Proceedings of the Indian National Science Academy (2026) 92:1046–1055
https://doi.org/10.1007/s43538-025-00470-7

RESEARCH PAPER

Genetic diversity and haplotype analyses of *Anopheles stephensi* Liston 1901 using ITS2 and COI markers

Lipsa Routray[1] · Laxman Kumar Murmu[1] · Tapan Kumar Barik[1]

Received: 30 March 2025 / Accepted: 20 May 2025 / Published online: 31 May 2025
© Indian National Science Academy 2025

Abstract
Anopheles stephensi a prominent vector in urban India and an emerging vector in African countries, significantly contributes to malaria transmission. This study explores the genetic diversity, haplotype network, and phylogenetic analyses based on the Internal Transcribe Spacer 2 (ITS2) and Cytochrome Oxidase c subunit I (COI) publicly available GenBank repository. A total of 73 ITS2 and 64 COI sequences were used for analysis, out of 140 and 111 retrieved sequences, respectively. The phylogenetic analysis of ITS2 sequences through Bayesian approaches reveals two distinct clades, whereas, COI sequences reveal six distinct clades with posterior probability (pp) 100%. Haplotype network analysis for ITS2 shows four haplotypes with one shared haplotype and three private haplotypes and COI shows nine haplotypes with two shared haplotypes and seven private haplotypes. The genetic heterogeneity based on 73 ITS2 and 64 COI sequences, reveals low haplotype diversity (Hd) values of 0.532 and 0.435, respectively. Tajima's D value is found to be -1.53298 ($P > 0.10$) for COI sequences, suggesting an excess of rare mutations, while 4.91417 ($P < 0.001$), value for ITS2, indicates fewer rare mutations, implying balancing selection. These findings enhance the genetic diversity and population structure of *Anopheles stephensi*, which has potential implications for vector control strategies.

Keywords Haplotype · Phylogenetics · Tajima's D · Population structure

Introduction

Malaria has caused the most significant morbidity and mortality globally (WHO 2024). Of the 460 known *Anopheles* species, around 70–80 is responsible for malaria transmission (Mopuri et al. 2019). The disease threatens approximately 3 billion people globally (Venkatesan 2024). The WHO malaria report highlights 263 million malaria cases in 85 countries and regions where the disease is endemic, with an increase of 14 million cases over 2022 (WHO 2024). India accounts for a significant number of malaria cases in the subcontinent and has implemented national elimination policies and initiatives to eliminate malaria by 2030. However, the comeback of malaria is threatened by factors such as treatment failure, resistance, changes in temperature and ecology, vector range and spread, and changes in vector behaviour (Karunaweera et al. 2014).

Anopheles stephensi is a prominent vector in urban India and significantly contributes to malaria transmission in the regions (Tyagi 2023). *An. stephensi* also been reported from countries China, Afghanistan, Pakistan, Iran, Myanmar, Egypt, and Thailand (Dong et al. 2019; Ahmed et al. 2021; Tadesse et al. 2021; Ryan et al. 2023). The phylogenetic and genetic diversity of organisms reveals their evolutionary history and future prospects (Hu et al. 2021). The geographic distribution and evolutionary relationships of *An. stephensi* populations across the region may help understand evolution and genetic diversity, predict probable future genetic divergence, and devise mitigating strategies for vectors. Molecular markers are paramount in the study of evolution, genetic diversity, and population structure of vectors. Specifically, makers derived from ribosomal DNA and the mitochondrial genome are widely acclaimed for taxonomy and molecular phylogenetic analysis (Sallum et al. 2002; Singh et al. 2018). The multi-locus barcoding such as Cytochrome c Oxidase subunit I (COI), internal transcribed spacer 1(ITS1) and internal transcribed spacer 2(ITS2) are used for genetic studies, however ITS2 barcode shows better discriminatory resolution than the ITS1 in genetic studies

✉ Laxman Kumar Murmu
murmu.laxmankumar36@gmail.com

[1] Department of Zoology, Berhampur University, Bhanja Bihar, Ganjam, Odisha 760007, India

and thus, the ITS2 is employed complentarily to COI in taxonomy and genetic diversity studies(Wu et al.2023).The internal transcribed spacer 2 (ITS2) region and Cytochrome c Oxidase subunit I (COI) genes have been utilized to examine the genetic diversity and evolutionary relationship of mosquitoes (Djadid et al. 2003; Walton et al. 2007; Loaiza et al. 2013; Sum et al. 2014; Fang et al. 2017; Rathnayake et al. 2023; Askari et al. 2024; Muthukumarana et al. 2024). This study represents the first attempt to utilize *An. stephensi* ITS2 and COI sequence up-to-date data from the GenBank database for evaluating the haplotype network, phylogenetic relationships, and genetic diversity.

Materials and methods

Data collection

ITS2 and COI gene sequences of *An. stephensi* were mined from the NCBI public data repository, and 140 (Supplementary Table 1) and 111 (Supplementary Table 2) sequences were obtained, respectively. Table 1 presents the country-wise number of ITS2 and COI sequences. Four *Aedes aegypti* sequences were utilized as the outgroups for phylogenetic analysis (Supplementary Table 3).

Sequence alignment

MEGA v11 was employed for sequence alignment and base quality trimming (Tamura et al. 2021). The misalignment was determined through visual inspection while doing the sequence alignment using Clustal W in MEGA v11, and the misaligned sequences were deleted manually from the rest of the sequences. The duplication of the sequences was assessed based on the metadata, i.e., sequences from the same species from the same location. Accordingly, repeated similar sequences were excluded to obtain a filtered dataset for better analysis. Out of 140 sequences in the ITS2 sequence alignment, 67 were eliminated due to misalignment, shorter, and duplicated sequences (Supplementary Table 1).

