Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

High genetic diversity of circumsporozoite surface protein in Vietnamese Plasmodium malariae isolates

Tuan Cuong Vo
Gyeongsang National University
Huong Giang Le
Gyeongsang National University
Chau Van Khanh
Institute of Malariology, Parasitology, and Entomology Quy Nhon
Đăng Thùy Dương Nguyễn
Gyeongsang National University
Thu Hằng Nguyễn
Gyeongsang National University
Thi Minh Trinh Nguyen
Institute of Malariology, Parasitology, and Entomology Quy Nhon
Thi Lien Hanh Nguyen
Institute of Malariology, Parasitology, and Entomology Quy Nhon
Hong Quang Huynh
Institute of Malariology, Parasitology, and Entomology Quy Nhon
Byoung-Kuk Na
bkna@gnu.ac.kr

Gyeongsang National University

Research Article

Keywords: Plasmodium malariae, circumsporozoite surface protein, Vietnam, genetic diversity, natural selection

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10685787/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

High genetic diversity of circumsporozoite surface protein in Vietnamese *Plasmodium malariae* isolates

Tuấn Cường Võ[1], Hương Giang Lê[1], Chau Van Khanh[2], Đăng Thùy Dương Nguyễn[1,2], Thu Hằng Nguyễn[1], Nguyen Thi Minh Trinh[2], Nguyen Thi Lien Hanh[2], Huynh Hong Quang[2]*, Byoung-Kuk Na[1]*

[1]Department of Parasitology and Tropical Medicine, Department of Convergence Medical Science, and Institute of Medical Science, Gyeongsang National University College of Medicine, Jinju 52727, Republic of Korea

[2]Institute of Malariology, Parasitology, and Entomology Quy Nhon, Gia Lai Province, Vietnam

*Correspondence: Huynh Hong Quang (huynhquangimpe@yahoo.com)
Byoung-Kuk Na (E-mail: bkna@gnu.ac.kr)

E-mails:
Tuấn Cường Võ: vtcuong241@gmail.com
Hương Giang Lê: gianglee291994@gmail.com
Chau Van Khanh: cvkhanh01@gmail.com
Đăng Thùy Dương Nguyễn: duongnguyenhcm1103@gmail.com
Thu Hằng Nguyễn: thuhang3399@gmail.com
Nguyen Thi Minh Trinh: nguyenminhtrinh1983@gmail.com
Nguyen Thi Lien Hanh: lienhanhnguyen84@gmail.com

Huynh Hong Quang: huynhquangimpe@yahoo.com

Byoung-Kuk Na: bkna@gnu.ac.kr

Abstract

Background: *Plasmodium malariae* is an apicomplexan parasite that causes quartan malaria in humans but remains understudied despite its global distribution. Furthermore, growing evidence suggests that *P. malariae* infections are more prevalent than previously recognized. In this study, the genetic diversity of *P. malariae* circumsporozoite surface protein gene (*pmcsp*) in recently emerged Vietnamese *P. malariae* isolates was investigated.

Methods: DNA samples of 116 *P. malariae*-infected patients in Khanh Hoa Province, Vietnam, collected during 2023–2024 were used. Full-length *pmcsp* was amplified and sequenced, and the genetic diversity and natural selection profiles of Vietnamese and global *pmcsp* were analyzed.

Results: The two non-repeat terminal regions of Vietnamese *pmcsp* exhibited low genetic diversity, whereas the central repeat region (CRR) showed high genetic diversity due to variation in the number, composition, and arrangement of peptide repeat motif (PRM) species. Global *pmcsp* CRR also displayed extensive polymorphism in repeat length and PRM composition, though Vietnamese *pmcsp* CRR exhibited the greatest diversity. Although modest genetic differences were observed among *pmcsp* populations from Asia and Africa, no strong geographical bias was observed.

Conclusion: Vietnamese *pmcsp* exhibits substantial genetic diversity despite the limited spatial and temporal range of isolates. Global *pmcsp* diversity is greater than previously recognized. Given the limited genetic information currently available for the global *P. malariae* population, studies with larger numbers of global *P. malariae* isolates from different origins are required to better understand the genetic nature and evolutionary aspects of the gene.

Keywords: *Plasmodium malariae*, circumsporozoite surface protein, Vietnam, genetic diversity, natural selection

Introduction

Plasmodium malariae is an apicomplexan parasite that causes quartan malaria in humans. Although the global incidence of *P. malariae* infection is substantially lower than that of *P. falciparum* and *P. vivax*, *P. malariae* is found globally, including in Africa, South America, Asia, and the Western Pacific [1,2]. The infection is often asymptomatic and benign, which has resulted in this parasite being relatively neglected. However, it has been associated with chronic nephrotic syndrome, which can persist for years and result in death [3–5]. Nonetheless, *P. malariae* is often underdiagnosed due to low parasitemia and a lack of sensitivity of rapid diagnostic tests (RDTs). However, the development of novel molecular diagnostic techniques has increased interest in this parasite over the decades because they have shown that infections are more prevalent than previously understood [1]. While *P. malariae* is less common than *P. falciparum* and *P. vivax*, unlike these two parasites, it is not uniformly in decline and can cause significant regional outbreaks [1,6]. In addition, mixed infections with other *Plasmodium* species, particularly *P. falciparum*, are also common and present complex clinical, diagnostic, and therapeutic challenges [7]. Thus, *P. malariae* represents a persistent challenge to those attempting to eliminate malaria.

The population structure of malaria parasites is affected by factors such as transmission intensity, geographic isolation, human migration, and selective pressure by control interventions [8]. Understanding the genetic diversity and population structure of the parasites is essential for monitoring their local and regional transmission patterns and the effect of public health interventions, and informs the design of malaria control and elimination strategies. For these reasons, global molecular studies have been conducted to investigate the population structures of *P. falciparum* and *P. vivax*, particularly to identify potential targets for vaccine development and

93 antimalarial drug resistance genes [9–15]. However, these studies largely excluded *P. malariae*,
94 mainly due to its low endemicity and clinical sampling difficulties. A few studies have addressed
95 the genetic diversity of *P. malariae* genes, including apical membrane antigen-1 (*pmama-1*),
96 merozoite surface protein-1 (*pmmsp-1*), circumsporozoite surface protein (*pmcsp*), and
97 thrombospondin-related anonymous protein (*pmtrap*), in limited numbers of *P. malariae* isolates
98 [16–22]. Notably, these studies suggested substantial genetic diversity and regional
99 differentiation, and thus highlighted the need to investigate *P. malariae* genes to understand the
100 genetic nature of the global population of this parasite.
101 Circumsporozoite surface protein (CSP) is abundantly expressed on the surface of the
102 sporozoites of *Plasmodium* parasites and plays a pivotal role in sporozoite migration and
103 hepatocyte invasion [23], suggesting its value as a vaccine. *P. falciparum* CSP (PfCSP) was
104 subjected to a major component in the most advanced malaria vaccines, RTS,S/AS01 (RTS,S)
105 and R21/Matrix-M (R21), endorsed by the World Health Organization (WHO) [24,25]. *P.*
106 *malariae* CSP (PmCSP) shares structural features, consisting of three different regions, with
107 PfCSP and *P. vivax* CSP (PvCSP). These regions include two relatively well-conserved terminal
108 regions and a highly polymorphic central repeat region (CRR) [26]. Although studies have been
109 performed to investigate the genetic diversity of *pmcsp* in *P. malariae* isolates [18,19,22,27,28],
110 they have been limited by small sample numbers or have not provided sufficient information on
111 the genetic diversity and evolutionary characteristics of *pmcsp* in the global *P. malariae*
112 population.
113 In this study, the genetic diversity of *pmcsp* in recently emerged *P. malariae* isolates in Khanh
114 Hoa Province, Vietnam, was investigated to understand the genetic structure of this gene.
115 Comparative analysis was also conducted using the 169 global *pmcsp* sequences known to date

to probe the genetic nature and evolutionary aspects of the gene in the global *P. malariae* population.

Materials and methods

Sample collection

Genomic DNA samples were obtained from 116 dried blood spots of *Plasmodium malariae*-infected patients in Khanh Hoa Province, Vietnam, in 2023–2024 [21]. *P. malariae* mono-infection was confirmed by microscopic examination of thin and thick blood smears and validated using a species-specific polymerase chain reaction (PCR) for 18S ribosomal RNA [29]. Verbal informed consent was obtained from all participants prior to blood collection. The study protocol was reviewed and approved by the Ethics Committee of the Institute of Malariology, Parasitology and Entomology, Quy Nhon, Vietnam (Institutional Review Board approval number 27/VSR-NCDT).

Amplification of *pmcsp*

Full-length *pmcsp* was amplified using nested PCR. The two pairs of primers were designed based on the nucleotide sequences of *pmcsp* from a reference strain (PmUG01; Genbank ID XM_029004694.1). The sequences of the primers and thermal conditions used for amplification are shown in Supplementary Table S1. All amplifications were performed using Ex *Taq* DNA polymerase (Takara, Otsu, Japan), which possesses proofreading activity, to minimize inadequate nucleotide misincorporation. Amplicons were visualized on 1.5% agarose gel, and each amplicon was purified from the gel and cloned into T&A vector (Real Biotech Corporation, Banqiao City, Taiwan). When multiple-sized amplicons were present in a sample, implying potential multiclonal infections, each amplicon was cloned separately. Each ligation mixture was transformed into *Escherichia coli* DH5α, and positive clones harboring cloned *pmcsp* were

selected using colony PCR. The nucleotide sequences of cloned *pmcsp* were analyzed by automated Sanger sequencing using M13 forward and reverse primers (Genotech, Daejeon, Korea). Plasmids from at least three different clones of each transformant were sequenced to verify sequence fidelity. Nucleotide sequences were deposited in the GenBank database under accession numbers PZ397962–PZ398135.

