



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 한국인 미토콘드리아
전체 DNA 염기서열 변이 분석

연세대학교 대학원

의 과학과

김 보 민

차세대 염기서열 분석법을
이용한 한국인 미토콘드리아
전체 DNA 염기서열 변이 분석

연세대학교 대학원

의 과학과

김 보 민

차세대 염기서열 분석법을
이용한 한국인 미토콘드리아
전체 DNA염기서열변이 분석

지도교수 신 경 진

이 논문을 석사 학위논문으로 제출함

2019년 12월

연세대학교 대학원

의 과 학 과

김 보 민

김보민의 석사 학위논문을 인준함

심사위원 김 상 우 인

심사위원 신 경 진 인

심사위원 이 환 영 인

연세대학교 대학원

2019년 12월

감사의 글

어린시절 법유전학에 대한 호기심이 이어져 시작된 석사학위 과정이 어느덧 마무리를 짓게 되었습니다. 학위 과정 동안 수많은 우여곡절이 있었지만, 뒤돌아보면 모든 순간들이 각자의 의미가 있었던 것 같습니다. 매 순간 저보다 저를 더 많이 응원해주는 많은 분들이 있었기에 이렇게 한 단락을 잘 마무리할 수 있었습니다. 먼저, 늘 세심한 지도로 이끌어주신 신경진 교수님께 감사드립니다. 좋은 말씀들로 항상 관심있게 바라 봐주시는 이환영 교수님, 좋은 논문이 될 수 있도록 격려해주신 김상우 교수님께도 감사드립니다. 또한 많은 지도와 관심 주신 양우익 교수님과 박종필 교수님께도 감사드립니다. 학위 과정을 시작할 때부터 지금까지 너무나도 많은 도움주신 이은영 선생님, 권소연 선배님께도 감사드립니다. 옆자리에서 늘 본인의 일처럼 도와주시는 새롬언니, 항상 좋은 말씀 많이 해주시는 배수진 선생님, 제가 생각하는 바를 멋지게 실현시켜 주시는 전해인 선생님께도 감사드립니다. 또한 학위 과정 내내 동료이자 친구로 많은 날들을 함께 했던 미현이와 예림언니에게도 감사드립니다. 항상 같은 자리에서 응원해주는 사랑하는 우리 가족, 엄마, 아빠, 오빠 그리고 뽀미 덕분에 큰 힘이 되었습니다. 항상 버팀목이 되어준 은영, 지영이 그리고 멀리서도 항상 응원해주는 효진이와 수웅오빠도 감사드립니다. 새로운 소중한 인연들을 만나고, 또 늘 곁에 있던 사람들의 소중함을 알게 된 시간들이었습니다. 제가 받았던 응원들과 도움들을 잊지 않고, 저도 배운 것처럼 베풀며 노력하는 사람이 되도록 노력하겠습니다.

차 례

국문 요약.....	1
I. 서론	4
II. 재료 및 방법	8
1. 연구대상	8
2. DNA 추출 및 정량	9
3. CE 방법을 이용한 미토콘드리아 DNA 조절 영역 분석	10
4. 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 MPS 수행	12
가. 긴 범위 PCR을 이용한 미토콘드리아 DNA 전체 영역 증폭.....	12
나. MPS 라이브러리 (Library) 제작 및 수행.....	13
다. MPS 결과를 이용한 미토콘드리아 DNA 전체 영역 염기서열 변이 자료 분석	14
III. 결과	16
1. CE 방법을 이용한 미토콘드리아 DNA 조절 영역 분석 결과.....	16
2. Illumina Miseq 분석용 한국인 집단 시료의 미토콘드리	

아 DNA 전체 영역 라이브러리 제작 결과.....	18
3. 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 MPS 분석 결과.....	21
4. 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 염 기서열 변이 분석 결과.....	22
5. 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 염 기서열의 다형성 분석 결과.....	44
6. 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 점 이형질 변이 분석 결과.....	46
IV. 고찰	53
V. 결론	54
참고 문헌	66
ABSTRACT	71

그림 차례

그림 1. Mitochondrial DNA control region (CR) with hypervariable region (HV) commonly analyzed in forensic genetics.....	7
그림 2. T-brake in hypervariable region 1.....	17
그림 3. Two long-range PCR amplicon.....	19
그림 4. Verification of whole mtDNA MPS library.....	20
그림 5. The whole mtGenome coverage map obtained from MPS data	21
그림 6. The distribution of point heteroplasmy across the whole mtGenome of the Korean population	52

표 차례

표 1. Amplification and sequencing primers for analyzing mitochondrial DNA control region·····	11
표 2. Observed whole mtGenome haplotype and haplogroup of the Korean population ······	23
표 3. Comparison of haplotype diversity between control region and whole mtGenome in the Korean population ······	45
표 4. Observed point heteroplasmy across the whole mtGenome of the Korean population·····	47
표 5. The most frequently observed point heteroplasmy across the whole mtGenome in the Korean population ······	51
표 6. Comparison of discrimination capacity increment between the mitochondrial DNA control region and whole mtGenome in three U.S population and Korean population ······	56
표 7-1. The most frequently observed haplotype in the mitochondrial DNA control region·····	58
표 7-2. The whole mtGenome haplotype of the	

most frequently observed haplotype in the control region..... 58

표 8-1. The whole mtGenome haplotype observed three times in the Korean population 59

표 8-2. The whole mtGenome haplotype observed two times in the Korean population 60

국문요약

차세대 염기서열 분석법을 이용한 한국인 미토콘드리아 전체 DNA 염기서열변이 분석

법유전학 분야에서 미토콘드리아 DNA는 분해된 시료나 모근이 없는 머리카락과 같이 핵 DNA를 분석할 수 없는 경우에 동일 모계 확인을 위한 유용한 분석 대상이다. 그 동안의 미토콘드리아 DNA 분석은 Sanger 방식의 효율 한계 및 분석 가능한 시료 양의 제한으로 인해 조절 영역에만 집중되어 왔다. 하지만 미토콘드리아 DNA의 분석 범위를 전체 영역으로 확장하면, 조절 영역만을 대상으로 하였을 때보다 더 높은 식별력을 얻을 수 있음이 여러 연구를 통해 보고되었다.

한편, 최근 차세대 염기서열 분석법(Massively Parallel Sequencing, MPS)의 발전으로 미토콘드리아 전체 DNA를 효율적으로 분석하는 것이 가능해졌고, 기존의 Sanger 방법으로는 한계가 있었던 미토콘드리아 DNA의 이형질 변이를 분석하는 것도 보다 용이해졌다.

본 연구에서는 한국인 376명의 미토콘드리아 DNA 전체 영역을 긴 범위 PCR을 통해 두 개의 절편으로 나누어 증폭시키고, Nextera XT DNA Library preparation kit을 사용해 MPS 라이브러리를 생성하였다. MPS는 MiSeq 장비를 이용해 수행한 후, 생성된 자료는 GATK Mutect2 파이프라인을 이용해 분석되었다.

미토콘드리아 DNA의 분석 범위를 조절 영역에서 전체 영역으로 확장함으로써 조절 영역에 비해 추가로 58개의 하플로타입을 관찰할 수 있었고, 변별력 (Discrimination capacity)은 조절 영역 0.8138에서 전체 영역 0.9681로 증가하였다. 미토콘드리아 DNA 점 이형질 변이는 5%의 검출 한도로 분석되었으며 조절 영역 36개 및 암호 영역 94개로 총 130개가 관찰되었다. 가장 많이 관찰된 점 이형질 변이는 152번 위치로 총 5번 발견되었다. 미토콘드리아 DNA 길이 이형질 변이는 양방향 리드를 통합하는 기존의 MPS 자료 분석 방법으로는 한계가 있어, 단방향 리드 분석을 통해 특정화를 시도해 볼 필요가 있었다.

본 연구를 통해 미토콘드리아 DNA 식별력에 있어 전체 영역 분석의 효용성을 확인하였고, 점 이형질 변이에 대한 정보

를 축적함으로써 동일인 혹은 동일 모계 판단에 유용한 정보를 제공할 수 있었다. 미토콘드리아 DNA 전체 영역에 걸친 한국인 집단의 유전학적 통계량은 법과학 실무 기초 자료로 활용될 것으로 기대된다.

핵심되는 말: 차세대 염기서열 분석법, 전체 미토콘드리아 DNA, 한국인, 하플로타입, 이형질 변이

차세대 염기서열 분석법을 이용한 한국인 미토콘드리아 전체
DNA 염기서열변이 분석

<지도교수 신 경 진>

연세대학교 대학원 의과학과

김 보 민

I. 서론

세포 당 많은 복제 개수를 가지고 있는 미토콘드리아 DNA는 법 유전학 분야에서 유용한 분석대상으로 사용되고 있다.¹ 특히 분해된 시료나 모근이 없는 머리카락, 오래된 뼈와 같이 핵 DNA가 미량으로 존재하여 충분한 정보를 제공하지 못하는 경우에 미토콘드리아 DNA의 염기서열 분석은 동일 모계 확인을 위한 유용한 단서를 제공할 수 있다.² 뿐만 아니라, 미토콘드리아 DNA는 재조합 되지 않고 모계로만 유전되어 시료의 생물지리학적 기원을 추정하고자 할 때 중요한 정보를 제공할 수 있다.³ 그 동안 미토콘드리아 DNA 분석은 시료 양의 제한 및 Sanger 방식의 분석 효율 한계로 인해 전

체 16,569bp 중 단 1.1kbp에만 해당하는 조절 영역 (그림 1⁴)에만 집중되어 왔다.^{5,6} 최근, 차세대 염기서열 분석법 (Massively Parallel Sequencing, MPS)의 발전으로 미토콘드리아 DNA 전체 영역 염기서열을 고품질로 빠르게 분석하는 것이 가능해짐에 따라 이를 범유전학 분야에 적용하고자 하는 연구들이 다양하게 진행되었다.^{7,8} 또한 미토콘드리아 DNA 전체 영역 분석은 조절 영역 보다 더 많은 수의 염기서열 변이를 확인할 수 있어 시료 간의 식별력을 높일 수 있음이 여러 연구를 통해 보고되었다.^{9,10}

미토콘드리아 DNA의 고유한 특성인 이형질 변이를 분석할 때, Sanger 방식 기반의 모세관 전기 영동 (Capillary Electrophoresis, CE)은 관찰된 피크의 상대적 형광 단위 (Relative Florescent Unit, RFU)를 이용하여 시각적으로 판단해야 한다. 하지만 피크가 낮을 경우에는 CE의 노이즈(noise)와 점 이형질 변이를 정확히 구분하는 것이 매우 어렵다. 반면에 MPS는 리드 수 (read count)를 기반으로 분석되기 때문에 점 이형질 변이를 정확하게 구분할 수 있으며¹¹, 길이 이형질 변이 또한 그 양상을 살펴 특정화 할 수 있을 것으로 기대된다.¹² 뿐만 아니라, MPS는 한 번에 다량의 시료를 병행하게 분석할 수 있기 때문에 집단유전학적 통계량 산출과 같이 많은 시료를 분석해야 할 때 유용하게 사용될 수 있다. 현재까지 한국인 집단을 대상으로 수행된 미토콘드리아 DNA 선행 연구는 조

질 영역만을 대상으로 하고 있으며,¹³⁻¹⁵ Park¹⁰ 등은 MPS를 이용하여 한국인 미토콘드리아 전체 DNA 염기서열을 분석하였으나 특정 하플로그룹(Haplogroup) 만을 대상으로 하였기 때문에 법과학 실무에 기초 통계량으로 적용하는 것에 한계가 있다고 생각된다. 따라서 무작위로 선택된 한국인 집단을 대상으로 분석된 미토콘드리아 전체 DNA 염기서열의 분석 및 집단 유전학적 통계량에 관한 연구가 필요한 실정이다.

본 연구에서는 MPS를 이용하여 혈연관계가 없는 한국인 집단 376명을 대상으로 미토콘드리아 전체 DNA를 분석 후, 얻어진 정보로부터 염기서열 변이를 조사하여 개체 식별력을 평가하고자 하였다. 나아가 미토콘드리아 DNA의 점 이형질 변이에 대해 한국인 집단 내에서 빈도 및 양상을 조사하였고, MPS 자료를 대상으로 길이 이형질 변이를 효과적으로 분석할 수 있는 전략을 제시해보고자 하였다.

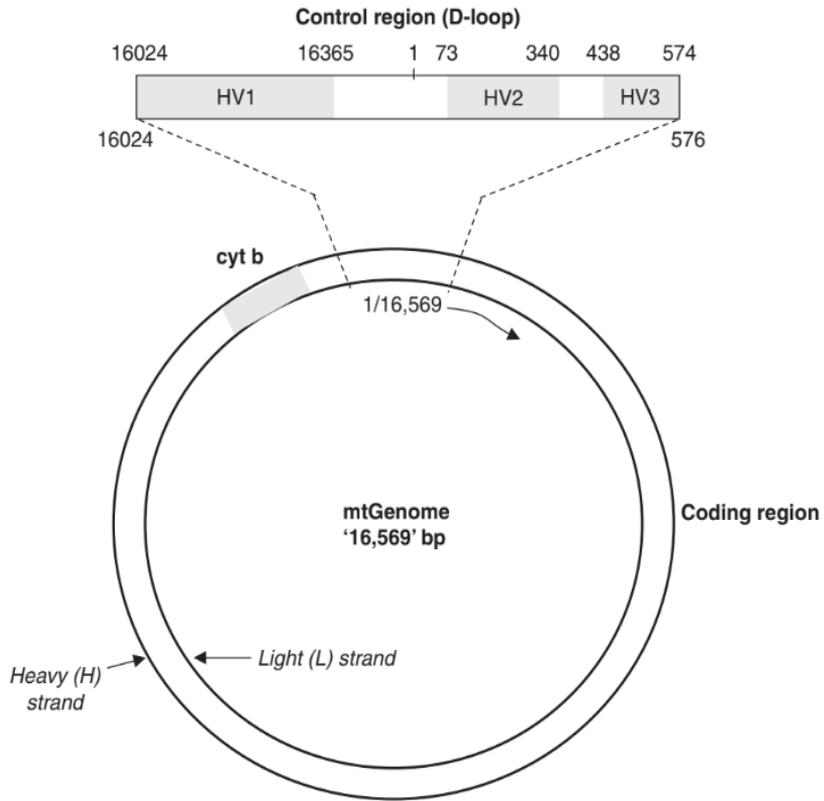


그림 1. Mitochondrial DNA control region (CR) with hypervariable region (HV) commonly analyzed in forensic genetics.

