Massively Parallel Sequencing Analysis of Whole Mitochondrial Genomes in Hair Shaft, Blood and Buccal Swab Samples

Bo Min Kim[1,2], Kyoung-Jin Shin[1,2]

[1]Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea
[2]Brain Korea 21 PLUS Project for Medical Science, Yonsei University, Seoul, Korea

YONSEI UNIVERSITY COLLEGE OF MEDICINE

1. Introduction

CHALLENGING

- Hair shafts
- Skeletal remains

REFERENCE

- Buccal swab
- Blood

YONSEI UNIVERSITY COLLEGE OF MEDICINE

1. Introduction

CHALLENGING

- **Hair shafts**
- **Skeletal remains**

- Insufficient amount of available DN
- High level of DNA fragmentation

→ ***Small amplicon strategy***

J Forensic Sci. 2001 Mar;46(2):247-53.

Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy.

Gabriel MN[1], Huffine EF, Ryan JH, Holland MM, Parsons TJ

Author infor

1 MtDNA Sec

Int J Legal Med (2008) 122:385–388
DOI 10.1007/s00414-008-0227-5

ORIGINAL ARTICLE

'Mitominis': multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples

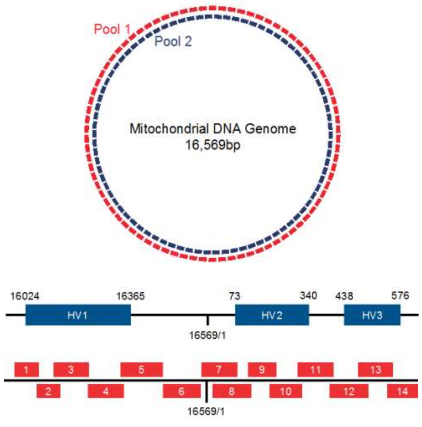
Cordula Eichmann · Walther Parson

YONSEI UNIVERSITY COLLEGE OF MEDICINE

1. Introduction

➢ **MPS strategy for challenging sample type**

Precision ID mtDNA Panels



ThermoFisher SCIENTIFIC

Precision ID mtDNA whole Genome Panel

- 163 bp average amplicon length
- 81 primer pairs
- 2-pool multiplex assay
- Degraded and compromised sample

Precision ID mtDNA Control Region Panel

- 153 bp average amplicon length
- 7 primer pairs

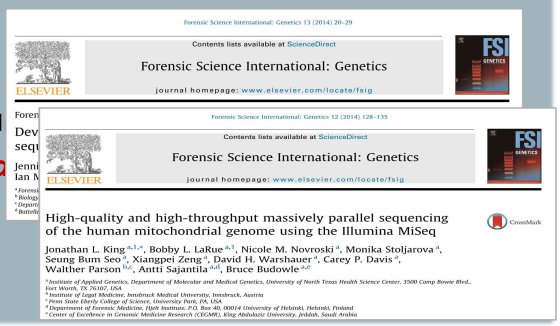
1. Introduction

REFERENCE

- **Buccal swab**
- **Blood**

- Often taken directly from a suspect
- Low level of DNA fragmentation
- Impacts of NUMTs (*nu*clear *mi*tochond

→ ***Long-range amplification stra***



Forensic Science International: Genetics 13 (2014) 20–29

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Forensic Science International: Genetics 12 (2014) 128–135

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq

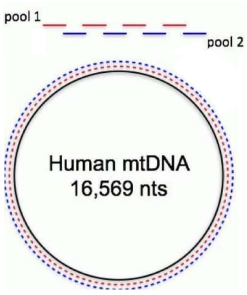
Jonathan L. King, Bobby L. LaRue, Nicole M. Novroski, Monika Stoljarova, Seung Bum Seo, Xiangpei Zeng, David H. Warshauer, Carey P. Davis, Walther Parson, Antti Sajantila, Bruce Budowle

YONSEI UNIVERSITY COLLEGE OF MEDICINE

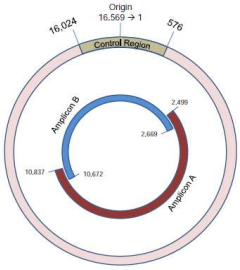
1. Introduction

➢ MPS strategy for forensic sample type

[Hair shafts]



[Reference]



Precision ID mtDNA Whole Genome Panel	Long-range amplification
- 163 bp average amplicon length	- 8,500 bp average amplicon length
- 81 primer pairs	- 2 primer pairs
- 2-pool multiplex assay	- Avoiding impacts of NUMTs
- Degraded and compromised sample	- Intact and high-quality sample

