DOI: 10.31276/VJSTE.2025.0008

Complete mitochondrial DNA sequences reveal the maternal genetic structure of Vietnamese Austroasiatic-speaking Bana, Hre, and Mnong populations

Dinh Huong Thao[1, 2], La Duc Duy[1], Nong Van Hai[1], Nguyen Thuy Duong[1*]

[1]*Institute of Biology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Nghia Do Ward, Hanoi, Vietnam*
[2]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Nghia Do Ward, Hanoi, Vietnam*

Received 27 February 2025; revised 19 September 2025; accepted 9 October 2025

Abstract:

Austroasiatic, a major ethnolinguistic family in Vietnam, accounts for 90% of the census population size. Apart from the majority Kinh group, 24 other Austroasiatic minorities account for 4.67% of the population, many of which remain understudied. Here, the genetic characteristics of three Vietnamese Austroasiatic ethnic groups (Bana, Hre, and Mnong) were investigated by analysing the complete mitochondrial DNA (mtDNA) of 119 individuals. A total of 6,361 polymorphic sites were identified, with Bana exhibiting the highest molecular diversity (53.64±3.51 variants per individual). Compared to hypervariable segments (HVS), complete mtDNA sequences enhanced phylogenetic resolution and accuracy. The maternal haplogroup profile consisted of 40 lineages, with 32 of the detected lineages being population specific, six shared between two populations, and two shared among all populations. The macro-haplogroup components included B, C, F, M, N, R, and Z, with 66-80% of each population assigned to haplogroups B and M. The most prevalent haplogroups were M20 (14.29%), B5a1a (6.72%), and M74b1 (6.72%). While lineages of haplogroup F were relatively more frequent in Bana and Hre, lineages of N were exclusive to Mnong, highlighting haplogroup distinctions within the Vietnamese Austroasiatic family of the Central Highlands. These findings contribute to reconstructing the complex history of migration and settlement in Vietnam.

Keywords: **Bana, Hre, mitochondrial DNA, Mnong, Vietnam.**

Classification numbers: **3.4, 3.5**

1. Introduction

Austroasiatic, a dominant language family of Southeast Asia (SEA), encompasses 167 languages spoken by approximately 117 million people and is categorised into 13 sub-branches [1]. The distribution of Austroasiatic languages extends from Northeast India to the Malay Peninsula [2]. All but three sub-branches are found in Mainland Southeast Asia: Aslian, Bahnaric, Katuric, Khmer, Khmuic, Monic, Pakanic, Palaungic, Pearic, and Vietic [1]. Compared to other major language families of the region, such as Austronesian in Island Southeast Asia or Tai-Kadai in the Thailand -Laos area, the migration routes of Austroasiatic speakers remain less clearly defined.

The origin of the Austroasiatic family continues to be debated, with proposed centres of dispersal including the Bay of Bengal [3], central/southern China or within the heartland of SEA itself [4]. The expansion of the Austroasiatic family may have been associated with the spread of rice agriculture [5]. However, these theories are primarily supported by archaeological, linguistic, and historical evidence, with limited corroboration from biological data [1, 6, 7].

In Vietnam, the Austroasiatic language family comprises 25 groups distributed across the northern and southern regions of the country. Predominant among them are the Kinh, who represent the majority of the Vietnamese population and are spread throughout the country. As such, the Kinh people have been the focus of various genetic and genomic studies [8-11]. Conversely, other Austroasiatic minorities remain largely understudied, with only a few groups, such as the Mnong, Khmer, and Mang, having received some attention [12-14]. The absence of a comprehensive genetic profile of the Austroasiatic-speaking communities in Vietnam is evident, underscoring the need for more extensive studies on this ethnolinguistic group.

*Corresponding author: Email: tdnguyen@ib.ac.vn

Mitochondrial DNA (mtDNA) serves as a widely used uniparental marker in population genetics. The 16,569 bp mitochondrial genome can be divided into two main regions: the coding and control regions. The coding region (np 577-16023), which constitutes 93% of the human mtDNA, contains 37 genes encoding 13 proteins, 22 tRNAs, and two rRNAs. The control region (np 1-576; np 16024-16569), where replication and transcription initiate, can be further subdivided into three parts: hypervariable segments (HVS) I, II, and III. These HVS regions mutate faster than the conserved coding region and are therefore often the target of sequencing efforts due to their high informativeness.

