

Climatic and Ecological Gradients Associated with MSP1 Genetic Diversity in *Plasmodium falciparum*: An Entropy-Based Population Analysis

Joseph O. Adebayo

jadebayo@unilorin.edu.ng

University of Ilorin

Oche A. George

University of Ilorin

Dorcas I. Adebayo

Anchor University Lagos

Evans C. Egwim

Federal University of Technology Minna

Simeon A. Adebisi

Benue State University

Caroline A. Okoli

University of Jos

Adedoyin Igunnu

University of Ilorin

Oluwatosin A. Adaramoye

University of Ibadan

Israel K. Olayemi

Federal University of Technology Minna

Olusola Ajibaye

Nigerian Institute of Medical Research

Oluyinka Iyiola

University of Ilorin

Moses Okpeku

University of KwaZulu-Natal

Article

Keywords: *Plasmodium falciparum*, Genetic diversity, Climatic gradients, MSP1 gene and Shannon entropy

Posted Date: July 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10158013/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Environmental heterogeneity may influence malaria transmission dynamics, yet its relationship with parasite population genetic variation within endemic regions remains incompletely understood. This study investigated whether climatic and ecological gradients are associated with measurable differences in *Plasmodium falciparum* genetic diversity across environmentally distinct regions of Vietnam using the merozoite surface protein 1 (*msp1*) marker.

Methods

A total of 145 publicly available *msp1* sequences retrieved from GenBank were analyzed across populations stratified by climatic zone, dry season duration, elevation, rainfall, humidity, and temperature. Genetic diversity was evaluated using allelic richness, heterozygosity, Shannon entropy-based partitioning, Nei's genetic distance, and principal coordinates analysis (PCoA). Diversity partitioning was used to quantify within- and among-population components of variation across environmental gradients.

Results

Across all environmental stratifications, most genetic variation was retained within populations (65–89%), whereas among-population differentiation remained modest. Elevation and rainfall exhibited comparatively larger entropy partitioning components, although overall differentiation remained low. In contrast, humidity and temperature showed minimal partitioning. Genetic distances were consistently low (Nei's $D \leq 0.008$), while genetic identities were high ($I \geq 0.992$), indicating substantial allelic overlap consistent with limited population subdivision. Populations from higher rainfall and mid-elevation zones showed relatively greater allelic richness, heterozygosity, and polymorphic loci, whereas other ecological groupings exhibited comparatively lower diversity and tighter clustering in ordination space.

Conclusion

Climatic gradients were associated primarily with variation in diversity magnitude rather than strong genetic subdivision among *P. falciparum* populations in Vietnam. Overall, parasite populations remained largely cohesive across ecological boundaries despite modest environmental differences in allelic diversity. Because analyses were based on the immune-selected *msp1* locus, findings should be interpreted as locus-specific patterns rather than evidence of genome-wide population structure.

1.0 INTRODUCTION

Malaria remains a major global health challenge, with *Plasmodium falciparum* responsible for the most severe morbidity and mortality. In 2022, an estimated 247 million malaria cases were reported worldwide, with the highest burden concentrated in tropical regions [1]. Although sustained investments in vector control, antimalarial therapies, and vaccine development have reduced transmission in several endemic areas, progress toward elimination remains uneven and vulnerable to ecological variability and parasite adaptability [2].

Environmental and climatic factors are fundamental determinants of malaria transmission intensity. Temperature, rainfall, humidity, and elevation influence mosquito abundance, vector survival, and the duration of parasite development within *Anopheles* mosquitoes [3–5]. Stable rainfall and moderate temperatures can sustain perennial transmission, whereas prolonged dry seasons or thermal extremes may reduce vector density and impose demographic bottlenecks [6, 7]. While numerous epidemiological studies have documented how climatic variability shapes malaria incidence, fewer investigations have examined how such ecological gradients relate to measurable differences in parasite genetic diversity within endemic regions.

Genetic diversity is a central component of malaria transmission dynamics. High allelic variation is typically associated with intense transmission, frequent recombination, and large effective population sizes, whereas reduced diversity may indicate bottlenecks, selective sweeps, or ecological constraints [8, 9]. Among the molecular markers used to assess parasite diversity, the merozoite surface protein 1 (*msp1*) gene remains one of the most extensively characterized. MSP1 exhibits high polymorphism, is under immune selection, and has been widely used to estimate transmission intensity, multiplicity of infection, and population connectivity across endemic settings [10]. Although whole-genome approaches now provide deeper resolution, single-locus markers such as *msp1* may still provide useful insight into allelic richness, heterozygosity, and diversity patterns in moderate-transmission settings, particularly where large-scale genomic resources are limited. However, because MSP1 is under balancing immune selection, variation at this locus may not fully reflect genome-wide neutral population structure.

Vietnam represents a geographically and climatically heterogeneous malaria-endemic country, encompassing tropical savanna, monsoon, and wet-dry climatic regimes, as well as marked gradients in elevation and seasonal rainfall [11]. These ecological contrasts provide a natural framework for

evaluating whether parasite populations exhibit measurable diversity differences across environmental boundaries within a single national setting [12]. However, systematic analyses linking fine-scale climatic variation to parasite genetic diversity within Vietnam remain limited.

Entropy-based diversity metrics provide an information-theoretic framework for quantifying allelic variability and overlap. Unlike classical fixation indices, entropy approaches emphasize proportional diversity partitioning and effective diversity estimates, thereby complementing conventional population genetic statistics rather than replacing them. Such approaches may be particularly useful for evaluating highly polymorphic loci where conventional differentiation statistics can become difficult to interpret.

In this study, we investigated whether climatic and ecological gradients are associated with differences in *P. falciparum* genetic diversity across multiple regions of Vietnam. Using *msp1* sequence data and Shannon entropy-based diversity partitioning, complemented by Nei's genetic distance and principal coordinates analysis, we assessed: (i) the distribution of genetic diversity within and among environmentally stratified populations; (ii) the extent of genetic overlap across climatic gradients; and (iii) whether specific environmental variables (temperature, rainfall, humidity, elevation, and dry season duration) are associated with detectable differences in diversity partitioning.

By explicitly quantifying both within- and among-population components of diversity, this study aimed to evaluate whether ecological heterogeneity is associated with measurable diversity differences or whether parasite populations remain largely cohesive across climatic zones. Understanding these relationships may provide useful context for interpreting malaria transmission dynamics and supporting climate-informed surveillance strategies in endemic settings. The selected provinces span both highland and coastal environments, providing natural climatic and ecological gradients for comparative analysis.

2.0 METHODOLOGY

2.1 Study Design and Sampling Framework

This study employed a retrospective population genetic design using publicly available *Plasmodium falciparum* merozoite surface protein 1 (*msp1*) coding sequences (CDS) retrieved from the NCBI GenBank database. The objective was to evaluate whether climatic and ecological gradients are associated with measurable differences in parasite genetic diversity within Vietnam.

A structured database query was conducted using the search string:

("Merozoite Surface Protein 1" OR "MSP1" OR "MSP-1") AND "Plasmodium falciparum"[Organism].

The initial search yielded 8,160 nucleotide records. Restriction to *P. falciparum* sequences returned 3,912 entries (3,756 genomic DNA/RNA and 156 mRNA sequences). No primary laboratory gene sequencing was performed as part of the present study. This study was a retrospective, secondary population-genetic analysis based on publicly available *Plasmodium falciparum msp1* sequences retrieved from NCBI GenBank. The sequence records were used for comparative genetic diversity, haplotype, and population-structure analyses. Therefore, no novel *msp1* sequence was generated, and no new sequence accession number was required for this study.

2.2 Inclusion and Exclusion Criteria

To ensure comparability and minimize bias, a priori inclusion criteria were applied:

1. Sequences must correspond to confirmed *msp1* coding or sub-coding regions.
2. Sequences must originate from field isolates (laboratory-adapted, vaccine, or reference strains were excluded).
3. Samples must fall within a contiguous five-year collection window to reduce long-term temporal confounding.
4. Sequence length must represent $\geq 80\%$ of the target polymorphic region.
5. Ambiguous bases (IUPAC codes) must constitute $\leq 2\%$ of sequence length after trimming.
6. Duplicate accession numbers and identical sequence entries were removed.

