FORENSIC SCIENCE JOURNAL SINCE 2002

Forensic Science Journal
2017;16(1):51-58
fsjournal.cpu.edu.tw
DOI:10.6593/FSJ.2017.1601.05

Applications of NGS on Mitochondrial DNA Analysis for Forensic Samples

Chia-Hung Huang [1], M.S. ; Tsun-Ying Huang [1], M.S. ; Fang-Chen Chung [1], M.S. ; Chu-Chun Hsu [1], M.S. ; Guan-Cheng Peng [1], B.S. ; Yi-Ting Chang [1], M.S. ; Chun-Yen Lin [1*], Ph.D.

[1] *Institute of Forensic Medicine, Ministry of Justice, Taiwan, ROC.*

Received: March 29, 2017; Accepted: November 29, 2017.

Abstract

Sequencing for forensic DNA identification is usually performed by using the Sanger method with capillary electrophoresis presently. However, it still has some problems by using this system such as determinations of the number of C for C-stretch and the sequence heteroplasmy in mitochondrial DNA D-loop region used in forensic identification. The new technology NGS is with the potential to resolve these problems. In this study, sequencing of the mitochondrial DNA for 39 forensic samples were successfully performed by NGS. Three samples of them were observed with the sequence heteroplasmy but not observed from sequencing by the Sanger method with capillary electrophoresis. Totally there were 14 types of C-stretch for HV1 and HV2 combinations. The results showed the feasibility for applications of NGS on mitochondrial DNA analysis for forensic samples. This is the first report for sequencing the forensic samples by using NGS in Taiwan.

Keywords: Forensic Science, Capillary Electrophoresis, Next-Generation Sequencing (NGS), Mitochondrial DNA (mtDNA)

Introduction

In forensic practice, mitochondrial DNA (mtDNA) sequencing by the Sanger method with capillary electrophoresis is usually used in determination of maternal relationships mostly based on the HV-1 and HV-2 regions of D-loop. However, sequencing by the Sanger method with capillary electrophoresis cannot precisely distinguish the noise and real signals, and cannot determine the relative abundance of each sequence for mtDNA heteroplasmy. Furthermore, the sequence interpretation from the electropherograms will also be interfered with the length heteroplasmy (usually observed in the C-stretch) due to the indels. However, the recently developed next-generation sequencing (NGS) was reported with the potential to resolve these problems [1-3]. NGS has been studied on forensic applications for Y-chromosome sequencing to differentiate the individuals of the same male-line relatives, whole transcriptome deep sequencing to identify the candidate genes expressed in skin, and human genome sequencing to differentiate identical twins for paternity testing et al. [4-6]. In this study, NGS was evaluated to be used for mtDNA sequencing of forensic samples.

Materials and methods

In this study, two primer pairs were used to amplify the mitochondrial genome from samples for NGS [7,8]. The amplification products were then digested and

*Corresponding author: Chun-Yen Lin, Institute of Forensic Medicine, Ministry of Justice, Taiwan, ROC.
E-mail: apollo591@gmail.com

labeled to enable simultaneous analysis of multiple samples with the mitochondrial genome sequencing platform. Sequences for NGS were compared with those obtained by the Sanger method with capillary electrophoresis. Furthermore, the preliminary identification of ethnicity for these sequences was also performed [9].

DNA extraction and quantification

The control DNA 9947A (Thermo Fisher Scientific, MA, USA) was used as the reference standard for NGS accuracy analysis. The 44 forensic samples, including 9 muscles tissue samples, 9 blood samples, 24 bone samples, and 2 tooth samples, were collected. Totally there were 45 samples analyzed in this study. DNA was extracted from each of the forensic samples by using the PureLink™ Genomic DNA Kit (Thermo Fisher Scientific) and quantified with the Quantifiler® Human DNA Quantification Kit (Thermo Fisher Scientific).

