ACADEMIA JOURNAL OF BIOLOGY 2025, 47(1): 85–95
DOI: 10.15625/2615-9023/20952

GENETIC VARIATION OF THE D-LOOP REGION OF THE KINH ETHNIC GROUP FROM THE NORTHERN, CENTRAL, AND SOUTHERN VIETNAM

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

[1]Institute of Biology, 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 11 February 2024; accepted 19 March 2025

ABSTRACT

The displacement-loop (D-loop) region is the 1122 bp sequence of the mitochondrial DNA (mtDNA) that plays an important role in mtDNA replication and transcription. This region is also a hotspot for mutations, making it a good candidate for genetic diversity studies, especially when combined with other unique characteristics of the mitochondria. The Kinh, which accounts for approximately 90% of the Vietnamese population, has been extensively used as a representative of the Vietnamese population. In this study, we sampled and sequenced the mtDNA of 31 Kinh individuals from Central Vietnam and 33 Kinh individuals from Southern Vietnam. In combination with 50 Northern Kinh mtDNA sequences from a previous study, the present study analyzed the diversity of the Kinh across Vietnam through the D-loop region. After aligning to the reference mtDNA sequence RSRS (Reconstructed Sapiens Reference Sequence), 134 unique variants were found in the D-loop region, including 34 variants found in all three populations, 25 variants found in two populations, and 75 variants unique to one Kinh population. Haplogroup calling using Haplogrep3 identified 60 haplogroups, with 23 out of 60 being absent in the Northern Kinh population. Interestingly, despite the genetic diversity of the Kinh, four haplogroups were found in all three populations (F1a1a, F1a2, B5a, M7b1a1+(16192T)). The pairwise genetic distance analysis showed that the Northern Kinh is distinguishable from the Central Kinh while the other pairing experienced the opposite trend. By sampling the Kinh group in three different regions, our results provided the basis for further genetic diversity analysis of the Kinh across Vietnam.

Keywords: D-loop, Kinh, mtDNA, variant, Vietnam.

Citation: La Duc Duy, Dinh Huong Thao, Nguyen Thuy Duong, 2025. Genetic variation of the D-loop region of the Kinh ethnic group from the northern, central, and southern Vietnam. *Academia Journal of Biology*, 47(1): 85–95. https://doi.org/10.15625/2615-9023/20952

[*]*Corresponding author email*: tdnguyen@igr.ac.vn

INTRODUCTION

Mitochondrial DNA (mtDNA) is a small circular DNA located within the mitochondria. It encodes for 37 genes, including two rRNA-encoding genes, 22 tRNA-encoding genes, and 13 genes encoding proteins of electron transport chain complexes (Anderson et al., 1981). With its characteristics such as high mutation rate (Brown et al., 1979), exclusive maternal inheritance (Hutchison et al., 1974), high copy number (Michaels et al., 1982), and the absence of recombination (Hagström et al., 2014; Merriwether et al., 1991), mtDNA has been used for human population studies (Pakendorf & Stoneking, 2005). In the 16,569 bases long circular DNA, the displacement-loop (D-loop) region, spanning from position 16024 to position 576, is the non-coding region of the mtDNA and plays an important role in mtDNA replication and transcription (Taanman, 1999). This region is also the hotspot for mutations, making it a good candidate for disease studies (Chen et al., 2023; Saha et al., 2021) as well as genetic diversity studies (Torroni et al., 1993; Pischedda et al., 2017; Duong et al., 2020; Watson et al., 1997; Irwin et al., 2008; Derenko et al., 2007).

Among the 54 ethnic groups in Vietnam, the indigenous Kinh accounts for approximately 90% of the population and speaks the official language of Vietnam (the 2019 Vietnam Population & Housing Census, accessed May 2024, https://www.gso.gov.vn). Therefore, Kinh ethnic group has been included in quite a few maternal molecular anthropology research using Vietnamese samples (Pischedda et al., 2017; Duong et al., 2020, 2018; Irwin et al., 2008; Abdulla et al., 2009; Macholdt et al., 2020; Li et al., 2007; Auton et al., 2015; Brandão et al., 2016; Ngoc et al., 2018). In previous studies, the Kinh was used mostly as a reference point for the Vietnamese people, and it was usually a piece of a bigger puzzle. To the authors' best knowledge, only one study has covered the diversity of the Kinh across Vietnam despite their ubiquitous presence (Pischedda et al., 2017). Therefore, in the present study, we sampled and sequenced the mtDNA of 31 Kinh individuals from Central Vietnam and 33 Kinh individuals from Southern Vietnam. The D-loop region of the study participants was then analyzed with the Northern Kinh from a previous study (Duong et al., 2018).

