

Multivariate Statistical Analysis of the Genetic Diversity of *Plasmodium falciparum* Isolates from Individuals Carrying the Major or Minor Sickle Cell Trait (HBAA, HBAS, and HBSS) in Abidjan (Côte d'Ivoire): Application of Principal Component Analysis, Jost's D Index and PERMANOVA

Egomli Stanislas Assohoun^{1,2,3}, Brou Médard Kouassi¹, Aristide Beranger Ako², Sherif Yacouba Ouattara², Alloh Albert Gnondjui², Offianan André Touré², Kablan Jérôme Adou¹, Ronan Jambou²

 ¹Mechanics and Computer Science Laboratory, Félix Houphouët-Boigny University, Abidjan, Ivory Coast
 ²Parasitology and Mycology Unit, Pasteur Institute (IPCI), Abidjan, Ivory Coast
 ³Laboratory of Environmental Science and Technology (LSTE), Jean Lorougnon Guédé University, Daloa, Ivory Coast Email: stanlasso@gmail.com

How to cite this paper: Assohoun, E.S., Kouassi, B.M., Ako, A.B., Ouattara, S.Y., Gnondjui, A.A., Touré, O.A., Adou, K.J. and Jambou, R. (2025) Multivariate Statistical Analysis of the Genetic Diversity of *Plasmodium falciparum* Isolates from Individuals Carrying the Major or Minor Sickle Cell Trait (HBAA, HBAS, and HBSS) in Abidjan (Côte d'Ivoire): Application of Principal Component Analysis, Jost's D Index and PERMANOVA. *Open Journal of Applied Sciences*, **15**, 1716-1731. https://doi.org/10.4236/ojapps.2025.156118

Received: May 4, 2025 **Accepted:** June 24, 2025 **Published:** June 27, 2025

Abstract

In malaria-endemic areas, the relatively high probability of hemoglobinopathies, especially sickle cell disease, raises important questions regarding the evolutionary interactions between host hemoglobin genotype and adaptive dynamics of *Plasmodium falciparum*. In fact, while the protection provided by heterozygous genotypic HbAS against malaria-resistant forms is now well established, the impact on the genetic structuring of parasitic populations remains relatively poorly understood. To do this, we carried out, as part of an exploratory approach, a population genomics analysis on 30 clinical isolates of P. falciparum from individuals of the HbAA, HbAS and HbSS genotypes in Côte d'Ivoire. Next-generation sequencing (NGS) data were analyzed using principal component analysis (PCA), permuted analysis of variance (PER-MANOVA), and Jost differentiation index (JOST'D) to assess genetic deviations. Coding and non-coding SNPs were then functionally annotated to identify evolutionary adaptation signatures. Genetic analyses show a significant structuring of isolates according to the hemoglobin genotype of the host, with moderate to strong differentiation between parasites from HbAS compared to Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/ **Open Access**

(00) (\mathbf{i})

those from homozygotes HbAA and HbSS. Discriminating SNPs are contained in genes associated with erythrocyte invasion, immune escape (clag3.2, acs7, rh5, var2csa, PfEMP1), as well as intergenic regions potentially involved in transcriptional regulation. Thus, this study validates an exploratory method to identify differentially expressed mutations according to sickle cell traits and supports the idea of the genetic adaptation of the parasite, controlled by the hemoglobin status of the patient. Moreover, the impact of the ervthrocyte environment on the variability of the P. falciparum genome has been highlighted. The data here opens the door to a suite of studies aimed at deepening the functional assessment of mutations and justifying the integration of hemoglobinopathies into the malaria control strategy in endemic regions. The results of this exploratory study will be consolidated by a larger print run of the DREPAL project (DREpanocytosis PALudism), promoted by the paludology mycology unit of the Institut Pasteur de Côte d'Ivoire.

Keywords

Plasmodium falciparum, Sickle Cell Anemia, Host-Parasite, Co-Evolution, SNP, PERMANOVA, ACP, Jost Index

1. Introduction

Despite the efforts of recent decades, malaria remains a major public health issue in sub-Saharan Africa. Caused mainly by the parasite Plasmodium falciparum, this infection is responsible for considerable morbidity and mortality. In 2023, there were about 263 million cases, resulting in 597,000 deaths, most of which are in children under 5 years of age [1]. Despite the use of coping strategies consisting of vector control, chemoprophylaxis and antimalarial treatments, the region continues to bear a disproportionate share of the global malaria burden [2].

