

# Đa dạng di truyền vùng D-loop ở hai dân tộc Cơ-tu và Rơ-măm

Lã Đức Duy<sup>1,2</sup>, Nông Văn Hải<sup>1</sup>, Nguyễn Thùy Dương<sup>1,2\*</sup>

<sup>1</sup>Viện Sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, 18 Hoàng Quốc Việt, phường Nghĩa Đô, Hà Nội, Việt Nam

<sup>2</sup>Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, 18 Hoàng Quốc Việt, phường Nghĩa Đô, Hà Nội, Việt Nam

Ngày nhận bài 6/6/2024; ngày chuyển phân biện 10/6/2024; ngày nhận phân biện 30/6/2024; ngày chấp nhận đăng 5/7/2024

## **Tóm tắt:**

Vùng D-loop hệ gen ty thể (mitochondrial DNA - mtDNA), vùng có tần suất xuất hiện đột biến cao nhất, thường được dùng để nghiên cứu khảo sát đa dạng di truyền trên các quần thể người khác nhau. Trong nghiên cứu này, chúng tôi thu thập và tách chiết DNA tổng số từ mẫu máu của 19 cá thể dân tộc Cơ-tu và 24 cá thể dân tộc Rơ-măm. Sau khi giải trình tự mtDNA của 43 cá thể và so sánh với trình tự mtDNA tham chiếu RSRS (Reconstructed sapiens reference sequence), 41 điểm đa hình khác nhau được phát hiện trên vùng D-loop, trong đó có 27 và 30 điểm lần lượt được tìm thấy trên dân tộc Cơ-tu và Rơ-măm. Kết quả định danh nhóm đơn bội sử dụng trình tự vùng D-loop và phần mềm Haplogrep3 cho thấy có 12 nhóm khác nhau, trong đó chỉ có F1a1a xuất hiện ở cả hai dân tộc nêu trên. Kết quả phân tích ma trận khoảng cách di truyền giữa các dân tộc trong nghiên cứu và hai dân tộc cùng ngữ hệ Việt Nam khác (Kinh và Mảng) cho thấy, Cơ-tu và Rơ-măm có khoảng cách di truyền cao nhất (0,14545). Đây là kết quả nghiên cứu đầu tiên về vùng D-loop của hai dân tộc Cơ-tu và Rơ-măm sinh sống tại Việt Nam. Các kết quả mới này đã bổ sung thêm thông tin di truyền theo dòng mẹ và các biến thể vùng D-loop ở hai dân tộc nghiên cứu.

**Từ khóa:** biến thể, D-loop, Việt Nam.

**Chỉ số phân loại:** 1.6, 3.1

## Genetic diversity in the D-loop region of the Co-tu and Ro-mam ethnic groups

Duc Duy La<sup>1,2</sup>, Van Hai Nong<sup>1</sup>, Thuy Duong Nguyen<sup>1,2\*</sup>

<sup>1</sup>Institute of Biology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Nghia Do Ward, Hanoi, Vietnam

<sup>2</sup>Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Nghia Do Ward, Hanoi, Vietnam

Received 6 June 2024; revised 30 June 2024; accepted 5 July 2024

## **Abstract:**

The D-loop region of mitochondrial DNA (mtDNA), with the highest mutation frequency in the mitochondrial genome, has been frequently used in studies on genetic diversity in different human populations. In this study, total DNA was collected and extracted from whole blood samples of 19 Co-tu and 24 Ro-mam individuals. After sequencing the mitochondrial DNA of 43 study subjects and aligning them to the mitochondrial DNA reconstructed sapiens reference sequence (RSRS), a total of 41 unique variants were obtained in the D-loop region, in which 27 and 30 variants were found in the Co-tu and Ro-mam populations, respectively. Haplogroup identification based on the D-loop region revealed 12 haplogroups, among which only haplogroup F1a1a was found in both ethnic groups using Haplogrep3. After analysing the genetic distances of the study populations and other previously published populations (Kinh and Mang), the results showed that the genetic distance between Co-tu and Ro-mam (0.14545) was the highest. This is the first study on the genetic diversity of the D-loop region of the Co-tu and Ro-mam ethnic groups living in Vietnam. These results provide more insights into the maternal genetic background and D-loop variants of the two studied ethnic groups.