MATERIALS AND METHODS

Study subjects

Peripheral blood was collected from Kinh individuals, including 31 Central Kinh samples and 33 Southern Kinh samples. Central Kinh and Southern Kinh individuals were recruited in Gialai province and Ho Chi Minh city, respectively. The recruited individuals were unrelated to each other, and their three-generation families must belong to the same ethnic group. All study participants have signed an informed consent document for blood collection. This study has required 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 from whole blood using GeneJET Whole Blood Genomic DNA Purification Mini Kit (ThermoFisher Scientific, USA). mtDNA libraries and their enrichment were described elsewhere (Maricic et al., 2010; Meyer & Kircher, 2010). The mtDNA libraries were sequenced on the Illumina platform (USA) using 150 bp paired-end reads. Then, the reads underwent quality control using FastQC and aligned to mtDNA reference sequence RSRS (Reconstructed Sapiens Reference Sequence) (Behar et al., 2012) with Burrows-Wheeler Alignment (BWA) (Li, 2013). After that, D-loop sequences were extracted and aligned with each other by MAFFT (Katoh & Standley, 2013).

Data analysis

A total of 114 D-loop sequences, including 64 Kinh individuals from our study, and 50 Kinh individuals from a previous study (Duong et al., 2018) were used for haplogroup calling using Haplogrep3 (Schönherr et al., 2023) and PhyloTree mtDNA Build 17.1 (van Oven & Kayser, 2009). Sequences used in

subsequent analyses were excluded missing nucleotides and the following regions: poly-C stretch of hypervariable segment 2 (HVS-II: nucleotide position (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 base-pair (bp) deletion-insertion (np 8,272-8,289); and poly-C stretch of hypervariable segment I (HVS-I; np 16,180-16,195). Fisher's exact test was used to test for significant differences in variant frequencies in the D-loop region of the Kinh individuals from different sampling locations. Pairwise genetic distance between two populations was calculated based on Φ_{ST} distances generated with ARLEQUIN 3.5.2.2 (Excoffier & Lischer, 2010). All p-values smaller than 0.05 were considered significant.

RESULTS

Genomic variants in the D-loop region

For the 64 newly sequenced Kinh samples, a total of 94 unique variants were identified, in which 61 and 68 of these unique variants were present in Central Kinh and Southern Kinh, respectively. When combining our dataset with 50 Northern Kinh samples, 134 unique single nucleotide polymorphisms (SNPs) were discovered in the D-loop region (Fig. 1). There were 34 variants appearing in all three Kinh populations, 25 variants that were found in 2 out of 3 Kinh populations, and 75 variants that were only found in one Kinh population (singletons). Out of the above singletons, only four variants are frequent (A16129G, T310C, A16183C, C16189T), and those variants were only found in Northern Kinh samples. Northern Kinh has the highest mean variant count (15.06 ± 0.32), with the number of variants ranging from 20 to 12 in each individual. On the other hand, Central and Southern Kinh have similar stats for mean variant per individual (11.87 ± 0.31 and 12.33 ± 0.26), maximum (both are at 16), and minimum number of variants per sample (9 and 8, respectively).

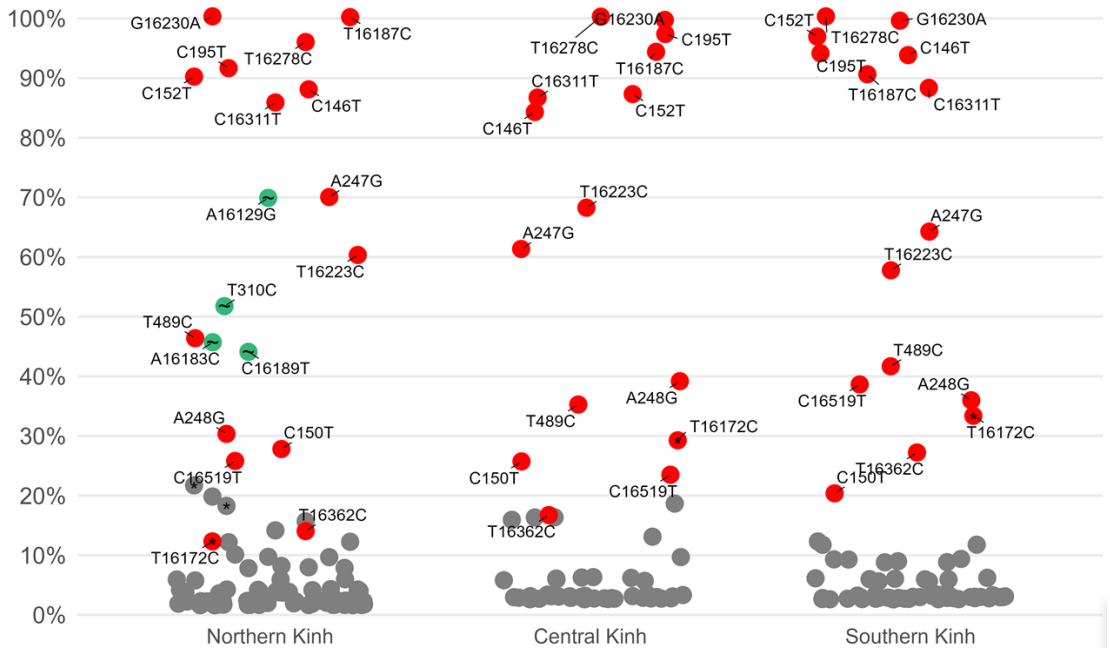

Figure 1. Genomic variants frequency of each Kinh population

Out of 134 unique variants, only G16230A was found in all 114 samples, seconded by T16278C, which was absent in 2/50 Northern Kinh individuals (Fig. 1). Fisher's exact test demonstrated a significant difference in the variant count among the

three Kinh populations for seven variants (C16189T, A16183C, T310C, A16129G, T16304C, A16182C, T16172C) in which 4 of them (C16189T, A16183C, T310C, A16129G) have p-value smaller than 0.0001.