After filtering, 145 unique *msp1* sequences met inclusion criteria and formed the final analytical dataset.

2.3 Geographic and Climatic Stratification

The study sites correspond to provinces located within central and south-central Vietnam. Dak refers to Dak Lak Province in the Central Highlands region, Gia refers to Gia Lai Province in the northern Central Highlands, and Phu refers to Phu Yen Province along the south-central coastal zone. The selected provinces span both highland and coastal environments, providing natural climatic and ecological gradients for comparative analysis.

Sequences were grouped into three geographically distinct regions within Vietnam:

- **Dak (n = 40)** – Tropical savanna (Aw)
- **Gia (n = 22)** – Tropical wet-dry (Aw)
- **Phu (n = 83)** – Tropical monsoon (Am)

These sample sizes are reported explicitly to clarify population representation per site.

For each isolate, metadata were extracted from GenBank records and associated publications, including collection location and year. When available, geocoordinates were validated using WorldClim v2.1 gridded climate normals (30 arc-second resolution). Climatic variables assigned to each location included:

- Mean annual temperature
- Annual precipitation
- Dry season duration
- Relative humidity
- Elevation

Climatic classification followed the Köppen–Geiger system. This stratification enabled environmental grouping independent of genetic analysis, reducing circular inference. Environmental categories were derived from the distribution of climatic values associated with the sampled locations rather than predefined ecological thresholds. Stratification was performed to preserve biologically meaningful contrasts among locations while ensuring sufficient sample representation within each category. Consequently, elevation, rainfall, humidity, temperature, and dry-season classes should be interpreted as study-specific environmental groupings rather than universally applicable ecological boundaries.

2.4 Sequence curation, deduplication, and codon-aware alignment

Following sequence retrieval, preprocessing and quality control were performed to ensure analytical consistency. Duplicate or redundant entries corresponding to the same isolate and sampling period were identified through sequence identity comparison and collapsed, retaining the longest high-quality representative sequence. Gene identity was verified using BLAST alignment against the *Plasmodium falciparum* MSP1 reference sequence (3D7 locus PF3D7_0930300) to exclude non-specific or misannotated records.

Sequences were screened for low-quality regions, excessive ambiguity, premature stop codons, and disruptive frameshifts. Sequences containing more than 2% ambiguous nucleotides were excluded, whereas sequences with $\leq 2\%$ ambiguity were retained with masked ambiguous positions. To minimize alignment artifacts associated with hypervariable regions, all sequences were trimmed to a standardized MSP1 coding segment relative to the 3D7 reference coordinates.

Codon-aware multiple sequence alignment was performed using MAFFT v7 with the G-INS-i algorithm [13]. The resulting alignment was subsequently refined using MACSE v2 [14] to preserve codon integrity and correct potential reading-frame inconsistencies. Alignment columns containing more than 30% gaps were excluded from downstream comparative analyses to reduce bias associated with poorly aligned regions.

To assess analytical robustness, a sensitivity alignment was additionally generated using PAL2NAL [15], which converts amino acid alignments into corresponding codon-based nucleotide alignments. This procedure was used solely to confirm that major diversity and population structure estimates were not dependent on alignment strategy. All software versions, alignment parameters, and processing scripts were documented to ensure reproducibility.

2.5 Summary diversity and population structure

2.5.1 Classical indices

Aligned haploid genotypes were converted to GenAIEx input format [16] to quantify fundamental population-genetic metrics, including the number of alleles (N_a), effective alleles (N_e), Shannon's information index (I), Nei's expected heterozygosity for haploids (H) [17], unbiased H (H_u), and the percentage of polymorphic sites. Principal Coordinates Analysis (PCoA) based on Nei's genetic distances was used to visualize clustering patterns by geographical region and ecological category. Cross-validation was carried out in R using the packages *adeigenet* [18], *pegas* [19], and *hierfstat* [20] to ensure concordance across computational environments.

2.5.2 Entropy-based partitioning

To complement classical diversity indices, diversity partitioning was also examined through information-theoretic measures of Shannon entropy following Sherwin [21] and Ma et al. [22]. The total genomic entropy (sH_γ) was decomposed into within-population (sH_α) and between-population (sH_β) components, allowing estimation of effective diversity as $\exp(sH)$ and of scaled differentiation (D') and genetic overlap ($O' = 1 - D'$). Statistical significance of among-group differentiation was assessed using 999 stratified permutations implemented in the *vegan* package [23].

2.5.3 Haplotype and phylogenetic analysis

Haplotypes were reconstructed in *pegas* [19], and summary statistics including haplotype diversity (H_d), number of segregating sites (S), Watterson's θ , and nucleotide diversity (π) were computed in DnaSP v6 [24] and cross-checked using PopGenome [25]. Haplotype relationships were visualized as median-joining networks generated with PopART [26]. Phylogenetic inference was performed using IQ-TREE 2 [27], which incorporates ultrafast

bootstrap approximation and automated model selection through ModelFinder [28]. The resulting trees provided insight into lineage structure and potential clustering of haplotypes by ecological zone.

2.5.4 Genetic differentiation and gene flow

Population subdivision was quantified using pairwise and global F_{ST} (Weir–Cockerham) and Jost's D statistics. Because *P. falciparum* is haploid during its asexual blood stages, F_{ST} values were converted to the number of migrants (N_m) under the haploid island model approximation, $N_m \approx (1 - F_{ST})/(2 \times F_{ST})$. For comparability with historical diploid frameworks, supplementary analyses additionally included N_m estimates derived using the diploid formulation, $N_m \approx (1 - F_{ST})/(4 \times F_{ST})$.

Because MSP1 is highly polymorphic and several pairwise F_{ST} estimates approached zero, inferred N_m values should be interpreted cautiously as theoretical indicators of low genetic differentiation rather than direct estimates of migration intensity. Extremely large N_m values may arise mathematically from near-zero differentiation at highly variable loci and therefore may not reflect actual biological migration rates.

Standardized G_{ST} measures were calculated to account for high within-population diversity, and 95% confidence intervals were estimated from 1,000 bootstrap replicates using the *hierfstat* package [20].

2.6 Neutrality, recombination, and selection

Tests of neutrality were conducted to assess deviations from mutation–drift equilibrium. Tajima's D , Fu & Li's D^* and F^* , and Fu's F_s were computed both per region and for pooled datasets using DnaSP [24] and PopGenome [25]. To account for possible recombination, we evaluated linkage breakpoints through PHI (pairwise homoplasy index) and GARD (genetic algorithm for recombination detection) [29]; recombinant segments were either excluded or analyzed separately.

Codon-based tests of selection were implemented in HyPhy using FEL and FUBAR models for pervasive selection [30] and MEME for episodic selection [31]. Gene-wide ω (dN/dS) ratios were estimated using PAML [32] under M0, M1a/M2a, and M7/M8 site models, with likelihood ratio tests determining significance. To control false discovery, Benjamini–Hochberg correction at $q = 0.10$ was applied to site-wise tests.

2.7 AMOVA and hierarchical structure

An analysis of molecular variance (AMOVA) was performed with 999 permutations to partition total genetic variance into hierarchical components: among climatic regions, among populations within regions, and within populations. The procedure was implemented in GenAlEx [16] and cross-checked in Arlequin v3.5 [33] for consistency of Φ -statistics (Φ_{RT} , Φ_{PR} , Φ_{PT}). These metrics provided insight into the hierarchical organization of genetic variation and its ecological correlates.

2.8 Isolation by distance and environment

To investigate the influence of spatial and environmental factors on genetic differentiation, great-circle geographic distances among sampling sites were computed using their latitude–longitude coordinates. Pairwise Mantel and partial-Mantel tests were then applied to correlate genetic and environmental distance matrices while controlling for geography, implemented via *ade4* [34] and *vegan* [23].

