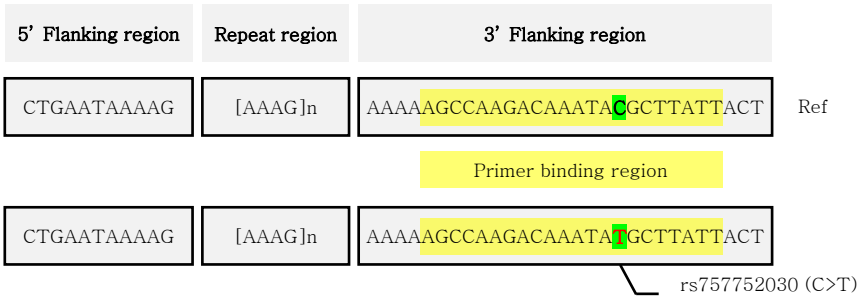

그림 9-2. Flanking region SNP observed within the 3' MPS primer binding site leading to genotype discordance in DYS576 marker

The 3' primer binding region where the SNP was observed was marked only (yellow box).

해당 SNP 으로 인하여 MPS 라이브러리를 생성하는 과정에 낮은 증폭 효율을 보여 MPS 결과 유전자형이 검출되지 않았다. 빈도가

연세대학교
YONSEI UNIVERSITY

낮기는 하지만 낮은 표지자 커버리지를 갖거나 프라이머 결합 부위의 SNP 등의 발생 가능성을 고려하였을 때, CE 기반 유전자형과의 역호환성 측면에서 해당 문제점들을 인식하여 해석에 유의해야 할 것으로 사료된다.

이번 연구에서는 연관되지 않은 남성들을 대상으로 염기 서열 분석을 통해 다양한 대립 유전자 변이를 확보하였고 이를 통해 식별력이 증가한 것을 확인하였다. 그러나 추후 연구에서는 아버지-아들 쌍과 같은 동일 부계 남성을 대상으로 하여 높은 돌연변이율을 가진 RM Y 염색체 STR 을 분석 한다면 개인 식별의 가능성을 확인해볼 수 있을 것이고, 이 표지자들의 염기 서열 기반 구조를 분석함으로써 부계의 특징을 지을 수 있는 변이를 확인해볼 수 있을 것으로 기대된다.

연세대학교
YONSEI UNIVERSITY

V. 결 론

이번 연구에서는 세 개의 아시아 인구 집단 (네팔, 미얀마, 파키스탄)의 261명 남성을 대상으로 11개의 RM Y 염색체 STR을 포함하여 총 33개의 Y 염색체 STR에 대한 염기 서열 변이를 확인하였다. 특히 11 RM Y 염색체 STR에서는 159개, 기존 22개 Y 염색체 STR에서는 81개의 염기서열 구조로 다양하게 구분되었다. 따라서, 다수의 RM Y-STR을 포함한 염기서열 분석을 법과학 실무에 적용하면 더 높은 염기서열 다양성을 확보하여 용의자와 가까운 부계 친척 관계인 남성이 용의자로 지목될 경우에 개인을 구분 짓는데 활용할 수 있을 것이다. 또한, 특정 인구 집단에서 상대적으로 빈도의 분포 차이를 보인 염기서열 구조들은 여러 문헌과 통계 자료를 통해 인구 집단 특이적일 가능성이 확인되었다. 따라서, 아시아 인구집단에서 조사된 대립유전자의 집단 특이적 분포는 법과학 실무에서 용의자의 인구 집단을 추정하여 그 범위를 좁히는데 유용한 정보를 제공할 수 있을 것이다. 본 연구는 현재 한국 사회에서 법유전학 정보가 많지 않은 주변 아시아 인구 집단을 대상으로 Y 염색체 STR의 염기서열 분석에 대한 효용성을 보여주었다. 다양성이 높은 RM Y 염색체 STR의 염기서열 분석으로 인한 대립 유전자 수의 증가와 분포를 제시하여 남성 개인식별과

연세대학교
YONSEI UNIVERSITY

생물지리학적 정보 추정의 가능성을 제시하였다. 본 연구를 기반으로 얻어진 주변 아시아 인구 집단의 Y 염색체 STR 염기서열에 대한 정보는 법과학 실무에서 MPS 적용을 통해 한국에서 체류하고 있는 주변 아시아 인구 집단 남성과 관련된 사건을 해결하는데 기초자료로서 활용될 것으로 기대된다.