2. Materials & Methods

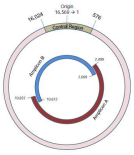
➢ MPS workflow of hair shafts

Sample collection	DNA extraction	MPS Library prep.	MPS system	MPS analysis
▪ 20 Korean males ▪ **≥ 1 year (R.T.)** ▪ **2nd proximal 2cm** ▪ in duplicate	▪ QIAamp DNA Investigator Kit	▪ **Precision ID mtDNA Whole Genome Panel kit**	▪ Ion chef system ▪ Ion S5 system ▪ Ion 520 chip	▪ Converge 2.1 software

YONSEI UNIVERSITY COLLEGE OF MEDICINE

2. Materials & Methods

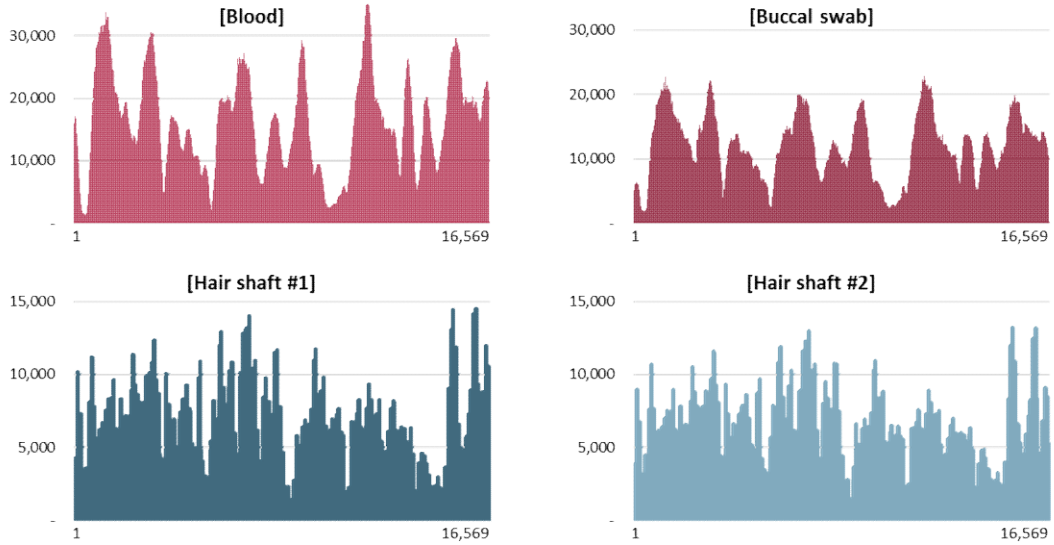
➢ MPS workflow of reference samples

Sample collection	DNA extraction	MPS Library prep.	MPS system	MPS analysis
				
▪ 20 Korean males ▪ Peripheral Blood ▪ Buccal swab	▪ QIAamp DNA Mini Kit	▪ **Long range PCR with 2 fragment** ▪ Nextera XT DNA Library Prep Kit	▪ MiSeq™ system ▪ MiSeq Kit v3	▪ GATK Mutect2

YONSEI UNIVERSITY COLLEGE OF MEDICINE

3. Results

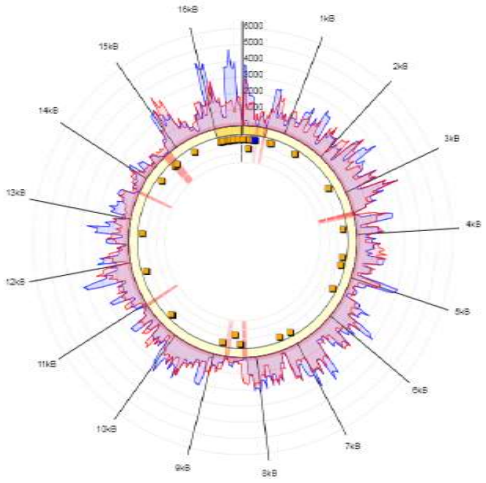
➢ **MPS coverage**



YONSEI UNIVERSITY COLLEGE OF MEDICINE

3. Results

➢ **MPS data analysis of hair shafts using Converge software**



[Circular coverage plot]

- Forward coverage
- Reverse coverage

- Deletion
- Insertion
- Mutation

YONSEI UNIVERSITY COLLEGE OF MEDICINE

3. Results

➢ MPS data analysis of hair shafts using Converge software

[Converge software result]