Sanger sequencing

The HV-1 and HV-2 regions were amplified from 0.5 ng DNA and by using the primer pairs L15997(5'-CACCATTAGCACCCAAAGCT-3')/H16401(5'-TGATTTCACGGAGGATGGTG-3')and L29(5'-GGTCTATCACCCTATTAACCAC-3')/H408(5'-CTGTTAAAAGTGCATACCGCCA-3') respectively. The PCR products were purified with the DNA Clean/Extraction Kit (GeneMark, Taipei, Taiwan), and then subjected to cycle sequencing by using one of the above mentioned primers and the BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific). The products were analyzed using the ABI PRISM 3130XL Genetic Analyzer (Thermo Fisher Scientific) and SeqScape® v2.6 software (Thermo Fisher Scientific). The Revised Cambridge Reference Sequence (rCRS, NC-012920.1 in GenBank) was used as the reference.

NGS

1. Long-range PCR

Human mitochondrial genome amplification was performed with modifications from the previous reports [7,8]. Two ng DNA was used for mitochondrial DNA amplification with the SequalPrep™ long PCR kit (Thermo Fisher Scientific). The PCR products were purified with the GeneMark® PCR Clean-Up Kit (GeneMark, Taipei, Taiwan) and quantified by using BioDrop µLITE spectrophotometer (OMNILAB, Bremen, Germany).

2. Library preparations

The procedures for fragmenting and labelling DNA were performed by using the commercial Illumina Nextera XT Kit (Illumina, San Diego, CA, USA) following the manufacturer's suggestions.

3. Quantification of the library products

Quantification of the library products were performed by using the KAPA Library Quantification Kit (Takara Bio, Shiga, Japan) following the manufacturer's suggestions and the 7500 Real-Time PCR System (Thermo Fisher Scientific).

4. Sequencing by NGS platform

Sequencing was performed by using the platform of MiSeq (Illumina). The sequencing data were analyzed with mtDNA Variant Analyzer® (Illumina).

Results and Discussion

Sequence analysis

Initially there were 44 forensic samples sequenced by NGS. Their storage time ranged from 1 to 2 years after collection. Totally there were 39 samples of them and the control 9947A successfully sequenced by NGS for mitochondrial whole genome. The 9947A sequences were consistent with the predicted sequences. It showed the accuracy of the sequencing method used in this study. Comparison with the rCRS (NC-012920.1), the mitochondrial HV-1 and HV-2 sequences different from the rCRS for all samples were recorded in Table 1. The different nucleotide numbers ranged from 7 (Sample 39) to 15 (Sample 34). There were three samples observed with the sequence heteroplasmy but not interpreted as a heteroplasmy from sequencing by the Sanger method with capillary electrophoresis (Table 2 and Fig. 1). The heteroplasmic sites were at 189, 189 and 16140 for samples 12, 17 and 33 respectively. The number and percentage of reads from NGS for each nucleotide were also showed in Table 2. These heteroplasmic sites were definitely identified.

Table 1 NGS results for mitochondrial HV-1 and HV-2 regions of 39 forensic samples and the standard DNA 9947A.