This epidemiology is accompanied by a high prevalence of sickle cell anemia, a common genetic in Africa [3]. This hereditary hemoglobinopathy, caused by a mutation responsible for the synthesis of hemoglobin S, is characterized by partial protection of the heterozygous HbAS carrier against severe forms of parasitosis [3].

This case of balanced natural selection has led to the hypothesis of the existence of reciprocal evolutionary pressures between the human host and the P. falcipa*rum* parasite [4].

While the host-protective effect of the HbS allele is well documented, its impact on the genetic diversity of *P. falciparum* and its structuring are relatively poorly understood. Few studies have examined the impacts resulting from the parasite's adaptation to the particular intracellular environment of HbS-containing red blood cells. This adaptation would impose particular constraints on the parasites due to the various pathophysiological aspects, in particular oxidative stress, membrane rigidity caused by HbS, and hypoxia [5].

In this study, we postulated that the hemoglobin genotype of the normal (HbAA), heterozygous (HbAS) or homozygous sickle cell (HbSS) subject, could contribute to modulating the profile of the *P. falciparum* parasite at the genomic level. Using next-generation sequencing data from 30 clinical isolates, he was identified SNPs and adopted a population genomics approach to track the genetic structuring of the parasite. To do this, the PCA [6] [7], Jost'D's index [8] and PERMANOVA [9] have been applied.

The present work highlights a genetic signature of host HbAS strains, characterized by a weak to moderate allelic differentiation compared to those infecting HbAA and HbSS hosts. The results are supported by multivariate analysis and identification of the most discriminating SNPs, located in genes functionally related to erythrocyte invasion and immune escape, indicating adaptive defense mechanisms specific to heterozygous blood environments.

This study is organized into three sections: (i) materials and methods, describing the experimental protocols, the bioinformatics tools used and the statistical approaches implemented; (ii) results and discussion, presenting and responding to the results obtained and implemented in the context of the existing literature; (iii) finally, conclusion and perspectives synthesizing the contribution of the present work to the understanding of host-parasite co-evolutions and observing the perspectives opened up in the context of future research and malaria control in areas with high prevalence rates of hemoglobinopathies.

2. Materials and Methods

2.1. Study Area and Sample Collection

The biological samples used in this study were collected from two sites in Abidjan. The first corresponds to the clinical hematology department of the University Hospital Center (CHU) of Yopougon. It is a reference center for the care of patients with sickle cell disease in Abidjan. The second is the Community-Based Urban Health Training (FSUCOM) of Anonkoua-kouté located in the commune of Abobo. This site was chosen for the recruitment of sickle cell trait (HbAS) carriers and the control group of healthy individuals with normal hemoglobin (HbAA). These two sites were selected because of their central role in the medical management and follow-up of patients with sickle cell disease (**Figure 1**).

The data used in this study are from a preliminary study conducted as part of the DREPAL project. The collection phase took place from May 2017 to January 2018. A total of 30 patients with febrile episodes suggestive of malaria had participated in this exploratory phase. The participants were divided into three groups of 10 individuals. These include patients with major sickle cell disease: HbSS, sickle cell trait: HbAS, and finally the non-sickle cell control population: HbAA. Sickle cell status was initially determined by hemoglobin electrophoresis and confirmed by the FRET (Fluorescence Resonance Energy Transfer) method. The strains of *Plasmodium falciparum* used were those purified from blood samples with EDTA tubes and stored at -80° C.



Figure 1. Study area.

