



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](#)

차세대 염기서열 분석법을
이용한 혼합시료의 Y-STR
유전자형 분석

연세대학교 대학원
의 과 학 과
권 소 연

차세대 염기서열 분석법을
이용한 혼합시료의 Y-STR
유전자형 분석

지도교수 신 경 진

이 논문을 석사 학위논문으로 제출함

2016 년 6 월

연세대학교 대학원

의 과 학 과

권 소 연

권소연의 석사 학위논문을 인준함

심사위원 이 환 영 인

심사위원 신 경 진 인

심사위원 김 현 석 인

연세대학교 대학원

2016년 6월

감사의 글

법의유전학 분야에 대해 아무 것도 모르고 오직 호기심만 가득하였던 출발이 엇그제 같은데 벌써 석사 학위과정을 마무리하게 되었습니다. 이 분야를 배우고 싶어 준비하던 시간들과 학위 과정을 되돌아 보면 힘들고 지쳤던 시간들도 있었지만 주변에 계신 분들의 격려와 도움으로 이 결실을 맺을 수 있었다고 생각합니다.

먼저 법의유전학을 공부할 수 있는 기회를 주시고 첫 걸음부터 지금까지 항상 세심하게 지도해주신 신경진 교수님께 감사 인사 드립니다. 끊임없는 칭찬으로 저에게 용기를 주시고 많은 조언과 관심 주신 이환영 교수님, 좋은 논문이 될 수 있도록 많은 조언을 주신 김현석 교수님께도 감사의 말씀 전합니다. 그리고 많은 지도와 관심 주신 양우익 교수님께도 감사 드립니다. 바쁘신 와중에도 저를 위해 항상 시간 내 주시고 같이 토론하고 고민해 주신 정상은 선생님, 옆에서 묵묵히 도와주신 이은희 선생님, 넘치도록 도와준 새롬씨, 4층 활력소 수빈씨 그리고 제가 귀찮을 만큼 많이 물어봐도 항상 친절하게 가르쳐 주시는 이은영 선생님에게 진심으로 감사 드립니다. 멀리 계셔도 언제나 든든함이 느껴지는 멋진 선배님들이신 박명진, 정옥희, 심정은, 최아진, 김은혜 선생님에게도 감사의 말을 전합니다. 대학원 생활에 많은 힘이 되어준 소라언니와 윤겸언니에게도 고마움을 전하고 싶습니다.

지금까지 제가 선택하는 길에 있어 늘 지지와 사랑을 아끼지 않으시는 부모님과 귀여운 동생들 그리고 큰이모, 큰이모부 모두 사랑하고 고맙습니다. 그리고 멀리서 지켜봐 주고 응원해 주는 친구들에게도 고마움을 전합니다.

저를 지켜 봐주고 도움 주신 모든 분들께 다시 한번 감사 드리며, 항상 노력하고 발전하는 사람이 되도록 하겠습니다. 감사합니다.

차례

국문요약	1
I. 서론	4
II. 재료 및 방법	9
1. 연구대상	9
2. DNA 추출 및 정량	10
가. 한국인 시료의 추출 및 정량	10
3. 혼합 시료 준비	11
가. 표준 남성 시료를 이용한 혼합 시료의 준비	11
나. 한국인 남성 시료를 이용한 혼합 시료의 준비	12
(1) PowerPlex [®] Y23를 이용한 한국인 남성 시료의 유전자 형 확보	12
(2) 두 명의 한국인 남성 시료를 이용한 혼합시료 조합 생성.....	12
(3) 두 명의 한국인 남성 혼합 시료 비율	14
(4) 두 명의 한국인 남성 시료를 이용한 혼합시료 제조	14
4. 연구 대상 Y-STR 표지자	15
5. 다중중합효소연쇄반응 체계	15
6. NGS 수행	19
가. 표준 남성 시료를 이용한 혼합시료의 NGS수행.....	19
(1) 표준 남성 시료를 이용한 혼합 시료의 라이브러리 (Library) 제작	19
(2) 표준 남성 시료를 이용한 혼합 시료 라이브러리의 정제 및 정도 관리 (Quality control)	21

(3) NGS 장비를 이용한 표준 남성 시료 혼합물 염기 서열 분석	22
나. 한국인 남성 시료를 이용한 혼합 시료의 NGS 수행	23
(1) 한국인 남성 시료를 이용한 혼합시료의 라이브러리 (Library) 제작	23
(2) 한국인 남성 시료를 이용한 혼합시료 라이브러리 정제 및 정도 관리 (Quality control)	24
7. NGS 결과를 이용한 Y-STR 유전자형 결정 및 자료 분석	25
(1) 표준 남성 시료 NGS 결과를 이용한 Y-STR 유전자형 결정 및 자료분석	25
(2) 한국인 남성 시료 NGS 결과를 이용한 Y-STR 유전자형 결정 및 자료 분석	27
8. NGS 및 CE 기법을 이용한 표준 남성 시료 분석 비교	28
9. NGS기반 Y-STR 분석 체계에서의 Stutter 특성 조사	29
10. NGS기반 Y-STR 분석 체계에서 발생하는 노이즈 조사	31
III. 결과	32
1. 표준 남성 시료 혼합물 결과 분석	32
가. 표준 남성 시료 혼합물을 이용한 CE 결과	32
나. Illumina Miseq 분석용 표준 남성시료 라이브러리 제작 결과	34
다. NGS 장비를 이용한 시퀀싱 결과	35
라. 표준 남성 시료를 이용한 NGS 결과 분석	35
2. 한국인 남성 혼합 시료 결과분석	38

가. PowerPlex [®] Y23를 이용한 한국인 남성 시료의 유전자형 결과	38
나. 두 명의 한국인 남성을 이용한 혼합 시료 제조 결과	40
(1) 두 명의 한국인 남성을 이용한 시료 조합	40
(2) 두 명의 한국인 남성 시료의 혼합 비율	43
다. Illumina Miseq 분석용 한국인 남성 시료 라이브러리 제작 결과	43
라. NGS 장비를 이용한 한국인 남성 시료 시퀀싱 결과	44
마. 한국인 남성 시료를 이용한 NGS 결과 분석	45
(1) 한국인 남성 단일 시료를 이용한 NGS 결과	45
(2) 한국인 남성 단일 시료 NGS 결과를 이용한 stutter 조사 결과	47
(3) 한국인 남성 단일 시료 NGS 결과를 이용한 노이즈 조사 결과	50
(4) Stutter 필터비율 및 노이즈 필터비율을 반영한 한국인 남성 혼합 시료 분석	52
(5) 한국인 남성 혼합 대립유전자형 결정 결과	53
(6) 한국인 남성 혼합 시료 혼합비율 추정 결과	67
 IV. 고찰	 71
V. 결론	83
 참고문헌	 85
ABSTRACT	90

그림 차례

그림 1. Average read counts for the 23 Y-STRs of the Y-STR multiplex system used in this study with selected 19 samples 46

그림 2-1. Boxplot of the percentage of minor components for mixture ratio estimation in mixed DNAs by 1:1 ~ 1:49 ratios in 9 Y-STR loci 68

그림 2-2. Boxplot of the percentage of minor components for mixture ratio estimation in mixed DNAs by 1:1 ~ 1:49 ratios in 9 Y-STR loci 69

그림 3. Boxplot of the percentage of minor components for mixture ratio estimation in mixed DNAs by the number of repeat difference in 1:1 ~ 1:14 70

표 차례

표 1. PCR primer sequences and final concentrations of Y-STR multiplex PCR systems for the first step of library preparation	16
표 2. Result of Y-STR genotyping in various range of mixture ratio of control DNAs from capillary electrophoresis	33
표 3. Allele determinations in various ranges of mixture ratio of control DNAs from next generation sequencing	37
표 4. Result of Y-STR genotyping with 20 unrelated Korean males using PowerPlex [®] Y23	39
표 5. The number of repeat difference of 21 Y-STRs with 11 sets of mixed DNAs using 13 unrelated Korean males	42
표 6. Stutter ratios of 21 Y-STR for the NGS analysis system used in this study	48
표 7. Stutter filter ratio of 21 Y-STR for the mixture interpretation using NGS analysis system	49

표 8. Noise ratios and noise filter ratio of 21 Y-STR for the NGS analysis system used in this study	51
표 9-1. Allele determinations in various range of mixture ratio of two Korean males of Set 1 (NGS001:NGS006) from next generation sequencing	55
표 9-2. Allele determinations in various range of mixture ratio of two Korean males of Set 2 (NGS003:NGS006) from next generation sequencing	56
표 9-3. Allele determinations in various range of mixture ratio of two Korean males of Set 3 (NGS005:NGS012) from next generation sequencing	57
표 9-4. Allele determinations in various range of mixture ratio of two Korean males of Set 4 (NGS001:NGS005) from next generation sequencing	58
표 9-5. Allele determinations in various range of mixture ratio of two Korean males of Set 5 (NGS007:NGS010) from next generation sequencing	59

- 표 9-6. Allele determinations in various range of mixture ratio of two Korean males of Set 6 (NGS007:NGS013) from next generation sequencing 60
- 표 9-7. Allele determinations in various range of mixture ratio of two Korean males of Set 7 (NGS003:NGS007) from next generation sequencing 61
- 표 9-8. Allele determinations in various range of mixture ratio of two Korean males of Set 8 (NGS008:NGS012) from next generation sequencing 62
- 표 9-9. Allele determinations in various range of mixture ratio of two Korean males of Set 9 (NGS002:NGS011) from next generation sequencing 63
- 표 9-10. Allele determinations in various range of mixture ratio of two Korean males of Set 10 (NGS003:NGS019) from next generation sequencing 64
- 표 9-11. Allele determinations in various range of mixture ratio of two Korean males of Set 11 (NGS008:NGS018) from next generation sequencing 65

표 10. The counts of allele determinations of 163 mixed DNAs combinations for 21 Y-STRs by mixture ratio	66
---	----

국문 요약

차세대 염기서열 분석법을 이용한 혼합시료의 Y-STR

유전자형 분석

법의유전학 분야에서 혼합시료에 대한 유전자분석은 꾸준히 연구되어 왔으나 여전히 도전이 많은 영역이다. 현재 혼합시료의 유전자 분석은 짧은 연쇄반복(Short tandem repeat, STR)의 중합효소연쇄반응 후 모세관 전기영동(Capillary electrophoresis, CE) 분석을 통해 관찰된 peak의 수, peak의 상대적 높이에 대한 정보를 바탕으로 하고 있으나 CE 기법의 특성에 따른 한계가 있다. 대용량 염기서열을 제공하는 차세대염기서열 분석법(Next Generation Sequencing; NGS)의 등장으로 법의유전학 영역에서 CE 한계점을 어느 정도 보완하며 혼합 시료 분석에서 활용 될 것으로 기대되고 있다. 그러나 현재까지 혼합 시료 분석에 NGS 기법을 이용하는

것이 실제로 유용한지에 대한 연구도 보고 된 바가 없으며 NGS 기법을 이용하기 위한 체계 및 기준이 필요하다.

이러한 목적을 위해서 본 연구에서는 PowerPlex[®] Y23의 Y 염색체 짧은 연쇄반복(Y chromosomal short tandem repeat, Y-STR) 표지자 및 1개 Y 염색체 단일 염기 다형성(Y chromosome single nucleotide polymorphism, Y-SNP) 을 대상으로 하는 NGS 분석용 다중 중합효소연쇄반응 체계를 바탕으로 한국인 남성 혼합시료를 이용하여 NGS 기법의 유용성을 평가해 보았다. 두 명의 한국인 남성 혼합 시료를 15가지의 비율로 11개 세트를 제조하였으며 20명의 한국인 남성 단일 시료를 포함하여 NGS를 수행하였다. 한국인 남성 단일 시료로 획득한 NGS 자료 분석을 통해 stutter 비율 및 노이즈 비율을 조사하였다. 조사된 노이즈 및 stutter 비율은 부산물 리드를 제거하기 위한 분석적 한계치 비율 및 해석적 한계치 비율로 도입하였으며, 이를 이용하여 혼합 시료의 대립유전자형을 결정하였다.

따라서 본 연구에서는 NGS 기반 Y-STR 체계에서 혼합 시료 분석에 필요한 한계치 비율이 제시되었다. 또한 혼합

시료 내 미량 시료 분석에 있어 각 Y-STR의 민감도가 다를 수 있음을 확인 하였고 대부분의 Y-STR은 1:14의 혼합 비율까지 두 시료의 대립유전자형 결정이 가능하였다. 그러나 NGS 리드 수를 이용한 혼합 비율 추정에는 한계가 있었고, PCR이 수반 된 STR의 NGS 분석에서도 STR의 stutter 생성과 미량 시료 증폭에서의 확률적 효과에 따른 영향을 고려하여 NGS 자료를 해석해야 함을 알 수 있었다.

본 연구를 통하여 NGS를 이용한 혼합 시료 분석에 대한 기초 자료 및 지침을 제시하여 NGS 분석을 혼합 시료 분석 적용에 대한 가능성을 보여주었다고 생각한다. 나아가 법의학 실무에서 접할 수 있는 혼합 시료 분석에 유용하게 적용할 수 있을 것으로 기대된다.

핵심되는 말: DNA 혼합물, Y 염색체 STR, 차세대 염기서열 분석법, 분석 한계치 비율, 해석 한계치 비율, stutter 비율

차세대 염기서열 분석법을 이용한 혼합 시료의 Y-STR
유전자형 분석

< 지도교수 신 경 진 >

연세대학교 대학원 의과학과

권 소 연

I. 서론

짧은연쇄반복(Short tandem repeat, STR)은 사람 DNA 내에 존재하는 특정 염기서열이 반복 되는 부분으로 이는 사람마다 반복서열의 수가 다르기 때문에 개인을 식별할 수 있어 범죄 현장이나 친자 검사에서 널리 이용 되고 있다.¹ 특히 Y 염색체 짧은연쇄반복(Y-chromosomal short tandem repeat, Y-STR)은 상염색체 STR과 다르게 아버지에서 아들로 Y 염색체가 재조합 없이 세대간으로 전달되는 특징이 있어 혈통을 추정할 수 있는 마커로 법과학 영역에서는 남성 자손과의 친자 검사와 인류학에서는 계통유전학적 조사에 유용하게 쓰인다. 특히 남성-

여성이 연루된 성범죄에서는 상염색체 STR과 더불어 남성을 구분하는 데에 추가적인 유전 정보를 제공할 수 있다.²

현재 법과학 영역에서 상염색체 STR 및 Y-STR 분석은 법과학 영역 표준 분석법인 모세관 전기영동(capillary electrophoresis, CE) 분석법을 이용한다. CE 분석법은 단일 시료 분석뿐만 아니라 혼합 시료 분석에서도 사용되는데, 단일 시료 분석과 비교하여 혼합 시료에 대한 해석은 법과학 분야에서의 DNA 분석에서 여전히 도전적인 과제로 남아 있다.³

일반적으로 DNA 혼합물 해석에 대한 원칙은 상염색체 STR 혼합물 해석에 맞춰 논의 되어 왔으며 Clayton 등에 의해 6단계로 이루어진 혼합물에 대한 해석이 SWGDAM(Scientific Working Group on DNA Analysis Methods) 및 ISFG(International Society of Forensic Genetics) 에 의해 권고 되고 있다.⁴⁻⁵ 현재 혼합 시료의 분석은 단일 시료를 이용한 STR 분석과 마찬가지로 CE 기법을 이용하여 분석하며 CE의 결과로 나타난 피크 수와 높이, 높이 비율을 조사하여 혼합 시료가 가지는 유전자형을 결정하고 통계적 도구를 이용하여 혼합 시료 분석을 시도하고 있다.^{3,6} CE 분석법을 이용한 혼합 시료 분석에서는 가장 먼저 혼합 시료가 가지고 있는 공여자의 전체 유전적 정보에 대한 파악이 필요하며⁷ 한 STR에서 관찰 된 peak 수를 조사하여 단일 시료와 혼합

시료를 구분한다. 측정된 피크 높이는 상대적 형광 단위(Relative florescent unit, RFU)를 이용하며 감지 된 DNA의 상대적인 총량을 계산하는데 peak가 매우 낮을 경우 peak인지 아닌지를 판단하기에 어려움이 있다. 따라서 노이즈, pull-up peak와 구분하여 실제 peak를 구분하기 위해 여러 가지 요소를 고려해야 한다. 또한 감지된 peak 높이를 이용하여 한계치 값을 적용하는데 CE 분석에서 사용되는 한계치 값에는 분석 한계치>Analytical threshold, AT)와 확률적 한계치(Stochastic threshold, ST)의 개념을 사용하며 이는 일반적으로 각 50 RFU, 250 RFU을 한계치로 한다.⁸ 하지만 CE 분석법에서는 같은 시료를 이용한 반복적인 실험에서 가변적인 값을 가지며 또한 다른 Genetic analyzer 장비를 사용 함에 따라서도 차이가 발생할 수 있다.⁹ 그리고 CE 분석으로는 혼합 시료 내에서 미량 혼합 된 시료의 참 유전자형을 나타내는 peak가 중합효소연쇄반응(Polymerase chain reaction, PCR)의 부산물로 생기는 stutter와 같은 크기를 가지는 경우도 있기 때문에 명확하게 구분이 되지 않는다는 한계점을 가진다.¹⁰

대용량의 염기서열에 대한 정보를 얻을 수 있는 차세대 염기서열 분석법(Next generation sequencing, NGS)의 등장으로 법과학 영역에서도 STR 분석에 있어 NGS를 적용하여 기존 CE 분석법의

한계를 보완하고 DNA 혼합물 분석에 적용 하려는 시도가 있어 왔다.¹¹⁻¹² 하지만 대부분의 연구들은 NGS 플랫폼에 대한 비교, 단일 시료를 대상으로 STR 영역을 포함하는 염기서열 변이를 보고하는 연구 중심으로 이루어져 왔다. 혼합 시료 분석에 대한 NGS 분석법의 적용은 특정 비율을 가진 혼합 시료 구분을 시도하는 연구만 보고 되어 있으며, NGS 분석법이 혼합 시료 분석에서 CE 분석법의 한계를 보완하며 더 유용한 분석법이 될 수 있다는 전망만 할 뿐 아직 구체적인 기준을 제시하는 연구는 없다. 따라서 혼합 시료 분석에 대한 NGS 분석법이 실질적으로 유용한지 구체적인 기준을 세워 평가 해 볼 필요성이 있다.