Environmental distance matrices were constructed from standardized variables, including mean annual temperature, total precipitation, relative humidity, elevation, and dry-season length. Additionally, a distance-based redundancy analysis (db-RDA) using the *vegan* function *capscale* quantified the proportion of genetic variance explained by ecological predictors. Model optimization relied on the corrected Akaike Information Criterion (AICc) and permutation testing with $n = 999$ to determine significance.

2.9 Ecological stratification and comparative framework

For comparative ecological analyses, sequences were stratified according to climatic type (savanna, wet–dry, monsoon), elevation tertiles, rainfall class, humidity level, and dry-season duration. For each grouping, within-group (α) and between-group (β) diversity components were computed alongside classical indices such as N_a , N_e , H , H_d , π , and θ . Pairwise F_{ST} and Jost's D were used to quantify differentiation across ecological categories.

To account for differences in sample size among strata, we employed coverage-based rarefaction and extrapolation implemented in the *iNEXT* package [35]. Each metric was recalculated on rarefied datasets with 1,000 stratified bootstrap replicates, allowing direct comparison of diversity across ecological gradients while minimizing sampling bias.

2.10 Reproducibility, software, and data management

All computational workflows—from sequence retrieval to final visualization—were scripted in R (v4.3+) and bash to ensure full reproducibility. Version control was achieved through YAML-pinned configurations and fixed random seeds. The analytical suite included BioEdit v7.2.5 for initial visual inspection [36], MAFFT v7 for multiple sequence alignment [13], MACSE v2 for codon-aware correction [14], PAL2NAL for alignment translation [15], IQ-TREE 2 for phylogenetic reconstruction [27], DnaSP v6 for polymorphism statistics [24], Arlequin v3.5 for AMOVA [33], HyPhy v2.5 and PAML 4.10 for selection analysis [30–32], PopART for haplotype networking [26], and R packages *adegenet* [18], *pegas* [19], *hierfstat* [20], *vegan* [23], *ade4* [34], *PopGenome* [25], and *iNEXT* [35].

.The raw sequence dataset, accession identifiers, population labels, ecological classifications, polymorphic nucleotide positions, haplotype assignments, and derived analytical outputs are provided in **Supplementary File 2**. The study exclusively analysed publicly available genetic sequence data retrieved from NCBI GenBank and did not involve new human participant recruitment, experimental animals, or newly generated laboratory sequencing data.

2.11 Sensitivity and robustness analyses

To evaluate the stability of our results, we re-computed population-structure metrics using the PAL2NAL-derived alignment [15] and repeated F_{ST} and AMOVA analyses after excluding any sequences flagged as recombinant by GARD [29]. Diversity indices were re-estimated following sample-size rarefaction, and the entire analysis was repeated on a narrower MSP1 core segment defined. For sensitivity analysis, diversity and population-structure metrics were recalculated using a conserved MSP1 core segment spanning nucleotide positions 1–45 relative to the *P. falciparum* 3D7 reference sequence (PF3D7_0930300). This region was selected because it was represented in all sequences after trimming and contained minimal alignment ambiguity, thereby reducing the influence of missing data and highly variable terminal regions. Across all sensitivity checks, principal results remained consistent, demonstrating robustness of inferred diversity and selection patterns to analytical choices.

2.12 Justification for MSP1 Marker Selection

The *msp1* gene was selected because:

- It exhibits high allelic polymorphism.
- It is under immune selection and reflects transmission intensity.
- It has been extensively validated as a marker for estimating parasite diversity and multiplicity of infection [8–10].

Although single-locus analyses cannot capture full genome-wide structure, *msp1* remains informative for detecting differences in allelic richness, heterozygosity, and population connectivity in moderate-transmission settings.

3.0 RESULTS

3.1 Genetic Structure Across Climatic and Environmental Gradients

Pairwise Shannon entropy partitioning and inferred gene flow estimates were evaluated across climatic region, dry season duration, elevation, humidity, rainfall, and temperature categories (Table 1). Across all environmental comparisons, within-population diversity (sHu) exceeded among-population entropy ($sHua$).

Among climatic zones, tropical savanna versus tropical wet–dry populations showed $sHu = 0.143$ and $sHua = 0.017$, while tropical wet–dry versus tropical monsoon populations showed $sHu = 0.078$ and $sHua = 0.018$ (Table 1). Comparisons within monsoon populations showed low differentiation values ($sHua = 0.003$). Several comparisons yielded very high inferred Nm estimates because differentiation values approached zero.

For dry season duration, the November–April category showed higher within-population diversity than the January–August category, whereas among-population entropy remained low (Table 1).

Elevation comparisons showed comparatively larger among-population entropy components than other environmental gradients. The 400–800 m category showed $sHua$ values ranging from 0.029 to 0.033, whereas the 650 m versus coastal/lowland comparison showed $sHua = 0.001$ (Table 1).

Humidity and temperature categories exhibited low among-population entropy values across comparisons. Rainfall categories showed intermediate partitioning values, with $sHua$ ranging from 0.015 to 0.018 in comparisons involving the 1800–2000 mm rainfall category (Table 1).

Table 1
Pairwise Shannon Diversity and Gene Flow Estimates Across Environmental Gradients

Environmental Factor	Comparison	sHu	sHua	e ^s Hua	Nm
Climatic Region	Savanna vs Wet–Dry	0.143	0.017	1.017	81.008
	Wet–Dry vs Monsoon	0.078	0.018	1.018	72.153
	Monsoon vs Monsoon	0.011	0.003	1.003	2945.043
Dry Season	Nov–Apr vs Jan–Aug	0.075	0.018	1.018	73.425
Elevation	400–800 vs 650	0.095	0.029	1.029	27.717
	400–800 vs Coastal	0.132	0.033	1.034	21.217
	650 vs Coastal	0.011	0.001	1.001	20,905.029
Humidity	High vs 82%	0.075	0.008	1.008	329.850
Rainfall	1800–2000 vs 1530	0.084	0.018	1.019	68.807
	1800–2000 vs 1500–2500	0.133	0.015	1.015	108.544
	1530 vs 1500–2500	0.011	0.001	1.001	20,244.554
Temperature	24–25 vs 24.5	0.095	0.005	1.005	1,033.433
	24–25 vs 25–27	0.086	0.007	1.007	451.513
	24.5 vs 25–27	0.004	0.000	1.000	230,839.305

3.2 Hierarchical Partitioning of Genetic Diversity

Hierarchical decomposition of Shannon information showed that most MSP1 diversity was retained within populations across all environmental stratifications (Table 2; Supplementary Figures S1–S6).

Across climatic zones, within-population diversity accounted for 73.62% of total Shannon information, whereas among-population diversity accounted for 26.38% (Table 2). Dry season duration showed a similar distribution, with 76.33% of variation occurring within populations and 23.67% among populations.

Elevation showed the largest among-population component, with 34.68% of total Shannon information attributable to among-population variation (Table 2). Humidity and temperature showed the smallest among-population components, accounting for 11.17% and 11.35% of total diversity, respectively.

Scaled diversity (D') values remained low across all environmental stratifications, whereas overlap values (O') remained high (Table 2).

3.3 Mean Diversity Profiles Across Environmental Groupings

Mean diversity indices across environmental categories are summarized in Table 3. Among climatic zones, the tropical savanna population showed the highest values for allelic richness, effective number of alleles, Shannon information index, heterozygosity, and percentage polymorphism. Lower values were observed in the tropical wet–dry and tropical monsoon populations (Table 3).

For dry season duration, the November–April category showed higher diversity indices and polymorphism than the January–August category (Table 3).

Elevation-associated diversity indices were highest in the 400–800 m category and lower in the 650 m and coastal/lowland categories. High-humidity populations showed higher allelic richness and polymorphism than the 82% humidity category.

Rainfall and temperature groupings showed similar distributions, with the highest diversity indices observed in the 1800–2000 mm rainfall and 24–25°C temperature categories, respectively (Table 3).

All sample counts were reported as whole numbers, and total sample size across analyses was $N = 145$.

