

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**BỘ Y TẾ**

**TRƯỜNG ĐẠI HỌC Y HÀ NỘI**



**TRẦN THỊ THÚY HẰNG**

**ĐÁNH GIÁ TÍNH ĐA HÌNH VÙNG HV1, HV2  
TRÊN DNA TY THỂ Ở MỘT SỐ DÂN TỘC VÀ  
BỆNH NHÂN UNG THƯ VÚ NGƯỜI VIỆT NAM**

**LUẬN ÁN TIẾN SĨ Y HỌC**

**HÀ NỘI - 2019**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**BỘ Y TẾ**

**TRƯỜNG ĐẠI HỌC Y HÀ NỘI**



**TRẦN THỊ THÚY HẰNG**

**ĐÁNH GIÁ TÍNH ĐA HÌNH VÙNG HV1, HV2  
TRÊN DNA TY THỂ Ở MỘT SỐ DÂN TỘC VÀ  
BỆNH NHÂN UNG THƯ VÚ NGƯỜI VIỆT NAM**

Chuyên ngành: Hóa Sinh Y học

Mã số: 62720112

**LUẬN ÁN TIẾN SĨ Y HỌC**

Người hướng dẫn khoa học:

PGS.TS Trần Văn Khánh

**HÀ NỘI – 2019**

## LỜI CẢM ƠN

Lời đầu tiên, cho phép tôi được bày tỏ sự kính trọng và lòng biết ơn sâu sắc với **PGS.TS. Trần Văn Khánh** - Trưởng Bộ môn Bệnh học phân tử, Phó Giám đốc Trung tâm nghiên cứu Gen và Protein, Trường Đại học Y Hà Nội người đã tận tình hướng dẫn, động viên giúp đỡ và tạo điều kiện thuận lợi cho tôi trong suốt quá trình nghiên cứu và viết luận án này. Những hướng dẫn, nhận xét và góp ý của Cô, đặc biệt là những gợi ý về hướng giải quyết vấn đề trong suốt quá trình nghiên cứu, thực sự là những bài học vô cùng quý giá đối với tôi không chỉ trong quá trình học tập nghiên cứu, viết luận án mà còn cả trong hoạt động chuyên môn sau này.

Tôi cũng xin gửi lời cảm ơn đến **GS.TS. Tạ Thành Văn** - Hiệu trưởng, Trưởng Bộ môn Hóa Sinh, Giám đốc Trung tâm nghiên cứu Gen và Protein, Trường Đại học Y Hà Nội, Thầy đã giúp đỡ, tạo mọi điều kiện tốt nhất và cho tôi những chỉ dẫn để tôi có thể hoàn thành tốt luận án của mình.

Tôi xin được gửi lời cảm ơn tới các Thầy, Cô trong Bộ môn Hóa Sinh, Trung tâm nghiên cứu Gen và Protein, Trường Đại Học Y Hà Nội cùng các anh, các chị, các em nghiên cứu viên, các bạn học viên của Bộ môn và Trung tâm đã giúp đỡ, góp ý, cho tôi những ý kiến quý báu, tạo mọi điều kiện tốt nhất cho tôi trong quá trình nghiên cứu và viết luận án của mình.

Tôi xin cảm ơn PGS.TS Nguyễn Đức Hình - chủ nhiệm đề tài cấp Nhà nước “Đánh giá đặc điểm di truyền người Việt Nam” thuộc đề tài nhiệm vụ Quỹ gen đã hỗ trợ kinh phí để tôi có thể hoàn thành luận án này.

Cuối cùng, tôi xin được bày tỏ lòng biết ơn sâu sắc tới bố, mẹ, chồng, con, gia đình và bạn bè đã luôn bên cạnh động viên, hỗ trợ và tạo mọi điều kiện thuận lợi để tôi có thể hoàn thành luận án này.

Hà Nội, ngày tháng năm 2019

**Trần Thị Thúy Hằng**

## **LỜI CAM ĐOAN**

Tôi là Trần Thị Thúy Hằng, nghiên cứu sinh khóa 34, Trường Đại học Y Hà Nội, chuyên ngành Hóa Sinh, xin cam đoan:

1. Đây là luận án do bản thân tôi trực tiếp thực hiện cùng với sự tham gia của một số đồng nghiệp dưới sự hướng dẫn của PGS.TS. Trần Văn Khánh.
2. Công trình này không trùng lặp với bất kỳ nghiên cứu nào khác đã được công bố tại Việt Nam.
3. Các số liệu và thông tin trong nghiên cứu là hoàn toàn chính xác, trung thực và khách quan, đã được xác nhận và chấp thuận của cơ sở nơi nghiên cứu.

Tôi xin hoàn toàn chịu trách nhiệm trước pháp luật về những cam kết này.

*Hà Nội, ngày tháng năm 2019*

**Tác giả luận án**

**Trần Thị Thúy Hằng**

## DANH MỤC CHỮ VIẾT TẮT

ATP	: Adenosine triphosphat
Bp	: Base pair (cặp bazơ)
ddNTP	: Dideoxyribonucleotid triphosphat
D-loop	: Displacement loop
DNA	: Deoxyribonucleic acid
dNTP	: Deoxyribonucleotid triphosphat
EDTA	: Ethylene Diamine Tetraacetic Acid
HV	: Hypervariable Region (vùng siêu biến)
kb	: Kilo base
MITOMAP	: A human Mitochondrial Genome Database
mtDNA	: mitochondrial DNA (DNA ty thể)
NCBI	: <i>National Center for Biotechnology Information</i> (Trung tâm quốc gia về thông tin công nghệ sinh học)
NST	: Nhiễm sắc thể
OXPHOS	: Oxidative phosphorylation
PCR	: Polymerase chain reaction
RFLP	: Restriction Fragment Length Polymorphism
RNA	: Ribonucleic acid
ROS	: Reactive oxygen speccies
SNP	: Single Nucleotide Polymorphisms

# MỤC LỤC

<b>ĐẶT VẤN ĐỀ .....</b>	<b>1</b>
<b>Chương 1: TỔNG QUAN TÀI LIỆU .....</b>	<b>3</b>
1.1. DNA ty thể .....	3
1.1.1. Ty thể.....	3
1.1.2. Cấu trúc DNA ty thể.....	3
1.1.3. Vùng HV1 và HV2 trên DNA ty thể.....	4
1.2. Đặc điểm di truyền DNA ty thể .....	7
1.3. Nghiên cứu tính đa hình đơn nucleotid .....	13
1.3.1. Đa hình đơn nucleotid .....	13
1.3.2. Đa hình trên vùng HV1 và HV2 của DNA ty thể .....	16
1.4. Sơ lược về bệnh ung thư vú.....	18
1.5. Đa hình gen ty thể và mối liên quan đến bệnh.....	19
1.6. Đa hình gen ty thể và bệnh ung thư vú.....	21
1.7. Một số đặc điểm dân tộc của người Việt Nam.....	27
1.7.1. Dân tộc Kinh .....	28
1.7.2. Dân tộc Mường.....	28
1.7.3. Dân tộc Chăm.....	29
1.7.4. Dân tộc Khmer .....	30
1.8. Tình hình nghiên cứu về DNA ty thể người Việt Nam .....	31
1.9. Một số phương pháp phát hiện đa hình thái gen ty thể.....	33
1.9.1. Kỹ thuật PCR .....	33
1.9.2. Kỹ thuật PCR - RFLP .....	34
1.9.3. Kỹ thuật giải trình tự gen .....	35
<b>Chương 2: ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU.....</b>	<b>40</b>
2.1. Đối tượng nghiên cứu .....	40
2.2. Thời gian và địa điểm nghiên cứu.....	41

2.3. Phương pháp nghiên cứu .....	41
2.4. Phương tiện nghiên cứu .....	41
2.4.1. Dụng cụ, trang thiết bị .....	41
2.4.2. Hoá chất .....	42
2.5. Kỹ thuật nghiên cứu .....	43
2.5.1. Tách chiết DNA từ máu ngoại vi .....	43
2.5.2. Phương pháp quang phổ .....	44
2.5.3. Điện di DNA sau tách chiết .....	45
2.5.4. Phản ứng PCR nhân đoạn gen HV1, HV2 .....	46
2.5.5. Giải trình tự vùng HV1 và HV2 .....	47
2.5.6. Phương pháp phân tích trình tự đoạn HV1 và HV2 .....	50
2.5.7. Phân tích mối liên quan giữa một số đa hình đơn nucleotid trên vùng HV1 với bệnh ung thư vú .....	50
2.6. Vấn đề đạo đức của đề tài .....	51
2.7. Sơ đồ nghiên cứu .....	52
<b>Chương 3: KẾT QUẢ NGHIÊN CỨU .....</b>	<b>53</b>
3.1. Kết quả giải trình tự gen vùng HV1, HV2 trên DNA ty thể .....	53
3.1.1. Tách chiết DNA tổng số .....	53
3.1.2. Kết quả khuếch đại đoạn gen HV1, HV2 của DNA ty thể .....	54
3.1.3. Kết quả giải trình tự vùng HV1, HV2 của DNA ty thể trên 4 dân tộc Kinh, Chăm, Mường, Khmer người Việt Nam .....	55
3.1.4. Kết quả giải trình tự vùng HV1 trên DNA ty thể ở bệnh nhân ung thư vú người Việt Nam .....	58
3.2. Kết quả phân tích đa hình vùng HV1 và HV2 trên DNA ty thể .....	60
3.2.1. Đa hình vùng HV1, HV2 trên DNA ty thể ở 4 dân tộc người Việt Nam (Kinh, Chăm, Khmer và Mường) được đối chiếu với trình tự chuẩn .....	60
3.2.2. Đa hình mới được phát hiện trên vùng HV1 và HV2 của DNA ty thể người Việt Nam .....	68

3.2.3. Các vị trí đa hình thường gặp trong các mẫu nghiên cứu .....	70
3.2.4. Tổng số đa hình trong các mẫu nghiên cứu .....	71
3.3. Phân nhóm SNP đặc trưng vùng HV1 và HV2 (phân chia các nhóm đơn bội mtDNA theo bộ SNP đặc trưng trên vùng HV1, HV2) ở 4 dân tộc Kinh, Chăm, Mường, Khmer Việt Nam.....	71
3.4. Tỷ lệ một số SNP trên vùng HV1 của DNA ty thể ở nhóm bệnh nhân bị ung thư vú và nhóm nữ bình thường .....	85
<b>Chương 4: BÀN LUẬN.....</b>	<b>90</b>
4.1. Đặc điểm mẫu nghiên cứu.....	90
4.2. Phân tích tính đa hình vùng HV1 và HV2 của DNA ty thể trên một số dân tộc người Việt Nam bằng phương pháp giải trình tự gen.....	91
4.3. Đánh giá tính đa hình vùng HV1 của mtDNA ở bệnh nhân ung thư vú	110
<b>KẾT LUẬN .....</b>	<b>119</b>
<b>DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ LIÊN QUAN ĐẾN LUẬN ÁN.....</b>	<b>121</b>
<b>TÀI LIỆU THAM KHẢO</b>	
<b>PHỤ LỤC</b>	



## DANH MỤC BẢNG

Bảng 1.1:	Phân chia các nhóm đơn bội DNA ty thể dựa vào vị trí đa hình đặc trưng trên vùng HV1 và HV2.....	12
Bảng 3.1:	Các dạng SNP trên vùng HV1 và HV2 của DNA ty thể ở 517 mẫu thuộc 4 dân tộc Kinh, Mường, Khmer, Chăm Việt Nam ...	62
Bảng 3.2:	Bảng các vị trí trên vùng HV1 có nhiều hơn một loại đa hình...	64
Bảng 3.3:	Bảng các vị trí trên vùng HV2 có nhiều hơn một loại đa hình...	65
Bảng 3.4:	Các dạng SNP trên vùng HV1 của DNA ty thể chỉ gặp ở 1 trong 4 dân tộc Kinh, Mường, Khmer, Chăm .....	66
Bảng 3.5:	Các dạng SNP trên vùng HV2 của DNA ty thể chỉ gặp ở 1 trong 4 dân tộc Kinh, Mường, Khmer, Chăm Việt Nam .....	67
Bảng 3.6:	Phân chia nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên vùng HV1, HV2 và các dạng SNP trên vùng HV1, HV2 của một số mẫu nghiên cứu đại diện cho dân tộc Chăm Việt Nam. .	72
Bảng 3.7:	Phân chia nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên vùng HV1, HV2 và các dạng SNP trên vùng HV1, HV2 của một số mẫu nghiên cứu đại diện cho dân tộc Kinh Việt Nam.....	74
Bảng 3.8:	Phân chia nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên vùng HV1, HV2 và các dạng SNP trên vùng HV1, HV2 của một số mẫu nghiên cứu đại diện cho dân tộc Khmer Việt Nam.....	76
Bảng 3.9:	Phân chia nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên vùng HV1, HV2 và các dạng SNP trên vùng HV1, HV2 của một số mẫu nghiên cứu dân tộc Mường Việt Nam.....	78
Bảng 3.10:	Số lượng haplotypes HV1/HV2 mtDNA của 517 mẫu nghiên cứu thuộc 4 dân tộc Kinh, Chăm, Mường, Khmer Việt Nam .....	80

Bảng 3.11: Bảng tần suất theo các nhóm đơn bội .....	81
Bảng 3.12: Bảng một số vị trí đa hình thường gặp trên vùng HV1 và HV2.	70
Bảng 3.13. Bảng tỷ lệ một số đa hình hay gặp trên vùng HV1 của mtDNA ở bệnh nhân ung thư vú.....	85
Bảng 3.14. Tỷ lệ một số SNP trên vùng HV1 mtDNA của nhóm ung thư vú và nhóm nữ bình thường.....	87
Bảng 4.1: Bảng tỷ lệ một số nhóm đơn bội mtDNA phổ biến ở Việt Nam và một số nước ở châu Á .....	98
Bảng 4.2: So sánh tỷ lệ một số đa hình trên vùng HV1 của mtDNA với một số nghiên cứu khác.....	102
Bảng 4.3: So sánh tỷ lệ một số đa hình trên vùng HV2 của mtDNA với một số nghiên cứu khác.....	103

## DANH MỤC BIỂU ĐỒ

- Biểu đồ 3.1: Biểu đồ biểu thị tỷ lệ các nhóm đơn bội mtDNA phổ biến của 4 dân tộc Kinh, Mường, Chăm và Khmer ..... 83
- Biểu đồ 3.2: Biểu đồ tỷ lệ các nhóm đơn bội chiếm tỷ lệ cao theo từng dân tộc của 4 dân tộc Kinh, Mường, Chăm, Khmer người Việt Nam..... 84
- Biểu đồ 3.3. Tỷ lệ một số SNP trên vùng HV1 mtDNA của 2 nhóm ..... 86
- Biểu đồ 3.4: Biểu thị mối tương quan giữa SNP trên vùng HV1 của DNA ty thể với bệnh ung thư vú. .... 89

## DANH MỤC HÌNH

Hình 1.1:	Sơ đồ cấu tạo chi tiết của DNA ty thể người .....	4
Hình 1.2:	Hình trình bày vị trí hai vùng HV1 và HV2 trên DNA ty thể.....	5
Hình 1.3:	Sơ đồ mô tả quá trình di cư của các nhóm đơn bội DNA ty thể người.	10
Hình 1.4:	Sơ đồ phân nhóm đơn bội mtDNA theo các vị trí đa hình đặc trưng .....	11
Hình 1.5:	Mô phỏng hiện tượng đa hình đơn nucleotid .....	15
Hình 1.6:	Các biến đổi đồng hoán và dị hoán trên mtDNA .....	17
Hình 1.7:	Phát hiện đột biến C16147T trong khối u vú .....	24
Hình 1.8:	Hình ảnh đa hình 16290insT và 16293delA vùng HV1 mtDNA trên bệnh nhân ung thư vú người Bangladesh .....	25
Hình 1.9:	Kết quả giải trình tự được so sánh với trình tự DNA ty thể chuẩn của người .....	33
Hình 1.10:	Quy trình giải trình tự theo phương pháp ddNTP .....	37
Hình 1.11:	Hình ảnh giải trình tự một số SNP vùng gen ty thể HV1, HV2	38
Hình 3.1:	Ảnh điện di DNA tổng số trên gel agarose 1,5% .....	53
Hình 3.2:	Hình ảnh điện di sản phẩm PCR của vùng HV1 .....	54
Hình 3.3:	Hình ảnh điện di sản phẩm PCR của vùng HV2 .....	54
Hình 3.4:	Hình ảnh giải trình tự SNP C16260T và SNP T16298C trên vùng HV1 của DNA ty thể .....	55
Hình 3.5:	Hình ảnh giải trình tự SNP T16189C, G16213A và SNP T16217C trên vùng HV1 của DNA ty thể.....	56
Hình 3.6:	Hình ảnh giải trình tự SNP T16311C trên vùng HV1 của mtDNA	56
Hình 3.7:	Hình ảnh giải trình tự SNP T152C trên vùng HV2 của mtDNA....	57
Hình 3.8:	Hình ảnh giải trình tự SNP (249DelA, A263G) trên vùng HV2 của DNA ty thể .....	57

Hình 3.9:	Hình ảnh giải trình tự SNP (A263G, 309insC, 315insC) trên vùng HV2 của DNA ty thể .....	58
Hình 3.10:	Hình trình bày toàn bộ các vị trí và loại đa hình trên vùng HV1 và HV2 của DNA ty thể ở bốn dân tộc Kinh, Chăm, Khmer, Mường người Việt Nam được đối chiếu với trình tự chuẩn.....	61
Hình 3.11:	Hình ảnh giải trình tự SNP (16038DelA) của vùng HV1.....	68
Hình 3.12:	Hình ảnh giải trình tự SNP mới (G16084C) trên vùng HV1 của DNA ty thể .....	69
Hình 3.13:	Hình ảnh giải trình tự SNP (A16515C) trên vùng HV1 mtDNA....	69

## ĐẶT VẤN ĐỀ

Trong những năm qua, sự phát triển mạnh mẽ của sinh học phân tử và công nghệ gen đã đem lại nhiều thành tựu khoa học. Một trong những thành tựu quan trọng đó là việc giải trình tự hoàn chỉnh hệ gen người. Công trình này đã mở ra những triển vọng hết sức to lớn đối với lĩnh vực y học. Các nhà khoa học đã tìm được bản chất của hàng ngàn gen có liên quan đến bệnh tật, đưa ra được các phương pháp chẩn đoán, điều trị mới, nhanh và có hiệu quả. Hệ gen người gồm có hai phần: hệ gen nhân (hệ gen nhiễm sắc thể) và hệ gen tế bào chất (hệ gen ty thể). Hệ gen trong nhân có kích thước lớn khoảng 3,2 tỷ bp, trong khi đó hệ gen ty thể có kích thước 16569 bp, nhỏ hơn hệ gen trong nhân rất nhiều lần. Tuy nhiên, hệ gen ty thể với các đặc điểm di truyền theo dòng mẹ, số lượng bản sao lớn và không tái tổ hợp nên việc nghiên cứu hệ gen ty thể không những có ý nghĩa quan trọng trong chẩn đoán, điều trị các bệnh di truyền ty thể mà còn có ý nghĩa trong nghiên cứu mối quan hệ di truyền, tiến hóa của quần thể người.

Vùng HV1, HV2 (Hypervariable region 1, 2) là đoạn DNA nằm trong vùng điều khiển của DNA ty thể. Đây là vùng có tần số đột biến cao nhất trong hệ gen ty thể người [1]. Hiện nay, người ta đã thống kê được trên 150 bệnh di truyền khác nhau do DNA ty thể quyết định [2]. Do đó, đã có nhiều nghiên cứu tập trung tìm hiểu về mối liên hệ của nó với các loại bệnh như bệnh di truyền, bệnh về cơ, bệnh thần kinh, bệnh chuyển hóa, lão hóa, bệnh ung thư trong đó có ung thư vú. Các nghiên cứu gần đây đã xác định được nhiều biến đổi của DNA ty thể có liên quan với bệnh ung thư vú, bao gồm những thay đổi về số lượng bản sao [3] [4], biến đổi mức độ biểu hiện và hoạt động của các tiểu đơn vị của chuỗi hô hấp và các đột biến điểm của DNA ty thể [5]. Tuy nhiên, các kết quả nghiên cứu vẫn còn gây nhiều

tranh cãi và có nhiều điều chưa được làm sáng tỏ. Các bệnh do rối loạn DNA ty thể thường được biểu hiện đa dạng liên quan đến rối loạn quá trình tổng hợp protein, có thể đơn thuần chỉ là các biểu hiện của sự đột biến điểm các nucleotid hoặc có thể liên quan đến các đa hình đơn nucleotid (Single Nucleotid Polymorphisms - SNP). Như vậy, tính đa hình/ đột biến của DNA ty thể có liên quan đến nhiều loại bệnh tật khác nhau, các kết quả nghiên cứu về tính đa hình/đột biến của DNA ty thể là cơ sở khoa học cần thiết cho những nghiên cứu về các bệnh lý DNA ty thể.

DNA ty thể với những đặc điểm di truyền ưu thế của mình đã nhanh chóng trở thành đối tượng được nghiên cứu rộng rãi trong nhiều lĩnh vực đặc biệt là lĩnh vực y học, sinh học phân tử và di truyền học... [6] [7]. Trong đó, vùng HV1 và HV2 với tốc độ tiến hóa nhanh, nhiều điểm đa hình, nhiều loại đột biến, nên các thông tin về trình tự, đa hình của vùng này được quan tâm nghiên cứu nhiều nhất. Vì vậy, chúng tôi tiến hành nghiên cứu đề tài: **“Đánh giá tính đa hình vùng HV1, HV2 trên DNA ty thể ở một số dân tộc và bệnh nhân ung thư vú người Việt Nam”**. Với ba mục tiêu chính:

- 1. Xác định tỷ lệ một số SNP (Single Nucleotid Polymorphisms) vùng HV1, HV2 trên DNA ty thể ở 4 dân tộc Kinh, Chăm, Mường, Khmer người Việt Nam.**
- 2. Phân nhóm SNP đặc trưng vùng HV1, HV2 của DNA ty thể trên 4 dân tộc Kinh, Chăm, Mường, Khmer người Việt Nam.**
- 3. Đánh giá một số SNP vùng HV1 của DNA ty thể trên bệnh nhân ung thư vú.**

## **Chương 1**

### **TỔNG QUAN TÀI LIỆU**

#### **1.1. DNA ty thể**

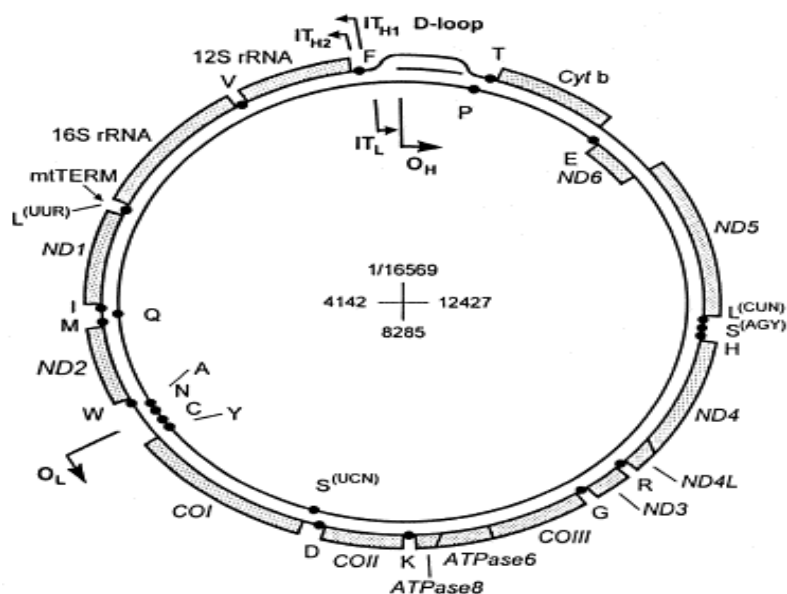
##### ***1.1.1. Ty thể***

Ty thể là bào quan có hình trụ dài, toàn bộ cấu trúc của ty thể được bao bọc bởi hai lớp màng cấu tạo bởi protein và lớp phospholipid kép. Không gian bên trong chứa chất nền, DNA ty thể, ribosom... Ty thể là trung tâm hô hấp của tế bào, là nơi sản sinh ra ATP, cung cấp năng lượng cho tế bào. Ty thể có hệ gen độc lập nên có khả năng tự sinh sản bằng cách phân đôi ty thể mẹ để sinh ra các ty thể con [8].

##### ***1.1.2. Cấu trúc DNA ty thể***

DNA ty thể có cấu trúc sợi kép, mạch vòng không liên kết với protein histon, nằm trong chất nền của ty thể và chiếm khoảng 5% tổng số DNA của mỗi tế bào. Trình tự DNA ty thể hoàn chỉnh của người đầu tiên đã được Anderson và các tác giả công bố năm 1981 [9], năm 1999 Andrews và các tác giả đã chỉnh sửa lại trình tự này và hiện nay được gọi là “trình tự chuẩn Cambridge đã chỉnh sửa - rCRS” [10]. DNA ty thể có kích thước 16569 bp, mã hóa cho 37 gen, 13 gen mã hóa cho 13 chuỗi polypeptide tham gia vào chuỗi hô hấp tế bào, 22 gen mã hóa cho 22 RNA vận chuyển, 2 gen mã hóa cho 2 RNA ribosom. Có thể thấy hệ gen ty thể người khá gọn và hầu hết đều tham gia vào mã hóa gen. Vùng duy nhất không mã hóa ở DNA ty thể là vùng điều khiển D-loop (Displacement loop) có chứa hai vùng HV1 (Hypervariable region 1) và HV2 (Hypervariable region 2).





**Hình 1.1: Sơ đồ cấu tạo chi tiết của DNA ty thể người [11]**

Chú thích: Phân tử DNA mạch vòng, kép, kích thước 16569 bp. Vị trí của các gen mã hóa cho 13 chuỗi polypeptid, 2 RNA ribosom (12S và 16S) và 22 RNA vận chuyển được trình bày tương ứng trên hình trên. Vùng điều khiển D-loop chứa các điểm khởi đầu sao chép, promoter của chuỗi nặng, nhẹ và chứa vùng siêu biến HV1 và HV2.

### 1.1.3. Vùng HV1 và HV2 trên DNA ty thể

#### 1.1.3.1. Vùng điều khiển D-loop

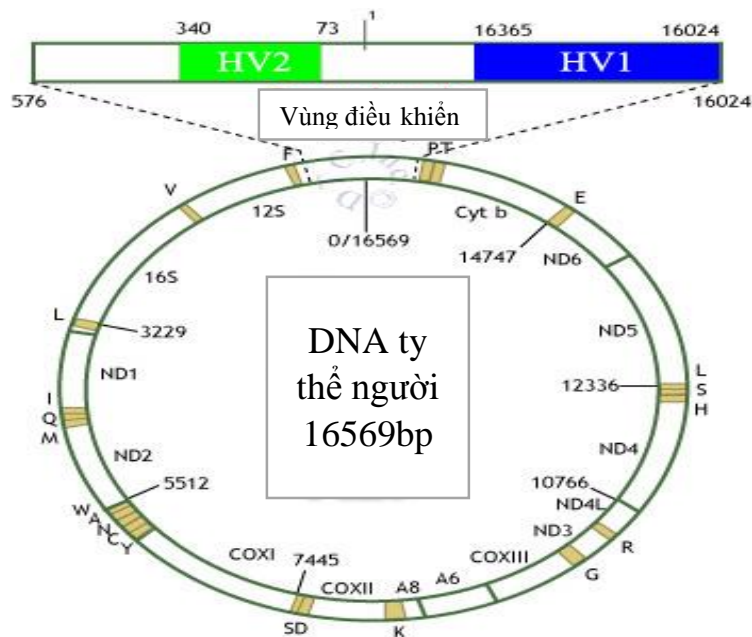
Vùng điều khiển D-loop có kích thước 1121bp, nằm từ vị trí 16024-16569/0-576 và giữa hai gen tRNA vận chuyển cho Phenylalanin và Prolin, chiếm khoảng 7% tổng lượng DNA ty thể, chứa các trình tự khởi đầu cho quá trình tái bản DNA ty thể và các đoạn điều khiển cho quá trình phiên mã của các gen chức năng trong vùng được mã hóa [9]. Đây là vùng được xem là có nhiều đột biến nhất với tần số đột biến cao hơn các vùng khác của hệ gen ty thể khoảng 4,4 lần [12].

Trình tự hoàn chỉnh vùng điều khiển D - loop trên DNA ty thể của nhiều dân tộc thuộc các chủng tộc người khác nhau, ở các châu lục khác nhau trên thế

giới được công bố trong (<http://www.mitomap.org>), cho đến nay đã có vài nghìn trình tự vùng D - loop đã được công bố. Có trên 300 trình tự hoàn chỉnh của hệ gen ty thể người thuộc các chủng tộc khác nhau được nghiên cứu và đăng ký trong các ngân hàng trình tự gen quốc tế EMBL/Genbank/DDBJ (<http://www.genpat.uu.se/mtDB/index.html>). Các công bố này cho thấy trình tự hệ gen ty thể nói chung và vùng điều khiển D-loop nói riêng của các chủng tộc và dân tộc khác nhau có những khác biệt nhất định [13]. Số lượng các trình tự D-loop cũng như trình tự toàn bộ hệ gen ty thể của các cá thể người thuộc các dân tộc khác nhau trên thế giới được giải mã vẫn tăng lên không ngừng.

### 1.1.3.2. Vùng HV1, HV2 của DNA ty thể

Năm 1981, Anderson và cộng sự đã xác định được hai đoạn DNA trong vùng D-loop được gọi là vùng siêu biến 1 (HV1- Hypervariable region 1) và vùng siêu biến 2 (HV2- Hypervariable region 2) [9]. Với tần số đột biến cao, nhiều điểm đa hình nên hai vùng này được tập trung nghiên cứu nhiều hơn cả, đặc biệt là vùng HV1.



**Hình 1.2: Hình trình bày vị trí hai vùng HV1 và HV2 trên DNA ty thể**

(theo [www.dnatestingcentre.com](http://www.dnatestingcentre.com))

Ở một số cá thể, hai vùng HV1 và HV2 có các đoạn lặp lại liên tiếp các nucleotid Cytosin, thường được gọi là các đoạn poly C. Trên vùng HV1 đoạn poly C nằm từ vị trí 16183 - 16193 còn trên vùng HV2 đoạn poly C nằm từ vị trí 303 - 327. Ở trình tự mtDNA hoàn chỉnh của người đầu tiên được công bố năm 1981 [9], đoạn poly C của vùng HV1 được ngắt quãng bởi nucleotid Thymin ở vị trí 16189. Tuy nhiên, rất nhiều trình tự DNA ty thể có đột biến T16189C tạo thành chuỗi có 10 nucleotid Cytosin liên tiếp. Đột biến T16189C được xem là đột biến có tốc độ cao nhất trong hệ gen ty thể người [14]. Các nghiên cứu cho thấy, trong các tế bào, có nhiều loại DNA ty thể có chiều dài và trình tự đoạn poly C khác nhau, đây là hiện tượng “dị tế bào chất” đoạn poly C. Tỷ lệ phần trăm các phân tử DNA ty thể mang các độ dài đoạn poly C khác nhau là ổn định ở mỗi cá thể, được di truyền theo dòng mẹ và được tạo mới trong quá trình phát triển [15].

Việc phân tích các đa hình di truyền DNA ty thể nhằm làm sáng tỏ mối quan hệ di truyền giữa các cá thể, đồng thời nghiên cứu mối liên quan giữa DNA ty thể với các bệnh di truyền theo dòng mẹ. Đa số các nghiên cứu dựa trên tính đa hình của vùng điều khiển D-loop. Các nghiên cứu cũng cho thấy có mối liên hệ nhất định giữa các bệnh với trạng thái “dị tế bào chất” ở DNA ty thể của các bệnh nhân [16]. Mặc dù, các nghiên cứu về hai vùng HV1 và HV2 được thực hiện nhiều trong những năm gần đây, nhưng người ta vẫn chưa tìm thấy mối liên quan giữa vùng HV1 và vùng HV2 của DNA ty thể. Các nghiên cứu khác nhau đánh giá tốc độ đột biến liên quan tới di truyền của vùng HV1 và HV2 vẫn còn gây tranh cãi. Do DNA ty thể không tái tổ hợp nên toàn bộ phân tử DNA có một lịch sử tiến hóa chung. Tuy nhiên, hai vùng HV1 và HV2 lại có tốc độ đột biến khác nhau và nếu sự khác nhau trong tốc độ đột biến này đủ lớn thì các yếu tố đa hình của hai vùng này có thể phản ánh được các quá trình tiến hóa khác nhau [17].

Có rất nhiều loại đột biến trong vùng điều khiển của DNA ty thể: đột biến thay thế, mất đoạn, thêm đoạn nhưng hay gặp nhất là loại đột biến thay thế nucleotid. Tốc độ đột biến của DNA ty thể cao nhất ở vùng điều khiển D-loop, đặc biệt là ở hai vùng HV1 và HV2, tốc độ này phụ thuộc vào vị trí các nucleotid khác nhau. Một số vị trí nucleotid trong vùng điều khiển bị đột biến thường xuyên hơn các vị trí khác được gọi là điểm nóng đột biến (mutational hotspots) [1].

Việc nghiên cứu hệ gen ty thể, giải mã trình tự nucleotid vùng điều khiển D-loop cũng như các gen khác của DNA ty thể, dẫn đến việc giải mã toàn bộ hệ gen ty thể của nhiều đại diện dân tộc người khác nhau trên thế giới, cùng với các nghiên cứu về đặc điểm, tính đa hình của vùng HV1 và HV2 sẽ cung cấp số liệu cần thiết cho các nghiên cứu về y học, di truyền học và nhiều lĩnh vực có liên quan khác.

## **1.2. Đặc điểm di truyền DNA ty thể**

Hệ gen người gồm có hai phần: hệ gen trong nhân và hệ gen ty thể. Việc sử dụng hệ gen trong nhân làm đối tượng nghiên cứu có một số nhược điểm như: tần số đột biến thấp, được di truyền từ cả bố và mẹ, mặt khác lại bị phân ly qua các thế hệ. Vì vậy, gen ty thể với các đặc điểm di truyền theo dòng mẹ, tần số đột biến cao, số lượng bản sao nhiều và không tái tổ hợp là thế mạnh nên DNA ty thể nhanh chóng trở thành đối tượng được nghiên cứu rộng rãi trong nhiều lĩnh vực đặc biệt là lĩnh vực y học, di truyền. Ngoài ra, thông tin về các trình tự nucleotid trên DNA ty thể còn có ý nghĩa quan trọng trong chẩn đoán, điều trị các bệnh di truyền ty thể.

DNA trong nhân có cấu trúc mạch thẳng thường không bền, dễ bị phân hủy, làm cho việc phân tích gặp nhiều khó khăn. Trong khi đó, DNA ty thể mạch vòng, có kích thước nhỏ, nằm trong tế bào chất, bền theo thời gian trong

các mô khó phân hủy như mô xương, răng và tóc, nên việc tách chiết, thu nhận DNA ty thể dễ dàng hơn, đặc biệt trong những trường hợp phân tích các mẫu bệnh phẩm sau nhiều năm [18].

Ở động vật có vú 99,99% DNA ty thể được di truyền theo dòng mẹ [19], [20]. Trong quá trình thụ tinh, tinh trùng chỉ đóng góp hệ gen trong nhân cho hợp tử chứ không đóng góp hệ gen ty thể. Ty thể của tinh trùng bị loại bỏ ngay sau khi vào trứng có thể do quá trình phân hủy protein phụ thuộc ubiquitin [21]. Vì vậy, trong đa số các trường hợp DNA ty thể đều được nhận từ mẹ. Điều này giúp cho việc nghiên cứu xác định một số bệnh di truyền theo dòng mẹ. Hiện tượng tái tổ hợp của DNA ty thể rất hiếm khi hoặc hầu như không xảy ra do đó có thể coi DNA ty thể không có sự tái tổ hợp, vì vậy DNA ty thể gần như hoàn toàn giống DNA ty thể mẹ ban đầu [22].

Các đặc điểm của hệ gen ty thể như không có histon bảo vệ, phân bố gần chuỗi phosphoryl oxy hóa, nơi mà các gốc tự do được tạo ra trong quá trình oxy hóa, đã làm cho khả năng bị đột biến của DNA ty thể cao hơn DNA trong nhân rất nhiều lần. Đặc biệt, các đột biến của DNA ty thể thường trung tính, đặc hiệu theo quần thể [23]. Tốc độ đột biến thay thế trong hệ gen ty thể gấp khoảng 10-100 lần tốc độ đột biến trung bình của hệ gen nhân [24]. Tốc độ đột biến của vùng điều khiển cao hơn vùng mã hóa, cao nhất ở hai vùng HV1, HV2 [25]. Tốc độ đột biến này có thể là do trong quá trình sao chép có lỗi và ty thể thiếu cơ chế để sửa chữa các sai sót như ở trong nhân [26].

Mặt khác, ty thể là nơi luôn diễn ra quá trình oxy hóa, sản sinh ra các chất oxy hóa mạnh như các gốc tự do. Các chất này sẽ tác động đến hệ gen của ty thể và phát sinh ra các đột biến. Do các đột biến là ngẫu nhiên nên bất kỳ nucleotid nào trong hệ gen ty thể cả vùng mã hóa và vùng không mã hóa đều có thể bị đột biến. Ngoài ra, mỗi tế bào có hàng trăm, hàng nghìn DNA ty

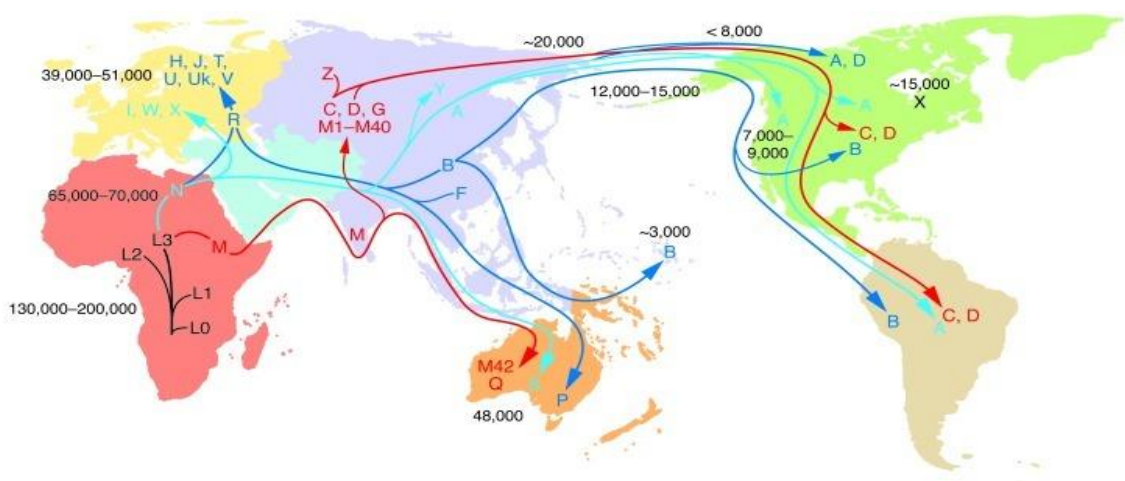
thể nên các đột biến gen ty thể có hại có thể xảy ra ở tất cả các mô, sinh dưỡng và sinh dục. Các đột biến ở các mô sinh dưỡng làm giảm việc sản xuất năng lượng cho tế bào. Các đột biến ở các tế bào sinh dục của mẹ sẽ được di truyền cho thế hệ sau tạo nên các đa hình DNA ty thể [27]. Và cũng bởi mỗi tế bào có nhiều DNA ty thể nên các gen bị đột biến có thể cùng tồn tại với các gen không bị đột biến tạo nên hiện tượng không đồng nhất. DNA ty thể có thể ở dạng đồng nhất khi tất cả các bản sao của DNA ty thể là như nhau. Trong nhiều trường hợp các đột biến dạng không đồng nhất không gây ra những biểu hiện lâm sàng hay những biểu hiện hóa sinh cho tới khi nó đạt tới ngưỡng đột biến [28].

Do DNA ty thể được di truyền theo dòng mẹ nên nó tích lũy các đột biến và phát tán theo dòng mẹ. Hơn nữa, nhờ đặc tính chọn lọc gần như vô tính nên các dạng DNA ty thể khác nhau không bị loại bỏ trong quá trình chọn lọc và do đó chúng trở nên phổ biến do sự trôi dạt di truyền. Chính vì vậy, đã tạo nên tính đa hình của DNA ty thể, tạo nên các nhóm kiểu đơn của DNA ty thể có quan hệ với nhau hay còn gọi là nhóm đơn bội (Haplogroup).

Năm 1987 Cann và đồng tác giả đã nghiên cứu đa hình DNA ty thể bằng kỹ thuật RFLP ở 147 cá thể thuộc các chủng tộc khác nhau trên thế giới. Nghiên cứu này chỉ ra rằng tất cả các mẫu DNA ty thể được phân tích đều bắt nguồn từ một tổ tiên chung được dự đoán là sống cách đây khoảng 200.000 năm, thuộc khu vực cận Sahara ở Đông Phi [29]. Kể từ đây, bùng nổ một cuộc cách mạng nghiên cứu về hệ gen ty thể người, được ứng dụng rộng rãi trong các nghiên cứu về mối quan hệ di truyền, hình sự, pháp y và về y học. Mặc dù, hệ gen ty thể chỉ chiếm một phần rất nhỏ trong hệ gen người nhưng các số liệu về DNA ty thể đã góp phần làm sáng tỏ cơ chế phát sinh một số bệnh giúp cho việc chẩn đoán và điều trị bệnh có hiệu quả.

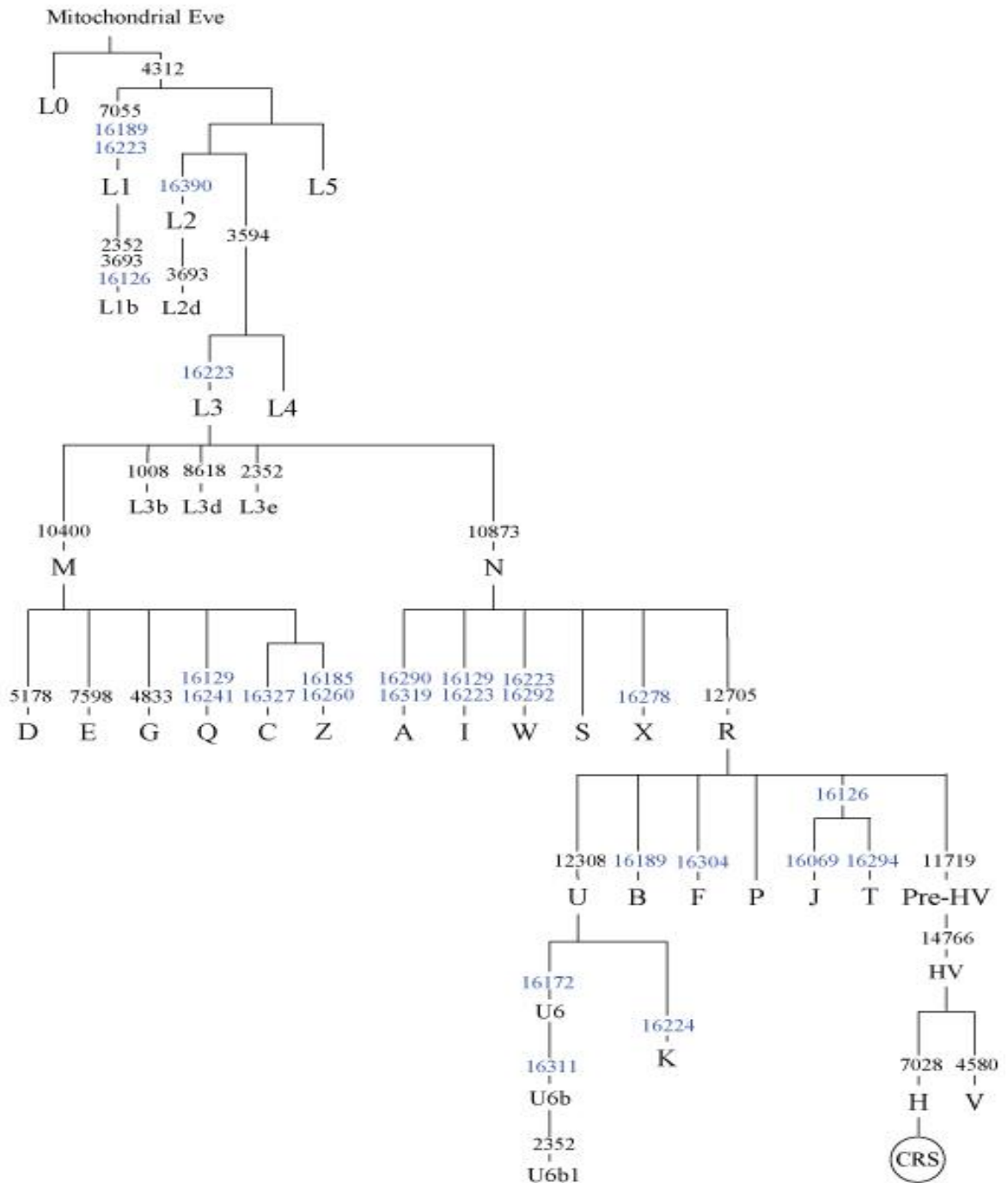
Từ các nghiên cứu đầu tiên về đa hình DNA ty thể người sử dụng kỹ thuật RFLP và trình tự đoạn HV1, HV2 trên vùng điều khiển D-loop của DNA ty thể, người ta thấy rằng các loại DNA ty thể khác nhau có thể xếp vào các nhóm đơn bội khác nhau dựa trên sự có mặt hay vắng mặt các băng khi cắt DNA ty thể bằng một enzym cắt giới hạn nào đó và dựa trên trình tự đặc trưng của vùng điều khiển. Theo đó mỗi nhóm đơn bội có một bộ các đa hình đơn nucleotid (SNP) đặc trưng. Việc phân loại và sắp xếp các nhóm đơn bội sẽ giúp phân tích các dòng DNA ty thể theo phả hệ, từ đó có thể dựng lại được nguồn gốc, mối quan hệ di truyền cho một bệnh nào đó có liên quan với DNA ty thể.

DNA ty thể được phân thành bốn nhóm đơn bội đối với người Phi là L0, L1, L2 và L3 trong đó ba nhóm L0, L1, L2 chiếm 76% dân số châu Phi. Ở Đông Bắc Phi, hai dòng DNA ty thể M và N xuất hiện từ nhóm L3 khoảng 65.000 năm trước đây và là tổ tiên của tất cả các dòng DNA ty thể trên lục địa Á, Âu. Ở châu Âu, nhóm L3 và N tạo nên chín nhóm đơn bội là H, I, J, K, T, U, V, W, X. Ở châu Á, hai nhóm đơn bội lớn là M và N phân ly tạo nên các nhóm đơn bội nhỏ hơn A, B, C, D, F, M, G, Z [30].



**Hình 1.3: Sơ đồ mô tả quá trình di cư của các nhóm đơn bội DNA ty thể người [30]**

Việc phân loại DNA ty thể theo các nhóm đơn bội khá phức tạp, dưới đây là sơ đồ phân nhóm đơn bội dựa theo các đa hình đặc trưng trên DNA ty thể được công bố trên Genbase *Tutorials* tháng 7 năm 2015.



**Hình 1.4: Sơ đồ phân nhóm đơn bội mtDNA theo các vị trí đa hình đặc trưng [31]**



Theo Yao và cộng sự năm 2002, tác giả đã phân chia nhóm đơn bội (Haplogroup) các dân tộc người Trung Quốc dựa vào các vị trí đa hình đặc trưng trên hai vùng HV1 và HV2 [32]. Cụ thể về các nhóm đơn bội và vị trí đa hình đặc trưng của chúng trên vùng HV1 và HV2 được trình bày ở bảng sau:

**Bảng 1.1: Phân chia các nhóm đơn bội DNA ty thể dựa vào vị trí đa hình đặc trưng trên vùng HV1 và HV2 [32]**

Haplogroup	SNP trên HV1 (16000+)	SNP trên HV2 (73, 263)	Haplogroup	SNP trên HV1 (16000+)	SNP trên HV2 (73, 263)
B4	189, 217		G2	278, 362	
B4a	189, 217, 261		M	223	
B4b	136, 189, 217		M7		146, 199
B5	189		M7b	297	150, 199
B5a	140, 189, 266		M7b1	129, 192, 297	150, 199
D4	362		M7c	295	146, 199
D5	189, 362	150	M9	234	153
F	304	249DelA	M10	311	
F1a	129, 172, 304	249DelA	N9a	257A, 261	150
F2a	291, 304	249DelA	R9a	298, 355, 362	249DelA
F1b	189, 304	249DelA	Z	185, 260, 298	152, 249DelA
M8a	184, 298, 319		C	327	249DelA
G2a	227, 278, 362		D4a	129, 362	152
D5a	189, 266, 362	150	D4b	319, 362	
A	290, 319		Y	126, 231, 266A	146
B5b	140, 189, 243		F1c	111, 129, 304	152, 249DelA

Chú thích: 249DelA: mất A ở vị trí 249 trên vùng HV2.

Ngày nay, với sự phát triển của các phương pháp PCR và phương pháp xác định trình tự DNA tự động, việc nghiên cứu hệ gen ty thể, đọc trình tự nucleotid các vùng HV1, HV2 của vùng điều khiển D-loop cũng như các gen chức năng khác của ty thể, dẫn đến giải mã toàn bộ hệ gen ty thể của nhiều đại diện dân tộc khác nhau sẽ cung cấp số liệu các bộ SNP để phân loại các nhóm đơn bội chính xác và chi tiết hơn [23].

### **1.3. Nghiên cứu tính đa hình đơn nucleotid**

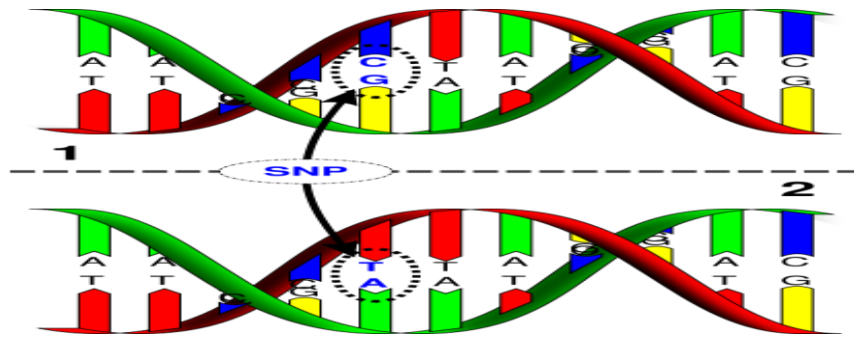
#### ***1.3.1. Đa hình đơn nucleotid (Single Nucleotide Polymorphisms-SNP)***

Đa hình đơn nucleotid (Single Nucleotide Polymorphisms - SNP) là loại biến đổi di truyền phổ biến nhất, đại diện cho sự khác biệt một nucleotid trong hệ gen người. Sự khác biệt cho mỗi cá thể được tạo bởi sự đa hình của các gen. Hiện tượng đa hình đơn nucleotid là sự khác nhau về trình tự DNA xảy ra khi một nucleotid đơn A, T, G, C ở trong bộ gen (hay trong các trình tự được phân lập khác) khác nhau giữa các cá thể của một loài hay giữa các cặp nhiễm sắc thể của một người. Bộ gen người có khoảng 3 tỉ base, trong đó có khoảng 10 triệu vị trí base mà tại đó SNPs xảy ra tương đối thường xuyên. Năm 2003 dự án HapMap (nghiên cứu về các SNPs) đã được khởi động trên toàn cầu. Mục tiêu của dự án HapMap là xác định các mô hình phổ biến của sự biến đổi trình tự DNA trong hệ gen của con người và thông tin này có thể tra cứu được dễ dàng. HapMap sẽ cho phép khám phá các biến thể của chuỗi DNA ảnh hưởng đến bệnh, tạo điều kiện phát triển các công cụ chẩn đoán, và tăng cường khả năng để lựa chọn các mục tiêu cho can thiệp điều trị [33].

Các nhà nghiên cứu hy vọng từ việc lập bản đồ SNPs và tần số xuất hiện của chúng trong các quần thể khác nhau, họ có thể tăng tác dụng của thuốc điều trị theo từng dạng di truyền khác biệt. Vì có rất nhiều SNP, nên

mục đích chính của HapMap là sắp xếp chúng sao cho các phân tích nói trên được dễ dàng hơn. Các nhà di truyền cố gắng tìm kiếm mối quan hệ giữa các SNP với nhau, nghĩa là khi một SNP nào đó xuất hiện thì sẽ luôn có SNP thứ hai thứ ba cũng được tìm thấy. Bước kế tiếp là giải mã dữ liệu để lập bản đồ SNP và nghiên cứu mối liên hệ giữa chúng, từ đó có thể làm sáng tỏ nguyên nhân chính xác của một số rối loạn di truyền. Mặc dù, sự khác biệt di truyền giữa các quần thể khác nhau ở các khu vực địa lý khác nhau thực sự rất nhỏ thì vẫn tồn tại một khuynh hướng di truyền thể hiện sự khác biệt rõ ràng giữa các nhóm tộc người. Ví dụ, xét một nhóm SNP, chúng có thể liên kết theo một cách nào đó ở quần thể người châu Á, nhưng chúng lại liên kết theo cách khác khi xem xét trên chủng người châu Âu. Ngoài ra khi xem xét toàn bộ hệ gen thì sự khác biệt giữa cá thể vẫn nhiều hơn sự khác biệt giữa nhóm tộc người, tuy thế khuynh hướng so sánh sự khác biệt giữa các chủng người vẫn tỏ ra hữu hiệu trong việc nghiên cứu các thuốc điều trị bệnh đặc hiệu. Tuy nhiên, không phải tất cả mọi người trong cùng một chủng tộc người nào đó đều có chung kiểu mẫu SNP, do đó sự hiện diện hay vắng mặt của những marker di truyền nào đó là thực sự quan trọng [33].

