

## ĐA HÌNH VÙNG D-LOOP HỆ GEN TY THỂ CỦA CÁC CÁ THỂ DÂN TỘC KINH VÀ MĂNG CÙNG TRONG NHÓM NGŨ HỆ NAM Á

Nguyễn Thy Ngọc<sup>1,2</sup>, Nguyễn Bảo Trang<sup>1,3</sup>, Nguyễn Quang Huy<sup>4</sup>, Nguyễn Đăng Tôn<sup>1,5</sup>, Nguyễn Thùy Dương<sup>1,5</sup>, ✉

<sup>1</sup>Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

<sup>2</sup>Trường Đại học Khoa học và Công nghệ Hà Nội, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

<sup>3</sup>Cục Sở hữu trí tuệ, Bộ Khoa học và Công nghệ

<sup>4</sup>Trường Đại học Khoa học tự nhiên, Đại học Quốc gia Hà Nội

<sup>5</sup>Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

✉Người chịu trách nhiệm liên lạc. E-mail: tdnguyen@igr.ac.vn

Ngày nhận bài: 25.12.2017

Ngày nhận đăng: 22.6.2018

### TÓM TẮT

Hai đoạn siêu biến HV-I và HV-II của vùng D-loop là các đoạn DNA có nhiều đa hình nhất trong hệ gen ty thể người, có ý nghĩa quan trọng trong các nghiên cứu về di truyền học quần thể cũng như tiến hóa. Với mục đích khảo sát các đa hình các nucleotide đơn (SNPs) thuộc vùng D-loop của hệ gen ty thể, chúng tôi đã thu thập mẫu máu của 81 cá thể người Việt Nam thuộc hai dân tộc Kinh và Măng. Sử dụng kỹ thuật PCR và các cặp mồi được thiết kế đặc hiệu, hai đoạn siêu biến HV-I và HV-II với kích thước tương ứng 693 bp và 689 bp đã được khuếch đại thành công. Tiếp đó, chúng tôi đã tiến hành đọc trình tự nucleotide trực tiếp sản phẩm PCR tinh sạch. Sau khi so sánh trình tự DNA thu được với trình tự genome ty thể tham chiếu rCRS công bố trên Genbank (NC\_012920.1), kết quả đã phát hiện 96 điểm đa hình ở nhóm người Kinh và 36 điểm đa hình ở nhóm người Măng. Trong số các điểm đa hình này, 16 điểm có tần suất thu được khác nhau có ý nghĩa thống kê giữa 2 nhóm ( $p < 0,05$ ). Ngoài ra, 8/16 đa hình, T146C, T199C, A16182C, T16217C, T16297C, T16140C, A16183C và T16189C xuất hiện khá phổ biến ở nhóm người Kinh nhưng ít hoặc không xuất hiện ở nhóm người Măng và bốn đa hình, C151T, A16162G, A16269G và T16271C xuất hiện phổ biến ở người Măng nhưng ít xuất hiện ở người Kinh. Kết quả nghiên cứu này cho thấy mặc dù cùng thuộc nhóm ngữ hệ Nam Á, hệ gen ty thể giữa 2 dân tộc trong cộng đồng các dân tộc Việt Nam vẫn tồn tại những khác biệt đáng chú ý.

**Từ khóa:** Hệ gen ty thể, D-loop, HV-I, HV-II, dân tộc Kinh, dân tộc Măng

### MỞ ĐẦU

Trong các bào quan của một tế bào, ty thể được coi là trung tâm tổng hợp năng lượng cho mọi hoạt động của tế bào. Chúng đóng vai trò tổng hợp nên hầu hết các năng lượng nội bào thông qua quá trình phosphoryl hóa oxy hóa và vận chuyển electron giữa các lớp màng ty thể. Một tế bào có trung bình khoảng 4000 ty thể. Các ty thể có vật chất di truyền riêng, dưới dạng mạch DNA vòng kép, và tồn tại độc lập với hệ gen nhân của tế bào chủ. Ở người, genome ty thể có kích thước 16569 bp, chứa 37 gen gồm: 2 gen mã hóa cho rRNA, 22 gen mã hóa cho tRNA, và 13 gen mã hóa cho các tiểu đơn vị enzyme có tham gia vào quá trình phosphoryl hóa oxy hóa, chuyển hóa năng lượng từ các chất dinh dưỡng hấp thu thành

dạng mà tế bào có thể sử dụng là adenosine triphosphate (ATP) (Taanman, 1999). Vùng duy nhất không mã hóa ở DNA ty thể người là vùng điều hòa, hay còn gọi là vùng D-loop, chiếm khoảng 7% chiều dài ty thể, chứa các trình tự promoter cho quá trình phiên mã và trình tự khởi đầu sao chép của hai chuỗi nặng và nhẹ.

Vùng D-loop được cấu thành từ 2 đoạn trình tự siêu biến (hypervariable region) HV-I và HV-II với kích thước lần lượt là 546 bp và 576 bp. Đây là vùng có tần suất xuất hiện đa hình cao nhất so với các vùng khác của ty thể. Nghiên cứu khảo sát các biến thể đa hình trên vùng gen này trên các quần thể người khác nhau có ý nghĩa quan trọng trong việc nghiên cứu lịch sử tiến hóa của các dân tộc (Kawabe *et al.*, 2014; Xu & Hu, 2015; Yao *et al.*, 2002), cũng

như trong giám định hài cốt, giám định pháp y, điều tra tội phạm (Chemale *et al.*, 2013).