Out of 111 sequences in the COI sequence alignment, 47 were eliminated due to misalignment, shorter, and duplicated comparable sequences (Supplementary Table 2). As a result, 64 COI and 73 ITS2 sequences were acquired. The final lengths of the sequences aligned with COI were 342 bp and ITS2 were 333 bp.

Phylogenetic analysis

The trimmed ITS2 and COI data generated after alignment were saved in FASTA format and converted to the Phylip alignment (.phy) and Nexus alignment (.nex) file formats using the Geneious Prime 2022.1.1 program. Partition Finder v1.1.2 programm was employed to obtain optimal evolutionary models for COI and ITS2 sequences (Table 2) and *Aedes aegypti* was used as the outgroup (Guindon et al. 2010; Lanfear et al. 2012, 2016). MrBayes V 3.2.7 software

Table 2 Best scheme model of evolution for ITS2 & COI sequences

Gene	Fragment length	Coding positions	Best models of evolution
ITS2	333 bp	1–333	TRN + I
		2–333	TRN + I
		3–333	TRN + I
COI	342 bp	1–342	HKY + G
		2–342	F81
		3–342	HKY + G

Table 1 Number of ITS2 and COI sequences of *Anopheles stephensi* from each country

Country	Number of ITS2 sequences	Percentage (%)	Number of COI sequences	Percentage (%)
India	103	73.57	29	26.13
Iran	19	13.57	5	4.50
Sri Lanka	11	7.86	18	16.22
Kenya	3	2.15	–	–
Saudi Arabia	2	1.43	9	8.11
UK	1	0.71	6	5.40
USA	1	0.71	2	1.80
Canada	–	–	30	27.03
Pakistan	–	–	9	8.11
Brazil	–	–	1	0.90
United Arab Emirates	–	–	1	0.90
South Africa	–	–	1	0.90
Total	140	100	111	100

was used for Bayesian phylogenetic analysis (Ronquist et al. 2015). "The tree topology was edited and visualized using FigTree version 1.4.4" (Rambaut 2018).

Haplotype network construction

The regional distribution of *An. stephensi* haplotypes for the ITS2 and COI sequences was determined by the MJ approach. DNA sequence polymorphism analysis was conducted with DnaSP (v6.12.03) and Network 10.2 software (Rozas et al. 2017). Haplotype frequencies were calculated using Arlequin v3.5.2.2.

Genetic diversity analysis

Genetic diversity, including haplotype diversity, haplotype analysis, nucleotide diversity, and neutrality tests, was assessed for the COI and ITS2 sequences of *An. stephensi* using DnaSP v6.12.03 software. The analysis was performed by running a FAS alignment file in the software.

Determination of genetic differentiation

Arlequin v3.5.2.2 software was utilized to assess the genetic differentiation of *An. stephensi* populations by analyzing fixation index values for both COI and ITS2 sequences.

Results

Phylogenetic analysis

The phylogenetic tree of ITS2 sequences obtained using the Bayesian approach reveals that *An. stephensi* individuals were separated into two distinct clades (clade 1 and 2) with posterior probability (pp) 100% (Fig. 1, Supplementary Table 4). Each clade evolved independently with distinct genetic structures. Clade 1 consists of 38 Indian species, while Clade 2 includes 35 species from India, Iran, Kenya, the UK, and Saudi Arabia. Clade 1 has a longer branch length than Clade 2, suggesting greater divergence from its most recent common ancestor (MRCA). Both clades have a posterior probability of 100% (nodes P and Q), indicating statistically similar branching patterns.

The phylogenetic analysis of COI sequences reveals a distinct topology compared to the ITS2 sequences. The phylogenetic tree of COI sequences reveals the separation of *An. stephensi* individuals into six clades with a pp value of 100% in node P (Fig. 2, Supplementary Table 5). Six distinct clades were identified, each evolved independently and possessing a unique genetic structure. Clade 1 represents Indian and Sri Lankan species with a posterior probability of 87.65% at node Q. clade 2 includes species from Saudi Arabia with a pp value of 84.08% at node R. Clade 2 was further divided into two clades with a pp value of 86.92%. The clade 3 and 4 includes Indian species with a pp value of 90.29% and 87.57% at nodes S and T respectively. The clade 5 consists only 2 species from UK with pp value of 89% at node U. Clade 6 with species from Sri Lanka, India, Canada, USA, Pakistan and UK, was derived from the main branch. As the posterior probability value of clade 3 is close to 100% than clade 1,2,4, and 5, the clade 3 branching pattern can be considered statistically more trustworthy than others. The Clade 6 branch length exceeded all other branch lengths implying the species of clade 6 might have more genetic change or divergence than the species of all other clades from their MRCA. Similarly, the branch lengths of clades 1 and 2 were greater than the branch lengths of clades 3,4, and 5, which indicates that the species from clades 1 and 2 might have more divergence than species from clades 3,4, and 5.

Construction of haplotype network

A haplotype network was constructed using 73 ITS2 sequences of *An. stephensi* (Fig. 3). A total of four haplotypes were identified from India, Iran, Kenya, Saudi Arabia, and UK (Supplementary Table 6 and Supplementary Table 7). The haplotype H_2 was the only shared haplotype, while the other three were private haplotypes. India (Hd = 0.364) had the greatest number of private haplotypes (H_1 and H_4), and there are more sequences there than in other countries. There was only one private haplotype (H_3) from Iran. There were no private haplotypes from Kenya, Saudi Arabia and UK. Haplotype H_1 (India) and H_3 (Iran) were the most distant, with H_1 and H_2 showing the highest number of mutations. In contrast, H_2 and H_3 were the closest, with the fewest mutations. There were 5 mutations present in between H_2 and H_4 haplotypes. One median vector named mv1 was present in the haplotype network of ITS2 sequences.

