



Original Article

# Genetic Diversity and Haplogroup Distribution of Three Ethnic Groups Tay, Thai, and Nung

Nguyen Thuy Duong<sup>1,2,\*</sup>, Nguyen Phuong Anh<sup>1</sup>, Nong Van Hai<sup>1,2</sup>

<sup>1</sup>*Institute of Genome Research, Vietnam Academy of Science and Technology,  
18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam*

<sup>2</sup>*Graduate University of Science and Technology, Vietnam Academy of Science and Technology,  
18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam*

Received 09 February 2023

Revised 14 November 2023; Accepted 14 May 2024

**Abstract:** The mitochondrial genome, with notable characteristics such as maternal inheritance, non-recombination, and high mutation rate, is very important in population genetics and evolution research. To study the genetic diversity and the distribution of haplogroups of Tai - Kadai language group, the mitochondrial genomes of 108 men from three ethnic groups, Tay, Thai, and Nung, were reanalyzed. After comparing the obtained mitochondrial genomes with the Reconstructed Sapiens Reference Sequence (RSRS) published on Genbank (NC\_012920), the results revealed 341, 246, and 256 variants in the Tay, Thai, and Nung ethnic groups, respectively, of which 109 variants were present in all three ethnic groups. Nucleotide diversity ( $\pi$ ) and haplotype were highest in the Thai ethnic group, 0.0023 and 0.989, respectively. The genetic distances between each ethnic group pair (Tay-Nung, Tay-Thai, and Thai-Nung), based on  $F_{ST}$  values, were 0.03101, 0.00447, and 0.03282, respectively. Haplogroup analysis showed that 108 studied individuals were assigned to 39 different sub-haplogroups, belonging to two macro-haplogroups M and N. The most frequent haplogroups in Tay, Thai and Nung were B4 (19.4%), F1 (16.7%), and M7b (19.4%), respectively. The study provides data on mitochondrial genomes of the Tai - Kadai language family in Vietnam, thereby contributing to the study on this language family's genetic structure.

**Keywords:** Haplotype, haplogroup, Nung, mitochondrial DNA, Tay, Thai, Vietnam.

\* Corresponding author.

E-mail address: [tdnguyen@igr.ac.vn](mailto:tdnguyen@igr.ac.vn)

<https://doi.org/10.25073/2588-1140/vnunst.5527>

# Đa dạng di truyền và phân bố nhóm đơn bội ở ba dân tộc Tày, Thái và Nùng

Nguyễn Thuỳ Dương<sup>1,2,\*</sup>, Nguyễn Phương Anh<sup>1</sup>, Nông Văn Hải<sup>1,2</sup>

<sup>1</sup>*Viện Nghiên cứu Hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam,  
18 Hoàng Quốc Việt, Cầu Giấy, 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, Cầu Giấy, Hà Nội, Việt Nam*

Nhận ngày 09 tháng 02 năm 2023

Chỉnh sửa ngày 14 tháng 11 năm 2023; Chấp nhận đăng ngày 14 tháng 5 năm 2024

**Tóm tắt:** Hệ gen ty thể (mtDNA) với các đặc điểm nổi bật như di truyền theo dòng mẹ, không tái tổ hợp và tỷ lệ đột biến cao đóng vai trò quan trọng trong nghiên cứu di truyền quần thể và tiến hóa. Để nghiên cứu sự đa dạng di truyền cũng như phân bố các nhóm đơn bội của ngữ hệ Tai - Kadai, chúng tôi đã phân tích hệ gen ty thể của 108 cá thể nam thuộc ba nhóm dân tộc Tày, Thái và Nùng. Sau khi so sánh trình tự các hệ gen ty thể với trình tự hệ gen ty thể tham chiếu RSRS đã công bố trên Genbank (NC\_012920), chúng tôi tìm thấy 341, 246 và 256 điểm đa hình lần lượt ở dân tộc Tày, Thái và Nùng, trong đó có 109 điểm đa hình xuất hiện ở cả ba dân tộc. Đa dạng nucleotide ( $\pi$ ) và kiểu gen đơn bội (haplotype) được tìm thấy cao nhất ở dân tộc Thái với giá trị tương ứng là 0,0023 và 0,989. Khoảng cách di truyền giữa các dân tộc theo cặp Tày-Nùng, Tày-Thái và Thái-Nùng dựa trên chỉ số  $F_{ST}$  lần lượt là 0,03101; 0,00447 và 0,03282. Phân tích nhóm đơn bội (haplogroup) cho thấy 108 cá thể nghiên cứu thuộc 39 nhóm khác nhau đều thuộc về hai nhóm đơn bội lớn là M và N. Trong đó, ba dân tộc Tày, Thái và Nùng có số lượng cá thể nhiều nhất thuộc các nhóm đơn bội tương ứng B4 (19,4%), F1 (16,7%) và M7b (19,4%). Nghiên cứu này cung cấp dữ liệu toàn bộ hệ gen ty thể của ngữ hệ Tai - Kadai ở dân tộc Việt Nam, từ đó hỗ trợ nghiên cứu về cấu trúc di truyền của họ ngôn ngữ này.