STR의 PCR 과정에서는 중합효소의 불이행으로 인하여 생기는 부산물로 stutter가 생성 된다.¹³⁻¹⁴ Stutter는 특정 시료가 가지는 유전자형을 N이라고 하였을 때, 두 반복 단위가 적은 유전자형 N-2, 한 반복수가 적은 유전자형 N-1, 한 반복 수를 더 가지는 유전자형을 N+1이라고 표현하며 일반적으로 세 가지 유형의 stutter를 고려한 다.¹³ 일반적으로 N-1형의 stutter가 가장 많이 관찰 되며, CE 분석에서 stutter는 N-1의 peak 높이/N형의 peak 높이로 계산하여 일반적으로 15%를 넘지 않으면 stutter로 간주하여 유전자형 결정에서 제외되고 그 이상일 경우 고유의 유전자형일 가능성을 고려한다.⁸ Stutter는 관련된 많은 선행

연구¹³⁻¹⁴ 를 통해 CE 기법을 이용하여 각 STR 별로 stutter의 특성이 연구 되어 있다. 하지만 NGS 분석에도 적용하기 위해서는 NGS 기법을 통한 stutter의 특성을 조사하는 것이 필요 하다. 또한 혼합 시료 분석에서 혼합된 유전형을 정확히 판단 하기 위해서는 stutter를 배제하여 정확한 참 유전자형을 파악하는데 각 STR의 stutter 비율 정보는 필수적이다.

본 연구에서는 혼합 시료를 대상으로 Y-STR의 유전자형을 결정 하는데 NGS 기법을 이용하였으며 이를 통해 NGS 분석법의 유효성과 유용성을 평가 하고자 하였다. 다양한 비율로 혼합된 남성-남성 혼합 시료를 대상으로 존재하는 미량 시료의 NGS 분석용 최소 분석 한계치(Analytical threshold)를 제시하고, 유전자형을 결정하기 위한 해석적 한계치(Interpretational threshold)에 대한 기준을 제시하고자 하였다. 뿐만 아니라 21개의 Y-STR 유전자 좌 별 최소 감지 한계치, 분석적 한계치를 조사해 보고자 하였다. 그리고 NGS를 통해 생성된 리드 수를 역으로 계산하여 혼합 비율을 추정해 보고자 한다. 또한 NGS 기법을 통한 각 Y-STR stutter의 특성을 조사하고, 혼합시료 분석에 반영하고자 하였다. 마지막으로 NGS 분석법에서 관찰되는 염기서열변이 정보를 이용하여 혼합 시료를 분석에서 공여자 구분 가능성을 확인 해 보고자 하였다.

II. 재료 및 방법

1. 연구대상

본 연구에서는 Y-STR을 대상으로 혼합 시료에 대한 NGS 분석법의 유효성과 효용성을 평가 하고자 하였다. 혼합 시료에 대한 참고 자료를 만들기 위해 남성 표준 시료 2800M (Promega, Madison, WI, USA), 9948 (Promega)를 이용하여 여러 비율로 혼합하여 사용하였다. 또한 혼합 시료 분석 모델을 만들기 위하여 한국인 집단을 대상으로 혈연 관계가 없는 한국인 20명을 표본으로 무작위 선정하고, 이들로부터 2.0 ml 주사기를 이용하여 2.0 ml의 혈액을 채취하여 즉시 EDTA 용액이 포함된 혈액보관 튜브에 넣고 DNA를 추출할 때까지 4℃ 냉장고에 보관하였다. 본 연구는 연세대학교 의료원 세브란스병원 연구심의위원회와 유전자연구 생명윤리심의위원회 승인을 받아 수행 되었다(승인번호 4-2015-1068).

2. DNA 추출 및 정량

가. 한국인 시료의 추출 및 정량

보관된 시료는 QIAamp[®] DNA Mini Kit (QIAGEN, Hilden, Germany)를 이용하여 제조사의 지시에 따라 DNA를 추출하였다. 추출된 시료는 NanoDrop 1000 spectrophotometer (Thermo Fisher scientific, Waltham, MA, USA)를 이용하여 정량을 측정 후 -20°C 에 보관하였다. NanoDrop 1000 spectrophotometer (Thermo Fisher scientific)을 이용하여 측정된 시료는 측정된 농도를 바탕으로 $10.0\text{ng}/\mu\text{l}$ 로 희석한 후 정확한 농도를 이용하여 혼합 시료를 제조하고자 Quantifiler[®] Trio DNA Quantification Kit (Thermo Fisher scientific)를 이용하여 2차적으로 정량을 측정하였으며 정량을 측정한 후 -20°C 에 보관하였다.

3. 혼합 시료 준비

가. 표준시료를 이용한 혼합 시료의 준비

본 연구는 혼합 시료 분석을 위해 NGS 기법을 적용해 보고 CE 기법과 비교하고자 비율이 다른 혼합 시료를 준비한다. 혼합 시료는 남성 표준시료 2800M (Promega), 9948 (Promega)를 사용하였고 정확한 비율로 혼합된 시료를 얻고자 Quantifiler[®] Trio DNA Quantification Kit (Thermo Fisher scientific)를 이용하여 제조사의 지시에 따라 정량하였다. 측정된 정량 값을 이용하여 1:1, 1:3, 1:6, 1:9, 1:14, 1:19, 1:29, 1:49, 1:99의 비율로 혼합하였다. 각 비율의 혼합 시료에서 미량 시료는 500 pg/μl, 250 pg/μl, 143 pg/μl, 100 pg/μl, 67 pg/μl, 50 pg/μl, 33 pg/μl, 20 pg/μl, 10 pg/μl 의 농도를 가지며, 혼합된 DNA의 총 농도는 1 ng/μl가 되도록 하였다.

나. 한국인 남성 시료를 이용한 혼합 시료의 준비

(1) PowerPlex[®] Y23를 이용한 한국인 남성 시료의 유전자형 확보

혈연관계가 없는 20명의 한국인 남성시료 중 두 명의 시료를 혼합 하여 DYS385를 제외한 21개의 Y-STR을 고르게 분석 하기 위하여 먼저 PowerPlex[®] Y23를 이용하여 대립유전자형을 확보하였다.

(2) 두 명의 한국인 남성 시료를 이용한 혼합 시료 조합 생성

PowerPlex[®] Y23를 이용하여 확보한 대립유전자형을 바탕으로 혼합 시료를 제조하였다. 일차적으로 20명의 시료를 이용하여 두 명의 시료를 단 한번씩 혼합하면 190개 조합이 형성된다. 그 중 stutter의 영향을 덜 받는 두 대립유전자를 구성하기 위하여 두 시료의 대립유전자 형의 차이가 2이상 나는 Y-STR 유전자 좌의 개수가 많을수록 우선 순위로 배치하여 혼합 시료 세트를 선택하였다. 우선 순위를 통해 선택한 혼합 시료 세트 내 조합에서 분석에 포함되지 않는 유전자 좌가 생길 경우에는 우선 순위가

낮은 혼합 시료 세트를 제외한다. 그 다음 분석이 필요한 유전자좌가 포함된 세트를 포함시키되 같은 조건의 세트 중에서는 우선순위에 있는 세트를 선택 한다. 최종적으로 DYS385, Y-M175 표지자를 제외한 21개 Y-STR이 모두 고르게 혼합되도록 제조하였으며 21개 Y-STR 모두에서 두 남성의 대립유전자 형이 같은 경우가 적어도 한 번 포함되고, 두 대립유전자형 차이가 1인 대립유전자 형이 복수 개, 두 대립유전자형 차이가 2인 대립유전자 형이 복수 개 이상 포함되도록 하였다. DYS392와 DYS481은 두 대립유전자형 차이가 3인 대립유전자형이 복수 개 이상 포함되도록 하였다.

혼합 시료에 사용되는 두 명의 한국인 남성 시료에 대하여 정확한 비율로 혼합된 시료를 얻고자 Quantifiler[®] Trio DNA Quantification Kit (Thermo Fisher scientific)로 측정된 정량 값을 이용하였다. 또한 Quantifiler[®] Trio DNA Quantification Kit (Thermo Fisher scientific)로 측정된 정량 값을 검증하기 위해 선택된 조합에 사용 되는 두 남성 DNA가 1:1로 혼합된 혼합물을 CE 분석법을 이용하여 peak 균형 맞춤을 통해 농도의 정확성을 검증하였다.

(3) 두 명의 한국인 남성 혼합 시료 비율

표준 남성 혼합 시료의 결과를 참고하여 1:99 비율을 제외한 1:1, 1:3, 1:6, 1:9, 1:14, 1:19, 1:29, 1:49비율을 선정하였다. 또한 3:1, 6:1, 9:1, 14:1, 19:1, 29:1, 49:1의 비율을 포함하여 15개의 혼합 비율을 선정하였다.

(4) 두 명의 한국인 남성 시료를 이용한 혼합 시료 제조

차세대 염기 서열을 이용하여 혼합 시료의 분석 모델을 제시하기 위하여 조합에 사용되는 모든 한국인 남성 시료를 각 2 ng/ μ l, 1 ng/ μ l, 200 pg/ μ l, 50 pg/ μ l의 농도로 희석하여 1:1, 1:3, 1:6, 1:9, 1:14, 1:19, 1:29, 1:49비율의 혼합 시료를 제조하는 데 사용하였다. 또한 두 남성 DNA A, B 혼합물에 대하여 A:B 1:1, 1:3, 1:6, 1:9, 1:14, 1:19, 1:29, 1:49 비율 및 3:1, 6:1, 9:1, 14:1, 19:1, 29:1, 49:1의 비율로 혼합하였다. 각 비율의 미량 시료는 500 pg/ μ l, 250 pg/ μ l, 143 pg/ μ l, 100 pg/ μ l, 67 pg/ μ l, 50 pg/ μ l, 33 pg/ μ l, 20 pg/ μ l의 농도를 가지며 혼합된 DNA의 총 농도는 1 ng/ μ l가 되도록 하였다.

4. 연구 대상 Y-STR 표지자

현재 법과학 영역에서 Y-STR 분석 키트로 널리 사용되는 제품 중 가장 많은 Y 염색체 marker를 포함하는 CE기반의 상용 PowerPlex[®] Y23 Kit (Promega)에서 분석되는 23개의 Y-STR (DYS19, DYS385, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS481, DYS533, DYS549, DYS570, DYS576, DYS635, DYS643, Y-GATAH4) 중 두 개 이상의 대립 유전자형을 갖는 DYS385를 제외한 21 개의 Y-STR을 분석 대상으로 하였다.

5. NGS 분석용 다중중합효소연쇄반응 체계

선행 연구에서 구축된 24개 Y 염색체 표지자를 동시에 분석할 수 있는 NGS 분석용 다중중합효소연쇄반응 체계를 사용하였다. 이 체계는 Nextera XT library preparation kit에서 사용한 indexing PCR 방법을 차용하여 두 단계의 PCR로 구성되어 있으며 본 연구에서도 기존에 구축된 NGS 분석 체계를 사용한다. 사용한 primer 염기서열과 농도는 표 1에 표시하였다.

표 1. PCR primer sequences and final concentrations of Y-STR multiplex PCR systems for the first step of library preparation

Locus		Primer sequences* (5'→3')	Conc. (μM)
M175	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtgatttaaactctctgaatcaggc	0.35
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtgataaccttttttctactgataccttt	0.35
DYS19	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGccactctgggtaagagagagtgt	1.5
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGttcactatgactactgagtttctgtt	1.5
DYS392	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtaaactaccaatcccattcct	0.75
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtttgtattttaaaagccaagaagg	0.75
DYS393	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtgtgtcttctacttgttcaa	0.45
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGaaactcaagtccaaaaatgagg	0.45
DYS438	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtggggaatagttgaacggtaa	0.55
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGgcaacaagagtgaactccatt	0.55
DYS448	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGagaaaaggagatagagacatgga	0.6
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtgcccggtctggaaatttat	0.6
DYS570	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGagctgaaatgcagatattccc	0.35
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGgctgtgtcctccaagtctct	0.35
DYS643	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtgccctggtaaactactgtgc	0.6
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtccccccaaaatttctactga	0.6

DYS385	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGggaaggagaaagaaagtaaaaaagaa	2.0
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGttccaattacatagtcctcctttct	2.0
DYS439	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGacataggtggagacagatagatga	0.35
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGggcttggaattcttttaccca	0.35
DYS458	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGgcaacaggaatgaaactcaa	0.55
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGcccaaagtctggcattacaa	0.55
DYS481	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGcagaaggttgaagactcaa	0.33
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGaggaatgtggctaacgctgt	0.33
DYS533	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGcttctacctatcatctttctagc	0.6
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtcagttcttaactcaaccaaaaa	0.6
DYS549	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGgtccccttttcatttgtga	0.9
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGgcaattaggtaggtaaagaggaaga	0.9
DYS635	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtgcttctcactttgcatagaa	0.5
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGgtggaaccagcccaaatatc	0.5

	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG <u>cc</u> aaactctcatctgtattatctatgta	0.2
DYS389	R	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG <u>cc</u> aaactctcatctgtattatctatgtg	1.5
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGgatagattgatagaggaggga	1.5
DYS390	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGgtgtataactcagaacaagaaaga	0.4
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtgcattttggtacccata	0.4
DYS391	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGttcaatcatacacccatatctgtc	0.3
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtgaagcaattgcatagag	0.3
DYS437	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGgactatggcgtgagtgcat	0.4
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGgataagtagatagacattcacaga	0.4
DYS456	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGctgttggtggaccttgata	0.5
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGactcagcccaaaacttctaaa	0.5
DYS576	F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGgcgtattgtcttggctttt	0.35
	R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGcatagcaagacctcatctgaa	0.35

*The Y chromosome marker-specific sequences were shown in small letter and read sequences were shown in capital letter. Platform-specific sequences for ligating indices were indicated with underline.

6. NGS 수행

가. 표준 남성시료를 이용한 혼합 시료의 NGS 수행

(1) 표준 남성시료를 이용한 혼합 시료의 라이브러리(library) 제작

혼합 시료를 이용한 NGS 참고 자료를 생성하기 위하여 표준 남성시료 2800M과 9948을 이용한 혼합 시료에 대한 NGS를 수행하였다. 이를 위해 Nextera XT library preparation kit에서 사용한 indexing PCR 방법을 차용하여 두 단계의 PCR로 이루어지는 자체 구축한 NGS 라이브러리 제작법을 사용하였다. 두 단계의 PCR 과정 중 1차 중합효소연쇄반응은 표 1의 고안된 Nextera primer를 이용하여 Y-STR 영역을 포함하는 증폭 산물을 얻는다. 표 1의 최종 농도에 해당하는 primer와 1 ng의 주형 DNA 및 5.0 U AmpliTaq Gold DNA Polymerase (Applied Biosystems)가 포함되도록 하였고, 2 μ l의 Gold ST*R 10x Buffer (Promega)를 추가하였다. 혼합물의 최종 부피는 멸균된 3차 증류수로 20 μ l가 되도록 맞추어 첨가한 후 잘 섞어 주었다. 중합효소연쇄반응 전 혼합물은 Veriti[®] 96-Well Thermal Cycler (Applied Biosystems)에

장착하여, 95 °C에서 11분간 변성시키고 94 °C에서 20초, 60 °C에서 1분 30초, 72°C에서 30초의 조건으로 30회 중합반응 후 최종적으로 72 °C에서 7분간 반응시킨 후 4 °C에서 잠시 보관하였다.