The study identified nine haplotypes (H_1–H_9) from eight countries—India, Saudi Arabia, Sri Lanka, USA, Canada, UK, UAE, and Pakistan—based on 64 COI sequences (Fig. 4, Supplementary Table 6 and Supplementary Table 8). H_1 and H_3 were shared haplotypes, H_1: found in India and Sri Lanka. H_3: shared among India, Canada, Pakistan, Sri Lanka, USA, UK, and UAE. Private haplotypes were highest in India (H_2, H_5), Saudi Arabia (H_6, H_7), UK (H_8, H_9), USA had the fewest (H_4), Sri Lanka and Canada had none. The greatest genetic distance was between H_3–H_7 and H_4–H_6, with five mutations each.

Determination of genetic diversity

The genetic heterogeneity of *An. stephensi*, based on 73 ITS2 and 64 COI sequences, reveals low haplotype

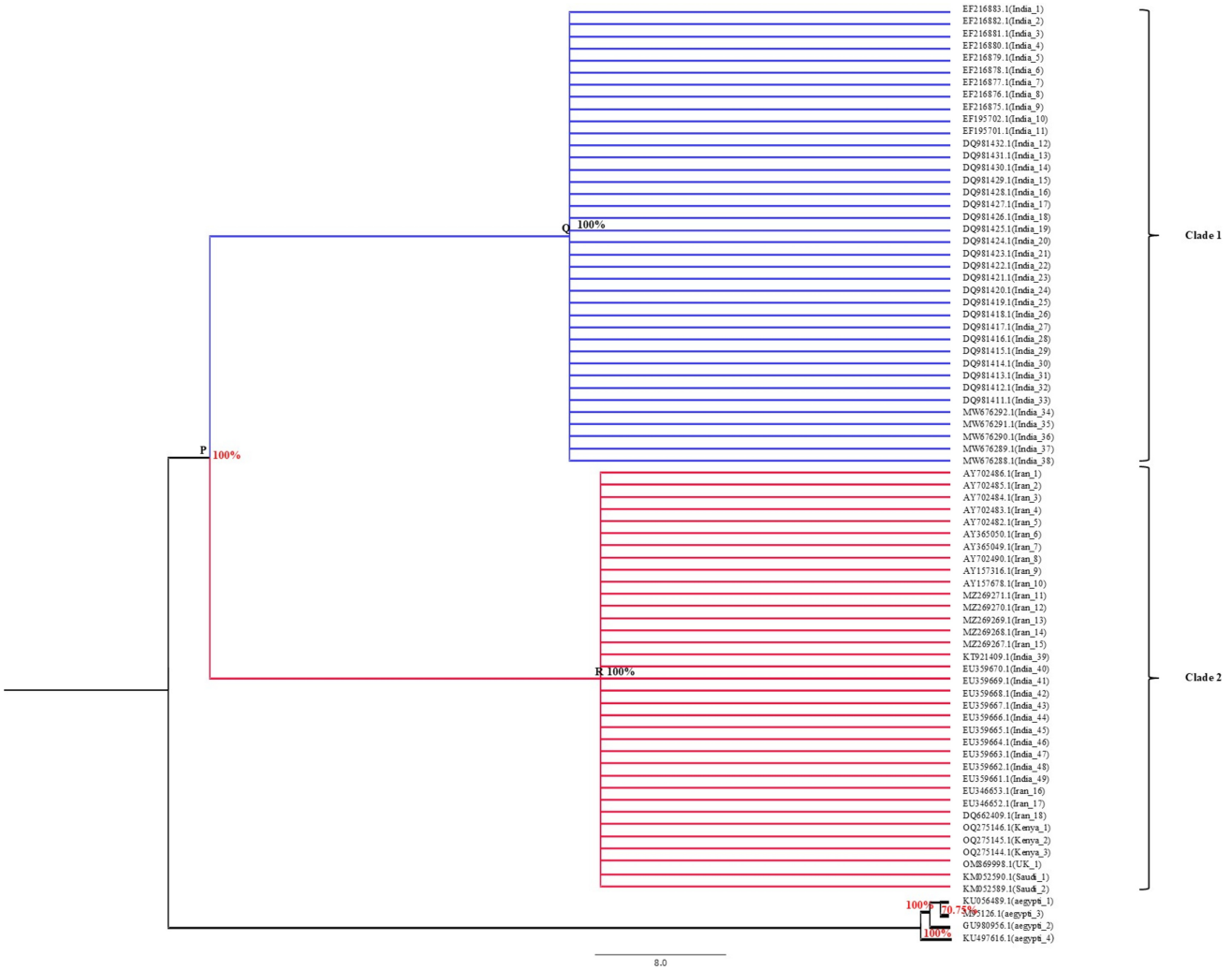


Fig. 1 Phylogenetic analysis based on ITS2 sequences of *Anopheles stephensi* species using Bayesian likelihood tree generated by MrBayes v 3.2.7_win 64 software: 77 sequences, 333 characters, 1,640,000 generations, 2 mcmc runs. Percentages on node labels are Bayesian posterior probabilities (pp). *Aedes aegypti* was used as the outgroup. GenBank accession numbers of ITS2 sequences included in clade 1, 2 and outgroup were listed in Supplementary Table 4

diversity (Hd) values of 0.532 and 0.435, respectively. The average number of nucleotide differences (k) and nucleotide diversity (π) are lower in COI sequences compared to ITS2 (Table 3).Tajima's D value is negative for COI sequences, suggesting an excess of rare mutations, while a positive value for ITS2 indicates fewer rare mutations, implying balancing selection. Fu and Li's D and F values are positive for both markers, with statistically significant *P*-values for ITS2 but non-significant *P*-values for COI. Fu's Fs statistic is positive for ITS2 but negative for COI sequences. The G + C content at coding positions is 56.7% across 333 sites for ITS2 and 32.1% across 342 sites for COI.