Từ năm 2004 trở lại đây, một số các nghiên cứu đa hình trên vùng D-loop ở một số các dân tộc người Việt Nam đã được thực hiện. Năm 2005, Huỳnh Thị Thu Huệ và đồng tác giả đã công bố công trình đầu tiên về phân tích trình tự vùng D-loop của 5 cá thể người Việt Nam thuộc 3 dân tộc Kinh, Tày và H'Mông (Mông). Năm 2008, Nguyễn Đăng Tôn và đồng tác giả đã phát hiện 73 đa hình mới xuất hiện khi nghiên cứu khảo sát tần số các nhóm đơn bội (haplogroup) của ty thể ở 78 mẫu nghiên cứu của các cá thể người Việt Nam thuộc 3 dân tộc Kinh, Tày và Mường (Nguyễn Đăng Tôn *et al.*, 2008). Năm 2016, Đỗ Mạnh Hưng và đồng tác giả đã phát hiện được 79 nhóm đơn bội khi phân tích trình tự đoạn HV-II của 169 cá thể người Việt Nam thuộc 3 dân tộc Kinh, Mường, Jarai và Ê đê (Hung *et al.*, 2016). Năm 2017, Trần Thị Thúy Hằng và các tác giả khác đã giải trình tự đoạn HV-I và HV-II của 100 cá thể thuộc hai dân tộc Kinh và Mường và phát hiện được 12 điểm đa hình phổ biến trên cả hai dân tộc này (Trần Thị Thúy Hằng *et al.*, 2017). Gần đây, nghiên cứu đa hình vùng D-loop của 622 cá thể người Việt Nam thuộc 6 dân tộc: Kinh, Tày, Nùng, Mông, Khơ me và Thái đã phát hiện rất nhiều điểm tương đồng giữa hệ gen ty thể người Việt với các dân tộc láng giềng (Pischedda *et al.*, 2017). Tuy nhiên hiện chưa có một nghiên cứu nào về trình tự hoàn chỉnh của vùng điều khiển D-loop của trên DNA ty thể của cá thể người dân tộc Mảng. Trong 54 dân tộc Việt Nam, người Mảng là dân tộc có dân số ít nhất khoảng 4000 người, với địa bàn sinh sống hiểm trở và tương đối cách biệt so với các dân tộc khác (Ban chỉ đạo Tổng điều tra dân số và nhà ở Trung ương, 2009). Trong nghiên cứu này, chúng tôi tiến hành phân tích trình tự hai đoạn gen siêu biến HV-I và HV-II trên hệ gen ty thể của các cá thể người thuộc dân tộc Mảng và so sánh với nhóm các cá thể người thuộc dân tộc Kinh. Đây là hai dân tộc cùng thuộc ngữ hệ Nam Á, dân tộc Kinh thuộc nhánh ngôn ngữ Việt Mường, còn dân tộc Mảng thuộc nhánh ngôn ngữ Môn – Khơ me, với nhiều điểm chung giống nhau về phong tục tập quán. Tuy nhiên, về đa dạng di truyền có thể tồn tại những khác biệt. Nghiên cứu này sẽ mở rộng hơn cơ sở dữ liệu về hệ gen ty thể của các dân tộc người Việt Nam.

## NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

### Nguyên liệu

Đối tượng nghiên cứu là mẫu máu ngoại vi thu

được từ các cá thể người Việt Nam khỏe mạnh thuộc hai dân tộc gồm 51 cá thể người Kinh và 37 cá thể người Mảng. Các cá thể được chọn lựa là những người không có cùng huyết thống, có ông bà bố mẹ cùng thuộc một dân tộc và đồng ý tham gia cung cấp mẫu máu cho nghiên cứu.

Các cặp mồi đặc hiệu để khuếch đại 2 đoạn gen HV-I và HV-II có trình tự như sau: HV-I-F(5'-CCACCATAGCACCCAAAGC-3'), HV-I-R (5'-CAGCGTCTCG

CAATGCTATC-3'), HV-II-F (5'-GCCTAAATAGCCCACACGTTC-3') và HV-II-R (5'-TTTATGGGGTGATGTGAGCC-3').

### Phương pháp nghiên cứu

#### Xác định các đa hình thuộc vùng HV-I và HV-II bằng phương pháp sinh học phân tử

DNA tổng số được tách chiết từ máu bằng kit GeneJET Whole Blood Genomic DNA Purification (Thermo Fisher Scientific Inc., Vilnius, Lithuania) theo quy trình của nhà sản xuất, được kiểm tra độ tinh sạch bằng phương pháp đo OD 260nm/280nm và điện di trên gel agarose 1,5%. Nồng độ của DNA tổng số được xác định bằng máy đo quang phổ Nanodrop Eppendorf Biospectrometer® basic. Phản ứng PCR khuếch đại các trình tự HV-I và HV-II có thành phần bao gồm: 1X đệm PCR; 2,5mM dNTP; 0,2μM cho mỗi mồi tương ứng; 1 unit Taq polymerase; 50ng DNA và H<sub>2</sub>O thêm vào đến tổng thể tích 20μl; chạy trên chu trình nhiệt: 95°C-5 phút; 40 chu kỳ (95°C- 45 giây, 55°C-30 giây, 72°C-1 phút); 72°C-8 phút; giữ sản phẩm sau phản ứng ở 4°C. Sản phẩm PCR được điện di kiểm tra trên gel agarose 1,5% và tinh sạch bằng GeneJET PCR Purification Kit (Thermo Scientific), sau đó được giải trình tự theo phương pháp Sanger sequencing sử dụng ABI Big Dye Terminator v3.1 Sequencing Standard Kit (Applied Biosystems, CA) trên máy giải trình tự ABI 3500 Genetic Analyzer (Applied Biosystems). Dữ liệu trình tự của các cá thể tham gia vào nghiên cứu được so sánh với trình tự DNA ty thể tham chiếu rCRS (Revised Cambridge Reference Sequence of the Human Mitochondrial DNA), Genbank ID NC\_012920.1 (Andrews *et al.*, 1999) bằng phần mềm BioEdit.

