LENGTH AND SEQUENCE HETEROPLASMY IN mtDNA D-LOOP REGION OF INDIVIDUALS FROM KINH POPULATION IN SOUTHERN VIETNAM

Nguyen Thi Lan Huong*, Nguyen Ngoc Hoa Binh, Dang Thanh Xuan, Mai Quang Truong

Forensic Medicine Center of Ho Chi Minh City, HCMC, Vietnam

SUMMARY

In forensic science, the control region of human mitochondrial DNA (mtDNA), mainly the hypervariable regions I (HVI), II (HVII), and III (HVIII), which are located at positions 16024 to 576, are the most extensively studied. These regions provide a high index of polymorphism. However, high mutation rates in mtDNA can cause heteroplasmy, complicating the interpretation of mtDNA results. In spite of this, there are currently few studies that describe this issue. To gain further insights in mtDNA heteroplasmy, this work investigated the presence of heteroplasmy (length and sequence) in the hypervariable regions (HVI, HVII and HVIII) on D-loop segment of the mtDNA from 396 unrelated healthy Kinh individuals using standard Sanger sequencing method with separation by capillary electrophoresis. All the subjects displayed length and sequence heteroplasmies in the HVI, HVII and HVIII regions. From position 16180 to 16195, 15 patterns of polycytosine or C-stretch were observed (125 samples, 31.56%), 221 Kinh individuals (55.81%) were found to have similar pattern to the reference, and 8 patterns (50 samples, 12.63%) didn't have polycytosine but were different from the reference. From position 303 to 315, 8 patterns were observed, whereas 7CT6C and 8CT6C were the most frequently found. From position 514 to 523, there were 4 motifs of $(AC)_n$, including $(AC)_4$ to $(AC)_7$. From position 568 to 573, there were 6 samples (1.52%) with insertion of +2C, +3C, and +4C at np 573. In addition, there were 8 samples (2.02%) with sequence heteroplasmy, in which 7 positions were detected. The frequency of heteroplasmy was also calculated. The nomenclature of variants was established according to EMPOP guidelines. The study provided a new perspective with important consequences in medical, evolutionary and forensic fields.

Keywords: Control region, heteroplasmy, Kinh population, mitochondrial DNA.

INTRODUCTION

In forensic science, the nucleotide sequence analysis of the control region of human mtDNA, mainly in hypervariable regions I (HVI), II (HVII) and III (HVIII), has been categorized as a useful tool for personal identification and maternal testing. HVI is located at position 16,024 to 16,365, HVII at position 73 to 340 and HVIII at position 438 to 576 (Butler, 2011). These regions provide a high index of polymorphism. Due to the huge number of mtDNA copies existing in one human body and high mutation rate of mtDNA, all individuals are expected to exhibit some level of heteroplasmy, in which some copies of that individual's mtDNA have a different sequence than the others (Melton, 2004). In other words, heteroplasmy is the presence of more than one type of mtDNA in an individual (Melton, 2004). Two or more mtDNA populations may occur between cells in an individual, within a single cell, or within a single mitochondrion (Butler, 2011). Length heteroplasmy often occurs around the homopolymeric C-stretches while sequence heteroplasmy is typically detected by the presence of two nucleotides at a single site, showing up as overlapping peaks in a sequence electropherogram (Butler, 2011). Heteroplasmy has also been reported to remain stable over time in the same individuals and thus be inherited rather than age related (Lagerström-Fermér *et al*., 2001). Although heteroplasmy can sometimes complicate the interpretation of mtDNA results, the presence of heteroplasmy at identical sites can improve the probability of a match (Ivanov *et al*., 1996).

Understanding the levels of human mtDNA heteroplasmy is vital in several fields, especially in forensic science. However, currently there are few works that focus on this issue. To gain further insights in mtDNA heteroplasmy, this work investigated the presence of heteroplasmy (length and sequence) in the hypervariable regions (HVI, HVII and HVIII) on the D-loop segment of the mtDNA.