2.2. DNA Extraction and Library Preparation for Illumina Sequencing

Genomic DNA was extracted from whole blood samples from patients infected with *Plasmodium falciparum*. To improve the yield of the parasitic DNA and reduce host contamination, the samples were previously enriched by saponin treatment followed by selective lysis of uninfected erythrocytes and successive washes with PBS 1X. DNA extraction from the samples was achieved with the DNA Mini Kit, QIAamp DNA Mini Kit (Qiagen). This protocol was carried out like the manufacturer's protocol, however, the information received was that it was necessary to expand the protocol for samples with low parasitaemia. Then, 200 μ L of blood was incubated in AL and proteinase K solution at 56°C for 30 min. The genomic DNA was precipitated and purified in silica columns and different buffer solutions. Finally, genomic DNA was eluted in 100 µL of nuclease-free ultrapure water and stored at -20° C for further analyses. DNA quality was assessed by spectrophotometry with the A260/A280 between 1.8 and 2.0. Quantitation was done in the InvitrogenTM QubitTM 4 Fluorometer by the dsDNA High Sensitivity Assay Kit (ThermoFisherTM). Integrity by 1% agarose gel electrophoresis. The libraries were prepared for sequencing with 100 ng of total DNA per sample with the Nextera XT DNA Library Preparation Kit (Illumina[®]). This protocol had a tagging strategy linked to technology. The complex DNA fragment index is added with i5 and i7 adapters at the ends to identify each sample.

After purification by AMPure XP magnetic beads (Beckman Coulter) and verification of fragment sizes between 350 - 500 bp by capillary electrophoretic measurement (Agilent Bioanalyzer[®] or Tapestation[®]), the libraries were quantized QubitTM. The libraries were eventually assembled into a pool and standardized with 4 nM and then sent to the Institut Pasteur in Paris for sequencing.

The sequencing was carried out with the Illumina MiSeq in 2×300 bp pairedend mode. The data generated was then used for downstream bioinformatics analyses, including variant detection and comparative genomic evaluations.

2.3. Bioinformatics Analysis

The analysis began with quality control of the raw sequences using FastQC [10] to evaluate parameters such as Phred scores, the distribution of read lengths, and the presence of contaminants. The results were synthesized and visualized with MultiQC [11], providing an overview of sequence quality and facilitating the identification of anomalies to be corrected before subsequent analyses. The sequences were cleaned using Trim Galore (https://github.com/FelixKrueger/TrimGalore) which removes low-quality adapters and bases. This ensures that the data is reliable for the analyzed.

The process of extracting the parasitic sequences was done in two steps via BWA [12] and Samtools [13]: an alignment was performed on the human reference genome (HG19) to remove the human sequences, then a second alignment was made with the genome of *Plasmodium falciparum* (3D7) to retain only the parasitic sequences.

The BAM files (alignment file), have been marked with Picard [14], to correct the distortions associated with duplicate reads.

SNPs were then detected with FreeBayes [15], and then filtered with VCFtools [16] by applying strict quality criteria (DEP > 5, QUAL > 5). The results were recorded in a VCF (Variant Call Format) file containing detailed information about the extracted SNPs.

Finally, the functional annotation of the variants was executed with VEP (Variant Effect Predictor) [17], determining the potential impact of mutations on genes and proteins.

2.4. Genotypes Based SNP Matrix

The multivariate statistical analyses performed in this study are based on the construction of a coded matrix of SNPs. Genotypes extracted from generated bioinformatics VCF files were organized to obtain a matrix in which each row represents an individual while each column represents a SNP. In order to simplify the use of this matrix for downstream analyses, the genotypes, usually represented by traits such as (0/0) for the major homozygous genotype, (0/1) or (1/0) for the heterozygous genotype, and (1/1) for the minor homozygous genotype, were transformed into numerical values: 1 for (0/0), 2 for (0/1) or (1/0), and 3 for (1/1).

Let G be this coded matrix of genotypes of dimension $n \times m$, where n represents the number of individuals in the study, m represents the number of SNPs (Single Nucleotide Polymorphisms) analyzed and represents the coded genotype of individual *i* for SNPj, with $G_{ij} \in \{1, 2, 3\}$.

$$G = \begin{pmatrix} G_{11} & G_{12} & \cdots & G_{1m} \\ G_{21} & G_{22} & \cdots & G_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ G_{n1} & G_{n2} & \cdots & G_{nm} \end{pmatrix}_{n \times m}$$
(1)

2.5. Multivariate Statistical Analysis

Among the multivariate statistical methods used in our study, we should mention:

- Principal Component Analysis (PCA), performed on a matrix of coded SNPs for which an additional variable, the sickle cell status of the individual.
- PERMANOVA, used to verify the statistical significance of genetic differences between groups represented by sickle cell status.
- finally, the pair Jost D-index was used to detract significant SNPs and then evaluate the genetic differentiation between sickle cell status pairs.

All of these methods have been implemented with R, version 4.0.3 and RStudio, version 1.3.1056.