Determination of genetic differentiation

Table 4 presents the percentage of variation and variance components for ITS2 sequences, comparing population differences. The results indicate that genetic variation and variance components are greater between populations than within populations, suggesting a higher degree of differentiation among populations. The pairwise fixation index (F_{ST}) is used to measure genetic differentiation between populations. Table 5 presents the F_{ST} values for *An. stephensi* populations based on ITS2 sequences, indicating the degree of genetic divergence among them. F_{ST} levels often fall into one of the following ranges: 0–0.05 for modest, 0.05–0.15 for

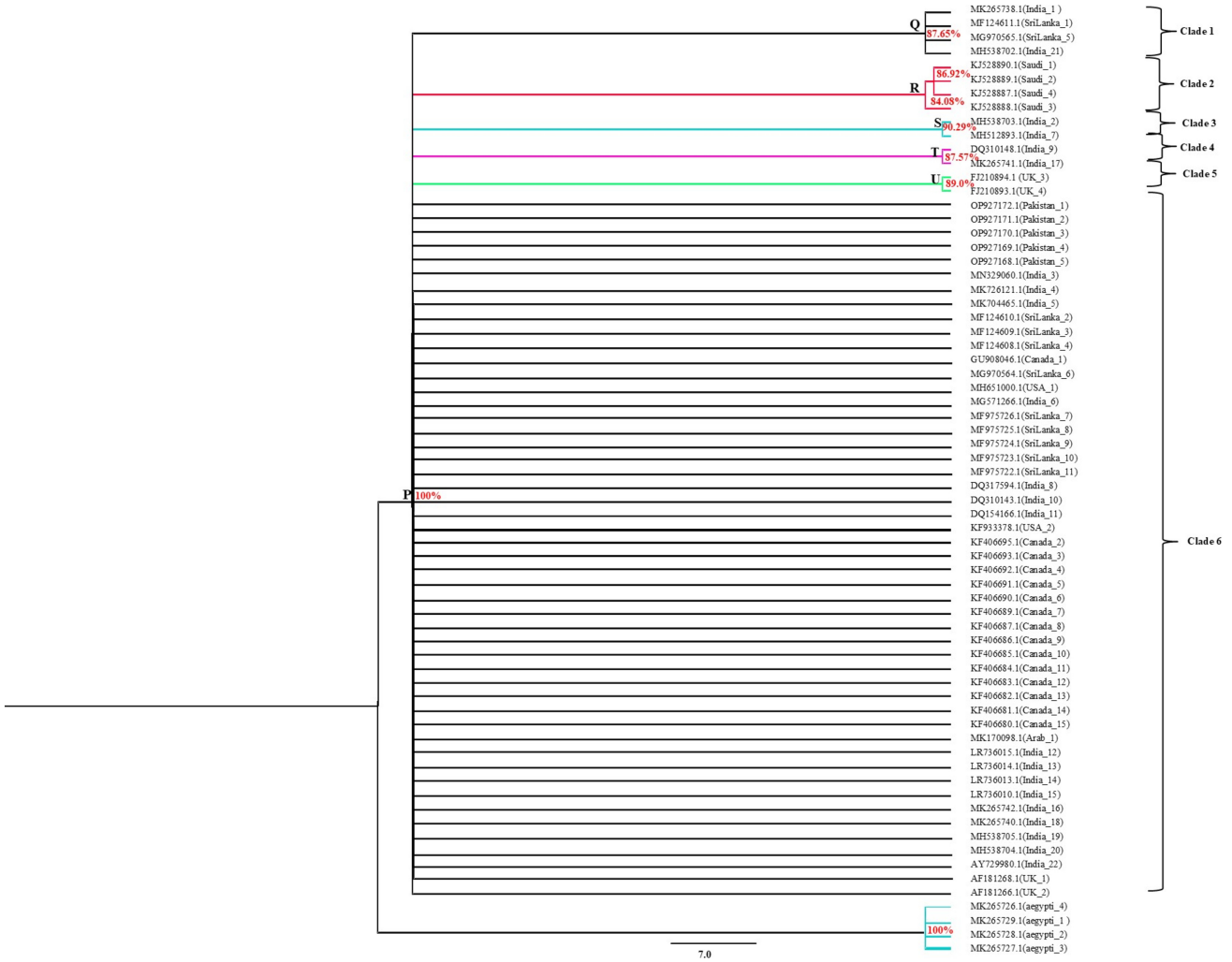


Fig. 2 Phylogenetic analysis based on COI sequences of *Anopheles stephensi* species using Bayesian tree generated by MrBayes v 3.2.7_win 64 software: 68 sequences, 342 characters, 1,640,000 generations, 2 mcmc runs. Percentage on node labels are Bayesian posterior probabilities (pp). *Aedes aegypti* was considered as the outgroup. GenBank accession numbers of COI sequences included in clade 1, 2, 3, 4, 5, 6 and outgroup were listed in Supplementary Table 5

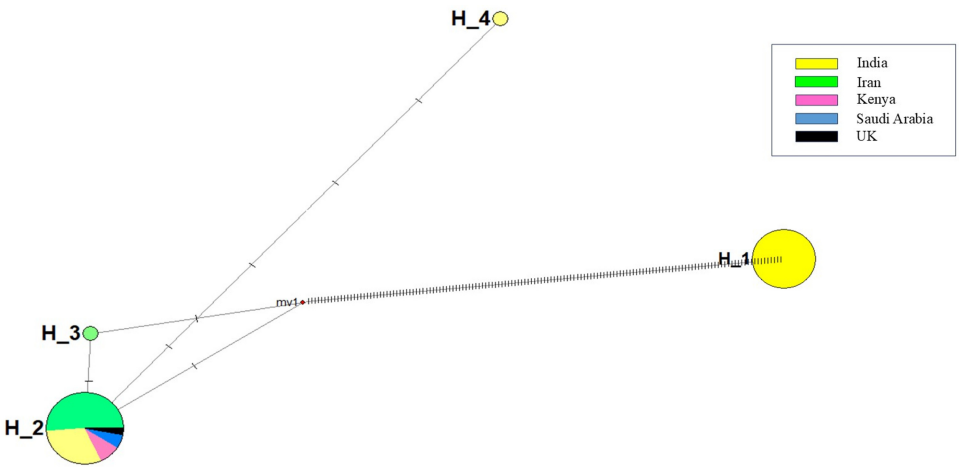


Fig. 3 Four haplotypes of the median-joining haplotype network of *Anopheles stephensi* species' ITS2 sequences were produced using Network 10.2 and DnaSP v6.12.03 software. The various geographical regions are indicated by the colour. The relative frequency of sequences associated with each haplotype is indicated by the size of a circle. The number of mutations is part of the distance between two haplotypes. Red coloured dot indicates median vector