MATERIALS AND METHODS

Materials

The buccal cells were collected by sterile cotton swabs from 396 unrelated healthy Kinh individuals. This information was ascertained through inquiry before sampling. The data were collected from cases examined at the Forensic Medicine Center of Ho Chi Minh City from 2018 to 2022.

Methods

Extraction of genomic DNA: Genomic DNA was extracted from buccal swabs using the QIAamp DNA Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. The quantity of recovered DNA was determined using Invitrogen™ Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, USA).

PCR amplification, DNA purification, and sequencing:

PCR amplification: The extended D-loop region (about 1.3 kb in size, np 15973 to 632) was amplified with the primer sets of 15973F (5'-AACTCCACCATTAGCACCCAAAG-3') and 632R (5'-GTGAGCCCGTCTAAACATT-3') (Sigma, Germany). PCR was performed in 25 µL of a volume containing 6.75 µL of DNA solution (1-5 ng), 2.5 µL of each 5 µM primer, 0.75 µl DMSO 10%, and 12.5 µL Phusion® Hot Start Flex 2X Master Mix (Thermo Fisher Scientific, USA). Thermal cycling was carried out on the Veriti™ 96-Well Thermal Cycler (Thermo Fisher Scientific, USA), began with 30 s at 98°C, followed by 30 cycles of 5 s at 98°C, 10 s at 55°C, and 30 s at 72°C, and the final extension at 72°C for 5 minutes.

Size selection and purification: All PCR products were analyzed by electrophoresis on a 1.5% agarose gel and visualized by GelRed™ Nucleic Acid Stain (Thermo Fisher Scientific, USA) staining under a UV transilluminator (Wealtec, USA). The PCR products were purified using QIAquick® Gel Extraction Kit (Qiagen, Germany).

Sequencing: The purified amplicons were sequenced by Sanger method using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, USA) with 6 primers on the capillary electrophoresis benchtop instrument ABI 3130 Genetic Analyzer (Applied Biosystems, USA).

Statistical analysis: The data were preliminarily treated by Sequencing Analysis 5.4 software. The nucleotide sequences were analyzed and aligned with the revised Cambridge Reference Sequences (rCRS) using Sequencher 5.4.5 software. The frequencies of each heteroplasmy were facilitated by dividing the number of times the heteroplasmy observed in a population by the total number of examined samples at a particular genetic locus in the D-loop region.

RESULTS

Nucleotide variations in C-stretch at position 16180-16195 in HVI

All of the HVI variable sequences in 396 Kinh individuals were analyzed and compared with the rCRS using Sequencher 5.4.5 software and rechecked by the expanded rCRS (Figure 1). Some homopolymeric sequences were observed in some individuals in the control region between position 16180 and position 16195. There were 56 samples where nucleotide insertions occurred, causing the sequence of the 16180-16195 region to extend to 17 and 18 nucleotides instead of 16 nucleotides like the rCRS sequence.