SNP là một hiện tượng phổ biến, được coi là hậu quả của những đột biến điểm thay thế một cặp nucleotid. Để phân biệt đột biến và SNP thì người ta dựa vào tần số xuất hiện trong cộng đồng. Nếu những đột biến xuất hiện > 1% trong quần thể dân cư thì được coi là SNP. Theo dữ liệu của NCBI tính đến tháng 6/2012 có khoảng gần 19 triệu SNPs trong bộ gen người. Các đa hình đơn nucleotid có tính chủng tộc. Cùng một SNP nhưng tỷ lệ các biến thể (variant) trong quần thể khác nhau giữa các tộc người, thậm chí có và không có, trong bộ gen của những tộc người khác nhau.



**Hình 1.5: Mô phỏng hiện tượng đa hình đơn nucleotid [34]**

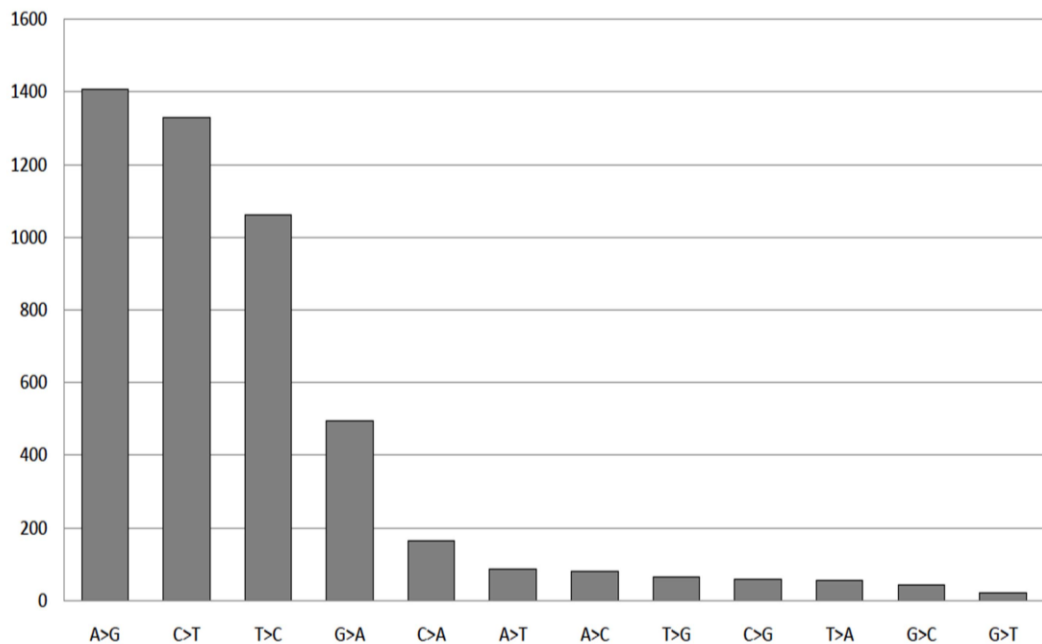
Các dạng đa hình đơn nucleotid có thể xuất hiện ở vùng mã hóa, vùng không mã hóa protein hoặc ở vùng giữa hai khu vực mã hóa và không mã hóa. SNP có thể làm thay đổi mã di truyền hoặc không làm thay đổi mã di truyền. Có những SNP làm thay đổi acid amin, có thể dẫn đến ảnh hưởng tới cấu trúc và chức năng phân tử protein, những SNP không làm thay đổi mã di truyền được gọi là các Silent SNP. Các Silent SNP không làm thay đổi trình tự acid amin, nhưng nếu nằm trên các vùng chức năng quan trọng cũng có thể gây ra các tác động đến chức năng sinh học của gen đó.

Các SNP được ứng dụng trong xác định huyết thống, điều tra tội phạm, pháp y, xác định mối quan hệ di truyền, xây dựng cây phát sinh tiến hóa quần thể giữa các quần thể người... Các SNP cũng được sử dụng như một dấu ấn sinh học giúp các nhà khoa học xác định được các gen có liên quan đến bệnh tật hoặc để theo dõi sự di truyền của một bệnh. Các SNP có thể là các dạng đột biến gen tạo nên sự đa dạng về mặt di truyền giữa các cá thể loài người, hoặc có thể là các yếu tố có nguy cơ cao gây phát sinh, đáp ứng với các tác nhân gây bệnh, đáp ứng với thuốc điều trị... Do đó, việc nghiên cứu các SNP là vô cùng quan trọng, cụ thể đã có nhiều SNP đang được sử dụng làm công cụ hữu ích trong những lĩnh vực như y học, dược học, hình sự, và nghiên cứu di truyền, tiến hóa người.

### ***1.3.2. Đa hình trên vùng HV1 và HV2 của DNA ty thể***

Cũng giống như hệ gen trong nhân, hệ gen ty thể mang những trình tự đa hình. Đa hình là sự khác biệt về chuỗi DNA giữa những cá thể, các nhóm, hoặc các quần thể. Nó bao gồm đa hình đơn (SNP), các chuỗi lặp, thêm đoạn, mất đoạn và tái tổ hợp. Đa hình gen có thể là kết quả của quá trình tiến hóa hoặc được tạo nên bởi yếu tố bên ngoài (như virus hoặc chiếu xạ). Cho đến nay đã có rất nhiều nghiên cứu về tính đa hình gen ty thể, hầu hết các nghiên cứu tập trung vào các đa hình đơn nucleotid (SNP) đặc biệt trên hai vùng HV1 và HV2 thuộc vùng điều khiển D-loop. Các vị trí đa hình hay gặp trên vùng HV1 đã được công bố trên MITOMAP như vị trí 16189 (T-C), 16192 (C-T), 16304 (T-C),... trên vùng HV2 như 73(A-G), 263(A-G), 309 (thêm C), vị trí 310 (mất T) hoặc 310 (T-C), 315 (thêm C)...

Các đa hình hay gặp nhất là sự thay thế nucleotid, có thể là các đột biến đồng hoán (transition) - đột biến thay thế nucleotid có cùng gốc purin (A-G) hoặc pyrimidin (C-T); hoặc là các đột biến dị hoán (transversion) - đột biến thay thế nucleotid có gốc purin thành pyrimidin hoặc ngược lại (A-T, C-G). Theo nghiên cứu của tác giả Shamnamole K và cộng sự năm 2013 khi nghiên cứu tổng hợp trên dữ liệu DNA ty thể từ 675 quần thể gồm 5231 cá nhân cho thấy hầu hết các đa hình là các thay thế đồng hoán, gặp 4293 thay thế đồng hoán (transition) trong khi thay thế dị hoán chỉ là 575, [35].



**Hình 1.6: Các biến đổi đồng hoán và dị hoán trên mtDNA [35]**

*Chú thích: Tần số gặp transitions - đột biến thay thế nucleotid có cùng gốc purin (A-G) hoặc pyrimidin (C-T): (4293) cao hơn transversions - đột biến thay thế nucleotid có gốc purin thành pyrimidin hoặc ngược lại (A-T, C-G): (575)*

Trong những năm qua, các thành tựu về công nghệ di truyền phân tử đã mở ra một kỷ nguyên mới cho các nghiên cứu về mối liên quan giữa bệnh tật với tính đa hình và đột biến DNA. Các nhà khoa học đã ứng dụng các chỉ thị DNA: RFLP, nhiễm sắc thể Y, DNA ty thể, SNP,... trong nhiều lĩnh vực như sinh học, dược học và đặc biệt trong y học. Nhờ vậy, mà một số bệnh đã được làm sáng tỏ về cơ chế gây bệnh, mức độ nguy cơ, sự đáp ứng với thuốc điều trị và tiên lượng bệnh. Những nghiên cứu gần đây, cho thấy đa hình trên gen ty thể có liên quan đến nhiều bệnh như: các bệnh ung thư, các bệnh liên quan đến quá trình lão hóa, chuyển hóa, bệnh di truyền ty thể. Thêm vào đó, việc nghiên cứu các SNP trên DNA ty thể cũng đã làm sáng tỏ sự khác biệt về trình tự của bộ gen ty thể có ý nghĩa quan trọng không chỉ trong nghiên cứu lịch sử mẫu hệ của con người hiện đại mà còn có ý nghĩa trong việc xác

định các mối quan hệ về chủng tộc và các vùng địa lý khác nhau. Những khác biệt về trình tự DNA ty thể ở những vùng mã hóa còn cho biết quá trình chọn lọc tự nhiên có ảnh hưởng quyết định đến quá trình tiến hóa của bộ gen ty thể [36], [37]. Các nghiên cứu cũng cho thấy trình tự DNA ty thể của các chủng tộc người khác nhau có những khác biệt nhất định, vì vậy DNA ty thể được coi là công cụ hữu hiệu trong các nghiên cứu về di truyền, tiến hóa và mối quan hệ giữa các quần thể người.

#### **1.4. Sơ lược về bệnh ung thư vú**

##### *Tỷ lệ mắc ung thư vú ở Việt Nam*

Ung thư vú là loại ung thư phổ biến nhất ở phụ nữ cả ở các nước phát triển và các nước đang phát triển. Tại Việt nam, theo số liệu ghi nhận ung thư năm 2012, ung thư vú đứng hàng đầu ở nữ giới với tỷ lệ mắc chuẩn theo tuổi trung bình trong cả nước là 29,9/100.000 dân. Ước tính năm 2020, con số này là 38,1/100.000 [38].

##### *Các yếu tố nguy cơ gây ung thư vú*

Cho đến nay, có nhiều yếu tố nguy cơ được cho là nguyên nhân gây ung thư vú, tuy nhiên, cơ chế tác động của các yếu tố này khiến tế bào bình thường trở thành tế bào ung thư vẫn chưa được biết đến đầy đủ. Nguyên nhân có thể là do các yếu tố di truyền hoặc là các yếu tố môi trường, hoặc là sự kết hợp của cả di truyền và môi trường.

Theo Hiệp hội Ung thư Hoa Kỳ các yếu tố nguy cơ ung thư vú liên quan đến di truyền bao gồm: tiền sử gia đình có người mắc ung thư vú; do di truyền liên quan đến đột biến các gen ức chế khối u *BRCA1*, *BRCA2*, *TP53*; u xơ vú lành tính; các yếu tố về sinh sản và nội tiết như mức độ hormon nội sinh cao, tuổi có kinh nguyệt sớm hoặc mãn kinh muộn, tình trạng mang thai và cho con bú [39].

Các yếu tố nguy cơ liên quan đến môi trường và lối sống bao gồm: sử dụng hormon sau mãn kinh, ít vận động, chế độ ăn nhiều chất béo, ít chất xơ, béo phì, uống rượu, hút thuốc lá và sử dụng thuốc tránh thai. Một số yếu tố khác như phơi nhiễm với phóng xạ, ô nhiễm môi trường cũng làm tăng nguy cơ mắc ung thư vú. Bên cạnh đó, tỉ lệ mắc và tử vong vì ung thư vú tăng theo độ tuổi và nữ giới có nguy cơ mắc ung thư vú cao hơn so với nam giới.

*Các xét nghiệm cận lâm sàng dùng trong chẩn đoán ung thư vú*

Hiện nay, các xét nghiệm dùng trong chẩn đoán, theo dõi điều trị ung thư vú bao gồm: kháng nguyên ung thư CA15-3; kháng nguyên ung thư CA27-29; thụ thể oestrogen (*ER*) và thụ thể progesterone (*PR*); thụ thể của yếu tố tăng trưởng thượng bì (*HER2/NEU*). Ngoài ra, để sàng lọc và chẩn đoán sớm ung thư vú thì cần thiết phải thăm khám vú lâm sàng định kỳ, chụp nhũ ảnh và siêu âm.

Tỷ lệ tử vong của bệnh ung thư vú cao nguyên nhân chủ yếu được xác định là do bệnh được chẩn đoán và phát hiện vào giai đoạn muộn. Theo báo cáo của tổ chức y tế thế giới năm 2012 cho biết nếu bệnh được phát hiện sớm thì sẽ giảm được nguy cơ tử vong xuống 1/3 so với tỷ lệ mắc, từ đó cho thấy việc chẩn đoán và phát hiện sớm ung thư vú đóng một vai trò hết sức quan trọng trong việc kéo dài khả năng sống cho bệnh nhân.

Những biến đổi di truyền được xem là một trong những nguyên nhân gây nên sự tiến triển ung thư vú. Bên cạnh những đột biến xảy ra trong nhân thì đột biến trên DNA ty thể đặc biệt là vùng điều khiển D-loop cũng được quan sát thấy ở nhiều loại ung thư trong đó có ung thư vú [5].

### **1.5. Đa hình gen ty thể và mối liên quan đến bệnh**

Những đặc điểm của hệ gen ty thể được công bố từ đầu những năm 1980, và tới năm 1988 những đột biến đầu tiên có liên quan tới các bệnh đã được tìm thấy. Bệnh của DNA ty thể ảnh hưởng nhiều đến hệ thống thần kinh



trung ương, cơ xương, tim, thận, gan, tụy, vú, tuyến nội tiết và hệ thống hô hấp. Tùy thuộc vào vị trí các tế bào bị ảnh hưởng mà các triệu chứng có thể bao gồm mất kiểm soát hoạt động cơ, yếu cơ, đau, rối loạn dạ dày - ruột, bệnh tim, bệnh gan, tiểu đường, co giật, bệnh thị giác, thính giác, chậm phát triển, tăng quá trình lão hóa và bệnh ung thư. Các bệnh được mô tả khá rõ dựa trên các biểu hiện lâm sàng, hình thái và hóa sinh nhưng bệnh khó được nhận ra bởi biểu hiện lâm sàng của bệnh rất biến đổi và khởi đầu bệnh diễn ra rất âm thầm, đặc biệt trong giai đoạn đờ ỉa trẻ còn nhỏ [36].

Hiện nay, do sự tiến bộ của các phương pháp sinh học phân tử nên việc xác định các bệnh lý ty thể đã có nhiều khả quan. Các nghiên cứu trong những năm gần đây cho thấy, bệnh ty thể là một trong những bệnh liên quan đến đột biến gen phổ biến ở người. Theo một nghiên cứu ở Đông Bắc nước Anh có 12,48/100.000 cá nhân ở người lớn và trẻ em có bệnh DNA ty thể hoặc có nguy cơ phát triển bệnh DNA ty thể [40].

Đã có hơn 250 đột biến DNA ty thể gây ra các bệnh ở người được phát hiện. Các bệnh này có thể xuất hiện ở bất kỳ giai đoạn nào trong cuộc đời, từ đờ ỉa bé mới sinh đến người trưởng thành ở mọi lứa tuổi. Ngoài ra, đột biến DNA ty thể được di truyền theo dòng mẹ nên chẩn đoán cho một người có thể được dùng để chẩn đoán cho nhiều thế hệ trong gia đình [41].

Sự tích lũy các loại oxy phản ứng (ROS) và các tác nhân oxy hóa có liên quan đến nhiều bệnh lý, bao gồm các bệnh thoái hóa thần kinh, tiểu đường, ung thư và lão hóa sớm. ROS và các gốc tự do gây ra các đột biến gen, tăng sự hình thành và tích lũy các đột biến DNA ty thể ở các mô trong quá trình lão hóa. Ví dụ, các loại oxy phản ứng có nguồn gốc từ ty thể (ROS) kích hoạt phản ứng tổn thương DNA ty thể và apoptosis (chết theo chương trình của tế bào) có thể thúc đẩy xơ phổi và các bệnh phổi thoái hóa, ung thư phổi và bệnh phổi tắc nghẽn mãn tính...[42].

Y học tiến hóa cho rằng các biến thể di truyền cho phép tổ tiên của chúng ta thích ứng với sự đa dạng của môi trường đang có một ảnh hưởng sâu sắc đến những khuynh hướng về bệnh tật hiện nay. Các đột biến DNA ty thể kết hợp với việc kiểm soát môi trường và lượng calo không giới hạn, đang ngày càng dẫn đến sự mất cân bằng năng lượng cá nhân. Sự mất cân bằng giữa di truyền DNA ty thể và lượng calo đang thúc đẩy các bệnh của cuộc sống hiện đại như bệnh béo phì, đái tháo đường, bệnh thoái hóa thần kinh, bệnh tim mạch và ung thư [43]. Nhiều nghiên cứu đã chỉ ra rằng đa hình trên gen ty thể cũng liên quan đến các bệnh chuyển hóa, lão hóa như: Đái tháo đường [44], Alzheimer [45], Parkinson [46]...

Ngoài ra, những thông tin về trình tự DNA ty thể cũng là công cụ hữu hiệu trong việc giám định gen, xác định huyết thống [47]. Các đột biến DNA ty thể gây bệnh bao gồm đột biến điểm, đột biến mất đoạn và đột biến thêm đoạn, đã có nhiều nghiên cứu về các đột biến DNA ty thể gây nên các bệnh di truyền ty thể, thường gặp như: hội chứng Leigh, bệnh thần kinh thị giác Leber, bệnh LHON, hội chứng Kearns-Sayre, hội chứng Pearson, hội chứng MERRF, NARP...[48].

### **1.6. Đa hình gen ty thể và bệnh ung thư vú**

Các đột biến của DNA ty thể từ lâu đã được cho là có liên quan với quá trình tạo khối u bởi vì các tế bào cần sử dụng nhiều năng lượng để sinh trưởng và tăng sinh dưới các điều kiện hạn chế.

Ty thể có chức năng quan trọng trong chết theo chương trình của tế bào (apoptosis), quá trình sinh học thiết yếu trong đó tế bào chết theo một phương thức được kiểm soát [49]. Apoptosis cũng đóng vai trò chủ yếu trong sự phát triển của ung thư và trong đáp ứng của tế bào đối với các chất chống ung thư. Bên cạnh đó, DNA ty thể rất dễ bị tổn thương bởi quá trình oxy hóa do nằm

gắn với vị trí sản sinh ra các gốc tự do chứa oxy (ROS - reactive oxygen species), do đó tỉ lệ đột biến của DNA ty thể tăng từ 10 đến 100 lần so với tỉ lệ đột biến của DNA nhân [50]. Mặt khác, không giống với DNA nhân, DNA ty thể không có protein histon bảo vệ và cơ chế sửa chữa DNA kém hiệu quả dẫn đến dễ xảy ra các sai sót trong quá trình sao chép. Ước tính tỉ lệ đột biến ở DNA ty thể cao hơn khoảng 19 - 220 lần so với gen nhân [51].

Do có nhiều bản sao DNA ty thể trong một tế bào nên các phân tử DNA ty thể bị đột biến có thể cùng tồn tại với phân tử DNA ty thể bình thường, hiện tượng này được gọi là trạng thái dị tế bào chất (heteroplasmy). Ngược lại ở trạng thái đồng tế bào chất (homoplasmy) chỉ có một dạng DNA tồn tại trong tế bào. Mức độ biểu hiện bệnh của các đột biến DNA ty thể chủ yếu phụ thuộc vào tỷ lệ giữa phân tử DNA bị đột biến với phân tử DNA bình thường. Khi mức độ dị tế bào chất vượt qua một ngưỡng nhất định từ 50% - 90% tùy thuộc vào loại đột biến và loại mô, thì sẽ biểu hiện ra lâm sàng [52].

Bên cạnh đó, một đặc điểm của các tế bào ung thư là phải tạo ra một lượng lớn ATP (adenosin triphosphat) đáp ứng nhu cầu năng lượng của chúng và để tổng hợp các nucleotid, lipid và protein mới cần cho tăng sinh nhanh chóng. Ty thể tạo ra năng lượng cho tế bào thông qua 2 con đường chính: phosphoryl oxy hóa (OXPHOS - *Oxidative phosphorylation*) và chu trình acid citric, trong đó con đường phosphoryl oxy hóa - nguồn tạo năng lượng chính của các sinh vật hiếu khí, tạo ra năng lượng cao gấp 10 lần so với chu trình acid citric và con đường đường phân (glycolysis).

Các nguyên nhân gây ảnh hưởng đến quá trình tổng hợp năng lượng của con đường phosphoryl oxy hóa có thể dẫn đến giảm sản xuất năng lượng, tăng sản xuất các loại oxy phản ứng (ROS), thay đổi trạng thái oxy hóa khử và thay đổi cân bằng nội môi canxi. Sự suy giảm hiệu quả của con đường

phosphoryl oxy hóa có thể lần lượt gây nhiều sinh học ty thể, tăng sản xuất các loại oxy phản ứng, suy giảm chức năng ty thể,... dẫn đến sự gia tăng tổn thương mtDNA và đột biến soma làm suy giảm thêm chức năng của ty thể. Khi chức năng ty thể giảm xuống dưới ngưỡng năng lượng sinh học của mô, các triệu chứng bệnh sẽ xảy ra, đó là các bệnh thoái hóa, chuyển hóa, lão hóa và ung thư [53].

Ở các tế bào bình thường, con đường phosphoryl oxy hóa tạo ra khoảng 90% năng lượng ATP trong khi con đường đường phân chỉ chiếm khoảng 10% [54]. Otto Warburg là người đầu tiên quan sát thấy rằng các tế bào ung thư tăng sinh tiêu thụ rất nhiều glucose và giải phóng ra lactat thay vì  $\text{CO}_2$  [55]. Điều này có nghĩa là các tế bào ung thư sử dụng con đường phosphoryl oxy hóa ít hơn (tạo ra khoảng 50% ATP từ con đường phosphoryl oxy hóa và 50% ATP từ đường phân). Đặc biệt là sự thay đổi này xảy ra thậm chí khi có đầy đủ oxy cung cấp cho trao đổi năng lượng của ty thể, hiện tượng này được gọi là “Đường phân hiếu khí” hoặc “Hiệu ứng Warburg” [56].

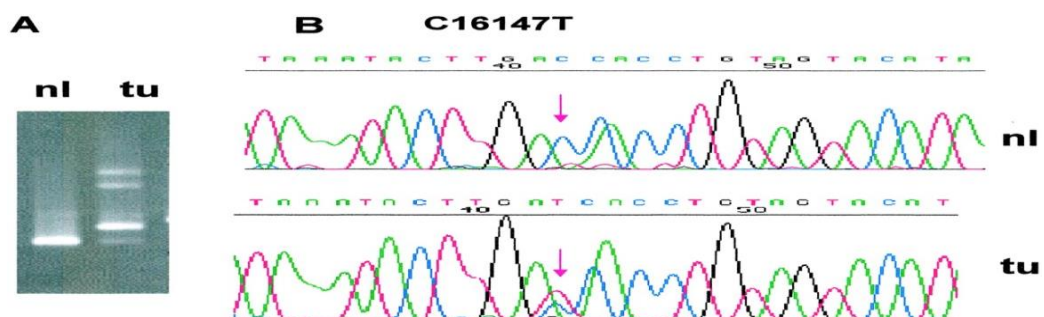
Tuy nhiên, một số nghiên cứu lại cho thấy con đường phosphoryl oxy hóa chiếm khoảng từ 40 - 80% của ATP tổng số được tạo ra trong tế bào trong môi trường giàu glucose [57]. Ngoài ra, thiếu oxy hoặc thiếu oxy máu gián đoạn và/hoặc thiếu glucose góp phần vào việc chuyển đổi sang sử dụng đường phân hoặc hô hấp không phụ thuộc glucose ở các tế bào ung thư để tạo ra đủ lượng ATP yêu cầu [58].

Cho đến nay, vai trò của các đột biến DNA ty thể ảnh hưởng đến chức năng của con đường phosphoryl oxy hóa đã được chứng minh rõ ràng. Tác động của DNA ty thể đến quá trình phát sinh ung thư hoặc tiến triển thành ác tính có thể bao gồm một số thay đổi trong DNA ty thể mà trước tiên là giảm số

bản sao DNA ty thể, sau đó là giảm biểu hiện của các gen ty thể [59] hoặc biến đổi hoạt tính enzym của ty thể [60], hoặc có thể là các đột biến soma, đột biến dòng mầm của DNA ty thể [61]. Sự thiếu hụt của phosphoryl hóa oxy hóa làm giảm nguồn cung cấp năng lượng và thúc đẩy sản sinh các loại oxy phản ứng, kích ứng các đột biến và phá hủy oxy hóa đối với DNA ty thể.

Việc giải mã toàn bộ hệ gen ty thể người đã giúp xác định được một số biến đổi của DNA ty thể liên quan đến nhiều bệnh ung thư khác nhau, bao gồm ung thư vú [62], ung thư buồng trứng [63], ung thư biểu mô dạ dày [64], ung thư gan [65], ung thư tụy [66], ung thư tuyến tiền liệt [67], ung thư phổi [68]...

Trong những năm gần đây, nhiều tác giả đã tập trung nghiên cứu về đa hình/đột biến của DNA ty thể đặc biệt là đa hình/đột biến trên vùng HV1 của DNA ty thể với bệnh ung thư vú. Nghiên cứu của Tan và cộng sự các tác giả đã tìm thấy 14 trong số 19 khối u vú (74%) có ít nhất một đột biến mtDNA soma, 27 đột biến soma đã được tìm thấy, và 22 trong số đó xảy ra trong khu vực vùng điều khiển D-loop [62]. Dưới đây là hình ảnh phát hiện đột biến soma trong khối u vú:



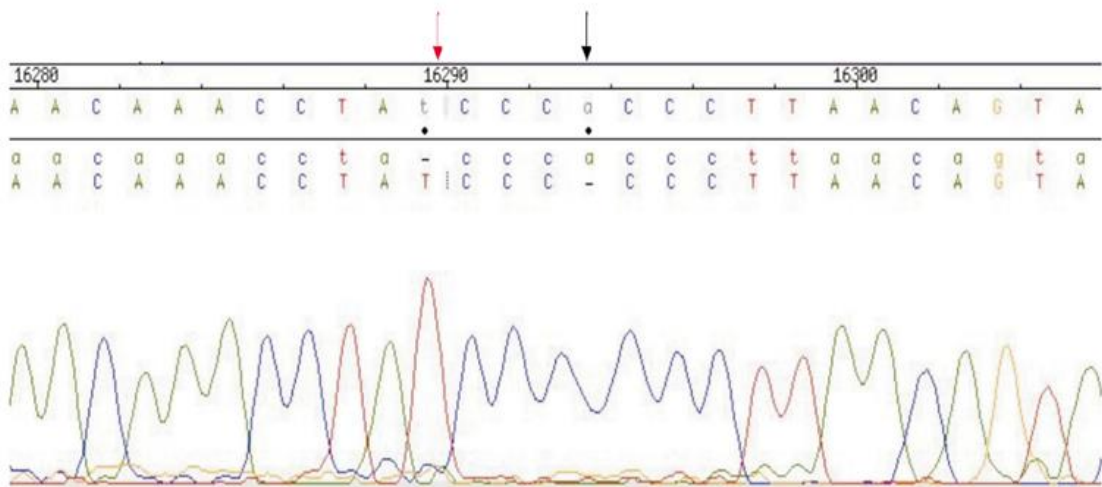
**Hình 1.7: Phát hiện đột biến C16147T (trên vùng HV1 của DNA ty thể) trong khối u vú [62]**

*A: phát hiện ra đột biến dị hợp tử trong khối u*

*nl: bình thường; tu: khối u.*

*B: phân tích trình tự cho thấy đột biến dị hợp tử C16147T/C.*

Biến đổi theo hướng giảm số bản sao DNA ty thể cũng được xác định thấy ở bệnh ung thư vú [69]. Trong nghiên cứu của Sultana và cộng sự năm 2012, cho thấy hai đa hình thường gặp nhất trên vùng HV1 của DNA ty thể ở bệnh nhân ung thư vú là SNP 16290insT và 16293delA gặp trong 95% và 75% trường hợp bệnh, nhưng chỉ gặp dưới 5% số người ở nhóm chúng [70].



**Hình 1.8: Hình ảnh đa hình 16290insT và 16293delA vùng HV1 mtDNA trên bệnh nhân ung thư vú người Bangladesh [70].**

Năm 2011, Chuanzhong Ye và cộng sự khi nghiên cứu phát hiện đột biến tại vị trí đa hình *MnII* nằm giữa các nucleotid 16.106 và 16.437 của vùng HV1 trên vùng điều khiển D-loop của DNA ty thể ở bệnh nhân ung thư vú nguyên phát, tác giả cho thấy đột biến tại vị trí *MnII* có thể đóng vai trò trong sinh bệnh học của ung thư vú [71]. Fang và cộng sự khi nghiên cứu các đa hình đặc trưng, các nhóm đơn bội DNA ty thể trên 104 bệnh nhân ung thư vú ở người Trung Quốc cho kết quả nguy cơ bị ung thư vú của haplogroup M tăng hơn các nhóm đơn bội khác với (OR = 1,77; 95%CI (1,03-3,07); p = 0,04). Tác giả cho rằng sự kết hợp giữa đa hình

đơn nucleotid của DNA ty thể với các yếu tố trong nhân có thể đóng vai trò trong sự hình thành khối u. Và sự kết hợp của SNP T16362C và G16129A trong vùng HV1 có ảnh hưởng đến sự sao chép của DNA ty thể [72].

Zhu và cộng sự khi nghiên cứu trên bệnh nhân ung thư vú đã phát hiện được 54 đột biến trong vùng điều khiển D-loop của DNA ty thể. Các đột biến ở vị trí 16293 trên vùng HV1 và ở vị trí 204, 207 trên vùng HV2 của DNA ty thể có thể gợi ý nguy cơ mắc bệnh ung thư vú [73]. Đột biến DNA ty thể cũng đã được phát hiện trong các mô ung thư vú, có 42 đột biến ở vùng điều khiển D-loop của DNA ty thể được tìm thấy trong các mô ung thư vú [74].

Đột biến soma của DNA ty thể đã được chứng minh trong các khối u khác nhau, bao gồm cả ung thư vú. Khi phân tích các đột biến soma ở vùng D-loop, đột biến xóa 4.977 bp phổ biến và số lượng bản sao mtDNA trong ung thư vú từ 60 bệnh nhân Đài Loan. Kết quả cho thấy có 30% bệnh nhân ung thư vú có đột biến soma ở vùng điều khiển D-loop của mtDNA. Sự xuất hiện của đột biến trong vùng D-loop có liên quan đến tuổi khởi phát bệnh ( $\geq 50$  tuổi,  $p = 0,042$ ), bệnh nhân bị đột biến trong vùng D-loop của mtDNA có tỷ lệ sống sót thấp hơn đáng kể so với những người không bị đột biến. Khi phân tích hồi quy đa biến tác giả chỉ ra rằng đột biến trong vùng điều khiển D-loop của mtDNA là một dấu hiệu quan trọng độc lập với các biến lâm sàng khác và nó có thể được sử dụng để đánh giá tiên lượng của bệnh nhân [69]. Số lượng đột biến soma trong vùng điều khiển D-loop của mtDNA là một chỉ số tiên lượng xấu với bệnh ung thư vú [75].

Ngoài ra, nhiều nghiên cứu cũng đã chỉ ra mối liên quan giữa đột biến DNA ty thể với nhiều loại ung thư khác nhau. Năm 2002, Okochi và cộng sự đã phát hiện các đột biến soma trên vùng D-loop của DNA ty thể trong bệnh ung thư gan [76]. Nghiên cứu của Shi và cộng sự năm 2002 cũng chỉ ra rằng,

tần số đột biến nucleotid vùng D-loop của các mô ung thư buồng trứng cao hơn nhiều so với mô bình thường [77]. Số bản sao DNA ty thể tăng ở ung thư tuyến giáp [78], ung thư phổi [79], ung thư đại trực tràng [80]. Biến đổi theo hướng giảm số bản sao DNA ty thể được xác định thấy ở bệnh nhân ung thư biểu mô tế bào gan [81], ung thư buồng trứng [82]. Khi phân tích gen ATPase6, CytB, ND1, and D310 của mtDNA trong ung thư bàng quang các tác giả đã thấy gặp phổ biến là các đa hình: G8697A, G14905A, C15452A, and A15607G [83].

Như vậy, có thể thấy DNA ty thể có liên quan đến nhiều bệnh ung thư trong đó có ung thư vú, do đó việc nghiên cứu về DNA ty thể nói chung và vùng HV1 của DNA ty thể nói riêng trên bệnh nhân ung thư vú người Việt Nam là cần thiết và có ý nghĩa khoa học.

### **1.7. Một số đặc điểm dân tộc của người Việt Nam**

Việt Nam là một quốc gia đa dân tộc (54 dân tộc anh em). Dân tộc Kinh chiếm 87% dân số còn lại là dân tộc ít người phân bố rải rác trên địa bàn cả nước. Các dân tộc thiểu số có số dân trên một triệu người gồm: dân tộc Mường, dân tộc Khmer, dân tộc Tày, dân tộc Mông, dân tộc Thái. Hình thái cư trú xen kẽ giữa các dân tộc ở Việt Nam ngày càng gia tăng. Các dân tộc không có lãnh thổ riêng, không có nền kinh tế riêng, sự thống nhất giữa các dân tộc và quốc gia trên mọi mặt của đời sống xã hội ngày càng được củng cố.

Do điều kiện tự nhiên, xã hội và hậu quả của các chế độ áp bức bóc lột trong lịch sử nên trình độ phát triển kinh tế - văn hóa giữa các dân tộc còn chênh lệch, khác biệt. Cùng với nền văn hóa cộng đồng mỗi dân tộc trong gia đình các dân tộc Việt Nam có đời sống văn hóa mang bản sắc riêng, góp phần làm phong phú thêm nền văn hóa của cả cộng đồng.



Dựa theo ngôn ngữ, 54 dân tộc Việt Nam được chia thành 8 nhóm. Trong nghiên cứu này chúng tôi lựa chọn 4 dân tộc thuộc 3 nhóm ngôn ngữ: dân tộc Kinh, Mường thuộc nhóm ngôn ngữ Việt-Mường, dân tộc Khmer thuộc nhóm ngôn ngữ Môn-Khmer, dân tộc Chăm thuộc họ ngôn ngữ Nam Đảo. Bởi vì, dân tộc Kinh là dân tộc đa số, dân tộc Mường và dân tộc Khmer là dân tộc thiểu số có số dân lớn, bên cạnh đó dân tộc Khmer, dân tộc Chăm là dân tộc còn theo chế độ mẫu hệ.

### ***1.7.1. Dân tộc Kinh***

Dân tộc Kinh có tên gọi khác là Việt. Người Kinh cư trú khắp cả nước, nhưng đông nhất là các vùng đồng bằng và thành thị (Hà Nội, TP. HCM), dân số khoảng 73.594.427 người, chiếm 86,83% dân số toàn quốc.

Người Kinh (người Việt) có tiếng nói và chữ viết riêng. Tiếng Việt nằm trong nhóm ngôn ngữ Việt - Mường (ngữ hệ Nam Á).

Người Kinh là dân tộc đa số tại Việt Nam, tuy nhiên ở một số tỉnh người Kinh lại là dân tộc thiểu số (ví dụ ở Hòa Bình người Mường là dân tộc chiếm đa số khoảng 64% dân số toàn tỉnh).

Theo truyền thống của người Kinh thì người đàn ông là trụ cột gia đình. Các con đều lấy họ bố và được coi là nối nghiệp tông đường. Dòng họ của bố là "họ nội" còn dòng họ bên mẹ là "họ ngoại".

### ***1.7.2. Dân tộc Mường***

Dân tộc Mường có khoảng trên 1.200.000 người, sống chủ yếu ở khu vực miền núi phía Bắc, tập trung đông nhất ở tỉnh Hòa Bình và các huyện miền núi tỉnh Thanh Hóa (người dân tộc Mường ở tỉnh Hòa Bình chiếm khoảng 64% dân số toàn tỉnh).

Tiếng Mường thuộc nhóm ngôn ngữ Việt - Mường (ngữ hệ Nam Á).

Người Mường có quan hệ rất gần với người Kinh. Các nhà dân tộc học và ngôn ngữ cho rằng người Mường và người Kinh có nguồn gốc chung là người Việt-Mường cổ. Vào thời kỳ ngàn năm bắc thuộc thì bộ phận người cư trú ở miền núi ít bị Hán hóa, bảo tồn lối sống cổ đến nay là người Mường. Người Mường sống định canh định cư ở miền núi, nơi có nhiều đất sản xuất, lúa nước là cây lương thực chủ yếu. Người Mường ở nhà sàn, sống tập trung thành làng xóm ở chân núi, bên sườn đồi, nơi đất thoải gần sông suối, mỗi làng có khoảng vài chục nóc nhà.

### **1.7.3. Dân tộc Chăm**

Người Chăm còn gọi là người Chăm, người Chiêm (Chiêm Thành) là một dân tộc đã từng có một quốc gia độc lập, hùng mạnh trong lịch sử, có nền văn hóa phát triển, và là hậu duệ của các cư dân nền văn hóa Sa Huỳnh thời kỳ đồ sắt.

Ở Việt Nam người Chăm có mối liên hệ gần gũi với các dân tộc cùng thuộc họ ngôn ngữ Nam Đảo như Gia Rai, Ê Đê, Ra Glai và Chu Ru.

Người Chăm được xác định là cư dân bản địa ở khu vực duyên hải Nam Trung Bộ Việt Nam và đã có quá trình định cư lâu đời ở khu vực này. Trải qua hàng ngàn năm, dưới những biến cố lịch sử, xã hội người Chăm không còn chỉ cư trú tập trung ở khu vực duyên hải Nam Trung Bộ mà đã phân bố rộng rãi ở khắp các tỉnh phía Nam Việt Nam và một số các quốc gia khác. Hiện nay, theo Tổng điều tra dân số và nhà ở năm 2009, người Chăm ở Việt Nam có dân số khoảng 161.729 người, sống rải rác ở các tỉnh phía Nam trong đó riêng hai tỉnh Ninh Thuận và Bình Thuận chiếm khoảng trên 60%. Do đặc điểm cư trú, tính chất tôn giáo và sắc thái văn hóa mang tính vùng miền, người Chăm ở Việt Nam được chia thành 3 nhóm cộng đồng chính là: Chăm H'roi, Chăm Ninh Thuận - Bình Thuận, và Chăm Nam Bộ.

Chế độ mẫu hệ và tín ngưỡng nữ thần vẫn tồn tại ở người Chăm. Phong tục Chăm quy định con theo họ mẹ, họ bên mẹ được xem là gần (họ nội). Chỉ con gái được thừa kế tài sản, người con gái út được thừa kế nhà tự để thờ cúng ông bà và phải nuôi dưỡng cha mẹ già.

#### **1.7.4. Dân tộc Khmer**

Người Khmer ở Việt Nam (hay còn gọi là Khmer Krom, Khơ-me Crôm, Khơ-me hạ, Khơ-me dưới) là một bộ phận dân tộc Khmer sống ở đồng bằng sông Cửu Long Việt Nam. Phần lớn người Khmer sống tập trung ở Campuchia.

Tiếng nói và chữ viết của người Khmer thuộc nhóm ngôn ngữ Môn-Khmer (ngữ hệ Nam Á).

Theo Tổng điều tra dân số và nhà ở năm 2009, người Khmer ở Việt Nam có dân số 1.260.600 người, riêng người Khmer ở Sóc Trăng chiếm 30,7% dân số toàn tỉnh và chiếm 31,5% tổng số người Khmer tại Việt Nam.

Xã hội người Khmer vẫn còn tồn tại nhiều tàn dư mẫu hệ. Gia đình nhỏ một vợ một chồng, ở riêng và là đơn vị kinh tế độc lập, có nơi 3-4 thế hệ sống chung trong một nhà. Hôn nhân thường do cha mẹ xếp đặt, có sự thoả thuận của con cái, cưới xin trải qua 3 bước: làm mối, dạm hỏi và lễ cưới, được tổ chức ở bên nhà gái. Sau đó, người con trai phải ở bên nhà vợ một thời gian. Trải qua ít năm hoặc khi có con, họ ra ở riêng, nhưng vẫn cư trú bên ngoại.

Người Khmer có rất nhiều họ khác nhau. Những họ do triều Nguyễn trước đây đặt ra như: Danh, Kiên, Kim, Sơn, Thạch. Những họ tiếp thu từ người Việt và người Hoa như: Trần, Nguyễn, Dương, Trương, Mã, Lý...

### **1.8. Tình hình nghiên cứu về DNA ty thể người Việt Nam**

Hiện nay, các nghiên cứu về hệ gen ty thể nói chung đã được triển khai, tuy nhiên việc nghiên cứu về trình tự, đa hình trên vùng HV1 và HV2 của DNA ty thể ở dân tộc người Việt Nam vẫn còn là vấn đề mới. Nghiên cứu đầu tiên về DNA ty thể của người Việt Nam do Ballinger SW và cộng sự thực hiện năm 1992, sử dụng phương pháp PCR- RFLP với số lượng mẫu nhỏ và không rõ nguồn gốc nên kết quả thu được còn hạn chế [84]. Năm 1999, Ivanova cùng cộng sự nghiên cứu về sự đa hình DNA ty thể của 50 người Kinh sống tại Hà Nội, đã đưa ra kết luận người Kinh có nguồn gốc kép từ người Trung Quốc và các nhóm quần thể người Thái - Indonesia [85].

Năm 2002, Oota cùng các tác giả khác đã có một công trình đăng trên tạp chí quốc tế công bố về đoạn HV1 thuộc vùng D-loop trên hệ gen ty thể của 35 cá thể người Việt Nam sống tại Hoa Kỳ [86]. Năm 2003 đã có hai nhóm nghiên cứu công bố về ứng dụng phân tích trình tự đoạn HV1 vào việc giám định hài cốt liệt sỹ [87] và nghiên cứu chỉ thị phân tử vùng D-loop trên một số bệnh nhân ung thư vú, kết quả phân tích đã cho thấy có đột biến điểm và đột biến mất đoạn 280bp trong vùng D-loop của bệnh nhân ung thư vú người Việt Nam [88]. Phan Văn Chi và cộng sự (2006) khi nghiên cứu giải mã hệ gen ty thể các tộc người Việt Nam đã tìm thấy một số biến đổi của DNA ty thể ở người Việt Nam thuộc vùng D-loop [89].

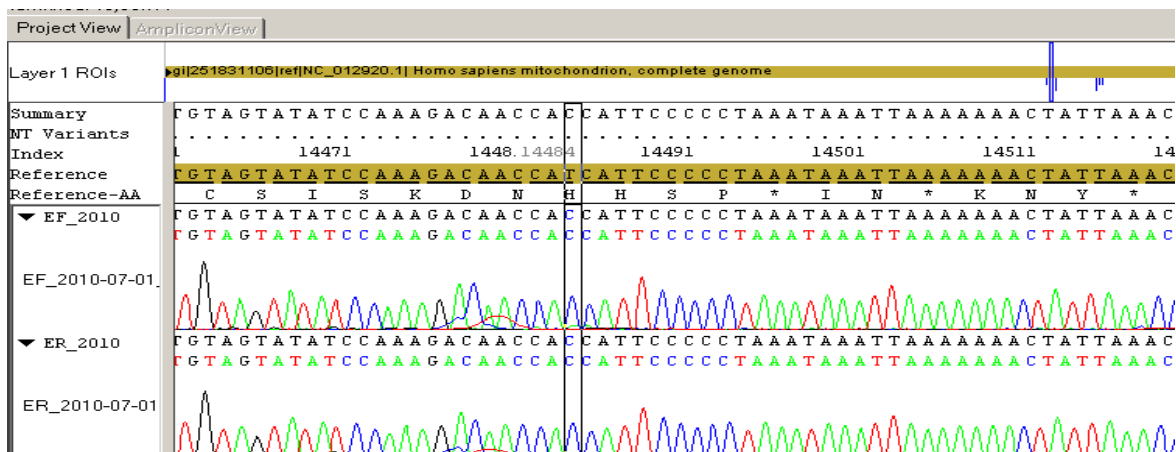
Năm 2008, Nguyễn Đăng Tôn và các tác giả khác đã sử dụng phương pháp giải trình tự DNA để nghiên cứu đa hình kiểu đơn bội DNA ty thể của 78 cá thể người Việt Nam thuộc ba dân tộc Kinh, Tày, Mường [90]. Nhóm tác giả đã xác định được 1043 điểm đa hình, tương ứng với sự thay đổi nucleotid ở 176 vị trí. Cũng trong năm 2008 nhóm nghiên cứu của Nguyễn Hữu Nghĩa đã sử dụng phương pháp giải trình tự tách dòng đã xác định được 14 vị trí đột

biến trên vùng D-loop ở một người lao động thường xuyên tiếp xúc với bức xạ ion hóa [91].

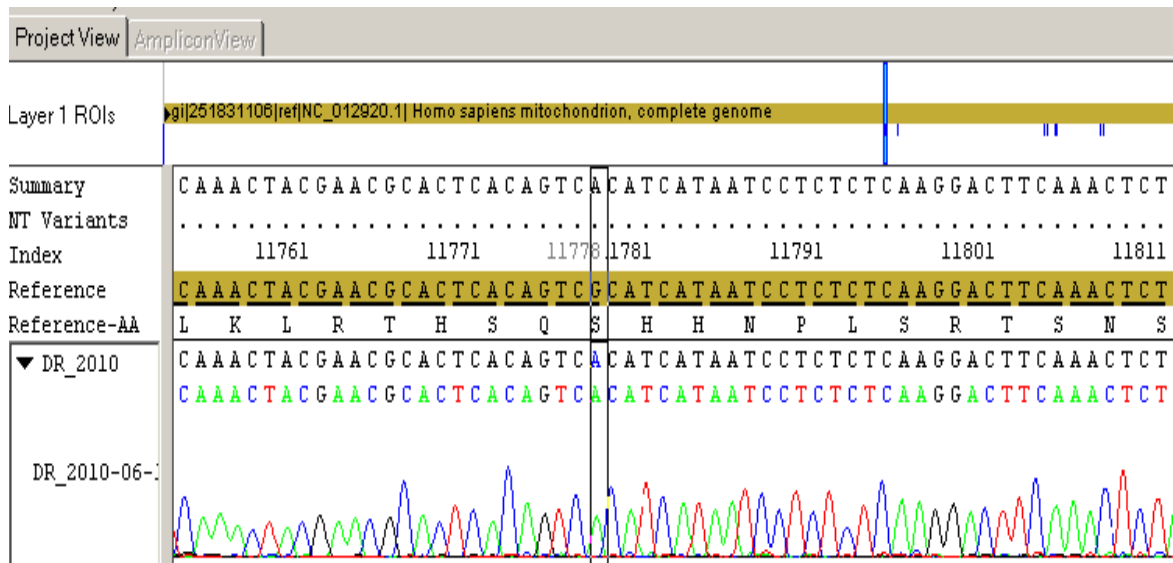
Chu Văn Mẫn và cộng sự (2009) đã sử dụng phương pháp PCR-RFLP kết hợp với giải trình tự DNA đã xác định thấy đột biến A3243G có mặt ở bệnh nhân và mẹ bệnh nhân trong một gia đình có người mắc hội chứng MELAS [92].

Năm 2010 Phạm Hùng Vân và cộng sự đã sử dụng phương pháp giải trình tự trực tiếp DNA ty thể tách chiết từ máu bệnh nhân bị bệnh LEBER, và đã xác định thấy 9/12 ca có đột biến DNA ty thể [93].

Một số nghiên cứu khác về mtDNA và bệnh ty thể như nghiên cứu trên bệnh nhân mang hội chứng MERRF [94], nghiên cứu phát hiện đột biến mtDNA trên bệnh nhân mắc bệnh thần kinh thị giác Leber: phát hiện được 18 trường hợp mang đột biến, trong đó 14 mẫu mang đột biến G11778A và 4 mẫu mang đột biến T14484C (hình dưới) [95], nghiên cứu phát hiện mất đoạn 9 bp trên gen ty thể trong bệnh não cơ [96]...



Mẫu giải trình tự có điểm đột biến T14484C



*Mẫu giải trình tự có điểm đột biến G11778A*

**Hình 1.9: Kết quả giải trình tự được so sánh với trình tự DNA ty thể chuẩn của người [95]**

Một vài nghiên cứu về mối liên quan giữa DNA ty thể với bệnh ung thư cũng đã được thực hiện trong những năm gần đây như là bệnh ung thư đại trực tràng [97]. Tuy nhiên, số liệu thu được vẫn chỉ mang tính cá thể và chưa có ý nghĩa thống kê, đại diện cho các dân tộc Việt Nam.

## **1.9. Một số phương pháp phát hiện đa hình thái gen ty thể**

### **1.9.1. Kỹ thuật PCR (Polymerase Chain Reaction)**

Kỹ thuật PCR được một nhà khoa học người Mỹ tên là Kary Mullis phát minh vào năm 1985.

#### *Nguyên tắc của phản ứng PCR*

Nguyên tắc của phản ứng PCR dựa trên cơ sở tính chất biến tính, hồi tính của DNA và nguyên lý tổng hợp DNA nhờ hoạt tính của các DNA polymerase. Với bốn loại nucleotid, mỗi xuôi và mỗi ngược và enzym DNA polymerase xúc tác sự tổng hợp một mạch DNA mới từ mạch DNA khuôn.

Phản ứng PCR là một chuỗi nhiều chu kỳ nối tiếp nhau, mỗi chu kỳ gồm ba giai đoạn: biến tính, bắt cặp, tổng hợp (kéo dài). Sau mỗi chu kỳ các chuỗi đôi DNA mới tạo thành sẽ tiếp tục được dùng để làm các DNA khuôn cho sự tổng hợp các DNA mới trong chu kỳ tiếp theo. Sản phẩm cuối của phản ứng PCR là đoạn DNA chuỗi đôi có chiều dài bằng khoảng cách giữa hai đoạn gen mô, và hai đầu tận cùng của sản phẩm được xác định bởi đầu tận cùng 5' của hai đoạn gen mô [98], [99].

Số lượng chu kỳ mỗi phản ứng PCR phụ thuộc vào số lượng khuôn DNA ban đầu, sau mỗi chu kỳ sẽ làm tăng gấp đôi lượng mẫu của lần trước. Như vậy sau  $n$  chu kỳ, từ một DNA đích đã nhân bản thành  $2^n$  bản sao. Nhờ vậy đủ số lượng DNA để có thể tách ra khi điện di và có thể phát hiện được sau khi nhuộm và để có thể tạo dòng hoặc giải trình tự. Trong quá trình thực hiện phản ứng PCR, ở những chu kỳ sau lượng khuôn tăng nhưng lượng môi và dNTP (deoxyribonucleotid triphosphat) tự do giảm, enzym DNA polymerase hoạt động yếu dần. Do đó, cần tính toán hàm lượng môi, dNTP và enzym để đảm bảo phản ứng PCR cho kết quả tốt nhất [100], [101].

### ***1.9.2. Kỹ thuật PCR - RFLP (Polymerase Chain Reaction - Restriction Fragment Length Polymorphism)***

#### *Nguyên lý*

Kỹ thuật đa hình độ dài các phân đoạn DNA được cắt giới hạn là kỹ thuật được phát triển nhờ nguyên lý của phản ứng PCR và sự phát hiện các enzym cắt giới hạn. Những đột biến gen như thay thế, mất, thêm, đảo cặp nucleotid tạo ra những sai khác rất nhỏ ở DNA, gần như không làm thay đổi kích thước đoạn DNA nên không thể phân biệt được bằng điện di. Bởi vậy, khi sử dụng các enzym cắt giới hạn để cắt các đoạn DNA có kích thước gần như nhau thành các phân đoạn nhỏ hơn, có kích thước khác nhau thì có thể

xác định được bằng điện di. Sản phẩm PCR đặc hiệu sẽ được tinh sạch và xử lý với enzym cắt giới hạn đã được lựa chọn nhằm tạo ra những phân đoạn DNA nhỏ hơn. Sản phẩm cắt bằng enzym sẽ được điện di trên gel agarose hoặc polyacrylamid, số lượng và chiều dài các đoạn DNA này sẽ giúp phát hiện các trường hợp đột biến.

Quá trình thực hiện kỹ thuật PCR-RFLP gồm các giai đoạn như sau: tiến hành phản ứng PCR với cặp mồi thích hợp để khuếch đại đoạn DNA cần phân tích sự đa hình, phân cắt đoạn DNA này bằng một enzym cắt giới hạn thích hợp, điện di xác định sự khác nhau của các băng DNA và đưa ra kết luận [102].

Kỹ thuật này đặc biệt có giá trị trong việc xác định các SNP vì có thể xác định được các trạng thái đồng hợp hay dị hợp. Do đó, kỹ thuật PCR-RFLP được ưu tiên lựa chọn trong việc phân tích các SNP có vị trí cắt đặc hiệu của enzym. Tuy nhiên kỹ thuật này chỉ cho phép xác định có SNP đã biết trước và phải có trình tự đặc hiệu cho enzym cắt giới hạn.

### ***1.9.3. Kỹ thuật giải trình tự gen***

Giải trình tự gen (DNA sequencing) là phương pháp xác định vị trí sắp xếp của các nucleotid trong phân tử DNA. Hiện nay, người ta thường sử dụng hai phương pháp giải trình tự đó là phương pháp dideoxynucleotid và giải trình tự bằng máy tự động. Ngoài ra, giải trình tự thế hệ mới được đưa vào sử dụng với ưu điểm cho phép xác định đồng thời nhiều trình tự DNA đích trong một lần chạy cho phép tăng hiệu suất và giảm chi phí thực hiện.

#### ***1.9.3.1. Giải trình tự theo phương pháp dideoxynucleotid***

Phương pháp này do Sanger, Smith và Coulson phát hiện ra năm 1977. Nguyên lý của phương pháp này như sau: Dideoxynucleotid (ddNTP) là một phân tử nhân tạo, cấu trúc của nó tương tự như phân tử deoxynucleotid



(dNTP), tuy nhiên ở carbon số 3 của đường deoxyribo không phải là nhóm hydroxyl (-OH) mà là (-H).

Trong quá trình tổng hợp mạch đơn bổ sung, một dNTP tự do gắn vào chuỗi đang tổng hợp bằng liên kết phosphodiester giữa 5' phosphat với nhóm 3' hydroxyl của nucleotid cuối cùng của chuỗi. Tuy nhiên, nếu một ddNTP được gắn vào đầu 3' của chuỗi đang tổng hợp thì sự tổng hợp DNA sẽ dừng lại do không hình thành được liên kết phosphodiester với nucleotid tiếp theo.

*Quy trình giải trình tự theo phương pháp dideoxynucleotid gồm các bước:*

a) Đoạn DNA cần giải trình tự được chèn vào một vector tại một vị trí đã biết trình tự.

