

# Whole genome sequencing to inform the epidemiology of *Plasmodium falciparum* malaria in the elimination setting of Malaysia

Received: 20 July 2025

Accepted: 11 February 2026

Published online: 20 February 2026

Cite this article as: Tan M.K.I., Billows N., Divis P.C.S. *et al.* Whole genome sequencing to inform the epidemiology of *Plasmodium falciparum* malaria in the elimination setting of Malaysia. *BMC Genomics* (2026). <https://doi.org/10.1186/s12864-026-12653-7>

Mark K. I. Tan, Nina Billows, Paul C. S. Divis, Cyrus Daneshvar, Jonathan Edgeworth, Janet Cox Singh, Susana Campino & Taane G. Clark

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

## Whole genome sequencing to inform the epidemiology of *Plasmodium falciparum* malaria in the elimination setting of Malaysia

### Authors:

Mark KI Tan <sup>1,2,3</sup>	<a href="mailto:Mark.Tan@lshtm.ac.uk">Mark.Tan@lshtm.ac.uk</a>
Nina Billows <sup>1</sup>	<a href="mailto:Nina.Billows@lshtm.ac.uk">Nina.Billows@lshtm.ac.uk</a>
Paul C. S. Divis <sup>4</sup>	<a href="mailto:pcsimon@unimas.my">pcsimon@unimas.my</a>
Cyrus Daneshvar <sup>5</sup>	<a href="mailto:cyrus.daneshvar@nhs.net">cyrus.daneshvar@nhs.net</a>
Jonathan Edgeworth <sup>2,3</sup>	<a href="mailto:Jonathan.Edgeworth@nhs.net">Jonathan.Edgeworth@nhs.net</a>
Janet Cox Singh <sup>6,*</sup>	<a href="mailto:jcs26@st-andrews.ac.uk">jcs26@st-andrews.ac.uk</a>
Susana Campino <sup>1,*</sup>	<a href="mailto:Susana.campino@lshtm.ac.uk">Susana.campino@lshtm.ac.uk</a>
Taane G. Clark <sup>1,7,*</sup>	<a href="mailto:taane.clark@lshtm.ac.uk">taane.clark@lshtm.ac.uk</a>

1 Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine (LSHTM), London, WC1E 7HT, UK

2 Department of Infectious Diseases, Faculty of Life Sciences and Medicine, King's College London, London, UK

3 Centre for Clinical Infection and Diagnostics Research, Department of Infection, Guy's & St. Thomas' NHS, London, SE1 7EH, UK

4 Malaria Research Centre, Faculty of Medicine and Health Sciences, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia

5 University Hospitals Plymouth NHS Trust, Plymouth, UK

6 School of Medicine, University of St Andrews, North Haugh, St Andrews, KY16 9TF, UK

7 Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, LSHTM, London, WC1E 7HT, UK

\* joint authors

Corresponding:

Professor Taane G. Clark

Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, UK

[Taane.clark@lshtm.ac.uk](mailto:Taane.clark@lshtm.ac.uk)

## **ABSTRACT**

### **Background**

Imported malaria cases, driven by human migration and travel, pose a significant challenge to malaria elimination efforts. Genomic approaches have become essential for distinguishing between local transmission and imported infections. The state of Sarawak, Malaysia, provides a pertinent example of a malaria-eliminating setting under pressure from *Plasmodium* parasite importation.

### **Results**

In this study, we analysed 21 *Plasmodium falciparum* isolates obtained from archived whole blood samples collected between 2008 and 2010 and compared them to 9,518 publicly available isolates from Central Africa (518), East Africa (849), Horn of Africa (25), Oceania (349), South America (75), South Asia (404) Southeast Asia (5,182) and West Africa (2,116). By applying nanopore sequencing and population genomic analyses, we found that most of the cases (n=13/15) likely originated from endemic regions outside Malaysia, supported by patient travel histories and high multiplicity of infection levels. These findings and drug resistance profiles are consistent with the historical epidemiology of the suspected source regions. Notably, two cases showed genomic evidence of origins inconsistent with the patients' reported travel histories, underscoring the limitations of traditional epidemiological methods. Identity-by-descent analysis revealed clustering in only two cases, indicating that the majority of infections were likely isolated introductions rather than evidence of sustained local transmission.