Here, we analysed the complete mitochondrial genomes of 119 Austroasiatic individuals from three populations - Bana, Hre, and Mnong. Haplogroup profiles based on whole mtDNA sequences were assessed in each population. Nucleotide variant distributions across different regions of mtDNA were examined to explore molecular diversity. This research provides new insights into the genetic structure of underrepresented Austroasiatic minorities and contributes to the broader understanding of Vietnam's population history.

2. Materials and methods

2.1. Sample information and data collection

Our dataset included complete mtDNA sequences obtained from the GenBank database (accession numbers: PP654481-PP654599) [15]. The research received ethical approval (No. 2-2019/NCHG-HĐĐĐ) from the Institutional Review Board of the Institute of Genome Research, Vietnam Academy of Science and Technology. Following the original study design - which examined both mtDNA and the male-specific region of the Y chromosome - only male participants were included. This sampling strategy does not affect the representativeness of mtDNA haplogroups or bias the maternal inheritance pattern.

The dataset included 119 unrelated males from three Vietnamese Austroasiatic populations: Bana (36 individuals), Hre (30 individuals), and Mnong (53 individuals), all of whom provided written informed consent. Each participant self-identified as belonging to the same ethnicity for at least three generations. Sampling locations included Kon Ray, Kon Tum (Bana); Kon Plong, Kon Tum (Hre); and Lak, Dak Lak (Mnong).

2.2. Genetic analyses

Nucleotide variants across multiple mtDNA segments (coding and HVS regions) were identified by aligning reads against the Reconstructed Sapiens Reference Sequence (RSRS) using an in-house algorithm. Positions with missing nucleotides (Ns) were excluded, as were eight additional sites: the poly-C stretch of hypervariable segment II (HVS-II; np 303-317), CA-repeat (np 514-523), C-stretch 1 (np 568-573), 12S rRNA (np 956-965), historical site (np 3,107), C-stretch 2 (np 5,895-5,899), 9 bp deletion/insertion (np 8,272-8,289), and the poly-C stretch of hypervariable segment I (HVS-I; np 16,180-16,195).

Variant distributions across the three populations were visualised using the R package "ggplot2" [16]. The HVS-I (np 16,024-16,383) and HVS-II (np 57-372) regions, along with the complete mtDNA sequences, were analysed using HaploGrep3 [17] to assign haplogroups, with references based on PhyloTree mtDNA tree Build 17 [18]. Correspondence analysis (CA) was computed based on haplogroup frequencies in R using the libraries "vegan" [19] and "ca" [20]. Median-joining networks were constructed and visualised by HapNetworkView [21].

3. Results

3.1. Variant distribution

The molecular diversity of each population was further examined by visualising variant distributions across different mtDNA regions (Fig. 1). Among the 119 mtDNA sequences, 6,361 variants were detected (average 53.45±3.40 variants per sample), of which 73.32% were located in the coding region and the remainder in the control region. Specifically, the HVS-I (np 16,024-16,383), HVS-II (np 57-372), HVS-III (np 438-574), and other non-coding sequences accounted for 12.42, 8.63, 4.87, and 0.76% of all variants, respectively.

In HVS-I, the Bana distribution was skewed toward the upper portion of the curve, while those of Hre and Mnong were more evenly dispersed. In HVS-II, Hre displayed the most compact curve, centralised around the median; Bana's variants were concentrated in the upper range, whereas Mnong's were distributed more evenly. Variant ranges in the HVS-III region were similar across all three populations; however, Bana and Hre showed wider distributions in the upper range, while Mnong's distribution was more symmetrical around the median. In the coding region, the distribution curves of all three populations were highly similar, with Mnong exhibiting the broadest range (Fig. 1).