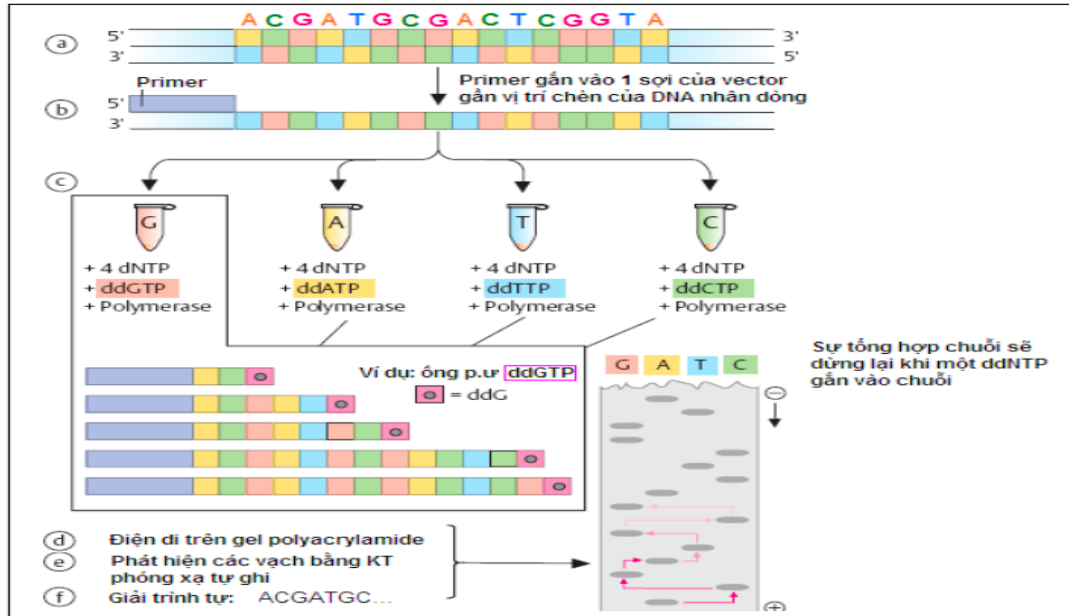
b) Sử dụng một đoạn mồi bổ sung với trình tự của vector và gắn vào đầu 3' của vector gần vị trí chèn DNA.

c) Phản ứng được tiến hành trong 4 ống nghiệm, mỗi ống được cung cấp thêm DNA polymerase, 4 loại dNTP tự do (một trong bốn loại dNTP này được đánh dấu phóng xạ) và một trong bốn dideoxynucleic (ddATP, ddCTP, ddGTP, ddTTP). Nồng độ của mỗi ddNTP được điều chỉnh thận trọng, thường chỉ dùng một lượng nhỏ, khoảng 1%. Trong quá trình tổng hợp chuỗi, enzym DNA polymerase sẽ gắn các dNTP vào để kéo dài chuỗi. Do hàm lượng rất thấp nên thỉnh thoảng mới có 1 dideoxynucleotid được gắn vào chuỗi và lúc đó phản ứng tổng hợp chuỗi bị dừng lại. Như vậy, trong mỗi ống phản ứng chứa các mạch đơn DNA có chiều dài khác nhau, và ở đầu 3' của các đoạn DNA chứa một ddNTP tương ứng đã cho vào ống đó.

d) Tiến hành điện di 4 ống phản ứng trên 4 hàng của một gel polyacrylamid để phân tách các đoạn DNA.

e) Do một trong 4 dNTP có đánh dấu phóng xạ nên khi dùng kỹ thuật phóng xạ tự ghi, các mạch đơn trên gel điện di sẽ tạo các vạch sáng trên phim X quang.

f) Tổng hợp kết quả ở cả 4 ống, thu được các mạch đơn có kích thước hơn kém nhau 1 nucleotid. Đoạn DNA ngắn nhất sẽ di chuyển xa nhất. Đoạn DNA dài nhất sẽ di chuyển gần nhất, phân tích tương tự, chúng ta sẽ biết được trình tự đoạn mạch đơn mới được tổng hợp và xác định được trình tự của mạch DNA khuôn [103].



**Hình 1.10: Quy trình giải trình tự theo phương pháp ddNTP [103]**

### 1.9.3.2. Giải trình tự bằng máy tự động

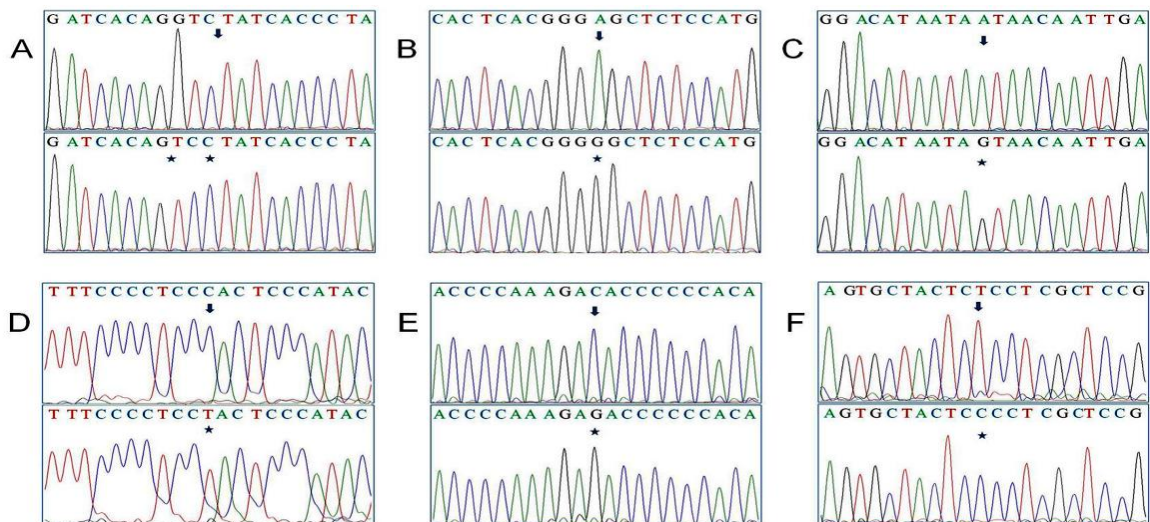
Ngày nay, việc giải trình tự được thực hiện dễ dàng nhờ có sự hỗ trợ của máy giải trình tự tự động. Máy giải trình tự hoạt động dựa theo nguyên lý của phương pháp Sanger có cải biến. Trong đó các ddNTP không được đánh dấu phóng xạ mà được đánh dấu bằng chất huỳnh quang có màu khác nhau cho mỗi loại ddNTP để các vạch điện di của các mạch đơn DNA được tổng hợp ra phát sáng khi đi qua một chùm tia sáng laser.

Cấu tạo của một máy tự động giải trình tự gồm hai phần chính yếu, đó là: phần điện di với gel polyacrylamid và phần phát hiện các vạch điện di. Phần điện di polyacrylamid có thể là một bản gel hay là một ống mao quản

chứa gel. Phần phát hiện vạch điện di là những con mắt cảm quang và một chùm tia laser đi qua trước nó.

Nguyên tắc hoạt động của máy là trong suốt quá trình điện di, mỗi khi có một vạch điện di đi qua chùm tia laser thì vạch điện di sẽ phát sáng lên và sự phát sáng này sẽ được con mắt cảm quang ghi nhận và lưu lại thành một đỉnh cường độ sáng trong biểu đồ. Từ biểu đồ của các đỉnh cường độ sáng này, máy sẽ so dòng của các đỉnh tương ứng với các màu để cuối cùng phân tích thành trình tự của đoạn DNA. Với các máy thế hệ mới sau này, người ta có thể dùng 4 màu huỳnh quang khác nhau để đánh dấu 4 loại ddNTP, nhờ vậy phản ứng giải trình tự có thể thực hiện được chỉ trong một ống nghiệm và khi giải trình tự chỉ cần điện di trên một hàng mà không cần phải trên 4 hàng khác nhau như trước đây [103].

Trình tự gen được đối chiếu và so sánh với trình tự gen trên Genbank. Phương pháp giải trình tự gen cho phép phát hiện tất cả các đột biến, đặc biệt là các đột biến điểm, do đó kỹ thuật này được áp dụng cho việc phát hiện một số SNP không có vị trí cắt enzyme giới hạn của gen.



**Hình 1.11: Hình ảnh giải trình tự một số SNP vùng HV1, HV2 mtDNA [104]**

(A) 12insC; (B) A37G; (C) A238G; (D) C463T; (E) C566G (F) T16445C

### 1.9.3.3. Giải trình tự gen thế hệ mới (next-generation sequencing -NGS)

Giải trình tự gen thế hệ mới được coi là thế hệ tiếp theo của phương pháp giải trình tự trực tiếp. Phương pháp này còn được gọi nhiều tên khác nhau như giải trình tự thông lượng cao (high-throughput sequencing), giải trình tự chuyên sâu (deep sequencing), giải trình tự thế hệ thứ 2 (second-generation sequencing). Kỹ thuật này cho phép giải trình tự đồng thời nhiều đoạn gen với tốc độ nhanh và giảm chi phí đáng kể so với việc giải trình tự theo phương pháp truyền thống.

*Nguyên lý chung của NGS gồm 3:*

- Chuẩn bị thư viện (Library preparation): Các thư viện được tạo ra bằng sự phân cắt ngẫu nhiên của DNA cần giải trình tự, sau đó gắn với các adaptor.
- Khuếch đại vô tính (Clonal Amplification): Thư viện được khuếch đại bằng PCR (emulsion PCR, bridge PCR).
- Giải trình tự (Sequencing): DNA được giải trình tự dựa trên các cách tiếp cận khác nhau (Pyrosequencing, Sequencing-by-ligation, Sequencing-by-synthesis).

Mặc dù giải trình tự là kỹ thuật tương đối mất thời gian nhưng đây vẫn là tiêu chuẩn vàng để phát hiện các đột biến và xác định các SNP. Phương pháp được sử dụng phổ biến nhất vẫn là giải trình tự trực tiếp sản phẩm PCR đoạn DNA đích bằng máy giải trình tự gen tự động. Trong những năm gần đây, chất lượng của các dữ liệu giải trình tự DNA được cải thiện đáng kể. Tuy nhiên, tình trạng xuất hiện các đỉnh không đồng đều hay những đỉnh nhiễu gây khó khăn cho việc đọc trình tự tự động. Đặc biệt là các đỉnh dị hợp tử do 2 base đa hình khác nhau cùng chiếm giữ 1 vị trí. May mắn là trình tự các đoạn DNA xác định thường giống nhau giữa các cá thể nên khi so sánh có thể xác định được đỉnh dị hợp tử hay đồng hợp tử, từ đó dễ dàng phát hiện các đột biến hay các đa hình thái của gen. Ngoài ra chiều cao của các đỉnh trong 1 mẫu hỗn hợp DNA từ nhiều cá thể có thể giúp ước tính tần suất của các alen nhóm phân tích.

## Chương 2

### ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

#### 2.1. Đối tượng nghiên cứu

- 517 người bình thường khỏe mạnh thuộc 4 dân tộc: Kinh, Mường, Chăm và Khmer: 206 mẫu người dân tộc Kinh (lấy mẫu ở Hà Nội, TP HCM), 100 mẫu người dân tộc Mường (lấy mẫu ở Hòa Bình), 113 mẫu người dân tộc Chăm (lấy mẫu ở Bình Thuận) và 98 mẫu người dân tộc Khmer (lấy mẫu ở Sóc Trăng). Lấy máu ngoại vi để phân tích nghiên cứu.

Tiêu chuẩn chọn mẫu: Là người bình thường, khỏe mạnh thuộc 1 trong 4 dân tộc Kinh, Mường, Chăm, Khmer (cụ thể có mẹ và bà ngoại là người cùng một dân tộc). Tình nguyện tham gia nghiên cứu.

Các địa điểm chọn lấy mẫu là những nơi có số dân của các dân tộc trên chiếm tỷ lệ cao (dân tộc Kinh tập trung đông nhất ở Hà Nội và TP. HCM, dân tộc Mường tập trung đông nhất ở Hòa Bình chiếm 64% dân số toàn tỉnh, dân tộc Chăm sống rải rác ở các tỉnh phía Nam trong đó riêng Bình Thuận chiếm khoảng trên 30%, dân tộc Khmer ở Sóc Trăng chiếm 30,7% dân số toàn tỉnh và chiếm khoảng 31,5% tổng số người Khmer tại Việt Nam), bản đồ thu mẫu 4 dân tộc Kinh, Mường, Chăm và Khmer (Phụ lục 1).

Bảo quản mẫu: mẫu máu sau khi lấy được bảo quản ở  $-80^{\circ}\text{C}$  cho đến khi phân tích.

- 186 bệnh nhân nữ bị ung thư vú đã được chẩn đoán xác định dựa vào kết quả giải phẫu bệnh, tại bệnh viện K Trung ương. Tiêu chuẩn loại trừ: loại trừ các trường hợp ung thư di căn từ nơi khác đến, bệnh nhân không đồng ý tham gia nghiên cứu. Mẫu máu ngoại vi được lấy và lưu giữ bảo quản tại Trung tâm Nghiên cứu Gen và Protein, Trường Đại học Y Hà Nội đến khi phân tích.

Mẫu đối chứng được chọn từ tất cả các mẫu người nữ/517 mẫu người bình thường khỏe mạnh của 4 nhóm dân tộc ở trên (có 255 mẫu nữ/517 mẫu).

Công thức tính cỡ mẫu:

$$n = Z^2_{1-\alpha/2} \frac{p(1-p)}{d^2}$$

Trong đó:

n: Cỡ mẫu cho mỗi một nhóm nghiên cứu

Z: Hệ số tin cậy (với  $\alpha = 0,05$ , độ tin cậy 95%),  $Z = 1,96$ .

p: Ước tính tỉ lệ gặp đa hình gen ty thể trong quần thể, chọn  $p = 0,5$ .

d: Độ chính xác tuyệt đối mong muốn,  $d = 0,1$ .

Áp dụng công thức trên tính được  $n=97$  (cho mỗi nhóm).

## 2.2. Thời gian và địa điểm nghiên cứu

- Thời gian nghiên cứu: Từ năm 2014 đến năm 2019

- Địa điểm nghiên cứu: Bộ môn Hóa Sinh, Trung tâm nghiên cứu Gen và Protein - Trường Đại học Y Hà Nội.

## 2.3. Phương pháp nghiên cứu

Mô tả cắt ngang, nghiên cứu bệnh có nhóm đối chứng.

## 2.4. Phương tiện nghiên cứu

### 2.4.1. Dụng cụ, trang thiết bị

- Ống lấy máu chống đông EDTA.
- Pipet, đầu côn các loại.
- Ống Eppendorf 1,5 ml và ống Facol.
- Máy Gene Amp PCR System 9700 (USA).
- Tủ lạnh sâu:  $-30^{\circ}\text{C}$ ;  $-80^{\circ}\text{C}$  (SANYO).
- Máy điện di: Mupid (Nhật Bản).

- Máy soi gel và chụp ảnh tự động: Chemidoc EQ-Bio-Rad (USA).
- Máy ly tâm lạnh Beckman và máy ly tâm để bàn Eppendorf (Đức).
- Lò vi sóng (Samsung).
- Máy đọc trình tự gen ABI Prism 3100 Genetic Analyzer (USA).
- Tủ ẩm.

#### **2.4.2. Hoá chất**

##### **\* Hóa chất dùng để tách chiết DNA (Promega, Madison, WI, USA):**

- Dung dịch Lysis buffer.
- Dung dịch K.
- Dung dịch SDS 10%.
- Proteinase K (10mg/mL).
- Dung dịch phenol: chloroform: isoamyl với tỷ lệ 25: 24: 1
- Dung dịch chloroform: isoamyl với tỷ lệ 24 : 1
- Ethanol 100% và ethanol 70%.
- Sodium acetate 3M, pH=5,2.
- Dung dịch hòa tan DNA để bảo quản.

##### **\* Hoá chất để thực hiện kỹ thuật PCR (Invitrogen, Carlsbad, California, USA):**

- 10x buffer
- Gold Taq chứa: 4 loại dNTP, Taq polymerase, MgCl<sub>2</sub>.

Mũi xuôi và mũi ngược:

HV1-F: 5' - CTC CAC CAT TAG CAC CCA AAG C -3'

HV1-R: 5' - CCT GAA GTA GGA ACC AGA TG -3'

HV2-F: 5' - GGT CTA TCA CCC TAT TAA CCA C -3'

HV2-R: 5' - CTG TTA AAA GTG CAT ACC GCC A -3'

**\* Hoá chất để điện di sản phẩm PCR trên gel agarose**

- Dung dịch đệm TBE (Tris Borate EDTA) 1X gồm: Tris base, boric acid và EDTA (pH 8,0)
- Agarose.
- Thang chuẩn DNA 100 bp (Marker 100bp).
- Dung dịch ethidium bromide 10 mg/mL.

**2.5. Kỹ thuật nghiên cứu**

**2.5.1. Tách chiết DNA từ máu ngoại vi**

DNA được tách chiết từ tế bào bạch cầu máu ngoại vi bằng phương pháp sử dụng phenol/chloroform theo quy trình chuẩn.

*Quy trình:*

- Mẫu máu đông lạnh được ủ ở bể ổn nhiệt 37°C trong 30 phút.
- Cho 0,5 ml máu toàn phần chống đông bằng EDTA vào ống Eppendorf 1,5 ml sau đó cho thêm vào 0,5 ml dung dịch Lysis buffer rồi để trên đá trong 10 phút.
- Ly tâm 8000 vòng/ phút trong 10 phút ở 4°C, bỏ dịch nổi và thu cặn. lặp lại quá trình này 4 lần.
- Cho 0,5 ml dung dịch K vào, ly tâm 8000 vòng/ phút trong 10 phút ở 4°C, bỏ dịch nổi và thu cặn.
- Cho 0,5 ml Lysis buffer; 12,5 µl SDS 10%; 10 µl Proteinase K; ủ ở 56°C trong 2÷ 3 giờ.
- Cho 0,5 mL phenol: chloroform: isoamyl, ly tâm 10000 vòng/phút trong 10 phút ở 4 °C, hỗn hợp được chia làm 3 phần:
  - Lớp dung dịch phía trên có chứa DNA.
  - Lớp ở giữa là cặn tế bào.
  - Lớp dưới cùng là dịch chiết.



Hút lấy phần dịch chứa DNA phía trên cùng và tiến hành lặp lại các bước trên một lần nữa sẽ đảm bảo không còn tạp chất trong mẫu.

- Cho 0,5 ml chloroform: isoamyl, ly tâm mẫu 10.000 vòng/phút trong 10 phút ở 4 °C. Hút lấy phần dịch trên cùng và tiến hành lặp lại một lần nữa.

- Rửa DNA bằng 1 ml ethanol 100%, cho thêm 50 µl sodium acetate, để lạnh ở -20°C trong 4 giờ.

- Ly tâm 13.000 vòng/phút trong 20 phút ở 4°C, bỏ dịch nổi, thu rửa.

- Rửa DNA bằng ethanol 70%. Rửa DNA được hoà tan bằng 50 ml nước tinh khiết hoặc TE.

Phân tử DNA thu được sẽ được kiểm tra độ tinh sạch bằng phương pháp đo mật độ quang ở bước sóng 260/280 nm và điện di trên gel agarose 1,5%.

- + Nếu  $OD_{260nm}/OD_{280nm} = 1,8 - 2$  thì DNA được coi là tinh sạch.

- + Chất lượng của DNA tốt khi điện di cho vạch rõ nét, băng gọn. Ngược lại, phân tử DNA bị đứt gãy, lẫn nhiều protein hoặc các tạp chất khác thì hình ảnh điện di là một vệt trải dài không tạo thành băng gọn.

### **2.5.2. Phương pháp quang phổ**

\* Nguyên lý đo mật độ quang học

- Acid nucleic có phổ hấp thụ cực đại ở bước sóng 260 nm là do sự có mặt của base purin và base pyrimidin. Giá trị mật độ quang ở bước sóng 260nm ( $OD_{260nm}$ ) của các mẫu cho phép xác định nồng độ acid nucleic có trong mẫu nghiên cứu.

- Protein hấp thụ ánh sáng tử ngoại ở bước sóng 280nm do sự hấp thụ của các acid amin nhân thơm và dị vòng, bao gồm: tyrosin, tryptophan và phenylalanin.

- Dựa vào tỷ lệ  $OD_{260nm}/OD_{280nm}$  để kiểm tra độ tinh sạch của DNA. Nếu  $OD_{260nm}/OD_{280nm} = 1,8 \div 2$  thì DNA được coi là tinh sạch.

\* *Tiến hành đo mật độ quang học*

Sử dụng 2 $\mu$ l DNA xác định nồng độ và độ tinh sạch bằng phương pháp đo OD trên máy Nanodrop 1000.

### **2.5.3. Điện di DNA sau tách chiết**

*Nguyên lý:*

Ở pH =8, acid nucleic mang điện tích âm; dưới tác dụng của dòng điện một chiều, acid nucleic sẽ di chuyển về cực dương. Các đoạn DNA có kích thước khác nhau sẽ di chuyển với tốc độ khác nhau. Tùy vào mục đích có thể sử dụng các chất giá khác nhau để điện di acid nucleic, nhưng phổ biến nhất là sử dụng gel agarose. Thông thường người ta sử dụng nồng độ gel là 1,5%. Bản gel điện di được nhuộm bằng ethidium bromide sẽ phát sáng.

Chất lượng của DNA tốt khi vạch điện di rõ nét, băng gọn. Ngược lại, phân tử DNA bị đứt gãy, lẫn nhiều protein hoặc các tạp chất khác thì hình ảnh điện di là một vệt trải dài không tạo thành băng gọn.

*Cách chuẩn bị gel agarose 1,5%*

- Cân 1,5 g agarose hoà tan trong 10 mL boric acid EDTA (TBE). Sử dụng lò vi sóng để agarose tan, sau khi agarose tan hết để nguội 55-60°C, đổ vào khuôn gel. Để bản gel đông lại và ổn định hoàn toàn mới sử dụng.

- Gỡ lược, đặt bản gel vào bể điện di chứa sẵn đệm TAE 1X sao cho đệm ngập hoàn toàn bản gel.

- Pha dung dịch TBE 10X (Tris; acid boric; EDTA).

Tris 0,89 M; acid boric 0,89 M; EDTA 2,02 M

*Tiến hành kỹ thuật điện di*

- Đổ bản gel agarose 1,5% với 6 hoặc 12 giếng rồi ngâm vào bể điện di.
- Nạp 5 µl sản phẩm khuếch đại PCR vào các giếng trên bản gel.
- Nạp 2,5 µl Marker loại 100 bp vào 1 giếng còn lại.
- Sử dụng máy điện di 80-100V (Mupid- Nhật), điện di ở hiệu điện thế 100V khoảng 30 phút.
- Sau điện di, gel được ngâm vào ethidium bromide 5 phút, rửa qua nước cất và đưa vào soi dưới đèn tử ngoại, chụp ảnh dưới ánh sáng tử ngoại, nhận định hình ảnh và chụp hình lưu trữ.

**2.5.4. Phản ứng PCR (polymerase chain reaction) nhân đoạn gen HV1, HV2**

*Quy trình:* DNA sau khi được tách chiết và kiểm tra nồng độ, độ tinh sạch đạt yêu cầu được sử dụng làm khuôn mẫu để khuếch đại đoạn gen HV1 và HV2 bằng máy PCR Eppendorf. Tất cả các mẫu DNA được tách chiết đều được pha loãng về nồng độ khoảng 40 ng/µl để thực hiện phản ứng PCR.

- Thành phần của phản ứng PCR: 10X đệm PCR; 2,5mM dNTP; môi xuôi và ngược; 0,5unit *Taq* polymerase; DNA và H<sub>2</sub>O, tổng thể tích 20µl.

<b>Thành phần</b>	<b>Thể tích (µl)</b>
Nước cất PCR	11,9
Buffer 10X	2,0
dNTP (2,5 mM)	2,0
Môi xuôi	0,5
Môi ngược	0,5
<i>Taq</i> polymerase (5 u/µl)	0,1
DNA	3,0
<b>Tổng</b>	<b>20,0</b>

- Chu trình nhiệt của phản ứng PCR: 94°C - 5 phút; 30 chu kỳ [94°C- 30 giây, 54°C - 30 giây, 72°C - 30 giây]; 72°C - 5 phút; bảo quản mẫu ở 15°C.

- Sản phẩm PCR được điện di cùng với thang chuẩn 100bp trên gel agarose 1,5%. Các băng DNA được nhuộm ethidium bromide và chụp ảnh bằng hệ thống máy EC3 Imaging system. Phản ứng PCR đạt yêu cầu khi chỉ có 1 băng duy nhất với mỗi đường chạy, băng rõ nét, kích thước so trên thang DNA chuẩn tương ứng với kích thước tính toán lý thuyết.

### **2.5.5. Giải trình tự vùng *HVI* và *HV2* (DNA sequencing)**

- Tinh sạch DNA sử dụng Kit QIAGEN (Hilden, Germany).

- Giải trình tự gen theo quy trình và sử dụng phương pháp BigDye terminator sequencing (Applied Biosystems, Foster city, USA). Tiến hành trên máy 3100-Avant Genetic Analyzer của hãng ABI-PRISM. So sánh với trình tự GenBank để xác định đa hình.

#### ***Quy trình thực hiện:***

##### ***Tinh sạch DNA (sử dụng Kit QIAGEN)***

- Cho 500 µl dung dịch PB vào ống chứa 10 µl DNA, lắc đều để ở nhiệt độ phòng 30 phút.

- Hút hết dịch cho qua cột tinh sạch (do hãng Qiagen cung cấp).

- Quay ly tâm 10000 v/p trong 30 giây, đổ bỏ dịch ở đáy ống.

- Cho tiếp 750 µl PE vào cột.

- Ly tâm 10000 v/p trong 30 giây, đổ bỏ dịch phía đáy ống.

- Ly tâm tiếp thêm 60 giây để loại bỏ hết dịch trong cột.

- Lấy cột ra, cho vào ống eppendorf 1,5 mL.

- Cho 50 µl dung dịch EB (gồm 10 mM Tris-Cl, pH 8,5).

- Để ở nhiệt độ phòng trong 5 phút, ly tâm 10000 v/p trong 60 giây.
- Dịch thu được ở đáy ống là dịch chứa DNA đã được tinh sạch.

*Quy trình thực hiện giải trình tự gen:*

Thực hiện theo qui trình và sử dụng phương pháp BigDye terminator sequencing (Applied Biosystems, Foster city, USA).

+ *Giai đoạn 1:* Chuẩn bị master mix cho phản ứng PCR.

- Chuẩn bị effendorf 0,2ml đã đánh dấu sẵn thứ tự các mẫu.
- Chuẩn bị hóa chất để thực hiện phản ứng.
- Làm tan hoàn toàn hóa chất, trộn đều sau đó ly tâm nhẹ để toàn bộ dịch trên nắp ống rơi xuống.
- Tiến hành pha master mix theo bảng sau:

<b>Thành phần</b>	<b>Thể tích (μl)</b>
Sản phẩm sau PCR đã được tinh sạch	1,0
Big Dye Buffer 1X	3,0
Big Dye terminator V3.1 (2,5X)	2,0
Nước cất PCR	13,0
Mồi đơn (5pmol/μl)	1,0
<b>Tổng</b>	<b>20</b>

*Chú ý:*

- Toàn bộ khâu chuẩn bị master mix phải được thực hiện trên khay đá.
- Các hóa chất phải được làm tan và trộn đều trước khi sử dụng.
- Big dye 2,5X phải được bảo quản tránh ánh sáng.
- Mồi DNA thực hiện hai phản ứng với mồi xuôi và mồi ngược.

+ *Giai đoạn 2: Chạy PCR Sequencing.*

- Sau khi chuẩn bị master mix cho phản ứng xong, ly tâm nhanh các ống PCR để toàn bộ dịch dính trên thành và nắp ống xuống dưới và làm tan bọt.

- Xếp các ống effendorf vào máy PCR.

- Chọn chương trình nhiệt đã được cài đặt sẵn trong máy theo chu trình đã được tối ưu hóa.

- Kiểm tra lại toàn bộ chu trình nhiệt:

<b>Chu trình</b>	<b>Biến tính</b>	<b>Bắt cặp</b>	<b>Tổng hợp</b>
1	95°C - 2 phút		
2 – 25	95°C - 5 giây	50°C - 10 giây	60°C - 4 phút
Bảo quản ở 10°C			

- Ấn nút “Start” cho máy bắt đầu chạy chương trình.

- Sau khi chạy xong chương trình đưa máy về chế độ nghỉ và tắt máy

- Lấy mẫu ra khỏi máy PCR

- Các sản phẩm sau khi được khuếch đại bằng Big dye kit sẽ được tinh sạch bằng big Dye temination để loại bỏ toàn bộ big dye thừa và đem đọc trên máy giải trình tự gen ABI 3100 vant (Applied Biosystem)

- Các nucleotid trên gen sẽ được biểu hiện bằng các đỉnh (peak) với 4 màu tương đương với 4 loại nucleotid A,T,G,C.

Kết quả giải trình tự gen được đối chiếu và so sánh với trình tự chuẩn J01415 trên GenBank (National center for biotechnology information, NCBI) để xác định đa hình trên hai vùng HV1 và HV2 của DNA ty thể.

### ***2.5.6. Phương pháp phân tích trình tự đoạn HV1 và HV2***

Giải trình tự gen từ vị trí 15.974 đến vị trí 16.517 có chứa trình tự vùng HV1 và từ vị trí 8 đến vị trí 431 có chứa trình tự vùng HV2. Vùng HV1 được xác định là từ vị trí 16.024 đến 16.365 và trình tự cho vùng HV2 được xác định từ vị trí 73 đến 340. Mỗi mẫu được giải trình tự theo cả hai hướng (5' và 3') để tránh nhiễu và xác định trình tự gen cũng như phát hiện các điểm đa hình là đầy đủ nhất.

Trình tự vùng HV1 và HV2 mtDNA của các mẫu nghiên cứu được so sánh với trình tự chuẩn Cambridge (rCRS) đã được công bố trên ngân hàng dữ liệu về DNA ty thể (<http://mitomap.org>) bằng phần mềm sinh học chuyên dụng (CLC Main Workbench 6.0.1). Đồng thời cũng so sánh các đặc điểm đa hình của các mẫu nghiên cứu với các đặc điểm đa hình đã được công bố trên MITOMAP (A Human Mitochondrial Genome Database).

Xác định độ đa dạng di truyền (genetic diversity) và xác suất trùng lặp ngẫu nhiên giữa hai cá thể trong quần thể nghiên cứu được tính theo công thức:  $h = (1 - \sum x^2) / n(n-1)$ ; trong đó:  $h$  là độ đa dạng di truyền,  $\sum x^2$  là xác suất trùng lặp ngẫu nhiên giữa hai cá thể trong quần thể,  $x$  là tần suất xuất hiện của haplotype trong quần thể và  $n$  là số mẫu, theo Fumio Tajima năm 1989 [105].

### ***2.5.7. Phân tích mối liên quan giữa một số đa hình đơn nucleotid (SNP) trên vùng HV1 với bệnh ung thư vú***

Phương pháp thống kê và kiểm định Chi-Square ( $\chi^2$ ) trên phần mềm SPSS 12.0.1 được sử dụng để đánh giá tỷ lệ các SNP và mối tương quan giữa chúng với ung thư vú.

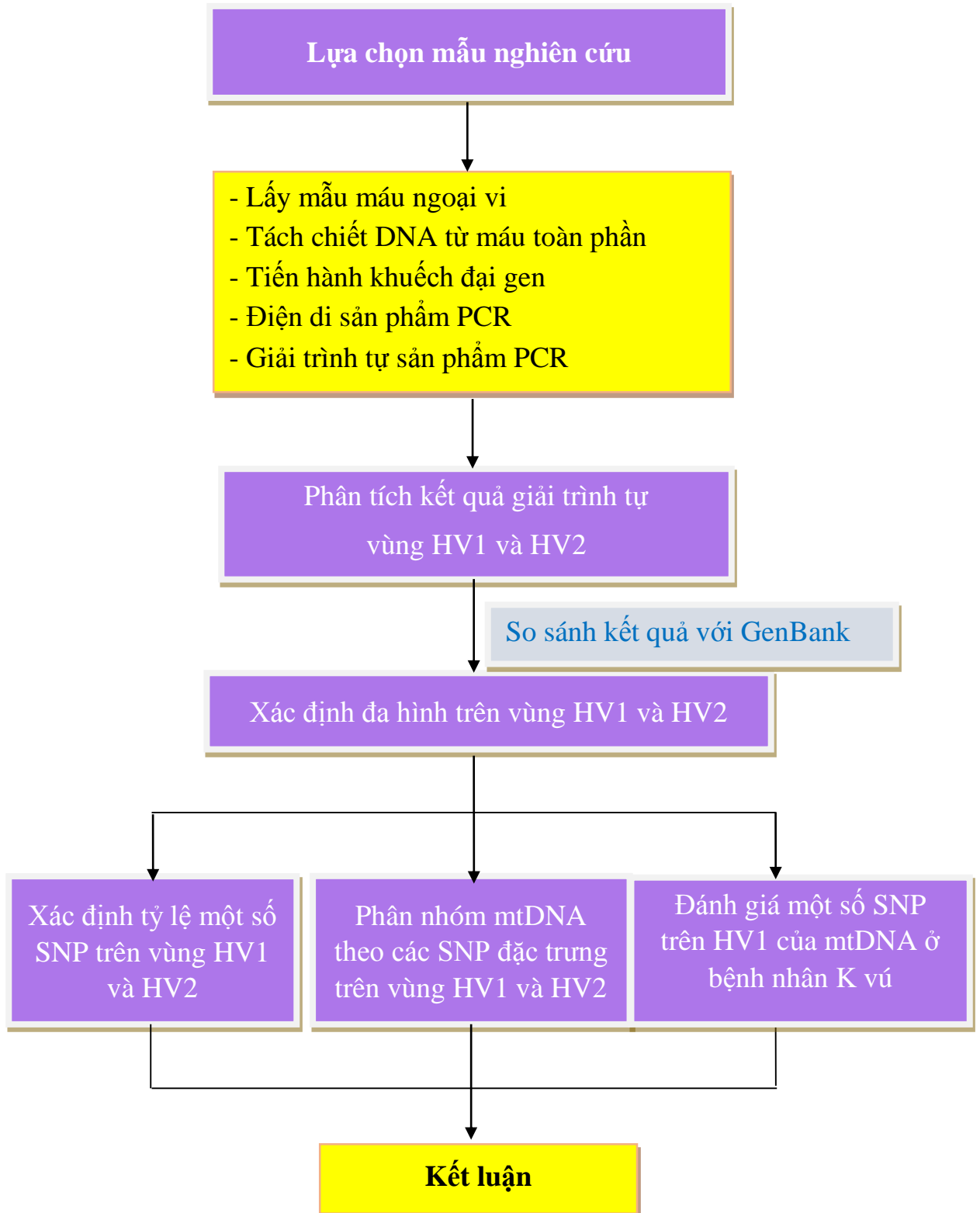
Xác định mối liên quan giữa một số SNP của vùng HV1 mtDNA bằng phân tích tương quan. Ước tính nguy cơ gây bệnh được biểu thị bằng tỉ số odds (Odds ratio-OR) và khoảng tin cậy 95% (95% confidence interval-95% CI).

## **2.6. Vấn đề đạo đức của đề tài**

- Các đối tượng tham gia nghiên cứu là hoàn toàn tự nguyện và có quyền rút khỏi nghiên cứu khi không muốn tham gia nghiên cứu.
- Các thông tin liên quan đến người tham gia nghiên cứu được đảm bảo bí mật.
- Các kỹ thuật thao tác trên người tham gia nghiên cứu được đảm bảo đúng chuyên môn.
- Đề tài nghiên cứu này được thực hiện hoàn toàn vì mục đích khoa học chứ không vì mục đích nào khác.
- Đề tài đã được chấp thuận thông qua Hội đồng đạo đức trong nghiên cứu y sinh học Trường Đại học Y Hà Nội.



## 2.7. Sơ đồ nghiên cứu



## Chương 3

### KẾT QUẢ NGHIÊN CỨU

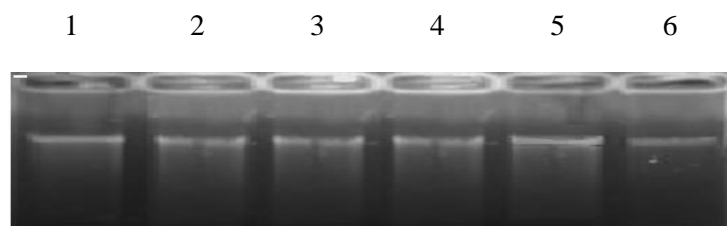
#### 3.1. Kết quả giải trình tự gen vùng HV1, HV2 trên DNA ty thể

##### 3.1.1. Tách chiết DNA tổng số

Bước đầu tiên trong hầu hết các nghiên cứu sinh học phân tử là thu nhận được DNA tổng số với một lượng đủ lớn và tinh sạch DNA để dùng cho các thí nghiệm tiếp sau.

Sau khi tách chiết, tinh sạch, các mẫu DNA được kiểm tra nồng độ và độ tinh sạch bằng cách đo phổ hấp thụ tử ngoại ở hai bước sóng 260 nm, 280 nm trên máy Nanodrop 1000. Các mẫu DNA được tách chiết có chỉ số A260/280 dao động trong khoảng 1,81- 2.02 với nồng độ dao động trong khoảng từ 37,0 ng/μl đến 295,0 ng/μl. Các mẫu DNA đều có đường biểu diễn độ hấp thụ quang trơn tru, không gãy khúc, gập góc, đỉnh hấp thụ tương ứng với bước sóng 260 nm.

Kết quả điện di DNA tổng số từ các mẫu máu được thể hiện trong hình sau:



**Hình 3.1: Ảnh điện di DNA tổng số trên gel agarose 1,5%**

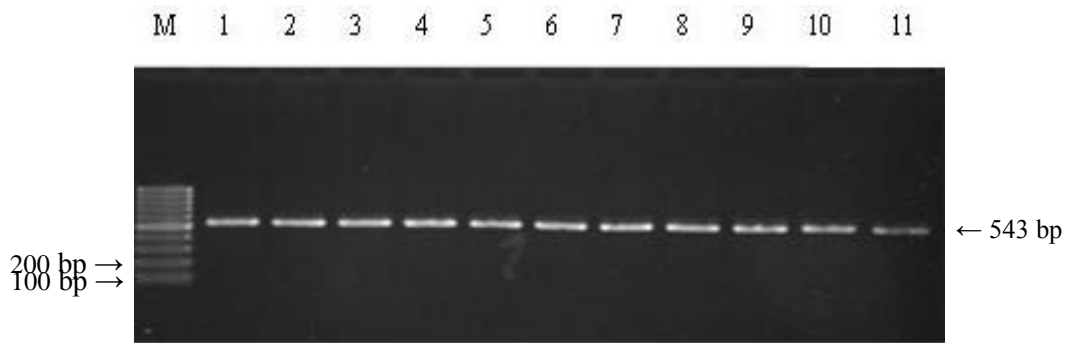
Giếng 1, 2, 3: DNA tổng số tách từ mẫu M10, M12, M14

Giếng 4, 5, 6: DNA tổng số tách từ mẫu KN8, KN9, KN10

Nhận xét: tất cả các đường chạy đều xuất hiện một băng DNA sáng rõ nét, không đứt gãy. Các mẫu DNA được tách chiết tương đối tinh sạch và nồng độ đạt yêu cầu cho phản ứng PCR.

### 3.1.2. Kết quả khuếch đại đoạn gen HV1, HV2 của DNA ty thể

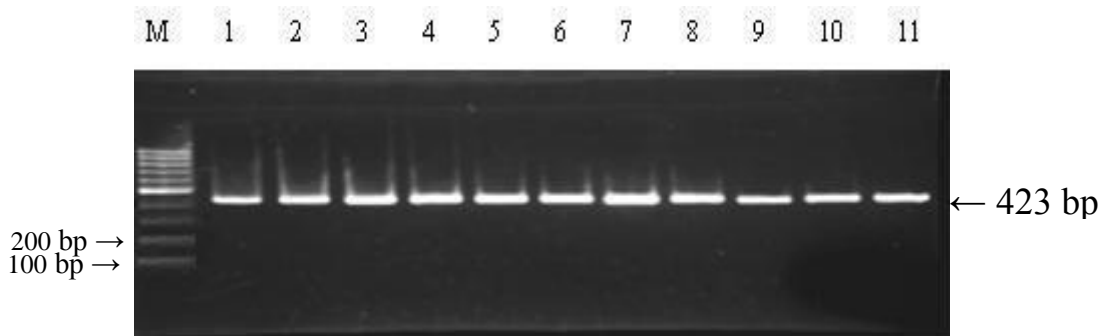
Sử dụng hai cặp môi đặc hiệu với quy trình tối ưu để khuếch đại hai vùng gen HV1 và HV2 sản phẩm khuếch đại được điện di trên gel agarose 1,5%. Kết quả điện di được thể hiện như trong hình sau:



**Hình 3.2: Hình ảnh điện di sản phẩm PCR của vùng HV1**

*M: Thang chuẩn DNA 100bp; Giếng 1-11: Sản phẩm khuếch đại vùng HV1*

Nhận xét: đã khuếch đại được một đoạn gen có kích thước 543bp (từ vị trí 15975-16517 của DNA ty thể) có chứa vùng siêu biến HV1. Sản phẩm PCR đặc hiệu, rõ nét ở tất cả các mẫu nghiên cứu, chỉ gồm một băng có kích thước phù hợp với các tính toán lý thuyết. Như vậy, chúng tôi đã khuếch đại thành công vùng HV1.



**Hình 3.3: Hình ảnh điện di sản phẩm PCR của vùng HV2**

*M: Thang chuẩn DNA 100bp; Giếng 1-11: Sản phẩm khuếch đại vùng HV2*

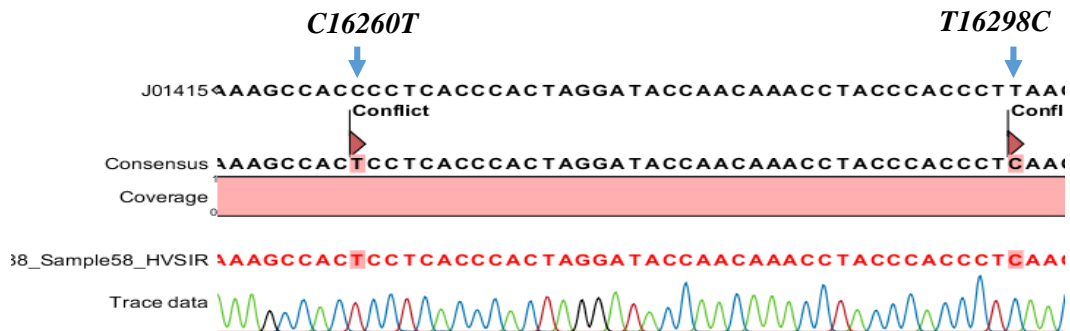
Nhận xét: đã khuếch đại được một đoạn gen có kích thước 423bp (từ vị trí 8-431 của DNA ty thể) có chứa vùng siêu biến HV2. Ảnh điện di cho thấy sản phẩm PCR đặc hiệu, rõ nét ở tất cả các mẫu nghiên cứu, chỉ gồm một băng có kích thước phù hợp với các tính toán lý thuyết. Như vậy, chúng tôi đã khuếch đại thành công vùng HV2.

### 3.1.3. Kết quả giải trình tự vùng HV1, HV2 của DNA ty thể trên 4 dân tộc Kinh, Chăm, Mường, Khmer người Việt Nam

Trong nghiên cứu này, sản phẩm PCR của vùng HV1 và HV2 của DNA ty thể được tiến hành giải trình tự gen trên máy 3100-Avant Genetic Analyzer của hãng ABI-PRISM. So sánh với trình tự chuẩn trên GenBank (trình tự chuẩn gen ty thể - J01415) để xác định đa hình.

Một số kết quả giải trình tự gen hai vùng siêu biến HV1 và HV2 được thể hiện ở các hình sau:

a) Đa hình trên vùng HV1 của DNA ty thể:

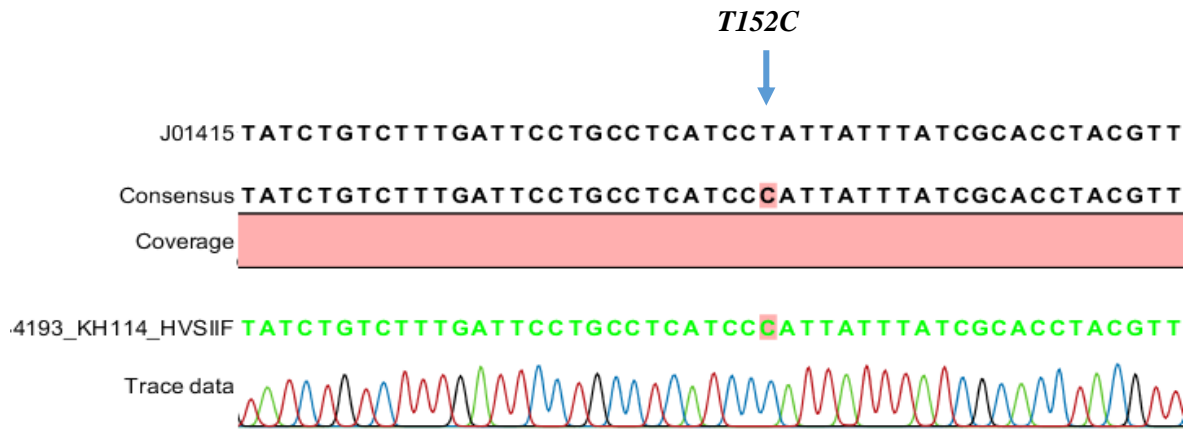


**Hình 3.4: Hình ảnh giải trình tự SNP C16260T và SNP T16298C trên vùng HV1 của DNA ty thể**

Nhận xét: hình ảnh trên là đa hình đơn nucleotid vùng HV1 (SNP C16260T và SNP T16298C) của DNA ty thể ở một mẫu nghiên cứu. Hình ảnh cũng cho thấy tín hiệu các đỉnh cao, rõ, không bị nhiễu, tín hiệu đường nền thấp.

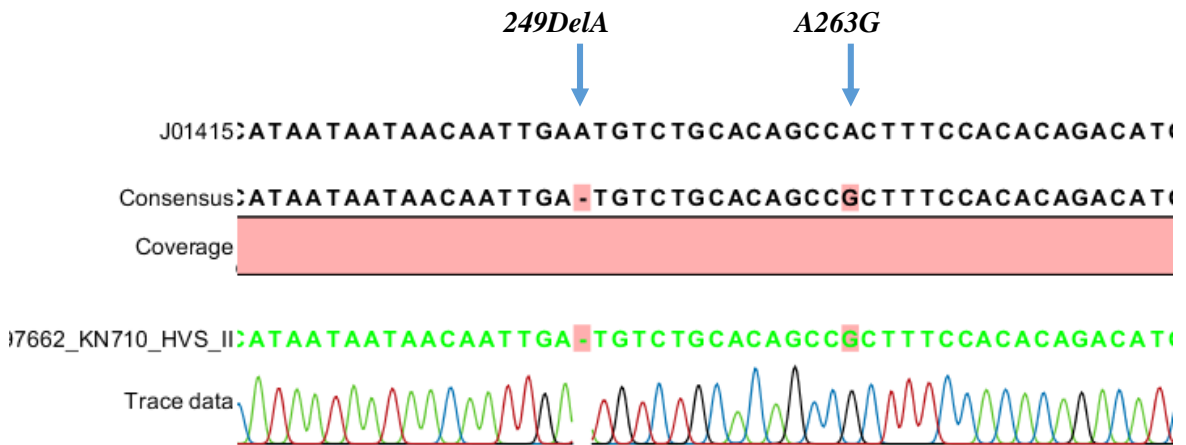


b) Đa hình trên vùng HV2 của DNA ty thể:



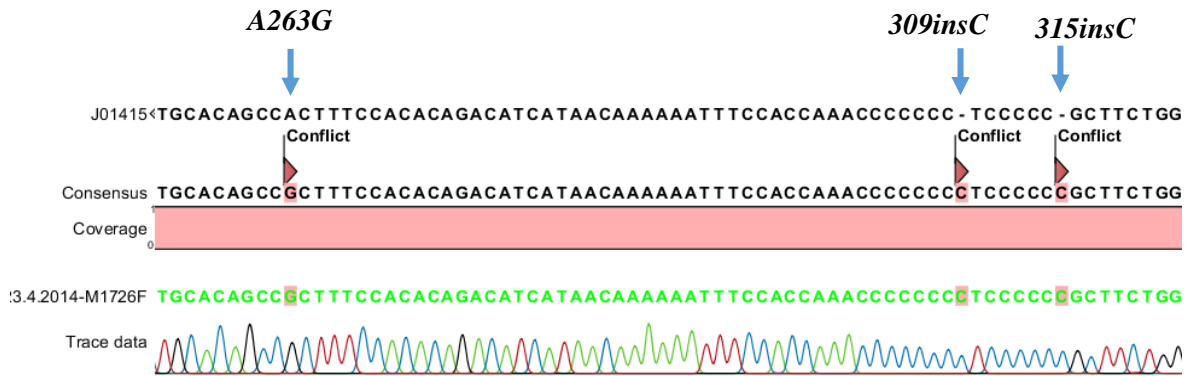
**Hình 3.7: Hình ảnh giải trình tự SNP T152C trên vùng HV2 của DNA ty thể**

Nhận xét: Kết quả giải trình tự SNP T152C trên vùng HV2 của DNA ty thể, hình ảnh trên cho thấy tín hiệu các đỉnh cao rõ không bị nhiễu và tín hiệu đường nền thấp.



**Hình 3.8: Hình ảnh giải trình tự SNP (249DelA, A263G) trên vùng HV2 của DNA ty thể**

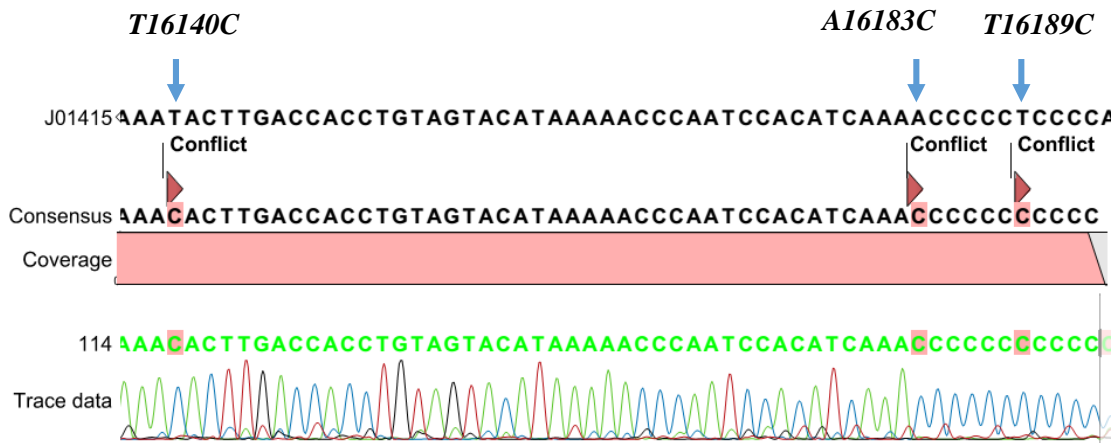
Nhận xét: Kết quả giải trình tự vùng gen ty thể HV2 ở mẫu nghiên cứu trên thấy đa hình tại vị trí 249 và 263 (SNP 249DelA và A263G).



**Hình 3.9: Hình ảnh giải trình tự SNP (A263G, 309insC, 315insC) trên vùng HV2 của DNA ty thể**

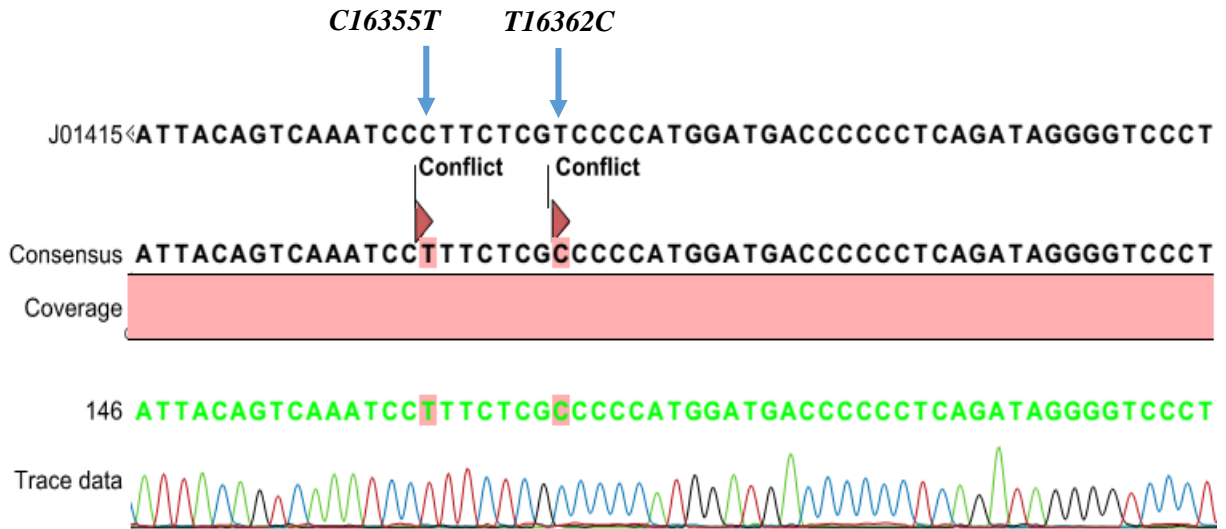
**Nhận xét:** Ở mẫu nghiên cứu này khi giải trình tự vùng HV2 thấy có đa hình tại vị trí 263 (A-G), đa hình tại vị trí 309 (thêm C) và đa hình tại vị trí 315 (thêm C).