다음으로 2차 중합효소연쇄반응에서는 증폭을 수행하고자 혼합물에 Illumina사의 Nextera XT index kit v2를 사용하여 다중인식자 (Index) 1, 2를 각 증폭산물 양 쪽에 붙여 각 시료가 분류 될 수 있도록 어댑터를 부착한다. 1차 중합효소연쇄반응에서 증폭된 산물을 3차 증류수를 이용하여 증류수 99 μ l에 증폭산물 1 μ l를 혼합하여 1/100의 비율로 희석 한 증폭 산물을 준비하였다. 2차 중합효소연쇄반응을 위하여 각 시료 마다 다른 조합의 2 μ l의 Index 1, 2가 포함되도록 하였고, 1/100으로 희석한 1 μ l의 1차 증폭산물 그리고 2.0 U AmpliTaq Gold DNA Polymerase (Applied Biosystems)와 2 μ l의 Gold ST*R 10x Buffer (Promega)를 추가하였다. 혼합물의 최종 부피는 멸균된 3차 증류수로 20 μ l가 되도록 맞추어 첨가한 후 잘 섞어주었다. 중합효소연쇄반응 전 혼합물은 Veriti[®] 96-Well Thermal Cycler (Applied Biosystems)에 장착하여, 95 °C에서 15분간 변성시키고 94 °C에서 20초, 59 °C에서 1분 30초, 72 °C에서 30초의 조건으로 17회 중합반응

후 최종적으로 60 °C에서 7분간 반응시킨 후 4 °C에서 잠시 보관하였다.

두 단계의 PCR을 통해 만들어진 증폭산물을 확인하기 위해 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA)를 통해 확인한다. 200bp ~ 400bp 크기의 증폭산물의 존재 유무를 확인하고, 측정된 농도를 확인하였다. 각 시료를 모두 5 ng/ μ l의 농도로 표준화 시켜 준비하여, 12개씩 5.0 μ l로 혼합하여 준비하였다.

(2) 표준 남성 시료를 이용한 혼합 시료 라이브러리의 정제 및 정도 관리(quality control)

9가지 비율로 혼합된 두 표준 남성시료 혼합물에 대하여 어댑터와 증폭산물의 크기를 고려하여 원하는 라이브러리만을 얻고, Primer 이량체 및 PCR을 통해 생성된 부산물을 제거하기 위해 Ampure Bead (Beckman Coulter, Brea, CA, USA)를 이용한 정제를 수행한다. 150 bp 미만의 작은 절편을 제거할 수 있도록 증폭산물의 양에 비하여 1.2x의 Ampure Bead (Beckman Coulter)를 이용하여 제조사의 지시에 따라 정제하였다. Ampure

Bead (Beckman Coulter)의 비율은 관련 연구 논문¹⁷을 참고하여 사용하였다.

제작된 라이브러리를 정량 하기 위해 KAPA Library Quantification Kit (KAPA Biosystems, Woburn, MA, USA)를 이용하여 제조사의 지시에 따라 라이브러리 정량을 수행하였다. 정량 값에 따라 10 nM의 농도로 동량 섞어 준비한다.

(3) NGS 장비를 이용한 표준 남성 시료 혼합물 염기서열 분석

10 nM의 농도로 동량 섞인 단일 시료 및 혼합시료 라이브러리를 대상으로 250 bp씩 양 끝을 읽는 (Paired-end read) Illumina Miseq Reagent v2 Kit (Illumina)를 이용하여 제조사의 지시에 따라 분석하였다.

나. 한국인 남성 시료를 이용한 혼합 시료의 NGS 수행

(1) 한국인 남성 시료를 이용한 혼합 시료의 라이브러리(library) 제작

한국인 남성시료를 이용하여 제조한 15가지 비율의 165개 혼합시료 및 20개 한국인 남성 단일시료에 대한 NGS를 수행하였다. II-6-가-(1)에서 설명한 표준 남성 혼합시료에 대한 라이브러리 제작에 사용한 체계와 같은 체계를 사용하였다. 그러나 표준 남성시료의 NGS를 수행결과를 참고하여 라이브러리 제작 프로토콜 간소화 및 stutter의 영향을 최소화 하기 위해 PCR cycle 조건을 조정하여 시행하였다. 두 단계의 PCR 과정 중 1단계 PCR은 II-6-가-(1)에서 사용한 조건에서 PCR cycle을 30회에서 27회로 3회 줄여 시행하였다. 1차 PCR에서 증폭된 산물은 3차 증류수를 이용하여 증류수 9 μ l에 증폭산물 1 μ l를 혼합하여 1/10의 비율로 희석 한 증폭 산물을 준비하였다. 다음으로 2차 PCR에서도 II-6-가-(1)에서 사용한 조건을 사용하였으나 PCR cycle을 17회에서 15회로 2회 줄여 수행하였다.

두 단계의 PCR을 통해 만들어진 증폭산물을 확인하기 위해 2100 bioanalyzer (Agilent Technologies)를 통해 확인한다. 200bp

~ 400bp 크기의 증폭산물의 존재 유무를 확인하고, 측정된 농도를 확인하였다. 각 시료를 모두 4 ng/ μ l의 농도로 표준화 시켜 준비하여, 185개의 시료에서 30개씩 4.0 μ l로 혼합한 1개 세트와 31개씩 4.0 μ l로 혼합한 5개 세트 모두 6개의 세트를 준비한다.

(2) 한국인 남성시료를 이용한 혼합시료의 라이브러리의 정제 및 정도 관리(Quality control)

한국인 남성시료 혼합물을 이용해 제작된 라이브러리를 정제하기 위해서 II-6-가-(2) 에서 사용한 AMPure Bead (Beckman Coulter)를 이용한 정제 및 KAPA Library Quantification Kit (KAPA Biosystems)를 이용하여 제작된 라이브러리를 정량하였다. 정량 값에 따라 6개 세트를 10 nM의 농도로 동량 섞어 185개를 모두 섞어 준비한다.

7. NGS 결과를 이용한 Y-STR 유전자형 결정 및 자료 분석

(1) 표준 남성 시료 NGS 결과를 이용한 Y-STR 유전자형 결정 및 자료 분석

NGS 플랫폼으로부터 시퀀싱 후 얻어지는 FASTQ 형식으로 저장된 표준 남성시료 자료는 Warshauer 등¹⁸ 이 법과학 영역에서 주로 사용되는 STR 대립유전자형을 결정하기 위해 개발한 STRait Razor 프로그램을 사용하여 NGS 분석 결과를 우선적으로 확인하였다. STRait Razor 프로그램을 통해 각 STR의 유전자형을 결정하고 각 유전자형에 분석된 리드 수를 알 수 있었다. 또한 좀 더 세부적인 염기 서열정보를 관찰하기 위하여 관련 연구¹⁹에서 사용한 BWA program,²⁰ Bowtie2 program²¹ 및 SAMtools,²² VarScan²³을 이용한 NGS 자료 분석도 병행하여 Y-STR 유전자형과 염기서열 변이를 조사하였다.

BWA program 및 Bowtie2 program을 이용하기 위해서 필요한 참조서열(reference sequence) 자료는 human genome GRCh37/hg19에 존재하는 모든 STR 유전자 좌 및 STRbase (<http://www.cstl.nist.gov/biotech/strbase>)를 통해 현재까지

알려진 Y-STR 유전자 좌의 반복영역 및 염기서열 정보를 모두 포함하여 제작하였다. 또한 Y-STR의 주변 영역 (flanking region)의 염기서열 변이를 조사하기 위하여 500~550bp를 포함하여 참조 서열을 제작하였으며 주변 영역을 포함함으로써 어떠한 primer 조합을 통해 얻어진 리드도 참조서열 자료와 정렬이 이루어 질 수 있도록 하였다. 위에 설명한 모든 정보를 포함한 참조 서열은 Microsoft® Excel® program을 이용하여 제작하였다.

FASTQ 형식의 NGS 시퀀싱 자료는 제작된 참조서열을 이용하고 BWA program 및 Bowtie2 program을 이용하여 분석하였다. 분석된 자료는 SAM 형식으로 제공되고 C#언어로 코딩 한 프로그램을 이용하여 HTML 형식으로 변환하여 결과를 확인하였다.

또한 각 Y-STR 유전자에 대한 염기서열을 바탕으로 한 반복단위와 염기서열 변이 확인은 SAM 형식의 파일을 SAMtools²⁰을 이용하여 BAM 형식의 파일로 변환하여 Integrative Genomics Viewer (IGV)²⁴ 프로그램을 통해 확인하였다. 이러한 NGS 자료 분석 과정은 연세유전체센터의 연구자와 긴밀하게 협의하여 진행하였다.

(2) 한국인 남성 시료 NGS 결과를 이용한 Y-STR 유전자형 결정
및 자료 분석

NGS 플랫폼으로부터 시퀀싱 후 얻어지는 FASTQ 형식으로 저장된 한국인 남성 시료 자료는 일차적으로 STRait Razor¹⁸ 프로그램을 사용하여 NGS 분석 결과를 우선적으로 확인하였다. 세부적인 분석을 위해서는 Java언어를 이용하여 만든 프로그램을 이용하였다. 이 분석법은 참조서열을 사용하지 않고 사용된 primer의 염기서열 위치를 바탕으로 5' 방향의 주변 영역 염기서열의 위치, 3' 방향의 주변 영역 염기서열의 위치, 각 유전자좌의 반복 염기서열을 이용하여 config파일을 제작한다. Config 파일에 입력된 정보를 이용하여 먼저 23개 Y-STR 및 Y-M175에 해당되는 리드를 선별한다. 선별된 리드는 각 Y-STR 각각의 특정 반복 염기서열을 이용하여 반복 수를 도출해 낸다. 유전자형 결과는 text 파일로 제공되며 Microsoft[®] Excel[®] program을 이용하여 자료를 확인 및 분석하였다.

Microsoft[®] Excel[®] program을 이용하여 size 기반의 유전자형 분석을 진행하였다. 각 Y-STR의 대립유전자형 범위를 고려하여 범위를 벗어나 결정된 유전자형은 제외하였다.

8. NGS 및 CE 기법을 이용한 남성 표준시료 분석 비교

분석 가능한 최소한계치를 조사하기 위해 사용한 2800M (Promega)과 9948 (Promega)의 모든 비율(1:1, 1:3, 1:6, 1:9, 1:14, 1:19, 1:29, 1:49, 1:99)의 혼합 시료에 대해 NGS 분석 결과 및 CE 결과에서 Y-STR에 대한 프로파일을 비교하여 Y-STR 유전자형을 이중 확인 한다. CE 결과에서는 분석적 한계치 100 RFU를 적용하고, 이론적으로 각 비율에서 관찰 가능한 peak height를 기준으로 관찰된 유전자형을 조사하였다. NGS분석을 통한 결과에서도 CE 결과 분석과 같은 방식으로 분석적 한계 리드 수를 적용하고 각 비율에서 관찰 가능한 리드 수를 기준으로 관찰된 유전자형을 조사하였다. 두 분석법을 통해 각 비율에서 관찰된 유전자형을 비교하였다.

9. NGS기반 Y-STR 분석 체계에서의 Stutter 특성 조사

NGS 분석을 통해 확인된 20개 한국인 남성 단일시료의 21개 Y-STR (Y-M175와 DYS385ab를 제외)에서 stutter 비율을 조사하였다. 주 대립유전자형을 N, 한 반복 수 적은 대립유전자형을 N-1, 두 반복 수 적은 대립유전자형을 N-2, 세 반복 수 적은 대립유전자형을 N-3, 한 반복 수 많은 대립유전자형을 N+1이라고 지정하였다. 3염기 반복 단위를 가지는 DYS392, DYS481은 주 대립유전자 형이 가지는 리드 수와 비교하여 N-1, N-2, N-3, N+1이 차지하는 리드수의 비율을 계산하였으며 식1-1, 식1-2, 식1-3, 식1-4에 따라 계산하였다. DYS392, DYS481을 제외한 19개 Y-STR stutter 비율은 주 대립유전자 형이 가지는 리드 수와 비교하여 N-1, N-2, N+1이 차지하는 리드수의 비율을 계산하였으며 식1-1, 식1-2, 식1-4에 따라 계산하였다. 3염기 반복 단위, 4염기 반복 단위, 5염기 반복 단위, 6염기 반복 단위로 구분하여 stutter의 평균 비율을 조사하였다. 또한 단순 반복 서열과 복합 반복 서열을 가지는 Y-STR을 구분하여 조사하였다.

$$N - 1 \text{ stutter 비율}(\%) = \frac{\text{한 반복 수 적은 대립유전자형 } (N - 1) \text{ 리드수}}{\text{주 대립유전자형 } (N) \text{ 리드수}}$$

식 1-1. 한 반복 수 적은 대립유전자 형 stutter 비율

$$N - 2 \text{ stutter 비율}(\%) = \frac{\text{두 반복 수 적은 대립유전자형 } (N - 2) \text{ 리드수}}{\text{주 대립유전자형 } (N) \text{ 리드수}}$$

식1-2. 두 반복 수 적은 대립유전자 형 stutter 비율

$$N - 3 \text{ stutter 비율}(\%) = \frac{\text{세 반복 수 적은 대립유전자형 } (N - 3) \text{ 리드수}}{\text{주 대립유전자형 } (N) \text{ 리드수}}$$

식1-3. 세 반복 수 적은 대립유전자 형 stutter 비율

$$N + 1 \text{ stutter 비율}(\%) = \frac{\text{한 반복 수 많은 대립유전자형 } (N + 1) \text{ 리드수}}{\text{주 대립유전자형 } (N) \text{ 리드수}}$$

식1-4. 한 반복 수 많은 대립유전자 형 stutter 비율

10. NGS 기반 Y-STR 분석 체계에서의 발생하는 노이즈 조사

Java 언어로 제작된 STR 분석 프로그램을 이용하여 분석한 한국인 남성 시료 Y-STR 유전자형 결과를 바탕으로 21개 Y-STR (Y-M175와 DYS385ab를 제외)에서 노이즈 비율을 조사하였다. 노이즈는 결정된 유전자형 및 stutter에 해당하는 리드를 제외한 리드로 정의한다.²⁶ 노이즈 비율은 각 Y-STR에 대한 총 리드 수에 대하여 결정된 대립유전자형 및 stutter가 차지하는 리드 수를 제외한 리드 수 비율을 말하며 식 2에 따라 계산하였다.

노이즈 비율(%)

$$= \frac{Y-STR \text{의 총 리드수} - (\text{대립유전자형 리드수} + \text{stutter 리드수})}{Y-STR \text{의 총 리드수}}$$

식2. 21개 Y-STR에서 노이즈가 차지하는 비율

III. 결과

1. 표준 남성시료 혼합물 결과 분석

가. 표준 남성시료 혼합물을 이용한 CE 결과

Powerplex[®] Y23 키트를 이용하여 남성 표준시료 혼합시료는 9948(Promega): 2800M (Promega) 로 1:1, 1:3, 1:6, 1:9, 1:14, 1:19, 1:29, 1:49, 1:99의 비율로 혼합하여 PowerPlex[®] Y23 키트를 이용하여 CE 분석하였다. 분석적 한계치를 100 RFU로 적용하여 그 이하의 peak height를 가지는 유전자 형은 분석에서 제외하였으며 100 RFU 이상인 peak 중에 stutter는 제거하였다. 2800M과 9948은 23개의 Y-STR 중 11개 Y-STR(DYS385, DYS389I, DYS437, DYS438, DYS458, DYS481, DYS570, DYS576, DYS635, DYS643, Y-GATAH4)에서 두 개 이상의 혼합 시료 유전자형이 관찰 되었다 (표 2). 1:19 혼합 비율에서부터 유전자 소실이 관찰 되었다.