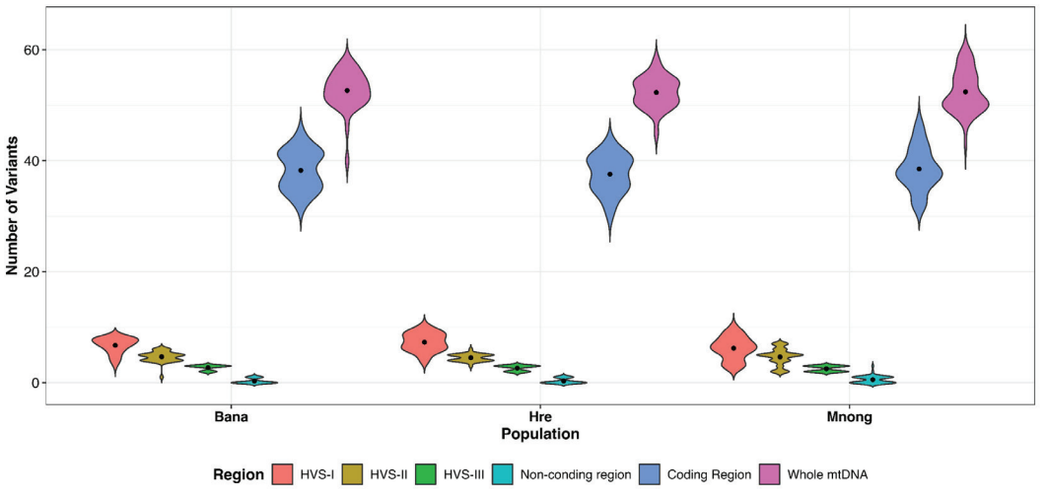

Fig. 1. Variant distributions across different mtDNA regions in the Bana, Hre, and Mnong populations.

3.2. Haplogroup classification

In our dataset of 119 samples, a total of 38 maternal lineages were identified using HVS sequences in HaploGrep3, while 40 lineages were detected when analysing complete mtDNA genomes. When comparing lineage components, 58.82% of the samples (70/119) displayed identical haplogroups between the two approaches, whereas the remaining 41.18% (49/119) showed discrepancies (Table 1). Among these, 17 samples exhibited a change in lineage assignment between the two methods. Of the 49 samples with differing results, 17 displayed distinct lineage classifications, while 32 had mitogenome-based haplogroups that were more refined than those derived from HVS-based analysis. In terms of ethnolinguistic groups, Bana (n=36), Hre (n=30), and Mnong (n=53) each contained 18, 10, and 21 samples, respectively, that yielded differing results when comparing different mtDNA segments.

Table 1. Differences between haplogroups assigned by HVS and complete mtDNA.

Number	Haplogroup by HVS-I and -II	Haplogroup by whole mtDNA	Number of samples	Frequency (n=119)
1	B4a1a5a	B4a1c+146T!	1	0.84%
2	B4b1	B4b1a	1	0.84%
3	B4b2	B4b1a	1	0.84%
4	B5a	B5a1a	8	6.72%
5	B5a	B5a1b1	1	0.84%
6	B5a1d	B5a1c	1	0.84%
7	C4b3	C7	1	0.84%
8	C4b8a	C7	4	3.36%
9	F1a1	F1a1d	3	2.52%
10	F1a1a	F1a1a1	4	3.36%
11	L3f2	M23'75	3	2.52%
12	L3x1	M73	5	4.20%
13	M12	M12a1b	1	0.84%
14	M17c1a1	M17c1a1a	2	1.68%
15	M50	M13b1	1	0.84%
Total	14	13	49	41.18%

"n" denotes the total number of samples in the study population.

A total of 40 mitogenome-based haplogroups were classified into seven macro-haplogroups (B, C, F, M, N, R, and Z) (Table 2). Two lineages (M20 and B5a1a) were shared among all three populations, while six (M73b, C7, R9b1a1a, M9b, F1a1a1, and F1a1d) were shared between two populations. The remaining 32 lineages were population-specific, 17 of which were singletons. Thirty-two representatives of macro-haplogroups M, B, and N together accounted for 82.35% of the total samples. The most widely distributed haplogroups were M20 (14.29%), M74b1 (6.72%), and B5a1a (6.72%).

Within each population, the Mnong exhibited 22 haplogroups, of which 11 were singletons. The most frequent haplogroups were M74b1 (15.09%), N9a6 (11.32%), and B5a1a (9.43%). Notably, B5a1a was the only lineage from haplogroup B present in all three populations, whereas M74b1 and N9a6 were exclusive to Mnong. The Hre samples were classified into 14 haplogroups, including seven singletons, with B5a1d (20%) and M20 (20%) being predominant. While B5a1d was absent in both Mnong and Bana, M20 was shared across all three populations.

