ACADEMIA JOURNAL OF BIOLOGY 2024, 46(3): 63–72
DOI: 10.15625/2615-9023/18604

THE GENETIC STRUCTURES OF THE CHURU, EDE AND GIARAI UNRAVELLED BY COMPLETE MITOCHONDRIAL DNA

Dinh Huong Thao[1,2], Tran Huu Dinh[1], Nguyen Thuy Duong[1,*]

[1]Institute of Genome Research, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Ha Noi, Vietnam
[2]Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Ha Noi, Vietnam

Received 19 May 2024; accepted 4 September 2024

ABSTRACT

Vietnam, a nation with a rich and complex history of migration and settlement, is home to 5 fundamental language families: Austroasiatic (AA), Tai-Kadai (TK), Austronesian (AN), Sino-Tibetan (ST) and Hmong-Mien (HM). Among them is the Austronesian, a language family substantial in island Southeast Asia (ISEA) but marginal in mainland counterpart (MSEA), constituted five Vietnamese ethnolinguistic groups. Here, we analyzed the control region, and the complete mitochondrial DNA (mtDNA) of 121 individuals from 3 AN-speaking populations (Churu, Ede, and Giarai). To explore the molecular diversity, the sequences were aligned against the Reconstructed Sapiens Reference Sequence (RSRS). The quantification and distribution of nucleotide variations resulted in 6,369 variants in our dataset in which the control region and coding region retained 1,707 and 4,662 variants, respectively. Churu harbored the most diversity (54.6 ± 2.8 variants/person), followed by Giarai (52.2 ± 3.3 variants/person), and Ede (51.1 ± 5.3 variants/person). Both the control region and whole mtDNA were input to Haplogrep3 to call haplogroups, resulting in 47.11% of our samples having their haplogroup changed from 17 whole mtDNA lineages to 16 different control region lineages. The haplogroup profile derived from whole mtDNA included 31 unique clades, in which only B5a1d was shared among three groups, and 23/31 lineages were present exclusively in a single population. The haplogroup component of each minority also revealed that all 3 AN groups had the majority of their samples attributed to the macrohaplogroups M, B, and F, with the disparity fixed in their underlying sublineages. This study increased the knowledge wealth of the genetic characteristics of AN speakers in the region from a different analysis approach, and highlighted the contribution of variants in different complete mtDNA, providing insight to reconstruct a comprehensive genetic architecture of Vietnam.

Keywords: Churu, Ede, Giarai, mtDNA, Vietnam.

Citation: Dinh Huong Thao, Tran Huu Dinh, Nguyen Thuy Duong, 2024. The genetic structures of the Churu, Ede and Giarai unravelled by complete mitochondrial DNA. *Academia Journal of Biology*, 46(3): 63–72. https://doi.org/10.15625/2615-9023/18604

[*]*Corresponding author email*: tdnguyen@igr.ac.vn; https://orcid.org/0000-0001-8691-9138

INTRODUCTION

Vietnam is the homeland to 54 officially recognized ethnic groups, belonging to 5 language families: Austroasiastic (AA), Sino-Tibetan (ST), Thai-Kadai (TK), Hmong-Mien (HM) and Austronesian (AN). The general consensus reported that 85.32% of the national population were the AA Kinh, leaving the remaining 14.68% divided into 53 ethnolinguistic groups (General Statistics Office, 2019). Many of these minorities either resided in reclusive areas or/and had diminished populations. As such, the enormous diversity of Vietnamese people, especially in the biological aspect, required immediate measures to be preserved and understood. Among these underrepresented were the Austronesian speakers, whose traces of arrival could be found prior to the establishment of the Champa kingdom around 500 BCE (Vickery, 2011).