Genetic diversity analysis of Vietnamese and global *pmcsp*

The nucleotide and amino acid sequences of Vietnamese *pmcsp* were annotated using the Editseq and Seqman programs in DNASTAR software (DNASTAR, Madison, WI, USA). For annotation, the *pmcsp* sequence (Genbank ID: XM_029004694.1) was used as a reference. Multiplicity of infection (MOI) was determined by tallying alleles identified in each isolate. Segregating sites (S), haplotypes (H), haplotype diversity (Hd), nucleotide diversity (π), and average pairwise nucleotide differences (K) within a population were analyzed using DnaSP version 5.10.00 [30]. Synonymous (dS) and non-synonymous (dN) substitution rates were calculated using Nei and Gojobori's method with the Jukes-Cantor correction, employing 1,000 bootstrap replications in MEGA6 [31]. A Z-test was used to assess significance—defined as a P value of < 0.05. Neutrality metrics, such as Tajima's D, Fu and Li's D, and Fu and Li's F, were calculated using DnaSP version 5.10.00 [30]. Results were regarded significant for P values $<$ 0.05. The haplotype network for the *pmcsp* population was generated using PopART 1.7 software (http://popart.otago.ac.nz) employing the Median-Joining method. Bayesian analysis to determine the lineage of Vietnamese *pmcsp* was conducted using the STRUCTURE program [32]. The quantity of clusters (K), varying from 1 to 10, was assessed with a burn-in period of 10,000 steps and 20,000 Markov chain Monte Carlo iterations. The best K value was ascertained using StructureSelector [32,33]. The genetic diversity and natural selection profiles of global

pmcsp populations were examined by incorporating 169 *pmcsp* sequences from Thailand, Myanmar, Lao PDR, Bangladesh, Malaysia, China, Kenya, Cameroon, Uganda, and Côte d'Ivoire (Supplementary Table S2). Countries with fewer than 5 *pmcsp* sequences were not analyzed individually because of limited statistical power, but were included in the intracontinental analysis.

Results

Amplification of Vietnamese *pmcsp*

Amplification of *pmcsp* in Vietnamese *P. malariae* isolates exhibited amplicons varying from 663 to 1371 base pairs. Of the 116 samples, 47 showed differently sized multiple amplicons—40 had 2, 5 had 3, and 2 had 5 differently sized amplicons. All amplicons were successfully cloned and sequenced. A total of 174 different *pmcsp* sequences were obtained from the 116 specimens—an estimated MOI of 1.5.

Polymorphic patterns of the N- and T-terminal regions in Vietnamese *pmcsp*

Amino acid changes in the N-terminal (N-T) and C-terminal (C-T) of Vietnamese *pmcsp* were analyzed. The N-T was generally well-conserved in the gene. Compared to the reference sequence (PmUG01; XM_029004694.1), only two amino acid substitutions, Q58R and N96S, were observed, each occurring in two sequences, while 170 sequences shared the amino acid sequence of the reference. This resulted in 3 distinct haplotypes (H1, H2, and H3), among which H1 was dominant (Fig. 1). The enzyme-cutting motif ^{97}KLKQP^{101} [34] was well conserved across all sequences; C-T also exhibited few amino acid changes (Fig. 1). The most remarkable substitution was E295G, which was identified in all Vietnamese *pmcsp* sequences. Three amino acid changes, S301P, L313S, and I335M, were found in 2, 5, and 2 sequences, respectively, but

these were not detected concomitantly. Based on these amino acid substitution patterns, the C-T of Vietnamese *pmcsp* clustered into 4 different haplotypes (H1–H4), of which H1 was prevalent.

Polymorphic patterns of the N- and C-T regions in the global *pmcsp*

Like Vietnamese *pmcsp*, the global *pmcsp* population also exhibited highly conserved amino acid patterns in the N-T and C-T. Amino acid substitutions were identified in *pmcsp* from Myanmar, Thailand, Lao PDR, Kenya, and Cameroon, but only minor polymorphisms were observed in important regions, conserved region I (CR I; ^{93}AVENKLKQ^{100}) in the N-T and conserved region II (CR II; ^{303}GPSEEHIKNYLESIRNS^{319}), Th2R (^{320}ITEEWSPCSVTCG^{332}), and Th3R (^{339}RKVDAKNKKPAELVLSDLE^{357}) in the C-T (Fig. 2). No clear geographical lineage of these amino acid polymorphisms was identified among the sequences.

Genetic diversity and natural selection of N-T and C-T in Vietnamese and global *pmcsp*

The genetic diversity and natural selection trends of the N-T and C-T in global *pmcsp* were analyzed (Fig. 3; Table 1). Both the N-T and C-T of Vietnamese *pmcsp* had low Hd and π values; Hd (N-T: 0.045 ± 0.022, C-T: 0.164 ± 0.038) and π (N-T: 0.0002 ± 0.0001, C-T: 0.0007 ± 0.0002). Global *pmcsp* exhibited various Hd and π values. Myanmar *pmcsp* exhibited the highest values of Hd at both N-T (0.913 ± 0.022) and C-T (0.805 ± 0.059). Thai *pmcsp* revealed the highest π at N-T (0.0126 ± 0.0027), while Myanmar *pmcsp* had the highest π at C-T (0.0196 ± 0.0042). African *pmcsp* had non-significantly higher Hd and π values than Asian *pmcsp*.

Global *pmcsp* exhibited similar natural selection patterns. Tajima's D values of global *pmcsp* N-T ranged from –1.887 (Malaysia) to 0.064 (Myanmar). Global *pmcsp* C-T also exhibited similar Tajima's D profiles, ranging from –1.983 (Cameroon) to –0.697 (Thailand). Tajima's D values for Asian *pmcsp* and African *pmcsp* also showed negative values at N-T and C-T; Asian *pmcsp*, N-T (–1.963, $P < 0.05$) and C-T (–2.416, $P < 0.01$), and African *pmcsp*, N-T (–1.846, $P < 0.05$)

and C-T (–2.151, $P < 0.05$). These findings suggest that purifying selection or selective sweep forces might have acted in the N-T and C-T of global *pmcsp* populations. Sliding window plots of π and Tajima's D provided more detailed information on these values across the N-T and C-T (Fig. 4). The N-T showed a low level of nucleotide diversity, except for the latter part, including CR I, which harbored ^{97}KLKQP^{101}, whereas the C-T exhibited a greater nucleotide diversity than the N-T, although no synchronized pattern was detected in the global *pmcsp* C-T. Negative Tajima's D values were detected in all global populations across the N-T and C-T, except for Myanmar and Thailand in the latter part of the N-T and C-T.

Haplotype network and population structure of the N-T and C-T of global *pmcsp*

Haplotype network analysis for each region was conducted to further investigate the genetic characteristics of the N-T and C-T in the global *pmcsp* population (Fig. 5). The N-T haplotype network exhibited 42 distinct haplotypes (H1–H42), among which H1 was the most dominant, shared by *pmcsp* sequences from a major proportion of Vietnam (170/174) and eight other countries, including PmUG01. The second major haplotype was H4, represented by sequences from five countries: Thailand, Myanmar, Lao PDR, Bangladesh, and Kenya. No geographical clustering was found in the N-T. On the other hand, the haplotype network of *pmcsp* C-T exhibited a geographically differentiated pattern. A total of 39 different haplotypes (H1–H39) were identified, of which H1 and H2 were the most dominant. H1, the largest haplotype, was predominantly represented by sequences from Asian countries, while H2 was predominantly represented by sequences from African countries and PmUG01. Interestingly, most Malaysian *pmcsp* C-T sequences clustered into the unique haplotype H24, although a few grouped into H2. Structure analysis revealed that global *pmcsp* N-T and C-T had two different origins, K1 and K2 ($\Delta K = 9.21$ for N-T and $\Delta K = 47.39$ for C-T), respectively (Fig. 5). The Vietnamese and

Malaysian *pmcsp* N-Ts were highly uniform, being predominantly K2, while neighboring Southeast countries, such as Thailand, Myanmar, and Lao PDR, revealed different patterns of mixed K1 and K2. Meanwhile, African countries displayed distinct patterns, in which K2 was highly dominant in Cameroon, Côte d'Ivoire, and Uganda, and a mixed K1 and K2 pattern was detected in Kenya. For C-T, K1 was generally dominant in Asian countries, whereas K2 was dominant in African countries, suggesting a partial genetic bias by continent.

Polymorphism patterns of the CRR in Vietnamese *pmcsp*

The CRR of Vietnamese *pmcsp* exhibited an extremely high level of polymorphic features (Fig. 6), and various PRMs contributed to its polymorphic patterns. Vietnamese *pmcsp* CRR contained 17 different PRMs. Aside from the two major PRMs, NDAG and NAAG, 15 minor PRMs, including NVAG, TAAG, NAAS, DAAG, NATG, NTAG, NAAD, NAAE, NEAG, SAAG, SDAG, NSAG, NAVG, NGAG, and NAAR, were identified as CRR building blocks. Unusual sequences of NAAAG and AAG identified in the reference sequence (PmUG01) were not found in the Vietnamese *pmcsp* CRR. Different numbers, compositions, and arrangements of these PRMs generated a large number of polymorphisms in Vietnamese *pmcsp* CRR, resulting in 106 distinct haplotypes (H1–H106). Each haplotype was constructed of a different number of PRMs, ranging from 1 to 62. H5 and H25 were the most dominant haplotypes, and were shared by 28 (16.1%) and 24 (13.8%) sequences, respectively, while the remaining 104 haplotypes consisted of fewer than 3 sequences, usually just one sequence.

Polymorphism patterns of the CRR in the global *pmcsp*

Global *pmcsp* also exhibited highly polymorphic features in its CRR due to different arrangements of various PRM species. Vietnamese *pmcsp* CRR showed the greatest size polymorphisms of 46 different lengths, followed by Thai, Myanmar, and Kenyan *pmcsp* CRRs

with 13, 12, and 12, respectively (Fig. 7A and 7B). Bangladeshi, Chinese, Côte d'Ivoire, and Ugandan CRRs revealed size homogeneities, but these were not significant, as only one or two sequences were available for each country. The shortest CRR, which harbored only 1 PRM, was detected in Vietnamese *pmcsp*, while the longest, with 85 PRMs, was found in Kenyan *pmcsp* (Fig. 7A and 7B). The median length of PRMs in the Asian *pmcsp* CRR (median: 53) was longer than that in the African *pmcsp* CRR (median: 51). PRM species were also highly diverse in the global *pmcsp* CRR (Fig. 7C). A total of 32 different PRMs were identified in the global *pmcsp* CRR, of which two PRMs, NAAG and NDAG, were the most common and were shared across the global *pmcsp* CRR at high frequencies. The other 30 PRMs were detected in the global *pmcsp* CRR with varying distributions and country-specific frequencies. Seven PRMs, including NVAG, TAAG, NAAS, DAAG, NSAG, NAAD, and NAAE, were found in Vietnamese *pmcsp* alone. NAPG was commonly found in the CRRs of Asian countries, including Myanmar, Thailand, Lao PDR, and Bangladesh, but not in Vietnam. NEAG and NDEG were also common PRMs in Southeast Asian countries, but their frequencies were low. Overall, PRM species were more diverse in the Asian *pmcsp* CRR than in the African *pmcsp* CRR.