### **3.1.4. Kết quả giải trình tự vùng HV1 trên DNA ty thể ở bệnh nhân ung thư vú người Việt Nam**



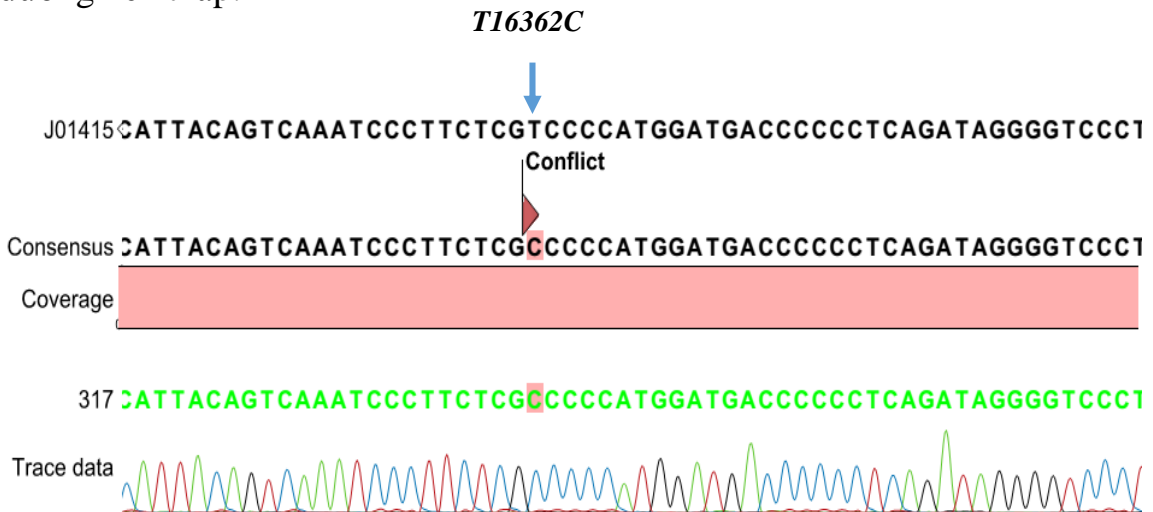
**Hình 3.10: Hình ảnh giải trình tự SNP (T16140C, A16183C, T16189C) trên vùng HV1 của DNA ty thể ở bệnh nhân ung thư vú**

**Nhận xét:** Kết quả giải trình tự vùng gen ty thể HV1 ở mẫu bệnh nhân ung thư vú ở trên thấy đa hình T16140C, A16183C, T16189C.



**Hình 3.11: Hình ảnh giải trình tự SNP (C16355T, T16362C) trên vùng HV1 của DNA ty thể ở bệnh nhân ung thư vú**

Nhận xét: hình ảnh trên là đa hình đơn nucleotid vùng HV1 (SNP C16355T và SNP T16362C) của DNA ty thể ở một mẫu nghiên cứu bệnh nhân ung thư vú. Hình ảnh cũng cho thấy tín hiệu các đỉnh cao, rõ, không bị nhiễu, tín hiệu đường nền thấp.



**Hình 3.12: Hình ảnh giải trình tự SNP T16362C vùng HV1 của DNA ty thể trên một bệnh nhân ung thư vú**

Nhận xét: Kết quả giải trình tự vùng HV1 thấy đa hình tại vị trí 16362 (T-C) của một mẫu nghiên cứu bệnh nhân ung thư vú cụ thể.



### 3.2. Kết quả phân tích đa hình vùng HV1 và HV2 trên DNA ty thể

#### 3.2.1. Đa hình vùng HV1, HV2 trên DNA ty thể ở 4 dân tộc người Việt Nam (Kinh, Chăm, Khmer và Mường) được đối chiếu với trình tự chuẩn

<b>HV1</b>						
16021	CTGTTCTTTC	ATGGGGAAGC	AGATTTGGGT	ACCACCCAAG	TATTGACTCA	CCCATCAACA
		G-	A	G C	G TC T C	
16081	ACCGCTATGT	ATTCGTACA	TTACTGCCAG	CCACCATGAA	TATTGTACGG	TACCATAAAT
	C C	TCC T	C T T	T CA C	C C C A	C C
				G		A A
				A		
16141	ACTTGACCAC	CTGTAGTACA	TAAAAACCCA	ATCCACATCA	AAACCCCCTC	CCCATGCTTA
	A TTCT	AC C	AGGG GTT	C TC CA	CCCATTT C	TTCC G
		G	C		-T +C	+C
			-			
16201	CAAGCAAGTA	CAGCAATCAA	CCCTCAACTA	TCACACATCA	ACTGCAACTC	CAAAGCCACC
	T	G C	AT CTG	T TC CG	CTGTG	TG CTCAT CT TA TT
						A
16261	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA	CAGTACATAG	TACATAAAGC
	TTC A	GT CG AGA TA	A G TTCGT	TTTTT CCG	TG CT GA C	C AT
		G		G		
	T			+C		
16321	CATTTACCGT	ACATAGCACA	TTACAGTCAA	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC
	C CC TT	G G	GT T	CTTTCC	C T C	
16381	TCAGATAGGG	GTCCCTTGAC	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAAGAGTGCT
	C	A A	GT		G	A
16441	ACTCTCCTCG	CTCCGGGCC	ATAACACTTG	GGGGTAGCTA	AAGTGAAGT	TATCCGACAT
			T G	-		AG
			-			
16501	CTGGTTCCTA	CTTCAGGGTC	ATAAAGCCTA	AATAGCCCAC	ACGTTCCCCT	TAAATAAGAC
	AC TCA	AAT				
<b>HV2</b>						
1	GATCACAGGT	CTATCACCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT	TTGGTATTTT
					T	A CCC
						G
61	CGTCTGGGGG	GTATGCACGC	GATAGCATTG	CGAGACGCTG	GAGCCGGAGC	ACCCTATGTC
	AACT A	G	C	GAC	A	
121	GCAGTATCTG	TCTTTGATTC	CTGCCTCATC	CTATTATTTA	TCGCACCTAC	GTTCAATATT
	C C C		A C T	TCG AC		G
181	ACAGGCGAAC	ATACTTACTA	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA
	TG A G	G-C TCG	GCC A G	GC C	A G	GG G
		T				
241	ACAATTGAAT	GTCTGCACAG	CCACTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACCA

	G	-	A	GA	G	T	T	G	T	--C-	G
			G							-	
301	AACCCCCCT	CCCCCGCTTC	TGGCCACAGC	ACTTAAACAC	ATCTCTGCCA	AACCCCAAAA					
	CA	T-C	T	+CATCG	G	GA	T	-	-	G-	
		+C-		C							
		+CC		+CC							
361	ACAAAGAACC	CTAACACCAG	CCTAACCCAGA	TTTCAAATTT	TATCTTTTGG	CGGTATGCAC					
	G	G		G						A	
	-			-							
421	TTTTAACAGT	CACCCCCCAA	CTAACACATT	ATTTTCCCCT	CCCACTCCCA	TACTACTAAT					
481	CTCATCAATA	CAACCCCCGC	CCATCCTACC	CAGCACACAC	ACACCGCTGC	TAACCCCA					

**Hình 3.13: Hình trình bày toàn bộ các vị trí đa hình và loại đa hình trên vùng HV1, HV2 của DNA ty thể ở bốn dân tộc Kinh, Chăm, Khmer, Mường người Việt Nam được đối chiếu với trình tự chuẩn**

*Chú thích: Các vị trí đa hình trên hai vùng HV1 và HV2 của DNA ty thể được so sánh và đối chiếu với trình tự chuẩn Cambridge đã sửa đổi. Trình tự chuẩn của DNA ty thể được đánh số vị trí ở đầu dòng và được in thường. Các vị trí nucleotid và các dạng thay đổi so với trình tự chuẩn được in nghiêng ngay phía dưới trình tự chuẩn, xóa nucleotid (-), thêm nucleotid (+).*

**Nhận xét:** Hình trên cho thấy có nhiều đa hình trên hai vùng HV1, HV2 của DNA ty thể được xác định: có 172 vị trí đa hình trên vùng HV1 và 89 vị trí đa hình trên vùng HV2, phần lớn là đa hình thay thế nucleotid, ngoài ra còn có đa hình thêm nucleotid, đa hình mất nucleotid. Một số vị trí có nhiều loại đa hình như ở vị trí 16166 trên vùng HV1 có 3 loại đa hình A16166G, A16166C, 16166DelA, vị trí 316 trên vùng HV2 có 2 loại đa hình G316A, G316C.

**Bảng 3.1: Các dạng SNP trên vùng HV1 và HV2 của DNA ty thể ở 517 mẫu thuộc 4 dân tộc Kinh, Mường, Khmer, Chăm người Việt Nam**

Loại đa hình	Vị trí
Vùng HV1	
C-stretch polymorphism	A16181C, A16182C, A16183C, 16188insC, T16189C, 16193insC
Transitions (tổng số 136 loại SNP)	A16037G, G16042A, A16051G, C16069T, C16071T, T16075C, T16086C, T16092C, T16093C, C16095T, T16102C, C16104T, C16108T, C16111T, T16124C, T16126C, G16129A, T16136C, T16140C, G16145A, C16147T, C16148T, C16150T, G16153A, T16154C, T16157C, A16162G, A16163G, A16164G, A16166G, C16167T, C16168T, T16172C, C16174T, T16178C, C16184T, C16185T, C16186T, C16187T, T16189C, C16192T, C16193T, T16195C, C16201T, A16207G, T16209C, G16213A, C16214T, T16217C, C16218T, A16219G, C16221T, C16223T, T16224C, A16227G, T16231C, C16232T, A16233G, C16234T, A16235G, C16239T, A16240G, C16242T, T16243C, G16244A, C16245T, T16249C, C16250T, C16256T, C16259T, C16260T, C16261T, C16262T, T16263C, C16266T, A16269G, C16270T, T16271C, A16272G, G16274A, A16275G, C16278T, A16284G, C16286T, C16287T, T16288C, A16289G, C16290T, C16291T, C16292T, A16293G, C16294T, C16295T, T16297C, T16298C, A16299G, A16300G, C16301T, A16302G, T16304C, A16309G, G16310A, T16311C, G16319A, C16320T, T16324C, T16325C, C16327T, C16328T, A16335G, A16339G, A16343G, C16344T, C16348T, T16352C, C16353T, C16354T, C16355T, T16356C, T16357C, T16362C, C16365T, T16381C, G16390A, G16391A, A16399G, C16400T, G16438A, C16465T, G16496A, A16497G, T16512C, C16514T, G16516A, G16518A, C16520T
Transversions	A16054C, C16067G, A16070C, G16084C, A16091T, C16111G, C16111A, A16113C, C16114A, A16116C, A16122C, T16136A,

(tổng số 43 loại SNP)	T16140A, A16149C, T16157G, T16161A, A16166C, A16175C, C16179A, A16181C, A16182C, A16183C, C16184A, A16194C, C16197G, A16226C, A16241C, C16257A, C16259A, C16266A, C16266G, T16276A, C16279A, C16282A, A16293T, A16305T, A16317C, A16322C, A16367C, C16426G, C16511A, A16515C, T16519A
Del	16038DelA, 16166DelA, 16183DelA, 16469DelT, 16474DelG
<b>Vùng HV2</b>	
C-stretch polymorphism	309insC, 309 insCC, 315insC, 315insCC
Transitions (tổng số 72 loại SNP)	C43T, G53A, T55C, A56G, T57C, G62A, T63C, C64T, G66A, A73G, T89C, A93G, G94A, G103A, T125C, T127C, T131C, G143A, T146C, C150T, C151T, T152C, A153G, T159C, A178G, C182T, A183G, G185A, A189G, A193G, C194T, T195C, C198T, T199C, A200G, A202G, T204C, G207A, A210G, A214G, T217C, G225A, A227G, A234G, A235G, A237G, A244G, A249G, G251A, A259G, G260A, A263G, C271T, C273T, A281G, C285T, T293C, A297G, C308T, T310C, C311T, G316A, T317C, T318C, A326G, A328G, G329A, C332T, A351G, A368G, A374G, A385G,
Transversions (tổng 11 SNP)	A56C, C61A, A95C, T146A, T158A, G203C, A215C, A302C, C303A, G316C, T319G
Del	194DelC, 249DelA, 291DelA, 292DelT, 293DelT, 294DelT, 309DelC, 310DelT, 334DelT, 337DelA, 352DelA, 368DelA, 385DelA

*Chú thích: Transition: đa hình thay thế nucleotid có cùng gốc purin (A-G) hoặc pyrimidin (C-T); Transversion: đa hình thay thế nucleotid có gốc purin thành pyrimidin hoặc ngược lại (A-T, C-G); C-stretch polymorphism: đa hình vùng lặp nucleotid Cystein; Del: đa hình xóa nucleotid, ins: đa hình thêm nucleotid.*

**Nhận xét:** bảng 3.1 cho thấy tính đa hình cao của hai vùng HV1, HV2 mtDNA ở 4 dân tộc Kinh, Chăm, Mường, Khmer người Việt Nam. Các đa hình thường gặp là đa hình thay thế nucleotid trong đó chủ yếu gặp đa hình thay thế đồng hoán (transition) -thay thế cùng gốc purin (A-G) hoặc pyrimidin (C-T). Trên vùng HV1 có 136 SNP là thay thế đồng hoán (transition) và 43 SNP là thay thế dị hoán (transversion). Trên vùng HV2 có 72 SNP là thay thế đồng hoán và 11 SNP là thay thế dị hoán. Ngoài ra còn gặp đa hình mất nucleotid (5 SNP trên vùng HV1, 13 SNP trên vùng HV2), đa hình thêm nucleotid (2 SNP trên vùng HV1 và 4 SNP trên vùng HV2), nucleotid thêm vào thường là nucleotid C trong khi nucleotid mất đi thường là nucleotid A. Như vậy, nghiên cứu đã phát hiện được 286 loại SNP, đa hình thay thế nucleotid chiếm 91,6% (262/286), đa hình thêm nucleotid chiếm 2,1% (6/286) và đa hình mất nucleotid chiếm 6,3% (18/286).

**Bảng 3.2: Bảng các vị trí trên vùng HV1 có nhiều hơn một loại đa hình**

STT	Vị trí trên vùng HV1	Thay đổi nucleotid		
1	16136	T-C	T-A	
2	16140	T-C	T-A	
3	16157	T-C	T-G	
4	16183	A-C	DelA	
5	16184	C-A	C-T	
6	16259	C-T	C-A	
7	16266	C-A	C-G	C-T
8	16111	C-T	C-G	C-A
9	16166	A-G	A-C	DelA
10	16293	A-G	A-T	insC

**Nhận xét:** Trên vùng HV1 thấy 10 vị trí có từ hai loại đa hình trở lên trong đó 4 vị trí có 3 loại đa hình (vị trí 16111, 16166, 16266, 16293). Như vậy, có thể thấy đây là đoạn DNA có tính đa hình cao, nhiều điểm nóng đột biến.

**Bảng 3.3: Bảng các vị trí trên vùng HV2 có nhiều hơn một loại đa hình**

STT	Vị trí trên vùng HV2	Thay đổi nucleotid		
1	56	A-C	A-G	
2	194	C-T	DelC	
3	249	A-G	DelA	
4	293	T-C	DelT	
5	310	T-C	DelT	
6	315	insC	insCC	
7	316	G-A	G-C	
8	368	A-G	DelA	
9	385	A-G	DelA	
10	309	insC	DelC	insCC

Nhận xét: Kết quả giải trình tự vùng HV2 DNA ty thể của 517 mẫu nghiên cứu thuộc 4 dân tộc Kinh, Chăm, Mường và Khmer người Việt Nam được đối chiếu với trình tự chuẩn cho thấy có 10 vị trí trên vùng HV2 có từ hai loại đa hình trở lên trong đó vị trí 309 có 3 loại đa hình (309insC, 309insCC, 309DelC).

**Bảng 3.4: Các dạng SNP trên vùng HVI của DNA ty thể chỉ gặp ở 1 trong 4 dân tộc Kinh, Mường, Khmer, Chăm**

STT	SNP chỉ gặp ở dân tộc Kinh	SNP chỉ gặp ở dân tộc Mường	SNP chỉ gặp ở dân tộc Khmer	SNP chỉ gặp ở dân tộc Chăm
1	C16071T	C16069T	G16042A	G16153A
2	C16114A	G16084C	A16054C	T16178C
3	T16161A	A16091T	A16070C	C16201T
4	A16164G	C16095T	T16102C	C16224T
5	C16187T	A16113C	C16150T	C16232T
6	16188insC	A16116C	C16186T	C16270T
7	A16227G	A16122T	G16244A	C16320T
8	C16250T	T16154C	A16275G	C16348T
9	C16286T	A16194C	T16276A	
10	A16302G	T16195C	C16282A	
11	A16305T	C16197G	A16322C	
12	A16317C	A16233G	T16352C	
13	C16328T	A16240G	C16353T	
14	A16339G	C16279A	T16381C	
15	C16365T	A16299T	G16391A	
16	C16511A	A16300G	C16400T	
17	T16512C	G16310A	G16438A	
18	C16514T	A16367C		
19	A16515C	C16426G		
20	G16516A	16474DelG		
21	C16520T			

Nhận xét: có 66 loại đa hình trên vùng HV1 chỉ gặp ở các mẫu thuộc 1 trong 4 dân tộc trong đó có 21 loại đa hình chỉ thấy ở dân tộc Kinh, 20 loại SNP chỉ có ở dân tộc Mường, 17 loại SNP chỉ có ở dân tộc Khmer và ít nhất là dân tộc Chăm có 8 loại SNP chỉ gặp ở dân tộc này.

**Bảng 3.5: Các dạng SNP trên vùng HV2 của DNA ty thể chỉ gặp ở 1 trong 4 dân tộc Kinh, Mường, Khmer, Chăm Việt Nam**

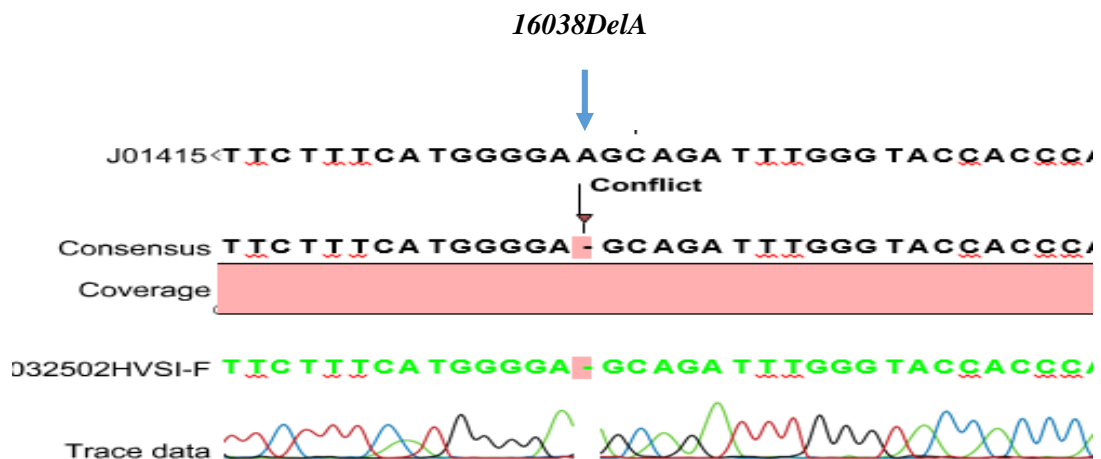
STT	SNP chỉ gặp ở dân tộc Kinh	SNP chỉ gặp ở dân tộc Mường	SNP chỉ gặp ở dân tộc Khmer	SNP chỉ gặp ở dân tộc Chăm
1	T89C	C43T	T63C	G62A
2	A93G	G53A	G66A	C64T
3	A95C	T55C	A178G	G94A
4	T127C	T158A	A259G	A193G
5	G203C	A237G	C273T	A202G
6	A227G	A302C	291DelA	A215C
7	G251A	C303A	292DelT	
8	G260A	A328G	T293C/DelT	
9	C271T	334DelA	294DelT	
10	A281G	A351G	A297G	
11	C285T	352DelA	A326G	
12	C311T		C418A	
13	C317T			

Nhận xét: có 42 loại đa hình trên vùng HV2 chỉ gặp ở các mẫu thuộc 1 trong 4 dân tộc (6 loại SNP ở dân tộc Chăm, 11 loại SNP ở dân tộc Mường, 12 loại SNP ở dân tộc Khmer, 13 loại SNP ở dân tộc Kinh).



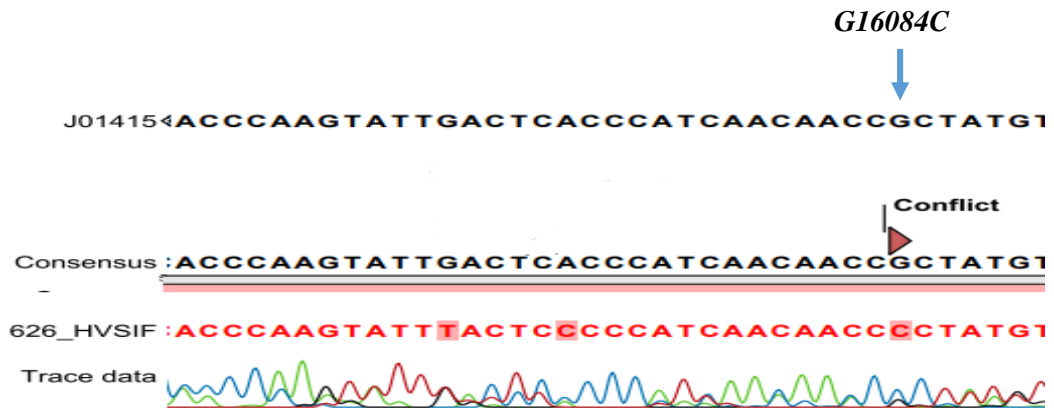
### 3.2.2. Đa hình mới được phát hiện trên vùng HV1 và HV2 của DNA ty thể người Việt Nam

Toàn bộ các đa hình được tìm thấy trên hai vùng HV1 và HV2 của 517 mẫu nghiên cứu được kiểm tra trên Mitomap (<https://www.mitomap.org>), từ đó chúng tôi đã phát hiện được 3 SNP mới chưa được công bố trên Mitomap đó là SNP 16038DelA (công bố trên Mitomap nucleotid A tại vị trí 16038 chuyển thành nucleotid G hoặc T hoặc thêm nucleotid A hoặc thêm C); SNP G16084C (công bố trên Mitomap nucleotid G tại vị trí 16084 chuyển thành nucleotid A hoặc T hoặc thêm nucleotid G) và SNP A16515C (SNP đã công bố trên Mitomap là 16515DelA). Hai SNP mới 16038DelA và G16084C được tìm thấy ở dân tộc Mường và SNP A16515C được tìm thấy ở dân tộc Kinh.



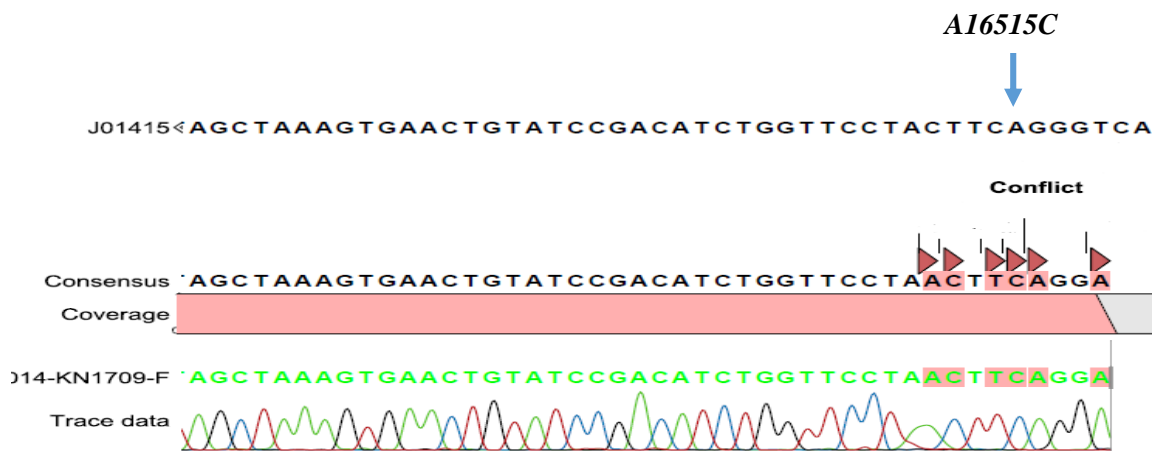
**Hình 3.14: Hình ảnh giải trình tự SNP (16038DelA) của vùng HV1**

Nhận xét: Kết quả giải trình tự vùng HV1 của DNA ty thể ở một mẫu nghiên cứu trên dân tộc Mường thấy đa hình 16038DelA, đây là loại đa hình mới chưa thấy công bố trên Mitomap, công bố trên Mitomap nucleotid A tại vị trí 16038 chuyển thành nucleotid G hoặc T hoặc thêm nucleotid A hoặc thêm nucleotid C).



**Hình 3.15: Hình ảnh giải trình tự SNP mới (G16084C) trên vùng HV1 của DNA ty thể**

Nhận xét: Kết quả giải trình tự vùng HV1 của DNA ty thể ở mẫu nghiên cứu trên thấy đa hình đơn nucleotid G16084C là đa hình mới chưa thấy công bố trên Mitomap.



**Hình 3.16: Hình ảnh giải trình tự SNP (A16515C) trên vùng HV1 mtDNA**

Nhận xét: Ở mẫu nghiên cứu này khi giải trình tự vùng HV1 của DNA ty thể thấy trên một đoạn gen ngắn có tới 6 SNP (C16511A, T16512C, C16514T, A16515C, G16516A, T16519A) trong đó SNP A16515C là SNP mới chưa thấy công bố trên Mitomap.

### 3.2.3. Các vị trí đa hình thường gặp trong các mẫu nghiên cứu

Chúng tôi cũng đã thống kê được một số đa hình thường gặp trên hai vùng HV1, HV2 của DNA ty thể ở 517 mẫu nghiên cứu. Bảng 3.12 dưới đây thể hiện các vị trí đa hình thường gặp:

**Bảng 3.6: Bảng một số vị trí đa hình thường gặp trên vùng HV1 và HV2**

<b>Vị trí đa hình trên HV1</b>	<b>Chăm (n=113)</b>	<b>Kinh (n=206)</b>	<b>Khmer (n=98)</b>	<b>Mường (n=100)</b>	<b>Tổng</b>	<b>Tỷ lệ % (n=517)</b>
<i>G16129A</i>	26	82	33	30	171	33
<i>T16172C</i>	13	50	16	30	109	21
<i>A16183C</i>	52	61	29	29	171	33
<i>T16189C</i>	64	72	34	33	203	39
<i>T16217C</i>	22	27	9	8	66	12,7
<i>C16223T</i>	41	80	52	39	212	41
<i>T16304C</i>	23	65	33	28	149	28,8
<i>T16311C</i>	14	26	13	8	61	11,8
<b>Vị trí đa hình trên HV2</b>	<b>Chăm (n=113)</b>	<b>Kinh (n=206)</b>	<b>Khmer (n=98)</b>	<b>Mường (n=100)</b>	<b>Tổng</b>	<b>Tỷ lệ % (n=517)</b>
<i>A73G</i>	113	206	98	98	515	99,6
<i>C150T</i>	22	53	21	25	121	23,4
<i>T152C</i>	22	27	16	11	76	14,7
<i>T199C</i>	10	38	13	24	85	16,4
<i>249delA</i>	11	63	21	33	128	24,7
<i>A263G</i>	113	206	98	100	517	100
<i>309insC</i>	72	115	45	60	292	56,4
<i>315insC</i>	112	200	94	92	498	96,3

Nhận xét: Các vị trí đa hình hay gặp nhất trên vùng HV1 là: C16223T với tần suất gặp 41%, T16189C với tần suất gặp 39%, G16129A, A16183C với tần suất gặp 33%. Các vị trí đa hình hay gặp trên vùng HV2 là: 309insC với tần suất gặp 56,4%, 315insC với tần suất gặp 96,3%, A73G với tần suất gặp 99,6%, đặc biệt là đa hình A263G gặp ở 100% các mẫu nghiên cứu.

### **3.2.4. Tổng số đa hình trong các mẫu nghiên cứu**

Sau khi kiểm tra, so sánh với trình tự chuẩn chúng tôi đã thu được kết quả tổng số các đa hình của các mẫu nghiên cứu, kết quả cho thấy đa số các mẫu nghiên cứu đều có nhiều đa hình so với trình tự chuẩn, có 401/517 mẫu nghiên cứu có số lượng đa hình từ 10 trở lên, mẫu có số lượng đa hình nhiều nhất là 21 vị trí đa hình (có 1 mẫu thuộc dân tộc Chăm), mẫu có số lượng đa hình ít nhất là 6 đa hình (có 8 mẫu). Đây là một số lượng đa hình khá lớn và điều này phù hợp với các nghiên cứu trước đây cho rằng hai vùng siêu biến HV1 và HV2 có tần số đột biến cao nhất trong hệ gen ty thể. Số liệu thu được về các vị trí đa hình trên 2 vùng HV1, HV2 của DNA ty thể ở 4 dân tộc Kinh, Chăm, Mường, Khmer người Việt Nam đã phản ánh tốc độ đột biến rất cao của hai vùng HV1, HV2. Kết quả này cũng sẽ cung cấp những số liệu quý giá cho những nghiên cứu tiếp theo.

### **3.3. Phân nhóm SNP đặc trưng vùng HV1 và HV2 (phân chia các nhóm đơn bội mtDNA theo bộ SNP đặc trưng trên vùng HV1, HV2) ở 4 dân tộc Kinh, Chăm, Mường, Khmer Việt Nam**

Việc phân loại DNA ty thể theo các nhóm đơn bội (Haplogroup) các mẫu nghiên cứu thuộc 4 dân tộc Kinh, Mường, Chăm và Khmer là rất phức tạp. Do đó, trong nghiên cứu này chúng tôi tiến hành phân nhóm đơn bội DNA ty thể dựa theo các nghiên cứu gần đây trên thế giới [106], [107], [108], và đặc biệt là các nghiên cứu về phân loại nhóm đơn bội mtDNA của các dân tộc ở các nước châu Á, cụ thể dựa theo cách phân loại của Yao năm 2002 khi phân nhóm đơn bội các dân tộc người Trung Quốc [32]. Sau khi phân tích kết quả chúng tôi đã tiến hành phân chia các nhóm đơn bội DNA ty thể của các dân tộc Việt Nam dựa trên các SNP đặc trưng ở vùng HV1 và HV2 của DNA ty thể. Cụ thể các nhóm đơn bội DNA ty thể của 4 dân tộc Kinh, Chăm, Mường, Khmer người Việt Nam được thể hiện ở các bảng dưới đây (bảng 3.6, 3.7, 3.8, 3.9):

**Bảng 3.7: Phân chia nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên vùng HV1, HV2 và các dạng SNP trên vùng HV1, HV2 của một số mẫu nghiên cứu đại diện cho dân tộc Chăm Việt Nam**

Mẫu	Haplogroup (Số lượng mẫu)	HV1(16000+)	HV2
Cham07	B4 (2)	C147T, A183C, C184A, <b>T189C</b> , <b>T217C</b> , T362C, C465T, T519A	A73G, A263G, 315C
Cham38	B4a (5)	T093C, C174T, A182C, A183C, <b>T189C</b> , <b>T217C</b> , <b>T261C</b> , T519A	A73G, T146C, A153G, A263G, 315C 522-523d
Cham08	B4b (1)	<b>T136C</b> , A183C, <b>T189C</b> , <b>T217C</b> , A241C, T519A	A73G, T146C, A263G, 309CC, 315C
Cham31	B4c (14)	G129A, <b>T140C</b> , A166C, A182C, <b>A183C</b> , <b>T189C</b> , <b>T217C</b> , <b>G274A</b> , A335G, T519A	A73G, <b>C150T</b> , A263G, 309CC, 315C
Cham29	B4g (2)	A181C, <b>A182C</b> , <b>A183C</b> , <b>T189C</b> , C201T, G213A, <b>T217C</b> , C232T, C270T, <b>C292T</b> , T519A	A73G, T195C, A263G, 309C, 315C, 522-523d
Cham03	B5a (18)	<b>T140C</b> , G153A, T178C, A183C, <b>T189C</b> , A207G, <b>C266A</b> , T519A	42G, A73G, G103A, <b>A210G</b> , A263G, 309CC, 315C, 522-523d
Cham94	B5b (3)	C067G, <b>T140C</b> , <b>A183C</b> , <b>T189C</b> , <b>T243C</b> , T519A	A73G, A103G, T152C, T204C, A263G, 315C, 522-523d
Cham52	C (1)	T086C, C223T, <b>T298C</b> , <b>C327T</b> , T357C, T519A	C64T, A73G, 249DelA, A263G, 309C, 315C
Cham85	D4a (2)	T093C, <b>G129A</b> , <b>C223T</b> , T263C, <b>T362C</b> , T519A	A73G, T152C, A263G, 309C, 315C, 523CA
Cham69	D4e (1)	C167T, <b>C223T</b> , C320T, <b>T362C</b>	A73G, G94A, A263G, 309C, 315C
Cham33	E (1)	<b>C223T</b> , <b>C291T</b> , <b>T362C</b> , <b>G390A</b> , T519A	A73G, <b>A193G</b> , A263G, 309C, 315C
Cham51	F (1)	T189C, G274A, <b>T304C</b> , T519A	A73G, <b>249delA</b> , A263G, 309CC, 315C, 522-523d
Cham61	F1a (8)	<b>G129A</b> , <b>T172C</b> , <b>T304C</b> , T362C, T519A	A73G, <b>249delA</b> , A263G, 315C, 522-523d
Cham06	M (15)	G129A, T209C, <b>C223T</b> , T325C	A73G, T152C, A200G, A263G, 315C
Cham58	M12 (1)	G129A, <b>C223T</b> , <b>C234T</b> , C261T, C262T, G274A, <b>C290T</b>	A73G, G143A, T152C, A263G, 309C, 315C, T318C
Cham65	M7b (3)	<b>C223T</b> , <b>T297C</b> , T311C	A73G, <b>C150T</b> , <b>T199C</b> , T204C, A263G, 309CC, 315C

Cham97	M7b1 (3)	<b>G129A</b> , T189C, <b>C192T</b> , <b>C223T</b> , <b>T297C</b>	A73G, <b>C150T</b> , <b>T199C</b> , A202G, A263G, 309C, 315C, C332T
Cham20	M7c (3)	T075C, <b>C223T</b> , 293T, C295T, T519A	A73G, <b>T146C</b> , T152C, <b>T199C</b> , A263G, 309C, 315C, 522-523d
Cham98	M8a (2)	C184A, T189C, <b>C223T</b> , <b>T298C</b> , <b>G319A</b>	A73G, A263G, 315C
Cham17	M9a (1)	<b>C223T</b> , C256T, <b>T311C</b> , T362C, T519A	A73G, T125G, T146C, A153G, A263G, 309C, 315C
Cham10	M9b (1)	<b>A051G</b> , <b>T209C</b> , <b>C223T</b> , <b>T362C</b> , T519A	A73G, A153G, A263G, 315C
Cham80	N21 (1)	G129A, <b>C193T</b> , <b>C223T</b> , <b>T325C</b> , T519A	A73G, C150T, <b>T195C</b> , A263G, 315C, 337d, 522-523d
Cham09	N9a (10)	T093C, T140C, T189C, <b>C223T</b> , <b>257A</b> , <b>C261T</b> , C292T, T519A	A73G, <b>C150T</b> , A263G, 309C, 315C
Cham57	R (2)	C256T, C290T, <b>T304C</b> , C465T	A73G, T195C, A263G, 309C, 315C, 522-523d
Cham63	R11 (2)	T086C, A182C, <b>A183C</b> , <b>T189C</b> , <b>T311C</b> , G390A, A399G, T519A	A73G, <b>A189G</b> , T215C, A263G, 309CCC, 315C
Cham04	R9b (10)	C221T, T249C, T288C, C301T, <b>T304C</b> , <b>G390A</b> , T519A	A73G, A263G, 315C, G329A
<b>Tổng số</b>	<b>113 mẫu</b>		

*Kết quả giải trình tự của các mẫu nghiên cứu được so sánh với trình tự chuẩn. Phân nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên 2 vùng HV1 và HV2 của mtDNA. Các SNP in đậm là các SNP đặc trưng cho nhóm đơn bội tương ứng.*

**Nhận xét:** bảng 3.6 là bảng chi tiết toàn bộ các SNP trên một số mẫu dân tộc Chăm, các mẫu này đại diện cho dân tộc Chăm, được phân loại vào các nhóm đơn bội mtDNA dựa trên các SNP đặc trưng của vùng HV1 và HV2. 113 mẫu dân tộc Chăm được phân loại vào 26 nhóm đơn bội, trong đó nhóm B4c, M, B5a là 3 nhóm đơn bội chiếm tỷ lệ cao nhất lần lượt là 14/113, 15/113 và 18/113 mẫu nghiên cứu dân tộc Chăm. Cham: dân tộc Chăm.

**Bảng 3.8: Phân chia nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên vùng HV1, HV2 và các dạng SNP trên vùng HV1, HV2 của một số mẫu nghiên cứu đại diện cho dân tộc Kinh Việt Nam**

Mẫu	Haplogroup (Số lượng mẫu)	HV1(16000+)	HV2
KN23	A (3)	38delA, T86C, G129A, T209C, <b>C223T</b> , A272G, <b>C290T</b> , T519A	A73G, <b>T152C</b> , G225A, 249del, A263G, 315C, G316A
KN20	B (1)	<b>T189C</b> , C223T, C278T	A73G, C150T, A263G, 315C
KN3	B4 (13)	C147T, A183C, C184A, <b>T189C</b> , <b>T217C</b> , A235G, C294T	A73G, A263G, 309C, 315C
Khin37	B4a (4)	C168T, A182C, A183C, <b>T189C</b> , <b>T217C</b> , <b>C261T</b> , T311C, T519A	A73G, T146C, A263G, 309CCC, 315C, 522-523d
Khin33	B4b (3)	T126C, <b>T136A</b> , A183C, <b>T189C</b> , <b>T217C</b> , C260T, C287T, C325T, T519A	A73G, A200G, A263G, 315C, 522-523d
Khin26	B4c (6)	C147T, C168T, <b>A183C</b> , C184A, <b>T189C</b> , <b>T217C</b> , C234T, A235G, T519A	A73G, A263G, 309C, 315C
Khin5	B4g (4)	T093C, A181C, <b>A182C</b> , <b>A183C</b> , <b>T189C</b> , G213A, <b>T217C</b> , C242T, C261T, C287T, <b>C292T</b> , C301T, C355T, T519A	61A, A73G, A263G, 309C, 315C, 522-523d
Khin3	B5 (1)	<b>T140C</b> , C187T, <b>T189C</b> , C256T, C266G, T519A	A73G, A93G, A210G, A263G, 315C, 522-523d
Khin13	B5a (23)	<b>T140C</b> , A183C, <b>T189C</b> , T243C, <b>C266A</b> , T311C, T519A	A73G, <b>A210G</b> , A263G, 309C, 315C, 522-523d
Khin60	B5b (1)	<b>T140C</b> , A183C, <b>T189C</b> , <b>T243C</b> , T311C, T519A	A73G, G103A, T204C, A263G, 309CC, 315C, 522-523d
Khin47	B6 (1)	T093, <b>C179A</b> , A182C, A183C, <b>T189C</b>	A73G, C150T, A263G, 309C, 315C
Khin15	C (6)	T189C, <b>C223T</b> , <b>T298C</b> , C327T, T519A	A73G, <b>249delA</b> , A263G, 309C, 315C
KN59	D4 (1)	38del, 188C, 193C, <b>C223T</b> , C234T, T311C, <b>T362C</b>	A73G, C150T, C151T, T152C, A263G, C285T, T310C, 315C
KN11	D5 (1)	<b>T189C</b> , <b>C223T</b> , <b>T362C</b>	A73G, <b>C150T</b> , 309C, 315C
Khin85	D4a (3)	C111G, <b>G129A</b> , <b>C223T</b> , <b>T362C</b>	A73G, <b>T152C</b> , A263G, 309C, 315C
Khin95	D5a (1)	T092C, A164G, <b>A182C</b> , <b>A183C</b> , <b>T189C</b> , <b>C223T</b> , <b>C266T</b> , <b>T362C</b>	A73G, <b>C150T</b> , A263G, 309CC, 315C, 522-523d
Khin10	D5b (2)	T092C, C148T, A183C, <b>T189C</b> , <b>C223T</b> , <b>T362C</b> , T519A	A73G, <b>C150T</b> , T152C, G185A, A263G, 309C, 315C, 522-523d
Khin2	F (3)	T157C, C256T, <b>T304C</b> , A335G	A73G, <b>249delA</b> , A263G, 315C
Khin29	F1a (43)	<b>G129A</b> , A162G, <b>T172C</b> , <b>T304C</b> , A399G, T519A	A73G, <b>249delA</b> , A263G, 315C, 522-523d
Khin21	F1b (3)	<b>A183C</b> , <b>T189C</b> , C292T, <b>T304C</b> , T519A	A73G, <b>249delA</b> , A263G, 315C, 522-523d
Khin19	F1c (1)	<b>C111T</b> , <b>G129A</b> , C266T, <b>T304C</b> , T519A	A73G, <b>T152C</b> , <b>249delA</b> , A263G, 309C, 315C, 522-523d
Khin75	F3a (1)	T093C, T249C, <b>T298C</b> , <b>C355T</b> , <b>T362C</b> , G390A	A73G, T152C, G207A, <b>249delA</b> , A263G, 309CC, 315C
KN79	G2a (1)	<b>C223T</b> , A227G, <b>C278T</b> , <b>T362C</b>	A73G, A263G, 309C, 315C

Khin105	M (8)	<b>C223T</b> , C295T, T519A	A73G, T146C, T199C, A263G, 315C, 522-523d
KN 25	M10 (6)	<b>C223T</b> , C260G, <b>T311C</b>	A73G, A263G, 315C
Khin4	M12 (1)	C148T, <b>C223T</b> , <b>C234T</b> , C261T, <b>C290T</b> , T519A	A73G, A263G, 309C, 315C, T318C
Khin1	M7a (2)	T086C, G129A, <b>T209C</b> , <b>C223T</b> , A272G, T519A	A73G, T152C, 249delA, A263G, 315C, G316A, 522-523d
Khin73	M7b (7)	C185T, <b>C223T</b> , <b>T297C</b>	A73G, <b>C150T</b> , <b>T199C</b> , T204C, A263G, C271T, 309CC, 315C
Khin20	M7b1 (20)	<b>G129A</b> , T189C, C192T, <b>C223T</b> , <b>T297C</b> , T356C	A73G, T131C, <b>C150T</b> , <b>T199C</b> , A263G, 309CC, 315C, C332T, 522-523d
Khin42	M7c (7)	<b>C223T</b> , C295T, T519A	A73G, <b>T146C</b> , <b>T199C</b> , A263G, 315C, 522-523d
KN84	M8a (1)	38delA, C184T, <b>C223T</b> , <b>T298C</b> , <b>G319A</b> , A343G	A73G, A263G, 309C, 315C
KN10	M9 (3)	C223T, <b>C234T</b> , T271C, C344T, T362C	A73G, <b>A153G</b> , A263G, 315C
Khin80	M9a (4)	T093C, <b>C223T</b> , <b>C234T</b> , T271C, <b>T362C</b>	A73G, <b>A153G</b> , A263G, 315C
Khin90	N (1)	<b>C223T</b> , T263C, <b>G274A</b> , <b>T311C</b> , <b>A343G</b> , T357C, T519A	A73G, <b>G103A</b> , C151T, T152C, G260A, A263G, 315C
Khin6	N9a (1)	<b>C223T</b> , <b>C257A</b> , <b>C261T</b> , T311C	A73G, <b>C150T</b> , A263G, 315C
KN40	R (7)	38delA, T288C, <b>T304C</b> , G390A, G518A, T519G	A73G, G143A, T146C, A183G, T204C, A263G, 309C, 315C
Khin40	R9 (3)	<b>T304C</b> , A335G, T362C	A73G, C150T, T152C, A263G, 309C, 315C
Khin28	R9b (4)	A284G, <b>T304C</b> , A309G, <b>G390A</b> , T519A	A73G, A183G, A227G, A263G, 315C, 522-523d
Khin104	Z (5)	<b>C185T</b> , <b>C223T</b> , <b>C260T</b> , <b>T298C</b> , A317C	A73G, C151T, <b>T152C</b> , <b>249delA</b> , A263G, 309C, 315C
<b>Tổng số</b>	<b>206 mẫu</b>		

*Kết quả giải trình tự của các mẫu nghiên cứu được so sánh với trình tự chuẩn. Phân nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên 2 vùng HV1 và HV2 của mtDNA. Các SNP in đậm là các SNP đặc trưng cho nhóm đơn bội tương ứng. Khin: dân tộc Kinh lấy mẫu ở Hà Nội, KN: dân tộc Kinh lấy mẫu ở TPHCM.*

**Nhận xét:** bảng 3.7 là bảng chi tiết các SNP trên một số mẫu dân tộc Kinh, các mẫu này đại diện cho dân tộc Kinh, được phân loại vào các nhóm đơn bội mtDNA dựa trên các SNP đặc trưng của vùng HV1, HV2. 206 mẫu dân tộc Kinh được phân loại vào 39 nhóm đơn bội, trong đó nhóm M7b1, B5a, F1a là 3 nhóm đơn bội chiếm tỷ lệ cao nhất, lần lượt là 20/206, 23/206, 43/206 mẫu dân tộc Kinh.



**Bảng 3.9: Phân chia nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên vùng HV1, HV2 và các dạng SNP trên vùng HV1, HV2 của một số mẫu nghiên cứu đại diện cho dân tộc Khmer Việt Nam**

Mẫu	Haplogroup (Số lượng mẫu)	HV1(16000+)	HV2
Kh 040	A (1)	T93C, <b>C223T</b> , C234T, <b>C290T</b> , A293C, <b>G319A</b>	A73G, <b>A235G</b> , A263G, A297G, 315C
Kh 026	B (12)	<b>T189C</b> , C214A, C223T, G274A, T276A, C282A, T311C	A73G, T146C, A263G, 315C, C418A
Kh 054	B4 (5)	A166C, A183C, <b>T189C</b> , <b>T217C</b> , A235G,	A73G, A263G, 309C, 315C, G207A
Kh 033	B4a (2)	A182C, A183C, <b>T189C</b> , <b>T217C</b> , <b>T261C</b>	A73G, T146C, A263G, 309C, 315C
Kh 055	B4b (2)	<b>T136A</b> , A183C, <b>T189C</b> , <b>T217C</b> , A235G	A73G, A263G, 309C, 315C
Kh 070	B5a (8)	<b>T140C</b> , C179A, A181C, A182C, A183C, <b>T189C</b> , C261T, <b>C266A</b> , A335G	A73G, T152C, <b>A210G</b> , A263G, T310C, G316C, 318C
Kh 001	D4 (8)	<b>C223T</b> , C259T, G274A, T311C, <b>T362C</b> , T381C, G518A	T63C, C64T, G66A, A73G, T146C, A263G, 315C
Kh 66	D5 (1)	T136A, A175C, A182C, A183C, <b>T189C</b> , <b>C223T</b> , <b>T362C</b>	A73G, <b>C150T</b> , A263G, 309C, 315C
Kh 005	F1a (14)	C108T, <b>G129A</b> , A162G, <b>T172C</b> , C239T, <b>T304C</b> , C327T	A73G, <b>249DelA</b> , A263G, 309C, 315C
Kh 048	F1b (2)	T136A, T140A, A183C, A184C, <b>T189C</b> , <b>T304C</b> , G390A	A73G, <b>249DelA</b> , A263G, 309C, 315C
Kh 092	G2 (2)	C214A, C223T, C256T, <b>C278T</b> , <b>C362T</b>	A73G, T152C, A263G, 309C, 315C
Kh 023	M (18)	38delA, G129A, T209C, <b>C223T</b> , A272G, A322C	T58G, 56G, A73G, T152C, G225A, 249DelA, A263G, 315C, G316A
Kh 086	M10 (4)	<b>C223T</b> , T263C, G274A, <b>T311C</b> , A343G, T357C	A73G, A263G, T310C, G316C, T319G
Kh 091	M7b (3)	C192T, C223T, <b>T297C</b>	A73G, <b>C150T</b> , <b>T199C</b> , A263G, 309C, 315C

Kh 049	M7b1 (7)	<b>G129A</b> , T189C, <b>C192T</b> , C223T, <b>T297C</b> , T217C, A235G	A73G, <b>C150T</b> , <b>T199C</b> , A263G, 309C, 315C
Kh 007	M7c (3)	G145A, C291T, <b>C295T</b> , T304C	A73G, A210G, A263G, 315C
Kh 025	R (2)	T249C, T288C, <b>T304C</b> , C344T	A73G, T217C, A263G, 315C, G329A
Kh 067	R9a (1)	C260T, <b>T298C</b> , <b>C355T</b> , <b>T362C</b>	A73G, <b>249DelA</b> , A263G, 309C, 315C
Kh 35	U5a (1)	T93C, G129A, <b>C256T</b> , <b>T357C</b> , A399G	A73G, T131C, C150T, T199C, A263G, 315C
Kh 14	Z (2)	<b>C185T</b> , <b>C223T</b> , <b>C260T</b> , <b>T298C</b>	A73G, <b>T152C</b> , <b>249DelA</b> , A263G, 315C
<b>Tổng</b>	<b>98 mẫu</b>		

*Kết quả giải trình tự của các mẫu nghiên cứu được so sánh với trình tự chuẩn. Phân nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên 2 vùng HV1 và HV2 của mtDNA. Các SNP in đậm là các SNP đặc trưng cho nhóm đơn bội tương ứng. Kh: dân tộc Khmer*

Nhận xét: bảng 3.8 là các nhóm đơn bội mtDNA đại diện cho dân tộc Khmer được phân nhóm dựa trên các SNP đặc trưng của vùng HV1 và HV2. 98 mẫu người dân tộc Khmer được phân loại vào 20 nhóm đơn bội, trong đó nhóm B, F1a, M là 3 nhóm đơn bội chiếm tỷ lệ cao nhất, lần lượt là 12/98, 14/98, 18/98 mẫu dân tộc Khmer.

**Bảng 3.10: Phân chia nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên vùng HV1, HV2 và các dạng SNP trên vùng HV1, HV2 của một số mẫu nghiên cứu dân tộc Mường Việt Nam**

Mẫu	Haplogroup (Số lượng mẫu)	HV1(16000+)	HV2
M1	B (8)	T93C, T178C, A182C, A183C, <b>T189C</b> , T271C, G274A	A73G, T146C, A263G, 315C
M8	B4 (4)	A182C, A183C, <b>T189C</b> , <b>T217C</b> , G274A, T304C, G310A, 474delG	A73G, A263G, 309CC, 315C
M2	B4a (4)	A182C, A183C, <b>T189C</b> , G213A, <b>T217C</b> , <b>C261T</b> , C292T	A73G, A263G, C308T, 310delT
M41	B5 (1)	T92C, G129A, <b>T140C</b> , A182C, A183C, <b>T189C</b> , A194C, C197G	A73G, A210G, A263G, T310C
M15	B5a (7)	A129G, <b>T140C</b> , A182C, A183C, <b>T189C</b> , <b>C266A</b> , A399G	A73G, <b>A210G</b> , A263G, T310C, G316C
M26	C (5)	A182C, A183C, T189C, <b>C223T</b> , <b>T298C</b> , C327T, 469delT	A73G, T146C, A237G, <b>249delA</b> , A263G, C303A
M74	D4 (4)	C111T, T126C, T140A, T172C, A183C, T189C, <b>C223T</b> , <b>T362C</b>	A73G, A263G, 309CC, 315C, 302C, 334delA
M39	F1a (16)	C108T, <b>G129A</b> , A126G, <b>T172C</b> , <b>T304C</b>	A73G, <b>249delA</b> , A263G, 309CC, 315C
M22	F1b (1)	A51G, <b>T189C</b> , A269G, C299T, A300G, <b>T304C</b>	A73G, C150T, T195C, A214G, <b>249delA</b> , A263G, T310C
M93	F2 (2)	<b>T304C</b> , C465T	A73G, <b>249delA</b> , A263G, 309C, 315C
M63	F2a (1)	T92A, <b>C291T</b> , <b>T304C</b>	A73G, A214G, <b>249delA</b> , A263G, 309CC, 315C
M3	G2 (6)	C69T, T172C, C223T, A235G, <b>C278T</b> , C291A, T298C, <b>T362C</b>	A73G, T146C, C150T, T199C, A263, 309CC, 315C
M69	M (5)	C193T, <b>C223T</b>	A73G, C150T, T195C, A263G, 309CC, 315C, 337delA
M24	M10 (2)	<b>C223T</b> , C256T, A299G, <b>T311C</b>	A73G, A263G, 315C
M38	M7 (6)	T311C, T356C	A73G, <b>T146A</b> , <b>T199C</b> , A263G, 315C
M31	M7b (4)	<b>C223T</b> , <b>T297C</b>	A73G, <b>C150T</b> , <b>T199C</b> , T204C, A263G, 309C, 315C

M4	M7b1 (10)	<b>G129A, C192T, C223T, T297C</b>	A73G, <b>C150T, T199C</b> , A263G, 309C, 315C
M98	M7c (1)	C295T, G319A	A73G, <b>T146C, T199C</b> , A263G, 315C
M28	M8a (2)	C184T, T189C, <b>C223T, T298C, G319A</b>	A73G, T195C, A263G, 309CC, 315C
M34	N9a (2)	T172C, T189C, C201T, <b>C223T, C257A, C261T</b>	A73G, <b>C150T</b> , T195C, A263G, 309C, 315C
M7	R (3)	T124C, C148T, A183G, <b>T304C</b> , A309G, G390A, 474delG	A73G, A263G, T310C
M23	R9a (6)	T93C, C111T, C192T, T249C, <b>T298C, C355T, T362C</b> , G390A	A73G, G207A, 249delA, A263G, 309C, 315C
<b>Tổng</b>	<b>100 mẫu</b>		

*Kết quả giải trình tự của các mẫu nghiên cứu được so sánh với trình tự chuẩn. Phân nhóm đơn bội mtDNA dựa trên các SNP đặc trưng trên 2 vùng HV1 và HV2 của mtDNA. Các SNP in đậm là các SNP đặc trưng cho nhóm đơn bội tương ứng. M: dân tộc Mường.*

**Nhận xét:** bảng 3.9 là bảng chi tiết toàn bộ các SNP trên một số mẫu dân tộc Mường, các mẫu này là đại diện cho 100 mẫu dân tộc Mường được phân loại vào 22 haplogroups dựa trên các SNP đặc trưng vùng HV1, HV2, trong đó 3 nhóm đơn bội chiếm tỷ lệ cao nhất là nhóm B, M7b1, F1a, lần lượt là 8/100, 10/100, 16/100 mẫu nghiên cứu dân tộc Mường.