Sample	Nucleotide different from rCRS														
1	73G	93G	210G	263G	315.1C	16140C	16187T	16189C	16256T	16266G					
2	73G	93G	263G	315.1C	16129R	16182C	16183C	16189C	16217C	16261T	16356C				
3	73G	195C	263G	309.1C	315.1C	16171G	16189C	16223T	16311C	16362C					
4	73G	189G	199C	263G	309.1C	315.1C	16223T	16227G	16278T	16362C					
5	73G	152C	263G	315.1C	16129A	16223T	16249C	16278T	16284G	16311C	16362C				
6	73G	249DEL	263G	315.1C	16051G	16092A	16291T	16304C	16311C						
9	73G	152C	249DEL	263G	309.1C	315.1C	16111T	16129A	16266T	16304C					
10	73G	195C	263G	309.1C	315.1C	16051G	16223T	16304C	16362C						
11	73G	146A	182T	199C	263G	309.1C	315.1C	16287T							
12	73G	152C	189R	225R	249DEL	263G	309.1C	315.1C	16129A	16209C	16223T	16272G			
13	73G	249DEL	263G	309.1C	315.1C	16129A	16148T	16169T	16172C	16304C					
15	73G	152C	193G	263G	309.1C	315.1C	16182C	16183C	16189C	16213A	16217C	16261T	16295T	16299G	
16	73G	93G	210G	263G	315.1C	16140C	16187T	16189C	16256T	16266G					
17	73G	150T	152C	189R	263G	309.1C	315.1C	16148T	16182C	16183C	16189C	16223T	16362C		
18	73G	150T	152C	263G	315.1C	16304C	16335G	16362C							
19	73G	263G	309.1C	309.2C	315.1C	16147T	16183C	16184A	16189C	16217C	16235G				
20	73G	150T	263G	309.1C	315.1C	16129A	16148Y	16223T	16257A	16261T	16311Y	16325C			
9947A	93G	195C	214G	263G	309.1C	309.2C	315.1C	16311C							
23	73G	152C	235G	263G	315.1C	16148T	16223T	16290T	16319A	16362C					
24	73G	146C	150T	151T	152C	182T	217C	263G	315.1C	16188.1C	16193.1C	16362C			
25	73G	146C	263G	309.1C	315.1C	16182C	16183C	16189C	16217C	16223T	16261T				
26	73G	185A	249DEL	263G	309.1C	309.2C	315.1C	16092A	16291T	16304C					
27	73G	194T	263G	315.1C	16188.1C	16223T	16278T	16362C							
28	73G	199C	263G	309.1C	315.1C	16129R	16223T	16278T	16290T	16293G	16354T				
29	73G	234R	263G	315.1C	16136C	16175G	16183C	16189C	16217C	16218T	16311C	16324C			
33	73G	146C	150T	263G	309.1C	315.1C	16140Y	16182C	16183C	16189C	16217C	16274A	16311C	16335G	
34	73G	150T	152C	263G	309.1C	309.2C	315.1C	16124C	16148T	16182C	16183C	16189C	16223T	16287T	16362C
35	73G	150T	263G	309.1C	315.1C	16093Y	16129A	16223T	16257A	16261T					
37	73G	143R	183G	263G	309.1C	315.1C	16189C	16288C	16304C	16309G					
38	73G	146C	263G	307DEL	308DEL	309DEL	312.1A	16182C	16183C	16189C	16217C				
39	73G	146C	207A	263G	309.1C	315.1C	16231C								
40	73G	146C	199C	263G	309.1C	315.1C	16223T	16249C	16319A						
41	73G	263G	315.1C	16051G	16093C	16183C	16189C	16220G	16356C						
42	73G	93G	204C	210G	263G	315.1C	16140C	16183C	16189C	16266G	16362C				
44	73G	150T	263G	309.1C	309.2C	315.1C	16223T	16257A							
45	73G	195C	263G	309.1C	315.1C	16051G	16189C	16223T	16256T	16362C					
Q	73G	150T	263G	309.1C	315.1C	16223T	16311C	16325C	16362C						
G	73G	146Y	263G	309.1C	309.2C	315.1C	16182C	16183C	16189C	16217C	16261T	16354T			
X	73G	153G	263G	309.1C	315.1C	16223T	16234T	16248T	16265C	16316G	16362C				
Y	73G	249DEL	263G	309.1C	315.1C	16136C	16154C	16183C	16189C	16217C	16218T				

Table 2 Comparisons of the sequences from NGS and the Sanger method with capillary electrophoresis.