Table 2. Haplogroup profiles in three Austroasiatic populations using complete mtDNA genomes.

Number	Haplogroup	Bana (n=36)	Hre (n=30)	Mnong (n=53)	Overall (n=119)
1	B4a1c+146T!	-	-	1.89%	0.84%
2	B4b1a	5.56%	-	-	1.68%
3	B4c2	5.56%	-	-	1.68%
4	B4g1	-	3.33%	-	0.84%
5	B5a1	-	3.33%	-	0.84%
6	B5a1a	5.56%	3.33%	9.43%	6.72%
7	B5a1b1	-	-	1.89%	0.84%
8	B5a1c	-	-	1.89%	0.84%
9	B5a1d	-	20.00%	-	5.04%
10	C7*	8.33%	6.67%	-	4.20%
11	F1a1a	-	6.67%	-	1.68%
12	F1a1a1	5.56%	-	3.77%	3.36%
13	F1a1d	5.56%	3.33%	-	2.52%
14	M12a1b	-	-	1.89%	0.84%
15	M12b1a2	-	-	1.89%	0.84%
16	M12b1b	-	3.33%	-	0.84%
17	M12b2a	-	-	3.77%	1.68%
18	M13b1	-	3.33%	-	0.84%
19	M17c1a1a	5.56%	-	-	1.68%
20	M20	27.78%	20.00%	1.89%	14.29%
21	M21b2	-	-	1.89%	0.84%
22	M23'75	8.33%	-	-	2.52%
23	M24a	2.78%	-	-	0.84%
24	M50a2	-	-	1.89%	0.84%
25	M51b1b	-	-	7.55%	3.36%
26	M68a1a	-	-	1.89%	0.84%
27	M73	-	-	9.43%	4.20%
28	M73b	5.56%	10.00%	-	4.20%
29	M74b1	-	-	15.09%	6.72%
30	M7b1a1f	-	-	3.77%	1.68%
31	M7c1c2	2.78%	-	-	0.84%
32	M9a1a3	-	6.67%	-	1.68%
33	M9b	5.56%	6.67%	-	3.36%
34	N21+195T!	-	-	3.77%	1.68%
35	N22	-	-	7.55%	3.36%
36	N9a6	-	-	11.32%	5.04%
37	R22	-	-	1.89%	0.84%
38	R23	-	-	1.89%	0.84%
39	R9b1a1a	5.56%	-	3.77%	3.36%
40	Z3c*	-	3.33%	-	0.84%

"n" denotes the total number of individuals in each population; haplogroup labels without an asterisk include all downstream subhaplogroups, whereas haplogroup labels with an asterisk exclude all downstream subhaplogroups.

The Bana samples contained 14 lineages, including two singletons. M20 (27.78%) was the most common haplogroup in Bana, followed by C7 and M23'75 (8.33% each). M20 and B5a1a were the only lineages observed in all three ethnic groups, with M20 reaching high frequencies in Bana (27.78%) and Hre (20%). Haplogroups shared between two populations (C7*, F1a1d, F1a1a, and M9a'b) or those with high frequencies in the dataset (N9a6, M74b1, M51b1b) tended to cluster near the centre of the CA plot. By contrast, several haplogroups associated with Mnong (M12b1a2, N22, M50a2, R22, B5a1b1, M21b2, and M7b1a1f) were positioned toward the periphery of the plot.

Haplotype network analysis, visualised using HapNetworkView (Fig. 2), revealed radial clusters within clades M and B, characterised by large nodes with multiple one-step connections, suggesting recent population expansion. Macrohaplogroup N was confined to Mnong, while haplogroup F formed scattered, mixed, and moderately sized clusters (Fig. 2A). Shared haplotypes between Bana and Hre highlighted close genetic connections between these two groups (Fig. 2B).