AN is a vast language family of more than 1200 dialects, stretching from Madagascar of Eastern Africa, South East Asia (SEA), to Eastern Island on the far east of the Pacific (Eberhard et al., 2023). In Vietnam, they are Cham, Churu, Ede, Giarai, and Raglay, constituting 1.32% of the nation's demography (General Statistics Office, 2019). Modern Austronesian communities in Vietnam (VN-AN) mostly occupied the mountainous area of the Central Highland and coastline of the South Central. Being the most populous AN nation in MSEA, Vietnam had Cham, Ede, and Giarai explored on various degrees. The first VN-AN ethnic to be examined was the Cham in a study of mitochondrial DNA (mtDNA) hypervariable segments (HVS) by Peng et al. (2010). Since then, both the uniparental markers (Y-chromosome and mtDNA) and the genome-wide data of the Giarai-I and Ede-I were unfolded (Duong et al., 2018; Liu et al., 2020; Macholdt et al., 2020). So far, the comprehensive picture of this ethnolinguistic family stayed patchy, urging for more evidences to fill in the missing pieces.

MtDNA has long been a preferred uniparental marker to study evolution and population genetics. Its structure could be further divided into sub-regions: the coding (range: 577–16,023) and the control (range: 1–576; 16,024–16,569). Packed with encoded genes, the former is highly conservative, while the latter retained fast mutational rate. Embedded within control region are HVS -I (range: 16,024–16,383), -II (range: 57–372) and -III (range: 438–574), three particular sites accounted for most variables. As such, variants in the control region have been routinely used to define many branches on the phylogenetic tree. With the advancement of next-generation sequencing (NGS), sequencing whole mtDNA became less resource-consuming, providing a more accurate haplogroup profile and, therefore, a finer phylogenetic resolution. In this study, we analyzed the genetic characteristics of 121 males from 3 VN-AN indigenous tribes (Churu, Ede, and Giarai). Nucleotide variants were used for the first time to assess the diversity on the molecular level. To determine the importance of different mtDNA regions, sequences of the control region and complete mtDNA were implemented to extract haplogroup information. The dataset present here would provide details on the maternal genetic structures of individual minorities as well as the AN family in VN.

MATERIALS AND METHODS

Sample information

Whole blood samples were obtained from 121 males of 3 VN-AN populations (Churu, Ede, and Giarai). All participants consenting to donate blood were unrelated and self-identified to have at least three generations of the same ethnicity. The sampling locations were Lam Dong (Churu), Dak Lak (Ede), and Kontum (Giarai). This study received ethical approval from the Institutional Review Board of the Institute of Genome Research, Vietnam Academy of Science and Technology (No: 2-2019/NCHG-HĐĐĐ).

To distinguish between different sets of samples from the same ethnicity, the Ede and Giarai in this study were labeled with -II, and the ones in Duong et al., 2018 were labeled

with -I. Furthermore, the Cham and Giarai in Cambodia were referred to as CB-Cham and CB-Giarai (Kloss-Brandstätter et al., 2021; Zhang et al., 2013), and the Cham in Vietnam was named VN-Cham (Peng et al., 2010).

mtDNA sequencing

Genomic DNAs were extracted by GeneJET Whole Blood Genomic DNA Purification Mini Kit (ThermoFisher Scientific, USA) following the manufacturer's protocol. Construction of genomic libraries and capture-enrichment for mtDNA were performed using the method by Maricic et al. (2010). The libraries were sequenced on Illumina platform. The reads generated by sequencing were undergone quality control and processed as described previously, then were aligned to the Reconstructed Sapiens References Sequence (RSRS) (Behar et al., 2012), using an in-house alignment program. Multiple sequence alignment was performed using MAFFT (Katoh & Standley, 2013). The mitogenome sequences of 121 samples were available in GenBank (Thao et al., 2024).