2.5.1. Principal Component Anaysis (PCA)

Prior to the analysis, the data were standardized to ensure that the variance of each SNP was zero (0) and the mean was one (1).

$$G_{cr} = \left(\frac{G_{ij} - \overline{G}_j}{s_j}\right)_{1 \le i \le n, 1 \le j \le m}$$
(2)

where $\overline{G}_j = \sum_{i=1}^m G_{ij}$ is the mean of the SNP *j*, $s_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (G_{ij} - \overline{G}_j)^2}$ the standard deviation of the SNP i and C_j the standard matrix

standard deviation of the SNP j and G_{cr} the standardized matrix.

The PCA was carried out using Σ , the covariance matrix calculated on normalized data, where *n* is the number of individuals, and G_{cr}^{T} is the transpose of the standardized matrix of *G*.

$$\Sigma = \frac{1}{n+1} G_{cr}^{\mathrm{T}} G_{cr} \tag{3}$$

This method simply consists of a decomposition of the covariance matrix into eigenvalues and eigenvectors, carried out via spectral decomposition, that is to say, the resolution of the following equation:

$$\Sigma v_j = \lambda_j v_j \tag{4}$$

where v_j is the eigenvector which is associated with the eigenvalue λ_j of the j^{th} principal component.

This leads to the diagonal matrix of eigenvalues Λ and the matrix V of eigenvectors:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}, \quad V = (v_1, v_2, \cdots, v_p)$$
(5)

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

PCA reduces the dimensionality of the data, but maintains its genetic variability among groups based on:

- the principal components (PC): axes defined by the eigenvectors.
- the Z matrix: coordinates of individuals in the reduced space of principal components.

$$Z = G_{cr}V \tag{6}$$

It thus facilitates the visualization of individuals and their structuring in the factorial space. In addition, it contributes to the identification of atypical individuals and clarifies the links between genotypes and clinical status.

2.5.2. PERMANOVA

PERMANOVA (Multivariate Analysis of Variance by Permutation) tests for significant differences between groups. It does this in multidimensional space using data permutations and does not require the use of assumptions such as normalities and homogeneities of variances. By applying random permutations of groups, this technique based on distance matrices makes it possible to evaluate the observed differences between groups in a more robust and flexible way.

Using a Euclidean distance, the first step is to calculate the distance matrix D between individuals.

$$D_{ij} = \sqrt{\sum_{k=1}^{m} \left(G_I - G_{jk}\right)^2}$$
(7)

where each element D_{ij} represents the distance between individuals *i* and *j*.

$$D = \begin{pmatrix} 0 & D_{12} & D_{13} & \cdots & D_{1n} \\ D_{21} & 0 & D_{23} & \cdots & D_{2n} \\ D_{31} & D_{32} & 0 & \cdots & D_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D_{n1} & D_{n2} & D_{n3} & \cdots & 0 \end{pmatrix}$$
(8)

The first step of the PERMANOVA involves the calculation of two inter and intra group variances, respectively.

The within-group variance for each pair of groups k and l is calculated as follows:

$$S_{w_k}^2 = \frac{1}{n_k (n_k - 1)} \sum_{i, j \in k} D_{ij}$$
(9)

where: $S_{w_k}^2$ is the within-group variance for group k, n_k is the number of individuals in group k and D_{ii} is the distance between individuals i and j.

The inter-group variance is calculated as follows:

$$S_{B_{kl}}^{2} = \frac{1}{n_{k}n_{l}} \sum_{i \in k, j \in l} D_{ij}$$
(10)

where: n_k and n_l are the number of individuals in groups k and l and D_{ij} is the distance between individuals *i* (from group k) and *j* (from group l).

The F statistic of this test is defined by the ratio of inter-group variance to

within-group variance.

$$F = \frac{S_B^2}{S_w^2} \tag{11}$$

PERMANOVA tests the significance of F by performing random permutations and recalculating the previous statement each time. The p-value is then gleaned from the percentage of permutations for which F is equal to or greater than the observed value.

If the p-value is less than an arbitrarily chosen value, it is customary to reject the assumption that there are no significant genetic differences between the study groups.

2.5.3. Jost's D

Jost's D measures genetic differentiation between two groups based on the allele frequencies of SNPs. To calculate this index, we need to know the allele frequencies of each SNP in each group.