Discussion

Intraspecies sequence variations in the *csp* gene suggest extensive genetic diversity in global *P. falciparum* and *P. vivax* populations [35–37]. The N-T and C-T regions of *pfcsp* and *pvcsp* have been reported to be relatively well conserved, with low levels of genetic diversity across global populations [35–38]. Similarly, the N-T and C-T of global *pmcsp* revealed low levels of genetic diversity, although minor polymorphisms have been observed in CR I in the N-T and CR II, Th2R, and Th3R in the C-T of the global *pmcsp*. Both the N-T and C-T of Vietnamese *pmcsp* showed the lowest level of genetic diversity among the global population, as evidenced by low

Hd (0.045 ± 0.022 for N-T and 0.164 ± 0.038 for C-T) and π (0.0002 ± 0.0001 for N-T and 0.0007 ± 0.0002 for C-T). In particular, Vietnamese *pmcsp* N-T and C-T exhibited much lower genetic diversity and milder natural selection than the genes of Thailand, Myanmar, or Lao PDR [19]. Although N-T and C-T showed low genetic diversity in the Vietnamese and global *P. malariae* populations, C-T exhibited slightly higher nucleotide diversity than N-T. The two T cell epitopes, Th2R and Th3R, are known to be the main targets of host immune responses, and this has resulted in substantial genetic diversity and a positive natural selection force in both *pfcsp* and *pvcsp* due to host immune attack [36–38]. However, the two regions in Vietnamese and global *pmcsp* revealed low genetic diversity and evidence of negative natural selection. This could be due to the small number of *pmcsp* sequences examined, or to a lack of host immune pressure due to the limited effects of natural selection forces. However, given the low transmission intensity of *P. malariae,* this should be investigated further. Such limited genetic variation may obscure geographical structuring, making it difficult to detect clear population differentiation in global *pmcsp*. Due to the low genetic diversity of N-T across global *pmcsp*, the haplotype network did not show a clear geographical bias among populations. However, Vietnamese and Malaysian *pmcsp* N-Ts formed a distinct cluster separated from other Asian *pmcsp* N-Ts. Structure analysis of global *pmcsp* N-T also indicated that Vietnamese and Malaysian *pmcsp* N-Ts share the African *pmcsp* N-T pattern rather than the Asian pattern. Meanwhile, the haplotype network and structure analyses of global *pmcsp* C-T exhibited a partial geographical bias between Asian and African populations, with mild genetic segregation between the two populations. Despite these genetic biases within global *pmcsp* C-T, the region was highly conserved in the global *pmcsp* population at the amino acid level. Given that the C-T contains putative T cell epitopes [36–38], the low genetic diversity of the region in the global

pmcsp population suggests its potential as a vaccine candidate for *P. malariae*, though further studies are needed to evaluate its potential as a vaccine candidate.

Unlike the N-T and C-T of global *pmcsp*, its CRR exhibited considerable polymorphic character. Size polymorphisms, attributed to different numbers and arrangements of various PRM species, were major factors driving these polymorphisms. Overall, the Asian *pmcsp* population had a longer CRR than the African population, although the longest CRR was identified in Kenyan *pmcsp*. Vietnamese *pmcsp* CRR exhibited the greatest size-polymorphism features among global *pmcsp*, which was attributed to the different copy numbers of PRMs. This CRR consisted of more diverse species of PRMs, including 7 novel species (NVAG, TAAG, NAAS, DAAG, NSAG, NAAD, and NAAE) not previously reported, which rendered the greatest genetic heterogeneity. Unlike the CRRs of *pfcsp* and *pvcsp*, no clear evidence for deletion or insertion of strain-specific sequences was identified in the CRR of global *pmcsp*. Two types of PRMs, NAAG and NDAG, were predominant and commonly detected in the global *pmcsp* examined. Meanwhile, the minor PRMs exhibited uneven distributions across global *pmcsp*, resulting in genetic complexity in global *pmcsp* CRR. NAPG was proposed as a unique genetic bias that differentiates the Asian population—Bangladesh, Myanmar, Lao PDR, and Thailand—from the African population [19], but it was not identified in the Vietnamese population. This PRM was also not found in Malaysian *pmcsp* [18]. NAAG was also reported as a feature differentiating the Asian and African populations, as the Asian population was found to carry fewer copies of NAAG [19]. However, the Vietnamese population had a similar prevalence to African populations, and Malaysian *pmcsp* showed a high frequency of PRM copy-number variants similar to the African population [18]. Given the limited number of global *pmcsp* sequences, it is difficult to determine whether these PRMs genuinely act as genetic markers that differentiate

Asian and African *pmcsp* populations. Their potential value in defining population structure therefore requires reconsideration, and further studies are warranted.

The high genetic diversity identified in Vietnamese *pmcsp* CRR is an interesting finding, given the population homogeneity of the *P. malariae* samples collected in restricted areas and the limited collection period. Due to heterogeneity in CRR, Vietnamese *pmcsp* exhibited significant size polymorphisms and genetic variation within the population, resulting in a high MOI of 1.5. The reason for this high genetic diversity of Vietnamese *pmcsp* is unclear, but the rapid transmission of *P. malariae* in restricted endemic areas over a short period and the high prevalence of gametocytes identified in patients [6] might enhance genetic recombination within the parasite genome. A similarly high MOI value of 1.23 in Vietnamese *pmama-1* [21] supports this notion, but further studies with other genetic markers, such as microsatellites, are required to validate this possibility.

This study has several limitations that warrant consideration. First, the number of global *pmcsp* sequences analyzed was small, and the times when sequences were generated varied. Second, the number of sequences from each country was unequal, undermining statistical reliability. In particular, only 1 or 2 sequences from Bangladesh, China, Uganda, and Côte d'Ivoire were included in the analysis. Therefore, it is recommended that the results on genetic diversity and neutrality for individual countries and the global *pmcsp* be interpreted cautiously.

Conclusion

This study is the first investigation conducted on the genetic diversity of *pmcsp* in Vietnamese *P. malariae* isolates. Despite the spatiotemporal homogeneity of the *P. malariae* isolates analyzed, Vietnamese *pmcsp* exhibited a high degree of genetic diversity. As was anticipated, the N-T and C-T of Vietnamese *pmcsp* showed low genetic diversity, but the CRR exhibited high diversity,

resulting in large-size polymorphisms and sequence variations. A comparative analysis of currently available global *pmcsp* sequences suggested substantial genetic diversity of the gene within the global *P. malariae* population, mainly due to CRR. However, no clear genetic bias by continent was identified. This study suggests that the genetic diversity of *pmcsp* in the global *P. malariae* population is greater than previously recognized and highlights the need for larger and more comprehensive studies using larger numbers of global *P. malariae* isolates from different origins to improve the understanding of the genetic nature and evolutionary aspects of the *pmcsp* gene in this parasite.

List of abbreviations

CRR: Central repeat region

CSP: Circumsporozoite surface protein

C-T: C-terminal

dN: Non-synonymous substitution

dS: Synonymous substitution

H: Haplotype

Hd: Haplotype diversity

***K*:** Average pairwise nucleotide differences

K: Quantity of clusters

MOI: Multiplicity of infection

N-T: N-terminal

PCR: Polymerase chain reaction (PCR)

PRM: Peptide repeat motif

S: Segregating sites

Th: T-helper cell epitope

π: Nucleotide diversity

Acknowledgments

The authors thank the staff in the Tropical Diseases Clinical and Treatment Research Department, Institute of Malariology, Parasitology, and Entomology Quy Nhon for their contribution and technical support in field study. This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-02413635).

Authors contributions

TCV, HHQ, and BKN conceptualized the study; TCV, HGL, CVK, NTMT, and NTLH performed the experiments; TCV, HGL, DTDN, THN, and BKN interpreted the results; TCV, HHQ, and BKN designed the study; TCV and BKN wrote the original manuscript; HGL, CVK, DTDN, THN, MTMT, NTLH, and HHQ reviewed and edited the manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (RS-2025-02413635 and RS-2026-25486632).

Availability of data and materials

The nucleotide sequences obtained in this study were deposited in the GenBank database with accession numbers PZ397962–PZ398135.

Ethics declaration

The study protocol was approved by the Institutional Review Board, Institute of Malariology, Parasitology and Entomology Quy Nhon, Vietnam (27/VSR-NCDT). Verbal informed consent was obtained from each participant.

Consent for publication

Not applicable.

Competing interest

The authors declare no conflicts of interest.