4. Discussion

In this study, the distributions of nucleotide variants revealed genetic diversity at the molecular level. The control region was found to be more compact than the coding region. Specifically, the HVS sequences (813 bp) accounted for only 4.91% of the total mtDNA length yet contained 25.92% of all variants. Among the hypervariable segments, HVS-III - the most recently defined - displayed the highest variant occurrence per nucleotide (2.26 variants/nucleotide), whereas HVS-II showed the lowest (1.74 variants/nucleotide). Prior to the advent of next-generation sequencing (NGS), the HVS region was the primary target for sequencing in mtDNA studies due to the large amount of genetic information retrievable from a relatively short segment. However, given the considerable length of the coding region (15,447 bp), it still contains a substantial amount of genetic data. Moreover, HVS data alone cannot achieve the same level of accuracy and phylogenetic resolution as complete mtDNA when assigning haplogroups, as demonstrated by the discrepancies between HVS-based and complete mtDNA-based classifications (Table 1). With the development and increasing accessibility of NGS, sequencing the entire mitochondrial genome has become the gold standard for maternal lineage analysis.

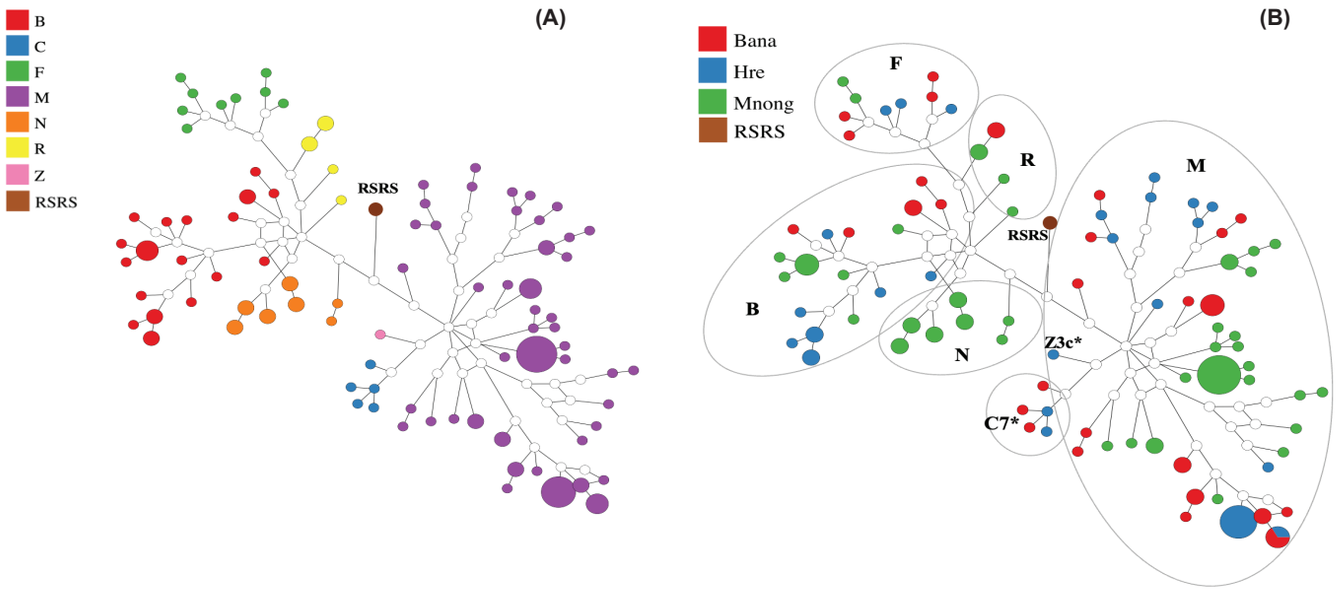

Fig. 2. Haplotype network analysis classified by (A) macrohaplogroups and (B) ethnic groups.

Across populations, Bana exhibited the highest mean number of variants per individual (53.64±3.51), followed by Mnong (53.43±3.69) and Hre (53.30±2.77). Haplogroup compositions reflected substantial diversity within the dataset: macrohaplogroups M and B predominated (ranging from 66-80% in each population), while macrohaplogroup N appeared exclusively in Mnong (22.64%). Macrohaplogroup F ranked as the third most common in Bana (11.11%) and Hre (10%). Mnong also showed the largest number of population-specific haplogroups (18 lineages). The sample sizes for each population (Bana: n=36; Hre: n=30; Mnong: n=53) are consistent with standard practice in uniparental marker studies [22-25]. While common haplogroups (≥5% frequency) are typically captured within this range, rarer lineages may remain underrepresented or undetected. Therefore, the absence of a haplogroup in our dataset should not be interpreted as a true absence within the population. This limitation should be considered when studying minority groups with modest population sizes.