### **Conclusion**

Overall, our results highlight the power of malaria genomics in discerning imported from locally acquired cases and emphasise its critical role in

maintaining malaria elimination, particularly in regions situated along major migration and labour exchange corridors.

Keywords: *Plasmodium falciparum*, malaria, nanopore sequencing technology, imported diseases, Malaysia, molecular epidemiology, population surveillance

ARTICLE IN PRESS

## INTRODUCTION

Malaria, caused by *Plasmodium* parasites and transmitted by *Anopheles* mosquitoes, has a high global burden, with over 282 million estimated cases and 610,000 associated deaths in 2024 alone (1). Despite these challenges, progress towards elimination has been significant. A total of 47 countries have been certified malaria-free by the World Health Organization (WHO) - a designation granted to countries that demonstrate interruption of local malaria transmission for at least three consecutive years and maintain a robust surveillance system to prevent re-establishment (1). Malaria control measures, such as the use of insecticides, insecticide-treated nets, and antimalarial medications like artemisinin-based combination therapies, have significantly reduced the disease burden. However, challenges, including zoonotic malaria, drug resistance, and climate change, are hindering further progress toward elimination (1).

Malaria elimination is focused on controlling the local transmission of human-host adapted *Plasmodium* species. Paths to malaria elimination can be undermined by the importation of cases from endemic regions into increasingly immune-naïve populations, formerly endemic for malaria. Importation can be driven by migration, travel, and re-introduction through asymptomatic parasite reservoirs (2). A good example of malaria importation is southeastern Iran, where indigenous transmission was limited, but has experienced malaria resurgence linked to importation from neighbouring countries (3). In Malaysia, where *Plasmodium falciparum* transmission has been largely interrupted through sustained control efforts, the risk of re-establishment persists due to the influx of migrant workers from malaria-endemic countries (4). Similarly, countries such as Cambodia, Nepal, and Bhutan, despite substantial declines in

malaria incidence, continue to face challenges from the sporadic detection of indigenous *P. falciparum* cases in areas previously considered malaria-free (2).

National malaria control programmes (NMCPs) are using genomics tools to investigate outbreaks (5). For example, imported cases of malaria into non-endemic countries, including the UK, are being investigated by applying genomics-based tools to identify *Plasmodium* species, drug resistance markers and the likely source of infections (6). Whole-genome (WGS) or targeted amplicon sequencing data can provide insights into drug resistance, multiplicity of infection linked to transmission intensity, and the geographical source of infections that can identify importation events, all supported by informatics tools (6). Population genomic analysis elucidates population structure and ancestry, transmission events and parasite relatedness through identity-by-descent (IBD), and intervention effectiveness (7). Decreasing sequencing costs, including developments in Oxford Nanopore Technologies (ONT) portable sequencing platforms, as well as selective whole genome amplification (SWGA) of DNA for low parasitaemia infections (8), are leading to NMCPs investing in sequencing technologies to improve genomic resolution to characterise spatial and temporal trends in drug resistance markers and population structure (9).

Within the WHO Western Pacific Region, malaria cases have declined from 2,678,000 in 2000 to 1,747,000 in 2023 (1). Malaysia has contributed significantly to this reduction and, as of 2018, has reported zero indigenous cases of non-zoonotic malaria (10). Excluding cases of zoonotic malaria caused by *P. knowlesi*, monoclonal *P. vivax* infections were the predominant malaria cases reported in Malaysia before elimination (11). However, as expected *P. falciparum* was associated with a higher average case fatality rate, 1.41%

compared to just 0.04% for *P. vivax* (11). Imported malaria made up 17.7% of all cases in Malaysia from 2013 to 2017 (11). This is partly due to Malaysia's reliance on migrant workers, especially from countries in the wider Asia region (12). The majority of the migrant worker population originates from neighbouring high-risk countries, such as Indonesia and Thailand, where multiple drug-resistant malaria strains, including artemisinin-resistant *P. falciparum* from the Greater Mekong Subregion, have emerged (13, 14). These populations represent a potential reservoir for the reintroduction of malaria.