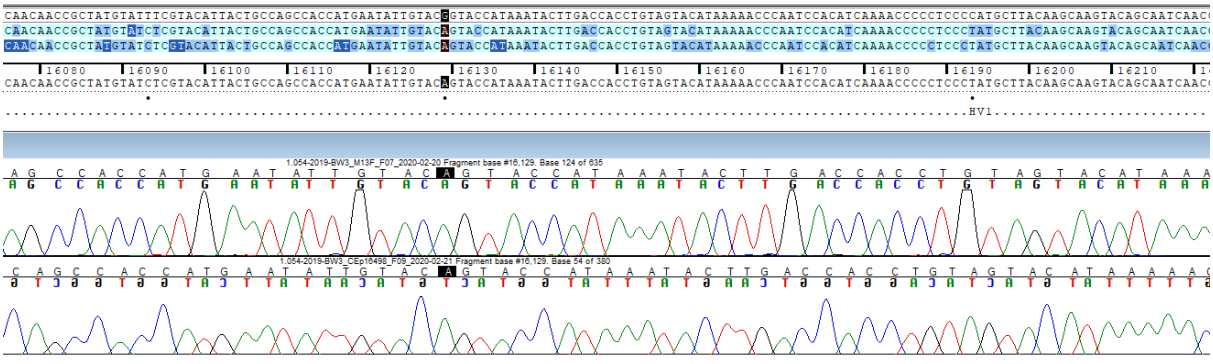

Figure 1. Sequence analysis of the mitochondrial D-loop region

The electropherogram shows the differences between the sample and the rCRS (top row). Each sample was sequenced more than once to check for accuracy. Dots indicate where differences occurred.

Table 1. Sequence types of polyC1 (position 16180 to 16195) in the mtDNA control region for 396 Kinh individuals

(Ref: AAAACCCCCTCCCCAT, bold; Base change: red, bold, underline; Symbol ":" means absence of one nucleotide)

Sequence	SNP	Number of Observations	Frequency
AAAACCCCCTCCCCAT	rCRS	**221**	**0.5581**
AAAACCCCC**C**CCCCAT	16189C	12	0.0303
AAAACCCCC**C**CCCC**C**AT	16189C, 16193.1C	2	0.0051
AAAACCCCC**C**CCCC**CC**AT	16189C, 16193.1C, 16193.2C	5	0.0126
AAAACCCCC**C**CCC**T**CAT	16189C, 16192.1T	3	0.0076

AAAACCCCC**C**CC**T**CAT	16189C, 16192T	4	0.0101
AAAACCCCCTCCC**T**AT	16193T	2	0.0051
AAAACCCCCTCC**T**CAT	16192T	28	0.0707
AAAACCCC**TC**CCC**T**CAT	16188T, 16189C, 16192.1T	1	0.0025
AAAACC**T**CC**C**CCCCAT	16186T, 16189C	1	0.0025
AAAAC**T**CCCTCCCCAT	16185T	10	0.0253
AAAA**T**CCCC**C**CCCCAT	16184T, 16189C	1	0.0025
AAAA**T**CCCCTCCCCAT	16184T	1	0.0025
AAA**C**CCCCC**C**CCCCAT	16183C, 16189C	3	0.0076
AAA**C**CCCCC**C**CCCC**C**AT	16183C, 16189C, 16193.1C	45	0.1136
AAA**C**CCCCC**C**CCC**T**CAT	16183C, 16189C, 16192.1T	1	0.0025
AAA**C**CCCCC**C**CC**T**CAT	16183C, 16189C, 16192T	1	0.0025
AAA**G**CCCCCTCCCCAT	16183G	2	0.0051
AA**CC**CCCCC**C**CCCCAT	16182C, 16183C, 16189C	39	0.0985
AA**G**ACCCCCTCCCCAT	16182G	5	0.0126
A**CCC**CCCCC**C**CCCCAT	16181C, 16182C, 16183C, 16189C	4	0.0101
GA**CC**CCCCC**C**CCCCAT	16180G, 16182C, 16183C, 16189C	1	0.0025
AAA**:**CCCCC**A**CCCCAT	16183delA, 16189A	1	0.0025
AAA**CA**CCCCC**C**CCCCAT	16183C, 16184A, 16189C	3	0.0076

In this study, it was revealed that the polycytosine stretch (C-stretch or polyC) observed in the HVI region had a total of 24 patterns of nucleotide sequence variations (Table 1). 15 patterns of C-stretch were observed (125 samples, 31.56%), whereas 221 Kinh individuals (55.81%) were found to be similar to the reference and 8 patterns (50 samples, 12.63%) didn't have polycytosine but were different from the reference. Interestingly, C-stretches with the nucleotide transitions from T to C at position 16189 were observed in 125 Kinh individuals (31.56%). In the C-stretches region, repetitions of cytosine with more than 9 nucleotides were observed in 112 Kinh individuals. Additionally, nucleotide transition (A → G) at position 16180 was found in one Kinh individual and one individual with A deletion at position 16183 was identified.