**Keywords:** D-loop, variants, Vietnam.

**Classification numbers:** 1.6, 3.1

\*Tác giả liên hệ: Email: tdnguyen@igr.ac.vn

## 1. Đặt vấn đề

Ngữ hệ Nam Á (NHNA) là một trong những ngữ hệ lớn nhất trên thế giới bao gồm 167 ngôn ngữ được nói khắp Nam Á, Đông Á, và Đông Nam Á [1]. Tại Nam Á, ngôn ngữ thuộc NHNA được phân loại vào nhánh phụ Monic và là một trong bốn ngữ hệ chính ở miền Nam và miền Trung Ấn Độ. Tại Đông Á, NHNA được nói rải rác ở phía Nam Trung Quốc, chẳng hạn như tiếng Bolyu và Bagan [2]. Ở Đông Nam Á, các bằng chứng về khảo cổ và ngôn ngữ học chỉ ra rằng, NHNA là ngữ hệ xuất hiện đầu tiên ở vùng Đông Nam Á lục địa (Mainland Southeast Asia - MSEA), sau đó với sự xuất hiện của các hệ ngôn ngữ khác như Nam Đảo (Austronesian, NHND), Thái - Kadai (Tai - Kadai, NHTK), Hán - Tạng (Sino - Tibetan, NHHT) và Mông - Miên (H'mong - Mien, NHHM), NHNA bị cô lập và phân tán rải rác ở Đông Nam Á [3, 4]. Vào thời điểm hiện tại, ngôn ngữ được nói chủ yếu ở vùng Đông Nam Á hải đảo (Island Southeast Asia - ISEA) thuộc về NHND còn ở MSEA thì NHNA vẫn là hệ ngôn ngữ được sử dụng nhiều nhất [1, 5]. Tại MSEA, tiếng Khmer của Campuchia và tiếng Việt của Việt Nam là hai ngôn ngữ thuộc NHNA và cũng được sử dụng bởi một số dân tộc thiểu số ở Thái Lan, Lào, Myanmar và vùng bán đảo của Malaysia. Với xấp xỉ 126 triệu người sử dụng NHNA [1], ngữ hệ này đứng thứ 8 về số lượng người nói trên thế giới và thứ 3 tại Đông Nam Á, tuy nhiên nguồn gốc và con đường phát tán của NHNA hiện vẫn là vấn đề chưa được giải quyết [5-7].

Với vị trí địa lý chiến lược kết nối bán đảo Đông Dương và ISEA, Việt Nam trở thành khu vực có bề dày lịch sử phong phú và phức tạp, được tạo nên bởi nhiều làn sóng di cư của loài người hiện đại từ nhiều hướng khác nhau. Đồng thời, Việt Nam cũng sở hữu số lượng người nói NHNA nhiều nhất trên thế giới, với khoảng 90% dân số nói ngôn ngữ thuộc ngữ hệ này (Tổng điều tra dân số và nhà ở năm 2019, www.gso.gov.vn). Trong 25 dân tộc sử dụng NHNA, dân tộc Kinh chiếm khoảng 85% số lượng người sử dụng ngôn ngữ và 24 dân tộc thiểu số còn lại chiếm khoảng 5%. Cho đến nay, các nghiên cứu nhân học phân tử (molecular anthropology) về các dân tộc NHNA ở Việt Nam còn ít, đặc biệt là các nghiên cứu trên các dân tộc thiểu số [8, 9]. Các dân tộc thiểu số thuộc NHNA ở Việt Nam sống thưa thớt, có phần tách biệt trải dài khắp cả nước, dẫn đến sự đa dạng di truyền của các dân tộc này có thể không được tìm thấy ở quần thể người Kinh, đặc biệt là các dân tộc di cư vào Việt Nam từ các nước lân cận mang theo nguồn gen có thể không có ở quần thể người bản địa.