Although non-zoonotic malaria has been eliminated in Malaysia, the zoonotic species *P. knowlesi* continues to pose a public health challenge, particularly in East Malaysia, where 2,879 related cases were reported in 2023 (1). The transmission of *P. knowlesi* is positively associated with forest cover and historical forest loss, implicating deforestation driven by agriculture and forestry as key drivers of transmission from the macaque monkey host to humans (15). The presence of well-established competent vectors underscores the real risk of re-introducing *P. falciparum* malaria transmission in the country, and the fragility of sustaining malaria elimination. The continued decline in malaria cases can lead to reduced public awareness, diminished health system preparedness, and potentially waning population immunity (16, 17). Given the severe health consequences posed by *P. falciparum*, particularly among immunologically naïve populations (18), preventing its re-establishment, especially of drug-resistant strains, is an urgent public health priority. To inform such efforts, it is crucial to understand the historical *P. falciparum* population structure and dynamics. This requires the integration of WGS, population genomic analyses (7, 19), and advanced AI-based profiling approaches (20).

We performed ONT WGS and population genomics analysis on *P. falciparum* isolates from 21 patients with travel histories, sampled at two tertiary hospitals in Sarawak, Malaysia, between 2008 and 2010. Our objective was to assess the utility of genomic data in reconstructing the historical epidemiology of *P. falciparum* cases in the region. This work contributes to the development of a genomic reference database and analytical framework to support future whole-genome or amplicon-based surveillance efforts for malaria control in Malaysia and the wider region.

## METHODS

### *Sample collection, DNA extraction, and sequencing*

A total of twenty-one whole blood samples (WBS) were collected from patients diagnosed with *P. falciparum* malaria and treated at Hospital Sarikei and Hospital Sibu in Sarawak, Malaysia, between 2008 and 2010 (21). Nationwide, 21,330 malaria cases were diagnosed during this period. Informed consent, including that to use residual samples for related studies, was obtained from all participants, and individual travel histories were recorded.

Initial case identification and species determination were performed by microscopy and subsequently confirmed by quantitative PCR (qPCR). Genomic DNA was extracted from all twenty-one samples using the Monarch® Genomic DNA Purification Kit (NEB, UK), which demonstrated broadly superior performance compared to the DNeasy® Blood and Tissue Kit (Qiagen, UK) in an initial evaluation of twelve samples (**Figure S1, Table S1**). The resulting DNA underwent SWGA using EquiPhi29™ DNA polymerase (Thermo Fisher Scientific) and primer sets previously described (22). DNA concentration was measured using Qubit™ fluorometry, and sequencing libraries were prepared

with the Native Barcoding Kit 24 V14 (SQK-NBD114.24, ONT, UK), following the manufacturer's protocol. Isolates were multiplexed and sequenced on R10.4.1 flow cells for 24 to 48 hours via The Applied Genomics Centre at the London School of Hygiene & Tropical Medicine (genomics.lshtm.ac.uk).

### *Bioinformatics and genotypic profiling*

Raw signal sequence data (.pod5 files) underwent basecalling and demultiplexing using *dorado* (v0.7.3) software (ONT, UK) with the superior accuracy model (dna\_r10.4.1\_e8.2\_400bps\_sup@v5.0.0). Raw reads were trimmed using the in-built function in *dorado*, and alignment to the *P. falciparum* 3D7 (PF3D7; PlasmoDB; (23)) reference genome was performed using *minimap2* software (v2.1.1-r341) (24). Raw sequence reads were also quality-checked by *pycoQC* (v2.5.2) and *NanoStat* (v1.6.0) (25). Assembly depth across the core *P. falciparum* genome (26), excluding subtelomeric regions, was quantified using *mosdepth* software (v0.3.3) (27). Isolates with less than 50% genome coverage at 5-fold coverage depth were removed from further analysis.

Raw *fastq* files were analysed using *Malaria-Profiler* software (v.0.0.10) (6), leading to predictions of likely geographical region source (Oceania, Southeast Asia, or Central Africa) and genotypic drug resistance. Raw sequencing data are available from the ENA archive (see **Table S2** for a list of accession numbers). Variant calling of ONT data was performed by *Clair3* (v1.0.10) software with model r1041\_e82\_400bps\_sup\_v500 to create a GVCF output (28). The SNP data obtained using Clair3 in GVCF format were merged with a database of 9,518 *P. falciparum* isolates from Central Africa (518), East Africa (849), Horn of Africa (25), Oceania (349), South America (75), South Asia (404), Southeast Asia (5,182) and West Africa (2,116) that were previously sequenced using an