II. 재료 및 방법

1. 연구대상

본 연구에서는 MPS를 이용해 한국인 집단의 미토콘드리아 DNA 전체영역 염기서열을 분석하고자 혈연관계가 없는 376명의 한국인을 표본으로 무작위로 선정하였다. 총 376개의 시료 중 249개 시료는 Lee¹⁴등의 논문에서 사용된 시료 중 선택되었으며, 20개 시료는 Kim¹⁶등의 논문에서 사용된 시료를 대상으로 하였다. 나머지 107개 시료는 서울대학교 의과대학 시료은행(Seoul, Republic of Korea)으로부터 분양 받아 이 들로부터 혈액 또는 구강 상피 세포를 채취하였다. 혈액은 주사기를 이용하여 2.0 ml를 채취한 후, 즉시 EDTA 용액이 포함된 혈액 보관 튜브에 분주하였다. 구강 상피 세포는 피험자의 볼 안쪽을 소독된 면봉으로 2-4회 문질러 얻었다. 모든 시료는 DNA를 추출할 때까지 4℃ 냉장고에 보관하였다. 본 연구는 연세대학교 의료원 세브란스병원 연구심의위원회와 (승인번호: 4-2014-1025) (승인번호: 4-2015-1068) 서울대학교병원 연구심의위원회의 승인(승인번호: 1404-068-572)을 받아 수행되었다.

2. DNA 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 농도를 측정한 후 -20℃에 보관하였다. 시료는 1ng/ul로 희석하여 다음 단계 실험에서 사용되었다.

3. CE 방법을 이용한 미토콘드리아 DNA 조절 영역 분석

MPS 분석에 앞서 먼저 Sanger 기반의 CE 방법으로 미토콘드리아 DNA 조절 영역내 과변이 영역을 분석하였다. CE 분석을 위한 증합효소연쇄반응(PCR) 혼합물은 2.0ul의 Gold ST★R 10X PCR Buffer (Promega Corp., Madison, WI, USA), 0.6 uM의 프라이머 (primer) (표 1), 1.5 U의 AmpliTaq Gold DNA Polymerase (Thermo Fisher Scientific) 와 1ng의 DNA를 혼합하여 총 20ul로 준비하였다. 혼합물을 잘 섞어준 뒤, Veriti® 96-Well Thermal Cycler (Thermo Fisher Scientific)에 장착하여, 95°C에서 11분간 변성시키고 94°C에서 20초, 56°C에서 30초, 72°C에서 30초의 조건으로 35회 증합 반응 후 최종적으로 72°C에서 7분간 반응시킨 후 4°C에서 잠시 보관하였다. PCR 생성물은 은 2.0ul의 ExoSAP-IT™ PCR Product Cleanup Reagent (Thermo Fisher Scientific)를 첨가하여 37°C에서 45분, 80°C에서 15분 간 반응하여 부산물을 제거하였다. Sanger 방법을 이용한 염기서열 분석은 Macrogen, Inc (Seoul, Republic of Korea)에서 수행되었으며, 조절 영역 PCR 및 염기서열분석에 사용된 프라이머의 염기서열은 표1에 나열하였다. 시퀀싱 결과를 바탕으로 Seqscape v2.6 등을 이용해 참조 염기서열¹⁷(revised Cambridge Reference Sequence, rCRS)에 대비하여

염기서열 변이를 얻었다. 얻어진 염기서열 변이는 국제법유전학회 (International Society for Forensic Genetics, ISFG)가 제안한 형태¹⁸에 맞게 기록되었다. 확보된 미토콘드리아 DNA 조절 영역의 염기서열 변이를 바탕으로 EMPOP mtDNA database v4¹⁹를 통해 하플로그룹을 지정하였다.

표 1. Amplification and sequencing primers for analyzing mitochondrial DNA control region.

Primer	Usage	Sequence (5'→3')
F15975	Amplification & Sequencing	CTC CAC CAT TAG CAC CCA AA
R635	Amplification & Sequencing	GAT GTG AGC CCG TCT AAA CA
F16524	Sequencing	AAG CCT AAA TAG CCC ACA CG
R042	Sequencing	AGA GCT CCC GTG AGT GGT TA

4. 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 MPS 수행

가. 긴 범위 PCR을 이용한 미토콘드리아 DNA 전체 영역 증폭

한국인 집단 시료의 미토콘드리아 DNA 전체 영역 염기서열을 분석을 위해 미토콘드리아 전체 DNA를 두 개의 긴 범위로 나누어 PCR을 수행하였다. 긴 범위 PCR 혼합물은 2.0ul의 10X LA PCR Buffer II with Mg²⁺ (Takara Bio, Inc., Kusatu, Shiga, Japan)와 2.5 mM의 dNTP (Takara Bio, Inc.), 1.5 U의 Takara LA Taq HS (Takara Bio, Inc.)을 비롯해 Fendt²⁰ 등에 의해서 입증된 0.5uM의 프라이머와 함께 1ng의 DNA를 혼합하여 총 20ul로 준비하였다. 혼합물을 잘 섞어준 뒤, Veriti[®] 96-Well Thermal Cycler (Thermo Fisher Scientific)에 장착하여, 94℃에서 1분간 변성시키고 94℃에서 20초, 60℃에서 30초, 68℃에서 9분의 조건으로 26-28회 중합 반응 후 최종적으로 72℃에서 10분간 반응시킨 후 4℃에서 잠시 보관하였다. PCR을 통해 만들어진 증폭 산물은 Agilent 12000 kit (Agilent Technologies, Palo Alto, CA, USA)를 이용하여 Agilent 2100 Bioanalyzer (Agilent)를 통해 확인되었다. 각 시료당 두 개의 증폭 산물 존재 유무를 확인한 후 농도를 측정하였고, 다음 단계 실험을 위하여 1ng/ μ l로 희석하여 준비하였다.

나. MPS 라이브러리 (Library) 제작 및 수행

긴 범위 PCR로 얻은 미토콘드리아 전체 DNA 증폭산물 각각 1.25ng를 대상으로 Nextera XT library preparation kit (Illumina, San Diego, CA, USA)를 사용해 제조사의 지시에 따라 MPS 라이브러리를 제작하였다. 생성된 라이브러리는 KAPA Library Quantification Kit (KAPA Biosystems, Woburn, MA, USA)를 이용하여 7500 Real-Time PCR system (Thermo Fisher Scientific)에서 제조사의 지시에 따라 정량되었다. 정량된 라이브러리를 20nM의 농도로 동량 섞은 후, 0.6배의 AMPure Bead (Beckman Coulter, Brea, CA, USA)를 이용하여 제조사의 지시에 따라 정제하여 원하는 크기의 라이브러리만을 얻고 부산물을 제거하였다. 정제된 라이브러리는 다시 Agilent 12000kit (Agilent) 와 Agilent 2100 Bioanalyzer (Agilent)를 사용하여 확인하였다. 그 후 KAPA Library Quantification Kit (KAPA Biosystems)를 이용하여 정량하고, 측정 값에 따라서 최종적으로 10nM의 농도로 동량 섞어 준비하였다. 준비된 미토콘드리아 전체 DNA 라이브러리를 대상으로 Illumina MiSeq Reagent v3 Kit (Illumina) 와 MiSeq (Illumina)장비를 이용하여 300bp씩 양 끝을 읽는 (Paired-end read) 방법으로 제조사의 지시에 따라 분석하였다.

다. MPS 결과를 이용한 미토콘드리아 DNA 전체 영역 염기서열 변이 자료 분석

MiSeq (Illumina) 장비로부터 얻은 MPS결과는 FASTQ 파일 형태로 생성되었다. 얻어진 FASQ file은 Cutadapt v 1.18²¹을 이용하여 어댑터 시퀀스를 제거하고, read quality 20이상이 되도록 트리밍 (trimming)하였다. 정돈된 FASTQ file은 BWA-mem²²을 이용하여 rCRS에 맞춰 정렬하였다. 다만 참조 염기 서열에 대한 일반적인 정렬 방식은 원형의 미토콘드리아 DNA로부터 선형의 리드로 구성할 때 중단점 (Break point)이 생기므로, 이를 포함하는 조절 영역의 커버리지에 영향을 미칠 수 있다. 따라서 이러한 문제점을 보완하기 위해 조절 영역과 암호 영역을 분리하여 각각 참조 염기 서열과 회전된 참조 염기 서열에 맞추어 정렬하였다. 이 후 GATK²³ Base Recalibration 모듈을 사용하여 base quality를 보정하고, GATK MuTect2의 미토콘드리아 DNA 변이 호출 모드를 이용해 염기서열 변이를 얻었다. 시퀀싱 오류로 인한 불확실한 변이는 GATK의 FilterMutectCalls 모듈을 사용하여 필터링(filtering)하였다. 염기서열 변이는 커버리지가 100X 이상인 것에 한하여 호출하였고, 얻어진 염기서열 변이는 국제법유전학회 (International Society for Forensic Genetics, ISFG)가 제안한 형태¹⁸에 맞게 기

록되었다. 점 이형질 변이는 전체 커버리지가 400X 이상이며, 소수 변이의 커버리지가 전체 커버리지의 5%이상인 것에 한하여 호출하였고, 길이 이형질 변이는 본 연구의 보고 대상에서 제외되었다. 각 시료로부터 얻은 미토콘드리아 DNA 전체 염기서열 변이를 바탕으로 EMPOP mtDNA database v4(<http://empop.online>)¹⁹에 입력하여 하플로그룹을 지정하였다.

III. 결과

1. CE 방법을 이용한 미토콘드리아 DNA 조절 영역 분석 결과

본 연구에서는 MPS 분석에 앞서 한국인 집단 시료의 미토콘드리아 DNA 전체 염기서열 변이 결과의 신뢰도를 확보하기 위해 Sanger 기반의 CE 방법으로 미토콘드리아 DNA 조절 영역 내 과변이 영역을 분석하였다. Sanger 방법을 이용해 분석된 조절 영역 염기서열 변이는 MPS 결과에서 모두 동일하게 관찰되어 확보된 자료의 완결성을 확인할 수 있었다. 특히 염기서열 변이를 전기영동도를 통해 시각적으로 확인함으로써 과변이 영역1과 과변이 영역2의 T염기 브레이크(break)이 소실 여부를 보다 정확히 파악할 수 있었다 (그림 2).

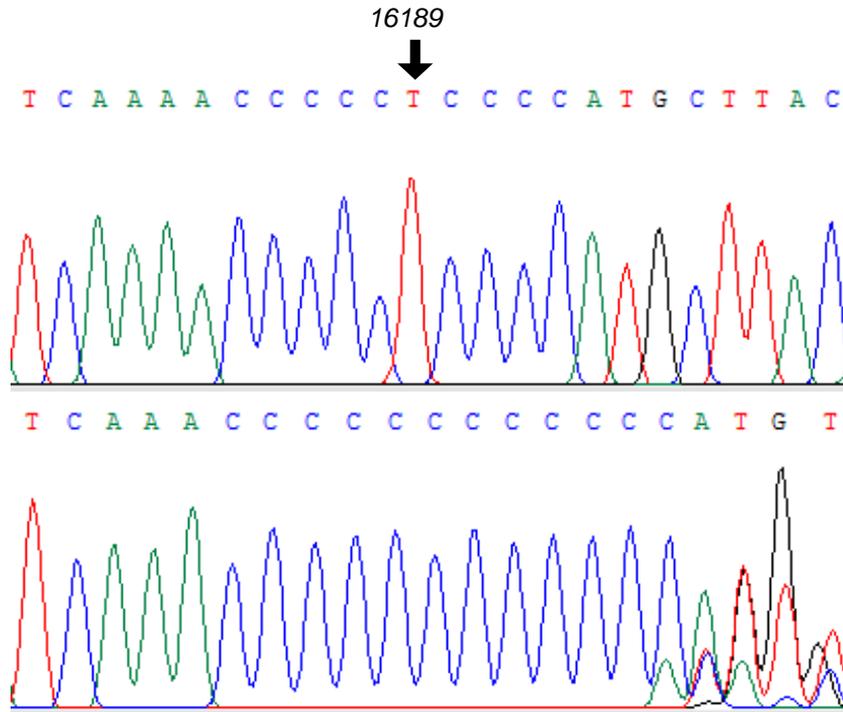


그림 2. T-brake in hypervariable region 1. The T-brake in hypervariable region 1 usually located on np 16189. The loss of the T-brake is described in the electropherogram below.

2. Illumina MiSeq 분석용 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 라이브러리 제작 결과

본 연구에서는 한국인 집단 시료의 미토콘드리아 전체 DNA에 대해 긴 범위 PCR을 수행하였다. 2100 bioanalyzer (Agilent Technologies)로 PCR 생성물 확인 결과 참조 염기서열을 기준으로 각각 8,379bp와 8,604bp 크기의 두 개 절편으로 증폭 산물이 생성되었음을 확인하였다 (그림 3). 얻어진 증폭산물로부터 Nextera XT DNA library preparation kit (Illumina)를 이용해 MPS 라이브러리를 제작하였다. 그 결과 800~1500bp 내에 라이브러리가 성공적으로 형성되었으며, 약 800~1000bp 부근에서 가장 높은 농도를 갖는 것으로 확인되었다 (그림 4). 미토콘드리아 DNA 전체 영역 라이브러리는 KAPA quantification kit (KAPA Biosystems)을 이용한 정량을 통해 적절하게 생성되었음을 확인하였다.

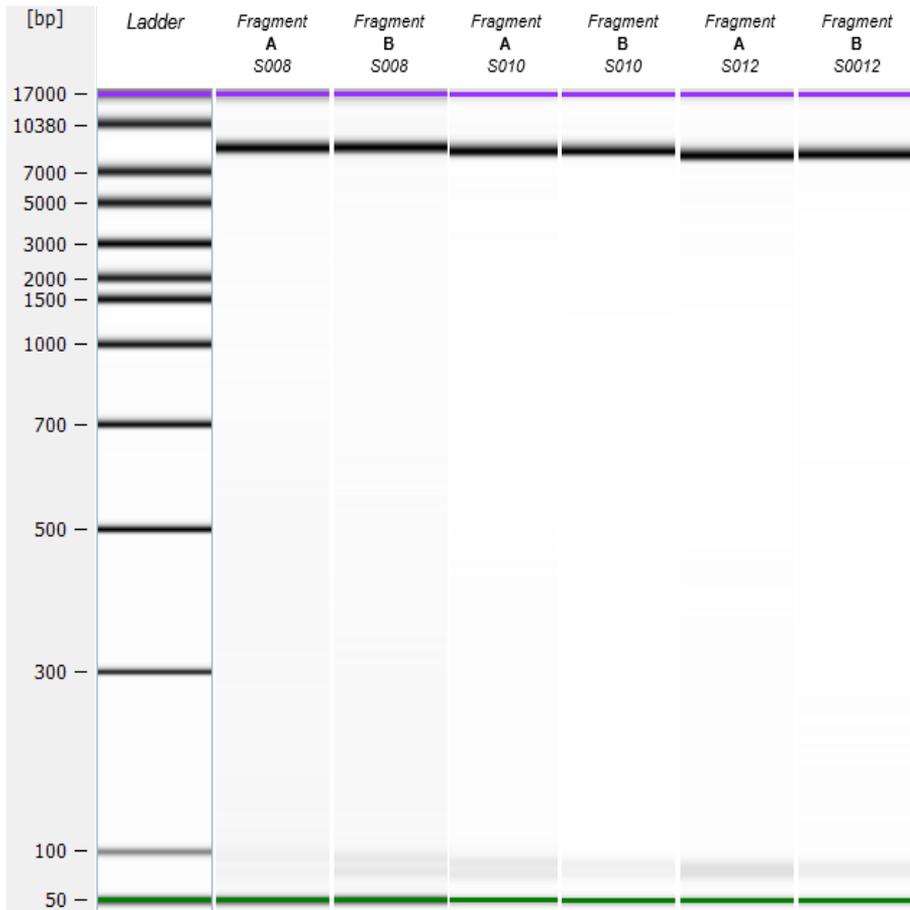


그림 3. Two long-range PCR amplicon. Each of the two fragments were verified by the 2100 bioanalyzer (S008, S010, and S012).