References

1. Culleton R, Pain A, Snounou G. *Plasmodium malariae*: the persisting mysteries of a persistent parasite. Trends Parasitol. 2023;39:113–25. https://doi.org/10.1016/j.pt.2022.11.008

2. Shankar H, Kumar G, Ahmed N, Florentin A. *Plasmodium malariae* is an overlooked malaria parasite with emerging challenges. Commun Med (Lond). 2026;6:95. https://doi.org/10.1038/s43856-025-01360-1

3. Brouwer EE, van Hellemond JJ, van Genderen PJ, Slot E, van Lieshout L, Visser LG, et al. A case report of transfusion-transmitted *Plasmodium malariae* from an asymptomatic non-immune traveller. Malar J. 2013;12:439. https://doi.org/10.1186/1475-2875-12-439

4. Scuracchio P, Vieira SD, Dourado DA, Bueno LM, Colella R, Ramos-Sanchez EM, et al. Transfusion-transmitted malaria: case report of asymptomatic donor harboring *Plasmodium malariae*. Rev Inst Med Trop Sao Paulo. 2011;53:55–9. https://doi.org/10.1590/s0036-46652011000100010

411 5. Langford S, Douglas NM, Lampah DA, Simpson JA, Kenangalem E, Sugiarto P, et al.
412 *Plasmodium malariae* Infection Associated with a High Burden of Anemia: A Hospital-Based
413 Surveillance Study. PLoS Negl Trop Dis. 2015;9:e0004195.
414 https://doi.org/10.1371/journal.pntd.0004195

415 6. Khanh CV, Lê HG, Võ TC, Quang NX, Nguyen DV, Dung NCT, et al. Unprecedented large
416 outbreak of *Plasmodium malariae* malaria in Vietnam: Epidemiological and clinical
417 perspectives. Emerg Microbes Infect. 2025;14:2432359.
418 https://doi.org/10.1080/22221751.2024.2432359

419 7. Mayxay M, Pukrittayakamee S, Newton PN, White NJ. Mixed-species malaria infections in
420 humans. Trends Parasitol. 2004;20:233–40. https://doi.org/10.1016/j.pt.2004.03.006

421 8. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, et al.
422 Microsatellite markers reveal a spectrum of population structures in the malaria parasite
423 *Plasmodium falciparum*. Mol Biol Evol. 2000;17:1467–82.
424 https://doi.org/10.1093/oxfordjournals.molbev.a026247

425 9. Lê HG, Naw H, Kang J-M, Võ TC, Myint MK, Htun ZT, et al. Molecular Profiles of Multiple
426 Antimalarial Drug Resistance Markers in *Plasmodium falciparum* and *Plasmodium vivax* in the
427 Mandalay Region, Myanmar. Microorganisms. 2022;10:2021.
428 https://doi.org/10.3390/microorganisms10102021

429 10. Lê HG, Võ TC, Kang J-M, Van Khanh C, Trinh NTM, Hanh NTL, et al. Molecular Profiles
430 of Antimalarial Drug Resistance in Plasmodium Species from Asymptomatic Malaria Carriers in

Gia Lai Province, Vietnam. Microorganisms. 2025;13:2101. https://doi.org/10.3390/microorganisms13092101

11. Võ TC, Lê HG, Kang J-M, Trinh NTM, Quang HH, Na B-K. Genetic polymorphism of merozoite surface protein 1 and merozoite surface protein 2 in the Vietnam *Plasmodium falciparum* population. BMC Infect Dis. 2024;24:1216. https://doi.org/10.1186/s12879-024-10116-6

12. Zaib K, Khan A, Khan MU, Ullah I, Võ TC, Kang J-M, et al. Genetic structure of apical membrane antigen-1 in *Plasmodium falciparum* isolates from Pakistan. Parasites Hosts Dis. 2024;62:302–12. https://doi.org/10.3347/PHD.24028

13. Nguyễn KO, Kang J-M, Võ TC, Lê HG, Akhtar S, Nguyễn TH, et al. Genetic polymorphisms of merozoite surface protein-3α in *Plasmodium vivax* isolates from Pakistan. Acta Trop. 2025;272:107904. https://doi.org/10.1016/j.actatropica.2025.107904

14. Võ TC, Lê HG, Kang J-M, Minh Trinh NT, Cho M, Goo Y-K, et al. Genetic diversity and natural selection of cell-traversal protein for ookinetes and sporozoites (CelTOS) in *Plasmodium falciparum* isolates from Vietnam. Gene. 2025;968:149731. https://doi.org/10.1016/j.gene.2025.149731

15. Nguyễn TH, Lê HG, Võ TC, Nguyễn ĐTD, Nguyễn KO, Cho M, et al. Molecular surveillance of antifolate drug resistance markers in *Plasmodium vivax* from Khyber Pakhtunkhwa province, northwest Pakistan. Acta Trop. 2025;264:107583. https://doi.org/10.1016/j.actatropica.2025.107583

16. Srisutham S, Saralamba N, Sriprawat K, Mayxay M, Smithuis F, Nosten F, et al. Genetic diversity of three surface protein genes in *Plasmodium malariae* from three Asian countries. Malar J. 2018;17:24. https://doi.org/10.1186/s12936-018-2176-x

17. Rodi M, Kawecka K, Stephan L, Berner L, Medina MS, Lalremruata A, et al. Genetic diversity of *Plasmodium malariae* in sub-Saharan Africa: a two-marker genotyping approach for molecular epidemiological studies. Front Cell Infect Microbiol. 2024;14:1405198. https://doi.org/10.3389/fcimb.2024.1405198

18. Phang WK, Bukhari FDM, Zen LPY, Jaimin JJ, Dony JJF, Lau YL. Genetic polymorphism of circumsporozoite protein (CSP) in *Plasmodium malariae* isolates from Malaysia. Parasitol Int. 2022;87:102519. https://doi.org/10.1016/j.parint.2021.102519

19. Saralamba N, Mayxay M, Newton PN, Smithuis F, Nosten F, Archasuksan L, et al. Genetic polymorphisms in the circumsporozoite protein of *Plasmodium malariae* show a geographical bias. Malar J. 2018;17:269. https://doi.org/10.1186/s12936-018-2413-3

20. Guimarães LO, Wunderlich G, Alves JMP, Bueno MG, Röhe F, Catão-Dias JL, et al. Merozoite surface protein-1 genetic diversity in *Plasmodium malariae* and *Plasmodium brasilianum* from Brazil. BMC Infect Dis. 2015;15:529. https://doi.org/10.1186/s12879-015-1238-8

21. Nguyễn ĐTD, Lê HG, Võ TC, Nguyễn TH, Kang J-M, Nguyễn KO, et al. Genetic polymorphism and natural selection of apical membrane antigen-1 in *Plasmodium malariae* isolated from Vietnam. BMC Infect Dis. 2026;26:476. https://doi.org/10.1186/s12879-026-12935-1

22. Lo E, Nguyen K, Nguyen J, Hemming-Schroeder E, Xu J, Etemesi H, et al. *Plasmodium malariae* Prevalence and csp Gene Diversity, Kenya, 2014 and 2015. Emerg Infect Dis. 2017;23:601–10. https://doi.org/10.3201/eid2304.161245

23. Coppi A, Natarajan R, Pradel G, Bennett BL, James ER, Roggero MA, et al. The malaria circumsporozoite protein has two functional domains, each with distinct roles as sporozoites journey from mosquito to mammalian host. J Exp Med. 2011;208(2):341–56; https://doi.org/10.1084/jem.20101488

24. Sinnis P, Fidock DA. The RTS,S vaccine-a chance to regain the upper hand against malaria? Cell. 2022;185:750–4. https://doi.org/10.1016/j.cell.2022.01.028

25. Moorthy V, Binka F. R21/Matrix-M: a second malaria vaccine? Lancet. 2021;397:1782–3. https://doi.org/10.1016/S0140-6736(21)01065-5

26. Lal AA, de la Cruz VF, Campbell GH, Procell PM, Collins WE, McCutchan TF. Structure of the circumsporozoite gene of *Plasmodium malariae*. Mol Biochem Parasitol. 1988;30:291–4. https://doi.org/10.1016/0166-6851(88)90099-0

27. Qari SH, Collins WE, Lobel HO, Taylor F, Lal AA. A study of polymorphism in the circumsporozoite protein of human malaria parasites. Am J Trop Med Hyg. 1994;50:45–51. https://doi.org/10.4269/ajtmh.1994.50.45

28. Tahar R, Ringwald P, Basco LK. Heterogeneity in the circumsporozoite protein gene of *Plasmodium malariae* isolates from sub-Saharan Africa. Mol Biochem Parasitol. 1998;92:71–8. https://doi.org/10.1016/s0166-6851(97)00226-0

29. Snounou G, Viriyakosol S, Jarra W, Thaithong S, Brown KN. Identification of the four human malaria parasite species in field samples by the polymerase chain reaction and detection of a high prevalence of mixed infections. Mol Biochem Parasitol. 1993;58(2):283–92. https://doi.org/10.1016/0166-6851(93)90050-8

30. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451–2. https://doi.org/10.1093/bioinformatics/btp187

31. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30:2725–9. https://doi.org/10.1093/molbev/mst197

32. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol. 2005;14(8):2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x

33. Li Y-L, Liu J-X. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour. 2018;18:176–7. https://doi.org/10.1111/1755-0998.12719

34. Rathore D, Sacci JB, de la Vega P, McCutchan TF. Binding and invasion of liver cells by *Plasmodium falciparum* sporozoites. Essential involvement of the amino terminus of circumsporozoite protein. J Biol Chem. 2002;277:7092–8. https://doi.org/10.1074/jbc.M106862200

35. Võ TC, Lê HG, Kang J-M, Moe M, Naw H, Myint MK, et al. Genetic polymorphism and natural selection of circumsporozoite protein in Myanmar *Plasmodium vivax*. Malar J. 2020;19(1):303. https://doi.org/10.1186/s12936-020-03366-7

36. Lê HG, Kang JM, Moe M, Jun H, Thái TL, Lee J, et al. Genetic polymorphism and natural selection of circumsporozoite surface protein in *Plasmodium falciparum* field isolates from Myanmar. Malar J. 2018;17(1):361. https://doi.org/10.1186/s12936-018-2513-0

37. Võ TC, Lê HG, Kang J-M, Trinh NTM, Van Khanh C, Cho M, et al. Genetic diversity and natural selection of circumsporozoite surface protein in Vietnam *Plasmodium falciparum* isolates. Malar J. 2025;24:320. https://doi.org/10.1186/s12936-025-05585-2

38. Võ TC, Trinh NTM, Lê HG, Kang J-M, Yoo WG, Quang HH, et al. Genetic Diversity of Circumsporozoite Surface Protein of *Plasmodium vivax* from the Central Highlands, Vietnam. Pathogens. 2022;11:1158. https://doi.org/10.3390/pathogens11101158

Figure Legends

Fig. 1. Amino acid changes in the N-T and C-T of Vietnamese *pmcsp*. The positions of amino acid residues are marked based on the reference sequence, PmUG01 (GenBank ID: XM_029004694.1). Amino acid changes identified in Vietnamese *pmcsp* compared to the reference sequence are marked in red. Dots indicate residues that match the reference sequence. The conserved KLKQP motif, which is critical for sporozoite invasion of mosquito salivary glands and subsequent hepatocyte binding in *Plasmodium* [34], is underlined. The number of sequences corresponding to each haplotype is shown on the right. N-T, N-terminal; CRR, central repeat region; C-T, C-terminal.