Sample	5001_HIonXpress_085																								
lose haplogroup	Y1																								
Regions	-16569																								
Position	Ref	Sample	Variant	Var Freq	Type	ead Coverage	ead Coverage	lele Coverage	lele Coverage	Cover	G%	A%	T%	C%	N%	ins%	del%	olymorphism	ontrol Regio	State	Frequency	Artefact	ar Strand Bi	ead Strand Bi	EMPOP
73	A	G	G	99.8	SNP	3616	2865	6466	3607	2859	99.8	0.2	0	0	0	0.1	0	73G	73G	confirmed	99.8	True variant	0.5	0.6	unchecked
146	T	C	C	98.6	SNP	1885	1378	3218	1858	1360	0	0	0.1	98.6	0	0.6	1.3	146C	146C	confirmed	98.6	True variant	0.5	0.6	unchecked
152	T	T	Y	34.6	SNP	1883	1393	1132	649	483	0	0	65.4	34.6	0	1.2	0	152Y	152Y	possible	34.6	Point Heteroplasmy	0.5	0.6	unchecked
263	A	G	G	99.9	SNP	433	438	870	432	438	99.9	0	0	0	0	0.6	0.1	263G	263G	confirmed	99.9	True variant	0.5	0.5	unchecked
309	T	+	C	42.5	INS	365	798	494	171	323	0	0	0	0	0	54.9	0	309.1C	309.1C	likely	42.5	Length Heteroplasmy	0.5	0.7	confirmed
315	G	+	C	75.6	INS	357	802	876	225	651	0	0	0	0	0	85.7	0	315.1C	315.1C	likely	75.6	True variant	0.6	0.7	confirmed
709	G	A	A	98.6	SNP	3087	2871	5874	3040	2834	1.3	98.6	0	0	0	1.4	0.1	709A		confirmed	98.6	True variant	0.5	0.5	unchecked
750	A	G	G	99.4	SNP	791	1842	2618	789	1829	99.4	0	0	0	0	0.7	0.5	750G		confirmed	99.4	True variant	0.5	0.7	unchecked
1438	A	G	G	99.5	SNP	1186	709	1885	1185	700	99.5	0.5	0	0	0	0.4	0	1438G		confirmed	99.5	True variant	0.5	0.6	unchecked
2706	A	G	G	98.6	SNP	1644	1486	3086	1617	1469	98.6	1.4	0	0	0	0.6	0	2706G		confirmed	98.6	True variant	0.5	0.5	unchecked
3834	G	A	A	100	SNP	303	696	999	303	696	0	100	0	0	0	0	0	3834A		confirmed	100	True variant	0.5	0.7	unchecked
4562	A	G	G	97.3	SNP	1771	1386	3071	1710	1361	97.3	2.2	0	0	0	0.3	0.5	4562G		confirmed	97.3	True variant	0.5	0.6	unexpected
4769	A	G	G	99	SNP	1013	1136	2127	1011	1116	99	0	0	0	0	0.5	1	4769G		confirmed	99	True variant	0.5	0.5	unchecked
5417	G	A	A	99.8	SNP	176	441	616	175	441	0.2	99.8	0	0	0	1.3	0	5417A		confirmed	99.8	True variant	0.5	0.7	unchecked
7028	C	T	T	97.9	SNP	2217	1001	3149	2170	979	0	0	97.9	2.1	0	0.1	0	7028T		confirmed	97.9	True variant	0.5	0.7	unchecked
7325	A	G	G	99.6	SNP	1764	1952	3701	1762	1939	99.6	0.1	0	0	0	0.2	0.3	7325G		confirmed	99.6	True variant	0.5	0.5	unexpected
8392	G	A	A	99.1	SNP	167	388	550	164	386	0.9	99.1	0	0	0	0.2	0	8392A		confirmed	99.1	True variant	0.5	0.7	unchecked
8555	T	T	Y	25.1	SNP	791	552	337	208	129	0	0	74.9	25.1	0	0.3	0	8555Y		unclear	25.1	Point Heteroplasmy	0.5	0.6	unexpected
8860	A	G	G	99.9	SNP	429	848	1276	428	848	99.9	0.1	0	0	0	0.2	0	8860G		confirmed	99.9	True variant	0.5	0.7	unchecked
10379	A	G	G	97	SNP	953	1527	2406	909	1497	97	2.9	0	0	0	0.3	0.1	10379G		confirmed	97	True variant	0.5	0.6	unchecked
10398	A	G	G	99	SNP	1073	1746	2790	1061	1729	99	1	0	0	0	0	0	10398G		confirmed	99	True variant	0.5	0.6	unchecked
11719	G	A	A	99.1	SNP	3040	1028	4032	3019	1013	0.9	99.1	0	0	0	0.3	0	11719A		confirmed	99.1	True variant	0.5	0.7	unchecked
12705	C	T	T	98.1	SNP	1985	564	2500	1947	553	0	0	98.1	1.9	0	0.2	0	12705T		confirmed	98.1	True variant	0.5	0.8	unchecked

YONSEI UNIVERSITY
COLLEGE OF MEDICINE

3. Results

➢ MPS data analysis of hair shafts using Converge software

[Converge software result]