Nucleotide variations in C-stretch at position 303-315 in HVII

In this study, 8 patterns were observed at position 303-315 in HVII region (Table 2). Notably, there was no sample which was identical to the rCRS. The insertions were observed within a stretch of seven Cs between 303 and 309 (+C, +2C, +3C) or within five Cs between 311 and 315 (+C) in the control region. At positions 303-309, 151 sequences with an insertion of a single C residue, 76 sequences with an insertion of two C residues and 5 sequences with an insertion of three C residues were identified. Finally, at 311-315, 391 control regions with 6 cytidine residues were found. It was notable that at position 303-309, there were 5 sequences with one C deletion and one sequence with 3C deletion. Additionally, nucleotide transition (T → C) at position 310 which generated polycytosine were found in four Kinh individuals.

Table 2. Sequence variations from position 303 to position 315 in HVII of mitochondrial DNA for 396 Kinh individuals
(Ref: CCCCCCCTCCCCC or 7CT5C, bold; Base change: red, bold, underline; Symbol ":" means absence of one nucleotide)

Patterns	Sequence	SNP	Number of Observations	Frequency
7CT5C	**CCCCCCCTCCCCC**	**rCRS**	**0**	**0**
13C	CCCCCCC**C**CCCCC	310C	4	0.0101
4CT6C	CCCC**:::**TCCCCC**C**	307delC, 308delC, 309delC, 315.1C	1	0.0025
6CT5C	CCCCCC**:**TCCCCC	309delC	1	0.0025
6CT6C	CCCCCC**:**TCCCCC**C**	309delC, 315.1C	4	0.0101
7CT6C	CCCCCCCTCCCCC**C**	315.1C	154	0.3889
8CT6C	CCCCCCC**C**TCCCCC**C**	309.1C, 315.1C	151	0.3813
9CT6C	CCCCCCC**CC**TCCCCC**C**	309.1C, 309.2C, 315.1C	76	0.1919
10CT6C	CCCCCCC**CCC**TCCCCC**C**	309.1C, 309.2C, 309.3C, 315.1C	5	0.0126

Length variation with C-stretch at position 568-573 in HVIII

From position 568-573, there were 6 samples (1.52%) (Table 3) with insertion of +2C (3 samples), +3C (1 sample), and +4C (2 samples) at np 573. When the polyC has 10Cs or more, the downstream sequence was not recorded accurately because the signal was unclear, which was caused by an out-of-phase phenomenon during Sanger sequencing

Table 3. Nomenclature of SNPs from position 566 to position 576 in HVIII
(Ref: CACCCCCCACA, bold; Base change: red, bold, underline)

Sequence	SNP	Number of Observations
CACCCCCCACA	**rCRS**	**390**
CACCCCCC**CC**ACA	573.1C, 573.2C	3
CACCCCCC**CCC**ACA	573.1C, 573.2C, 573.3C	1
CACCCCCC**CCCC**ACA	573.1C, 573.2C, 573.3C, 573.4C	2

AC repeats at nucleotide position 513-522 in HVIII

$(AC)_n$ is a kind of length polymorphism in which the n represents the number of AC dinucleotide repeats (from position 513 to position 522 in the rCRS). Some previous studies recorded it as $(CA)_n$ motif. A total of four repeat motifs [$(AC)_4$ to $(AC)_7$] were detected: 195 sequences with $(AC)_4$, 196 with $(AC)_5$, 4 with $(AC)_6$ and 1 with $(AC)_7$ repeats (Table 4). $(AC)_4$ and $(AC)_5$ exhibited the most frequent occurrence whereas $(AC)_7$ was the rarest.