표 2. Result of Y-STR genotyping in various range of mixture ratio of control DNAs from capillary electrophoresis

Y-STR	CE result								
	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	1:99
DYS385	11, 13, 14, 16	11, 13, 14, 16	11, 13, 14, 16	11, 13, 14, 16	11, 13, 14, 16	11, 13, 14, 16	11, 13, 14, 16	13, 16	13, 16
DYS389I	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	14
DYS437	14, 15	14, 15	14, 15	14, 15	14, 15	14	14, 15	14, 15	14
DYS438	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9	9	9
DYS458	17, 18	17, 18	17, 18	17, 18	17, 18	17	17	17, 18	17
DYS481	22, 24	22, 24	22, 24	22, 24	21, 22	22, 24	21, 22	21, 22	21, 22
DYS570	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17	17
DYS576	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	18	18
DYS635	21, 23	21, 23	21, 23	21, 23	21, 23	21	21, 23	21	21, 24
DYS643	10, 11	10, 11	10, 11	10, 11	10, 11	10	10	10	10
Y-GATAH4	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11	11	11

나. Illumina Miseq 분석용 표준 남성시료 라이브러리 제작 결과

본 연구에서 사용한 단일 시료인 표준 DNA 2800M, 9948, 그리고 1:1, 1:3, 1:6, 1:9, 1:14, 1:19, 1:29, 1:49, 1:99의 비율로 혼합된 DNA 시료의 증폭산물로부터 두 단계의 PCR을 통해 라이브러리 제작을 수행하였으며 2100 bioanalyzer (Agilent Technologies)로 라이브러리를 확인 한 결과 200 bp ~ 400bp 내에 작은 절편들이 성공적으로 생성되었음을 확인하였다.

다. NGS 장비를 이용한 시퀀싱 결과

제작된 라이브러리는 Miseq 장비를 이용하여 시퀀싱을 수행 한 후 획득한 리드 수는 총 2,870,208개 였다. 단일 시료인 9948, 2800M은 각 각 231,540, 235,024 개의 리드수를 얻었다. 1:1 혼합시료는 188,744개, 1:3 혼합시료는 214,230개, 1:6 혼합시료는 253,840개, 1:9 혼합시료는 249,054개, 1:14 혼합시료는 235,266개, 1:19 혼합시료는 241,048개, 1:29 혼합시료는 247,898개, 1:49 혼합시료는 239,966개, 1:99 혼합시료는 239,966개의 리드를 얻었다.

라. 표준 남성 시료를 이용한 NGS 결과 분석

23개 Y-STR 대립유전자 좌에서 리드 수 100을 최소 한계치로 하여 그 이상의 리드 수를 가지는 대립 유전자형 중에 stutter는 비율을 적용하여 제거하였다.²³ 적용 된 stutter 비율은 250명의 한국인 남성 단일시료 결과를 바탕으로 계산된 비율을 사용하였다. 12개의 Y-STR에서 두 개 이상의 대립유전자 형을 관찰하였다 (표 3). DYS392에서는 2800M과 9948이 동일한 유전자 형을 갖는 Y-STR임에도 불구하고, 대립유전자형 획득이 관찰되어 두 개 이상의

대립유전자 형이 관찰 되었다. 1:3 혼합 비율에서 대립유전자 획득이 관찰 되었으며, 1:29 혼합 비율에서부터 대부분의 Y-STR에서 대립유전자형 소실이 관찰 되었다.

표 3. Allele determinations in various ranges of mixture ratio of control DNAs from next generation sequencing

Y-STR	NGS result									
	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	1:99	
DYS385	11, 13, 14, 16	11, 13, 14, 16	11, 13, 14, 16	11, 13, 14, 16	11, 13, 14, 16	11, 13, 14, 16	11, 13, 14, 16	13, 16	11, 13, 14, 16	13, 16
DYS389I	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14
DYS392*	13	<u>10, 12, 13,</u>	13	13	13	13	13	13	13	13
DYS437	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14
DYS438	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9
DYS458	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18
DYS481	22, 24	22, 24	22, 24	22, 24	22, 24	22, 24	22, 24	22, 24	22, 24	22, 24
DYS570	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17	17
DYS576	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	18	18
DYS635	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23
DYS643	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11
Y-GATAH4	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11	11, 12

*Y-STRs observed same allele determinations in both 9948 and 2800M. Allele-drop-in was indicated with underline.

2. 한국인 남성 혼합시료 결과 분석

가. PowerPlex[®] Y23를 이용한 한국인 남성 시료의 유전자형 결과

두 명의 한국인 남성 시료 혼합물을 제조 하기 위하여 먼저 20명의 한국인 남성 시료를 PowerPlex[®] Y23를 이용하여 확보하였고 표 4에 정리하였다.

표 4. Result of Y-STR genotyping with 20 unrelated Korean males using PowerPlex® Y23

	DYS 19	DYS 385	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 437	DYS 438	DYS 439	DYS 448	DYS 456	DYS 458	DYS 481	DYS 533	DYS 549	DYS 570	DYS 576	DYS 635	DYS 643	Y GATA H4
NGS001	15	10, 17	13	30	23	10	11	13	14	10	11	20	15	16	24	12	11	17	17	21	9	10
NGS002	14	13, 14, 19	12	28	23	10	14	12	15	11	12	19	15	16	22	11	12	18	19	21	11	12
NGS003	15	11, 17	12	29	23	10	11	14	14	10	11	19	15	18	24	12	12	17	18	21	9	11
NGS004	15	10, 19	15	30	23	10	13	13	14	13	12	18	15	17	23	11	11	19	18	21	13	11
NGS005	15	10, 20	14	30	22	10	13	13	14	13	12	18	15	17	21	11	13	19	19	20	12	12
NGS006	17	14, 14	12	27	25	10	13	12	15	10	13	19	14	17	26	10	13	19.3	17	23	10	12
NGS007	14	11, 13	14	31	23	11	16	15	14	11	10	19	14	17	20	11	12	16	16	20	11	12
NGS008	15	13, 17	13	29	24	11	14	13	14	10	12	18	14	21	23	10	12	18	18	21	12	10
NGS009	16	10, 18	13	28	23	10	13	13	14	13	12	18	14	17	23	11	12	17	18	21	12	11
NGS010	16	10, 18	14	29	23	10	13	13	14	13	12	18	15	17	23	11	12	18	17	21	13	11
NGS011	15	12, 17	14	30	24	10	13	14	14	10	11	18	15	15	23	11	12	21	18	22	12	12
NGS012	17	12, 18	12	28	24	10	13	12	14	10	12	20	15	18	23	11	12	18	17	23	10	12
NGS013	16	15, 21	12	28	24	10	13	12	15	10	11	20	15	16	23	11	13	18	18	22	11	12
NGS014	15	10, 21	15	30	22	10	13	13	14	13	14	18	15	20	21	11	12	20	19	21	12	12
NGS015	16	10, 15	14	29	23	10	13	13	14	13	11	18	15	17	23	11	12	18	18	21	13	11
NGS016	15	10, 19	14	28	23	10	13	13	14	13	12	18	15	16	24	11	12	18	17	21	13	11
NGS017	16	10, 18	14	28	23	10	13	13	14	13	12	18	15	18	23	11	12	17	17	21	13	11
NGS018	16	15, 19	12	30	24	10	13	12	15	10	11	20	14	16	24	11	14	19	18	21	11	13
NGS019	15	13, 17	12	30	23	10	12	13	15	10	11	19	15	15	23	10	12	19	19	21	11	12
NGS020	15	10, 18	14	29	23	10	13	13	14	13	12	18	15	17	23	11	13	18	18	21	13	11

나. 두 명의 한국인 남성을 이용한 혼합 시료 제조 결과

(1) 두 명의 한국인 남성을 이용한 시료 조합

21개 Y-STR을 고르게 분석하기 위하여 PowerPlex[®] Y23 결과를 이용하여 모든 Y-STR이 고른 혼합 횟수를 포함하게 하였으며 표 5에 정리하였다. 최종적으로 DYS385, Y-M175 표지자를 제외한 21개 Y-STR에서 두 남성의 같은 대립유전자 형이 적어도 한 번 혼합되고, 두 대립유전자형 차이가 1인 대립유전자 형이 세 개 이상, 두 대립유전자형 차이가 2 이상인 대립유전자 형이 세 개 이상 포함 되도록 하였다. 일반적으로 stutter 비율이 높은 3염기 반복 수를 가지는 DYS481과 DYS392는 두 대립유전자형 차이가 3 이상인 대립유전자 형이 세 개 이상 포함 되도록 하였다. 20명의 한국인 남성 PowerPlex[®] Y23 결과에서 DYS391은 10, 11 DYS437은 14, 15 DYS456은 14, 15의 각 두 개의 대립유전자형만 관찰되어 두 대립유전자형 차이가 나는 조합은 포함되지 않았다.

결과적으로 20명의 한국인 남성 중 13명의 한국인 남성(NGS001, NGS002, NGS003, NGS005, NGS006, NGS007, NGS008, NGS010, NGS011, NGS012, NGS013, NGS018, NGS019)이 혼합 시료

제조에 사용되었으며 두 명의 남성 시료가 섞인 11개 세트의 조합을 사용하였다.

표 5. The number of repeat difference of 21 Y-STRs with 11 sets of mixed DNAs using 13 unrelated Korean males

No. of repeat difference ^a	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391 ^c	DYS 392 ^b	DYS 393	DYS 437 ^c	DYS 438	DYS 439	DYS 448	DYS 456 ^c	DYS 458	DYS 481 ^b	DYS 533	DYS 549	DYS 576	DYS 635	DYS 643	Y GATA H4
0	2	3	1	6	1	1	5	5	3	3	5	1	1	4	5	1	2	2	1	3
1	3	4	3	5	4	6	6	3	5	4	6	7	3	4	3	3	5	4	4	4
2	6	4	5	0	3	3	0	1	3	4	0	0	3	3	3	4	4	4	5	3
3	0	0	2	0	2	1	0	2	0	0	0	2	3	0	0	1	0	1	1	1
4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Total	9	8	10	5	9	10	6	6	8	8	6	9	10	7	6	8	9	9	10	8

^aThe number of repeat differences between two samples.

^bDYS392 and DYS481 included more than one sample which has the alleles with three repeat differences.

^cDYS391, DYS437, and DYS456 could not make combinations with more than one repeat difference.

(2) 두 명의 한국인 남성 시료의 혼합 비율

표준 남성 혼합 시료 결과를 참고하여 미량 시료의 유전자형 확인이 어려운 1:99 비율은 제외하였다. 두 명의 한국인 남성 시료를 A시료: B시료라고 하였을 때 1:1, 1:3, 1:6, 1:9, 1:14, 1:19, 1:29, 1:49, 3:1, 6:1, 9:1, 14:1, 19:1, 29:1, 49:1 의 비율로 혼합하여 각 세트가 15개의 비율을 가지도록 제조 하였다.

다. Illumina Miseq 분석용 한국인 남성 시료 라이브러리 제작 결과

20명의 한국인 남성 단일 시료와 11개 세트의 15개 혼합 비율을 가지는 두 명의 한국인 남성 혼합 시료를 포함하여 총 185개 시료에 대한 Illumina Miseq 분석용 라이브러리를 제작하였다. DNA 시료의 증폭산물로부터 두 단계의 PCR을 통해 라이브러리 제작을 수행하였으며 2100 bioanalyzer (Agilent Technologies)로 라이브러리를 확인 한 결과 200 bp ~ 400 bp 내에 작은 절편들이 성공적으로 생성되었음을 확인하였다.

라. NGS 장비를 이용한 한국인 남성 시료 시퀀싱 결과

제작된 라이브러리는 Miseq 장비를 이용하여 시퀀싱을 수행 한 후 획득한 리드 수는 총 42,440,028개 였다. 185개 시료 중 3개 시료는 리드 수가 222, 276, 26로 리드 수를 확보 하지 못하였다. 3개를 제외한 182개의 평균 리드수는 총 233,184개 였다.

마. 한국인 남성 시료를 이용한 NGS 결과 분석

(1) 한국인 남성 단일 시료를 이용한 NGS 결과

NGS002, NGS008은 리드 수가 확보되지 못했으므로 분석에서 제외하고, 18개 단일시료를 이용하여 Java 기반 프로그램을 통해 분석하고, 이를 통해 결정된 유전자 형은 PowerPlex[®] Y23를 이용한 한국인 남성 시료의 유전자형 결과와 비교하여 동일한 유전자형을 가지는 것을 확인하였다.

18개 한국인 남성 단일시료를 이용하여 23개 Y-STR 및 Y-M175의 평균 리드 수를 계산하였고 그림 1에 정리하였다. DYS448의 평균 리드수가 가장 낮았으며 DYS392의 평균 리드수가 가장 높았다.

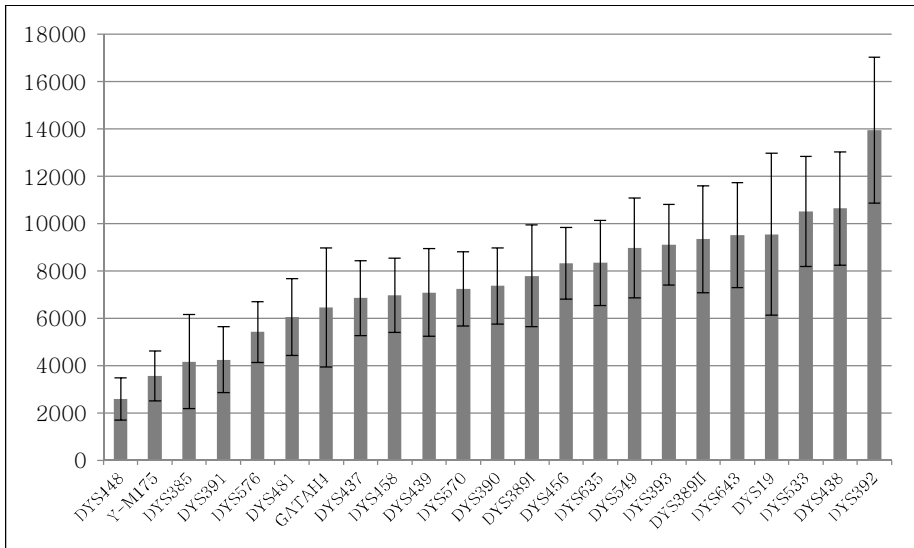


그림 1. Average read counts for the 23 Y-STRs of the Y-STR multiplex system used in this study with selected 18 samples. Error bars represent standard deviation.

(2) 한국인 남성 단일시료 NGS 결과를 이용한 stutter 조사 결과

한국인 남성 단일시료의 NGS 자료 분석을 통해 확인된 18개 단일시료에 대하여 결정된 대립유전자형을 기준으로 21개 Y-STR에서의 stutter 비율을 조사하였다. II-9에서 설명한 N-1, N-2, N-3, N+1 4가지 stutter 비율을 조사하여 표 6에 정리하였다. 3염기 반복 단위, 4염기 반복 단위, 5염기 반복 단위, 6염기 반복 단위로 구분하여 stutter의 평균 비율을 조사하였다. 각 반복 단위 염기수에 해당하는 N-1의 stutter 평균 비율은 23.16%, 12.06%, 6.64%, 7.38%로 조사되었다. 또한 단순 반복 서열과 복합 반복 서열을 가지는 Y-STR을 구분하여 조사하여 표 6에 표시하였다.

반복 단위 수에 따라 정리한 결과를 보면 3염기 반복 단위를 가지는 DYS392, DYS481의 N-1 stutter 비율이 가장 높으며 두 유전자 좌 중에서는 DYS481의 stutter 비율이 높게 나타난 것을 확인할 수 있었다.

Stutter 리드를 제거하기 위하여 stutter 필터 비율을 계산하여 정리하였다 (표 7). Stutter 필터 비율은 stutter 평균비율 + stutter 비율 표준편차 x 3이다. 이 개념은 Gaussian distribution⁹을 바탕으로 제시된 것이며 CE 기반 상용키트인 PowerPlex[®] Y23 개발 연구²⁵를 참고하였다.