Let $p_{A,j}$ be the frequency of the major allele for the *j* SNP in group A, and $p_{B,j}$ the frequency of the major allele for the *j* SNP in the B group. The Jost's D for each SNP is expressed as follows:

$$D_{j} = \frac{\left| p_{A,j} - p_{B,j} \right|}{1 - \left| p_{A,j} - p_{B,j} \right|}$$
(12)

where D_i is the genetic differentiation for SNPj, measured on a scale of 0 to 1.

The Jost's D average is calculated for all the *m* SNPs analyzed to calculate this index.

$$D_{\text{global}} = \frac{1}{m} \sum_{j=1}^{m} D_j \tag{13}$$

The overall Jost's D and SNPs Jost's D evaluate genetic differentiation between groups and SNPs. A high overall Jost's D, close to 1, indicates strong differentiation while values between 0.05 and 0.30 suggest moderate to substantial differentiation. SNPs with Jost's D \geq 0.15 are needed to identify the SNPs responsible for the difference between groups.

3. Results and discussion

3.1. Results

3.1.1. Descriptive Analysis of SNPs

Variant detection carried out as part of the bioinformatics analysis identified a total of 376,320 SNPs among the 30 *Plasmodium falciparum* strains studied.

Figure 2 shows a Venn diagram illustrating the distribution of SNPs according to the genotypic groups AA, AS, and SS. From this diagram, it can be seen that, of the 376,320 SNPs, 41,997 are specific to group AA, 18,289 to group AS and 17,276 to group SS. In addition, 11,873 SNPs are shared between the AA and AS groups, 13,464 between AA and SS, and 3878 between SA and SS, while 16,269 SNPs are common to all three groups. These results show a strong genetic heterogeneity.

They underline on the one hand the intrinsic variability of each genotype and, on the other hand, the existence of a genetic basis common to the three groups.



Figure 2. Venn diagram showing the distribution of SNPs according to the genotypes (AA, AS, and SS).

3.1.2. Results of Principal Component Analysis (PCA)

The Principal Component Analysis was performed on the matrix of coded SNPs, obtained from the data from this study, including sickle cell status as an additional variable. Applying the Kaiser criteria, the slope graph and the proportion of variance explained, the first seven components were kept (**Figure 3**).





Visualization in the first three factorial planes shows clear differentiations between the genotypic groups. The two individuals in group AA, particularly separated from their group, seem to be atypical. Two major genetic structures could be distinguished, one composed exclusively of AS individuals and the other of the group grouping AA and SS with a clear genetic differentiation between AS and other genotypes (**Figures 4-6**).



Figure 4. Representation of individuals in the first factorial plane.



Figure 5. Representation of individuals in the second factorial plane.



Figure 6. Representation of individuals in the third factorial plane.

3.1.3. PERMANOVA and JOST'D Results

The PERMANOVA method, with a p-value of 0.0436 (less than 0.05) and an Fstatistic of 2.4, indicates that the HbAA, HbAS, and HbSS genotypes affect the genetic divergence between the groups of individuals studied in this analysis.

The analysis of the pairwise Jost's D moderated between the genotypic groups HbAA and HbAS, as well as between HbAS and HbSS. Indeed, the Jost's D levels between HbAA and HbAS, and between HbAS and HbSS are 0.19 and 0.17, respectively. Which implies an intermediate status for the HbAS group, which is different compared to the homozygous AA and SS groups. Also, it is low between HbAA and HbSS (0.04), which means that HbAA is close to SS, but is different from HbAS. The analysis of Jost's D thus confirms that genotypic status here is an important factor in the genetic structuring of individuals.

Comparative analysis using Jost'D by SNP for HbAS and HbAA strains revealed 241 SNPs with significant genetic differentiation and associated medium to high Jost's D values > 0.15. A subset of these non-synonymous SNPs is located in genes functionally related to erythrocyte invasion and immune response modulation (PF3D7_0300200 - clag3.2, PF3D7_0100100 - acs7, PF3D7_1200600 - var2csa, PF3D7_0800100 - var1csa, PF3D7_0400400 - var).

On the other hand, the SS and AA analysis identified two significant SNPs, located in intergenic regions. Even if they are not present in coding sequences, the two variants could play a regulatory role, regulating the expression of adjacent genes involved in the virulence, metabolism and adaptation of the parasite. These results draw attention to the potential role of non-coding elements in adaptive processes, particularly in the context of SS red blood cells, which are marked by high oxidative stress and disruption of membrane structure.