Table 2

Hierarchical Shannon Partitioning and Scaled Diversity/Overlap Across Environmental Stratifications in *Plasmodium falciparum* MSP1 Populations

Environmental stratification	Source	DF	G-test (Chi-Sq)	Shannon information (sH)	% of total sH	exp(sH)	Scaled diversity (D')	Scaled overlap (O' = 1 - D')	P(rand ≥ data)
Climatic zones	Among pops	2	5.751	0.020	26.377	1.020	0.032	0.968	0.010
	Within pops	142	16.051	0.055	73.623	1.057	0.055	0.945	1.000
	Total	144	21.802	0.075	100.000	1.078	0.073	0.927	–
Dry season duration	Among pops	1	5.161	0.018	23.671	1.018	0.040	0.960	0.010
	Within pops	143	16.641	0.057	76.329	1.059	0.056	0.944	1.000
	Total	144	21.802	0.075	100.000	1.078	0.073	0.927	–
Elevation	Among pops	2	7.561	0.026	34.679	1.026	0.042	0.958	0.010
	Within pops	142	14.242	0.049	65.321	1.050	0.049	0.951	1.000
	Total	144	21.802	0.075	100.000	1.078	0.073	0.927	–
Humidity	Among pops	1	2.435	0.008	11.168	1.008	0.017	0.983	0.100
	Within pops	143	19.367	0.067	88.832	1.069	0.065	0.935	0.920
	Total	144	21.802	0.075	100.000	1.078	0.073	0.927	–
Rainfall	Among pops	2	5.386	0.019	24.703	1.019	0.030	0.970	0.030
	Within pops	142	16.416	0.057	75.297	1.058	0.056	0.944	0.980
	Total	144	21.802	0.075	100.000	1.078	0.073	0.927	–
Temperature	Among pops	2	2.474	0.009	11.349	1.009	0.014	0.986	0.190
	Within pops	142	19.328	0.067	88.651	1.069	0.066	0.934	0.820
	Total	144	21.802	0.075	100.000	1.078	0.073	0.927	–

Hierarchical decomposition of MSP1 genetic diversity using Shannon information (sH) across environmental stratifications. Across all factors, within-population diversity accounts for most of the total information, while among-population components are smaller but sometimes statistically significant (permutation probability). DF = degrees of freedom; G-test = likelihood-ratio statistic for differentiation; sH = Shannon information contribution; % of total sH = proportional partitioning; exp(sH) = effective diversity; D' = scaled genetic differentiation; O' = scaled allelic overlap (1 - D'); P(rand ≥ data) = permutation-based probability of observing the partitioning by chance.

Table 3
Mean Genetic Diversity Indices and Polymorphism Across Environmental Groupings of *Plasmodium falciparum* MSP1 Populations

Environmental grouping	Population category	N	Na (mean $\pm$ SE)	Ne (mean $\pm$ SE)	I (mean $\pm$ SE)	h (mean $\pm$ SE)	uh/Uh (mean $\pm$ SE)	% polymorphic loci (%P)
Climatic zones	Tropical savanna (Aw)	40.000	1.844 $\pm$ 0.055	1.102 $\pm$ 0.010	0.179 $\pm$ 0.015	0.089 $\pm$ 0.008	0.092 $\pm$ 0.008	84.44
	Tropical wet and dry (Aw)	22.000	1.111 $\pm$ 0.047	1.018 $\pm$ 0.009	0.028 $\pm$ 0.013	0.015 $\pm$ 0.007	0.015 $\pm$ 0.007	11.11
	Tropical monsoon (Am)	83.000	1.044 $\pm$ 0.031	1.001 $\pm$ 0.001	0.003 $\pm$ 0.002	0.001 $\pm$ 0.001	0.001 $\pm$ 0.001	4.44
	Total (as reported)	48.333	1.333 $\pm$ 0.041	1.040 $\pm$ 0.006	0.070 $\pm$ 0.009	0.035 $\pm$ 0.005	0.036 $\pm$ 0.005	33.33 $\pm$ 25.63
Dry season duration	Nov–Apr	40.000	1.844 $\pm$ 0.055	1.102 $\pm$ 0.010	0.179 $\pm$ 0.015	0.089 $\pm$ 0.008	0.092 $\pm$ 0.008	84.44
	Jan–Aug	105.000	1.156 $\pm$ 0.055	1.004 $\pm$ 0.002	0.011 $\pm$ 0.004	0.004 $\pm$ 0.002	0.004 $\pm$ 0.002	15.56
	Mean polymorphism (as reported)	–	–	–	–	–	–	50.00 $\pm$ 34.44
Elevation	400–800 m	22.000	1.844 $\pm$ 0.055	1.191 $\pm$ 0.018	0.270 $\pm$ 0.022	0.152 $\pm$ 0.013	0.159 $\pm$ 0.014	84.44
	650 m	83.000	1.133 $\pm$ 0.051	1.005 $\pm$ 0.002	0.012 $\pm$ 0.005	0.005 $\pm$ 0.002	0.005 $\pm$ 0.002	13.33
	Coastal/Lowland	40.000	1.044 $\pm$ 0.031	1.002 $\pm$ 0.002	0.005 $\pm$ 0.004	0.002 $\pm$ 0.002	0.002 $\pm$ 0.002	4.44
	Mean polymorphism (as reported)	–	–	–	–	–	–	34.07 $\pm$ 25.32
Humidity	High	83.000	1.956 $\pm$ 0.031	1.052 $\pm$ 0.004	0.114 $\pm$ 0.007	0.049 $\pm$ 0.003	0.049 $\pm$ 0.004	95.56
	82%	62.000	1.044 $\pm$ 0.031	1.001 $\pm$ 0.001	0.004 $\pm$ 0.003	0.001 $\pm$ 0.001	0.001 $\pm$ 0.001	4.44
	Mean polymorphism (as reported)	–	–	–	–	–	–	50.00 $\pm$ 45.56
Rainfall	1800–2000 mm	40.000	1.844 $\pm$ 0.055	1.102 $\pm$ 0.010	0.179 $\pm$ 0.015	0.089 $\pm$ 0.008	0.092 $\pm$ 0.008	84.44
	1530 mm	83.000	1.111 $\pm$ 0.047	1.004 $\pm$ 0.002	0.010 $\pm$ 0.005	0.004 $\pm$ 0.002	0.004 $\pm$ 0.002	11.11
	1500–2500 mm	22.000	1.044 $\pm$ 0.031	1.004 $\pm$ 0.003	0.008 $\pm$ 0.006	0.004 $\pm$ 0.003	0.004 $\pm$ 0.003	4.44
	Total (as reported)	48.333	1.333 $\pm$ 0.041	1.037 $\pm$ 0.005	0.066 $\pm$ 0.009	0.032 $\pm$ 0.005	0.033 $\pm$ 0.005	33.33 $\pm$ 25.63
Temperature	24–25°C	83.000	1.956 $\pm$ 0.031	1.052 $\pm$ 0.004	0.114 $\pm$ 0.007	0.049 $\pm$ 0.003	0.049 $\pm$ 0.004	95.56
	24.5°C	22.000	1.000 $\pm$ 0.000	1.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.00
	25–27°C	40.000	1.044 $\pm$ 0.031	1.002 $\pm$ 0.002	0.005 $\pm$ 0.004	0.002 $\pm$ 0.002	0.002 $\pm$ 0.002	4.44
	Total (as reported)	48.333	1.333 $\pm$ 0.041	1.018 $\pm$ 0.002	0.040 $\pm$ 0.005	0.017 $\pm$ 0.002	0.017 $\pm$ 0.002	33.33 $\pm$ 31.14

Mean genetic diversity indices ($\pm$ SE) and polymorphism of MSP1 across environmental groupings. Diversity metrics (Na, Ne, Shannon's I, heterozygosity) and % polymorphic loci summarize allelic richness and variability within each ecological category, enabling direct comparison of within-group diversity patterns across climatic, seasonal, and environmental gradients. N = sample size; Na = mean number of alleles per locus; Ne = effective number of alleles; I = Shannon's Information Index; h = expected heterozygosity; uh/Uh = unbiased expected heterozygosity; %P = percentage of polymorphic loci. "Total (as reported)" values are reproduced exactly as presented in the manuscript outputs.