Genetic analyses

To locate the nucleotide variants on multiple mtDNA segments (coding and control region), reads were aligned against RSRS using an in-house algorithm. Positions with missing nucleotide (Ns) and other 8 sites were excluded: 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 3,107); C-stretch 2 (np 5,895–5,899); 9 bp deletion/insertion (np 8,272–8,289); and poly-C stretch of hypervariable segment 1 (HVS-I; np 16,180–16,195). The distribution of variants across three populations was visualized by the R package "ggplot2". The control region (1–576 bp; 16,024–16,569 bp) and entire mtDNA sequences were implemented to classify haplogroups via HaploGrep3 (Weissensteiner et al., 2016) with PhyloTree mtDNA tree Build 17 (van Oven & Kayser, 2009). The correspondence analysis (CA) was computed based on haplogroup frequencies in R via libraries "vegan v2.6-4" (Oksanen et al, 2022) and "ca v0.71.1" (Nenadic & Greenacre, 2007).

RESULTS

Variants distribution

We screened 6369 variants in our sample set, in which the control and the coding region took a portion of 73.2% and 26.8%, respectively. In term of population group, Churu had the highest number of variants per individual (54.6 ± 2.8 variants/person). Giarai-II was the second, with 52.2 ± 3.3 variants/person. Ede-II had the least variants, only 51.1 ± 5.3 variants/person.

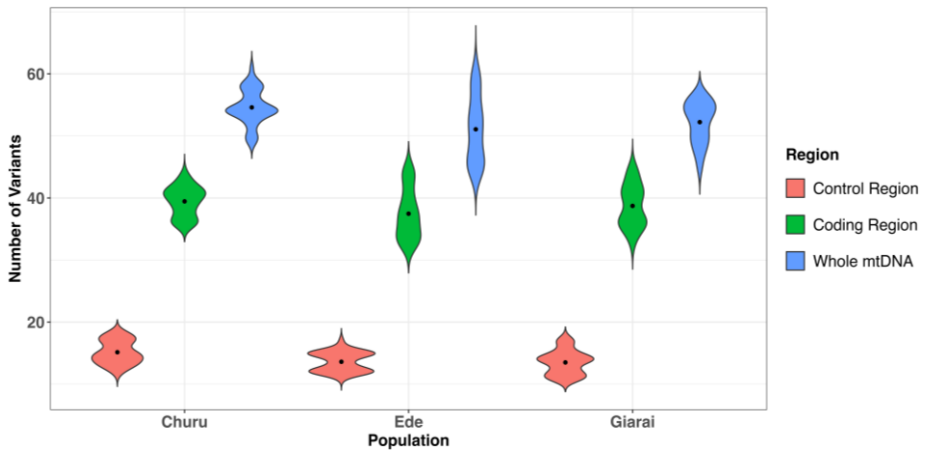

Figure 1. Variant distribution across the complete mitochondrial sequences of Churu, Ede-II, and Giarai-II. Different mitochondrial DNA regions were color-labeled: red is the control region, green is the coding region, blue is the entire mitogenome. Black dot denoted the median values.

The distribution of variants in different mitogenomes was visualized on the violin plot (Figure 1). In the control region, the distribution curve of Churu was broader from the median to the lower portion. In Ede-II it was skewed at the median point, dividing the curve into two noticeable parts. The curve in Giarai-II was the opposite of that in Churu: it was wider from the median point to the upper portion. In the coding region, the median was highest among the Churu (Figure 1), followed by the Giarai-II and Ede-II. Churu had the broadest area around the median value; Ede-II and Giarai-II had thinner and more prolonged tips. When comparing the whole mtDNA, Ede-II had the most elongated distribution. In Giarai-II, the upper portion was wider and shorter than the lower portion. In Churu, the most extended part was centralized around the median point, with more outliers on the top.

Haplogroup classification

To evaluate the significance of variants in coding and control regions, the sequences of the later were aligned to RSRS to call haplogroups. Details of the differentiation between using whole mtDNA and control region sequences to classify haplogroups were listed in Table 1 below. Overall, 43.8% of our samples had their haplogroups changed, from 17 whole mtDNA to 16 control region haplogroups. The number of unique polymorphic sites were 98 in the control regions and 441 in the entire mitogenomes of 121 individuals. Notably, 14 out of 15 M71 + 151T (assigned using whole mtDNA) switched to D6a1 (assigned using control region sequences). All F1a1a1 samples defined by variants in whole mtDNA were corresponded to F1a1a defined by those in the control region.

