

BỘ GIÁO DỤC  
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC  
VÀ CÔNG NGHỆ VIỆT NAM

**HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ**

---



**Trần Hữu Định**

**NGHIÊN CỨU HỆ GEN TY THỂ NGƯỜI VIỆT NAM THUỘC  
BA DÂN TỘC CÔNG, HOA VÀ SÁN DÌU**

**LUẬN VĂN THẠC SĨ SINH HỌC**

*Hà Nội - 2024*

BỘ GIÁO DỤC  
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC  
VÀ CÔNG NGHỆ VIỆT NAM

**HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ**



**Trần Hữu Định**

**NGHIÊN CỨU HỆ GEN TY THỂ NGƯỜI VIỆT NAM THUỘC  
BA DÂN TỘC CÔNG, HOA VÀ SÁN DÌU**

**LUẬN VĂN THẠC SĨ SINH HỌC**

**Ngành:** Sinh học thực nghiệm

**Mã số:** 8 42 01 14

NGƯỜI HƯỚNG DẪN KHOA HỌC:  
PGS. TS. Nguyễn Thủy Dương

*Hà Nội - 2024*

## LỜI CAM ĐOAN

Tôi là Trần Hữu Định, học viên cao học khoá 2022B, khoa Công nghệ sinh học, Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, xin cam đoan:

1. Đây là luận án do bản thân tôi thực hiện dựa trên những tài liệu và số liệu do chính bản thân tôi tìm hiểu và nghiên cứu, dưới sự hướng dẫn của PGS. TS. Nguyễn Thuỳ Dương.

2. Những kết quả thu được trong luận văn là mới, trung thực và chưa từng được công bố trong bất kỳ một công trình nào khác.

**Tác giả luận văn ký và ghi rõ họ tên**



**Trần Hữu Định**

## LỜI CẢM ƠN

Đầu tiên, tôi xin được bày tỏ sự kính trọng và biết ơn sâu sắc tới PGS. TS. Nguyễn Thuỳ Dương, Trưởng phòng Hệ gen học người, Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã tận tình hướng dẫn, định hướng, tạo mọi điều kiện thuận lợi nhất, cũng như động viên tôi trong suốt quá trình học tập, nghiên cứu và thực hiện luận văn thạc sĩ tại phòng Hệ gen học người, Viện Nghiên cứu hệ gen.

Luận văn được thực hiện dưới sự hỗ trợ kinh phí của đề tài nghiên cứu độc lập cấp Quốc gia “Xây dựng cơ sở dữ liệu hệ gen biến thể ty thể và nhiễm sắc thể Y của một số dân tộc người Việt Nam” chủ nhiệm bởi PGS. TS. Nguyễn Thuỳ Dương trong khoảng thời gian 2019 - 2024.

Tôi cũng xin được cảm ơn sự các cán bộ nhân viên cũng như các sinh viên thuộc phòng Hệ gen học người, Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã tận tình hỗ trợ tôi trong suốt thời gian thực hiện luận văn.

Tôi xin cảm ơn Ban Lãnh Đạo, phòng Đào tạo, các phòng chức năng của Học viện Khoa học và Công nghệ, cũng như các thầy giáo, cô giáo đã tận tình giảng dạy và tạo mọi điều kiện thuận lợi nhất để tôi có thể hoàn thành được luận văn này.

Đặc biệt học viên xin được bày tỏ sự biết ơn với bố mẹ, những người thân trong gia đình và bạn bè đồng nghiệp đã luôn bên cạnh khích lệ và động viên trong suốt quá trình học tập, nghiên cứu và hoàn thành luận văn thạc sĩ.

**Tác giả luận văn ký và ghi rõ họ tên**



**Trần Hữu Định**

## MỤC LỤC

LỜI CAM ĐOAN .....	i
LỜI CẢM ƠN.....	ii
MỤC LỤC .....	iii
DANH MỤC BẢNG .....	vi
DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ.....	vii
MỞ ĐẦU .....	1
CHƯƠNG 1. TỔNG QUAN NGHIÊN CỨU .....	3
1.1. TY THỂ VÀ HỆ GEN TY THỂ.....	3
1.1.1. Cấu trúc ty thể.....	3
1.1.2. Chức năng của ty thể .....	4
1.1.3. Hệ gen ty thể.....	5
1.1.3.1. Vùng điều khiển.....	5
1.1.3.2. Vùng mã hoá.....	6
1.2. NGŨ HỆ HÁN - TẠNG TẠI VIỆT NAM.....	7
1.3. TÌNH HÌNH NGHIÊN CỨU HỆ GEN TY THỂ .....	8
1.3.1. Thế giới.....	8
1.3.2. Việt Nam.....	11
CHƯƠNG 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU.....	13
2.1. ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU .....	13
2.1.1. Đối tượng nghiên cứu .....	13
2.1.2. Phạm vi nghiên cứu .....	13
2.1.3. Đạo đức trong nghiên cứu .....	13
2.2. PHƯƠNG PHÁP NGHIÊN CỨU .....	13
2.2.1. Thu thập và tách chiết DNA tổng số từ các mẫu nghiên cứu.....	13
2.2.3. Thiết lập và làm giàu thư viện .....	15
2.2.3.1. Chuẩn bị DNA đầu vào.....	15
2.2.3.2. Tạo đầu bằng cho các đoạn DNA .....	15
2.2.3.3. Gắn adapter.....	16
2.2.3.4. Làm đầy adapter (fill-in adapter).....	17
2.2.3.5. Indexing PCR.....	17
2.2.3.6. Gộp chung các sản phẩm indexing PCR.....	18

2.2.3.7. Làm giàu thư viện bằng phản ứng lai - bắt giữ.....	18
2.2.3.8. Khuếch đại thư viện DNA .....	19
2.2.4. Giải trình tự hệ gen biến thể ty thể và tiền xử lý dữ liệu .....	19
2.2.5. Phân tích hệ gen biến thể ty thể.....	19
CHƯƠNG 3. KẾT QUẢ VÀ THẢO LUẬN.....	21
3.1. KẾT QUẢ TÁCH CHIẾT DNA TỔNG SỐ .....	21
3.2. KẾT QUẢ THIẾT LẬP VÀ GIẢI TRÌNH TỰ THƯ VIỆN HỆ GEN TY THỂ ..	21
3.2.1. Kết quả cắt phân đoạn DNA tổng số .....	21
3.2.2. Kết quả làm đầy adapter .....	22
3.2.3. Kết quả indexing PCR .....	23
3.2.4. Kết quả đánh giá chất lượng thư viện DNA sau tinh sạch bằng DNA chip .....	23
3.2.5. Kết quả giải trình tự thư viện hệ gen ty thể .....	24
3.3. ĐẶC ĐIỂM DI TRUYỀN VÀ CÁC BIẾN THỂ TRÊN HỆ GEN TY THỂ DÂN TỘC CÔNG, HOA VÀ SÁN DÌU .....	27
3.3.1. Đặc điểm chung .....	27
3.3.2. Biến thể trên vùng không mã hoá.....	30
3.3.3. Biến thể trên vùng mã hoá.....	31
3.3.2.1. Biến thể trên họ gen NADH dehydrogenase subunit (ND).....	31
3.3.2.2. Biến thể trên họ gen Cytochrome c oxidase subunit (COX).....	35
3.5.2.3. Biến thể trên họ gen ATP Synthase Membrane Subunit (ATP).....	35
3.3.2.4. Biến thể trên gen Cytochrome b (CYTB).....	36
3.3.2.5. Biến thể trên gen mã hoá rRNA .....	36
3.4. SO SÁNH THỐNG KÊ BIẾN THỂ.....	37
3.4.1. Các dân tộc Hán - Tạng Việt Nam.....	37
3.4.2. Các dân tộc thuộc ngữ hệ khác tại Việt Nam .....	39
3.4.3. Các dân tộc Hán - Tạng ở các nước lân cận .....	42
KẾT LUẬN VÀ KIẾN NGHỊ .....	46
4.1. KẾT LUẬN .....	46
4.2. KIẾN NGHỊ.....	47
DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ.....	48
DANH MỤC TÀI LIỆU THAM KHẢO.....	49
PHỤ LỤC .....	PL1

## DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ CÁI VIẾT TẮT

Chữ viết tắt	Tiếng Anh	Tiếng Việt
95% CI	95% Confident Interval	Khoảng tin cậy 95%
ADP	Adenosine diphosphate	Adenosine diphosphate
ATP	Adenosine triphosphate	Adenosine triphosphate
BAM	Binary alignment and map	Dữ liệu giải trình tự bộ gen nén dưới dạng nhị phân
CJ	Cristae junctions	Mối nối giữa các mào
DNA	Deoxyribonucleic acid	Axit deoxyribonucleic
H	Haplotype diversity	Chỉ số đa dạng haplotype
HSP	Heavy strand promoter	Vùng khởi động chuỗi nặng
HVS	Hypervariable region	Vùng siêu biến
IBM	Inner boundary membrane	Màng ranh rới bên trong
IMM	Inner mitochondrial membrane	Màng trong ty thể
IMS	Intermembrane space	Vùng giữa hai màng
LHON	Leber Hereditary Optic Neuropathy	Liệt thần kinh thị giác di truyền Leber
LSP	Light strand promoter	Vùng khởi động chuỗi nhẹ
mtDNA	Mitochondrial DNA	Hệ gen ty thể
OMM	Outer mitochondrial membrane	Màng ngoài ty thể
OR	Odds ratio	Chỉ số nguy cơ
OXPHOS	Oxidative phosphorylation	Phosphoryl hoá oxy hoá
PCR	Polymerase chain reaction	Phản ứng chuỗi polymerase
RFLP	Restriction fragment length polymorphism	Đa hình chiều dài đoạn cắt giới hạn
SAM	Sequence alignment map	File đóng hàng trình tự
SPRI	Solid Phase Reversible Immobilization	Phương pháp cố định thuận nghịch pha rắn
TAS	Termination-associated sequence	Đoạn trình tự kết thúc
rRNA	Ribosomal ribonucleic acid	Axit ribonucleic riboxom
tRNA	Transfer ribonucleic acid	Axit ribonucleic vận chuyển

## **DANH MỤC BẢNG**

Bảng 2.1. Trình tự nucleotide của các adapter sử dụng trong quá trình gắn adapter ...	16
Bảng 3.1. Số lượng biến thể xuất hiện trên dân tộc Công, Hoa và Sán Dìu.....	29
Bảng 3.2. Dữ liệu biến thể mtDNA được sử dụng trong các phân tích so sánh.....	37



## DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

Hình 1.1. Cấu trúc của bào quan ty thể.....	3
Hình 1.2. Chuỗi truyền điện tử trong con đường phosphoryl hoá oxy hoá (OXPHOS)	4
Hình 1.3. Cấu trúc hệ gen ty thể (mtDNA) ở người [8] .....	5
Hình 3.1. Điện di đồ DNA tổng số của 69 cá thể thuộc 3 dân tộc Công, Hoa và Sán Dìu trên gel agarose 0,8%.....	21
Hình 3.2. Điện di đồ DNA tổng số của 69 mẫu sau khi cắt phân đoạn trên gel agarose 2%.....	22
Hình 3.3. Kết quả làm đầy adapter. ....	22
Hình 3.4. Điện di đồ sản phẩm indexing PCR trên gel agarose 2%.....	23
Hình 3.5. Kết quả đánh giá chất lượng thư viện bằng DNA chip.....	24
Hình 3.6. Kết quả đánh giá chất lượng trình tự thô của 69 mẫu.....	25
Hình 3.7. Kết quả lắp ráp trình dưới dạng BAM sử dụng phần mềm IGV. ....	27
Hình 3.8. Số lượng biến thể trên hệ gen ty thể của dân tộc Công, Hoa và Sán Dìu.....	28
Hình 3.9. Sự phân bố các biến thể phổ biến trên hệ gen ty thể của ba dân tộc Công, Hoa và Sán Dìu.....	30
Hình 3.10. Số lượng các biến thể được tìm thấy trên họ gen và gen của hệ gen ty thể trong nghiên cứu. ....	31
Hình 3.11. Các biến thể phổ biến trên họ gen <i>ND</i> có liên quan tới bệnh ty thể. ....	32
Hình 3.12. So sánh thống kê một số biến thể nổi bật ở ba dân tộc Công, Hoa và Sán Dìu với các dân tộc khác thuộc ngữ hệ Hán - Tạng ở Việt Nam.....	38
Hình 3.13. Số lượng biến thể khác biệt so với các dân tộc Hán - Tạng. ....	39
Hình 3.14. Số lượng biến thể khác biệt của các dân tộc Hán - Tạng so với các dân tộc thuộc các ngữ hệ khác.....	40
Hình 3.15. Một số biến thể khác biệt nổi bật giữa các nhóm dân tộc thuộc năm ngữ hệ tại Việt Nam.....	41
Hình 3.16. Số lượng biến thể khác biệt giữa các dân tộc Hán - Tạng của Việt Nam so với các nước lân cận .....	43
Hình 3.17. Một số biến thể khác biệt nổi bật giữa nhóm dân tộc Hán - Tạng Việt Nam so với các nước lân cận.....	44

## MỞ ĐẦU

Ty thể là một bào quan quan trọng ở sinh vật nhân thực với chức năng chính là sản sinh năng lượng cho tế bào dựa vào con đường phosphoryl hoá oxy hoá các chất dinh dưỡng để tạo nên các phân tử adenosine triphosphate (ATP). Bào quan ty thể sở hữu vật chất di truyền riêng (hệ gen ty thể - mtDNA), tách biệt với hệ gen nhân dưới dạng một phân tử DNA kép mạch vòng và hoàn toàn không có vùng intron. Một số đặc điểm quan trọng của mtDNA có thể kể đến như di truyền theo dòng mẹ, tỷ lệ đột biến cao, không tái tổ hợp và khả năng sao chép cao.

Hiện nay, việc sử dụng trình tự hệ gen biến thể ty thể để đánh giá cấu trúc di truyền giữa các quần thể khác nhau cũng như nghiên cứu bệnh học ty thể vẫn còn đang là mối quan tâm của các nhóm nghiên cứu trong nước và trên thế giới. Trong đó, vùng không mã hoá của ty thể với số lượng đa hình nhiều nhất trên mtDNA đã được nghiên cứu chuyên sâu về nhân chủng học tiến hoá, di cư của loài người và nghiên cứu về đa hình cá thể phục vụ cho khoa học giám định hình sự và pháp y. Các đa hình trên vùng mã hoá của mtDNA lại được chú ý hơn trong các nghiên cứu về nguyên nhân hoặc mối tương quan với bệnh ty thể như bệnh liệt thần kinh thị giác di truyền Leber (LHON), hội chứng Leigh,... hoặc với các kiểu hình như thích ứng với môi trường cao nguyên, trường thọ, chịu lạnh,... Nhìn chung, các nghiên cứu trên hệ gen biến thể ty thể trước đây đã chỉ ra rằng, mỗi quần thể người khác nhau lại có những đặc điểm di truyền đặc trưng, và không thể áp dụng cơ sở dữ liệu của quần thể này cho một quần thể khác.

Ở Việt Nam, với sự đa dạng của 54 dân tộc, các nghiên cứu về lĩnh vực trên lại càng phải được chú trọng hơn nữa. Trong hai thập kỉ trở lại đây, các nghiên cứu về hệ gen ty thể đã bước đầu giải thích được sự đa dạng di truyền của các dân tộc thuộc 5 ngữ hệ lớn ở Việt Nam. Trong đó, ngữ hệ Hán - Tạng đã được nghiên cứu chuyên sâu với 5 dân tộc đại diện thuộc nhánh ngôn ngữ Tạng - Miến. Tuy nhiên, hệ gen ty thể của các dân tộc thuộc nhánh ngôn ngữ Hán của ngữ hệ này lại chưa được nghiên cứu toàn bộ hệ gen ty thể trước đây. Chính vì vậy, chúng tôi tiến hành nghiên cứu “Nghiên cứu hệ gen ty thể người Việt Nam thuộc ba dân tộc Cống, Hoa và Sán Dìu” nhằm cung cấp thông tin di truyền hệ gen ty thể của hai dân tộc Hoa và Sán Dìu thuộc nhánh ngôn ngữ Hán cũng như bổ sung thêm dân tộc Cống thuộc nhánh Tạng - Miến vào bức tranh di truyền tổng thể của ngữ hệ Hán - Tạng. Nghiên cứu này sẽ giúp bổ sung và hoàn thiện dữ liệu về hệ gen biến thể ty thể của người Việt Nam, từ đó làm cơ sở tham chiếu cho các nghiên cứu về hệ gen ty thể sau này.

**Mục tiêu đề tài**

1. Giải trình tự hệ gen ty thể của các cá thể thuộc ba nhóm dân tộc Việt Nam gồm Cống, Hoa và Sán Dìu;
2. Phân tích được những biến thể của ba nhóm dân tộc này Cống Hoa và Sán Dìu.

**Nội dung nghiên cứu**

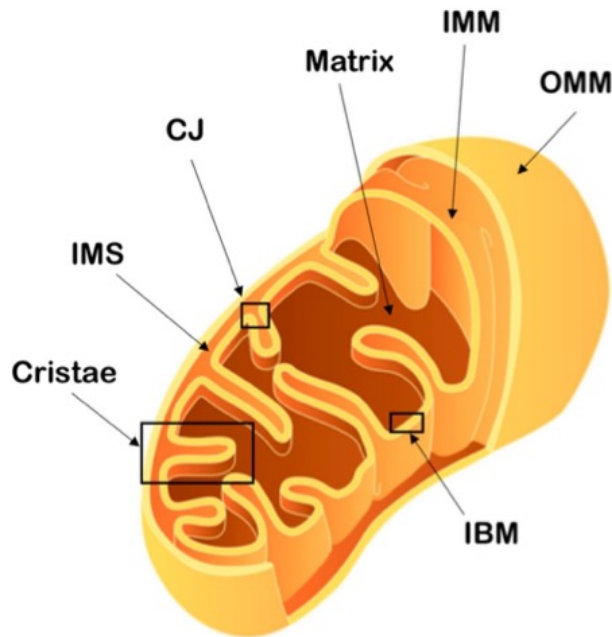
1. Thu thập mẫu máu và tách chiết DNA tổng số của các cá thể thuộc ba dân tộc Cống, Hoa và Sán Dìu;
2. Giải trình tự hệ gen ty thể của cá thể thuộc ba dân tộc Cống, Hoa và Sán Dìu;
3. Phân tích các biến thể trên hệ gen ty thể các dân tộc Cống, Hoa và Sán Dìu;
4. So sánh các biến thể trên hệ gen ty thể của các dân tộc trong nghiên cứu với các dân tộc khác ở Việt Nam và các nước lân cận.

## CHƯƠNG 1. TỔNG QUAN NGHIÊN CỨU

### 1.1. TY THỂ VÀ HỆ GEN TY THỂ

#### 1.1.1. Cấu trúc ty thể

Ty thể là một bào quan ở các tế bào nhân thực trong đó, mỗi tế bào người có đến hàng trăm ty thể với khoảng 2-10 bản sao của hệ gen ty thể tồn tại. Ty thể cấu tạo bởi lớp kép phospholipid với màng trong (OMM) và màng ngoài (IMM) chia ty thể thành hai vùng riêng biệt là chất nền (matrix) và vùng giữa hai màng (IMS) [1] (Hình 1.1). Hai màng này có sự khác biệt về hình dáng, thành phần lipid, trong đó, màng ngoài có thành phần lipid giống với màng tế bào của sinh vật nhân thực, còn màng trong lại giống với màng tế bào của vi khuẩn [2]. Màng trong gấp khúc tạo thành các mào (cristae) bên trong chất nền. Trên các mào này có chứa rất nhiều các protein, bộ máy phosphoryl hoá oxy hoá (OXPHOS) và nguyên nhân cho việc gấp khúc này là làm tăng diện tích tiếp xúc cho quá trình sản sinh năng lượng. Phần của màng trong không nhô ra trong chất nền mà chạy song song với ngoài được gọi là màng ranh giới bên trong (IBM). Mào ty thể và màng ranh giới này được kết nối thông qua các cấu trúc hình ống hẹp, được gọi là các mối nối giữa các mào (cristae junctions - CJ).



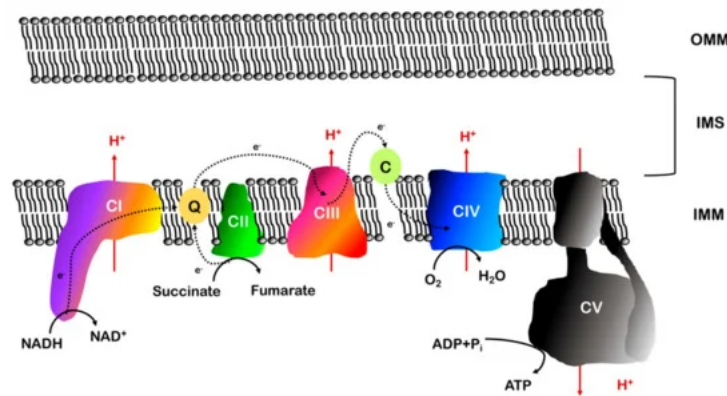
**Hình 1.1. Cấu trúc của bào quan ty thể.** OMM – Outer mitochondrial membrane: Màng ngoài ty thể, IMM – Inner mitochondrial membrane: Màng trong ty thể, IBM – Inner boundary membrane: màng ranh giới bên trong, CJ – cristae junctions: mối nối giữa các mào, IMS – Intermembrane space: vùng giữa hai màng, mào và chất nền ty thể [3].

Ngoài ra, màng trong và màng ngoài ty thể cũng có khác biệt lớn về tính thấm. Cụ thể, màng ngoài cho phép các ion và các phân tử nhỏ di chuyển qua các kênh anion

dựa trên điện thế, nhưng ở màng trong, chỉ  $H_2O$ ,  $O_2$  và  $CO_2$  là có thể tự do đi qua màng [4]. Hiện tượng này được gọi là tính thấm có chọn lọc, nó cho phép sự hình thành của gradient nồng độ giữa bên trong và bên ngoài màng, đây là nguyên lý của quá trình sản sinh ATP và sự điều chỉnh nghiêm ngặt nồng độ của các ion khác như canxi được sử dụng chủ yếu trong tín hiệu tế bào [5].

### 1.1.2. Chức năng của ty thể

Ty thể được đặc trưng bởi chức năng sản sinh ra năng lượng ATP mà tế bào có thể hấp thụ từ các chất dinh dưỡng thông qua một chuỗi truyền điện tử ở quá trình phosphoryl hoá oxy hoá (OXPHOS). Hệ thống này tạo năng lượng thông qua sự phối hợp hoạt động của 5 phức hệ gắn màng để oxy hoá các phân tử NADH và  $FADH_2$  để phosphoryl hoá adenosine diphosphate (ADP) thành ATP.



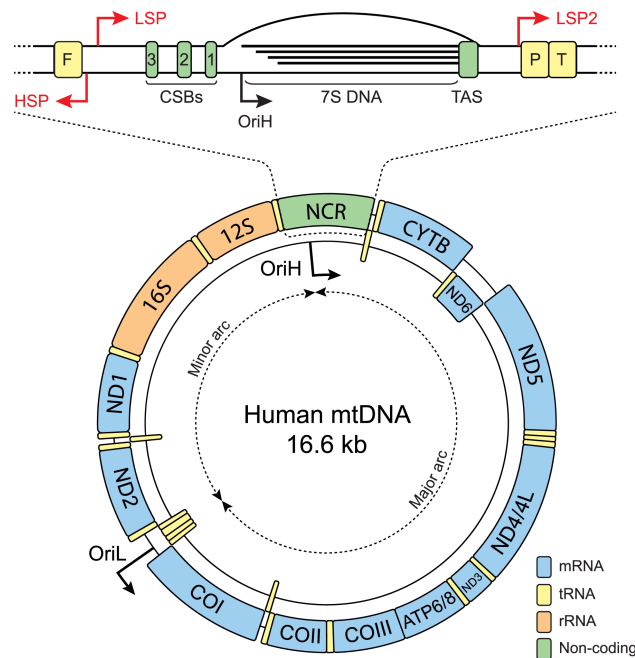
**Hình 1.2. Chuỗi truyền điện tử trong con đường phosphoryl hoá oxy hoá (OXPHOS).** CI - CV: Phức hệ I - IV; Q: Coenzyme Q; C: Cytochrome c; OMM – Outer mitochondrial membrane: Màng ngoài ty thể, IMM – Inner mitochondrial membrane: Màng trong ty thể; IMS – Intermembrane space: vùng giữa hai màng [3].

Hệ thống này bao gồm 4 enzyme trong chuỗi truyền điện tử (phức hợp I-IV); 2 chất mang điện tử di động (coenzyme Q - CoQ, và cytochrome c) (Hình 1.2). Các phân tử NADH và  $FADH_2$  được tạo ra từ các chu trình Krebs và  $\beta$ -oxy hoá axit béo sẽ được sử dụng trong quá trình OXPHOS để tạo ra các phân tử ATP. NADH gắn vào phức hợp I và  $FADH_2$  gắn vào phức hợp II và cả 2 đều bị oxy hoá thành  $NAD^+$  và  $FAD^+$ . Điện tử từ hai phân tử này sẽ được chuyển lần lượt vào phức hợp I và II, và sau đó được chuyển sang coenzyme Q. Các điện tử này sẽ được tiếp tục chuyển đến phức hợp III, cytochrome c và cuối cùng là phức hợp IV, ở phức hợp này các điện tử kết hợp với  $O_2$  và tạo thành phân tử  $H_2O$ . Trong toàn bộ quá trình trên, khi các điện tử được truyền từ trung tâm khử này sang trung tâm khử khác, các proton hay  $H^+$  sẽ được bơm từ bên trong chất nền ra vùng giữa hai màng thông qua các phức hệ CI, CIII và CIV. Từ đó tạo nên một gradient giữa bên trong và bên ngoài màng trong ty thể và được sử dụng bởi phức hợp V để bơm proton ngược lại vào trong chất nền và hỗ trợ tạo nên ATP. Bên chức năng này, ty thể

còn tham gia vào nhiều quá trình trong tế bào, bao gồm chuyển hoá trong tế bào, sản sinh nhiệt, tổng hợp và oxy hoá lipid, hô hấp tế bào, duy trì cân bằng nội môi [6].

### 1.1.3. Hệ gen ty thể

Bào quan ty thể sở hữu vật chất di truyền riêng (hệ gen ty thể - mtDNA), tách biệt với hệ gen nhân dưới dạng một phân tử DNA kép mạch vòng và không sở hữu vùng intron (Hình 1.3). Một số đặc điểm quan trọng của mtDNA có thể kể đến như di truyền theo dòng mẹ, tỷ lệ đột biến cao, không tái tổ hợp và khả năng sao chép cao [7]. Ở người, hệ gen ty thể có kích thước 16.596 bp và bao gồm 2 vùng chính là vùng không mã hoá (vùng điều khiển) dài khoảng 1.100 bp và vùng mã hoá khoảng 15.000 bp. mtDNA bao gồm hai mạch là chuỗi nặng giàu các nucleotide G và A (purin) và chuỗi nhẹ giàu các nu C và T (pyrimidine). Trong đó, phần lớn thông tin di truyền của mtDNA được mã hoá trên chuỗi nặng và chỉ có gen *ND6* và 8 tRNA được mã hoá trên chuỗi nhẹ.



Hình 1.3. Cấu trúc hệ gen ty thể (mtDNA) ở người [8]

#### 1.1.3.1. Vùng điều khiển

Vùng không mã hoá hay vùng điều khiển đóng vai trò là trung tâm điều hoà cho quá trình tái bản và phiên mã của mtDNA nhờ chứa điểm khởi đầu cho quá trình tái bản và các vùng khởi động cho quá trình phiên mã [9]. Cụ thể, đoạn này chứa hai vùng khởi động cho hai mạch của ty thể, bao gồm vùng khởi động chuỗi nặng (HSP) và vùng khởi động chuỗi nhẹ (LSP), và một điểm khởi đầu cho quá trình tái bản DNA ( $O_H$ ) cho chuỗi nặng ở ngay phía dưới HSP [8]. Tuy nhiên, không phải tất cả các quá trình tái bản bắt đầu ở  $O_H$  đều được hoàn thiện, mà thay vào đó, khoảng 95% các quá trình này được kết thúc ngay sau khi kéo dài được 650 nucleotide ở đoạn trình tự kết thúc (TAS). Những đoạn ngắn này được gọi là 7S DNA [10]. Những trình tự DNA ngắn hình thành bằng

cách này sẽ tiếp tục gắn với chuỗi nhẹ và khiến cho chuỗi nặng có nguồn gốc từ mẹ bị chiếm chỗ, và cuối cùng hình thành nên một cấu trúc loop gồm 3 mạch được gọi là D-loop. Vai trò chức năng của vùng D-loop cũng như cách mà quá trình tái bản kết thúc ở vùng TAS vẫn chưa được nghiên cứu cụ thể [11]. Bên cạnh điểm khởi đầu  $O_H$  ở vùng điều khiển, ty thể còn sở hữu một điểm bắt đầu tái bản cho chuỗi nhẹ  $O_L$  nằm ở vùng mã hoá tRNA, cách  $O_H$  khoảng 11.000 bp. Toàn bộ vùng điều khiển này có thể được chia ra thành ba vùng siêu biến khác nhau (Hypervariable region I - HVS-I, HVS-II và HVS-III). Nhiều nghiên cứu trước đây đã báo cáo rằng các biến thể được tìm thấy ở vùng này là đặc trưng cho các quần thể khác nhau hoặc có liên quan đến một số bệnh ty thể như buồng trứng đa nang, bệnh tăng nhãn áp góc mở nguyên phát,...

### 1.1.3.2. Vùng mã hoá

Vùng mã hoá bao gồm 37 gen chức năng, gồm 2 gen mã hoá cho rRNA, 22 gen mã hoá cho tRNA và 13 gen mã hoá cho các protein khác nhau (Hình 1.3). Những protein này tham gia cấu thành nên các phức hợp I, III, IV và V trong quá trình phosphoryl hoá oxy hoá có chức năng vận chuyển điện tử từ các chất dinh dưỡng thành dạng năng lượng mà tế bào có thể hấp thụ là ATP.

Phức hợp I là phức hợp lớn nhất trong số 4 phức hợp, có dạng hình chữ L và được tạo nên bởi 44 polypeptide khác nhau, được cấu trúc thành 3 domain cấu trúc: module P nằm trên màng trong ty thể; module N và Q nhô ra chất nền ty thể. Phần module P có chứa 7 tiểu phần protein mã hoá bởi 7 gen trên ty thể (NADH dehydrogenase subunit - *ND1*, *ND2*, *ND3*, *ND4*, *ND4L*, *ND5* và *ND6*). Trong đó, protein ND1 hình thành trung tâm khử cho ubiquinone; ND2, ND4 và ND5 được cho là đóng vai trò trong quá trình bơm proton. Cho đến hiện tại, các đột biến trên các gen ty thể mã hoá cho phức hệ I đã được báo cáo ở tương đối nhiều bệnh với các kiểu hình khác trong, bao gồm LHON, MELAS, hội chứng Leigh, ....

Phức hệ III là thành phần trung tâm của chuỗi truyền điện tử với chức năng xúc tác quá trình truyền điện tử từ ubiquinol (Coenzyme Q đã bị khử) đến cytochrome c và sử dụng năng lượng để vận chuyển proton từ bên trong màng trong ty thể ra bên ngoài. Phức hệ này bao gồm 11 tiểu phần, trong đó chỉ có một tiểu phần duy nhất mã hoá bởi gen *cytochrome b* (*CYTB*) trên ty thể. Các bệnh ty thể liên quan đến sự suy giảm của phức hệ III thường không phổ biến và phần lớn đều đến từ các đột biến trên gen *CYTB* [3]. Cụ thể, các biến thể trên gen này được phát hiện có liên quan tới các bệnh như vô sinh nam, hội chứng MELAS,...

Phức hệ IV là enzyme thứ ba và cũng là enzyme cuối cùng trong chuỗi truyền điện tử. Phức hệ này tiếp nhận điện tử từ cytochrome c đã bị khử và chuyển chúng cho oxy để tạo thành nước. Nó cấu tạo bởi 13 tiểu phần, trong đó 3 tiểu phần được mã hoá

bởi 3 gen trong ty thể là *COX1*, *COX2*, và *COX3* [3]. Protein COX1 chứa 3 phần gồm cytochrome a<sub>3</sub> và CuB, hình thành nên trung tâm hai nhân liên kết với oxy, và cytochrome a [3]. Protein COX2 liên kết chặt chẽ với trung tâm Cu<sub>A</sub> của phức hệ IV và protein COX3 không tham gia vào chức năng xúc tác. Các bệnh liên quan đến sự suy giảm chức năng của phức hệ IV chỉ phổ biến sau bệnh liên quan đến phức hệ I [12]. Những bệnh phổ biến nhất có liên quan tới sự suy giảm chức năng của phức hệ IV bao gồm các bệnh ảnh hưởng tới cơ xương, và các bệnh có tính hệ thống như hội chứng Leigh. Tuy nhiên, các đột biến trên các gen mã hoá cho phức hệ IV thường ít được tìm thấy ở các bệnh nhân hơn so với các phức hệ khác.

Phức hệ V là một enzyme xúc tác cho phản ứng tổng hợp ATP từ ADP và phosphate. Ở người, phức hệ này cấu tạo bởi 29 protein, trong đó chỉ có 2 protein (ATP6 và ATP8) là mã hoá bởi hai gen cùng tên trên mtDNA [13]. Cho đến nay, rất ít các biến thể có liên quan tới bệnh được tìm thấy trên các gen mã hoá cho phức hệ V [3], phần lớn được tìm thấy trên hai gen ty thể là *ATP6* và *ATP8* có liên quan tới các bệnh như yếu cơ thần kinh, viêm võng mạc sắc tố và bệnh cơ tim phì đại,...

## 1.2. NGŨ HỆ HÁN - TẠNG TẠI VIỆT NAM

Việt Nam là một đất nước nằm ở vị trí quan trọng của khu vực Đông Nam Á với đường biên giới tiếp giáp với Trung Quốc ở phía Bắc, Lào ở phía Tây Bắc, Campuchia ở phía Tây Nam và biển Đông ở phía Đông. Theo thống kê của Tổng cục thống kê vào năm 2023, Việt Nam có diện tích vào khoảng 331.000 km<sup>2</sup> và là nơi cư trú của khoảng 100 triệu người (<https://www.gso.gov.vn>), chủ yếu tập trung ở khu vực đồng bằng. Đây là một đất nước đa dạng dân tộc với 54 dân tộc khác nhau, thuộc về 5 ngữ hệ lớn trên thế giới, bao gồm Nam Á, Nam Đảo, Mông - Miên, Thái - Kadai, và Hán - Tạng. Phần lớn người dân Việt Nam là thuộc ngữ hệ Nam Á (chiếm khoảng 89,9% tổng dân số) với dân tộc Kinh là dân tộc đông dân nhất (chiếm 85,3% dân số). Các người dân còn lại cư trú ở vùng núi, vùng sâu vùng xa thuộc về bốn ngữ hệ còn lại, chiếm lần lượt 5,9% ở Thái - Kadai, 2,1% người Mông - Miên, 1,2% người Hán - Tạng và 0,9% người Nam Đảo.