Để nghiên cứu về đa dạng di truyền, đặc biệt làm sáng tỏ các con đường di cư, sự giao thoa giữa các dân tộc với nhau và sự khác biệt di truyền giữa các dân tộc khác nhau, hệ gen ty thể thường được lựa chọn làm đối tượng nghiên cứu [10, 11]. Trong mtDNA, vùng D-loop (vị trí nucleotide 1-576 và 16.024-16.569 của mtDNA) là vùng không mã hóa và đóng vai trò quan trọng trong việc điều hòa quá trình tái bản và phiên mã của mtDNA [12]. Vùng D-loop cũng là vùng có

tần suất xuất hiện đột biến cao nhất trên mtDNA [13], do đó vùng này thường xuyên được sử dụng để nghiên cứu khảo sát đa dạng di truyền theo dòng mẹ trên các quần thể người khác nhau [11]. Vì vậy, trong nghiên cứu này, trình tự vùng D-loop trên mtDNA của 43 cá thể thuộc hai dân tộc nói NHNA tại Việt Nam (Cơ-tu, Rơ-măm) đã được giải trình tự và phân tích. Ngoài ra, vùng D-loop của hai dân tộc Kinh và Mảng thuộc NHNA đã công bố trước đây cũng được sử dụng để phân tích, so sánh sự khác biệt giữa các dân tộc trong cùng ngữ hệ [8].

## 2. Đối tượng và phương pháp nghiên cứu

### 2.1. Đối tượng

Mẫu máu ngoại vi được thu từ các cá thể thuộc NHNA người Việt gồm 19 cá thể thuộc dân tộc Cơ-tu (Thừa Thiên Huế) và 24 cá thể thuộc dân tộc Rơ-măm (Kon Tum). Mẫu được chọn không có quan hệ huyết thống và có ít nhất ba thế hệ trong gia đình đều thuộc một dân tộc. Các cá thể tham gia nghiên cứu đều ký vào giấy đồng ý tự nguyện cho máu phục vụ nghiên cứu. Nghiên cứu này được thông qua bởi Hội đồng Đạo đức trong nghiên cứu sinh của Viện Nghiên cứu Hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam (Số 9-2019/NCHG-HĐĐĐ).

### 2.2. Giải trình tự vùng D-loop hệ gen ty thể

DNA từ bộ gen được chiết xuất bằng GeneJET Whole Blood Genomic DNA Purification Mini Kit (ThermoFisher Scientific, Hoa Kỳ) theo hướng dẫn của nhà sản xuất. Việc xây dựng thư viện bộ gen và làm giàu mtDNA đã được thực hiện như mô tả ở nghiên cứu khác [14]. Các thư viện đã được giải trình tự trên nền tảng NovaSeq 6000 (Illumina, Hoa Kỳ) với các đoạn đọc kép có chiều dài 150 bp. Các đoạn đọc được kiểm tra chất lượng bởi FastQC và được đóng hàng với trình tự tham chiếu Sapiens được sửa đổi (Reconstructed Sapiens Reference Sequence - RRSRS) [15] bằng Burrows-Wheeler Alignment (BWA). Sau đó, các trình tự D-loop sẽ được đóng hàng với nhau bằng MAFFT [16].

### 2.3. Phân tích số liệu

Trình tự vùng D-loop của 43 cá thể thuộc nghiên cứu này kết hợp với 87 trình tự đã được nghiên cứu bao gồm 50 cá thể Kinh và 37 cá thể Mảng [8, 9] được sử dụng để xác định các nhóm đơn bội bằng phần mềm Haplogrep3 [17] và cây PhyloTree mtDNA phiên bản 17.1 [18]. Khoảng cách di truyền giữa các dân tộc Cơ-tu, Rơ-măm, Kinh và Mảng được tính dựa trên chỉ số  $\Phi$ ST sử dụng phần mềm Arlequin [19]. Sự phân bố của các nhóm đơn bội trong bốn nhóm dân tộc được thể hiện dưới dạng biểu đồ phân tích tương quan (Correspondence analysis - CA) dựa vào tần suất xuất hiện của các nhóm đơn bội. Biểu đồ CA được vẽ trong R [20] sử dụng các gói hỗ trợ “vegan” và “ca”. Kiểm định Fisher’s exact (2 phía) được sử dụng để so sánh tỷ lệ xuất hiện đa hình giữa hai dân tộc Cơ-tu và Rơ-măm. Tất cả giá trị p nhỏ hơn 0,05 đều được coi là có ý nghĩa.