3.4 Pairwise Genetic Relatedness and Ordination-Based Population Structure

3.4.1 Nei’s Genetic Distance and Identity Across Environmental Groupings

Pairwise Nei’s genetic distance and identity estimates are summarized in Table 4. Across all environmental stratifications, Nei’s genetic distances remained low, whereas genetic identity values remained close to unity.

Among climatic zones, the largest standard genetic distance was observed between tropical savanna and tropical wet–dry populations ($D = 0.003$), whereas lower distances were observed between tropical wet–dry and tropical monsoon populations ($D = 0.002$) (Table 4).

Elevation comparisons involving the 400–800 m category showed the highest observed distances ($D = 0.008$), whereas the 650 m versus coastal/lowland comparison showed $D = 0.000$ (Table 4).

Humidity, rainfall, and temperature comparisons also showed low genetic distances and high genetic identity values across groups (Table 4).

3.4.2 Principal Coordinates Analysis (PCoA)

Principal Coordinates Analysis (PCoA) plots illustrating MSP1 genetic relationships across environmental categories are presented in Fig. 1A–F.

In climatic-zone analyses, broader dispersion was observed in the tropical savanna group relative to tropical wet–dry and tropical monsoon groups (Fig. 1A). Similar patterns were observed across dry season duration, elevation, humidity, rainfall, and temperature stratifications (Fig. 1B–F).

Substantial overlap among groups was observed across all ordination plots.

Table 4
Combined Nei’s Genetic Distance/Identity Summary Across Environmental Stratifications of *Plasmodium falciparum* MSP1 Populations

Environmental stratification	Population 1	Population 2	Nei D	Nei I	Nei uD	Nei ul
Climatic zones	Tropical savanna (Aw)	Tropical wet and dry (Aw)	0.003	0.997	0.002	0.998
	Tropical wet and dry (Aw)	Tropical monsoon (Am)	0.002	0.998	0.001	0.999
	Tropical monsoon (Am)	Tropical monsoon (Am)	0.001	0.999	0.000	1.000
Dry season duration	Nov–Apr	Jan–Aug	0.002	0.998	0.001	0.999
Elevation	400–800 m	650 m	0.008	0.992	0.003	0.997
	400–800 m	Coastal/Lowland	0.008	0.992	0.003	0.997
	650 m	Coastal/Lowland	0.000	1.000	0.000	1.000
Humidity	High	82%	0.001	0.999	0.000	1.000
Rainfall	1800–2000 mm	1530 mm	0.002	0.998	0.001	0.999
	1800–2000 mm	1500–2500 mm	0.002	0.998	0.001	0.999
	1530 mm	1500–2500 mm	0.000	1.000	0.000	1.000
Temperature	24–25°C	24.5°C	0.000	1.000	0.000	1.000
	24–25°C	25–27°C	0.001	0.999	0.000	1.000
	24.5°C	25–27°C	0.000	1.000	0.000	1.000

Standard and unbiased Nei’s genetic distance (D , uD) and genetic identity (I , ul) across environmental stratifications. Across all gradients, genetic distances were extremely low ($D \leq 0.008$; $uD \leq 0.003$) and identities approached unity ($I \geq 0.992$; $ul \geq 0.997$), indicating high allelic continuity and limited population subdivision across climatic and ecological groupings. Nei D and Nei I represent standard genetic distance and identity; Nei uD and Nei ul are unbiased estimates that adjust for sampling effects. Lower distance and higher identity indicate stronger genetic similarity.

3.5 Haplotype Diversity and Climatic/Ecological Barcoding Profiles

Barcoding analysis of *Plasmodium falciparum* populations revealed a total of 12 unique haplotypes (H1–H12) differentiated by specific nucleotide substitutions across the barcode region. These haplotypes were further characterized based on four ecological parameters: temperature, rainfall, humidity, and elevation (Table 5).

Most haplotypes (e.g., H1, H2, H3, H5, H7, H8) were associated with medium temperature, medium to low rainfall, high humidity, and medium elevation, representing the dominant barcode category: T-Medium_R-Medium_H-High_E-Medium. This cluster suggests that moderate thermal and rainfall conditions in humid, mid-elevation regions are conducive to haplotype richness and diversification.

In contrast, haplotypes such as H4, H6, H9, H10, and H11 were consistently found in high-temperature, medium-rainfall, high-humidity, and low-elevation settings, forming the distinct barcode group: T-High_R-Medium_H-High_E-Low. These variants likely reflect environment-specific adaptation, potentially favored by lowland microclimates.

Interestingly, H7 and H8 were the only haplotypes observed under low rainfall, although both were still associated with medium temperature, high humidity, and medium elevation. These two shared near-identical barcode profiles, suggesting recent divergence or local expansion within a niche experiencing reduced precipitation.

Haplotype H12, with a unique sequence and matching the common medium environmental signature, reinforces the observed trend that mid-range temperature and elevation with persistent humidity support broader haplotype diversity.

Table 5
Haplotype Sequences and Their Associated Climatic and Ecological Barcodes for *P. falciparum* Populations.

haplotype_id	haplotype	Temp	Rainfall	Humidity	Elevation	Barcode
H1	CAAGTGCTGGGGGCTTCGTGGCTGGCTTCTGGGCACAGACTCTAT	Medium	Medium	High	Medium	T-Medium_R-Medium_H-High_E-Medium
H10	TAAGTGCTGGGGGCTTCGTGGCTGGCTTCTGGGTACAGACTCTGT	High	Medium	High	Low	T-High_R-Medium_H-High_E-Low
H11	TAAGTGCTGGGGGTTTCGGGGTTGGCTTCTGGGCACAGACTCTGT	High	Medium	High	Low	T-High_R-Medium_H-High_E-Low
H12	TGGTTACTGGTACTTGGTTACTTTCTTTCTGTCTATTACTCGCGT	Medium	Medium	High	Medium	T-Medium_R-Medium_H-High_E-Medium
H2	TAAGGGGCTCGGGTTGGGTGGCCTCTTGGCGTCTATTACTCGCGT	Medium	Medium	High	Medium	T-Medium_R-Medium_H-High_E-Medium
H3	TAAGTGCTGGGGGCATCGTGGCTGGCTTCTGGGCACAGACTCTGT	Medium	Medium	High	Medium	T-Medium_R-Medium_H-High_E-Medium
H4	TAAGTGCTGGGGGCTTCGTGGCTGGCCTCTGGGCACAGACTCTGT	High	Medium	High	Low	T-High_R-Medium_H-High_E-Low
H5	TAAGTGCTGGGGGCTTCGTGGCTGGCTTCTAGGCACAGACTCTGT	Medium	Medium	High	Medium	T-Medium_R-Medium_H-High_E-Medium
H6	TAAGTGCTGGGGGCTTCGTGGCTGGCTTCTGGGCACAGACTCTAT	High	Medium	High	Low	T-High_R-Medium_H-High_E-Low
H7	TAAGTGCTGGGGGCTTCGTGGCTGGCTTCTGGGCACAGACTCTGA	Medium	Low	High	Medium	T-Medium_R-Low_H-High_E-Medium
H8	TAAGTGCTGGGGGCTTCGTGGCTGGCTTCTGGGCACAGACTCTGT	Medium	Low	High	Medium	T-Medium_R-Low_H-High_E-Medium
H9	TAAGTGCTGGGGGCTTCGTGGCTGGCTTCTGGGCGCAGACTCTGT	High	Medium	High	Low	T-High_R-Medium_H-High_E-Low

4.0 DISCUSSION

4.1 Overall Population Structure

This study evaluated the relationship between climatic heterogeneity and MSP1 genetic diversity in *Plasmodium falciparum* populations across environmentally distinct regions of Vietnam. Across all environmental stratifications examined, the dominant pattern was the retention of most genetic variation within populations rather than between ecological groups. Shannon partitioning analyses consistently showed that within-population diversity accounted for the majority of total genetic information, while among-population components remained comparatively small.

Pairwise genetic differentiation estimates further supported this pattern. Nei's genetic distances were uniformly low across climatic and ecological comparisons, and genetic identity values approached unity in most pairwise analyses. Similarly, overlap indices remained high across all environmental gradients. Although some comparisons yielded statistically detectable differentiation, the magnitude of divergence remained limited overall.