Table 1. Whole mtDNA and control region haplogroup differentiation

Haplogroups		Number of samples	Percentage (%)
Whole mtDNA	Control region		
B5a1a	B5a	4	3.31
B5a1b1	B5a	2	1.65
B5a1c	B5a1d	1	0.83
C7	C4c1b	1	0.83
F1a1a1	F1a1a	8	6.61
M7c1a	M7c1a1b	1	0.83
M21b	M7c	6	4.96
M71+151T	D6a1	14	11.57
M73b	M73’79	5	4.13
M74	D4h4a	1	0.83
M74	D4p	1	0.83
M74b2	M43a1	3	2.48
M7b1a1b	M7b1a1	1	0.83
M7b1a1f	M7b1a1a	1	0.83
M9a’b	D4l1a	1	0.83
N21a	N21	2	1.65
N7b	M5	1	0.83

From the complete mitochondrial genomes of 121 VN-AN individuals, 31 haplogroups were stratified into seven macro-haplogroups: B (15.70%), C (0.83%), D (1.65%), F (16.53%), M (52.89%), N (6.61%), R (5.79%) (Table 2). Macro-haplogroups B, F and M and their sub-branches were predominant, accounting for 85.12% of our dataset. A total of 23 assigned lineages appearred in a single ethnic group only, including 10 singletons. A total of 14 haplogroups were assigned to the Churu, of which 4 were singleton. The most

frequent were M12b1a2 and R22 (16.67% each), F1a1a, and M76 (14.29% each). There were 12 haplogroups arising in the Ede-II; three of those lineages were singleton. The most common were M71+151 (34.88%), F1a1a1 (16.28%) and B5a1d (13.95%). In the Giarai-II, 3 out of 14 lineages were singleton. The most widely distributed were M21 and M73b (13.89% each), B5a1d, and F1a1d (11.11% each). The distribution was further visualized on the haplogroup frequency-based CA plot (Figure 2) indicating B5a1d as the only lineage shared among all three populations. Striking outliers were separated by M73b (Giarai-II), M7c1a (Churu), B5a1b1 (Giarai-II) and M12a1b (Churu).

Table 2. Haplogroup composition and distribution in Churu, Ede, and Giarai. Haplogroup denoted with * was the only representative of its macrohaplogroup in this dataset

Haplogroup (s)	Churu (n = 42)	Ede-II (n = 43)	Giarai-II (n = 36)	Total
B	14.29%	16.28%	16.67%	15.70%
B5a1a	7.14%	2.33%	-	3.31%
B5a1b1	-	-	5.56%	1.65%
B5a1c	2.38%	-	-	0.83%
B5a1d	4.76%	13.95%	11.11%	9.92%
C7*	-	-	2.78%	0.83%
D5a2a1*	-	-	5.56%	1.65%
F	16.67%	20.93%	11.11%	16.53%
F1a1a	14.29%	4.65%	-	6.61%
F1a1a1	2.38%	16.28%	-	6.61%
F1a1d	-	-	11.11%	3.31%
M	42.86%	58.14%	58.33%	52.89%
M12a1b	2.38%	-	-	0.83%
M12b1a2	16.67%	-	-	5.79%
M20	2.38%	-	8.33%	3.31%
M21b	-	2.33%	13.89%	4.96%
M21b2	4.76%	2.33%	-	2.48%
M71+151T	-	34.88%	-	12.40%
M71a2	-	2.33%	-	0.83%
M73b	-	-	13.89%	4.13%
M74	-	-	5.56%	1.65%
M74b1	-	11.63%	-	4.13%
M74b2	-	-	8.33%	2.48%
M76	14.29%	-	2.78%	5.79%
M7b1a1a	-	-	2.78%	0.83%
M7b1a1b	-	2.33%	-	0.83%
M7b1a1f	-	2.33%	-	0.83%
M7c1a	2.38%	-	-	0.83%
M9a'b	-	-	2.78%	0.83%
N	9.52%	4.65%	5.56%	6.61%
N21a	-	-	5.56%	1.65%
N22	7.14%	-	-	2.48%
N7b	2.38%	-	-	0.83%
N8	-	4.65%	-	1.65%
R22*	16.67%	-	-	5.79%