### 3. Kết quả và bàn luận

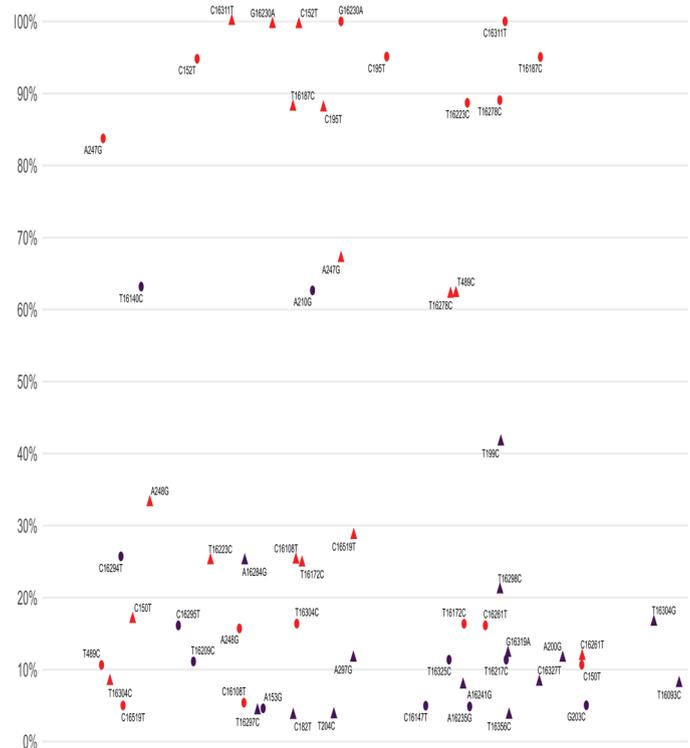
#### 3.1. Tần suất xuất hiện các điểm biến đổi ở vùng D-loop

Trong 43 cá thể tham gia nghiên cứu, chúng tôi phát hiện được 41 điểm biến đổi khác nhau ở vùng D-loop với tất cả các biến đổi đều là điểm đa hình nucleotide đơn (Single nucleotide polymorphism - SNP) (hình 1). Số lượng biến thể tối đa và tối thiểu xuất hiện trong một cá thể của mỗi dân tộc tương đối đồng đều (bảng 1). Hai điểm biến đổi C16311T và G16230A xuất hiện trên tất cả các mẫu trong nghiên cứu, đứng thứ hai là C152T được tìm thấy ở tất cả cá thể của dân tộc Rơ-măm và 95% cá thể của dân tộc Cơ-tu. Kết hợp với dữ liệu biến thể của 81.124 trình tự vùng điều khiển mtDNA thu thập từ Mitomap [20], hầu hết những điểm biến đổi xuất hiện với tần suất cao trong nghiên cứu hiện tại cũng được phát hiện ở phần lớn các trình tự trong Mitomap, ví dụ như G16230A (79.785/81.124), A247G (78.859/81.124), C16311T (67.049/81.124) và T16223C (40.381/81.124). Hai biến đổi T489C và T16140C xuất hiện ở ít trình tự hơn lần lượt là 12.479 và 2512. Ngoài ra, chỉ có hai điểm (T489C, T16140C) đã được phát hiện với tần suất cao trong vùng D-loop của các quần thể người Việt Nam thuộc NHNA đã được nghiên cứu trước đó [21].

**Bảng 1. Số lượng các điểm biến đổi tìm thấy trong nghiên cứu.**

Thứ tự	Dân tộc	Số lượng cá thể	Số lượng biến thể ở vùng D-loop		
			Giá trị tối đa giữa các cá thể	Giá trị tối thiểu giữa các cá thể	Giá trị trung bình ± độ lệch chuẩn
1	Cơ-tu	19	12	9	10,63±0,45
2	Rơ-măm	24	13	8	10,25±0,45