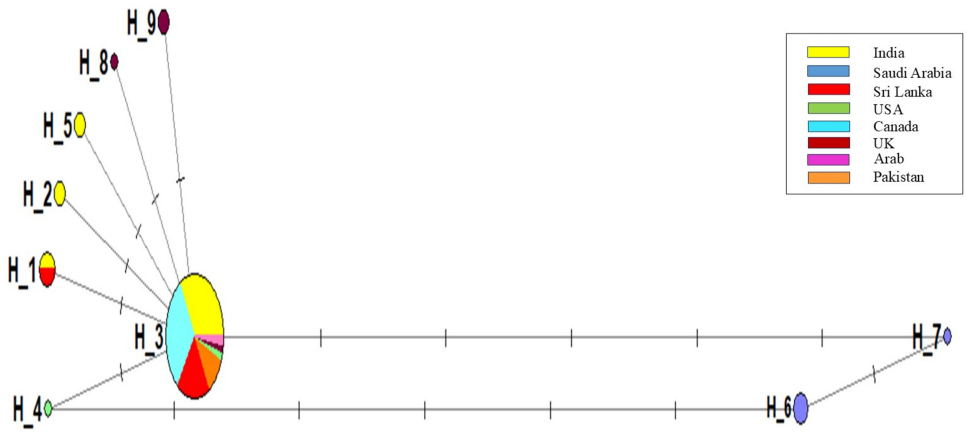


Fig. 4 Using DnaSP v6.12.03 and Network 10.2 software, a median-joining haplotype network of COI sequences of *Anopheles stephensi* species (9 haplotypes) was created. The size of a circle represents the relative frequency of sequences belonging to each haplotype, while the colour denotes various geographic regions

Table 3 Genetic diversity indices including sample size (x), number of variable sites (S), haplotypes (h), haplotype diversity (Hd), variance of Haplotype diversity (v), nucleotide diversity (π), the average number of nucleotide differences (k), Tajima's D, Fu and Li's D, Fu and Li's F, Fu's Fs statistics calculated using ITS2 and COI sequences of *Anopheles stephensi*

Sl. No	Genetic diversity indices	Region	
		ITS2	COI
	Sample size	73	64
	number of variable sites (S)	222	11
	Haplotypes (h)	4	9
	Haplotype diversity (Hd)	0.532	0.435
	Variance of Haplotype diversity (v)	0.00048	0.00595
	Nucleotide diversity (π)	0.33459	0.00307
	Average number of nucleotide differences (k)	111.419	1.049
	Tajima's D	4.91417* (P < 0.001)	−1.53298** (P > 0.10)
	Fu and Li's D	2.30009* (P < 0.02)	0.82841** (P > 0.10)
	Fu and Li's F	3.87999* (P < 0.02)	0.03122** (P > 0.10)
	Fu's Fs statistics	108.229	−3.454

*Significant P < 0.05

**Non-significant P > 0.10

Table 4 List of sums of squares, variance components and percentage of variation among populations and within populations of ITS2 sequences of *Anopheles stephensi* species

Source of variation	d.f	Sum of squares	Variance components	Percentage of variation
Among Populations	4	2131.798	56.96966 Va	67.34
Within Populations	68	1879.271	27.63634 Vb	32.66
Total	72	4011.068	84.60599	

Fixation Index F_{ST}: 0.67335

Table 5 Pairwise fixation index (F_{ST}) values among populations based on ITS2 sequences of *An. stephensi*

	India	Iran	Kenya	UK	Saudi Arabia
India	0.00000				
Iran	0.69234	0.00000			
Kenya	0.61312	−0.19531	0.00000		
UK	0.54150	−1.00000	0.00000	0.00000	
Saudi Arabia	0.59453	−0.33043	0.00000	0.00000	0.00000

Matrix of significant F_{ST} P values Significance Level = 0.0500

moderate, and 0.15–0.25 for great, levels above 0.25 suggest significant genetic difference.

Table 6. depicts the proportion of variations and variance components between and within populations of COI sequences, showing that the percentage of variation and variance components were higher between COI sequence populations than within populations. Table 7 displays the pairwise fixation index (F_{ST}) values, highlighting the genetic differentiation among *An.stephensi* populations based on COI sequences. The F_{ST} values were ranging from − 1.00000 to 0.98435.

Table 6 List of sum of squares, variance components and percentage of variation among populations and within populations of COI sequences of *Anopheles stephensi*

Source of variation	d.f. (Degree of Freedom)	Sum of squares	Variance components	Percentage of variation
Among Populations	7	22.956	0.43335 Va	70.63
Within Populations	56	10.091	0.18019 Vb	29.37
Total	63	33.047	0.61355	

Fixation Index FST: 0.70631

Table 7 Pairwise fixation index (F_{ST}) values among populations based on COI sequences of *An. stephensi*

	India	Canada	Pakistan	Sri Lanka	USA	Saudi Arabia	UK	United Arab Emirates
India	0.00000							
Canada	0.02295	0.00000						
Pakistan	– 0.07056	0.00000	0.00000					
Sri Lanka	– 0.01418	0.14063	– 0.00406	0.00000				
USA	0.19022	0.77612	0.47368	0.30229	0.00000			
Saudi Arabia	0.91423	0.98435	0.96259	0.93778	0.88539	0.00000		
UK	0.29282	0.57619	0.28571	0.33013	0.11111	0.87179	0.00000	
United Arab Emirates	– 0.90476	0.00000	0.00000	– 0.80000	– 1.00000	0.91304	– 0.55556	0.00000