**Từ khóa:** Haplotype, haplogroup, Nùng, hệ gen ty thể, Tày, Thái, Việt Nam.

## 1. Mở đầu

Việt Nam có vị trí địa lý quan trọng ở lục địa Đông Nam Á, là cửa ngõ đến các nước, các đảo và quần đảo trong khu vực. Do đó, nước ta có sự đa dạng rất cao về mặt sắc tộc. Nơi đây là địa bàn cư trú từ lâu đời của cộng đồng 54 dân tộc anh em thuộc năm ngữ hệ (hay họ ngôn ngữ): i) Nam Á (Austroasiatic); ii) Thái - Kadai (Tai - Kadai); iii) Mông - Miên (H'mong - Mien); và iv) Hán - Tạng (Sino - Tibetan) và Nam Đảo (Austronesian) [1]. Trong đó, nhánh Tai của ngữ hệ Tai - Kadai gồm 8 dân tộc là Tày, Thái,

Nùng, Giáy, Lào, Lự, Sán Chay và Bố Y, cư trú tập trung ở các tỉnh vùng Đông Bắc và Tây Bắc Việt Nam như Lai Châu, Lạng Sơn, Điện Biên, Cao Bằng, Sơn La,...

Ngữ hệ Tai - Kadai là một họ đa dạng phân bố ở miền Nam Trung Quốc, Đông Bắc Ấn Độ và phần lớn Đông Nam Á, với cộng đồng người di cư ở Bắc Mỹ và châu Âu [2]. Đây là một trong những họ chính ở lục địa Đông Nam Á, có khoảng 95 ngôn ngữ thuộc họ này với khoảng 93 triệu người nói ở 6 quốc gia: Trung Quốc, Thái Lan, Lào, Myanmar, Ấn Độ và Việt Nam [3]. Họ Tai - Kadai được cho rằng bắt nguồn từ Đông Nam Trung Quốc 2500 năm trước và sau đó phát triển tới Đông Nam Á trong khoảng từ 1000 - 2000 năm trước [4, 5]. Tai - Kadai được tạo thành từ hai nhóm chính:

\* Tác giả liên hệ.

Địa chỉ email: tdnguyen@igr.ac.vn

<https://doi.org/10.25073/2588-1140/vnunst.5527>

Thái và Kadai. Nhóm Thái bao gồm hai ngôn ngữ chính là tiếng Thái và tiếng Lào, ngôn ngữ quốc gia của đất nước Thái Lan và Lào, những người thuộc hai đất nước này chiếm hơn một nửa dân số Tai - Kadai. Tiếng Thái và tiếng Lào có quan hệ mật thiết với tiếng Choang, ngôn ngữ của nhóm thiểu số lớn nhất ở Trung Quốc. Những ngôn ngữ quan trọng khác trong nhóm Thái bao gồm Kam và Sui, với hàng triệu người nói [2]. Nhóm Kadai bao gồm những ngôn ngữ ít được biết đến hơn, một vài ngôn ngữ trong số đó chỉ có vài trăm người nói thông thạo. Phần lớn các ngôn ngữ Tai - Kadai không có hệ thống chữ viết của riêng họ, đặc biệt là các ngôn ngữ Kadai [2].

Các dân tộc Tày, Thái và Nùng đều là những dân tộc thuộc họ Tai - Kadai. Dân tộc Tày có mặt ở Việt Nam từ rất sớm, có thể từ nửa cuối thiên niên kỷ thứ nhất trước Công nguyên. Người Tày là dân tộc có dân số đứng thứ hai Việt Nam, với khoảng 1,845,492 người (Theo số liệu Điều tra 53 dân tộc thiểu số năm 2019) sinh sống chủ yếu ở vùng núi thấp miền núi và vùng trung du Bắc Bộ, đông nhất là ở các tỉnh Cao Bằng và Lạng Sơn [1]. Một số nhóm địa phương thuộc dân tộc Tày bao gồm người Ngạn, Pa Dí, Thu Lao, Phén và Thổ hóa [6]. Cũng trong nhóm ngôn ngữ Tày - Thái này, dân tộc Thái là dân tộc có nguồn gốc lâu đời ở Việt Nam, hiện nay sống tập trung ở các tỉnh Sơn La và Nghệ An, với dân số khoảng 1,820,950 người (Theo số liệu Điều tra 53 dân tộc thiểu số năm 2019), cao thứ ba Việt Nam sau người Kinh và Tày. Các nhóm, ngành lớn của người Thái tại Việt Nam bao gồm Tay Đón (Thái Trắng), Tay Đăm (Thái Đen), Tay Đèng (Thái Đỏ) và Tay Dọ (Thái Yo) cùng một số khác nhỏ hơn [7]. Dân tộc Thái đã có mặt ở miền Tây Bắc Việt Nam trên 1000 năm, có nguồn gốc từ những người Thái đã di cư từ vùng đất thuộc tỉnh Vân Nam, Trung Quốc bây giờ. Một trong những dân tộc thuộc nhóm ngôn ngữ Tày - Thái khác là dân tộc Nùng. Dân tộc Nùng phần lớn từ Quảng Tây (Trung Quốc) di cư sang Việt Nam cách đây khoảng 200 - 300 năm [8]. Người Nùng sống tập trung ở các tỉnh đông bắc Bắc Bộ, nhiều nhất ở tỉnh Lạng Sơn và Cao Bằng. Theo số liệu Điều tra 53 dân tộc