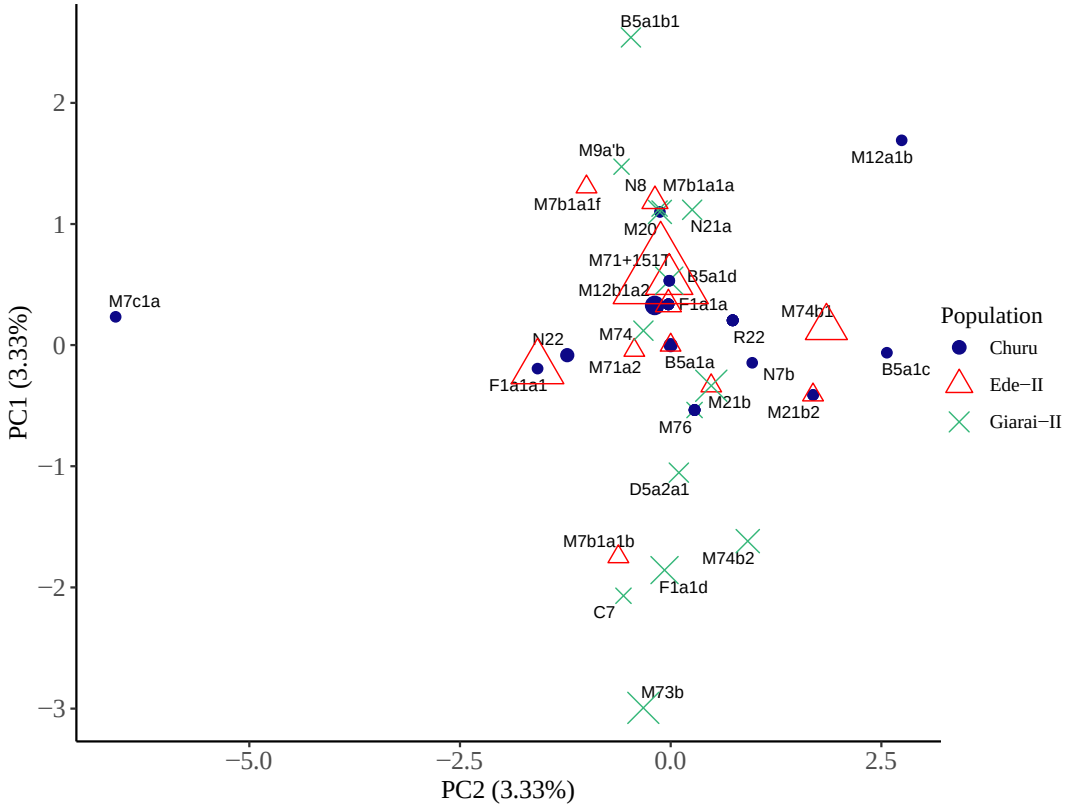

Figure 2. Haplogroup frequencies based-CA plot of Churu, Ede-II, and Giarai-II. The size of the haplogroup label is proportional to the number of samples possessing that haplogroup.

DISCUSSION

On the molecular level, Churu possessed the highest mean number of variants per person (54.6 ± 2.8), followed by Giarai-II (52.2 ± 3.3) and Ede-II (51.2 ± 5.3). The distribution of variants across different mtDNA regions (Figure 1) indicated that the majority of Churu had a number of variants that were close to median values. On the other hand, Giarai-II and especially Ede-II had more widespread distribution patterns, suggesting many Ede-II and Giarai-II had either extremely low or high number of variants. In terms of mtDNA sub-structures, the average number of variants in the control region was about 4 times higher than that of whole mtDNA, underlining the amount of data compacted in this limited size region. However, more than half (52.54%) of the polymorphic sites in this study were in the coding region, which possibly created difference between using the control region sequences and whole mtDNA to determine haplogroups. A further inspection into the haplogroup taxonomy rule explained that the reason for discordance was the location of the critical mutations defining a particular haplogroup: clades with defining mutations in the coding region could not be called by using solely the control region (van Oven & Kayser, 2009). Therefore, 17 lineages defined by whole mtDNA are not identical to those defined by the sequences (Table 1).

