Vietnam Journal of Biotechnology **23**(1): 1-12, 2025. DOI: 10.15625/vjbt-21323

THE MATERNAL GENETICS STRUCTURE OF THE CHAM-RAGLAY-COHO-MA DISCLOSED BY WHOLE MITOCHONDRIAL DNA

Huong Thao Dinh[1,2], Van Hai Nong[1] and Thuy Duong Nguyen[1,✉]

[1]*Institute of Genome Research, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Nghia Do, Cau Giay, Hanoi, Vietnam*

[2]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Nghia Do, Cau Giay, Hanoi, Vietnam*

[✉]*To whom correspondence should be addressed. Email: tdnguyen@igr.ac.vn*

Received: 12.08.2024
Accepted: 14.02.2025

ABSTRACT

Cham, Raglay, Coho, and Ma were native people residing in Central and Southern Vietnam. Their lifestyle and traditional customs revealed the closeness in contact. Here, the mitogenomes of 126 individuals of these four aforementioned groups were analyzed on a molecular level and evaluated the maternal haplogroup architecture. A total of 445 unique variants were screened 6807 times in this dataset, showing a mean variant count of 54.02 ± 3.77 variants per person. Out of 445 unique variants, 88 were present in all individuals, and 137 were shared between at least two populations. The remaining 220 were exclusive to a single population, in which Raglay retained the most (88 variants), and Coho had the least (22 variants). Ma and Coho shared the highest number of unique variants: 163/445 were identified in both. Pair-wise genetic distance was the highest between Raglay and Coho (F_{st} = 0.13659) and the lowest between Ma and Coho. The maternal haplogroup profile encompassed 30 representatives, stratified into six macro-haplogroups (B, E, F, M, N and R). M, B, and F accounted for 94.45% of the dataset, and the most prevalent branches were B5a1b1 (11.11%), M24b (10.32%), and F1a1a1 (7.94%). The similarities between highlander populations were shown in the macrohaplogroup components, as well as regional features, while the diversity was expressed in unique variants and specific lineages, providing further understanding of the genetic structure of these underrepresented ethnic groups.

Keywords: Cham, Raglay, Coho, Ma mitogenomes, Austronesian, Austroasiatic

INTRODUCTION

Vietnam, with a consensus size of 98.21 million (General Statistics Office, 2021), comprises 54 ethnicities that can be stratified into five primary ethnolinguistic families: Austroasiatic (AA), Austronesian (AN), Thai-Kadai (TK), Sino-Tibetan (ST) and Hmong-Mien (HM). Notable here is AA, possessing the most diverse ethnicities (25 officially recognized groups) and accounting for 89% of the Vietnamese population. Being the most ancient language family in the region, AA was distributed widely and

blotchy, as later arrivals of non-AA speakers overlaid it. While the populous AA groups, such as Kinh or Muong, inhabit throughout the country, other AA minorities, such as Coho and Ma are more localized and occupy only certain areas. Ma and Coho populations in 2019 ranked 15th (0.21%) and 29th (0.05%), respectively (General Statistic Office, 2019). They were among the earliest inhabitants of Central Highland, sharing the same residential locations. These two maintained an intimate relationship, reflecting in many overlapping social traditions, especially in linguistic patterns (Tinh, 2020; Dang, 2024).

On the other hand, AN, being the main ethnolinguistic family on the island of Southeast Asia (ISEA), is represented by five minorities in Vietnam: Cham, Churu, Ede, Gia Rai, and Raglay. As the most influential Austronesian group (ranked 17th and 0.19% in consensus size), the Cham had a well-documented history, indicating their settlement in Vietnam as early as 500 BCE (General Statistics Office, 2021; General Statistic Office, 2019; Vickery, 2011). Their ancestors established the Champa empire, promoting their civilization over Central Vietnam in both lowland and highland parts (Vickery, 2011). As such, their heritage could be traced throughout the region, embedded in the folktales, customs, and architectures of multiple AA and AN community. To this day, the modern Cham continues to maintain stable economic and social alliances with local tribes. Raglay (ranked 18th and 0.15% in consensus size) was mistakenly considered a sub-branch of Cham people who migrated to higher altitudes of land, partly owing to their linguistic affiliation (Vuong, 2020). While culture and language can be transmitted horizontally between groups of people, genes are inherited vertically from parents to offspring, which could result in a mismatch between socio-cultural and genetic patterns.