thiểu số năm 2019, dân số của người Nùng là 1,083,298 người, có quan hệ gần gũi với người Tày và người Tráng ở Trung Quốc. Cả ba dân tộc Tày, Thái, Nùng đều ở nhà sàn, biết tận dụng địa hình vùng thung lũng nơi có điều kiện tự nhiên thuận lợi, để cấy lúa nước kết hợp với làm nương rẫy, sáng tạo ra chiếc cối giã gạo, con quay cùng hệ thống mương, phai, lái, lín đưa nước về ruộng. Các nghề thủ công cũng khá phát triển ở các dân tộc này, như rèn, dệt thổ cẩm với nhiều loại hoa văn đẹp, tinh tế và độc đáo [1].

Hệ gen ty thể đã trở thành một yếu tố quan trọng trong việc nghiên cứu về di truyền, lịch sử tiến hoá của các dân tộc [9]. Từ năm 2018 trở lại đây, Việt Nam đã có một số nghiên cứu về hệ gen ty thể trên các cá thể thuộc cộng đồng dân tộc người Việt Nam. Các nghiên cứu này chủ yếu tập trung vào việc so sánh khác biệt giữa các dân tộc thuộc các nhóm ngữ hệ với nhau [10-12]. Nhằm nghiên cứu đa dạng di truyền và phân bố các nhóm đơn bội trong cùng ngữ hệ, chúng tôi đã phân tích toàn bộ hệ gen ty thể của 108 cá thể thuộc ba dân tộc Tày, Thái, Nùng trong ngữ hệ Tai - Kadai, từ đó so sánh đánh giá mức độ đa dạng quần thể và phân bố nhóm đơn bội của các dân tộc này với các quần thể người khác trong cùng khu vực lân cận.

## 2. Đối tượng và phương pháp

### 2.1. Đối tượng

Tổng số 108 mẫu máu ngoại vi của các cá thể nam khỏe mạnh thuộc các dân tộc Tày (47 mẫu), Thái (24 mẫu) và Nùng (37 mẫu) được thu thập và bảo quản ở nhiệt độ 4 °C. Các cá thể được lựa chọn không cùng huyết thống và có ít nhất ba đời cùng thuộc cùng dân tộc. Các cá thể tham gia đều hiểu rõ mục đích của nghiên cứu và đồng ý tham gia cung cấp mẫu máu cho nghiên cứu này. Nghiên cứu này đã được thông qua bởi Hội đồng đạo đức 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

Tách chiết và giải mã hệ gen ty thể.

DNA tổng số được tách chiết từ mẫu máu toàn phần, sử dụng kit GeneJET Whole Blood Genomic DNA Purification (Thermo Fisher Scientific, USA) theo hướng dẫn của hãng sản xuất. Nồng độ và chất lượng của DNA được kiểm tra bằng điện di trên gel agarose 0,8% và dựa trên độ hấp thụ ánh sáng ở các bước sóng 230, 260 và 280 nm trên máy Nanodrop Lite (ThermoFisher Scientific, USA). Từ DNA tổng số, mtDNA được khuếch đại, phân mảnh thành các đoạn ngắn và nối với các đoạn tiếp hợp (adapters) để thiết lập thư viện. Sau đó, thư viện mtDNA được giải trình tự trên hệ thống máy giải trình tự thế hệ mới Illumina [13]. Các đoạn đọc ngắn được đóng hàng và căn chỉnh dựa trên trình tự tham chiếu (Reconstructed Sapiens Reference sequence - RSRS) [14, 15] sử dụng ngôn ngữ lập trình R và phần mềm MAFFT [16].

**Phân tích số liệu**

Các nhóm đơn bội DNA ty thể (mtDNA) được xác định bằng phần mềm HaploGrep2 ([www.haplogrep.uibk.ac.at](http://www.haplogrep.uibk.ac.at)) [17] và cơ sở dữ liệu của PhyloTree mtDNA tree Build 17 (<http://www.phylotree.org>). Phân tích tương ứng (Correspondence Analysis - CA) của các nhóm đơn bội được thực hiện sử dụng ngôn ngữ lập trình R và thư viện “vegan” - “ca”. Các phân tích mức độ đa dạng hệ gen ty thể (Analysis of Molecular Variance - AMOVA) ở các dân tộc được tiến hành bằng phần mềm Arlequin phiên bản 3.5.2.2 [18].