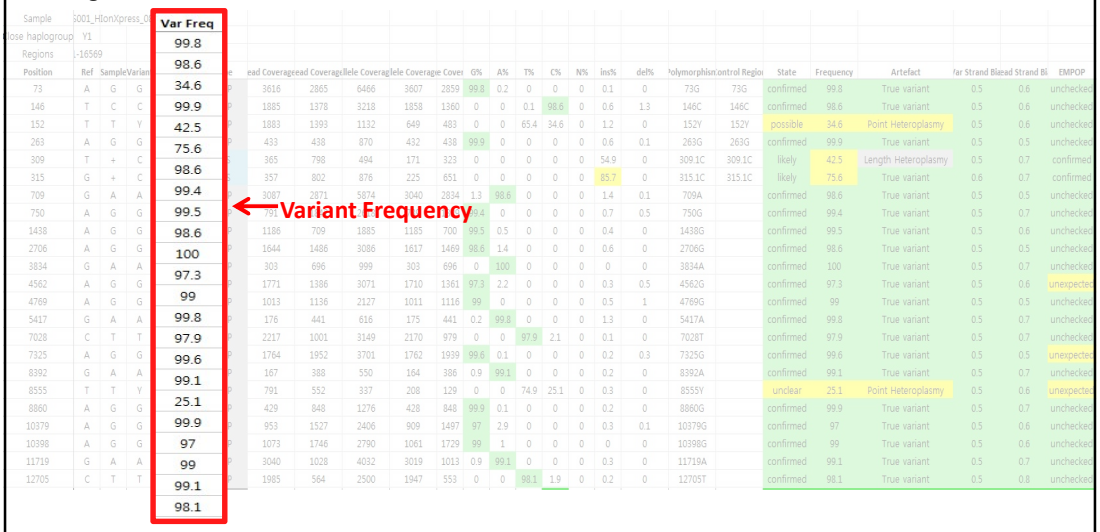
Close haplogroup Y1 ← Close haplogroup based on EMPOP

YONSEI UNIVERSITY
COLLEGE OF MEDICINE

3. Results

➢ **MPS data analysis of hair shafts using Converge software**

[Converge software result]

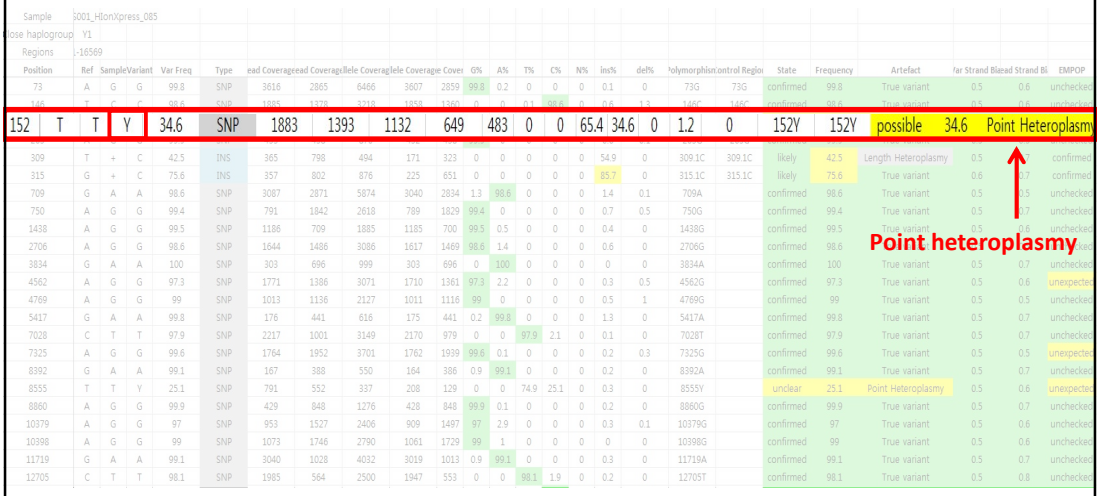


YONSEI UNIVERSITY COLLEGE OF MEDICINE

3. Results

➢ **MPS data analysis of hair shafts using Converge software**

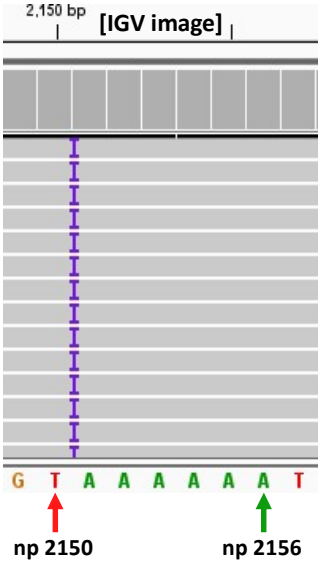
[Converge software result]



YONSEI UNIVERSITY COLLEGE OF MEDICINE

3. Results

➢ **MPS data analysis of hair shafts using Converge software**



[Converge 2.1 result]

Position	Ref	Sample	Variant	Type	…	Polymorphism
2156	T	+	A	INS	…	**2156.1A**

→ Automatically calling according to forensic nomenclature *(ISFG guidelines)*

[GATK Mutect2 result]