Macrohaplogroup M was the most widely distributed in this study, comprising more than half of all samples. Half of the detected lineages (20/40) were descendants of haplogroup M, of which 85% were singletons. M20 was the most frequent haplogroup, identified in all three populations, with an overall frequency of 14.29%. It was the most common lineage in Bana (27.78%) and Hre (20%), but occurred in only one Mnong individual (1.89%). M20 has been reported among Indian and Eurasian populations, suggesting gene flow between southern and eastern Asia [26, 27]. Within Southeast Asia, this lineage has been observed in Indonesia [28], Vietnam [12], Cambodia [29], Thailand, and Laos [30]. Another notable haplogroup is M74b1, which dates back to approximately 12.6-10 kya (thousand years ago) [31]. To date, M74b1 has not been reported in other Vietnamese Austroasiatic populations but has been identified in several groups across both continental and island Southeast Asia, including Austroasiatic and Tai-Kadai speakers in Thailand [30], Sino-Tibetan in Myanmar [32], or the Austronesian in Vietnam [12], the Philippines [33], and Indonesia [34]. The diversity of M clades across Vietnam and its neighbouring countries suggests a region-specific expansion history for this haplogroup.

Lineages of macrohaplogroup B was the second most common, representing 19.32% of the dataset. Nine B lineages were identified and classified into the B4 and B5 sub-haplogroups, eight of which were restricted to a single population. The only lineage shared among all three populations was B5a1a, previously reported in Vietnam [12], Thailand, Laos [30], and Cambodia [29]. Together with B5a1d, this haplogroup was presumed to have coalesced in the middle of the Holocene, dating back to approximately 8 kya [31]. A noteworthy representative of the B4 sub-haplogroup is B4c2, which appeared in two Bana individuals (5.56%). Other studies also detected B4c2 in South China [22], Vietnam [12-13], Laos [31], Thailand [30], indicating a late Pleistocene origin in Mainland Southeast Asia.

Macrohaplogroup N is the third most common in our dataset (10.08%). Interestingly, three N lineages (N21+195T!, N22, N9a6) were exclusive to the Mnong. Unlike its direct ancestor N9a, which is more widespread in East Asia [35-37], N9a6 appeared in low frequency in Southeast Asia populations such as Laos, Thailand [38], Indonesia [39] and Malaysia [31]. Its descendants, N9a6a and N9a6b, are found at high frequency in Malaysia, suggesting a local diversification of this branch [28]. N22, which is associated with the expansion of Austronesian speakers into Malaysia [40], has low frequencies in Thailand [25-30], Laos [41] and the Philippines [42]. The fragmented distribution of the N22 clade, alongside a deep coalescent age of ~27.3 kya [28], suggests multiple overlapping waves of migration.

The presence of numerous population-specific lineages and diverse macrohaplogroups reflects both ancient divergence and localised demographic processes. Future research should integrate complete mtDNA and Y-chromosome data, along with genome-wide SNP genotyping, to develop a more comprehensive model of Vietnam's demographic history and its connections to broader Southeast Asian migrations.

5. Conclusions

This study highlights the maternal genetic characteristics of 119 male individuals from three Austroasiatic populations (Bana, Hre, and Mnong), with a focus on mitogenomic substructure. The non-coding region contained a higher proportion of variants than the coding region. Among the three groups, Bana exhibited the highest average number of variants per individual (53.64±3.51). Forty haplogroups were identified, predominantly belonging to macrohaplogroups B (nine lineages) and M (20 lineages), many of which - such as M20, B4c2, and B5a1a - are characteristic of the Southeast Asian region and not specific to any linguistic family, reflecting genetic continuity among local populations.

Distinctive patterns emerged across the three populations: Mnong possessed the highest number of unique haplogroups, including three sub-branches of macrohaplogroup N that were absent in Hre and Bana. Vietnam's modern genetic landscape reflects a complex history of settlement and migration. Insights into these underrepresented Austroasiatic minorities expand current knowledge of the Austroasiatic linguistic family and contribute to clarifying the still-unresolved peopling history of this region.

CRediT author statement

Dinh Huong Thao: Data analysis and Writing; La Duc Duy: Data curation and Data analysis; Nong Van Hai: Reviewing and Supervision; Nguyen Thuy Duong: Conceptualisation and Revising.