**3. Kết quả**

**3.1. Đa dạng di truyền nucleotide và haplotype hệ gen ty thể của các cá thể dân tộc**

Kết quả so sánh hệ gen ty thể của 108 cá thể với trình tự hệ gen ty thể tham chiếu RSRS đã công bố trên Genbank (NC\_012920) đã tìm thấy 341, 246 và 256 điểm đa hình lần lượt ở dân tộc Tày, Thái và Nùng, trong đó có 109 điểm đa hình xuất hiện ở cả ba dân tộc. Từ đó có thể thấy, số liệu thu được từ ba dân tộc trên có các trình tự có sự đa dạng lớn, phản ánh tốc độ đột biến cao của hệ gen ty thể. Dựa trên các điểm sai khác này trong hệ gen ty thể, 89 trình tự hệ gen ty thể khác biệt (haplotype) đã được

xác định. Đa dạng về kiểu gen đơn bội (H) thấp nhất ở dân tộc Nùng (H = 0,982) và cao nhất ở dân tộc Thái (H = 0,989). Cả hai giá trị về sự đa dạng nucleotide ( $\pi$ ) và số lượng chênh lệch trung bình theo cặp (MPD) cũng thấp nhất ở nhóm dân tộc Nùng (với giá trị tương ứng là 0,0021 và 34,2), trong khi giá trị lớn nhất được quan sát thấy ở nhóm dân tộc Thái (tương ứng là 0,0023 và 38,2) (Bảng 1). Kết quả so sánh khoảng cách di truyền theo cặp hai dân tộc cho thấy Tày - Thái có khoảng cách di truyền thấp nhất ( $F_{ST} = 0,00447$ ), tiếp đến là Tày - Nùng ( $F_{ST} = 0,03101$ ) và Thái - Nùng ( $F_{ST} = 0,03282$ ).

Bảng 1. Sự đa dạng di truyền ở nhóm Tai - Kadai

Dân tộc	Số mẫu	Số lượng haplotypes	H	$\pi$	MPD
Tày	47	40	0,988	0,0022	35,8
Thái	24	21	0,989	0,0023	38,2
Nùng	37	28	0,982	0,0021	34,2

Chú ý: H: đa dạng kiểu gen đơn bội,  $\pi$ : đa dạng nucleotide, MPD: số lượng chênh lệch trung bình theo cặp.

**3.2. Phân nhóm đơn bội**

Kết quả định danh các nhóm đơn bội sử dụng trình tự toàn bộ hệ gen ty thể cho thấy 108 cá thể thuộc ba dân tộc Tày, Thái và Nùng được phân thành 39 nhóm đơn bội khác nhau (Bảng 2). Sơ đồ trực quan tương ứng của toàn bộ 39 nhóm đơn bội được thể hiện ở Hình 1.

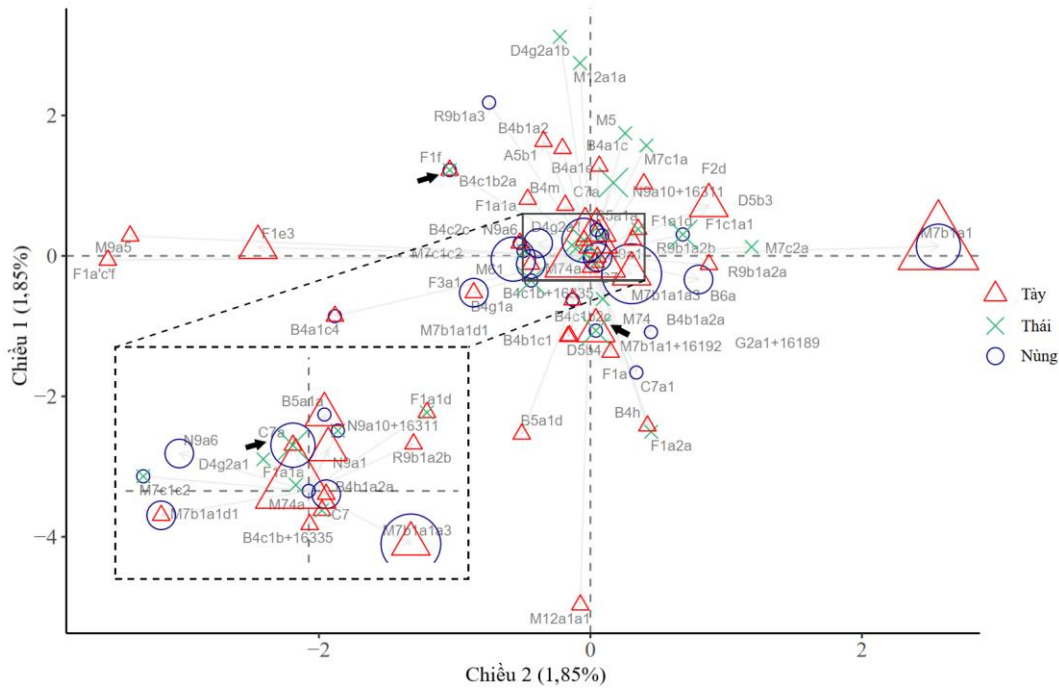
Toàn bộ 39 nhóm đơn bội này thuộc hai nhóm đơn bội lớn (macro-haplogroup) M và N, trong đó phần lớn thuộc ba nhóm M7b, F1 và B4 với tỷ lệ phần trăm trên tổng 108 cá thể tương ứng là 19,4% (21/108), 16,7% (18/108) và 19,4% (21/108) (Bảng 2). Nhóm macro-haplogroup M gồm các nhóm đơn bội M, C, G, và D, chiếm 45,4% tổng số mẫu nghiên cứu (49/108) và được phân vào 17 nhóm đơn bội nhỏ (sub-haplogroup) khác nhau. Trong đó, nhóm sub-haplogroup M7b1a1 xuất hiện nhiều ở cả ba dân tộc Tày, Thái và Nùng với tỉ lệ tương ứng là 19,1% (9/47), 8,3% (2/24) và 27% (10/37) (Bảng 2).