Whole mtDNA haplogroup classification revealed the macro-haplogroups B, F, and M accounted for 85.12% of our samples (Table 2). Being the most widely distributed in the Asia continent, those three kept their own distinct subsets of lineages in East-, South- and SE- Asia (Underhill & Kivisild, 2007). As an extensive macrohaplogroup whose direct ancestral was the Out-of-Africa L3, M made up more than half

(17/31) of the detected lineages in this study, with only 4 (M20, M21b, M21b2, and M76) being shared between 2 populations. M7b1a1b and M7c1a were also two singletons attributed to the outstanding Churu and Ede-II samples on the CA plot (Figure 2). Firstly described in Ede-I and Giarai-I, M71 + 151T was considered to be a signature of VN-AN (Duong et al., 2018). It was absent in Churu and Giarai-II but present in Ede-II at a considerable proportion (34.88%). This discordance could be partially influenced by the geographical difference of acquired samples: both Ede-I and -II originated in Dak Lak province; Giarai-I and -II came from two different places (Gia Lai and Kontum, respectively) (Duong et al., 2018).

F and B were descended from macrohaplogroup N, which was a sister clade to macrohaplogroup M. All 4 of the assigned lineages here were of B5a lineages, in which B5a1d was the only haplogroup shared by the three ethnics. Within the island AN-speakers, B5a1d was detected in the Filipinos Abaknon (3.33%) and several Indonesian groups (ranging from 0.85-5.56%) (Delfin et al., 2014; Kusuma et al., 2015) but did not occurred in Taiwan (Delfin et al., 2014; Tätte et al., 2021; Trejaut et al., 2005). On the continent, B5a1d was found in Ede-I (8.33%), Giarai-II (3.33%), CB-Giarai (7.84%), CB-Cham (4.92%) (Duong et al., 2018; Kloss-Brandstätter et al., 2021; Zhang et al., 2013). Overall, B5a and its subclades were relatively rare in oceanic areas and more common on the mainland, AN, and non-AN alike. They reached high frequencies in Laos (Bodner et al., 2011), Cambodia (Zhang et al., 2013), Thailand (Kutanan et al., 2017), and Myanmar (Summerer et al., 2014). B4a and its sub-branches (including the 'Polynesian motif'-B4a1a1a), which were more prevalent in ISEA than in MSEA (Matsumura et al., 2018), were virtually absent in our dataset. F lineages in our study comprised of F1a1d, F1a1a1, and F1a1a. While the first one was exclusive to Giarai-II (11.11%), both F1a1a1 and F1a1a were shared between Churu and Ede-II but lacking in Giarai-II. Rarely occurring in non-AN populations, F1a1d appeared in several Taiwanese-AN groups (Ko et al., 2014) and reached the peak of 14.67% in Taiwanese-AN Yami (Tätte et al., 2021). F1a1a was well-characterized in SEA, particularly dominant in Myanmar (Summerer et al., 2014), and Cambodia (Zhang et al., 2013). F1a1a1 was less common in published studies, showing up in Ede-I, Cham-CB, and non-AN CB populations (Duong et al., 2018; Kloss-Brandstätter et al., 2021; Zhang et al., 2013).