Fig. 2. Polymorphism patterns in the N-T and C-T of global *pmcsp*. The N-T and C-T of global *pmcsp* exhibited high genetic conservation with only a few amino acid substitutions. CR I, conserved region I; CR II, conserved region II; T-helper cell epitope 2R, Th2R; T-helper cell epitope 3R, Th3R.

Fig. 3. Haplotype diversity, nucleotide diversity, and natural selection in global *pmcsp*. The values for individual countries and continents are represented by different colors. Tajima's D value for Cameroon is not presented, as the population had no natural selection signal.

Fig. 4. Sliding window plots. The plots depict nucleotide diversity (π) and Tajima's D values across the two terminal regions of global *pmcsp*. A window size of 10 bp and a step size of 5 bp were applied. Patterns of π and Tajima's D for each country and continent are presented as lines of different colors. The positions of the CR I, CR II, Th2R, and Th3R regions in the N-T and C-T are marked by color bars.

Fig. 5. Haplotype network and structure analysis of global *pmcsp*. Haplotype networks of N-T and C-T of global *pmcsp*, along with the PmUG01 (GenBank ID: XM_029004694.1) reference

556 sequence. Country sequences are represented by different colors. Node sizes indicate the
557 frequencies of specific haplotypes. The lengths of lines connecting nodes reflected the number of
558 base pair substitutions between haplotypes. In the structure analysis, global *pmcsp* sequences
559 were grouped into two distinct clusters, K1 and K2. Bars depict the proportion of the respective
560 cluster.

561 **Fig. 6. Size polymorphism profiles in the CRR of Vietnamese *pmcsp*.** The CRR of
562 Vietnamese *pmcsp* exhibited considerable polymorphic characteristics attributed to different
563 numbers, arrangements, and combinations of 19 distinct PRM species. A total of 106 distinct
564 haplotypes were identified in Vietnamese *pmcsp* CRR, among which H5 and H25 were the most
565 dominant. Each PRM is represented by a different color. The number of sequences
566 corresponding to each haplotype is provided as a heatmap on the left.

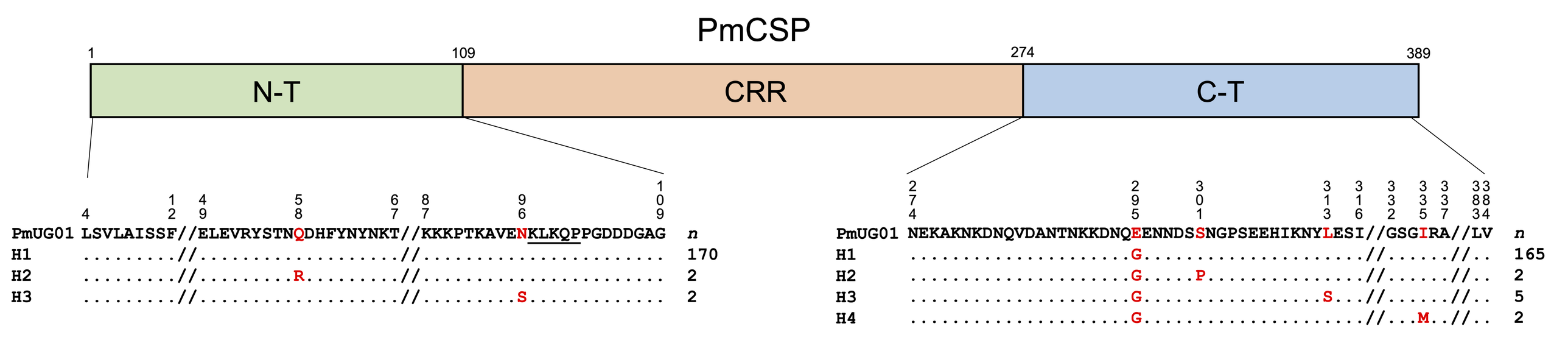
567 **Fig. 7. Polymorphic characters of global *pmcsp* CRR. A** Frequencies of the number of PRMs
568 in global *pmcsp* CRR. Global *pmcsp* CRR was constructed with different numbers of PRMs. **B**
569 Size polymorphisms of the CRR in global *pmcsp*. Variations in the number of PRMs led to size
570 polymorphisms in *pmcsp* populations. **C** Composition of PRMs in the global *pmcsp* population.
571 A total of 32 different PRMs were identified in global *pmcsp*. These PRMs were distributed in
572 the global population with different patterns and frequencies, but NAAG and NDAG were
573 common and consistently identified in the global population. The proportions of each PRM in
574 individual countries are provided.

575

576 **Supplementary information**

577 **Supplementary Table S1.** Primer sequences and thermal conditions for pmcsp amplification

578 **Supplementary Table S2.** Global *pmcsp* sequences analyzed in this study


PmCSP
1
109
274
389
N-T
CRR
C-T
PmUG01 LSVLAISSF//ELEVRYSTNQDHFYNYNKT//KKKPTKAVENKLKQPPGDDDGAG n
H1 .....//.....//..... 170
H2 .....//.....R.....//..... 2
H3 .....//.....//.....S..... 2
PmUG01 NEKAKNKDNQVDANTNKKDNQEENNDSSNGPSEEHIKNYLESI//GSGIRA//LV n
H1 .....G.....//.....//.. 165
H2 .....G.....P.....//.....//.. 2
H3 .....G.....S...//.....//.. 5
H4 .....G.....//...M..//.. 2

N-T

C-T

CR I

CR II

Th2R

Th3R

PmUG01

Vietnam (*n* = 174)

Myanmar (*n* = 40)

Thailand (*n* = 44)

Lao PDR (*n* = 4)

Malaysia (*n* = 26)

Bangladesh (*n* = 1)

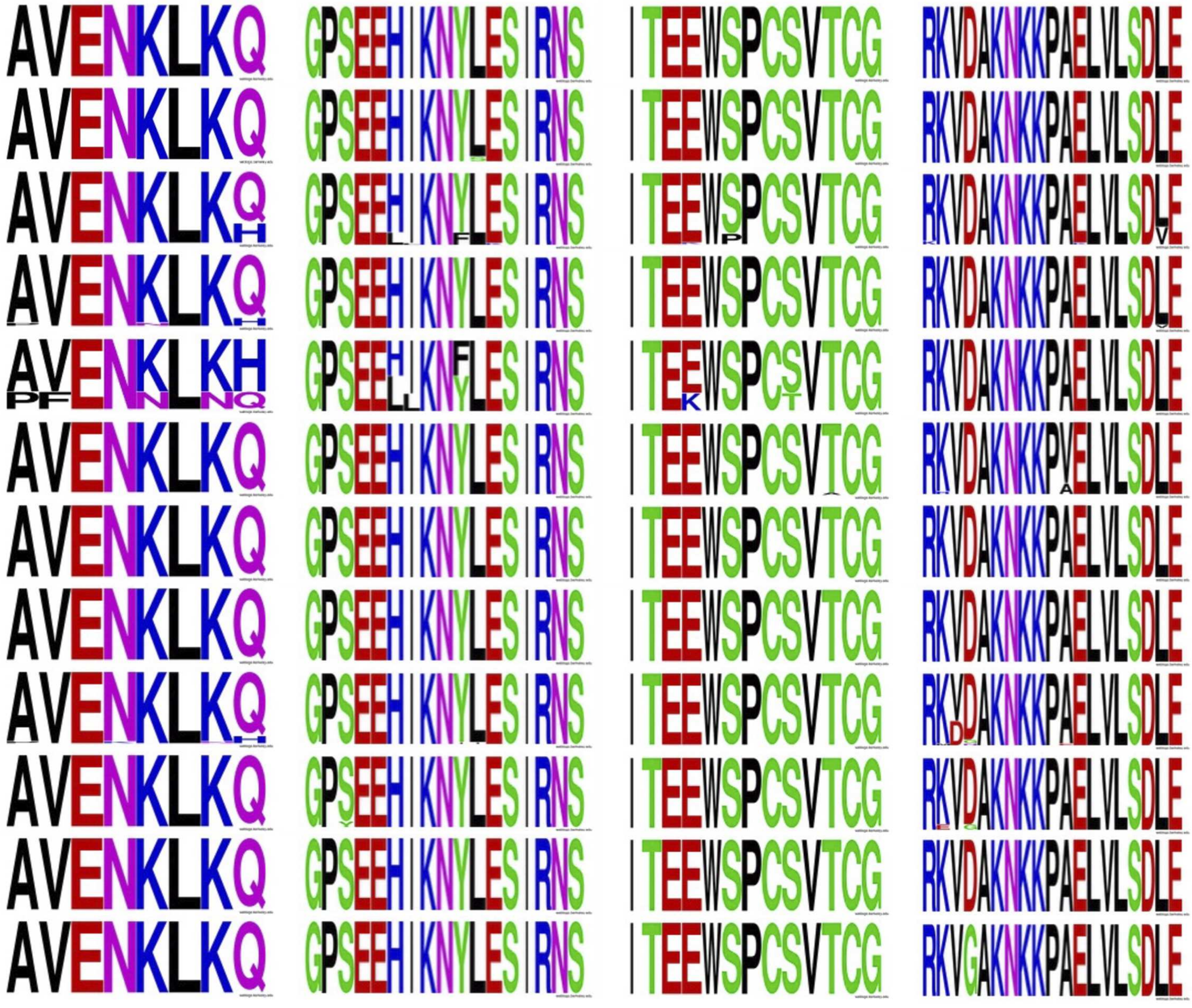
China (*n* = 2)