Phân nhóm đơn bội DNA ty thể dựa vào các SNP đặc trưng trên vùng HV1, HV2 của mtDNA ở 517 mẫu nghiên cứu (206 mẫu dân tộc Kinh, 100 mẫu dân tộc Mường, 113 mẫu dân tộc Chăm và 98 mẫu dân tộc Khmer) thuộc 4 dân tộc: Kinh, Mường, Chăm và Khmer được thể hiện ở các bảng 3.6, 3.7, 3.8, 3.9 ở trên. Tổng số 517 mẫu của 4 dân tộc trên đã được phân loại vào 50 haplogroups mtDNA. Chi tiết toàn bộ đa hình và phân nhóm đơn bội của 517 mẫu nghiên cứu được trình bày ở Phụ lục 2.

Từ kết quả chi tiết toàn bộ đa hình trên vùng HV1, HV2 của mtDNA ở 517 mẫu nghiên cứu (Phụ lục 2), chúng tôi đã thống kê được tổng số haplotype HV1/HV2 mtDNA của 517 mẫu nghiên cứu và số lượng mẫu trên mỗi một haplotype HV1/HV2 mtDNA (bảng 3.10) dưới đây:

**Bảng 3.11: Số lượng haplotypes HV1/HV2 mtDNA của 517 mẫu nghiên cứu thuộc 4 dân tộc Kinh, Chăm, Mường, Khmer Việt Nam**

Số mẫu/Haplotype HV1/HV2	Số lượng haplotypes	Tổng số mẫu
1 mẫu/1 haplotype	393	393
2 mẫu/1 haplotype	27	54
3 mẫu/1 haplotype	10	30
4 mẫu/1 haplotype	2	8
5 mẫu/1 haplotype	4	20
6 mẫu/1 haplotype	2	12
Tổng	<b>438</b>	<b>517</b>

Như vậy, kết quả giải trình tự của 517 mẫu nghiên cứu đã xác định được 438 haplotypes, trong đó 393 haplotypes là duy nhất (tương ứng với 393 mẫu) và 45 haplotypes (cho 124 mẫu còn lại) xuất hiện ở nhiều hơn một cá thể (từ 2 đến 6 mẫu trên 1 haplotype HV1/HV2 mtDNA). Từ đó đã xác định được độ đa dạng di truyền (genetic diversity) và xác suất trùng lặp ngẫu nhiên giữa hai cá thể theo công thức  $h = (1 - \sum x^2) / n(n-1)$  của Fumio Tajima năm 1989 [105], cho kết quả: sự đa dạng di truyền là 99,83% và xác suất trùng lặp ngẫu nhiên của hai cá thể có cùng một haplotype HV1/HV2 mtDNA là 0,37%.

Dựa vào các SNP đặc trưng trên vùng HV1, HV2 của DNA ty thể (bảng 3.6 đến bảng 3.9) chúng tôi đã phân loại được 517 mẫu nghiên cứu của 4 dân tộc Kinh, Mường, Chăm, Khmer vào 50 nhóm đơn bội hoặc dưới đơn bội (chi tiết tại Phụ lục 2) với tần số xuất hiện của từng nhóm như bảng sau:

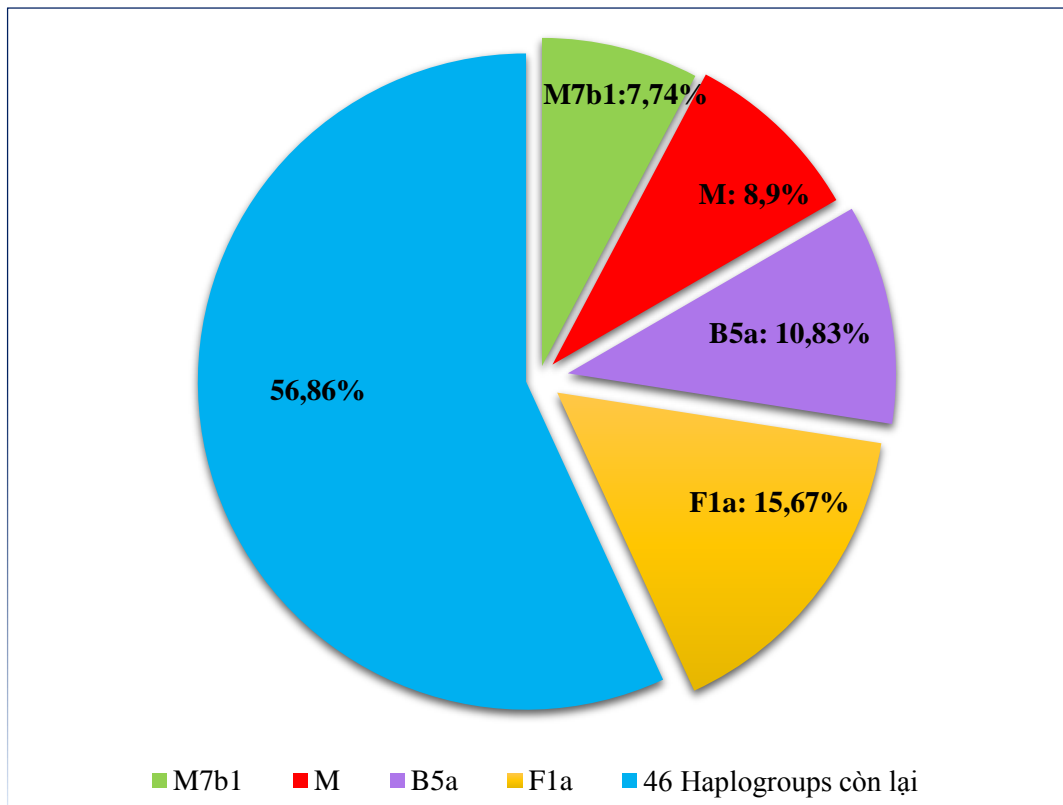
**Bảng 3.12: Bảng tần suất theo các nhóm đơn bội (Haplogroup)**

Nhóm đơn bội (Haplogroup)	Kinh	Khmer	Chăm	Mường	Tổng	Tỷ lệ % (n=517)
A	3	1			4	0,77
B	1	12		8	21	4,06
B4	13	5	2	4	24	4,64
B4a	4	2	5	4	15	2,9
B4b	3	2	1		6	1,16
B4c	6		14		20	3,86
B4g	4		2		6	1,16
B5	1			1	2	0,39
B5a	23	8	18	7	56	10,83
B5b	1		3		4	0,7
B6	1				1	0,19
C	6		1	5	12	2,32
D4	1	8		4	13	2,51
D4a	3		2		5	0,97
D4e			1		1	0,19
D5	1	1			2	0,39
D5a	1				1	0,19
D5b	2				2	0,39
E			1		1	0,19
F	3		1		4	0,77
F1a	43	14	8	16	81	15,67
F1b	3	2		1	6	1,16
F1c	1				1	0,19
F2				2	2	0,39
F2a				1	1	0,19

F3a	1				1	0,19
G2		2		6	8	1,56
G2a	1				1	0,19
<i>M</i>	8	<i>18</i>	<i>15</i>	5	46	8,9
M10	6	4		2	12	2,32
M12	1		1		2	0,39
M7				6	6	1,16
M7a	2				2	0,39
M7b	7	3	3	4	17	3,29
<i>M7b1</i>	20	7	3	<i>10</i>	40	<i>7,74</i>
M7c	7	3	3	1	14	2,71
M8a	1		2	2	5	0,97
M9	3				3	0,58
M9a	4		1		5	0,97
M9b			1		1	0,19
N	1				1	0,19
N9a	1		10	2	13	2,51
N21			1		1	0,19
R	7	2	2	3	14	2,71
R9	3				3	0,58
R9a		1		6	7	1,35
R9b	4		<i>10</i>		<i>14</i>	<i>2,70</i>
R11			2		2	0,39
U5a		1			1	0,19
Z	5	2			7	1,35
<b>Tổng</b>	<b>206</b>	<b>98</b>	<b>113</b>	<b>100</b>	<b>517</b>	<b>100</b>

Nhận xét: nghiên cứu đã phân loại 438 haplotypes mtDNA của 517 mẫu nghiên cứu vào 50 nhóm đơn bội (haplogroup) dựa theo các SNP đặc trưng trên hai vùng HV1 và HV2. Bảng trên cho thấy sự phân chia nhóm đơn bội của các dân tộc Việt Nam tập trung vào 4 nhóm haplogroup F1a, M, B5a, M7b1. Những nhóm đơn bội có ở cả 4 dân tộc: B4, B4a, B5a, M, F1a, M7b,

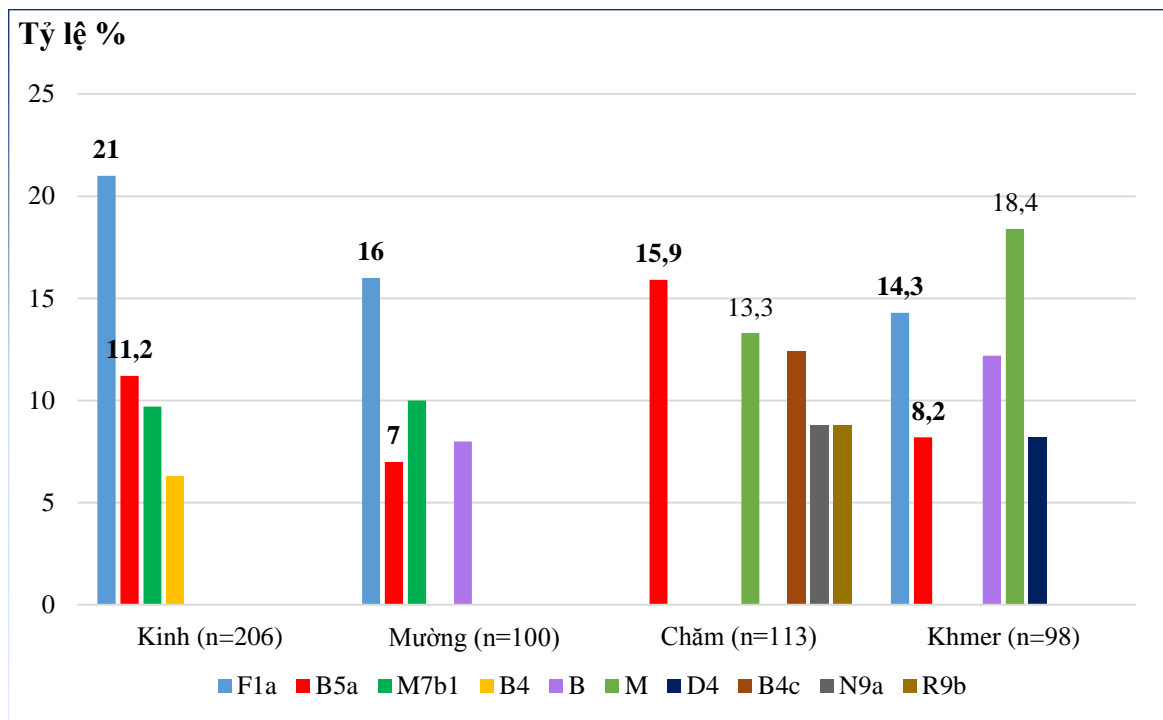
M7b1, M7c. 12 nhóm đơn bội có tần số xuất hiện thấp nhất (chỉ có 1 cá thể trong tổng số 517 mẫu nghiên cứu của cả bốn dân tộc) là nhóm B6, D4e, D5a, E, F1c, F2a, F3a, G2a, N, M9b, N21, và nhóm U5a. Các nhóm đơn bội chỉ có ở 1 dân tộc là nhóm U5a chỉ có ở dân tộc Khmer, nhóm D4e, E, N21, R11 chỉ có ở dân tộc Chăm, nhóm B6, D5a, D5b, F1c, F3a, G2a, M7a, M9, N, R9 chỉ có ở dân tộc Kinh, nhóm F2, F2a, M7 chỉ có ở dân tộc Mường.



**Biểu đồ 3.1: Biểu đồ biểu thị tỷ lệ các nhóm đơn bội mtDNA phổ biến của 4 dân tộc Kinh, Mường, Chăm và Khmer**

Nhận xét: Từ biểu đồ trên cho thấy với 4 nhóm đơn bội phổ biến F1a, M, B5a, M7b1 của người Việt Nam đã chiếm đến gần 50%, trong đó nhóm đơn bội F1a (81/517 mẫu nghiên cứu) chiếm tỷ lệ cao nhất 15,67%, nhóm B5a (56/517 mẫu nghiên cứu) chiếm 10,83%, nhóm M (46/517 mẫu nghiên cứu) chiếm 8,9%, nhóm M7b1 (40/517 mẫu nghiên cứu) chiếm 7,74%. Đây có thể là các nhóm đơn bội đại diện cho các dân tộc Việt Nam.





**Biểu đồ 3.2: Biểu đồ tỷ lệ các nhóm đơn bội chiếm tỷ lệ cao theo từng dân tộc của 4 dân tộc Kinh, Mường, Chăm, Khmer người Việt Nam**

**Nhận xét:** Khi xét riêng từng dân tộc, 4 nhóm đơn bội chiếm tỷ lệ cao của dân tộc Kinh là F1a, B5a, M7b1 và B4 trong đó nhóm F1a chiếm tỷ lệ cao nhất là 21% (43/206 mẫu dân tộc Kinh). Đối với dân tộc Mường nhóm F1a cũng chiếm tỷ lệ cao nhất là 16% (16/100 mẫu dân tộc Mường), sau đó lần lượt là nhóm M7b1, B và nhóm B5a. Biểu đồ trên cũng cho thấy có sự gần gũi giữa dân tộc Kinh với dân tộc Mường hơn so với dân tộc Chăm và dân tộc Khmer. Các nhóm đơn bội phổ biến ở dân tộc Chăm là B5a, M, B4c, N9a, R9b trong đó haplogroup B5a chiếm tỷ lệ cao nhất (18/113) 15,9%. Ở dân tộc Khmer haplogroup M chiếm tỷ lệ cao nhất (18/98) 18,4%, sau đó là các nhóm F1a, B, B5a và nhóm D4.

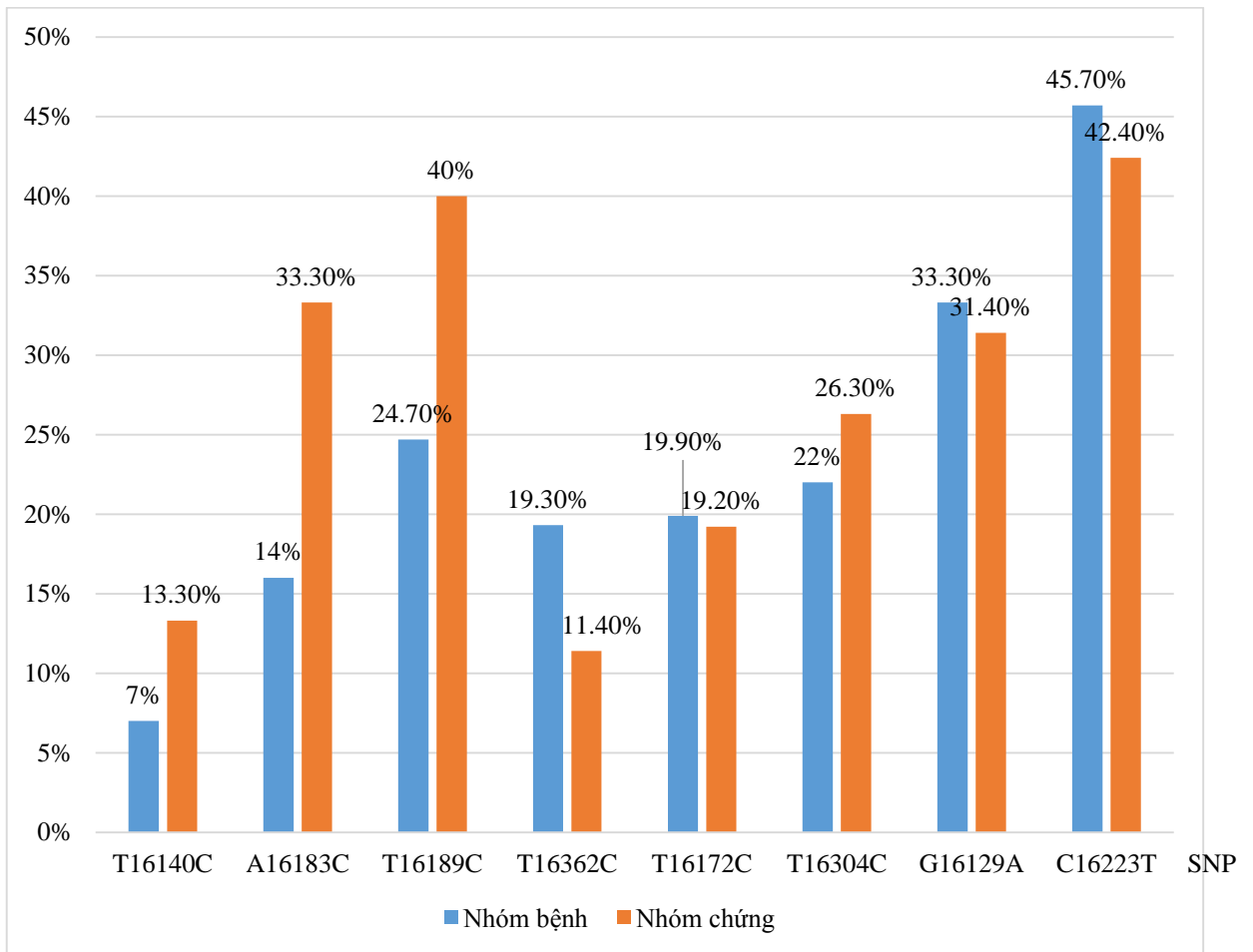
### 3.4. Tỷ lệ một số SNP trên vùng HV1 của DNA ty thể ở nhóm bệnh nhân bị ung thư vú và nhóm nữ bình thường

Để tìm hiểu về đa hình DNA ty thể trên bệnh nhân ung thư vú chúng tôi đã giải trình tự vùng HV1 của 186 mẫu bệnh nhân bị ung thư vú. Kết quả được đối chiếu với trình tự chuẩn trên Genbank để phát hiện đa hình. Từ đó phân tích so sánh với 255 mẫu nữ đối chứng (được chọn từ tất cả các mẫu nữ trên 517 mẫu người bình thường của 4 dân tộc ở trên, có 255 mẫu nữ/517 mẫu). Chúng tôi đã thống kê được 184 loại đa hình trên vùng HV1 của 186 mẫu bệnh nhân bị ung thư vú trong đó chủ yếu là đa hình thay thế nucleotid. Mẫu có nhiều đa hình nhất là 26 đa hình, mẫu có ít đa hình nhất là 1 đa hình. Một số loại SNP thấy hay gặp ở các mẫu nghiên cứu (bảng 3.13)

**Bảng 3.13. Bảng tỷ lệ một số đa hình hay gặp trên vùng HV1 của mtDNA ở bệnh nhân ung thư vú**

Loại SNP	Số lượng (n=186)	Tỷ lệ %
<b>G16129A</b>	62	33,3
<b>T16172C</b>	37	19,9
<b>A16183C</b>	26	14,0
<b>T16189C</b>	46	24,7
<b>C16223C</b>	85	45,7
<b>T16304C</b>	41	22,0
<b>T16362C</b>	36	19,4

Nhận xét: Bảng trên cho thấy có nhiều SNP hay gặp trên vùng HV1 ở bệnh nhân ung thư vú, trong đó đa hình C16223T gặp với tỷ lệ cao nhất trong số các loại SNP (85/186) mẫu bệnh nhân) chiếm 45,7%.



***Biểu đồ 3.3. Tỷ lệ một số SNP trên vùng HV1 mtDNA của 2 nhóm***

Nhận xét: SNP T16140C, A16183C, T16189C và T16304C ít gặp ở nhóm bệnh hơn nhóm chứng. SNP T16362C gặp ở nhóm bệnh nhiều hơn so với nhóm chứng. SNP T16172C, G16129A và C16223T là những đa hình có tỷ lệ gặp gần như nhau ở cả nhóm bệnh và nhóm chứng.

**Bảng 3.14. Tỷ lệ một số SNP trên vùng HV1 mtDNA của nhóm ung thư vú và nhóm nữ bình thường**

SNP	Nhóm K vú (n= 186)		Nhóm chứng (n= 255)		p	OR	95%CI
	Số lượng	Tỷ lệ %	Số lượng	Tỷ lệ %			
<b>T16140C</b>	13	7%	34	13,3%	0,033	0,488	0,25-0,954
<b>T16172C</b>	37	19,9%	49	19,2%	0,859	1,044	0,649-1,681
<b>A16183C</b>	26	14%	85	33,3%	<i>0,0001</i>	0,325	0,199-0,53
<b>T16189C</b>	46	24,7%	102	40%	<i>0,001</i>	0,493	0,325-0,748
<b>C16223T</b>	85	45,7%	108	42,4%	0,485	1,145	0,783-1,68
<b>T16362C</b>	36	19,3%	29	11,3%	<b>0,02</b>	<b>1,87</b>	<b>1,1-3,18</b>
<b>T16140C, T16189C</b>	10	5,4%	32	12,5%	0,011	0,396	0,189-0,827
<b>A16183C, T16189C</b>	24	12,9%	45	17,6%	<i>0,0001</i>	0,296	0,179-0,489
<b>C16223T, T16362C</b>	29	15,6%	16	6,3%	<b>0,001</b>	<b>2,759</b>	<b>1,45-5,25</b>

Nhận xét:

SNP T16172C, T16189C, A16183C, C16223T đều chiếm tỷ lệ cao ở cả nhóm bệnh và nhóm chứng (đều có tỷ lệ > 15%). SNP T16140C, A16183C và T16189C ở nhóm chứng chiếm tỷ lệ cao hơn nhóm bệnh (lần lượt là 13,3%, 33,3% và 40% so với 7%, 14% và 24,7%). SNP T16362C ở nhóm bệnh

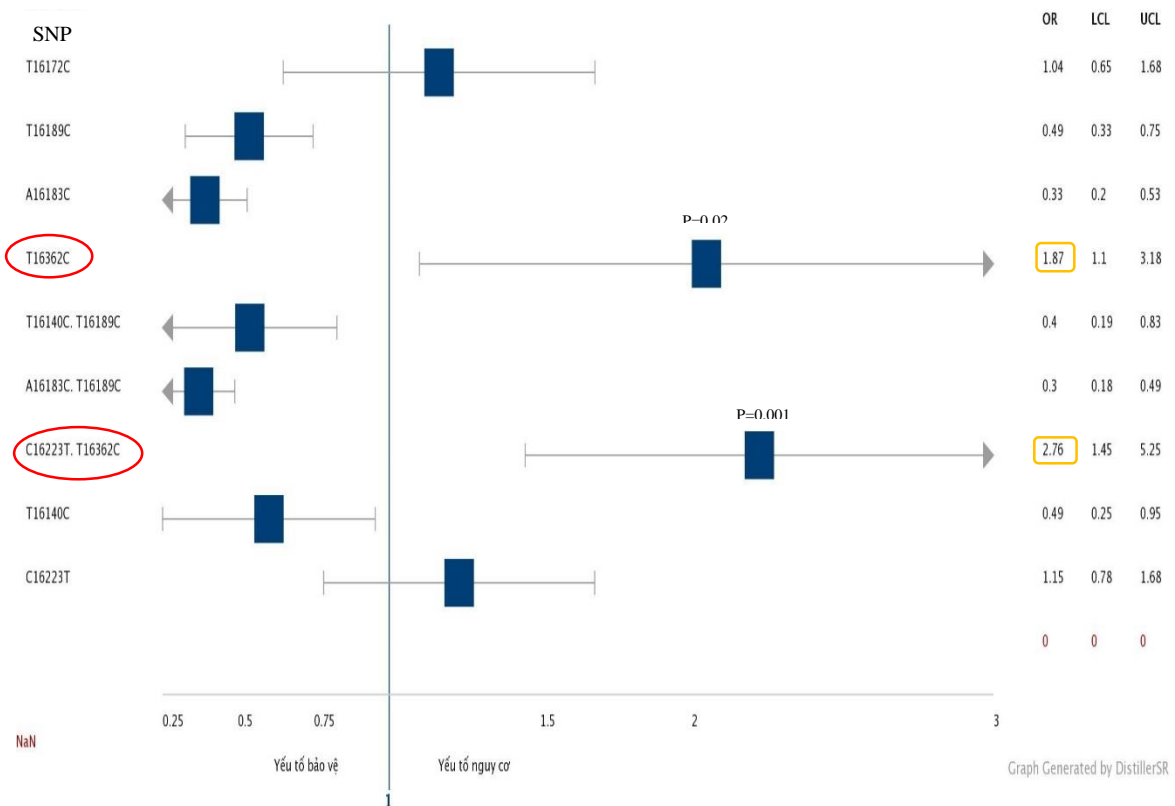
chiếm tỷ lệ cao hơn nhóm chứng (chiếm 19,3% so 11,4%). Haplotype mtDNA HV1/HV2 T16140C, T16189C (có đồng thời cả 2 SNP) và haplotype A16183C, T16189C ở nhóm bệnh có tỷ lệ gặp thấp hơn nhóm chứng (5,4% so với 12,5% và 12,9% so với 17,6%). Haplotype C16223T, T16362C ở nhóm bệnh có tỷ lệ gặp cao hơn nhóm chứng (15,6% so với 6,3%).

Tỷ lệ gặp SNP T16362C ở nhóm bệnh cao hơn nhóm chứng và có sự khác biệt có ý nghĩa thống kê với  $p=0,02$ ,  $OR=1,87$  và 95% CI (1,1-3,18) (bảng 3.5). Khi có đồng thời cả 2 SNP C16223T và T16362C (haplotype C16223T, T16362C) có khả năng bị ung thư vú tăng gấp 2,759 lần, sự khác biệt có ý nghĩa thống kê với  $p=0,001$ , 95% CI (1,45-5,25).

SNP A16183C, T16189C hoặc khi có đồng thời có cả A16183C và T16189C hoặc T16140C và T16189C khả năng ít mắc bệnh ung thư vú hơn, sự khác biệt có ý nghĩa thống kê với  $p < 0,05$  lần lượt là  $p=0,0001$ ,  $p=0,001$ ,  $p=0,0001$  và  $p=0,011$  (bảng 3.5).

Như vậy, có thể thấy có các đa hình trong vùng HV1 của DNA ty thể có thể là yếu tố nguy cơ của bệnh ung thư vú nhưng cũng có những đa hình mang ý nghĩa bảo vệ với bệnh ung thư vú. Biểu đồ dưới đây biểu diễn mối tương quan của các SNP vùng gen ty thể HV1 với bệnh ung thư vú:

## Nguy cơ ung thư vú



**Biểu đồ 3.4: Biểu thị mối tương quan giữa SNP trên vùng HV1 của DNA ty thể với bệnh ung thư vú.**

**Nhận xét:** Biểu đồ 3.4 cho thấy một số SNP và haplotype HV1/HV2 của DNA ty thể có thể có ảnh hưởng đến bệnh ung thư vú. Trong đó ảnh hưởng rõ nhất là SNP T16362C và haplotype C16223T, T16362C là yếu tố nguy cơ với bệnh ung thư vú với OR lần lượt là 1,87 và 2,76 và  $p < 0,05$ . SNP T16189C, A16183C, T16140C và 2 haplotype T16140C, T16189C và A16183C, T16189C có thể là yếu tố bảo vệ đối với bệnh ung thư vú.

## **Chương 4**

### **BÀN LUẬN**

#### **4.1. Đặc điểm mẫu nghiên cứu**

Việt Nam là một quốc gia đa văn hoá với 54 dân tộc anh em, tổng dân số ước tính đạt 90.493.352 người [109] đứng thứ 3 ở Đông Nam Á (sau Indonexia và Philipin), đứng thứ 8 trong khu vực Châu Á, và đứng thứ 14 trong số những nước đông dân nhất khu vực và thế giới. Dân số Việt Nam đã tăng thêm khoảng 4,64 triệu người trong vòng 5 năm kể từ tổng điều tra dân số năm 2009. Do vậy, việc nghiên cứu đặc điểm và nguồn gốc của mỗi dân tộc cũng như mối liên quan giữa các dân tộc là một việc làm cần thiết, nhất là trong điều kiện phát triển kinh tế xã hội hiện nay, việc di cư giữa các vùng, việc kết hôn giữa các dân tộc ngày càng trở nên phổ biến.

Trong nghiên cứu này, chúng tôi lựa chọn các mẫu nghiên cứu được lấy từ các cá thể người Việt Nam trưởng thành, khỏe mạnh thuộc bốn dân tộc Kinh, Mường, Chăm, Khmer, những địa điểm được lựa chọn để lấy mẫu là nơi có số dân của các dân tộc trên tập trung đông nhất. Bởi vì, dân tộc Kinh là dân tộc chủ yếu có số dân lớn nhất cả nước (chiếm 87%) mẫu nghiên cứu được lấy ở Hà Nội và thành phố Hồ Chí Minh. Dân tộc Mường (mẫu nghiên cứu được lấy ở Hòa Bình - nơi có số người dân tộc Mường chiếm tỷ lệ cao nhất và chiếm đến 64% dân số toàn tỉnh) và dân tộc Khmer (mẫu nghiên cứu được lấy ở Sóc Trăng - nơi có tới 30,7% dân số toàn tỉnh là người Khmer và tổng số người Khmer ở đây cũng chiếm 31,5 tổng số người Khmer cả nước) đây là hai dân tộc đại diện cho các dân tộc anh em có số dân tương đối lớn trong 53 dân tộc còn lại (hơn một triệu dân). Dân tộc Chăm (mẫu nghiên cứu được lấy ở Bình Thuận - nơi có tới 30% dân số người dân tộc Chăm) là đại diện cho nhóm các dân tộc thiểu số có dân số ít (chỉ có trên 150.000 dân) [110].

Ở Việt Nam trong những năm gần đây, một số khía cạnh nhân chủng học như khảo cổ, nhân trắc, văn hoá - xã hội, ngôn ngữ... đã được quan tâm nghiên cứu nhiều, những nghiên cứu này cho thấy các tộc người Kinh, Mường, Chăm và Khmer có những điểm tương đồng. Tuy nhiên, khía cạnh về gen học chưa được quan tâm nghiên cứu, chưa có nhiều nghiên cứu về gen học đánh giá mối liên quan giữa các dân tộc Việt Nam. Hơn nữa, Việt Nam là quốc gia có nhiều dân tộc nên việc nghiên cứu đặc điểm về gen học của tất cả các dân tộc là rất khó khăn và tốn kém nên hầu hết các nghiên cứu mới chỉ nghiên cứu trên một số ít dân tộc và số lượng mẫu hạn chế.

Nghiên cứu của chúng tôi đánh giá sự đa dạng di truyền vùng HV1 và HV2 của DNA ty thể ở bốn dân tộc ở Việt Nam (dân tộc Kinh, dân tộc Mường, dân tộc Khmer và dân tộc Chăm) nhằm khảo sát đặc điểm về gen học của một số dân tộc người Việt Nam, để từ đó có thể đưa ra những bằng chứng khoa học về nguồn gốc, đặc điểm riêng của các chủng tộc người này, cũng như có thể định hướng xem xét về mối liên quan giữa đa hình DNA ty thể với một số loại bệnh, hướng tới việc bảo tồn và lưu trữ nguồn gen phục vụ cho công tác bảo vệ và chăm sóc sức khỏe người Việt Nam.

#### **4.2. Phân tích tính đa hình vùng HV1 và HV2 của DNA ty thể trên một số dân tộc người Việt Nam bằng phương pháp giải trình tự gen**

Sản phẩm PCR của vùng HV1 và HV2 được tiến hành giải trình tự tự động trên máy 3100-Avant Genetic Analyzer của hãng ABI-PRISM. Kết quả giải trình tự được phân tích trên phần mềm CLC Main Workbench 6.0.1. So sánh với trình tự chuẩn trên GenBank để xác định đa hình.

Nghiên cứu về đa hình đơn nucleotid ngoài những giá trị trong các nghiên cứu phát sinh chủng tộc của loài người, còn có những giá trị thiết thực đối với nghiên cứu về sức khỏe con người thông qua việc cung cấp một nguồn



dữ liệu các đa hình di truyền liên quan đến bệnh và các đáp ứng của cá thể với các tác nhân trong điều trị. Từ đó các nhà khoa học có thể rút ra được nhiều thông tin về nguồn gốc bệnh và các cách dự phòng bệnh, chẩn đoán và điều trị các bệnh đó. Các bệnh phổ biến như ung thư, tim mạch, tiểu đường, rối loạn chuyển hóa, hen phế quản... thường do tác động tổng hợp của nhiều nhân tố bao gồm cả di truyền và môi trường.

Theo một giả thuyết về đa hình phổ biến - bệnh phổ biến, nguy cơ mắc các bệnh phổ biến bị ảnh hưởng bởi các đa hình di truyền xuất hiện tương đối phổ biến trong quần thể. Tuy rằng chưa có nhiều bằng chứng chứng minh cho giả thuyết này, nhưng ngày càng nhiều các đa hình di truyền được phân bố rộng trong hệ gen có liên quan tới các bệnh phổ biến đã được phát hiện, bao gồm các đa hình liên quan tới các bệnh ung thư, tự miễn, tâm thần phân liệt, tiểu đường, đột quỵ và tim mạch... Các đa hình di truyền đóng vai trò trong kéo dài tuổi thọ hoặc khả năng kháng bệnh có thể sẽ được xác định, đưa tới những phương pháp điều trị mới với nhiều lợi ích hơn.

Trình tự di truyền của các cá thể người tương đồng với nhau tới 99,9%, chỉ khác nhau một phần rất nhỏ 0,1%. Những khác biệt trong từng nucleotid (SNP), về cơ bản chính là dạng phổ biến nhất của đa hình di truyền. Khi so sánh các nhiễm sắc thể của hai người hoàn toàn không có quan hệ họ hàng gần gũi với nhau, có thể thấy rằng các trình tự DNA của họ có thể giống nhau tới hàng trăm nucleotid. Tuy nhiên, trung bình trên mỗi 1200 nucleotid trình tự sẽ có 1 nucleotid sai khác nhau. Ví dụ, ở một người, tại một vị trí nào đó trong trình tự DNA có thể là nucleotid A, trong khi đó, ở một người khác có thể là G, hoặc cũng có thể bị mất nucleotid đó, hoặc thêm một hoặc một số nucleotid khác. Mỗi dạng sai khác nhau như vậy tại một vùng trên nhiễm sắc thể được gọi là một allen, và tập hợp các allen trên các nhiễm sắc thể của một người được gọi là một kiểu gen. Bằng cách xác định hầu hết 10 triệu SNP ước

tính xuất hiện phổ biến trong hệ gen người, dự án HapMap quốc tế đã xác định cơ sở của phần lớn tính đa dạng di truyền của loài người [33].

### **Phân nhóm đơn bội (haplogroup) DNA ty thể:**

Đối với các nhà di truyền học, các SNP đóng vai trò như những chỉ thị để định vị các gen trong trình tự DNA. Các đa hình di truyền nằm gần nhau thường có xu hướng được di truyền cùng nhau. Những nghiên cứu gần đây trên thế giới đã phân loại các nhóm đơn bội của DNA ty thể dựa vào các vị trí đa hình đặc trưng trên DNA ty thể mà đặc biệt là các vị trí đa hình trên vùng HV1 và HV2. Nghiên cứu của Yao và cộng sự năm 2002 trên 263 người thuộc 6 dân tộc người Trung Quốc đã phân loại được các nhóm đơn bội theo các vị trí đa hình đặc trưng trên vùng HV1 và HV2 của DNA ty thể. Cụ thể Yao đã phân loại được 42 nhóm đơn bội và dưới đơn bội trong đó nhóm D4 chiếm tỷ lệ cao nhất (27/263), nhóm A chiếm (18/263), nhóm F1a chiếm (15/263), các nhóm có tỷ lệ thấp nhất là G2, T1, D, N, B5 (chỉ có 1/263 cá thể nghiên cứu) [32]. Nghiên cứu của chúng tôi đã xác định được 438 haplotypes HV1/HV2 mtDNA của 517 mẫu nghiên cứu thuộc 4 dân tộc Kinh, Chăm, Mường, Khmer, phân loại thành 50 nhóm đơn bội và dưới đơn bội trong đó nhóm F1a chiếm tỷ lệ cao nhất (81/517 mẫu nghiên cứu tương ứng 15,67%), nhóm B5a (56/517 mẫu nghiên cứu tương ứng 10,83%), nhóm M (46/517 mẫu nghiên cứu tương ứng 8,9%), nhóm M7b1 (40/517 mẫu nghiên cứu tương ứng 7,74%), các nhóm đơn bội có tần số xuất hiện thấp nhất (chỉ có 1 cá thể trong tổng số 517 mẫu nghiên cứu của cả bốn dân tộc) là nhóm B6, D4e, D5a, E, F1c, F2a, F3a, G2a, N, M9b, N21, và nhóm U5a (bảng 3.11). Có được kết quả như vậy có lẽ do số lượng mẫu trong nghiên cứu của chúng tôi lớn hơn so với nghiên cứu của Yao và cộng sự. Có sự tương đồng về các nhóm đơn bội giữa hai nghiên cứu: nhóm đơn bội F1a đều chiếm tỷ lệ cao, nhóm G2 đều chiếm tỷ lệ thấp, điều này có lẽ là do các

dân tộc Việt Nam và các dân tộc Trung Quốc đều là những dân tộc ở Châu Á có quan hệ chủng tộc và địa lý gần gũi.

Gần đây, nghiên cứu của Nguyễn Thùy Dương và cộng sự trên 609 người Việt Nam thuộc 5 nhóm hệ ngôn ngữ phổ biến của Việt Nam cũng cho thấy tỷ lệ nhóm F1 là 19,38%, nhóm M7 là 9,36% và nhóm B5 là 7,22% [111]. Kết quả này cũng tương đồng với kết quả nghiên cứu của chúng tôi khi phân loại được nhóm F1a là 15,67%, nhóm B5a là 10,83%, nhóm M7b1 là 7,74%.

Năm 2009 Han Jun Jin và cộng sự khi nghiên cứu trên 445 cá nhân, được thu thập từ bảy nhóm dân cư Đông Á (Hàn Quốc, Hàn Quốc-Trung Quốc (người gốc Hàn Quốc hiện đang sống ở Trung Quốc), Mông Cổ, Mãn Châu, Hán (Bắc Kinh), Việt Nam và Thái Lan) cho kết quả: haplogroup chiếm tỷ lệ cao nhất của 4 nhóm người Hàn Quốc, Hàn Quốc-Trung Quốc (người gốc Hàn Quốc hiện đang sống ở Trung Quốc), người Mãn Châu, người Hán (Bắc Kinh) đều là haplogroup D4 (lần lượt là 44/185, 11/51, 8/40, 5/40), của nhóm người Thái Lan là haplogroup F1b (8/40), nhóm người Mông Cổ là haplogroup C (8/47). Ở nhóm người Việt Nam (42 mẫu) được phân loại vào 23 haplogroup trong đó haplogroup F1a có tỷ lệ cao nhất 10/42 (23,8%) [112]. Trong nghiên cứu của chúng tôi đã phân loại được 50 haplogroups mtDNA, tỷ lệ haplogroup F1a trên 517 mẫu trong nghiên cứu của chúng tôi cũng chiếm tỷ lệ cao nhất phù hợp với nghiên cứu trên là 81/517 mẫu (15,67%). Còn lại các haplogroup khác M7b1 (40/517) chiếm 7,74%, M (46/517) chiếm 8,9%, B5a (56/517) chiếm 10,83% và các haplogroup có tần số xuất hiện thấp nhất (chỉ có 1 cá thể trong tổng số 517 mẫu nghiên cứu của cả bốn dân tộc) là nhóm B6, D4e, D5a, E, F1c, F2a, F3a, G2a, N, M9b, N21, và nhóm U5a (bảng 3.11). Như vậy, có thể thấy mỗi chủng tộc người khác nhau có những nhóm haplogroup mtDNA phổ biến khác nhau và những

chúng tộc người có mối quan hệ gần gũi có thể có cùng một nhóm haplogroup mtDNA phổ biến.

Khi xét riêng từng dân tộc, 3 nhóm đơn bội phổ biến của dân tộc Kinh là F1a, B5a, M7b1 trong đó nhóm F1a chiếm tỷ lệ cao nhất là 21%. Đây cũng là 3 nhóm đơn bội chiếm tỷ lệ cao của dân tộc Mường và nhóm F1a cũng chiếm tỷ lệ cao nhất là 16% (biểu đồ 3.2). Như vậy, có thể thấy giữa dân tộc Kinh và dân tộc Mường dường như có quan hệ gần gũi hơn điều này có lẽ là do dân tộc Kinh và dân tộc Mường cùng có chung nguồn gốc là người Việt - Mường cổ và cùng thuộc nhóm ngôn ngữ Việt - Mường. Giữa dân tộc Chăm và Khmer cũng có sự tương đồng về các nhóm đơn bội phổ biến: ở dân tộc Chăm là nhóm B5a, M, B4c, N9a trong đó haplogroup B5a chiếm tỷ lệ cao nhất 15,9%, ở dân tộc Khmer là nhóm M, F1a, B, B5a (biểu đồ 3.2). Điều này cũng phù hợp với nghiên cứu của Peng và cộng sự khi so sánh người Chăm với một số dân tộc khác ở Đông Nam Á khác cho thấy rằng người Chăm có mối quan hệ gần gũi hơn với quần thể người Môn- Khmer [113].

Một nghiên cứu khác năm 2017 trên 622 người Việt Nam (399 người Kinh, 62 người Tày, 115 người Mông, 23 người Hoa, 21 người Nùng, 1 người Chăm, 1 người Thái) cũng cho thấy biến thể di truyền quan sát được ở người Việt Nam rất phù hợp với các mẫu mtDNA được quan sát ở Đông Nam Á (được coi là khu vực đa hình và đa hình nhất của Châu Á) và tính đa dạng mtDNA rất cao trên toàn lãnh thổ Việt Nam [114]. Tác giả cũng đã chỉ ra tất cả người Việt Nam mang haplotypes Đông Nam Á, có sự phân tầng địa lý và dân tộc vừa phải trong đó phần lớn mtDNA người dân Việt Nam thuộc nhóm đơn bội M7 (20%), R9, F (27%). Các nhóm haplogroup khác như A, B, C và D được đại diện ở lãnh thổ Việt Nam nhưng với tần suất thấp hơn, ngoại trừ haplogroup B ở Lào Cai (chiếm 41%), haplogroup M phổ biến hơn ở miền Bắc và ở miền Trung so với miền Nam [114]. Kết quả này cũng phù hợp với

ngiên cứu của chúng tôi khi cho thấy haplogroup M, M7b1 và haplogroup B cũng là những nhóm đơn bội phổ biến của 4 dân tộc Kinh, Mường, Chăm và Khmer. Irwin và cộng sự (2008) nghiên cứu trên người Việt Nam ở Hà Nội cho thấy tỷ lệ của các nhóm haplogroup N, A, C và D thấp ( $< 5\%$ ) [115]. Kết quả này cũng phù hợp với nghiên cứu của chúng tôi với tỷ lệ haplogroup A là 0,77%, B là 4,06%, C là 2,32%, N là 0,19% (đều  $< 5\%$ ) (bảng 3.12).

Những năm gần đây trên thế giới cũng đã có nhiều nghiên cứu phân tích về mtDNA nhằm làm sáng tỏ các mối quan hệ di truyền tiến hóa của các dân tộc, các quần thể người dựa trên các đặc trưng của mtDNA. Bodner và cộng sự khi nghiên cứu trên 214 mẫu người Lào đã phân loại được 64 nhóm đơn bội trong đó 47 haplotypes HV1/HV2 mtDNA (chiếm 28%) của các mẫu Lào đã được tìm thấy trong quần thể Việt Nam. Nghiên cứu cũng chỉ ra không có sự khác biệt đáng kể về cấu trúc di truyền được tìm thấy giữa mẫu dân số Lào và Việt Nam, điều này có thể chỉ ra dòng gen mở rộng do di cư giữa hai quốc gia. Các nhóm đơn bội phổ biến nhất ở Lào là B5a (12%), F1a (17,3%), C7 và M7b1 (mỗi nhóm 6%) [116]. Kết quả này tương đồng với nghiên cứu của chúng tôi khi cho kết quả các nhóm đơn bội phổ biến ở người Việt Nam là nhóm F1a (15,67%), nhóm M7b1 (7,74%), nhóm B5a (10,83%) (biểu đồ 3.1). Có sự tương đồng này có lẽ do Lào và Việt Nam là 2 quốc gia Đông Nam Á, có quan hệ gần gũi về mặt vị trí địa lý, chính trị, có sự di cư, hợp tác về nhiều mặt của đời sống xã hội nên có sự tương đồng về mặt dân tộc.

Một nghiên cứu về DNA ty thể trên một số lượng mẫu lớn 1054 người thuộc 14 dân tộc người Campuchia, thông qua vùng điều khiển D-loop và khu vực mã hóa, tổng số 1.000 mẫu trong số 1.054 mẫu nghiên cứu (chiếm 94,88%) của các cá thể người Campuchia đã được phân loại thành 69 nhóm haplogroup đã biết ở các quốc gia Đông Nam Á và Đông Á, 54 cá thể còn lại được phân thành tám nhóm haplogroup mới được xác định (M59, M69, M78,

N7, M68, M68a, M3d, M3d1) chiếm 5,12% tổng số. Nhìn chung, các nhóm đơn bội chiếm ưu thế được quan sát ở dân số Campuchia là nhóm B5a 288/1054 (27,3%), F1a 188/1054 (17,8%), M12b (8,25%), R22 (5,79%) và B4 (5,60%) [117]. Kết quả này cũng phù hợp với nghiên cứu của chúng tôi nhóm F1a và nhóm B5a là hai nhóm chiếm tỷ lệ cao nhất và cũng có nhiều phần phù hợp với kết quả về phân nhóm đơn bội trên dân tộc Khmer khi cho kết quả các nhóm đơn bội phổ biến ở dân tộc Khmer là nhóm B4 (5%), B5a (8,2%), B (12,2%) và nhóm F1a là 14,3% (bảng 3.11, biểu đồ 3.2). Có sự tương đồng như vậy có lẽ do dân tộc Khmer là dân tộc chiếm đa số ở Campuchia và người Khmer ở Việt Nam hay ở Campuchia đều thuộc ngữ tộc Môn - Khmer trong ngữ hệ Nam Á.

Từ các phân tích ở trên cho thấy một số nhóm đơn bội là phổ biến ở các nước châu Á (bảng 4.1), trong đó đặc biệt là 3 quốc gia ở Đông Nam Á (Việt Nam, Lào, Campuchia) có những đặc điểm di truyền mtDNA là gần gũi hơn cả khi hai haplogroup mtDNA B5a và F1a là phổ biến nhất ở 3 quốc gia này. Trong khi Trung Quốc, Hàn Quốc dường như cũng có đặc điểm di truyền mtDNA gần gũi hơn so với các quốc gia còn lại vì đều có haplogroup D4 là nhóm đơn bội phổ biến nhất, riêng ở Malaysia thì haplogroup M là phổ biến nhất. Như vậy, có thể thấy mặc dù các quốc gia khác nhau có nhiều chủng tộc người khác nhau nhưng cũng vẫn có thể có một số đặc điểm chung.

**Bảng 4.1: Bảng tỷ lệ một số nhóm đơn bội mtDNA phổ biến ở Việt Nam và một số nước ở châu Á**

Nghiên cứu	Quốc gia	Haplogroup					
		B4a	B5a	D4	F1a	M	M7b1
Yao và cs [32]	Trung Quốc (n=263)	9/263	6/263	<b>27/263</b>	<b>15/263</b>	10/263	6/263
Han-Jun Jin và cs [112]	Hàn Quốc (n=185)	11/185	2/185	<b>44/185</b>	8/185	1/185	1/185
Rashid và cs [118]	Malaysia (n=248)	10/248	12/248	1/248	22/248	<b>31/248</b>	27/248
Boner và cs [116]	Lào (n=214)	7/214	<b>26/214</b>	6/214	<b>37/214</b>	17/214	13/214
Zhang và cs [117]	Campuchia (n=1054)	52/1054	<b>288/1054</b>		<b>188/1054</b>		
<b>Nghiên cứu này (2019)</b>	Việt Nam (n=517)	15/517	<b>56/517</b>	13/517	<b>81/517</b>	46/517	40/517

Một nghiên cứu khác khi nghiên cứu trên 208 người Trung Quốc cho thấy tỷ lệ haplogroup D là nhóm đơn bội phổ biến nhất (25,96%). Các nhóm đơn bội phổ biến ở miền bắc Trung Quốc (A, C, D, G, M8, Y và Z) và các nhóm đơn bội phổ biến ở miền nam Trung Quốc (B, F, M7, N9 và R9) lần lượt chiếm 48,56% và 46,63% [119]. Năm 2018, Xu và cộng sự khi giải trình tự các trình tự vùng HV1, HV2 và phân tích RFLP trên vùng mã hóa của 315 cá thể được thu thập tại tỉnh Liêu Ninh, Trung Quốc cũng cho tỷ lệ nhóm đơn

bội phổ biến nhất là nhóm D4 (12,93%). Nghiên cứu cũng đã chỉ ra rằng khoảng cách di truyền của quần thể Liêu Ninh từ nhóm Tây Tạng là rất xa, trong khi đó từ nhóm Miao là tương đối gần [120]. Trong nghiên cứu của chúng tôi nhóm D4 chỉ chiếm tỷ lệ 2,51% (13/517), sự khác nhau này có thể mang tính đặc trưng của các dân tộc giữa các nước khác nhau.

Nghiên cứu giải trình tự vùng điều khiển mtDNA từ 407 cá nhân Rumani (nằm ở khu vực đông nam châu Âu) tác giả thấy có 306 haplotypes mtDNA khác nhau và đã phân loại thành các nhóm đơn bội chủ yếu bao gồm các dòng Tây Âu Á H (31,7%), U (12,8%), J (10,8%), R (10,1%), T (9,1%), N (8,1%), HV (5,4%), K (3,7%), HV0 (4,2%), ngoài ra còn có một phần nhỏ haplogroup M châu Á (3,4%) và haplogroup L châu Phi (0,7%) [121]. Nghiên cứu của chúng tôi cho kết quả các haplogroup chiếm tỷ lệ cao nhất là F1a (15,7%), B5a (10,83%), M (8,9%) và M7b1 (7,74%) đây cũng là những nhóm đơn bội phổ biến ở châu Á (bảng 4.1). Trong khi đó nhóm U5a, nhóm N chiếm tỷ lệ thấp nhất (1/517 cá thể). Điều này cho thấy các nhóm haplogroup DNA ty thể giữa các quần thể người châu Âu và châu Á có những đặc điểm khác biệt rõ rệt.

Năm 2016 Đỗ Mạnh Hưng và cộng sự khi nghiên cứu vùng HV2 trên mtDNA của 169 người khỏe mạnh thuộc bốn dân tộc Kinh, Mường, Jarai và Ê-đê Việt Nam (40 người Kinh, 47 người Mường, 34 người Ê-đê, 48 người Jarai). Kết quả cho thấy haplogroup mtDNA các mẫu nghiên cứu phần lớn thuộc về 3 haplogroup R, B và F, trong đó các cá thể người Kinh và người Mường có sự đa dạng hơn về thành phần haplogroup so với người Jarai và người Ê-đê. Người Kinh và người Mường có sự tương đồng rất lớn về thành phần các haplogroup. Sự đa dạng về các phân nhóm đơn bội của người Kinh và người Mường cũng là đặc trưng của các quần thể người sinh sống trên



phần lục địa của khu vực Đông Nam Á và Đông Á. Haplogroup M ở người Kinh chỉ chiếm 3% số mẫu cá thể người Kinh [122]. Kết quả nghiên cứu trên cũng có nhiều điểm phù hợp với nghiên cứu của chúng tôi khi thấy rằng có sự tương đồng về các haplogroup mtDNA ở dân tộc Kinh và dân tộc Mường (ba nhóm đơn bội phổ biến của dân tộc Kinh và Mường đều là F1a, B5a, M7b1 trong đó haplogroup F1a chiếm tỷ lệ cao nhất lần lượt là 21%, 16%), và haplogroup M ở người dân tộc Kinh cũng chỉ chiếm 3,8% (8/206 mẫu người Kinh). Tuy nhiên, trong nghiên cứu của chúng tôi phân nhóm các haplogroup mtDNA chi tiết và cụ thể hơn điều này có lẽ do chúng tôi nghiên cứu trên cả hai vùng HV1 và HV2 mtDNA trong khi tác giả chỉ nghiên cứu trên vùng HV2 của DNA ty thể và cỡ mẫu trong nghiên cứu của chúng tôi cũng lớn hơn nhiều so với nghiên cứu trên.

#### **Tính đa hình trên vùng HV1 và HV2 của DNA ty thể:**

Trong nghiên cứu này, chúng tôi đã phát hiện được 3 loại SNP mới trên vùng HV1 của DNA ty thể người Việt Nam chưa thấy công bố trên Mitomap đó là SNP 16038DelA (công bố trên Mitomap nucleotid A tại vị trí 16038 biến đổi thành AA/AC/G/T); SNP G16084C (công bố trên Mitomap nucleotid G tại vị trí 16084 chuyển thành A/GG/T) và SNP A16515C (SNP đã công bố trên Mitomap là 16515DelA). Hai đa hình đơn nucleotid mới là SNP 16038DelA và SNP G16084C (hình 3.14, 3.15) được tìm thấy trên dân tộc Mường còn SNP A16515C (hình 3.16) được tìm thấy ở dân tộc Kinh. Đây có thể là những SNP đặc trưng cho dân tộc Việt Nam bởi vì dân tộc Kinh và dân tộc Mường là dân tộc sống chủ yếu tại Việt Nam và cũng chiếm số lượng dân số lớn nhất Việt Nam.