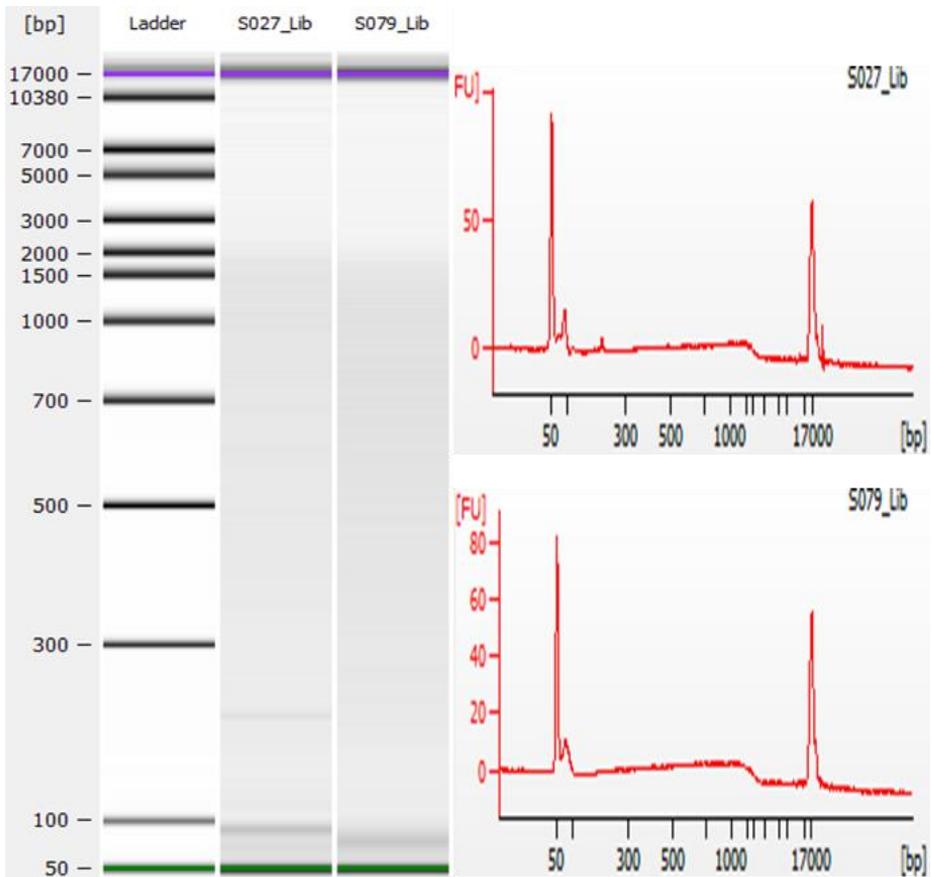


그림 4. Verification of whole mtDNA MPS library. The libraries were constructed by Nextera XT DNA library preparation kit using the 2100 bioanalyzer (S007, S079).

3. 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 MPS 분석 결과

생성된 라이브러리는 MiSeq (Illumina) 장비를 이용하여 MPS 분석되었다. 총 376 개의 한국인 집단 시료는 미토콘드리아 DNA 전체 영역에 걸쳐 평균 7569X 의 커버리지를 보였다 (그림 5). 염기서열 변이가 호출된 모든 영역에서 100X 이상의 커버리지가 확보되었고, 연속된 C 가 포함된 과변이 영역 2 (300~350bp)의 커버리지는 다른 영역에 비해 다소 낮게 관찰되었다.

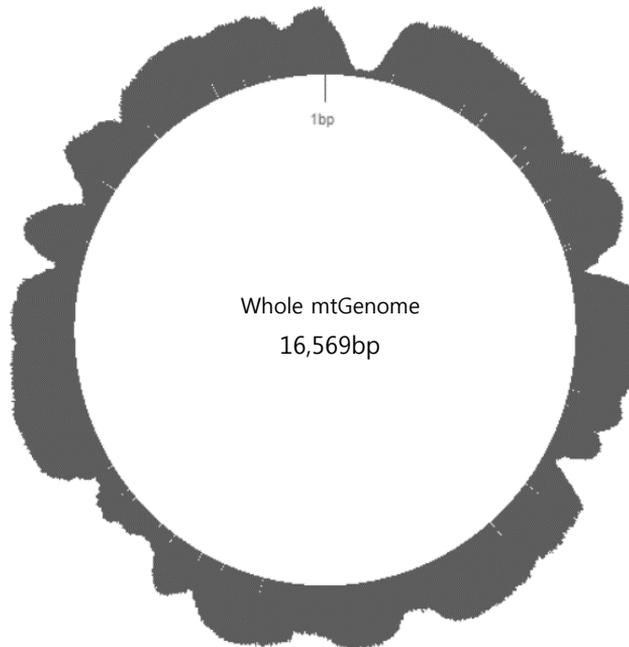


그림 5. The whole mtGenome coverage map obtained from MPS data.

4. 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 염기서열 변이 분석 결과

본 연구에서 얻어진 MPS 결과는 GATK Mutect2 파이프라인을 이용해 분석되었다. 미토콘드리아 DNA 길이 이형질 변이를 제외한 모든 염기 서열 변이는 국제범유전학회에서 제시한 방법에 맞게 기록되었다. 미토콘드리아 DNA의 분석 범위를 조절 영역에서 전체 영역으로 확장함으로써 훨씬 더 많은 염기서열 변이를 얻을 수 있었으며, 이들로부터 EMPOP DNA database v4를 이용해 세밀한 하플로그룹을 지정할 수 있었다. 총 376개의 분석대상 중 127개 시료의 하플로그룹, 암호 영역의 삽입과 결실 (indel) 및 전체 영역 하플로타입을 표 2에 정리하였다.

표 2. Observed whole mtGenome haplotype and haplogroup of the Korean population

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
1	S001	M8a2c	960DEL	73G 152C 263G 315.1C 489C 750G 960DEL 1438G 2706G 2835T 3290C 4715G 4769G 5297T 6179A 7028T 7196A 8584A 8684T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 12705T 14470C 14766T 14783C 15043A 15244G 15301A 15326G 15487T 16223T 16298C 16319A
2	S002	M10a1b	3172.1C	73G 152C 263G 315.1C 489C 709A 750G 1438G 2706G 3172.1C 4140T 4769G 5102G 7028T 7250G 8701G 8793C 8856A 8860G 9540C 9770C 10398G 10400T 10646A 10873C 11024T 11719A 12549T 12705T 13152G 14502C 14766T 14783C 15040T 15043A 15071C 15218G 15301A 15326G 16066G 16093C 16223T 16311C
3	S003	M7b1a1a1	-	73G 150T 199C 263G 315.1C 489C 750G 1438G 2706G 4048A 4071T 4164G 4769G 5351G 5460A 6455T 6680C 7028T 7684C 7853A 8701G 8860G 9540C 9824C 10345C 10398G 10400T 10873C 11719A 12405T 12705T 12811C 14766T 14783C 15002A 15043A 15301A 15326G 16129A 16182C 16189C 16223T 16297C 16298C
4	S004	G2a1e	-	73G 263G 315.1C 489C 709A 750G 1438G 2706G 4769G 4833G 5108C 5601T 7028T 7600A 8541A 8701G 8860G 9377G 9540C 9575A 10398G 10400T 10873C 11719A 12705T 13563G 14200C 14569A 14766T 14783C 15043A 15301A 15326G 16051G 16150Y 16223T 16278T 16362C
5	S005	F4a1a	-	73G 146C 152C 207A 249DEL 263G 315.1C 750G 1438G 2706G 3290C 3699G 3970T 4769G 5263T 6386Y 6392C 6852A 7028T 7861C 8860G 9007G 10310A 10915C 11038G 11719A 12153T 12396C 12408C 12630A 13602C 13928C 14766T 15326G 15670C 16093C 16207G 16304C 16311C 16362C 16399G 16497G
6	S007	F1b1a	-	73G 152C 249DEL 263G 315.1C 523DEL 524DEL 629C 750G 1438G 2706G 3970T 4732G 4769G 5147A 6392C 6962A 7028T 8860G 10310A 10609C 10976T 11719A 12406A 12633T 12882T 13928C 14476A 14766T 15326G 16129A 16182C 16183C 16189C 16232A 16249C 16304C 16311C 16344T 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
7	S008	D4a3b	-	73G 152C 263G 315.1C 489C 750G 1189C 1438G 2706G 3010A 3206T 4769G 4883T 5178A 5466G 7028T 7912A 8414T 8473C 8701G 8860G 9254G 9540C 10398G 10400T 10873C 11719A 12705T 14668T 14766T 14783C 14979C 15040T 15043A 15301A 15326G 16093C 16129A 16223T 16249C 16362C
8	S009	M9a1a1a	-	73G 263G 315.1C 489C 750G 1041G 1438G 2706G 3394C 4491A 4769G 7028T 8701G 8860G 9242G 9540C 10398G 10400T 10873C 11719A 11963A 12705T 13731G 14308C 14766T 14783C 15043A 15301A 15326G 16223T 16234T 16316G 16362C
9	S010	M7c1a2a1	-	73G 146C 199C 263G 315.1C 489C 523DEL 524DEL 750G 1438G 2706G 3882A 4071T 4769G 4850T 4917R 5442C 6053T 6365C 6455T 7028T 7961C 8538C 8701G 8860G 9540C 9824C 10398G 10400T 10873C 11665T 11719A 12091C 12705T 12804C 13269G 14182Y 14755T 14766T 14783C 15043A 15301A 15326G 16223T 16294T 16295T 16519C
10	S012	D4j+ 16311	-	73G 263G 315.1C 489C 750G 769A 1438G 2706G 3010A 4769G 4883T 5178A 7028T 7080Y 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11696A 11719A 12705T 14587G 14668T 14766T 14783C 15043A 15301A 15326G 15758R 16171G 16223T 16291T 16311C 16362C
11	S014	D411a1	-	73G 263G 315.1C 489C 750G 1310T 1438G 2706G 3010A 3535C 4769G 4883T 5147A 5178A 7028T 8414T 8701G 8860G 9355G 9540C 10398G 10400T 10427A 10873C 11719A 12705T 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16362C 16368C
12	S015	N9a1	-	73G 150T 200G 263G 315.1C 750G 789C 1438G 2706G 4386C 4769G 5231A 5417A 7028T 7124G 8860G 11719A 12007A 12358G 12372A 12705T 14766T 15326G 16111T 16129A 16223T 16257A 16261T