So far, the academic interest has been underlined in the genetic characteristics of the well-documented Cham (Peng *et al.*, 2010). More recent studies touched on local groups like Giarai and Ede (Macholdt *et al.*, 2020; Duong *et al.*, 2018; Liu *et al.*, 2020). Given the broad spectrum of these minorities has yet to be inspected, we analyzed the mitogenomes of the Cham, Coho, Ma, and Raglay on both the molecular scale and maternal haplogroup components. The insights of these four Central minorities could elucidate any possible affinity between their genetic profile, contributing to the overall genetic architecture of the Vietnamese population.

MATERIALS AND METHODS

Subject and ethnic approval

Peripheral blood was obtained from 126 Vietnamese males belonging to two AN-speaking groups, Raglay (n = 37) and Cham (n = 17) and two AA-speaking groups, the Coho (n = 46) and Ma (n = 26), between the period of December 30th, 2019 and November 30th, 2022. Sampling locations were based in Lam Dong (Ma and Coho), Khanh Hoa (Raglay) and Binh Phuoc (Cham). Written informed consents were acquired from all blood donors. All participants were unrelated and self-identified as having at least three generations of the same ethnicity. This study received ethical approval from the Institutional Review Board of the Institute of Genome Research, Vietnam Academy of Science and Technology (No: 9-2019/NCHG-HĐĐĐ).

mtDNA sequencing

Genomic DNA was extracted using the GeneJET Whole Blood Genomic DNA Purification Mini Kit (ThermoFisher Scientific, USA). Construction of genomic libraries and capture enrichment for mtDNA were performed as described previously (Maricic *et al.*, 2010). Paired-end sequences of 150 bp length were generated using genomic libraries on the Illumina platform, and the reads underwent quality control and were processed as described previously (Arias *et al.*, 2018). An in-house alignment program was applied to align reads to the Reconstructed *Sapiens* Reference Sequence (RSRS) (Behar *et al.*, 2012), followed by multiple sequence alignments done by MAFFT v7.490 (Katoh and Standley, 2013). The complete mitogenome sequences were available in GenBank (Thao *et al.*, 2024).

Genetic analyses

Unique variants in each individual were identified via HaploGrep2 (Weissensteiner *et al.*, 2016). Variant assessments were performed via an in-house algorithm. Whole mtDNA (1-16569 bp) was divided into two regions: the D-loop (1-576; 16024-16569) and the coding region (577-16023). Embedded in the D-loop are three HVS regions: -I (16024-16383), -II (57-372), and -III (438-574).

HaploGrep2 identified the maternal haplogroup here with PhyloTree mtDNA tree Build 17 (Van Oven and Kayser, 2009). Except for haplogroup classification, subsequent analyses excluded positions with missing nucleotides (Ns) and the following sites: poly-C stretch of hypervariable segment 2 (HVS-II; nucleotide positions (np) 303-317); CA-repeat (np 514-523); C-stretch 1 (np 568-573), 12S rRNA (np 956-965), historical site (np 3107), C-stretch 2 (np 5895-5899), 9 bp deletion/insertion (np 8272-8289), and poly-C stretch of hypervariable segment 1 (HVS-I; np 16180-16195). The correspondence analysis (CA) was computed based on haplogroup frequencies in R via libraries "vegan v2.6-4" (Oksanen *et al.*, 2012) and "ca v0.71.1" (Nenadic and Greenacre, 2007). The pairwise genetic distance (Φ_{ST} distances) value, generated by arranging the populations in pairs, was performed by Arlequin version 3.5.2.2 (Excoffier and Lischer, 2010) and visualized on a heat plot via the R package "ggplot2 3.5.1" (Wickham, 2016).

RESULTS

Nucleotide variant distribution

Whole mitogenomes of 126 samples accounted for 445 unique variants, appearing 6807 times in this dataset. Out of 445 variants, 75.51% were screened in the coding region and the remaining were in the D-loop. Further investigation into the sub-structures of the D-loop showed that HVS-I, -II, and -III had 70, 26, and 8 unique variants, respectively. In the coding region, there were 336 variants, of which 204 are synonymous, 83 are non-synonymous, and 49 are located in the tRNA coding genes.

The mean variant count per person in the dataset is 54.02 ± 3.77 variants, with the highest in the Cham (54.76 ± 3.61 variants), and the lowest in the Raglay (53.08 ± 4.41 variants). The Ma and Coho each had 54.35 ± 3.44 and 54.33 ± 3.40 variants/person, respectively. A violin plot illustrating the distribution of variants (Figure 1) showed that the whole mtDNA density curves of Ma and Cham were quite similar, with more

variants concentrated above the median value. While both Raglay and Coho's density curves had more variants below the median, the former had a more stretching range of variants. In the coding region, the Raglay curve was the most skewed: 75.68% of variants were located above the median. On the other hand, Coho was more skewed toward the lower portion. Ma and Cham were more evenly spread. The variant range in the D-loop section of Cham, Coho, and Ma was compacted, while that of Raglay was more extended.