Illumina platform, obtained from the Pf7 data release (29). The merge was performed using GATK's GenomicsDBImport (30). In total, 1,178 isolates (22.69%) were sequenced and assembled between 2008 and 2010, with a further 7,847 sequenced after 2010. The resulting variant file was filtered through an in-house pipeline ([LSHTMPathogenSeqLab/fastq2matrix](https://github.com/LSHTMPathogenSeqLab/fastq2matrix)). In brief, GATK variant filtration (30; best practices), biallelic SNP selection, and *bcftools* (31) consequence calling for gene annotation, and 1% minor allele frequency filtering were applied (**Table S3**). The filtered variant file was taken forward for population genomic analysis, which excluded SNPs with high genotype missingness (threshold of 40%).

#### *Population genomic analyses*

The population structure of the isolates was analysed using principal component analyses based on a SNP distance matrix generated using PLINK (v1.90b6.21) software (32) (<https://github.com/LSHTMPathogenSeqLab/malaria-hub>).

Multiplicity of infection (MOI) was estimated by *Moimix* (v0.0.2) (33), which calculates Wright's inbreeding coefficient ( $F_{WS}$ ), with values  $> 0.95$  indicating monoclonality. Identity-by-descent (IBD) was estimated from SNP data using *hmmIBD* (v2.1.0), and was restricted to samples classified as monoclonal ( $F_{WS} > 0.95$ ) (34). Sample clusters were inferred for isolates that shared a minimum of 90% identity. IBD was estimated between Malaysian ( $n = 6$ ) and monoclonal samples from Pf7 ( $n = 7,134$ ).

## **RESULTS**

#### *Patient and isolate characteristics*

Twenty-one WBS containing *P. falciparum* were included in this study, collected from patients at Sibul (n = 17) and Sarikei (n = 4) district hospitals in Sarawak,

Malaysia, in 2008 (n=11), 2009 (n=7) and 2010 (n = 3). The patients ranged in age from 19 to 52 years and were predominantly male (19/21; 90.5%), with the majority belonging to the Dayak indigenous communities (13/21; 61.9%) (**Table 1**). Parasitaemia levels varied from 0.01% to 8.8%, with a mean of 1.74% and a median of 1.0%. Travel history data indicated that eight patients had recently returned from Africa (Democratic Republic of the Congo (n = 2), Ghana (n = 1), South Africa (n = 1), other (n = 4)) and five from Oceania (Papua New Guinea (n = 4), Western New Guinea (n = 1)). Six patients reported no recent travel, and two others provided no travel history.

#### *Sequencing and genotypic profiles*

The twenty-one *P. falciparum* DNA samples were sequenced using the ONT PromethION platform. The median sequencing depth across genome positions was 9.12-fold. The median percentage of the core genome covered at a minimum depth of 5-fold was 72.8% (range: 0.3% to 98.3%) (**Table S4**). Of the 21 samples, 15 met the predefined threshold of  $\geq 50\%$  genome coverage at 5-fold depth with  $\leq 40\%$  missing genotypes and were included in downstream population genomic analyses.

Drug resistance profiling was performed using *Malaria-Profiler* software.

Twelve of the 15 isolates harboured one or more drug resistance-associated genotypes (**Table 2, Table S5**). Chloroquine resistance was the most frequently observed, with ten isolates carrying at least one mutation in the *pfcr*t gene, and five of these also possessed mutations in *pfmdr*1. The most common *pfcr*t mutation was A220S (8/10 isolates), while N86Y was the dominant *pfmdr*1 variant (4/5). Resistance-associated mutations in *pfppk-dhps* (sulfadoxine) and *pfdhfr-ts* (pyrimethamine) were identified in seven

isolates each. The most prevalent markers were *pfpppk-dhps* A437G and *pfdhfr-ts* S108N. Five isolates (Pf\_06, Pf\_07, Pf\_11, Pf\_12, and Pf\_17) carried genotypes associated with resistance to both chloroquine and sulfadoxine-pyrimethamine (SP). In contrast, three isolates (Pf\_09, Pf\_15, Pf\_20) displayed no known resistance mutations and were classified as drug-sensitive. No artemisinin resistance-associated genotypes were detected.