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
13	S017	D4c1a	191.1A	73G 191.1A 194T 199C 207A 263G 315.1C 489C 750G 1438G 2706G 2766T 3010A 3391A 4769G 4883T 5178A 7028T 7375R 8414T 8701G 8860G 9540C 9755A 9977C 10398G 10400T 10873C 11719A 12705T 14668T 14766T 14783C 15043A 15301A 15326G 16245T 16362C
14	S018	Y1	-	73G 146C 263G 315.1C 709A 750G 1438G 2706G 3834A 4562G 4769G 5417A 7028T 8392A 8803G 8860G 10379G 10398G 11719A 12705T 14178C 14693G 14766T 15326G 16126C 16231C 16266T 16319A 16399G 16519C
15	S019	D4j+ 16311	-	73G 263G 315.1C 489C 750G 769A 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11696A 11719A 12705T 13225A 14668T 14766T 14783C 15043A 15301A 15326G 16171G 16223T 16291T 16311C 16362C
16	S020	G3a2a	-	73G 143A 152C 263G 315.1C 489C 709A 750G 1438G 2706G 4769G 4833G 5108C 6086C 6221C 7028T 7621C 8701G 8860G 9540C 10398G 10400T 10873C 10969T 11719A 12705T 14569A 14766T 14783C 15043A 15301A 15326G 15746G 16189C 16223T 16265C 16274A 16362C 16465T
17	S021	M7c1a3	5752DEL	73G 146C 199C 263G 315.1C 489C 523DEL 524DEL 750G 1438G 2706G 3882A 4071T 4769G 4850T 5442C 5752DEL 6455T 7028T 8701G 8860G 9540C 9824C 10398G 10400T 10873C 11665T 11719A 12091C 12705T 14766T 14783C 15043A 15301A 15326G 16223T 16295T 16319A 16519C
18	S022	A5a	2156.1A	73G 195C 235G 263G 315.1C 523DEL 524DEL 663G 750G 1438G 1736G 2156.1A 2706G 4248C 4655A 4769G 4824G 7028T 8563G 8794T 8860G 11536T 11647T 11719A 12705T 14766T 15326G 15519C 16187T 16223T 16290T 16319A 16519C
19	S023	G2a1c2	-	73G 263G 315.1C 489C 709A 750G 1041G 1438G 2706G 4769G 4833G 5108C 5601T 7028T 7337A 7600A 8701G 8860G 9377G 9540C 9575A 10398G 10400T 10873C 11719A 12705T 13563G 14200C 14569A 14766T 14783C 15043A 15301A 15326G 16183C 16189C 16194G 16195C 16223T 16227G 16278T 16362C 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
20	S025	D4e1a1	-	73G 94A 194T 263G 315.1C 489C 750G 1438G 2706G 3010A 3316A 4769G 4883T 5178A 5964C 7028T 7269A 8414T 8701G 8860G 9536T 9540C 10398G 10400T 10873C 11215T 11719A 12705T 14207R 14470C 14668T 14766T 14783C 15043A 15301A 15326G 15924G 16093C 16176T 16223T 16362C
21	S026	M7c1a4a	5899.1C	73G 146C 199C 263G 315.1C 489C 523DEL 524DEL 750G 1438G 2706G 3882A 4071T 4769G 4850T 5054A 5442C 5899.1C 6455T 7028T 8701G 8860G 9540C 9797C 9824C 10398G 10400T 10873C 11665T 11719A 11815A 12091C 12705T 12906A 14766T 14783C 15043A 15301A 15326G 16172Y 16173T 16223T 16295T 16362C 16519C
22	S027	Y1a	-	73G 146C 263G 315.1C 750G 1438G 2706G 3834A 4769G 5417A 5979A 7028T 7933G 8392A 8860G 10398G 11719A 12705T 14178C 14693G 14766T 15326G 16126C 16231C 16256Y 16266T 16519C
23	S028	Z3	-	73G 143A 152C 249DEL 263G 315.1C 489C 750G 1438G 2706G 3618C 4715G 4769G 6752G 7028T 7196A 8584A 8701G 8860G 9090C 9540C 10208C 10398G 10400T 10873C 11719A 11864C 12705T 12810G 13359C 13620C 14766T 14783C 15043A 15301A 15326G 15487T 15784C 15820T 16185T 16209C 16223T 16260T 16298C
24	S029	M7a1a1	5899.1C	73G 263G 315.1C 489C 523DEL 524DEL 750G 1438G 2626C 2706G 2772T 4386C 4769G 4958G 5899.1C 6455T 7028T 8701G 8860G 9540C 9824C 10398G 10400T 10589A 10873C 11017C 11084G 11719A 11722C 12705T 12771A 14364A 14766T 14783C 15043A 15301A 15326G 16209C 16223T 16324C
25	S030	F1g	-	73G 152C 249DEL 263G 315.1C 523DEL 524DEL 750G 1438G 2389T 2706G 3398C 3970T 4769G 6392C 6827C 6962A 7028T 8285T 8860G 8902A 10310A 10609C 11719A 12406A 12882T 13928C 14071G 14766T 15326G 16092C 16183C 16189C 16304C 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
26	S031	M7b1a1h	-	73G 150T 199C 263G 315.1C 489C 524.1A 524.2C 750G 1438G 2706G 4048A 4071T 4164G 4769G 5330T 5351G 5460A 6455T 6680C 7028T 7684C 7853A 8701G 8860G 9540C 9824C 10397G 10398G 10400T 10873C 11719A 12405T 12705T 12811C 14766T 14783C 15043A 15301A 15326G 16086C 16129A 16192T 16223T 16297C
27	S032	D4g2	-	73G 263G 298T 315.1C 489C 750G 1438G 1555R 2706G 3010A 3398C 4769G 4883T 5178A 5231A 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 12705T 13104G 14668T 14766T 14783C 15043A 15301A 15326G 16189C 16223T 16362C 16519C
28	S033	D4i1	-	73G 263G 315.1C 489C 750G 1438G 2706G 3010A 4769G 4883T 5178A 5773A 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 12705T 14668T 14766T 14783C 15043A 15299C 15301A 15326G 16179T 16223T 16263C 16294T 16362C 16519C
29	S034	D4t	-	73G 315.1C 489C 750G 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 12705T 13356C 13928A 14668T 14766T 14783C 15043A 15301A 15326G 15901G 16223T 16362C
30	S035	D4b2b2	-	73G 194T 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9440T 9540C 9824A 10398G 10400T 10873C 11719A 12358G 12705T 14668T 14766T 14783C 15043A 15301A 15326G 16172C 16234T 16362C 16519C
31	S036	G1a1a	-	73G 150T 263G 315.1C 489C 709A 750G 1438G 2706G 4769G 4793G 4833G 5108C 7028T 7867T 8200C 8701G 8860G 9540C 9950C 10398G 10400T 10873C 11719A 11914A 12705T 14569A 14766T 14783C 15043A 15301A 15323A 15326G 15497A 15860G 16223T 16325C 16362C 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
32	S037	M8a2c	-	73G 152C 263G 315.1C 489C 750G 1438G 2706G 2835T 3290C 4715G 4769G 5297T 6179A 7028T 7196A 8584A 8684T 8701G 8860G 8950R 9540C 10398G 10400T 10873C 11719A 12705T 14470C 14766T 14783C 15043A 15244G 15301A 15326G 15487T 16183C 16189C 16223T 16298C 16319A
33	S038	G1a1a	-	73G 93G 150T 263G 315.1C 489C 709A 750G 1438G 2706G 4769G 4793G 4833G 5108C 7028T 7867T 8200C 8701G 8860G 9540C 10398G 10400T 10873C 11719A 11914A 12366G 12705T 14569A 14766T 14783C 15043A 15301A 15323A 15326G 15497A 15860G 16223T 16325C 16343G 16362C 16519C
34	S039	M7c1	-	73G 146C 199C 263G 315.1C 489C 523DEL 524DEL 750G 1438G 2706G 3483A 4071T 4769G 4850T 5442C 6455T 7028T 7279C 7299G 7859A 8280G 8701G 8860G 9540C 9824C 10398G 10400T 10873C 11665T 11719A 12091C 12501A 12564T 12705T 14766T 14783C 15043A 15301A 15326G 16223T 16295T 16319A 16519C
35	S041	D4b2b1	-	73G 152C 194T 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 5632T 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10398G 10400T 10873C 11719A 12705T 14605G 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16362C 16519C
36	S042	D4	-	73G 152C 263G 315.1C 489C 750G 1438G 2706G 3010A 3603T 3764T 4769G 4883T 5178A 7028T 7762A 8343R 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 12351C 12705T 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16274A 16362C
37	S044	M9a1a1a	-	73G 263G 315.1C 489C 750G 980C 1041G 1438G 2706G 3394C 4491A 4769G 7028T 8701G 8860G 9242G 9540C 10398G 10400T 10873C 11719A 11854C 11963A 12705T 13731G 13928A 14308C 14766T 14783C 15043A 15301A 15326G 16094C 16223T 16234T 16280R 16316G 16362C 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
38	S045	Z4	15944DEL 5746.1G	73G 151Y 152C 249DEL 263G 315.1C 489C 750G 1438G 2706G 4715G 4769G 4991A 5746.1G 6326T 6752G 7028T 7196A 8292A 8584A 8701G 8860G 9090C 9540C 10398G 10400T 10873C 11719A 12705T 13818C 14766T 14783C 15043A 15301A 15326G 15475G 15487T 15784C 15944DEL 16185T 16223T 16260T 16298C 16362C 16519C
39	S047	F1b1a1a1	-	73G 152C 249DEL 263G 315.1C 523DEL 524DEL 750G 1438G 1927A 2706G 3970T 4705C 4732G 4769G 5049T 5147A 6392C 6962A 7028T 8860G 10310A 10609C 10976T 11719A 12406A 12633T 12882T 13928C 14476A 14766T 15326G 15954G 16129A 16182C 16183C 16189C 16232A 16249C 16304C 16311C 16344T 16519C
40	S048	G2b2	-	73G 263G 315.1C 489C 709A 750G 1282A 1438G 2706G 2833R 3593C 4769G 4833G 5108C 5601T 6932G 7028T 7705C 8279Y 8701G 8860G 8877C 8896A 9540C 10398G 10400T 10873C 11002G 11719A 12705T 13563G 14569A 14766T 14783C 15043A 15301A 15326G 16223T 16362C 16519C
41	S049	M9a1a1b	-	73G 153G 263G 315.1C 489C 750G 1041G 1438G 2706G 3394C 4491A 4769G 5951G 7028T 8701G 8860G 9115G 9242G 9540C 10398G 10400T 10873C 11719A 12705T 13434G 14308C 14766T 14783C 15043A 15301A 15326G 16223T 16234T 16300G 16316G 16362C
42	S050	B4a4	8281-8289 DEL	73G 152C 188R 193G 263G 315.1C 523DEL 524DEL 709A 750G 1438G 2706G 4769G 5465C 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 9123A 11719A 12088T 14751T 14766T 15326G 16182C 16183C 16189C 16217C 16261T 16299G 16519C
43	S051	D4j3	-	73G 263G 315.1C 489C 750G 1438G 2706G 3010A 4769G 4883T 5178A 6488C 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10463C 10873C 11696A 11719A 12705T 14668T 14766T 14783C 15043A 15301A 15326G 16184T 16223T 16311C 16362C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
44	S052	M10a1a1b1	3172.1C	73G 146C 152C 263G 315.1C 489C 523DEL 524DEL 709A 750G 1438G 2706G 3172.1C 3866C 4140T 4769G 7028T 7250G 8701G 8793C 8856A 8860G 9540C 10398G 10400T 10529G 10646A 10873C 11002G 11719A 12549T 12630A 12633T 12634R 12705T 13135A 13152G 14502C 14766T 14783C 15040T 15043A 15071C 15218G 15301A 15326G 16193T 16223T 16311C 16357C 16497G
45	S053	D4j	-	73G 263G 315.1C 489C 603G 750G 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8414T 8701G 8860G 9540C 10110G 10398G 10400T 10873C 11696A 11719A 12705T 14668T 14766T 14783C 15043A 15301A 15326G 15670C 16223T 16293G 16362C
46	S054	F1a1	-	73G 249DEL 263G 315.1C 523DEL 524DEL 548T 750G 1438G 2416C 2706G 3970T 4086T 4769G 5821A 5984G 6392C 6962A 7028T 8601G 8860G 9053A 9548A 10310A 10609C 11653G 11719A 12406A 12882T 13269G 13759A 13928C 14766T 14870G 15326G 16129A 16162G 16172C 16304C 16519C
47	S055	G1a1a1	-	73G 150T 263G 315.1C 489C 709A 750G 827G 1438G 1822C 2706G 3736A 4769G 4793G 4833G 5108C 7028T 7867T 8200C 8701G 8860G 9540C 10398G 10400T 10873C 11719A 11914A 12705T 13386C 13434R 14569A 14766T 14783C 15043A 15301A 15323A 15326G 15497A 15860G 16075C 16223T 16325C 16362C 16519C
48	S056	M7a1a9	5899.1C	73G 151T 263G 315.1C 489C 523DEL 524DEL 750G 1438G 2626C 2706G 2772T 4048A 4386C 4769G 4958G 5899.1C 6455T 7028T 8701G 8860G 9540C 9824C 10398G 10400T 10873C 11017C 11084G 11719A 12705T 12771A 14025C 14218C 14364A 14766T 14783C 15043A 15301A 15326G 15394C 16209C 16223T 16324C 16519C
49	S057	D4b2b	-	73G 194T 263G 315.1C 489C 523DEL 524DEL 709A 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 5320G 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10398G 10400T 10873C 11719A 12705T 14668T 14766T 14783C 15043A 15301A 15326G 16111T 16186T 16223T 16362C 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
50	S058	M7b1a1a1b	-	73G 150T 199C 263G 315.1C 489C 750G 1438G 2706G 3567T 4048A 4071T 4113A 4164G 4769G 5351G 5460A 6455T 6680C 7028T 7684C 7853A 8701G 8860G 9540C 9824C 10345C 10353T 10398G 10400T 10873C 11719A 12405T 12705T 12811C 14766T 14783C 15043A 15301A 15326G 16129A 16189C 16223T 16297C 16298C 16519C
51	S059	F1a1b	-	64T 73G 249DEL 263G 315.1C 523DEL 524DEL 750G 1438G 2706G 3970T 4086T 4769G 6392C 6962A 7028T 8860G 9053A 9548A 10310A 10609C 11719A 12406A 12882T 13759A 13928C 14002G 14766T 15326G 16129A 16162G 16172C 16304C 16390A 16519C
52	S060	D5b1b2	-	73G 150T 263G 315.1C 456T 489C 681C 750G 1048T 1107C 1438G 2706G 3308C 3759G 4769G 4883T 5147A 5153G 5178A 5301G 6253C 7028T 7403G 8701G 8860G 9180G 9540C 10397G 10398G 10400T 10873C 11719A 12705T 14766T 14783C 15043A 15301A 15326G 15724G 16183C 16189C 16223T 16357C 16362C 16519C
53	S061	B4a2b1a	8281-8289 DEL	73G 263G 315.1C 523DEL 524DEL 750G 1438G 2706G 4769G 5465C 6620C 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 9069G 9123A 9812T 10876G 11719A 13413G 14766T 15326G 16182C 16183C 16189C 16217C 16261T 16360T 16519C
54	S062	M7b1a1a1	-	73G 150T 199C 263G 315.1C 489C 750G 1438G 2706G 4048A 4071T 4164G 4769G 5351G 5460A 6455T 6680C 7028T 7684C 7853A 8701G 8860G 9540C 9824C 10345C 10398G 10400T 10873C 11719A 12405T 12705T 12811C 14766T 14783C 15002A 15043A 15301A 15326G 16129A 16183C 16189C 16223T 16297C 16298C
55	S063	F3a+ 207	-	73G 114T 204C 207A 249DEL 263G 315.1C 750G 1438G 2706G 3434G 3970T 4769G 5585A 5913A 5978G 6392C 7028T 8860G 10310A 10320A 11065G 11719A 12621T 13928C 14766T 15326G 15586C 15854G 16093C 16176Y 16187T 16298C 16355T 16362C 16390A 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
56	S064	D4h1c1	-	73G 152C 263G 315.1C 489C 750G 1243C 1438G 2706G 3010A 3336C 3644C 4769G 4883T 5048C 5178A 7028T 7181T 7673G 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11068G 11719A 12705T 13914A 14470C 14668T 14766T 14783C 15043A 15301A 15326G 16174T 16223T 16311C 16317G 16362C
57	S065	F1c	-	73G 195C 249DEL 263G 315.1C 523DEL 524DEL 709A 750G 1438G 2308G 2706G 3970T 4232C 4769G 6392C 6962A 7028T 8860G 9053A 10256C 10310A 10454C 10609C 11719A 12406A 12882T 13759A 13928C 14364A 14766T 15326G 16093C 16129A 16172C 16304C 16519C
58	S066	Y1	-	73G 146C 263G 315.1C 709A 750G 1438G 2706G 3834A 4562G 4769G 5417A 7028T 8392A 8860G 10379G 10398G 11719A 12705T 14178C 14693G 14766T 15326G 16126C 16231C 16266T 16319A 16399G 16519C
59	S067	D4q	-	73G 200G 234R 263G 315.1C489C 593C 750G 1438G 2706G 3010A 4769G 4883T 5178A 6266G 7028T 8410T 8414T 8701G 8860G 9540C 10398G 10400T 10490C 10873C 11719A 11902A 12705T 14017T 14668T 14766T 14783C 15043A 15301A 15326G 15581R 16223T 16256T 16311C 16519C
60	S068	B4a	8281-8289 DEL	73G 263G 315.1C 523DEL 524DEL 750G 1438G 2706G 4769G 5465C 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 9123A 11719A 14766T 15326G 15926T 16148T 16154C 16182C 16183C 16189C 16217C 16261T 16519C
61	S070	D5b1b2	-	73G 150T 263G 315.1C 456T 489C 681C 750G 1048T 1107C 1438G 2706G 3308C 3759G 4769G 4883T 5147A 5153G 5178A 5301G 6253C 7028T 7403G 8701G 8860G 9180G 9540C 10397G 10398G 10400T 10873C 11719A 12705T 14766T 14783C 15043A 15301A 15326G 15724G 16183C 16189C 16223T 16357C 16362C 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
62	S071	F2b1	-	73G 195C 249DEL 263G 315.1C 523DEL 524DEL 750G 1005C 1438G 1824C 2706G 3970T 4769G 4811G 6392C 6962A 7028T 7828G 8860G 10310A 10535C 10586A 11719A 12338C 13708A 13928C 14766T 15218G 15326G 16092A 16291T 16304C
63	S072	A5a	2156.1A	73G 235G 263G 315.1C 523DEL 524DEL 663G 750G 1438G 1736G 2156.1A 2706G 4248C 4655A 4769G 4824G 7028T 8563G 8598C 8794T 8860G 11536T 11647T 11719A 12705T 14766T 15326G 16187T 16189C 16223T 16290T 16319A
64	S073	Z4a1a1	15944DEL	73G 152C 249DEL 263G 315.1C 489C 750G 1438G 2706G 3520G 4715G 4769G 5492C 5894G 6752G 7028T 7196A 7270C 8584A 8654C 8701G 8860G 9090C 9540C 10398G 10400T 10873C 12705T 14766T 14783C 15043A 15301A 15326G 15475G 15487T 15784C 15944DEL 16185T 16189C 16193DEL 16223T 16260T 16294T 16298C 16302G
65	S075	C7a1	-	73G 146C 249DEL 263G 315.1C 489C 750G 1438G 2706G 3552A 4715G 4769G 5821A 6338G 7028T 7196A 7382A 7853A 8584A 8701G 8860G 9540C 9545G 10398G 10400T 10873C 11719A 11914A 12705T 12957C 13263G 14318C 14766T 14783C 14978G 15043A 15301A 15326G 15487T 16223T 16298C 16309G 16327T 16519C
66	S076	F1b1a	-	73G 146C 249DEL 263G 315.1C 523DEL 524DEL 750G 1438G 2706G 3970T 4732G 4769G 5147A 6392C 6962A 7028T 8860G 9090C 10310A 10609C 10976T 11719A 12406A 12633T 12882T 13928C 14476A 14766T 15326G 16129A 16182C 16183C 16189C 16232A 16249C 16304C 16311C 16344T 16519C
67	S077	Z4	15944DEL	73G 151T 152C 249DEL 263G 315.1C 489C 750G 1438G 2706G 4715G 4769G 6752G 7028T 7196A 8584A 8701G 8860G 9090C 9540C 10398G 10400T 10873C 11719A 12705T 14766T 14783C 15043A 15301A 15326G 15475G 15487T 15784C 15944DEL 16223T 16260T 16298C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
68	S078	D5a1	-	68A 73G 150T 263G 309DEL 315.1C 489C 750G 752T 1107C 1438G 2706G 3496T 3864C 4769G 4883T 5178A 5301G 7028T 8701G 8860G 9180G 9540C 10397G 10398G 10400T 10745T 10873C 11719A 11944C 12026G 12705T 13708A 14766T 14783C 15043A 15301A 15326G 16182C 16183C 16189C 16223T 16362C 16390A 16519C
69	S079	B4c1a1b	8281-8289 DEL	73G 204C 263G 315.1C 709A 750G 1119C 1284C 1438G 2706G 3497T 4769G 5979R 6416G 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 10310A 11719A 12028C 14133G 14203G 14766T 15326G 15346A 16093C 16183C 16189C 16217C 16311C 16519C
70	S080	G2b1a	-	73G 263G 315.1C 489C 592T 709A 750G 1438G 2706G 4769G 4833G 4853A 5108C 5601T 7028T 8701G 8860G 8877C 9038C 9540C 10398G 10400T 10873C 11151T 11719A 12705T 13215Y 13563G 14569A 14766T 14783C 15043A 15301A 15326G 16129A 16223T 16319A 16356C 16362C
71	S081	B4c1a1b	8281-8289 DEL	73G 204C 263G 315.1C 523DEL 524DEL 709A 750G 1119C 1438G 2706G 3497T 4769G 5020Y 5021Y 6416G 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 10310A 11719A 14133G 14203G 14766T 15326G 15346A 16093C 16183C 16189C 16217C 16274A 16311C 16519C
72	S082	D4b2b1	-	73G 152C 194T 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10398G 10400T 10873C 11407T 11719A 12705T 14605G 14668T 14766T 14783C 15043A 15119A 15301A 15326G 15884A 16223T 16362C 16519C
73	S083	Z4	15944DEL	73G 151T 152C 249DEL 263G 315.1C 489C 750G 1438G 2706G 4715G 4769G 6752G 7028T 7196A 8584A 8701G 8860G 9090C 9540C 9545G 10398G 10400T 10873C 11719A 12705T 14766T 14783C 15043A 15301A 15326G 15475G 15487T 15784C 15944DEL 16185T 16223T 16260T 16298C 16390A