The ordination analyses were consistent with these findings. Principal Coordinates Analysis demonstrated substantial overlap among environmental groupings, with broader dispersion primarily observed in populations exhibiting higher allelic richness and heterozygosity. Together, these findings suggest that *P. falciparum* populations across Vietnam retain substantial allelic continuity despite climatic and ecological heterogeneity.

These observations are consistent with previous studies reporting limited parasite subdivision in endemic settings where recombination, vector movement, and human mobility contribute to continued genetic connectivity among populations [10–12, 39].

4.2 Ecological Gradients and Diversity Magnitude

Although strong population subdivision was not observed, variation in diversity magnitude was detected across several environmental gradients. Elevation and rainfall showed comparatively larger among-population entropy components than humidity and temperature, while still retaining high within-population diversity overall.

Mid-elevation populations exhibited higher allelic richness, heterozygosity, and polymorphism than lower-elevation categories. Similarly, higher-rainfall populations displayed greater diversity indices than lower-rainfall groups. In contrast, humidity and temperature showed comparatively limited differentiation across environmental categories.

Seasonality also appeared to influence diversity magnitude. Populations from shorter dry-season regions exhibited higher allelic richness and heterozygosity than populations exposed to prolonged dry conditions. This pattern is broadly consistent with ecological studies linking transmission stability to maintenance of parasite diversity [40–44].

The relatively larger among-population entropy components associated with elevation and rainfall may reflect their influence on local transmission ecology. Elevation affects temperature regimes, vector abundance, and the duration of parasite development within mosquito hosts, potentially generating differences in transmission stability between highland and lowland environments (Gething et al., 2011; Ohm et al., 2018; Sulaiman et al., 2025). Similarly, rainfall influences the availability and persistence of mosquito breeding habitats, which can affect vector density and opportunities for parasite recombination (Paaijmans et al., 2013; Kar et al., 2014).

In areas experiencing more favorable transmission conditions, larger effective parasite populations may maintain greater allelic richness and diversity, whereas less favorable environments may experience periodic reductions in diversity associated with transmission bottlenecks. Sexual reproduction and recombination of *Plasmodium* parasites within mosquito vectors contribute substantially to the generation and maintenance of parasite genetic diversity, particularly in areas of sustained transmission where parasite populations remain large (Bousema & Drakeley, 2011; Hellewell et al., 2025). Nevertheless, extensive human movement and parasite gene flow across ecological zones likely limit the accumulation of strong genetic differentiation, explaining why entropy partitioning detected modest diversity differences without substantial population subdivision. Similar patterns have been observed in malaria-endemic regions where frequent parasite migration and high transmission connectivity maintain genetic mixing despite environmental heterogeneity (Manske et al., 2012; Amambua-Ngwa et al., 2019).

The diversity patterns observed in Vietnam are broadly comparable to reports from several moderate-transmission settings in Southeast Asia and the Pacific, where substantial allelic diversity is often accompanied by limited population subdivision. Studies from Thailand and Papua New Guinea have similarly reported extensive sharing of *Plasmodium falciparum* alleles across geographically separated populations despite ecological differences, reflecting ongoing gene flow and parasite connectivity (Pumpaibool et al., 2009; Schultz et al., 2010). In contrast, some high-transmission regions of sub-Saharan Africa frequently exhibit greater allelic richness and more complex infection structures due to larger effective parasite populations and higher recombination rates (Pumpaibool et al., 2009). The comparatively low differentiation observed in the present study therefore appears consistent with a setting in which environmental heterogeneity influences local diversity profiles without generating strong barriers to parasite connectivity.

Despite these differences in diversity magnitude, absolute genetic distances across all environmental gradients remained low. Therefore, the observed ecological patterns are more consistent with variation in allelic richness and polymorphism than with strong ecological subdivision of parasite populations.

4.3 Entropy-Based Diversity Metrics

This study applied Shannon entropy partitioning as a complementary framework for evaluating MSP1 diversity across environmental gradients. Unlike conventional fixation indices that primarily quantify allele-frequency divergence, entropy-based metrics emphasize proportional diversity partitioning and effective diversity components.

The entropy framework allowed simultaneous quantification of within-population and among-population diversity contributions while also providing scaled overlap measures across environmental categories. In the present study, entropy analyses consistently showed high overlap values and

dominant within-population diversity components across all stratifications.

The entropy-based results were generally concordant with classical genetic distance analyses. Both approaches identified low overall differentiation and substantial allelic continuity across climatic and ecological groupings. These findings suggest that entropy partitioning may provide a useful complementary perspective for evaluating diversity patterns in highly polymorphic loci such as MSP1, particularly when conventional differentiation estimates approach zero.

4.4 Limitations of MSP1-Based Population Inference

Several limitations should be considered when interpreting these findings. First, MSP1 is a highly polymorphic antigenic locus that is subject to balancing immune selection. Consequently, variation at this marker may not fully reflect genome-wide neutral population structure.

Second, recombination and immune-mediated selection may contribute to maintenance of allelic diversity independently of environmental subdivision. This is particularly relevant in endemic settings where repeated exposure and transmission intensity can sustain extensive polymorphism at antigenic loci.

Third, several pairwise comparisons produced extremely high inferred N_m values because differentiation estimates approached zero. These estimates should therefore be interpreted cautiously as theoretical indicators of low differentiation rather than direct biological estimates of migration intensity.

How would results differ using neutral genomic markers?

If neutral genomic markers were analyzed instead of MSP1, stronger evidence of population subdivision might be detected. Unlike MSP1, which is maintained by balancing immune selection and antigen-driven diversification, neutral loci are influenced primarily by migration, genetic drift, recombination, and demographic processes (Conway et al., 2000; Volkman et al., 2012). Genome-wide studies have demonstrated that parasite populations may appear similar at antigenic loci while exhibiting measurable differentiation across neutral regions of the genome, thereby revealing patterns of population structure and connectivity that are less affected by host immune selection (Orjuela-Sánchez et al., 2010; Barry et al., 2009). For surveillance programs, this distinction is important because MSP1 diversity may be more informative for monitoring transmission intensity and antigenic variation, whereas neutral genomic markers may provide greater resolution for tracking parasite movement, population connectivity, and local transmission networks (Volkman et al., 2012; Orjuela-Sánchez et al., 2010).

Finally, this study relied on a single-locus framework. Although MSP1 remains informative for describing allelic diversity and population variability, broader genome-wide analyses would provide higher-resolution assessment of population structure, demographic history, and adaptive processes.

4.5 Implications for Malaria Surveillance

The present findings indicate that climatic and ecological gradients in Vietnam are associated with variation in MSP1 diversity magnitude, while overall population subdivision remains limited. These patterns suggest that environmental heterogeneity may influence local diversity profiles without generating strong fragmentation of parasite populations.

From a surveillance perspective, the persistence of high allelic overlap across environmental regions highlights the importance of coordinated monitoring strategies across endemic zones. Continued integration of climatic information with molecular surveillance may improve understanding of transmission heterogeneity and population dynamics in moderate-transmission settings.

Relationship to Genome-Wide Population Genomic Studies

Recent genome-wide investigations of *Plasmodium falciparum* have demonstrated that parasite populations may exhibit patterns of structure that are not fully captured by highly polymorphic antigenic loci such as MSP1. Whole-genome sequencing studies have revealed fine-scale population subdivision, signatures of local adaptation, and regional differences in gene flow that can remain undetected when analyses are restricted to immune-selected markers.

Because MSP1 is subject to balancing immune selection, allelic diversity may be maintained across geographically separated populations even when neutral genomic regions show greater differentiation. Consequently, the high allelic overlap observed in the present study should not be interpreted as evidence of complete genetic homogeneity across Vietnam. Rather, the findings suggest that environmental gradients are associated with variation in MSP1 diversity magnitude while maintaining substantial sharing of antigenic variants.

The concordance between low differentiation estimates and high allelic overlap nevertheless provides useful epidemiological information. Maintenance of diverse MSP1 alleles across ecological settings may reflect ongoing parasite connectivity, sustained transmission, and shared selective pressures imposed by host immunity. Future studies integrating whole-genome sequencing with environmental data would provide a more comprehensive assessment of the extent to which climatic heterogeneity influences both adaptive and neutral components of parasite genetic variation.