Each point represents the frequency of the corresponding variant. To avoid cluttering the figure with variants' names, we only showed the names of variants with a frequency higher than 25% in at least one population. Each point is color-coded based on the appropriate variant's frequency and distribution, with a variant appearing in all three populations as red, a variant only found in one population and has a frequency higher or equal to 25% as green, and a variant with a frequency less than 25% (> 0%) in all three populations as grey. For the p-value, the "*" symbol indicates that the p-value is greater than 0.0001 but smaller than 0.05, while the "~" symbol shows that the p-value is smaller than 0.0001.

Haplogroup distribution

A total of 39 haplogroups were found in 64 Kinh individuals from Central and Southern Vietnam, among which five lineages were detected in both populations, 15 were in the Central, and 19 were in the Southern Kinh. When combining with the data of Northern Kinh, we identified 60 haplogroups, of which 23 were not present in the Northern group, four (F1a1a, F1a2, B5a, M7b1a1+(16192T)) were shared among three populations, 13 haplogroups that were found in two populations, and 43 haplogroups found in only one population (Fig. 2). The highest haplogroup frequency was only 13%, as the majority of haplogroups were at less than 4%. In addition, most of their haplogroups are not shared, with only 9 out of 60 haplogroups being found in both Northern Kinh and Central Kinh, 11 out of 60 between Northern Kinh and Southern Kinh, and 1 out of 60 between Central Kinh and Southern Kinh.

Each point represents the frequency of the corresponding haplogroup. Each point is color-coded based on haplogroup distribution, with haplogroups being found in all three populations as red, haplogroups being found in two populations as yellow, and haplogroups being found in only one population as green. Three Kinh populations are distinguished by shape, with Southern Kinh as a circle, Northern Kinh as a triangle, and Central Kinh as a cross.

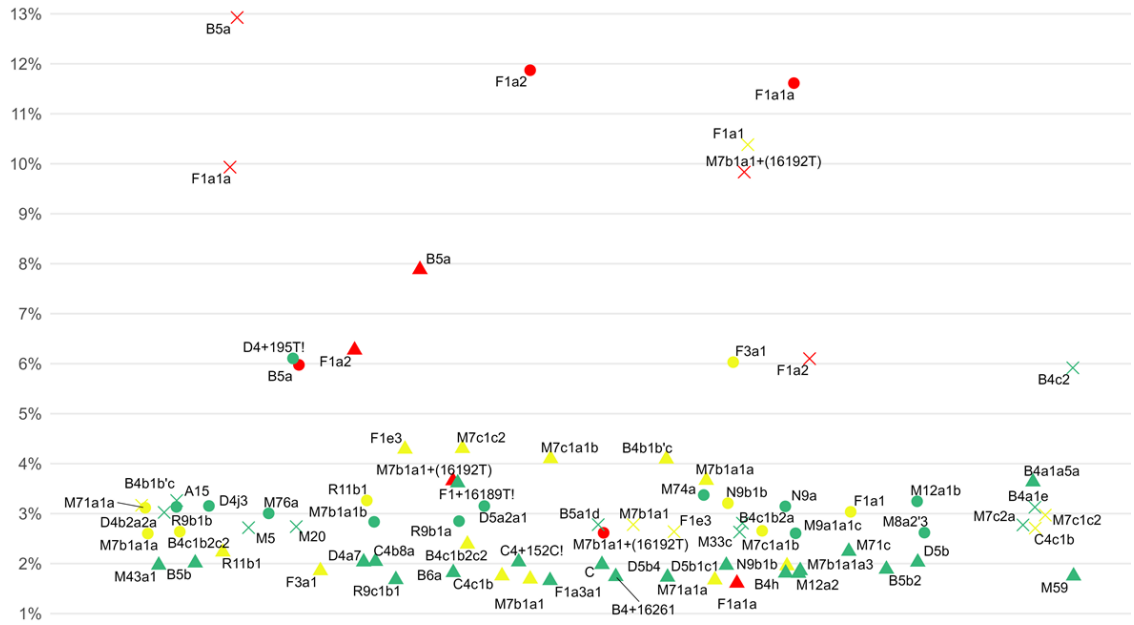

Figure 2. Haplogroup distribution in three Kinh populations

Pairwise genetic distance among the three Kinh populations

To evaluate the genetic relationship among the three Kinh, ARLEQUIN was used to calculate the matrices of pairwise Φ_{ST} distances based on the D-loop sequences (Table 1). The pairwise genetic distances are relatively small with only one significant pairing spotted between Northern and Central Kinh. The distance between Northern Kinh and Central Kinh is rather similar to the corresponding distances between Northern Kinh and Vietnamese populations in the Central Highlands (Giarai & Ede) (Macholdt et al., 2020).