Ở các điểm biến đổi có tần suất thấp hơn, tần suất xuất hiện của điểm biến đổi giữa các quần thể khác nhau có sự khác biệt rõ ràng (hình 1). Cụ thể hơn, điểm biến đổi T16233C xuất hiện ở đa số người Cơ-tu (89%) trong khi chỉ xuất hiện ở 25% người Rơ-măm. Ngược lại, T489C chỉ xuất hiện ở số ít người Cơ-tu (11%) nhưng lại được tìm thấy ở phần lớn người Rơ-măm (62%). Sự khác biệt rõ ràng nhất là hai điểm biến đổi A210G và T16140C, khi hai điểm này có ở phần lớn người Cơ-tu với tần suất đều là 63% nhưng lại không xuất hiện ở người Rơ-măm. Đối với những điểm biến đổi xuất hiện ở tần suất thấp hơn 30% (phân tích trên từng dân tộc), đa số biến đổi chỉ được tìm thấy ở một dân tộc ví dụ như T16298C, C16294T, T16304G, A200G. Kết quả của kiểm định Fisher's exact cho thấy, có 7 biến đổi có tần suất xuất hiện khác biệt đáng kể giữa hai dân tộc (T16140C, A210G, T16223C, T489C, T199C, C16294T và A16284G), trong đó T16140C, A210G và T16223C có giá trị  $p < 0,0001$  (dữ liệu không được thể hiện).



**Hình 1. Tần suất xuất hiện của các biến đổi trong hai dân tộc Cơ-tu và Rơ-măm.** Mỗi điểm tương ứng với tần suất xuất hiện của các biến đổi. Màu đỏ: biến đổi xuất hiện ở cả hai nhóm dân tộc; màu tím: biến đổi chỉ xuất hiện ở một dân tộc; điểm hình tròn: biến đổi chỉ xuất hiện ở dân tộc Cơ-tu; điểm hình tam giác: biến đổi chỉ xuất hiện ở dân tộc Rơ-măm.

#### 3.2. Sự phân bố nhóm đơn bội

Tổng cộng 12 nhóm đơn bội thuộc 3 nhóm đơn bội lớn (M, N9, R) đã được xác định từ 43 cá thể thuộc 2 dân tộc Cơ-tu và Rơ-măm thông qua phần mềm Haplogrep3 (bảng 2). Phần lớn các nhóm đơn bội ở vùng D-loop của hai dân tộc này thuộc nhóm đơn bội lớn R và M, chiếm 40/43 số lượng nhóm đơn bội trong bộ mẫu nghiên cứu. Trong đó, nhóm đơn bội F1a1a là nhóm duy nhất xuất hiện ở cả hai dân tộc. Trong số các nhóm đơn bội chỉ xuất hiện ở một dân tộc, có bốn nhóm đơn bội chỉ được phát hiện trên một cá thể (B5a1d, B4m, B4c2, M7b1a1). Nhóm đơn bội B5a xuất hiện với tần suất cao nhất chiếm 25% số lượng nhóm đơn bội mẫu nghiên cứu và chỉ có ở người Cơ-tu (58%) (bảng 2). Đặc biệt, khác với người Rơ-măm, vốn gen ty thể của người Cơ-tu phần lớn là thuộc macro-nhóm đơn bội R. Đây có thể là kết quả của hiện tượng người sáng lập (Founder effect) hoặc trôi dạt di truyền (Genetic drift) kết hợp với việc sống tách biệt trong thời gian dài. Còn đối với dân tộc Rơ-măm, mặc dù dân tộc này có số lượng nhóm đơn bội ít hơn nhưng lại có số lượng nhóm đơn bội lớn nhiều hơn với sự xuất hiện của nhóm đơn bội lớn N.