Bảng 2. Các nhóm đơn bội (haplogroup) được định danh của 3 dân tộc Tày, Thái, Nùng

Macro-haplogroup	Haplogroup				Dân tộc				
					Tày (n=47)	Thái (n=24)	Nùng (n=37)		
M	M5						1 (4,2%)		
	M7	M7b'c	M7b	M7b1a1	M7b1a1	4 (8,5%)		3(8,1%)	
					M7b1a1 + 16192	2 (4,3%)	2 (8,3%)	1 (2,7%)	
			M7b1a1a3	3 (6,4%)		6 (16,2%)			
			M7c		M7c1		2 (8,3%)	1 (2,7%)	
					M7c2		1 (4,2%)		
	M8	CZ	C	C7	C7	1 (2,1%)	1 (4,2%)		
					C7a	1 (2,1%)	2 (8,3%)	3 (8,1%)	
					C7a1			1 (2,7%)	
	M9	M9a'b	M9a	M9a5		1 (2,1%)			
	M12'G	M12	M12a	M12a1		1 (2,1%)	1 (4,2%)		
		G	G2	G2a1				1 (2,7%)	
	M13'46'61	M61						3 (8,1%)	
	M42'74	M74	M74a				1 (4,2%)	1 (2,7%)	
M80'D	D	D4	D4g2a1			2 (8,3%)			
				D5b3	2 (4,3%)				
		D5b4	1 (2,1%)						
N	N9	N9a	N9a1'3	N9a1		2 (4,3%)			
			N9a6				2 (5,4%)		
			N9a10			1 (4,2%)	1 (2,7%)		
	A	A5	A5b1			1 (2,1%)			
	R	R9	F	R9b	R9b1		2 (4,3%)		1 (2,7%)
						F1	F1a'c'f	F1a	8 (17%)
				F1c				2 (8,3%)	1 (2,7%)
				F1f	1 (2,1%)			1 (4,2%)	1 (2,7%)
				F1e	F1e3	1 (2,1%)			
				F2	F2d		1 (2,1%)		
		F3	F3a	F3a1	1 (2,1%)		2 (5,4%)		
		R11'B	B	B4	B4a	B4a1c	1 (2,1%)		
						B4alc4	1 (2,1%)		1 (2,7%)
						B4ale		2 (8,3%)	
B4g					B4g1		2 (8,3%)	1 (2,7%)	
B4h					1 (2,1%)				
B4m		1 (2,1%)							
B4b	B4b1	3 (6,4%)		3 (8,1%)					
B4c	B4c1	3 (6,4%)							
	B4c2	1 (2,1%)		1 (2,7%)					
B5	B5a	B5a1	3 (6,4%)		1 (2,7%)				
B6	B6a				2 (5,4%)				

Bên cạnh đó, nhóm macro-haplogroup N gồm các nhóm đơn bội N, A, R, F và B, chiếm 54,6% tổng số mẫu nghiên cứu (59/108) và phân vào 22 sub-haplogroup. Đặc biệt, hai nhóm đơn bội B4 và F1 xuất hiện nhiều ở ba dân tộc. Tần suất của nhóm B4 trên ba dân tộc Tày, Thái và Nùng lần lượt là 23,4% (11/47); 16,7% (4/24) và 16,2% (6/37) trong khi nhóm

F1 xuất hiện với tần số tương ứng là 21,3% (10/47); 25% (6/24) và 5,4% (2/37) (Bảng 2). Ba sub-haplogroup M7b1a1+16192, C7a và F1f được tìm thấy ở cả ba dân tộc (Bảng 2, Hình 1). Ngoài ra, nhóm đơn bội A chỉ xuất hiện trên một cá thể dân tộc Tày trong khi nhóm G chỉ có ở một cá thể Nùng (Bảng 2, Hình 1).



Hình 1. Sơ đồ phân tích tương ứng (Correspondence Analysis - CA) dựa trên tần số nhóm đơn bội haplogroup của 108 cá thể thuộc ba dân tộc Tày, Thái và Nùng.