#### Phân tích thống kê các đa hình gen tìm được

Tần suất của từng đa hình phát hiện được trên 2 đoạn gen HV-I và HV-II của các cá thể thuộc 2 dân tộc Kinh và Mảng được so sánh với nhau để tìm ra các đa hình khác biệt có ý nghĩa thống kê. Số liệu

thu thập được phân tích thống kê sử dụng phần mềm SPSS bản 23.0 (IBM company, Chicago, United States), phần mềm excel bản Office 2013 (Microsoft Corp, Washington DC, US). Kiểm định Fisher-exact (2 phía) được sử dụng để so sánh tỷ lệ tần suất đa hình giữa 2 dân tộc Kinh và Mảng. Mức ý nghĩa được lấy với  $\alpha = 0,05$ .

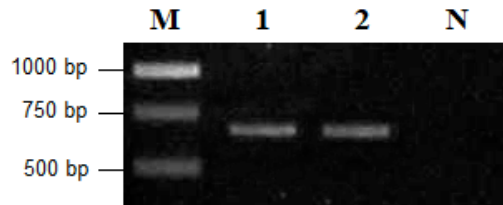
Vấn đề đạo đức trong Nghiên cứu Y sinh học:

Công trình này đã được Hội đồng đạo đức trong Nghiên cứu Y sinh học của Viện Nghiên cứu hệ gen thông qua và đồng ý cho phép tiến hành lấy mẫu, nghiên cứu trên người theo quyết định số 4-2015/NCHG-HĐĐĐ.

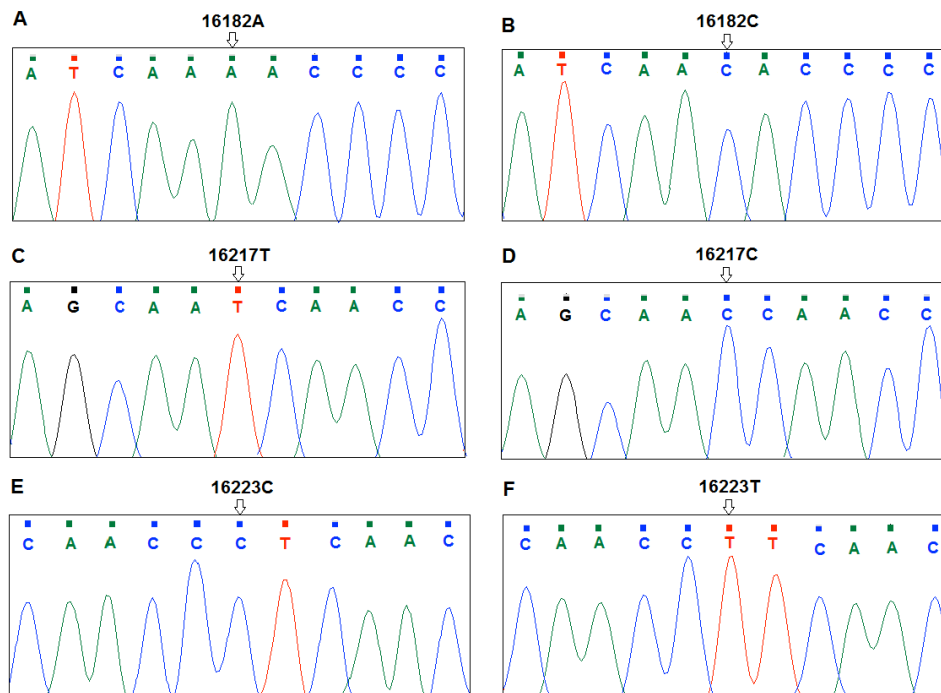
## KẾT QUẢ NGHIÊN CỨU

### Khuếch đại đoạn gen chứa trình tự gen HV-I và HV-II

DNA tổng số tinh sạch từ máu được sử dụng để tổng hợp hai đoạn HV-I và HV-II bằng cách sử dụng các cặp mồi đặc hiệu. Kết quả cho thấy sản phẩm PCR của hai đoạn HV-I và HV-II thu được là các băng đặc hiệu có kích thước khoảng 700 bp. PCR thu được sản phẩm là 1 băng đặc hiệu sắc nét với kích thước phù hợp với tính toán trên lý thuyết (Hình 1). Sản phẩm PCR của hai đoạn HV-I và HV-II được tinh sạch cho phản ứng đọc trình tự tiếp theo.



Hình 1. Sản phẩm PCR khuếch đại đoạn gen HV-I và HV-II và trên gel agarose 1,5%. M: DNA ladder 1kb (Thermo), 1: Sản phẩm PCR khuếch đại đoạn gen HV-I (kích thước lý thuyết 693 bp), 2: Sản phẩm PCR khuếch đại đoạn gen HV-II (kích thước lý thuyết 689 bp), N: Đối chứng âm (không có DNA khuôn).