#Chrom	Position	Ref	Alt
chrM	2150	T	TA

3. Results

➢ **Observed point heteroplasmy of whole mtGenomes**

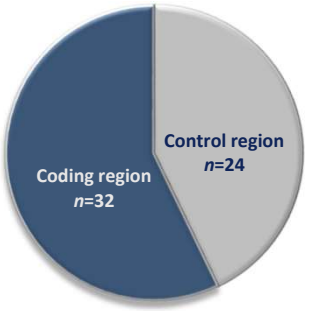
Sample ID	rCRS Position	rCRS Nucleotide	Hair shaft#1	Hair shaft#2	Blood	Buccal swab
Sample 001	152	T	C	T/C(34.6)	T/C(22.0)	C/T(33.4)
	3834	G	A/G(5.4)	A	A	A
	8555	T	T	T/C(25.1)	T	T
	15364	C	C/T(10.4)	C	C	C
	16266	C	T/C(5.6)	T	T	T
	16319	G	A/G(8.3)	A/G(7.8)	A	A
Sample 004	709	G	A/G(41.8)	G	G	G
	16320	C	C/T(10.1)	C	C/T(24.9)	C/T(16.8)
Sample 006	41	C	C/T(20.7)	C	C	C
	11847	G	G/A(13.6)	G	G	G
Sample 007	16150	C	T	T	T	T/C(5.8)
Sample 008	15373	A	A	A/G(32.2)	A	A
	16362	T	C	C/T(7.7)	C	C
Sample 009	8517	G	G	G	G/C(6.4)	G
	16187	C	T	T/C(7.5)	T	T
Sample 012	11969	G	A/G(5.7)	A	A	A
	16192	C	T/C(7.0)	T	T	T
Sample 013	16311	T	C	C/T(45.2)	C	C
Sample 014	8473	T	C/T(19.2)	C/T(14.3)	C	C
	16362	T	C/T(6.4)	C/T(5.8)	C	C
Sample 015	204	T	T	T	T/C(13.6)	T/C(49.0)
	499	G	A/G(7.6)	A	A	A
	2831	G	A/G(5.7)	A	A	A
	4820	G	A/G(12.3)	A/G(8.0)	A	A
	15034	A	G/A(5.9)	A/G(29.9)	A/G(15.6)	A/G(9.9)
	15236	A	G/A(5.7)	G	G	G
Sample 016	14016	G	A/G(11.4)	A/G(10.3)	A	A
Sample 017	930	G	A/G(23.0)	A/G(30.9)	A/G(28.3)	A/G(23.0)
	15279	T	T	T/C(33.1)	T	T
	16103	A	A	A/G(33.0)	G/A(19.1)	G/A(29.7)
Sample 018	15262	T	C/T(27.1)	C/T(18.4)	C	C
Sample 019	9947	G	G/A(45.6)	G	G	G
	10644	G	G/A(37.3)	G	G/A(18.6)	G/A(19.3)
Sample 020	1200	G	G	G/A(40.0)	G	G
	16162	A	A/G(13.6)	A	A	A

3. Results

➢ Observed point heteroplasmy of whole mtGenomes

- ✓ Variant calling threshold **>100X**
- ✓ Point heteroplasmy threshold **> 400X**
- ✓ Minor nucleotides of PHP were observed in **> 5%** of the total coverage

- Total 56 PHPs
- 15/20 donors across 4 tissue sets
- PHPs per donor up to 5,3,2,2 in hair shaft #1, hair shaft#2, blood, buccal swab
- Control region : Coding region = 24 : 32

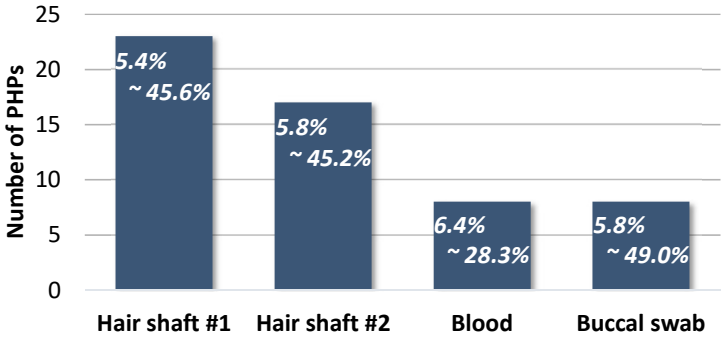


YONSEI UNIVERSITY COLLEGE OF MEDICINE

3. Results

➢ Observed point heteroplasmy of whole mtGenomes

Sample type	Total	Hair shaft #1	Hair shaft #2	Blood	Buccal swab
No. of PHPs	56	23	17	8	8
Proportion range(%)	5.4~49.0	5.4~45.6	5.8~45.2	6.4~28.3	5.8~49.0