연세대학교
YONSEI UNIVERSITY

참고 문헌

1. Butler JM. Fundamentals of Forensic DNA Typing. 1st ed. USA. Academic Press, 2009.

2. Jobling MA, Pandya A, Tyler-Smith C. The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 1997;110(3):118-24.

3. Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 2003;4(8):598-612.

4. Ballantyne KN, Keerl V, Wollstein A, Choi Y, Zuniga SB, Ralf A, et al. A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. Forensic Sci Int Genet 2012;6(2):208-18.

5. Ballantyne KN, Ralf A, Aboukhalid R, Achakzai NM, Anjos MJ, Ayub Q, et al. Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. Hum Mutat 2014;35(8):1021-32.

연세대학교
YONSEI UNIVERSITY

6. Adnan A, Ralf A, Rakha A, Kousouri N, Kayser M. Improving empirical evidence on differentiating closely related men with RM Y-STRs: A comprehensive pedigree study from Pakistan. Forensic Sci Int Genet 2016;25:45-51.

7. Rakha A, Oh YN, Lee HY, Hussain S, Waryah AM, Adnan A, et al. Discriminating power of rapidly mutating Y-STRs in deep rooted endogamous pedigrees from Sindhi population of Pakistan. Legal Med (Tokyo) 2018; 34: 17-20.

8. Kwon SY, Lee HY, Kim EH, Lee EY, Shin KJ. Investigation into the sequence structure of 23 Y chromosomal STR loci using massively parallel sequencing. Forensic Sci Int Genet 2016;25:132-41.

9. Novroski NMM, King JL, Churchill JD, Seah LH, Budowle B. Characterization of genetic sequence variation of 58 STR loci in four major population groups. Forensic Sci Int Genet 2016;25:214-26.

10. Phillips C, Devesse L, Ballard D, van Weert L, de la Puente M, Melis S, et al. Global patterns of STR sequence variation: Sequencing the CEPH human genome diversity panel for 58

연세대학교
YONSEI UNIVERSITY

forensic STRs using the Illumina ForenSeq DNA Signature Prep Kit. Electrophoresis 2018;39(21):2708-24.

11. Kim SY, Lee HC, Chung U, Ham SK, Lee HY, Park SJ, et al. Massive parallel sequencing of short tandem repeats in the Korean population. Electrophoresis 2018;39(21):2702-7.

12. Gettings KB, Borsuk LA, Steffen CR, Kiesler KM, Vallone PM. Sequence-based U.S. population data for 27 autosomal STR loci. Forensic Sci Int Genet 2018;37:106-15.

13. Kwon Y-L, Kim BM, Lee EY, Shin K-J. Massively parallel sequencing of 25 autosomal STRs including SE33 in four population groups for forensic applications. Scientific Reports 2021;11(1):4701.

14. Lee EY, Shin KJ, Rakha A, Shin JE, Park MJ, Kim NY, et al. Analysis of 22 Y chromosomal STR haplotypes and Y haplogroup distribution in Pathans of Pakistan, Forensic Sci Int Genet 2014;11:111-16.

15. Ha HH, Nguyen TH, Tran LH, Nguyen HTH, Hoang H, Chu HH, Genetic characteristics of 23 Y-chromosomal STRs in the Kinh

연세대학교
YONSEI UNIVERSITY

population in Northern Vietnam, Int J Legal Med 2019;133:1403-4.

16. Woerner AE, King JL, Budowle B. Fast STR allele identification with STRait Razor 3.0. Forensic Sci Int Genet 2017;30:18-23.