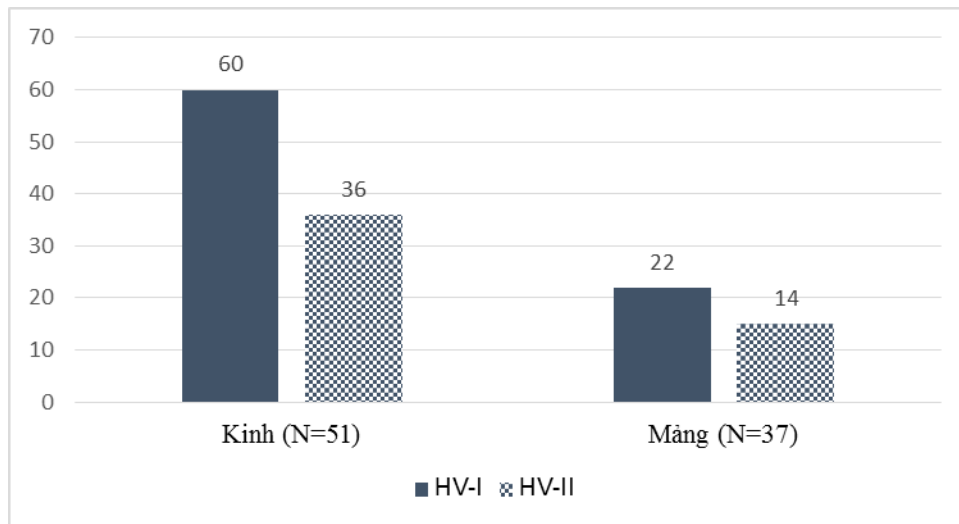


Hình 2. Kết quả giải trình tự một số mẫu đại diện chứa một số đa hình trên đoạn HV-I. A, C, E: Các đoạn trình tự của đoạn HV-I giống với trình tự tham chiếu rCRS đọc từ sản phẩm PCR của các cá thể tương ứng Kinh15, Mang307 và Mang280, B, D, F: Các đoạn trình tự của đoạn HV-I khác với trình tự tham chiếu rCRS đọc từ sản phẩm PCR của các cá thể tương ứng Kinh12, Mang05 và Mang276.

**Xác định trình tự các đoạn gen HV-I và HV-II**

Để khẳng định chính xác sản phẩm PCR là các đoạn HV-I và HV-II, sản phẩm PCR sau khi được tinh sạch được tiếp tục xác định trình tự nucleotide. Sau khi phân tích và so sánh trình tự hai đoạn HV-I và HV-II với trình tự rCRS trên GenBank (NC\_012920.1), kết quả thu được 106 điểm đa hình khác nhau, trong đó ở người Kinh có 60 đa hình thuộc đoạn HV-I và 36 đa hình thuộc đoạn HV-II (Hình 3). Ở dân tộc Kinh, có nhiều đa hình lần đầu tiên phát hiện thấy khi so với các nghiên cứu trước đây, đặc biệt là 6 đa hình phổ biến (có tần số allele từ

0,1 trở lên) ở trên nhóm người Kinh bao gồm: T152C (tần số allele  $f = 0,1$ ); T199C ( $f = 0,2$ ); T16140C ( $f = 0,18$ ); T16297C ( $f = 0,12$ ); T16311C ( $f = 0,14$ ) và T16362C ( $f = 0,14$ ) (Bảng 2). Ở người Mảng, số điểm đa hình được phát hiện ít hơn: 22 đa hình thuộc trên đoạn gen HV-I và 14 đa hình trên đoạn HV-II (Hình 3). Các đa hình phát hiện trong nghiên cứu này đã được liệt kê chi tiết ở Bảng 1, trong đó có 18 đa hình đã phát hiện ở người Kinh trong các nghiên cứu trước đây. Từ các đa hình này chúng tôi chọn ra các đa hình phổ biến (có tần suất đa hình ở ít nhất 1 trong 2 dân tộc nghiên cứu từ 10% trở lên) để phục vụ các phân tích tiếp theo.



**Hình 3.** Biểu đồ thống kê đa hình trên 2 đoạn HV-I và HV-II của vùng D-loop phát hiện được ở 2 nhóm cá thể người Kinh và người Mảng khi so sánh với trình tự tham chiếu rCRS.

**Bảng 1.** Các vị trí đa hình phát hiện được tại 2 đoạn HV-I và HV-II ở quần thể nghiên cứu (gồm 51 người Kinh và 37 người Mảng).

Vị trí trên hệ ty thể	Vùng gen	rCRS	Đa hình	Số cá thể mang đa hình		Đối chiếu với dân tộc Kinh ở các nghiên cứu khác
				Dân tộc Kinh	Dân tộc Mảng	
63	HV-II	T	C	0	1	
64	HV-II	C	T	0	1	
66	HV-II	G	A	0	1	
73	HV-II	A	G	51	37	2 (100%) (Huỳnh Thị Thu Huệ <i>et al.</i> , 2005)
103	HV-II	G	A	1	0	
114	HV-II	C	T	1	0	
131	HV-II	T	C	1	0	
143	HV-II	G	A	1	0	
146	HV-II	T	C	5	0	2 (100%) (Huỳnh Thị Thu Huệ <i>et al.</i> , 2005)