3.2. Discussion

This study sought to further investigate the effect of the hemoglobin status of the

human host on the genetic structuring of *Plasmodium falciparum* in a malariaendemic context, using a population genomics approach. The conclusions obtained provide solid new evidence concerning the adaptive dynamics of the parasite in response to the erythrocyte pressures imposed by sickle cell anemia.

The joint use of PCA, PERMANOVA and Jost's differentiation index led to the discovery of internal genetic patterning patterns. This combination of multivariate methods was essential to estimate the effect of host hemoglobin status on *P. falciparum* diversity. PCA showed a partial but very clear separation of HbAS-related strains and PERMANOVA confirmed the important contribution of host genotype to genomic variance. Such a methodological choice was, therefore, consistent with that currently used in the analysis of population genomics data to explore adaptation signals under complex epidemiological conditions [18]-[22].

All the discriminating mutations identified in this study, including those affecting genes known for their role in erythrocyte invasion and modulation of the immune response of the host clag3.2, acs7, rh5, var2csa, var1csa, PfEMP1, conclude that host-parasite co-evolution occurs during a *Plasmodium falciparum* - human host interaction. This adaptive theme is supported by an evolutionary drift between *Plasmodium falciparum* and the human host. These processes evolve over the long term in endemic areas where the parasite is subject to specific selection pressures that exclusively concern hemoglobin status. In addition, a significant proportion of SNPs could also be found in intergenic vegetation and can possibly affect the regulation of transcription in the parasite [6] [23]-[26].

In the context of molecular surveillance of malaria, these observations have important practical implications. They imply the need to take into account the genetic diversity of human populations in the design of malaria control strategies. In addition, practices confirm the value of integrating data into common databases, such as MalariaGEN [27] or PlasmoDB [28], for longitudinal monitoring of adaptive variants. This is to be more foreseeable in relation to the risks of vaccine or therapeutic escape. Notably, the variability of diversity in the multigene families of var and rifin could reduce the effect of some interventions based on antigenic recognition.

Although the present work is not directly in line with the specific initial objectives of the DREPAL project, this study can prove to be a relevant valorization of the material and biological resources that have been generated in this framework. As a result of this additional focus, access to the data has been partially restricted. However, it can be noted that even if the sample size obtained is relatively small (n = 30), the main objectives of the study have nevertheless been achieved: it is to develop an accessible, reproducible methodological approach capable of generating exploratory data on mutations in the *Plasmodium falciparum* genome in sickle cell hosts. At this stage, the work is mainly functionally descriptive. The biological implications of the identified SNPs, in particular their potential effects on protein expression, structure or function, will be investigated by means of experimental analyses in subsequent work. In this context, the integration of multiomics data

(transcriptomics, proteomics, metabolomics), coupled with molecular dynamics simulation methods, seems very promising in refining the functional characterization of the identified mutations. This type of approach has already been successfully used to understand the mechanisms of adaptation of the parasite to different hostile cellular environments. In addition, evolutionary-specific experimental models with HbAA, HbAS, and HbSS erythrocytes are needed to test the adaptation hypothesis. Finally, better integration of epigenomic and immune data would be richer in information on the underlying mechanisms of host-parasite coevolution.

4. Conclusion

Overall, this work revealed a strong genetic structuring of *Plasmodium falciparum* as a function of the host's hemoglobin state, identifying signatures of adaptation to disturbed intracellular environments related to sickle cell disease. Through the use of rigorous statistical methods such as PCA, PERMANOVA and Jost's D, coupled with a focus on coding and non-coding SNP data, a number of differentiating mutations in relevant functional genes were found. This suggests that genetic plasticity in the parasite is influenced by selective pressures resulting from the high levels of HbAS and HbSS red blood cells, in conjunction with a host/parasite coevolution model. In addition to the fundamental aspect, this information could have direct applications in the field of malaria genomics and the improvement of control strategies, particularly in regions heavily affected by haemoglobinopathies. In addition, the incorporation of this knowledge into global molecular monitoring programs and the deployment of large-scale functional and multi-omics approaches will be essential to further identify adaptive mechanisms. In the long term, these initiatives could pave the way for more targeted control strategies that take into account the genetic variability of the parasite and its host.