Trên thế giới, người Hán - Tạng phân bố chủ yếu ở Trung Quốc, Myanmar và Thái Lan. Ở Việt Nam, hiện nay có khoảng 1 triệu người người Hán - Tạng, thuộc về 9 dân tộc Hán - Tạng, sinh sống chủ yếu ở vùng núi phía Bắc của Việt Nam [14]. Những dân tộc này được phân làm hai nhánh ngôn ngữ khác nhau, bao gồm nhánh Hán (Hoa, Sán Dìu và Ngái) và nhánh Tạng - Miên (Hà Nhì, Phù Lá, Si La, La Hủ, Lô Lô, Cống).

Dân tộc Hoa thuộc nhánh Hán có số lượng rơi vào khoảng 750.000 người, là dân tộc có số lượng người nhiều nhất ở ngữ hệ này [14]. Người dân tộc Hoa sinh sống trải khắp Việt Nam. Tổ tiên của người dân tộc Hoa ở Việt Nam được cho là đến từ Trung



Quốc [15], nơi có số lượng người Hán chiếm trên 90% dân số Trung Quốc. Họ di chuyển đến Việt Nam vào nhiều thời điểm khác nhau trong tiến trình lịch sử, đặc biệt là vào lúc trước và sau thời điểm chiến tranh thế giới thứ II. Phần lớn trong số họ là người bản địa của các tỉnh phía Nam Trung Quốc như Quảng Đông, Quảng Tây, Phúc Kiến, và Chiết Giang [15].

Dân tộc Sán Dìu thuộc nhánh Hán có khoảng 183.000 người, chiếm 2,2% số người Hán - Tạng ở Việt Nam [14] và chỉ xếp sau dân tộc Hoa về số lượng người. Họ cư trú chủ yếu ở các vùng trung du của một số tỉnh miền Bắc như Thái Nguyên, Vĩnh Phúc và Tuyên Quang. Nghiên cứu trước đây cho rằng người Sán Dìu có nguồn gốc từ tỉnh Quảng Đông, Trung Quốc và di cư đến Việt Nam vào khoảng thế kỷ 17 [15].

Dân tộc Công thuộc nhánh Tạng - Miến có dân số vào khoảng 2.700 người [14]. Họ có số lượng người tương đối thấp trong ngữ hệ Hán - Tạng, là dân tộc thiểu số (chỉ nhiều hơn dân tộc Si La và Ngái). Người Công phân bố chủ yếu ở một vài xã thuộc huyện Mường Tè, tỉnh Lai Châu cũng như các khu vực dọc theo biên giới Việt Nam - Lào - Trung Quốc [15].

### **1.3. TÌNH HÌNH NGHIÊN CỨU HỆ GEN TY THỂ**

#### **1.3.1. Thế giới**

Với sự phát triển vượt bậc của khoa học công nghệ cũng như kỹ thuật giải trình tự trong vài thập kỷ qua, các nghiên cứu về hệ gen người ngày càng được đẩy mạnh. Qua đó, hàng triệu các chỉ thị trên toàn bộ hệ gen đã được xác định phục vụ cho việc phân tích chuyên sâu về đặc điểm di truyền ở người. Thuật ngữ “Variome” hay “Hệ biến thể gen” đã được ra đời và dùng để chỉ tất cả các biến thể gen được tìm thấy ở các quần thể khác nhau của cùng một loài. Ở người, hệ biến thể gen này bao gồm 2 thành phần: (1) hệ gen nhân, có kích thước khoảng 3,2 tỷ base pair (bp) và (2) hệ gen ty thể hay hệ gen biến thể ty thể, có kích thước chỉ hơn 16 kbp. Tuy nhiên, để nghiên cứu toàn bộ hệ gen lại đòi hỏi một nguồn kinh phí đầu tư rất lớn và lâu dài, do đó hệ gen biến thể ty thể đã được chọn để tập trung nghiên cứu trước.

Hệ gen ty thể ở người đã và đang được sử dụng trong nhiều lĩnh vực nghiên cứu khác nhau, bao gồm nhân chủng học, bệnh học ty thể và nghiên cứu về cái tính trạng thích nghi với môi trường. Một trong những hướng nghiên cứu lâu đời nhất sử dụng mtDNA là nghiên cứu về nguồn gốc tiến hoá của loài người. Trong đó, nghiên cứu nổi tiếng nhất là của Cann và các cộng sự vào năm 1987 [16], đã chỉ ra rằng tất cả hệ gen ty thể của người hiện đại đều bắt nguồn từ một quần thể châu Phi vào khoảng 200.000 năm trước. Sau đó, những nghiên cứu tiếp nối đã sử dụng những biến thể được tích lũy và phát tán theo dòng mẹ trên mtDNA để xây dựng nên các nhóm đơn bội đặc trưng cho

các quần thể và khu vực khác nhau trên thế giới [17]. Theo đó, và khoảng 65.000 năm trước, nhóm đơn bội L3 đặc trưng cho người châu Phi đã di cư từ vùng Đông Phi tới lục địa châu Á - Âu [18] và tiếp tục mở rộng tới Nam Á và Đông Nam Á trước khi phân nhánh thành 2 nhóm đơn bội ty thể M và N, là tổ tiên của tất cả các nhóm đơn bội khác được tìm thấy bên ngoài Châu Phi [19]. Ngày nay, nghiên cứu sử dụng các biến thể định danh cho một số nhóm đơn bội nhất định đã cho thấy có thể có hiện tượng trôi dạt di truyền gây đây và một số sự phát tán cổ xưa từ Myanmar đến Trung Quốc vào khoảng 10 đến 25 nghìn năm trước [20]. Bên cạnh đó, một nghiên cứu khác trên 4.004 biến thể định danh nhóm đơn bội ở 21.668 người Hán ở Trung Quốc đã cho thấy sự khác biệt về mặt di truyền có ý nghĩa thống kê giữa các nhóm người ở 3 hệ thống sông chính là sông Dương Tử, sông Hoàng Hà và sông Châu, và những nhóm người này đóng góp cho sự phổ biến của các nhóm đơn bội D4, B4 và M7 ở lưu vực những sông này [21].

Bên cạnh hướng nghiên cứu kể trên, các biến thể và đột biến trên mtDNA cũng được tập trung nghiên cứu liên quan đến bệnh khác nhau, bao gồm những bệnh đặc trưng cho ty thể như liệt thần kinh thị giác di truyền Leber (LHON), bệnh MELAS, hội chứng Leigh và những bệnh có kiểu hình lâm sàng phổ biến khác như bệnh tim mạch, bệnh chuyển hoá, bệnh thần kinh, covid,... Đặc biệt ở bệnh LHON, khoảng 90% các bệnh nhân mắc bệnh này mang ít nhất 1 trong 3 đột biến trên ba gen ty thể: 3460A trên gen *ND1*, 11778A trên gen *ND4* và 14484C trên gen *ND6* [22]. Ngoài ra, cũng có nghiên cứu cho rằng gen *ND6* là “điểm nóng” của các đột biến gây ra LHON [23]. Đột biến 13513A trên gen *ND5* được tìm thấy ở 26 trên 33 bệnh nhân mắc hội chứng MELAS và hội chứng Leigh [24]. Hiện nay, các nghiên cứu về các biến thể mtDNA liên quan đến bệnh ty thể thường tập trung vào từng quần thể hoặc từng khu vực khác nhau [25-27]. Năm 2021, nghiên cứu khác về sự liên quan của các biến thể mtDNA với nguy cơ mắc bệnh COVID-19 ở người Hồ Bắc (Trung Quốc) đã cho thấy rằng các biến thể mtDNA phổ biến C5178a và A249d/T6392C/G10310A có thể góp phần vào khả năng chống lại sự phát triển nghiêm trọng của bệnh ( $p < 0,05$ ), trong khi A4833G, A4715G, T3394C và G5417A/C16257a/C16261T có thể làm tăng nguy cơ phát triển bệnh COVID-19 nghiêm trọng ( $p < 0,05$ ) [25]. Sau đó, vào năm 2022, nghiên cứu trên mtDNA của nhóm 146 người Daur thuộc ngữ hệ Thái - Kadai ở Trung Quốc, đã tìm thấy 71 biến thể được báo cáo có liên quan tới bệnh ty thể, trong đó 17 biến thể có tần số tối thiểu 0,05 (chiếm 23,94%) [26]. Phần lớn các biến thể này nằm ở vùng mã hoá với biến thể 10398G xuất hiện phổ biến nhất (chiếm 74,66%), 11696A và 6962A là hai biến thể xuất hiện ít nhất (chiếm 6,16%). Nghiên cứu gần đây nhất của Kutanana [27] trên 82 người mắc 3 loại bệnh tim mạch khác nhau (bệnh cơ tim phì đại - HCM, hội chứng QT dài - LQTS, hội chứng Brugada - BrS) ở phía Đông Bắc Thái Lan đã cho thấy số lượng biến thể mtDNA

trên gen rRNA cao hơn đáng kể được tìm thấy ở những bệnh nhân mắc HCM và BrS ( $p < 0,001$ ) so với những người mắc LQTS hoặc nhóm đối chứng. Ba đa hình trên gen mã hoá tRNA (5618C, 5631A và 4392T) được dự đoán làm thay đổi cấu trúc thứ cấp của tRNA và có thể dẫn đến thay đổi trong chức năng của tRNA [27]. Ngoài ra, nhóm đơn bội R9c có liên quan tới HCM (OR = 62,42; 95% CI = 6,892 - 903,4;  $p = 0,0032$ ), trong khi đó, nhóm đơn bội M12c có liên quan tới LQTS (OR = 32,93; 95% CI = 5,784 - 199,6;  $p = 0,0039$ ).

Ngoài ra, với vai trò trung tâm của các gen mtDNA trong sinh lý của tế bào thông qua hệ thống OXPHOS, các biến thể trên mtDNA lại càng được chú ý trong những nghiên cứu ảnh hưởng của chúng tới sự thích nghi của con người với điều kiện môi trường sống [17]. Cụ thể, nghiên cứu trước đây đã chỉ ra rằng hai biến thể 10398A (A114T) trên gen *ND3* và 8701A (A59T) trên gen *ATP6* làm thay đổi điện thế màng cũng như khả năng điều chỉnh ion canxi [28], do đó có khả năng ảnh hưởng tới hiệu quả của quá trình tạo năng lượng ATP và thích nghi tốt hơn với điều kiện khí hậu lạnh. Tương tự với nghiên cứu này, nghiên cứu về các biến thể ở gen *ATP6* trên quần thể người Siberia ở Bắc Á cho thấy chúng có liên quan tới sự thích nghi với điều kiện khắc nghiệt do giảm hoạt động của chuỗi vận chuyển điện tử, từ đó tạo ra nhiều nhiệt hơn cho cơ thể [29]. Ở ngữ hệ Hán - Tạng, nhiều nghiên cứu đã được thực hiện trên nhóm người Tây Tạng ở Trung Quốc về khả năng thích nghi với môi trường cao nguyên liên quan đến mtDNA. Nghiên cứu của Kang vào năm 2013 trên người Sherpa ở Tây Tạng đã tìm thấy 2 biến thể là 4216C và 3745A trên gen *ND1* định danh cho 2 nhóm đơn bội C4a3b1 và A4e3a đặc trưng cho quần thể này và đề xuất rằng chúng có ảnh hưởng nhất định đến các quy định trao đổi chất và có thể tham gia vào quá trình thích nghi [30]. Tiếp đó, một nghiên cứu khác trên người Tây Tạng vào năm 2016 đã cho thấy rằng các đột biến gây bệnh xảy ra thường xuyên hơn ở 22 nhánh đặc trưng của vùng cao nguyên so với 6.857 nhóm đơn bội của tất cả 36.914 trình tự đối chứng ( $p = 4,87 \times 10^{-8}$ ) [31]. Hơn nữa, số lượng đột biến trung bình có thể gây bệnh ( $3,18 \pm 1,27$ ) có thể xảy ra ở người vùng cao cao hơn đáng kể so với nhóm đối chứng ( $2,82 \pm 1,40$ ) ( $p = 1,89 \times 10^{-4}$ ). Từ đó, nghiên cứu này đã đưa ra giả thuyết rằng các biến thể này có thể đóng một vai trò trong việc thích nghi với tình trạng thiếu oxy [31].

Nhìn chung, các nghiên cứu trên hệ gen biến thể ty thể trước đây đã chỉ ra rằng, mỗi quần thể người khác nhau lại có những đặc điểm di truyền đặc trưng, thể hiện bằng sự phân bố tần suất allele các locus DNA trong mỗi quần thể là khác nhau và không thể áp dụng cơ sở dữ liệu của quần thể này cho một quần thể khác. Qua đó, việc tiến hành nghiên cứu trên các hệ gen ty thể của các nhóm người khác nhau là cần thiết để bổ sung vào cơ sở dữ liệu biến thể giúp cho các nghiên cứu về đặc điểm di truyền quần thể và di

truyền bệnh học sau này.

### 1.3.2. Việt Nam

Các nghiên cứu về một số vùng của hệ gen ty thể của người Việt Nam đã được đẩy mạnh từ khoảng hai thập kỉ trở lại [32-40]. Năm 2005, Huỳnh Thị Thu Huệ và đồng tác giả đã công bố công trình đầu tiên về trình tự vùng điều khiển của 5 cá thể người Việt Nam thuộc ba dân tộc Kinh, Tày và H'Mông [39]. Năm 2008, nhóm nghiên cứu của Nguyễn Đăng Tôn đã phát hiện 73 đa hình mới khi nghiên cứu khảo sát tần số của các nhóm đơn bội ty thể ở 78 cá thể người Việt Nam thuộc ba dân tộc Kinh, Tày và Mường [38]. Sau đó, vào năm 2016, Đỗ Mạnh Hưng và các cộng sự đã phát hiện được 79 nhóm đơn bội khi phân tích đoạn HVS-II của 169 cá thể người dân tộc Kinh, Mường, Gia Rai và Ê đê ở Việt Nam [36]. Năm 2018, khi phân tích vùng D-loop của các mẫu dân tộc Kinh và Mảng, Nguyễn Thy Ngọc và các cộng sự đã tìm thấy 8 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 [34]. Kết quả nghiên cứu này cho thấy mặc dù cùng thuộc nhóm ngữ hệ Nam Á, hệ gen ty thể giữa 2 dân tộc trong cộng đồng các dân tộc Việt Nam vẫn tồn tại những khác biệt đáng chú ý. Tiếp đó, nghiên cứu về vùng D-loop của trình tự toàn bộ hệ gen ty thể của 119 mẫu dân tộc Kinh thuộc nhánh ngôn ngữ Việt – Mường và 2 dân tộc thuộc nhánh ngôn ngữ Tạng – Miến (Lô Lô và La Hủ) đã tìm thấy lần lượt 23, 13 và 24 đa hình khác biệt khi so sánh từng cặp (Kinh – Lô Lô, Kinh – La Hủ, Lô Lô – La Hủ) [35]. Khi tính toán khoảng cách di truyền giữa 3 dân tộc này cho thấy chúng có khoảng cách di truyền tương đối gần nhau dù thuộc hai ngữ hệ khác nhau (0,0092 - 0,0101). Năm 2020, nghiên cứu trên vùng HVS-I và HVS-II của 517 người thuộc các dân tộc Kinh, Khmer, Mường và Chăm cũng đã tìm thấy các đa hình phổ biến trên từng nhóm người [37]. Ngoài ra, các nghiên cứu trên vùng mã hóa cũng đã được tiến hành. Trong đó, nổi bật nhất phải kể đến nghiên cứu của Nguyễn Minh Tùng và các cộng sự khi đã giải trình tự thành công hai gen *ND5* và *ND6* ở 54 người dân tộc Giarai và Êđê sinh sống ở tỉnh Tây Nguyên và bước đầu cho thấy tính đa hình trình tự hai gen mã hóa này [41]. Nghiên cứu phát hiện và định lượng đột biến G11778A trên gen *ND6* liên quan đến hội chứng liệt thần kinh thị giác di truyền Leber (LHON) bằng cách sử dụng kết hợp kỹ thuật PCR-RFLP và real-time PCR cũng đã được thực hiện trên nhóm 149 bệnh nhân tại bệnh viện Nhi Trung Ương. Kết quả đã phát hiện một bệnh nhân mang đột biến G11778A ở dạng không đồng nhất với tỷ lệ đột biến là  $2,71 \pm 0,12\%$  [42].

Bên cạnh những nghiên cứu từng vùng của hệ gen ty thể đã kể trên, các nghiên cứu về toàn bộ hệ gen ty thể cũng đã được triển khai. Năm 2008, lần đầu tiên tại Việt Nam đã giải mã thành công hệ gen ty thể của người Việt Nam, gồm 10 cá thể thuộc 3 dân tộc Kinh, Tày, Mường (mỗi dân tộc 3 cá thể) và 1 bệnh nhân mắc bệnh liên quan đến gen ty thể [43]. Sau đó, vào năm 2018, nhóm nghiên cứu của TS. Nguyễn Thùy Dương đã thực hiện giải mã toàn bộ hệ gen ty thể của 609 cá thể thuộc 17 dân tộc (Kinh, Mảng, Tày, Nùng, Thái, Cờ Lao, La Chí, Gia Rai, Ê Đê, Pà Thên, Hmong, Dao, La Hủ, Hà Nhì, Phù Lá, Lô Lô và Si La) của 5 nhóm ngôn ngữ tại Việt Nam [32]. Phân tích cây phát sinh chủng loại của tổng số 2742 trình tự, bao gồm 609 trình tự người Việt Nam cùng với 2133 trình tự khác từ các dân tộc cư trú trên lục địa DNA (bao gồm: Thái Lan, Lào, Campuchia, Myanmar, Tây Malaysia, Nam Trung Quốc) và Đài Loan, đã phát hiện được 111 dòng nhánh (lineages) mới DNA ty thể của Việt Nam [32]. Đặc biệt, nhóm đã phát hiện được đỉnh cao của sự đa dạng mtDNA khoảng 2,5-3 ngàn năm về trước, trùng với nền Văn hóa Đông Sơn, và do vậy có thể liên quan đến sự mở rộng nền văn hoá theo xu hướng nông nghiệp. Đây là bằng chứng khoa học mới, đầu tiên trên thế giới về di truyền phân tử người liên quan đến khoảng thời gian ra đời của nền văn hoá Đông Sơn [32]. Ngoài ra, nghiên cứu về ngữ hệ Mông – Miến gồm 3 dân tộc H'mông, Pà Thên, và Dao cho thấy dân tộc Dao có mức độ đa dạng di truyền cao nhất ( $H=0.957$ ) và dân tộc Pà Thên có mức độ đồng đều về mặt di truyền cao nhất ( $H=0.900$ ) [33]. Phân tích về di truyền quần thể cũng chỉ ra rằng. Ngoài ra, khi xét về khoảng cách di truyền thì dân tộc Dao và H'mông tương đối khác nhau, trong khi đó H'mông và Pà Thên lại có mối quan hệ gần gũi [33]. Công bố gần đây về hệ gen ty thể của 369 cá thể thuộc hai ngữ hệ Nam Á (Ba Na, M'Nông, H'Re, Cơ Ho và Mạ) và Nam Đảo (Gia Rai, Ê đê, Chu Ru, Raglay và Chăm) đã cho thấy sự giao thoa giữa dân tộc Gia Rai của Việt Nam và một số nhóm người ở Đài Loan (Rukai, Paiwan và Bunun) [44]. Tuy nhiên, nhìn chung, nhóm dân tộc Nam Đảo của Việt Nam có sự gần gũi lớn hơn với các nhóm dân tộc Nam Á ở xung quanh, cho thấy rằng có thể có sự chuyển đổi về mặt ngôn ngữ giữa hai nhóm dân tộc thuộc hai ngữ hệ này. Đây là nghiên cứu hoàn chỉnh đầu tiên về hệ gen ty thể của toàn bộ 5 dân tộc thuộc ngữ hệ Nam Đảo của Việt Nam, cho thấy rằng sự phát triển của nhóm dân tộc này trên lục địa Đông Nam Á là sự kết hợp của cả việc giao thoa văn hoá và di cư [44].

## **CHƯƠNG 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU**

### **2.1. ĐỐI TƯỢNG VÀ PHẠM VI NGHIÊN CỨU**

#### **2.1.1. Đối tượng nghiên cứu**

Đối tượng nghiên cứu trong đề tài là các mẫu máu ngoại vi của 69 cá thể người dân tộc thiểu số Việt Nam, trong đó có 17 cá thể người dân tộc Cống, 26 cá thể người dân tộc Hoa và 26 cá thể người dân tộc Sán Dìu được thu thập lần lượt tại tỉnh Lai Châu, Thành phố Hồ Chí Minh và tỉnh Thái Nguyên. Các cá thể được lựa chọn cho nghiên cứu là những người khoẻ mạnh, không dùng huyết thống trong vòng 3 đời cũng như có cả 4 ông bà nội ngoại đều cũng thuộc một dân tộc.

#### **2.1.2. Phạm vi nghiên cứu**

Đề tài nghiên cứu được thực hiện tại phòng Hệ gen học người, Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam trong thời gian từ năm 2022 - 2024. Nghiên cứu được hỗ trợ kinh phí từ đề tài nghiên cứu cấp Quốc gia “Xây dựng cơ sở dữ liệu hệ gen biến thể ty thể và nhiễm sắc thể Y của một số dân tộc người Việt Nam” chủ nhiệm bởi PGS. TS. Nguyễn Thuỳ Dương trong khoảng thời gian 2019 – 2024.

#### **2.1.3. Đạo đức trong nghiên cứu**

Những đối tượng trong nghiên cứu đều được kí xác nhận vào đơn tự nguyện tham gia và đồng ý cung cấp mẫu máu cho nghiên cứu trước khi được thu thập mẫu máu. Việc tiến hành thu thập mẫu máu cũng như thực hiện nghiên cứu trên đối tượng người đã được đồng ý thông qua bởi Hội đồng đạo đức trong nghiên cứu y sinh của Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam (Số: 9-2019/NCHG-HĐĐĐ).

### **2.2. PHƯƠNG PHÁP NGHIÊN CỨU**

#### **2.2.1. Thu thập và tách chiết DNA tổng số từ các mẫu nghiên cứu**

Các mẫu máu ngoại vi sau khi thu thập được bảo quản ở điều kiện  $-80^{\circ}\text{C}$ . DNA tổng số của 69 cá thể người dân tộc được tách chiết từ mẫu máu đã thu thập bằng bộ kit thương mại Exgene™ Blood SV Mini (GeneAll, Korea) theo hướng dẫn của nhà sản xuất, gồm các bước như sau:

- Bước 1: Các mẫu máu được giữ đông nhanh ở nhiệt độ  $37^{\circ}\text{C}$  trong bể ổn nhiệt và giữ trên đá trước khi sử dụng;
- Bước 2: Chuyển 200  $\mu\text{L}$  máu toàn phần vào ống Eppendorf 1,5 mL có chứa sẵn 20  $\mu\text{L}$  proteinase K và mix đều bằng máy vortex;
- Bước 3: Thêm 200  $\mu\text{L}$  dung dịch đệm BL vào ống Eppendorf 1,5 mL ở trên và trộn đều bằng máy vortex để được dung dịch đồng nhất;
- Bước 4: Ủ mẫu ở nhiệt độ  $56^{\circ}\text{C}$  cho đến khi tế bào tan hoàn toàn;

- Bước 5: Thêm 200  $\mu\text{L}$  cồn (96 - 100%) và mix nhẹ nhàng bằng pipet cho đến khi được dung dịch đồng nhất;
- Bước 6: Chuyển hỗn hợp trên vào cột được cung cấp kèm theo bộ kit. Ly tâm cột ở tốc độ 10.000 vòng/ phút trong vòng 1 phút. Loại bỏ phần dịch thu được và đặt cột vào ống thu mới;
- Bước 7: Thêm 500  $\mu\text{L}$  đệm BW (đã được bổ sung cồn 96 - 100%) vào cột và ly tâm ở tốc độ 10.000 vòng/ phút trong vòng 1 phút. Loại bỏ phần dịch thu được và đặt cột vào ống thu mới;
- Bước 8: Thêm 700  $\mu\text{L}$  đệm TW (đã được bổ sung cồn 96 - 100%) vào cột và ly tâm ở tốc độ tối thiểu 12.000 vòng/ phút trong vòng 3 phút. Loại bỏ phần dịch thu được và đặt cột vào ống thu mới;
- Bước 9: Ly tâm cột ở tốc độ tối thiểu 12.000 vòng/ phút trong vòng 2 phút để loại bỏ tất cả dịch còn lại trên màng của cột;
- Bước 10: Đặt cột vào ống Eppendorf 1,5 mL mới và thêm 200  $\mu\text{L}$  đệm TE vào giữa màng để thu DNA. Ủ cột ở nhiệt độ phòng trong vòng 10 phút và ly tâm ở tốc độ 10.000 vòng/ phút trong vòng 1 phút để thu DNA.

### 2.2.2. Kiểm tra chất lượng DNA tổng số

Mẫu DNA sau khi tách chiết được kiểm tra chất lượng bằng phương pháp điện di trên gel agarose 0,8% theo các bước sau:

- Bước 1: Hòa 0,8 g agarose trong 100 mL TAE 1X, đun sôi trong lò vi sóng cho tan hoàn toàn. Để cho nhiệt độ hạ xuống khoảng 50°C, đổ vào khuôn có lắp đặt sẵn rãnh lược thích hợp. Đợi cho thạch đông và ổn định trong khoảng 30 phút. Sau đó bỏ rãnh lược ra và đặt gel vào bể điện di. Đổ đệm TAE 1X sao cho mức đệm cao hơn mặt gel từ 1 - 2 mm;
- Bước 2: Trộn một lượng mẫu thích hợp với đệm tra mẫu (2X). Điện di với dòng điện một chiều có hiệu điện thế 150V, trong thời gian khoảng 20 phút;
- Bước 3: Nhuộm gel bằng Ethidium bromide 10  $\mu\text{g}/\text{mL}$  trong khoảng 5 phút và rửa sạch với nước 1X trong khoảng 3 phút;
- Bước 4: Quan sát và chụp ảnh trên máy GelDoc UVP.

Tiếp tục tiến hành đo DNA bằng máy quang phổ kế NanoDrop™ One (Thermo Fisher) theo các bước sau:

- Bước 1: Lấy 2  $\mu\text{L}$  đệm TE để làm mẫu đối chứng;
- Bước 2: Sử dụng 2  $\mu\text{L}$  mỗi mẫu để xác định nồng độ DNA ở bước sóng 260 nm và 280 nm;

- Bước 3: Ghi lại kết quả nồng độ ADN cho mỗi mẫu nghiên cứu sau mỗi lần đo. Các mẫu DNA được bảo quản và lưu trữ ở tủ lạnh  $-20^{\circ}\text{C}$ .

### 2.2.3. Thiết lập và làm giàu thư viện

#### 2.2.3.1. Chuẩn bị DNA đầu vào

Mẫu DNA tổng số ban đầu sẽ được pha loãng về nồng độ  $10\text{ ng}/\mu\text{L}$  sử dụng dung dịch đệm TE. Sau đó,  $80\ \mu\text{L}$  dung dịch DNA tổng số đã pha loãng sẽ được chuyển vào ống Eppendorf  $0,5\ \text{mL}$  để tiến hành thí nghiệm phân cắt DNA tổng số sử dụng máy siêu âm Bioruptor-ucd-200 TM-EX (Diagenode) với ít nhất 3 chu kỳ siêu âm. Sản phẩm phân cắt DNA tổng số sẽ được điện di kiểm tra trên gel agarose 2%. Ngoài ra, các mẫu DNA đối chứng cũng được chuẩn bị và sử dụng từ bước tạo đầu bằng với mục đích kiểm tra và đánh giá mức độ hiệu quả trong quy trình chuẩn bị thư viện DNA. Đối chứng âm là dung dịch đệm TE và đối chứng dương là đoạn DNA hoặc sản phẩm PCR có kích thước khoảng  $200\ \text{bp}$ . Các mẫu DNA tổng số sau khi được làm đứt gãy nhờ siêu âm sẽ được chuyển vào một đĩa PCR cùng với các mẫu đối chứng ở trên.

#### 2.2.3.2. Tạo đầu bằng cho các đoạn DNA

Phản ứng tạo đầu bằng cho các đoạn DNA được thực hiện trên đĩa mẫu PCR với tổng thể tích của một phản ứng là  $70\ \mu\text{L}$ , bao gồm:  $7,12\ \mu\text{L}\ \text{H}_2\text{O}$ ;  $7,0\ \mu\text{L}$  đệm Tango (10X);  $0,28\ \mu\text{L}$  dNTPs (25 mM);  $0,7\ \mu\text{L}$  ATP (100 mM);  $3,5\ \mu\text{L}$  T4 polynucleotide kinase ( $10\text{U}/\mu\text{L}$ );  $1,4\ \mu\text{L}$  T4 DNA polymerase ( $5\text{U}/\mu\text{L}$ ) và  $50\ \mu\text{L}$  dung dịch DNA sau khi phân cắt bằng siêu âm. Các phản ứng này được sử dụng kèm thêm 1 đối chứng âm và 1 đối chứng dương. Ủ phản ứng tạo đầu bằng ở  $25^{\circ}\text{C}$  trong thời gian 15 phút, sau đó giữ phản ứng ở  $12^{\circ}\text{C}$  trong thời gian 5 phút. Tinh sạch các sản phẩm sau khi tạo đầu bằng Solid Phase Reversible Immobilization (SPRI) theo các bước sau:

- Bước 1: Cân bằng các hạt từ (bead) SPRI ở nhiệt độ phòng trong 30 phút, tránh ánh sáng.
- Bước 2: Trộn đều các hạt từ SPRI thành một dung dịch đồng nhất bằng máy vortex. Sau đó, bổ sung  $126\ \mu\text{L}$  dung dịch hạt từ SPRI vào đĩa mẫu. Đậy các ống của đĩa mẫu bằng một tấm dán nhiệt (thermo film).
- Bước 3: Trộn đều bằng máy vortex để làm đồng nhất dung dịch. Để ở nhiệt độ phòng trong 5 phút.
- Bước 4: Ly tâm nhẹ đĩa mẫu bằng máy ly tâm. Caren thận bỏ tấm dán nhiệt (thermo film) khỏi đĩa mẫu và đặt đĩa mẫu lên giá từ tính (magnetic plate). Để ở nhiệt độ phòng 3 phút.



- Bước 5: Hút bỏ phần dung dịch trong các giếng của đĩa mẫu. Rửa các hạt từ với 150  $\mu$ L dung dịch cồn (ethanol) 70% trong khi vẫn giữ đĩa mẫu trên giá từ tính. Ủ 1 phút ở nhiệt độ phòng và loại bỏ dung dịch ethanol.
- Bước 6: Tiếp tục rửa các hạt từ với 150  $\mu$ L dung dịch cồn (ethanol) 70% thêm một lần nữa. Làm khô các hạt từ ở nhiệt độ phòng trong thời gian 20 phút.
- Bước 7: Nhấc đĩa mẫu ra khỏi giá từ tính. Bổ sung 22  $\mu$ L dung dịch EBT để tách DNA khỏi các hạt từ.
- Bước 8: Đậy các ống của đĩa mẫu bằng một tấm dán nhiệt. Vortex đĩa mẫu vài giây để dung dịch trở nên đồng nhất. Ly tâm nhẹ đĩa mẫu trong máy ly tâm.
- Bước 9: Đặt đĩa mẫu trở lại giá từ tính và bỏ tấm dán nhiệt khỏi đĩa mẫu. Để ở nhiệt độ phòng 3 phút để kéo các hạt từ lên thành của các giếng trong đĩa mẫu.
- Bước 10: Chuyển 22  $\mu$ L dung dịch sang một đĩa mới và bảo quản đĩa mẫu ở  $-20^{\circ}\text{C}$ .

### 2.2.3.3. Gắn adapter

Quá trình gắn adapter sử dụng 3 loại adapter, bao gồm: Sol\_MP\_P5, Sol\_MP\_P7 và Sol\_MP\_M5P7\_comp (Bảng 2.1).

**Bảng 2.1. Trình tự nucleotide của các adapter sử dụng trong quá trình gắn adapter**

Tên adapter	Trình tự nucleotide
Sol_MP_P5	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCG*A*T*C*T
Sol_MP_P7	Biotine-G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCG*A*C*T*T
Sol_MP_M5P7_comp	A*G*A*T*CGGAA*G*A*G*C

Ghi chú: \* thể hiện sự sửa đổi PTO (phosphorothioate).

Adapter P5 có nồng độ 200  $\mu$ M bao gồm: 20  $\mu$ L Sol\_MP\_P5 (500  $\mu$ M); 20  $\mu$ L Sol\_MP\_P5P7\_comp (500  $\mu$ M); 5  $\mu$ L Hyb buffer (10X); 5  $\mu$ L dung dịch TE (1X). Adapter P7 có nồng độ 500  $\mu$ M bao gồm: 20  $\mu$ L Sol\_MP\_P7 (500  $\mu$ M); 20  $\mu$ L Sol\_MP\_P5P7\_comp (500  $\mu$ M); 5  $\mu$ L Hyb buffer (10X); 5  $\mu$ L dung dịch TE (1X). Ủ adapter P5 và P7 ở  $95^{\circ}\text{C}$  trong 1 phút. Sau đó giảm nhiệt độ xuống  $14^{\circ}\text{C}$  với tốc độ giảm nhiệt  $0,1^{\circ}\text{C}/\text{s}$ . Adapter mix được chuẩn bị bằng cách trộn adapter P5 và adapter P7 theo tỷ lệ thể tích 1:1.

Quá trình gắn adapter được thực hiện với phản ứng có tổng thể tích 40  $\mu$ L, bao gồm các thành phần: 10  $\mu$ L  $\text{H}_2\text{O}$ ; 4  $\mu$ L T4 DNA ligase bufer (10X); PEG-4000 (50%), Adapter mix (100  $\mu$ M), T4 DNA ligase (5U/  $\mu$ L) và 20  $\mu$ L sản phẩm đã được tạo đầu bằng trước đó. Đậy đĩa mẫu bằng một tấm thermo film và đảo đều bằng máy vortex. Ly tâm nhẹ đĩa mẫu bằng máy ly tâm và ủ đĩa mẫu ở  $60^{\circ}\text{C}$  trong 60 phút. Tinh sạch sản phẩm gắn adapter bằng SPRI giống như đã miêu tả ở trên.