표 6. Stutter ratios of 21 Y-STR for the NGS analysis system used in this study

Y-STR locus	Repeat motif	Ave		SD		Ave		SD		Repeat type
		N-3	N-2	N-1	N+1					
DYS392	3bp	0.0056	0.0026	0.0211	0.0069	0.1605	0.0236	0.0686	0.0115	Simple
DYS481		0.0092	0.0029	0.0578	0.0131	0.3028	0.0331	0.0551	0.0079	Simple
DYS19	4bp	-	-	0.0100	0.0023	0.1232	0.0119	0.0110	0.0025	Compound
DYS389I		-	-	0.0067	0.0038	0.0824	0.0213	0.0079	0.0029	Simple
DYS389II		-	-	0.0272	0.0056	0.2090	0.0137	0.0147	0.0043	Compound
DYS390		-	-	0.0114	0.0028	0.1307	0.0110	0.0071	0.0050	Compound
DYS391		-	-	0.0067	0.0027	0.0962	0.0113	0.0087	0.0020	Simple
DYS393		-	-	0.0109	0.0041	0.1260	0.0142	0.0199	0.0041	Simple
DYS437		-	-	0.0032	0.0016	0.0701	0.0179	0.0086	0.0031	Compound
DYS439		-	-	0.0058	0.0019	0.0892	0.0110	0.0177	0.0064	Simple
DYS456		-	-	0.0100	0.0017	0.1234	0.0104	0.0275	0.0056	Simple
DYS458		-	-	0.0163	0.0058	0.1537	0.0219	0.0128	0.0031	Simple
DYS533		-	-	0.0077	0.0019	0.1158	0.0093	0.0139	0.0022	Simple
DYS549		-	-	0.0082	0.0027	0.1231	0.0159	0.0173	0.0042	Simple
DYS570		-	-	0.0117	0.0036	0.1336	0.0162	0.0216	0.0046	Simple
DYS576		-	-	0.0118	0.0045	0.1302	0.0183	0.0198	0.0045	Simple
DYS635		-	-	0.0087	0.0029	0.1121	0.0203	0.0124	0.0066	Compound
GATAH4		-	-	0.0079	0.0023	0.1101	0.0154	0.0122	0.0046	Simple
DYS438	5bp	-	-	0.0042	0.0023	0.0639	0.0141	0.0035	0.0013	Simple
DYS643		-	-	0.0046	0.0027	0.0688	0.0080	0.0051	0.0037	Simple
DYS448	6bp	-	-	0.0082	0.0072	0.0738	0.0101	0.0019	0.0022	Compound

Ave, average; SD, standard deviation

표 7. Stutter filter ratio of 21 Y-STR for the mixture interpretation
using NGS analysis system

Locus	N-3	N-2	N-1	N+ 1
	Ave + 3SD	Ave + 3SD	Ave + 3SD	Ave + 3SD
DYS19		0.0168	0.1588	0.0185
DYS389I		0.0182	0.1465	0.0167
DYS389II		0.0441	0.2500	0.0276
DYS390		0.0197	0.1636	0.0222
DYS391		0.0149	0.1301	0.0148
DYS392	0.0135	0.0418	0.2314	0.1030
DYS393		0.0233	0.1687	0.0323
DYS437		0.0080	0.1239	0.0178
DYS438		0.0111	0.1063	0.0073
DYS439		0.0114	0.1221	0.0369
DYS448		0.0297	0.1041	0.0085
DYS456		0.0151	0.1546	0.0444
DYS458		0.0337	0.2194	0.0219
DYS481	0.0178	0.0973	0.3962	0.0792
DYS533		0.0133	0.1436	0.0205
DYS549		0.0162	0.1710	0.0300
DYS570		0.0226	0.1822	0.0354
DYS576		0.0253	0.1851	0.0332
DYS635		0.0174	0.1731	0.0322
DYS643		0.0127	0.0928	0.0161
GATAH4		0.0149	0.1562	0.0259

Ave, average; SD, standard deviation

(3) 한국인 남성 단일시료 NGS 결과를 이용한 노이즈 조사 결과

한국인 남성 단일시료 Y-STR 유전자형 결과를 바탕으로 21개 Y-STR (Y-M175와 DYS385ab를 제외)에서 노이즈 비율을 조사하여 표 8에 정리하였다. 노이즈 비율은 각 Y-STR에 배정된 리드 수에서 결정된 대립유전자형 및 stutter가 차지하는 리드 수를 제외한 리드 수가 차지하는 비율이다. DYS392의 노이즈 비율이 가장 높은 것을 확인할 수 있었다.

CE 기법을 이용한 STR 분석에서 발생하는 노이즈는 조사된 노이즈 peak height의 평균과 표준 편차 값의 세배를 더한 값을 사용하여 관찰된 peak와 노이즈를 구분하였다.²⁷ 이 개념을 NGS 자료분석에 도입하기 위하여 표 8에 노이즈 평균비율 + 노이즈 표준편차 x 3로 노이즈 필터비율 추가적으로 정리하였다.

노이즈 필터값은 혼합시료 NGS 결과 분석에서 분석에 포함 또는 불포함 하는 리드를 선별하기 위한 값으로 각 시료 및 각 Y-STR에 배정되는 리드수가 시료마다 다르게 제공되므로 절대값이 아닌 비율로 제시하였다.

표 8. Noise ratios and noise filter ratio of 21 Y-STR for the NGS analysis system used in this study

Locus	Noise ratio		Noise filter ratio
	AVE	SD	AVE+ 3SD
DYS19	0.0006	0.0008	0.0029
DYS389I	0.0007	0.0010	0.0037
DYS389II	0.0006	0.0009	0.0031
DYS390	0.0005	0.0008	0.0028
DYS391	0.0005	0.0003	0.0015
DYS392	0.0021	0.0021	0.0084
DYS393	0.0008	0.0024	0.0081
DYS437	0.0003	0.0002	0.0008
DYS438	0.0009	0.0015	0.0053
DYS439	0.0004	0.0007	0.0025
DYS448	0.0012	0.0015	0.0056
DYS456	0.0005	0.0004	0.0017
DYS458	0.0007	0.0016	0.0056
DYS481	0.0014	0.0015	0.0059
DYS533	0.0006	0.0015	0.0052
DYS549	0.0006	0.0013	0.0046
DYS570	0.0009	0.0018	0.0063
DYS576	0.0007	0.0019	0.0063
DYS635	0.0004	0.0005	0.0019
DYS643	0.0008	0.0011	0.0040
GATAH4	0.0013	0.0015	0.0058

(4) Stutter 필터비율 및 노이즈 필터비율을 반영한 한국인 남성
혼합 시료 분석

본 연구를 통해 21개 Y-STR의 stutter 필터비율(표 7)과 노이즈 필터비율과 표준편차(표 8)를 이용하여 한국인 남성 혼합시료를 분석하였다.

각 시료에서 21개 Y-STR에 배정된 리드 수에 노이즈 필터비율을 곱하여 계산된 노이즈 리드 수는 기준 값이 되어 그 값보다 같거나 낮은 리드 수를 가지는 대립유전자형을 모두 제외하였다. 다음으로 각 세트의 21개 Y-STR에서 알고 있는 대립유전자형을 중심으로 stutter 필터비율을 적용하여 stutter에 해당되는 리드를 제거하였다. 또한 stutter 필터비율 적용 후 남은 리드가 노이즈 리드 수 보다 같거나 낮은 리드 수를 가지는 대립유전자형은 제거하였다. Stutter 비율은 DYS392, DYS481에서는 N-1, N-2, N-3, N+1 의 stutter 필터비율을 적용하였으며, 위 두 유전자 좌를 제외한 19 Y-STR에서는 N-2, N-1, N+1에 stutter 필터비율을 적용하였다.

위와 같은 방법으로 각 시료 및 21개 Y-STR에 반영하여 163개 한국인 남성 혼합시료를 분석하였다. 혼합시료 내 두 시료가 동일한 대립유전자형을 갖는 Y-STR 유전자 좌는 분석에서 제외하였다.

(5) 한국인 남성 혼합시료 대립유전자형 결정 결과

Stutter 필터비율 및 노이즈 필터비율을 반영하여 대립유전자형을 결정하였으며 표 9-1 ~ 표 9-11에 정리하였다.

각 시료가 가지는 21 Y-STR에 대한 대립유전자형을 비교하였고, 예상하지 못한 대립유전자형은 각 표에서 밑줄로 표시하였다. 각 혼합세트에서 21 Y-STR 중 DYS390, DYS392, DYS456, DYS481, DYS570, DYS576, DYS635, DYS643 유전자 좌에서 실제 시료가 가지지 않는 대립유전자형이 하나 이상 관찰되었다.

163개의 혼합시료 대립유전자형 결정 결과를 종합하여 표 10으로 정리하였다. 1:3, 1:6, 1:9, 1:14, 1:19, 1:29, 1:49비율은 각각 3:1, 6:1, 9:1, 14:1, 19:1, 29:1, 49:1을 포함하여 정리하였다. 유전자 좌, 혼합 비율 별로 혼합시료 내 두 시료가 다른 유전자형을 가지는 경우의 수를 유전자형을 획득한 경우/전체 경우의 수로 정리하였다. 이미 알고 있는 두 시료의

대립유전자형을 대립유전자형 결정 결과와 비교하고, 유전자 형을 획득한 비율이 100%일 경우에는 흰색, 90%이상, 80%이상, 70%이상, 60%이상, 50%이상, 50%이하로 점점 진한 색으로 표시하였다.

각 Y-STR마다 다른 혼합비율에서 대립유전자형 소실을 관찰할 수 있었다. DYS458, DYS635의 경우에는 1:6 비율부터 대립유전자형 소실이 관찰 되었으며, DYS19의 경우에는 1:19 비율까지 대립유전자형 소실 없이 모든 획득 된 것을 알 수 있다. 각 혼합 비율에서 두 시료의 대립유전자형이 모두 관찰 된 Y-STR 비율을 종합적으로 살펴보면 1:1, 1:3에서는 21개 Y-STR에서 모두 혼합된 두 대립유전자형을 관찰하였으며(100%), 1:6에서는 99.50%, 1:9에서는 96.75%, 1:14에서는 91.37%, 1:19에서는 82.00%, 1:29에서는 59.06%, 1:49에서는 49.47%로 관찰되었다.

표 9-1. Allele determinations in various range of mixture ratio of two Korean males of Set 1 (NGS001:NGS006) from next generation sequencing

Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	19:1	29:1	49:1
DYS19	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15	15, 17
DYS389I	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	13	13	13	13
DYS389II	27, 30	27, 30	27, 30	27, 30	27, 30	27, 30	27, 30	27, 30	27, 30	27, 30	27, 30	27, 30	27, 30	27, 30	27, 30
DYS390	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25
DYS392	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11
DYS393	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	13
DYS437	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15
DYS439	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13
DYS448	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20
DYS456	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	15	15
DYS458	16, 17	16, 17	16, 17	16, 17	17	17	17	17	16, 17	16, 17	16, 17	16, 17	16, 17	16, 17	16, 17
DYS481	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26
DYS533	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	12
DYS549	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13
DYS570	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3
DYS635	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23
DYS643	9, 10	9, 10	9, 10	9, 10	9, 10	9, 10	9, 10	10	9, 10	9, 10	9, 10	9, 10	9, 10	9	9, 10
GATAH4	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12

Unexpected alleles were indicated with underline.

표 9-2. Allele determinations in various range of mixture ratio of two Korean males of Set 2 (NGS003:NGS006) from next generation sequencing

Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	19:1	29:1	49:1
DYS19	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17
DYS389II	27, 29	27, 29	27, 29	27, 29	27, 29	27, 29	27, 29	27, 29	27, 29	27, 29	27, 29	27, 29	27, 29	27, 29	29
DYS390	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25	23, 25
DYS392	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	13	13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13
DYS393	12, 14	12, 14	12, 14	12, 14	12	12, 14	12	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14
DYS437	14, 15	14, 15	14, 15	15	14, 15	14, 15	15	15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15
DYS439	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	13	13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13
DYS456	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14	14	14, 15	14, 15	14, 15	14, 15	14, 15	15	14, 15
DYS458	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17	17, 18	17, 18	17, 18	17, 18	17, 18	18	18	18
DYS481	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24, 26	24
DYS533	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10	10	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12
DYS549	12, 13	12, 13	12, 13	13	13	12, 13	12, 13	13	12, 13	12, 13	12, 13	12, 13	12, 13	12	12
DYS570	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	19.3	19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3	17, 19.3
DYS576	17, 18	17, 18	17, 18	17, 18	<u>11, 14, 17, 18</u>	17, 18	17	17	17, 18	17, 18	17, 18	17, 18	17, 18	18	18
DYS635	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23
DYS643	9, 10	9, 10	9, 10	9, 10	9, 10	10	10	10	9, 10	9, 10	9, 10	9, 10	9, 10	9	9, 10
GATAH4	11, 12	11, 12	11, 12	11, 12	11, 12	12	12	12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11

Unexpected alleles were indicated with underline.

표 9-3. Allele determinations in various range of mixture ratio of two Korean males of Set 3 (NGS005:NGS012) from next generation sequencing

Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	19:1	29:1	49:1
DYS19	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17
DYS389I	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14
DYS389II	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30
DYS390	22, 24	22, 24	22, 24	22, 24	22, 24	22, 24	24	22, 24	22, 24	22, 24	22, 24	22, 24	22, 24	22, 24	22, 24
DYS392	13, 13	13, 13	13, 13	13, 13	13, 13	13, 13	13, 13	13, 13	<u>6</u> , 13, 13	<u>6</u> , 13, 13	13, 13	13, 13	13, 13	13, 13	13, 13
DYS393	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	13	13
DYS438	10, 13	10, 13	10, 13	10, 13	10	10, 13	10	10	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13
DYS448	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18	20
DYS458	17, 18	17, 18	17, 18	18	18	18	18	18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18
DYS481	21, 23	21, 23	21, 23	23	21, 23	23	23	23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23
DYS549	12, 13	12, 13	12, 13	12, 13	12, 13	12	12, 13	12, 13	12, 13	12, 13	13	12, 13	13	13	12
DYS570	18, 19	18, 19	18, 19	18, 19	18, 19	18	18, 19	18	18, 19	18, 19	18, 19	18, 19	18, 19	18, 19	18
DYS576	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	<u>11</u> , 17, 19
DYS635	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23
DYS643	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	12	10, 12	10, 12

Unexpected alleles were indicated with underline.

표 9-4. Allele determinations in various range of mixture ratio of two Korean males of Set 4 (NGS001:NGS005) from next generation sequencing

Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	19:1	29:1	49:1
DYS389I	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	14	14	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	13
DYS390	22, 23	22, 23	22, 23	22, 23	22, 23	22, 23	22, 23	22	22, 23	22, 23	22, 23	23	23	23	23
DYS392	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11
DYS438	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13
DYS439	11, 12	11, 12	11, 12	11, 12	12	12	12	12	11, 12	11, 12	11, 12	11, 12	11, 12	11	11
DYS448	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	20	20
DYS456	15, 15	15, 15	15, 15	15, 15	15, 15	15, 15	15, 15	15, 15	15, 15	15, 15	15, 15	<u>15, 15</u> <u>17</u>	15, 15	15, 15	15, 15
DYS458	16, 17	16, 17	16, 17	16, 17	16, 17	17	17	17	16, 17	16, 17	16, 17	16, 17	16, 17	16	16, 17
DYS481	21, 24	21, 24	21, 24	21, 24	21, 24	21, 24	21, 24	21, 24	21, 24	21, 24	21, 24	21, 24	21, 24	<u>8</u> <u>21, 24</u>	21, 24
DYS533	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	12
DYS549	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13
DYS570	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19
DYS576	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19
DYS635	20, 21	20, 21	20, 21	20, 21	20, 21	20	20	20	20, 21	20, 21	21	20, 21	21	21	21
DYS643	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12	9, 12
GATAH4	10, 12	10, 12	10, 12	10, 12	10, 12	12	12	12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12

Unexpected alleles were indicated with underline.

표 9-5. Allele determinations in various range of mixture ratio of two Korean males of Set 5 (NGS007:NGS010) from next generation sequencing

Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	19:1	29:1	49:1
DYS19	14, 16	14, 16	14, 16	14, 16	14, 16	14, 16	16	14, 16	14, 16	14, 16	14, 16	14, 16	14, 16	14, 16	14, 16
DYS389II	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	31
DYS391	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11
DYS392	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16
DYS393	13, 15	13, 15	13, 15	13, 15	13, 15	13, 15	13, 15	13	13, 15	13, 15	13, 15	13, 15	13, 15	13, 15	13, 15
DYS438	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13
DYS439	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12
DYS448	18, 19	18, 19	18, 19	18, 19	18, 19	18, 19	18	18	18, 19	18, 19	18, 19	18, 19	18, 19	18, 19	19
DYS456	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14
DYS481	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23
DYS570	16, 18	16, 18	16, 18	16, 18	18	16, 18	18	18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18
DYS576	16, 17	16, 17	16, 17	17	17	17	17	17	16, 17	16, 17	16, 17	16, 17	16, 17	16, 17	16, 17
DYS635	20, 21	20, 21	20, 21	21	21	21	21	21	20, 21	20, 21	20, 21	20, 21	20, 21	20, 21	20, 21
DYS643	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13	11, 13
GATAH4	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	12	12

Unexpected alleles were indicated with underline.