Matrix of significant F_{ST} P values Significance Level = 0.0500

Discussion

Malaria vector surveillance, cryptic species identification, and mapping the population structure and gene flow of vectors across different regions are crucial for mitigating the vector-borne diseases. Understanding the dispersion and genetic diversity of malaria vectors may help design effective control strategies and monitor the spread of vectors. Molecular marker ITS2 and COI have been crucial for understanding Anopheles mosquitoes' genetic diversity, phylogenetics, and population structure. Askari et al. (2024) analyzed *An. hyrcanus* in northern Iran, identifying distinct haplotypes and potential subspecies differentiation. Rathnayake et al. (2023) examined the *An. culicifacies* species complex throughout Asia, revealed several clades with notable genetic heterogeneity. Carter et al. (2019) also employed ITS2 and COI sequences for anopheles species identification in eastern Ethiopia and reported that ITS2 effectively distinguished species, whereas COI struggled with closely related taxa in the *An. gambiae* complex. Zhang et al. (2021) studied the molecular phylogeny of *An. hyrcanus* based on the ITS2 and COI sequences and revealed phylogenetic relationship, evolutionary history, and population dynamics.

Phylogenetic tree based on ITS2 Bayesian analysis revealed two distinct clades (Clade 1 and Clade 2) of *An. stephensi*. Clade 1 consists of 38 sequences from India, while Clade 2 consists of 35 sequences reported from Saudi Arabia, Iran, Kenya, the UK, and India. Clade 1 may have undergone greater genetic divergence from their most recent common ancestor (MRCA) than Clade 2 because of the clade's longer branch length. The genetic and geographic distinction between species found in India and those found in a number of other countries is indicated by the clustering patterns of Clades 1 and 2. This could indicate potential evolutionary or ecological reasons for the genetic diversity among these populations. The topology of the COI phylogenetic tree, which separated *An. stephensi* individuals into six clades, was more intricate than that of the ITS2 tree. The 100% posterior probability values provided statistically strong support for each clade's distinct genetic structure at branching locations. While most of the species in Clade 1 are from India and Sri Lanka, Saudi Arabian species are part of Clade 2. A pp value of 86.92% in Clade 2 to further subdivided into two groups. The pp values of Clades 3 and 4's Indian species were significantly higher, at 87.57% and 90.29%, respectively. Clade 5 had a pp value of 89% and contained two species from the UK. Species from India, Sri Lanka, Canada, the United States, Pakistan, and the United Kingdom made formed Clade 6, the most genetically diverse group. Clade 6 had the greatest divergence, according to the branch length study, indicating significant genetic changes within its species in contrast to other clades. Furthermore, Clades 1 and 2 had longer branches than

Clades 3, 4, and 5, indicating a higher degree of species divergence within these clades. Clade 3 demonstrated the most statistically stable branching pattern among all the clades, highlighting the strength of its evolutionary links with a pp value that was closest to 100%.The ITS2 analysis clearly discriminated between Indian and non-Indian species, suggesting that geography significantly influences genetic organization. In contrast, the COI analysis revealed more complex relationships with clades—such as Clade 6, which included species from multiple continents—comprising species from different geographical regions. This suggests that the COI marker captures a wider range of genetic diversity, which may represent modern gene flow between populations and past biogeographical events. The ITS2 marker provides a complete picture of genetic differentiation than the COI.

The haplotype network analysis based on the COI and ITS2 sequences revealed that COI exhibits more haplotype networks than ITS2. The COI had two shared haplotypes (H_1 and H_3) but ITS2 had one shared haplotype (H_2) with H_3 showing a wide distribution across multiple countries.The COI network displayed greater complexity and longer branch lengths, indicating higher mutation rates or a longer evolutionary history of COI sequences. The existence of a median vector (mv1) in the ITS2 network increases the likelihood of missing intermediate sequences or ancestral forms, but the absence of median vectors in the COI network suggests a more clear representation of haplotype connections.The analysis shows low Hd values for ITS2 and COI (Supplementary Tables 7 and 8) in *An. stephensi*. The higher nucleotide diversity and polymorphic sites in ITS2 evidential of greater sequence variable suggests higher evolutionary dynamics, but COI has more haplotypes, and low nucleotide diversity may be indicative of recent mutation in the conserved region (Fernando et al. (2020).

Positive Tajima's D and Fu's F values of ITS2 show an absence of rare alleles, indicating a bottleneck or balancing selection. Negative Tajima's D and positive Fu's F values of COI suggest that COI evolves more neutrally, as expected from mitochondrial genes, which typically show lower polymorphism (Panhuis et al. 2006). The highest F_{ST} values of Saudi Arabia-India, Saudi Arabia-Sri Lanka Saudi Arabia-Pakistan, and Saudi Arabia-India in the COI study, as well as India-Iran and India-Kenya in the ITS2 analysis, indicated a significant genetic divergence across populations, according to the pairwise FST values. Given the great geographic separation between the two sites, a high level of genetic divergence might be attained (Mayoke et al. 2020). The genetic closeness between the populations reduced as the distance between them increased. This implies the spatially constrained gene flow that results in isolation by distance (Jensen et al. 2005).

Conclusion

The significance of utilizing *An. stephensi* species ITS2 and COI sequences were emphasized to comprehend mosquito vector genetic diversity and evolutionary linkages. Six clades were found in phylogenetic analysis using COI sequences, compared to just two in the ITS2. While the COI study showed more intricate associations with clades, like Clade 6, which comprised species from different continents. The ITS2 analysis distinguished Indian and non-Indian species. This implies that the COI marker captures a greater variety of genetic variability. The COI network showed longer distances and greater overall complexity, indicating that the COI may have undergone or experienced higher mutation rates. The low genetic diversity was observed in *An. stephensi* indicates this mosquito population likely has a relatively small effective population size over time. Thus, uniform control approaches may not be equally successful for all populations of a particular species, It is important to consider the genetic diversity of mosquito-borne disease vectors while implementing efficient vector control strategies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43538-025-00470-7.