#### 2.2.3.4. *Làm đầy adapter (fill-in adapter)*

Quá trình làm đầy adapter được thực hiện với phản ứng có tổng thể tích 40  $\mu\text{L}$ , bao gồm các thành phần: 14,1  $\mu\text{L}$   $\text{H}_2\text{O}$ ; 4,0  $\mu\text{L}$  Thermopol buffer (10X); 0,4  $\mu\text{L}$  dNTPs (25 mM); 1,5  $\mu\text{L}$  Bst polymerase (8U/  $\mu\text{L}$ ) và 20  $\mu\text{L}$  sản phẩm đã gắn adapter trước đó. Đậy đĩa mẫu bằng một tấm thermo film và đảo đều bằng máy vortex. Ly tâm nhẹ đĩa mẫu bằng máy ly tâm và ủ đĩa mẫu ở 37°C trong 20 phút. Tinh sạch sản phẩm gắn adapter bằng SPRI giống như đã miêu tả ở trên. Mẫu đối chứng dương và đối chứng âm sau khi làm đầy adapter được điện di kiểm tra trên gel agarose 2% để đánh giá tính hiệu quả của quá trình gắn adapter.

Phương pháp PCR định lượng (qPCR) sẽ được sử dụng để xác định số chu kỳ cho phản ứng index PCR ở bước tiếp theo. Từ kết quả của qPCR, số chu kỳ được chọn sẽ phản ánh mức độ tích lũy của sản phẩm PCR ở điểm bắt đầu đi vào pha cân bằng ở toàn bộ các mẫu định lượng. Phản ứng PCR định lượng bao gồm các thành phần: 7,0  $\mu\text{L}$   $\text{H}_2\text{O}$ ; 1,0  $\mu\text{L}$  môi xuôi (Sol\_iPCR\_P5); 1,0  $\mu\text{L}$  môi ngược (Sol\_iPCR\_P7); 10,0  $\mu\text{L}$  Maxima Mastermix và 1,0  $\mu\text{L}$  sản phẩm làm đầy adapter. Cụ thể, 1  $\mu\text{L}$  của sản phẩm làm đầy adapter từ 8 mẫu ngẫu nhiên sẽ được sử dụng. Sau đó chuẩn bị các mẫu pha loãng với hai tỷ lệ (1:10) và (1:100); đối chứng âm và mẫu qPCR chuẩn với các nồng độ giảm dần  $10^9$ ,  $10^8$ ,  $10^7$ ,  $10^6$ ,  $10^5$ ,  $10^4$ ,  $10^3$ ,  $10^2$ .

#### 2.2.3.5. *Indexing PCR*

Phản ứng index PCR được thực hiện với mục đích gắn các trình tự nucleotide của môi index vào các đoạn DNA ở các mẫu thư viện DNA. Bên cạnh đó, các đoạn DNA này cũng được khuếch đại lên nhiều lần trong phản ứng PCR. Mỗi mẫu thư viện DNA sẽ được gắn với một cặp môi index riêng biệt. Với việc thiết kế 96 trình tự nucleotide môi index P5 (P5\_iPCR\_LP1 - P5\_iPCR\_LP96) và 96 trình tự nucleotide môi index P7 (P7\_iPCR\_LP301 - P7\_iPCR\_LP396), chúng ta có thể tiến hành giải mã trình tự của hàng triệu thư viện DNA cùng một lần.

Phản ứng indexing PCR được thực hiện bằng cách chuẩn bị một ống mastermix có tổng thể tích 37  $\mu\text{L}$ , bao gồm các thành phần: 26,0  $\mu\text{L}$   $\text{H}_2\text{O}$ ; 10,0  $\mu\text{L}$  đệm Herculase (5X); 0,5  $\mu\text{L}$  dNTPs (25 mM) và 0,5  $\mu\text{L}$  Herculase polymerase. Chuyển 37  $\mu\text{L}$  thể tích mastermix này vào từng giếng có sẵn 10  $\mu\text{L}$  sản phẩm đã làm đầy adapter và thêm vào 1,5  $\mu\text{L}$  môi indexing PCR P5 và P7 vào từng ống tương ứng trên đĩa mẫu. Điều kiện của phản ứng indexing PCR như sau: 95°C/2 phút, n chu kỳ của (95°C/20s, 60°C/20s, 72°C/30s), 72°C/3 phút. Trong đó, số chu kỳ n được xác định bằng phản ứng định lượng sản phẩm làm đầy adapter ở phía trên. Sản phẩm indexing PCR được kiểm tra trên gel agarose 2% và sau đó được tinh sạch bằng SPRI như đã mô tả ở trên.

### 2.2.3.6. Gộp chung các sản phẩm indexing PCR

Các sản phẩm indexing PCR sẽ được định lượng bằng phương pháp đo quang phổ bằng máy Nanodrop. Dựa trên kết quả đo, thực hiện tính toán thể tích cần sử dụng để mỗi mẫu có được khoảng 200 ng sản phẩm indexing PCR. Sau đó, chuyển các mẫu với thể tích đã được tính toán vào chung một ống Eppendorf 0,5 mL.

### 2.2.3.7. Làm giàu thư viện bằng phản ứng lai - bắt giữ

Để thực hiện phản ứng lai - bắt giữ, trước hết cần cố định DNA bait trên M-270 Streptavidin bead, theo các bước sau:

- Chuẩn bị một ống Eppendorf 0,5 mL bao gồm: 3  $\mu$ L DNA bait, 7  $\mu$ L H<sub>2</sub>O, và 10  $\mu$ L đệm BWT (1X). Ủ ở 95°C trong 1 phút và ngay lập tức đưa lên đá sau ủ.
- Chuẩn bị một ống Eppendorf 0,5 mL khác bao gồm: 7  $\mu$ L hạt từ để lên giá từ. Rửa hạt từ với lần lượt 200  $\mu$ L đệm BWT (1X) và 200  $\mu$ L đệm TET (1X). Sau đó, rửa giải (elute) hạt từ với 50  $\mu$ L đệm TET và đặt trực tiếp lên đá.
- Chuyển toàn bộ 20  $\mu$ L thể tích DNA bait vào ống chứa hạt từ ở trên và ủ ở nhiệt độ phòng trong vòng 20 phút. Rửa hai lần với dung dịch đệm BWT (1X) và để trên giá từ trong 2 phút và loại bỏ dịch huyền phù. Tiếp tục rửa hai lần với đệm BWT (1X) và sau mỗi lần đều ủ ở 40°C trong 2 phút. Rửa giải hỗn hợp DNA bait và hạt từ trong 50  $\mu$ L đệm TET.

Chuẩn bị hỗn hợp dung dịch lai với thể tích 52  $\mu$ L bao gồm: 17,8  $\mu$ L sản phẩm gộp chung (2  $\mu$ g); 0,75  $\mu$ L BO4\_CSH (200  $\mu$ M); 0,75  $\mu$ L BO6\_CSH (200  $\mu$ M); 0,75  $\mu$ L BO8.P5.part1.R (200  $\mu$ M); 0,75  $\mu$ L BO10.P5.part2.R (200  $\mu$ M); 5,2  $\mu$ L agilent blocking agent (10X) và 26  $\mu$ L agilent hybridization buffer (2X). Thể tích của từng thành phần phản ứng có thể được thay đổi dựa theo tỷ lệ thể tích của sản phẩm gộp chung. Ủ dung dịch lai trong 3 phút ở 95°C và tiếp tục ủ lắng dung dịch lai ở 37°C trong 30 phút. Đặt hỗn hợp bait, hạt từ và đệm TET lên giá từ và loại bỏ dịch huyền phù. Bổ sung dung dịch lai vào hỗn hợp trên và ủ ở lò lai với nhiệt độ 65°C trong khoảng 2 ngày.

Sản phẩm lai sau khi ủ sẽ được đặt trên giá từ và rửa 4 lần với 200  $\mu$ L đệm BWT (1X). Bổ sung 200  $\mu$ L đệm HWB đã được làm nóng trước ở 60°C vào ống và ủ ở 60°C trong vòng 2 phút. Đặt lên giá từ và loại bỏ dịch huyền phù. Rửa với 200  $\mu$ L đệm BWT (1X), loại bỏ dịch huyền phù. Bổ sung 100  $\mu$ L đệm TET và chuyển hỗn hợp dung dịch sang ống Eppendorf 0,5 mL mới. Chuyển lên giá từ và loại bỏ dịch huyền phù. Thêm 30  $\mu$ L đệm EBT và ủ ở 95°C trong 3 phút và chuyển lên đá. Chuyển toàn bộ phần dịch huyền phù vào ống Eppendorf 0,5 mL mới. Thư viện sau khi làm giàu sẽ được định lượng bằng phản ứng qPCR bao gồm các thành phần như sau: 7,0  $\mu$ L H<sub>2</sub>O, 1,0  $\mu$ L môi Sol\_bridge\_P5 (10  $\mu$ M), 1,0  $\mu$ L môi Sol\_bridge\_P7 (10  $\mu$ M), 10,0  $\mu$ L DyNAmo mastermix (2X) và 1  $\mu$ L thư viện đã được làm giàu. Điều kiện qPCR được thực hiện

như sau: 95°C/ 15 phút, 40 chu kỳ của (94°C/ 30s, 58°C/30s, 72°C/ 1 phút), 72°C/ 10 phút.

### **2.2.3.8. Khuếch đại thư viện DNA**

Phản ứng khuếch đại thư viện DNA có tổng thể tích 50 µL bao gồm các thành phần sau: 26,0 µL H<sub>2</sub>O; 10,0 µL đệm Herculase (5X); 0,5 µL dNTPs (25 mM); mỗi P5 Bridge (10 µM); mỗi P7 Bridge (10 µM); 0,5 µL Herculase polymerase và 10 µL sản phẩm thư viện được làm giàu. Điều kiện của phản ứng khuếch đại như sau: 95°C/ 2 phút, n chu kỳ của (98°C/ 20s, 60°C/ 20s, 72°C/30s) và 72°C/3 phút. Trong đó, n chu kỳ được tính toán dựa trên kết quả định lượng thư viện sau khi làm giàu ở trên. Thư viện sau khi khuếch đại sẽ được tinh sạch bằng SPRI như đã mô tả ở trên. Thư viện sau khi tinh sạch sẽ được mang đi kiểm tra nồng độ và kích thước bằng DNA chip trước khi được mang đi giải trình tự.

### **2.2.4. Giải trình tự hệ gen biến thể ty thể và tiền xử lý dữ liệu**

Thư viện DNA được giải trình tự trên hệ thống giải trình tự gen thế hệ mới Illumina với đoạn đọc hai đầu (pair-end) có kích thước 150 bp. Công cụ FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>) sẽ được sử dụng để đánh giá chất lượng dữ liệu thông qua file định dạng fastq sau khi giải trình tự và được biểu diễn dưới dạng các đồ thị bằng công cụ MultiQC [45]. Tiếp đó, công cụ như Trimmomatic [46] sẽ được dùng để loại bỏ những trình tự không cần thiết hoặc có thể làm nhiễu kết quả như: trình tự adapter, nhưng đoạn trình tự có chất lượng đọc kém, những đoạn không bắt cặp (singleton) với kiểu giải trình tự pair-end và những đoạn trình tự gây nhiễu như đoạn lặp A hoặc lặp T. Các đoạn đọc còn lại sẽ được dóng hàng với trình tự tham chiếu của ty thể revised Cambridge Reference Sequence (rCRS) trên GenBank (NC\_012920.1) sử dụng công cụ BWA [47]. File đầu ra của công cụ này dưới định dạng SAM sẽ được chuyển thành file có định dạng nhị phân BAM bằng công cụ Samtools (<http://www.htslib.org>) để tiến hành các nghiên cứu tiếp theo. Sau khi dóng hàng, các đoạn lặp lại tạo ra trong quá trình xây dựng thư viện sẽ được đánh dấu và loại bỏ bằng công cụ Picard 1.42 (<https://github.com/broadinstitute/picard>). Cuối cùng, khung trình tự (consensus sequence) của hệ gen ty thể của từng mẫu sẽ được xây dựng sử dụng công cụ iVar 1.4 [48].

### **2.2.5. Phân tích hệ gen biến thể ty thể**

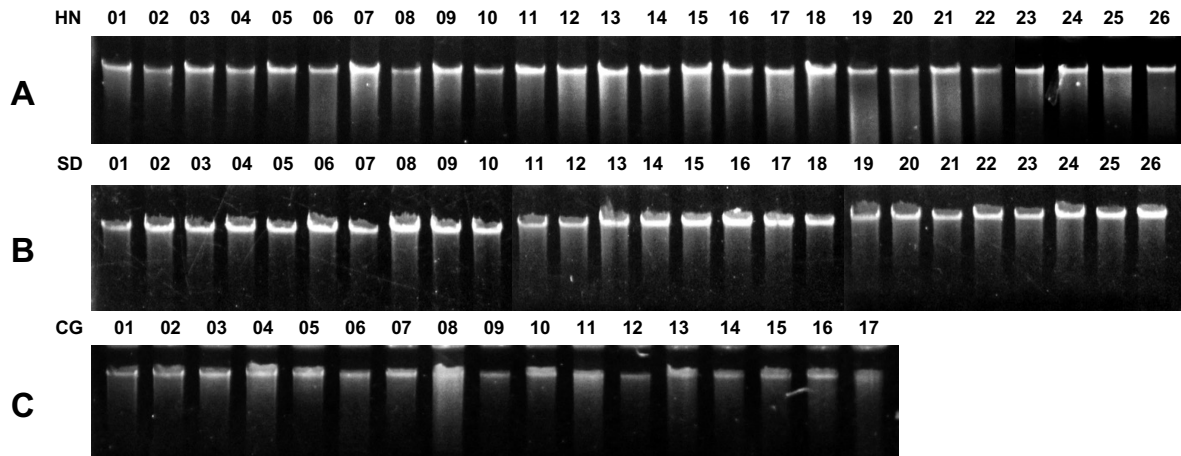
Các biến thể trên mtDNA sẽ được phát hiện bằng công cụ Haplogrep2 [49] với trình tự tham chiếu rCRS trên GenBank (NC\_012920.1) và cây phân loại Phylotree 17 [50]. Những biến thể thu được sẽ được mang đi đối chiếu với cơ sở dữ liệu biến thể hệ gen ty thể MITOMAP (<https://www.mitomap.org/MITOMAP>) để xác định các biến thể mới cũng như những biến thể đã được báo cáo có liên quan tới bệnh ty thể. Ảnh hưởng

của các điểm biến thể trên vùng mã hoá lên cấu trúc và chức năng của protein được dự đoán *in silico* bằng các công cụ dự đoán tin sinh, bao gồm SIFT [51], Polyphen2 [52], PROVEAN [53], MutPred2 [54] và MitImpact [55]. Với các phân tích so sánh, chúng tôi thu thập 784 trình tự hệ gen ty thể (mtDNA) của 27 dân tộc Việt Nam thuộc các ngữ hệ Hán - Tạng, Nam Á, Nam Đảo, Mông - Miên, Thái - Kadai của nghiên cứu trước đây [32, 44] và 1210 trình tự mtDNA từ 47 quần thể dân tộc Hán - Tạng ở các nước hoặc vùng lân cận, bao gồm Thái Lan, Myanmar, Trung Quốc, Đài Loan, Tây Tạng, Nepal và Ấn Độ [20, 30, 31, 56-67] (Phụ lục 2). Tần số của các biến thể trên mtDNA của các dân tộc trong nghiên cứu được so sánh với nhau và để tìm ra các biến thể khác biệt có ý nghĩa thống kê sử dụng package “stats” trong R v4.2.2 [68]. Kiểm định Fisher’s exact (2 phía) được sử dụng để so sánh tỷ lệ tần số biến thể theo cặp giữa các dân tộc trong nghiên cứu. Kiểm định Mann-Whitney được sử dụng để so sánh sự phân bố của các biến thể trên các vùng gen khác nhau cũng như sự khác biệt giữa các dân tộc thuộc ngữ hệ khác nhau. Mức ý nghĩa được lấy với  $\alpha = 0,05$ . Các đồ thị được biểu diễn bằng package “ggplot2” trong R.

## CHƯƠNG 3. KẾT QUẢ VÀ THẢO LUẬN

### 3.1. KẾT QUẢ TÁCH CHIẾT DNA TỔNG SỐ

DNA tổng số của 69 cá thể thuộc 3 dân tộc Công, Hoa và Sán Dìu sau khi tách chiết được điện di kiểm tra chất lượng trên gel agarose 0,8% (Hình 3.1). Kết quả điện di cho thấy các băng DNA tổng số sáng rõ và sắc nét.



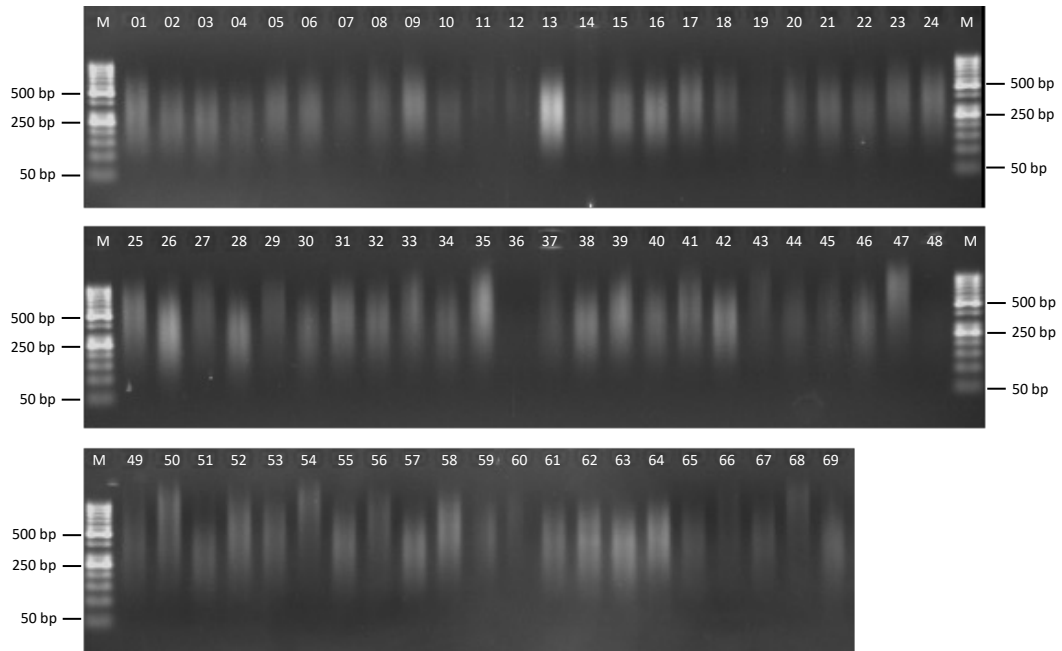
**Hình 3.1. Điện di đồ DNA tổng số của 69 cá thể thuộc 3 dân tộc Công, Hoa và Sán Dìu trên gel agarose 0,8%.**(A) Dân tộc Hoa; (B) Dân tộc Sán Dìu; (C) Dân tộc Công.

Tất cả các mẫu DNA sau khi điện di kiểm tra sẽ được tiếp tục xác định nồng độ và độ tinh sạch bằng máy đo quang phổ Nanodrop<sup>TM</sup> One. Kết quả cho thấy các mẫu DNA đều có nồng độ cao (30,2 - 164,8 ng/ $\mu$ L) và giá trị Optical density (OD) để chỉ độ tinh sạch ở mức tốt (1,7 - 1,9) (Phụ lục 1). Kết quả điện di cũng như đo quang phổ đều cho thấy các mẫu DNA tổng số có chất lượng tốt và sẽ được tiến hành pha loãng về nồng độ 10 ng/ $\mu$ L để phục vụ cho quá trình làm đứt gãy DNA tổng số ở bước tiếp theo.

### 3.2. KẾT QUẢ THIẾT LẬP VÀ GIẢI TRÌNH TỰ THƯ VIỆN HỆ GEN TY THỂ

#### 3.2.1. Kết quả cắt phân đoạn DNA tổng số

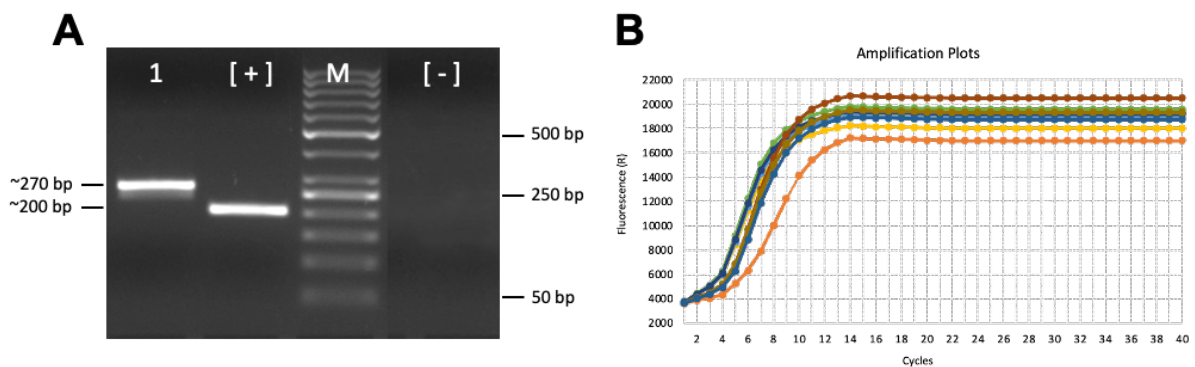
DNA tổng số sau khi pha loãng sẽ được tiến hành cắt phân đoạn bằng máy siêu âm Bioruptor-ucd-200 TM-EX (Diagnode). Sản phẩm sau khi cắt phân đoạn được điện di kiểm tra trên gel agarose 2% (Hình 3.2). Kết quả điện di cùng với thang DNA chuẩn 50 bp cho thấy DNA tổng số được cắt phân đoạn thành công, có kích thước rơi vào khoảng 200 bp đến 400 bp, tạo nguyên liệu đầu vào cho các thí nghiệm tiếp theo bao gồm tạo đầu bằng, gắn và làm đầy adapter.



**Hình 3.2.** Điện di đồ DNA tổng số của 69 mẫu sau khi cắt phân đoạn trên gel agarose 2%. M: thang DNA chuẩn 50 bp; 01 – 69: mẫu DNA 01 đến 69.

### 3.2.2. Kết quả làm đầy adapter

Quá trình làm đầy adapter được thực hiện trên 69 mẫu DNA được làm đứt gãy ở trên cùng với một đối chứng dương và một đối chứng âm. Để kiểm tra kết quả làm đầy, đối chứng dương và đối chứng âm sẽ được điện di kiểm tra trên gel agarose 2% cùng với thang DNA chuẩn 50 bp và một sản phẩm đối chứng không được gắn adapter có kích thước khoảng 200 bp (Hình 3.3A). Kết quả điện di cho thấy mẫu đối chứng dương đã được gắn và làm đầy adapter thành công với kích thước lớn hơn kích thước mẫu đối chứng chưa được gắn adapter cũng như mẫu đối chứng âm không bị tạp nhiễm.



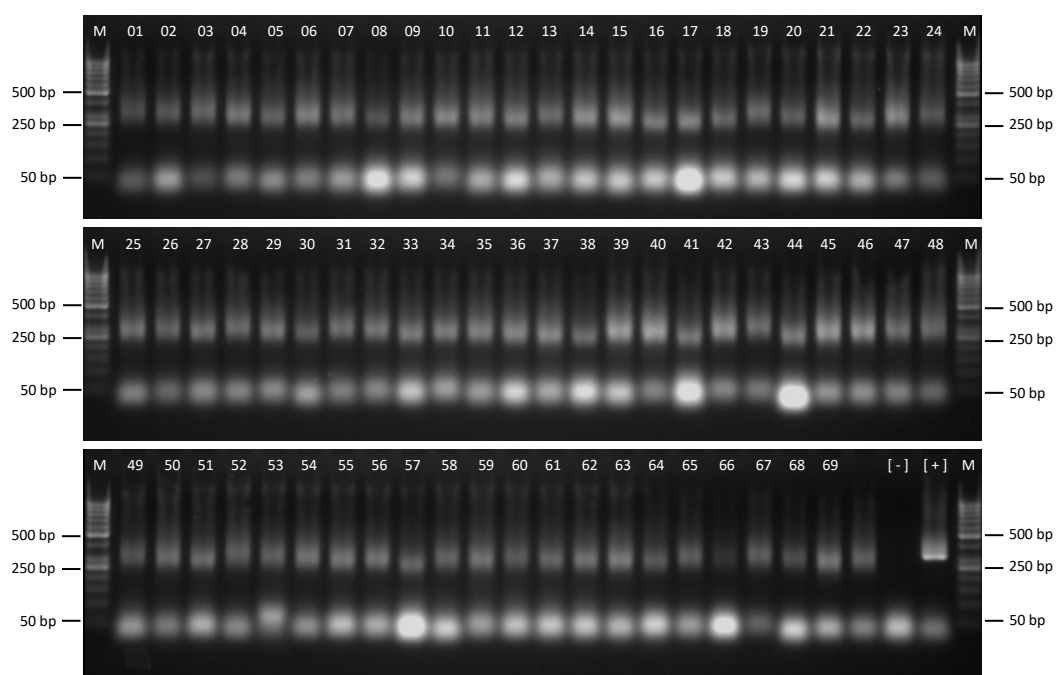
**Hình 3.3.** Kết quả làm đầy adapter. (A) Điện di đồ đối chứng dương, đối chứng âm và mẫu đối chứng sau khi làm đầy adapter trên gel agarose 2% cùng với thang DNA chuẩn 50 bp; (B) Kết quả qPCR định lượng sản phẩm làm đầy adapter.

Sau đó, 8 mẫu bất kỳ sau làm đầy adapter sẽ được sử dụng để chạy qPCR với 2 lần lặp lại cùng các mẫu tiêu chuẩn để xác định được số chu kỳ cần phải sử dụng trong

phản ứng indexing PCR tiếp theo. Hình 3.3B mô tả kết quả đại diện của phản ứng qPCR với cài đặt “denaturing curve” cho kết quả xác định số chu kỳ cho phản ứng indexing PCR là 13 chu kỳ.

### 3.2.3. Kết quả indexing PCR

Phản ứng indexing PCR được tiến hành với số chu kỳ được ghi nhận ở thí nghiệm trên. Kết quả điện di kiểm tra chất lượng indexing PCR trên gel agarose 2% được thể hiện ở hình 3.4. Kết quả cho thấy sản phẩm indexing PCR xuất hiện một dải băng tập trung ở vùng 200 - 400 bp giống với lý thuyết. Sau đó, các sản phẩm này sẽ được mang đi kiểm tra chất lượng bằng máy đo quang phổ Nanodrop™ One. Kết quả cho thấy tất cả các sản phẩm indexing đều có nồng độ cao và đủ tiêu chuẩn cho các bước tiếp theo.



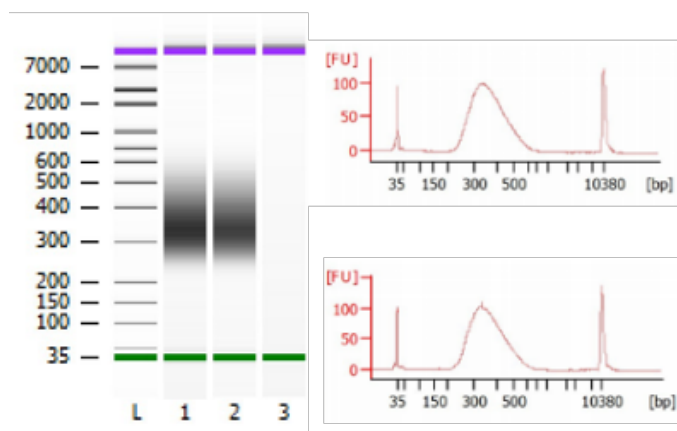
**Hình 3.4. Điện di đồ sản phẩm indexing PCR trên gel agarose 2%.** M: thang DNA chuẩn 50 bp; 01 – 69: mẫu DNA 01 đến 69; “-”: đối chứng âm; “+”: đối chứng dương.

### 3.2.4. Kết quả đánh giá chất lượng thư viện DNA sau tinh sạch bằng DNA chip

Các sản phẩm indexing PCR sau khi kiểm tra nồng độ sẽ được gộp chung thành một sản phẩm duy nhất sao cho khối lượng gộp chung của mỗi mẫu vào khoảng 200 ng. Sản phẩm gộp chung này sẽ được sử dụng để làm giàu thư viện DNA thông qua phản ứng lai - bắt giữ với DNA bait và hạt từ. Thư viện sau khi làm giàu sẽ được định lượng bằng phản ứng qPCR và cho ra kết quả khoảng 9 chu kỳ cho thí nghiệm khuếch đại thư viện. Sau đó, thư viện này sẽ được mang đi tinh sạch và được đánh giá chất lượng, đo chính xác nồng độ của thư viện DNA, cũng như xác định kích thước (bp) trung bình của các đoạn DNA thông qua DNA chip (Hình 3.5). Kết quả phân tích cho thấy kích thước của thư viện rơi vào khoảng 316 bp và nồng độ phân tử là 25 nmol/L. Kết quả này cho



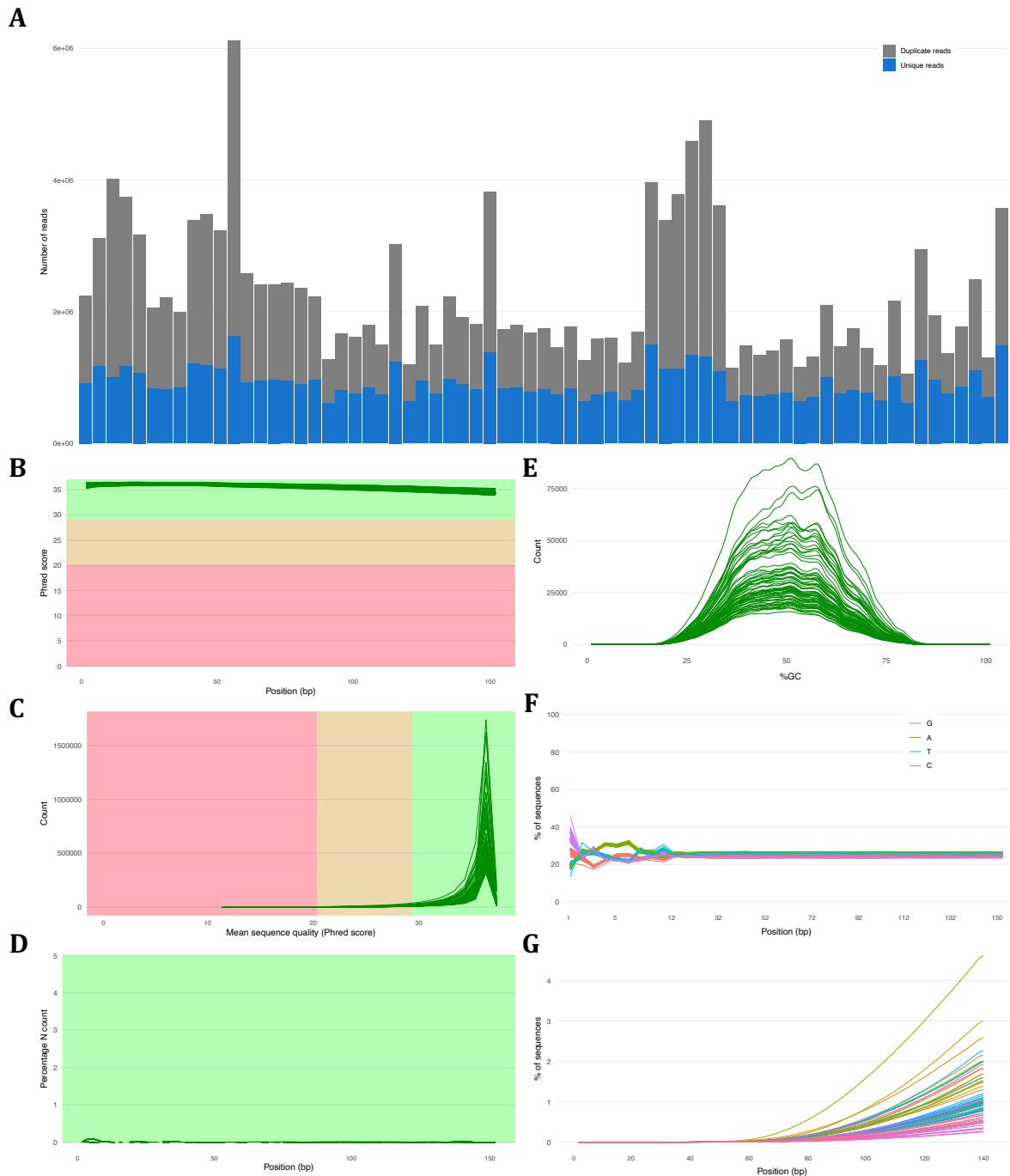
thấy chất lượng của thư viện phù hợp cho đầu vào của hệ thống giải trình tự gen thế hệ mới Illumina. Do đó, có thể thấy thư viện DNA đã được thiết lập và làm giàu tốt, đáp ứng chất lượng để tiến hành giải trình tự.



**Hình 3.5. Kết quả đánh giá chất lượng thư viện bằng DNA chip.** (A) Điện di đồ các mẫu thư viện DNA; (B) Biểu đồ kết quả mẫu thư viện DNA.