146	HV-II	T	A	1	0	1 (50%) (Huỳnh Thị Thu Huệ <i>et al.</i> , 2005)
150	HV-II	C	T	14	15	Không xác định tần suất (Trần Thị Thủy Hằng, 2017)
151	HV-II	C	T	2	15	
152	HV-II	T	C	5	6	
153	HV-II	A	G	1	0	
185	HV-II	G	A	2	0	
189	HV-II	A	G	2	0	
195	HV-II	T	C	4	3	
199	HV-II	T	C	10	0	
200	HV-II	A	G	0	2	
204	HV-II	T	C	5	0	
207	HV-II	G	A	2	0	
210	HV-II	A	G	5	1	
215	HV-II	A	G	1	0	
226	HV-II	T	C	1	0	
234	HV-II	A	G	1	0	
235	HV-II	A	G	0	2	
249	HV-II	A	G	1	0	
255	HV-II	G	A	1	0	
263	HV-II	A	G	50	37	50 (100%) (Trần Thị Thủy Hằng, 2017)
279	HV-II	T	C	1	0	
310	HV-II	T	C	40	35	Không xác định tần suất (Nguyễn Đăng Tôn <i>et al.</i> , 2008)
318	HV-II	T	C	1	0	
384	HV-II	A	G	1	0	
456	HV-II	C	T	4	0	
463	HV-II	C	T	1	0	
479	HV-II	A	G	1	0	
489	HV-II	T	C	23	21	Không xác định tần suất (Nguyễn Đăng Tôn <i>et al.</i> , 2008)
499	HV-II	G	A	2	0	
504	HV-II	T	C	1	0	
507	HV-II	T	C	1	0	
523	HV-II	A	G	1	0	
574	HV-II	A	C	1	0	
16075	HV-I	T	C	1	0	
16086	HV-I	T	C	1	0	
16092	HV-I	T	C	3	0	
16093	HV-I	T	C	7	2	1 (50%) (Huỳnh Thị Thu Huệ <i>et al.</i> , 2005)
16108	HV-I	C	T	2	0	
16111	HV-I	C	T	1	0	
16129	HV-I	G	A	15	15	Không xác định tần suất (Trần Thị Thủy Hằng, 2017)
16136	HV-I	T	C	3	0	
16140	HV-I	T	C	9	1	
16147	HV-I	C	T	1	0	

16148	HV-I	C	T	1	0	
16157	HV-I	T	C	1	0	
16162	HV-I	A	G	1	11	
16166	HV-I	A	G	1	0	
16172	HV-I	T	C	6	11	Không xác định tần suất (Trần Thị Thúy Hằng, 2017)
16174	HV-I	C	T	1	0	
16175	HV-I	A	G	2	0	
16182	HV-I	A	C	9	0	1 (50%) (Huỳnh Thị Thu Huệ <i>et al.</i> , 2005)
16183	HV-I	A	C	24	1	1 (50%) (Huỳnh Thị Thu Huệ <i>et al.</i> , 2005)
16189	HV-I	T	C	29	1	1 (50%) (Huỳnh Thị Thu Huệ <i>et al.</i> , 2005)
16192	HV-I	C	T	2	0	
16209	HV-I	T	C	1	0	
16217	HV-I	T	C	6	0	1 (50%) (Huỳnh Thị Thu Huệ <i>et al.</i> , 2005)
16218	HV-I	C	T	2	0	
16222	HV-I	C	T	0	1	
16223	HV-I	C	T	20	23	Không xác định tần suất (Trần Thị Thúy Hằng, 2017)
16234	HV-I	C	T	2	1	
16243	HV-I	T	C	2	0	
16244	HV-I	G	A	0	1	
16249	HV-I	T	C	2	0	
16260	HV-I	C	T	1	0	
16261	HV-I	C	T	5	0	2 (100%) (Huỳnh Thị Thu Huệ <i>et al.</i> , 2005)
16263	HV-I	T	C	1	0	
16266	HV-I	C	A	5	1	
16269	HV-I	A	G	1	15	
16271	HV-I	T	C	2	15	
16273	HV-I	G	A	1	0	
16274	HV-I	G	A	1	0	
16278	HV-I	C	T	2	0	
16290	HV-I	C	T	1	2	
16293	HV-I	A	T	1	0	
16294	HV-I	C	T	1	0	
16295	HV-I	C	T	3	0	
16296	HV-I	C	T	1	0	
16297	HV-I	T	C	6	0	
16298	HV-I	T	C	5	0	
16300	HV-I	A	G	2	0	
16301	HV-I	C	G	1	0	
16304	HV-I	T	C	11	14	Không xác định tần suất (Trần Thị Thúy Hằng, 2017)
16305	HV-I	A	G	2	0	
16311	HV-I	T	C	7	12	
16319	HV-I	G	A	0	2	

16327	HV-I	C	T	4	0
16335	HV-I	A	G	1	0
16344	HV-I	C	T	0	1
16354	HV-I	C	T	2	0
16355	HV-I	C	T	1	0
16356	HV-I	T	C	1	0
16362	HV-I	T	C	7	10
16381	HV-I	T	C	0	2
16390	HV-I	G	A	3	0
16391	HV-I	G	A	1	0
16399	HV-I	A	G	1	0
16482	HV-I	A	G	1	0
16519	HV-I	T	C	38	15

2 (100%) (Huỳnh Thị Thu Huệ *et al.*, 2005)

**Phân tích thống kê so sánh đa hình giữa hai dân tộc Kinh và Mảng**

Trong các đa hình này, sau khi phân tích thống kê sử dụng kiểm định Fisher exact test, chúng tôi nhận thấy có 16 đa hình có tần suất khác nhau có ý nghĩa thống kê ở 2 nhóm dân tộc Kinh và Mảng ( $p < 0,05$ ) bao gồm: T146C/A, C151T và T199C thuộc đoạn gen HV-II và T16140C, A16162G, T16172C, A16282C, A16183C, T16189C, T16217C, C16223T, A16269G, T16271C, T16297C, và

T16519C thuộc đoạn gen HV-I (Bảng 2).