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
74	S084	D4b2b	-	73G 194T 263G 315.1C 489C 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 5567Y 7028T 8020A 8059T 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10398G 10400T 10873C 11719A 12705T 13470G 14133G 14668T 14766T 14783C 15043A 15301A 15326G 15747C 16193T 16223T 16362C 16519C
75	S085	D4a1	-	73G 152C 263G 315.1C 489C 750G 1438G 2706G 3010A 3206T 4769G 4883T 5178A 7028T 7775A 8414T 8473C 8701G 8860G 9540C 10398G 10400T 10410C 10873C 11719A 12705T 14668T 14766T 14783C 14979C 15043A 15301A 15326G 16129A 16223T 16362C 16519C
76	S087	B4a1c1a1	8281-8289 DEL	73C 75A 89C 263G 315.1C 523DEL 524DEL 709A 750G 1438G 2706G 4769G 5020Y 5021Y 5465C 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 9123A 9254G 10238C 11719A 14766T 15292T 15326G 16182C 16183C 16189C 16217C 16261T 16519C
77	S088	D4b2a2	-	73G 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8020A 8251A 8414T 8701G 8860G 8964T 9540C 9824A 10104T 10398G 10400T 10873C 11719A 12705T 14668T 14766T 14783C 15043A 15301A 15326G 15524G 16129A 16223T 16362C
78	S089	D4g1	-	73G 263G 315.1C 489C 750G 1438G 2706G 3010A 4343G 4769G 4883T 5178A 7028T 8414T 8860G 9540C 10398G 10400T 10873C 11719A 12705T 13104G 13512G 14668T 14766T 14783C 15043A 15301A 15326G 15518T 16223T 16278T 16362C 16519C
79	S090	D4n	-	73G 263G 315.1C 489C 750G 1438G 2706G 3010A 3421A 4769G 4883T 5178A 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 12705T 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16355A 16362C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
80	S091	D4h1a2	-	73G 146C 183G 263G 315.1C 489C 750G 827G 1009T 1438G 2706G 3010A 3336C 3644C 4769G 4883T 5048C 5178A 6962A 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10700R 10873C 11177T 11719A 12396C 12705T 13135A 14668T 14766T 14783C 15043A 15301A 15326G 16174T 16223T 16362C
81	S092	D4	-	73G 227R 263G 315.1C 489C 750G 1438G 2706G 3010A 4769G 4883T 5178A 7028T 7080C 7754A 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11437C 11719A 12705T 13153G 14668T 14766T 14783C 14857A 15043A 15301A 15326G 15884A 16223T 16311C 16362C
82	S093	G3a2+ 152	-	73G 143A 152C 204C 263G 315.1C 489C 709A 750G 1438G 2706G 3426G 4769G 4833G 5108C 7028T 7621C 8701G 8854A 8860G 9540C 10398G 10400T 10455G 10873C 11719A 12705T 14569A 14766T 14783C 15043A 15301A 15326G 15746G 16223T 16274A 16292T 16362C
83	S094	B5b2a2	8281-8289 DEL	73G 204C 207A 263G 315.1C 523DEL 524DEL 709A 750G 1438G 1598A 2706G 4769G 4895G 7028T 7269A 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8584A 8829T 8860G 9950C 10398G 11719A 12361G 14766T 15223T 15326G 15508T 15662G 15758G 15850C 15851G 15927A 16111T 16140C 16183C 16189C 16234T 16243C 16463G 16519C
84	S095	A5a	2156.1A	73G 235G 263G 315.1C 523DEL 524DEL 663G 750G 1438G 1736G 2156.1A 2706G 4248C 4655A 4769G 4824G 7028T 8537G 8563G 8598C 8794T 8860G 11536T 11647T 11719A 12705T 14766T 15326G 16058T 16187T 16223T 16290T 16319A
85	S097	B5b2a2	8281-8289 DEL	73G 131C 204C 207A 228R 263G 315.1C 523DEL 524DEL 709A 750G 1438G 1598A 2706G 4769G 4895G 7028T 7247T 7678C 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8584A 8829T 8860G 9950C 10398G 11719A 12361G 14766T 15223T 15326G 15508T 15662G 15758G 15850C 15851G 15927A 16111T 16140C 16183C 16189C 16234T 16243C 16463G 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
86	S098	F1b1f	-	73G 249DEL 263G 315.1C 523DEL 524DEL 750G 1438G 2706G 3970T 4732G 4769G 5147A 6392C 6962A 7028T 8860G 10045C 10310A 10609C 10976T 11719A 12406A 12633T 12882T 13928C 14476A 14766T 15326G 16183C 16189C 16232A 16249C 16304C 16311C 16519C
87	S100	D4b2b1	-	73G 194T 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10398G 10400T 10873C 11719A 12705T 14605G 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16362C 16519C
88	S101	F1d1	-	73G 146C 204Y 249DEL 263G 315.1C 516T 523DEL 524DEL 750G 1438G 1734T 2706G 3970T 4769G 5628C 6392C 6962A 7028T 7738C 7861C 8860G 10310A 10609C 11719A 12406A 12882T 13135A 13251T 13928C 14766T 15326G 15402T 16189C 16304C 16519C
89	S102	N9a2a	-	73G 150T 263G 315.1C 750G 1438G 2706G 4769G 5231A 5417A 7028T 8860G 11719A 12358G 12372A 12705T 14766T 15067C 15326G 16172C 16223T 16257A 16261T 16497G
90	S103	M7a1a7	5899.1C	73G 263G 315.1C 489C 750G 1438G 2626C 2706G 2772T 4386C 4769G 4958G 5899.1C 6455T 6926T 7028T 8701G 8860G 9540C 9824C 10398G 10400T 10873C 11017C 11080Y 11084G 11719A 12705T 12771A 14364A 14766T 14767Y 14783C 15043A 15301A 15326G 15884A 16209C 16223T 16291T 16324C
91	S104	F1b1d	-	73G 153G 249DEL 263G 315.1C 523DEL 524DEL 750G 1438G 2706G 3970T 4732G 4769G 5147A 6392C 6962A 7028T 8860G 10310A 10609C 10976T 11719A 12406A 12633T 12882T 13928C 14476A 14766T 15326G 15459T 16182C 16183C 16189C 16232A 16249C 16304C 16311C 16519C
92	S105	A5a1a	2156.1A	73G 188R 235G 263G 315.1C 523DEL 524DEL 663G 750G 1438G 1736G 2156.1A 2706G 4248C 4655A 4769G 4824G 7028T 8563G 8794T 8860G 10801A 11075C 11536T 11647T 11719A 12705T 14766T 14944T 15326G 16187T 16223T 16290T 16319A

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
93	S106	Z4	15944DEL 5746.1G	73G 151T 152C 249DEL 263G 315.1C 489C 750G 1438G 2706G 4715G 4769G 4991A 5746.1G 6752G 7028T 7196A 8292A 8584A 8701G 8860G 9090C 9540C 10398G 10400T 10873C 11719A 12171G 12705T 14766T 14783C 15043A 15301A 15326G 15475G 15487T 15784C 15944DEL 16111T 16185T 16223T 16232T 16260T 16298C 16362C 16519C
94	S107	D5a1	-	68A 73G 150T 263G 309DEL 315.1C 489C 750G 752T 1107C 1438G 2706G 3496T 4769G 4883T 5178A 5301G 7028T 8701G 8860G 9180G 9540C 10397G 10398G 10400T 10873C 11719A 11944C 12026G 12040G 12705T 13708A 14766T 14783C 15043A 15301A 15326G 16182C 16183C 16189C 16223T 16362C 16390A 16519C
95	S108	D4g1c	-	73G 235G 263G 315.1C 489C 709A 750G 1438G 2706G 3010A 4343G 4769G 4883T 5178A 7028T 8347G 8414T 8860G 9540C 10398G 10400T 10873C 11719A 12076R 12705T 13104G 14668T 14766T 14783C 15043A 15301A 15326G 15518T 16223T 16278T 16362C
96	S109	D5a2a+ 160 92	-	73G 150T 263G 315.1C 489C 523DEL 524DEL 750G 752T 1107C 2706G 4769G 4883T 5178A 5301G 6596T 7028T 8701G 8860G 9180G 9540C 10397G 10398G 10400T 10873C 11719A 11944C 12026G 12705T 14766T 14783C 15043A 15301A 15326G 16092C 16129R 16172C 16182C 16183C 16189C 16223T 16266Y 16362C
97	S110	G1a1	-	73G 150T 263G 315.1C 489C 709A 750G 1438G 2706G 4742C 4769G 4833G 5108C 7028T 7867T 8200C 8701G 8860G 9540C 10398G 10400T 10873C 11233Y 11719A 12523G 12705T 14569A 14766T 14783C 15043A 15301A 15323A 15326G 15497A 15860G 16223T 16325C 16362C 16519C
98	S111	M9a1a1c1a	-	73G 153G 263G 315.1C 489C 750G 1041G 1438G 2706G 3394C 4491A 4769G 7028T 7142C 7861C 8701G 8860G 9242G 9540C 10398G 10400T 10873C 11719A 12705T 14308C 14417G 14766T 14783C 15043A 15301A 15326G 16223T 16234T 16291T 16316G 16362C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
99	S112	B4c1a1	8281-8289 DEL	73G 263G 315.1C 709A 750G 1119C 1438G 2706G 3497T 4769G 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 10310A 11719A 14133G 14766T 15326G 15346A 16183C 16189C 16217C 16311C 16519C
100	S113	D4a1b	-	73G 152C 263G 315.1C 489C 750G 1438G 2706G 3010A 3206T 4769G 4883T 5178A 7028T 8137T 8414T 8473C 8701G 8860G 9540C 10398G 10400T 10410C 10873C 10961T 11719A 12705T 13651G 14668T 14766T 14783C 14979C 15043A 15301A 15326G 15924G 16129A 16172C 16223T 16362C 16519C
101	S114	D4h1c1	-	73G 152C 263G 315.1C 489C 750G 1438G 2706G 3010A 3336C 3644C 4769G 4883T 5048C 5178A 7028T 7181T 7673G 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 12705T 13914A 14470C 14668T 14766T 14783C 15043A 15301A 15326G 16174T 16223T 16311C 16317G 16362C
102	S115	Z3d	-	73G 146C 152C 247A 249DEL 263G 315.1C 489C 709A 750G 1438G 2706G 2858G 4715G 4769G 6752G 7028T 7196A 8584A 8701G 8860G 9090C 9540C 10208C 10398G 10400T 10873C 11719A 11782T 12684A 12705T 14766T 14783C 15043A 15301A 15326G 15487T 15784C 16185T 16189C 16193DEL 16223T 16260T 16261Y 16298C
103	S116	B4a1c3b	8281-8289 DEL	73G 89Y 263G 309DEL 315.1C 523DEL 524DEL 709A 750G 1438G 2706G 4218C 4769G 4907C 5465C 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 9123A 10238C 10679G 11719A 14766T 15326G 16182C 16183C 16189C 16194C 16195C 16217C 16261T 16519C
104	S117	B5b2a2	8281-8289 DEL	73G 204C 207A 263G 315.1C 523DEL 524DEL 709A 750G 1438G 1598A 2706G 4769G 4895G 7028T 7269A 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8584A 8829T 8860G 9950C 10398G 11719A 12361G 14766T 15223T 15326G 15508T 15662G 15758G 15850C 15851G 15927A 16111T 16140C 16183C 16189C 16234T 16243C 16463G 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
105	S118	Y1	-	73G 263G 315.1C 709A 750G 1438G 2706G 3834A 4562G 4769G 5417A 7028T 8392A 8860G 10379G 10398G 11719A 12705T 14178C 14693G 14766T 15326G 16126C 16231C 16266T 16319A 16399G 16519C
106	S119	B4m	8281-8289 DEL	73G 263G 315.1C 523DEL 524DEL 750G 1438G 2706G 4769G 6863G 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 10398G 11719A 14410A 14766T 15326G 15381T 16150T 16182C 16183C 16189C 16217C 16240G 16261T 16287T
107	S120	B5b2a1	8281-8289 DEL	73G 131C 199C 204C 263G 315.1C 523DEL 524DEL 709A 750G 1438G 1598A 2706G 4769G 4895G 5744A 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8584A 8829T 8860G 9205C 9950C 10398G 11437C 11719A 12361G 14766T 15223T 15326G 15508T 15662G 15850C 15851G 15927A 16102C 16111T 16140C 16182C 16183C 16189C 16234T 16243C 16291T 16463G 16519C
108	Sample 001	Y1	-	73G 146C 152Y 263G 315.1C 709A 750G 1438G 2706G 3834A 4562G 4769G 5417A 7028T 7325G 8392A 8860G 10379G 10398G 11719A 12705T 14178C 14693G 14766T 15326G 16051G 16126C 16231C 16266T 16319A 16399G 16519C
109	Sample 002	G3a2+ 152	-	73G 143A 152C 204C 263G 315.1C 489C 709A 750G 1438G 2706G 3426G 4769G 4833G 5108C 7028T 7621C 8701G 8854A 8860G 9540C 10398G 10400T 10873C 11719A 12705T 14569A 14766T 14783C 15043A 15301A 15326G 15746G 16223T 16274A 16292T 16362C
110	Sample 003	B4h1	8281-8289 DEL	73G 263G 315.1C 523DEL 524DEL 750G 1438G 2363G 2706G 4703C 4769G 5093C 5783A 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 11719A 12361G 13158G 13269G 14766T 15326G 16129A 16182C 16183C 16189C 16214T 16261T