YONSEI UNIVERSITY COLLEGE OF MEDICINE

4. Discussions

➢ **Inverted major nucleotide according to the tissue type**

Sample ID	rCRS position	rCRS nucleotide	Hair shaft #1	Hair shaft #2	Blood	Buccal swab
Sample 001	152	T	C	T/C(34.6)	T/C(22.0)	C/T(33.4)
Sample 004	709	G	A/G(41.8)	G	G	G
Sample 015	15034	A	G/A(5.9)	A/G(29.9)	A/G(15.6)	A/G(9.9)
Sample 017	16103	A	A	A/G(33.0)	G/A(19.1)	G/A(29.7)

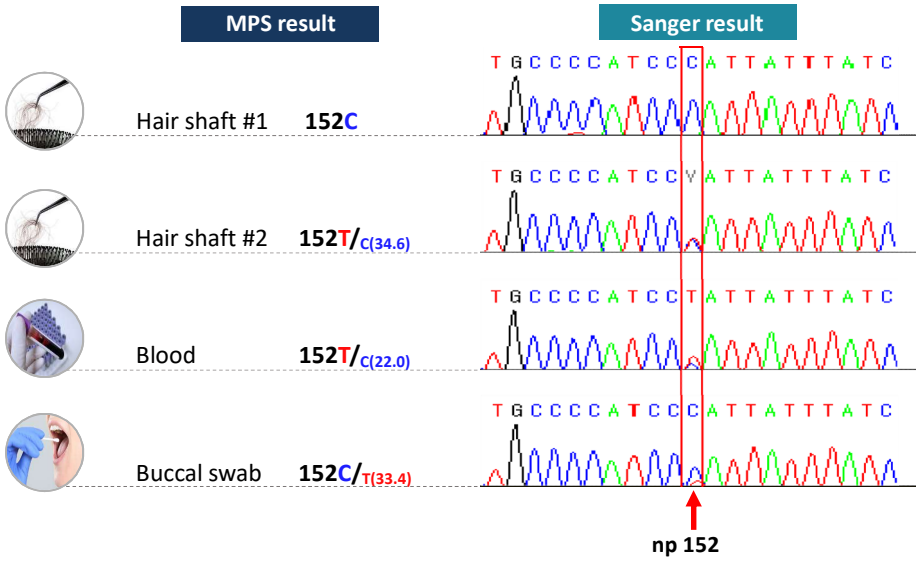
**Minor variants are indicated using smaller font size with their proportion stated in parentheses.*

YONSEI UNIVERSITY COLLEGE OF MEDICINE

4. Discussions

➢ **Inverted major nucleotide according to the tissue type**

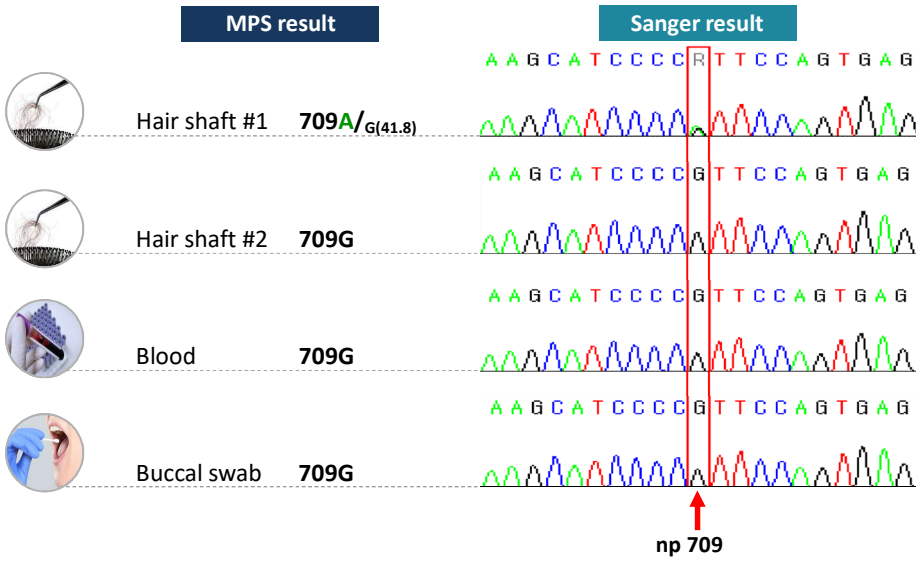
[Sanger sequencing results of *Sample 001*]