#### *Genotypic source, travel histories, and population structure*

Based on the AI-driven geographical prediction model implemented in *Malaria-Profiler* software, the fifteen *P. falciparum* isolates were classified as originating from Central Africa (n = 4), Southeast Asia (n = 5), and Oceania (n = 6) (**Table 2**). When compared with recorded patient travel histories, the predicted geographical origins were concordant for all but two isolates. Isolate Pf\_05 was predicted to originate from Southeast Asia, despite the patient having a travel history to Africa. Conversely, isolate Pf\_06, from a patient with no reported travel history, was classified as originating from Central Africa.

To further contextualise the origin of these isolates, we positioned them within a global *P. falciparum* reference dataset (n = 9,518). A total of 4,685,356 SNPs were identified from the initial dataset, including 4,391,371 biallelic SNPs across 14 nuclear chromosomes. Following filtering, 1,785,208 biallelic SNPs remained for further analysis. Principal component analysis (PCA) using multidimensional scaling revealed the expected clustering of isolates by geographic region, with some overlap between isolates from Oceania and Southeast Asia. The Malaysian isolates clustered by their *Malaria-Profiler* predicted origins (**Figure 1**), reinforcing the validity of the AI-based classifications. Notably, the overlapping positions of some Southeast Asian and

Oceanian isolates may reflect the geographic proximity of Sarawak to neighbouring regions, including the Indonesian state of Kalimantan, and the potential for cross-border malaria transmission or shared parasite ancestries.

#### *Malaysian isolates presented with potential multiclonal infections*

To investigate the likely transmission intensity at the sites of infection, the within-host diversity metric  $F_{WS}$  was estimated. Values greater than 0.95 were considered indicative of monoclonal infections. The estimated  $F_{WS}$  scores ranged from 0.70 to 0.97, with a mean of 0.91, suggesting that many infections likely originated from areas with moderate to high transmission. Evidence of polyclonality was observed in nine of the fifteen Malaysian isolates, each exhibiting  $F_{WS}$  values  $\leq 0.95$  (**Figure S2, Table 2**).

Stratified by predicted geographical origin, most monoclonal isolates were assigned to Southeast Asia (4 of 6), with the remainder from Oceania and Central Africa. The predominance of monoclonality among Southeast Asian isolates aligns with known low transmission intensities and declining *P. falciparum* prevalence in the region, particularly in countries such as Thailand and Vietnam, which were actively pursuing malaria elimination during the study period (35). In contrast, the higher prevalence of polyclonal infections among isolates associated with Central Africa and Oceania is consistent with historically and currently high malaria transmission intensities and large *P. falciparum* parasite reservoirs in those regions (1).

*falciparum* prevalence in the region, particularly in countries such as Thailand and Vietnam, which were actively pursuing malaria elimination during the study period (35). In contrast, the higher prevalence of polyclonal infections among isolates associated with Central Africa and Oceania is consistent with historically and currently high malaria transmission intensities and large *P. falciparum* parasite reservoirs in those regions (1).

#### *Identity-by-descent analysis*

Segments of IBD were subsequently used to assess genetic relatedness between monoclonal isolates from Malaysia (n = 6) and those from eight geographical

regions (n = 7,134) (**Figure 2**). Overall, comparisons between samples from Malaysia revealed high pairwise IBD fractions (median: 0.11, range: 0.07-0.94). Samples from Malaysia showed low IBD with isolates from the same timeframe in the database (median pairwise IBD < 0.1 across all eight regions; **Figure 2**). This was higher than other regions including West Africa (median: 0.02, range: 0-1), Central Africa (median: 0.03, range: 0-1), East Africa (median: 0.03, range: 0-1), Southeast Asia (median: 0.05, range: 0-1) and South Asia (median: 0.06, range: 0-0.99), but similar to Oceania (median: 0.16, range: 0.05-1), consistent with six samples predicted to be from this region (**Figure 2**). Two isolates, Pf\_08 and Pf\_13 had the highest IBD proportions (90.3%), suggesting possible shared common ancestry. The two patients from whom these isolates were obtained reported no travel history, were of Malaysian ethnicity (Bidayuh and Malay), and genomic analyses classified the isolates as originating from Southeast Asia. This result supports the likelihood that the infections were acquired locally, suggesting ongoing transmission within the region. Although both isolates were genotypically resistant to chloroquine, differences in their resistance genotypes point to potentially distinct drug selective pressures.