Chúng tôi cũng đã thống kê ra 16 vị trí đa hình hay gặp trên vùng HV1 và HV2 của DNA ty thể (đều gặp trên 10% ở các mẫu nghiên cứu) trong đó

có 12 vị trí đa hình có tỷ lệ gặp trên 20% (bảng 3.12) đó là các SNP: C16223T (212/517) chiếm tỷ lệ 41% mẫu nghiên cứu, T16189C (203/517) chiếm tỷ lệ 39%, G16129A và A16183C (171/517) chiếm tỷ lệ 33%, T16304C (149/517) chiếm tỷ lệ 28,8%, T16172C (109/517) chiếm tỷ lệ 21%. Trên vùng HV2 là A263G (517/517) gặp 100% mẫu nghiên cứu, A73G (515/517) chiếm tỷ lệ 99,6%, 315insC (498/517) chiếm tỷ lệ 96,3%, 309insC (292/517) chiếm tỷ lệ 56,4%, 249DelA (128/517) chiếm tỷ lệ 24,7%, C150T (121/517) chiếm tỷ lệ 23,4%. Đặc biệt đa hình A263G gặp ở 100% các mẫu nghiên cứu. Kết quả này cũng phù hợp với kết quả trong nghiên cứu của Nguyễn Đăng Tôn và cộng sự khi nghiên cứu trên 78 cá thể người Việt Nam đã thống kê được 10 vị trí đa hình hay gặp giống như trong nghiên cứu của chúng tôi và cũng có tỷ lệ tương tự đó là các đa hình T16172C (22/78), A16183C (23/78), T16189C (31/78), C16223T (26/78), T16304 (26/78), C150T (20/78), 249DelA (25/78), A263G (75/78), 315C (24/78), A73G (78/78) trong đó đa hình A73G gặp ở 100% mẫu nghiên cứu, đa hình A263G gặp ở 96,2% mẫu nghiên cứu [90]. Tuy nhiên, nghiên cứu của chúng tôi đã chỉ ra nhiều đa hình hay gặp trên vùng HV1, HV2 hơn có thể do cỡ mẫu nghiên cứu của từng dân tộc và cả 4 dân tộc của chúng tôi đều lớn hơn.

Nghiên cứu của Nguyễn Thy Ngọc và cộng sự năm 2018 cho thấy đa hình phổ biến trên dân tộc Kinh và dân tộc Mảng cũng phù hợp với nghiên cứu của chúng tôi như A73G (88/88 mẫu nghiên cứu của cả hai dân tộc), A263G (87/88), C150T (29/88), G16129A (30/88), T16172C (17/88), A16183 (25/88), T16189C (30/88), C16223T (43/88) trong đó đa hình A73G gặp ở 100% và A263G gặp ở 98,9% mẫu nghiên cứu của cả hai dân tộc [123]. Các đa hình trên vùng HV1, HV2 mtDNA 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 Mảng và bốn đa hình C151T, A16162G, A16269G và T16271C xuất hiện phổ biến ở người Mảng nhưng ít xuất hiện ở người Kinh. Mặc dù cùng thuộc nhóm ngữ hệ Nam Á, hệ gen ty thể giữa 2 dân tộc Kinh và Mảng trong cộng đồng các dân tộc Việt Nam vẫn tồn tại những khác biệt đáng chú ý [123]. Trong nghiên cứu của chúng tôi các đa hình T199C (38/206 mẫu dân tộc Kinh), T16217C (27/206), A16183C (61/206) và T16189C (72/206) cũng là các đa hình xuất hiện phổ biến ở dân tộc Kinh (bảng 3.6). Ngoài ra chúng tôi cũng đã thống kê được 21 loại đa hình trên vùng HV1 và 13 loại đa hình trên vùng HV2 chỉ gặp ở dân tộc Kinh mà không thấy gặp ở dân tộc khác (bảng 3.4, 3.5)

**Bảng 4.2: So sánh tỷ lệ một số đa hình trên vùng HV1 của DNA ty thể với một số nghiên cứu khác**

Nghiên cứu	Quốc gia	Tỷ lệ %				
		SNP trên vùng HV1 (16000+)				
		T172C	A183C	T189C	T217C	C223T
Nguyễn Đăng Tôn và cs [90]	Việt Nam	28,2	29,5	39,7		33,3
Nguyễn Thy Ngọc và cs [123]	Việt Nam	19,3	28,4	34,1	19,3	48,9
Tuladhar và cs [124]	Malaysia					37,22
Fang và cs [125]	Trung Quốc	11,7	24,8	37,6	12,4	59,5
<b>Nghiên cứu này (2019)</b>	Việt Nam	<b>21</b>	<b>33</b>	<b>39</b>	<b>12</b>	<b>41</b>

**Bảng 4.3: So sánh tỷ lệ một số đa hình trên vùng HV2 của DNA ty thể với một số nghiên cứu khác**

Nghiên cứu	Quốc gia	Tỷ lệ % SNP trên vùng HV2			
		A73G	T150C	249DelA	A263G
Nguyễn Đăng Tôn và cs [90]	Việt Nam	100	25,6	32,1	96,2
Nguyễn Thy Ngọc và cs [123]	Việt Nam	100	32,9		98,8
Yao và cs [32]	Trung Quốc	100			100
Tuladhar và cs [124]	Malaysia	100			100
Fang và cs [125]	Trung Quốc		21	22,6	
<b>Nghiên cứu này (2019)</b>	Việt Nam	<b>99,6</b>	<b>23,4</b>	<b>24,7</b>	<b>100</b>

Một số vị trí đa hình hay gặp ở trên (bảng 4.2, 4.3) đều khá phổ biến và gặp ở nhiều quần thể người khác nhau trên thế giới. Như vậy có thể thấy một số đa hình hay gặp trên vùng HV1 và HV2 mtDNA ở các dân tộc Việt Nam là phổ biến, không phải là đặc điểm riêng hay có tính chất vùng miền.

Nghiên cứu của Bhinu Shova Tuladhar và cộng sự khi giải trình tự vùng gen ty thể HV1 và HV2 của 103 mẫu người Malaysia cũng cho thấy đa hình tại vị trí A73G, A263G và 315insC gặp ở 100% mẫu nghiên cứu, đa hình C16223T chiếm 37,22% [124]. Kết quả này phù hợp với nghiên cứu của chúng tôi khi xác định được đa hình A73G gặp ở 99,6%, A263G gặp ở 100%, và 315insC gặp ở 96,3%, A16223C gặp ở 41% mẫu nghiên cứu.

Đa hình tại vị trí T16189C đã tạo nên vùng lặp nucleotid Cystein (poly C) từ vị trí 16183- 16193 trên vùng siêu biến HV1. Nghiên cứu năm 2009 của Chen và cộng sự cho thấy tỷ lệ đa hình T16189C thuộc vùng lặp nucleotid

Cystein trên vùng HV1 là 25,14% [126] và sự đa dạng di truyền trong vùng lặp nucleotid Cystein của vùng HV1 có ý nghĩa quan trọng trong việc giám định hình sự gen và di truyền quần thể. Tỷ lệ đa hình T16189C trên người Tunisia được xác định là 29% [127]. Nghiên cứu của chúng tôi lại cho kết quả tỷ lệ đa hình T16189C là 39%. Như vậy, có thể thấy có sự khác nhau về tính đa hình đơn nucleotid giữa các dân tộc người ở các vùng địa lý khác nhau.

Ở dân tộc Kinh loại đa hình hay gặp nhất trên vùng HV1 của DNA ty thể là G16129A (82/206) chiếm 39,8%, ở dân tộc Chăm loại đa hình hay gặp nhất là T16189C (64/113) chiếm 56,6%, trong khi ở dân tộc Khmer và dân tộc Mường loại đa hình hay gặp nhất cùng là C16223T lần lượt là (52/98) chiếm 53% và (39/100) chiếm 39%, (bảng 3.6), cần có thêm nghiên cứu trên nhiều dân tộc khác để xác định các đa hình đặc trưng cho từng dân tộc. Nghiên cứu của Nguyễn Thy Ngọc và cộng sự cho thấy đa hình T146C, T16297C, T199C trên dân tộc Kinh lần lượt chiếm tỷ lệ 5/51 (chiếm 9,8%), 6/51 (chiếm 11,8%), 10/51 (19,6%) và không có cá thể người Mảng nào mang các đa hình trên. Đa hình hay gặp nhất trên vùng HV1 ở dân tộc Kinh là T16189C (29/51) chiếm 56,5%, trên dân tộc Mảng là C16223T (23/37) chiếm 62,2%. Trên vùng HV2 ngoài hai loại đa hình A73G và A263G thì loại đa hình hay gặp nhất của cả hai dân tộc Kinh và Mảng là C150T có tỷ lệ lần lượt là 27,5% và 45,5% [123].

Nghiên cứu của Đỗ Mạnh Hưng và cộng sự năm 2016 trên 169 người khỏe mạnh thuộc bốn dân tộc dân tộc Kinh (40 cá thể), Mường (47 cá thể), Ê-đê (34 cá thể) và Jarai (48 cá thể) cho kết quả những vị trí đa hình hay gặp trên vùng HV2 của DNA ty thể là A73G, C150T, T152C, 249delA và A263G, đặc biệt có một đoạn poly C lớn từ vị trí 303 đến 315 với 1 nucleotide T được xen vào tại vị trí 310 [122]. Đây cũng là các loại đa hình hay gặp trên vùng HV2 mtDNA trong nghiên cứu của chúng tôi (bảng 3.12) Tác giả cũng đã đi

đến kết luận vùng gen ty thể HV2 là một trình tự có tính đa hình rất lớn và có ý nghĩa nhất định với việc nghiên cứu di truyền của các quần thể người. Cả 4 nhóm cá thể Kinh, Mường, Ê-đê và Jarai người Việt Nam đều có sự tương đồng di truyền với các quần thể người đang sinh sống trong khu vực Đông Nam Á và Đông Á [122].

Một số các nghiên cứu khác cũng đã chỉ ra các vị trí đa hình hay gặp trên vùng HV1 và HV2 tương đồng với nghiên cứu của chúng tôi như nghiên cứu của Fang và cộng sự năm 2014 cho thấy một số các đa hình trên vùng HV1, HV2 của mtDNA ở nhóm người bình thường (420 người Trung Quốc) là: C150T (chiếm 21%), T152C (chiếm 24,5%), 249De1A (chiếm 22,6%), G16129A (chiếm 20%), A16183C (chiếm 24,8%), T16189C (chiếm 37,6%), và C16223T (chiếm 59,5) được phát hiện rất phổ biến tại trong hệ gen ty thể người Hán miền Nam Trung Quốc [125].

Sự thay thế nucleotid là phổ biến nhất chiếm 91,6% và chủ yếu là các thay thế đồng hoán, trong khi đó thêm nucleotid là 2,1% và xóa nucleotid là 6,3% (bảng 3.1). Đa hình tại vị trí A263G gặp ở 100% mẫu nghiên cứu và đa hình A73G gặp ở 99,6% mẫu nghiên cứu. Kết quả này phù hợp với kết quả nghiên cứu của Yao và cộng sự năm 2002 trên 263 người thuộc 6 dân tộc người Trung Quốc cũng cho thấy đa hình tại hai vị trí này là 100% [32]. Trong nghiên cứu của Nguyễn Đăng Tôn và cộng sự cũng cho kết quả tương tự tỷ lệ đa hình A73G gặp ở 100% mẫu nghiên cứu (78/78) và tỷ lệ đa hình A263G gặp ở 96% (75/78) mẫu nghiên cứu.

Trong nghiên cứu này, chúng tôi cũng đã xác định được 172 vị trí đa hình trên vùng HV1 trong đó có 66 vị trí đa hình chỉ gặp ở các mẫu thuộc 1 trong 4 dân tộc. Trên vùng HV2 xác định được 89 vị trí đa hình, 42 vị trí đa hình chỉ gặp ở các mẫu thuộc 1 trong 4 dân tộc. Có thể thấy sự đa hình của

các mẫu nghiên cứu tập trung nhiều nhất trên vùng HV1 (hình 3.13). Kết quả nghiên cứu của chúng tôi cũng phù hợp với kết quả nghiên cứu của Bhinu Shova Tuladhar và cộng sự khi giải trình tự vùng HV1 và HV2 mtDNA của 103 mẫu người Malaysia cũng xác định được tổng số đa hình trên vùng HV1 là 91 và tổng số đa hình trên HV2 là 52, điều này cho thấy sự đa hình ở hai vùng HV1 và HV2 là rất cao. Tác giả cũng nhận thấy hầu hết là đa hình thay thế pyrimidin (C-T) trên vùng HV1 là 35,77%, đa hình thay thế purin (A-G) trên vùng HV2 là 64,51% [124].

Sự đa hình ở hai vùng HV1 và HV2 mtDNA hầu hết là các đa hình đơn nucleotid (SNP). Các đa hình có thể là các thay thế đồng hoán (transition) - thay thế nucleotid có cùng gốc purin (A-G) hoặc pyrimidin (C-T); hoặc là các thay thế dị hoán (transversion)-thay thế nucleotid có gốc purin thành pyrimidin hoặc ngược lại (A-T, C-G). Các đa hình thay thế đồng hoán thường gặp hơn sự thay thế dị hoán. Trên vùng HV1 có 136 loại SNP là thay thế đồng hoán (transition) và 43 SNP là thay thế dị hoán (transversion). Trên vùng HV2 có 72 loại SNP là thay thế đồng hoán (transition) và 11 SNP là thay thế dị hoán (transversion). Các đa hình khác ít gặp hơn, đa hình mất nucleotid (5 SNP trên vùng HV1, 13 SNP trên vùng HV2), đa hình thêm nucleotid (2 SNP trên vùng HV1 và 4 SNP trên vùng HV2), nucleotid thêm vào thường là nucleotid C trong khi nucleotid mất đi thường là nucleotid A (bảng 3.1). Kết quả này cũng tương tự như trong nghiên cứu của Nguyễn Đăng Tôn và cộng sự [90]. Kết quả nghiên cứu của chúng tôi cũng phù hợp với nghiên cứu của tác giả Tuladhar và cộng sự khi giải trình tự vùng HV1, HV2 DNA ty thể người Malaysia cũng cho thấy đa hình thay thế đồng hoán (71 SNP trên vùng HV1, 29 SNP trên vùng HV2) gặp nhiều hơn đa hình thay thế dị hoán (17 trên vùng HV1 và 9 trên vùng HV2). Về đa hình thêm nucleotid tác giả cũng chỉ gặp 2 loại SNP thêm nucleotid trên vùng HV1 và 7 loại SNP thêm nucleotid

trên vùng HV2, đa hình mất nucleotid gặp 1 loại trên vùng HV1 và 7 loại trên vùng HV2 [124].

Chúng tôi cũng nhận thấy có một số đa hình chỉ gặp ở dân tộc này mà không thấy gặp ở dân tộc khác (bảng 3.4, bảng 3.5), có 66 đa hình trên vùng HV1 chỉ gặp ở các mẫu thuộc 1 trong 4 dân tộc trong đó có 21 loại đa hình chỉ thấy ở dân tộc Kinh, 20 loại SNP chỉ có ở dân tộc Mường, 17 loại SNP chỉ có ở dân tộc Khmer và ít nhất là dân tộc Chăm có 8 loại SNP chỉ gặp ở dân tộc này. Trên vùng HV2 có 42 loại đa hình chỉ gặp ở các mẫu thuộc 1 trong 4 dân tộc (6 loại SNP ở dân tộc Chăm, 11 loại SNP ở dân tộc Mường, 12 loại SNP ở dân tộc Khmer, 13 loại SNP ở dân tộc Kinh). Nhìn chung, dân tộc Kinh và Mường có nhiều đa hình riêng trên cả hai vùng HV1 và HV2 trong khi dân tộc Chăm có ít đa hình đa hình riêng nhất. Điều này cho thấy giữa 4 dân tộc Kinh, Chăm, Mường, Khmer trong cộng đồng các dân tộc Việt Nam vẫn có những đặc điểm di truyền khác biệt đáng chú ý.

Gần đây, một nghiên cứu của Nguyễn Thy Ngọc và cộng sự (năm 2018) trên 81 cá thể người Việt Nam thuộc 2 dân tộc Kinh và dân tộc Mảng đã phát hiện 96 điểm đa hình trên vùng HV1 và HV2 ở nhóm người dân tộc Kinh (trong đó có 60 đa hình thuộc vùng HV1 và 36 đa hình thuộc vùng HV2) và 36 điểm đa hình trên vùng HV1 và HV2 ở nhóm người dân tộc Mảng (22 đa hình thuộc trên vùng HV1 và 14 đa hình trên vùng HV2). Các 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 Mảng và bốn đa hình: C151T, A16162G, A16269G và T16271C xuất hiện phổ biến ở người Mảng nhưng ít xuất hiện ở người Kinh. Điều đó cho thấy 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 [123]. Nghiên cứu của chúng tôi phát hiện được số lượng các đa hình trên vùng HV1 và HV2 nhiều



hơn có lẽ do số lượng mẫu nghiên cứu và số dân tộc trong nghiên cứu của chúng tôi cũng lớn hơn nhiều so với nghiên cứu trên. Một nghiên cứu khác khi giải trình tự vùng HV2 cho kết quả có 52 vị trí đa hình trên vùng HV2, các trình tự này xuất hiện nhiều điểm nóng đa hình như tại đa hình A73G, C150T, T152C, 249delA và A263G [122].

Phân tích trình tự nucleotid của vùng HV1 trên DNA ty thể của 100 cá thể người Thái Lan các tác giả đã xác định được 85 vị trí đa hình trên vùng HV1, chủ yếu là đa hình thay thế nucleotid. Các đa hình phổ biến ở là đa hình tại vị trí 16223 (gặp ở 48% mẫu nghiên cứu), đa hình tại vị trí 16304 (31%), đa hình tại vị trí 16332 (30%), đa hình tại vị trí 16129 (26%). Chuyển từ nucleotid T sang C (chiếm 43,7%) là sự thay thế thường xuyên nhất. Thêm nucleotid được tìm thấy ở hai vị trí 16188 và vị trí 16193. Đã xác định được 13% các mẫu nghiên cứu có đoạn polyC từ vị trí 16180 đến vị trí 16193 [128]. Kết quả này cũng phù hợp với nghiên cứu của chúng tôi khi cho thấy tỷ lệ đa hình tại vị trí 16223 là 41%, đa hình tại vị trí 16304 là 28,8% và đa hình tại vị trí 16129 là 33%, và hai vị trí đa hình thêm nucleotid C trên vùng HV1 cũng là vị trí 16188 và 16193 (bảng 3.1 và bảng 3.6). Điều này có thể là do Thái Lan và Việt Nam cũng là hai nước có quan hệ gần gũi về mặt địa lý và giao lưu văn hóa.

Khi phân tích đa hình trên hai vùng HV1 và HV2 mtDNA của 125 mẫu người dân tộc Bai ở Trung Quốc cũng cho thấy sự đa hình rất cao của hai vùng này đã xác định được 86 vị trí đa hình trên vùng HV1 và 40 vị trí đa hình trên vùng HV2 [129]. Nghiên cứu của chúng tôi xác định được nhiều đa hình trên cả hai vùng HV1 và HV2 so với nghiên cứu trên (172 vị trí đa hình trên vùng HV1 và 89 vị trí đa hình trên vùng HV2) có thể do cỡ mẫu và số dân tộc trong nghiên cứu của chúng tôi lớn hơn nhiều so với nghiên cứu trên (517 mẫu thuộc 4 dân tộc so với 125 mẫu).

***DNA ty thể với độ đa dạng di truyền và xác suất trùng khớp ngẫu nhiên giữa hai cá thể:***

Khi phân tích sự đa dạng di truyền của người Việt Nam dựa trên hai vùng HV1 và HV2 của DNA ty thể, chúng tôi thấy độ đa dạng di truyền là 99,83% và xác suất trùng khớp ngẫu nhiên giữa hai cá thể (cùng haplotype mtDNA) là 0,37%. Rashid và cộng sự năm 2010, khi nghiên cứu trên người Malaysia cho kết quả lần lượt là 99,47% và 0,93% [118]. Trong một nghiên cứu về dân số Nhật Bản sự đa dạng di truyền và xác suất trùng khớp ngẫu nhiên giữa hai cá thể được ước tính là 99,69% và 0,40% [130]. Nghiên cứu trên với quần thể người Deng ở Tây Tạng Trung Quốc, cho kết quả độ đa dạng di truyền và xác suất trùng khớp ngẫu nhiên lần lượt là 99,16% và 0,84% [131]. Vanecek và cộng sự khi nghiên cứu trên 93 người Séc cho thấy có 85 haplotypes, 78 haplotypes là duy nhất, đa dạng di truyền và xác suất khớp ngẫu nhiên lần lượt là 99,9% và 1,2% [132]. Năm 2017, Rakha và cộng sự khi phân tích vùng điều khiển D-loop của 319 người Hazara ở Pakistan cho kết quả có 189 haplotypes là duy nhất, đa dạng di truyền và xác suất khớp ngẫu nhiên lần lượt là 99,45% và 0,85% [133]. Như vậy, từ những kết quả ở trên cho thấy sự đa dạng di truyền vùng HV1, HV2 của DNA ty thể người Việt Nam có thể cao hơn các chủng tộc người khác ở châu Á. Mặc dù nghiên cứu mới được thực hiện trên bốn dân tộc trong số 54 dân tộc người Việt Nam nhưng đã cho thấy sự đa dạng di truyền của người Việt Nam, điều này làm nổi bật tính đa dạng di truyền của các dân tộc Việt Nam.

***Số lượng đa hình trên vùng HV1, HV2 của DNA ty thể:***

Các kết quả thu được về các vị trí đa hình, số lượng đa hình trên vùng HV1 và HV2 mtDNA đã phản ánh tốc độ đột biến rất cao của hai vùng siêu biến này. Đa số các mẫu nghiên cứu đều có nhiều đa hình so với trình tự

chuẩn, có 401/517 mẫu nghiên cứu có từ 10 loại đa hình trở lên chiếm 77,5%, mẫu có nhiều đa hình nhất là 21 loại SNP (1 mẫu thuộc dân tộc Chăm), mẫu có ít đa hình nhất là 6 loại SNP (gặp ở 8 mẫu nghiên cứu). Đây là một số lượng đa hình khá lớn tuy nhiên chúng phù hợp với các nghiên cứu trước đây cho rằng hai vùng siêu biến HV1 và HV2 có tần số đột biến cao nhất trong hệ gen ty thể.

Nghiên cứu năm 1999 của nhóm R. Ivanova trên 50 người Việt Nam đã phát hiện được 20 đa hình vị trí các enzym cắt trên DNA ty thể [85]. Năm 2005, Huỳnh Thị Thu Huệ và cộng sự đã giải trình tự hai vùng HV1 và HV2 trên 2 mẫu dân tộc Kinh, 2 mẫu dân tộc Tày và một mẫu dân tộc H'Mông, đã phát hiện 26 vị trí đa hình trên vùng HV1 và 5 vị trí đa hình trên vùng HV2 [134]. Năm 2016, một nghiên cứu khác trên vùng HV2 của 4 nhóm người dân tộc Kinh, Mường, Ê-đê, Jarai cũng cho thấy số lượng đa hình ở từng mẫu không giống nhau, chủ yếu nằm trong khoảng từ 5 đến 7 đa hình trên mỗi trình tự, trong đó mẫu có nhiều đa hình nhất là 11 và mẫu có đa hình ít nhất là 3 [122]. Nghiên cứu của chúng tôi cho kết quả số lượng đa hình trên hai vùng HV1 và HV2 lớn hơn có lẽ do số lượng mẫu nghiên cứu của chúng tôi lớn hơn nhiều so với các nghiên cứu trên.

#### **4.3. Đánh giá tính đa hình vùng HV1 của mtDNA ở bệnh nhân ung thư vú**

##### ***Đa hình gen ty thể và mối liên quan đến một số bệnh:***

Những năm gần đây, có nhiều nghiên cứu trên thế giới đã nghiên cứu phân tích trình tự của hệ gen ty thể đặc biệt là vùng điều khiển D-loop với hai vùng HV1 và HV2 nhằm nghiên cứu mối quan hệ di truyền tiến hóa của các dân tộc cũng như mối liên quan với các loại bệnh tật khác nhau. Saldana-Rivera năm 2018 khi nghiên cứu mối liên quan của đa hình T16189C trên vùng HV1 của mtDNA với hội chứng chuyển hóa ở người Mexico đã thấy tần

số của đa hình T16189C cao hơn ở nhóm người mắc hội chứng chuyển hóa với 21/65 (chiếm 32,3%) ở nhóm bệnh so với 3/35 (chiếm 8,5%) ở nhóm chứng (với OR 5.0909, 95% CI 1.3977–18.5424,  $p = 0.0136$ ), từ đó cho thấy đa hình này là yếu tố rủi ro có thể xảy ra đối với hội chứng chuyển hóa trong dân số Mexico [135]. Trong nghiên cứu này cũng cho thấy tỷ lệ đa hình T16189C là 25% thấp hơn so với nghiên cứu của chúng tôi khi cho tỷ lệ đa hình T16189C là 39%. Tỷ lệ đa hình T16189C trong nghiên cứu của chúng tôi cũng cao hơn so với kết quả nghiên cứu của Park và cộng sự khi nghiên cứu trên người châu Á cho thấy tỷ lệ đa hình T16189C là 30% [136]. Đa hình T16189C gây ra sự thay đổi liên kết với protein vùng DNA bị ảnh hưởng dẫn đến giảm sự sao chép mtDNA [136]. Trong một nghiên cứu khác tác giả thấy vai trò của đa hình T16189 liên quan đến hệ thống phòng thủ chống oxy hóa vì tăng stress oxy hóa ở bệnh nhân tiểu đường [137]. Ngoài ra đa hình T16189C cũng làm thay đổi số lượng bản sao mtDNA trong tế bào máu [138].

Một nghiên cứu khi phân tích tổng hợp bao gồm tổng số 11.794 trường hợp đái tháo đường type 2 và 14.465 người thuộc nhóm chứng, phân tích kết hợp cho thấy nguy cơ tương đối của bệnh đái tháo đường type 2 đối với biến thể mtDNA mang đoạn poly-C từ vị trí 16184 đến 16193 ở dân số châu Âu có liên quan với sự gia tăng nguy cơ mắc bệnh đái tháo đường type 2 [139], [140]. Đa hình T16189C trên vùng HV1 của DNA ty thể cũng được chứng minh có liên quan đến việc tăng nguy cơ mắc bệnh đái tháo đường type 2 ở người châu Á [136].

Mối liên quan giữa đa hình T16189C cũng được xác định có tần số cao hơn ở bệnh nhân nhồi máu sớm (trước 55 tuổi ở nam giới) so với bệnh nhân bị nhồi máu lần đầu ở tuổi trên 55 (24,1% so với 12,5%;  $p = 0,008$ ), có thể do đa hình mtDNA có ảnh hưởng lên quá trình phosphoryl oxy hóa và sản xuất các loại oxy phản ứng trong ty thể do đó góp phần gây xơ vữa động mạch

vành [141]. Tỷ lệ đa hình T16189C cao hơn đáng kể ở những bệnh nhân mắc bệnh mạch vành (21,6% so với 11,8% của nhóm chứng) cũng như ở những bệnh nhân mắc bệnh đái tháo đường type 2 (19,4% so với 11,8% của nhóm chứng). Sự kết hợp của bệnh mạch vành với đa hình T16189C vẫn có ý nghĩa cao sau khi điều chỉnh về tuổi, giới tính và chỉ số khối cơ thể (BMI) [142].

Các đa hình C16270T và C16320T trên vùng HV1 DNA ty thể liên quan có ý nghĩa (với  $p=0,02$  và  $p=0,03$ ) đến việc tăng nguy cơ mắc bệnh đái tháo đường type 2 ở bệnh nhân người Ma-rốc [143]. Haplotype HV1/HV2 DNA ty thể A16183C, T16189C, C16192T, C16270T và SNP T195C cao hơn đáng kể ở những bệnh nhân mắc khối u ác tính so với nhóm chứng. Các đa hình A302insCC và T310C trên vùng HV2 có liên quan đáng kể với kích thước khối u, trong khi đa hình T16519C trên vùng HV1 có liên quan với sự di căn của khối u. Các biến thể mtDNA có thể liên quan đến nguyên nhân u ác tính và sinh bệnh học u ác tính [144].

Đột biến DNA ty thể cũng được tìm thấy ở bệnh nhân u não, trên 49 mẫu mô khối u não được đánh giá có 25/49 trường hợp (chiếm 51%) có đột biến soma của vùng điều khiển D-loop [145]. Đột biến soma ở vị trí 310 trên vùng HV2 có thể được dùng làm dấu ấn sinh học có giá trị để chẩn đoán ung thư phổi Trung Quốc [146]. Đa hình trong vùng điều khiển D-loop của DNA ty thể cũng được nghiên cứu như một yếu tố nguy cơ và là dấu ấn sinh học để chẩn đoán và phát hiện sớm bệnh ung thư bạch cầu lympho ác tính [147]. Liên quan giữa DNA ty thể và ung thư cổ tử cung cũng được Ines Badano và cộng sự nghiên cứu, kết quả cho thấy phụ nữ mang DNA ty thể nguồn gốc châu Phi có khả năng bị mắc tổn thương ở cổ tử cung cao gần gấp ba lần so với những người mang DNA ty thể của người Mỹ bản địa hoặc người châu Âu [148].

Đa hình T152C của trên vùng HV2 của DNA ty thể cũng được chứng minh có liên quan đến bệnh bạch cầu cấp dòng tủy. Đa hình T152C xảy ra thường xuyên hơn ở những bệnh nhân mắc bệnh bạch cầu tủy cấp tính. Người mang đa hình T152C có nguy cơ mắc bệnh bạch cầu cấp dòng tủy cao gấp 1,752 lần với người không mang đa hình T152C. Đa hình T152C có thể tham gia vào cơ chế sinh bệnh học của bệnh bạch cầu cấp dòng tủy [149]. Silkjaer và cộng sự khi giải trình tự toàn bộ DNA ty thể ở 56 mẫu bệnh nhân mắc bệnh bạch cầu cấp dòng tủy và 14 mẫu đối chứng. Nghiên cứu này xác định rằng đa hình T16311C trên vùng HV1 DNA ty thể là đa hình thường xuyên nhất trong vùng HV1 và có xu hướng liên quan đến bất thường nhiễm sắc thể [150]. Kết quả nghiên cứu của Anudishi Tyagi và cộng sự cho thấy một số đa hình trong vùng điều khiển D-loop của DNA ty thể có thể có ảnh hưởng đến các gen được mã hóa của ty thể trong bệnh bạch cầu tủy cấp tính ở trẻ em [151].

Nghiên cứu của Ya-Fang Chen năm 2015 trên dân tộc Hán ở Trung Quốc chỉ ra rằng haplogroup B có thể có nguy cơ mắc Parkinson thấp hơn trong khi haplogroup D có thể dẫn đến nguy cơ mắc Parkinson cao hơn ở những người dưới 50 tuổi [152].

Một số nghiên cứu khác cũng đã chứng minh rằng đa hình DNA ty thể có liên quan đến quá trình phát sinh bệnh và tiên lượng bệnh. Vùng điều khiển D-loop của mtDNA rất quan trọng đối với việc điều hòa bộ gen của ty thể, các SNP trong khu vực này có thể ảnh hưởng đến sự sao chép mtDNA và dẫn đến thay đổi chuỗi vận chuyển điện tử, chịu trách nhiệm giải phóng sản phẩm ROS cao, gây ảnh hưởng đến bộ gen nhân, khởi đầu và thúc đẩy ung thư. Năm 2015, Hu Weng-xing và cộng sự khi nghiên cứu đa hình di truyền trong vùng điều khiển D-loop của DNA ty thể trên bệnh nhân bị ung thư phổi không tế bào nhỏ. Kết quả cho thấy việc phân tích đa hình di truyền trong vùng điều khiển D-loop của ty thể có thể được sử dụng để xác định các phân

nhóm bệnh nhân bị ung thư phổi không tế bào nhỏ có nguy cơ cao khởi phát sớm. Đó là các đa hình G200A, G207A trên HV2 và đa hình T16362C trên HV1 được xác định là các dấu hiệu dự đoán cho tuổi bắt đầu bị ung thư phổi không tế bào nhỏ [153]. Một nghiên cứu khác năm 2016 trên bệnh nhân ung thư tuyến giáp cũng đã xác định đa hình T16362C trên vùng HV1 mtDNA có mối liên quan với ung thư tuyến giáp [154].

Hàm lượng DNA ty thể trong các bệnh ung thư cũng được nhiều tác giả nghiên cứu. Số bản sao DNA ty thể thấp hơn ở những bệnh nhân bị ung thư vú tiến triển [155]. Hàm lượng mtDNA cao là một yếu tố tiên lượng xấu ở bệnh nhân ung thư đại tràng giai đoạn tiến triển [156]. Số lượng bản sao mtDNA thấp có liên quan và có thể dự đoán thêm khả năng sống sót ở ung thư tuyến tiền liệt khu trú [157]. Hàm lượng mtDNA trong huyết tương giảm có liên quan đến tiềm năng di căn của khối u và tiên lượng không thuận lợi ở bệnh nhân ung thư phổi [158].

#### ***Đa hình gen ty thể và bệnh ung thư vú:***

Nghiên cứu mối liên quan giữa DNA ty thể và các bệnh ung thư thì ung thư vú là bệnh được nhiều tác giả quan tâm nghiên cứu nhất. Nageswara và cộng sự khi phân tích giải trình tự toàn bộ vùng điều khiển D-loop của DNA ty thể ở 213 bệnh nhân ung thư vú và 207 người ở nhóm đối chứng thuộc miền nam Ấn Độ đã xác định được 7 đột biến mới và 170 đa hình trên vùng điều khiển D-loop của DNA ty thể. Các đa hình chủ yếu nằm trên vùng HV1 (chiếm 60%). Các tác giả cũng nhận thấy những haplotype có đoạn poly-C không bị gián đoạn có tần số cao hơn đáng kể ở bệnh nhân ung thư vú so với nhóm chứng ( $p = 0,0019$ ). Ngoài ra, tần số của 310insC cũng tăng ở bệnh nhân ung thư vú giai đoạn tiến triển. Các tác giả cũng đã xác định được tỷ lệ đa hình C16223T trên vùng HV1 ở bệnh nhân ung thư vú là 45,5% kết quả này cũng

phù hợp với nghiên cứu của chúng tôi khi cho tỷ lệ của đa hình này là 45,7%. Tuy nhiên, với hai đa hình T16362C và T16189C thì có sự khác nhau giữa nghiên cứu của tác giả và nghiên cứu của chúng tôi. Tỷ lệ đa hình T16189C, T16362C ở bệnh nhân ung thư vú trong nghiên cứu trên là 17,8% và 10,8%, còn trong nghiên cứu của chúng tôi hai SNP này chiếm tỷ lệ cao hơn lần lượt là 24,7% và 19,3%. Đa hình T16189C được chứng minh có mối liên quan đến tăng nguy cơ di căn ở bệnh nhân ung thư vú gặp ở 38/213 (17,8%) bệnh nhân ung thư vú so với 16/207 (7,7%) người nhóm chứng. Điều này khác biệt với nghiên cứu của chúng tôi khi cho thấy tỷ lệ đa hình T16189C ở bệnh nhân ung thư vú thấp hơn có ý nghĩa (với  $p = 0,001$ ) so với nhóm chứng (lần lượt là 24,7% và 40%). Ngoài ra, trong nghiên cứu của chúng tôi đã cho thấy đa hình T16362C là yếu tố tăng nguy cơ với bệnh ung thư vú người Việt Nam tăng gấp 1,87 lần ( $p = 0,02$  OR là 1,87 và 95% CI (1,1 - 3,18) và khi có đồng thời 2 loại SNP C16223T và T16362C nguy cơ ung thư vú tăng gấp gần 2,8 lần (với  $p = 0,001$ , OR là 2,759, 95% CI 1,45 - 5,15). Như vậy, các loại đa hình gen ty thể có thể ảnh hưởng khác nhau trên các dân tộc người khác nhau. Đa hình/đột biến trên vùng điều khiển D-loop của DNA ty thể có thể tạo thành các yếu tố nguy cơ hoặc cũng có thể là các yếu tố bảo vệ cho sự phát triển ung thư vú. Phân tích các thay đổi di truyền trên vùng điều khiển D-loop của mtDNA có thể giúp xác định bệnh nhân ung thư vú có nguy cơ tiến triển xấu, do đó giúp điều chỉnh các quyết định điều trị trong bệnh ung thư vú [104].

Một nghiên cứu khác khi nghiên cứu trên bệnh nhân ung thư vú người Italy đã tìm được 107 đa hình trong vùng điều khiển D-loop, trong đó SNP T16189C chiếm tỷ lệ cao hơn ở nhóm người khỏe mạnh (40% so với 13,9% của nhóm bệnh nhân ung thư vú,  $p = 0,03$ ) [5]. Kết quả này phù hợp với nghiên cứu của chúng tôi SNP T16189C cũng chiếm tỷ lệ cao hơn ở nhóm chúng (40% so với 24,7% ở nhóm bệnh nhân ung thư vú, với  $p = 0,001$ ). Tuy



nhiên trong nghiên cứu của chúng tôi đã tìm thấy 184 loại SNP trên vùng HV1 của bệnh nhân ung thư vú nhiều hơn so với nghiên cứu của tác giả có thể do số mẫu nghiên cứu của chúng tôi lớn hơn và cũng có thể do sự khác biệt giữa 2 dân tộc người Việt Nam ở châu Á và người Italy thuộc các chủng tộc người châu Âu.

Nghiên cứu trên 50 bệnh nhân ung thư vú người Krud ở Iran tác giả đã tìm thấy đột biến 16342delT trên vùng HV1 và C182T, 194insT, 285insA trên vùng HV2 ở trong các khối u, trong khi đột biến 302insC, C309T và C16069T được tìm thấy trong cả mô khối u và các mô bình thường xung quanh. Nghiên cứu đã kết luận rằng sự thay đổi di truyền mtDNA có thể liên quan đến sự khởi phát và tiến triển của khối u vú [159].

Khi phân tích trình tự vùng điều khiển D-loop của bệnh nhân ung thư vú người Ba Lan bằng phương pháp RFLP, các tác giả đã phát hiện tỷ lệ đa hình vùng HV1 và HV2 của DNA ty thể T239C, A263G và C16207T ở bệnh nhân ung thư vú cao hơn đáng kể và tỷ lệ đa hình A73G, C150T, A16183C, T16189C, C16223T, T16362C thấp hơn đáng kể so với nhóm chứng. Các đa hình trong vùng điều khiển D-loop được xác định có thể liên quan đến việc tăng nguy cơ phát triển ung thư vú [160]. Trong nghiên cứu của chúng tôi hai loại haplotype A16183C, T16189C và haplotype T16140C, T16189C ở bệnh nhân ung thư vú cũng thấp hơn (có ý nghĩa thống kê với  $p < 0,05$ ) so với nhóm người khỏe mạnh (12,9% so với 17,6%,  $p = 0,001$  và 5,4% so với 12,5%,  $p = 0,022$ ). Haplotype C16223C, T16362C ở nhóm bệnh nhân ung thư vú cao hơn so với nhóm người khỏe mạnh (15,6% so với 6,3%,  $OR = 2,759$  và 95%CI 1,45 - 5,25). Trong một nghiên cứu năm 2015 Jie Shen và cộng sự đã chứng minh tăng số lượng bản sao mtDNA có liên quan đến tăng nguy cơ ung thư vú gấp 1,32 lần, dị hợp chiều dài vùng HV1 và HV2 có liên quan đến sự gia tăng nguy cơ ung thư vú tăng gấp 2,01 và 1,63 lần [3].

Từ các kết quả trên có thể gợi ý rằng một số loại đa hình DNA ty thể có thể là yếu tố nguy cơ phát sinh bệnh ung thư vú nhưng cũng có một số loại đa hình là yếu tố bảo vệ khỏi sự phát triển của ung thư vú ở phụ nữ.

Việc nghiên cứu DNA ty thể còn được ứng dụng trong nhiều lĩnh vực khác trong đó phải kể đến ứng dụng trong pháp y, nhận dạng, mối quan hệ di truyền tiến hóa. Khi phân tích trình tự vùng điều khiển HV1 của mtDNA được tiến hành trên một quần thể gồm 400 cá thể người Iraq đã xác định 117 nhóm đơn bội khác nhau của mtDNA và các vị trí đa hình 16037, 16075, 16104 và 16201 trên vùng HV1 tìm thấy trong nghiên cứu được cho là nguồn phù hợp cho mục đích nhận dạng trong tương lai [161]. Nghiên cứu của Wibhu Kutanan và cộng sự năm 2018, đã đi đến kết luận các phân tích DNA ty thể là thông tin trong việc làm sáng tỏ quan điểm di truyền trong địa lý và dân số liên quan đến ngôn ngữ [162]. Một nghiên cứu khác khi phân tích 439 trình tự mtDNA của các cá nhân thuộc 40 nhóm dân tộc ở Tây Bắc Amazonia đã cho thấy mtDNA có tính đa dạng di truyền cao với sự chia sẻ rộng rãi các kiểu đơn bội mtDNA giữa các nhóm dân tộc [163]. Jing-Hua Meng và cộng sự đã chứng minh rằng người Maonans sống tại khu tự trị Huân Giang Trung Quốc có mối quan hệ di truyền chặt chẽ với cộng đồng dân tộc Gelao ở Lào và Trung Quốc [164]. Các dữ liệu trong vùng điều khiển của DNA ty thể là cơ sở để làm sáng tỏ cho các nghiên cứu về pháp y, nghiên cứu về di truyền tiến hóa của con người [165].

Như vậy, chúng tôi đã giải trình tự hoàn chỉnh hai vùng HV1 và HV2 của DNA ty thể ở 517 mẫu nghiên cứu thuộc bốn dân tộc Kinh, Mường, Chăm, Khmer (206 mẫu dân tộc Kinh; 98 mẫu dân tộc Khmer, 113 mẫu dân tộc Chăm, 100 mẫu nghiên cứu dân tộc Mường) và giải trình tự vùng HV1 của 186 mẫu bệnh nhân ung thư vú người Việt Nam. Nghiên cứu đã xác định được 172 vị trí đa hình trên HV1 và 89 vị trí đa hình trên vùng HV2, có thể

thấy sự đa hình của các mẫu tập trung nhiều nhất trên vùng HV1. Sự đa hình ở hai vùng HV1 và HV2 hầu hết là các đa hình đơn nucleotid, đã tính được tỷ lệ một số đa hình hay gặp trên hai vùng HV1 và HV2 của DNA ty thể. Xác định được 438 haplotypes trong đó 393 haplotypes (cho 393 mẫu nghiên cứu) là duy nhất, 45 haplotypes (cho 124 mẫu còn lại) gặp ở nhiều hơn một cá thể. Xác định được độ đa dạng di truyền và xác suất khớp ngẫu nhiên của DNA ty thể, phân loại DNA ty thể của 517 mẫu nghiên cứu thành 50 haplogroups dựa theo các vị trí đa hình (SNP) đặc trưng trên hai vùng HV1 và HV2 của DNA ty thể. Đánh giá được một số đa hình trên vùng HV1 ở bệnh nhân ung thư vú người Việt Nam. Các kết quả này rất có ý nghĩa trong việc nghiên cứu di truyền quần thể và tiến hóa của các dân tộc Việt Nam. Kết quả cho thấy bộ gen ty thể của Việt Nam rất đa dạng nhưng vẫn có những đặc điểm riêng biệt khác với các quần thể khác. Dữ liệu này sẽ giúp hiểu rõ hơn về sự đa dạng trong bộ gen của ty thể người Việt Nam cũng như đóng góp thêm vào dữ liệu của các biến thể ty thể toàn cầu.

Trên cơ sở kết quả nghiên cứu này, chúng tôi sẽ tiếp tục nghiên cứu, đánh giá sự đa dạng di truyền DNA ty thể của các dân tộc khác ở Việt Nam nhằm khảo sát đặc điểm về gen học của các dân tộc người Việt Nam cũng như xác định được cơ sở dữ liệu của DNA ty thể để làm sáng tỏ cho các nghiên cứu về pháp y, nghiên cứu về di truyền tiến hóa của con người, nghiên cứu về mối liên quan giữa DNA ty thể với một số loại bệnh lý đặc biệt là các bệnh về ung thư, các bệnh về rối loạn chuyển hóa.

## KẾT LUẬN

### **1. Xác định được tỷ lệ một số SNP vùng HV1, HV2 của DNA ty thể ở 4 dân tộc Kinh, Chăm, Mường, Khmer người Việt Nam**

- Xác định được 172 vị trí đa hình trên vùng HV1 và 89 vị trí đa hình trên vùng HV2. 286 loại SNP đã được phát hiện trong đó có 3 SNP mới trên vùng HV1 chưa được công bố trên Mitomap: 16038DelA, G16084C và A16515C.

- Tỷ lệ một số đa hình hay gặp trên vùng HV1 là: G16129A, A16183C: đều là 33%, T16189C: 39% và C16223T: 41%.

- Tỷ lệ một số đa hình hay gặp trên vùng HV2 là: 309insC: 56,4%, 315insC: 96,3%, A73G: 99,6%, đặc biệt là đa hình A263G gặp ở 100% các mẫu nghiên cứu.

### **2. Phân nhóm SNP đặc trưng vùng HV1, HV2 của DNA ty thể trên 4 dân tộc Kinh, Mường, Chăm, Khmer người Việt Nam**

- Xác định được 438 haplotypes mtDNA, trong đó có 393 haplotypes là duy nhất và 45 haplotypes xuất hiện ở nhiều hơn một cá thể. Sự đa dạng di truyền là 99,83% và xác suất trùng lặp ngẫu nhiên của hai cá thể cùng một haplotype HV1/HV2 mtDNA là 0,37%.

- Phân loại DNA ty thể người Việt Nam dựa trên các SNP đặc trưng vùng HV1, HV2 thành 50 nhóm đơn bội trong đó 4 nhóm đơn bội chiếm tỷ lệ cao nhất là F1a, B5a, M, M7b1 lần lượt là 15,67%, 10,83%, 8,9% và 7,74%. Các nhóm đơn bội có tần số xuất hiện thấp nhất (1/517) là nhóm B6, D4e, D5a, E, F1c, F2a, F3a, G2a, N, M9b, N21 và nhóm U5a.

- Nhóm F1a là nhóm đơn bội phổ biến nhất ở dân tộc Kinh (chiếm 21%) và dân tộc Mường (chiếm 16%). Hai nhóm đơn bội phổ biến nhất ở dân tộc Chăm và Khmer lần lượt là nhóm B5a (15,9%) và nhóm M (18,4%).

### **3. Đánh giá một số SNP vùng HV1 của DNA ty thể trên bệnh nhân ung thư vú.**

- Tỷ lệ SNP T16362C ở nhóm bệnh nhân ung thư vú cao hơn nhóm chứng và có sự khác biệt có ý nghĩa thống kê với  $p < 0,05$  ( $p = 0,02$ ),  $OR = 1,87$ , 95% CI (1,1-3,18).

- Tỷ lệ haplotype T16223C, T16362C ở nhóm bệnh nhân ung thư vú cao hơn nhóm chứng và có sự khác biệt có ý nghĩa thống kê với  $p < 0,05$  ( $p = 0,001$ ),  $OR = 2,759$ , 95% CI (1,45-5,25).

- Tỷ lệ SNP A16183C, T16189C và hai loại haplotype A16183C, T16189C, haplotype T16140C, T16189C ở nhóm bệnh nhân ung thư vú thấp hơn nhóm chứng, sự khác biệt có ý nghĩa thống kê với  $p < 0,05$  lần lượt là  $p = 0,0001$ ,  $p = 0,001$ ,  $p = 0,0001$  và  $p = 0,011$ .

## **DANH MỤC CÁC CÔNG TRÌNH ĐÃ CÔNG BỐ LIÊN QUAN ĐẾN LUẬN ÁN**

1. Thi Thuy Hang Tran, Duc Hinh Nguyen, Van Khanh Tran, Quy Linh Nguyen, Hong Anh Trinh, Long Hoang Luong, Van Anh Tran, Le Anh Tuan Pham, Thu Thuy Nguyen, Van Bang Nguyen, Thinh Huy Tran, and Thanh Van Ta (2019). Variation of Mitochondrial DNA HV1 and HV2 of the Vietnamese Population, *Advances in Experimental Medicine and Biology*, 3038542
2. Trần Thị Thúy Hằng, Trần Huy Thịnh, Trần Vân Khánh, (2017). Đa hình thái đơn nucleotid vùng gen ty thể HV1 và HV2 trên người dân tộc Kinh và dân tộc Mường Việt Nam, *Tạp chí Nghiên cứu Y học*, số 106 (1): 33 - 40.
3. Trần Thị Thúy Hằng, Trần Huy Thịnh, Trần Vân Khánh (2017). Phân tích trình tự vùng gen ty thể HV1 và HV2 trên dân tộc Chăm Việt Nam, *Tạp chí Nghiên cứu Y học*, số 108 (3): 8 - 15
4. Nguyễn Thu Thúy, Trần Thúy Hằng, Tạ Thành Văn, Nguyễn Đức Hinh, Trần Vân Khánh (2014). Đa dạng di truyền đoạn siêu biến 1 và đoạn siêu biến 2 trên DNA ty thể của dân tộc Mường ở Việt Nam, *Tạp chí Nghiên cứu Y học, Phụ trương* 91 (5): 24 - 29
5. Trần Thị Thúy Hằng, Trần Huy Thịnh, Bùi Thị Minh Phượng, Trần Vân Khánh (2015). Đa hình gen ty thể HV1 và HV2 trên một nhóm người dân tộc Kinh ở Việt Nam, *Tạp chí Nghiên cứu Y học*, số 96 (4): 23 - 30
6. Trần Thị Thúy Hằng, Trần Huy Thịnh, Trần Vân Khánh (2017). Đa hình đơn nucleotid vùng gen ty thể HV1 và HV2 trên dân tộc Khmer Việt Nam, *Tạp chí Y học Việt Nam*, (458)9: 203 -209.

## TÀI LIỆU THAM KHẢO

1. Stoneking, M., (2000). Hypervariable Sites in the mtDNA Control Region Are Mutational Hotspots, *Am. J. Hum. Genet*, **67**: p. 1029-1032.
2. Dimauro S, D.G., (2005). Mitochondrial DNA and disease. *Annals of Medicine*, **37**(3): p. 222-23.
3. Jie Shen, et al (2015). Peripheral blood mitochondrial DNA copy number, length heteroplasmy and breast cancer risk: a replication study. *Carcinogenesis*, **36**(11): p. 1307-1313.
4. Thyagarajan, B., et al., (2013). Mitochondrial DNA copy number is associated with breast cancer risk. *PloS one*, **8**(6): p. e65968-e65968.
5. Tommasi S., P.F., Stefania Weigl, et al, (2014). Mitochondrial DNA variants and risk of familial breast cancer: An exploratory study,. *International Journal Of Oncology*, **44**: p. 1691-1698.
6. Levin BC, C.H., Reeder DJ, (1999). A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis, and mutation detection. *Genomics*, **55**(2): p. 135-46.
7. Xiao-na Li, J.-l.Z., Jun Yao, Yue Dong, Zhang-sen Shi, (2017). Mitochondrial DNA control region sequences may differentiate Yanbian Koreans in China from other Asian populations,. *Annals of Human Biology*, **44**(5): p. 464-466.
8. Hiễn, N.N., (2005). *Sinh học đại cương*. Chương 2. Nhà xuất bản đại học quốc gia Hà Nội
9. Anderson S, B.A., Barrell BG, et al (1981) *Sequence and organization of the human mitochondrial genome*. *Nature*, 290(5806), 457-465. *Nature*, (1981). **290**(5806): p. 457-465.
10. Andrews Richard M, I.K., Patrick F. Chinnery, (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA,. *Nature Genetics*, **23**: p. 147.

11. Taanman, J.-W., (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta*. **1410**: p. 103-123.
12. Michael D. Sorenson, R.C.F., (1996). Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. *Proc Natl Acad Sci U S A*, **93**(26): p. 15239-15243
13. Ingman M, G.U., (2003). Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.*, **13**(7): p. 1600-1606.
14. Liou CW, L.T., Huang FM, Chen TL, Lee CF, et al. (2004). Association of the mitochondrial DNA 16189 T to C variant with lacunar cerebral infarction: evidence from a hospital-based case-control study. *Ann N Y Acad Sci*. **1011**: p. 317-324.
15. Malik S, S.H., Pramoongjago 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): p. 402-411.
16. R. G. Boles, C.L., M. Ito, (2003). Severe reversible cardiomyopathy in four unrelated infants associated with mitochondrial DNA D-Loop heteroplasmy. *Pediatr. Cardiol*, **24**: p. 484-487.
17. Salas A, L.V., Calafell F, Bertranpetit J, Carracedo A, (2000). mtDNA hypervariable region II (HVII) sequences in human evolution studies. *Eur J Hum Genet*, **8**(12): p. 964-974.
18. Imad H, A.F., Cheah Y, et al, (2013). Discovery of Three Newly Described Single Nucleotide Polymorphisms in Mitochondrial DNA Hypervariable Region I (HVI) and Estimation of Variants and Haplotypes Encompassing Nucleotide Positions 16024-16365. *J Forensic Res*. **5**(1).



19. Giles RE, B.H., Cann HM, Wallace DC, (1980). Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci.USA*, 77(11), 6715-6719., **77**(11): p. 6715-6719.
20. Schwartz M, V.J., (2003). New patterns of inheritance in mitochondrial disease. *Biochem. Biophys. Res. Commun.* **310**: p. 247-251.
21. Manfredi G, T.D., Papadopoulou LC, Pallotti F, Schon EA., (1997). The fate of human sperm-derived mtDNA in somatic cells. *Am J Hum Genet.* **61**(4): p. 953-960.
22. Brigitte Pakendorf, M.S., (2005). Mitochondrial DNA and human evolution. *Annu. Rev. Genomics Hum. Genet*, **6**: p. 165-183.
23. Ingman, M., et al., (2000). Mitochondrial genome variation and the origin of modern humans. *Nature*, **408**: p. 708.
24. Brown WM, G.M.J., Wilson AC, (1979). Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A*, **76**(4): p. 1967-1971.
25. Aquadro CF, G.B., (1983). Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics*, **103**(2): p. 287-312.
26. Bogenhagen, D.F., (1999). Repair of mtDNA in vertebrates. *Am J Hum Genet*, **64**(5): p. 1276-1281.
27. Wallace DC, B.M., Lott MT, (1999). Mitochondrial DNA variation in human evolution and disease. *Gene*, **238**(1).
28. Eric Dufour, M.T., Fredrik Hansson Sterky, (2008). Age-associated mosaic respiratory chain deficiency causes trans-neuronal degeneration. *Human Molecular Genetics*, **17**(10): p. 1418-1426.
29. Cann, R.L., M. Stoneking, and A.C. Wilson, (1987). Mitochondrial DNA and human evolution. *Nature*, **325**: p. 31.
30. Wallace, D.C., (2013). A mitochondrial bioenergetic etiology of disease. *The Journal of Clinical Investigation*, **123**(4): p. 1405-1412.
31. <http://www.genebase.com/learning/article/65>.