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
111	Sample 004	N9a1	-	73G 150T 263G 315.1C 750G 1438G 2706G 4386C 4769G 5231A 5417A 7028T 8860G 11719A 12007A 12358G 12372A 12705T 13966G 14766T 15326G 16111T 16129A 16223T 16257A 16261T 16320Y
112	Sample 005	D4g2b1	-	73G 195C 263G 298T 315.1C 489C 504C 523DEL 524DEL 750G 1438G 2706G 3010A 4131G 4769G 4883T 5178A 5231A 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 12705T 13104G 13359A 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16362C 16519C
113	Sample 006	G2a5+ 152	-	73G 152C 263G 315.1C 489C 573.1C 709A 750G 1438G 2706G 4769G 4833G 5108C 5601T 7028T 7600A 7844G 8701G 8860G 9377G 9540C 9575A 10398G 10400T 10873C 11719A 12705T 13383T 13563G 14569A 14766T 14783C 14861A 15043A 15301A 15326G 15562G 16223T 16227G 16234T 16278T 16309G 16362C
114	Sample 007	B4d3a1	8281-8289 DEL	73G 151T 197G 263G 315.1C 546G 750G 827G 1438G 2706G 3083C 4769G 5084G 5418C 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 9120G 10256C 11719A 11914A 12135T 13942G 14766T 15326G 15535T 15930A 16150T 16183C 16185T 16189C 16193DEL 16217C 16234T 16519C
115	Sample 008	D4b2b1	-	73G 194T 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10398G 10400T 10873C 11719A 12705T 14605G 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16362C 16519C
116	Sample 009	A5a1a	2156.1A	73G 235G 263G 315.1C 523DEL 524DEL 663G 750G 1438G 1736G 2156.1A 2706G 4248C 4655A 4769G 4824G 7028T 8517S 8563G 8794T 8860G 10801A 11536T 11647T 11719A 12705T 14766T 14867T 14944T 15326G 16187T 16223T 16290T 16319A

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
117	Sample 010	D4h1c1	-	73G 152C 263G 315.1C 489C 750G 1438G 2706G 3010A 3336C 3644C 4769G 4883T 5048C 5178A 7028T 7181T 7673G 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 12705T 13914A 14470C 14668T 14766T 14783C 15043A 15301A 15326G 16174T 16223T 16311C 16317G 16362C
118	Sample 011	B5b3a	5899.1C 8281-8289 DEL	73G 103A 152C 189G 199C 203A 204C 263G 315.1C 523DEL 524DEL 709A 750G 1438G 1598A 2626C 2706G 4769G 5263T 5899.1C 6962A 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8584A 8829T 8856A 8860G 9950C 10103G 10398G 11719A 11881T 12361G 13477A 14766T 15223T 15326G 15508T 15662G 15851G 15927A 16136C 16140C 16170G 16183C 16189C 16243C 16256T 16519C
119	Sample 012	C4a1b	2232.1A	73G 249DEL 263G 315.1C 489C 750G 1438G 2232.1A 2706G 3552A 4715G 4769G 5846T 6023A 6026A 7028T 7196A 7999C 8584A 8701G 8860G 9389G 9540C 9545G 10398G 10400T 10873C 11719A 11914A 11969A 12672G 12705T 13263G 14034C 14180C 14318C 14766T 14770T 14783C 15043A 15204C 15301A 15326G 15487T 16129A 16192T 16223T 16327T 16519C
120	Sample 013	M10a1b	3172.1C	73G 263G 315.1C 489C 709A 750G 1438G 2706G 3172.1C 4035G 4140T 4769G 7028T 7250G 8701G 8793C 8856A 8860G 9540C 10398G 10400T 10646A 10873C 11719A 12549T 12705T 13152G 14178C 14502C 14766T 14783C 15040T 15043A 15071C 15212G 15218G 15301A 15326G 16066G 16223T 16311C
121	Sample 014	D4a1a1	-	73G 152C 263G 315.1C 489C 750G 1438G 2706G 3010A 3206T 4769G 4883T 5178A 5261A 7028T 8414T 8473C 8701G 8860G 8988G 9371T 9540C 10398G 10400T 10410C 10873C 11719A 12705T 14668T 14766T 14783C 14979C 15043A 15301A 15314A 15326G 16129A 16223T 16362C 16519C

Continued)

No.	Sample ID	Haplogroup*	CodR indel*	Whole mtGenome haplotype
122	Sample 015	B4b1a1	8281-8289 DEL	73G 146C 199C 202G 204Y 207A 263G 315.1C 499A 750G 827G 1438G 2706G 2831A 4117C 4769G 4820A 6023A 6413C 7028T 8206A 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 11719A 13590A 14766T 15034R 15236G 15326G 15497A 15535T 16136C 16183C 16189C 16217C 16284G 16298C 16519C
123	Sample 016	N9a7	-	73G 150T 263G 315.1C 750G 1438G 2706G 4769G 5231A 5417A 7028T 8860G 11719A 12358G 12372A 12705T 14016A 14766T 15326G 16223T 16257A 16261T 16291T
124	Sample 017	B4a1b1	8281-8289 DEL	73G 263G 315.1C 523DEL 524DEL 750G 930R 1438G 2706G 4769G 5004C 5465C 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 9123A 9932A 10238C 10915C 11719A 13143C 14766T 15326G 16103R 16182C 16183C 16189C 16217C 16261T 16268T 16288C 16311C 16519C
125	Sample 018	A15c	-	73G 150T 152C 235G 315.1C 523DEL 524DEL 663G 750G 1438G 1736G 2706G 4248C 4769G 4824G 7028T 8459G 8794T 8860G 9052G 11719A 12705T 13111C 14067T 14766T 15262C 15326G 16223T 16290T 16311C 16319A 16362C
126	Sample 019	Z3	-	73G 150T 249DEL 263G 315.1C 489C 750G 1438G 2706G 4715G 4769G 5291C 5460A 5985A 6752G 7028T 7196A 8584A 8701G 8860G 9090C 9540C 10208C 10398G 10400T 10644R 10873C 11719A 12705T 13105G 14750G 14766T 14783C 15043A 15301A 15326G 15487T 15784C 16185T 16223T 16260T 16294T 16298C
127	Sample 020	D5a2a1b	-	73G 150T 263G 315.1C 489C 523DEL 524DEL 750G 752T 1107C 2706G 3528T 4769G 4883T 5178A 5301G 7028T 8387A 8701G 8860G 9180G 9540C 10397G 10398G 10400T 10873C 11719A 11944C 12026G 12705T 14766T 14783C 15043A 15301A 15326G 16092C 16164G 16172C 16182C 16183C 16189C 16223T 16266T 16362C 16519C

*Haplogroup was determined using EMPOP mtDNA database v4 (<http://empop.online>). Point heteroplasmy was indicated with IUPAC code. Length heteroplasmy regions were excluded. *CodR = Coding region indel

5. 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 염기서열 다형성 분석 결과

본 연구에서는 376명의 한국인 집단 시료를 분석한 MPS자료를 기반으로 미토콘드리아 DNA의 조절 영역과 전체 영역의 하플로타입(Haplotype) 다형성을 비교하였다. 분석 대상 시료 중 미토콘드리아 DNA 조절 영역과 전체 영역에서 동일한 하플로타입을 공유하는 시료의 수를 표 3에 각각 정리하였다.

미토콘드리아 DNA 조절 영역에서는 6명이 동일한 하플로타입을 공유하는 그룹이 1개 있었으며, 4명이 동일한 하플로타입을 공유하는 그룹이 5개, 3명이 동일한 하플로타입을 공유하는 그룹이 10개, 2명이 동일한 하플로타입을 공유하는 그룹이 30개, 그리고 중복되지 않는 유일한 하플로타입의 수는 260개로 전체 시료의 수 중 69.1%를 차지하였다. 미토콘드리아 DNA 전체 영역에서는 3명이 동일한 하플로타입을 공유하는 그룹이 2개 있었으며, 2명이 동일한 하플로타입을 공유하는 그룹이 8개, 그리고 유일한 하플로타입의 수는 354개로 전체 시료의 수 중 94.1%를 차지하였다.

미토콘드리아 DNA 조절 영역과 전체 영역의 하플로타입 다형성(Haplotype diversity)는 Nei와 Tajima등의 계산법²⁴에 따라 각각 0.9985, 0.9998로 산출되었다. 미토콘드리아 DNA 변별력

(Discrimination capacity)는 전체 시료의 수 중 관찰된 하플로타입의 수로 계산되었으며, 조절 영역에서 0.8138 및 전체 영역에서 0.9681로 산출되었다.

표 3. Comparison of haplotype diversity between control region and whole mtGenome in the Korean population

Number of Haplotypes Matching	Control Region	Whole mtGenome
1	260 (69.1%)	354 (94.1%)
2	30	8
3	10	2
4	5	0
6	1	0
Number of distinct haplotypes	306	364
Haplotype diversity (HD*)	0.9985	0.9998
Discrimination capacity (DC*)	0.8138	0.9681

*HD= $n(1-\sum Pi^2)/(n-1)$

*DC= Total number of observed haplotype/ Total number of samples

6. 한국인 집단 시료의 미토콘드리아 DNA 전체 영역 점 이형질 변이 분석 결과

본 연구에서는 미토콘드리아 DNA의 점 이형질 변이를 전체 커버리지가 400X 이상이며 소수 변이의 커버리지가 전체 커버리지의 5%이상인 것에 한하여 호출하였다. 총 130개의 점 이형질 변이가 관찰되었으며, 영역별로 살펴보았을 때는 조절 영역에서 36개 및 암호 영역에서 94개로 확인되었다. 전체 376명의 시료 중 110명의 시료에서 하나 이상의 점 이형질 변이가 관찰되었으며 한 명의 개인(K050)에게서 최대 3개의 점 이형질 변이가 확인되었다 (표 4). 미토콘드리아 DNA 전체 영역 중 가장 많은 점 이형질 변이를 보인 곳은 152번 위치로, 총 5번 관찰되었다. 이어서 146번 위치에서 4번, 204번 위치에서 3번 발견되어 상위 3개의 점 이형질 변이가 모두 조절 영역에 위치하는 것으로 확인되었다 (그림 6²⁵). 점 이형질 변이가 2번 관찰된 위치는 188, 234, 5020, 5021, 15119, 15530 그리고 16093번 이었고, 나머지 위치에서는 한 번씩 발견되었다 (표 5).

표 4. Observed point heteroplasmy across the whole mtGenome of the Korean population

No.	Sample ID	Position	Reference*	Variant	Variant frequency*(%)
1	K006	14118	A	G	8.6
2	K036	903	T	C	6.7
3	K039	16177	A	G	7.2
4	K040	4124	T	C	28.8
5	K041	152	T	C	10.9
6	K043	13879	T	C	21.7
7	K046	7854	T	C	8.3
8	K047	152	T	C	9.1
9	K049	8476	C	A	24.4
10	K050	195	T	C	12.8
11	K050	13881	C	T	15.6
12	K050	16302	A	G	94.3
13	K052	204	T	C	12.1
14	K058	15603	A	G	23.1
15	K068	16093	T	C	85.5
16	K073	1106	C	T	18.2
17	K074	16362	T	C	92.4
18	K079	15530	T	C	67.8
19	K084	152	T	C	65.4
20	K085	4068	C	A	24.4
21	K086	14750	A	G	33.3
22	K089	9997	T	C	8.0
23	K096	9881	T	C	31.7
24	K098	10680	G	A	22.5
25	K099	3460	G	A	5.5
26	K102	14766	C	T	99.9
27	K112	16342	T	C	24.5
28	K122	15663	T	C	19.0
29	K125	215	A	G	13.4
30	K136	146	T	C	23.6
31	K137	146	T	C	22.6
32	K140	12279	A	G	25.2
33	K141	7270	T	C	6.8
34	K146	15530	T	C	76.8
35	K148	6900	A	G	26.6
36	K148	9182	G	A	25.7

Continued)

No.	Sample ID	Position	Reference*	Variant	Variant frequency*(%)
37	S004	16150	C	T	5.0
38	S005	6386	C	T	77.1
39	S010	4917	A	G	15.8
40	S010	14182	T	C	12.4
41	S012	15758	A	G	7.2
42	S017	7375	A	G	34.0
43	S025	14207	G	A	50.5
44	S026	16172	T	C	21.0
45	S027	16256	C	T	6.1
46	S032	1555	A	G	48.4
47	S037	8950	G	A	23.1
48	S042	8343	A	G	49.6
49	S044	16280	A	G	94.7
50	S045	151	C	T	75.6
51	S048	2833	A	G	7.4
52	S048	8279	T	C	11.3
53	S050	188	A	G	11.1
54	S052	12634	A	G	16.4
55	S055	13434	A	G	7.6
56	S063	16176	C	T	12.3
57	S067	234	A	G	7.6
58	S067	15581	A	G	5.5
59	S079	5979	G	A	29.2
60	S080	13215	T	C	60.1
61	S081	5020	T	C	8.7
62	S081	5021	T	C	6.2
63	S084	5567	T	C	33.3
64	S087	5020	T	C	5.4
65	S087	5021	T	C	5.3
66	S091	10700	A	G	11.6
67	S092	227	A	G	7.2
68	S097	228	G	A	84.6
69	S101	204	T	C	5.6
70	S103	11080	T	C	16.4
71	S103	14767	T	C	34.8
72	S105	188	A	G	93.8
73	S109	16129	G	A	9.7