17. Huszar TI, Jobling MA, Wetton JH. A phylogenetic framework facilitates Y-STR variant discovery and classification via massively parallel sequencing. Forensic Sci Int Genet 2018;35:97-106.

18. Phillips C, Gettings KB, King JL, Ballard D, Bodner M, Borsuk L, et al. "The devil's in the detail": Release of an expanded, enhanced and dynamically revised forensic STR Sequence Guide. Forensic Sci Int Genet 2018;34:162-9.

19. Watahiki H, Fujii K, Fukagawa T, Mita Y, Kitayama T, Mizuno N. Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups. Forensic Sci Int Genet 2019;41:e1-e7.

20. Verogen. ForenSeq DNA Signature Prep Reference Guide. https://verogen.com/wp-content/uploads/2020/08/forenseq-dna-signature-prep-reference-guide-VD2018005-c.pdf.

연세대학교
YONSEI UNIVERSITY

21. Promega. PowerSeq® 46GY System. https://www.promega.kr/products/forensic-dna-analysis-mps/target-amplification-and-library-prep/powerseq-46gy-system/?catNum=PS4600.

22. Park MJ, Lee HY, Chung U, Kang SC, Shin KJ. Y-STR analysis of degraded DNA using reduced-size amplicons. Int J Legal Med 2007;121(2):152-7.

23. Moon MH, Hong SR, Shin KJ. Sequence variations of 31 Y-chromosomal short tandem repeats including 9 rapidly mutating Y-chromosomal short tandem repeats analyzed by massively parallel sequencing in three U.S. population groups and Korean population. J Korean Med Sci 2022;37(6) In progress.

24. Wendt FR, King JL, Novroski NMM, Churchill JD, Ng J, Oldt RF, et al. Flanking region variation of ForenSeq™ DNA Signature Prep Kit STR and SNP loci in Yavapai Native Americans. Forensic Sci Int Genet 2017;28:146-54.

연세대학교
YONSEI UNIVERSITY

ABSTRACT

Sequence variation analysis of Y chromosomal STRs using massively parallel sequencing in three Asian populations

Mi Hyeon Moon

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Kyoung-Jin Shin)

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. Massively parallel sequencing (MPS) can identify sequence variants in repeat and flanking regions of targeted STRs and analyze a lot of markers simultaneously, so can

연세대학교
YONSEI UNIVERSITY

lead in much more increased diversity of alleles and increased discrimination power. In this study, sequence variations of 33 Y-STRs including 11 RM Y-STRs (DYF387S1, DYF399S1, DYF404S1, DYS449, DYS518, DYS526, DYS570, DYS576, DYS612, DYS626 and DYS627) were investigated using MPS method for 261 males in Myanmar, Nepal, and Pakistan populations. We presented sequence-based allele frequencies, distributions, and gain in the number of alleles across Asian populations. Three samples showed genotype discordance between CE and MPS data by MPS primer binding site mutation. There were 240 alleles increases by MPS compared to CE-based alleles (287 to 527). The DYS449 marker showed the largest increase in the number of alleles by repeat region variation (a 4.17-foled increase) followed by the DYS518 marker. The highest increase in the number of alleles by single nucleotide polymorphisms in flanking region showed in the DYF399S1 marker followed by the DYS481 marker. Compared to 22 conventional Y-STRs, 11 RM Y-STRs were distinguished in

연세대학교
YONSEI UNIVERSITY

much more diverse sequence. Relative difference in alleles' distribution among populations showed in the DYS393 and DYS635 marker, which divided into South Asian (Nepal and Pakistan) and Southeast Asian (Myanmar) population groups. The flanking region polymorphisms in specific population groups were also observed. The usefulness of sequence analysis for Y-STRs was demonstrated in Asian populations. The increase in the number of alleles and distributions of alleles obtained by MPS of RM Y-STR was presented, suggesting the possibility of male individualization and biogeographic inference. It is expected that sequence variation of Y-STR in the Asian populations through this study will be used as the reference database for solving Asian male-related forensic cases.

Key words: Y chromosomal STR, RM Y-STR, massively parallel sequencing, sequence variations, Asian populations