### 3.3. Môi quan hệ di truyền giữa các dân tộc

Để phân tích quan hệ di truyền giữa các dân tộc trong cùng NHNA, hai dân tộc Rơ-măm và Cơ-tu đã được phân tích cùng với hai dân tộc Kinh và Mảng. Phần mềm Arlequin đã được sử dụng để tính ma trận khoảng cách  $\Phi_{ST}$  dựa vào trình tự vùng D-loop của các cá thể thuộc bốn dân tộc trên (bảng 3). Kết quả phân tích ma trận  $\Phi_{ST}$  cho thấy, giữa người Cơ-tu và người Rơ-măm có khoảng cách di truyền lớn nhất (0,14545). Khoảng cách giữa người Kinh và người Rơ-măm (0,06368) lớn hơn khoảng cách giữa người Kinh với người Cơ-tu (0,03558). Ngoài ra, khoảng cách di truyền giữa dân tộc Kinh đối với dân tộc khác thấp hơn khoảng cách di truyền giữa các dân tộc còn lại với nhau. Điều này phù hợp với thực trạng phân bố khắp cả nước với số lượng cá thể nhiều nhất của dân tộc Kinh, do đó dân tộc Kinh trở thành dân tộc có khả năng tương tác và trao đổi thông tin di truyền cao nhất với các dân tộc khác tại Việt Nam.

**Bảng 3. Ma trận khoảng cách  $\Phi_{ST}$  cho vùng D-loop giữa 4 dân tộc Cơ-tu, Rơ-măm, Mảng và Kinh.**

	Cơ-tu	Rơ-măm	Mảng	Kinh
Cơ-tu	0,00000	0,14545 (0,00198±0,0004)	0,13284 (0,00010±0,0001)	0,03558 (0,03693±0,0019)
Rơ-măm	0,14545 (0,00198±0,0004)	0,00000	0,10575 (0,00000±0,0000)	0,06368 (0,00059±0,0002)
Mảng	0,13284 (0,00010±0,0001)	0,10575 (0,00000±0,0000)	0,00000	0,02642 (0,01426±0,0013)
Kinh	0,03558 (0,03693±0,0019)	0,06368 (0,00059±0,0002)	0,02642 (0,01426±0,0013)	0,00000

Giá trị trong ngoặc đơn là giá trị p.

Phân bố của các nhóm đơn bội trong bốn dân tộc Cơ-tu, Rơ-măm, Kinh và Mảng bằng biểu đồ tương quan CA cho thấy, những nhóm đơn bội xuất hiện với tần suất cao như M73'79 và M71a1a ở các dân tộc tương ứng (Rơ-măm và Mảng) và xuất hiện ở ít nhất hai dân tộc (B5a và F1a1a) sẽ nằm ở gần phía giữa của biểu đồ, còn những nhóm đơn bội xuất hiện ở tần suất thấp như M59, D4a7 và B4c1b2c2 sẽ nằm ở phía ngoài của biểu đồ (hình 2). Mặc dù khoảng cách di truyền giữa quần thể Kinh đối với hai quần thể Cơ-tu và Rơ-măm thấp nhưng ba quần thể này lại không có nhiều nhóm đơn bội chung, thể hiện tính đặc trưng của người Cơ-tu và Rơ-măm. Sự đa dạng di truyền của các dân tộc NHNA ở Việt Nam cũng được thể hiện ở biểu đồ này với hầu hết các nhóm đơn bội chỉ tồn tại ở một dân tộc.

### 4. Kết luận

Trong nghiên cứu này, vùng D-loop của 43 cá thể từ hai quần thể người Cơ-tu và Rơ-măm thuộc NHNA tại Việt Nam đã được giải trình tự và phân tích. Hai điểm biến đổi (C16311T và G16230A) được tìm thấy ở tất cả các cá thể và

có 7 SNP có tần suất xuất hiện khác biệt đáng kể giữa hai dân tộc (T16140C, A210G, T16223C, T489C, T199C, C16294T và A16284G). 12 nhóm đơn bội phát hiện ở dân tộc Cơ-tu và Rơ-măm thuộc những nhóm đơn bội lớn thường gặp ở DNA gồm M, B và N. Thành phần nhóm đơn bội giữa hai dân tộc này tương đối khác biệt chỉ có một nhóm đơn bội chung là F1a1a. Điều này phản ánh mức độ đa dạng di truyền cao giữa hai dân tộc thiểu số Cơ-tu và Rơ-măm. Khi kết hợp với dữ liệu của dân tộc Kinh với Mảng đã công bố, khoảng cách giữa người Cơ-tu và Rơ-măm lớn hơn so với khoảng cách giữa hai dân tộc này đối với dân tộc Kinh và Mảng. Những phát hiện này sẽ mang đến cái nhìn sâu sắc mới về nhóm NHNA, bổ sung thêm dữ liệu quan trọng cho các nghiên cứu tiếp theo về ngữ hệ này nói riêng và các ngữ hệ khác của dân tộc Việt Nam nói chung.