## DISCUSSION

Malaysia is no longer considered a malaria-endemic country with zero non-zoonotic indigenous cases reported for 6 consecutive years (2018-2023) (1). Population mobility poses a significant challenge to preventing malaria re-establishment and threatens Malaysia's recent successes, particularly in East Malaysia, through disease reintroduction via importation-mediated outbreaks derived from the global demand for skilled workers in agricultural, forestry and industrial sectors (36). In this study, we leveraged long-read nanopore sequencing to perform retrospective genomic surveillance from archived WBS.

Through this analysis, it was possible to elucidate the drug resistance prevalence, demographics of potential imported malaria cases, and the approximate regional transmission intensities within that time frame; all of which are vital towards securing malaria elimination, especially in preventing re-introduction. Such knowledge has proven important towards malaria control programmes, including tailoring drug strategies (13).

Geographical classification was derived from the genomic information from the Malaysian archived isolates and identified potential case importations. Frequent travels for employment have often been cited as a major mode for malaria influx from high malaria transmission zones (37, 38). Most cases were found to cluster closely with their geographical counterpart, particularly for those of possible Central African origins. Other studies conducted in Malaysia further reinforce these findings (10, 12, 39), with Papua New Guinea, Indonesia, Nigeria and Pakistan being sources of imported cases (10). Two cases were contradictory to the patients' travel histories but could be explained by exposure to local strains or inaccurate self-reported travel. The high accuracy of malaria genomic-based geographical profiling, which can distinguish local from imported infections, provides a valuable complement to traditional epidemiological approaches that rely on self-reported travel histories - data that are often incomplete or unreliable, yet remain the primary source guiding immediate clinical management and public health responses (38). As sequencing data continues to grow, genomic-based predictions can be refined to the country level, helping to resolve ambiguities between Southeast Asian and Oceanian isolates. Crucially, the detection of imported drug-resistant parasites, including *pfkelch13* mutations that vary regionally, will provide critical insights for targeted infection control efforts. Similarly, our estimates of polyclonality, or multiplicity

of infection, reflected the transmission intensity in the inferred source countries at the time of collection (40), and further supported regional geographical classifications, such as the higher transmission levels observed in Africa.

From the drug resistance profiles, the majority of the Malaysian isolates in this study were found to carry resistance genotypes to chloroquine, sulfadoxine, and pyrimethamine, including presently common mutant alleles in *pfmdr1* and *pfcr1* genes. The drug resistance profiles from this study follow the documented malaria epidemiology and national drug policies of the regions (41).

Interestingly, no artemisinin resistance genotypes were found in the isolates despite their strong presence throughout Southeast Asia during the collection period (42). However, the sample size is small, and there are biological and fitness cost barriers hindering the spread of artemisinin resistance (43), especially through importation.

Importation events can also be corroborated through ancestry clustering using IBD analysis, which revealed no genetic relatedness among the isolates except for Pf\_08 and Pf\_13, suggesting that these two may belong to a common transmission cluster. Although the sampling was not exhaustive, the overall findings suggest that most of the sequenced isolates likely represent independent transmission events over the study period. Similar patterns have been reported in other studies, where *P. falciparum* population genetic heterogeneity is more often associated with importation, whereas genetic relatedness typically reflects ongoing local transmission (40). The lack of clustering between the isolates and publicly available sequences is likely due to the absence of representative samples from the actual source regions of the

infections. This underscores the need for broader genomic surveillance in malaria-endemic regions, particularly among at-risk groups.

Moreover, whilst we have successfully demonstrated the utility of ONT sequencing to inform the epidemiology and surveillance of *P. falciparum* during the pre-elimination stage in Malaysia, future improvements in ONT sequencing yield and genomic coverage will be essential to enhance such approaches in the future. This may explain the minor underlying differences in geographical classification and travel history reported. Additionally, the nanopore-sequenced *P. falciparum* genomes were compared to the curated MalariaGEN Pf7 dataset, composed of Illumina-sequenced *P. falciparum* isolates. Although care was taken to select a relevant dataset, the use of different sequencing platforms and convenience sampling could affect geographical classification and IBD analysis. Overcoming these limitations, increasing genome coverage, as well as spatiotemporal and geographical resolution of training datasets, will enhance ONT-based surveillance efforts in the future.