Trong số 16 đa hình này, có 8 đa hình chỉ phát hiện được ở nhóm người Kinh mà không thấy xuất hiện ở nhóm người Mảng (T146C, T199C, A16182C, T16217C, T16297C) hoặc xuất hiện rất ít với chỉ 1 cá thể đơn lẻ (T16140C, A16183C, T16189C). Ngược lại, có 4 đa hình phổ biến ở người Mảng nhưng lại có tần suất xuất hiện khá thấp ở người Kinh là C151T, A16162G, A16269G, và T16271C.

**Bảng 2.** Các đa hình phổ biến (frequency  $\geq 0,1$ ) có tần suất khác nhau giữa dân tộc Kinh và Mảng.

Đa hình	Dân tộc Kinh (Số cá thể = 51)	Dân tộc Mảng (Số cá thể = 37)	Giá trị p <sup>(*)</sup>
T146C	5 (9,8%)	0 (0%)	0,038
C151T	2 (3,9%)	15 (45,5%)	0,000017
T199C	10 (19,6%)	0 (0%)	0,004
T16140C	9 (17,6%)	1 (2,7%)	0,04
A16162G	1 (2%)	11 (29,7%)	0,00022
T16172C	6 (11,8%)	11 (29,7%)	0,035
A16182C	9 (17,6%)	0 (0%)	0,006
A16183C	24 (47,1%)	1 (2,7%)	0,000002
T16189C	29 (56,9%)	1 (2,7%)	0,0000004
T16217C	6 (11,8%)	0 (0%)	0,038
C16223T	20 (39,2%)	23 (62,2%)	0,034
A16269G	1 (2%)	15 (40,5%)	0,000003
T16271C	2 (3,9%)	15 (40,5%)	0,00002
T16297C	6 (11,8%)	0 (0%)	0,038
T16519C	38 (74,5%)	15 (40,5%)	0,002

Ghi chú: (\*) Giá trị p được tính bằng kiểm định Fisher exact 2 phía.

## THẢO LUẬN

Đa hình T146C và T16297C đã được phát hiện ở một số dân tộc nhánh ngữ hệ Kravet (thuộc ngữ hệ Bahnar) tại Thái Lan và Campuchia (Zhang *et al.*, 2013). Trong nghiên cứu này chúng tôi đã phát hiện 5/51 (chiếm 9,8%) và 6/51 (chiếm 11,8%) cá thể người Kinh mang đa hình T146C và T16297C, tuy nhiên không có cá thể người Mảng nào mang các đa hình trên.

Đa hình T199C được tìm thấy ở 10/51 người dân tộc Kinh (19,6%) nhưng không thấy xuất hiện ở các cá thể người dân tộc Mảng. Đa hình này cũng xuất hiện ở quần thể người Trung Quốc trên đảo Hải Nam (Yang *et al.*, 2011). Các đa hình T146C, T14640C, A16183C, T16189C và T16217C được phát hiện rất phổ biến tại trong hệ gen ty thể người Hán miền Nam Trung Quốc (Fang *et al.*, 2014) cũng được phát hiện với tần số tương đối lớn ở quần thể dân tộc Kinh nhưng rất ít xuất hiện ở dân tộc Mảng.

Một số đa hình khác chỉ xuất hiện tại nhóm người Kinh nhưng không xuất hiện hoặc xuất hiện rất ít chỉ ở 1 cá thể đơn lẻ trong nhóm người Mảng tại nghiên cứu này như: A16182C, T16140C, T16189C. Các công trình nghiên cứu cũng đã chỉ ra các đa hình này đã xuất hiện ở các quần thể lân cận như: A16182C ở quần thể người Indonesia (Malik *et al.*, 2002), T16140C ở quần thể người Đài Loan (Liou *et al.*, 2012), T16189C ở các quần thể người Tunisia và vùng Địa Trung Hải (Hsoune *et al.*, 2016; Mueller *et al.*, 2011; Palmieri *et al.*, 2011). Tuy nhiên, đối với 4 điểm đa hình xuất hiện phổ biến ở người Mảng mà ít xuất hiện ở người Kinh (C151T, A16162G, A16269G, T16271C), chỉ có A16162G là được phát hiện ở quần thể người Thái Lan (Sangthong *et al.*, 2014).

Các bằng chứng trên cho thấy mặc dù cùng thuộc nhóm ngữ hệ Nam Á và có nhiều đặc điểm về văn hóa, phong tục tập quán giống nhau, giữa người Mảng và người Kinh vẫn có nhiều điểm khác biệt về mặt di truyền. Trong nghiên cứu này, hệ gen ty thể, đặc biệt là vùng D-loop ở người Kinh có chứa nhiều đa hình được phát hiện ở nhiều quần thể khác trên thế giới (Trung Quốc, Thái Lan, Campuchia, Đài Loan, Indonesia...) thì ở người Mảng các đa hình này đều hiếm hoặc hầu như không xuất hiện. Có thể giải thích hiện tượng này do lịch sử hình thành và phát triển của các dân tộc. Trong khi người Kinh từ sớm đã có nền văn hóa giao lưu rộng rãi với các dân tộc khác, với khu vực sinh sống trải rộng và qua đó kết hôn, chia sẻ vật chất di truyền

với các dân tộc khác trong khu vực thì người Mảng chỉ tập trung sinh sống chủ yếu tại tỉnh Lai Châu và các tỉnh vùng cao Tây Bắc và ít khi kết hôn với người ngoại tộc (Ban chỉ đạo Tổng điều tra dân số và nhà ở Trung ương, 2009).