### 3.2.5. Kết quả giải trình tự thư viện hệ gen ty thể

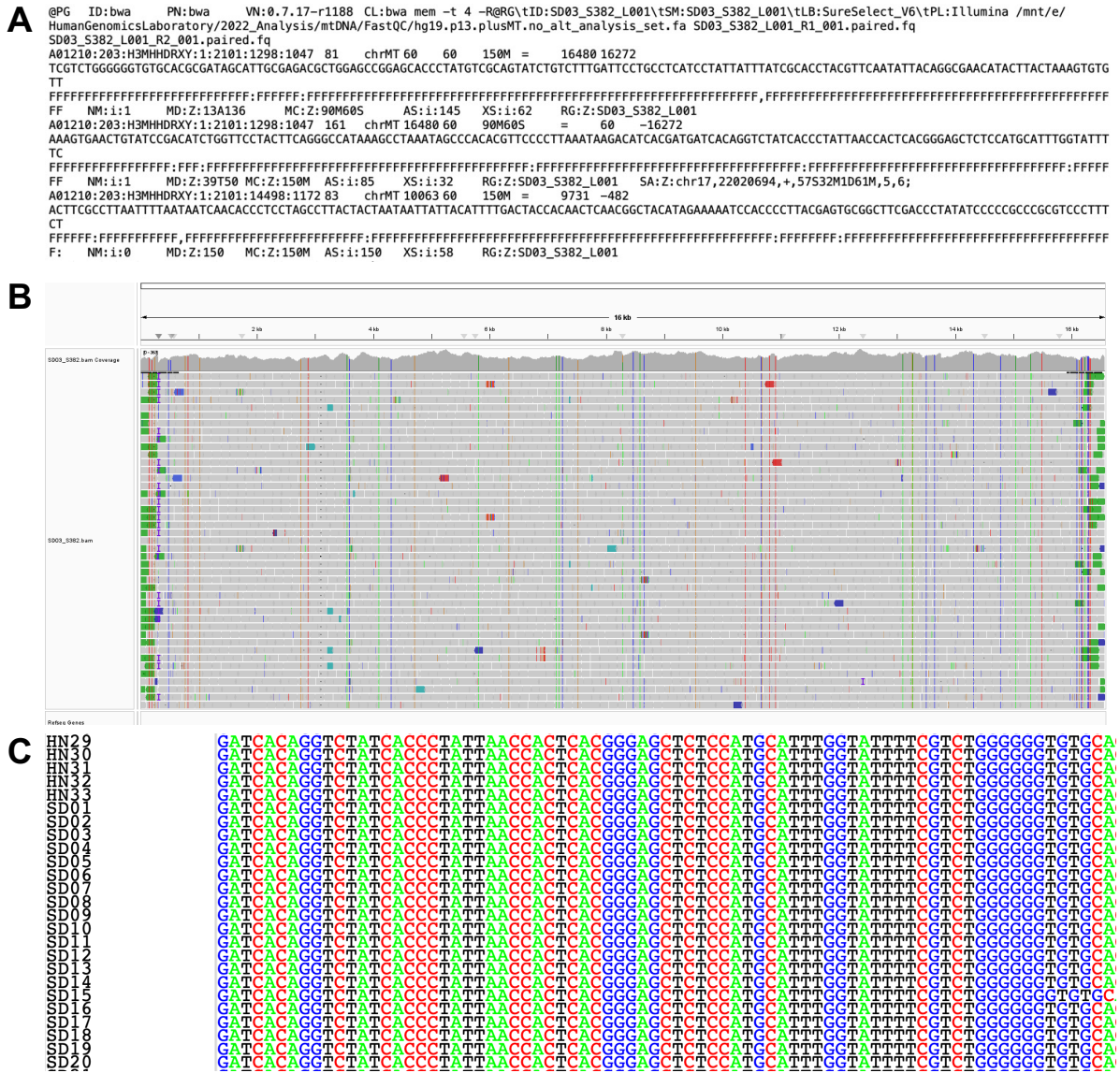
Kết quả giải trình tự trên hệ thống giải trình tự Illumina là tệp dữ liệu thô có định dạng FASTQ và sẽ được đánh giá chất lượng bằng công cụ FastQC. Toàn bộ các tiêu chí đánh giá chất lượng sẽ được biểu diễn bằng công cụ MultiQC (Hình 3.6). Mức độ bao phủ của các đoạn đọc trình tự của 69 mẫu được thể hiện ở hình 3.6A với độ dài của các đoạn đọc đều là 150 bp. Trong đó, độ bao phủ sâu (depth coverage) của các mẫu trải dài từ 92X cho tới 1438X với giá trị trung bình là  $209X \pm 132X$ . Phần lớn, các đoạn đọc đều có chất lượng trung bình trên từng vị trí giảm dần từ đầu cho tới cuối của mỗi đoạn đọc nhưng nhìn chung đều ở ngưỡng tốt với giá trị PHRED vào khoảng 35 (Hình 3.6B). Kết quả đánh giá chất lượng của từng đoạn đọc cũng cho chất lượng tương đối tốt với giá trị trung bình của các đoạn đọc ở ngưỡng điểm PHRED cao (30 - 37) (Hình 3.6C). Hàm lượng nu N ở các đoạn đọc gần như không xuất hiện trên các đoạn đọc (Hình 3.6D). Hàm lượng %GC ở mỗi đoạn đọc được thể hiện ở hình 3.6E, với hàm lượng GC trung bình trên cả 69 mẫu là khoảng 50%. Hình 3.6F cho thấy tỷ lệ xuất hiện của từng nu A/T/G/C qua mỗi vị trí của đoạn đọc. Kết quả cho thấy ở 10 bp đầu tiên có sự khác biệt giữa sự xuất hiện của các nu thành phần do tín hiệu nhiễu từ quá trình tạo thư viện DNA, tuy nhiên ở những phần còn lại của đoạn đọc đều cho thấy sự đồng đều về thành phần của các nu này. Ngoài ra, kết quả đánh giá chất lượng cũng cho thấy rằng hàm lượng adapter xuất hiện ở các mẫu là tương đối cao (Hình 3.6G), nên trước khi sử dụng cho các nghiên cứu tiếp theo thì những trình tự thô này cần phải trải qua quá trình tiền xử lý bằng công cụ Trimmomatic để tạo dữ liệu tinh là các trình tự mtDNA.



**Hình 3.6. Kết quả đánh giá chất lượng trình tự thô của 69 mẫu.** (A) Số lượng các đoạn đọc ở 69 mẫu nghiên cứu. Màu xanh thể hiện các đoạn đọc chỉ xuất hiện ở một mẫu và màu xám thể hiện các đoạn đọc bị trùng nhau ở các mẫu; (B) Chất lượng trung bình trên từng vị trí của đoạn đọc; (C) Chất lượng trung bình của từng đoạn đọc; (D) Hàm lượng nu N ở từng vị trí trên đoạn đọc; (E) Hàm lượng GC của từng đoạn đọc; (F) Hàm lượng các nucleotide A, T, G, C trên từng vị trí của đoạn đọc; và (G) Hàm lượng adapter xuất hiện trong thư viện.

Tất cả các file trình tự mtDNA thô dưới định dạng FASTQ sau khi được đánh giá bằng công cụ FastQC sẽ được tiến hành loại bỏ những đoạn trình tự kém chất lượng

(Phred <30), adapter, các đoạn không bắt cặp,... bằng phần mềm Trimmomatic để đảm bảo cho tính chính xác cho các phân tích sau này. Sau đó, những đoạn đọc (read) từ các file FASTQ này sẽ được dóng hàng với trình tự tham chiếu rCRS bằng công cụ “bwa mem” với kết quả đầu ra thường được lưu dưới định dạng file SAM (Hình 3.7A). Đây là một loại file định dạng lưu lại các ký tự thông tin liên quan đến quá trình dóng hàng và các thông số của các đoạn đọc được phát hiện có vị trí tương ứng trong hệ gen tham chiếu. Cụ thể, file này có chứa 11 cột khác nhau, bao gồm cột 1 là tên đoạn truy vấn, cột 2 là các cờ nhị phân, cột 3 là tên đoạn tham chiếu, cột 4 là vị trí của base khớp đầu tiên, cột 5 là chất lượng dóng hàng, cột thứ 6 là mã CIGAR, cột thứ 7 là tên đoạn tham chiếu được dóng hàng của đoạn đọc tiếp theo (kí hiệu “=” khi trùng tên với đoạn tham chiếu ở cột 3), cột thứ 8 là vị trí của đoạn này, cột thứ 9 là độ dài đoạn khuôn (template), cột thứ 10 là trình tự của đoạn và cột thứ 11 là chất lượng base. Thông thường, các file SAM này sẽ chiếm nhiều khoảng trống trong hệ thống lưu trữ nên dữ liệu dóng hàng thường được chuyển trực tiếp thành các file dữ liệu nhị phân với định dạng BAM trước khi tiến hành các bước tiếp theo. File BAM này sẽ được tiếp tục xử lý bằng công cụ Samtools với các cài đặt như sort, fixmate, rmdup, flagstat và merge. Ngoài ra, công cụ Picard cũng được sử dụng để xác định, đánh dấu và loại bỏ những trình tự lặp lại ở file BAM sinh ra do quá trình khuếch đại PCR và có thể gây ra các kết quả phát hiện biến thể dương tính giả. Các file BAM sau khi xử lý có thể được quan sát bằng phần mềm IGV (Integrative Genomics Viewer). Hình 3.7B thể hiện các đoạn đọc trình tự của mẫu đại diện SD03 được quan sát trên phần mềm IGV với các biến thể được thể hiện dưới các màu khác nhau. Các tệp tin này sau đó được sử dụng để tạo thành 69 trình tự mtDNA hoàn chỉnh bằng công cụ iVar, phục vụ cho các nghiên cứu chuyên sâu tiếp theo (Hình 3.7C).



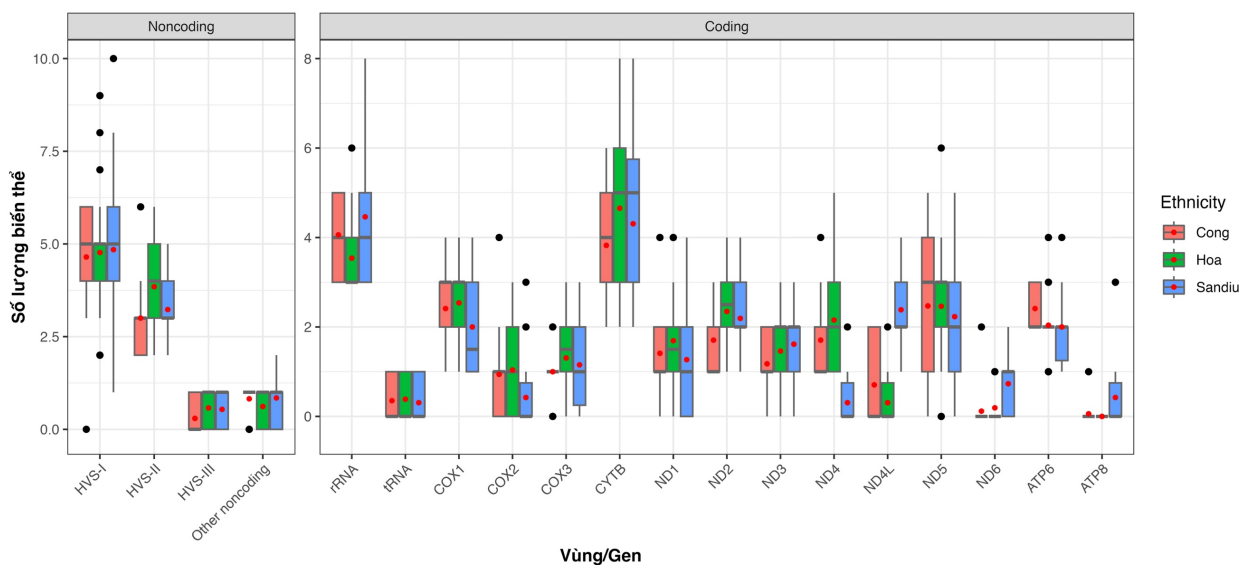
**Hình 3.7. Kết quả lắp ráp trình tự dưới dạng BAM sử dụng phần mềm IGV. (A)** Kết quả lắp ráp trình tự toàn bộ hệ gen ty thể của mẫu SD03 dưới dạng file SAM; **(B)** Kết quả lắp ráp trình tự toàn bộ hệ gen ty thể của mẫu SD03 dưới dạng file BAM; **(C)** Kết quả xây dựng trình tự mtDNA hoàn chỉnh của một số mẫu nghiên cứu dưới dạng FASTA được quan sát bằng phần mềm Sequotron.

**3.3. ĐẶC ĐIỂM DI TRUYỀN VÀ CÁC BIẾN THỂ TRÊN HỆ GEN TY THỂ DÂN TỘC CÔNG, HOA VÀ SÁN DÌU**

**3.3.1. Đặc điểm chung**

Các trình tự mtDNA của 69 cá thể sau khi lắp ráp hoàn chỉnh sẽ được so sánh với trình tự tham chiếu rCRS bằng phần mềm HaploGrep2 và cây phân loại Phylotree 17. Kết quả cho thấy số lượng biến thể xảy ra nhiều nhất ở dân tộc Sán Dìu với trung bình  $40,2 \pm 5,2$  biến thể trên một cá thể, theo sau là dân tộc Hoa với  $36 \pm 6,2$  và dân tộc Công với  $33,1 \pm 6,6$  biến thể. Trên vùng không mã hoá, số lượng biến thể tập trung nhiều nhất

ở vùng siêu biến HVS-I với số lượng biến thể trung bình rơi vào khoảng 5 biến thể ở cả 3 dân tộc (Hình 3.8). Vùng này có số lượng biến thể ở các mẫu nghiên cứu rất khác nhau, trong đó cá thể có số lượng biến thể cao nhất là ở dân tộc Sán Dìu với 10 biến thể và thấp nhất là ở dân tộc Công với không biến thể nào được ghi nhận. Trên vùng mã hoá, dân tộc Công và Sán Dìu có số lượng biến thể trung bình xuất hiện nhiều nhất ở vùng rRNA (12S và 16S) với lần lượt  $4,06 \pm 0,83$  và  $4,46 \pm 1,48$  biến thể, và cao hơn một cách đáng kể khi với dân tộc Hoa ( $3,54 \pm 0,81$  biến thể;  $p < 0,05$ ). Dân tộc Hoa có số lượng biến thể nhiều nhất trên gen *CYTB* với  $4,65 \pm 1,90$  biến thể. Số lượng biến thể thấp nhất ở dân tộc Công và Hoa xuất hiện ở gen *ATP8* với chỉ 1 biến thể xuất hiện trên 17 cá thể dân tộc Công và 26 cá thể dân tộc Hoa, trong khi đó, số lượng biến thể thấp nhất ở dân tộc Sán Dìu được tìm thấy ở vùng mã hoá cho các tRNA và gen *ND4L* với chỉ trung bình 0,31 biến thể xuất hiện. Những gen mã hoá còn lại chỉ xuất hiện ít hơn 2,5 biến thể ở cả 3 dân tộc.



**Hình 3.8. Số lượng biến thể trên hệ gen ty thể của dân tộc Công, Hoa và Sán Dìu.** Chấm tròn màu đỏ biểu diễn số lượng biến thể trung bình. Chấm tròn màu đen biểu diễn điểm dữ liệu nằm cách xa các điểm dữ liệu khác. Màu cam đại diện cho dân tộc Công, màu xanh lá cây cho dân tộc Hoa và màu nước biển cho dân tộc Sán Dìu.

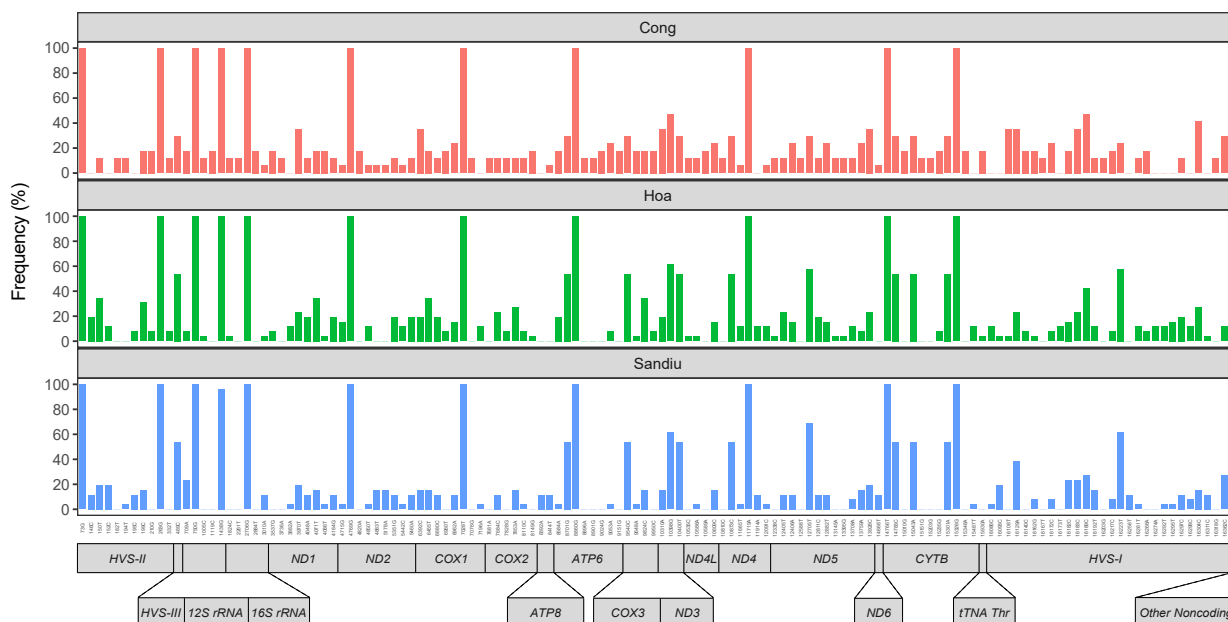
Bên cạnh đó, khi xét theo các vị trí xuất hiện biến thể, chúng tôi xác định được 488 biến thể phân bố ở 484 vị trí trên toàn bộ hệ gen ty thể của 69 mẫu (Bảng 3.1). Trong đó, xuất hiện 4 vị trí đa alen nằm trong vùng siêu biến I và II thuộc vùng không mã hoá, bao gồm các vị trí 146 (T>A/C), 16092 (T>A/C), 16266 (C>A/T), 16293 (A>G/T). Trên cả 3 quần thể, biến thể đồng hoán (462) chiếm số lượng vượt trội và cao gấp 17,7 lần so với các biến thể dị hoán (26). Ở biến thể đồng hoán, loại biến thể đồng hoán phổ biến nhất là T>C, theo sau là A>G, C>T và G>A (91 - 140 biến thể), trong khi đó, ở biến thể dị hoán thì loại C>A là loại phổ biến nhất với 10 biến thể. Ngoài ra, ba

loại biến thể dị hoán là A>T, C>G và T>G chỉ xuất hiện ở ít nhất một trong hai dân tộc là Hoa và Sán Dìu và hoàn toàn không xuất hiện ở dân tộc Cống. Bên cạnh đó, khi so sánh với cơ sở dữ liệu MITOMAP (<https://www.mitomap.org/MITOMAP>), chúng tôi đã nhận thấy hai biến thể 4555T và 9467G xuất hiện lần lượt ở một cá thể thuộc dân tộc Hoa và Sán Dìu, chưa từng báo cáo trên cơ sở dữ liệu này. Trong đó, biến thể 9467G được dự đoán là có khả năng gây bệnh bởi một số phần mềm dự đoán *in silico* như SIFT, Polyphen2 và MillImpact.

**Bảng 3.1. Số lượng biến thể xuất hiện trên dân tộc Cống, Hoa và Sán Dìu**

Dạng biến thể	Toàn bộ quần thể	Cống	Hoa	Sán Dìu
<b>Tổng số</b>	488	160	254	302
<b>Đồng hoán</b>				
<b>A&gt;G</b>	117	36	59	68
<b>G&gt;A</b>	91	34	48	60
<b>C&gt;T</b>	114	38	55	70
<b>T&gt;C</b>	140	45	79	87
<b>Tổng số</b>	462	153	241	285
<b>Dị hoán</b>				
<b>A&gt;C</b>	4	2	3	3
<b>A&gt;T</b>	3	-	2	2
<b>C&gt;A</b>	10	3	4	6
<b>C&gt;G</b>	1	-	-	1
<b>G&gt;C</b>	2	1	1	2
<b>T&gt;A</b>	4	1	2	2
<b>T&gt;G</b>	2	-	1	1
<b>Tổng số</b>	26	7	13	17

Khi phân tích sâu hơn về hệ gen ty thể của 69 cá thể thuộc 3 dân tộc Cống, Hoa và Sán Dìu, chúng tôi phát hiện được 134 biến thể trong số 488 biến thể kể trên có tần số xuất hiện phổ biến  $\geq 10\%$  ở ít nhất một trong ba dân tộc (Hình 3.9), trong đó 41 biến thể nằm trên vùng điều khiển và 93 biến thể nằm trên vùng mã hóa. Ở những biến thể này, mười một biến thể, gồm hai biến thể thuộc vùng điều khiển (73G và 263G) và chín biến thể thuộc vùng mã hoá (750G, 1438G, 2706G, 4769G, 7028T, 8860G, 11719A, 14766T và 15326G), được tìm thấy ở gần như tất cả 69 cá thể thuộc 3 dân tộc nghiên cứu (tần số  $\sim 100\%$ ). Những biến thể kể trên cũng được tìm thấy ở tần số tương đối cao ( $\geq 90\%$ ) ở các nước Đông Nam Á như Việt Nam, Thái Lan và Lào [34, 35, 37, 69]. Năm biến thể trong những biến thể kể trên (73G, 2706G, 7028T, 11719A, 14766T) cũng được ghi nhận thuộc về nhánh đơn bội M đặc trưng cho khu vực Châu Á [50].

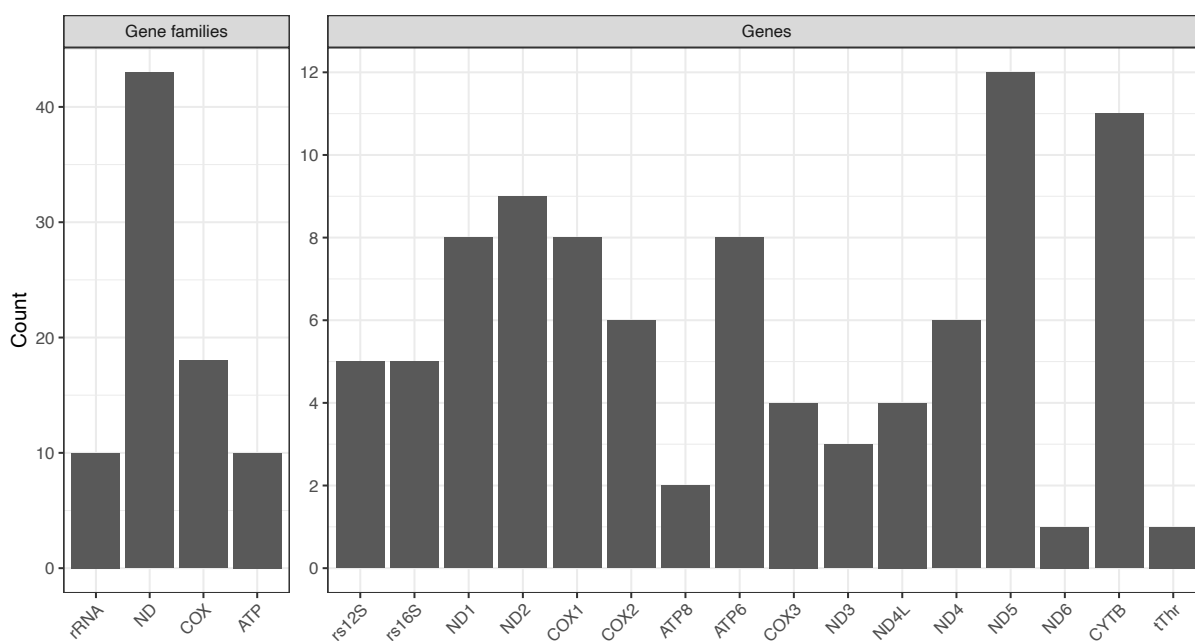


**Hình 3.9. Sự phân bố các biến thể phổ biến trên hệ gen ty thể của ba dân tộc Công, Hoa và Sán Dìu.** Màu cam đại diện cho dân tộc Công, màu xanh lá cây cho dân tộc Hoa và màu xanh nước biển cho dân tộc Sán Dìu.

### 3.3.2. Biến thể trên vùng không mã hoá

Khi phân tích sâu hơn về các biến thể trên vùng không mã hoá, chúng tôi nhận thấy một số biến thể phổ biến trên cả ba nhóm dân tộc nghiên cứu, trong đó có thể kể đến biến thể 16192T trên HVS-I xuất hiện phổ biến ở dân tộc Công, Hoa và Sán Dìu (11,7% - 15,4%). Biến thể này cũng được tìm thấy ở dân tộc Si La (9,68%) nhưng không xuất hiện ở các dân tộc Hà Nhi, Phù Lá, La Hủ và Lô Lô trong ngữ hệ Hán - Tạng. Biến thể 489C trên đoạn siêu biến III (HVS-III) xuất hiện ở 5/17 cá thể người Công (chiếm 29,4%), 14/26 cá thể mỗi dân tộc Hoa và Sán Dìu (chiếm 53,8%) là biến thể đặc trưng cho các nhóm dân tộc Đông Nam Á [70]. Ngoài ra, chúng tôi cũng tìm thấy những biến thể đặc trưng cho từng dân tộc trong mẫu nghiên cứu. Cụ thể, các biến thể 182T và 194T chỉ xuất hiện ở dân tộc Công, biến thể 16173T chỉ xuất hiện ở dân tộc Hoa, và biến thể 16256T chỉ xuất hiện ở dân tộc Sán Dìu với tần số 11,8%. Các biến thể này đều không được tìm thấy hoặc tìm thấy với tần số thấp ở những dân tộc Việt Nam khác [32, 35]. Bên cạnh đó, trên vùng không mã hoá, nghiên cứu hiện tại tìm thấy 11 biến thể được báo cáo có liên quan tới bệnh ty thể.

### 3.3.3. Biến thể trên vùng mã hoá



**Hình 3.10.** Số lượng các biến thể được tìm thấy trên họ gen và gen của hệ gen ty thể trong nghiên cứu.

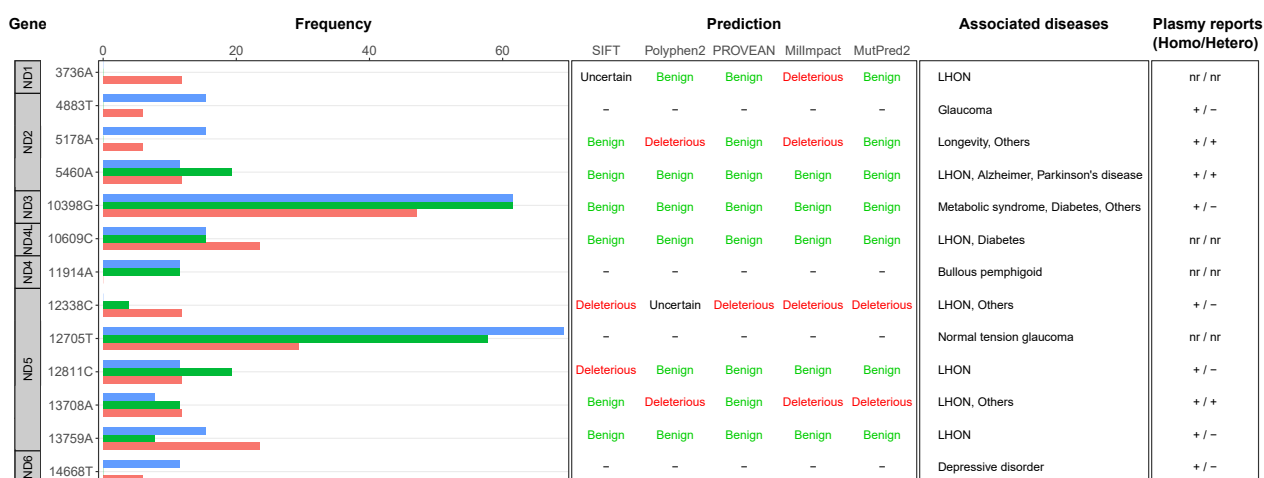
Phần lớn các biến thể (93/135, chiếm 68,9%) trên mDNA được tìm thấy trên vùng mã hoá thuộc ba họ gen và một gen khác nhau, bao gồm họ gen *NADH dehydrogenase subunit (ND)*, họ gen *Cytochrome c oxidase subunit (COX)*, họ gen *ATP Synthase Membrane Subunit (ATP)* và gen *Cytochrome b (CYTB)*. Biến thể nhiều nhất được tìm thấy ở họ gen *ND* với 43 biến thể, theo sau là họ gen *COX* với 18 biến thể, gen *CYTB* với 11 biến thể, họ gen *ATP* và *rRNA* với 10 biến thể và chỉ 1 biến thể được tìm thấy ở vùng mã hoá cho *tRNA* (Hình 3.10). Khi xét đến các gen trong những họ gen kể trên, chúng tôi nhận thấy gen *ND5* và *CYTB* là hai gen có số lượng biến thể lớn nhất với lần lượt 12 và 11 biến thể. Những gen còn lại đều có dưới 9 biến thể với *ND6* và gen mã hoá cho *tRNA-Thr* là hai gen có số biến thể thấp nhất với chỉ một biến thể. Khi so sánh với cơ sở dữ liệu DNA ty thể MITOMAP (<https://www.mitomap.org/MITOMAP>), nghiên cứu đã xác định được 18 biến thể trên vùng mã hoá có tần số phổ biến, đã được ghi nhận có liên quan tới bệnh ty thể.

#### 3.3.2.1. Biến thể trên họ gen *NADH dehydrogenase subunit (ND)*

Họ gen *NADH dehydrogenase subunit* trên hệ gen ty thể là họ gen có số lượng gen lớn nhất với 7 gen, bao gồm *ND1*, *ND2*, *ND3*, *ND4*, *ND4L*, *ND5* và *ND6*. Số lượng biến thể trên mỗi gen này dao động khác nhau, trong đó, nhiều nhất là gen *ND5* với 12 biến thể và thấp nhất là gen *ND6* với chỉ 1 biến thể được tìm thấy. Ngoài ra, nghiên cứu cũng nhận thấy được trên họ gen này có 5 biến thể, bao gồm 3736A, 4820A, 10589A, 10810C và 12588C nằm lần lượt trên các gen *ND1*, *ND2*, *ND4L*, *ND4* và *ND5* có tần số



dao động từ 11,8% đến 17,6% chỉ xuất hiện trên dân tộc Cống mà không hề xuất hiện ở dân tộc Hoa và Sán Dìu. Một số biến thể được báo cáo có liên quan tới bệnh ty thể được thể hiện ở hình 3.11.



**Hình 3.11. Các biến thể phổ biến trên họ gen ND có liên quan tới bệnh ty thể.** Màu cam đại diện cho dân tộc Cống, màu xanh lá cây cho dân tộc Hoa và màu xanh nước biển cho dân tộc Sán Dìu. “+”: xuất hiện dạng đồng nhất (homoplasmy) hoặc dạng không đồng nhất (heteroplasmy); “-”: không xuất hiện dạng đồng nhất (homoplasmy) hoặc dạng không đồng nhất (heteroplasmy); “nr”: chưa được báo cáo.

Trên gen *ND1*, nghiên cứu xác định được tổng cộng 8 biến thể có tần số phổ biến ở ít nhất một trong 3 dân tộc nghiên cứu. Trong đó, biến thể làm biến đổi từ nucleotide C sang T ở vị trí 3970 là biến thể xuất hiện phổ biến nhất ở cả ba dân tộc nghiên cứu (chiếm 19,2 - 35,3%). Nghiên cứu trước đây đã chỉ ra rằng kiểu haplotype giữa biến thể dạng wildtype ở vị trí này (3970C) và 3010G trên vùng mã hoá 16S rRNA xuất hiện với tần suất cao hơn ở người Tây Tạng khi so sánh với những người Hán Trung Quốc ở những vùng thấp hơn ( $p < 0,05$ ) [71]. Kết quả này cho thấy rằng kiểu haplotype này có thể đóng góp trong việc thích nghi với môi trường cao nguyên. Ngoài ra, biến thể không đồng nghĩa 3736A (V144I) chỉ xuất hiện ở dân tộc Cống (chiếm 11,8%) là biến thể duy nhất trên gen *ND1* được báo cáo có liên quan đến bệnh ty thể (Hình 3.11). Cụ thể, biến thể này đã được báo cáo có liên quan tới sự khởi phát của bệnh liệt thần kinh thị giác di truyền Leber (LHON) ở một gia đình người Trung Quốc [72], với chỉ số bảo thủ (conservation index - CI) là 100 ở 44 loài khác nhau.

*ND2* là gen có số lượng biến thể phổ biến thứ hai của họ gen *ND* với 9 biến thể. Hai biến thể đáng chú ý nhất trên gen này có thể kể đến 4883T và 5178A, xuất hiện với tần số tương đối lớn ở dân tộc Sán Dìu (15,4%), nhưng lại xuất hiện ít ở dân tộc Cống (5,9%) và không được ghi nhận ở dân tộc Hoa (Hình 3.11). Trong một nghiên cứu vào năm 2014 trên bệnh glycoma tăng nhãn áp, biến thể 4883T đã được báo cáo có tần suất

xuất hiện cao hơn ở nhóm bệnh nhân khi so sánh với nhóm đối chứng khoẻ mạnh (28,4% vs 15,3%,  $p = 0,002$ ) [73]. Nghiên cứu này cũng chỉ ra rằng những bệnh nhân mang kiểu gen T của biến thể này sẽ biểu hiện giá trị sai lệch trung bình của trường thị giác kém hơn so với người mang kiểu gen C ( $p = 0,009$ ). Biến thể còn lại, 5178A, đã được báo cáo xuất hiện nhiều hơn một cách đáng kể ở người Nhật Bản sống lâu trăm tuổi so với quần thể người bình thường ( $p < 0,05$ ) [74]. Nghiên cứu khác cũng đề xuất rằng những người sở hữu alen A của biến thể này có khả năng kháng lại các bệnh khởi phát ở tuổi trưởng thành hơn những người có alen C [75] và thay đổi sự tiêu thụ rượu hoặc trà xanh dựa vào chức năng của thận eGFR [76, 77]. Ngoài ra cũng có những nghiên cứu chỉ ra sự liên quan giữa hiện tượng già hoá về mặt thể chất của chức năng phổi và chức năng thận ở quần thể người Nhật Bản với biến thể kể trên ( $p < 0,05$ ) [78, 79].

Gen *ND3* là một trong những gen ngắn nhất của họ gen *ND* với kích thước 345 bp. Nghiên cứu hiện tại đã tìm thấy 3 biến thể trên gen này có tần số tương đối cao ở cả ba dân tộc Cống, Hoa và Sán Dìu là 10310A, 10398G và 10400T (chiếm từ 15,4% - 61,5%). Trong đó, biến thể 10398G nằm ở đầu carboxyl của gen *ND3*, được tìm thấy rất phổ biến ở cả 3 dân tộc Cống, Hoa và Sán Dìu ( $\geq 47,1\%$ ) (Hình 3.11), được báo cáo chỉ tồn tại các quần thể sinh sống ở độ cao 2500 - 4000m so với mực nước biển [80]. Biến thể này làm thay đổi amino acid Threonine thành Alanine, từ đó làm giảm sự liên kết trong quá trình phosphoryl hóa oxy hóa và khiến sản sinh ra nhiều nhiệt hơn, từ đó thích nghi với môi trường lạnh ở địa hình cao [80]. Ngoài ra, theo cơ sở dữ liệu ty thể EMPOP (<https://empop.online>), đây là một biến thể được tìm thấy phổ biến ở các nước châu Á như Thái Lan, Trung Quốc, Nhật Bản, tuy nhiên, nó cũng đã được báo cáo có liên quan tới một số bệnh ở những quần thể này như bệnh tiểu đường [81], bệnh Parkinson [82] và bệnh chuyển hoá [83].