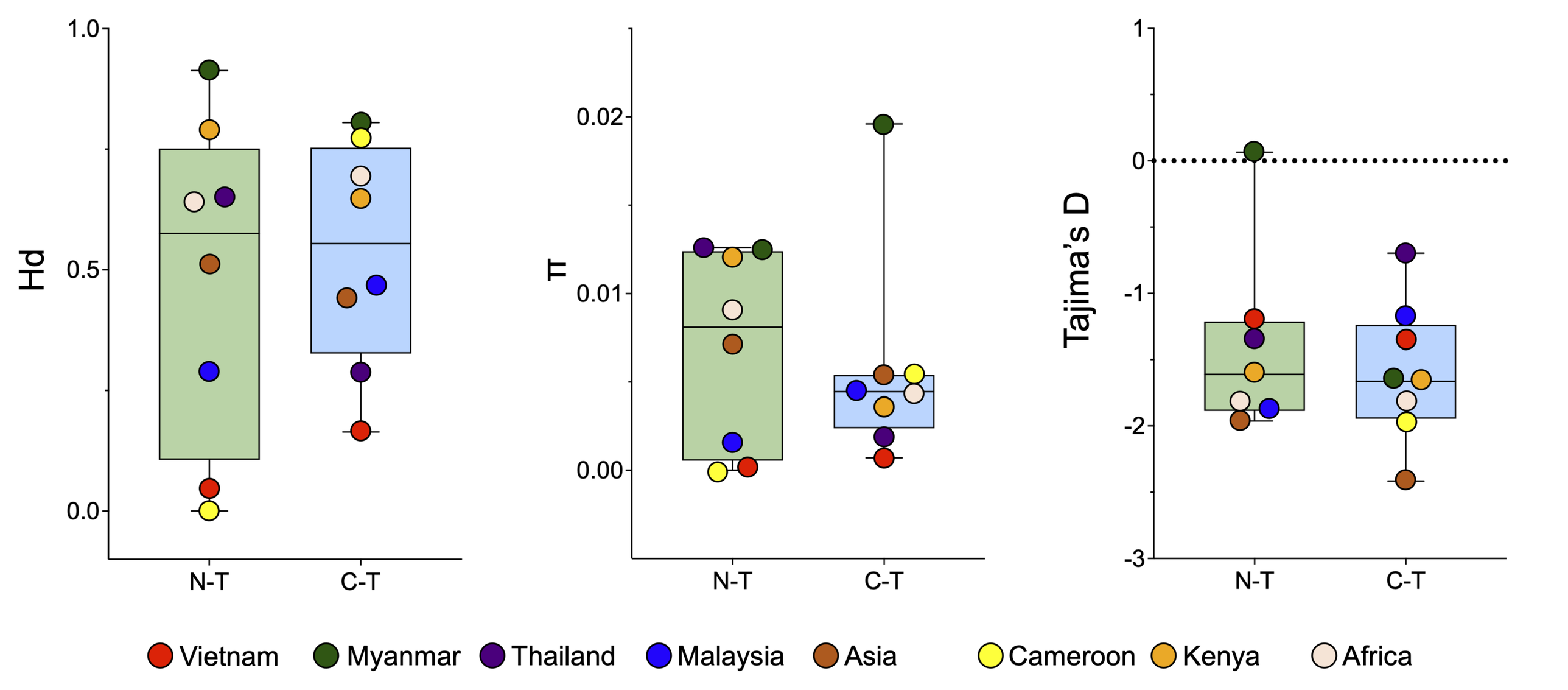
Kenya (*n* = 39)

Cameroon (*n* = 12)

Côte d'Ivoire (*n* = 1)

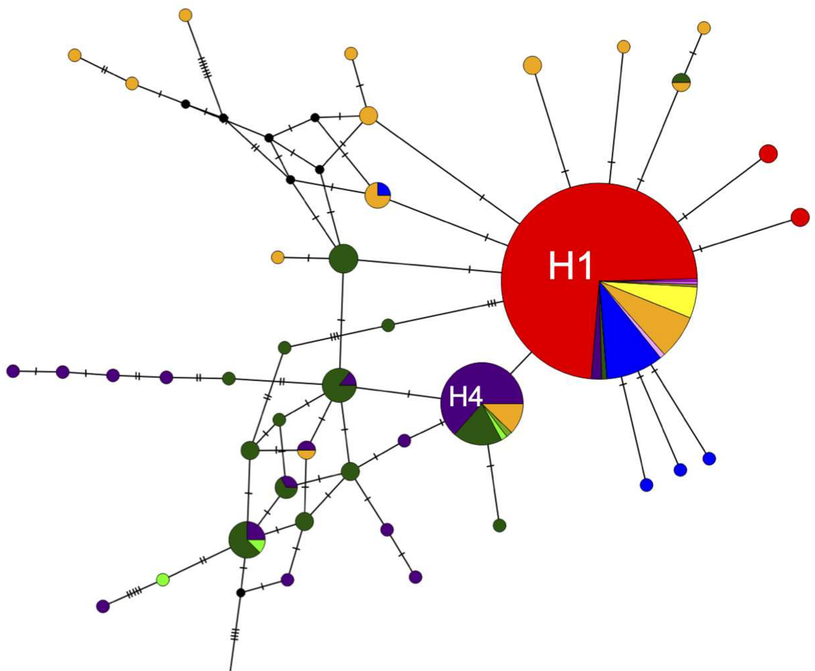
Uganda (*n* = 1)




Hd
1.0
0.5
0.0
N-T
C-T
π
0.02
0.01
0.00
N-T
C-T
Tajima's D
1
0
-1
-2
-3
N-T
C-T
Vietnam
Myanmar
Thailand
Malaysia
Asia
Cameroon
Kenya
Africa

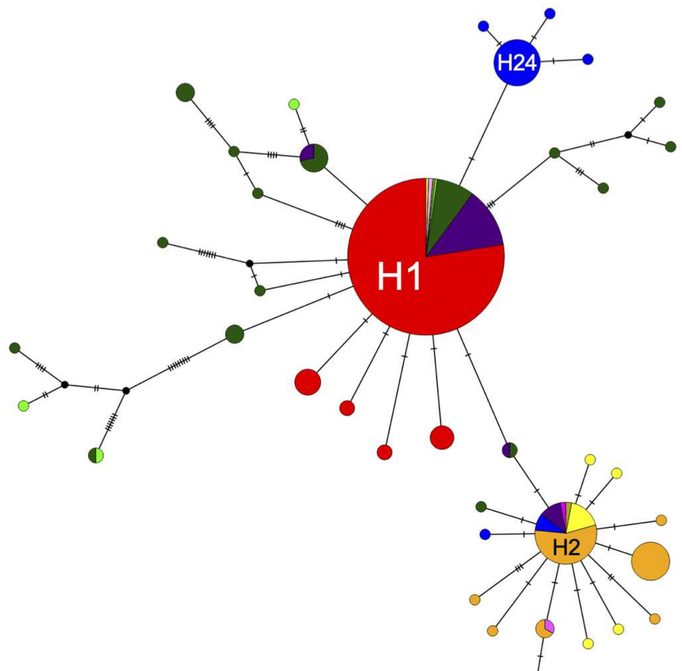
N-T
C-T
0.10
0.05
0.00
π
100
200
300
900
1000
1100
Nucleotide position
Vietnam
Thailand
Myanmar
Malaysia
Asia
Kenya
Cameroon
Africa
CR I
CR II
Th2R
Th3R
2
1
0
-1
-2
Tajima's D