Table 1. Matrices of pairwise Φ_{ST} distances of the three Kinh populations

	Northern Kinh	Central Kinh	Southern Kinh
Northern Kinh	0.00000		
Central Kinh	0.01769 (0.03920 ± 0.0019)	0.00000	
Southern Kinh	0.01219 (0.08296 ± 0.0029)	0.00000 (0.90694 ± 0.0027)	0.00000

Note: the value inside the parentheses is the p-value with a standard deviation.

DISCUSSION

Although the Kinh ethnic group has been studied extensively (Duong et al., 2020, 2018; Irwin et al., 2008; Abdulla et al., 2009; Macholdt et al., 2020; Li et al., 2007; Auton et al., 2015), our study is the first study to analyze the genetic diversity of Kinh group in which the participants were stratified based on sampling locations. Upon evaluating the variants in the D-loop region of 114 samples, 36 unique variants were not present in the Northern Kinh, 1/36 was shared between Central Kinh and Southern Kinh (G16319A), 14 variants unique to Central Kinh, and 21 variants were only found in Southern Kinh. The results showed that the differences in variant frequencies among the three Kinh mostly came from singletons, which result in the highest level of diversity. Second, Fisher's exact test for pairwise Kinh populations revealed that all of the significant values came from Northern Kinh pairings (data not shown) in accordance with pairwise Φ_{ST} distances (Table 1). When comparing to the MITOMAP database of 81,124 mtDNA control region sequences (Lott et al., 2013), these new variants are also present at low frequency with the highest count at 7080 (rank 25th) for G16319A that was found in one Central Kinh individual and two Southern Kinh individuals.

The diversity of D-loop variants in turn leads to the diversity of the haplogroups identified (Fig. 2). The 23 haplogroups that were not present in Northern Kinh samples fall into the B (B4c1b2a, B5a1d, B4c2, B4a1e), M* (M20, M5, M33c, M12a1b, M76a, M74a), D (D4b2a2a, D4j3, D5a2a1, D4+195T!), R9 (F1a1, R9b1a, R9b1b), A (A15), N9 (N9a), M7 (M7c2a, M7b1a1b), M9 (M9a1a1c), and M8 (M8a2'3) lineages though only A and M9 lineages were not found in Northern Kinh individuals. The aforementioned lineages distributed differently across Asia, namely haplogroup A and haplogroup B, can mostly be found in northern and eastern Asia (Derenko et al., 2007), haplogroup D is widespread in Northeast Asia and Central Asia (Pimenoff et al., 2008; Comas et al., 2004), macro-haplogroup M (including M*, M7, M8, M9) is frequent across all Asia (Kutanan et al., 2017; Rajkumar et al., 2005), haplogroup N9a can be found in Central Asia, East Asia, and Southeast Asia (Derenko et al., 2007; Tanaka et al., 2004; Wen et al., 2005) with peak frequency in Malaysia (Duong et al., 2018). Haplogroup R9 is predominant in Southeast/East Asia and less common in South Asia (Woravatin et al., 2023; Jaisamut et al., 2023). Other haplogroups found in Northern Kinh, such as R*, N9, and C, were missing in the Central Kinh, while only haplogroup C was not found in the Southern Kinh. After filtering with the published haplogroup database of the Kinh ethnic group retrieved from multiple studies (Pischedda et al., 2017; Li et al., 2007; Auton et al., 2015;

Brandão et al., 2016; Peng et al., 2011, 2010), there are six haplogroups (D4b2a2a, M33c, D4j3, D5a2a1, D4+195T, and M8a2'3) that are novel for the Kinh, in which haplogroup D5a2a1 is found in Vietnamese Ha Nhi group (Duong et al., 2018) and other several Asia countries such as Thailand (Kutanan et al., 2017), China (Auton et al., 2015), India (Chandrasekar et al., 2009), and Taiwan (Ko et al., 2014) while haplogroup M33c is only found in Laos (Kutanan et al., 2017) and China (Auton et al., 2015).

Differences in genomic variant frequencies, haplogroup composition, and significance of genetic distance suggest that the Northern Kinh is distinguishable from the Central Kinh while it is only marginally diverged from the Southern Kinh. With the gradual migration of the lowland Kinh to the Central Highlands in the 20th century (Tam & Linh, 2022), the gene pool of the Central Kinh could be the result of the founder effect where they admixed with the locals, especially when the migration was done in waves. Furthermore, the fact that Central Kinh was similar to Southern Kinh might indicate that they both admixed with similar minority groups and/or the gene pool of the Central Highlands people was brought to southern plains when new economic opportunities arrived after the unity of Vietnam in 1975. The previous suggestion was based on the geographical closeness between the Central Highlands and the southern plains. However, due to the size of the Kinh population, further analysis covering a bigger population proportion should be carried out before any concrete conclusion is made.

CONCLUSION

In this study, we sequenced the mtDNA of 64 Kinh individuals from Central and Southern Vietnam and analyzed the D-loop region of those sequences in combination with 50 published Northern Kinh mtDNA sequences. This is the first research to study genetic variation within the Kinh ethnic group using sampling location as a variable. Although the three studied populations came from the same ethnic group, their haplogroup makeup and D-loop variants are distinguishable from novel haplogroups and variants spotted in the newly sequenced samples. These results provided the basis for further genetic diversity analysis of the Kinh across Vietnam.