Gen *ND4L* là gen ngắn nhất trong 7 gen của họ *ND* với kích thước chỉ khoảng 290 bp. Nghiên cứu hiện tại xác định được 4 biến thể trên gen này, bao gồm 10535C, 10586A, 10589A, 10609C, trong đó 3 biến thể đầu tiên chủ yếu được tìm thấy ở dân tộc Cống (chiếm ít nhất 11,7%) và biến thể 10609C xuất hiện phổ biến ở cả 3 dân tộc trong nghiên cứu (15,4% - 23,5%). Biến thể không đồng nghĩa ở vị trí 10609 này làm thay đổi amino acid methionine (M) thành threonine (T) nhưng được dự đoán không có khả năng gây ra bệnh bằng các công cụ dự đoán như SIFT, PROVEAN, MitImpact, và MutPred2 (Hình 3.11). Tuy nhiên, nghiên cứu trước đây cho rằng biến thể sự thay thế amino acid này tác động tới sự tương tác vốn có với các amino acid xung quanh, làm ảnh hưởng đến độ dài của loop gần điểm biến đổi và cuối cùng làm thay đổi sự ổn định của cả chuỗi xoắn [84]. Nó góp phần thay đổi hình dạng của loop phía dưới, hạn chế sự xâm nhập của các phân tử nước. Biến thể này kết hợp cùng với 10676G được dự đoán có thể làm

gián đoạn con đường vận chuyển proton từ bên trong ra ngoài màng trong quá trình sản xuất ATP. Ngoài ra, biến thể này còn gây ra sự gia tăng các loài oxy phản ứng (reactive oxygen species - ROS) do quá trình siêu phân cực hoặc rối loạn chức năng của các protein phức tạp trong chuỗi hô hấp. Nghiên cứu thực nghiệm trước đây sử dụng tế bào ở người cho thấy đột biến 10609C là yếu tố làm giảm khả năng mắc bệnh đa hồng cầu ở độ cao (OR = 0,391, 95%CI: 0,191 - 0,800,  $p < 0,01$ ) [85]. Nghiên cứu này cũng chỉ ra rằng trong điều kiện thiếu oxy, tế bào không mang biến thể này có khả năng tạo ra nhiều ROS nội bào hơn so với tế bào mang biến thể ( $p < 0,01$ ). Bên cạnh đó, biến thể này cũng được báo cáo ở các bệnh nhân mắc các bệnh LHON [86].

Gen *ND4* có kích thước 1377 bp, trên gen này nghiên cứu đã tìm được 6 biến thể có tần suất xuất hiện phổ biến ở ít nhất 1 trong 3 dân tộc nghiên cứu, tuy nhiên, chỉ có biến thể 11914A được báo cáo có liên quan tới bệnh ty thể (Hình 3.11). Biến thể 11914A xuất hiện với tần số 11,5% ở người Hoa và Sán Dìu nhưng không xuất hiện ở người Cống. Biến thể này được báo cáo có liên quan tới bệnh pemphigoid bong nước (BP) ở quần thể người Đức gồm 262 bệnh nhân mắc BP và 278 người đối chứng khoẻ mạnh (OR = 4,2418; 95% CI: 1,3521 - 13,3077;  $p = 0,0132$ ) [87]. Gen *ND4* cũng được cho là một trong những gen đáng chú ý ở bệnh LHON, đặc biệt là biến thể 11778A được tìm thấy với tần suất cao hơn ở các bệnh nhân mắc LHON người Trung Quốc so với những bệnh nhân không phải người Đông Á ( $p < 0,0001$ ) [88].

Gen *ND5* là gen có kích thước lớn nhất ở họ gen *ND* (1811 bp), với tổng cộng 12 biến thể phổ biến được tìm thấy trên gen này nhưng chỉ 5 biến thể được báo cáo có liên quan đến bệnh trên cơ sở dữ liệu MITOMAP, bao gồm 12338C, 12705T, 12811C, 13708A và 13759A. Biến thể C12705T là biến thể có tần số xuất hiện lớn nhất, chiếm 69,2% ở dân tộc Sán Dìu, 57,7% ở dân tộc Hoa và 29,4% ở dân tộc Cống. Đây được báo cáo là yếu tố di truyền có khả năng bảo vệ khỏi bệnh Glacom nhãn áp [89]. Bốn biến thể còn lại đều được báo cáo có liên quan tới bệnh LHON trên cơ sở dữ liệu MITOMAP (<https://www.mitomap.org/MITOMAP>). Trong đó, biến thể 12338C làm thay đổi amino acid và được dự đoán có khả năng gây bệnh bởi 4 phần mềm khác nhau, bao gồm SIFT, PROVEAN, MitImpact, và MutPred2 (Hình 3.11). Nghiên cứu trước đây cho rằng biến thể này làm giảm biểu hiện mRNA của gen *ND5* và ngăn chặn quá trình trưởng thành của RNA, qua đó, có thể đóng vai trò quan trọng trong việc tăng khả năng biểu hiện lâm sàng của bệnh gây ra bởi đột biến *ND4* 11696A trên một gia đình người Trung Quốc mắc LHON [90].

Gen *ND6* trên ty thể là được cho là điểm nóng của các đột biến gây ra LHON [23]. Tuy nhiên, trong nghiên cứu hiện tại, chỉ có 1 biến thể trên gen *ND6* là 14668T được tìm thấy ở 3 người Sán Dìu (chiếm 11,5%) và 1 người Cống (chiếm 5,9%), được

báo cáo có liên quan tới hội chứng trầm cảm trên cơ sở dữ liệu MITOMAP (Hình 3.11). Biến thể này cũng không được tìm thấy ở hai dân tộc Việt Nam là Giarai và Ê đê sinh sống ở vùng Tây Nguyên [41].

### 3.3.2.2. *Biến thể trên họ gen Cytochrome c oxidase subunit (COX)*

Họ gen *COX* bao gồm 3 gen là *COX1*, *COX2*, và *COX3* đều nằm trên chuỗi nặng của hệ gen ty thể. Nghiên cứu đã tìm thấy lần lượt 8, 6 và 4 biến thể có tần số phổ biến ở ba dân tộc nghiên cứu trên những gen này (Hình 3.8). Một số biến thể đáng chú ý có thể kể đến như hai biến thể 7076G trên gen *COX1* và 7681A trên gen *COX2* đều xuất hiện với tần số 11,7% ở dân tộc Công nhưng lại không xuất hiện ở dân tộc Hoa và Sán Dìu. Khi so sánh với cơ sở dữ liệu biến thể mtDNA MITOMAP, nghiên cứu cũng nhận thấy không có biến thể nào trên hai gen *COX2* và *COX3* được báo cáo có liên quan tới bệnh ty thể. Tuy nhiên, ở gen *COX1*, nghiên cứu đã tìm được hai biến thể (6392A và 6362A) xuất hiện với tần số cao ở cả ba dân tộc ( $\geq 11,5\%$ ) được báo cáo có liên quan tới bệnh ty thể. Trong đó, biến thể 6392A được báo cáo có liên quan đến việc giảm mức độ nặng của bệnh Covid ở quần thể người Hán ở miền Trung Trung Quốc (OR = 0,590, 95% CI: 0,428 - 0,814,  $p = 0,001$ ) [25]. Biến thể còn lại là 6962A (L356G) được báo cáo là đóng vai trò trong quá trình biểu hiện của đột biến 15927A được tìm thấy ở bệnh nhân người Hán (Trung Quốc) mắc bệnh tim mạch vành [91].

### 3.5.2.3. *Biến thể trên họ gen ATP Synthase Membrane Subunit (ATP)*

Họ gen *ATP* bao gồm 2 gen *ATP6* và *ATP8*. Trên gen *ATP8*, chỉ có 2 biến thể (8392A và 8414T) xuất hiện, trong khi đó, trên gen *ATP6* có đến 8 biến thể. Hai biến thể trên gen *ATP8* xuất hiện phổ biến ở dân tộc Sán Dìu (chiếm 11,5%) nhưng lại không xuất hiện ở hai dân tộc còn lại, ngoại trừ biến thể 8414T xuất hiện với tần số thấp ở dân tộc Công (5,9%). Biến thể 8701G trên gen *ATP6* có tần số cao nhất (chiếm 29,4% ở dân tộc Công và 53,8% ở dân tộc Hoa và Sán Dìu), và làm thay đổi amino acid Threonine thành Alanine. Kazuno et al., cho rằng dạng wildtype của biến thể này (8701A) đóng vai trò quan trọng trong sinh lý bệnh Parkinson bằng cách ảnh hưởng tới pH của chất nền ty thể và động lực học canxi nội bào [28]. Tuy nhiên, biến thể này lại được báo cáo không có liên quan tới bệnh Parkinson ở quần thể người Mexico ( $p < 0,05$ ) [92]. Bên cạnh đó, nghiên cứu cũng xác định được bốn biến thể chỉ xuất hiện ở dân tộc Công mà không xuất hiện ở các dân tộc còn lại, bao gồm 8896A và 8901G (chiếm 11,8%) và 9024G và 9151G (chiếm 17,6%). Những biến thể này đều chưa từng được công bố có liên quan hoặc gây ra bệnh ty thể nào trước đây. Ngoài ra, nghiên cứu trên quần thể người Siberia ở Bắc Á cho thấy gen *ATP6* xuất hiện hiện tượng chọn lọc âm tính ưu tiên các biến thể đồng nghĩa hơn các biến thể khác nghĩa [29]. Các biến thể trên gen này

cũng được cho là có liên quan đến việc thích nghi với điều kiện khí hậu khắc nghiệt [29].

#### **3.3.2.4. Biến thể trên gen *Cytochrome b (CYTB)***

Trên gen *CYTB*, nghiên cứu đã tìm được 11 biến thể có tần suất phổ biến ở ít nhất một trong 3 dân tộc Công, Hoa và Sán Dìu. Trong đó, bốn biến thể chỉ được tìm thấy ở dân tộc Công mà không được tìm thấy ở dân tộc Sán Dìu và Hoa, bao gồm: 15010G và 15346A (chiếm 17,6%), 15151G và 15203G (chiếm 11,8%). Ba biến thể đồng hoán (14783C, 15043A và 15301A) không làm thay đổi amino acid được tìm thấy với tần số 29,4% ở dân tộc Công và 53,8% ở dân tộc Hoa và Sán Dìu. Các biến thể này được dự đoán ít ảnh hưởng tới chức năng của protein và ít có khả năng gây bệnh, tuy nhiên, trên cơ sở dữ liệu ClinVar, chúng lại được báo cáo liên quan tới bệnh ung thư vú có tính chất gia đình. Bên cạnh đó, cũng trên gen này, biến thể 14766T làm thay đổi amino acid Threonine thành Isoleucine ở vị trí thứ 7 trên chuỗi polypeptide mã hoá bởi gen *CYTB* được tìm thấy trên tất cả các cá thể ở 3 dân tộc. Biến thể này được dự đoán có khả năng gây bệnh với điểm SIFT là 0,03, tuy nhiên một số phần mềm khác như Polyphen2, CADD, MutPred2, MitImpact đều cho rằng biến thể này không có khả năng gây bệnh. Trên thực tế, biến thể này được ghi nhận là có liên quan tới các bệnh phù phổi ở địa hình cao trên người Ấn Độ [93].

#### **3.3.2.5. Biến thể trên gen mã hoá rRNA**

Trên hệ gen ty thể có hai gen mã hoá cho hai loại rRNA, bao gồm 12S rRNA và 16S rRNA. Bên cạnh 3 biến thể xuất hiện ở gần như tất cả các cá thể thuộc 3 dân tộc nghiên cứu, chúng tôi cũng tìm thấy 7 biến thể khác có tần số tương đối phổ biến (> 11%), bao gồm 709A, 1005A, 1119C nằm trên gen mã hoá 12S rRNA và 1824C, 2261T, 2884T và 3010A nằm trên gen mã hoá 16S rRNA. Trong đó, phần lớn các biến thể này nằm trên dân tộc Công và ít hoặc không xuất hiện ở hai dân tộc Hoa và Sán Dìu, ngoại trừ biến thể 709A xuất hiện ở cả dân tộc Công (chiếm 17,6%) và dân tộc Sán Dìu (chiếm 23,1%) cũng như biến thể 3010A xuất hiện ở 11,5% người Sán Dìu nhưng ít xuất hiện ở hai dân tộc còn lại. Biến thể 3010A được đặc biệt chú ý vì là biến thể định danh một số nhóm đơn bội có liên quan tới bệnh LHON [94]. Ngoài ra, với vị trí nằm gần trung tâm vận chuyển ribosomal peptidyl liên kết với kháng sinh linezolid, Pacheu-Grau và cộng sự [95] đã báo cáo rằng các tế bào cybrid chứa biến thể 3010A có số lượng sản phẩm dịch mã ty thể thấp hơn đáng kể, tỷ lệ COX1/succinate dehydrogenase subunit A thấp hơn và tỷ lệ giữa phức hệ IV với hoạt động của citrate synthase thấp hơn sau khi điều trị bằng linezolid. Biến thể này cũng được dự đoán có làm ảnh hưởng tới cấu trúc thứ cấp cũng như ba chiều của 16S rRNA bằng các phần mềm dự đoán *in silico* [96]. Nhiều nghiên cứu bệnh - chứng cũng đã được thực hiện trên biến thể 3010A và cho

nhiều kết quả khác nhau. Cụ thể, biến thể này được cho rằng không có liên quan tới bệnh tăng nhãn áp góc mở sơ cấp ( $p > 0,05$ ) [97], hội chứng nôn mửa theo chu kỳ ở trẻ em ( $p > 0,05$ ) [98] nhưng lại có liên quan tới các bệnh như hội chứng mệt mỏi mãn tính ( $p < 0,05$ ) [99]. Bên cạnh biến thể này, hai biến thể khác nằm trên vùng gen mã hoá cho 12S rRNA (1119C và 1005C) cũng được báo cáo có liên quan tới bệnh tiểu đường [100] và mất thính lực gây ra bởi aminoglycoside [101].

### 3.4. SO SÁNH THỐNG KÊ BIẾN THỂ

Để có một cái nhìn tổng thể hơn về sự phân bố biến thể cũng như sự giống và khác nhau trong các dân tộc Hán - Tạng Việt Nam, giữa các dân tộc Hán - Tạng và các dân tộc thuộc bốn ngữ hệ khác của Việt Nam cũng như với các dân tộc Hán - Tạng ở các nước lân cận, nghiên cứu thực hiện phân tích sử dụng 3 bộ dữ liệu khác nhau với các dữ liệu biến thể mtDNA đã được công bố trên thế giới. Số lượng trình tự cũng như biến thể tìm được của mỗi bộ dữ liệu được mô tả chi tiết ở bảng 3.2.

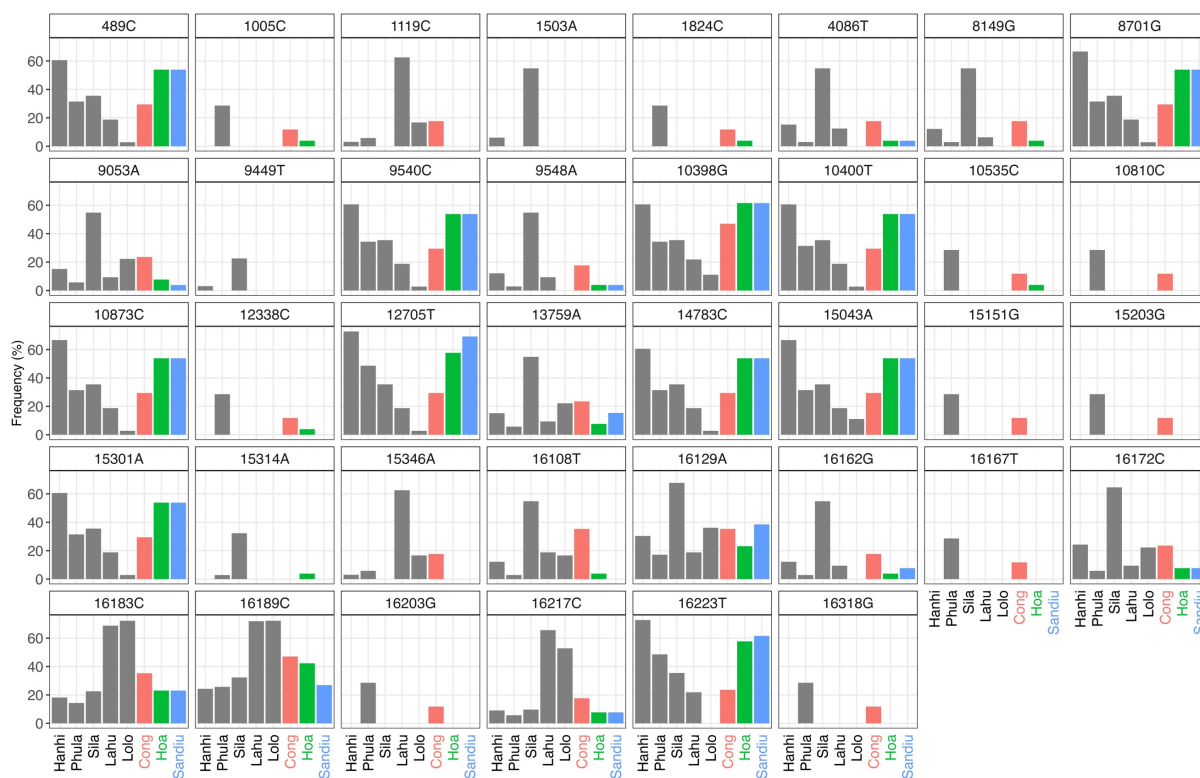
**Bảng 3.2. Dữ liệu biến thể mtDNA được sử dụng trong các phân tích so sánh**

Bộ dữ liệu	Số lượng trình tự sử dụng	Số lượng quần thể	Số biến thể tìm được	Số biến thể khác biệt theo cặp
1	236	8	765	177
2	880	30	1605	468
3	1446	40	2664	1024

#### 3.4.1. Các dân tộc Hán - Tạng Việt Nam

Phân tích so sánh sự phân bố các biến thể trong các quần thể thuộc ngữ hệ Hán - Tạng của Việt Nam sử dụng bộ dữ liệu số 1 ở bảng 3.2, bao gồm 69 mẫu dân tộc Công, Hoa và Sán Dìu trong nghiên cứu hiện tại và 167 mẫu thuộc 5 dân tộc Hán - Tạng của Việt Nam (Hà Nhì, Phù Lá, Si La, La Hủ và Lô Lô). Trong tổng số 765 biến thể tìm được, nghiên cứu xác định được 177 biến thể có sự khác biệt về thống kê giữa các cặp dân tộc sử dụng kiểm định Fisher's exact ( $p < 0,05$ ). Trong đó, dân tộc Lô Lô là dân tộc có số lượng biến thể khác biệt nhiều nhất so với các dân tộc Công, Hoa và Sán Dìu với lần lượt 44, 63 và 54 biến thể khác biệt. Các dân tộc còn lại (Hà Nhì, Phù Lá, Si La, và La Hủ) có ít nhất 10 biến thể khác biệt so với 3 dân tộc trong nghiên cứu, ngoại trừ Sán Dìu chỉ có 4 biến thể khác biệt với dân tộc Hà Nhì. Dân tộc Công có lần lượt 2 và 3 biến thể khác biệt với hai dân tộc Hoa và Sán Dìu, trong khi đó, hai dân tộc này không có bất kỳ điểm biến thể nào có khác biệt về mặt thống kê. Cụ thể, biến thể 16108T xuất hiện nhiều ở dân tộc Công (chiếm 35,2%) nhưng xuất hiện ít ở dân tộc Hoa (chiếm 3,8%) và không xuất hiện ở dân tộc Sán Dìu ( $p < 0,05$ ) (Hình 3.12). Ngược lại, biến thể 16223T

lại xuất hiện rất phổ biến ở Hoa và Sán Diu với tần số lần lượt là 57,7% và 61,5%, nhưng chỉ xuất hiện với tần số 23,5% ở dân tộc Cống ( $p < 0,05$ ).



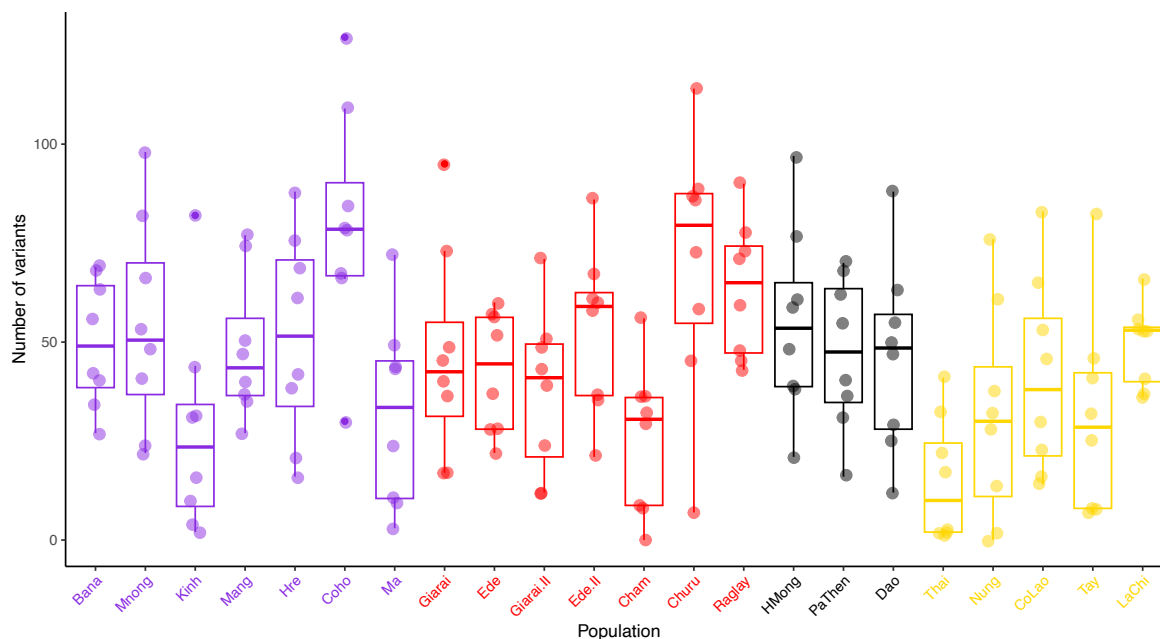
**Hình 3.12. So sánh thống kê một số biến thể nổi bật ở ba dân tộc Cống, Hoa và Sán Diu với các dân tộc khác thuộc ngữ hệ Hán - Tạng ở Việt Nam.** Tần số xuất hiện của các biến thể trên dân tộc Cống, Hoa và Sán Diu lần lượt được kí hiệu bằng các cột màu cam, màu xanh lá cây và màu xanh da trời.

Ngoài ra, kết quả phân tích cũng chỉ ra rằng, có sự khác biệt giữa dân tộc La Hủ và Lô Lô với dân tộc Hoa và Sán Diu ( $p < 0,05$ ). Cụ thể, một số biến thể (489C, 8701G, 9540C, 10398G, 10400T, 10873C, 12705T, 14783C, 15043A, 15301A, 16223T) xuất hiện ít ở La Hủ và Lô Lô nhưng lại xuất hiện nhiều ở Sán Diu và dân tộc Hoa ( $p < 0,05$ ). Nhìn chung, các đa hình này không phổ biến ở các nước Đông Nam Á như Thái Lan và Lào [69]. Ngược lại, các biến thể (1119C, 15346A, 16183C, 16189C, 16217C) lại phổ biến ở La Hủ và Lô Lô (tần số  $\geq 50\%$ ) nhưng ít xuất hiện ở dân tộc Hoa và Sán Diu ( $p < 0,05$ ). Những biến thể này cũng xuất hiện với tần số khoảng 20% ở hai dân tộc Phù Lá và Hà Nhì, tuy nhiên, chúng lại xuất hiện phổ biến ở người Kinh ( $\sim 50\%$ ) [34, 37]. Bên cạnh đó, nghiên cứu cũng tìm thấy mười biến thể 1005C, 1824C, 10535C, 10810C, 12338C, 15151G, 15203G, 16167T, 16203G và 16318G xuất hiện với tần số cao (28,6%) ở dân tộc Phù Lá và mười hai biến thể 1503A, 4086T, 8149G, 9053A, 9449T, 9548A, 13759A, 15314A, 16108T, 16129A, 16162G, 16172C ở quần thể người Si La (chiếm 22,6% - 67,7%) khác biệt có ý nghĩa thống kê với ít nhất 6 trong 7 các dân tộc Hán - Tạng còn lại ( $p < 0,05$ ). Tương tự như vậy, dân tộc La Hủ và Lô Lô cũng sở hữu

lần lượt 15 và 6 biến thể khác biệt. Sự khác biệt giữa chỉ thị mtDNA của các dân tộc thuộc ngữ hệ Hán - Tạng ở Việt Nam cũng đã được quan sát ở nghiên cứu trước đây [102].

### 3.4.2. Các dân tộc thuộc ngữ hệ khác tại Việt Nam

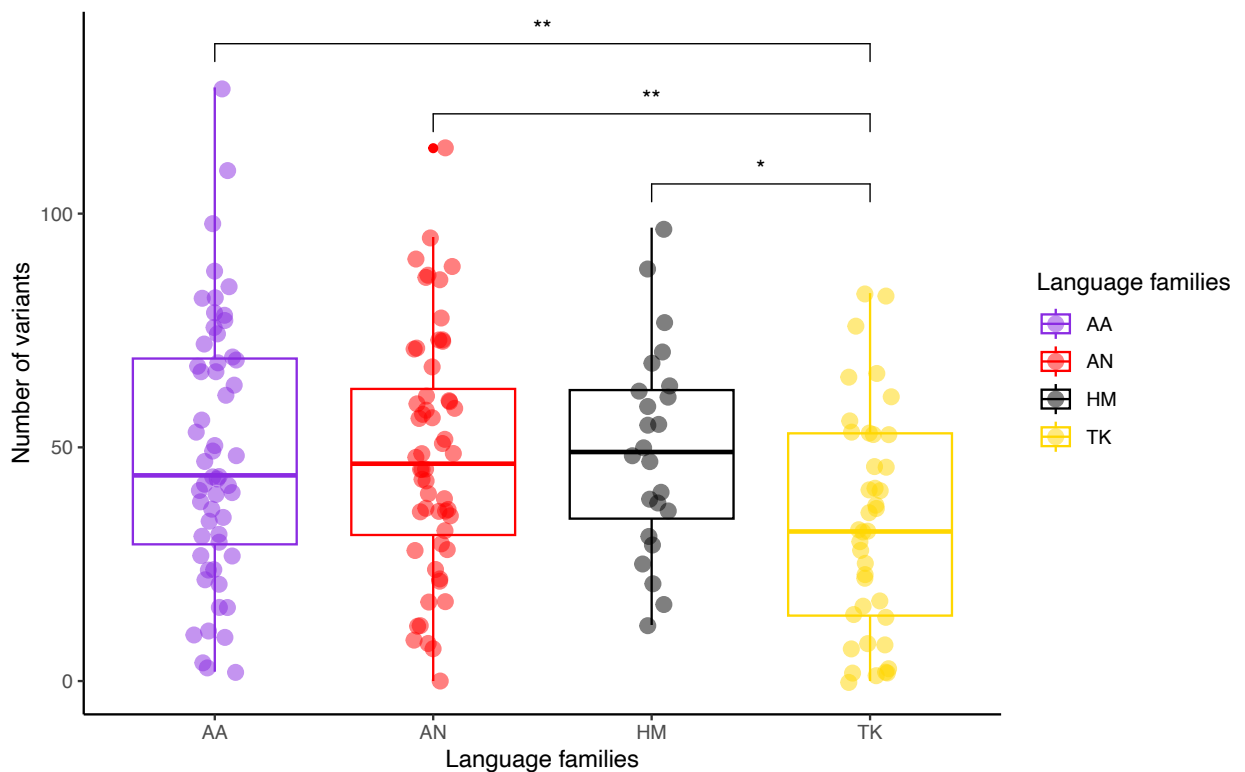
Bên cạnh ngữ hệ Hán - Tạng, Việt Nam còn có sự xuất hiện của 4 ngữ hệ lớn khác, bao gồm Nam Á, Nam Đảo, Mông - Miền và Thái - Kadai, do đó, nghiên cứu cũng thực hiện phân tích sự phân bố của các biến thể trong toàn bộ năm ngữ hệ kể trên bằng bộ số liệu 2 (Bảng 3.2). Cụ thể, bên cạnh các trình tự trong nghiên cứu hiện tại, phân tích sẽ sử dụng thêm 811 trình tự mtDNA thuộc 22 dân tộc của 4 ngữ hệ của Việt Nam, bao gồm ngữ hệ Nam Á (Ba Na, M'Nông, Kinh, Mảng, H'Rê, Cơ Ho, Mạ), Nam Đảo (Giarai, Giarai-II, Ê Đê, Ê Đê-II, Chăm, Chu Ru, Raglay), Mông - Miền (H'Mông, Pà Thên, Dao) và Thái - Kadai (Thái, Nùng, Cờ Lao, Tày, La Chí) (Phụ lục 2). Qua đó, kết quả đã tìm thấy 1605 biến thể ở tất cả các mẫu nghiên cứu, với 468 biến thể được xác định có khác biệt về thống kê giữa từng cặp dân tộc ( $p < 0,05$ ). Dân tộc Cơ Ho thuộc ngữ hệ Nam Á và Chu Ru thuộc ngữ hệ Nam Đảo là hai dân tộc số lượng biến thể trung bình khác biệt nhất so với các dân tộc Hán - Tạng với lần lượt 80,0 và 69,9 biến thể (Hình 3.13). Bên cạnh đó, các dân tộc Thái thuộc ngữ hệ Thái - Kadai, Kinh thuộc ngữ hệ Nam Á và Chăm thuộc ngữ hệ Nam Đảo là những dân tộc có số lượng biến thể khác biệt thấp nhất với chỉ lần lượt 15; 27,5 và 27,8 biến thể. Những dân tộc còn lại đều có ít nhất 30 biến thể khác biệt.



**Hình 3.13. Số lượng biến thể khác biệt so với các dân tộc Hán - Tạng.** Bốn ngữ hệ của Việt Nam bao gồm Nam Á, Nam Đảo, Mông - Miền và Thái - Kadai được kí hiệu lần lượt bằng các màu tím, đỏ, đen và vàng.



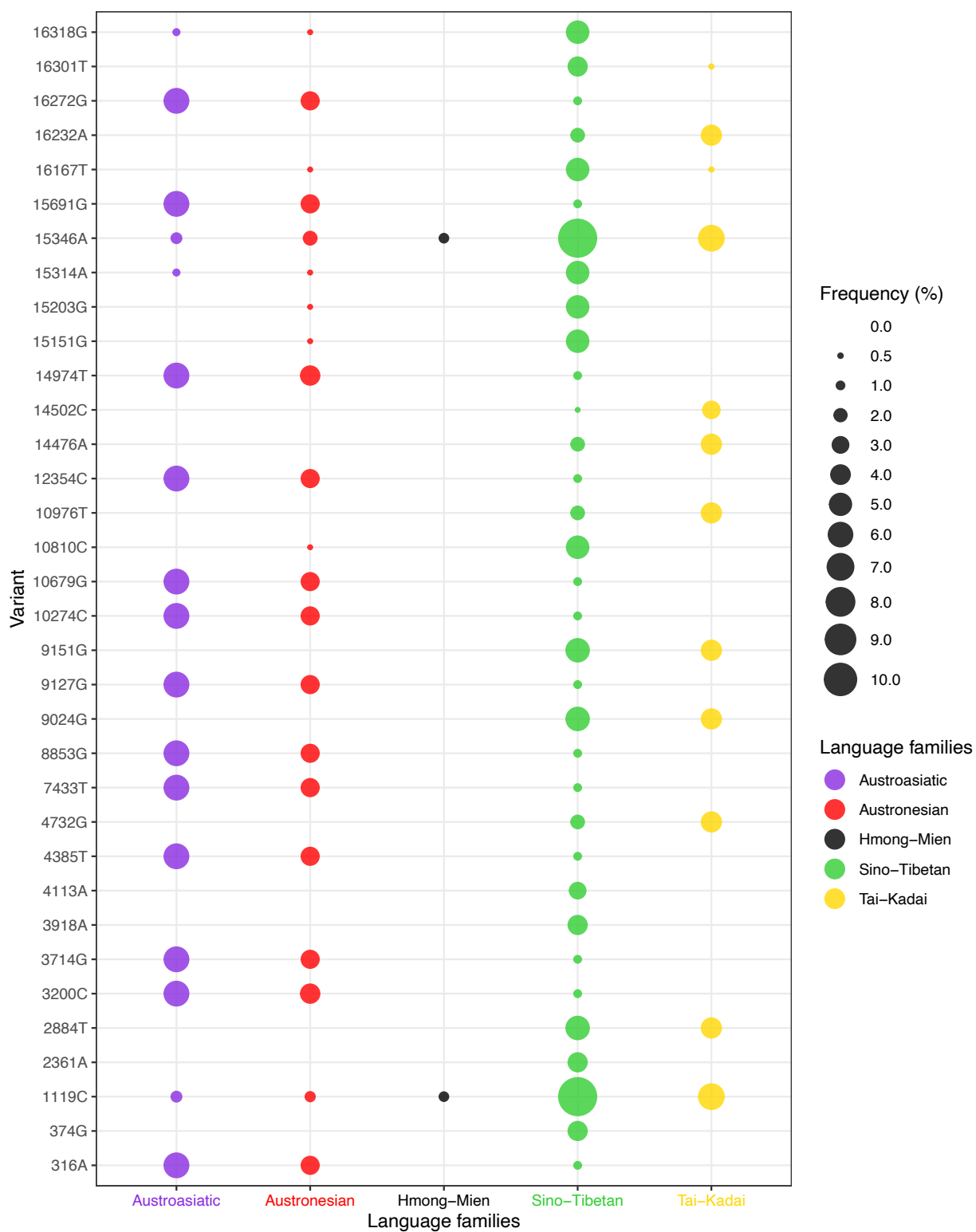
Khi phân tích tổng thể theo năm ngữ hệ lớn, kết quả đã cho thấy số lượng biến thể khác biệt giữa ngữ hệ Hán - Tạng và các ngữ hệ Nam Á, Nam Đảo, Mông - Miền, Thái - Kadai lần lượt là 198, 201, 160 và 162 biến thể. Cụ thể, một dân tộc thuộc ngữ hệ Thái - Kadai có khoảng 25 biến thể khác biệt với một dân tộc Hán - Tạng, tuy nhiên, các dân tộc thuộc ba ngữ hệ còn lại lại có số lượng biến thể khác biệt với các dân tộc Hán - Tạng nhiều hơn một cách đáng kể, rơi vào khoảng 32 - 34 biến thể trên một ngữ hệ ( $p < 0,05$ ) (Hình 3.14).



**Hình 3.14. Số lượng biến thể khác biệt của các dân tộc Hán - Tạng so với các dân tộc thuộc các ngữ hệ khác.** Bốn ngữ hệ của Việt Nam bao gồm Nam Á, Nam Đảo, Mông - Miền và Thái - Kadai được kí hiệu lần lượt bằng các màu tím, đỏ, đen và vàng.