Acknowledgements The corresponding authors thank the authorities of Berhampur University, Odisha, India, for providing the Seed fund grant (LNo.1722/R&Dcell/BU/2024, dt. 24.03.2024) with the necessary support and facilities. We acknowledge and thank all authors whose publicly available data deposited in GenBank have been utilized in this manuscript.

Author contributions All authors contributed to the conception and design of the work. LR: Data collection; Statistical interpretation, Writing draft manuscript; LKM: Conceptualization, critical analysis, and updating of the manuscript. TKB: Critical analysis and updating of the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of interests There is no conflict of interest among the authors for the publication.

Ethical approval Ethical approval was not required as this study involved the use of already published data.

References

Ahmed, A., Khogali, R., Elnour, M.B., Nakao, R., Salim, B.: Emergence of the invasive malaria vector *Anopheles stephensi* in Khartoum State. Central Sudan. Parasites Vectors **14**, 511 (2021). https://doi.org/10.1186/s13071-021-05026-4

Askari, F., Paksa, A., Shahabi, S., Saeedi, S., Sofizadeh, A., Vahedi, M., Soltani, A.: Population genetic structure and phylogenetic analysis of *Anopheles hyrcanus* (Diptera: Culicidae) inferred

from DNA sequences of nuclear ITS2 and the mitochondrial COI gene in the northern part of Iran. BMC Infect. Dis. **24**, 724 (2024). https://doi.org/10.1186/s12879-024-09626-0

Carter, T.E., Yared, S., Hansel, S., Lopez, K., Janies, D.: Sequence-based identification of *Anopheles* species in eastern Ethiopia. Malar. J. **18**, 135 (2019). https://doi.org/10.1186/s12936-019-2768-0

Djadid, D.N., Sanati, M.H., Zare, M., Hassanzehi, A.: rDNA-ITS2 Identification of *Anopheles pulcherrimus* (Diptera: Culicidae): Genetic differences and phylogenetic relation with other Iranian vectors and its implications for malaria control. Iran. Biomed. J. **7**(1), 1–6 (2003)

Dong, X.S., Zhou, H.N.: A catalog of the mosquito of the Yunnan, vol. 5. Yunnan Science and Technology Press, Kunming (2019)

Fang, Y., Shi, W.Q., Zhang, Y.: Molecular phylogeny of *Anopheles hyrcanus* group members based on ITS2 rDNA. Parasites Vectors. **10**, 417 (2017). https://doi.org/10.1186/s13071-017-2351-x

Fernando, H.S.D., Hapugoda, M., Perera, R., Black, W.C., Silva, B.G.D.N.K.D.: Mitochondrial metabolic genes provide phylogeographic relationships of global collections of *Aedes aegypti* (*Diptera*: Culicidae). PLoS One. **15**(7), e0235430 (2020). https://doi.org/10.1371/journal.pone.0235430

Geneious 2022.1.1. https://www.geneious.com.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O.: New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 30. Syst. Biol. **59**(3), 307–321 (2010). https://doi.org/10.1093/sysbio/syq010

Hu, Y., Fan, H., Chen, Y., Chang, J., Zhan, X., Wu, H., Zhang, B., Wang, M., Zhang, W., Yang, L., Hou, X., Shen, X., Pan, T., Wu, W., Li, J., Hu, H., Wei, F.: Spatial patterns and conservation of genetic and phylogenetic diversity of wildlife in China. Sci. Adv. **7**(4), eabd725 (2021). https://doi.org/10.1126/sciadv.abd5725

Jensen, J.L., Bohonak, A.J., Kelley, S.T.: Isolation by distance, web service. BMC Genom. Data. **6**, 13 (2005). https://doi.org/10.1186/1471-2156-6-13

Karunaweera, N.D., Galappaththy, G.N., Wirth, D.F.: On the road to eliminate malaria in Sri Lanka: Lessons from history, challenges, gaps in knowledge and research needs. Malar. J. **13**(59), 1–10 (2014). https://doi.org/10.1186/1475-2875-13-59

Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S.: PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. **29**(6), 1695–1701 (2012). https://doi.org/10.1093/molbev/mss020

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B.: PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. **34**(3), 772–773 (2016). https://doi.org/10.1093/molbev/msw260

Loaiza, J.R., Scott, M.E., Bermingham, E., Sanjur, O.I., Rovira, J.R., Dutari, L.C., Linton, Y.M., Bickersmith, S., Conn, J.E.: Novel genetic diversity within Anopheles punctimacula s.l.: Phylogenetic discrepancy between the Barcode cytochrome c oxidase I (COI) gene and the rDNA second internal transcribed spacer (ITS2). Acta Trop. **128**(1), 61–69 (2013). https://doi.org/10.1016/j.actatropica.2013.06.012

Mayoke, A., Muya, S.M., Bateta, R., Mireji, P.O., Okoth, S.O., Onyoyo, S.G., Auma, J.E., Ouma, J.O.: Genetic diversity and phylogenetic relationships of tsetse flies of the palpalis group in Congo Brazzaville based on mitochondrial COX1 gene sequences. Parasites Vectors. **13**, 253 (2020). https://doi.org/10.1186/s13071-020-04120-3

Mopuri, R., Mutheneni, S.R., Kumaraswamy, S., Kadiri, M.R., Upadhyayula, S.M., Naish, S.: An epidemiological and spatiotemporal analysis to identify high risk areas of malaria in Visakhapatnam district of Andhra Pradesh, India, 1999–2015. Spat. Inf. Res. **27**(6), 659–672 (2019). https://doi.org/10.1007/s41324-019-00267-z

Muthukumarana, S.L., Wedage, M.M., Rathnayake, S., De Silva, K.N.: Phylogenetic, phylogeographic and divergence time analysis of *Anopheles subpictus* species complex using ITS2 and COI sequences. Asian Pac J Trop Med **17**(5), 214–225 (2024). https://doi.org/10.4103/apjtm.apjtm_790_23

Panhuis, T.M., Swanson, W.J.: Molecular evolution and population genetic analysis of candidate female reproductive genes in *Drosophila*. Genetics **173**(4), 2039–2047 (2006). https://doi.org/10.1534/genetics.105.053611

Rambaut, A.: FigTree ver 1.4.4. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh (2018).