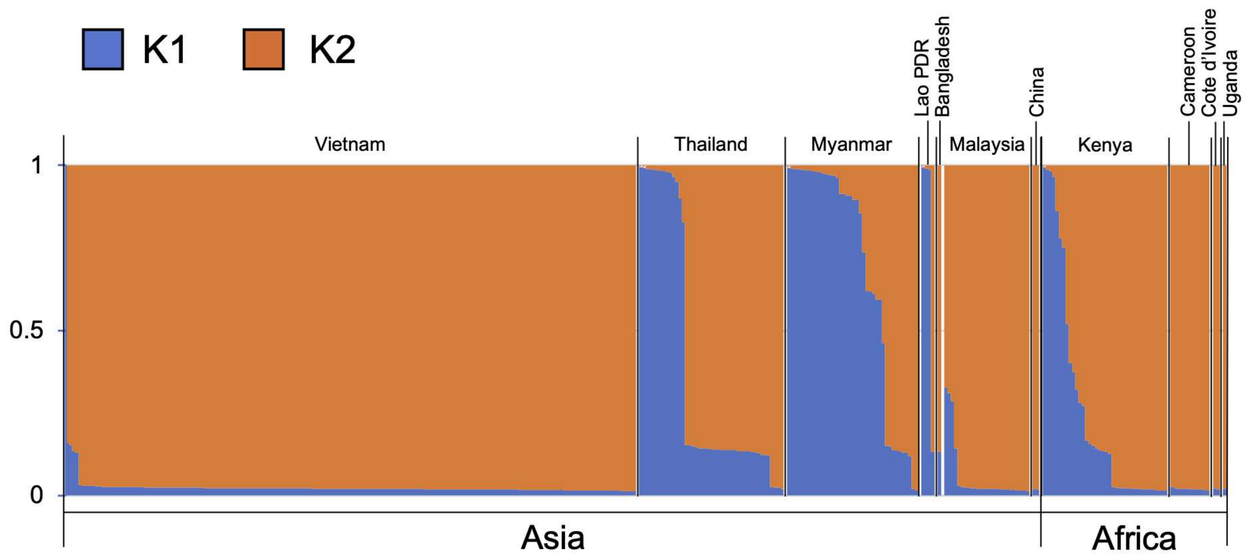
N-T

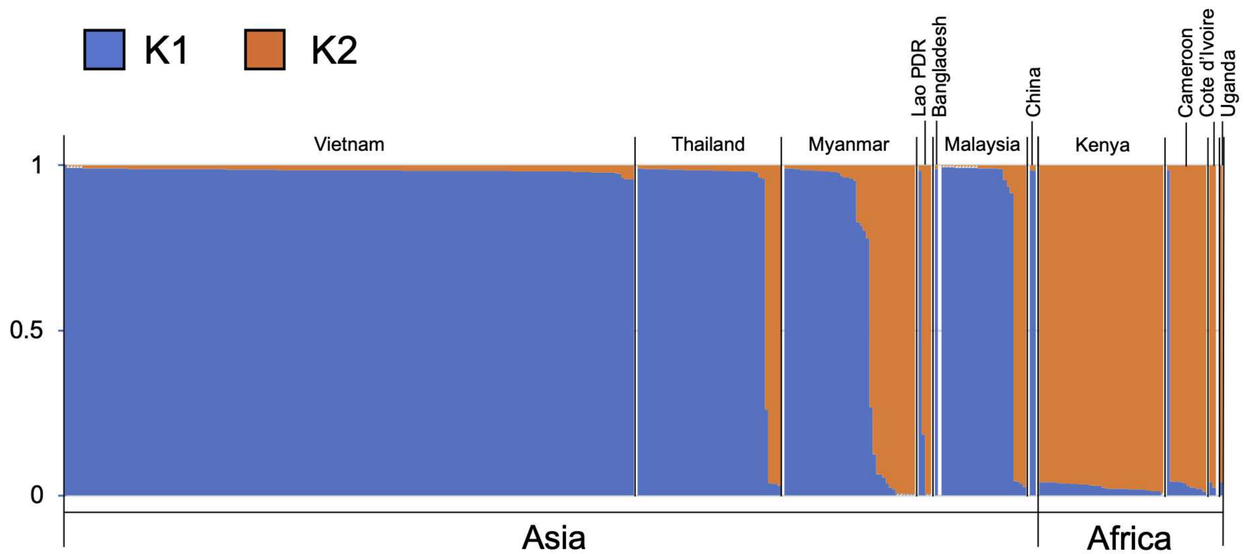


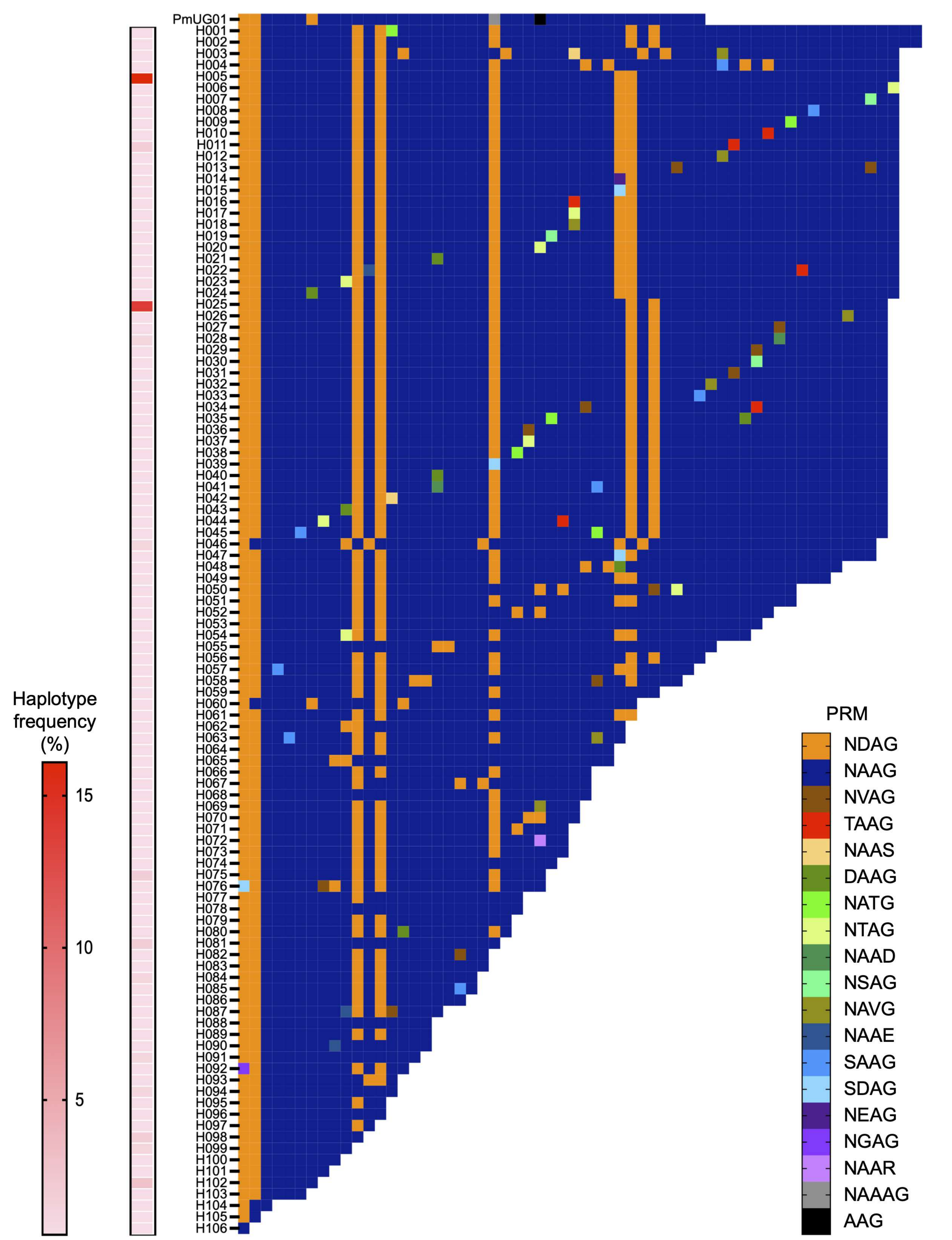
Vietnam
Thailand
Myanmar
Lao PDR
Bangladesh
Malaysia
China
Kenya
Cameroon
Côte d'Ivoire
Uganda
PmUG01