COMPETING INTERESTS

The authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] P. Sidwell, S. Paul, J. Mathias (2021), "11 classification of MSEA Austroasiatic languages", *The Languages and Linguistics of Mainland Southeast Asia*, pp.179-206, DOI: 10.1515/9783110558142-011.

[2] D.M. Eberhard, G.F. Simons, C.D. Fennig (2025), *Ethnologue: Languages of The World*, 28th ed, SIL International.

[3] G.V. Driem (2001), *Languages of The Himalayas: Volume 2*, DOI: 10.1163/9789004492530.

[4] P. Sidwell, R. Blench (2011), *The Austroasiatic Urheimat: The Southeastern Riverine Hypothesis*, Australian National University, pp.315-343.

[5] M. Robbeets (2017), *Language Dispersal Beyond Farming*, John Benjamins Publishing Company, pp.1-23, DOI: 10.1075/z.215.01rob.

[6] P.J. Piper, H. Matsumura, D. Bulbeck (2017), *New Perspectives in Southeast Asian and Pacific Prehistory*, Australian National University Press, 384pp.

[7] P.S. Bellwood (2005), *First Farmers: The Origins of Agricultural Societies*, Blackwell Pub, 360pp.

[8] A. V. Trieu, S. Djoulah, C.T. Thi, et al. (1997), "HLA-DR and -DQB1 DNA polymorphisms in a Vietnamese Kinh population from Hanoi", *European Journal of Immunogenetics*, **24(5)**, pp.345-356, DOI: 10.1046/j.1365-2370.1997.d01-107.x.

[9] M.S. Peng, H.H. Quang, K.P. Dang, et al. (2010), "Tracing the Austronesian footprint in mainland Southeast Asia: A perspective from Mitochondrial DNA", *Molecular Biology and Evolution*, **27(10)**, pp.2417-2430, DOI: 10.1093/molbev/msq131.

[10] M.D. Do, L.G.H. Le, V.T. Nguyen, et al. (2020), "High-resolution HLA typing of HLA-A, -B, -C, -DRB1, and -DQB1 in Kinh Vietnamese by using next-generation sequencing", *Frontiers in Genetics*, **11**, DOI: 10.3389/fgene.2020.00383.

[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] 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)**, DOI: 10.1038/s41598-018-29989-0.

[13] M. Busson, A.V. Trieu, P. Labelle, et al. (2002), "HLA-DRB1 and DQB1 allele distribution in the Muong population exposed to malaria in Vietnam", *Tissue Antigens*, **59(6)**, pp.470-474, DOI: 10.1034/j.1399-0039.2002.590603.x.

[14] S. Pischedda, R.B. Arca, A.G. Carballa, et al. (2017), "Phylogeographic and genome-wide investigations of Vietnam ethnic groups reveal signatures of complex historical demographic movements", *Scientific Reports*, **7(1)**, DOI: 10.1038/s41598-017-12813-6.

[15] D.H. Thao, T.H. Dinh, S. Mitsunaga, et al. (2024), "Investigating demic versus cultural diffusion and sex bias in the spread of Austronesian languages in Vietnam", *PLOS ONE*, **19(6)**, DOI: 10.1371/journal.pone.0304964.

[16] H. Wickham (2016), *ggplot2: Elegant Graphics for Data Analysis*, Springer International Publishing: Imprint: Springer, 260pp.

[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] J. Oksanen, F.G. Blanchet, R. Kindt, et al. (2012), *Vegan: Community Ecology Package*, Research Gate.

[20] O. Nenadic, M. Greenacre (2007), "Correspondence analysis in R, with two- and three-dimensional graphics: The Ca Package", *Journal of Statistical Software*, **20(3)**, DOI: 10.18637/jss.v020.i03.

[21] L. Chi, Y. Dong, R. Wang, et al. (2025), "HapNetworkView: A tool for haplotype network exploration and visualization", *BMC Genomics*, **26(1)**, DOI: 10.1186/s12864-025-11206-8.

[22] S. Lippold, H. Xu, A. Ko, et al. (2014), "Human paternal and maternal demographic histories: Insights from high-resolution Y chromosome and mtDNA sequences", *Investigative Genetics*, **5(1)**, DOI: 10.1186/2041-2223-5-13.