#### 4. Thảo luận

Ba nhóm đơn bội B4, M7b và F1 xuất hiện nhiều nhất ở 108 cá thể dân tộc thuộc nhóm Tai - Kai với tỷ lệ phần trăm tương ứng là 19,4%, 19,4% và 16,7%. Các nhóm này cũng là những nhóm đơn bội phổ biến xuất hiện ở các quần thể sống tại lục địa Đông Nam Á khác như Lào, Thái Lan, Trung Quốc, Campuchia,... [5, 19-21]. Trong đó, nhóm B là một trong những nhóm đơn bội phổ biến nhất ở phía bắc và phía đông châu Á [22], xuất hiện nhiều thứ hai ở Việt Nam [23] và phổ biến trên khắp lục địa Đông Nam Á, đặc biệt là miền Bắc Việt Nam, miền Bắc Thái Lan và Đài Loan [5, 19-21]. Nhóm đơn bội F, một trong những nhóm phổ biến nhất ở khắp Châu Á, với tần suất dao động từ 31 đến

77%. Nhóm M là một macro-haplogroup được tìm thấy với tần suất cao trên toàn châu Á (bao gồm cả lục địa Đông Nam Á) [5, 23-26] và Việt Nam [23]. Trong đó, nhóm M7 xuất hiện phổ biến nhất, các nhóm này có tần suất cao nhất ở vùng Đông Bắc dọc theo bờ biển Vịnh Bắc Bộ và tần suất thấp nhất ở vùng Tây Bắc Bộ và vùng cực Nam của Việt Nam (Đồng bằng sông Cửu Long) [27]. Nhóm đơn bội M7 cũng có tần suất cao nhất ở miền đông Thái Lan và miền bắc Đài Loan, đây cũng là nhóm đơn bội xuất hiện nhiều nhất trong nhóm cá thể người dân tộc Nùng của chúng tôi với 29,7%.

Các nhóm đơn bội A5, M9, D5, F2 chỉ tìm thấy ở các cá thể người Tày, trong khi các nhóm M5, D4 chỉ tìm thấy ở các cá thể người

Thái và các nhóm G2, M61 và B6 chỉ xuất hiện ở dân tộc Nùng, trong khi ba M7b1a1+16192, C7a và F1f xuất hiện ở cả ba dân tộc, do đó phản ánh những đặc điểm chung và riêng về vật chất di truyền của nhánh Tai của ngữ hệ Tai - Kadai ở Việt Nam. Nghiên cứu năm 2007 trên 774 cá thể thuộc 30 quần thể người Daic ở Trung Quốc và Việt Nam, thuộc ngữ hệ Tai - Kadai đã giải trình tự vùng HVS-1 trên mtDNA để tìm hiểu về mức độ đa dạng của DNA ty thể ở vùng Đông Nam Á [28]. Quần thể người Daic ở Trung Quốc đứng thứ hai sau người Hán, là một nhóm dân tộc lớn ở Nam Trung Quốc, với khoảng 80 triệu người Thái ở Nam Trung Quốc đã di cư đến Thái Lan, Lào, Việt Nam, Myanmar và Ấn Độ. Kết quả nghiên cứu cho thấy các nhóm đơn bội phổ biến nhất trong những quần thể người Daic này là nhóm B4a, F1a, và M7b1 cũng tương đồng với nghiên cứu của chúng tôi. Đây đều là những nhóm đơn bội phổ biến ở phía Nam Trung Quốc, từ đó khẳng định những quần thể người Daic này là nhóm điển hình ở phương Nam [28]. Ngữ hệ Tai - Kadai khá phổ biến ở Đông Nam Á, trong đó có 94,4% người Thái Lan và 69,6% người Lào nói ngôn ngữ này. Nghiên cứu khác năm 2017 đã giải trình tự 1234 hệ gen ty thể từ 51 nhóm quần thể thuộc ngữ hệ Tai - Kadai và ngữ hệ Nam Á từ Thái Lan và Lào cho thấy có 761 kiểu gen đơn bội (haplotype) thuộc 212 nhóm đơn bội khác nhau và có sự khác biệt di truyền đáng kể giữa các mẫu khác nhau từ cùng một nhóm dân tộc [5]. Các nhóm Tai - Kadai có sự đồng nhất về mặt di truyền hơn các nhóm ngữ hệ Nam Á. Kết quả này cũng ủng hộ quan điểm về sự phổ biến của ngữ hệ Tai - Kadai ở Thái Lan bằng cách truyền bá nhân khẩu học, với những người Tai - Kadai hiện nay chủ yếu thuộc nhóm nói tiếng Tai - Kadai từ miền nam Trung Quốc, những người đã di cư xuống phía nam từ 1000 - 2000 năm trước [5].

## 5. Kết luận

Nghiên cứu này đã phân tích đa dạng di truyền và phân bố nhóm đơn bội dựa trên toàn bộ hệ gen ty thể của 108 cá thể thuộc ba tộc người cùng thuộc ngữ hệ Tai - Kadai. Tổng số

89 haplotype khác nhau thuộc 39 nhóm đơn bội đã được xác định, phần lớn các cá thể của cả ba dân tộc thuộc ba nhóm đơn bội B4 (19,4%), F1 (16,7%) và M7b (19,4%). Đa dạng nucleotide ( $\pi$ ) ở ba dân tộc Tày, Thái, Nùng lần lượt là 0,0022; 0,0023 và 0,0021 trong khi đa dạng về kiểu gen đơn bội (H) lần lượt là 0,988; 0,989 và 0,982. Các kết quả nghiên cứu cho thấy mức độ đa dạng di truyền giữa các dân tộc thuộc cùng một ngữ hệ, do đó đóng góp vào các nghiên cứu chuyên sâu hơn về đa dạng hệ gen ty thể ở cộng đồng các dân tộc Việt Nam nói chung và các dân tộc thuộc ngữ hệ Tai - Kadai nói riêng.