C-T

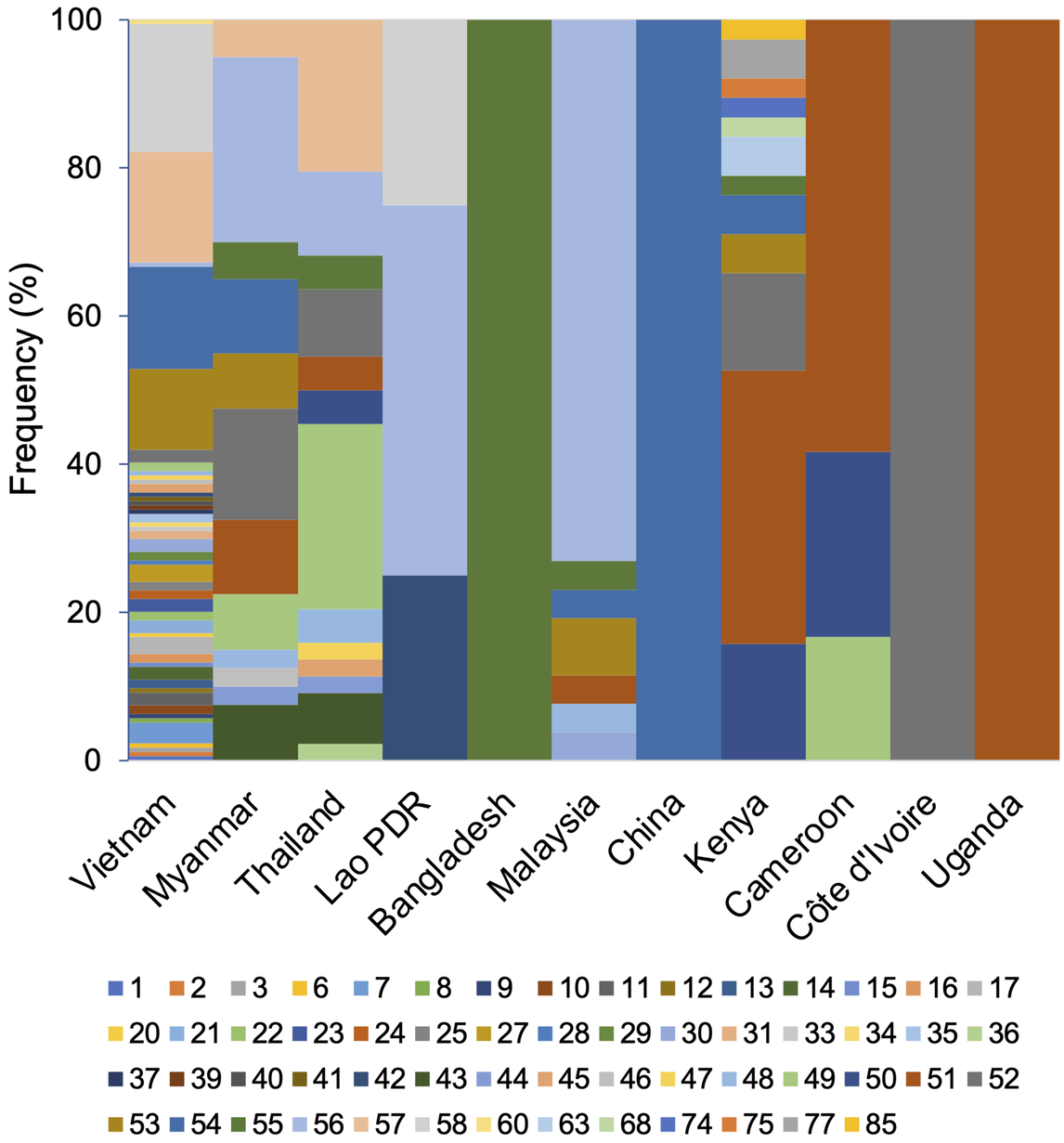




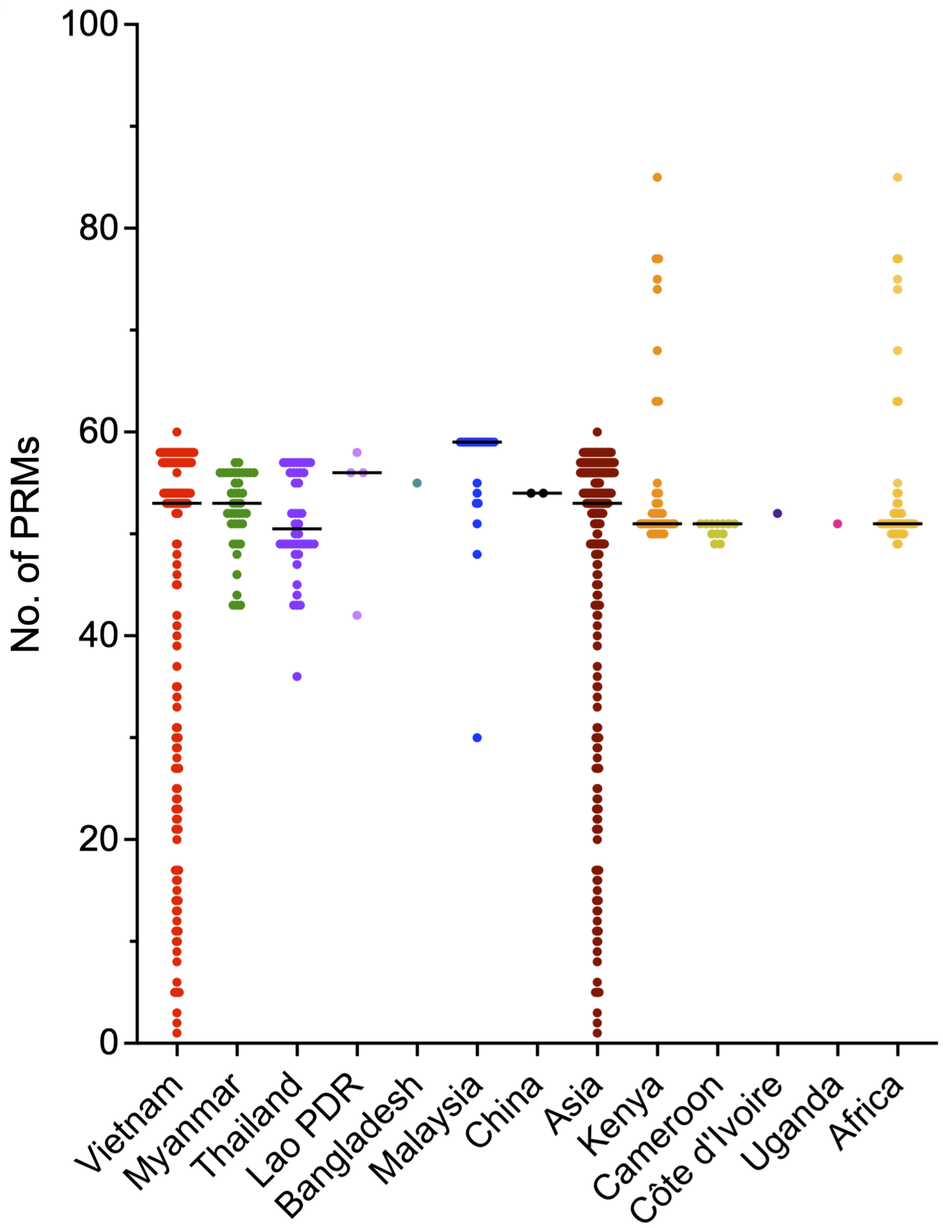



PmUG01
H001
H002
H003
H004
H005
H006
H007
H008
H009
H010
H011
H012
H013
H014
H015
H016
H017
H018
H019
H020
H021
H022
H023
H024
H025
H026
H027
H028
H029
H030
H031
H032
H033
H034
H035
H036
H037
H038
H039
H040
H041
H042
H043
H044
H045
H046
H047
H048
H049
H050
H051
H052
H053
H054
H055
H056
H057
H058
H059
H060
H061
H062
H063
H064
H065
H066
H067
H068
H069
H070
H071
H072
H073
H074
H075
H076
H077
H078
H079
H080
H081
H082
H083
H084
H085
H086
H087
H088
H089
H090
H091
H092
H093
H094
H095
H096
H097
H098
H099
H100
H101
H102
H103
H104
H105
H106
Haplotype frequency (%)
15
10
5
PRM
NDAG
NAAG
NVAG
TAAG
NAAS
DAAG
NATG
NTAG
NAAD
NSAG
NAVG
NAAE
SAAG
SDAG
NEAG
NGAG
NAAR
NAAAG
AAG

A



B



C

	PmUG01	Vietnam	Myanmar	Thailand	Lao PDR	Malaysia	Bangladesh	China	Kenya	Cameroon	Côte d'Ivoire	Uganda
NDAG	7.32	12.97	10.72	9.60	10.38	11.08	10.91	12.96	9.84	11.30	11.54	11.76
NAAG	87.80	86.06	67.35	77.01	20.28	87.07	83.64	85.19	89.41	88.53	88.46	88.24
NVAG		0.16										
TAAG		0.11										
NAAS		0.03										
DAAG		0.09										
NATG		0.07		0.18								
NTAG		0.11				0.07						
NSAG		0.04										
NAVG		0.09								0.17		
NAAD		0.04										
NAAE		0.04										
SAAG		0.11				0.07						
SDAG		0.05				0.07						
NEAG		0.01	0.10	0.31								
NGAG		0.01	0.10			0.07						
NAAR		0.01							0.05			
NAPG			21.11	12.32	68.40		5.45					
NPPG			0.38									
NEPG			0.19	0.04								
NAAE			0.05						0.38			
NDPG				0.13	0.47							
NDEG				0.27	0.47	1.56		1.85				
IDAG				0.04								
NTPG				0.04								
IAAG				0.04								
NAGG									0.05			
NDAE									0.05			
NA									0.05			
NAAAG	2.44											
AAG	2.44											
Unidentified									0.19			

Table 1. Nucleotide diversity and neutrality in the N-T and C-T of the global *pmcsp*

Origin	Region	*n*	*K*	S	Eta	H	Hd ± SD	π ± SD	dN – dS	Tajima's D (*P* value)	Fu & Li's D (*P* value)	Fu & Li's F (*P* value)
Vietnam	N-T	174	0.05	2	2	3	0.045 ± 0.022	0.0002 ± 0.0001	0.000	–1.213 (*P* > 0.1)	0.646 (*P* > 0.1)	0.068 (*P* > 0.1)
	C-T		0.17	4	4	5	0.164 ± 0.038	0.0007 ± 0.0002	–0.001	–1.376 (*P* > 0.1)	0.895 (*P* > 0.1)	0.172 (*P* > 0.1)
Thailand	N-T	44	2.42	18	18	15	0.649 ± 0.082	0.0126 ± 0.0027	0.003	–1.334 (*P* > 0.1)	–1.068 (*P* > 0.1)	–1.369 (*P* > 0.1)
	C-T		0.46	3	3	4	0.289 ± 0.086	0.0019 ± 0.0006	–0.002	–0.697 (*P* > 0.1)	0.905 (*P* > 0.1)	0.497 (*P* > 0.1)
Myanmar	N-T	40	2.40	10	10	15	0.913 ± 0.022	0.0125 ± 0.0011	0.000	0.064 (*P* > 0.1)	–0.342 (*P* > 0.1)	–0.250 (*P* > 0.1)
	C-T		4.83	37	39	16	0.805 ± 0.059	0.0196 ± 0.0042	–0.002	–1.661 (0.05 < *P* < 0.1)	–1.500 (*P* > 0.1)	–1.846 (*P* > 0.1)
Malaysia	N-T	26	0.31	4	4	5	0.289 ± 0.115	0.0016 ± 0.0007	–0.001	–1.887 (*P* < 0.05)	–2.869 (*P* < 0.05)	–2.996 (*P* < 0.05)
	C-T		1.12	7	7	6	0.465 ± 0.116	0.0046 ± 0.0014	–0.005	–1.192 (*P* > 0.1)	–1.406 (*P* > 0.1)	–1.562 (*P* > 0.1)
Asia*	N-T	291	1.36	29	30	33	0.512 ± 0.035	0.0071 ± 0.0008	0.003	–1.963 (*P* < 0.05)	–3.428 (*P* < 0.02)	–3.388 (*P* < 0.02)
	C-T		1.33	47	51	28	0.439 ± 0.037	0.0054 ± 0.0010	–0.001	–2.416 (*P* < 0.01)	–2.865 (*P* < 0.05)	–3.205 (*P* < 0.02)
Kenya	N-T	38	2.33	17	19	14	0.785 ± 0.064	0.0121 ± 0.0032	–0.013	–1.611 (0.05 < *P* < 0.1)	–1.538 (*P* > 0.1)	–1.844 (*P* > 0.1)
	C-T		0.93	9	9	7	0.644 ± 0.055	0.0038 ± 0.0008	0.005	–1.667 (0.05 < *P* < 0.1)	–2.970 (*P* < 0.05)	–3.002 (*P* < 0.05)
Cameroon	N-T	12	0	0	0	1	0.000	0.0000	0.000	ND	ND	ND
	C-T		1.33	8	8	7	0.773 ± 0.128	0.0054 ± 0.0015	–0.002	–1.983 (*P* < 0.05)	–2.446 (*P* < 0.02)	–2.641 (*P* < 0.02)
Africa**	N-T	52	1.74	17	19	14	0.638 ± 0.075	0.0091 ± 0.0026	–0.010	–1.846	–1.836	–2.182

								($P < 0.05$)	($P > 0.1$)	($0.05 < P < 0.1$)
	C-T	1.07	16	16	13	0.694 ± 0.056	0.0043 ± 0.0007	0.003	–2.151 ($P < 0.05$)	–4.605 ($P < 0.02$)	–4.460 ($P < 0.02$)

*Included sequences from Vietnam (n = 174), Thailand (n = 44), Myanmar (n = 40), Malaysia (n = 26), Lao PDR (n = 4), Bangladesh (n = 1), and China (n = 2).

**Included sequences from Kenya (n = 38), Cameroon (n = 12), Côte d'Ivoire (n = 1), and Uganda (n = 1).

n, number of analyzed sequences; K, average number of nucleotide differences; S, number of segregating sites; Eta, total number of mutations; H = number of haplotypes; Hd, haplotype diversity; π, observed average pairwise nucleotide diversity; SD, standard deviation; dN, nonsynonymous mutation; dS, synonymous mutation; ND, not determined

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1.docx
- SupplementaryTableS2.docx