32. Yao, Y.-G., et al., (2002). Phylogeographic Differentiation of Mitochondrial DNA in Han Chinese. *American Journal of Human Genetics*, **70**(3): p. 635-651.
33. The International HapMap, C., (2003). The International HapMap Project. *Nature*, **426**: p. 789.
34. <http://knowgenetics.org/snps/>.
35. Shamnamole K, S.J., Vinod Scaria, (2013). MitoLSDB: A Comprehensive Resource to Study Genotype to Phenotype Correlations in Human Mitochondrial DNA Variations,. *Plos One*, **8**(4): p. e60066.
36. DiMauro S, S.E., (2003). Mitochondrial respiratory-chain diseases. *New Engl J Med*, **348**(26): p. 2656-2668.
37. Li YC, K.A., Fahima T et al (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol*, **11**(12): p. 2453-2465.
38. Bùi Diệu, Trần Văn Thuán và cộng sự, (2012). Gánh nặng bệnh ung thư và chiến lược phòng chống ung thư quốc gia đến năm 2020. *Tạp chí Ung thư học - Hội thảo quốc gia phòng chống ung thư lần thứ 6*, p. 13-19.
39. Society, A.C., (2013). *Breast Cancer Facts & Figures 2013-2014*. Atlanta: American Cancer Society.
40. Chinnery PF, J.M., Wardell TM, Singh-Kler R, Hayes C at el, (2000) The epidemiology of pathogenic mitochondrial DNA mutations. *Annlas of Neruology*, **48**(2): p. 188-193.
41. Solano, A., et al., (2001). Genetic diseases of the mitochondrial DNA in humans. *Salud Publica Mex*, **43**(2): p. 151-61.
42. Seok-Jo Kim, P.C., Renea P. Jablonski, David B. Williams, and David W. Kamp, (2015). The role of mitochondrial DNA in mediating alveolar epithelial cell apoptosis and pulmonary fibrosis. *Int J Mol Sci*, **16**(9): p. 21486-21519.

43. Wallace, D.C., (2005). A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine. *Annu Rev Genet*, **39**: p. 359-407.
44. Lowell BB, S.G., (2005). Mitochondrial dysfunction and type 2 diabetes. *Science*, **307**: p. 384-387.
45. Chagnon P, et al. (1999). Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population. *Am J Med Genet*, **85**: p. 20-30.
46. Walt JM, et al (2003). Mitochondrial polymorphisms significantly reduce the risk of Parkinson disease. *Am J Hum Genet*. **72**: p. 804-811.
47. Kelly N Owens, a.e., (2002). Genomic sequencing in the service of human rights. *International Journal of Epidemiology*, **31**: p. 53-58.
48. Mitomap.org. *MITOMAP*. (2012).
49. Wang, X., (2001). The expanding role of mitochondria in apoptosis. *Genes Dev*, **15**(22): p. 2922-2933.
50. Cenk Aral, A.O., (2007). Mitochondrial DNA and cancer. *Marmara Medical Journal*. **20**(2): p. 127-136.
51. Makiko S, F., Henning Usadel, Otavia L, Caballero, (2000). Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. *Science*, **287**(5460): p. 2017-2019.
52. Abbott J.A., F.C.S., Robey-Bond S.M, (2014). Transfer RNA and human disease. *Front Genet*, **5**: p. 158.
53. Wallace, D.C., (2011). Bioenergetic Origins of Complexity and Disease. *Cold Spring Harb Symp Quant Biol*, **76**: p. 1-16.
54. Rolfe D.F., B.G.C., (1997). Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol Rev*, **77**(3): p. 731-758.
55. Warburg O, W.F., Negelein A, (1926). The metabolism of tumors in the body. *J Gen Physiol*, 1926. **8**(6): p. 519-530.

56. Dumas J.F., R.D., Servais S, (2012). *Mitochondria and cancer*. Cellular Bioenergetics in Health and Diseases: New Perspectives in Mitochondrial Biology, p. 115-147.
57. Moreno-Sanchez R., R.-E.S., Marin-Hernandez A., (2007). Energy metabolism in tumor cells. *FEBS J*, **274**(6): p. 1393-1418.
58. Chiche J., B.-H.M.C., Pouyssegur J, (2010). Tumour hypoxia induces a metabolic shift causing acidosis: a common feature in cancer. *J Cell Mol Med*, **14**(4): p. 771-794.
59. Kulawiec M., S.A., Desouki M.M., Still I., Matsui S., Bakin A., Singh K.K, (2008). Tumorigenic transformation of human breast epithelial cells induced by mitochondrial DNA depletion. *Cancer Biol Ther*, **7**(11): p. 1732-1743.
60. Isidoro A., M.M., Fernandez P.L et al (2004). Alteration of the bioenergetic phenotype of mitochondria is a hallmark of breast, gastric, lung and oesophageal cancer. *Biochem J*, **378**(1): p. 17-20.
61. Brandon M., B.P., Wallace D.C, (2006). Mitochondrial mutations in cancer. *Oncogene*, **25**(34): p. 4647-4662.
62. Tan D.J., B.R.K., Wong L.J, (2002). Comprehensive scanning of somatic mitochondrial DNA mutations in breast cancer. *Cancer Res*, **62**(4): p. 927-976.
63. Liu V.W., S.H.H., Cheung A.N., Chiu P.M., Leung T.W., Nagley P., Wong L.C., Ngan H.Y, (2001). High incidence of somatic mitochondrial DNA mutations in human ovarian carcinomas. *Cancer Res*, **61**(16): p. 5998-6001.
64. Wu C.W., Y.P.H.a.e., (2005). Mitochondrial DNA mutations and mitochondrial DNA depletion in gastric cancer. *Genes Chromosomes Cancer*, **44**(1): p. 19-28.
65. Pen-Hui Yin a, C.-C.W.b., Jin-Ching Lin (2010). Somatic mutations of mitochondrial genome in hepatocellular carcinoma. *Mitochondrion*, **10**: p. 174-182.

66. Jones J.B., S.J.J., Hempen P.M., Parmigiani G., Hruban R.H., Kern S.E, (2001). Detection of mitochondrial DNA mutations in pancreatic cancer offers a “mass”-ive advantage over detection of nuclear DNA mutations. *Cancer Res*, **61**(4): p. 1299-1304.
67. Petros J.A., B.A.K., Ruiz-Pesini E, et al (2005). mtDNA mutations increase tumorigenicity in prostate cancer. *Proc Natl Acad Sci U S A*, **102**(3): p. 719-724.
68. Suzuki M., T.S., Miyajima K., et al, (2003). Alterations in the mitochondrial displacement loop in lung cancers. *Clin Cancer Res*, **9**(15): p. 5636-5641.
69. Tseng LM, Y.P., Chi CW, Hsu CY, Wu CW, Lee LM, Wei YH, Lee HC, (2006). Mitochondrial DNA mutations and mitochondrial DNA depletion in breast cancer. *Genes Chromosomes Cancer*, **45**(7): p. 629-638.
70. Gazi Nurun Nahar Sultana, A.R., Abu Din Ahmed Shahinuzzaman, Rowshan Ara Begum, and Chowdhury Faiz Hossain, (2012). Mitochondrial DNA mutations – candidate biomarkers for breast cancer diagnosis in Bangladesh. *Chin J Cancer*, **31**(9): p. 449-454.
71. Ye, C., et al., (2010). Mutations in the mitochondrial DNA D-loop region and breast cancer risk. *Breast cancer research and treatment*, **119**(2): p. 431-436.
72. Fang, H., et al., (2010). Cancer type-specific modulation of mitochondrial haplogroups in breast, colorectal and thyroid cancer. *BMC Cancer*, **10**(1): p. 421.
73. Zhu, W., et al., (2005). Mitochondrial DNA mutations in breast cancer tissue and in matched nipple aspirate fluid. *Carcinogenesis*. **26**(1): p. 145-152.
74. CAI, F.F., et al., (2011). Mutations of Mitochondrial DNA as Potential Biomarkers in Breast Cancer. *Anticancer Research*, **31**(12): p. 4267-4271.

75. Kuo, S.-J., et al., (2010). Number of somatic mutations in the mitochondrial D-loop region indicates poor prognosis in breast cancer, independent of *TP53* mutation. *Cancer Genetics and Cytogenetics*, **201**(2): p. 94-101.
76. Okochi O, H.K., Uemura T, Inoue S, Takeda S, Kaneko T, Nakao A, (2002). Detection of mitochondrial DNA alterations in the serum of hepatocellular carcinoma patients. *Clin Cancer Res*, **8**(9): p. 2875-2878.
77. Shi Y, Y.F., Lu W, et al (2002). Histologic classification in 10,288 cases of ovarian malignant tumors in China. *Chinese J Obstet Gynecol*, **37**: p. 97-100.
78. Mambo E, C.A., Xing M, Tallini G, Haugen BR, Yeung SC, Sukumar S, Sidransky D. (2005). Tumor-specific changes in mtDNA content in human cancer. *International Journal Cancer*, **116**(6): p. 920-924.
79. Matthew R. Bonner, M.S., Chin-San Liu, Margaret DiVita, Xingzhou He, and Qing Lan (2009). Mitochondrial DNA Content and Lung Cancer Risk. *Lung cancer*, **63**(3): p. 331-334.
80. Pei-Ching Lin, J.-K.L., Shung-Haur Yang, Huann-Sheng Wang, Anna Fen-Yau Li, Shih-Ching Chang, (2008). Expression of  $\beta$ -F1-ATPase and mitochondrial transcription factor A and the change in mitochondrial DNA content in colorectal cancer: clinical data analysis and evidence from an in vitro study. *International Journal of Colorectal Disease*, **23**(12): p. 1223-1232.
81. Huang XW, Z.Q., Chen DZ, Zhang LS, (2005). Mutations in the D-loop region of mitochondrial DNA and the ROS level in the tissue of hepatocellular carcinoma. *Yi Chuan*, **27**(1), 14-20, **27**(1): p. 14-20.
82. Y Wang, V.W.S.L., W C Xue, A N Y Cheung and H Y S Ngan (2006). Association of decreased mitochondrial DNA content with ovarian cancer progression. *British Journal of Cancer*, **95**: p. 1087-1091.

83. Ahmet Ilter Guney, D.S.E., Hasan Huseyin Tavukcu, Gulsah Koc, Deniz Kirac, Korkut Ulucan, Dilara Javadova, and Levent Turkeri, (2012). Detection of mitochondrial mutations in nonmuscle invasive bladder cancer. *Genetic Testing and Molecular Biomarkers 2012*. **16**: p. 672-678.
84. Ballinger SW, S.T., Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC, (1992). Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics*, **130**(1): p. 139-152.
85. Ivanova R., A.V.T., et al (1999). Mitochondrial DNA polymorphism in the Vietnamese population. *European Journal of Immunogenetics*, (26): p. 417-422.
86. Oota H, K.T., Jin F, Yuasa I, Wang L, Stoneking M, (2002). Extreme mtDNA homogeneity in continental Asian populations. *Am J Phys Anthropol*, **118**(2): p. 146-153.
87. Lê Quang Huân, T.T.M.L., Vũ Thị Thu (2003). Nghiên cứu giám định phả hệ bằng kỹ thuật DNA. Những vấn đề nghiên cứu cơ bản trong khoa học sự sống. *Báo cáo khoa học Hội nghị toàn quốc lần thứ 2, Nghiên cứu cơ bản trong sinh học, nông nghiệp, y học*, **Huế 25-26/7/2003**(917-919).
88. Đái Duy Ban, Hoàng Minh Châu (2003). Bước đầu nghiên cứu ung thư vú bệnh nhân Việt Nam bằng phương pháp sinh học phân tử sử dụng chỉ thị di truyền hệ gen ty thể đoạn D-loop. Những vấn đề nghiên cứu cơ bản trong khoa học sự sống. *Báo cáo khoa học Hội nghị toàn quốc lần thứ 2, Nghiên cứu cơ bản trong sinh học, nông nghiệp, y học*, **Huế 25-26/7/2003**: p. 825-829.
89. Phan Văn Chi và cộng sự (2006), Nghiên cứu giải mã genome ty thể các tộc người Việt Nam và định hướng ứng dụng. *Báo cáo đề tài cấp Nhà nước*, 2006. **mã số KC-04-25**.
90. Nguyễn Đăng Tôn, Vũ Hải Chi và cộng sự (2008). Đa hình đơn bội DNA ty thể của các cá thể người Việt Nam. *Tạp chí Công nghệ sinh học*, **6**(4): p. 579-590.

91. Nguyễn Hữu Nghĩa, Tống Quang Vinh và cộng sự (2008). Đột biến hệ gen ty thể vùng D-loop ở người lao động thường xuyên tiếp xúc với bức xạ ion hóa. *Tạp chí Y học Việt Nam*, **351**: p. 29-34.
92. Mẫn, C.V., (2009). Tìm hiểu bệnh ty thể ở người bằng phương pháp sinh học phân tử. *Báo cáo đề tài DH Quốc gia Hà Nội*, mã số **QT-08-31**.
93. Phạm Hùng Vân, Võ Quang Hồng Điềm và cộng sự (2010). Các trường hợp đầu tiên về bệnh lý thần kinh nhãn cầu LEBER xác định bằng kỹ thuật giải trình tự DNA ty thể tách chiết từ bạch cầu của bệnh nhân để phát hiện các đột biến gây bệnh. *Kỷ yếu Hội nghị Sinh học phân tử và hóa sinh y học*, p. 285-289.
94. Khánh, B.T., (2015). *Luận văn Thạc sỹ, Trường Đại học khoa học tự nhiên.*, in *Trường Đại học khoa học tự nhiên*. 2015.
95. H.H.N và cs. (2014). Phát hiện đột biến DNA ty thể trong bệnh lý thần kinh thị giác di truyền Leber bằng kỹ thuật giải trình tự gen. *Y Học TP. Hồ Chí Minh*, **18**(1): p. 81-85.
96. Trương Thị Huệ, Lê Ngọc Yên, Phạm Thị Vân Anh, Ngô Diễm Ngọc, Phan Tuấn Nghĩa và cộng sự (2012). Phát hiện mất đoạn 9 bp trên gen ty thể ở một số bệnh nhân nghi mắc bệnh não cơ. *Tạp chí sinh học*, **34**(2): p. 246-252.
97. Tú, N.T.N., (2012). *Phân tích đột biến gen tRNA và ND3 của DNA ty thể ở bệnh nhân ung thư đại trực tràng*, Luận văn Thạc sỹ, in Đại học khoa học tự nhiên. Đại học khoa học tự nhiên.
98. Văn, T.T., (2010). *PCR và một số kỹ thuật Y sinh học phân tử*. Nhà xuất bản Y học.
99. McPherson MJ, H.B., Taylor GR, (1995). Optimizing PCR”, PCR 2- A practical approach. *Oxford University Press*. 1-22.
100. Thanh, K.t.H.u., (2006). *Một số phương pháp cơ bản sử dụng trong kỹ thuật gen*. Nhà xuất bản Khoa học và kỹ thuật.
101. Quyền Đình Thi và cộng sự (2008). *Những kỹ thuật PCR và ứng dụng trong phân tích DNA*. 2008, Nhà xuất bản Khoa học tự nhiên và công nghệ.



102. Rasmussen, H.B., (2012). Restriction Fragment Length Polymorphism Analysis of PCR-RFLP and Gel Electrophoresis-Valuable Tool for Genotyping and Genetic Fingerprinting. *Gel Electrophoresis – Principles and Basics*, 2012: p. 315-334.
103. Koolman J (2005), D.s., Color Atlas of Biochemistry, 2nd edition, pp.260-2.
104. Nageswara Rao Tipiriseti, et al (2014). Mitochondrial Control Region Alterations and Breast Cancer Risk: A Study in South Indian Population. *Plos one*, **9**(1): p. e85363.
105. Fumio Tajima, (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism,. *Genetics*, **123**,: p. 585-595.
106. Hansi Weissensteiner, D.P., Anita Kloss-Brandstatter, Lukas Forer., H.-J.B. Gunther Specht, Florian Kronenberg, Antonio Salas and, and S. Schonherr, (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, **44**: p. 58-63.
107. Hwan Young Lee, I.S., Eunho Ha, Sung-Bae Cho, Woo Ick Yang and a.K.-J. Shin, (2008). mtDNAManager: a Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. *BMC Bioinformatics*. **9**: p. 483.
108. LongFan, Y.-G., (2011). MitoTool: A web server for the analysis and retrieval of human mitochondrial DNA sequence variations. *Mitochondrion*, **11**(2): p. 351-356.
109. kê, T.c.T., (2015). *Điều tra dân số và nhà ở giữa kỳ thời điểm 1/4/2014: Các kết quả chủ yếu*. (Hà Nội, tháng 9 năm 2015).
110. Kết quả toàn bộ Tổng điều tra Dân số và Nhà ở Việt Nam năm 2009, p.I.B.t.h., Tổng cục Thống kê Việt Nam, <http://www.gso.gov.vn/default.aspx?tabid=512&ItemID=10798>.
111. Duong, N.T., et al., (2018). Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia. *Scientific Reports*, **8**(1): p. 11651.

112. Jin, H.-J., C. Tyler-Smith, and W. Kim, (2009). The Peopling of Korea Revealed by Analyses of Mitochondrial DNA and Y-Chromosomal Markers. *PLoS ONE*, **4**(1): p. e4210.
113. Min-Sheng Peng, H.H.Q., Khoa Pham Dang et al, (2010). Tracing the Austronesian Footprint in Mainland Southeast Asia: A Perspective from Mitochondrial DNA. *Mol. Biol. Evol*, **27**(10): p. 2417-2430.
114. S. Pischedda, et al (2017). Phylogeographic and genome-wide investigations of Vietnam ethnic groups reveal signatures of complex historical demographic movements. *Scientific Reports*, **7**(12630).
115. Irwin JA, et al (2008). Mitochondrial control region sequences from a Vietnamese population sample. *Int J Legal Med*, **122**(3): p. 257-259.
116. Bodner, M., et al., (2011). Southeast Asian diversity: first insights into the complex mtDNA structure of Laos. *BMC Evolutionary Biology*, **11**(1): p. 49.
117. Zhang, X., et al., (2013). Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. *Nature Communications*, **4**: p. 2599.
118. Rashid, N., et al., (2010). *Sequence polymorphisms of mtDNA HV1, HV2, and HV3 regions in the Malay population of Peninsular Malaysia*. Vol. 124. 415-26.
119. X.K. et al (2015). Population data of mitochondrial DNA HVS-I and HVS-II sequences for 208 Henan Han Chinese. *Legal Medicine*, **17**(4): p. 287-294.
120. Xu, F.-L., et al., (2018). Characterization of mitochondrial DNA polymorphisms in the Han population in Liaoning Province, Northeast China. *Mitochondrial DNA Part A*, **29**(2): p. 250-255.
121. Turchi, C., et al., (2016). The mitochondrial DNA makeup of Romanians: A forensic mtDNA control region database and phylogenetic characterization. *Forensic Science International: Genetics*, **24**: p. 136-142.

122. Đỗ Mạnh Hưng, Phạm Nhật Khôi và cộng sự (2016). Sự đa dạng di truyền vùng HV2 hệ gen ty thể của một số nhóm người Việt. *Tạp chí sinh học*, **38**(2): p. 243-249.
123. Nguyễn Thy Ngọc, Nguyễn Quang Huy và cộng sự, (2018). Đa hình vùng D-LOOP hệ gen ty thể của các cá thể dân tộc Kinh và mường cùng trong nhóm ngữ hệ Nam Á. *Tạp chí Công nghệ Sinh học*, **16**(2): p. 231-240.
124. al, B.S.T.e., (2014). Sequence polymorphism of mitochondrial dna hypervariable regions i and ii in malay population of malaysia. *Scientific World*, **12**(12): p. 24-29.
125. al, H.F.e., (2014). Role of mtDNA Haplogroups in the Prevalence of Knee Osteoarthritis in a Southern Chinese Population. *International Journal of Molecular Sciences*. **15**: p. 2646-2659.
126. Chen, F., et al., (2009). Sequence-length variation of mtDNA HVS-I C-stretch in Chinese ethnic groups. *Journal of Zhejiang University. Science. B*, **10**(10): p. 711-720.
127. Hsouna, S., et al., (2016). Study of the T16189C variant and mitochondrial lineages in Tunisian and overall Mediterranean region. *Mitochondrial DNA Part A*, **27**(2): p. 1558-1563.
128. Sangthong, P., A. Jansom, and N. Chinnabanchonchai, (2015). Sequence analysis of mitochondrial DNA hypervariable region I in Thai individuals. *Australian Journal of Forensic Sciences*, **47**(3): p. 345-354.
129. C.F.e., al, (2015). Single nucleotide polymorphisms of mitochondrial DNA HVS-I and HVS-II in Chinese Bai ethnic group. *Electrophoresis*, **36**(6): p. 930-936.
130. Sekiguchi K, et al (2008). Mitochondrial DNA population data of HV1 and HV2 sequences from Japanese individuals,. *Legal Medicine* (Tokyo, Japan), **10**(5): p. 284-286.

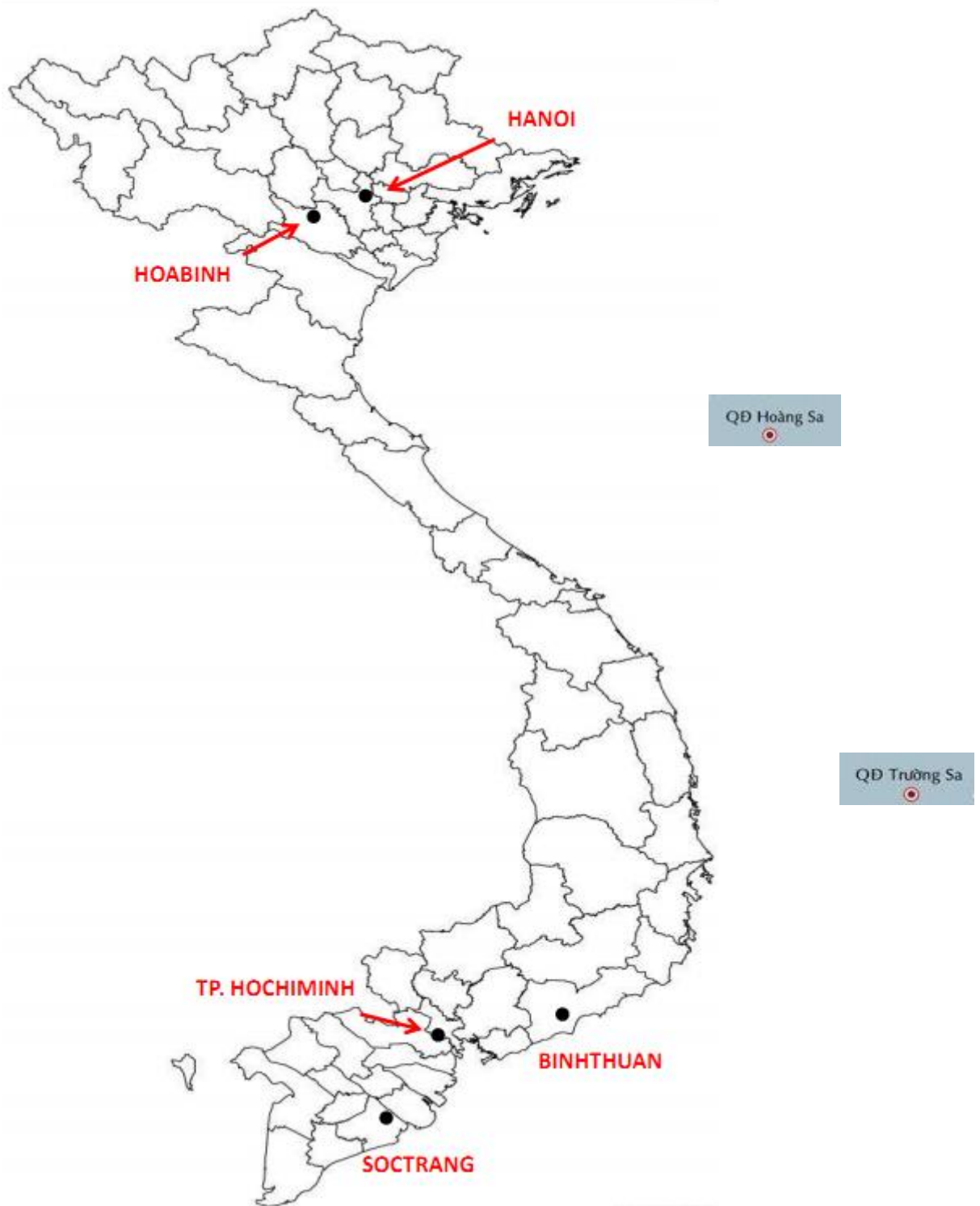
131. Kang L, et al (2009). Sequence polymorphisms of the mitochondrial DNA HVR I and HVR II regions in the Deng populations from Tibet in China., *Zhonghua Yi Xue Yi Chuan Xue Za Zhi*, **26**(6): p. 690-695.
132. Vanecek T., Vorel F., and S.M. et al (2004). *Mitochondrial DNA D-loop hypervariable regions*. Vol. 118.
133. Rakha, A., et al., (2017). *MtDNA sequence diversity of Hazara ethnic group from Pakistan*. Vol. 30.
134. Huỳnh Thị Thu Huệ, Nguyễn Đăng Tôn và cộng sự (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): p. 15-22.
135. al, E.S.-R.e., (2018) Mitochondrial T16189C Polymorphism Is Associated with Metabolic Syndrome in the Mexican Population. *Disease markers*, p. 3981315. doi:10.1155/2018/3981315.
136. Park KS, et al (2008). A mitochondrial DNA variant at position 16189 is associated with type 2 diabetes mellitus in Asians. *Diabetologia*, **51**(4): p. 602-608.
137. Tsu-Kung Lin, S.-D.C., Pei-Wen Wang et al, (2005). Increased oxidative damage with altered antioxidative status in type 2 diabetic patients harboring the 16189 T to C variant of mitochondrial DNA. *Annals of the New York Academy of Sciences*, **1042**(1): p. 64-69.
138. C. W. Liou, T.K.L., J. B. Chen et al. (2010). Association between a common mitochondrial DNA D-loop polycytosine variant and alteration of mitochondrial copy number in human peripheral blood cells., *Journal of Medical Genetics*, **47**(11): p. 723-728.
139. Zheng Ye, et al (2013). The association of the mitochondrial DNA OriB variant (16184–16193 polycytosine tract) with type 2 diabetes in Europid populations. *Diabetologia*, **56**(9): p. 1907-1917.
140. Skuratovskaia D. A., et al (2017). The association of the mitochondrial DNA oriB variants with metabolic syndrome. *Biomeditsinskaya Khimiya*, **63**(6): p. 533-538.

141. Golubenko M, et al (2015). Association of mitochondrial DNA polymorphism with myocardial infarction and prognostic signs for atherosclerosis. *Molecular Biology*, **49**(6): p. 867-874.
142. Mueller E., et al (2011). The mitochondrial T16189C polymorphism is associated with coronary artery disease in Middle European populations. *PloS one*, **6**(1): p. e16455.
143. Hicham Charoute, R.K., Safaa Bounaceur, et al, (2018). Novel variants of mitochondrial DNA associated with Type 2 diabetes mellitus in Moroccan population. *Mitochondrial DNA A DNA Mapp Seq Anal*, **29**(1): p. 9-13.
144. Ebner S., et al (2011). Mitochondrial haplogroups, control region polymorphisms and malignant melanoma: a study in middle European Caucasians. *PLoS One*, **6**(12): p. e27192.
145. Abdul Aziz Mohamed Yusoff, et al (2017). Detection of somatic mutations in the mitochondrial DNA control region D-loop in brain tumors: The first report in Malaysian patients,. *Oncology letters*, **14**: p. 5179-5188.
146. Chen, X.-Z., et al., (2016). Mitochondrial D310 instability in Chinese lung cancer patients. *Mitochondrial DNA Part A*, **27**(2): p. 1177-1180.
147. Haitham Ahmed Yacoub et al, (2014). Novel Mutations in the Displacement Loop of Mitochondrial DNA are Associated with Acute Lymphoblastic Leukemia: A Genetic Sequencing Study, *Asian Pacific Journal of Cancer Prevention*, **15**: p. 9283-9289.
148. Ines Badano, et al (2018). Mitochondrial DNA ancestry, HPV infection and the risk of cervical cancer in a multiethnic population of northeastern Argentina Ines Badan. *PLoS One*, **13**(1): p. e0190966,.
149. JUAN ZHOU, et al (2017). Sequence variations of mitochondrial DNA D-loop region in patients with acute myeloid leukemia,. *Oncology Letters*, **14**: p. 6269-6276.

150. Silkjaer T et al, (2013). Characterization and prognostic significance of mitochondrial DNA variations in acute myeloid leukemia. *Eur J Haematol*, **90**(5): p. 385-396.
151. AnudishiTyagi, et al (2018). Pattern of mitochondrial D-loop variations and their relation with mitochondrial encoded genes in pediatric acute myeloid leukemia. *Mutation Reseach*, **810**: p. 13-18.
152. Chen, Y.F., et al (2018), Mitochondrial DNA Haplogroups and the Risk of Sporadic Parkinson's Disease in Han Chinese. Chinese medical journal. *Chinese medical journal*, **128**(13): p. 1748-1754.
153. Hu W.X, et al (2018). Single nucleotide polymorphisms in the mitochondrial displacement loop and age-at-onset of non-small cell lung cancer. *Genetics and Molecular Research*, **14**(1): p. 2512-2517.
154. Xingyun Su, W.W., Guodong Ruan et al, (2016). A Comprehensive Characterization of Mitochondrial Genome in Papillary Thyroid Cancer,. *Int. J. Mol. Sci.*, **17**: p. 1594-1608.
155. Marjolein J.A. Weerts, et al (2016). Mitochondrial DNA content in breast cancer: Impact on in vitro and in vivo phenotype and patient prognosis. *Oncotarget*, **7**(20): p. 29166-29176.
156. Yun Wang et al, (2016). High copy number of mitochondrial DNA predicts poor prognosis in patients with advanced stage colon cancer,. *Int J Biol Markers*; 31(4): e382-e388, 2016. **31**(4): p. e382-e388.
157. Huakang Tu et al (2015). Mitochondrial DNA copy number in peripheral blood leukocytes and the aggressiveness of localized prostate cancer Huakang Tu. *Oncotarget*, **6**(39): p. 41988-41996.
158. Jianhua Chen et al (2018). Clinical application of plasma mitochondrial DNA content in patients with lung cancer, *Oncology Letters*, **16**(6): p. 7074-7081.
159. Babak Rahmani, et al (2012). Mutation Screening in the Mitochondrial D-Loop Region of Tumoral and Non-tumoral Breast Cancer in Iranian Patients. *Acta Medica Iranica*, **50**(7): p. 447-453.

160. Anna M. Czarnecka, T.K., Katarzyna Plak, et al, (2010). Mitochondrial genotype and breast cancer predisposition. *Oncology Reports*, **24**(6): p. 1521-1534.
161. A.e., et al (2015). Detection of novel polymorphisms in the mitochondrial DNA D-Loop hypervariable region HVI from 400 healthy unrelated individuals from central and North-central Iraq Ameer. *Academic Journals*, **14**(14): p. 1186-1184.
162. Wibhu Kutanan, et al (2018). New insights from Thailand into the maternal genetic history of Mainland Southeast Asia. *European Journal of Human Genetics*. **26**: p. 898-911.
163. A.L. et al (2018). High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia. *Am J Phys Anthropol.*, **165**(2): p. 238-255.
164. Meng, J.-H., et al., (2017). Investigation of control region sequences of mtDNA in a Chinese Maonan population. *Mitochondrial DNA Part A*, **28**(3): p. 350-354.
165. Bhatti, S., et al., (2018). Genetic perspective of uniparental mitochondrial DNA landscape on the Punjabi population, Pakistan. *Mitochondrial DNA Part A*, **29**(5): p. 714-726.

**PHỤ LỤC 1: BẢN ĐỒ THU MẪU 4 DÂN TỘC KINH, CHĂM,  
MƯỜNG, KHMER NGƯỜI VIỆT NAM**





### PHỤ LỤC 3: PHÂN LOẠI NHÓM ĐƠN BỘI DNA TY THỂ DỰA TRÊN SNP CỦA VÙNG HV1 VÀ HV2

(theo mtDNAMANAGER - a forensic mtDNA tool - <http://mtmanager.yonsei.ac.kr>)

#### East Asian Haplogroup

Haplogroup	HV1	HV2	HV3 etc.
D4/G	16223-16362		489
D1	16223-16325-16362		489
D4a	16129-16223-16362	152	(16519), 489
D4a3	16129-16223-16249-16362	152	16519, 489
D4a4	16129-16223-16294-16362	152	16519, 489
D4b1	16223-16319-16362		489-523d-524d
D4b2a	16223-16355-16362	280	489-523d-524d
D4b2a	16223-16362	199	489-523d-524d
D4b2b	(16223)-16362	194	16519, 489-523d-524d
D4c	16245-16362		489
D4e1	16223-16362	94	489
D2	16129-16223-16271-16362		489
D4e*	16223-16256-16362		489
D4g1	16223-16278-16362		489-573.pC
D4g2	16223-16362	195-298	489
D4g2a	(16223)-16274-16362	298	489
D4h1	16174-16223-16362	146-183	489
D4h*	16174-16223-16311-16362	152	489
D4h3	16223-16301-16342-16362		489
D4i	16223-16294-16362		489
D4j1a	16068-16223-16362		489
D4j2	16223-16291-16362	152	489
D4j3	16184-16223-16311-16362		489
D4j*	16223-16293-16362		489
D4k	16192-16223-@16362	195	489
D4l	16223-16362		16368, 489
D4m1	16244-16362		489
D4m2	16042-16223-16362		489

D4n	16223-16355A-16362		489
D4o	16223-16290-16362	195	489
D4p	16223-16362	195-198	489
D4q	16223-16256-16362	200	489
D5	16189-16223-16362	150	489
D5a1	16182C-16183C-16189-16223-16362	150-309d	16390, 68, 489
D5a2	16182Y-16183C-16189-16223-16266-16362	150	489-523d-524d
D5a3	16182C-16183C-16189-16223-16360-16362	150	489
D5b	16189-16223-16362	150	456-489
D5b1a	16167-16189-16223-16362	150	456-489
D5b1b	16189-16216-16223-16362	150	16519, 456-489
D5c	16188.1C-16193.1C-16362	150-(151)-152	489
D5d	16148-16189-16223-16362	150	489
D6a	16189-16223-16274-16362	146	489
D6c	16189-16223-16311-16362	152	489
G1a1	16223-16325-16362	150	16519, 489
G1a2	16184-16223-16290-16362	150	489
G1a3	16184-16214-16223-16362	150	489
G1b	16129-16223-16362		16017, 489
G2a	16189-16223-16278-16362		489
G2a1	16223-16227-16278-(16362)		489
G2a2	16051-16150-16223-16278-16362		489
G2a3	(16189)-16223-16278-16362	260	489
G2a4	16204-16223-16278-16362	146-152	489
G2b	16172-16223-16362	263	489
G2c	16093-16223-16278-16303-16311-16362		64, 489
G3a	16223-16274-(16362)	143-(152)	489
G3b1	16223-16274-16362	195	489
G4	16114A-16223-16362	191.1A	489
M7a	16209-16223		489
M7a1	16209-16223-16324		489-(523d-524d)
M7a2	16140-16209-16223	146	489
M7b	(16129)-16223-16297	150-199	489

M7b1	16129-16192-16223-16297	150-199	489
M7b2	16129-16189-16223-16297-16298	150-199	489
M7b4	16129-16223-16297	150-199-204	489
M7b3	16086-16223-16297-16324	199	489
M7d	16129-16152-16179-16192-16223	199	489
M7c	16223	146-199	16519, 489-523d-524d
M7c1	16223-16295	(146)-199	16519, 489-523d-524d (or 513d-514d)
M7c3	16223	146A-199	489-523d-524d
M8	16223-16298		489
M8a	16223-16298-16319		489
M8a1	16223-16298-16319-16311		489
M8a2	16184-16223-16298-16319		489
CZ	16223-16298	249d	489
C	(16223)-16298-16327	249d	16519, 489
C1	(16223)-(16298)-16325-16327	249d-290d-291d	16519, 489-523d-524d
C1a	16223-16298-16325-16327-16356	93-249d-290d- 291d	16519, 489-523d-524d
C1b	(16223)-(16298)-16325-16327	249d-290d-291d	16519, 489-493-523d- 524d
C1c	16223-16298-16325-16327	249d-290d-291d	16519, 489-523d-524d
C1d	16051-(16223)-(16298)-16325-16327	194-249d-290d- 291d	16519, 489-523d-524d
C1d1c	16051-16188-16223-(16298)-16325-16327- 16362	(194)-249d- 290d-291d	16519, 489-523d-524d
C4a1	16129-16223-16298-16327	195-249d	16519, 489
C4a2'3'4	16223-16298-16327-16357	249d	16519, 489
C4b1/C7a	16223-16298-16327	146-249d	16519, 489
C5	16223-16288-16298-16327	249d	16519, 489
C5a	16223-16261-16288-16298-@16327	249d	16519, 489
C5b1	16148-16223-16288-16298-16327	249d	16519, 489
C7a1	16223-16298-16327	146-153-249d	16519, 489
pre-Z	16223-16260-16298	152-249d	489

Z	16185-16223-16260-16298	152-249d(or 247d)	489
Z*	16185-16223-16260-16294-16298	150-249d	489
M9a*	16223-16234-16362	153	489
M9a	16223-16234-16316-16362		489
M9d	16158-16223-16234-16362	150-152-153	16519, 489
M9c	16223-16234-16271-16362	153	489
M9b	16051-16209-16223-16362		489-573.pC
M10	16223-16311		489-573.pC
M10a	16129-16223-16311-16357		16497, 489, 523d-524d-573.pC
M10b	16066-16223-16311		489-573.pC
M11	16223	215-318-326	489
M12	16223-16234-16290		489
M13a	16145-16188(or 16188d)-16223	152	489
F	16304	249d	
F1ac	16129-16304	249d	16519, 523d-524d
F1a	16129-16172-16304	249d	16519, 523d-524d
F1a1	(16129)-16162-16172-(16304)	249d	(16519), 523d-524d
F1a2	16129-16172-16304	249d	16465-16519, 523d- 524d
F1a3	16129-16172-16284-16304-16311	249d	16390-16519, 523d- 524d
F1c	16111-16129-16304	152-249d	16519, 523d-524d
F1bde	(16183C)-16189-16304	249d	16519, 523d-524d
F1b	(16182C)-16183C-16189-16232A-16249- 16304-16311	249d	(16519), 523d-524d
F1d	16189-16304	146-249d	16519, 523d-524d
F1e	16189-16304-16355	249d	16519, 523d-524d
F2*	16304	235-249d	
F2a	16291-16304	249d	
F2b	16203-16304	249d	
F3	16298-16304-16362	249d	

F3a	16298-16304-16355-16362	249d	
F3b	16220C-16298-16304-16362	249d	
F4a	16207-16304-16362	146-249d	16399
F4b	16218-16304-16311	249d	573.pC
B4	16183C-16189-16217		
B4a	16182C-16183C-16189-16217-16261		(16519), 523d-524d
B4a1a	16182C-16183C-16189-16217-16261	146	(16519), 523d-524d
B4a1b	16182C-16183C-16189-16217-16261- 16288-16311		(16519), 523d-524d
B4g	16182C-16183C-16189-16217-16261-16292		16519, 523d-524d
B4b	16183C-16189-16217		16519, 499
B4b1	16136-16183C-16189-16217		16519, 499
B2a	16111-16183C-16189-16217		16483-16519, 499
B2c	16182C-16183C-16189-16217		16519, 499
B2f	16183C-16189-16217-(16298)	114G	16519, 499
B4d1	16183C-16189-16217	316	
B4d3	16183C-16185-16189d-16217-16234	(151)	16519, 546
B4c1a	16183C-16189-16217-16311		
B4c1b	16140-16183C-16189-16217-16274	150	
B4c1c	16183C-16189-16217-16311	150-195-214	
B4c2	16147-16184A-16189-16217-16235		
B4f	16168-16172-16183C-16189-16217-16249- 16325	200	16390
B5	16140-16189		16519, 523d-524d
B5a	16140-16189-16266R	210	16519, 523d-524d
B5a2	16129-16140-16187-16189-16266R	93-210	16519, 523d-524d
B5b	16140-16183C-16189-16243		16519, 523d-524d (or 513d-514d)
B5b2	16111-16140-16183C-16189-16234-16243		(16463)-16519, 523d-524d
B5b3	16140-16183C-16189-16243-16256	103-189-199- 203-204	16519, 523d-524d
R11	16183C-16189-16311	185-189	16519

A	16223-16290-16319	235	
A4	16223-16290-16319-16362	235	523d-524d
A2	16111-16223-16290-16319-(16362)	146-(235)	64, 523d-524d
A4a	16223-16249-16290-16319-16362	(235)	523d-524d
A4b	16189-16223-16290-16319-16362	235	523d-524d
A4c	16223-16290-16319-16362	200-235	523d-524d
A4d	16223-16290-16319-16362	151-235	523d-524d
A5a	16187-16223-16290-16319	235	523d-524d
A5b	16126-(16223)-16235-16290-16319	235	523d-524d
A5c	16129-16213-16223-16290-16319	152-235	
A7	16051-16223-16290-16319	152-235	523d-524d
A8	16223-16242-16290-16319	146-152-(235)	64, 523d-524d
A11	16223-16290-16293C-16319	152-235	523d-524d
N9a	16223-16257A-16261	150	
N9a1	16111-16129-16223-16257A-16261	150	
N9a3	16129-16223-16257A-16261	150	
N9a2	16172-16223-16257A-16261	150	
N9a2ab	16172-16223-16257A-16261	150	16497
N9a4	16145-16172-16223-16245-16257A-(16261)	150	524.1A-524.2C
N9a5	16172-16189-16209-16223-16257A-16261	150	
N9a6	16223-16257A-16261-16292	150	16519
N9b	16183C-16189-16223		
N9b1b	16129-16183C-16189-16223		16390-16519
N9b1c	16183C-16189-16223	94	16519
N9b2	16183C-16189-16223-16294-16309		16519
Y	16126-16231		16519
Y1	16126-16231-16266	146	
Y2	16126-16231-16311		482

**PHỤ LỤC 2: CHI TIẾT TOÀN BỘ ĐA HÌNH TRÊN VÙNG HV1, HV2 CỦA DNA TY THỂ Ở 517 MẪU NGHIÊN CỨU THUỘC 4 DÂN TỘC KINH, CHĂM, MUÔNG, KHMER NGƯỜI VIỆT NAM VÀ PHÂN NHÓM ĐƠN BỘI THEO CÁC SNP ĐẶC TRƯNG TRÊN VÙNG HV1 VÀ HV2 CỦA 4 DÂN TỘC TRÊN**

Dân tộc	Mẫu	Haplogroup	HV1(16000+)	HV2
Chăm	Cham07	B4	C147T, A183C, C184A, T189C, T217C, T362C, C465T, T519A	A73G, A263G, 315C
	Cham101	B4	A182C, A183C, T189C, T261C, C348T, T519A	A73G, A263G, 309CC, 315C
	Cham38	B4a	T093C, C174T, A182C, A183C, T189C, T217C, T261C, T519A	A73G, T146C, A153G, A263G, 315C 522-523d
	Cham39	B4a	A182C, A183C, T189C, T217C, T261C, T519A	A73G, T146C, A263G, 315C, 522-523d
	Cham53	B4a	A182C, A183C, T189C, T217C, C218T, T224C, T261C	A73G, T146C, T152C, A263G, 315C, 522-523d
	Cham75	B4a	A182C, A183C, T189C, T217C, C218T, T224C, T261C	A73G, T146C, T152C, A263G, 315C, 522-523d
	Cham77	B4a	A182C, A183C, T189C, T217C, C218T, T224C, T261C	A73G, T146C, T152C, A263G, 315C, 522-523d
	Cham08	B4b	T136C, A183C, T189C, T217C, A241C, T519A	A73G, T146C, A263G, 309CC, 315C
	Cham31	B4c	G129A, T140C, A166C, A182C, A183C, T189C, T217C, G274A, A335G, T519A	A73G, C150T, A263G, 309CC, 315C
	Cham43	B4c	T140C, A182C, A183C, T189C, T217C, G274A, T311C, T519A	A73G, T146C, C150T, A263G, 309CC, 315C
	Cham11	B4c	C147T, A183C, C184A, T189C, T217C, A235G, T519A	A73G, A263G, 309C, 315C
	Cham19	B4c	C147T, A183C, C184A, T189C, T217C, A235G, T362C, C465T, T519A	A73G, A263G, 315C
	Cham21	B4c	C147T, A183C, C184A, T189C, T217C, A235G, T519A	A73G, A263G, 309C, 315C
	Cham22	B4c	C147T, A183C, C184A, T189C, T217C, T362C, C465T, T519A	A73G, A263G, 315C
	Cham30	B4c	C147T, A183C, C184A, T189C, T217C, A235G, T519A	A73G, A263G, 309C, 315C
	Cham35	B4c	C147T, T183C, C184A, T189C, C201T, G213A, T217C, C232T, C270T, C292T, T519A	A73G, A263G, 315C
	Cham40	B4c	C147T, T183C, C184A, T189C, T217C, T362C, C465T, T519A	A73G, A263G, 315C
	Cham54	B4c	C147T, A183C, C184A, T189C, T217C, A235G, T519A	A73G, A263G, 309C, 315C
	Cham66	B4c	C147T, A183C, C184A, T189C, T217C, A235G, T519A	A73G, A263G, 309C, 315C

Cham71	B4c	C147T, A183C, C184A, T189C, T217C, A235G, T519A	A73G, A263G, 309C, 315C
Cham78	B4c	C147T, A183C, C184A, T189C, T217C, A235G, T362C, C465T, T519A	A73G, A263G, 315C
Cham83	B4c	C147T, A183C, C184A, T189C	A73G, A263G, 315C
Cham29	B4g	A181C, A182C, A183C, T189C, C201T, G213A, T217C, C232T, C270T, C292T, T519A	A73G, T195C, A263G, 309C, 315C, 522-523d
Cham105	B4g	A181C, A182C, A183C, T189C, G213A, T217C, C261T, C292T, T519A	61A, 62A, A73G, A263G, 308-309d, 315C, 522-523d
Cham03	B5a	T140C, G153A, T178C, A183C, T189C, A207G, C266A, T519A	42G, A73G, G103A, A210G, A263G, 309CC, 315C, 522-523d
Cham12	B5a	T140C, A183C, T189C, C266A, C291T, T519A	A73G, A210G, A263G, 309C, 315C, 455T, 522-523d
Cham18	B5a	T140C, A183C, T189C, C266A, C291T, T519A	A73G, A210G, A263G, 309C, 315C, 455T, 522-523d
Cham25	B5a	T140C, A182C, A183C, T189C, C266A, T519A	A73G, A210G, A263G, 309CC, 315C, 522-523d
Cham28	B5a	T140C, A183C, T189C, C266A, T519A	A73G, A210G, A263G, 309CC, 315C, 522-523d
Cham32	B5a	T140C, A182C, A183C, T189C, C261T, C266A, T519A	T57C, 61A, A73G, T152C, A210G, A263G, 309CC, 315C, A368G
Cham42	B5a	T140C, A182C, A183C, T189C, C261T, C266A, T519A	A73G, T152C, A210G, A263G, 309CC, 315C, 522-523d
Cham46	B5a	T140C, A182C, A183C, T189C, C266A, T519A	A73G, A210G, A263G, 309C, 315C, 522-523d
Cham48	B5a	T140C, A183C, T189C, C257A, C266A, T519A	A73G, A210G, A263G, 309C, 315C, 522-523d
Cham59	B5a	G129A, T140C, A182C, A183C, T189C, C266A, T519A	A73G, A210G, A263G, 309C, 315C, 522-523d
Cham67	B5a	T140C, C148T, A183C, T189C, T243C, C266A, T519A	A73G, A210G, A263G, 309C, 315C, 522-523d
Cham81	B5a	T140C, A183C, T189C, C257A, C266A, T519A	A73G, A210G, A263G, 309C, 315C, 522-523d
Cham86	B5a	T140C, A183C, T189C, C266A, C291T, T519A	A73G, A210G, A263G, 309C, 315C, 455C, 522-523d
Cham91	B5a	T140C, A183C, T189C, C266A, T519A	A73G, A210G, A263G, 309C, 315C, 522-523d
Cham103	B5a	T140C, G153A, T178C, A183C, T189C, A207G, C266A, T519A	42G, A73G, G103A, A210G, A263G, 309C, 315C
Cham106	B5a	T140C, C148T, A183C, T189C, T243C, C266A, T519A	A73G, A210G, A263G
Cham108	B5a	T140C, G153A, T178C, A183C, T189C, A207G, C266A, T519A	42G, A73G, G103A, A189G, A210G, A263G, 309CC, 315C, 522-523d



Cham115	B5a	T140C, A183C, T189C, C266A, T304C, T519A	A73G, A210G, A263G, 315C, 522-523d
Cham94	B5b	C067G, T140C, A183C, T189C, T243C, T519A	A73G, A103G, T152C, T204C, A263G, 315C, 522-523d
Cham27	B5b	C104T, C111T, T140C, A182C, A183C, T189C, C234T, T243C, C291T, C463T, T519A	A73G, T131C, A263G, 309CC, 315C, 522-523d
Cham99	B5b	C104T, C111T, T140C, A182C, A183C, T189C, C234T, T243C, C291T, C463T, T519A	A73G, T131C, A263G, 309CC, 315C, 522-523d
Cham52	C	T086C, C223T, T298C, C327T, T357C, T519A	C64T, A73G, 249DelA, A263G, 309C, 315C
Cham84	D4a	G129A, T209C, C223T, A272G, T362C, T519A	A73G, T152C, 249delA, A263G, 309C, 315C, G316C, 522-523d
Cham85	D4a	T093C, G129A, C223T, T263C, T362C, T519A	A73G, T152C, A263G, 309C, 315C, 523CA
Cham69	D4e	C167T, C223T, C320T, T362C	A73G, G94A, A263G, 309C, 315C
Cham33	E	C223T, C291T, T362C, G390A, T519A	A73G, A193G, A263G, 309C, 315C
Cham51	F	T189C, G274A, T304C, T519A	A73G, 249delA, A263G, 309CC, 315C, 522-523d
Cham16	F1a	G129A, T172C, T304C	A73G, 249delA, A263G, 315C, 522-523d
Cham61	F1a	G129A, T172C, T304C, T362C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Cham68	F1a	G129A, A162G, T172C, T189C, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Cham88	F1a	T093C, G129A, A162G, T172C, T304C, A399G, T519A	A73G, 249delA, A263G, 315C, 522-523d
Cham45	F1a	C108T, G129A, A162G, T172A, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Cham49	F1a	C108T, G129A, A162G, T172A, A183C, T189C, T304C, T519A	A73G, 249delA, A263G, 309C, 315C, 522-523d
Cham50	F1a	C108T, G129A, A162G, T172A, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Cham89	F1a	C108T, G129A, A162G, T172A, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Cham01	M	G145A, A181C, T189C, C192T, C262T, T304C, G390A, T519A	A73G, A210G, A263G, 309C, 315C
Cham06	M	G129A, T209C, C223T, T325C	A73G, T152C, A200G, A263G, 315C
Cham15	M	C223T, C278T, T311C, C354T, T519A	A73G, T199C, A263G, 309CC, 315C
Cham23	M	T093C, T209C, C223T, T224C, T263C, C278T, G319A	A73G, T146C, C150T, C151T, A263G, 309CC, 315C, 522-523d

Cham24	M	G145A, A181C, T189C, C192T, C262T, T304C, T311C, G390A, T519A	A73G, A210G, A263G, 309C, 315C
Cham47	M	A129G, A183C, T189C, G213A, C218T, C223T, G274A, T519A	A73G, C150T, A263G, 309CC, 315C, 522-523d
Cham55	M	A129G, A183C, T189C, G213A, C218T, C223T, G274A, T519A	A73G, C150T, A263G, 309CC, 315C, 522-523d
Cham60	M	A129G, A183C, T189C, G213A, C218T, C223T, G274A, T519A	A73G, C150T, A263G, 309CC, 315C, 522-523d
Cham70	M	G145A, A181C, A182C, C223T, C291T, T304C, T519A	A73G, A210G, A263G, 309C, 315C
Cham74	M	G145A, A181C, T189C, 192T, C262T, T304C, G390A, T519A	A73G, A210G, A263G, 309C, 315C
Cham82	M	G129A, T209C, C223T, T325C	A73G, T152C, A200G, A263G, 315C
Cham95	M	G129A, T209C, C223T, T325C	A73G, T152C, A200G, A263G, 315C
Cham96	M	G129A, T209C, C223T, T325C	A73G, T152C, A200G, A263G, 315C
Cham100	M	G129A, T209C, C223T, T325C	A73G, T152C, A200G, A263G, 315C
Cham102	M	T093C, T209C, C223T, T224C, T263C, C278T, G319A	A73G, T146C, C150T, C151T, A263G, 309CC, 315C, 522-523d
Cham58	M12	G129A, C223T, C234T, C261T, C262T, G274A, C290T	A73G, G143A, T152C, A263G, 309C, 315C, T318C
Cham65	M7b	C223T, T297C, T311C	A73G, C150T, T199C, T204C, A263G, 309CC, 315C
Cham97	M7b1	G129A, T189C, C192T, C223T, T297C	A73G, C150T, T199C, A202G, A263G, 309C, 315C, C332T
Cham92	M7b1	G129A, T189C, T297C	A73G, C150T, T199C, T204C, A263G, 309C, 315C
Cham114	M7b1	T093C, G129A, A183C, T189C, C223T, T297C, T311C	A73G, C150T, T199C, T204C, A263G, 309CC, 315C
Cham112	M7b	C223T, C278T, C354T, G390A, T519A	A73G, T199C, A263G, 309C, 315C
Cham107	M7b	C223T, C278T, T311C, C354T, T519A	A73G, T199C, A263G, 319CC, 315C
Cham20	M7c	T075C, C223T, 293T, C295T, T519A	A73G, T146C, T152C, T199C, A263G, 309C, 315C, 522-523d
Cham44	M7c	T075C, C223T, 293T, C295T, T519A	A73G, T146C, T152C, T199C, A263G, 309C, 315C, 522-523d
Cham113	M7c	C223T, C295T, T362C, T519A	A73G, T146C, T199C, A263G, 309CC, 315C, 522-523d
Cham87	M8a	C184A, C223T, T298C, G319A	A73G, A263G, 315C
Cham98	M8a	C184A, T189C, C223T, T298C, G319A	A73G, A263G, 315C
Cham17	M9a	C223T, C256T, T311C, T362C, T519A	A73G, T125G, T146C, A153G, A263G, 309C, 315C