## Lời cảm ơn

Công trình được hoàn thành với sự tài trợ từ đề tài thuộc Bộ Khoa học và Công nghệ mã số DTDL.CN-XNT-60/19. Chúng tôi cũng xin chân thành cảm ơn những người đã tham gia hiến mẫu cho nghiên cứu này.

## Tài liệu tham khảo

- [1] D. N. Van, C. T. Son, L. Hung, Ethnic Minorities in Vietnam, 5<sup>th</sup> ed, The Gioi Publishers, 2014, pp. 1-2.
- [2] A. V. N. Diller, J. A. Edmondson, Y. Luo, The Tai-Kadai Languages, Routledge, 2004.
- [3] D. M. Eberhard, G. F. Simons, C. D. Fennig (Eds.). Ethnologue: Languages of the world (24<sup>th</sup> ed.). SIL International, Dallas, 2021.
- [4] P. Pittayaporn, Layers of Chinese Loanwords in Proto-Southwestern Tai as Evidence for the Dating of the Spread of Southwestern Tai, *Manusya: Journal of Humanities*, 2014, pp. 47-68, <https://doi.org/10.1163/26659077-01703004>.
- [5] W. Kutanan, J. Kampuansai, M. Srikumool, D. Kangwanpong, S. Ghirotto, A. Brunelli, M. Stoneking, 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, 2017, pp. 85-98, <https://doi.org/10.1007/s00439-016-1742-y>.
- [6] D. N. Van, C. T. Son, L. Hung, Ethnic Minorities in Vietnam, 5<sup>th</sup> ed, The Gioi Publishers, 2014, pp. 121-125.
- [7] D. N. Van, C. T. Son, L. Hung, Ethnic Minorities in Vietnam, 5<sup>th</sup> ed, The Gioi Publishers, 2014, pp. 126-130.

- [8] D. N. Van, C. T. Son, L. Hung, *Ethnic Minorities in Vietnam*, 5<sup>th</sup> ed, The Gioi Publishers, 2014, pp. 131-135.
- [9] R. L. Cann, M. Stoneking, A. C. Wilson, *Mitochondrial DNA and Human Evolution*, Nature, Vol. 325, No. 6099, 1987, pp. 31-36, <https://doi.org/10.1038/325031a0>.
- [10] D. M. Hung, N. H. Ha, P. N. Khoi, V. P. Nhung, N. V. Phong, N. T. Duong, N. V. Hai, N. D. Ton, Genetic Variation of Mitochondrial Sequence-hv2 in Vietnamese Populations, *Academia Journal of Biology*, Vol. 38, No. 2, 2016, pp. 243-249, <https://doi.org/10.15625/0866-7160/v38n2.7071> (in Vietnamese).
- [11] N. T. Ngoc, N. B. Trang, N. Q. Huy, N. D. Ton, N. T. Duong, 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, 2018, pp. 231-240, <https://doi.org/10.15625/1811-4989/16/2/13432> (in Vietnamese).
- [12] T. T. H. Tran, D. H. Nguyen, V. K. Tran, Q. L. Nguyen, H. A. Trinh, L. H. Luong, V. A. Tran, L. A. T. Pham, T. T. Nguyen, V. B. Nguyen, T. H. Tran, T. V. Ta, Variation of Mitochondrial DNA HV1 AND HV2 of the Vietnamese Population, *Advances in Experimental Medicine and Biology*, Vol. 1292, No. 2020, pp. 37-63, [https://doi.org/10.1007/5584\\_2018\\_301](https://doi.org/10.1007/5584_2018_301).
- [13] M. Meyer, M. Kircher, Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing, *Cold Spring Harb Protoc*, Vol. 2010, No. 6, 2010, pp. 5448, <https://doi.org/10.1101/pdb.prot5448>.
- [14] L. Arias, C. Barbieri, G. Barreto, M. Stoneking, B. Pakendorf, High-Resolution Mitochondrial DNA Analysis Sheds Light on Human Diversity, Cultural Interactions, and Population Mobility in Northwestern Amazonia, *American Journal of Biological Anthropology*, Vol. 165, No. 2, 2018, pp. 238-255, <https://doi.org/10.1002/ajpa.23345>.
- [15] D. M. Behar, M. V. Oven, S. Rosset, M. Metspalu, E. L. Loogvali, N. M. Silva, T. Kivisild, A. Torroni, R. Villems, A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root, *American Journal of Human Genetics*, Vol. 90, No. 4, 2012, pp. 675-684, <https://doi.org/10.1016/j.ajhg.2012.03.002>.
- [16] K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability, *Molecular Biology and Evolution*, Vol. 30, No. 4, 2013, pp. 772-780, <https://doi.org/10.1093/molbev/mst010>.
- [17] H. Weissensteiner, D. Pacher, A. K. Brandstatter, L. Forer, G. Specht, H. J. Bandelt, F. Kronenberg, A. Salas, S. Schonherr, HaploGrep 2: Mitochondrial Haplogroup Classification in the Era of High-throughput Sequencing, *Nucleic Acids Research*, Vol. 44, No. W1, 2016, pp. 58-63, <https://doi.org/10.1093/nar/gkw233>.
- [18] L. Excoffier, 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, 2010, pp. 564-567, <https://doi.org/10.1111/j.1755-0998.2010.02847.x>.
- [19] M. S. Peng, H. H. Quang, K. P. Dang, A. V. Trieu, H. W. Wang, Y. G. Yao, Q. P. Kong, Y. P. Zhang, Tracing the Austronesian Footprint in Mainland Southeast Asia: A Perspective from Mitochondrial DNA, *Molecular Biology and Evolution*, Vol. 27, No. 10, 2010, pp. 2417-2130, <https://doi.org/10.1093/molbev/msq131>.
- [20] X. Zhang, X. Qi, Z. Yang, B. Serey, T. Sovannary, L. Bunnath, H. S. Aun, H. Samnom, H. Zhang, Q. Lin, M. V. Oven, H. Shi, B. Su, Analysis of Mitochondrial Genome Diversity Identifies New and Ancient Maternal Lineages in Cambodian Aborigines, *Nature Communications*, Vol. 4, No. 2013, pp. 2599, <https://doi.org/10.1038/ncomms3599>.
- [21] M. Bodner, B. Zimmermann, A. Rock, A. K. Brandstatter, D. Horst, B. Horst, S. Sengchanh, T. Sanguansermisri, J. Horst, T. Kramer, P. M. Schneider, W. Parson, Southeast Asian Diversity: First Insights into the Complex mtDNA Structure of Laos, *BMC Ecology and Evolution*, Vol. 11, No. 2011, pp. 49, <https://doi.org/10.1186/1471-2148-11-49>.
- [22] M. Derenko, B. Malyarchuk, T. Grzybowski, G. Denisova, I. Dambueva, M. Perkova, C. Dorzhu, F. Luzina, H. K. Lee, T. Vanecek, R. Villems, I. Zakharov, Phylogeographic Analysis of Mitochondrial DNA in Northern Asian Populations, *American Journal of Human Genetics*, Vol. 81, No. 5, 2007, pp. 1025-1041, <https://doi.org/10.1086/522933>.
- [23] N. T. Duong, E. Macholdt, N. D. Ton, L. Arias, R. Schroder, N. V. Phong, V. T. T. B. Thuy, N. H. Ha, H. T. T. Hue, N. T. Xuan, K. T. P. Oanh, L. T. T. Hien, N. H. Hoang, B. Pakendorf, M. Stoneking, N. V. Hai, Complete Human mtDNA Genome Sequences from Vietnam and the