Table 4. Length variation from position 513 to position 522 in HVIII region of mitochondrial DNA for 396 Kinh individuals
(Ref: ACACACACAC or $(AC)_5$, bold)

Sequence	Number of Observations	Frequency
ACACACAC or $(AC)_4$	195	0.4924
ACACACACAC or $(AC)_5$	**196**	**0.4949**
ACACACACACAC or $(AC)_6$	4	0.0101
ACACACACACACAC or $(AC)_7$	1	0.0025

Sequence heteroplasmy

There were 8 samples (2.02%) with sequence heteroplasmy: Y (T/C) and R(A/G) (Figure 2). Among the 8 individuals who possessed detectable sequence heteroplasmy, only one sample exhibited mixed bases at 2 positions (16111 C>Y, 198 C>Y), the remaining samples were heteroplasmic at one position. The sequence included 16111 C>Y, 16247 A>R, 16295 C>Y, 16299 A>R, 16311 T>Y, 198 C>Y, and 215 A>R. These SNP occurred at both HVI and HVII regions. The total number of sequence heteroplasmic positions recorded was 7 positions, of which 5 positions (71.43%) occur in the HVI region, the remaining 2 positions (28.53%) occur in the HVII region. No sequence heteroplasmy was detected in the remaining regions on the D-loop. Among the 7 heteroplasmic positions, position 215 was observed three times, the others occurred only once (Table 5).

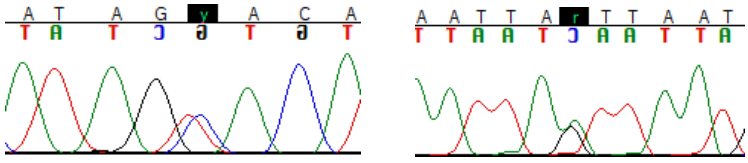

Figure 2. Sequence heteroplasmy at 16311Y (T/C) and 215R (A/G)

Table 5. Positions along the control region at which heteroplasmy was observed

Nucleotide position	Number of Observations	rCRS	Base change
198	1	C	Y
215	3	A	R
16111	1	C	Y
16247	1	A	R
16295	1	C	Y
16299	1	A	R
16311	1	T	Y

DISCUSSION

PolyC in HVI

The nucleotide transition (T>C) at position 16189, the nucleotide transversion (A>C) at position 16181 and/or 16182 and/or 16183 and the nucleotide C insertion at position 16193 resulted in the repetition of cytosine by more than 10 times. This region carries valuable genetic markers that could discriminate one ancestral population from another. The HVI C-stretch was observed in the Thai, Chinese Tu, Asian (Chinese, Taiwanese, Korean, and Japanese), European, and Amerindian Populations (Sangthong *et al*., 2015). This homopolymeric stretch of cytosines create problems for DNA polymerases as they synthesize the complementary strand to the mtDNA template present in the PCR reaction (Butler, 2011). The impact of a 16189 T to C transition on the sequencing result downstream of the C-stretch region could be seen in Figure 3.

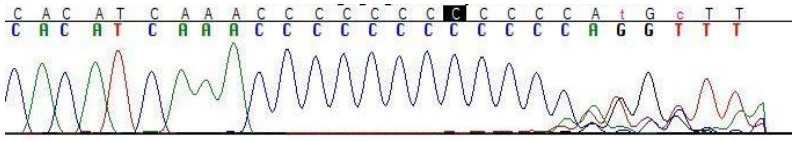

Figure 3. A sample with the C-stretch in HVI

Frequently, approximately 15% of Europeans get a T-to-C transition at position 16189, relative to the CRS (Bendall *et al*., 1995). In the phylogenetic context, position 16189 was commonly observed in different mtDNA lineages (Finnila *et al*., 2001). This position was among the 50 fastest evolving sites (Irwin *et al*., 2009). 16189 T>C transition in the HVI region could be associated with disease (Bendall *et al*., 1995). Chinnery *et al*. (2005) suggested that the 16184-16193 polyC tract has a very small role in the pathophysiology of type 2 diabetes (Chinnery *et al*., 2005) while Liou *et al*. (2010) reported that the mtDNA 16189 variant could cause alteration of mtDNA copy number in human blood cells (Liou *et al*., 2010).