Phân tích chi tiết về 468 biến thể kể trên đã tìm thấy một số những biến thể nổi bật giữa các nhóm ngữ hệ (Hình 3.15). Trong đó, nghiên cứu đã tìm thấy mười hai biến thể ở ngữ hệ Hán - Tạng có tần số cao hơn đáng kể so với các ngữ hệ còn lại ( $p < 0,05$ ). Cụ thể, bốn biến thể 374G, 2361A, 3918A và 4113A tồn tại với tần số khoảng 3% trên toàn bộ ngữ hệ Hán - Tạng nhưng hoàn toàn không xuất hiện ở các ngữ hệ khác (Hình 3.14). Tám biến thể còn lại, bao gồm 1119C, 15346A, 10810C, 15151G, 15203G, 16167T, 16318G, 15314A đều là những biến thể xuất hiện với tần số cao ở một hoặc nhiều dân tộc Hán - Tạng và xuất hiện ít ở những dân tộc thuộc các ngữ hệ còn lại ( $p < 0,05$ ). Cụ thể, hai biến thể 1119C và 15346A tồn tại ở 62,5% cá thể người dân tộc La Hủ, năm biến thể 10810C, 15151G, 15203G, 16167T và 16318G tồn tại ở 28,6% cá thể

dân tộc Phù Lát và 11,8% ở dân tộc Cống, và biến thể 15314A xuất hiện ở 32,3% người dân tộc Si La.



**Hình 3.15. Một số biến thể khác biệt nổi bật giữa các nhóm dân tộc thuộc năm ngữ hệ tại Việt Nam.** Năm ngữ hệ của Việt Nam bao gồm Nam Á, Nam Đảo, Mông - Miền,

Hán - Tạng và Thái - Kadai được kí hiệu lần lượt bằng các màu tím, đỏ, đen, xanh lá cây và vàng.

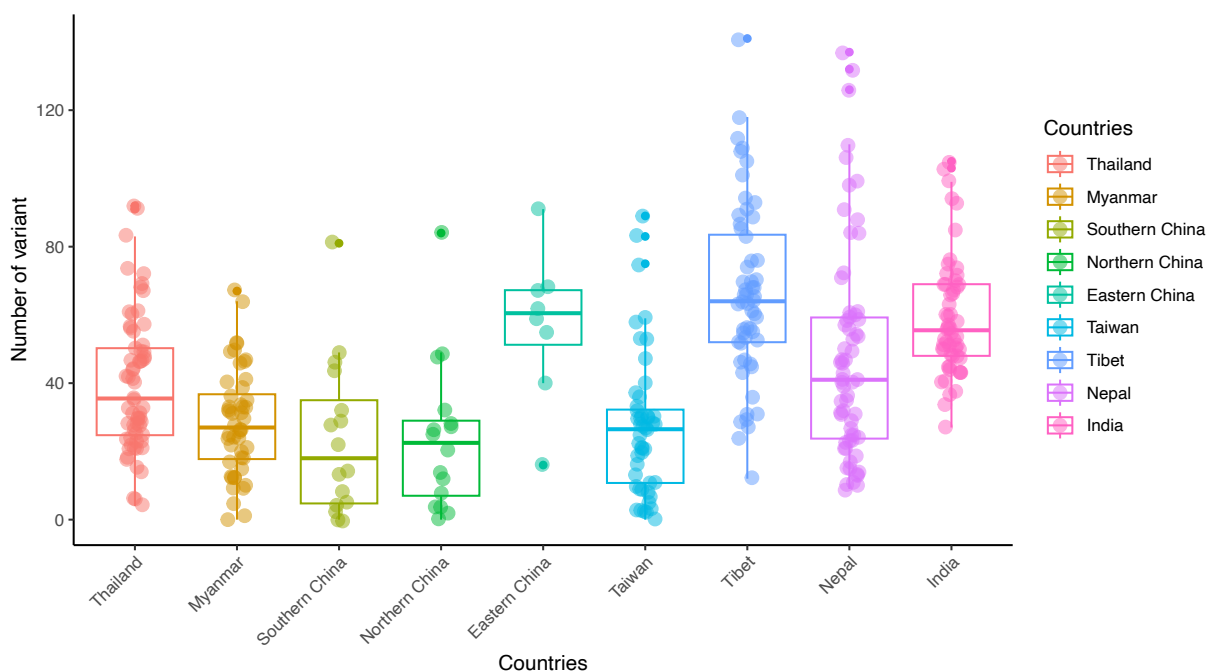
Với sự khác biệt thấp của các dân tộc thuộc ngữ hệ Thái - Kadai và Hán - Tạng đã được chỉ ra ở trên, nghiên cứu cũng đã tìm thấy mười chín biến thể chỉ xuất hiện riêng biệt ở hai ngữ hệ này. Trong đó, mười biến thể, bao gồm 4140T, 4823C, 8793C, 9729T, 12549T, 12976T, 13152G, 15040T, 15071C, 15514C là những biến thể chỉ được chia sẻ giữa dân tộc Cờ Lao (chiếm 17,6%) và dân tộc Sán Dìu (chiếm 3,8%) và hoàn toàn không xuất hiện ở những dân tộc khác trong Việt Nam ( $p < 0,05$ ). Tuy nhiên, khi xét trên toàn bộ nhóm ngôn ngữ thì những biến thể này lại không có ý nghĩa thống kê ( $p > 0,05$ ). Ngoài ra, chín biến thể khác cũng chỉ phân bố ở hai ngữ hệ này và khác biệt so với các ngữ hệ khác ( $p < 0,05$ ), bao gồm 2884T, 4732G, 9024G, 9151G, 10976T, 14476A, 14502C, 16232A, 16301T. Trong đó ba biến thể 2884T, 9024G, 9151G chỉ xuất hiện ở La Chí (25%), Cống (17,6%), La Hủ (12,5%), và Lô Lô (16,7%); bốn biến thể 4732G, 10976T, 14476A, 16232A xuất hiện ở La Chí (25%), Lô Lô (11,1%) và Sán Dìu (3,8%); biến thể 14502C xuất hiện ở Sán Dìu (3,8%), Cờ Lao (17,6%) và Tày (2,1%); và biến thể 16301T xuất hiện ở Hoa (3,8%), Lô Lô (22,2%) và Cờ Lao (2,9%). Nghiên cứu trước đây về mtDNA trên các nhóm người Hán - Tạng và Thái - Kadai cũng cho thấy có sự chia sẻ haplotype giữa các dân tộc thuộc hai ngữ hệ này [44, 102].

Bên cạnh đó, với số lượng biến thể khác biệt trung bình lớn nhất với ngữ hệ Hán - Tạng, nghiên cứu cũng tiến hành phân tích các biến thể xuất hiện ở ngữ hệ Nam Á. Kết quả cho thấy mười ba biến thể 316A, 3200C, 3714G, 4385T, 7433T, 8853G, 9127G, 10274C, 10679G, 12354C, 14974T, 15691G, 16272G, xuất hiện với tần số khoảng 6% và cao hơn hẳn những ngữ hệ khác ( $p < 0,05$ ). Trong đó, dân tộc Ba Na là dân tộc chiếm số lượng cá thể mang các biến thể nhiều nhất (27,7%) và khác biệt có ý nghĩa thống kê với tất cả các dân tộc khác ( $p < 0,05$ ). Những biến thể này cũng có tồn tại với tần số lớn ở một số dân tộc khác, bao gồm H'Rê (20%) thuộc Nam Á và Giarai (13,3%), Giarai-II (8,1%) thuộc Nam Đảo. Những biến thể này có thể thuộc các haplotype được chia sẻ giữa các dân tộc Ba Na, H'Rê, Giarai và Giarai-II như đã được miêu tả ở nghiên cứu trước đây [44]. Nghiên cứu này cũng cho thấy rằng các dân tộc Nam Á và Nam Đảo này cư trú ở vùng Tây Nguyên có sự giao thoa về mặt di truyền, tuy nhiên chúng lại không có haplotype nào được chia sẻ với các dân tộc Hán - Tạng.

### 3.4.3. Các dân tộc Hán - Tạng ở các nước lân cận

Nghiên cứu cũng tiến hành so sánh các dân tộc Hán - Tạng của Việt Nam với các dân tộc thuộc cùng ngữ hệ ở các nước lân cận. Cụ thể, chúng tôi sử dụng 1377 trình tự trình tự toàn bộ mtDNA thuộc 47 quần thể ở 6 đất nước hoặc một vùng có khu vực địa lý gần với Việt Nam, bao gồm Thái Lan, Myanmar, Đài Loan, Trung Quốc, Tây Tạng,

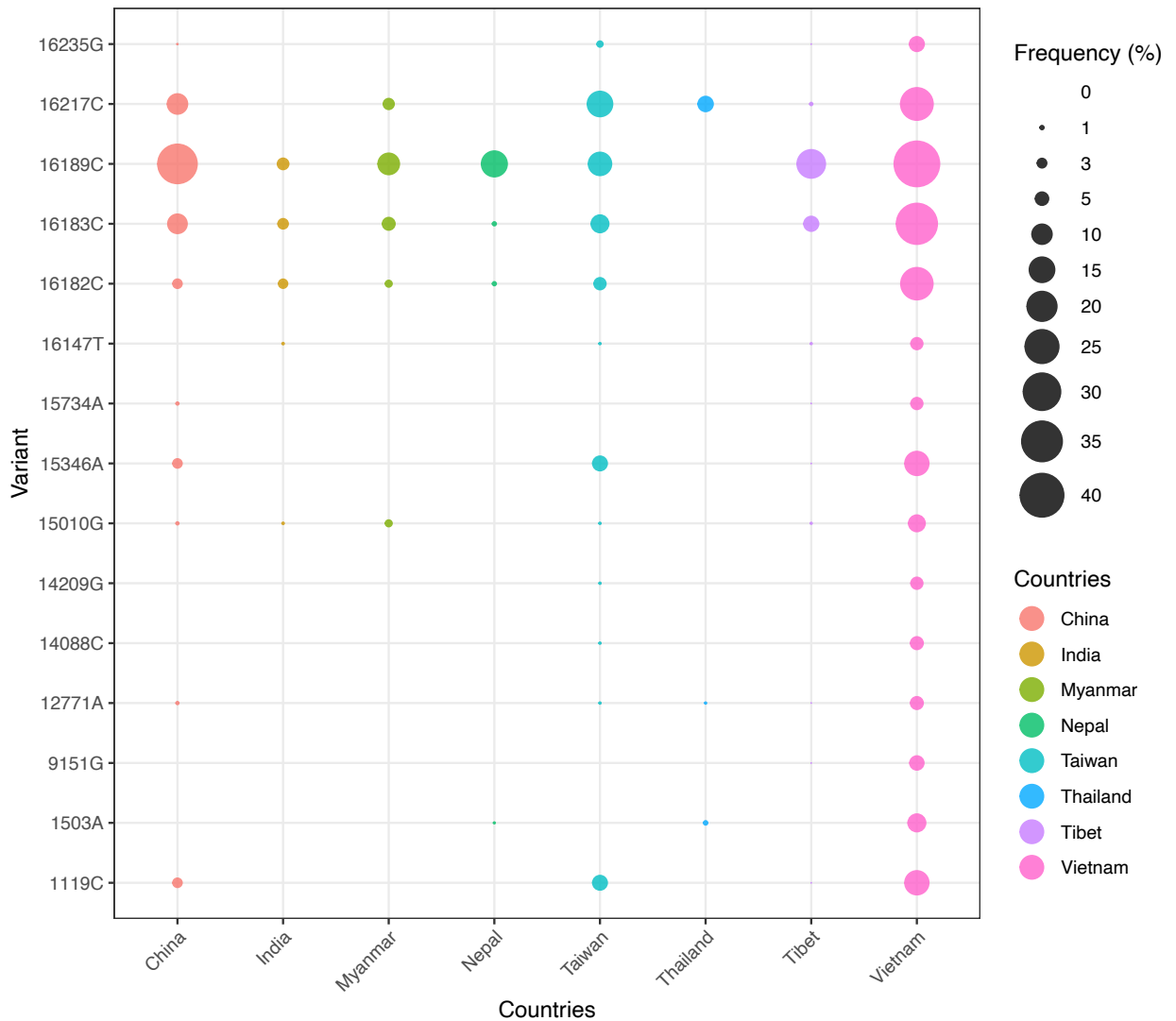
Nepal và Ấn Độ. Kết quả đã phát hiện được có tổng cộng 2664 biến thể tồn tại trong toàn bộ dữ liệu phân tích, trong đó có 1024 biến thể có sự phân bố không đồng đều giữa 6 đất nước kể trên ( $p < 0,05$ ). Trong đó, Đài Loan, Myanmar và Trung Quốc là 3 nước có số lượng biến thể trung bình khác biệt so với Việt Nam thấp nhất với 27,1 - 30,5 biến thể, trong khi đó, các đất nước khác đều có từ lượng biến thể khác biệt từ 39 trở lên (Hình 3.16). Tây Tạng và Ấn Độ là hai nơi có số biến thể khác biệt với Việt Nam nhiều hơn đáng kể so với các đất nước khác ( $p < 0,05$ ), với 66,2 và 59,1 biến thể.



**Hình 3.16. Số lượng biến thể khác biệt giữa các dân tộc Hán - Tạng của Việt Nam so với các nước lân cận**

Khi tiếp tục chia Trung Quốc ra thành 3 vùng nhỏ là Đông, Nam và Bắc Trung Quốc, thì chúng tôi nhận thấy người Hán ở Thượng Hải (phía Đông Trung Quốc) có số lượng biến thể khác biệt so với Việt Nam là 57,3 biến thể, lớn hơn nhiều so với các nhóm dân tộc ở các nước khác trong trong khu vực Đông Nam Á và Đông Á ( $p < 0,05$ ) (Hình 3.16). Ngược lại, các dân tộc Hán ở phía Bắc và Nam Trung Quốc lại có số lượng biến thể khác biệt thấp nhất khi so với gần như tất cả các dân tộc ở các nước lân cận (ngoại trừ Myanmar và Đài Loan) với chỉ 23,9 và 23,5 biến thể khác biệt ( $p < 0,05$ ). Các nghiên cứu trước đây cũng cho rằng các dân tộc Hán - Tạng của Việt Nam (Hoa, Sán Dìu, Hà Nhì, Lô Lô và Si La) có nguồn gốc từ Trung Quốc và di cư tới Việt Nam vào các thời điểm khác nhau trong lịch sử [15]. Ngoài ra, kết quả cũng chỉ ra rằng các dân tộc ở Thái Lan có số biến thể trung bình khác biệt nhiều so với Việt Nam, và nhiều hơn đáng kể so với Trung Quốc (39,7 vs. 30,5;  $p < 0,05$ ). Nghiên cứu trên mtDNA người Hán - Tạng của Thái Lan và Việt Nam cũng cho thấy các dân tộc thuộc ngữ hệ này có sự không đồng nhất về mặt di truyền cũng như dân tộc La Hủ của Việt Nam khác với

dân tộc La Hủ của Thái Lan [57]. Bên cạnh đó, nghiên cứu trước trên các đa hình toàn bộ hệ gen cũng cho thấy rằng dân tộc La Hủ của Việt Nam giống người La Hủ Trung Quốc về mặt di truyền hơn là với người La Hủ ở Thái Lan [103, 104].



**Hình 3.17. Một số biến thể khác biệt nổi bật giữa nhóm dân tộc Hán - Tạng Việt Nam so với các nước lân cận**

Phân tích chi tiết về 1024 biến thể kể trên, chúng tôi xác định được 19 biến thể xuất hiện cao hơn ở các dân tộc Việt Nam so với các dân tộc ở những đất nước khác ( $p < 0,05$ ). Trong đó, 4 biến thể chỉ tồn tại ở Việt Nam với tần số 4,2 - 5,5%, bao gồm 2884T, 9024G, 9030T, 11527T. Khi đối chiếu với cơ sở dữ liệu MITOMAP (<https://www.mitomap.org/MITOMAP>), chúng tôi nhận thấy những biến thể này chỉ được báo cáo trên quần thể người Việt Nam và chưa từng được báo cáo trên bất kỳ quần thể nào khác trên thế giới. Cụ thể, 3 biến thể đầu tiên được chia sẻ giữa các dân tộc Thái - Kadai và Hán - Tạng (tần số 12,5% - 25%), trong khi đó, biến thể 11527T là biến thể tồn tại đặc trưng cho nhóm người Hán - Tạng với tần số lớn nhất xuất hiện ở dân tộc Lô Lô (chiếm 22,2%). Tần số của 15 biến thể còn lại trên các nước trong nghiên cứu được

thể hiện ở hình 3.17. Phần lớn các biến thể này đều xuất hiện ở 3 nước là Việt Nam, Trung Quốc và Đài Loan (9/15, chiếm 60%). Hai biến thể 16183C và 16189C được tìm thấy ở tất cả các nước trong khu vực Đông Á, Nam Á và Đông Nam Á (trừ Thái Lan), trong đó, tần số lớn nhất của hai biến thể này xuất hiện ở Việt Nam và Trung Quốc (chiếm 9,4% - 43,2%).

## KẾT LUẬN VÀ KIẾN NGHỊ

### 4.1. KẾT LUẬN

Qua quá trình nghiên cứu, luận văn “Nghiên cứu hệ gen ty thể người Việt Nam thuộc ba dân tộc Công, Hoa và Sán Dìu” đã thu được một số kết luận như sau:

1. Nghiên cứu đã thu thập và tách chiết DNA tổng số thành công 17 cá thể dân tộc Công, 26 cá thể dân tộc Hoa và 26 cá thể dân tộc Sán Dìu;
2. Nghiên cứu đã giải trình tự thành công hệ gen ty thể của 69 mẫu thuộc ba dân tộc trong nghiên cứu;
3. Nghiên cứu đã xác định được 488 biến thể trên ba dân tộc trong nghiên cứu, trong đó:
  - 2 biến thể 4555T và 9567G chưa từng báo cáo ở các cơ sở dữ liệu trên thế giới;
  - 29 biến thể trong số đó đã được báo cáo có liên quan tới bệnh ty thể;
4. Nghiên cứu đã thực hiện so sánh thống kê các biến thể giữa các dân tộc ngữ hệ Hán - Tạng, giữa các dân tộc Hán - Tạng Việt Nam và các dân tộc thuộc 4 ngữ hệ khác của Việt Nam cũng như các dân tộc Hán - Tạng của các nước lân cận, trong đó:
  - Các dân tộc thuộc ngữ hệ Nam Á, Nam Đảo và Mông - Miên có số lượng biến thể trung bình khác biệt so với các dân tộc thuộc ngữ hệ Hán - Tạng nhiều hơn đáng kể so với ngữ hệ Thái – Kadai;
  - Bốn biến thể chỉ xuất hiện ở các dân tộc Việt Nam và chưa từng được báo cáo trên bất kì quần thể nào khác trên thế giới.

## **4.2. KIẾN NGHỊ**

Tiếp tục nghiên cứu sâu hơn về đặc điểm di truyền quần thể cũng như cấu trúc quần thể của những mẫu dân tộc Cống, Hoa và Sán Dìu bằng cách kết hợp phân tích các kết quả phân tích của các chỉ thị khác như nhiễm sắc thể Y, STR và đa hình toàn bộ hệ gen.



**DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ**

Định, T., Thủy, V., & Dương, N. (2024). Phân tích biến thể hệ gen ty thể các cá thể dân tộc Sán Dìu sinh sống tại tỉnh Thái Nguyên. *TNU Journal of Science and Technology*, 229(13), 296-303. doi: <https://doi.org/10.34238/tnu-jst.10545>.

## DANH MỤC TÀI LIỆU THAM KHẢO

1. Kuhlbrandt, W., 2015, Structure and function of mitochondrial membrane protein complexes, *BMC Biology*, 13, pp. 89.
2. Cavalier-Smith, T., 2006, Origin of mitochondria by intracellular enslavement of a photosynthetic purple bacterium, *Proceedings of the Royal Society B: Biological Sciences*, 273(1596), pp. 1943-1952.
3. Protasoni, M., Zeviani, M., 2021, Mitochondrial Structure and Bioenergetics in Normal and Disease Conditions, *International Journal of Molecular Sciences*, 22(2), pp. 586.
4. Colombini, M., 1989, Voltage gating in the mitochondrial channel, VDAC, *The Journal of Membrane Biology*, 111(2), pp. 103-111.
5. De Stefani, D., Raffaello, A., Teardo, E., Szabo, I., Rizzuto, R., 2011, A forty-kilodalton protein of the inner membrane is the mitochondrial calcium uniporter, *Nature*, 476(7360), pp. 336-340.
6. Chen, W., Zhao, H., Li, Y., 2023, Mitochondrial dynamics in health and disease: mechanisms and potential targets, *Signal Transduction and Targeted Therapy*, 8(1), pp. 333.
7. Pakendorf, B., Stoneking, M., 2005, Mitochondrial DNA and human evolution, *Annual Review of Genomics and Human Genetics*, 6, pp. 165-183.
8. Bernardino Gomes, T.M., Vincent, A.E., Menger, K.E., Stewart, J.B., Nicholls, T.J., 2024, Mechanisms and pathologies of human mitochondrial DNA replication and deletion formation, *Biochemical Journal*, 481(11), pp. 683-715.
9. Taanman, J.W., 1999, The mitochondrial genome: structure, transcription, translation and replication, *Biochimica et Biophysica Acta*, 1410(2), pp. 103-123.
10. Bogenhagen, D., Clayton, D.A., 1978, Mechanism of mitochondrial DNA replication in mouse L-cells: kinetics of synthesis and turnover of the initiation sequence, *Journal of Molecular Biology*, 119(1), pp. 49-68.
11. Pereira, F., Soares, P., Carneiro, J., Pereira, L., Richards, M.B., Samuels, D.C., Amorim, A., 2008, Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region, *Molecular Biology and Evolution*, 25(12), pp. 2759-2770.
12. Ghezzi, D., Zeviani, M., 2018, Human diseases associated with defects in assembly of OXPHOS complexes, *Essays in Biochemistry*, 62(3), pp. 271-286.
13. He, J., Ford, H.C., Carroll, J., Douglas, C., Gonzales, E., Ding, S., Fearnley, I.M., Walker, J.E., 2018, Assembly of the membrane domain of ATP synthase in human mitochondria, *Proceedings of the National Academy of Sciences U S A*, 115(12), pp. 2988-2993.
14. General Statistics Office of Vietnam. "Completed results of the 2019 Vietnam population and housing census." [www.gso.gov.vn](http://www.gso.gov.vn).
15. Dang, N.V., Chu, T.S., Luu, H., 2017, *Ethnic Minorities in Vietnam*, The Gioi, pp.
16. Cann, R.L., Stoneking, M., Wilson, A.C., 1987, Mitochondrial DNA and human evolution, *Nature*, 325(6099), pp. 31-36.
17. Wallace, D.C., 2015, Mitochondrial DNA variation in human radiation and disease, *Cell*, 163(1), pp. 33-38.
18. Soares, P., Rito, T., Pereira, L., Richards, M.B., 2016, A Genetic Perspective on African Prehistory, *Africa from MIS 6-2: Population Dynamics and Paleoenvironments*, Springer Netherlands, Dordrecht, pp. 383-405.
19. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T., Endicott, P., Mastana, S., Papiha, S.S., Skorecki, K., Torroni, A., Villems, R., 2004, Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans, *BMC Genetics*, 5, pp. 26.

20. Li, Y.C., Wang, H.W., Tian, J.Y., Liu, L.N., Yang, L.Q., Zhu, C.L., Wu, S.F., Kong, Q.P., Zhang, Y.P., 2015, Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene, *Scientific Reports*, 5, pp. 9473.
21. Li, Y.C., Ye, W.J., Jiang, C.G., Zeng, Z., Tian, J.Y., Yang, L.Q., Liu, K.J., Kong, Q.P., 2019, River Valleys Shaped the Maternal Genetic Landscape of Han Chinese, *Molecular Biology and Evolution*, 36(8), pp. 1643-1652.
22. Yu-Wai-Man, P., Griffiths, P.G., Chinnery, P.F., 2011, Mitochondrial optic neuropathies - disease mechanisms and therapeutic strategies, *Progress in Retinal and Eye Research*, 30(2), pp. 81-114.
23. Chinnery, P.F., Brown, D.T., Andrews, R.M., Singh-Kler, R., Riordan-Eva, P., Lindley, J., Applegarth, D.A., Turnbull, D.M., Howell, N., 2001, The mitochondrial ND6 gene is a hot spot for mutations that cause Leber's hereditary optic neuropathy, *Brain*, 124(Pt 1), pp. 209-218.
24. Shanske, S., Coku, J., Lu, J., Ganesh, J., Krishna, S., Tanji, K., Bonilla, E., Naini, A.B., Hirano, M., DiMauro, S., 2008, The G13513A mutation in the ND5 gene of mitochondrial DNA as a common cause of MELAS or Leigh syndrome: evidence from 12 cases, *Archives of Neurology*, 65(3), pp. 368-372.
25. Wu, Y., Wang, X.H., Li, X.H., Song, L.Y., Yu, S.L., Fang, Z.C., Liu, Y.Q., Yuan, L.Y., Peng, C.Y., Zhang, S.Y., Cheng, W., Ma, H.C., Wang, L.F., Tang, J.M., Wang, Y.F., Ji, F.Y., 2021, Common mtDNA variations at C5178a and A249d/T6392C/G10310A decrease the risk of severe COVID-19 in a Han Chinese population from Central China, *Military Medical Research*, 8(1), pp. 57.
26. Jia, M., Li, Q., Zhang, T., Dong, B., Liang, X., Fu, S., Yu, J., 2022, Genetic Diversity Analysis of the Chinese Daur Ethnic Group in Heilongjiang Province by Complete Mitochondrial Genome Sequencing, *Frontiers in Genetics*, 13, pp. 919063.
27. Woravatin, W., Wongkomonched, R., Tassaneeyakul, W., Stoneking, M., Makarawate, P., Kutanant, W., 2024, Complete mitochondrial genomes of patients from Thailand with cardiovascular diseases, *PLoS One*, 19(7), pp. e0307036.
28. Kazuno, A.A., Munakata, K., Nagai, T., Shimozono, S., Tanaka, M., Yoneda, M., Kato, N., Miyawaki, A., Kato, T., 2006, Identification of mitochondrial DNA polymorphisms that alter mitochondrial matrix pH and intracellular calcium dynamics, *PLoS Genetics*, 2(8), pp. e128.
29. Malyarchuk, B.A., Derenko, M.V., 2023, Evaluating the role of selection in the evolution of mitochondrial genomes of aboriginal peoples of Siberia, *Vavilovskii Zhurnal Genet Selektsii*, 27(3), pp. 218-223.
30. Kang, L., Zheng, H.X., Chen, F., Yan, S., Liu, K., Qin, Z., Liu, L., Zhao, Z., Li, L., Wang, X., He, Y., Jin, L., 2013, mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands, *Molecular Biology and Evolution*, 30(12), pp. 2579-2587.
31. Kang, L., Zheng, H.X., Zhang, M., Yan, S., Li, L., Liu, L., Liu, K., Hu, K., Chen, F., Ma, L., Qin, Z., Wang, Y., Wang, X., Jin, L., 2016, MtDNA analysis reveals enriched pathogenic mutations in Tibetan highlanders, *Scientific Reports*, 6, pp. 31083.
32. Duong, N.T., Macholdt, E., Ton, N.D., Arias, L., Schroder, R., Van Phong, N., Thi Bich Thuy, V., Ha, N.H., Thi Thu Hue, H., Thi Xuan, N., Thi Phuong Oanh, K., Hien, L.T.T., Hoang, N.H., Pakendorf, B., Stoneking, M., Van Hai, N., 2018, Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia, *Scientific Reports*, 8(1), pp. 11651.
33. Huong Thao, D., Van Hai, N., Thuy Duong, N., 2022, Complete human mtDNA genome sequences revealed haplotype features of the Hmong-Mien language family in Vietnam, *Academia Journal of Biology*, 44(2), pp. 21-28.

34. Ngoc, N.T., Trang, N.B., Huy, N.Q., Ton, N.D., Duong, N.T., 2018, Single Nucleotide Polymorphisms in the D-Loop Region of the Mitochondrial Genomes of Individuals from Two Ethnic Groups Kinh and Mang of Austro-Asiatic Language Family, *Vietnam Journal of Biotechnology*, 16(2), pp. 231-240.
35. Duong, N.T., Phong, N.V., Ngoc, N.T., Hai, N.V., 2020, Study on genetic variations of the D-loop region in three Vietnamese ethnic groups Kinh, Lolo, and Lahu, *Vietnam Journal of Biotechnology*, 18(2), pp. 231-238.
36. Mạnh Hưng, Đ., Hải Hà, N., Nhật Khôi, P., Phương Nhung, V., Văn Phòng, N., Nguyễn, T.D., Văn Hải, N., Đăng Tôn, N., 2016, Sự đa dạng di truyền vùng HV2 hệ gen ty thể của một số nhóm người Việt, *Tạp Chí Sinh học*, 38(2), pp. 243-249.
37. Tran, T.T.H., Nguyen, D.H., Tran, V.K., Nguyen, Q.L., Trinh, H.A., Luong, L.H., Tran, V.A., Pham, L.A.T., Nguyen, T.T., Nguyen, V.B., Tran, T.H., Van Ta, T., 2020, Variation of Mitochondrial DNA HV1 AND HV2 of the Vietnamese Population, *Advances in Experimental Medicine and Biology*, 1292, pp. 37-63.
38. Đăng Tôn, N., Thị Tú Linh, N., Hải Chi, V., Thị Ngọc Diệp, T., Thị Kim Hương, Đ., Thị Tuyết, B., Hải Hà, N., Thị Thu Huệ, H., Thị Thu Hiền, L., Thị Phương Liên, T., Văn Chi, P., Văn Hải, N., 2008, Đa hình đơn bội DNA ty thể của các cá thể người Việt Nam, *Tạp chí Công nghệ Sinh học*, 6(4), pp. 579-590.
39. Thị Thu Huệ, H., Thị Thu Yên, H., Đăng Tôn, N., Thị Thu Hiền, L., Đình Cường, N., Văn Chi, P., Văn Hải, N., 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), pp. 15-22.
40. Thị Thuý Hằng, T., Huy Thịnh, T., Vân Khánh, T., 2017, Đa hình thái đơn nucleotide vùng gen ty thể HV1 và HV2 trên người dân tộc Kinh và dân tộc Mường của Việt Nam, *Tạp chí Nghiên cứu Y học*, 106(1), pp. 33-40.
41. Minh Tùng, N., Văn Ba, N., Đăng Tôn, N., 2021, Tính đa hình trình tự gen mã hóa ND5 và ND6 ở người dân tộc Giarai và Êđê sống ở Tây Nguyên, *F1000Research*, 504(1), pp. 20-25.
42. Bao Khanh, P., Van Minh, N., Thi Hong Loan, N., Van Anh, P., Ngoc Anh, L., Vu Hung, C., Tuan Nghia, P., 2017, Detection and quantitation of mitochondrial G11778A mutation of LHON syndrome in a Vietnamese patient with tentatively diagnosed mitochondrial disease, *VNU Journal of Science: Natural Sciences and Technology*, 33(2S), pp. 20-25.
43. Thị Minh Nguyệt, T., Thị Bích Thảo, L., Thị Huyền, B., Đình Minh, P., Thế Thành, T., Thị Ty, N., Bích Nhi, N., Diễm Hồng, Đ., Quang Huân, L., Đình Thi, Q., Đăng Tôn, N., Văn Hải, N., Văn Chi, P., 2008, Trình tự toàn bộ genome ty thể từ 9 cá thể người Việt Nam, *Tạp chí Công nghệ Sinh học*, 6(4A), pp. 569-578.
44. Thao, D.H., Dinh, T.H., Mitsunaga, S., Duy, D., Phuong, N.T., Anh, N.P., Anh, N.T., Duc, B.M., Hue, H.T.T., Ha, N.H., Ton, N.D., Hubner, A., Pakendorf, B., Stoneking, M., Inoue, I., Duong, N.T., Hai, N.V., 2024, Investigating demic versus cultural diffusion and sex bias in the spread of Austronesian languages in Vietnam, *PLoS One*, 19(6), pp. e0304964.
45. Ewels, P., Magnusson, M., Lundin, S., Kaller, M., 2016, MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics*, 32(19), pp. 3047-3048.
46. Bolger, A.M., Lohse, M., Usadel, B., 2014, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, 30(15), pp. 2114-20.
47. Li, H., Durbin, R., 2009, Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics*, 25(14), pp. 1754-1760.
48. Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul, L.M., Brackney, D.E., Grewal, S., Gurfield, N., Van Rompay, K.K.A., Isern, S., Michael, S.F., Coffey, L.L., Loman, N.J., Andersen, K.G., 2019, An amplicon-

- based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar, *Genome Biology*, 20(1), pp. 8.
49. Weissensteiner, H., Pacher, D., Kloss-Brandstatter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., Schonherr, S., 2016, HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing, *Nucleic Acids Research*, 44(W1), pp. W58-63.
  50. van Oven, M., Kayser, M., 2009, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Human Mutation*, 30(2), pp. E386-E394.
  51. Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., Ng, P.C., 2012, SIFT web server: predicting effects of amino acid substitutions on proteins, *Nucleic Acids Research*, 40(Web Server issue), pp. W452-457.
  52. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010, A method and server for predicting damaging missense mutations, *Nature Methods*, 7(4), pp. 248-249.
  53. Choi, Y., Chan, A.P., 2015, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels, *Bioinformatics*, 31(16), pp. 2745-2747.
  54. Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K.A., Lin, G.N., Nam, H.J., Mort, M., Cooper, D.N., Sebat, J., Iakoucheva, L.M., Mooney, S.D., Radivojac, P., 2020, Inferring the molecular and phenotypic impact of amino acid variants with MutPred2, *Nature Communications*, 11(1), pp. 5918.
  55. Castellana, S., Ronai, J., Mazza, T., 2015, MitImpact: an exhaustive collection of pre-computed pathogenicity predictions of human mitochondrial non-synonymous variants, *Human Mutation*, 36(2), pp. E2413-2422.
  56. Kutanan, W., Kampuansai, J., Brunelli, A., Ghirotto, S., Pittayaporn, P., Ruangchai, S., Schroder, R., Macholdt, E., Srikummool, M., Kangwanpong, D., Hubner, A., Arias, L., Stoneking, M., 2018, New insights from Thailand into the maternal genetic history of Mainland Southeast Asia, *European Journal of Human Genetics*, 26(6), pp. 898-911.
  57. Kutanan, W., Shoocongdej, R., Srikummool, M., Hubner, A., Suttipai, T., Srithawong, S., Kampuansai, J., Stoneking, M., 2020, Cultural variation impacts paternal and maternal genetic lineages of the Hmong-Mien and Sino-Tibetan groups from Thailand, *European Journal of Human Genetics*, 28(11), pp. 1563-1579.
  58. Summerer, M., Horst, J., Erhart, G., Weissensteiner, H., Schonherr, S., Pacher, D., Forer, L., Horst, D., Manhart, A., Horst, B., Sanguansermsri, T., Kloss-Brandstatter, A., 2014, Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar, *BMC Ecology and Evolution*, 14, pp. 17.
  59. Yao, Y.G., Kong, Q.P., Bandelt, H.J., Kivisild, T., Zhang, Y.P., 2002, Phylogeographic differentiation of mitochondrial DNA in Han Chinese, *The American Journal of Human Genetics*, 70(3), pp. 635-651.
  60. Ma, K., Li, H., Cao, Y., Zhao, X., Liu, W., Zhao, X., 2016, Haplotype diversity in mitochondrial genome in a Chinese Han population, *Journal of Human Genetics*, 61(10), pp. 903-906.
  61. Ko, Albert M.-S., Chen, C.-Y., Fu, Q., Delfin, F., Li, M., Chiu, H.-L., Stoneking, M., Ko, Y.-C., 2014, Early Austronesians: Into and Out Of Taiwan, *The American Journal of Human Genetics*, 94(3), pp. 426-436.
  62. Lin, M., Trejaut, J.A., 2023, Diversity and distribution of mitochondrial DNA in non-Austronesian-speaking Taiwanese individuals, *Human Genome Variation* 10(1), pp. 2.
  63. Qin, Z., Yang, Y., Kang, L., Yan, S., Cho, K., Cai, X., Lu, Y., Zheng, H., Zhu, D., Fei, D., Li, S., Jin, L., Li, H., Genographic, C., 2010, A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum, *American Journal of Physical Anthropology*, 143(4), pp. 555-569.