표 9-6. Allele determinations in various range of mixture ratio of two Korean males of Set 6 (NGS007:NGS013) from next generation sequencing

Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	19:1	29:1	49:1
DYS19	14, 16	14, 16	14, 16	14, 16	14, 16	14, 16	14, 16	16	14, 16	14, 16	14, 16	14, 16	14, 16	14, 16	14, 16
DYS389I	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12	12	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14
DYS389II	28, 31	28, 31	28, 31	28, 31	28, 31	28, 31	28	28	28, 31	28, 31	28, 31	28, 31	28, 31	28, 31	28, 31
DYS390	23, 24	23, 24	23, 24	23, 24	23, 24	23, 24	24	24	23, 24	23, 24	23, 24	23, 24	23, 24	23, 24	23, 24
DYS391	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10	10	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	11
DYS392	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	13	13, 16	13, 16	13, 16	13, 16	13, 16	13, 16	16
DYS393	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15	12, 15
DYS437	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	15	15	14, 15	14, 15	14, 15	14, 15	14, 15	14	14
DYS438	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	11	11	11
DYS439	10, 11	10, 11	10, 11	10, 11	10, 11	11	10, 11	11	10, 11	10, 11	10, 11	10	10	10, 11	10
DYS448	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20	19, 20
DYS456	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	15	15	14, 15	14, 15	14, 15	14, 15	14, 15	14	14
DYS458	16, 17	16, 17	16, 17	16, 17	16, 17	16, 17	16	16	16, 17	16, 17	17	17	17	17	17
DYS481	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	23	23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23	20, 23
DYS549	12, 13	12, 13	12, 13	12, 13	13	13	12, 13	13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13
DYS570	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16
DYS576	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	18	18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18
DYS635	20, 22	20, 22	20, 22	20, 22	20, 22	20, 22	20, 22	22	20, 22	20, 22	20, 22	20, 22	20, 22	20, 22	20, 22

Unexpected alleles were indicated with underline.

표 9-7. Allele determinations in various range of mixture ratio of two Korean males of Set 7 (NGS003:NGS007) from next generation sequencing

Locus	1:1	1:6	1:3	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	19:1	29:1	49:1
DYS19	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14	14	14, 15	14, 15	14, 15	14, 15	14, 15	15	14, 15
DYS389I	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14
DYS389II	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	31	31	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31	29, 31
DYS391	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11
DYS392	11, 16	11, 16	11, 16	11, 16	11, 16	11, 16	16	16	11, 16	11, 16	11, 16	11, 16	11, 16	11, 16	11, 16
DYS393	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	15	15	14, 15	14, 15	14, 15	14, 15	14, 15	14	14, 15
DYS438	10, 11	10, 11	10, 11	10, 11	10, 11	11	11	11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11
DYS439	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10	10, 11	10, 11	10, 11	10, 11	10, 11	11	10, 11
DYS456	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	15	14, 15
DYS458	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17	17	17, 18	18	18	18	18	18	17, 18
DYS481	20, 24	20, 24	20, 24	20, 24	20, 24	20, 24	<u>8</u> , 20	20	20, 24	20, 24	20, 24	20, 24	20, 24	20, 24	20, 24
DYS533	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11	11	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12
DYS570	16, 17	16, 17	16, 17	16, 17	16, 17	16	16	16	16, 17	16, 17	16, 17	16, 17	17	17	16, 17
DYS576	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18	16, 18
DYS635	20, 21	20, 21	20, 21	20, 21	20, 21	20, 21	<u>20</u> , <u>22</u>	<u>20</u> , <u>22</u>	20, 21	21	20, 21	21	21	21	20, 21
DYS643	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	11	11	9, 11	<u>9</u> , <u>11</u> <u>12</u>	9, 11	9, 11	9, 11	9	<u>9</u> , <u>11</u> <u>22</u>
GATAH4	11, 12	11, 12	11, 12	11, 12	12	12	12	12	11, 12	11, 12	11, 12	11, 12	11, 12	11	11, 12

Unexpected alleles were indicated with underline.

표 9-8. Allele determinations in various range of mixture ratio of two Korean males of Set 8 (NGS008:NGS012) from next generation sequencing

Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	19:1	29:1	49:1
DYS19	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17	15, 17
DYS389I	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12	12, 13	12, 13	12, 13	12, 13	13	13	13
DYS389II	28, 29	28, 29	28, 29	28, 29	28, 29	28, 29	28	28	28, 29	28, 29	28, 29	28, 29	29	29	29
DYS391	10, 11	10, 11	10, 11	10, 11	10, 11	10	10	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11
DYS392	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	13	13	13, 14	13, 14	13, 14	13, 14	14	14	14
DYS393	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12	12	12, 13	12, 13	12, 13	13	12, 13	13	13
DYS448	18, 20	18, 20	18, 20	18, 20	18, 20	20	20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20
DYS456	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	15	15	14, 15	14, 15	14, 15	14, 15	14	14, 15	14, 15
DYS458	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21
DYS533	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11
DYS570	18, 18	18, 18	18, 18	18, 18	<u>10, 13</u> <u>18, 18</u>	18, 18	18, 18	18, 18	18, 18	18, 18	18, 18	18, 18	18, 18	18, 18	18, 18
DYS576	17, 18	17, 18	17, 18	17, 18	17, 18	17	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18	17, 18
DYS635	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23	21, 23
DYS643	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12
GATAH4	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	12	12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12

Unexpected alleles were indicated with underline.

표 9-9. Allele determinations in various range of mixture ratio of two Korean males of Set 9 (NGS002:NGS011) from next generation sequencing

Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	19:1	29:1	49:1
DYS19	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	15	15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14
DYS389I	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14
DYS389II	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30	28, 30
DYS390	23, 24	23, 24	23, 24	23, 24	23, 24	23, 24	24	24	23, 24	23, 24	23, 24	23, 24	23, 24	23, 24	23
DYS392	13, 14	13, 14	13, 14	13, 14	13, 14	13	13	13	13, 14	13, 14	13, 14	13, 14	14	14	14
DYS393	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	14	14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14
DYS437	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	14	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15	15
DYS438	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11	10	10, 11	10, 11	10, 11	10, 11	11	10, 11	11	11
DYS439	11, 12	11, 12	11, 12	11, 12	11, 12	11	11	11	11, 12	11, 12	11, 12	11, 12	11, 12	12	12
DYS448	18, 19	18, 19	18, 19	18, 19	18, 19	18, 19	18	18, 19	18, 19	18, 19	18, 19	18, 19	18, 19	19	19
DYS458	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16	15	15, 16	15, 16	15, 16	16	16	16	16
DYS481	22, 23	22, 23	22, 23	23	23	23	23	23	22, 23	22, 23	22, 23	22	22, 23	22, 23	22, 23
DYS570	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21	18, 21
DYS576	18, 19	18, 19	18, 19	18, 19	18, 19	18, 19	18	18	18, 19	18, 19	18, 19	18, 19	18, 19	19	19
DYS635	21, 22	21, 22	21, 22	21, 22	22	22	22	22	21, 22	21, 22	21, 22	21, 22	21, 22	21, 22	21
DYS643	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	12	12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12

Unexpected alleles were indicated with underline.

표 9-10. Allele determinations in various range of mixture ratio of two Korean males of Set 10 (NGS003:NGS019) from next generation sequencing

Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	29:1	49:1
DYS389II	29, 30	29, 30	29, 30	29, 30	29, 30	29, 30	30	30	29, 30	29, 30	29, 30	29, 30	29, 30	29, 30
DYS390	23	23	23	23	23	23	23	23	23	23, <u>25</u>	23	23	23	23
DYS392	11, 12	11, 12	11, 12	<u>6</u> , 12	12	12	12	12	11, 12	11, 12	11, 12	11, 12	11	11
DYS393	13, 14	13, 14	13, 14	13, 14	13	13	13	13	13, 14	13, 14	13, 14	13, 14	14	14
DYS437	14, 15	14, 15	14, 15	14, 15	15	14, 15	15	15	14, 15	14, 15	14, 15	14, 15	14, 15	14, 15
DYS458	15, 18	15, 18	15, 18	15, 18	15, 18	15, 18	15, 18	15	15, 18	15, 18	15, 18	15, 18	15, 18	15, 18
DYS481	23, 24	23, 24	23, 24	23, 24	23, 24	23, 24	23	23	23, 24	23, 24	23, 24	24	24	24
DYS533	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12	10	10	10, 12	10, 12	10, 12	10, 12	10, 12	10, 12
DYS570	17, 19	17, 19	17, 19	17, 19	17, 19	17, 19	19	19	17, 19	17, 19	17, 19	17, 19	17, 19	17
DYS576	18, 19	18, 19	18, 19	18, 19	19	19	19	<u>4</u> , 19	18, 19	18, 19	18, 19	18, 19	<u>7</u> , 18, 19	18
DYS635	21	21	21	21	21	21	21	21	21	21, <u>23</u>	21	21	21	21
DYS643	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11	9, 11, <u>12</u>
GATAH4	11, 12	11, 12	11, 12	11, 12	11, 12	12	12	12	11, 12	11, 12	11, 12	11, 12	11	11

Unexpected alleles were indicated with underline.

표 9-11. Allele determinations in various range of mixture ratio of two Korean males of Set 11 (NGS008:018) from next generation sequencing

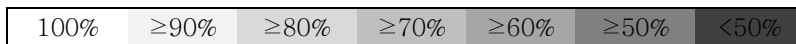
Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49	3:1	6:1	9:1	14:1	29:1	49:1
DYS19	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16	16	16	15, 16	15, 16	15, 16	15, 16	15	15
DYS389I	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12	12	12, 13	12, 13	12, 13	13	13	13
DYS389II	29, 30	29, 30	29, 30	29, 30	29, 30	30	29	29	29, 30	29, 30	29, 30	29, 30	29, 30	29, 30
DYS391	10, 11	10, 11	10, 11	10, 11	10, 11	10	10, 11	11	10, 11	10, 11	10, 11	10, 11	10, 11	11
DYS392	13, 14	13, 14	13, 14	13, 14	13, 14	13, 14	13	13	13, 14	13, 14	14	14	14	14
DYS393	12, 13	12, 13	12, 13	12, 13	12, 13	12, 13	12	12	12, 13	12, 13	12, 13	12, 13	13	13
DYS437	14, 15	14, 15	14, 15	14, 15	14, 15	15	15	15	14, 15	14, 15	14, 15	14, 15	14, 15	14
DYS439	11, 12	11, 12	11, 12	11, 12	11, 12	11	11	11	11, 12	11, 12	11, 12	11, 12	11, 12	12
DYS448	18, 20	18, 20	18, 20	18, 20	18, 20	18, 20	20	20	18, 20	18, 20	18, 20	18, 20	18, 20	18
DYS458	16, 21	16, 21	16, 21	16, 21	16, 21	16, 21	16, 21	16	16, 21	16, 21	16, 21	16, 21	16, 21	16, 21
DYS481	23, 24	23, 24	23, 24	24	24	24	24	24	23, 24	23, 24	23, 24	23, 24	23, 24	23
DYS533	10, 11	10, 11	10, 11	10, 11	11	11	11	11	10, 11	10, 11	10, 11	10, 11	10, 11	10, 11
DYS549	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14	14	14	12, 14	12, 14	12, 14	12, 14	12, 14	12, 14
DYS570	18, 19	18, 19	18, 19	18, 19	18, 19	18, 19	19	19	18, 19	18, 19	18, 19	18, 19	18, 19	18
DYS643	11, 12	11, 12	11, 12	11, 12	11, 12	11, 12	11	11	11, 12	11, 12	11, 12	11, 12	11, 12	12
GATAH4	10, 13	10, 13	10, 13	10, 13	10, 13	10, 13	13	10, 13	10, 13	10, 13	10, 13	10, 13	10	10, 13

Unexpected alleles were indicated with underline.

표 10. The counts of allele determinations of 163 mixed DNAs combinations for 21 Y-STRs by mixture ratios

Locus	1:1	1:3	1:6	1:9	1:14	1:19	1:29	1:49
DYS19	9/9	18/18	18/18	18/18	18/18	17/17	11/18	11/18
DYS389I	8/8	16/16	16/16	16/16	14/16	13/15	10/16	7/16
DYS389II	10/10	20/20	20/20	20/20	20/20	16/18	14/20	11/20
DYS390	6/6	12/12	12/12	12/12	11/12	11/12	8/12	7/12
DYS391	5/5	10/10	10/10	10/10	10/10	7/9	7/10	6/10
DYS392	10/10	20/20	20/20	18/20	18/20	14/18	10/20	9/20
DYS393	10/10	20/20	20/20	20/20	17/20	17/18	9/20	8/20
DYS437	6/6	12/12	12/12	11/12	11/12	9/10	7/12	3/12
DYS438	6/6	12/12	12/12	12/12	10/12	10/12	7/12	8/12
DYS439	8/8	16/16	16/16	16/16	14/16	10/15	9/16	6/16
DYS448	8/8	16/16	16/16	16/16	16/16	14/15	8/16	8/16
DYS456	6/6	12/12	12/12	12/12	12/12	11/12	5/12	5/12
DYS458	10/10	20/20	19/20	18/20	15/20	11/18	9/20	8/20
DYS481	10/10	20/20	20/20	17/20	16/20	15/18	13/20	10/20
DYS533	7/7	14/14	14/14	14/14	13/14	11/15	10/14	7/14
DYS549	6/6	12/12	12/12	10/12	10/12	8/11	9/12	6/12
DYS570	10/10	20/20	20/20	20/20	19/20	15/18	14/20	10/20
DYS576	9/9	18/18	18/18	17/18	16/18	14/17	11/18	10/18
DYS635	9/9	18/18	17/18	16/18	15/18	13/18	12/18	10/18
DYS643	10/10	20/20	20/20	20/20	20/20	16/18	13/20	12/20
GATAH4	8/8	16/16	16/16	16/16	15/16	10/16	5/16	7/16

The percentages of allele determination of 21 Y-STR loci by mixture ratios were indicated with color gradually getting dark informed below.



(6) 한국인 남성 혼합시료 혼합비율 추정 결과

두 시료의 대립유전자형의 조합에서 두 대립유전자형의 반복 수 차이가 0 또는 1 을 포함하는 DYS391, DYS437, DYS456 을 제외한 18 개 Y-STR 에서 1:1 비율부터 1:49 비율까지 결정된 대립유전자형이 가지는 리드 수를 이용하여 역으로 비율을 계산한 것을 단순 상자그림으로 그림 2-1, 2-2 로 나타내었다. 또한 21 개 Y-STR 중 90%에서 대립유전자형을 확인 한 1:14 의 비율까지 두 대립유전자형의 반복 수 차이 별 비율 추정을 하였고, 각 대립유전자형에 배정된 리드 수를 이용하였다. 단순 상자그림을 이용하였으며 그림 3 으로 나타내었다.

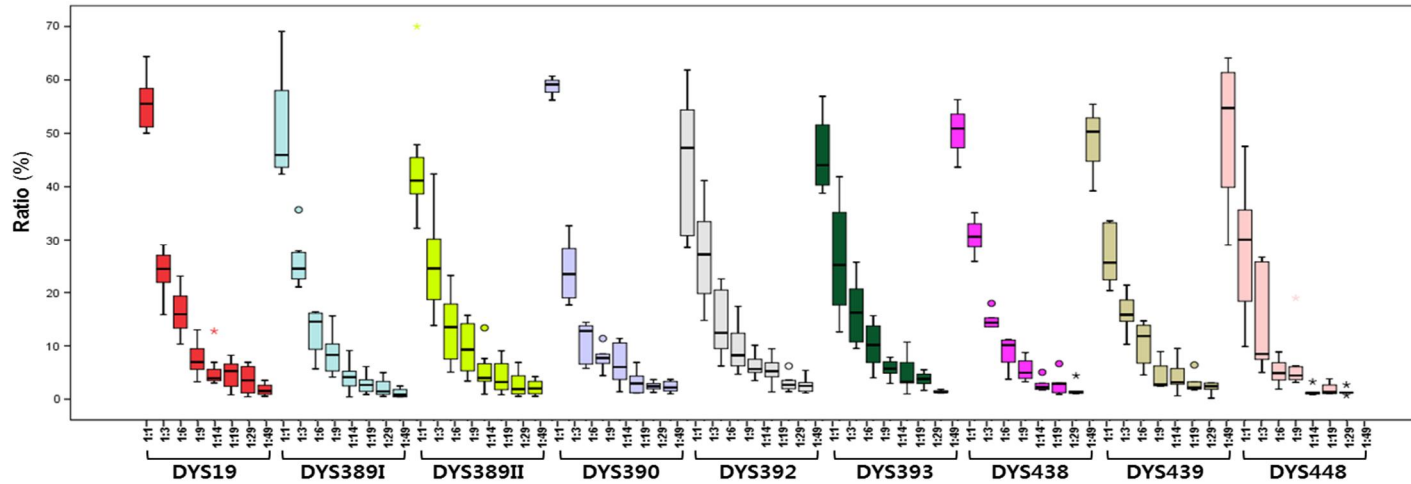


그림 2-1. Boxplot of the percentage of minor components for mixture ratio estimation in mixed DNAs by 1:1 ~ 1:49 ratios in 9 Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS392, DYS393, DYS438, DYS439, DYS448). The boxes illustrate the first, median and third quartiles, and the whiskers represent the minimum and maximum values.