### LỜI CẢM ƠN

Nghiên cứu được hoàn thành với sự đồng ý tham gia của những cá thể cho phép lấy mẫu và được tài trợ bởi nhiệm vụ khoa học và công nghệ quốc gia do Bộ Khoa học và Công nghệ quản lý (mã số ĐTĐL.CN.60/19). Lê Đức Duy được nhận học bổng của Chương trình học bổng đào tạo thạc sỹ, tiến sỹ trong nước của Quỹ Đổi mới sáng tạo Vingroup (VINIF), mã số VINIF.2023.ThS.027.

### TÀI LIỆU THAM KHẢO

- [1] D.M. Eberhard, G.F. Simons, C.D. Fennig (2023), *Ethnologue: Languages of The World*, Twenty-six, SIL International, 1272pp.
- [2] P. Sidwell, M. Jenny (2021), *The Languages and Linguistics of Mainland Southeast Asia*, De Gruyter Mouton, **8**, 983pp, DOI: 10.1515/9783110558142/EPUB.
- [3] P. Sidwell, R. Blench (2011), *The Austroasiatic Urheimat: The Southeastern Riverine Hypothesis*, Pacific Linguistics, 315pp.
- [4] P.S. Bellwood (2005), *First Farmers: The Origins of Agricultural Societies*, Wiley-Blackwell, 384pp.
- [5] M. Klammer (2019), "The dispersal of Austronesian languages in island South East Asia: Current findings and debates", *Language and Linguistics Compass*, **13(4)**, DOI: 10.1111/LNC3.12325.
- [6] M. Lipson, O. Cheronet, S. Mallick, et al. (2018), "Ancient genomes document multiple waves of migration in Southeast Asian prehistory", *Science*, **361(6397)**, pp.92-95, DOI: 10.1126/SCIENCE.AAT3188.
- [7] D. Tagore, F. Aghakhanian, R. Naidu, et al. (2021), "Insights into the demographic history of Asia from common ancestry and admixture in the genomic landscape of present-day Austroasiatic speakers", *BMC Biology*, **19(1)**, DOI: 10.1186/S12915-021-00981-X.
- [8] N.T. Duong, E. Macholdt, N.D. Ton, et al. (2018), "Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia", *Scientific Reports*, **8(1)**, pp.1-13, DOI: 10.1038/s41598-018-29989-0.