Sample no.	Nucleotide sequence*		Reads for A (%)	Reads for C (%)	Reads for G (%)	Reads for T (%)
	Sanger method with capillary electrophoresis	NGS				
12	189A	189R	869 (70.4)		366 (29.6)	
17	189A	189R	1045 (70.6)		436 (29.4)	
33	16140C	16140Y		1592 (79.1)		420 (20.9)

* The nucleotide sequence is named following the IUPAC rule.

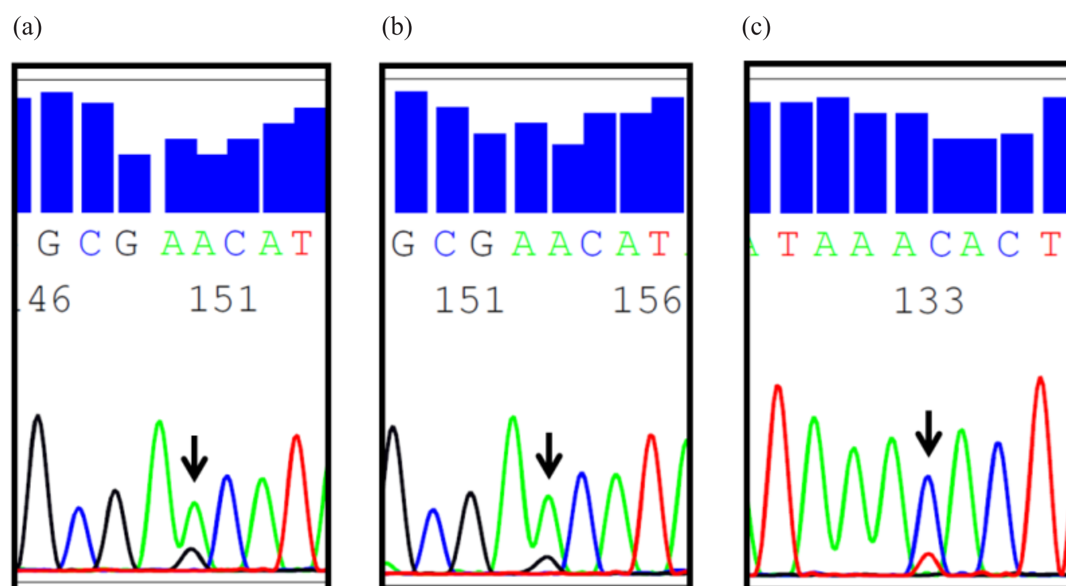

Fig. 1 Electropherograms for sample 12 (a), sample 17 (b) and sample 33(c). Arrows indicate the 189 site in (a), 189 site in (b) and 16140 site in (c).

For analysis of C-stretch in HV1 (16184-16193) and HV2 (303-315), totally there were 14 types for HV1 and HV2 combinations of the samples analyzed in this study (Table 3). The most popular type was type 10 (32.5%), and there were 6 types (type 2, 4, 5, 12, 13 and 14) observed only in one sample for each of them. The C-stretch for HV1, there were 37.5% samples observed without interruption by other nucleotides.

Table 3 C-stretch types for HV1 and HV2 of the samples sequenced by NGS in this study.

Type of the combination	HV1	HV2	Sample	Frequency (%)
1	C10	C8TC6	3,37,45	7.5
2	C11	C7TC5	2	2.5
3	C11	C7TC6	29,41,42	7.5
4	C11	C8TC6	Y	2.5
5	C12	C4TC2AC3	38	2.5
6	C12	C8TC6	15,17,25,33	10.0
7	C12	C9TC6	34,Q	5.0
8	C3TC6	C7TC6	1,16	5.0
9	C5TC4	C7TC6	5,6,18,23	10.0
10	C5TC4	C8TC6	4,9,10,11,12,13,20,28,35,39,40,G,X	32.5
11	C5TC4	C9TC6	9947A,26,44	7.5
12	C6TC4	C7TC6	27	2.5
13	C6TC5	C7TC6	24	2.5
14	CAC9	C9TC6	19	2.5

These results showed that mtDNA whole genome was sequenced successfully by NGS and the sequences from NGS were explicit.