Acknowledgements

We would like to thank the Paludology Unit and the Molecular Genetics Platform (PFGM) of the Institut Pasteur de Côte d'Ivoire for their methodological support and the provision of their study resources. We extend our sincere thanks to Drs. Diplo Bernadette and Akoguhi N'guessan Patrice for their comments and constructive criticisms that have strengthened this book. Our thanks also go to the National Computing Center of Côte d'Ivoire (CNCCI) for providing the IT infrastructure that simplified the analysis of NGS data.

Financing

This study received funding from Rotary International.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Authors' Contribution

SEA designed the bioinformatics analysis pipelines, analyzed the data, wrote and approved the final version of the manuscript. SYO contributed to the design of the pipelines and data analysis. ABA, BMK, and AAG analyzed the data and reviewed drafts of the manuscript. OAT, KJA, and RJ reviewed the drafts of the manuscript and approved its final version. Pipeline Data and Analysis The data and analysis pipelines generated for this study are available upon request to authors.

References

- [1] WHO (2025) World Malaria Report 2023. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023
- Weiss, D.J., Dzianach, P.A., Saddler, A., Lubinda, J., Browne, A., McPhail, M., et al. (2025) Mapping the Global Prevalence, Incidence, and Mortality of *Plasmodium Falciparum* and Plasmodium Vivax Malaria, 2000–22: A Spatial and Temporal Modelling Study. *The Lancet*, **405**, 979-990. <u>https://doi.org/10.1016/s0140-6736(25)00038-8</u>
- [3] Crompton, P.D., Traore, B., Kayentao, K., Doumbo, S., Ongoiba, A., Diakite, S.A.S., et al. (2008) Sickle Cell Trait Is Associated with a Delayed Onset of Malaria: Implications for Time-to-Event Analysis in Clinical Studies of Malaria. The Journal of Infectious Diseases, 198, 1265-1275. <u>https://doi.org/10.1086/592224</u>
- [4] Archer, N.M., Petersen, N., Clark, M.A., Buckee, C.O., Childs, L.M. and Duraisingh, M.T. (2018) Resistance to *plasmodium Falciparum* in Sickle Cell Trait Erythrocytes Is Driven by Oxygen-Dependent Growth Inhibition. *Proceedings of the National Academy of Sciences*, **115**, 7350-7355. <u>https://doi.org/10.1073/pnas.1804388115</u>
- [5] Gwarinda, H.B., Tessema, S.K., Raman, J., Greenhouse, B. and Birkholtz, L. (2023) Population Structure and Genetic Connectivity of *Plasmodium Falciparum* in Pre-Elimination Settings of Southern Africa. *Frontiers in Epidemiology*, **3**, Article 1227071. <u>https://doi.org/10.3389/fepid.2023.1227071</u>
- [6] Hotelling, H. (1933) Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, 24, 417-441. <u>https://doi.org/10.1037/h0071325</u>
- [7] Pearson, K. (1901) On Lines and Planes of Closest Fit to Systems of Points in Space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2, 559-572. <u>https://doi.org/10.1080/14786440109462720</u>
- [8] Verity, R. and Nichols, R.A. (2014) What Is Genetic Differentiation, and How Should We Measure It—G_{ST}, D, Neither or Both? *Molecular Ecology*, 23, 4216-4225. <u>https://doi.org/10.1111/mec.12856</u>
- [9] Anderson, M.J. (2017) Permutational Multivariate Analysis of Variance (PER-MANOVA). In: Kenett, R.S., Longford, N.T., Piegorsch, W.W. and Ruggeri, F., Eds., *Statistics Reference Online*, Wiley, 1-15. https://doi.org/10.1002/9781118445112.stat07841
- [10] Andrews, S. (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data. <u>https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</u>
- [11] Ewels, P., Magnusson, M., Lundin, S. and Käller, M. (2016) Multiqc: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. *Bioinformatics*,