Acknowledgements: This research was supported by the assistance of Prof. Ituro Inoue and the consent from blood donors. This research was funded by a grant from the Ministry of Science and Technology project DTDL/CN.60/19. La Duc Duy was funded by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation, code VINIF.2023. ThS.027.

REFERENCES

Anderson S., Bankier A. T., Barrell B. G., De Bruijn M. H. L., Coulson A. R., Drouin J., Eperon I. C., Nierlich D. P., Roe B. A., Sanger F., Schreier P. H., Smith A. J. H., Staden R., Young I. G., 1981. Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806): 457–465. doi: 10.1038/290457A0

Brown W. M., George M., Wilson A. C., 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 76(4): 1967–1971. https://doi.org/10.1073/PNAS.76.4.1967

Hutchison C. A., Newbold J. E., Potter S. S., Edgell M. H., 1974. Maternal inheritance of mammalian mitochondrial DNA. *Nature*, 251(5475): 536–538. https://doi.org/10.1038/251536A0

Michaels G. S., Hauswirth W. W., Laipis P. J., 1982. Mitochondrial DNA copy number in bovine oocytes and somatic cells. *Developmental Biology*, 94(1): 246–251. https://doi.org/10.1016/0012-1606(82)90088-4

Hagström E., Freyer C., Battersby B. J., Stewart J. B., Larsson N. G., 2014. No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. *Nucleic Acids Research*, 42(2): 1111–1116. https://doi.org/10.1093/NAR/GKT969

Merriwether D. A., Clark A. G., Ballinger S. W., Schurr T. G., Soodyall H., Jenkins T., Sherry S. T., Wallace D. C., 1991. The structure of human mitochondrial DNA variation. *Journal of Molecular Evolution*, 33(6): 543–555. https://doi.org/10.1007/BF02102807

Pakendorf B., Stoneking M., 2005. Mitochondrial DNA and human evolution. *Annual Review of Genomics and Human Genetics*, 6: 165–183. https://doi.org/10.1146/ANNUREV.GENOM.6.080604.162249

Taanman J. W., 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1410(2): 103–123. https://doi.org/10.1016/S0005-2728(98)00161-3

Chen Y., Wu W. J., Xing L. W., Zhang X. J., Wang J., Xia X. Y., Zhao R., Zhao R., 2023. Investigating the role of mitochondrial DNA D-loop variants, haplotypes, and copy number in polycystic ovary syndrome: implications for clinical phenotypes in the Chinese population. *Frontiers in Endocrinology*, 14: 1206995. https://doi.org/10.3389/FENDO.2023.1206995/BIBTEX

Saha S. K., Saba A. Al, Hasib M., Rimon R. Al, Hasan I., Alam M. S., Mahmud I., Nabi A. H. M. N., 2021. Evaluation of D-loop hypervariable region I variations, haplogroups and copy number of mitochondrial DNA in Bangladeshi population with type 2 diabetes. *Heliyon*, 7(7): e07573. https://doi.org/10.1016/J.HELIYON.2021.E07573

Torroni A., Schurr T. G., Cabell M. F., Brown M. D., Neel J. V., Larsen M., Smith D. G., Vullo C. M., Wallace D. C., 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *American Journal of Human Genetics*, 53(3): 563.

Pischedda S., Barral-Arca R., Gómez-Carballa A., Pardo-Seco J., Catelli M. L., Álvarez-Iglesias V., Cárdenas J. M., Nguyen N. D., Ha H. H., Le A. T., Martinón-Torres F., Vullo C., Salas A., 2017. Phylogeographic and genome-wide investigations of Vietnam ethnic groups reveal signatures of complex historical demographic movements. *Scientific Reports*, 7(1): 1–15. https://doi.org/10.1038/s41598-017-12813-6

Duong N.T., Phong N. Van, Ngoc N.T., Hai N. Van, 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): 231–238. https://doi.org/10.15625/1811-4989/18/2/15136

Watson E., Forster P., Richards M., Bandelt H. J., 1997. Mitochondrial footprints of human expansions in Africa. *American Journal of Human Genetics*, 61(3): 691. https://doi.org/10.1086/515503

Irwin J. A., Saunier J. L., Strouss K. M., Diegoli T. M., Sturk K. A., O'Callaghan J. E., Paintner C. D., Hohoff C., Brinkmann B., Parsons T. J., 2008. Mitochondrial control region sequences from a Vietnamese population sample. *International Journal of Legal Medicine*, 122(3): 257–259. https://doi.org/10.1007/S00414-007-0205-3

Derenko M., Malyarchuk B., Grzybowski T., Denisova G., Dambueva I., Perkova M., Dorzhu C., Luzina F., Hong K. L., Vanecek T., Villems R., Zakharov I., 2007. Phylogeographic Analysis of Mitochondrial DNA in Northern Asian Populations. *American Journal of Human Genetics*, 81(5): 1025. https://doi.org/10.1086/522933