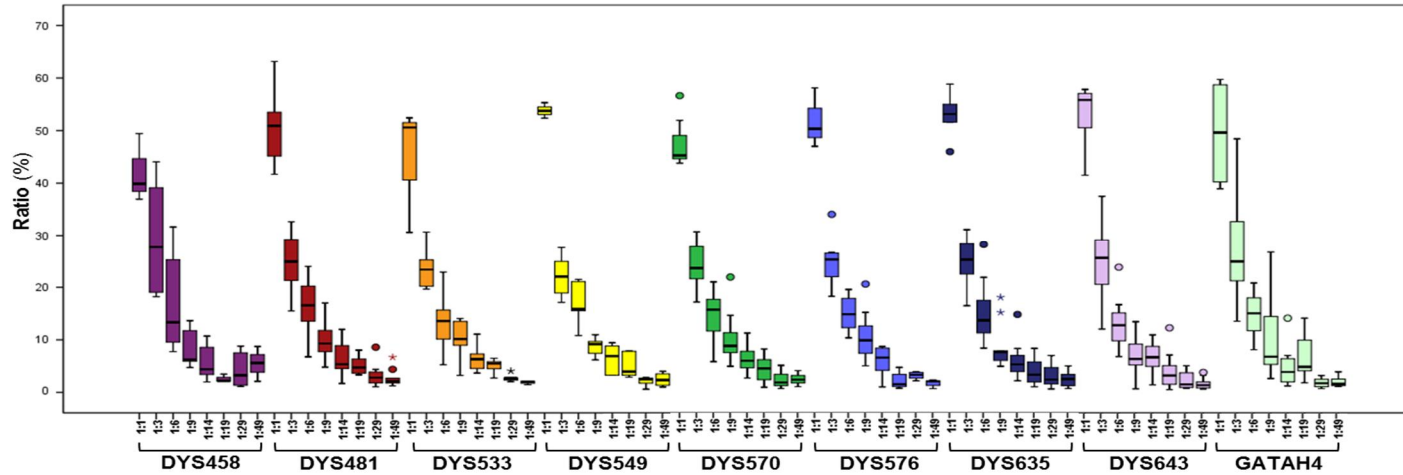


그림 2-2. Boxplot of the percentage of minor components for mixture ratio estimation in mixed DNAs by 1:1 ~ 1:49 ratios in 9 Y-STR loci (DYS458, DYS481, DYS533, DYS549, DYS570, DYS576, DYS635, DYS643, Y-GATAH4). The boxes illustrate the first, median and third quartiles, and the whiskers represent the minimum and maximum values.

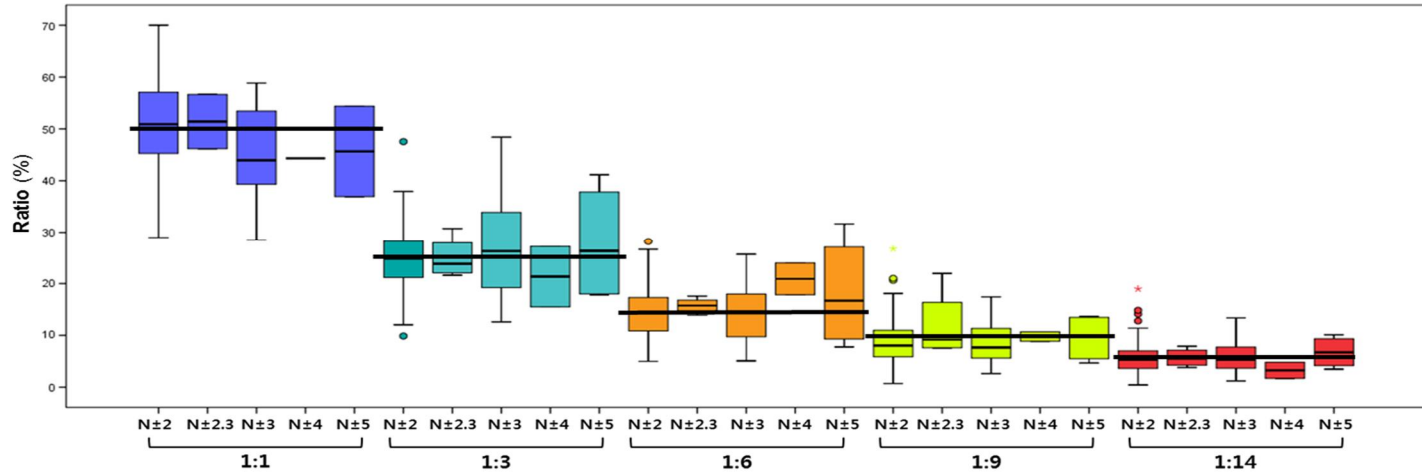


그림 3. Boxplot of the percentage of minor components for mixture ratio estimation in mixed DNAs by the number of repeat difference in 1:1 ~ 1:14. The boxes illustrate the first, median and third quartiles, and the whiskers represent the minimum and maximum values.

IV. 고 찰

현재 법의유전학영역의 STR 표준 분석법은 PCR 증폭 후 사이즈 기반의 분석법인 CE기법이 널리 쓰이고 있다. 이 기법은 단일 시료 분석뿐만 아니라 혼합 시료 분석에서도 이용되고 있다. 혼합 시료 분석에서는 일반적으로 혼합 시료 내 미량 시료의 정보를 알아내는 것에 의미가 있다. 하지만 CE를 이용하여 혼합 시료 내 미량 시료의 정보를 알기 위해서는 실제 peak를 구분하기 위한 노이즈, pull-up, dye-blob 등 고려할 요소가 많으며 또한 PCR 과정 중에 생성된 stutter가 미량 혼합 시료의 유전자형과 같은 크기를 가질 경우에는 구분이 어렵다. 그리고 RFU단위로 표현되는 peak는 상대적인 정량 값을 나타내며 CE 분석법에서는 같은 시료를 이용한 반복적인 실험과 다른 Genetic analyzer 장비를 사용 함에 따라서도 가변적인 값을 가질 수 있다.⁷ 이러한 CE 기법의 한계를 보완하고자 NGS 기법을 이용하여 혼합시료 분석을 시도하고 NGS 기법의 유용성 및 유효성을 조사하고자 하였다.

먼저 단일 시료를 이용하여 NGS 기법에서 관찰되는 stutter 비율과 노이즈 비율을 조사하여 혼합시료 분석에 반영하고자 하였다. 이를 이용하여 혼합 시료 분석에 유효한 분석적 한계치, 해석적 한계치를 제시하고자 하였다. 그리고 혼합 시료 내에서

관찰된 두 대립유전자형의 리드 수를 역으로 계산하여 혼합 비율을 추정해 보고자 하였다. 마지막으로 NGS의 최대 장점인 염기서열 정보를 이용한 혼합시료 분석을 하고자 하였다.

한국인 남성 혼합 시료 분석을 위한 실험을 설계하기 앞서 표준 남성 DNA를 이용한 예비실험을 실시하였다. 표준 남성 혼합시료는 CE 및 NGS 기법을 모두 이용하여 두 기법을 비교하였다. 여러 비율의 표준 남성 혼합 시료의 CE 결과는 일반적인 STR 분석 원칙에 의해 분석 되었다. 분석적 한계치를 100 RFU로 설정하여 그 이상의 RFU를 가지는 peak을 이용하여 대립유전자형을 결정하였으며 stutter는 CE 기법으로 조사된 비율을 적용²⁵하여 제거하였다. 이 결과에서는 9948(Promega): 2800M(Promega)이 1:19 혼합 비율에서부터 대립유전자형 소실이 관찰되어 확률적 효과(stochastic effect)가 나타나기 시작하는 것을 확인하였다. 1:19 혼합 비율에서의 미량 시료는 50 pg/μl의 농도를 가지며 이 농도부터 CE 기법으로 혼합 시료를 분석하는데 한계가 있음을 알 수 있었다.

표준 남성 혼합 시료의 NGS 결과 분석에서는 주 대립유전자형 리드수의 100을 기준으로 하여 그 이상의 리드 수를 가지는 대립유전자형을 결정하였다. 표 3을 살펴보면 DYS392에서는

stutter 위치에서 대립유전자형 획득이 관찰되었다. CE는 1:19, NGS는 1:29부터 대립유전자형 소실이 관찰되었는데 확률적 효과 (Stochastic effect)가 발생하는 미량 시료의 양이 다른 것이 아니라 더 높은 감도를 가지는 NGS 분석이 미량 시료 결과를 더 잘 보여주는 장비의 차이로 인한 것으로 볼 수 있다고 생각한다.

그러나 표준 남성 시료를 이용한 예비실험에서는 비교 가능한 Y-STR 수가 제한적이었으며 NGS 분석에서는 확실한 기준이 적용되지 않은 분석이었다. 또한 한 세트의 혼합시료를 이용한 분석으로는 NGS분석법의 유용성을 판단할 수 있는 근거가 부족하다. 따라서 분석 가능한 Y-STR을 모두 포함하고 더 많은 수의 시료를 이용하여 혼합 시료 분석에 대한 NGS 기법의 평가를 다음 단계의 연구에서 하고자 하였다.

한국인 남성 시료를 대상으로 20개 단일 시료와 165개의 혼합시료를 사용하여 Y-STR 분석을 진행하였다. 분석 대상 Y-STR 중 다중 대립유전자형을 가지는 DYS385ab는 단순한 Y-STR 혼합시료 분석 모델을 제시하기 위하여 제외하였다. DYS385ab를 제외한 21개 Y-STR을 혼합 시료 분석에서 고르게 포함하기 위하여 PowerPlex[®] Y23유전자형 결과를 참고하였다. 그러나 20명의 한국인 시료를 이용한 조합 설계에서 DYS391은 10, 11

DYS437은 14, 15 DYS456에서는 14, 15과 같이 두 개의 대립유전자형만이 관찰 되어 시료 조합 기준에 예외가 발생하였다. 각 Y-STR에 두 개의 대립유전자형이 아닌 다른 유전자 형을 가지는 시료가 있었다면 더 완성도 높은 모델을 제시할 수 있었을 거라 사료된다.

두 명의 한국인 남성 DNA A, B 혼합물에 대하여 1:1, 1:3, 1:6, 1:9, 1:14, 1:19, 1:29, 1:49 비율 및 3:1, 6:1, 9:1, 14:1, 19:1, 29:1, 49:1의 비율로 혼합하였다. 이를 통해 대립유전자형 차이가 1인 조합에서 예를 들어 A가 15, B가 14형을 가질 경우, A:B 1:3의 비율에서 A의 비율이 낮기 때문에 A의 15는 B의 14에 영향을 주지 않는다. 하지만 3:1인 경우에는 A의 비율이 높아지고 A의 15가 한 반복 수 작은 14의 stutter를 발생시키는 데 발생한 stutter는 14형을 주 대립유전자형으로 가지는 B의 리드 수에 영향을 주게 된다. 따라서 모든 조합에 1:3, 3:1과 같이 비율을 두 번 적용하면 stutter 영향을 받거나 받지 않는 두 경우를 모두 고려할 수 있게 된다. 또한 한 개 세트 조합에서 한 비율에 대해 두 개의 혼합 시료가 생성 되고 이중 반복 실험을 하게 되는 효과도 발생하여 혼합 비율을 효율적으로 적용할 수 있다.

정확한 혼합 시료 비율을 확인하기 위하여 Quantifiler[®] Trio DNA Quantification Kit (Thermo Fisher scientific)의 정량 값을

이용하여 시료를 1:1 혼합하였고, PowerPlex® Y23 Kit (Promega)을 통해 peak height를 확인하여 시료 간의 균형을 보았다. 이 방법으로 실제 혼합된 두 시료가 각 혼합 비율의 이론적인 값과의 차이를 확인하여 정량 값이 잘 적용되었는지 알 수 있었고, 오차가 클 경우에는 분석 대상에서 제외할 수 있었다.

현재 법의유전학 영역에서 NGS 자료 분석에 대해 확립된 분석 방법은 없고 아직까지 논의 되고 있는 상황이다. 이번 연구에서는 참조서열을 이용한 정렬(Alignment)분석 방법이 아닌 각 Y-STR 반복 서열 정보를 이용하여 시퀀싱 리드 형태를 특정하고 분류하는 방식을 이용하였다. 이 방법은 참조 서열을 이용한 정렬 분석 방법에서 가지는 한계점을 극복하는 방법으로 실제 리드가 가지는 특성을 더 잘 보여줄 수 있는 분석 방법이라고 생각한다. 하지만 이번 연구에서는 분석된 염기 서열 정보를 활용하지 못하고 리드 사이즈 정보를 이용한 분석만 이루어졌기 때문에 후속 연구로는 염기 서열을 포함한 분석을 통해 혼합 시료 분석에서 NGS 기법이 가지는 장점을 더욱 더 잘 활용할 수 있을 것이라 생각한다.

한국인 남성 단일 시료 및 혼합 시료 185개 NGS 시퀀싱 결과 중 3개 시료에서 현저히 낮은 리드 수를 가지는 것을 확인하고 이는 분석 대상에서 제외하였다. 리드 수를 확보하지 못한 3개 시료를

제외한 182개 시료에 대하여 자료 분석을 진행하였다. 리드 수 획득에 실패한 경우에는 다른 시료와 Index 중복으로 의심할 수 있다. 따라서 자료 분석에서 리드 수를 확보 하지 못한 3개 시료(NGS002, 10-19vs1, 11-19vs1)가 가지는 대립유전자형이 다른 시료에 포함되었는지 확인하였지만 구분할 수 없었다. 따라서 라이브러리 제작과정 중 표준화 단계에서나 마지막 단계인 모든 시료를 혼합하는 과정에서 포함되지 않은 것으로 사료된다.

단일 시료를 이용한 stutter를 조사하여 비율을 제시하였다(표 6). 반복 단위 염기수에 따른 stutter의 평균 비율은 일반적으로 염기 수가 증가할수록 stutter 발생 비율이 낮은 것으로 알려져 있다.²⁶ 그러나 6염기 반복 단위를 가지는 DYS448의 stutter 평균 비율이 7.38%(N-1), 0.82%(N-2)로 5염기 반복 단위를 가지는 DYS438, DYS643의 stutter 평균 비율인 6.64%(N-1), 0.44%(N-2)보다 높게 관찰 되었다. 이는 DYS448이 21개 Y-STR 중 가장 낮은 리드 수를 획득하였기 때문에 stutter 평균 비율 계산에서 오차의 영향을 받는 것으로 생각된다.

노이즈는 시료의 대립유전자형 및 stutter에 해당하는 리드를 제외된 위치에 존재하는 리드들 각 Y-STR 전체 리드 수에서 차지하는 비율을 나타낸 값이다. 이는 유전자형 결정에서 제거되어야 할 시퀀싱 에러나 대립유전자형 범위를 벗어나는

크기의 증폭 산물의 리드를 포함한다. 단일 시료 분석에서 시료의 대립유전자형 및 stutter 분석 다음으로 노이즈 비율을 조사하지만, 실제적으로 NGS 자료를 분석할 때 첫 번째로 고려되어야 할 부분이 노이즈이다. 노이즈 분석을 통해 대립유전자형 범위를 제한할 수 있으며, 범위 안에서 stutter가 고려되어 대립유전자형을 결정할 수 있게 된다.

조사된 stutter 및 노이즈 비율을 이용하여 분석적 한계치 및 해석적 한계치를 절대값이 아닌 비율로 제시하였다. 이는 각 Y-STR의 coverage가 완전히 균등하지 않기 때문에 한계치를 절대값으로 제시할 경우 coverage가 낮은 Y-STR에 맞추는 경우엔 대립유전자형 획득을 유발하거나 coverage가 높은 Y-STR에 맞추는 경우에는 대립유전자형 소실을 유발할 수 있다. 따라서 각 Y-STR의 coverage를 고려하여 stutter와 노이즈가 차지하는 비율을 이용하여 불필요한 리드를 제거하였다.