Continued)

No.	Sample ID	Position	Reference*	Variant	Variant frequency*(%)
74	S109	16266	C	T	27.6
75	S110	11233	T	C	78.8
76	S115	16261	C	T	35.6
77	S116	89	T	C	8.4
78	K006	14118	A	G	8.6
79	K036	903	T	C	6.7
80	K039	16177	A	G	7.2
81	K040	4124	T	C	28.8
82	K041	152	T	C	10.9
83	K043	13879	T	C	21.7
84	K046	7854	T	C	8.3
85	K047	152	T	C	9.1
86	K049	8476	C	A	24.4
87	K050	195	T	C	12.8
88	K050	13881	C	T	15.6
89	K050	16302	A	G	94.3
90	K052	204	T	C	12.1
91	K058	15603	A	G	23.1
92	K068	16093	T	C	85.5
93	K073	1106	C	T	18.2
94	K074	16362	T	C	92.4
95	K079	15530	T	C	67.8
96	K084	152	T	C	65.4
97	K085	4068	C	A	24.4
98	K086	14750	A	G	33.3
99	K089	9997	T	C	8.0
100	K096	9881	T	C	31.7
101	K098	10680	G	A	22.5
102	K099	3460	G	A	5.5
103	K102	14766	C	T	99.9
104	K112	16342	T	C	24.5
105	K122	15663	T	C	19.0
106	K125	215	A	G	13.4
107	K136	146	T	C	23.6
108	K137	146	T	C	22.6
109	K140	12279	A	G	25.2
110	K141	7270	T	C	6.8

Continued)

No.	Sample ID	Position	Reference*	Variant	Variant frequency*(%)
111	K146	15530	T	C	76.8
112	K148	6900	A	G	26.6
113	K148	9182	G	A	25.7
114	K150	4160	T	C	9.3
115	K151	1717	T	C	6.9
116	K167	234	A	G	6.9
117	K168	5859	T	C	24.1
118	K168	14248	C	T	25.1
119	K173	9196	G	A	52.1
120	K175	13753	T	C	64.7
121	K180	152	T	C	9.0
122	K180	15119	G	A	20.9
123	Sample001	152	T	C	22.0
124	Sample004	16320	C	T	24.9
125	Sample009	8517	G	C	6.4
126	Sample015	204	T	C	13.6
127	Sample015	15034	A	G	15.6
128	Sample017	16103	A	G	80.9
129	Sample017	930	G	A	71.7
130	Sample019	10644	G	A	18.6

*Reference= rCRS nucleotide

*Variant frequency= Variant nucleotide coverage/ Total coverage X 100

표 5. The most frequently observed point heteroplasmy across the whole mtGenome in the Korean population

Rank	Nucleotide position	Variant	Number of samples
1	152	Y	5
2	146	Y	4
3	204	Y	3
4	188	R	2
	234	R	2
	5020	Y	2
	5021	Y	2
	15119	R	2
	15530	Y	2
	16093	Y	2

IV. 고찰

법유전학 분야에서 미토콘드리아 DNA는 전체 16,569bp중 단 1.1kb에 해당하는 조절 영역만을 대상으로 Sanger 방법을 이용해 분석되어 왔다.^{6,26} 하지만 Sanger 방법은 시간과 노동을 많이 소모할 뿐 아니라, CE 노이즈로 인하여 미토콘드리아 DNA점 이형질 변이를 정확히 파악하기에 어려움이 있다. 최근 MPS 기법의 발전으로 미토콘드리아 DNA의 분석 범위를 전체 영역으로 확대하고, 동시에 높은 품질로 정밀하게 분석하는 것이 보다 용이해졌다.^{7,9,27} 본 연구에서는 무작위로 채취된 한국인 집단 시료를 대상으로 미토콘드리아 DNA 전체 영역을 MPS 방법을 이용해 분석하고 집단유전학적 통계량을 제시하고자 하였다.

MPS 분석에 앞서 Sanger 방법을 이용해 동일 시료의 미토콘드리아 DNA의 조절 영역을 분석해 참고자료로 사용하였다. 발견된 조절 영역의 염기서열 변이를 기반으로 미토콘드리아 DNA의 하플로그룹을 조사하였을 때, 염기서열 변이가 적은 시료의 경우에는 하플로그룹을 자세히 지정하는데 어려움이 있었다. 또한 전기영동도의 피크 높이가 낮을 경우, 미토콘드리아 DNA 점 이형질 변이를 CE의 노이즈와 구분하는 것이 매우 어려웠다. 그러나 과변이 영역 1 및 과변이 영역2에서 T염기가 소실된

경우에는 전기영동도를 통해 시각적으로 확인할 수 있어 변이 확인에 유용하였다.

미토콘드리아 DNA의 전체 영역은 긴 범위 PCR방법을 통해 단 두개의 절편으로 빠르고 간단하게 증폭할 수 있었다. 긴 범위 PCR 방법은 참조 염기서열을 기준으로 각각 8.4kbp 및 8.6kbp의 긴 증폭 산물을 생성하여 뉴마이트 (NUMT; nuclear mitochondrial pseudogene sequences)의 영향을 배제할 수 있었다.²⁸ 또한 MPS 라이브러리를 생성할 때, 35회 이상의 중합 반응으로 라이브러리를 생성한 이전의 연구들과 달리^{7,8,20,29} 본 연구에서는 26-28회만 이용하여 충분한 양의 라이브러리를 생성할 수 있었다. 또한, MPS 자료를 분석할 때, 원형의 미토콘드리아 DNA가 선형의 리드로 분석되는 점을 고려하여 조절 영역과 그 외의 암호 영역을 각각 나누어 정렬한 후 통합하여 분석하였다. 그 결과, 중단점을 포함하는 조절 영역에서의 MPS 자료의 손실을 방지할 수 있었다. 전체 영역에 걸쳐서 충분한 커버리지를 얻었으나, King⁷ 등이 이전 논문에서 언급한 바와 같이 연속된 C염기를 포함하는 조절 영역 내 과변이 영역 2(300~350bp)부근의 커버리지는 다소 낮았다.

미토콘드리아 DNA 전체 영역은 MPS를 이용해 분석 범위를 약 15배 확장함으로 훨씬 더 많은 수의 염기서열 변이가 관찰되었다. 앞서 Sanger 방법으로 얻어진 조절 영역 염기서열 변이를 MPS

결과에서 모두 동일하게 확인함으로써 확보된 MPS자료의 완결성을 확인할 수 있었다. 전체 영역에 걸친 하플로타입을 바탕으로 한국인 집단 시료의 미토콘드리아 DNA 염기서열 다형성을 분석한 결과, 전체 분석 대상 시료 376개 중 다른 시료와 중복되지 않는 유일한 하플로타입이 조절 영역에서는 260개로 약 69.1%에 그친 반면, 전체 영역에서는 354개로 약 94.1%로 약 25.0% 증가하였다. 본 연구에서 조사된 한국인 집단의 미토콘드리아 DNA의 변별력은 조절 영역(0.8138)에 비해 전체 영역(0.9681)에서 0.1543 증가하였으며, 이는 Just 등⁹에 의해 보고된 세계의 미국인 집단(African American, U.S. Caucasian 및 U.S. Hispanic) 자료와 함께 비교해보았을 때 가장 큰 증가율을 보이는 것으로 확인되었다 (표 6). 미토콘드리아 DNA 조절 영역에서의 변별력은 네 집단 중 한국인 집단에서 가장 낮았지만, 전체 영역에서의 증가율은 가장 높았다. 즉, 한국인 집단의 미토콘드리아 DNA 조절 영역에서는 많은 수의 시료가 중복되는 하플로타입을 갖지만, 이를 전체 영역으로 분석하면 많은 수의 하플로타입이 서로 분리될 수 있다는 것을 의미한다. 따라서 한국 법과학 실무에서 미토콘드리아 DNA 분석의 범위를 전체 영역으로 확장한다면 보다 높은 변별력을 확보할 수 있을 것으로 사료된다.

표 6. Comparison of discrimination capacity increment between the mitochondrial DNA control region and whole mtGenome in three U.S population and Korean population

	Control region	Whole mtGenome	Increment(Δ)
African American (n=170)			
<i>(Just et al.)</i>			
Haplotypes	148	169	21
Unique haplotypes	130	168	38
Discrimination capacity	0.8706	0.9941	0.1235
Haplotype diversity	0.9981	0.9999	0.0018
U.S. Caucasian (n=263)			
<i>(Just et al.)</i>			
Haplotypes	229	259	30
Unique haplotypes	211	255	44
Discrimination capacity	0.8707	0.9848	0.1141
Haplotype diversity	0.9978	0.9999	0.0021
U.S. Hispanic (n=155)			
<i>(Just et al.)</i>			
Haplotypes	141	147	6
Unique haplotypes	130	140	10
Discrimination capacity	0.9097	0.9484	0.0387
Haplotype diversity	0.9986	0.9992	0.0006
Korean (n=376)			
<i>(This study)</i>			
Haplotypes	306	364	58
Unique haplotypes	260	354	94
Discrimination capacity	0.8138	0.9681	0.1543
Haplotype diversity	0.9985	0.9998	0.0013

더 자세하게 살펴보면, 미토콘드리아 DNA의 조절 영역에서 가장 많이 관찰된 하플로타입은 표 7-1에 제시되어 있으며 이는 전체 분석 대상 시료 376개 중 6개가 동일한 것으로 관찰되었다. 미토콘드리아 DNA 전체 영역의 하플로타입을 살펴보았을 때, 앞서 언급된 6개 시료 중 동일한 하플로타입을 공유하는 시료는 3개로 줄어들었으며, 나머지 3개는 중복되지 않는 유일한 하플로타입으로 분리되었다 (표 7-2). 또한 이들 시료의 하플로그룹을 지정할 때, 조절 영역에서는 모두 D4b2b로 확인된 반면 전체 영역을 바탕으로 하플로그룹을 지정하였을 때는 각각 D4b2b1 또는 D4b2b3로 더욱 세밀한 지정이 가능하였다. 이와 같이 미토콘드리아 DNA 조절 영역만으로는 구분되지 않았던 시료들의 분리가 가능했다는 점에서 미토콘드리아 DNA 전체 영역 분석을 법과학 실무에 적용한다면 동일인 혹은 동일 모계 판정에 있어 보다 높은 식별력을 제공할 수 있을 것으로 사료된다.

표 7-1. The most frequently observed haplotype in the mitochondrial DNA control region

Control region haplotype	Haplotypes Matching	Haplogroup
73G 194T 263G 315.1C 489C 523DEL 524DEL 16223T 16362C 16519C	6	D4b2b

표 7-2. The whole mtGenome haplotype of the most frequently observed in the control region

Whole mtGenome haplotype*	Haplotypes Matching	Haplogroup
73G 194T 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10398G 10400T 10873C 11719A 12705T 14605G 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16362C 16519C	3	D4b2b1
73G 194T 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 6252G 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10398G 10400T 10873C 11719A 12705T 14605G 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16362C 16519C	1	D4b2b1
73G 194T 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10398G 10400T 10873C 11719A 12630A 12705T 14605G 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16362C 16519C	1	D4b2b1
73G 194T 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10373A 10398G 10400T 10873C 11719A 12705T 14325C 14668T 14766T 14783C 15043A 15301A 15326G 15331A 16223T 16362C 16519C	1	D4b2b3

*Additional variants of each sample were highlighted in bold.

본 연구에서 분석된 한국인 시료 중 미토콘드리아 DNA 전체 영역에서 동일한 하플로타입을 보인 시료는 각각 세 명씩 2그룹 (표 8-1), 두 명씩 8그룹 (표 8-2)이 있었다. 총 10개 그룹의 미토콘드리아 DNA 하플로그룹은 D4에 이어서 N9a 그룹에 가장 많이 속해 있었다. 이와 같은 맥락으로, MPS를 이용해 한국인 시료의 하플로그룹을 분석한 Park등¹⁰의 논문에서도 D4 그룹이 미토콘드리아 DNA의 분석 범위를 전체영역으로 확장함에 따라 가장 세밀하게 분석되는 하플로그룹으로 보고되었다.

표 8-1. The whole mtGenome haplotype observed three times in the Korean population

Whole mtGenome haplotype	Haplotypes Matching	Haplogroup
73G 150T 263G 315.1C 456T 489C 681C 750G 1048T 1107C 1438G 2706G 3308C 3759G 4769G 4883T 5147A 5153G 5178A 5301G 6253C 7028T 7403G 8701G 8860G 9180G 9540C 10397G 10398G 10400T 10873C 11719A 12705T 14766T 14783C 15043A 15301A 15326G 15724G 16183C 16189C 16223T 16357C 16362C 16519C	3	D5b1b2
73G 194T 263G 315.1C 489C 523DEL 524DEL 750G 1382C 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8020A 8414T 8701G 8860G 8964T 9296T 9540C 9824A 10398G 10400T 10873C 11719A 12705T 14605G 14668T 14766T 14783C 15043A 15301A 15326G 16223T 16362C 16519C	3	D4b2b1

표 8-2. The whole mtGenome haplotype observed three times in the Korean population

Whole mtGenome haplotype	Haplotypes Matching	Haplogroup
73G 146C 263G 315.1C 709A 750G 1438G 2706G 3834A 4562G 4769G 5417A 7028T 8392A 8860G 10379G 10398G 11719A 12705T 14178C 14693G 14766T 15326G 16126C 16231C 16266T 16319A 16399G 16519C	2	Y1
73G 150T 263G 315.1C 750G 1438G 2706G 4769G 5231A 5417A 7028T 8860G 11719A 12358G 12372A 12705T 14016A 14766T 15326G 15530Y 16223T 16257A 16261T 16291T	2	N9a7
73G 150T 263G 315.1C 750G 1438G 2706G 4769G 5231A 5417A 7028T 8860G 11719A 12358G 12372A 12705T 14766T 15067C 15326G 16172C 16223T 16257A 16261T 16497G	2	N9a2a
73G 152C 263G 315.1C 489C 750G 1438G 2706G 3010A 3336C 3644C 4769G 4883T 5048C 5178A 7028T 7181T 7673G 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11719A 12705T 13914A 14470C 14668T 14766T 14783C 15043A 15301A 15326G 16174T 16223T 16311C 16317G 16362C	2	D4h1c1
73G 204C 207A 263G 315.1C 523DEL 524DEL 709A 750G 1438G 1598A 2706G 4769G 4895G 7028T 7269A 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8584A 8829T 8860G 9950C 10398G 11719A 12361G 14766T 15223T 15326G 15508T 15662G 15758G 15850C 15851G 15927A 16111T 16140C 16183C 16189C 16234T 16243C 16463G 16519C	2	B5b2a2
73G 263G 315.1C 489C 603G 750G 1438G 2706G 3010A 4769G 4883T 5178A 7028T 8414T 8701G 8860G 9540C 10398G 10400T 10873C 11696A 11719A 12705T 14668T 14766T 14783C 15043A 15301A 15326G 15670C 16223T 16293G 16362C	2	D4j
73G 263G 315.1C 489C 750G 1041G 1438G 2706G 3394C 4491A 4769G 7028T 8701G 8860G 9242G 9540C 10398G 10400T 10873C 11719A 11963A 12705T 13731G 14308C 14766T 14783C 15043A 15301A 15326G 16223T 16234T 16316G 16362C	2	M9a1a1a
73G 263G 315.1C 523DEL 524DEL 709A 750G 1119C 1438G 2706G 3497T 4769G 7028T 8281DEL 8282DEL 8283DEL 8284DEL 8285DEL 8286DEL 8287DEL 8288DEL 8289DEL 8860G 10310A 11719A 14133G 14766T 15326G 15346A 16086C 16183C 16189C 16217C 16311C 16519C	2	B4c1a1a