Abdulla M. A., Ahmed I., Assawamakin A., Bhak J., Brahmachari S. K., Calacal G. C., Chaurasia A., Chen C. H., Chen J., Chen Y. T., Chu J., Cutiongco-de La Paz E. M. C., De Ungria M. C. A., Delfin F. C., Edo J., Fuchareon S., Ghang H., Gojobori T., Han J., Ho S. F., Hoh B. P., Huang W., Inoko H., Jha P., Jinam T. A., Jin L., Jung J., Kangwanpong D., Kampuansai J., Kennedy G.C., Khurana P., Kim H.L., Kim K., Kim S., Kim W.Y., Kimm K., Kimura

R., Koike T., Kulawonganunchai S., Kumar V., Lai P. S., Lee J. Y., Lee S., Liu E. T., Majumder P. P., Mandapati K. K., Marzuki S., Mitchell W., Mukerji M., Naritomi K., Ngamphiw C., Niikawa N., Nishida N., Oh B., Oh S., Ohashi J., Oka A., Ong R., Padilla C. D., Palittapongarnpim P., Perdigon H. B., Phipps M. E., Png E., Sakaki Y., Salvador J. M., Sandraling Y., Scaria V., Seielstad M., Sidek M. R., Sinha A., Srikummool M., Sudoyo H., Sugano S., Suryadi H., Suzuki Y., Tabbada K. A., Tan A., Tokunaga K., Tongsima S., Villamor L. P., Wang E., Wang Y., Wang H., Wu J. Y., Xiao H., Xu S., Yang J. O., Shugart Y. Y., Yoo H. S., Yuan W., Zhao G., Zilfalil B. A., 2009. Mapping human genetic diversity in Asia. *Science (New York, N.Y.)*, 326(5959): 1541–1545. https://doi.org/10.1126/SCIENCE.1177074

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): 1–13. https://doi.org/10.1038/s41598-018-29989-0.

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

Li H., Cai X., Winograd-Cort E. R., Wen B., Cheng X., Qin Z., Liu W., Liu Y., Pan S., Qian J., Tan C. C., Jin L., 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. *American Journal of Physical Anthropology*, 134(4): 481–488. https://doi.org/10.1002/AJPA.20690

Auton A., Abecasis G. R., Altshuler D. M., Durbin R. M., Bentley D. R., Chakravarti A., Clark A. G., Donnelly P., Eichler E. E., Flicek P., Gabriel S. B., Gibbs R. A., Green E. D., Hurles M. E., Knoppers B. M., Korbel J. O., Lander E. S., Lee C., Lehrach H., Mardis E. R., Marth G. T., McVean G. A., Nickerson D. A., Schmidt J. P., Sherry S. T., Wang J., Wilson R. K., Boerwinkle E., Doddapaneni H., Han Y., Korchina V., Kovar C., Lee S., Muzny D., Reid J. G., Zhu Y., Chang Y., Feng Q., Fang X., Guo X., Jian M., Jiang H., Jin X., Lan T., Li G., Li J., Li Y., Liu S., Liu X., Lu Y., Ma X., Tang M., Wang B., Wang G., Wu H., Wu R., Xu X., Yin Y., Zhang D., Zhang W., Zhao J., Zhao M., Zheng X., Gupta N., Gharani N., Toji L. H., Gerry N. P., Resch A. M., Barker J., Clarke L., Gil L., Hunt S. E., Kelman G., Kulesha E., Leinonen R., McLaren W. M., Radhakrishnan R., Roa A., Smirnov D., Smith R. E., Streeter I., Thormann A., Toneva I., Vaughan B., Zheng-Bradley X., Grocock R., Humphray S., James T., Kingsbury Z., Sudbrak R., Albrecht M. W., Amstislavskiy V. S., Borodina T. A., Lienhard M., Mertes F., Sultan M., Timmermann B., Yaspo M. L., Fulton L., Ananiev V., Belaia Z., Beloslyudtsev D., Bouk N., Chen C., Church D., Cohen R., Cook C., Garner J., Hefferon T., Kimelman M., Liu C., Lopez J., Meric P., O'Sullivan C., Ostapchuk Y., Phan L., Ponomarov S., Schneider V., Shekhtman E., Sirotkin K., Slotta D., Zhang H., Balasubramaniam S., Burton J., Danecek P., Keane T. M., Kolb-Kokocinski A., McCarthy S., Stalker J., Quail M., Davies C. J., Gollub J., Webster T., Wong B., Zhan Y., Campbell C. L., Kong Y., Marcketta A., Yu F., Antunes L., Bainbridge M., Sabo A., Huang Z., Coin L. J. M., Fang L., Li Q., Li Z., Lin H., Liu B., Luo R., Shao H., Xie Y., Ye C., Yu C., Zhang F., Zheng H., Zhu H., Alkan C., Dal E., Kahveci F., Garrison E. P., Kural D., Lee W. P., Leong W. F., Stromberg M., Ward A. N., Wu J., Zhang M., Daly M. J., DePristo M. A., Handsaker R. E., Banks E., Bhatia G., Del Angel G., Genovese G.,