Rathnayake, R.A.S., Wedage, W.M.M., Muthukumarana, L.S., De Silva, B.G.D.N.K.: Genetic diversity, phylogenetic and phylogeographic analysis of *Anopheles culicifacies* species complex using ITS2 and COI sequences. PLoS One. **18**(8), e02901782023 (2023). https://doi.org/10.1371/journal.pone.0290178

Ronquist, F., Teslenko, M., Mark, P.V.D., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P.: MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. **61**(3), 539–542 (2012). https://doi.org/10.1093/sysbio/sys029

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sánchez-Gracia, A.: DnaSP v6: DNA sequence polymorphism analysis of large datasets. Mol. Biol. Evol. **34**, 3299–3302 (2017). https://doi.org/10.1093/molbev/msx248

Ryan, S.J., Lippi, C.A., Villena, O.C., Singh, A., Murdock, C.C., Johnson, L.R.: Mapping current and future thermal limits to suitability for malaria transmission by the invasive mosquito *Anopheles stephensi*. Malar. J. **22**(1), 104 (2023). https://doi.org/10.1186/s12936-023-04531-4

Sallum, M.A.M., Schultz, T.R., Foster, P.G., Aronstein, K., Wirtz, R.A., Wilkerson, R.C.: Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences. Syst. Entomol. **27**(3), 361–382 (2002). https://doi.org/10.1046/j.1365-3113.2002.00182.x

Singh, D., Rai, M.K., Vashist, D.: Molecular studies on Indian mosquitoes (Diptera: Culicidae): A review. Int. J. Entomol. Res. **3**(2), 96–103 (2018)

Sum, J.S., Lee, W.C., Amir, A., Braima, K.A., Jeffery, J., Abdul-Aziz, N.M., Fong, M.Y., Lau, Y.L.: Phylogenetic study of six species of Anopheles mosquitoes in Peninsular Malaysia based on inter-transcribed spacer region 2 (ITS2) of ribosomal DNA. Parasites Vectors. **7**, 309 (2014). https://doi.org/10.1186/1756-3305-7-309

Tadesse, F.G., Ashine, T., Teka, H., Esayas, E., Messenger, L.A., Chali, W., Meerstein-Kessel, L., Walker, T., Wolde, B.S., Lanke, K., Heutink, R., Jeffries, C.L., Mekonnen, D.A., Hailemeskel, E., Tebeje, S.K., Tafesse, T., Gashaw, A., Tsegaye, T., Emiru, T., Simon, K., Bogale, E.A., Yohannes, G., Kedir, S., Shumie, G., Sabir, S.A., Mumba, P., Dengela, D., Kolaczinski, J.H., Wilson, A., Churcher, T.S., Chibsa, S., Murphy, M., Balkew, M., Irish, S., Drakeley, C., Gadisa, E., Baousema, T.: *Anopheles stephensi* mosquitoes as Vectors of *Plasmodium vivax* and *falciparum*, Horn of Africa, 2019. Emerg. Infect. Dis. **27**(2), 603–607 (2021). https://doi.org/10.3201/eid2702.200019

Tamura, K., Stecher, G., Kumar, S.: MEGA11: Molecular Evolutionary Genetics Analysis version 11. Mol. Biol. Evol. **38**(7), 3022–3027 (2021). https://doi.org/10.1093/molbev/msab120

Tyagi, B.K. (2023) *Anopheles stephensi* Liston 1901: Origin and Chorogeography—A New Hypothesis. In: Desert Malaria.

Springer, Singapore. https://doi.org/10.1007/978-981-19-7693-3_10.

Venkatesan, P.: The 2023 WHO World Malaria Report. Lancet Microbe. **5**(3), e214 (2024). https://doi.org/10.1016/S2666-5247(24)00016-8

Walton, C., Somboon, P., O'Loughlin, S.M., Zhang, S., Harbach, R.E., Linton, Y.M., Chen, B., Nolan, K., Doung, S., Fong, M.Y., Vythilingum, I., Mohammed, Z.D., Trung, H.D., Butlin, R.K.: Genetic diversity and molecular identification of mosquito species in the *Anopheles maculatus* group using the ITS2 region of rDNA. Infect. Genet. Evol. **7**(1), 93–102 (2007). https://doi.org/10.1016/j.meegid.2006.05.001

World Health Organization. World Malaria Report:2024. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2024(2024) (Accessed 25 Jan 2025).

Wu, J., Li, D., Boyd, B., Balan, R.K., George, S., Peacock, L., Pal, C.: Comparative performance of a multi-locus barcoding approach to enhance taxonomic resolution of New Zealand mosquitoes (Diptera: Culicidae). Austral. Entomol. **62**(1), 77–95 (2023). https://doi.org/10.1111/aen.12630

Zhang, C., Yang, R., Wu, L., Luo, C., Guo, X., Deng, Y., Zhou, H., Zhang, Y.: Molecular phylogeny of the *Anopheles hyrcanus* group (Diptera: Culicidae) based on rDNA–ITS2 and mtDNA–COII. Parasit. Vectors **14**, 454 (2021). https://doi.org/10.1186/s13071-021-04971-4

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.