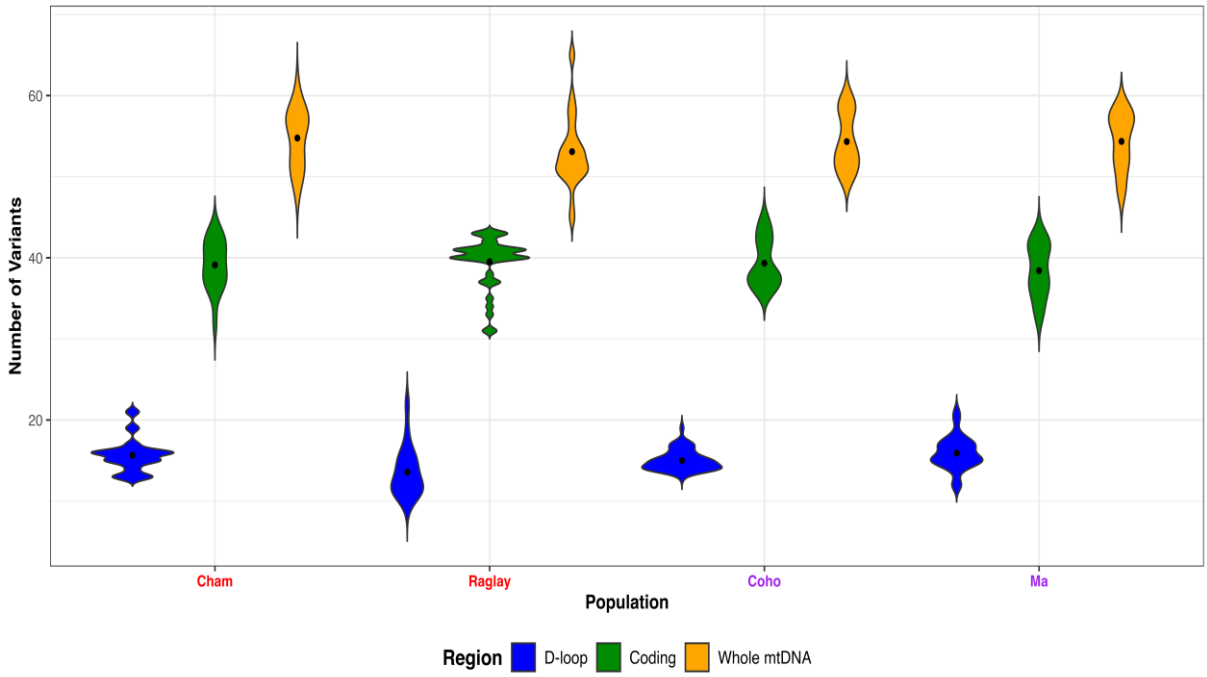

Figure 1. Variant distribution of different mtDNA regions among four populations. The language family is color-labeled: red-AN and purple-AA.

Genetic distances

To further assess how the nucleotide variants contributed to the genetic difference among populations, pairwise Φ_{ST} values were demonstrated on the heat plot (Figure 2). Raglay and Coho maintained the most genetic difference (0.13659), while Coho and Ma had the least (0.01613). Other considerable values belonged to Cham-Raglay (0.13361) and Ma-Raglay (0.10318).

Haplogroup profile

The maternal haplogroups of this dataset comprised 30 unique lineages, of which 14 appeared in 1 population, 12 were shared between 2 populations, three were shared among three populations, and only F1a1a1 was present in all four populations (Table 1). The Ma retained the most unique lineages (16), while the Cham had the least (10). The most common are B5a1b1 (11.11%), M24b (10.32%), F1a1a1 (7.94%) and M12b2a (7.94%). Other frequently distributed haplogroups included M68a1a, M12a1b, M12b2a, M51a, M24b, M21b and M7c1a.