64. Zhao, M., Kong, Q.P., Wang, H.W., Peng, M.S., Xie, X.D., Wang, W.Z., Jiayang, Duan, J.G., Cai, M.C., Zhao, S.N., Cidanpingcuo, Tu, Y.Q., Wu, S.F., Yao, Y.G., Bandelt, H.J., Zhang, Y.P., 2009, Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau, *Proceedings of the National Academy of Sciences U S A*, 106(50), pp. 21230-21235.
65. Basnet, R., Rai, N., Tamang, R., Awasthi, N.P., Pradhan, I., Parajuli, P., Kashyap, D., Reddy, A.G., Chaubey, G., Das Manandhar, K., Shrestha, T.R., Thangaraj, K., 2023, The matrilineal ancestry of Nepali populations, *Human Genetics*, 142(2), pp. 167-180.
66. Bhandari, S., Zhang, X., Cui, C., Bianba, Liao, S., Peng, Y., Zhang, H., Xiang, K., Shi, H., Ouzhuluobu, Baimakongzhuo, Gonggalanzi, Liu, S., Gengdeng, Wu, T., Qi, X., Su, B., 2015, Genetic evidence of a recent Tibetan ancestry to Sherpas in the Himalayan region, *Scientific Reports*, 5, pp. 16249.
67. Chandrasekar, A., Kumar, S., Sreenath, J., Sarkar, B.N., Urade, B.P., Mallick, S., Bandopadhyay, S.S., Barua, P., Barik, S.S., Basu, D., Kiran, U., Gangopadhyay, P., Sahani, R., Prasad, B.V., Gangopadhyay, S., Lakshmi, G.R., Ravuri, R.R., Padmaja, K., Venugopal, P.N., Sharma, M.B., Rao, V.R., 2009, Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor, *PLoS One*, 4(10), pp. e7447.
68. R Core Team. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing. <https://www.R-project.org/>.
69. Kutanan, W., Kampuansai, J., Srikummool, M., Kangwanpong, D., Ghirotto, S., Brunelli, A., Stoneking, M., 2017, Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages, *Human Genetics*, 136(1), pp. 85-98.
70. Vanichanukulyakit, J., Khacha-Ananda, S., Monum, T., Mahawong, P., Moophayak, K., Penkhrue, W., Khumpook, T., Thongsahuan, S., 2023, The Analysis of Genetic Polymorphism on Mitochondrial Hypervariable Region III in Thai Population, *Genes (Basel)*, 14(3), pp. 682.
71. Luo, Y., Gao, W., Liu, F., Gao, Y., 2011, Mitochondrial nt3010G-nt3970C haplotype is implicated in high-altitude adaptation of Tibetans, *Mitochondrial DNA*, 22(5-6), pp. 181-190.
72. Zou, Y., Jia, X., Zhang, A.M., Wang, W.Z., Li, S., Guo, X., Kong, Q.P., Zhang, Q., Yao, Y.G., 2010, The MT-ND1 and MT-ND5 genes are mutational hotspots for Chinese families with clinical features of LHON but lacking the three primary mutations, *Biochemical and Biophysical Research Communications*, 399(2), pp. 179-185.
73. Jeoung, J.W., Seong, M.W., Park, S.S., Kim, D.M., Kim, S.H., Park, K.H., 2014, Mitochondrial DNA variant discovery in normal-tension glaucoma patients by next-generation sequencing, *Investigative Ophthalmology & Visual Science*, 55(2), pp. 986-992.
74. Tanaka, M., Gong, J.S., Zhang, J., Yoneda, M., Yagi, K., 1998, Mitochondrial genotype associated with longevity, *Lancet*, 351(9097), pp. 185-186.
75. Kokaze, A., Ishikawa, M., Matsunaga, N., Yoshida, M., Satoh, M., Teruya, K., Masuda, Y., Honmyo, R., Uchida, Y., Takashima, Y., 2007, NADH dehydrogenase subunit-2 237 Leu/Met polymorphism modifies the effects of alcohol consumption on risk for hypertension in middle-aged Japanese men, *Hypertension Research: official journal of the Japanese Society of Hypertension*, 30(3), pp. 213-218.
76. Kokaze, A., Ishikawa, M., Matsunaga, N., Karita, K., Yoshida, M., Ohtsu, T., Ochiai, H., Shirasawa, T., Nanri, H., Hoshino, H., Takashima, Y., 2013, Unexpected combined effects of NADH dehydrogenase subunit-2 237 Leu/Met polymorphism and green tea consumption on renal function in male Japanese health check-up examinees: a cross-sectional study, *Journal of Negative Results in Biomedicine*, 12, pp. 17.

77. Kokaze, A., Ishikawa, M., Matsunaga, N., Karita, K., Yoshida, M., Shimada, N., Ohtsu, T., Shirasawa, T., Ochiai, H., Hoshino, H., Takashima, Y., 2013, Combined effect of mitochondrial DNA 5178 C/A polymorphism and alcohol consumption on estimated glomerular filtration rate in male Japanese health check-up examinees: a cross-sectional study, *BMC Nephrology*, 14, pp. 35.
78. Kokaze, A., Ishikawa, M., Matsunaga, N., Yoshida, M., Satoh, M., Teruya, K., Honmyo, R., Shirasawa, T., Hoshino, H., Takashima, Y., 2007, Longevity-associated mitochondrial DNA 5178 C/A polymorphism and its interaction with cigarette consumption are associated with pulmonary function in middle-aged Japanese men, *Journal of Human Genetics*, 52(8), pp. 680-685.
79. Ohtsu, I., Ishikawa, M., Matsunaga, N., Karita, K., Yoshida, M., Ochiai, H., Shirasawa, T., Yoshimoto, T., Minoura, A., Sai, S., Kokaze, A., 2019, Longevity-associated mitochondrial DNA 5178 C/A polymorphism modifies effect of aging on renal function in male Japanese health checkup examinees: an exploratory cross-sectional study, *Journal of Physiological Anthropology*, 38(1), pp. 12.
80. Verma, R.K., Kalyakulina, A., Giuliani, C., Shinde, P., Kachhvhah, A.D., Ivanchenko, M., Jalan, S., 2021, Analysis of human mitochondrial genome co-occurrence networks of Asian population at varying altitudes, *Scientific Reports*, 11(1), pp. 133.
81. Sharma, V., Sharma, I., Singh, V.P., Verma, S., Pandita, A., Singh, V., Rai, E., Sharma, S., 2014, mtDNA G10398A variation provides risk to type 2 diabetes in population group from the Jammu region of India, *Meta Gene*, 2, pp. 269-273.
82. Chu, Q., Luo, X., Zhan, X., Ren, Y., Pang, H., 2015, Female genetic distribution bias in mitochondrial genome observed in Parkinson's Disease patients in northern China, *Scientific Reports*, 5, pp. 17170.
83. Juo, S.H., Lu, M.Y., Bai, R.K., Liao, Y.C., Trieu, R.B., Yu, M.L., Wong, L.J., 2010, A common mitochondrial polymorphism 10398A>G is associated metabolic syndrome in a Chinese population, *Mitochondrion*, 10(3), pp. 294-299.
84. Destiarani, W., Mulyani, R., Yusuf, M., Maksum, I.P., 2020, Molecular Dynamics Simulation of T10609C and C10676G Mutations of Mitochondrial ND4L Gene Associated With Proton Translocation in Type 2 Diabetes Mellitus and Cataract Patients, *Bioinformatics and Biology Insights*, 14, pp. 1177932220978672.
85. Jiang, C., Cui, J., Liu, F., Gao, L., Luo, Y., Li, P., Guan, L., Gao, Y., 2014, Mitochondrial DNA 10609T promotes hypoxia-induced increase of intracellular ROS and is a risk factor of high altitude polycythemia, *PLoS One*, 9(1), pp. e87775.
86. Behbehani, R., Melhem, M., Alghanim, G., Behbehani, K., Alsmadi, O., 2014, ND4L gene concurrent 10609T>C and 10663T>C mutations are associated with Leber's hereditary optic neuropathy in a large pedigree from Kuwait, *The British Journal of Ophthalmology*, 98(6), pp. 826-831.
87. Russlies, J., Fahrnich, A., Witte, M., Yin, J., Benoit, S., Glaser, R., Gunter, C., Eming, R., Erdmann, J., Gola, D., Gupta, Y., Holtsche, M.M., Kern, J.S., Konig, I.R., Kiritsi, D., Lieb, W., Sadik, C.D., Sardy, M., Schauer, F., van Beek, N., Weidinger, A., Worm, M., Zillikens, D., Schmidt, E., Busch, H., Ibrahim, S.M., Hirose, M., 2019, Polymorphisms in the Mitochondrial Genome Are Associated With Bullous Pemphigoid in Germans, *Frontiers in Immunology*, 10, pp. 2200.
88. Zheng, Y., Wang, Y., Jiang, Y., Wang, J., Li, S., Xiao, X., Sun, W., Wang, P., Zhang, Q., Jia, X., 2024, Variant and clinical landscape of Leber hereditary optic neuropathy based on 1516 families with mtDNA variants in a tertiary centre, *The British Journal of Ophthalmology*, pp. 1318-1327.
89. Piotrowska-Nowak, A., Kosior-Jarecka, E., Schab, A., Wrobel-Dudzinska, D., Bartnik, E., Zarnowski, T., Tonska, K., 2019, Investigation of whole mitochondrial genome variation in normal tension glaucoma, *Experimental Eye Research*, 178, pp. 186-197.

90. Wei, L., Qiang, L., Aiqin, N., Li, J., 2017, The mitochondrial ND5 T12338C mutation may influence the phenotypic manifestation of Leber's hereditary optic neuropathy-associated ND4 G11696A mutation, *International Journal of Clinical and Experimental Pathology*, 10(6), pp. 7200-7206.
91. Qin, Y., Xue, L., Jiang, P., Xu, M., He, Y., Shi, S., Huang, Y., He, J., Mo, J.Q., Guan, M.X., 2014, Mitochondrial tRNA variants in Chinese subjects with coronary heart disease, *Journal of the American Heart Association*, 3(1), pp. e000437.
92. Garcia, S., Lopez-Hernandez, L., Davila-Maldonado, L., Cuevas-Garcia, C., Gallegos-Arreola, M., Alcaraz-Estrada, S., Cortes-Espinosa, L., Flores, C., Canto, P., Vazquez, R., 2019, Association of mitochondrial variants A4336G of the tRNAGln gene and 8701G/A of the MT-ATP6 gene in Mexicans Mestizos with Parkinson disease, *Folia Neuropathologica*, 57(4), pp. 335-339.
93. Sharma, S., Singh, S., Gupta, R.K., Ganju, L., Singh, S.B., Kumar, B., Singh, Y., 2019, Mitochondrial DNA sequencing reveals association of variants and haplogroup M33a2'3 with High altitude pulmonary edema susceptibility in Indian male lowlanders, *Scientific Reports*, 9(1), pp. 10975.
94. Gomez-Duran, A., Pacheu-Grau, D., Martinez-Romero, I., Lopez-Gallardo, E., Lopez-Perez, M.J., Montoya, J., Ruiz-Pesini, E., 2012, Oxidative phosphorylation differences between mitochondrial DNA haplogroups modify the risk of Leber's hereditary optic neuropathy, *Biochimica et Biophysica Acta*, 1822(8), pp. 1216-1222.
95. Pacheu-Grau, D., Gomez-Duran, A., Iglesias, E., Lopez-Gallardo, E., Montoya, J., Ruiz-Pesini, E., 2013, Mitochondrial antibiograms in personalized medicine, *Human Molecular Genetics*, 22(6), pp. 1132-1139.
96. Rovcanin, B., Jancic, J., Samardzic, J., Rovcanin, M., Nikolic, B., Ivancevic, N., Novakovic, I., Kostic, V., 2020, In silico model of mtDNA mutations effect on secondary and 3D structure of mitochondrial rRNA and tRNA in Leber's hereditary optic neuropathy, *Experimental Eye Research*, 201, pp. 108277.
97. Collins, D.W., Gudiseva, H.V., Trachtman, B., Bowman, A.S., Sagaser, A., Sankar, P., Miller-Ellis, E., Lehman, A., Addis, V., O'Brien, J.M., 2016, Association of primary open-angle glaucoma with mitochondrial variants and haplogroups common in African Americans, *Molecular Vision*, 22, pp. 454-471.
98. Veenin, K., Wattanasirichaigoon, D., Suktitipat, B., Noojareern, S., Lertrit, P., Tim-Aroon, T., Kaewsutthi, S., Treepongkaruna, S., 2022, Association of Mitochondrial DNA Polymorphisms With Pediatric-Onset Cyclic Vomiting Syndrome, *Frontiers in Pediatrics*, 10, pp. 876436.
99. Billing-Ross, P., Germain, A., Ye, K., Keinan, A., Gu, Z., Hanson, M.R., 2016, Mitochondrial DNA variants correlate with symptoms in myalgic encephalomyelitis/chronic fatigue syndrome, *Journal of Translational Medicine*, 14, pp. 19.
100. Guo, L.J., Oshida, Y., Fuku, N., Takeyasu, T., Fujita, Y., Kurata, M., Sato, Y., Ito, M., Tanaka, M., 2005, Mitochondrial genome polymorphisms associated with type-2 diabetes or obesity, *Mitochondrion*, 5(1), pp. 15-33.
101. Li, Z., Li, R., Chen, J., Liao, Z., Zhu, Y., Qian, Y., Xiong, S., Heman-Ackah, S., Wu, J., Choo, D.I., Guan, M.X., 2005, Mutational analysis of the mitochondrial 12S rRNA gene in Chinese pediatric subjects with aminoglycoside-induced and non-syndromic hearing loss, *Human Genetics*, 117(1), pp. 9-15.
102. Macholdt, E., Arias, L., Duong, N.T., Ton, N.D., Van Phong, N., Schroder, R., Pakendorf, B., Van Hai, N., Stoneking, M., 2020, The paternal and maternal genetic history of Vietnamese populations, *European Journal of Human Genetics*, 28(5), pp. 636-645.
103. Kutanan, W., Liu, D., Kampuansai, J., Srikummool, M., Srithawong, S., Shoocongdej, R., Sangkhano, S., Ruangchai, S., Pittayaporn, P., Arias, L., Stoneking, M., 2021,



Reconstructing the Human Genetic History of Mainland Southeast Asia: Insights from Genome-Wide Data from Thailand and Laos, *Molecular Biology and Evolution*, 38(8), pp. 3459-3477.

104. Liu, D., Duong, N.T., Ton, N.D., Van Phong, N., Pakendorf, B., Van Hai, N., Stoneking, M., 2020, Extensive Ethnolinguistic Diversity in Vietnam Reflects Multiple Sources of Genetic Diversity, *Molecular Biology and Evolution*, 37(9), pp. 2503-2519.
105. Kong, Q.P., Yao, Y.G., Sun, C., Bandelt, H.J., Zhu, C.L., Zhang, Y.P., 2003, Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences, *The American Journal of Human Genetics*, 73(3), pp. 671-676.

**PHỤ LỤC**

- I. Phụ lục nồng độ và độ tinh sạch của các mẫu dân tộc Cống, Hoa và Sán Dìu..... PL1
- II. Phụ lục thông tin chi tiết các trình tự sử dụng trong nghiên cứu ..... PL2

**I. Phụ lục nồng độ và độ tinh sạch của các mẫu dân tộc Công, Hoa và Sán Dìu**

STT	Mã mẫu	Dân tộc	Nồng độ (ng/ $\mu$ L)	A260/A280
1	HN01	Hoa	64,4	1,72
2	HN02	Hoa	53,4	1,83
3	HN03	Hoa	59,2	1,80
4	HN04	Hoa	45,7	1,88
5	HN05	Hoa	72,8	1,88
6	HN06	Hoa	61,8	1,77
7	HN07	Hoa	94,3	1,70
8	HN08	Hoa	87,1	1,78
9	HN09	Hoa	95,3	1,81
10	HN10	Hoa	37,5	1,82
11	HN11	Hoa	72,0	1,84
12	HN12	Hoa	80,7	1,84
13	HN13	Hoa	98,0	1,83
14	HN14	Hoa	66,9	1,83
15	HN15	Hoa	128,8	1,85
16	HN16	Hoa	73,2	1,91
17	HN17	Hoa	73,7	1,81
18	HN18	Hoa	94,3	1,84
19	HN19	Hoa	107,9	1,85
20	HN20	Hoa	90,5	1,86
21	HN21	Hoa	98,7	1,84
22	HN22	Hoa	67,8	1,86
23	HN23	Hoa	64,9	1,85
24	HN24	Hoa	80,4	1,83
25	HN25	Hoa	74,3	1,85
26	HN26	Hoa	63,9	1,84
27	SD01	Sán Dìu	61,4	1,84
28	SD02	Sán Dìu	101,0	1,83
29	SD03	Sán Dìu	80,6	1,83
30	SD04	Sán Dìu	96,3	1,82
31	SD05	Sán Dìu	68,9	1,83
32	SD06	Sán Dìu	113,7	1,84
33	SD07	Sán Dìu	62,5	1,81
34	SD08	Sán Dìu	123,4	1,84
35	SD09	Sán Dìu	87,2	1,82

## PL1

36	SD10	Sán Dìu	63,9	1,83
37	SD11	Sán Dìu	54,2	1,83
38	SD12	Sán Dìu	52,3	1,73
39	SD13	Sán Dìu	82,5	1,73
40	SD14	Sán Dìu	67,1	1,77
41	SD15	Sán Dìu	54,5	1,83
42	SD16	Sán Dìu	64,0	1,82
43	SD17	Sán Dìu	54,7	1,79
44	SD18	Sán Dìu	48,6	1,80
45	SD19	Sán Dìu	94,3	1,84
46	SD20	Sán Dìu	68,1	1,83
47	SD21	Sán Dìu	42,0	1,79
48	SD22	Sán Dìu	56,6	1,84
49	SD23	Sán Dìu	47,8	1,79
50	SD24	Sán Dìu	77,0	1,81
51	SD25	Sán Dìu	59,2	1,78
52	SD26	Sán Dìu	60,3	1,83
53	CG01	Cống	78,1	1,83
54	CG02	Cống	106,6	1,74
55	CG03	Cống	80,3	1,83
56	CG04	Cống	121,4	1,85
57	CG05	Cống	72,3	1,81
58	CG06	Cống	48,8	1,82
59	CG07	Cống	53,9	1,79
60	CG08	Cống	164,8	1,81
61	CG09	Cống	30,2	1,78
62	CG10	Cống	75,8	1,84
63	CG11	Cống	89,2	1,80
64	CG12	Cống	35,7	1,71
65	CG13	Cống	99,5	1,84
66	CG14	Cống	63,5	1,81
67	CG15	Cống	100,5	1,83
68	CG16	Cống	110,6	1,77
69	CG17	Cống	118,9	1,85

## II. Phụ lục thông tin chi tiết các trình tự sử dụng trong nghiên cứu

STT	Quần thể	Số lượng	Dân tộc	Ngữ hệ	Địa điểm thu mẫu				Nguồn
					Huyện	Tỉnh/ TP	Vùng	Nước	
1	Hoa	26	Hoa	Hán - Tạng	-	Hồ Chí Minh	-	Việt Nam	NC hiện tại
2	Sán Dìu	26	Sán Dìu	Hán - Tạng	Phú Lương	Thái Nguyên	-	Việt Nam	NC hiện tại
3	Cống	17	Cống	Hán - Tạng	-	Lai Châu	-	Việt Nam	NC hiện tại
4	Hà Nhì	33	Hà Nhì	Hán - Tạng	Mường Tè	Lai Châu	-	Việt Nam	[32]
5	Phù Lá	35	Phù Lá	Hán - Tạng	Xín Mần	Hà Giang	-	Việt Nam	[32]
6	Si La	31	Si La	Hán - Tạng	Mường Tè	Lai Châu	-	Việt Nam	[32]
7	La Hủ	32	La Hủ	Hán - Tạng	Mường Tè	Lai Châu	-	Việt Nam	[32]
8	Lô Lô	36	Lô Lô	Hán - Tạng	Mèo Vạc	Hà Giang	-	Việt Nam	[32]
9	Kinh	50	Kinh	Nam Á		Hà Nội	-	Việt Nam	[32]
10	Mảng	37	Mảng	Nam Á	Mường Tè	Lai Châu	-	Việt Nam	[32]
11	Ba Na	36	Ba Na	Nam Á	Kon Rẫy	Kon Tum	-	Việt Nam	[44]
12	Cơ Ho	46	Cơ Ho	Nam Á	Di Linh	Lâm Đồng	-	Việt Nam	[44]
13	H'Rê	30	H'Rê	Nam Á	Kon Plong	Kon Tum	-	Việt Nam	[44]
14	M'Nông	53	M'Nông	Nam Á	Lắk	Đắk Lắk	-	Việt Nam	[44]
15	Mạ	26	Mạ	Nam Á	Bảo Lâm	Lâm Đồng	-	Việt Nam	[44]
16	Giarai	27	Giarai	Nam Đảo	-	Gia Lai	-	Việt Nam	[32]
17	Ê Đê	24	Ê Đê	Nam Đảo	-	Đắk Lắk	-	Việt Nam	[32]
18	Giarai-II	37	Giarai-II	Nam Đảo	Sa Thầy	Kon Tum	-	Việt Nam	[44]
19	Ê Đê-II	43	Ê Đê-II	Nam Đảo	Buôn Ma Thuột	Đắk Lắk	-	Việt Nam	[44]
20	Chăm	17	Chăm	Nam Đảo	Đồng Phú	Bình Phước	-	Việt Nam	[44]
21	Chu Ru	44	Chu Ru	Nam Đảo	Đơn Dương	Lâm Đồng	-	Việt Nam	[44]

## PL2

22	Raglay	37	Raglay	Nam Đảo	Khánh Sơn	Khánh Hoà	-	Việt Nam	[44]
23	H'Mông	41	H'Mông	Mông - Miền	Điện Biên Phủ	Điện Biên	-	Việt Nam	[32]
24	Pà Thẻn	36	Pà Thẻn	Mông - Miền	Quảng Bình	Hà Giang	-	Việt Nam	[32]
25	Dao	43	Dao	Mông - Miền	Hoàng Su Phì	Hà Giang	-	Việt Nam	[32]
26	Tày	47	Tày	Thái - Kadai	Mường Khương	Lào Cai	-	Việt Nam	[32]
27	Nùng	37	Nùng	Thái - Kadai	Hoàng Su Phì	Hà Giang	-	Việt Nam	[32]
28	Thái	24	Thái	Thái - Kadai	Điện Biên Phủ	Điện Biên	-	Việt Nam	[32]
29	Cờ Lao	34	Cờ Lao	Thái - Kadai	Hoàng Su Phì	Hà Giang	-	Việt Nam	[32]
30	La Chí	36	La Chí	Thái - Kadai	Hoàng Su Phì	Hà Giang	-	Việt Nam	[32]
31	Lahu-Black	26	La Hủ	Hán - Tạng	Pang Mapha	Mae Hong Son	-	Thái Lan	[57]
32	Lahu-Red	26	La Hủ	Hán - Tạng	Pang Mapha	Mae Hong Son	-	Thái Lan	[57]
33	Lisu	29	Lisu	Hán - Tạng	Pang Mapha	Mae Hong Son	-	Thái Lan	[57]
34	PaduangKaren	25	Karen	Hán - Tạng	-	-	-	Thái Lan	[56]
35	PwoKaren	24	Karen	Hán - Tạng	-	-	-	Thái Lan	[56]
36	ShawKaren-1	25	Karen	Hán - Tạng	-	-	-	Thái Lan	[56]
37	ShawKaren-2	13	Karen	Hán - Tạng	-	-	-	Thái Lan	[56]
38	ShawKaren-3	26	Karen	Hán - Tạng	Pang Mapha	Mae Hong Son	-	Thái Lan	[57]
39	Bamar-1	23	Bamar	Hán - Tạng	-	-	-	Myanmar	[58]
40	Bamar-2	35	Bamar	Hán - Tạng	-	-	-	Myanmar	[20]
41	Chin	17	Chin	Hán - Tạng	-	-	-	Myanmar	[20]
42	Karen	12	Karen	Hán - Tạng	-	-	-	Myanmar	[58]
43	Naga	10	Naga	Hán - Tạng	-	-	-	Myanmar	[20]
44	Rakhine	11	Rakhine	Hán - Tạng	-	-	-	Myanmar	[20]
45	Han-Beijing	46	Hán	Hán - Tạng	-	Bắc Kinh	Bắc	Trung Quốc	1kGenome

## PL2

46	Northern Han-1	18	Hán	Hán - Tạng	-	Sơn Đông	Bắc	Trung Quốc	[105]
47	Han-Shanghai	145	Hán	Hán - Tạng	-	Thượng Hải	Đông	Trung Quốc	[60]
48	Southern Han-1	52	Hán	Hán - Tạng	-	-	Nam	Trung Quốc	1kGenome
49	Southern Han-2	17	Hán	Hán - Tạng	-	Quảng Đông	Nam	Trung Quốc	[105]
50	Hakka	45	Hakka	Hán - Tạng	Pingtung	Neipu	-	Đài Loan	[61]
51	Han-Taipei	15	Hán	Hán - Tạng	-	Đài Bắc	-	Đài Loan	[62]
52	Han-Taiwan	42	Hán	Hán - Tạng	-	-	-	Đài Loan	[62]
53	Matsu	16	Matsu	Hán - Tạng	LienChiang	Nangan	-	Đài Loan	[62]
54	Minnan	50	Minnan	Hán - Tạng	Kaohsiung	Kaohsiung	-	Đài Loan	[61]
55	Yunlin	20	Yunlin	Hán - Tạng	Yunlin	Shuilin	-	Đài Loan	[62]
56	Tibetan-2	31	Tây Tạng	Hán - Tạng	-	Shanan	-	Tây Tạng	[63]
57	Sherpa-Tibet	76	Sherpa	Hán - Tạng	-	Shigatse	-	Tây Tạng	[30]
58	Tibetan-1	23	Tây Tạng	Hán - Tạng	-	Shigatse	-	Tây Tạng	[64]
59	Tibetan-Deng	78	Tây Tạng	Hán - Tạng	-	Zayu	-	Tây Tạng	[31]
60	Tibetan-Lhoba	90	Tây Tạng	Hán - Tạng	-	Mainling	-	Tây Tạng	[31]
61	Tibetan-Monpa	17	Tây Tạng	Hán - Tạng	-	Gongbo'gyamda	-	Tây Tạng	[31]
62	Tibetan-Tingri	86	Tây Tạng	Hán - Tạng	-	Tingri	-	Tây Tạng	[31]
63	Bajracharya	10	Bajracharya	Hán - Tạng	-	Kathmandu	-	Nepal	[65]
64	Maharjan	22	Maharjan	Hán - Tạng	-	Kathmandu	-	Nepal	[65]
65	Manandhar	35	Manandhar	Hán - Tạng	-	Kathmandu	-	Nepal	[65]
66	Newa-mix	14	Newa	Hán - Tạng	-	Kathmandu	-	Nepal	[65]
67	Shakya	8	Shakya	Hán - Tạng	-	Kathmandu	-	Nepal	[65]
68	Sherpa-Nepal	79	Sherpa	Hán - Tạng	-	-	-	Nepal	[66]
69	Shrestha	7	Shrestha	Hán - Tạng	-	Kathmandu	-	Nepal	[65]

## PL2

70	Udaya	34	Udaya	Hán - Tạng	-	Kathmandu	-	Nepal	[65]
71	Dirang Monpa	30	Dirang Monpa	Hán - Tạng	-	-	-	Ấn Độ	[67]
72	Gallong	39	Gallong	Hán - Tạng	-	-	-	Ấn Độ	[67]
73	Lachungpa	23	Lachungpa	Hán - Tạng	-	-	-	Ấn Độ	[67]
74	Lepcha	20	Lepcha	Hán - Tạng	-	-	-	Ấn Độ	[67]
75	Shertukpen	15	Shertukpen	Hán - Tạng	-	-	-	Ấn Độ	[67]
76	Toto	28	Toto	Hán - Tạng	-	-	-	Ấn Độ	[67]
77	Wanchoo	22	Wanchoo	Hán - Tạng	-	-	-	Ấn Độ	[67]



## STUDY ON MITOCHONDRIAL VARIATIONS OF THE SANDIU ETHNIC GROUP LIVING IN THAI NGUYEN

Tran Huu Dinh<sup>1,2</sup>, Vo Thi Bich Thuy<sup>1</sup>, Nguyen Thuy Duong<sup>1,2\*</sup>

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

<sup>2</sup>Graduate University of Science and Technology - Vietnam Academy of Science and Technology

ARTICLE INFO	ABSTRACT
<b>Received:</b> 05/6/2024 <b>Revised:</b> 11/7/2024 <b>Published:</b> 17/7/2024	The mitochondrion is an organelle involved in cellular energy production. The human mitochondrial genome (mtDNA) is widely used in research on genetic diversities and mitochondrial genetic diseases in different populations. In this study, total DNA from 26 individuals of Sandiu ethnic living in Phu Luong district of Thai Nguyen province was used for whole mtDNA sequencing using next-generation sequencing. By comparing the obtained sequences with the reference mitochondrial genome rCRS (revised Cambridge Reference Sequence), an average of $40.2 \pm 5.2$ variants per sample was detected, of which 53.8% of Sandiu individuals possess the variant T489C in the hypervariable region III (HVS-III) specific for Southeast Asia ethnic groups. Further analysis of genetic distances between six Sino-Tibetan ethnics (Sandiu, Hanhi, Phula, Sila, Lahu, and Lolo) revealed that although these ethnics are living in the Northern part of Vietnam, the genetic distances varied among different populations (0.011 - 0.183). This is the first study on the mtDNA variation of the Sandiu ethnic in Vietnam. These results are valuable data, contributing to future studies on the population genetics and mitochondrial diseases of the Sandiu and ethnic minorities in Vietnam.

### KEYWORDS

Variant  
San Diu  
Population genetics  
Mitochondrial genome  
Vietnam

## PHÂN TÍCH BIẾN THỂ HỆ GEN TY THỂ CÁC CÁ THỂ DÂN TỘC SÁN DÌU SINH SỐNG TẠI TỈNH THÁI NGUYÊN

Trần Hữu Định<sup>1,2</sup>, Võ Thị Bích Thủy<sup>1</sup>, Nguyễn Thùy Dương<sup>1,2\*</sup>

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

<sup>2</sup>Học viện Khoa học và Công nghệ - Viện Hàn lâm Khoa học và Công nghệ Việt Nam

THÔNG TIN BÀI BÁO	TÓM TẮT
<b>Ngày nhận bài:</b> 05/6/2024 <b>Ngày hoàn thiện:</b> 11/7/2024 <b>Ngày đăng:</b> 17/7/2024	Ty thể là bào quan tham gia vào quá trình sản sinh năng lượng cho tế bào. Hệ gen ty thể (mitochondrial DNA - mtDNA) ở người được sử dụng rộng rãi trong nghiên cứu đa dạng di truyền quần thể và bệnh học. Trong nghiên cứu này, DNA tổng số tách chiết từ 26 cá thể người Sán Dìu sinh sống tại huyện Phú Lương, tỉnh Thái Nguyên được sử dụng để giải trình tự mtDNA bằng phương pháp giải trình tự gen thế hệ mới. Khi so sánh với trình tự hệ gen ty thể tham chiếu rCRS (revised Cambridge Reference Sequence), số lượng biến thể tìm thấy trên từng cá thể trung bình khoảng $40,2 \pm 5,2$ , trong đó 53,8% quần thể nghiên cứu mang biến thể T489C trên vùng siêu biến III (hypervariable region III - HVS-III) đặc trưng ở người Đông Nam Á. Phân tích khoảng cách di truyền giữa sáu dân tộc (Sán Dìu, Hà Nhì, Phù Lá, Si La, La Hủ và Lô Lô) cho thấy mặc dù các nhóm dân tộc này chỉ tập trung ở một số tỉnh phía Bắc Việt Nam nhưng lại có khoảng cách di truyền theo cặp dao động khác nhau (0,011 - 0,183). Đây là nghiên cứu đầu tiên về biến thể mtDNA của người Sán Dìu tại Việt Nam. Các kết quả nhận được là cơ sở dữ liệu hệ gen ty thể quan trọng, đóng góp vào các nghiên cứu về di truyền quần thể và di truyền bệnh học cho người dân tộc Sán Dìu nói riêng và các dân tộc người Việt Nam nói chung.

DOI: <https://doi.org/10.34238/tnu-jst.10545>

\* Corresponding author. Email: [tdnguyen@igr.ac.vn](mailto:tdnguyen@igr.ac.vn)

## 1. Giới thiệu

Ngữ hệ Hán - Tạng là một trong năm ngữ hệ lớn ở Việt Nam, có khoảng 1 triệu người sinh sống tại Việt Nam và thuộc về chín dân tộc thiểu số (Hà Nhì, Phù Lá, Si La, La Hủ, Lô Lô, Cống, Hoa, Sán Dìu và Ngái) [1]. Trong đó, dân tộc Sán Dìu có khoảng 183.000 người, chiếm 2,2% số người Hán - Tạng ở Việt Nam và chỉ xếp sau dân tộc Hoa về số lượng người [2]. Họ cư trú chủ yếu ở các vùng trung du của một số tỉnh miền Bắc như Thái Nguyên, Vĩnh Phúc và Tuyên Quang.