5.0 CONCLUSION

This study evaluated MSP1 genetic diversity across multiple climatic and ecological gradients in *Plasmodium falciparum* populations from Vietnam using entropy-based partitioning, genetic distance analysis, and ordination approaches. Across all environmental stratifications, most genetic variation was retained within populations, whereas among-population differentiation remained modest.

Elevation and rainfall showed comparatively larger among-population diversity components than humidity and temperature, while populations from higher-rainfall and mid-elevation regions exhibited greater allelic richness and heterozygosity. Nevertheless, genetic distances remained low and allelic overlap remained high across all ecological comparisons.

Overall, the findings indicate that climatic gradients in Vietnam are associated primarily with differences in diversity magnitude rather than strong population subdivision. Because analyses were based on the immune-selected MSP1 locus, these findings should be interpreted as locus-specific diversity patterns rather than evidence of genome-wide population structure.

Declarations

Acknowledgement

The authors are grateful to Tertiary Education Trust Fund (Nigeria) for the NRF grant given to them.

Authors' contributions

J.O.A. conceived the idea. O.A.G and D.I.A carried out the bench work while the manuscript was read by the authors.

Conflict of interest

There authors declare that there is no conflict of interest.

Funding declaration

The study was funded by a National Research Fund grant (TETF/ES/DR&D-CE/NRF2023/SETI/HSW/00412/VOL.1) jointly awarded by Tertiary Education Trust Fund (Nigeria) to the authors.

Data Availability Statement

All data generated or analysed during this study are included in this published article and its supplementary information files. The raw *Plasmodium falciparum* *msp1* sequence dataset, GenBank accession identifiers, population labels, climatic/ecological metadata, polymorphic nucleotide positions, haplotype assignments, and derived population-genetic outputs are provided in **Supplementary File 2**. The original source sequences were retrieved from the NCBI GenBank database, and the accession numbers are listed in Supplementary File 2. No new gene sequences were generated in this study. Additional derived datasets and analysis outputs are available from the corresponding author on reasonable request.

References

1. World Health Organization. *World Malaria Report 2023* (WHO, 2023).
2. Tessema, S. K. et al. Applying next-generation sequencing to track falciparum malaria transmission in sub-Saharan Africa. *Am. J. Trop. Med. Hyg.* **101** (4), 1261–1269 (2019).
3. Mobegi, V. A. et al. Population genetic structure of *Plasmodium falciparum* across a region of diverse transmission in West Africa. *Malar. J.* **11**, 223 (2012).
4. Ferreira, M. U. et al. The MSP1 gene as a molecular marker of *Plasmodium falciparum* diversity and evolution. *Mem. Inst. Oswaldo Cruz.* **99** (1), 1–12 (2004).
5. Manske, M. et al. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487** (7407), 375–379 (2012).
6. Tusting, L. S. et al. Environmental temperature and malaria transmission: evidence from a systematic review. *Lancet Planet. Health.* **4** (5), e172–e183 (2020).
7. Gething, P. W. et al. Climate change and the global malaria recession. *Nature* **465** (7296), 342–345 (2010).
8. Snounou, G., Viriyakosol, S., Jarra, W., Thaithong, S. & Brown, K. N. Identification of the four human malaria parasite species in field samples by the polymerase chain reaction and detection of high prevalence of mixed infections. *Mol. Biochem. Parasitol.* **58** (2), 283–292 (1993).
9. Putaporntip, C., Hughes, A. L. & Jongwutiwes, S. Low levels of variation in *Plasmodium falciparum* merozoite surface protein 1 gene (*msp1*) among isolates from Thailand. *Am. J. Trop. Med. Hyg.* **83** (6), 1231–1238 (2010).

10. Tatem, A. J. & Smith, D. L. International population movements and regional *Plasmodium falciparum* malaria elimination strategies. *Proc. Natl. Acad. Sci. U S A.* **107** (27), 12222–12227 (2010).
11. Wangdi, K. et al. Analysis of clinical malaria disease patterns and trends in Vietnam 2009–2015. *Malar. J.* **17** (1), 332 (2018).
12. Anders, K. *The epidemiology of dengue in southern Vietnam: distribution of disease and determinants of risk in an endemic setting* (Doctoral dissertation, Monash University). (2015).
13. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30** (4), 772–780 (2013).
14. Ranwez, V., Douzery, E. J., Cambon, C., Chantret, N. & Delsuc, F. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol. Biol. Evol.* **35** (10), 2582–2584 (2018).
15. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34** (suppl_2), W609–W612 (2006).
16. Smouse, R. P. P. & Peakall, R. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28** (19), 2537–2539 (2012).
17. Nei, M. Genetic distance between populations. *Am. Nat.* **106** (949), 283–292 (1972).
18. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24** (11), 1403–1405 (2008).
19. Paradis, E. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26** (3), 419–420 (2010).
20. Goudet, J. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes.* **5** (1), 184–186 (2005).
21. Sherwin, W. B. Entropy and information approaches to genetic diversity and its expression: genomic geography. *entropy*, **12**(7), 1765–1798. (2010).
22. Ma, Z., Li, L. & Zhang, Y. P. Defining individual-level genetic diversity and similarity profiles. *Sci. Rep.* **10** (1), 5805 (2020).
23. Oksanen, J. Vegan: community ecology package. *R package version.* **2**, 3 (2015).
24. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34** (12), 3299–3302 (2017).
25. Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E. & Lercher, M. J. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31** (7), 1929–1936 (2014).
26. Leigh, J. W., Bryant, D. & Nakagawa, S. POPART: full-feature software for haplotype network construction. *Methods Ecol. & Evolution*, **6**(9). (2015).
27. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37** (5), 1530–1534 (2020).
28. Kalyanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods.* **14** (6), 587–589 (2017).
29. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22** (24), 3096–3098 (2006).
30. Kosakovsky Pond, S. L. & Frost, S. D. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22** (5), 1208–1222 (2005).
31. Murrell, B. et al. S. L. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.*, **8**(7), e1002764. (2012).
32. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24** (8), 1586–1591 (2007).
33. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10** (3), 564–567 (2010).
34. Dray, S. & Dufour, A. B. The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).
35. Hsieh, T. C., Ma, K. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7** (12), 1451–1456 (2016).
36. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series* (Vol. 41, No. 41, pp. 95–98). (1999), January.
37. Aurrecochea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., ... Wang, H. (2009). PlasmoDB: a functional genomic database for malaria parasites. *Nucleic acids research*, **37**(suppl_1), D539–D543..
38. Wong, T. K., Ly-Trong, N., Ren, H., Baños, H., Roger, A. J., Susko, E., ... Minh, B. Q. (2025). IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models.
39. Rougeron, V. et al. Human–*Plasmodium falciparum* coevolution: an example of the Red Queen? *Philos. Trans. R Soc. Lond. B Biol. Sci.* **364** (1530), 2731–2740 (2009).
40. Kweka, E. J. et al. Anopheline larval habitats seasonality and ecological implications for malaria vector control in western Kenya. *Parasit. Vectors.* **5**, 199 (2012).
41. Koepfli, C. & Mueller, I. Malaria epidemiology at the clone level. *Trends Parasitol.* **33** (12), 974–985 (2017).