Cham10	M9b	A051G, T209C, C223T, T362C, T519A	A73G, A153G, A263G, 315C
Cham80	N21	G129A, C193T, C223T, T325C, T519A	A73G, C150T, T195C, A263G, 315C, 337d, 522-523d
Cham09	N9a	T093C, T140C, T189C, C223T, 257A, C261T, C292T, T519A	A73G, C150T, A263G, 309C, 315C
Cham14	N9a	T093C, T140C, T189C, C223T, 257A, C261T, C292T	A73G, C150T, A263G, 309C, 315C
Cham26	N9a	T093C, T189C, C223T, C256T, 257A, C261T, C292T, C354T, T519A	A73G, C150T, A263G, 309C, 315C
Cham64	N9a	T093C, T189C, C223T, C256T, 257A, C261T, C292T, C354T, T519A	A73G, C150T, A263G, 309C, 315C
Cham79	N9a	C111T, G129A, C223T, 257A, C260T, C261T	A73G, T146C, C150T, A263G, 309C, 315C
Cham02	N9a	T092C, G145A, T172C, C223T, C245T, 257A, C261T, T311C, T519A	A73G, C150T, T152C, A263G, 315C, 523CACA
Cham36	N9a	T092C, G145A, T172C, C223T, C245T, 257A, C261T, T311C, T519A	A73G, C150T, T152C, A263G, 315C, 523CACA
Cham72	N9a	T092C, G145A, T172C, C223T, C245T, 257A, C261T, T311C, T519A	A73G, C150T, T152C, A263G, 315C, 523CACA
Cham104	N9a	T092C, G145A, T172C, C223T, C245T, 257A, C261T, T311C, T519A	A73G, C150T, T152C, A263G, 315C, 523CACA
Cham109	N9a	T092C, G145A, T172C, C223T, C245T, 257A, C261T, T311C, T519A	A73G, C150T, T152C, A263G, 315C, 523CACA
Cham57	R	C256T, C290T, C465T	A73G, T195C, A263G, 309C, 315C, 522-523d
Cham93	R	C256T, C290T, C465T	A73G, T195C, A263G, 309C, 315C, 522-523d
Cham63	R11	T086C, A182C, A183C, T189C, T311C, G390A, A399G, T519A	A73G, A189G, T215C, A263G, 309CCC, 315C
Cham111	R11	T086C, A182C, A183C, T189C, T311C, G390A, A399G, T519A	A73G, A189G, T215C, A263G, 309CCC, 315C
Cham04	R9b	C221T, T249C, T288C, C301T, T304C, G390A, T519A	A73G, A263G, 315C, G329A
Cham13	R9b	C221T, T249C, T288C, C301T, T304C, G390A, T519A	A73G, A263G, 315C, G329A
Cham34	R9b	T249C, C259T, T288C, C301T, T304C, G390A, T519A	A73G, T152C, T195C, A263G, 309C, 315C, 523CA
Cham37	R9b	C192T, T304C, A309G, G390A, T519A	A73G, A183G, T204C, G207A, A263G, 309CC, 315C, 522-523d
Cham41	R9b	C192T, T304C, A309G, G390A, T519A	A73G, A183G, T204C, G207A, A263G, 309CC, 315C, 522-523d
Cham56	R9b	T124C, C148T, C290T, T304C, A309G, G390A, T519A	A73G, A263G, 309C, 315C
Cham62	R9b	C221T, T249C, T288C, C301T, T304C, G390A, T519A	A73G, A263G, 315C, G329A
Cham76	R9b	C221T, T249C, T288C, C301T, T304C, G390A, T519A	A73G, A263G, 315C, G329A

	Cham90	R9b	C192T, T304C, G309A, A390G, T519A	A73G, A183G, T204C, G207A, A263G, 309CC, 315C, 522-523d
	Cham110	R9b	C192T, C193T, T304C, G309A, A390G, T519A	A73G, A183G, T204C, G207A, A263G, 309CC, 315C, 522-523d
<i>Kinh</i>	Khin36	B4	G129A, A182C, A183C, T189C, T217C, T362C	A73G, A263G, 309CC, 522-523d
	Khin43	B4	A182C, A183C, T189C, T217C, C261T, T519A	A73G, A263G, 309C, 315C
	Khin65	B4	T092C, A182C, A183C, T189C, T217C, G274A, A289G, C301T, T519A	A73G, T152C, A183G, A244G, A263G, T310C, A374G
	Khin72	B4	T140C, A182C, A183C, T189C, T217C, G274A, A305T, A335G, T519A	A73G, C150T, T195C, A263G, 309CC, 315C
	Khin78	B4	A182C, A183C, T189C, T217C, G274A, A289G, C301T, T519A	A73G, A183G, A263G, 309CC, 315C, A374G, 522-523d
	Khin37	B4a	C168T, A182C, A183C, T189C, T217C, C261T, T311C, T519A	A73G, T146C, A263G, 309CCC, 315C, 522-523d
	Khin64	B4a	A182C, A183C, T189C, C234T, C256T, C261T, T519A	A73G, A263G, 309CC, 315C, 522-523d
	Khin87	B4a	A182C, A183C, T189C, T217C, A219G, C261T, C286T, T519A	A73G, T146A, A263G, 522-523d
	Khin33	B4b	T126C, T136A, A183C, T189C, T217C, C260T, C287T, C325T, T519A	A73G, A200G, A263G, 315C, 522-523d
	Khin16	B4c	C147T, C168T, A183C, C184A, T189C, T231C, A235G, T519A	A73G, A263G, 309C, 315C
	Khin17	B4c	C147T, C168T, A183C, C184A, T189C, T217C, T231C, A235G, T519A	A73G, A263G, 309C, 315C
	Khin26	B4c	C147T, C168T, A183C, C184A, T189C, T217C, C234T, A235G, T519A	A73G, A263G, 309C, 315C
	Khin34	B4c	A183C, T189C, T217C, C234T, T311C, A399G	A73G, C150T, T195A, A214G, A263G, 309CC, 315C
	Khin76	B4c	A181C, A182C, A183C, T189C, G213A, T217C, C292T, T311C, T519A	A73G, A263G, 309d, 315C, 522-523d
	Khin110	B4c	C147T, A183C, C184A, T189C, T217C, A235G, C362T, T519A	A73G, A263G, 309C, 315C
	Khin5	B4g	T093C, A181C, A182C, A183C, T189C, G213A, T217C, C242T, C261T, C287T, C292T, C301T, C355T, T519A	61A, A73G, A263G, 309C, 315C, 522-523d
	Khin27	B4g	A181C, A182C, A183C, T189C, G213A, T217C, C261T, C292T, T519A	A73G, T146C, T152C, A263G, 315C, 522-523d
	Khin49	B4g	A181C, A182C, A183C, T189C, C261T, C292T, T519A	A73G, C150T, A263G, 309CCC, 315C, 522-523d
Khin99	B4g	T093C, A181C, A182C, A183C, T189C, G213A, T217A, C242T, C250T, C261T, C287T, C292T, C301T, C355T, T519A	61A, A73G, A263G, 522-523d	
Khin3	B5	T140C, C187T, T189C, C256T, C266G, T519A	A73G, A93G, A210G, A263G, 315C, 522-523d	

Khin7	B5a	T140C, A182C, T189C, C266A	A73G, A210G, A263G, 309CC, 315C, 522-523d
Khin13	B5a	T140C, A183C, T189C, T243C, C266A, T311C, T519A	A73G, A210G, A263G, 309C, 315C, 522-523d
Khin31	B5a	T140C, A183C, T189C, T243C, C266A, T311C, T519A	A73G, A210G, A263G, 309CC, 315C, 522-523d
Khin35	B5a	T140C, C148T, A182C, A183C, T189C, T243C, C266A, T519A	A73G, A210G, A263G, 309CC, 315C, 522-523d
Khin48	B5a	T140C, A183C, T189C, T249C, C266A, T519A	A73G, A210G, A263G, 309C, 315C, 522-523d
Khin52	B5a	G129A, T140C, A183C, T189C, C266A, T519A	A73G, A210G, A263G, 309CC, 315C, 522-523d
Khin58	B5a	T140C, A183C, T189C, 193C, C266A, T519A	A73G, A210G, A263G, 309CC, 315C, 522-523d
Khin67	B5a	T140C, A183C, T189C, C234T, C266A, T519A	A73G, A210G, A263G, 309CC, 315C, 522-523d
Khin79	B5a	T140C, A183C, T189C, C266A, T519A	A73G, A210G, A263G, 309C, 315C, 522-523d
Khin81	B5a	T140C, A183C, T189C, C266A, T519A	A73G, A210G, A263G, 309CC, 315C, 522-523d
Khin84	B5a	T140C, A183C, T189C, C266A, T519A	A73G, A210G, A263G, 309C, 315C, 522-523d
Khin86	B5a	T140C, A183C, T189C, C266A, T304C, T519A	A73G, A210G, A263G, 315C, 522-523d
Khin98	B5a	T140C, A182C, A183C, T189C, C261T, C266A, T519A	A73G, A210G, A263G, 315C, 522-523d
Khin100	B5a	T140C, A183C, T189C, C266A, T519C	A73G, A210G, A263G, 309C, 315C, 522-523d
Khin60	B5b	T140C, A183C, T189C, T243C, T311C, T519A	A73G, G103A, T204C, A263G, 309CC, 315C, 522-523d
Khin47	B6	T093C, C179A, A182C, A183C, T189C	A73G, C150T, A263G, 309C, 315C
Khin15	C	T189C, C223T, T298C, C327T, T519A	A73G, 249delA, A263G, 309C, 315C
Khin97	C	T189C, C223T, T298C, C327T, T519A	A73G, 249delA, A263G, 309C, 315C
Khin85	D4a	C111G, G129A, C223T, T362C	A73G, T152C, A263G, 309C, 315C
Khin95	D5a	T092C, A164G, A182C, A183C, T189C, C223T, C266T, T362C	A73G, C150T, A263G, 309CC, 315C, 522-523d
Khin10	D5b	T092C, C148T, A183C, T189C, C223T, T362C, T519A	A73G, C150T, T152C, G185A, A263G, 309C, 315C, 522-523d
Khin18	D5b	A182C, A183C, T189C, C223T, T362C, T519A	A73G, C150T, A263G, 309CCC, 315C
Khin2	F	T157C, C256T, T304C, A335G	A73G, 249delA, A263G, 315C
Khin77	F	T075C, T172C, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d

Khin11	F1a	G129A, T172C, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Khin14	F1a	G129A, T172C, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Khin24	F1a	C108T, G129A, A162G, T172C, T304C, T519A	A73G, 249delA, A263G, 309C, 315C, 522-523d
Khin29	F1a	G129A, A162G, T172C, T304C, A399G, T519A	A73G, 249delA, A263G, 315C, 522-523d
Khin30	F1a	G129A, C218T, T304C, T311C	A73G, 249delA, A263G, 315C
Khin38	F1a	G129A, A162G, T172C, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Khin46	F1a	C108T, G129A, A162G, T172C, T304C, T519A	A73G, 249delA, A263G, 309C, 315C, 522-523d
Khin50	F1a	G129A, T172C, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Khin54	F1a	G129A, A162G, T172C, T304C, A399G, T519A	A73G, A200G, 249delA, A263G, 309C, 315C, 522-523d
Khin55	F1a	G129A, T172C, C287T, C295T, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Khin56	F1a	C108T, G129A, A162G, T172C, T304C, T519A	A73G, 249delA, A263G, 309CC, 315C, 522-523d
Khin61	F1a	G129A, A162G, T172C, C292T, T304C, T519A	A73G, T152C, 249delA, A263G, 315C, 522-523d
Khin63	F1a	G129A, T172C, C295T, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Khin69	F1a	G129A, A162G, T172C, T304C, T519A	A73G, 249delA, A263G, 309C, 315C, 522-523d
Khin74	F1a	G129A, T172C, C287T, C295T, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Khin96	F1a	C108T, G129A, A162G, T172C, T304C, T519A	A73G, 249delA, A263G, 309C, 315C, 522-523d
Khin106	F1a	C108T, G129A, T172C, T304C, C365T, T519A	A73G, 249delA, A263G, 309C, 315C, 522-523d
Khin109	F1a	C108T, G129A, A162G, T172C, T304C, T519A	A73G, 249delA, A263G, 309C, 315C, 522-523d
Khin8	F1a1	G129A, A162G, T172C, C292T, T304C, T519A	A73G, T152A, 249delA, A263G, 315C, 522-523d
Khin21	F1b	A183C, T189C, C292T, T304C, T519A	A73G, 249delA, A263G, 315C, 522-523d
Khin19	F1c	C111T, G129A, C266T, T304C, T519A	A73G, T152C, 249delA, A263G, 309C, 315C, 522-523d
Khin75	F3a	T093C, T249C, T298C, C355T, T362C, G390A	A73G, T152C, G207A, 249delA, A263G, 309CC, 315C
Khin23	M	C223T, C278T, T311C	A73G, C150T, G203C, A263G, 315C
Khin45	M	C223T, G274A	A73G, G185A, T195C, A263G, 309C, 315C

Khin105	M	C223T, C295T, T519A	A73G, T146C, T199C, A263G, 315C, 522-523d
Khin107	M10	T093C, G129A, C223T, T311C, T357C, A497G	A73G, T146C, A263G, 315C, 522-523d
Khin4	M12	C148T, C223T, C234T, C261T, C290T, T519A	A73G, A263G, 309C, 315C, T318C
Khin1	M7a	T086C, G129A, T209C, C223T, A272G, T519A	A73G, T152C, 249delA, A263G, 315C, G316A, 522-523d
Khin101	M7a	166d, T209C, C214T, C223T, C260T, T311C	A73G, A263G, 309C, 315C
Khin66	M7b	A235G, T311C, T356C, T519A,	A73G, T146A, T199C, A263G, 309C, 315C, 522-523d
Khin73	M7b	C185T, C223T, T297C	A73G, C150T, T199C, T204C, A263G, C271T, 309CC, 315C
Khin108	M7b	T136C, T189C, C223T, C278T, T311C	A73G, C150T, T152C, T199C, A263G, 315C
Khin9	M7b1	G129A, C192T, T297C	A73G, C150T, T199C, A263G, 309C, 315C
Khin12	M7b1	G129A, C192T, C223T, T297C	A73G, T131C, C150T, T199C, A263G, 309C, 315C
Khin20	M7b1	G129A, T189C, C192T, C223T, T297C, T356C	A73G, T131C, C150T, T199C, A263G, 309CC, 315C, C332T, 522-523d
Khin25	M7b1	G129A, C223T, T271C, T297C	A73G, C150T, T199C, A263G, 309C, 315C
Khin39	M7b1	G129A, C192T, C223T, C261T, T297C, T298C	A73G, C150T, T199C, A263G, 309C, 315C
Khin62	M7b1	G129A, C223T, T271C, T297C	A73G, C150T, T199C, A263G, 315C
Khin71	M7b1	G129A, T189C, T297C	A73G, C150T, T199C, T204C, A263G, 309C, T310C, C311T, 315C
Khin83	M7b1	G129A, C192T, C223T, T297C	A73G, C150T, T199C, A263G, 315C
Khin91	M7b1	G129A, C192T, C223T, T249C, T297C, T324C	A73G, C150T, T199C, A263G, 309C, 315C, C332T
Khin93	M7b1	G129A, A163G, T189C, C223T, T297C	A73G, C150T, T199C, T204C, A263G, 309C, 315C
Khin94	M7b1	G129A, G145A, C223T, T297C, T325C	A73G, C150T, T199C, A263G, 315C
Khin103	M7b1	G129A, C192T, C223T, T297C, T324C	A73G, C150T, T199C, A263G, 309C, 315C, C332T
Khin42	M7c	C223T, C295T, T519A	A73G, T146C, T199C, A263G, 315C, 522-523d
Khin41	M9a	C223T, C234T, T271C, T362C	A73G, C151T, A153G, A263G, 315C
Khin53	M9a	C223T, C234T, T356C, T362C	A73G, C150T, A153G, A263G, 309C, 315C, A385G

	Khin80	M9a	T093C, C223T, C234T, T271C, T362C	A73G, A153G, A263G, 315C
	Khin88	M9a	C223T, C234T, C287T, T362C	A73G, A153G, A263G, 309C, 315C
	Khin90	N	C223T, T263C, G274A, T311C, A343G, T357C, T519A	A73G, G103A, C151T, T152C, G260A, A263G, 315C
	Khin6	N9a	C223T, C257A, C261T, T311C	A73G, C150T, A263G, 315C
	Khin40	R9	T304C, A335G, T362C	A73G, C150T, T152C, A263G, 309C, 315C
	Khin44	R9	T093C, T157C, T304C	A73G, C151T, A263G, 315C
	Khin92	R9	T304C, T362C, T519A	A73G, A263G, 315C
	Khin28	R9b	A284G, T304C, A309G, G390A, T519A	A73G, A183G, A227G, A263G, 315C, 522-523d
	Khin59	R9b	T124C, C148T, T304C, A309G, G390A	A73G, T89C, T146C, A263G, 309CC, 315C
	Khin68	R9b	C239T, T304C, A309G, G390A, T519A	A73G, T152C, A263G, 309CC, 315C, 522-523d
	Khin82	R9b	T124C, C148T, T304C, A309G, G390A	A73G, T89C, T146C, A263G, 309C, 315C
	Khin22	Z	C185T, C223T, C260T, T298C	A73G, T152C, 249delA, A263G, 315C, 522-523d
	Khin32	Z	C185T, C223T, C260T, T298C	A73G, 249delA, A263G, 309C, 315C
	Khin70	Z	C185T, C223T, C260T, T298C, T519A	A73G, 249delA, A263G, 309C, 315C
	Khin104	Z	C185T, C223T, C260T, T298C, A317C	A73G, C151T, T152C, 249delA, A263G, 309C, 315C
<i>Kinh</i>	KN23	A	38delA, T86C, G129A, T209C, C223T, A272G, C290T, T519A	A73G, T152C, G225A, 249del, A263G, 315C, G316A
	KN33	A	T86C, G129A, T209C, C223T, A272G, C290T	A73G, T152C, G225A, 249del, A263G, 315C, G316A
	KN67	A	C071T, G129A, C223T, C234T, C262T, C263T, G274A, C290T	A73G, G143A, T152C, A263G, 309C, 315C, T317C
	KN20	B	T189C, C223T, C278T	A73G, C150T, A263G, 315C
	KN3	B4	C147T, A183C, C184A, T189C, T217C, A235G, C294T	A73G, A263G, 309C, 315C
	KN18	B4	C147T, A183C, C184A, T189C, T217C, A235G	A73G, A263G, 309C, 315C
	KN26	B4	A182C, A183C, T189C, T217C	A73G, A263G, 309CC, 315C
	KN31	B4	A183C, T189C, T217C	A73G, G207A, A263G, 315C
	KN35	B4	A183C, T189C, T217C, G274A, T311C	A73G, C150T, T195C, A263G, 309C, 315C



KN63	B4	A182C, A183C, T189C, T217C	A73G, A263G, 309C, 315C
KN68	B4	38delA, T140C, A182C, A183C, T189C, T217C, C242A, G274A, A335G	A73G, C150T, A263G, 309CC
KN100	B4	38delA, C147T, A183C, C184A, T189C, T217C, A235G, G518A	A73G, A263G, 309C, 315C
KN1	B4a	38delA, C67G, G129A, T157G, A182C, A183C, T189C, T217C, C261T, C354T	A73G, A263G, 315C
KN14	B4b	T136C, A175C, A183C, T189C, 193insC, T217C, C218T	A73G, A263G, 315C
KN37	B4b	T136C, A183C, T189C, T217C, C218T	A73G, A263G, 315C
KN27	B5a	T140C, A181C, A182C, A183C, T189C, C226A, A339G	A73G, A210G, A263G, 315C
KN48	B5a	T140C, A182C, A183C, T189C, C266A	A73G, A93G, A210G, A263G, 315C
KN51	B5a	T140C, A182C, A183C, T189C, 193inC, C266A	A73G, A210G, A263G, 309C, 315C
KN53	B5a	T140C, A182C, A183C, T189C, C266A	A73G, A210G, A263G, T310C, 315C
KN56	B5a	T93C, T140C, A183C, T189C, C266A	A73G, A210G, A263G, 309C, 315C
KN69	B5a	38delA, T140C, A183C, T189C, C266A	A73G, C150T, A210G, A235G, A263G, T310C, 315C
KN70	B5a	T140C, A181C, A182C, A183C, T189C, C266A	A73G, A210G, A263G, C308T, 310delT
KN71	B5a	T140C, A183C, T189C, 193insC, C266A	A73G, C150T, A210G, A235G, A263G, C308T, 309CC, 315C
KN92	B5a	T140C, T161A, A166C, A175C, A183C, T189C, C266A	A73G, A210G, A263G, 309C, 315C
KN52	C	T093C, T172C, C218T, C223T, T298C, C327T	A73G, 249del, A263G, 315C
KN83	C	C223T, T298C, C327T	A73G, T195C, C198T, 249DelA, A263G, 309CC, 315C
KN87	C	T172C, C218T, C223T, T298C, C327T	A73G, 249del, A263G, 315C
KN89	C	C223T, T298C, C327T, C328T	A73G, 249del, A263G, 309CC, 315C
KN59	D4	38del, 188C, 193C, C223T, C234T, T311C, T362C	A73G, C150T, C151T, T152C, A263G, C285T, T310C, 315C
KN46	D4a	038del, G129A, C223T, T362C	A73G, T152C, A263G, 309C, 315C
KN85	D4a	C111G, G129A, C223T, T362C	A73G, T152C, C182T, A263G, 309C, 315C
KN11	D5	T189C, C223T, T362C	A73G, C150T, 309C, 315C
KN78	F	A207G, T304C, A399G	A73G, T146C, T152C, 249del, A263G, A281G, 309C, 315C

KN8	F1a	C108T, G129A, A162G, T172C, T304C	A73G, 249del, A263G, 309C, 315C
KN9	F1a	T108C, G129A, A162G, T172C, T304C	A73G, C150T, 249del, A263G, 315C
KN17	F1a	G129A, T172C, T304C	A73G, C150T, 249DelA, A263G, 315C
KN29	F1a	G129A, A162G, T172C, T304C, A399G	A73G, 249DelA, A263G, 315C
KN36	F1a	C108T, G129A, A162G, T172C, T304C	A73G, 249DelA, A263G, 309CC, 315C
KN38	F1a	C108T, G129A, A162G, T172C, T304C, T311C	A73G, A93G, A95C, C150T, 249DelA, A263G, 315C
KN44	F1a	C108T, G129A, A162G, T172C, C214T, T304C	A73G, 249DelA, A263G, 315C
KN45	F1a	C108T, G129A, A162G, T172C, C214T, T304C	A73G, 249DelA, A263G, 315C
KN50	F1a	038DelA, G129A, T172C, T304C, C354T, G518A	A73G, T195C, 249DelA, A263G, 315C
KN55	F1a	C108T, G129A, A162G, T172C, C214T, T304C,	A73G, 249DelA, A263G, 315C
KN57	F1a	C108T, G129A, A162G, T172C, C214T, T304C, G518A, T519G	A73G, 249DelA, A263G, 315C
KN60	F1a	C108T, G129A, A162G, T172C, C214T, T304C	A73G, 249DelA, A263G, 315C
KN61	F1a	G129A, A162G, T172C, T304C,	A73G, 249DelA, A263G, 309C, 315C
KN64	F1a	G129A, T189C, T172C, T304C,	A73G, A263G, 315C
KN65	F1a	38delA, C108T, G129A, A162G, T172C, C214T, T304C	A73G, 249DelA, A263G, 315C
KN66	F1a	C108T, G129A, A162G, T172C, T304C	A73G, 249DelA, G251A, A263G, 309C, 315C
KN73	F1a	G129A, A162G, T172C, T304C, A399G	A73G, C150T, A210G, A235G, A263G, 309CC, 315C
KN74	F1a	G129A, A162G, T172C, T304C	A73G, 249DelA, A263G, 309CC, 315C
KN81	F1a	G129A, A162G, T172C, T243C, T304C, T311C	A73G, C150T, 249DelA, A263G, 309C, 315C
KN86	F1a	G129A, T172C, C294T, T304C	A73G, 249DelA, A263G, 315C
KN88	F1a	38delA, G129A, T172C, T304C, T311C, A497G, C514T, A515C, G516A, T519A	A73G, T152C, 249DelA, A263G, 315C
KN91	F1a	G129A, T172C, T304C	A73G, C150T, 249DelA, A263G, 315C
KN95	F1a	G129A, T172C, T304C	A73G, 249DelA, A263G, 309C, 315C
KN98	F1a	G129A, T172C, T304C, T311C	A73G, T152C, 249DelA, A263G, 315C

KN77	F1b	A182C, A183C, T189C, T304C	A73G, 249DelA, A263G, 309CC, 315C
KN90	F1b	A183C, T189C, T304C	A73G, 249DelA, A263G, 315C
KN79	G2a	C223T, A227G, C278T, T362C	A73G, A263G, 309C, 315C
KN21	M	C192T, C223T	A73G, C150T, 249DelA, A263G, 309C, 315C
KN28	M	C223T, A269G, T271C	A73G, C150T, A263G, 309C, 315C
KN32	M	C223T	A73G, T146C, T199C, A263G, 315C
KN72	M	T092C, C223T, A269G, T271C	A73G, C150T, C151T, A263G, 309C, 315C
KN82	M	C223T	A73G, T146C, T199C, A263G, 315C
KN24	M10	38delA, G129A, T172C, C174T, C223T, C234T, C290T, T311C, G518A, T519A	A73G, T125C, T127C, A263G, 309C, 315C
KN25	M10	C223T, C260G, T311C	A73G, A263G, 315C
KN34	M10	C167T, T311C	A73G, T146C, T199C, A263G, 315C
KN49	M10	T172C, C223T, T311C,	A73G, T146C, A263G, 315C
KN93	M10	A37G, C223T, T311C	A73G, A263G, 315C
KN6	M7b	38delA, G129A, C223T, T297C,	A73G, C150T, T199C, T204C, A263G, 309C, 315C
KN58	M7b	38delA, G129A, A183C, T189C, C223T, T297C	A73G, C150T, T199C, A263G, 315C
KN75	M7b	38delA, G129A, C223T, T297C, G518A, T519G, C520T	A73G, C150T, T159C, T199C, A263G, 309C, 315C
KN76	M7b	G129A, A182G, C223T, T297C, T356C	A73G, T146C, C150T, T199C, A234G, A263G, 309C, 315C
KN2	M7b1	38delA, G129A, C192T, C223T, T297C, T519A	A73G, C150T, T199C, A263G, 315C
KN4	M7b1	38delA, T86C, G129A, C192T, C223T, T297C, G518A	A73G, C150T, T199C, A263G, 315C
KN7	M7b1	38delA, T93C, G129A, C192T, C223T, T297C, G390A, G518A, T519G, C520T	A73G, C150T, T199C, A263G, 309C, 315C
KN12	M7b1	C114A, G129A, C192T, T297C	A73G, C150T, T199C, A263G, 315C
KN15	M7b1	G129A, C192T, C223T, T297C	A73G, C150T, T199C, A263G, 309C, 315C
KN30	M7b1	G129A, C192T, C223T, T297C	A73G, T146C, T199C, A263G, 315C
KN42	M7b1	G129A, C192T, C223T, T297C, G518A, T519G, C520T	A73G, T131C, C150T, T195C, A263G, 309C, 315C

	KN99	M7b1	G129, C192T, C223T, T297C	A73G, C150T, T199C, A263G, 309C, 315C
	KN16	M7c	C223T, C295T	A73G, T146C, T199C, A263G, 315C
	KN39	M7c	C223T, C291T, C295T	A73G, T146C, T199C, A263G, 315C
	KN54	M7c	C223T, C278T, C295T	A73G, T146C, T199C, A263G, 315C
	KN62	M7c	38delA, C223T, C295T, G518A, T519G,	A73G, T146C, T199C, A263G, 315C
	KN94	M7c	C223T, C261T, C295T, G518A	A73G, T146C, T199C, A263G, 309C, 315C
	KN97	M7c	T75C, C223T, A293T, C295T	A73G, T146C, T152C, T199C, A263G, 309CC, 315C
	KN84	M8a	38delA, C184T, C223T, T298C, G319A, A343G	A73G, A263G, 309C, 315C
	KN5	M9	T108C, G129A, T172C, C223T, C234T, C290T	A73G, G185A, A189G, A263G, 309C, 315C
	KN10	M9	C223T, C234T, T271C, C344T, T362C	A73G, A153G, A263G, 315C
	KN80	M9	G129A, T172C, C223T, C234T, C290T	A73G, A263G, 315C
	KN13	R	T93C, T157C, T304C	A73G, C150T, A263G, 309C, 315C
	KN19	R	38delA, T93C, T304C	A73G, T146C, T199C, A263G, 315C
	KN40	R	38delA, T288C, T304C, G390A, G518A, T519G	A73G, G143A, T146C, A183G, T204C, A263G, 309CC, 315C
	KN41	R	T124C, C148T, T304C, A309G, G390A	A73G, T146C, A263G, 309CC, 315C
	KN43	R	T304C, A309G, G390A, C511A, T512C, C514T, A515C, G516A, T519A	A73G, A183G, A263G, 315C
	KN47	R	C192T, T304C, A309G, G390A	A73G, T152C, A263G, 309C, 315C
	KN96	R	T288C, T304C, A309G, G390A	A73G, G143A, T146C, A183G, T204C, A263G, 315C
	KN22	Z	38delA, C185T, C223T, C260T, T298C, A302G, G518A, T519A	A73G, T195C, A263G, 309CC, 315C
<i>Khmer</i>	Kh 040	A	T93C, C223T, C234T, C290T, A293C, G319A	A73G, A235G, A263G, A297G, 315C
	Kh 026	B	T189C, C214A, C223T, G274A, T276A, C282A, T311C	A73G, T146C, A263G, 315C, C418A
	Kh 124	B	C04T, G042A, A183C, T189C, 193C, T209C, C223T, C291T, G390A	A73G, G207A, A263G, T310C, G316A, T318C, A326G
	Kh 052	B	C147T, A183C, C184A, T189C, C223T, A275G, G391A, G438A	A73G, A263G, 315C
	Kh 155	B	G129A, T140A, A149C, A183C, T189C, 193C, G213A, C218T, C223T, G274A	A73G, C150T, A263G, 309C, 315C

Kh 083	B	T189C, C223T, G274A, T311C	A73G, T146C, A263G, 315C
Kh 165	B	A183C, T189C, 193C, T209C, C223T, C291T, G390A,	A73G, G207A, A263G, 309C, 315C
Kh 089	B	A166G, A183C, T189C, 193C, C223T, A275G	A73G, A210G, A263G, 309C, 315C
Kh 037	B	T136A, A183C, T189C, 193C, T249C, T288C, A293G, T304C, C344T	A73G, T152C, A263G, 315C, G329A
Kh 056	B	T136A, A183C, T189C, 193C, T249C, T288C, A293G, T304C, C344T	A73G, T152C, A263G, 315C, G329A
Kh 151	B	T136A,A183C,T189C,G274A,T288C,T304C,T311C,G390A	A73G, T195C, A263G,315C,G329A
Kh 039	B	A054C, A70C, T136A, T140A, A183C, T189C, 193C, T249C, T288C, A293G, T304C, C344T	A73G,T152C,A263G, 315C, G329A
Kh 178	B	A183C, T189C, 193C, T249C, T288C, A293G, T304C, C344T	A73G, T152C, A263G,315C, G329A
Kh 093	B4	A183C, T189C, 193C, T217C, C234T, A309G, C354T	A73G, G207A, A263G, 309C, 315C
Kh 103	B4	T93C, C147T, C150T, A183C, C184A, T189C, T217C, A235G	A73G, A244G, A263G, 309C, 315C, 385DeIA
Kh 054	B4	A166C, A183C, T189C, T217C, A235G,	A73G, A263G, 309C, 315C, G207A
Kh 035	B4	T189C, T217C, G274A, T311C, A335G	A73G, T146C, C150T, C151T, A263G, 309C, 315C
Kh 072	B4	T93C, C147T, C149T, A183C, C184A, T189C, T217C, A235G	A73G, C150T, C151T, A263G, 309C, 315C
Kh 133	B4a	A182C, A183C, T189C, T217C, C261T	A73G, T146C, A153G, A263G, 309C, 315C
Kh 033	B4a	A182C, A183C, T189C, T217C, T261C	A73G, T146C, A263G, 309C, 315C
Kh 084	B4b	T136A, A183C, T189C, 193C, T249C, T288C, A293G, T304C, C344T	A73G, T152C, A263G, 315C, G329A
Kh 055	B4b	T136A, A183C, T189C, T217C, A235G	A73G, A263G, 309C, 315C
Kh 162	B5a	T140C, A183C, T189C, 193C, C266A	A73G, A210G, A263G, 309C, 315C
Kh 077	B5a	T140C, A183C, T189C, 193C, C266A, T311C	A73G, A210G, A263G, 309C, 315C
Kh 080	B5a	T140C, A183C, T189C, 193C, C266A	A73G, A210G, A263G, 309C, 315C
Kh 070	B5a	T140C, C179A, A181C, A182C, A183C, T189C, C261T, C266A, A335G	A73G, T152C, A210G, A263G, T310C, G316C, 318C
Kh 068	B5a	T140C, A183C, A182C, T189C, C266A	A73G, A210G, A263G, 315C
Kh 011	B5a	T140C, A182C, A183C, T189C, C266A	A73G, T152C, A210G, A263G, 309C, 315C

Kh 020	B5a	T140C, A182C, A183C, T189C,C266A	A73G, T152C, A210G, A263G, 309C, 315C
Kh 106	B5a	T140C, A182C, A183C, T189C, C266A	A73G, A210G, A263G, 309C, 315C
Kh 001	D4	C223T, C259T, G274A, T311C, T362C, T381C, G518A	T63C, C64T, G66A,A73G, T146C, A263G, 315C
Kh 116	D4	C104T, C223T, C287T, T362C, T469G	A73G, A200G, A263G, 309C, 315C
Kh004	D4	C223T, C259T, G274A, T311C, 362C	T63C, C64T, G66A, A73G, T146C, A263G, 315C
Kh 058	D4	C223T, T311C, T362C, C400T	A73G, T146C, A263G, 315C
Kh 018	D4	C223T, T362C	A73G, A153G, A183G, A263G, 315C
Kh 118	D4	T249C, T288C, C301T, T304C, T362C, G390A	A73G,A263G,315C, G329A
Kh125	D4	T249C, T288C, C301T, T304C, T362C, G390A	A73G, A263G, 309C, 315C
Kh 012	D4	T304C, T362C	A73G, A263G, 309C, 315C
Kh 166	D5	T136A, A175C, A182C, A183C, T189C, C223T, T362C	A73G, C150T, A263G, 309C, 315C
Kh 082	F1a	G129A, T172C, T304C, A309G	A73G, 249DelA, A263G, 315C
Kh 005	F1a	C108T, G129A, A162G, T172C, C239T, T304C, C327T	A73G, 249DelA, A263G, 309C, 315C
Kh 127	F1a	C108T, G129A, A162G, T172C, C245T, A284G, T304C	A73G, A214G, 249DelA, A263G, 309C, 315C
Kh 064	F1a	C108T, G129A, A162G, T172C, C256T, T304C	A73G, 249DelA, A263G, T293C, 309C, 315C
Kh 157	F1a	C108T, G129A, A162G, T172C, C259A, T304C	A73G, 249DelA, A263G, 315C
Kh 170	F1a	C108T, G129A, A162G, T172C, C266A, T304C	A73G, 249DelA, A263G, 315C
Kh 098	F1a	C108T, G129A, A162G, T172C, T304C	A73G, C150T, T195C, 249DelA, A263G, T310C, G316C
Kh 113	F1a	C108T, G129A, A162G, T172C, T304C	A73G, 249DelA, A263G, 315C
Kh 173	F1a	C108T, G129A, A162G, T172C, T304C, A335G	A73G, T217C, 249DelA, A263G, 315C
Kh 156	F1a	G129A, A162G, T172C, C214T, T304C	A73G, 249DelA, A263G, 315C
Kh 112	F1a	G129A, A162G, T172C, T304C	A73G, 249DelA, A263G, 315C
Kh 179	F1a	G129A, T172C, C295T, T304C	A73G, 249DelA, A263G, 315C
Kh 013	F1a	G129A, T172C, T304C	A73G, 249DelA, A263G, 315C

Kh 038	F1a	C108T, G129A, A162G, T172C, T304C	A73G, T195C, 249DelA, A263G, 309C, 315C
Kh 048	F1b	T136A, T140A, A183C, A184C, T189C, T304C, G390A	A73G, 249DelA, A263G, 309C, 315C
Kh 136	F1b	C186T, T189C, T209C, C242T, T304C	A73G, 249DelA, A263G, 309C, 315C
Kh 092	G2	C214A, C223T, C256T, C278T, C362T	A73G, T152C, A263G, 309C, 315C
Kh003	G2	C223T, T249C, C259T, C278T, T311C, T362C	A73G, A178G, A189G, A263G, 291DelA, 292-294delTTT, 315C
Kh 129	M	38delA, A219G, C223T, C290T, G518A	A73G, A263G, 315C, 385DelA
Kh 023	M	38delA, G129A, T209C, C223T, A272G, A322C	T58G, 56G, A73G, T152C, G225A, 249DelA, A263G, 315C, G316A
Kh 010	M	T086C, C223T, C234T, C278T, C294T, G518A	A73G, C150T, A263G, 315C
Kh 144	M	C108T, G129A, T172C, C174T, C223T, C234T, G244A, C290T	A73G, G185A, A263G, 309C, 315C
Kh 097	M	C108T, G129A, T172C, C223T, C234T, C290T	A73G, G185A, A189G, A263G, 309C, 315C
Kh 142	M	C223T, C239T, T263C, T381C	A73G, A263G, 309C, 315C
Kh 175	M	C223T, C278T, C294T	A73G, C150T, T152C, A263G, 315C
Kh 150	M	C223T, C290T, T304C	A73G, T159C, A263G, 309C, 315C
Kh 073	M	C223T, T271C	A73G, A244G, A263G, 309C, 315C
Kh 169	M	G129A, C223T	A73G, A214G, A263G, 309C, 315C
Kh 095	M	G129A, C223T, C278T, C294T, T304C	A73G, C150T, A263G, C273T, 315C, C418A
Kh 111	M	G129A, C223T, C278T, C294T, T304C	A73G, C150T, A263G, C273T, 315C
Kh 117	M	G129A, C223T, C290T	A73G, A263G, 315C
Kh 060	M	G145A, C192T, C223T, C291T, T304C	A73G, A210G, A263G, 315C
Kh 076	M	T124C, C193T, C223T	A73G, C150T, T195C, A263G, 315C, 337DelA
Kh 041	M	T124C, C223T, C234T, C261T, C290T	A73G, A263G, 309C, 315C, T318C
Kh 062	M	T86C, T129A, T209C, C223T, A272G	A73G, T152C, G225A, 249DelA, A263G, 315C, G316A
Kh 131	M	T93C, C148T, 183DelA, C223T	A73G, T152C, T195C, A263G, 309C, 315C
Kh 086	M10	C223T, T263C, G274A, T311C, A343G, T357C	A73G, A263G, T310C, G316C, T319G

	Kh 030	M10	C223T, T311C	A73G, A263G, C332T, 315C
	Kh 027	M10	T126C, T231C, T311C	A73G, T195C, A263G, 309C, 315C
	Kh 121	M10	C111A, T140C, T209C, C223T, T304C, T311C, T352C, C353T	A73G, A259G, A263G, 315C
	Kh 091	M7b	C192T, C223T, T297C	A73G, C150T, T199C, A263G, 309C, 315C
	Kh 009	M7b	G129A, C223T, T297C	A73G, C150T, T199C, T204C, A263G, 309C, 315C
	Kh 176	M7b	T102C, G129A, C223T, T297C	A73G, C150T, T159C, T199C, T204C, A263G, 315C
	Kh 022	M7b1	G129A, C192T, C223T, T297C	A73G, C150T, T199C, A263G, 315C
	Kh 024	M7b1	G129A, C192T, C223T, T297C	A73G, C150T, 194DelC, T199C, A263G, 315C
	Kh 071	M7b1	G129A, C192T, C223T, T297C	A73G, C150T, 194DelC, T199C, A263G, 315C
	Kh 139	M7b1	G129A, C192T, C223T, T297C	A73G, C150T, C182T, T199C, A263G, 315C
	Kh 065	M7b1	G129A, T189C, C192T, C223T, T297C	A73G, C150T, T199C, A263G, 309C, 315C
	Kh 049	M7b1	G129A, T189C, C192T, C223T, T297C, T217C, A235G	A73G, C150T, T199C, A263G, 309C, 315C
	Kh 159	M7b1	G129A, T192C, C223T, T297C	A73G, C150T, T199C, A263G, 315C
	Kh122	M7c	C223T, C295T	A73G, T146C, T199C, T204C, A263G, 309C, 315C
	Kh 168	M7c	T75C, C223T, C278T, A293T, C295T	A73G, T146C, T152C, T199C, A263G, 309C, 315C
	Kh 007	M7c	G145A, C291T, C295T, T304C	A73G, A210G, A263G, 315C
	Kh 107	R	A37G, G145A, A181G, C192T, C223T, C291T, T304C, G390A, G518A, T519A	A73G, A210G, A263G, 309C, 315C
	Kh025	R	T249C, T288C, T304C, C344T	A73G, T217C, A263G, 315C, G329A
	Kh 067	R9a	C260T, T298C, C355T, T362C	A73G, 249DelA, A263G, 309C, 315C
	Kh 135	U5a	T93C, G129A, C256T, T357C, A399G	A73G, T131C, C150T, T199C, A263G, 315C
	Kh 114	Z	C185T, C223T, C260T, T298C	A73G, T152C, 249DelA, A263G, 315C
	Kh 163	Z	C185T, C223T, C260T, T298C	A73G, T152C, A214G, 249DelA, A263G, 315C
<i>Mường</i>	M1	B	T93C, T178C, A182C, A183C, T189C, T271C, G274A	A73G, T146C, A263G, 315C
	M9	B	C168T, A182C, A183C, T189C, A194C	A73G, T146C, T217C, A263G, 309C, 315C



M10	B	C111T, T126C, T172C, A183C, T189C	A73G, G185A, A189G, T195C, A234G, A263G, 309C, 315C, A328G
M16	B	T93C, A183C, T189C,C426G	A73G, A263G, 315C
M35	B	A182C, A183C, T189C, C223T, C278T	A73G, T152C, A249G, A263G, 309C, 315C
M42	B	T93C, T178C, A182C, A183C, T189C, T271C	A73G, T146C, A263G, 315C
M46	B	A182C, A183C, T189C	A73G, A210G, A263G, T310C, G316C
M53	B	T154C, A182C, A183C, T189C	A73G, A263G, 309CC, 315C
M8	B4	A182C, A183C, T189C, T217C, G274A, T304C, G310A, 474delG	A73G, A263G, 309CC, 315C
M25	B4	A182C, A183C, T189C, T217C, A235G	A73G, T199C, A263G, 309C, 315C
M50	B4	A182C, A183C, T189C, 193CC, T217C, T231C, A235G, C291T, 469delT	A73G, A263G, 309CC, 315C
M56	B4	C147T, A183C, C184T, T189C, T217C, A235G	A73G, A263G, 315C
M2	B4a	A182C, A183C, T189C, G213A, T217C, C261T, C292T	A73G, A263G, C308T, 310delT
M86	B4a	A182C, A183C, T189C, T217C, A240G, C261T	A73G, A263G, 315C
M90	B4a	A182C, A183C, T189C, T217C, C261T	A73G, A263G, 309CC, 315C
M97	B4a	A182C, A183C, T189C, T217C, A240G, C261T	C43T, A73G, A263G, 315C
M41	B5	T92C, G129A, T140C, A182C, A183C, T189C, A194C, C197G	A73G, A210G, A263G, T310C
M12	B5a	G84C, C95T, A113C, A116C, A149C, A183C, T189C, 193C, C266A	A73G, A210G, A263G, 309C, 315C, A374G
M13	B5a	A91T, A113C, A122T, A182C, A183C, T189C, C266A	A73G, A210G, A263G, 309C, 315C
M14	B5a	A182C, A183C, T189C, C261T, C266A	A73G, T152C, A210G, A263G, 309C, 315C
M15	B5a	A129G, T140C, A182C, A183C, T189C, C266A, A399G	A73G, A210G, A263G, T310C, G316C
M30	B5a	A182C, A183C, T189C, C266A	A73G, A210G, A263G, 309C, 315C
M37	B5a	A182C, A183C, T189C, C226A, T304C	A73G, A210G, A263G, 315C
M64	B5a	T140C, A183C, T189C, A194C, T195C, C266A	A73G, A210G, A263G, 315C
M26	C	A182C, A183C, T189C, C223T, T298C, C327T, 469delT	A73G, T146C, A237G, 249delA, A263G, C303A

M32	C	C223T, A241C, G274A, T298C, C327T, G390A	A73G, T195C, 249delA, A263G, 309C, 315C
M71	C	C223T, G274A, T298C, C327T, G390A	A73G, T195C, 249delA, A263G, 309C, 315C
M76	C	T172C, C223T, T298C, C327T	A73G, 249delA, A263G, 315C
M80	C	38delA, T93C, T172C, C223T, T298C, T311C, C327T	A73G, T195C, 249delA, A263G, 315C, 368delA
M52	D4	C223T, T362C, A367C	A73G, C194T, A263G, 315C
M68	D4	C260T, T298C, T362C, A367C	T55C, T57C, A73G, T146C, C150T, T199C, A263G, 309CC, 315C
M70	D4	T304C, T362C	A73G, T195C, A263G, 309C, 315C
M74	D4	C111T, T126C, T140A, T172C, A183C, T189C, C223T, T362C	A73G, A263G, 309CC, 315C, 302C, 334delA
M5	F1a	G129A, T172C, T304C	T55C, A56C, T146C, C150T, T199C, A263G
M39	F1a	C108T, G129A, A126G, T172C, T304C	A73G, 249delA, A263G, 309CC, 315C
M43	F1a	T172C, T304C	A73G, C150T, T199C, T204C, A263G, 315C
M47	F1a	C108T, G129A, A162G, T172C, T304C	A73G, 249delA, A263G, 315C
M49	F1a	G129A, T172C, C295T, T304C	A73G, 249delA, A263G, 315C
M54	F1a	G129A, T172C, T304C, 469delT	A73G, A210G, A263G, 309CC, 315C
M60	F1a	G129A, T172C, T304C	A73G, 249delA, A263G, 315C
M65	F1a	G129A, A162G, T172C, T304C, A399G	A73G, T152C, 249delA, A263G, 309C, 315C
M67	F1a	T172C, T304C, C465T	A73G, 249delA, A263G, 309C, 315C
M77	F1a	38delA, G129A, T172C, C295T, T304C	A73G, 249delA, A263G, 315C
M81	F1a	38delA, G129A, T172C, T304C,	A73G, C198T, G207A, 249delA, A263G, 315C
M82	F1a	38delA, G129A, A162G, T172C, T304C	A73G, 249delA, A263G, 309C, 315C, 352delA
M84	F1a	38delA, G129A, T172C, C295T, T304C	A73G, 249delA, A263G, 309C, 315C

M85	F1a	C108T, G129A, A162G, T172C, T304C	A73G, C150T, T152C, T195C, 249delA, A263G, 315C
M89	F1a	G129A, T172C, T304C	A73G, 249delA, A263G, 309C, 315C
M96	F1a	C108T, G129A, A162G, T172C, T304C	A73G, 249delA, A263G, 309C, 315C
M22	F1b	A51G, T189C,A269G,C299T,A300G,T304C	A73G, C150T, T195C, A214G, 249delA, A263G, T310C
M72	F2	T304C	A73G, T195C, 249delA, A263G, 315C
M93	F2	T304C, C465T	A73G, 249delA, A263G, 309C, 315C
M63	F2a	T92A, C291T, T304C	A73G, A214G, 249delA, A263G, 309CC, 315C
M3	G2	C69T, T172C, C223T, A235G, C278T, C291A, T298C, T362C	A73G, T146C, C150T, T199C, A263, 309CC, 315C
M6	G2	38delA, C69T, T172C, C223T, C278T, C291A, T298C, T362C	T55C,A56C,T146C,C150T,T199C,A263G,309CC,315C,A351G
M18	G2	C69T, T172C, C223T, A233G, C278T, C291A, T298C, T362C, 474delG	A73G, T146C, C150T, T199C, A263G, 309C, 315C
M20	G2	C69T, T172C, C223T, C278T, C291A, T298C, T362C	A73G, T146C, C150T, T199C, A263G, 309C, 315C
M57	G2	T172C, C223T, C278T, C291A, T298C, T362C, A367C	G53A, T55C, A73G, T146C, C150T, A263G, 309C, 315C
M73	G2	C69T, T172C, C223T, T271C, G279A, C278T, C291A, T298C, T362C	A73G,C150T, T199C, A263G, 309CC, 315C
M51	M	G129A, T209C, C223T, A272G	A73G, T152C, A225G, 249delA, A263G, 315A, G316A
M55	M	C223T, A269G, T271C, 469delT	A73G, C150T, C151T, A263G, 309C, 315C
M69	M	C193T, C223T	A73G, C150T, T195C, A263G, 309CC, 315C, 337delA
M83	M	38delA, T86C, G129A, T209C, C223T, A272G	A73G, T152C, 249delA, A263G, 315C, G316A
M95	M	T86C, G129A, T209C, C223T, A272G	A73G, T152C, G225A, 249delA, A263G, 315C, G316A
M24	M10	C223T, C256T, A299G, T311C	A73G, A263G, 315C
M79	M10	A182C, A183C, T189C, C260G, T311C, G390A, A399G, C426G	A73G, A263G, 309CC, 315C
M11	M7	T172C, C223T, C291T, T311C	A73G, T146C, A263G, 309C, 315C

M38	M7	T311C, T356C	A73G, T146A, T199C, A263G, 315C
M40	M7	T356C	A73G, T146A, T199C, A263G, 309C, 315C
M45	M7	A163G, T172C, C223T, C291T, T311C	A73G, T146C, A263G, 315C
M78	M7	38delA, T311C, T356C	A73G, T146A, T199C, A263G, 315C, 368delA
M87	M7	T172C, C223T, C291T, T311C	A73G, T146C, A263G, 315C
M17	M7b	G129A, C223T, T297C	A73G, T146C, T199C, A263G, 309C, 315C
M31	M7b	C223T, T297C	A73G, C150T, T199C, T204C, A263G, 309C, 315C
M44	M7b	C223T, T297C, A299G	A73G, 249delA, A263G, 309C, 315C
M61	M7b	T297C, T304C, A367C, G390A, G496A	A73G, A183G, A263G, 309C, 315C
M4	M7b1	G129A, C192T, C223T, T297C	A73G, C150T, T199C, A263G, 309C, 315C
M21	M7b1	T86C, G129A, C192T, C223T, C297T	A73G, C150T, T152C, T199C, A263G, 315C
M27	M7b1	G129A, C192T, C223T, T297C	A73G, C150T, T199C, A263G, 309C, 315C
M29	M7b1	T86C, G129A, C192T, C223T, T297C	A73G, C150T, T152C, T199C, A263G, 315C
M33	M7b1	G129A, C192T, C223T, A289G, T297C	A73G, A263G, 309C, 315C
M36	M7b1	G129A, C192T, C223T, T297C	A73G, C150T, T199C, A263G, 309C, 315C
M59	M7b1	G129A, T189C, C192T, C223T, T297C, A367C, T304C	A73G, C150T, T199C, A263G, 309C, 315C
M66	M7b1	G129A, C192T, T297C	A73G, C150T, T199C, A263G, 315C
M91	M7b1	G129A, C192T, C223T, T297C, T324C	A73G, C150T, T199C, A263G, 309C, 315C, C332T
M92	M7b1	G129A, C192T, C223T, G274A, T297C, T324C	A73G, C150T, T152C, T199C, A263G, 315C, C332T
M98	M7c	C295T, G319A	A73G, T146C, T199C, A263G, 315C
M28	M8a	C184T, T189C, C223T, T298C, G319A	A73G, T195C, A263G, 309CC, 315C

M94	M8a	C184T, T189C, C223T, T298C, G319A	A73G, A263G, 309C, 315C
M34	N9a	T172C, T189C, C201T, C223T, C257A, C261T	A73G, C150T, T195C, A263G, 309C, 315C
M100	N9a	C223T, C257A, C261T, C292T	A73G, C150T, A263G, 309C, 315C
M7	R	T124C, C148T, A183G, T304C, A309G, G390A, 474delG	A73G, A263G, T310C
M19	R	C192T, T304C, A309G, G390A	A73G, T152C, A263G, 309C, 315C
M58	R	38delA, T124C, C148T, A183C, T304C, A309G, G390A	A73G, A263G, 309CC, 315C
M23	R9a	T93C, C111T, C192T, T249C, T298C, C355T, T362C, G390A	A73G, G207A, 249delA, A263G, 309C, 315C
M48	R9a	C111T, C192T, T249C, T298C, C355T, T362C, A367C, G390A	A73G, G207A, 249delA, A263G, 309C, 315C
M62	R9a	C260T, T298C, T362C, A367C	A73G, G207A, 249delA, A263G, 309CC, 315C
M75	R9a	C260T, T298C, C355T, T362C	A73G, T158A, G207A, 249delA, A263G, 309C, 315C
M88	R9a	T93C, C111T, C192T, T249C, T298C, C355T, T362C, G390A	A73G, G207A, 249delA, A263G, 309C, 315C
M99	R9a	T189C, C221T, G274A, C295T, T298C, G319A, C355T, T362C	A73G, A235G, 249delA, A263G, 309C, 315C

*Kết quả giải trình tự của các mẫu nghiên cứu được so sánh với trình tự chuẩn Cambridge đã sửa đổi, các đa hình thay thế, xóa, thêm nucleotid của từng mẫu được ghi lại. Phân nhóm đơn bội mtDNA các dân tộc Việt Nam dựa trên các SNP đặc trưng trên 2 vùng HV1 và HV2 của mtDNA. Cham: dân tộc Chăm, KN: dân tộc Kinh lấy mẫu ở TPHCM, Khin: dân tộc Kinh lấy mẫu ở Hà Nội, Kh: dân tộc Khmer, M: dân tộc Mường.*