Li H., Kashin S., McCarroll S. A., Nemesh J. C., Poplin R. E., Yoon S. C., Lihm J., Makarov V., Gottipati S., Keinan A., Rodriguez-Flores J. L., Rausch T., Fritz M. H., Stütz A. M., Beal K., Datta A., Herrero J., Ritchie G. R. S., Zerbino D., Sabeti P.C., Shlyakhter I., Schaffner S. F., Vitti J., Cooper D. N., Ball E. V., Stenson P. D., Barnes B., Bauer M., Cheetham R. K., Cox A., Eberle M., Kahn S., Murray L., Peden J., Shaw R., Kenny E. E., Batzer M. A., Konkel M. K., Walker J. A., MacArthur D. G., Lek M., Herwig R., Ding L., Koboldt D. C., Larson D., Ye K., Gravel S., Swaroop A., Chew E., Lappalainen T., Erlich Y., Gymrek M., Willems T. F., Simpson J. T., Shriver M. D., Rosenfeld J. A., Bustamante C. D., Montgomery S. B., De La Vega F. M., Byrnes J. K., Carroll A. W., DeGorter M. K., Lacroute P., Maples B. K., Martin A. R., Moreno-Estrada A., Shringarpure S. S., Zakharia F., Halperin E., Baran Y., Cerveira E., Hwang J., Malhotra A., Plewczynski D., Radew K., Romanovitch M., Zhang C., Hyland F. C. L., Craig D. W., Christoforides A., Homer N., Izatt T., Kurdoglu A. A., Sinari S. A., Squire K., Xiao C., Sebat J., Antaki D., Gujral M., Noor A., Ye K., Burchard E. G., Hernandez R. D., Gignoux C. R., Haussler D., Katzman S. J., Kent W. J., Howie B., Ruiz-Linares A., Dermitzakis E. T., Devine S. E., Kang H. M., Kidd J. M., Blackwell T., Caron S., Chen W., Emery S., Fritsche L., Fuchsberger C., Jun G., Li B., Lyons R., Scheller C., Sidore C., Song S., Sliwerska E., Taliun D., Tan A., Welch R., Wing M. K., Zhan X., Awadalla P., Hodgkinson A., Li Y., Shi X., Quitadamo A., Lunter G., Marchini J. L., Myers S., Churchhouse C., Delaneau O., Gupta-Hinch A., Kretzschmar W., Iqbal Z., Mathieson I., Menelaou A., Rimmer A., Xifara D. K., Oleksyk T. K., Fu Y., Liu X., Xiong M., Jorde L., Witherspoon D., Xing J., Browning B.L., Browning S. R., Hormozdiari F., Sudmant P. H., Khurana E., Tyler-Smith C., Albers C. A., Ayub Q., Chen Y., Colonna V., Jostins L., Walter K., Xue Y., Gerstein M. B., Abyzov A., Balasubramanian S., Chen J., Clarke D., Fu Y., Harmanci A. O., Jin M., Lee D., Liu J., Mu X. J., Zhang J., Zhang Y., Hartl C., Shakir K., Degenhardt J., Meiers S., Raeder B., Casale F. P., Stegle O., Lameijer E. W., Hall I., Bafna V., Michaelson J., Gardner E. J., Mills R. E., Dayama G., Chen K., Fan X., Chong Z., Chen T., Chaisson M. J., Huddleston J., Malig M., Nelson B. J., Parrish N. F., Blackburne B., Lindsay S. J., Ning Z., Zhang Y., Lam H., Sisu C., Challis D., Evani U. S., Lu J., Nagaswamy U., Yu J., Li W., Habegger L., Yu H., Cunningham F., Dunham I., Lage K., Jespersen J. B., Horn H., Kim D., Desalle R., Narechania A., Sayres M. A. W., Mendez F. L., Poznik G. D., Underhill P. A., Mittelman D., Banerjee R., Cerezo M., Fitzgerald T. W., Louzada S., Massaia A., Yang F., Kalra D., Hale W., Dan X., Barnes K.C., Beiswanger C., Cai H., Cao H., Henn B., Jones D., Kaye J. S., Kent A., Kerasidou A., Mathias R., Ossorio P. N., Parker M., Rotimi C. N., Royal C. D., Sandoval K., Su Y., Tian Z., Tishkoff S., Via M., Wang Y., Yang H., Yang L., Zhu J., Bodmer W., Bedoya G., Cai Z., Gao Y., Chu J., Peltonen L., Garcia-Montero A., Orfao A., Dutil J., Martinez-Cruzado J. C., Mathias R. A., Hennis A., Watson H., McKenzie C., Qadri F., LaRocque R., Deng X., Asogun D., Folarin O., Happi C., Omoniwa O., Stremlau M., Tariyal R., Jallow M., Joof F. S., Corrah T., Rockett K., Kwiatkowski D., Kooner J., Hien T. T., Dunstan S. J., ThuyHang N., Fonnie R., Garry R., Kanneh L., Moses L., Schieffelin J., Grant D.S., Gallo C., Poletti G., Saleheen D., Rasheed A., Brooks L. D., Felsenfeld A. L., McEwen J. E., Vaydylevich Y., Duncanson A., Dunn M., Schloss J. A., 2015. A global reference for human genetic variation. *Nature*, 526(7571): 68–74. https://doi.org/10.1038/nature15393