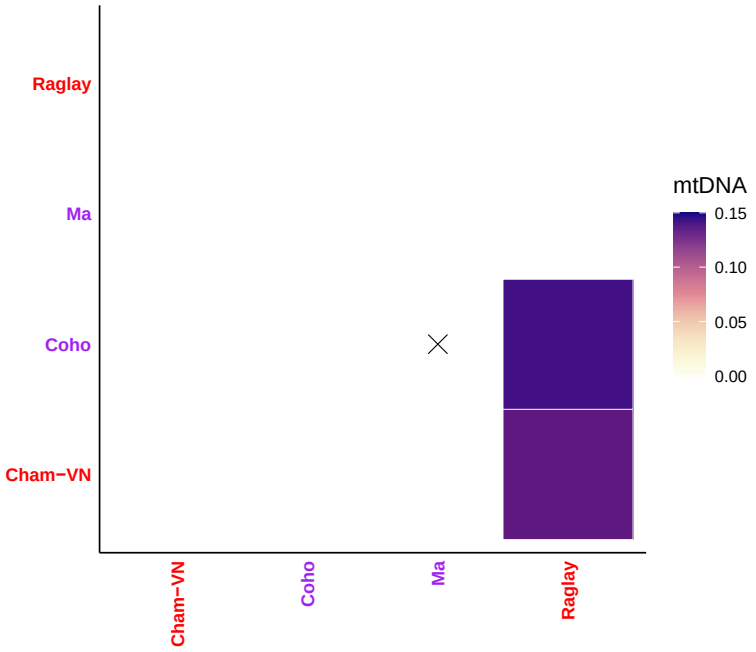

Figure 2. Genetic distance among populations based on pairwise values. The language family is color-labeled: Red-AN, Purple-AA. The cross symbol (X) denotes distance that is not significant (p > 0.05).

Out of 6 macro-haplogroups (B, E, F, M, N, R) detected in this dataset, M was the most predominant, taking account for 58.73% with 17 daughter clades. M lineages were dominant in Coho (58.70%) and Ma (57.69%) and reached the highest in Raglay (75.68%). Surprisingly, it was less significant in Cham (23.53%). Being descendants of macro haplogroup N, B (18.25%) and F (17.46%) were the second and third most common macro-haplogroups. B lineages were frequent among all populations except for Raglay, in which there was only 1 B5a1a sample (2.70%). F lineages were dominant in Cham (41.18%), especially with the two F1a1a and F1a1a1 (17.65% in each).

Table 1. Haplogroup frequencies of 126 individuals in Cham, Ma, Coho, and Raglay.

Haplogroup(s)	Cham (n=17)	Coho (n=46)	Ma (n=26)	Raglay (n=37)	Overall (n=126)
B	23.53%	28.26%	19.23%	2.70%	18.25%
B4c2	5.88%	2.17%	-	-	1.59%
B5a1a	5.88%	2.17%	-	2.70%	2.38%
B5a1b1	11.76%	21.74%	7.69%	-	11.11%
B5a1d	-	2.17%	11.54%	-	3.17%
E2a*	5.88%	-	-	-	0.79%

F	41.18%	10.87%	7.69%	21.62%	17.46%
F1a1	-	-	-	5.41%	1.59%
F1a1a	17.65%	4.35%	-	8.11%	6.35%
F1a1a1	17.65%	4.35%	7.69%	8.11%	7.94%
F1f	-	2.17%	-	-	0.79%
F2a	5.88%	-	-	-	0.79%
M	23.53%	58.70%	57.69%	75.68%	58.73%
M12a1b	-	-	11.54%	2.70%	3.17%
M12b1a2	5.88%	-	3.85%	-	1.59%
M12b2a	-	19.57%	3.85%	-	7.94%
M17a	-	-	-	2.70%	0.79%
M19	-	-	3.85%	-	0.79%
M21b	-	-	7.69%	16.22%	6.35%
M24b	-	-	-	35.14%	10.32%
M51a	-	15.22%	-	2.70%	6.35%
M51b	-	13.04%	3.85%	-	5.56%
M51b1b	-	-	3.85%	-	0.79%
M68a1a	-	10.87%	11.54%	-	6.35%
M71+151T	-	-	-	2.70%	0.79%
M72a	-	-	3.85%	2.70%	1.59%
M76	-	-	3.85%	-	0.79%
M77	-	-	-	5.41%	1.59%
M79	-	-	-	5.41%	1.59%
M7c1a	17.65%	-	-	-	2.38%
N	5.88%	-	11.54%	-	3.17%
N22	-	-	3.85%	-	0.79%
N9a6	5.88%	-	7.69%	-	2.38%
R23*	-	2.17%	3.85%	-	1.59%

Note: The asterisk * indicates that the lineage is the solely representative of its macro-haplogroup.

The distribution of these lineages was further visualized in the CA plot (Figure 3).