Tuy với cỡ mẫu còn khá hạn chế, nghiên cứu này đã bước đầu làm sáng tỏ một số điểm khác biệt trong đa dạng di truyền DNA ty thể thuộc vùng gen D-loop giữa người Kinh và người Mảng. Theo hiểu biết của chúng tôi, đây là công trình đầu tiên khảo sát về dữ liệu gen ty thể ở người dân tộc Mảng. Chúng tôi sẽ tiếp tục mở rộng cỡ mẫu để khẳng định chắc chắn hơn nữa về sự đa dạng di truyền trong vùng gen D-loop giữa dân tộc Kinh, dân tộc Mảng cũng như các dân tộc khác. Ngoài ra, nghiên cứu sẽ tiến hành khảo sát thêm các vùng gen khác thuộc ty thể nhằm khảo sát sự đa hình toàn bộ hệ genome ty thể ở các dân tộc người Việt Nam.

## KẾT LUẬN

Bằng phương pháp giải trình tự DNA, nghiên cứu này đã khảo sát các đa hình 2 đoạn siêu biến HV-I và HV-II thuộc hệ gen ty thể của 2 nhóm người dân tộc Kinh và dân tộc Mảng tại Việt Nam. Qua đó, tìm thấy rất nhiều đa hình lần đầu tiên phát hiện được ở cả 2 dân tộc Kinh và Mảng. Các phân tích thống kê cũng chỉ ra 16 điểm khác biệt có ý nghĩa thống kê trong vùng D-loop giữa các cá thể thuộc 2 dân tộc này. Các kết quả bước đầu này có ý nghĩa trong việc nghiên cứu sự đa dạng di truyền quần thể cũng như lịch sử hình thành và phát triển của các dân tộc Việt Nam nói chung và dân tộc Mảng nói riêng.

**Lời cảm ơn:** Công trình được hoàn thành với sự tài trợ từ đề tài thuộc Bộ Khoa học Công nghệ mã số DTĐL.CN-05/15 và đề tài cấp cơ sở thuộc Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Chúng tôi cũng xin chân thành cảm ơn những người đã tham gia hiến mẫu cho nghiên cứu này.

## TÀI LIỆU THAM KHẢO

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23(2): 147.

Ban chỉ đạo Tổng điều tra dân số và nhà ở Trung Ương (2009) Ban chỉ đạo Tổng điều tra dân số và nhà ở Trung



Ương. *Tổng điều tra dân số và nhà ở Việt Nam - Kết quả toàn bộ*: 3-5.

Chemale G, Paneto GG, Menezes MA, Freitas JM, Jacques GS, Cicarelli RM, Fagundes PR (2013) Development and validation of a D-loop mtDNA SNP assay for the screening of specimens in forensic casework. *Forensic Sci Int Genet* 7(3): 353-358.

Fang H, Liu X, Shen L, Li F, Liu Y, Chi H, Miao H, Lu J, Bai Y (2014) Role of mtDNA haplogroups in the prevalence of knee osteoarthritis in a southern Chinese population. *Int J Mol Sci* 15(2): 2646-2659.

Hsouna S, Ben Halim N, Lasram K, Meiloud G, Arfa I, Kerkeni E, Romdhane L, Jamoussi H, Bahri S, Ben Ammar S, Abid A, Barakat A, Houmeida A, Abdelhak S, Kefi R (2016) Study of the T16189C variant and mitochondrial lineages in Tunisian and overall Mediterranean region. *Mitochondrial DNA A DNA Mapp Seq Anal* 27(2): 1558-1563.

Hung DM, Ha NH, Khoi PN, Nhung VP, Phong NV, Duong NT, Hai NV, Ton ND (2016) Genetic variation of mitochondrial sequence-hv2 in Vietnamese populations. *Tạp chí Sinh học* 38(2): 243-249.

Huỳnh Thị Thu Huệ, Hoàng Thị Thu Yến, Nguyễn Đăng Tôn, Lê Thị Thu Hiền, Nguyễn Đình Cường, Phan Văn Chi, Nông Văn Hải (2005) Phân tích trình tự vùng điều khiển (D-loop) trên genome ty thể của 5 cá thể người Việt Nam. *Tạp chí Công nghệ Sinh học* 3(1): 15-22.

Kawabe K, Worawut R, Taura S, Shimogiri T, Nishida T, Okamoto S (2014) Genetic Diversity of mtDNA D-loop Polymorphisms in Laotian Native Fowl Populations. *Asian-Australas J Anim Sci* 27(1): 19-23.

Liou CW, Chen JB, Tiao MM, Weng SW, Huang TL, Chuang JH, Chen SD, Chuang YC, Lee WC, Lin TK, Wang PW (2012) Mitochondrial DNA coding and control region variants as genetic risk factors for type 2 diabetes. *Diabetes* 61(10): 2642-2651.

Malik S, Sudoyo H, Pramoonjago P, Suryadi H, Sukarna T, Njunting M, Sahiratmadja E, Marzuki S (2002) Nuclear mitochondrial interplay in the modulation of the homopolymeric tract length heteroplasmy in the control (D-loop) region of the mitochondrial DNA. *Hum Genet* 110(5): 402-411.