32, 3047-3048. https://doi.org/10.1093/bioinformatics/btw354

- [12] Li, H. and Durbin, R. (2009) Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25, 1754-1760. <u>https://doi.org/10.1093/bioinformatics/btp324</u>
- [13] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009) The Sequence Alignment/Map Format and SAMtools. *Bioinformatics*, 25, 2078-2079. <u>https://doi.org/10.1093/bioinformatics/btp352</u>
- [14] Broad Institute (2021) Picard: Tools for Working with High-Throughput Sequencing Data. <u>https://broadinstitute.github.io/picard/</u>
- [15] Garrison, E. and Marth, G. (2012) Haplotype-Based Variant Detection from Short-Read Sequencing. <u>https://doi.org/10.48550/ARXIV.1207.3907</u>
- [16] Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., *et al.* (2011) The Variant Call Format and VCFtools. *Bioinformatics*, 27, 2156-2158. <u>https://doi.org/10.1093/bioinformatics/btr330</u>
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., *et al.* (2016) The Ensembl Variant Effect Predictor. *Genome Biology*, **17**, Article No. 122. https://doi.org/10.1186/s13059-016-0974-4
- [18] Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies. *Nature Genetics*, **38**, 904-909. <u>https://doi.org/10.1038/ng1847</u>
- [19] Ping, Z., Yang, Z., Cheng, Q., Liwei, Z., Ruyang, Z., Jianwei, G., *et al.* (2015) Statistical Analysis for Genome-Wide Association Study. *The Journal of Biomedical Research*, 29, 285-297. <u>https://doi.org/10.7555/jbr.29.20140007</u>
- [20] Russell, J.T., Roesch, L.F.W., Ördberg, M., Ilonen, J., Atkinson, M.A., Schatz, D.A., et al. (2019) Genetic Risk for Autoimmunity Is Associated with Distinct Changes in the Human Gut Microbiome. Nature Communications, 10, Article No. 3621. https://doi.org/10.1038/s41467-019-11460-x
- [21] Mounger, J.M., van Riemsdijk, I., Boquete, M.T., Wagemaker, C.A.M., Fatma, S., Robertson, M.H., *et al.* (2022) Genetic and Epigenetic Differentiation across Intertidal Gradients in the Foundation Plant Spartina Alterniflora. *Frontiers in Ecology and Evolution*, **10**, Article 868826. <u>https://doi.org/10.3389/fevo.2022.868826</u>
- [22] Jost, L., Archer, F., Flanagan, S., Gaggiotti, O., Hoban, S. and Latch, E. (2018) Differentiation Measures for Conservation Genetics. *Evolutionary Applications*, **11**, 1139-1148. <u>https://doi.org/10.1111/eva.12590</u>
- [23] Shaul, O. (2017) How Introns Enhance Gene Expression. *The International Journal of Biochemistry & Cell Biology*, 91, 145-155. <u>https://doi.org/10.1016/j.biocel.2017.06.016</u>
- [24] Gupta, A.P. and Bozdech, Z. (2017) Epigenetic Landscapes Underlining Global Patterns of Gene Expression in the Human Malaria Parasite, *Plasmodium Falciparum*. *International Journal for Parasitology*, **47**, 399-407. https://doi.org/10.1016/j.ijpara.2016.10.008
- [25] Jabeena, C.A. and Rajavelu, A. (2019) Epigenetic Players of Chromatin Structure Regulation in *Plasmodium falciparum. ChemBioChem*, 20, 1225-1230. https://doi.org/10.1002/cbic.201800718
- [26] Taqi, M.M., Waseem, D., Ismatullah, H., Haider, S.A. and Faisal, M. (2016) In Silico Transcriptional Regulation and Functional Analysis of Dengue Shock Syndrome Associated SNPs in *PLCE1* and *MICB* Genes. *Functional & Integrative Genomics*, 16,

335-345. https://doi.org/10.1007/s10142-016-0489-9

- [27] Abdel Hamid, M.M., Abdelraheem, M.H., Acheampong, D.O., Ahouidi, A., Ali, M., Almagro-Garcia, J., *et al.* (2023) Pf7: An Open Dataset of *Plasmodium Falciparum* Genome Variation in 20,000 Worldwide Samples. *Wellcome Open Research*, 8, 22. https://doi.org/10.12688/wellcomeopenres.18681.1
- [28] Fraunholz, M.J. and Roos, D.S. (2003) PlasmoDB: Exploring Genomics and Post-Genomics Data of the Malaria Parasite, *Plasmodium falciparum. Redox Report*, 8, 317-320. <u>https://doi.org/10.1179/135100003225002961</u>