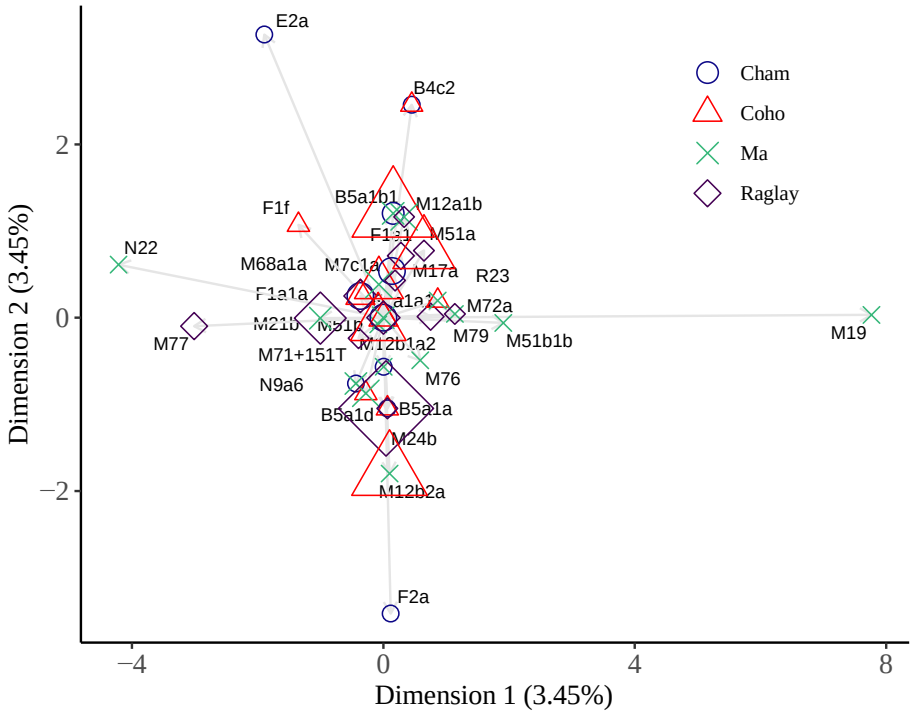

Figure 3. CA plot of haplogroups from 126 individuals in 4 Vietnamese populations (Cham, Coho, Ma and Raglay). The size of the symbol is proportional to the number of samples.

DISCUSSION

In this study, we analyzed the molecular variants and haplogroup profile of four Vietnamese minorities, two of which are linguistically affiliated with AN (Cham and Raglay) and the others with AA (Ma and Coho). The human mitogenome could be segmented into the D-loop and the coding region. D-loop, also known as the control region, contains 3 HVS regions, which are highly mutated. The coding region, taking 93% of the mitogenome length, was densely packed with 37 genes encoding for tRNA, rRNA, and peptide subunits of the respiratory chain complexes. Among them are 11 genes encoding for tRNA that do not carry any variant, suggesting their highly conservative nature. The coding region is more conservative than the D-loop; any mutations arising here are more likely to be non-synonymous (Guo *et al.*, 2023). In our dataset, the coding region had 60.71% variants that were non-synonymous; the rest were either synonymous or found in the tRNA-coding region. Pathogenic mutations of those genes were associated with clinical manifestations of multiple mitochondrial diseases, including Leber's Hereditary Optic Neuropathy (LHON), Leigh syndrome (LS), and mitochondrial encephalomyopathy with lactic acidosis and stroke-like episodes (MELAS) (Al-Kafaji *et al.*, 2022; Ng *et al.*, 2018; Danhelovska *et al.*, 2020; Lyu *et al.*, 2019).

In terms of population, 88 out of 445 unique variants were universal in all 126 individuals. Cham group retained 184 unique variants, in which portions of 6.52%, 7.61%, and 1.63% were shared pairwise with the Coho, Ma, and Raglay, respectively. Raglay possessed the highest number of nucleotide variants

specific to individual populations - 88/445 variants were observed solely in Raglay. It also exhibited considerable Φ_{ST} genetic distance, even with the AN-Cham (Figure 2). Coming from the AA language family and living in similar geographical locations, Ma and Coho are highly resembling in many socio-cultural elements (Dang, 2024). On the genetic aspects, these two shared 163 unique variants, 53 of which were absent in Raglay and Cham. The pair also had the least differentiation based on Φ_{ST} (Φ_{ST} = 0.01613).