CONCLUSION

Coming from the same ethnolinguistic family, Churu, Ede, and Giarai are three members of the Austronesian family in Vietnam. All of them held a similar macrohaplogroups component pattern. At the same time, each group exhibited distinctive features on both the molecular level and haplogroup distribution. The connection of the same language family could be responsible for the overlapping genetic elements among 3 populations, yet the influence of geographic location could be a significant factor. Austronesian speakers were distributed over the central part of Vietnam, a hub of diverse minorities. For the first time, the molecular diversity of 121 VN-AN was inspected from nucleotide variants, suggesting Churu is more heterogeneous than Ede and Giarai. Haplogroup resolution of our dataset was elevated upon using the whole mtDNA to assign haplogroup, signifying the advantage of extracting variants from both coding and control regions. This set of complete mitochondrial DNA provided a glimpse into the maternal genetic architecture of the Austronesians in Vietnam. Future studies on Y-chromosome markers and genome-wide data could further elucidate the holistic picture of Austronesian.

Acknowledgements: We express our gratitude to Prof. Inoue Ituro for making this work possible. We thank all sample donors for contributing to this research. This research was funded by the Ministry of Science and Technology, Vietnam (60/19-DTDL.CN-XNT). 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.

REFERENCES

Behar D. M., van Oven M., Rosset S., Metspalu, M., Loogväli E.-L., Silva N. M., Kivisild T., Torroni A., & Villems R., 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., Sengchanh S., Sanguansermsri T., Horst J., Krämer T., Schneider P. M. & Parson W., 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

Delfin F., Min-Shan Ko A., Li M., Gunnarsdóttir E. D., Tabbada K. A., Salvador J. M., Calacal G. C., Sagum M. S., Datar F. A., Padilla S. G., De Ungria M. C. A. & Stoneking M., 2014. Complete mtDNA genomes of Filipino ethnolinguistic groups: A melting pot of recent and ancient lineages in the Asia-Pacific region. *European Journal of Human Genetics*, 22(2): 228–237. https://doi.org/10.1038/ejhg.2013.122

Duong N. T., Macholdt E., Ton N. D., Arias L., Schröder R., Van Phong N., Thi Bich Thuy V., Ha N. H., Thi Thu Hue H., Thi Xuan N., Thi Phuong Oanh K., Hien L. T. T., Hoang N. H., Pakendorf B., Stoneking M. & Van Hai N., 2018. Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia. *Scientific Reports*, 8(1): 11651. https://doi.org/10.1038/s41598-018-29989-0

Eberhard D. M., Simons G. F., & Fennig C. D. (Eds.), 2023. Ethnologue: Languages of the World. (26th ed.). SIL International. http://www.ethnologue.com.

General Statistic Office, 2019. Completed Results of the 2019 Vietnam Population and Housing Census (p. 840) [Population Census]. 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.pdf

Katoh K. & 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., Kronenberg F., Sanguansermsri T., Horst J. & Weissensteiner H., 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

Ko A. M.-S., Chen C.-Y., Fu Q., Delfin F., Li M., Chiu H.-L., Stoneking M. & Ko Y.-C., 2014. Early Austronesians: Into and Out Of Taiwan. *The American Journal of Human Genetics*, 94(3): 426–436. https://doi.org/10.1016/j.ajhg.2014.02.003

Kusuma P., Cox M. P., Pierron D., Razafindrazaka H., Brucato N., Tonasso L., Suryadi H. L., Letellier T., Sudoyo H., & Ricaut F.-X., 2015. Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. *BMC Genomics*, 16(1): 191. https://doi.org/10.1186/s12864-015-1394-7

Kutanan W., Kampuansai J., Srikummool M., Kangwanpong D., Ghirotto S., Brunelli A. & Stoneking M., 2017. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai–Kadai languages. *Human Genetics*, 136(1): 85–98. https://doi.org/10.1007/s00439-016-1742-y

Liu D., Duong N. T., Ton N. D., Van Phong N., Pakendorf B., Van Hai N. & Stoneking M., 2020. Extensive Ethnolinguistic Diversity in Vietnam Reflects Multiple Sources of Genetic Diversity. *Molecular Biology and*