- [9] E. Macholdt, L. Arias, N.T. Duong, et al. (2020), “The paternal and maternal genetic history of Vietnamese populations”, *European Journal of Human Genetics*, **28(5)**, pp.636-645, DOI: 10.1038/s41431-019-0557-4.
- [10] S. Pischedda, R.B. Arca, A.G. Carballa, et al. (2017), “Phylogeographic and genome-wide investigations of Vietnamese ethnic groups reveal signatures of complex historical demographic movements”, *Scientific Reports*, **7(1)**, pp.1-15, DOI: 10.1038/s41598-017-12813-6.
- [11] N.T. Duong, N.V. Phong, N.T. Ngoc, et al. (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, DOI: 10.15625/1811-4989/18/2/15136.
- [12] J.W. Taanman (1999), “The mitochondrial genome: Structure, transcription, translation and replication”, *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1410(2)**, pp.103-123, DOI: 10.1016/S0005-2728(98)00161-3.
- [13] M. Stoneking (2000), “Hypervariable sites in the mtDNA control region are mutational hotspots”, *American Journal of Human Genetics*, **67(4)**, pp.1029-1032, DOI: 10.1086/303092.
- [14] T. Maricic, M. Whitten, S. Pääbo (2010), “Multiplexed DNA sequence capture of mitochondrial genomes using PCR products”, *PLOS ONE*, **5(11)**, DOI: 10.1371/Journal.pone.0014004.
- [15] D.M. Behar, M.V. Oven, S. Rosset, et al. (2012), “A “Copernican” reassessment of the human mitochondrial DNA tree from its root”, *American Journal of Human Genetics*, **90(4)**, DOI: 10.1016/J.AJHG.2012.03.002.
- [16] K. Katoh, D.M. Standley (2013), “MAFFT multiple sequence alignment software version 7: Improvements in performance and usability”, *Molecular Biology and Evolution*, **30(4)**, DOI: 10.1093/MOLBEV/MST010.
- [17] S. Schönherr, H. Weissensteiner, F. Kronenberg, et al. (2023), “Haplogrep 3 - An interactive haplogroup classification and analysis platform”, *Nucleic Acids Research*, **51(W1)**, pp.W263-W268, DOI: 10.1093/NAR/GKAD284.
- [18] M.V. Oven, M. Kayser (2009), “Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation”, *Human Mutation*, **30(2)**, pp.E386-E394, DOI: 10.1002/HUMU.20921.
- [19] L. Excoffier, H.E.L. Lischer (2010), “Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows”, *Molecular Ecology Resources*, **10**, pp.564-567, DOI: 10.1111/j.1755-0998.2010.02847.x.
- [20] R.C. Team (2021), “R: A language and environment for statistical computing”, <https://www.r-project.org/>, accessed 31 December 2021.
- [21] N.T. Ngoc, N.B. Trang, N.Q. Huy, et al. (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, DOI: 10.15625/1811-4989/16/2/13432.
- [22] W. Kutanan, J. Kampuansai, M. Srikummool, et al. (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, DOI: 10.1007/S00439-016-1742-Y.
- [23] K. Jaisamut, R. Pitiwararom, P. Sukawutthiya, et al. (2023), “Unraveling the mitochondrial phylogenetic landscape of Thailand reveals complex admixture and demographic dynamics”, *Scientific Reports*, **13(1)**, DOI: 10.1038/S41598-023-47762-W.
- [24] R. Rajkumar, J. Banerjee, H.B. Gunturi, et al. (2005), “Phylogeny and antiquity of M macrohaplogroup inferred from complete mtDNA sequence of Indian specific lineages”, *BMC Evolutionary Biology*, **5**, DOI: 10.1186/1471-2148-5-26.
- [25] W. Kutanan, J. Kampuansai, A. Brunelli, et al. (2018), “New insights from Thailand into the maternal genetic history of Mainland Southeast Asia”, *European Journal of Human Genetics*, **26(6)**, DOI: 10.1038/S41431-018-0113-7.
- [26] A.K. Brandstätter, M. Summerer, D. Horst, et al. (2021), “An in-depth analysis of the mitochondrial phylogenetic landscape of Cambodia”, *Scientific Reports*, **11(1)**, DOI: 10.1038/S41598-021-90145-2.
- [27] M. Derenko, B. Malyarchuk, T. Grzybowski, et al. (2007), “Phylogeographic analysis of mitochondrial DNA in northern Asian populations”, *American Journal of Human Genetics*, **81(5)**, DOI: 10.1086/522933.
- [28] M. Bodner, B. Zimmermann, A. Röck, et al. (2011), “Southeast Asian diversity: First insights into the complex mtDNA structure of Laos”, *BMC Evolutionary Biology*, **11(1)**, DOI: 10.1186/1471-2148-11-49.
- [29] W. Woravatin, M. Stoneking, M. Srikummool, et al. (2023), “South Asian maternal and paternal lineages in southern Thailand and the role of sex-biased admixture”, *PLOS ONE*, **18(9)**, DOI: 10.1371/journal.pone.0291547.
- [30] A. Auton, G.R. Abecasis, D.M. Altshuler, et al. (2015), “A global reference for human genetic variation”, *Nature*, **526(7571)**, pp.68-74, DOI: 10.1038/nature15393.
- [31] A.M.S. Ko, C.Y. Chen, Q. Fu, et al. (2014), “Early austronesians: Into and out of Taiwan”, *American Journal of Human Genetics*, **94(3)**, pp.426-436, DOI: 10.1016/J.AJHG.2014.02.003.
- [32] C. Hill, P. Soares, M. Mormina, et al. (2006), “Phylogeography and ethnogenesis of aboriginal Southeast Asians”, *Molecular Biology and Evolution*, **23(12)**, pp.2480-2491, DOI: 10.1093/MOLBEV/MSL124.