The haplogroup architecture of four populations exhibited striking features of SEA regions: representatives of M, B, and F contributed the most, with the widespread distribution of SEA-signature clades such as F1a and B5a daughter lineages (Jaisamut *et al.*, 2023; Woravatin *et al.*, 2023; Bodner *et al.*, 2011; Kutanan *et al.*, 2017). In particular, the Cham had a significant portion of B5a1b1 (11.76%) and F1a1a1 (17.65%), as well as F1a1a (17.65%), possibly reflecting the admixture of modern Cham with other Vietnamese ethnics. M7c1a, observed in the Cham at 17.65%, was reported in other VN ethnicities such as AN-Churu (Thao *et al.*, 2024), TK-Thai (Duong *et al.*, 2018), AA-Northern Kinh (Duong *et al.*, 2018). In the MSEA, M7c1a was described in other language families such as TK people on Hainan Island, China (Peng *et al.*, 2011), TK and -ST in Thailand and Lao (Kutanan *et al.*, 2017; Kutanan *et al.*, 2020). Ma and Coho had similar macrohaplogroup component patterns, with the addition of N clades in the former. Raglay had M lineages outnumbering those of F and B macrohaplogroups. Moreover, the unique lineages underlying haplogroup components varied in each population. The CA plot demonstrated all haplogroups that emerged in only one population: M19 and N22 in Ma; E2a and F2a in Cham (Figure 3). F2a was prevalent in the VN-HM Pathen and VN-ST Phula (Duong *et al.*, 2018) and occurred less frequently in Taiwanese-AN. On the other hand, M19 and N22 were reported in the AN of the Philippines and Indonesia (Scholes *et al.*, 2011; Hill *et al.*, 2006). E2a was detected not only in the Filipino-AN and Indonesian-AN (Gunnarsdóttir *et al.*, 2011a; Gunnarsdóttir *et al.*, 2011b) but also in the Oceania natives (Duggan *et al.*, 2014). These lineages are rare and specific to the island populations, suggesting a possible migration from ISEA to Vietnam. Another notable lineage is M24b - the most common lineage in Raglay (35.14%). M24b was absent from the other three populations in this dataset. In MSEA, it was found at low frequencies in several Thai groups (AA, TK) (Kutanan *et al.*, 2017), Cambodian-Cham (Kloss-Brandstätter *et al.*, 2021), and Burmese-ST (Li *et al.*, 2015). The pair-wise comparison demonstrated that Ma and Coho had the highest number of shared haplogroups, including several M lineages M12b2a, M51b, and M68a1a (Table 1). M68a1a was also detected in Cambodian-AA and Cambodian-AN (Kloss-Brandstätter *et al.*, 2021; Zhang *et al.*, 2013). M12b2a was found in VN-AA (Mnong) (Thao *et al.*, 2024) and M51b was found in Thai-AN (Woravatin *et al.*, 2023).

CONCLUSION

Cham, Raglay, Coho, and Ma were indigenous people of the Central Vietnam. Many cultural and social customs overlap, reflecting their close-knit relationships throughout history. In this dataset, we explored the mitogenomes of 126 males residing in Central Vietnam. On the molecular level, 88/454 identified variants were ubiquitously shared among four populations, 220/454 were unique to 1

population, and the rest were found between at least 2 populations. The overall haplogroup distribution features were of the SEA region, especially in Cham. At the same time, specific lineages preserved unique patterns, with Raglay having the greatest number of exclusive lineages. Raglay also demonstrated considerable genetic distance Φ_{ST} with other studied groups, while Ma and Coho displayed the least. Combining with the archeological and linguistic evidences, these findings provide insights into the genetic structure of underrepresented minorities, lessening the knowledge gap of maternal haplogroups.

ACKNOWLEDGEMENTS

We express our gratitude to all sample donors for contributing to this research. This research was funded by the Ministry of Science and Technology, Vietnam (DTDL.CN.60/19). Dinh Huong Thao was funded by Vingroup JSC and supported by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), Institute of Big Data, code VINIF. 2022.TS.116.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Al-Kafaji, G., Alharbi, M. A., Alkandari, H., Salem, A. H. and Bakhiet, M. (2022). Analysis of the entire mitochondrial genome reveals Leber's hereditary optic neuropathy mitochondrial DNA mutations in an Arab cohort with multiple sclerosis. *Scientific Reports*, *12*(1), 11099. https://doi.org/10.1038/s41598-022-15385-2

Arias, L., Barbieri, C., Barreto, G., Stoneking, M. and Pakendorf, B. (2018). High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia. *American Journal of Physical Anthropology*, *165*(2), 238-255. https://doi.org/10.1002/ajpa.23345

Behar, Doron M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, Nuno M., *et al.* (2012). A "copernican" reassessment of the human mitochondrial DNA tree from its root. *The American Journal of Human Genetics*, *90*(4), 675-684. https://doi.org/10.1016/j.ajhg.2012.03.002

Bodner, M., Zimmermann, B., Röck, A., Kloss-Brandstätter, A., Horst, D., Horst, B., *et al.* (2011). Southeast Asian diversity: first insights into the complex mtDNA structure of Laos. *BMC Evolutionary Biology*, *11*(1), 49. https://doi.org/10.1186/1471-2148-11-49