PolyC in HVII

In the HVII region, at position 303-315, a polycytosine tract with a single thymidine inserted at position 310 creates length polymorphism among individuals, as well as variation within an individual, associated with aging and cancer. This sequence was involved in forming a persistent RNA-DNA hybrid that led to the initiation of mtDNA heavy-strand replication (Frigi *et al*., 2009). The mtDNA length heteroplasmy in HVI and HVII regions played a role in determining mtDNA copy number (Zhao *et al*., 2010). In the study, most samples showed insertions of C residues, located within a stretch of seven Cs between 303 and 309 or within five Cs between 311 and 315. There were 4 haplotypes containing 13C in the region 303-315. Substitution of T to C has also been reported. Zahidin *et al* (2018) reported 2 cases of T to C substitution, leading to polyC in the HVII region with 16C and 17C (Zahidin *et al*., 2018). Similar to HVI, the phenomenon of C insertion in the region 303-315 also sometimes caused difficulty in sequencing (Figure 4).

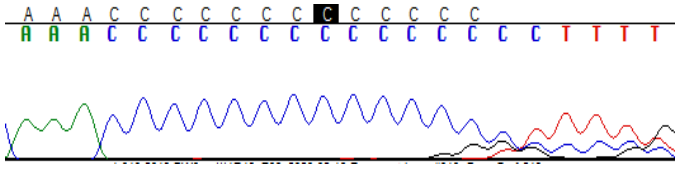

Figure 4. C-stretch region of HVII region between 309 and 315 with substitution of T to C at np 310

PolyC in HVIII

C-stretch in HVIII is considered as a hotspot for length heteroplasmy in EMPOP. Thus, it is ignored for calculating haplotype frequency and forensic genetic indices (Brandstätter *et al*., 2004). Accurate recording of the 6 Cs in this region was performed using CE, but this was not possible for longer than 10 Cs (Figure 5). Brandstätter *et al*. (2004) reported that polyC spanning nucleotides 568-573 were present in 5% of the individuals from Nairobi, Kenya. Irwin *et al*. (2009) recorded 3% of the samples showed length heteroplasmy with expansion of the C-stretch beyond the 6 Cs in most individuals.

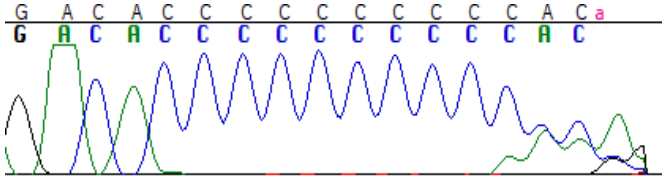

Figure 5. Insertion of +4C at np 573

$(AC)_n$ patterns in HVIII

Even though the HVI and HVII sequences were utilized in most forensic cases, it is remarkable that HVIII $(AC)_n$ repeat segment also plays a key role as an additional mtDNA marker in forensic investigations (Ivanov *et al*., 1996). In some previous studies, $(AC)_5$ was the most common motif among Pakistani, Iraqi, Malaysian, Korean, Germans, Japanese, and Indian Muslims, while $(AC)_4$ was common among Thais, Venezuelans, Cameroonians and Urali Kuruman.

Sequence heteroplasmy

The detection of mixed positions that may represent sequence heteroplasmy from Sanger-based data in this study relied primarily on repeated visual inspection of properly aligned electropherogram traces by experienced examiners. Shared sequence heteroplasmy between maternal relatives can increase the strength of the mtDNA evidence in a case of historical significance (Ivanov *et al*., 1996). Irwin *et al*. (2009) examined population-based mtDNA datasets developed via Sanger sequencing and identified sequence heteroplasmy in the control region in approximately 6% of individuals when buccal and blood cells were examined. The control region nucleotide positions at which sequence heteroplasmy has been most frequently observed are consistent with the positions with the highest substitution rates (Irwin *et al*., 2009). However, there are a few notable exceptions. The variable rates are primarily accounted for by differences in the tissue types, and populations examined. The incidence of sequence heteroplasmy in tissues with high metabolic activity (e.g. 79% in muscle) has been generally higher than in blood or blood-derived specimens (4-8%) (Naue *et al*., 2015).