Mueller EE, Eder W, Ebner S, Schwaiger E, Santic D, Kreindl T, Stanger O, Paulweber B, Iglseider B, Oberkofler H, Maier R, Mayr JA, Krempler F, Weitgasser R, Patsch W, Sperl W, Kofler B (2011) The mitochondrial T16189C

polymorphism is associated with coronary artery disease in Middle European populations *PLoS One*, 6(1), e16455.

Nguyễn Đăng Tôn, Nguyễn Thị Tú Linh, Vũ Hải Chi, Trần Thị Ngọc Diệp, Địch Thị Kim Hương, Bùi Thị Tuyết, Nguyễn Hải Hà, Huỳnh Thị Thu Huệ, Lê Thị Thu Hiền, Trần Thị Phương Liên, Phan Văn Chi, Nông Văn Hải (2008) Đa hình đơn bội DNA ty thể của các cá thể người Việt Nam. *Tạp chí Công nghệ Sinh học* 6(4): 579-590.

Palmieri VO, De Rasmio D, Signorile A, Sardanelli AM, Grattagliano I, Minerva F, Cardinale G, Portincasa P, Papa S, Palasciano G (2011) T16189C mitochondrial DNA variant is associated with metabolic syndrome in Caucasian subjects. *Nutrition* 27(7-8): 773-777.

Pischedda S, Barral-Arca R, Gomez-Carballa A, Pardo-Seco J, Catelli ML, Alvarez-Iglesias V, Cardenas JM, Nguyen ND, Ha HH, Le AT, Martinon-Torres F, Vullo C, Salas A (2017) Phylogeographic and genome-wide investigations of Vietnam ethnic groups reveal signatures of complex historical demographic movements. *Sci Rep* 7(1): 12630.

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

Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta* 1410(2): 103-123.

Trần Thị Thủy Hằng, Trần Huy Thịnh, Trần Văn Khánh (2017) Đa hình thái đơn nucleotid vùng gen ty thể HV1 và HV2 trên người dân tộc Kinh và dân tộc Mường của Việt Nam. *Tạp chí Nghiên cứu Y học* 106(1): 33-40.

Xu K, Hu S (2015) Population data of mitochondrial DNA HVS-I and HVS-II sequences for 208 Henan Han Chinese. *Leg Med (Tokyo)* 17(4): 287-294.

Yang K, Zheng H, Qin Z, Lu Y, Farina SE, Li S, Jin L, Li D, Li H (2011) Positive selection on mitochondrial M7 lineages among the Gelong people in Hainan. *J Hum Genet* 56(3): 253-256.

Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70(3): 635-651.

Zhang X, Qi X, Yang Z, Serey B, Sovannary T, Bunnath L, Seang Aun H, Samnom H, Zhang H, Lin Q, van Oven M, Shi H, Su B (2013) Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. *Nat Commun* 4: 2599.

## SINGLE NUCLEOTIDE POLYMORPHISMS IN THE D-LOOP REGION OF THE MITOCHONDRIAL GENOMES OF INDIVIDUALS FROM TWO ETHNIC GROUPS KINH AND MANG OF AUSTRO-ASIATIC LANGUAGE FAMILY

Nguyen Thy Ngọc<sup>1,2</sup>, Nguyen Bao Trang<sup>1,3</sup>, Nguyen Quang Huy<sup>4</sup>, Nguyen Dang Ton<sup>1,5</sup>, Nguyen Thuy Duong<sup>1,5</sup>

<sup>1</sup>*Institute of Genome Research, Vietnam Academy of Science and Technology*

<sup>2</sup>*University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology*

<sup>3</sup>*National Office of Intellectual Property of Vietnam, Ministry of Science and Technology*

<sup>4</sup>*University of Science, Vietnam National University, Hanoi*

<sup>5</sup>*Graduate University of Science and Technology, Vietnam Academy of Science and Technology*

### SUMMARY

The two hypervariable regions HV-I and HV-II in the D-loop region are the most diverse region in the mitochondrial human genome. These DNA regions have an important role in population genetics and human evolution research. In this study, we identified single nucleotide polymorphisms (SNPs) in the D-loop region of the human mitochondrial DNA genomes of individuals in two Vietnamese ethnic groups: Kinh and Mang. 81 blood samples from 50 Kinh and 31 Mang unrelated individuals were collected for genomic DNA extraction. Using PCR and specific primers, we successfully amplified the two HV-I and HV-II regions with respective sizes of 693 bp and 689 bp. The PCR products then were purified and sequenced. After alignment to the reference mitochondrial genome rCRS (NC\_012920.1), 96 SNPs were detected in the Kinh ethnic group, of which many polymorphisms were firstly found in this ethnic and 36 SNPs in the Mang ethnic group. Of the identified 132 SNPs, 16 SNPs were significantly different between the two ethnic groups ( $p < 0.05$ ). Eight out of the 16 SNPs (T146C, T199C, A16182C, T16217C, T16297C, T16140C, A16183C, T16189C) were frequently found in the Kinh ethnic group but not or only rarely in the Mang ethnic group. On the other hand, 4 SNPs (C151T, A16162G, A16269G, T16271C) were commonly found in the Mang population but rarely in the Kinh population. These findings suggest that although Kinh and Mang belong to one language family (Austro-Asiatic), there is a noticeable diversity in the mitochondrial genomes between these two Vietnamese ethnic groups.

**Keywords:** *mitochondrial genome, D-loop, HV-I, HV-II, Kinh, Mang*