한편, 미토콘드리아 DNA의 한 위치에서 두 개의 염기가 함께 존재하는 점 이형질 변이를 분석할 때, Sanger방법은 전기영동도의 피크의 높이를 바탕으로 시각적으로 판단하여야 한다. 따라서 소수 변이가 약 20%이상일 때만 확인이 가능하며, 피크가 높지 않을 경우에는 CE의 노이즈와 명확히 구별할 수 없다.¹⁸ 반면에 MPS 방법은 리드 수를 바탕으로 참조 염기와 변이 염기의 비율이 정확하게 측정되기 때문에 소수 변이의 검출 한도를 약 5%까지 낮출 수 있었다.¹¹

미토콘드리아 DNA 점 이형질 변이가 가장 많이 관찰된 상위 3개의 위치 (152, 146, 204번)는 모두 조절 영역 내에서도 과변이 영역 1 (Hypervariable region 1)에 포함되어 있었다. 이 위치들은 많은 연구^{30,31}에서 높은 돌연변이율로 인해 점 이형질 변이가 자주 관찰되는 곳으로 보고된 바 있어, 본 연구의 결과도 동일한 경향성을 보인 것으로 확인되었다. 점 이형질 변이는 각 조직의 초기발생단계에서 유래하는 배엽층(germ layer)에 따라 동일한 개인에서도 조직마다 다르게 나타날 수 있으며,³² 병목현상으로 인해 같은 조직 내에서도 서로 다른 양상을 보일 수 있다.³³ 이와 같은 점을 고려하여 점 이형질 변이를 더욱 유념해서 판단해야 한다.

동일한 개인에게서 유래된 세 가지 시료(혈액, 구강상피세포 및 머리카락의 미토콘드리아 DNA 전체 염기서열을 비교한 선행 연구¹⁶에서는 법과학 실무에서 점 이형질 변이의 정확한 검출의 중요성에 대해 보고되었다. Sanger방법으로 분석하였을 때 서로 다른 염기로 관찰되어 배제되었던 시료가 MPS 방법을 이용해 소수 변이를 5%의 검출 한도로 분석함으로써 두 시료가 배제되지 않음을 보여주었다. 따라서 법과학 실무에서 두 시료를 대조할 때, 점이형질 변이에 의해 잘못 배제되는 오류를 최소화하여 실무자의 올바른 판단에 큰 도움이 될 것으로 생각된다.

미토콘드리아 DNA 길이 이형질 변이는 리드 수 기반의 MPS 자료를 통해 보다 체계적으로 분류할 수 있을 것으로 기대된다. 기존의 Sanger 방법은 길이 이형질 변이가 발생하는 연속된 C 영역 이전까지는 프라이머의 방향에 따라 문제없이 읽을 수 있으나, 연속된 C 영역을 포함해 그 이후로는 급격하게 시퀀싱 품질이 하락하여 미토콘드리아 DNA 길이 이형질 변이를 확인하기 매우 어렵다. 반면에 MPS 자료는 각 리드에서의 C염기의 개수를 파악할 수 있어 그에 따라 미토콘드리아 DNA 길이 이형질 변이를 특정화할 수 있을 것으로 기대된다. Lin 등¹²의 논문에서 MPS를 이용해 미토콘드리아 DNA 이형질 변이에 대한 분석을 시도가 보고되었다. 하지만 자료 분석에 사용된 SEQ mapper³⁴를 이용하기

위해서는 가능한 길이 이형질 변이의 유형에 대해 미리 CE 방법을 이용해 참고자료를 생성해야 한다. 따라서 MPS 자료만으로도 길이 이형질 변이를 분석할 수 있는 방법이 필요하다. 양방향 리드에서 얻어진 자료를 통합하는 현재의 MPS방법으로는 연속된 C 염기의 개수를 명확하게 얻기 어려웠다. STRait razor³⁵는 MPS 자료를 바탕으로 길이 기반의 다형성(polymorphism)과 염기서열을 특정화하여 각 리드를 분석할 수 있기 때문에 이러한 점을 고려해 후속 연구에서 단방향 리드 분석을 시도한다면 미토콘드리아 DNA길이 이형질 변이를 보다 효율적으로 분석할 수 있을 것으로 기대된다.

V. 결 론

1. 본 연구에서는 혈연관계가 없는 한국인 376명의 미토콘드리아 DNA 전체 영역 (16,569 bp)을 긴 범위 PCR을 통해 두 개의 절편으로 나누어 증폭한 후, Nextera XT library preparation kit 및 MiSeq system을 사용하여 MPS 분석하였다.
2. 한국인에서 관찰된 미토콘드리아 DNA 조절 영역의 변별력 (Discrimination capacity)는 0.8138에서 분석 범위를 전체 영역으로 확장하였을 때는 0.9681로 증가하였다. 따라서 미토콘드리아 DNA 전체 영역 분석을 법과학 실무에 적용하면 더 높은 개체식별력을 확보할 수 있을 것이다.
3. 미토콘드리아 DNA 점 이형질 변이는 GATK Mutect2 파이프라인을 이용하여 5%의 검출 한도로 분석할 수 있었으며, 암호 영역에서 94개 및 조절 영역에서 36개로 총 130개 관찰되었다. 한국인에서 조사된 점 이형질 변이에 대한 정보는 동일인 혹은 동일 모계 판단에 유용한 정보를 제공할 수 있을 것이다.

4. 양방향 리드를 통합하는 MPS 자료 분석방법으로는 미토콘드리아 DNA 길이 이형질 변이 분석에 한계가 있었다. 추후 연구에서 단방향 리드 분석을 통한 길이 이형질 변이의 특정화를 시도해볼 필요가 있다.

본 연구는 미토콘드리아 DNA 전체 영역 분석에 대한 효용성을 보여주고 있으며, 미토콘드리아 DNA 이형질 변이에 대해서 보다 명확한 기준과 분석 가능성을 제시하였다. 본 연구를 기반으로 얻어진 한국인 집단의 미토콘드리아 DNA 전체 영역에 대한 정보는 법과학 실무에서 기초 통계 자료로서 활용될 것으로 기대된다.

참 고 문 헌

1. Robin ED, Wong R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol.* 1988;136:507-513.
2. Budowle B, Allard MW, Wilson MR, Chakraborty R. FORENSICS AND MITOCHONDRIAL DNA: Applications, Debates, and Foundations. *Annu Rev Genomics Hum Genet.* 2003;4(1):119-141.
3. Chaitanya L, Van Oven M, Weiler N, et al. Developmental validation of mitochondrial DNA genotyping assays for adept matrilineal inference of biogeographic ancestry at a continental level. *Forensic Sci Int Genet.* 2014;11(1):39-51.
4. Butler JM. *Fundamentals of Forensic DNA Typing.* Elsevier; 2010.
5. Irwin JA, Saunier JL, Strouss KM, et al. Development and expansion of high-quality control region databases to improve forensic mtDNA evidence interpretation. *Forensic Sci Int Genet.* 2007:154-157.
6. Chaitanya L, Van Oven M, Brauer S, et al. High-quality mtDNA control region sequences from 680 individuals sampled across the Netherlands to establish a national forensic mtDNA reference database. *Forensic Sci Int Genet.* 2016;21:158-167.
7. King JL, LaRue BL, Novroski NM, et al. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. *Forensic Sci Int Genet.* 2014;12:128-135.
8. McElhoe JA, Holland MM, Makova KD, et al. Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina

- MiSeq. *Forensic Sci Int Genet.* 2014;13:20-29.
9. Just RS, Scheible MK, Fast SA, et al. Full mtGenome reference data: Development and characterization of 588 forensic-quality haplotypes representing three U.S. populations. *Forensic Sci Int Genet.* 2015;14:141-155.
 10. Park S, Cho S, Seo HJ, Lee JH, Kim M, Lee SD. Entire Mitochondrial DNA Sequencing on Massively Parallel Sequencing for the Korean Population. *J Korean Med Sci.* 2017;587-592.
 11. Just RS, Irwin JA, Parson W. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Sci Int Genet.* 2015;18:131-139.
 12. Lin CY, Tsai LC, Hsieh HM, et al. Investigation of length heteroplasmy in mitochondrial DNA control region by massively parallel sequencing. *Forensic Sci Int Genet.* 2017;30:127-133.
 13. Lee HY, Yoo JE, Park MJ, Chung U, Kim CY, Shin KJ. East Asian mtDNA haplogroup determination in Koreans: Haplogroup-level control region SNP analysis and subhaplogroup-level control region sequence analysis. *Electrophoresis.* 2006;27(22):4408-4418.
 14. Lee HY, Yoo JE, Park MJ, Chung U, Shin KJ. Mitochondrial DNA control region sequences in Koreans: Identification of useful variable sites and phylogenetic analysis for mtDNA data quality control. *Int J Legal Med.* 2006;120(1):5-14.
 15. Scheible M, Kim SH, Sturk-Andreaggi K, Coble MD, Irwin JA. Mitochondrial control region variation in a Korean population sample. *Int J Legal Med.* 2014;128:745-746.
 16. Kim BM, Hong SR, Chun H, Kim S, Shin K. Comparison of whole mitochondrial genome variants between hair shafts and reference samples using massively parallel

sequencing. 2019.

17. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 1999;23:147.
18. Parson W, Gusmão L, Hares DR, et al. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Sci Int Genet.* 2014;13:134-142.
19. Parson W, Dür A. EMPOP-A forensic mtDNA database. *Forensic Sci Int Genet.* 2007;1:88-92.
20. Fendt L, Zimmermann B, Daniaux M, Parson W. Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. *BMC Genomics.* 2009;10(139).
21. Martin M. Cutadapt removes adapter sequences from high-throughput. *EMBnet.journal.* 2011;17(1):10-12.
22. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760. d
23. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinforma.* 2014;11(1110):1-42.
24. Nei M, Tajima F. DNA polymorphism detectable by restriction endonucleases. *Genetics.* 1981;97(1):145-163.
25. Holland MM. Molecular analysis of the human mitochondrial DNA control region for forensic identity testing. *Curr Protoc Hum Genet.* 2012;(SUPPL.74):1-23.
26. Lyons EA, Scheible MK, Sturk-Andreaggi K, Irwin JA, Just

- RS. A high-throughput Sanger strategy for human mitochondrial genome sequencing. *BMC Genomics*. 2013;14(881).
27. Parson W, Huber G, Moreno L, et al. Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples. *Forensic Sci Int Genet*. 2015;15:8-15.
 28. Cui H, Li F, Chen D, et al. Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. *Genet Med*. 2013;15(5):388-394.
 29. Peck MA, Brandhagen MD, Marshall C, Diegoli TM, Irwin JA, Sturk-Andreaggi K. Concordance and reproducibility of a next generation mtGenome sequencing method for high-quality samples using the Illumina MiSeq. *Forensic Sci Int Genet*. 2016;24:103-111.
 30. Santos C, Sierra B, Álvarez L, et al. Frequency and pattern of heteroplasmy in the control region of human mitochondrial DNA. *J Mol Evol*. 2008;67(2):191-200.
 31. Irwin JA, Saunier JL, Niederstätter H, et al. Investigation of heteroplasmy in the human mitochondrial DNA control region: A synthesis of observations from more than 5000 global population samples. *J Mol Evol*. 2009;68(5):516-527.
 32. Gallimore JM, McElhoe JA, Holland MM. Assessing heteroplasmic variant drift in the mtDNA control region of human hairs using an MPS approach. *Forensic Sci Int Genet*. 2018;32:7-17.
 33. Linch CA, Whiting DA, Holland MM. Human hair histogenesis for the mitochondrial DNA forensic scientist. *J Forensic Sci*. 2001;46(4):844-853..
 34. Lee JCI, Tseng B, Chang LK, Linacre A. SEQ Mapper: A

DNA sequence searching tool for massively parallel sequencing data. *Forensic Sci Int Genet.* 2017;26:66-69.

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

ABSTRACT

Analysis of whole mitochondrial DNA sequences using massively parallel sequencing in Koreans.

Bo min Kim

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

(Directed by Professor Kyoung-Jin Shin)

Mitochondrial DNA is a useful analytical tool for identification of the maternal lineage in cases where nuclear DNA cannot be analyzed, such as degraded samples or hair shafts in forensic casework. In the meantime, mitochondrial DNA analysis has been focused only on the control region due to the limitations of the Sanger-type sequencing and little quantity for the sample. However, it is well recognized that expanding the analysis range from the control region to whole mitochondrial DNA could increase the discrimination power. Recently, the development of Massively Parallel Sequencing (MPS) has made it possible to analyze the whole mitochondrial DNA more feasibly. It

also has been more sophisticated to identify the mitochondrial DNA heteroplasmy which is limited by the Sanger-type sequencing. In this study, the whole mitochondrial DNA of 376 Koreans was amplified with two fragments by long-range PCR. The MPS library was constructed using Nextera XT DNA library preparation kit and sequenced on MiSeq system. The obtained MPS data was analyzed using the GATK Mutect2 pipeline. An additional 58 haplotypes were observed by expanding the analysis range of mitochondria DNA from the control region to the whole mtGenome. The discrimination capacity was increased from 0.8138 in the control region to 0.9681 in the whole mtGenome. The point heteroplasmy(PHP) was analyzed with a 5% threshold. Total 130 PHPs were observed with 36 PHPs in control region and 94 PHPs in coding region. The most frequently observed PHP was located in nucleotide position 152. The analysis of length heteroplasmy was limited by current MPS data analysis methods, which analyzes with paired-end reads. Therefore, it was necessary to analysis with one-way read. The usefulness of analysis whole mtGenome was verified through this study. Also, the accumulation of mtGenome PHP data has facilitated the application of MPS in the comparative analysis. It is expected that the statistics of the Korean population including the whole mitochondrial DNA sequence

variation data will be used as the reference data for the analysis of mitochondrial DNA in the Korean population.

Key words: Massively Parallel Sequencing, whole mitochondrial DNA, Korean population, Haplotype, Heteroplasmy