Race analysis

In this study, haplotype classification for these samples were performed by using the HaploGrep [https://haplogrep.uibk.ac.at] to determine the race of them. The result was showed in Table 4. The positive control standard 9947A was classified to be H11b1, which is the common haplotype in Central Europe [10]. All the forensic samples collected in Taiwan were classified to be the haplotypes common in Asia [11]. The results are consistent with the prediction.

Table 4 Haplogroup classification of the samples analyzed in this study.

Haplogroup	Sample
A20	23
B4	15,19,25,29,33,38,Y
B5	1,16,42
B6	41
D4	5,27
D5	17,24,34
D6	3
E2a	10,45
F1	9,13
F2	6,26
G1a1a	G
G2a1f1	4
M7	11,40
M9a1a	X
M20	12
M73b	28
N9a	20,35,44
R0	2,Q
R9	18,37
Y1b	39
H11b1	9947A

Conclusions

This is the first study to sequence the forensic samples by using NGS in Taiwan. The results showed that mtDNA whole genome was sequenced successfully by NGS and the sequences from NGS were explicit. Furthermore, as NGS has the characteristics of high throughput, massively parallel sequencing and deep sequencing, it is more time-saving and the quantitative results are available. From the results of this study, it showed the feasibility for applications of NGS on mitochondrial DNA analysis for forensic samples.

Acknowledgements

The authors wish to express their gratitude to The Ministry of Justice for financial support (Science and Technology Program number 104-1301-05-05-06 and 105-1301-05-05-01).

References

1. Holland M-M, McQuillan M-R, O'Hanlon K-A. Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. Croat Med J 2011;52(3):299-313.
2. Sosa M-X, Sivakumar I-K, Maragh S, Veermachaneni V, Hariharan R, Parulekar M, Fredrikson K-M, Harkins T-T, Feldman A-B, Tata P, Ehret G-B, Chakrevarti A. Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency. PLoS Comput Biol 2012;8(10):e1002737.
3. Lin C-Y, Tsai L-C, Hsieh H-M, Huang C-H, Yu Y-J, Tseng B, Linacre A, Lee JC-I. Investigation of length heteroplasmy in mitochondrial DNA control region by massively parallel sequencing. Forensic Sci Int Genet 2017;30:127–133.
4. Xue Y, Tyler-Smith C. The hare and the tortoise: one small step for four SNPs, one giant leap for SNP-kind. Forensic Sci Int Genet 2010;4(2):59–61.
5. Hanson E, Haas C, Jucker R, Ballantyne J. Specific and sensitive mRNA biomarkers for the identification of skin in 'touch DNA' evidence. Forensic Sci Int Genet 2012;6(5):548-558.
6. Weber-Lehmann J, Schilling E, Gradl G, Richter D-C, Wiehler J, Rolf B. Finding the needle in the haystack: differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. Forensic Sci Int Genet 2014;9:42-46.
7. Gunnarsdóottir E-D, Li M, Bauchet M, Finstermeier K, Stoneking M. High-throughput sequencing of complete human mtDNA genomes from the Philippines. Genome Res 2011;21(1):1-11.
8. Stawski H, Bintz B-J, Burnside E-S, Wilson M. Preparing Next Generation Sequencing (NGS) Libraries of Human Mitochondrial DNA Using Illumina Nextera XT and NEBNext dsDNA Fragmentase Technology. Proceedings of the American Academy of Forensic Sciences. 2013.
9. Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, Kronenberg F. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat 2011;32(1):25-32.
10. Álvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintáns B, Zarrabeitia M-T, Cuscó I, Lareu M-V, García O, Pérez-Jurado L, Carracedo A, Salas A. New Population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. PLoS ONE 2009;4(4):e5112.
11. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 2009;30(2):E386-94.