42. Tusting, L. S., Bousema, T., Smith, D. L. & Drakeley, C. Measuring changes in *Plasmodium falciparum* transmission: precision, accuracy and costs of metrics. *Adv. Parasitol.* **84**, 151–208 (2014).
43. Midega, J. T. et al. Climate variability and malaria transmission in Kenya: a review. *Parasit. Vectors.* **5**, 112 (2012).
44. Bousema, T. et al. Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med.* **9** (1), e1001165 (2012).
45. Mobegi, V. A. et al. Population genetic structure of *Plasmodium falciparum* across a region of diverse endemicity in West Africa. *Malar. J.* **11**, 223 (2012).
46. Anderson, T. J. et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17** (10), 1467–1482 (2000).
47. Conway, D. J. Molecular epidemiology of malaria. *Clin. Microbiol. Rev.* **20** (1), 188–204 (2007).
48. Barry, A. E. et al. *Plasmodium falciparum* population structure and gene flow associated with anti-malarial drug resistance in Papua New Guinea. *PLoS One.* **8** (5), e61847 (2013).
49. Roper, C. et al. Seasonal changes in the structure of *Plasmodium falciparum* gametocyte populations and their relationship to mosquito infectivity. *Trop. Med. Int. Health.* **9** (6), 618–624 (2004).
50. Takala, S. L. & Branch, O. H. Evidence for intragenic recombination and selective sweeps in the *Plasmodium falciparum* msp1 gene. *Mol. Biochem. Parasitol.* **150** (1), 45–57 (2006).
51. Volkman, S. K. et al. A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.* **39** (1), 113–119 (2007).
52. Omumbo, J. A., Hay, S. I., Guerra, C. A. & Snow, R. W. The influence of urbanization on measures of *Plasmodium falciparum* infection prevalence in East Africa. *Acta Trop.* **93** (1), 11–21 (2005).
53. Guerra, C. A. et al. The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. *PLoS Med.* **5** (2), e38 (2008).
54. Ernst, K. C., Adoka, S. O., Kowuor, D. O., Wilson, M. L. & John, C. C. Malaria hotspot areas in a highland Kenya site are consistent in epidemic and non-epidemic years and are associated with ecological factors. *Malar. J.* **5**, 78 (2006).
55. Conway, D. J. Molecular epidemiology of malaria. *Clin. Microbiol. Rev.* **20** (1), 188–204 (2007).
56. Mobegi, V. A. et al. Population genetic structure of *Plasmodium falciparum* across a region of diverse endemicity in West Africa. *Malar. J.* **11**, 223 (2012).
57. Mita, T. et al. Temporal changes in the population structure of *Plasmodium falciparum* in Kilifi, Kenya. *Infect. Genet. Evol.* **11** (4), 683–689 (2011).
58. Anderson, T. J. et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17** (10), 1467–1482 (2000).
59. Lindsay, S. W. & Birley, M. H. Climate change and malaria transmission. *Ann. Trop. Med. Parasitol.* **90** (6), 573–588 (1996).
60. Paaijmans, K. P., Read, A. F. & Thomas, M. B. Understanding the link between malaria risk and climate. *Proc. Natl. Acad. Sci. USA.* **106** (33), 13844–13849 (2009).
61. Joy, D. A. et al. Local adaptation and vector-mediated population structure in *Plasmodium vivax* malaria. *Curr. Biol.* **18** (12), 930–936 (2008).
62. Barry, A. E. et al. *Plasmodium falciparum* population structure and gene flow associated with antimalarial drug resistance in Papua New Guinea. *PLoS One.* **8** (5), e61847 (2013).
63. Mita, T., Tanabe, K. & Kita, K. Spread and evolution of *Plasmodium falciparum* drug resistance. *Parasitol. Int.* **58** (3), 201–209 (2009).
64. Anderson, T. J. et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17** (10), 1467–1482 (2000).
65. Smith, D. L., Dushoff, J., Snow, R. W. & Hay, S. I. The entomological inoculation rate and *Plasmodium falciparum* infection in African children. *Nature* **438** (7067), 492–495 (2005).
66. Molineaux, L. & Gramiccia, G. *The Garki project: research on the epidemiology and control of malaria in the Sudan savanna of West Africa* (World Health Organization, 1980).
67. Reiter, P. Climate change and mosquito-borne disease. *Environ. Health Perspect.* **109** (Suppl 1), 141–161 (2001).
68. Githeko, A. K., Lindsay, S. W., Confalonieri, U. E. & Patz, J. A. Climate change and vector-borne diseases: a regional analysis. *Bull. World Health Organ.* **78** (9), 1136–1147 (2000).
69. Anderson, T. J. et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol. Biol. Evol.* **17** (10), 1467–1482 (2000).
70. Mita, T., Tanabe, K. & Kita, K. Spread and evolution of *Plasmodium falciparum* drug resistance. *Parasitol. Int.* **58** (3), 201–209 (2009).

Figures

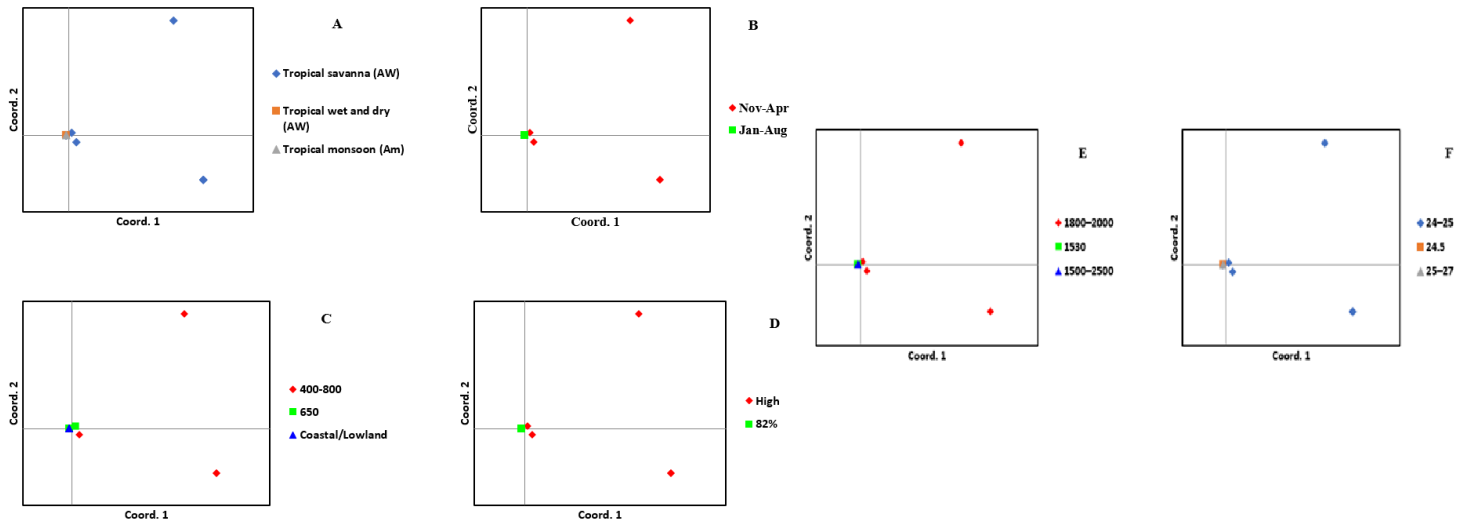


Figure 1

Principal Coordinates Analysis (PCoA) of *Plasmodium falciparum* MSP1 Genetic Structure Across Climatic and Ecological Gradients. Principal Coordinates Analysis (PCoA) plots based on pairwise Nei's genetic distance matrices illustrating the genetic relationships among *Plasmodium falciparum* populations stratified by environmental variables. Each point represents an individual genotype positioned according to its genetic distance from other samples along the first two principal coordinates (Coord. 1 and Coord. 2), which capture the largest proportion of total genetic variance. **(A)** Climatic zones (Tropical savanna, Tropical wet and dry, Tropical monsoon). **(B)** Dry season duration (November–April vs January–August). **(C)** Elevation bands (400–800 m, 650 m, Coastal/Lowland). **(D)** Humidity levels (High vs 82%). **(E)** Rainfall zones (1800–2000 mm, 1530 mm, 1500–2500 mm) and **(F)** Temperature ranges (24–25°C, 24.5°C, 25–27°C). Across all environmental stratifications, populations exhibiting higher allelic richness and heterozygosity demonstrate broader dispersion in ordination space, whereas low-diversity groups cluster tightly near the origin. The substantial overlap among clusters supports the overall low genetic differentiation (Nei's $D \leq 0.008$) and high allelic identity ($I \geq 0.992$) observed across climatic and ecological gradients.

Supplementary Files

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

- [RevisedSupplementaryfile1.docx](#)