Dang, H. G. (2024). The montagnard village: unique heritage of the Vietnamese central highlands. *The Russian Journal of Vietnamese Studies*, *7*(4), 71-81. https://doi.org/10.54631/VS.2023.74-623865

Danhelovska, T., Kolarova, H., Zeman, J., Hansikova, H., Vaneckova, M., Lambert, L., *et al.* (2020). Multisystem mitochondrial diseases due to mutations in mtDNA-encoded subunits of complex I. *BMC Pediatrics*, *20*(1), 41. https://doi.org/10.1186/s12887-020-1912-x

Duggan, A. T., Evans, B., Friedlaender, F. R., Friedlaender, J. S., Koki, G., Merriwether, D. A., *et al.* (2014). Maternal history of

Oceania from complete mtDNA genomes: contrasting ancient diversity with recent homogenization due to the Austronesian expansion. *American Journal of Human Genetics*, 721-733. https://doi.org/10.1016/j.ajhg.2014.03.014

Duong, N. T., Macholdt, E., Ton, N. D., Arias, L., Schröder, R., Van Phong, N., *et al.* (2018). Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia. *Scientific Reports*, *8*(1), 11651. https://doi.org/10.1038/s41598-018-29989-0

Excoffier, L. and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

General Statistic Office. (2019). Completed results of the 2019 Vietnam population and housing census. Vietnam: General Statistics Office. https://www.gso.gov.vn/wp-content/uploads/2019/12/Ket-qua-toan-bo-Tong-dieu-tra-dan-so-va-nha-o-2019.pdfpp

General Statistics Office. (2021). Infographic population, labour and employment in 2021. In: General Statistics Office.

Gunnarsdóttir, E. D., Li, M., Bauchet, M., Finstermeier, K. and Stoneking, M. (2011b). High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Research*, *21*(1), 1-11. https://doi.org/10.1101/gr.107615.110

Gunnarsdóttir, E. D., Nandineni, M. R., Li, M., Myles, S., Gil, D., Pakendorf, B., *et al.* (2011a). Larger mitochondrial DNA than Y-chromosome differences between matrilocal and patrilocal groups from Sumatra. *Nature Communications*, *2*(1), 228. https://doi.org/10.1038/ncomms1235

Guo, X., Xu, W., Zhang, W., Pan, C., Thalacker-Mercer, A. E., Zheng, H., *et al.* (2023). High-frequency and functional mitochondrial DNA mutations at the single-cell level. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(1), e2201518120. https://doi.org/10.1073/pnas.2201518120

Hill, C., Soares, P., Mormina, M., Macaulay, V., Meehan, W., Blackburn, J., *et al.* (2006). Phylogeography and Ethnogenesis of Aboriginal Southeast Asians. *Molecular Biology and Evolution*, *23*(12), 2480-2491. https://doi.org/10.1093/molbev/msl124

Jaisamut, K., Pitiwararom, R., Sukawutthiya, P., Sathirapatya, T., Noh, H., Worrapitirungsi, W., *et al.* (2023). Unraveling the mitochondrial phylogenetic landscape of Thailand reveals complex admixture and demographic dynamics. *Scientific Reports*, *13*(1), 20396. https://doi.org/10.1038/s41598-023-47762-w

Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772-780. https://doi.org/10.1093/molbev/mst010

Kloss-Brandstätter, A., Summerer, M., Horst, D., Horst, B., Streiter, G., Raschenberger, J., *et al.* (2021). An in-depth analysis of the mitochondrial phylogenetic landscape of Cambodia. *Scientific Reports*, *11*(1), 10816. https://doi.org/10.1038/s41598-021-90145-2

Kutanan, W., Kampuansai, J., Srikummool, M., Kangwanpong, D., Ghirotto, S., Brunelli, A., *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), 85-98. https://doi.org/10.1007/s00439-016-1742-y.

Kutanan, W., Shoocongdej, R., Srikummool, M., Hübner, A., Suttipai, T., Srithawong, S., *et al.* (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), 1563-1579. https://doi.org/10.1038/s41431-020-0693-x.

Li, Y.-C., Wang, H.-W., Tian, J.-Y., Liu, L.-N., Yang, L.-Q., Zhu, C.-L., *et al.* (2015). Ancient inland human dispersals from Myanmar into interior East Asia since the Late Pleistocene. *Scientific Reports*, *5*(1), 9473. https://doi.org/10.1038/srep09473.