- Phylogeography of Mainland Southeast Asia, *Scientific Reports*, Vol. 8, No. 1, 2018, pp. 11651, <https://doi.org/10.1038/s41598-018-29989-0>.
- [24] R. Rajkumar, J. Banerjee, H. B. Gunturi, R. Trivedi, V. K. Kashyap, Phylogeny and Antiquity of M Macrohaplogroup Inferred from Complete mt DNA Sequence of Indian Specific Lineages, *BMC Ecology and Evolution*, Vol. 5, 2005, pp. 26, <https://doi.org/10.1186/1471-2148-5-26>.
- [25] S. Maruyama, K. Minaguchi, N. Saitou, Sequence Polymorphisms of the Mitochondrial DNA Control Region and Phylogenetic Analysis of mtDNA Lineages in the Japanese Population, *International Journal of Legal Medicine*, Vol. 117, No. 4, 2003, pp. 218-225, <https://doi.org/10.1007/s00414-003-0379-2>.
- [26] J. Kampuansai, W. Kutanana, F. Tassi, M. Kaewgahya, S. Ghirotto, D. Kangwanpong, Effect of Migration Patterns on Maternal Genetic Structure: A Case of Tai-Kadai Migration from China to Thailand, *Journal of Human Genetics*, Vol. 62, No. 2, 2017, pp. 223-228, <https://doi.org/10.1038/jhg.2016.112>.
- [27] S. Pischedda, R. B. Arca, A. G. Carballa, J. P. Seco, M. L. Catelli, V. A. Iglesias, J. M. Cardenas, N. D. Nguyen, H. H. Ha, A. T. Le, F. M. Torres, C. Vullo, A. Salas, Phylogeographic and Genome-Wide Investigations of Vietnam Ethnic Groups Reveal Signatures of Complex Historical Demographic Movements, *Scientific Reports*, Vol. 7, No. 1, 2017, pp. 12630, <https://doi.org/10.1038/s41598-017-12813-6>.
- [28] H. Li, X. Cai, E. R. W. Cort, B. Wen, X. Cheng, Z. Qin, W. Liu, Y. Liu, S. Pan, J. Qian, C. C. Tan, L. Jin, Mitochondrial DNA Diversity and Population Differentiation in Southern East Asia, *American Journal of Biological Anthropology*, Vol. 134, No. 4, 2007, pp. 481-488, <https://doi.org/10.1002/ajpa.20690>.