Sequence heteroplasmy, which occurred in HVI and HVII, was similar to the study of Irwin *et al*., 2009. Position 215 is not reported as an evolutionary fast site. This may be due to unusual mechanisms affecting variation at this site. Position 215 is situated close to the origin of heavy strand replication (Irwin *et al*., 2009). Santos *et al*. (2008) reported that the 215G variant may induce a conformational change in the control region. Therefore, mutations at this position could affect mtDNA replication (Irwin *et al*., 2009). In contrast to position 215, positions 198, 16111 and 16311 in this study are considered some of the fastest evolving sites according to the study of Irwin *et al*. (2009).

CONCLUSION

The study reported on heteroplasmy in the mtDNA control region in the Kinh population. From the 396 samples analyzed, examples of length heteroplasmy caused by polycytosine stretches were found in HVI, HVII and HVIII regions. Sequence heteroplasmy occurred at both HVI and HVII regions. The presence of these variants can be visually identified in sequence electropherograms generated by capillary electrophoresis. Whereas heteroplasmy may complicate interpretation of the results, it actually becomes a “signature” of the donor's mtDNA haplotype, which can strengthen the identification of a sample. The analysis of heteroplasmy in D-loop is potentially useful for medical, evolutionary, and forensic purposes.

REFERENCE

Bendall KE, Sykes BC (1995). Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet* 57: 248-256.

Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, Parsons TJ (2004). Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int J Legal Med* 118: 294-306.

Butler JM (2011). Advanced topics in forensic DNA typing: methodology. *Academic Press*: 405.

Chinnery PF, Elliott HR, Patel S, Lambert C, Keers SM, Durham SE, McCarthy Hitman GA, Hattersley AT, Walker M (2005). Role of the mitochondrial DNA 16184–16193 poly-C tract in type 2 diabetes. *Lancet* 366: 1650-1651, DOI:10.1016/S0140-6736(05)67492-2.

Finnila S, Lehtonene MS, Majamaa K (2001). Phylogenetic network for European mtDNA. *Am J Hum Gen* 68: 1475-1484.

Frigi S, Ennafaa H, Cherni L, Gaaied AE (2009). Investigation of polymorphisms in hypervariable region II (HVII) sequences of human mitochondrial DNA in two North-African populations. *Anthropologie* 107-118.

Hua Zhao, Jie Shen, Leonard Medico, Mary Platek, Christine B Ambrosone (2010). Length heteroplasmies in human mitochondrial DNA control regions and breast cancer risk. *Int J Mol Epidemiol Genet* 1(3): 184-192.

Irwin JA, Saunier JL, Niederstätter H, Strouss KM, Sturk KA, Diegoli TM, Parsons TJ (2009). Investigation of heteroplasmy in the human mtDNA control region: A Synthesis of Observations from More Than 5000 Global Population Samples. *J Mol Evol* 68(5): 516-527.

Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996). Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet* 12: 417-420.

Lagerström-Fermér M *et al.* (2001). Heteroplasmy of the human mtDNA control region remains constant during life. *Am J Hum Genet* 68: 1299-1301.

Liou CW, Lin TK, Chen JB, Tiao MM, Weng SW, Chen SD, Chuang YC, Chuang JH, Wang PW (2010). Association between a common mitochondrial DNA D-loop polycytosine variant and alteration of mitochondrial copy number in human peripheral blood cells. *J Med Genet* 47: 723-728.

Melton T (2004). Mitochondrial DNA heteroplasmy, *Forensic Sci Rev* 16: 1-20.