Liu, D., Duong, N. T., Ton, N. D., Van Phong, N., Pakendorf, B., Van Hai, N., *et al.* (2020). Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity. *Molecular Biology and Evolution*, *37*(9), 2503-2519. https://doi.org/10.1093/molbev/msaa099.

Lyu, Y., Xu, M., Chen, J., Ji, Y., Guan, M.-X., Zhang, J. (2019). Frequency and spectrum of MT-TT variants associated with Leber's hereditary optic neuropathy in a Chinese cohort of subjects. *Mitochondrial DNA. Part B, Resources*, *4*(2), 2266-2280. https://doi.org/10.1080/23802359.2019.1627921.

Macholdt, E., Arias, L., Duong, N. T., Ton, N. D., Van Phong, N., Schröder, R., *et al.* (2020). The paternal and maternal genetic history of Vietnamese populations. *European Journal of Human Genetics*, *28*(5), 636-645. https://doi.org/10.1038/s41431-019-0557-4.

Maricic, T., Whitten, M. and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using pcr products. *PLoS ONE*, *5*(11), e14004. https://doi.org/10.1371/journal.pone.0014004.

Nenadic, O. and Greenacre, M. (2007). Correspondence analysis in R , with two- and three-dimensional graphics: the ca package. *Journal of Statistical Software*, *20*(3). https://doi.org/10.18637/jss.v020.i03.

Ng, Y. S., Lax, N. Z., Maddison, P., Alston, C. L., Blakely, E. L., Hepplewhite, P. D., *et al.* (2018). MT-ND5 mutation exhibits highly variable neurological manifestations at low mutant load. *Ebiomedicine*, 30, 86-93. https://doi.org/10.1016/j.ebiom.2018.02.010

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., *et al.* (2012). vegan: Community Ecology Package. https://github.com/vegandevs/vegan/issues.

Peng, M. S., Quang, H. H., Dang, K. P., Trieu, A. V., Wang, H. W., Yao, Y. G., *et al.* (2010). Tracing the austronesian footprint in mainland Southeast Asia: A perspective from mitochondrial DNA. *Molecular Biology and Evolution*, *27*(10), 2417-2430. https://doi.org/10.1093/molbev/msq131.

Peng, M.-S., He, J.-D., Liu, H.-X., Zhang, Y.-P. (2011). Tracing the legacy of the early Hainan Islanders - a perspective from mitochondrial DNA. *BMC Evolutionary Biology*, *11*(1), 46. https://doi.org/10.1186/1471-2148-11-46.

Scholes, C., Siddle, K., Ducourneau, A., Crivellaro, F., Järve, M., Rootsi, S., *et al.*

(2011). Genetic diversity and evidence for population admixture in Batak Negritos from Palawan. *American Journal of Physical Anthropology*, *146*(1), 62-72. https://doi.org/10.1002/ajpa.21544.

Thao, D. H., Dinh, T. H., Mitsunaga, S., Duy, D., Phuong, N. T., Anh, N. P.*, et al.* (2024). Investigating demic versus cultural diffusion and sex bias in the spread of Austronesian languages in Vietnam. *PLoS One*, *19*(6), e0304964. https://doi.org/10.1371/journal.pone.0304964.

Tinh, V. X. (2020). *Mon-Khmer group* (Tinh, V. X., Ed. 2nd ed., Vol. 3). National political publishing house.

Van Oven, M. and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*(2), E386-E394. https://doi.org/10.1002/humu.20921

Vickery, M. (2011). Champa revised. In the Cham of Vietnam: History, society and art (https://www.jstor.org/stable/j.ctv1qv2rd). NUS Press.

Vuong, X. T. (2020). *Part II: Sinitic and Austronesian groups* (Vuong, X. T., Ed. 2nd ed., Vol. 4). National political publishing house.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J.*, et al.* (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, *44*(W1), W58-W63. https://doi.org/10.1093/nar/gkw233

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd 2016 ed., Springer International Publishing: Imprint: Springer. https://link.springer.com/content/pdf/bfm:978-3-319-24277-4/1?pdf=chapter%20toc)

Woravatin, W., Stoneking, M. and Srikummool, M., Kampuansai, J., Arias, L., Kutanan, W. (2023). South Asian maternal and paternal lineages in southern Thailand and the role of sex-biased admixture. *Plos One*, *18*(9), e0291547. https://doi.org/10.1371/journal.pone.0291547

Zhang, X., Qi, X., Yang, Z., Serey, B., Sovannary, T., Bunnath, L.*, et al.* (2013). Analysis of mitochondrial genome diversity identifies new and ancient maternal lineages in Cambodian aborigines. *Nature Communications*, *4*(1), 2599. https://doi.org/10.1038/ncomms3599.