4. Discussions

➢ **Inverted major nucleotide according to the tissue type**

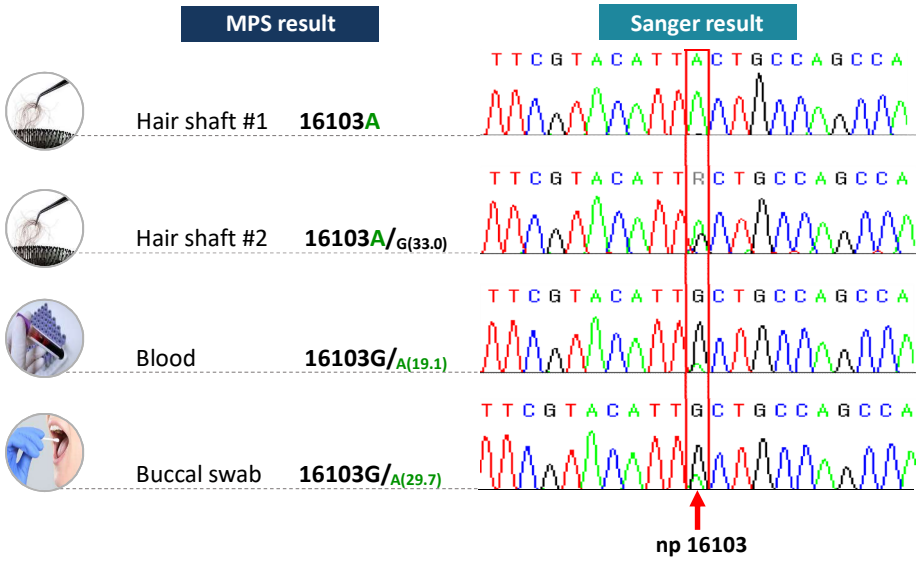
[Sanger sequencing results of *Sample 004*]



4. Discussions

➢ **Inverted major nucleotide according to the tissue type**

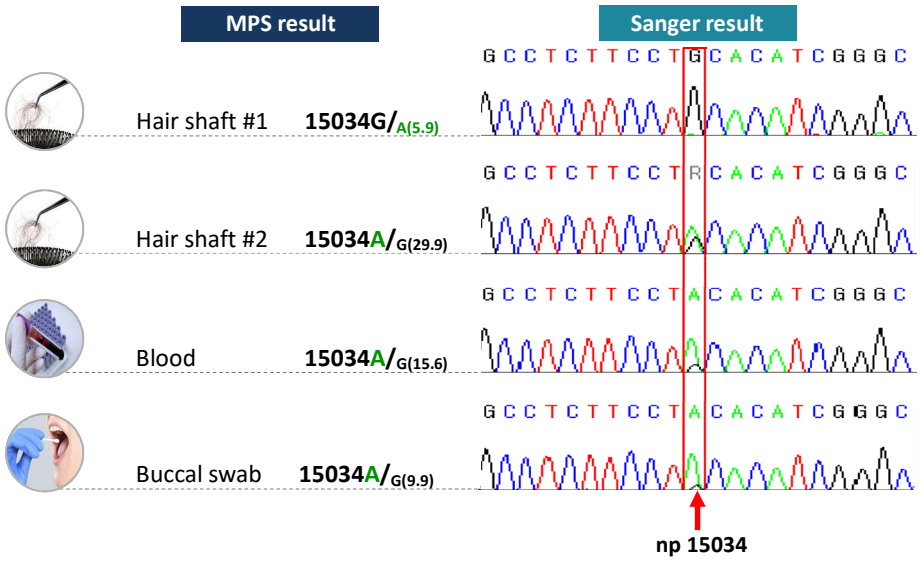
[Sanger sequencing results of *Sample 017*]



4. Discussions

➢ **Inverted major nucleotide according to the tissue type**

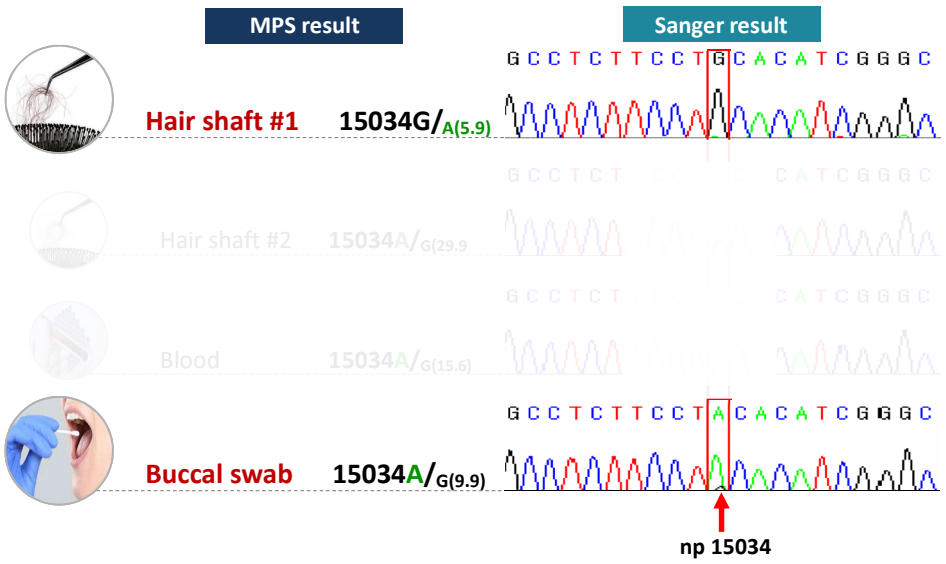
[Sanger sequencing results of *Sample 015*]