Naue J, Horer S, Sanger T, Strobl C, Hatzer-Grubwieser P, Parson W *et al.* (2015). Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. *Mitochondrion* 20: 82-94.

Sangthong P, Janso A, Chinnabanchonchai N (2015). Sequence analysis of mitochondrial DNA hypervariable region I in Thai individuals. *Aust J Forensic Sci* 47(3): 345-354.

Santos C, Sierra B, Alvarez L, Ramos A, Fernandez E, Nogues R, Aluha MP (2008). Frequency and pattern of heteroplasmy in the control region of human mitochondrial DNA. *J Mol Evol* 67: 191-200.

Zahidin MA, Omar WBW, Taib WRW, Japning JRR, Abdullah MT (2018). Sequence polymorphism and haplogroup data of the hypervariable regions on mtDNA in Semoq Beri population. *Data in Brief* 21: 2609-2615.

ĐA HÌNH TRÌNH TỰ VÙNG SIÊU BIẾN VÀ HIỆN TƯỢNG DỊ THỂ BỘ GEN TY THỂ CỦA CÁC CÁ THỂ DÂN TỘC KINH Ở MIỀN NAM VIỆT NAM

Nguyễn Thị Lan Hương*, Nguyễn Ngọc Hòa Bình, Đặng Thanh Xuân, Mai Quang Trường

Trung tâm Pháp y Thành phố Hồ Chí Minh, Việt Nam

TÓM TẮT

Trong khoa học pháp y, vùng điều khiển của DNA ty thể người (mtDNA), gồm các vùng siêu biến I (HVI), II (HVII) và III (HVIII), nằm ở vị trí 16024 đến 576, được nghiên cứu rộng rãi nhất. Những vùng này cung cấp chỉ số đa hình cao. Tuy nhiên, tỷ lệ đột biến cao ở mtDNA có thể gây ra hiện tượng dị thể, làm phức tạp việc giải thích kết quả mtDNA. Mặc dù vậy, hiện nay có rất ít nghiên cứu mô tả vấn đề này. Để hiểu rõ hơn về dị thể mtDNA, đề tài này đã nghiên cứu sự hiện diện của dị thể (chiều dài và trình tự) trong các vùng siêu biến (HVI, HVII và HVIII) trên đoạn D-loop của mtDNA từ 396 cá thể người Kinh khỏe mạnh không có quan hệ, sử dụng phương pháp giải trình tự Sanger tiêu chuẩn dựa trên điện di mao quản. Tất cả các đối tượng đều hiển thị dị thể về chiều dài và trình tự ở các vùng HVI, HVII và HVIII. Từ vị trí 16180 đến 16195, có 15 kiểu polycytosine (125 mẫu, 31,56%), trong khi 221 cá thể Kinh (55,81%) được phát hiện giống với rCRS và 8 kiểu (50 mẫu, 12,63%) không có polycytosine nhưng khác với trình tự tham chiếu. Từ vị trí 303 đến 315, 8 kiểu trình tự đã được quan sát, trong đó 7CT6C và 8CT6C chiếm tần số cao nhất. Từ vị trí 514 đến 523 có 4 mô típ $(AC)_n$, gồm $(AC)_4$ đến $(AC)_7$. Từ vị trí 568 đến 573, 6 mẫu (1,52%) có chèn +2C, +3C, +4C tại vị trí 573. Ngoài ra, có 8 mẫu (2,02%) dị thể về trình tự, trong đó 7 vị trí được phát hiện. Danh pháp của các biến thể theo EMPOP đã được đề cập. Nghiên cứu này cung cấp một góc nhìn mới với những hệ quả quan trọng trong các lĩnh vực y tế, tiến hóa và pháp y.

Từ khóa: Dân tộc Kinh, DNA ty thể, hiện tượng dị thể, vùng điều khiển.

* Author for correspondence: Tel: 0974487053; Email: ntlhuong226@gmail.com