Brandão A., Eng K. K., Rito T., Cavadas B., Bulbeck D., Gandini F., Pala M., Mormina M., Hudson B., White J., Ko T. M., Saidin M., Zafarina Z., Oppenheimer

S., Richards M. B., Pereira L., Soares P., 2016. Quantifying the legacy of the Chinese Neolithic on the maternal genetic heritage of Taiwan and Island Southeast Asia. *Human Genetics*, 135(4): 363. https://doi.org/10.1007/S00439-016-1640-3

Ngoc N. T., Trang N. B., Huy N. Q., Ton N. D., Duong N. T., 2018. Single nucleotide polymorphisms in the d-loop region of the mitochondrial genomes of individuals from two ethnic groups kinh and mang of austro-asiatic language family. *Vietnam Journal of Biotechnology*, 16(2): 231–240. doi: 10.15625/1811-4989/16/2/13432

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

Meyer M., Kircher M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010(6): pdb.prot5448. https://doi.org/10.1101/PDB.PROT5448

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. *American Journal of Human Genetics*, 90(4): 675. https://doi.org/10.1016/J.AJHG.2012.03.002

Li H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

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. https://doi.org/10.1093/MOLBEV/MST010

Schönherr S., Weissensteiner H., Kronenberg F., Forer L., 2023. Haplogrep 3 - an interactive haplogroup classification and analysis platform. *Nucleic Acids Research*, 51(W1): W263–W268. https://doi.org/10.1093/NAR/GKAD284

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

Excoffier L., 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: 564–567. *Evolutionary Bioinformatics Online*.

Lott M. T., Leipzig J. N., Derbeneva O., Michael Xie H., Chalkia D., Sarmady M., Procaccio V., Wallace D. C., 2013. mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Current Protocols in Bioinformatics*, 44(1): 1–23. https://doi.org/10.1002/0471250953.BI0123S44

Pimenoff V. N., Comas D., Palo J. U., Vershubsky G., Kozlov A., Sajantila A., 2008. Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *European Journal of Human Genetics : EJHG*, 16(10): 1254–1264. doi: 10.1038/EJHG.2008.101

Comas D., Plaza S., Wells R. S., Yuldaseva N., Lao O., Calafell F., Bertranpetit J., 2004. Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. *European Journal of Human Genetics : EJHG*, 12(6): 495–504. https://doi.org/10.1038/SJ.EJHG.5201160

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

Rajkumar R., Banerjee J., Gunturi H. B., Trivedi R., Kashyap V. K., 2005. Phylogeny and antiquity of M macrohaplogroup inferred from complete mt DNA sequence of Indian specific lineages. *BMC Evolutionary Biology*, 5: 26. https://doi.org/10.1186/1471-2148-5-26

Tanaka M., Cabrera V. M., González A. M., Larruga J. M., Takeyasu T., Fuku N., Guo L. J., Hirose R., Fujita Y., Kurata M., Shinoda K. I., Umetsu K., Yamada Y., Oshida Y., Sato Y., Hattori N., Mizuno Y., Arai Y., Hirose N., Ohta S., Ogawa O., Tanaka Y., Kawamori R., Shamoto-Nagai M., Maruyama W., Shimokata H., Suzuki R., Shimodaira H., 2004. Mitochondrial Genome Variation in Eastern Asia and the Peopling of Japan. *Genome Research*, 14(10a): 1832–50. https://doi.org/10.1101/GR.2286304

Wen B., Li H., Gao S., Mao X., Gao Y., Li F., Zhang F., He Y., Dong Y., Zhang Y., Huang W., Jin J., Xiao C., Lu D., Chakraborty R., Su B., Deka R., Jin L., 2005. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Molecular Biology and Evolution*, 22(3): 725–734. https://doi.org/10.1093/MOLBEV/MSI055

Woravatin W., Stoneking M., 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.

Jaisamut K., Pitiwararom R., Sukawutthiya P., Sathirapatya T., Noh H., Worrapitirungsi W., Vongpaisarnsin K., 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

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(46). https://doi.org/10.1186/1471-2148-11-46

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

Chandrasekar A., Kumar S., Sreenath J., Sarkar B. N., Urade B. P., Mallick S., Bandopadhyay S.S., Barua P., Barik S. S., Basu D., Kiran U., Gangopadhyay P., Sahani R., Prasad B. V. R., Gangopadhyay S., Lakshmi G. R., Ravuri R. R., Padmaja K., Venugopal P. N., Sharma M. B., Rao V. R., 2009. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PloS One*, 4(10): e7447. https://doi.org/10.1371/JOURNAL.PONE.0007447

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. *American Journal of Human Genetics*, 94(3): 426–36. https://doi.org/10.1016/J.AJHG.2014.02.003

Tam H. T., Linh D. P., 2022. How minor immigrants became the dominants: the case of the Kinh people migrating to the Central Highlands, Vietnam in the twentieth century. *Social Identities*, 28(5): 608–627. https://doi.org/10.1080/13504630.2022.2131512