4. Discussions

➢ **Inverted major nucleotide according to the tissue type**

[Sanger sequencing results of *Sample 015*]

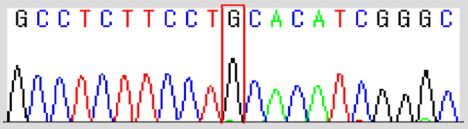


4. Discussions

What if..?

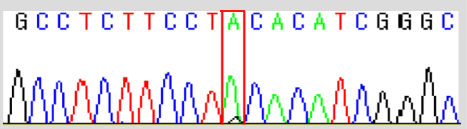
[Sanger sequencing results of *Sample 015***]**

At crime scene



Sanger sequencing result: **15034G**

Reference



Sanger sequencing result: **15034A**

4. Discussions

What if..?

[Sanger sequencing results of *Sample 015***]**

At crime scene

Reference

"Possibility of false exclusion"

Sanger sequencing result: **15034G**

Sanger sequencing result: **15034A**

4. Discussions

What if..?

[Sanger sequencing results of *Sample 015***]**

At crime scene

Sanger sequencing result: **15034G**

PHP threshold: approximately ~20%

Reference

Sanger sequencing result: **15034A**

PHP threshold: approximately ~20%

4. Discussions

What if..?

[Sanger sequencing results of *Sample 015***]**

At crime scene

Sanger sequencing result: **15034G**

MPS result: **15034G/**$_{A(5.9)}$

PHP threshold: 5%

Reference

Sanger sequencing result: **15034A**

MPS result: **15034A/**$_{G(9.9)}$

PHP threshold: 5%

5. Conclusions

- We successfully analyzed the whole mtGenome variants of 4 sets of tissue.
- Although **hair shafts** had been stored at room temperature **> 1 year**, their whole mtGenome variants were successfully obtained using the **Precision ID mtDNA Whole Genome Panel**.
- The whole mtGenome variants of **blood and buccal swabs** with 2 **LR-PCR** products were analyzed using the Nextera XT DNA Library Prep Kit.
- In 4 samples, the major nucleotide of PHP was **inverted at one nucleotide** position in the hair shaft and reference samples.
- The MPS data which provide a **high resolution** for PHP, **did not result in exclusion of that both haplotypes were derived from the same donor.**
- **Accumulation of mtGenome PHP data** will facilitate the application of MPS in the comparative analysis of mtGenome variants containing PHP to forensic casework.

5. Conclusions

International Journal of Legal Medicine
https://doi.org/10.1007/s00414-019-02205-y

ORIGINAL ARTICLE

Check for updates

Comparison of whole mitochondrial genome variants between hair shafts and reference samples using massively parallel sequencing

Bo Min Kim [1,2] · Sae Rom Hong [1,2] · Hein Chun [2,3] · Sangwoo Kim [2,3] · Kyoung-Jin Shin [1,2]

Received: 4 September 2019 / Accepted: 7 November 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Int J Legal Med. in press

donor. Accumulation of mtGenome PHP data will facilitate the application of MPS in the comparative analysis of mtGenome variants containing PHP to forensic casework.

Acknowledgments The authors would like to thank P.H. Kim and C.H. Lee from Thermo Fisher Scientific Korea for their generous support for this research.

9. King JL, LaRue BL, Novroski NM, Stoljarova M, Seo SB, Zeng X, Warshauer DH, Davis CP, Parson W, Sajantila A, Budowle B (2014) High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. Forensic Sci Int Genet 12:128–135
10. Gabriel MN, Huffine EF, Ryan JH, Holland MM, Parsons TJ (2001) Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy. J Forensic Sci 46:247–253

YONSEI UNIVERSITY COLLEGE OF MEDICINE

Acknowledgement

YONSEI UNIVERSITY COLLEGE OF MEDICINE

- Yonsei DNA Profiling Group
- Yonsei TGIL (Translational Genome Informatics Laboratory)

과학기술정보통신부 Ministry of Science and ICT

Ministry of Science and ICT, Republic of Korea
(Funding No.: NRF-2014M3A9E1069989)

BrainKorea21PLUS

Brain Korea 21 PLUS Project for
Medical Science, Yonsei University

Chia Hsuan Su, DongA science

Thank you!

KJSHIN@yuhs.ac

kbomin526@yuhs.ac

http://forensic.yonsei.ac.kr