Ty thể là một bào quan quan trọng ở động vật, tham gia vào nhiều quá trình trong tế bào, bao gồm chuyển hoá trong tế bào, sản sinh năng lượng, duy trì cân bằng nội môi, ... [3]. Bào quan này sở hữu vật chất di truyền dưới dạng DNA kép mạch vòng (hệ gen ty thể), tồn tại độc lập với hệ gen nhân của tế bào. Hệ gen ty thể (mtDNA) ở người có kích thước 16596 bp, bao gồm 2 vùng chính là vùng điều khiển (D-loop) và vùng mã hoá [4]. Vùng D-loop là vùng không mã hoá, với kích thước hơn 1100 bp, đóng vai trò điều hoà cho quá trình tái bản và phiên mã của mtDNA nhờ chứa điểm bắt đầu (origin) của quá trình tái bản cho mạch liên tục (leading strand) và vùng khởi động (promoter) cho cả chuỗi nặng và chuỗi nhẹ [5]. Vùng này bao gồm ba vùng siêu biến (Hypervariable region: HVS-I, HVS-II và HVS-III). Vùng mã hoá có kích thước khoảng 15000 bp với 37 gen chức năng, gồm 2 gen mã hoá cho rRNA, 22 gen mã hoá cho tRNA và 13 gen mã hoá cho các protein khác nhau. Những protein này tham gia cấu thành nên các phức hợp trong quá trình phosphoryl hoá oxy hoá và vận chuyển điện tử từ các chất dinh dưỡng thành dạng năng lượng mà tế bào có thể hấp thụ là ATP, trong đó có bảy gen (*NADH* dehydrogenase subunit: *ND1*, *ND2*, *ND3*, *ND4*, *ND4L*, *ND5* và *ND6*) mã hóa cho các tiểu phần của phức hệ I, ba gen mã hóa cho phức hợp IV (Cytochrome c oxidase subunit: *COX1*, *COX2* và *COX3*), hai gen mã hóa cho phức hợp V (ATP Synthase Membrane Subunit: *ATP6* và *ATP8*) và một gen mã hóa cho phức hợp III (Cytochrome b: *CYTB*). Hệ gen ty thể không có hiện tượng tái tổ hợp và di truyền theo dòng mẹ nên được sử dụng rất phổ biến trong nghiên cứu đa dạng quần thể và pháp y [6], [7]. Các nghiên cứu về mối liên quan của các biến thể trên mtDNA với các bệnh ty thể cũng được thực hiện trên nhiều quần thể người khác nhau [8]-[10].

Trong những năm gần đây, nhiều nghiên cứu về các biến thể trên vùng điều khiển của hệ gen ty thể đã được thực hiện trên nhiều nhóm dân tộc thiểu số người Việt Nam. Năm 2018, Nguyễn Thy Ngọc và cộng sự đã phân tích biến thể trên hai đoạn siêu biến HVS-I và HVS-II của các mẫu dân tộc Kinh và Mảng. Kết quả chỉ ra rằng hai dân tộc này dù cùng thuộc nhóm ngữ hệ Nam Á nhưng vẫn tồn tại những khác biệt biến thể đáng chú ý [11]. Nghiên cứu so sánh các biến thể xuất hiện phổ biến giữa các dân tộc Kinh, La Hủ và Lô Lô cũng trên hai đoạn siêu biến này cho thấy có từ 13 đến 24 biến thể khác biệt theo từng cặp [12]. Năm 2020, nghiên cứu trên vùng HVS-I và HVS-II của 517 người thuộc các dân tộc Kinh, Khmer, Mường và Chăm cũng đã tìm thấy các đa hình phổ biến trên từng nhóm người [13]. Ngoài ra, các nghiên cứu trên vùng mã hóa cũng đã được tiến hành. Trong đó, nổi bật nhất phải kể đến nghiên cứu của Nguyễn Minh Tùng và các cộng sự khi đã giải trình tự thành công hai gen *ND5* và *ND6* ở 54 người dân tộc Giarai và Êđê sinh sống ở tỉnh Tây Nguyên và bước đầu cho thấy tính đa hình trình tự hai gen mã hóa này [14]. Nghiên cứu phát hiện và định lượng đột biến G11778A trên gen *ND6* liên quan đến hội chứng liệt thân kinh thị giác di truyền Leber (LHON) bằng cách sử dụng kết hợp kỹ thuật PCR-RFLP và real-time PCR cũng đã được thực hiện trên nhóm 149 bệnh nhân tại bệnh viện Nhi Trung ương. Kết quả đã phát hiện một bệnh nhân mang đột biến G11778A ở dạng không đồng nhất với tỷ lệ đột biến là  $2,71 \pm 0,12\%$  [15]. Ngoài ra, các nghiên cứu trên toàn bộ mtDNA cũng được thực hiện trên một số dân tộc thuộc năm ngữ hệ của Việt Nam [7], [16]. Các nghiên cứu này đã cung cấp bằng chứng về di truyền phân tử đầu tiên có liên quan tới khoảng thời gian ra đời của nền văn hoá Đông Sơn [7] và giải thích được sự phát triển của nhóm ngữ hệ Nam Đảo ở Việt Nam [16].

Tuy nhiên, cho tới nay chưa có một nghiên cứu nào về trình tự hệ gen ty thể của người dân tộc Sán Dìu. Do đó, trong nghiên cứu này, chúng tôi thực hiện giải trình tự hệ gen ty thể của dân tộc này. Sau đó, hệ gen ty thể các cá thể dân tộc Sán Dìu được phân tích và so sánh với các dân tộc

khác thuộc cùng ngữ hệ đã được công bố trước đây gồm các dân tộc Hà Nhì, Phù Lá, Si La, La Hủ và Lô Lô.

## 2. Đối tượng và phương pháp nghiên cứu

### 2.1. Đối tượng nghiên cứu

Đối tượng nghiên cứu là các mẫu máu ngoại vi của 26 cá thể người Sán Dìu khỏe mạnh được thu thập tại huyện Phú Lương, tỉnh Thái Nguyên, năm 2022. Các cá thể được lựa chọn cho nghiên cứu là những người không cùng huyết thống trong vòng 3 đời cũng như có ông bà, bố mẹ thuộc cùng một dân tộc. Người tham gia đều được ký xác nhận vào đơn đồng ý cung cấp mẫu máu cho nghiên cứu trước khi thu thập mẫu máu. Việc tiến hành lấy mẫu và nghiên cứu trên người đã được Hội đồng đạo đức trong nghiên cứu y sinh học của Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đồng ý thông qua (Số: 9-2019/NCHG-HĐĐĐ).

### 2.2. Giải trình tự và phân tích hệ gen ty thể

DNA tổng số được tách chiết từ các mẫu máu bằng kit GeneJET Whole Blood Genomic DNA Purification (ThermoFisher Scientific Inc., USA) theo hướng dẫn của nhà sản xuất. Quá trình xây dựng thư viện giải trình tự, bắt giữ và làm giàu hệ gen ty thể được thực hiện như mô tả trong các nghiên cứu trước đây [17], [18]. Sau đó, thư viện được giải trình tự trên hệ thống NovaSeq 6000 (Illumina, CA, USA) với các đoạn đọc hai đầu có kích thước 150 bp. Kết quả trình tự thu được sẽ được kiểm tra chất lượng và xử lý như mô tả trong nghiên cứu trước đây [19]. Các biến thể trên mtDNA sẽ được phát hiện nhờ công cụ Haplogrep2 [20] với trình tự tham chiếu revised Cambridge Reference Sequence (rCRS) trên GenBank (NC\_012920.1) và cây phân loại Phylotree 17 [21]. Các biến thể tìm thấy trên dân tộc Sán Dìu được so sánh với các dân tộc thuộc cùng ngữ hệ Hán - Tạng gồm Hà Nhì, Phù Lá, Si La, Lô Lô và La Hủ [7]. Ảnh hưởng của các điểm biến thể lên cấu trúc và chức năng của protein được dự đoán *in silico* bằng các công cụ dự đoán tin sinh, bao gồm SIFT [22], PolyPhen2 [23], CADD [24], MutPred2 [25] và MitImpact [26].

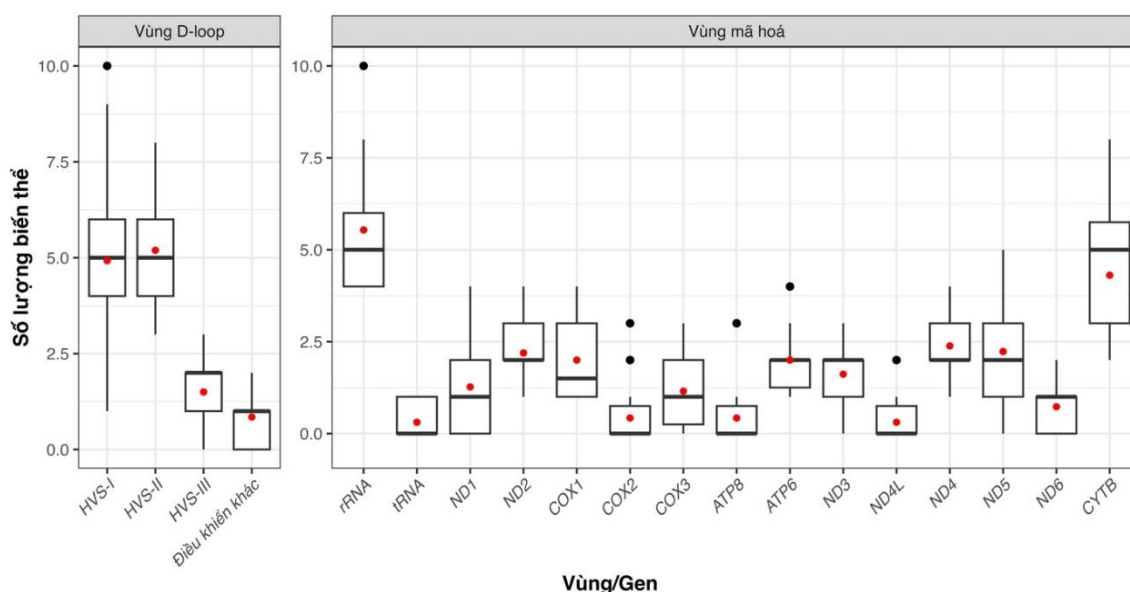
### 2.3. Phân tích thống kê các biến thể gen tìm được

Tần số của các biến thể trên mtDNA của các mẫu dân tộc Sán Dìu và các dân tộc khác được so sánh với nhau để tìm ra các biến thể khác biệt có ý nghĩa thống kê bằng cách sử dụng công cụ R v4.2.2 với package “stats” [27]. Kiểm định Fisher’s exact (2 phía) được sử dụng để so sánh tỷ lệ tần số biến thể theo cặp giữa các dân tộc trong nghiên cứu. Mức ý nghĩa được lấy với  $\alpha = 0,05$ . Các đồ thị được biểu diễn bằng package “ggplot2” trong R [28]. Khoảng cách di truyền giữa các dân tộc được tính toán bằng phần mềm Arlequin [29] với phép hoán vị 10000.

## 3. Kết quả và bàn luận

### 3.1. Đặc điểm di truyền hệ gen ty thể dân tộc Sán Dìu

Hệ gen ty thể của 26 cá thể dân tộc Sán Dìu được giải trình tự với độ bao phủ sâu trung bình 202X (92X - 480X). Số lượng base không xác định được trong quá trình giải trình tự (base N) xuất hiện ở ít hơn 1% vị trí trên tất cả các mẫu. Mỗi mẫu dân tộc Sán Dìu có số lượng biến thể trung bình là khoảng  $40,2 \pm 5,2$ . Trên vùng D-loop, số lượng biến thể tập trung nhiều nhất ở vùng siêu biến HVS-I và HVS-II với lượng biến thể trung bình lần lượt là  $4,9 \pm 2,0$  và  $5,2 \pm 1,3$  (Hình 1). Vùng HVS-I có số lượng biến thể ở các mẫu nghiên cứu rất khác nhau, trong đó cá thể có số lượng biến thể thấp nhất là 1 và cao nhất là 10. Trên vùng mã hoá, biến thể xuất hiện nhiều nhất ở vùng rRNA (12S và 16S) với trung bình  $5,5 \pm 1,3$  biến thể và thấp nhất ở các đoạn gen mã hoá cho 22 tRNA khác nhau với trung bình  $0,3 \pm 0,5$  biến thể trên một mẫu. Ở các gen mã hoá cho các protein chức năng, gen *CYTB* là gen mang nhiều biến thể nhất với  $4,3 \pm 1,9$  biến thể. Những vùng còn lại có ít biến thể hơn với 2,5 biến thể trên mỗi vùng (Hình 1).

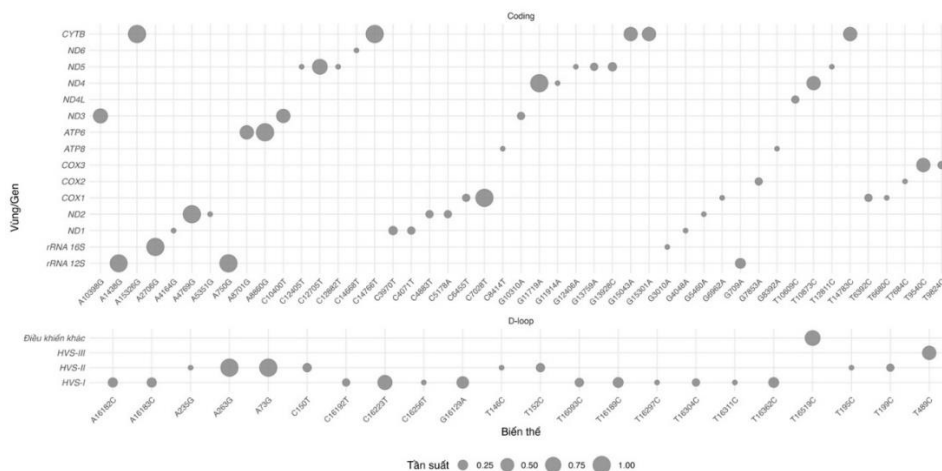


**Hình 1.** Số lượng biến thể trên hệ gen ty thể của dân tộc Sán Dìu chia theo từng vùng/gen. Chấm tròn màu đỏ thể hiện số lượng biến thể trung bình trên từng vùng của ty thể. Chấm tròn màu đen thể hiện điểm dữ liệu nằm cách xa các điểm dữ liệu khác (giá trị ngoại lai).

### 3.2. Các biến thể phổ biến trên hệ gen ty thể dân tộc Sán Dìu

Phân tích hệ gen ty thể của người Sán Dìu, chúng tôi phát hiện được 69 biến thể có tần số xuất hiện phổ biến  $\geq 10\%$ , trong đó 22 biến thể nằm trên vùng điều khiển và 47 biến thể nằm trên vùng mã hóa (Hình 2). Ở những biến thể đã nêu trên, mười biến thể, gồm hai biến thể thuộc vùng D-loop (A73G và A263G) và tám biến thể thuộc vùng mã hóa (A2706G, A750G, A4769G, C7028T, A8860G, G11719A, C14766T và A15326G), được tìm thấy ở tất cả các cá thể người Sán Dìu. Đặc biệt, năm biến thể trong số đó (A73G, A2706G, C7028T, G11719A, C14766T) thuộc về nhánh đơn bội M đặc trưng cho khu vực Châu Á và được tìm thấy nhiều ở các nước như Việt Nam, Thái Lan và Lào [11]-[13], [30].

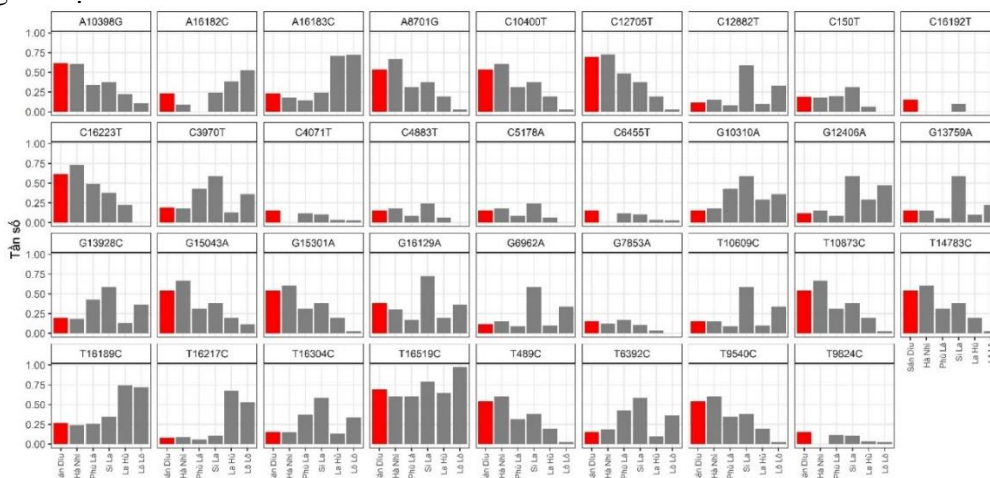
Khi phân tích sâu hơn về các biến thể trên vùng D-loop, chúng tôi nhận thấy biến thể T489C trên đoạn siêu biến III (HVS-III) xuất hiện ở 14/26 cá thể người Sán Dìu (chiếm 53,8%) là biến thể đặc trưng cho các nhóm dân tộc Đông Nam Á [31]. Biến thể C16192T trên đoạn siêu biến I (HVS-I) xuất hiện phổ biến ở dân tộc Sán Dìu (15,4%). Biến thể này cũng được tìm thấy ở dân tộc Si La (10,3%) nhưng không xuất hiện ở các dân tộc Hà Nhì, Phù Lá, La Hủ và Lô Lô trong ngữ hệ Hán - Tạng. Ngoài ra, phân tích biến thể trên các gen mã hóa cho protein chức năng cho thấy biến thể C12705T nằm trên gen ND5 là biến thể có tần số xuất hiện lớn nhất (69,2%). Đây là yếu tố di truyền có khả năng bảo vệ khỏi bệnh Glomcom nhãn áp [32]. Ngoài ra, ba biến thể đồng hoán (T14783C, G15043A và G15301A) không làm thay đổi amino acid được tìm thấy trên gen CYTB với tần số 53,8%. Các biến thể này được dự đoán ít ảnh hưởng tới chức năng của protein và ít có khả năng gây bệnh, tuy nhiên, trên cơ sở dữ liệu ClinVar, chúng lại được báo cáo liên quan tới bệnh ung thư vú có tính chất gia đình. Bên cạnh đó, cũng trên gen này, biến thể C14766T làm thay đổi amino acid Threonine thành Isoleucine ở vị trí thứ 7 trên chuỗi polypeptide mã hóa bởi gen CYTB được tìm thấy trên tất cả các cá thể. Biến thể này được dự đoán có khả năng gây bệnh với điểm SIFT là 0,03. Tuy nhiên một số phần mềm khác như Polyphen2, CADD, MutPred2, MitImpact đều cho rằng biến thể này không có khả năng gây bệnh. Trên thực tế, biến thể này được ghi nhận là có liên quan tới các bệnh phù phổi ở địa hình cao trên người Ấn Độ [33].



**Hình 2.** Sự phân bố và tần số các biến thể phổ biến trên hệ gen ty thể của người Sán Dìu

**3.3. So sánh thống kê các biến thể với các dân tộc cùng ngữ hệ**

Khi phân tích thống kê sử dụng kiểm định Fisher's exact, trong 69 biến thể phổ biến trên 10% tìm được, 35 biến thể có tần số khác nhau có ý nghĩa thống kê giữa dân tộc Sán Dìu và các dân tộc khác thuộc cùng ngữ hệ Hán - Tạng gồm Hà Nhì, Phù Lá, Si La, Lô Lô và La Hủ ( $p < 0,05$ ) (Hình 3). Hai dân tộc Hà Nhì và Phù Lá có số biến thể khác biệt tương đối ít với dân tộc Sán Dìu với số biến thể lần lượt là 5 và 7 biến thể. Ba dân tộc còn lại (La Hủ, Si La và Lô Lô) có số biến thể khác biệt tương ứng 17, 18 và 22 với dân tộc Sán Dìu. Cụ thể, một số biến thể (T489C, C16223T, A8701G, T9540C, A10398G, C10400T, T10873C, C12705T, T14783C, G15043A, G15301A) xuất hiện ít ở La Hủ và Lô Lô nhưng lại xuất hiện nhiều ở Sán Dìu ( $p < 0,05$ ). Nhìn chung, các đa hình này không phổ biến ở các nước Đông Nam Á như Thái Lan và Lào [30]. Các biến thể (A16183C, T16189C, T16217C) lại phổ biến ở La Hủ và Lô Lô (tần số  $\geq 50\%$ ) cũng xuất hiện ở dân tộc Sán Dìu với tần số khá lớn khoảng 20% ( $p < 0,05$ ). Những biến thể này xuất hiện với tần số khoảng 20% ở hai dân tộc Phù Lá và Hà Nhì cũng như ở người Hán Quảng Đông [34], tuy nhiên, chúng lại xuất hiện phổ biến ở người Kinh (~50%) [11], [13]. Kết quả so sánh cũng cho thấy năm biến thể G16129A, G13759A, C12882T, T10609C và G6962A xuất hiện với tần số cao ( $\geq 58\%$ ) ở quần thể người Si La và khác biệt có ý nghĩa thống kê với ít nhất 4 trong 5 các dân tộc Hán - Tạng còn lại.



**Hình 3.** So sánh thống kê các biến thể ở dân tộc Sán Dìu với các dân tộc khác thuộc ngữ hệ Hán - Tạng. Tần số xuất hiện của các biến thể trên dân tộc Sán Dìu được kí hiệu bằng cột có màu đỏ

### 3.4. Khoảng cách di truyền giữa dân tộc Sán Dìu và các dân tộc cùng ngữ hệ

Khoảng cách di truyền giữa 6 dân tộc (Sán Dìu, Hà Nhì, Phù Lá, Si La, Lô Lô và La Hủ) được mô tả ở bảng 1. Kết quả cho thấy các dân tộc này đều thuộc cùng ngữ hệ và cư trú ở các vùng trung du và miền núi phía Bắc nhưng lại có những sự khác biệt đáng chú ý về khoảng cách di truyền. Cụ thể, khoảng cách di truyền lớn nhất được tìm thấy ở dân tộc Hà Nhì và Lô Lô ( $\phi_{ST} = 0,183$ ) và khoảng cách di truyền nhỏ nhất ở dân tộc Hà Nhì và Sán Dìu ( $\phi_{ST} = 0,011$ ). Dân tộc Si La có khoảng cách di truyền lớn đối với 5 dân tộc còn lại (0,111 - 0,176). Dân tộc Sán Dìu có khoảng cách di truyền tương đối gần gũi với hai dân tộc Hà Nhì và Phù Lá với giá trị  $\phi_{ST}$  lần lượt là 0,011 và 0,042, nhưng lại có khoảng cách di truyền lớn hơn với ba dân tộc còn lại Si La, Lô Lô và La Hủ (0,104 - 0,151). Nghiên cứu về văn hóa và lịch sử trước đây cho rằng dân tộc Sán Dìu có nguồn gốc từ tỉnh Quảng Đông, Trung Quốc [1], tuy nhiên, chưa có bất kỳ một nghiên cứu nào về di truyền của dân tộc Sán Dìu. Bên cạnh đó, nghiên cứu gần đây về sự đa dạng các điểm đa hình mtDNA của các nhóm người Hán ở Trung Quốc đã cho thấy rằng nhóm người Hán ở Quảng Đông có sự gần gũi về mặt di truyền gần với các nhóm người ở phía Nam như Quảng Tây, Hồ Nam, Phúc Kiến (0,00019 - 0,00077) hơn so với các nhóm người phía Bắc như Bắc Kinh, Tân Cương và Ninh Hạ (0,00259 - 0,00416) [35]. Trong nghiên cứu này, chúng tôi đã giải trình tự hệ gen ty thể của một số cá thể thuộc dân tộc Sán Dìu và bước đầu phân tích khoảng cách di truyền của dân tộc này với các dân tộc khác trong cùng ngữ hệ. Tuy nhiên, các nghiên cứu và phân tích sâu trên hệ gen nhân cũng như nhiễm sắc thể Y của dân tộc Sán Dìu cần phải được thực hiện để làm sáng tỏ nguồn gốc và sự di cư của dân tộc này.

**Bảng 1.** Khoảng cách di truyền giữa dân tộc Sán Dìu và các dân tộc thuộc cùng ngữ hệ Hán - Tạng

	Sán Dìu	Hà Nhì	Phù Lá	Si La	Lô Lô	La Hủ
<b>Sán Dìu</b>						
<b>Hà Nhì</b>	0,011**					
<b>Phù Lá</b>	0,042	0,073**				
<b>Si La</b>	0,111**	0,119**	0,131**			
<b>Lô Lô</b>	0,151**	0,183**	0,145**	0,169**		
<b>La Hủ</b>	0,104**	0,129**	0,116**	0,176**	0,102**	

Ghi chú: \*\*: giá trị  $p < 0,01$

### 4. Kết luận

Trong nghiên cứu này, chúng tôi thực hiện giải trình tự hệ gen ty thể của 26 cá thể thuộc dân tộc Sán Dìu bằng phương pháp giải trình tự thế hệ mới. Đây là nghiên cứu đầu tiên về các biến thể trên hệ gen ty thể của người dân tộc Sán Dìu. Mặc dù cùng thuộc một ngữ hệ, đặc điểm hệ gen ty thể của người Sán Dìu vẫn có những khác biệt đáng chú ý với những dân tộc khác trong cùng ngữ hệ Hán - Tạng. Nghiên cứu này cung cấp các dữ liệu tham chiếu quan trọng, làm cơ sở tham khảo cho các nghiên cứu tiếp theo về di truyền quần thể và bệnh ty thể của dân tộc Sán Dìu cũng như các dân tộc thiểu số khác ở Việt Nam.

### Lời cảm ơn

Chúng tôi cũng xin chân thành cảm ơn những người đã tham gia hiến mẫu cho nghiên cứu. Công trình này được hoàn thành với sự tài trợ từ đề tài thuộc Bộ Khoa học và Công nghệ mã số ĐTĐL.CN.60/19. Trần Hữu Định được tài trợ bởi chương trình học bổng Thạc sĩ, Tiên sĩ trong nước của Quỹ đổi mới sáng tạo (VINIF), Viện Nghiên cứu Dữ liệu lớn mã số VINIF.2023.ThS.034.

### TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] N. V. Dang, T. S. Chu, and H. Luu, *Ethnic Minorities in Vietnam*. The Gioi (in English), 2017.
- [2] General Statistics Office of Vietnam, "Completed results of the 2019 Vietnam population and housing census.", Statistical Publishing House, 2019. [Online]. Available: <https://www.gso.gov.vn/>. [Accessed: Apr. 12, 2024].

- [3] W. Chen, H. Zhao, and Y. Li, "Mitochondrial dynamics in health and disease: mechanisms and potential targets," *Signal Transduction and Targeted Therapy*, vol. 8, no. 1, p. 333, 2023.
- [4] M. Habbane, J. Montoya, T. Rhouda *et al.*, "Human Mitochondrial DNA: Particularities and Diseases," *Biomedicines*, vol. 9, no. 10, p. 1364, 2021.
- [5] J. W. Taanman, "The mitochondrial genome: structure, transcription, translation and replication," *Biochim Biophys Acta*, vol. 1410, no. 2, pp. 103-123, 1999.
- [6] S. Oliveira, A. Hubner, A. M. Fehn *et al.*, "The role of matrilineality in shaping patterns of Y chromosome and mtDNA sequence variation in southwestern Angola," *European Journal of Human Genetics*, vol. 27, no. 3, pp. 475-483, 2019.
- [7] N. T. Duong, E. Macholdt, N. D. Ton *et al.*, "Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia," *Scientific Reports*, vol. 8, no. 1, p. 11651, 2018.
- [8] J. Liu, L. D. Wang, Y. B. Sun *et al.*, "Deciphering the signature of selective constraints on cancerous mitochondrial genome," *Molecular Biology and Evolution*, vol. 29, no. 4, pp. 1255-1261, 2012.
- [9] Y. Wu, X. H. Wang, X. H. Li *et al.*, "Common mtDNA variations at C5178a and A249d/T6392C/G10310A decrease the risk of severe COVID-19 in a Han Chinese population from Central China," *Military Medical Research*, vol. 8, no. 1, p. 57, 2021.
- [10] M. Venter, C. Tomas, I. S. Pienaar *et al.*, "MtDNA population variation in Myalgic encephalomyelitis/Chronic fatigue syndrome in two populations: a study of mildly deleterious variants," *Scientific Reports*, vol. 9, no. 1, p. 2914, 2019.
- [11] T. N. Nguyen, B. T. Nguyen, Q. H. Nguyen *et al.*, "Single Nucleotide Polymorphisms in the D-Loop Region of the Mitochondrial Genomes of Individuals from Two Ethnic Groups Kinh and Mang of Austro-Asiatic Language Family," *Vietnam Journal of Biotechnology*, vol. 16, no. 2, pp. 231-240, 2018.
- [12] T. D. Nguyen, V. P. Nguyen, T. N. Nguyen, and V. H. Nong, "Study on genetic variations of the D-loop region in three Vietnamese ethnic groups Kinh, Lolo, and Lahu," *Vietnam Journal of Biotechnology*, vol. 18, no. 2, pp. 231-238, 2020.
- [13] T. T. H. Tran, D. H. Nguyen, V. K. Tran *et al.*, "Variation of Mitochondrial DNA HV1 AND HV2 of the Vietnamese Population," *Advances in Experimental Medicine and Biology*, vol. 1292, pp. 37-63, 2020.
- [14] M. T. Nguyen, V. B. Nguyen, and D. T. Nguyen, "Mitochondrial NADH dehydrogenase subunit 5 and subunit 6 gene polymorphisms in Giarai and Ede ethnic people living in the central highlands," *Vietnam Medical Journal*, vol. 504, no. 1, pp. 20-25, 2021.
- [15] B. K. Phung, V. M. Nguyen, T. H. L. Nguyen *et al.*, "Detection and quantitation of mitochondrial G11778A mutation of LHON syndrome in a Vietnamese patient with tentatively diagnosed mitochondrial disease," *VNU Journal of Science: Natural Sciences and Technology*, vol. 33, no. 2S, pp. 20-25, 2017.
- [16] D. H. Thao, T. H. Dinh, S. Mitsunaga *et al.*, "Investigating demic versus cultural diffusion and sex bias in the spread of Austronesian languages in Vietnam," *PLoS One*, vol. 19, no. 6, 2024, Art. no. e0304964.
- [17] T. Maricic, M. Whitten, and S. Paabo, "Multiplexed DNA sequence capture of mitochondrial genomes using PCR products," *PLoS One*, vol. 5, no. 11, 2010, Art. no. e14004.
- [18] M. Meyer and M. Kircher, "Illumina sequencing library preparation for highly multiplexed target capture and sequencing," *Cold Spring Harbor Protocols*, vol. 2010, no. 6, 2010, Art. no. pdbprot5448.
- [19] L. Arias, C. Barbieri, G. Barreto *et al.*, "High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia," *American Journal of Physical Anthropology*, vol. 165, no. 2, pp. 238-255, 2018.
- [20] H. Weissensteiner, D. Pacher, A. Kloss-Brandstatter *et al.*, "HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing," *Nucleic Acids Research*, vol. 44, no. W1, pp. W58-63, 2016.
- [21] M. van Oven and M. Kayser, "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation," *Human Mutation*, vol. 30, no. 2, pp. E386-394, 2009.
- [22] N. L. Sim, P. Kumar, J. Hu, S. Henikoff *et al.*, "SIFT web server: predicting effects of amino acid substitutions on proteins," *Nucleic Acids Res*, vol. 40, no. Web Server, pp. W452-457, 2012.
- [23] I. A. Adzhubei, S. Schmidt, L. Peshkin *et al.*, "A method and server for predicting damaging missense mutations," *Nat Methods*, vol. 7, no. 4, pp. 248-249, 2010.
- [24] M. Kircher, D. M. Witten, P. Jain, B. J. O'Roak *et al.*, "A general framework for estimating the relative pathogenicity of human genetic variants," *Nat Genet*, vol. 46, no. 3, pp. 310-315, 2014.

- [25] V. Pejaver, J. Urresti, J. Lugo-Martinez *et al.*, "Inferring the molecular and phenotypic impact of amino acid variants with MutPred2," *Nat Commun*, vol. 11, no. 1, p. 5918, 2020.
- [26] S. Castellana, J. Ronai, and T. Mazza, "MitImpact: an exhaustive collection of pre-computed pathogenicity predictions of human mitochondrial non-synonymous variants," *Hum Mutat*, vol. 36, no. 2, pp. E2413-22, 2015.
- [27] R Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, 2022. [Online]. Available: <https://www.R-project.org/>. [Accessed Apr. 12, 2024].
- [28] W. Hadley, "ggplot2: Elegant Graphics for Data Analysis," Springer-Verlag New York, 2016. [Online]. Available: <https://ggplot2.tidyverse.org>. [Accessed Apr. 12, 2024].
- [29] L. Excoffier and H. E. Lischer, "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows," *Molecular Ecology Resources*, vol. 10, no. 3, pp. 564-567, 2010.
- [30] W. Kutanan, J. Kampuansai, M. Srikumool *et al.*, "Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages," *Human Genetics*, vol. 136, no. 1, pp. 85-98, 2017.
- [31] J. Vanichanukulyakit, S. Khacha-Ananda, T. Monum *et al.*, "The Analysis of Genetic Polymorphism on Mitochondrial Hypervariable Region III in Thai Population," *Genes (Basel)*, vol. 14, no. 3, p. 682, 2023.
- [32] A. Piotrowska-Nowak, E. Kosior-Jarecka, A. Schab *et al.*, "Investigation of whole mitochondrial genome variation in normal tension glaucoma," *Experimental Eye Research*, vol. 178, pp. 186-197, 2019.
- [33] S. Sharma, S. Singh, R. K. Gupta *et al.*, "Mitochondrial DNA sequencing reveals association of variants and haplogroup M33a2'3 with High altitude pulmonary edema susceptibility in Indian male lowlanders," *Scientific Reports*, vol. 9, no. 1, p. 10975, 2019.
- [34] F. Chen, S. Y. Wang, R. Z. Zhang *et al.*, "Analysis of mitochondrial DNA polymorphisms in Guangdong Han Chinese," *Forensic Science International: Genetics*, vol. 2, no. 2, pp. 150-153, 2008.
- [35] Y. C. Li, W. J. Ye, C. G. Jiang *et al.*, "River Valleys Shaped the Maternal Genetic Landscape of Han Chinese," *Molecular Biology and Evolution*, vol. 36, no. 8, pp. 1643-1652, 2019.