[23] F. Brisighelli, V.A. Iglesias, M. Fondevila, et al. (2012), "Uniparental markers of contemporary Italian population reveals details on its pre-Roman heritage", *PLOS ONE*, **7(12)**, DOI: 10.1371/journal.pone.0050794.

[24] N.R. Mesa, M.C. Mondragón, I.D. Soto, et al. (2000), "Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: Pre- and post-Columbian patterns of gene flow in South America", *Am. J. Hum. Genet.*, **67(5)**, pp.1277-1286, DOI: 10.1016/s0002-9297(07)62955-3.

[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)**, pp.898-911, DOI: 10.1038/s41431-018-0113-7.

[26] M. Lin, J.A. Trejaut (2023), "Diversity and distribution of mitochondrial DNA in non-Austronesian-speaking Taiwanese individuals", *Human Genome Variation*, **10(1)**, DOI: 10.1038/s41439-022-00228-3.

[27] M. Lipson, O. Cheronet, S. Mallick, et al. (2018), "Ancient genomes document multiple waves of migration in Southeast Asian prehistory", *Science (New York, N.Y.)*, **361(6397)**, pp.92-95, DOI: 10.1126/science.aat3188.

[28] T.A. Jinam, L.C. Hong, M.E. Phipps, et al. (2012), "Evolutionary history of continental Southeast Asians: "Early train" hypothesis based on genetic analysis of Mitochondrial and Autosomal DNA data", *Molecular Biology and Evolution*, **29(11)**, pp.3513-3527, DOI: 10.1093/molbev/mss169.

[29] 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.

[30] 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.

[31] T. Shoeib (2018), *Investigating The Modern Human Settlement of Mainland Southeast Asia Using Mitochondrial DNA*, University of Huddersfield.

[32] M. Summerer, J. Horst, G. Erhart, et al. (2014), "Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-ethnic population of Myanmar", *BMC Evolutionary Biology*, **14(1)**, DOI: 10.1186/1471-2148-14-17.

[33] E.D. Gunnarsdóttir, M. Li, M. Bauchet, et al. (2011a), "High-throughput sequencing of complete human mtDNA genomes from the Philippines", *Genome Research*, **21(1)**, pp.1-11, DOI: 10.1101/gr.107615.110.

[34] E.D. Gunnarsdóttir, M.R. Nandineni, M. Li, et al. (2011b), "Larger mitochondrial DNA than Y-chromosome differences between matrilocal and patrilocal groups from Sumatra", *Nature Communications*, **2(1)**, DOI: 10.1038/ncomms1235.

[35] S. Miura, A. Sasaki, S. Kasai, et al. (2022), "Association of mitochondrial DNA haplogroup and hearing impairment with aging in Japanese general population of the Iwaki Health Promotion Project", *J. Hum. Genet.*, **67(6)**, pp.369-375, DOI: 10.1038/s10038-022-01011-6.

[36] N. Fuku, K.S. Park, Y. Yamada, et al. (2007), "Mitochondrial haplogroup N9a confers resistance against type 2 diabetes in Asians", *Am. J. Hum. Genet.*, **80(3)**, pp.407-415, DOI: 10.1086/512202.

[37] H. Fang, N. Hu, Q. Zhao, et al. (2018), "mtDNA Haplogroup N9a increases the risk of type 2 diabetes by altering mitochondrial function and intracellular mitochondrial signals", *Diabetes*, **67(7)**, pp.1441-1453, DOI: 10.2337/db17-0974.

[38] 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.

[39] C. Hill, P. Soares, M. Mormina, et al. (2007), "A Mitochondrial stratigraphy for island Southeast Asia", *The American Journal of Human Genetics*, **80(1)**, pp.29-43, DOI: 10.1086/510412.

[40] F. Aghakhanian, B.P. Hoh, C.W. Yew, et al. (2022), "Sequence analyses of Malaysian indigenous communities reveal historical admixture between Hoabinhian hunter-gatherers and Neolithic farmers", *Scientific Reports*, **12(1)**, DOI: 10.1038/s41598-022-17884-8.

[41] 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.

[42] K.A. Tabbada, J. Trejaut, J.H. Loo, et al. (2010), "Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia?", *Molecular Biology and Evolution*, **27(1)**, pp.21-31, DOI: 10.1093/molbev/msp215.