N-1 위치에 생성되는 stutter는 평균 12.37%이며 표준 편차를 고려하여 제시된 해석적 한계치 비율은 평균 17.05%이다. 이는 혼합된 두 대립유전자형이 반복 수 차이가 1인 경우에는 1:4 이상으로 혼합 된 경우에는 해석이 불가능 하다는 것을 말해준다. N-1 stutter 비율이 가장 높은 DYS481의 경우에는 해석적 한계치 비율이 39.62%로 1:1로 혼합된 경우를 제외하면 해석에 어려움이

있음을 알 수 있다. 따라서 혼합 시료의 혼합된 두 대립유전자형의 반복 수 차이가 1인 Y-STR이 포함 된 경우에는 대립유전자형 결정에 stutter의 영향이 크기 때문에 혼합 시료 분석에 주의할 필요가 있다.

각 한계치 비율을 반영하여 리드를 필터하고 대립유전자형을 조사한 결과에서 DYS390, DYS392, DYS456, DYS481, DYS570, DYS576, DYS635, DYS643 유전자 좌에서 실제 시료가 가지지 않는 대립유전자형이 하나 이상 관찰되었으며, 이는 혼합된 시료가 가지는 대립유전자형 및 stutter 위치를 제외한 곳에서 나타났다. 첫 번째로, stutter 및 노이즈 비율을 조사에서 사용한 18개 단일 시료는 각 시료에 특성을 반영하기에 부족함이 있다. 두 번째로는 리드 필터를 위한 한계치 비율은 평균비율+표준편차x3을 이용하였는데, 이 값은 단일 시료 분석을 통해 도출한 평균 및 표준편차를 이용한 것이기 때문에 혼합 시료에 적용하는 데에 충분한지 검토해 볼 필요성이 있다. 세 번째는 CE 기법과 NGS 기법에서 관찰되는 stutter의 범위가 다른 것으로 보인다. 이는 CE 기법을 이용한 stutter 조사에서는 모든 Y-STR에 대하여 peak가 관찰 된 위치만 고려하여 N-2, N-1, N+1의 위치로 단정하고 있다. 그러나 NGS 분석에서는 한 개의 리드의 존재도 확인할 수 있기 때문에 관찰할 수 있는 stutter 발생 분포에 차이를 보인다. 따라서

NGS 기법을 이용한 분석에서는 먼저 각 Y-STR에 대한 stutter 범위를 조사해볼 필요가 있다.

혼합 시료의 결정된 대립유전자형을 각 Y-STR, 혼합 비율 별로 살펴보면 각 Y-STR 마다 대립유전자형 소실이 일어나는 혼합 비율이 다른 것을 알 수 있었다. DYS458, DYS635의 경우에는 1:6의 비율에서부터 대립유전자형 소실이 관찰 되었고 DYS19의 경우에는 1:19비율까지 대립유전자형을 완전히 얻을 수 있었다. 이는 동일한 PCR조건 내에서도 각 Y-STR마다 PCR 효율이 다르다는 것을 알 수 있다. 또한 1:14의 비율까지는 21개 Y-STR 중 90% 이상에서 대립유전자형을 획득할 수 있음을 확인하였고 1:29부터는 59.06%로 대립유전자형을 확인 가능한 Y-STR이 줄어드는 것을 통해 1:29 비율 즉, 33pg의 미량 시료를 이용한 PCR 과정에서 21개 Y-STR 중 절반에서 확률적 효과 영향이 크게 나타날 수 있음을 알 수 있었다.

혼합 시료에서 결정된 두 대립유전자형이 가지는 리드 수를 이용하여 각 Y-STR별로 혼합 비율을 역으로 추정해 보았다(그림 2-1, 2-2). 대립유전자형 반복 수 차이가 1을 포함하는 Y-STR의 경우에는 제외하고 반복 수 차이가 2 이상을 가지는 경우만을 포함하였다. 따라서 두 시료가 가지는 대립유전자형 반복 수의

차이가 0 또는 1만 포함하는 DYS391, DYS437, DYS456은 혼합 시료 추정에서 제외되었다. 세 개 Y-STR을 제외한 18개 Y-STR에 대한 단순 상자 그림을 통해 1:1 비율부터 1:49의 비율로 두 시료의 혼합 비율이 커짐에 따라 상자 그림 역시 낮은 비율로 떨어지는 것을 볼 수 있다. DYS458은 1:1, 1:3, 1:6 혼합 비율까지 서로 중첩 되지 않고 이론적인 비율인 50%, 25%, 14.3%에 근접하게 위치 한 것을 볼 수 있었다. 그러나 대부분의 Y-STR에서는 각 혼합 비율이 상자와 중첩되며 두 시료의 리드 수의 비율이 이상적인 위치에 분포하지 않음을 알 수 있다. 따라서 특정한 Y-STR에서 리드 수를 이용하여 역으로 혼합 비율을 계산하는 데는 어려움이 있는 것으로 보인다.

혼합 시료 내 두 시료의 대립유전자형의 반복 수 차이에 따라 혼합 비율을 추정하였다(그림 3). 반복 수 차이가 2 이상인 대립유전자 형을 포함하는 Y-STR을 수집하여 각 비율 별로 리드 수를 이용하여 혼합비율을 추정하였다. 표 10의 결과를 참고하여 21개 Y-STR 중 90% 이상의 Y-STR에서 대립유전자형을 얻을 수 있는 1:14 비율까지만 고려하여 나타내었다. 미량 시료가 차지하는 비율을 나타내었으며 대립유전자형 차이가 많이 날수록 이론적인 혼합 비율과 가깝게 나타날 것으로 예상하였으나 이론적인 비율 값과 차이를 보였다. 이 계산에서는 대립유전자형 반복 수 차이에

해당하는 시료의 수가 다르기 때문에 반복 수 차이에 따른 동등한 비교가 이루어 질 수 없었다. 대립유전자형 반복 수 차이에 따른 혼합 비율 추정의 유용성을 검증하기 위해서는 동일한 반복 수를 가지는 조합의 시료 수를 조절할 필요가 있다고 생각한다.

현재까지 CE 기법은 많은 연구를 통해 분석 방법 및 소프트웨어가 표준화 되어있다. 하지만 CE 기법 자체적으로 감지능력에 한계가 있고, 사이즈 정보 만을 제공하여 혼합 시료 분석에서는 효율적이지 않다. 이러한 한계를 보완하고자 NGS 기법을 혼합시료에 이용하였고 혼합 시료 내 미량 시료 확인은 CE가 1:9,²⁶ NGS는 1:14 비율에서도 식별이 가능함을 확인하였다. 그러나 NGS에서는 단일 시료 자료 분석에 조차 표준화된 방법이 정립되어 있지 않고, 분석 프로그램에 대한 가이드라인조차 제시되어 있지 않다. 또한 염기 서열 정보 기반 분석을 위한 프로그램이 없기 때문에 분석 프로그램 개발이 필요하다. 따라서 염기서열 분석 기반의 프로그램이 구축되어 stutter와 노이즈가 분석 된다면 정밀한 한계치 비율을 제시할 수 있고, 혼합 시료 분석에 대한 NGS의 유용성을 보다 분명하게 검증할 수 있을 것이라고 생각한다.

이상으로 본 연구에서는 사이즈 기반 단일 시료 NGS 자료 분석을 통하여 혼합 시료 분석에서 필요한 분석적 한계치 비율 및 해석적 한계치 비율을 도입하여 혼합 시료 분석에서의 NGS 기법의 유용성을 평가해 보았다. NGS는 정량적인 분석이 가능하기 때문에 단일 시료 분석을 통해 각 Y-STR의 stutter 및 노이즈 비율을 계산할 수 있었다. 이를 이용하여 분석적 한계치 및 해석적 한계치 비율을 제시하고 도입하였으나 한계치 비율을 이용해 필터 되지 않은 대립유전자형 획득이 발생하였으며, 대립유전자형 소실이 발생한 Y-STR 유전자 좌는 각 다른 비율에서 발생하였다. 이러한 점을 보완하기 위한 후속연구가 필요하다고 생각한다. 따라서 추후 연구에서는 본 연구를 기반으로 NGS로 생성된 염기 서열 정보를 이용한 자료 분석이 이루어져야 하며, 이를 통해 혼합 시료 분석에 더 유용한 기준 값을 제시할 수 있을 것 이라고 생각한다.

V. 결론

1. 단일 시료를 이용하여 구축된 NGS 기반 Y-STR 체계에서 관찰되는 stutter를 조사하였다. N-2 stutter의 평균 비율은 1.24%, N-1 stutter 평균 비율은 12.37%, N+1의 평균 비율은 1.75%이었으며, N-3 stutter의 평균비율은 DYS392, DYS481에서만 조사하였으며 0.74%이었다.

2. NGS 리드의 분석적 한계치(AT, analytical threshold)와 해석적 한계치(IT, interpretational threshold)은 절대값이 아닌 각 마커 별 최대 리드에 대한 상대 비율로 제시될 수 있다. AT는 평균 0.44%이며 범위는 0.08% ~ 0.84%으로 DYS437에서 가장 낮고, DYS392에서 가장 높다. Stutter 위치에 따른 IT의 평균은 N-3 stutter 에서 1.57%, N-2에서 2.42%, N-1에서 17.5%, N+1에서 3.07%이다.

3. 15가지 혼합비율을 가지는 한국인 남성 혼합시료에 대하여 분석적 한계치와 해석적 한계치를 적용하여 NGS 기법을 이용한 혼합 시료의 Y-STR 분석에서 각 Y-STR은 다른 민감도를 보였으며, 최소 민감 마커(1:6)는 DYS458과 DYS635였고, 최대

민감 마커(1:19)는 DYS19였다. 1:14의 혼합 비율까지 대부분의(평균 90% 이상) Y-STR에서 유효한 유전자형을 획득하는 것을 확인하였다.

4. PCR 증폭이 수반된 NGS 기반 STR 분석에서도 STR의 stutter 생성과 혼합 시료 내 미량 시료 증폭에서의 확률적 효과(Stochastic effect)가 동반되기 때문에 NGS 자료 해석에 있어 두 요소의 영향을 고려해야 한다.

본 연구를 통하여 제시된 NGS 기반 Y-STR 분석법은 NGS를 이용한 혼합 시료 분석에 필요한 기초 자료를 제시하고 있으며 가능성을 보여주었다. 이 연구를 기반으로 염기서열 정보를 이용한 심도 있는 연구를 추가적으로 진행한다면 법의학 실무에서 접할 수 있는 혼합 시료 분석에 더 유용한 정보를 제공할 수 있을 것으로 기대된다.

참고문헌

1. Butler JM. Forensic DNA Typing: Biology, Technology and Genetics of STR Markers. 2nd ed. New York, Elsevier Academic Press; 2005
2. Oh YN, Lee HY, Lee EY, Kim EH, Yang WI, Shin KJ. Haplotype and mutation analysis for newly suggested Y-STRs in Korean father-son pairs. Forensic Sci Int Genet. 2015;15:64-8
3. Ge J, Budowle B, Chakraborty R. Interpreting Y chromosome STR haplotype mixture. Legal Med. 2010;12:137-43
4. Clayton TM, Whitaker JP, Sparkes R, Gill P. Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Sci Int. 1998;91:55-70
5. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Sci Int. 2006;160(2-3):90-101
6. Andersen MM, Eriksen PS, Mogensen HS, Morling N. Identifying the most likely contributors to a Y-STR mixture using

the discrete Laplace method. *Forensic Sci Int Genet.* 2015; 15:76-83

7. Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA et al. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *J Forensic Sci.* 2009;54(4):810-21

8. Butler JM. *Advanced Topics in Forensic DNA Typing: Interpretation*, San Diego, AP press; Elsevier, 2014. p.38

9. Butler JM. *Advanced Topics in Forensic DNA Typing: Interpretation*, AP press, Elsevier, 2014. p.34

10. Butler JM. *Advanced Topics in Forensic DNA Typing: Interpretation*, AP press, Elsevier, 2014. p.76

11. Fordyce SL, Mogensen HS, Børsting C, Lagacé RE, Chang CW, Rajagopalan N, et al, Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGMTM, *Forensic Sci. Int. Genet.* 2015;14:132-40.

12. Kim EH, Lee HY, Yang IS, Jung SE, Yang WI, Shin KJ, Massive parallel sequencing of 17 commonly used forensic autosomal STRs and amelogenin with small amplicons, *Forensic Sci. Int. Genet.* 2016;22:1-7

13. Bright JA, Buckleton JS, Taylor D, Fernando MA, Curran JM. Modeling forward stutter: toward increased objectivity in forensic DNA interpretation. *Electrophoresis*. 2014;35:3251-57
14. Bright JA, Curran JM. Investigation into stutter ratio variability between different laboratories. *Forensic Sci Int Genet*. 2014;13:79-81
15. Butler JM, Hill CR. Scientific Issues with Analysis of Low Amounts of DNA. *Profiles in DNA*. 2010;13(1)
16. Butler JM. *Advanced Topics in Forensic DNA Typing: Interpretation*, AP press, Elsevier, 2014. p.137
17. Lee EY, Lee HY, Oh SY, Jung SE, Yang IS, Lee YH et al. Massively parallel sequencing of the entire control region and targeted coding region SNPs of degraded mtDNA using a simplified library preparation method, *Forensic Sci Int Genet*. 2016;22:37-43
18. Warshauer DH, Lin D, Hari K, Jain R, Davis C, Larue B, et al. STRait Razor: A length-based forensic STR allele-calling tool for use with second generation sequencing data. *Forensic Sci Int Genet*. 2013;7:409-17

19. Kim EH, Lee HY, Yang IS, Jung SE, Yang WI, Shin KJ. Massive parallel sequencing of 17 commonly used forensic autosomal STRs and amelogenin with small amplicons, *Forensic Sci. Int. Genet.* 2016;22:1-7
20. Li H1, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-60
21. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357-9.
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078-9.
23. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics.* 2009;25:2283-5
24. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14:178-92

25. Thompson JM, Ewing MM, Frank WE, Pogemiller JJ, Nolde CA, Koehler DJ et al. Developmental validation of the PowerPlex® Y23 System: a single multiplex Y-STR analysis system for casework and database samples. *Forensic Sci. Int. Genet.* 2013;7:240-250
26. Butler JM. *Advanced Topics in Forensic DNA Typing: Interpretation*, San Diego, AP press; Elsevier, 2014. p.33

ABSTRACT

Analysis of Y Chromosomal STR Profiles in Mixed DNA
using Next Generation Sequencing

Soyeun Kwon

*Department of Medical Science
The Graduate School, Yonsei University*

(Directed by Professor Kyoung-Jin Shin)

Mixed DNA profiling has been consistently studied in forensic genetics, however, it still remains as a challengeable area in DNA profiling. Currently, mixed DNA profiling using capillary electrophoresis (CE) is based on the number of peak, relative peak height resulted from multiplex PCR system. With advent of next generation sequencing (NGS) producing massive sequence information, mixed DNA analysis is also expected to receive support from NGS by making up the limitations of CE analysis. In the present study, the utility of NGS method for mixed DNA analysis was

evaluated using the multiplex PCR system for the amplification of 24 Y chromosomal markers including the PowerPlex[®] Y23 loci and the M175 marker. Eighteen single-source male DNAs and 11 combinations of two-people DNA mixture from 13 Korean males in various ranges of mixture ratios were investigated. Stutter ratios and noise ratios were determined from single-source male DNAs and used for filtering out artifact reads as analytical and interpretational threshold ratios, respectively. Using threshold ratios investigated here, we determined the genotypes of mixed DNAs.

In this study, the useful threshold ratios for mixed DNA analysis in Y-STR NGS system were provided. The resolution power of each Y-STR for minor component detection in mixed DNAs was different and most Y-STRs were determined by 1:14. However, it was difficult to estimate mixture ratios using read counts of NGS and the stutter generation of STR and stochastic effect of small

amount of DNA should be considered in PCR-based NGS analysis.

In the present study, it provided the guidelines for mixed DNA analysis of Y-STRs using NGS method and it raised the possibility of the NGS analysis for the mixed DNA profiling. Furthermore, it is expected that NGS method can be applied to mixed DNA analysis in forensic casework.

Key words: Mixed DNA, Y chromosomal STR, Next generation sequencing, analytical threshold ratio, interpretational threshold ratio, stutter ratio