Evolution, 37(9): 2503–2519. https://doi.org/10.1093/ molbev/msaa099

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

Maricic T., Whitten M. & 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

Matsumura H., Shinoda K., Shimanjuntak T., Oktaviana A. A., Noerwidi S., Octavianus Sofian H., Prastiningtyas D., Nguyen L. C., Kakuda T., Kanzawa-Kiriyama H., Adachi N., Hung H., Fan X., Wu X., Willis A. & Oxenham M. F., 2018. Cranio-morphometric and aDNA corroboration of the Austronesian dispersal model in ancient Island Southeast Asia: Support from Gua Harimau, Indonesia. *Plos one*, 13(6): e0198689. https://doi.org/10.1371/journal.pone.0198689

Nenadic O., Greenacre M., 2007. Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package. *Journal of Statistical Software*, 20(3): 1–13. doi: 10.18637/jss.v020.i03

Oksanen J., Simplson G. L., Blanchet F. G., Kindt R., Legendre P., Minchin P. R., O'Hara R. B., Solymos P., Stevens M. H. H, Szoecs E., Wagner H., Barbour M., Bedward M., Bolker B., Borcard D., Carvalho G., Chirico M., De Caceres M., Durand S., Evangelista H. B. A., FitzJohn R., Friendly M., Furneaux B., Hannigan G., Hill M. O., Lahti L., McGlinn D., Ouelette M-H., Cunha E. R., Smith T., Stier A., Braak C. J. F. T, Weedon J., 2022. Vegan: community ecology package, 2.6-2. R Foundation for Statistical Computing.

Peng M.-S., Quang H. H., Dang K. P., Trieu A. V., Wang H.-W., Yao Y.-G., Kong Q.-P. & Zhang Y.-P., 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

Phuong T. K., Lockhart B. M. & Vickery Michael (Eds.), 2011. Champa Revised. In *The Cham of Vietnam: History, Society and Art* (1–37). NUS Press. https://doi.org/10.2307/j.ctv1qv2rd

Summerer M., Horst J., Erhart G., Weißensteiner H., Schönherr S., Pacher D., Forer L., Horst D., Manhart A., Horst B., Sanguansermsri T. & Kloss-Brandstätter A., 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): 17. https://doi.org/10.1186/1471-2148-14-17

Tätte K., Metspalu E., Post H., Palencia-Madrid L., Luis J. R., Reidla M., Rea A., Tamm E., Moding E. J., de Pancorbo M. M., Garcia-Bertrand R., Metspalu M. & Herrera R. J., 2021. The Ami and Yami aborigines of Taiwan and their genetic relationship to East Asian and Pacific populations. *European Journal of Human Genetics*, 29(7): 1092–1102. https://doi.org/10.1038/s41431-021-00837-6

Thao D. H., Dinh T. H., Mitsunaga S., Duy L. D., Phuong N. T., Anh N. P., Anh N. T., Duc B. M., Thi Thu Hue H., Ha N. H., Ton N. D., Hübner A., Pakendorf B., Stoneking M., Inoue I., Duong N. T., Hai N. V., 2024, Investigating demic versus cultural diffusion and sex bias in the spread of Austronesian languages in Vietnam, *bioRxiv*. https://doi.org/10.1101/2024.03. 19.585662

Trejaut J. A., Kivisild T., Loo J. H., Lee C. L., He C. L., Hsu C. J., Li Z. Y. & Lin M., 2005. Traces of Archaic Mitochondrial Lineages Persist in Austronesian-Speaking Formosan Populations. *PLoS Biology*, 3(8): e247. doi: 10.1371/journal.pbio.0030247

Underhill P. A. & Kivisild T., 2007. Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. *Annual Review of Genetics*, 41(1): 539–564. https://doi.org/10.1146/annurev.genet.41.110306.130407

van Oven M. & 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

Weissensteiner H., Pacher D., Kloss-Brandstätter A., Forer L., Specht G., Bandelt H.-J., Kronenberg F., Salas A. & Schönherr S., 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

Zhang X., Qi X., Yang Z., Serey B., Sovannary T., Bunnath L., Seang Aun H., Samnom H., Zhang H., Lin Q., van Oven M., Shi H. & Su B., 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