Overall, this study demonstrates the utility of malaria genomics, specifically through long-read sequencing technology, in distinguishing imported *P. falciparum* cases from archived WBS. It also highlights the value of population genomic analyses in complementing traditional malaria epidemiological investigations. Although retrospective and small-scale in nature, the ability to differentiate imported cases reinforces Malaysia's malaria elimination trajectory and underscores the epidemiological relevance of such distinctions. These findings further emphasise the urgent need to integrate genomic data into elimination strategies, particularly to support surveillance among high-risk mobile populations, where patient travel histories are often incomplete or

unreliable. With continued advancements in sequencing throughput, genome coverage, and reductions in cost, genomic pipelines are becoming increasingly feasible for implementation, even in resource-limited settings. Importantly, the successful use of archived WBS illustrates the robustness of malaria genomics for retrospective analyses. The end-to-end workflow from genetic sequencing to population genomic analysis has demonstrated the operational practicality of third-generation sequencing technologies combined with open-source bioinformatics tools. As malaria elimination efforts progress, preventing the reintroduction and onward transmission of imported parasites will require that genomic surveillance becomes an integral component of both national and global malaria control strategies.

## **DECLARATIONS**

### *Ethics approval and consent*

Ethical approval for the study and recruitment process was granted by the Medical Research and Ethics Committee, Ministry of Health Malaysia, and the Ethics Committee of the Faculty of Medicine and Health Sciences, Universiti Malaysia Sarawak. Approval for the use of whole blood samples was also obtained from the University of St Andrews Teaching and Research Ethics Committee.

### *Consent for publication*

Not applicable.

### *Availability of data and materials*

Raw sequencing data is available from the ENA archive (PRJEB90161; see **Table S2** for a list of accession numbers). Details of the curated dataset are contained in **Table S6**.

### *Competing interests*

JE is employed part-time by Oxford Nanopore Technologies as VP Medical Affairs. All other authors have declared no conflicts of interest.

#### *Funding*

TGC and SC are funded by the UK Research and Innovation with the following grants: Biotechnology and Biological Sciences Research Council (BB/X018156/1), Medical Research Council (MR/R020973/1, MR/X005895/1), and Engineering and Physical Sciences Research Council (EP/Y018842/1). The funders did not have a role in the study design, data generation, collection, and analysis, decision to publish or preparation of the manuscript.

#### *Author Contributions*

JC-S, SC, and TGC conceived and directed the project. JC-S led the original study collection. PCSD, CD, and JE provided study data, context and resources. MKIT undertook sample processing and DNA extraction. MKIT performed sequencing. MKIT and NB performed bioinformatic and statistical analyses under the supervision of JC-S, SC, and TGC. All authors interpreted the results. MKIT wrote the first draft of the manuscript. All authors commented on and edited various versions of the draft manuscripts and approved the final manuscript. MKIT, NB, JC-S, SC, and TGC compiled the final manuscript.

#### *Acknowledgements*

We thank the patients and staff of Sibuh and Sarikei hospitals who participated in and supported the original study. This publication uses data from the MalariaGEN Plasmodium falciparum Community Project as described in 'An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide samples. MalariaGEN et al, Wellcome Open Research 2021642 DOI: 10.12688/wellcomeopenres.16168.1'.

## **REFERENCES**

1. World Health Organization. *World Malaria Report 2025*. World Health Organization; 2025. Accessed December 18, 2025.  
<https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2025>
2. Lek D, Shrestha M, Lhazeen K, et al. Malaria elimination challenges in countries approaching the last mile: a discussion among regional stakeholders. *Malar J*. 2024;23(1):401. doi:10.1186/s12936-024-05215-3
3. Alizadeh I, Sedaghat MM. Current challenges for controlling and eliminating malaria in Iran. *New Microbes and New Infections*. 2024;62:101464. doi:10.1016/j.nmni.2024.101464
4. Mohamed NA, Mansur F a. F, Abdul Rahman N. Imported *Falciparum* Malaria: A case series in a tertiary hospital. *Malays J Pathol*. 2020;42(1):107-110.
5. Okoth SA, Chenet SM, Arrospide N, et al. Molecular Investigation into a Malaria Outbreak in Cusco, Peru: *Plasmodium falciparum* BV1 Lineage is Linked to a Second Outbreak in Recent Times. *The American Society of Tropical Medicine and Hygiene*. 2016;94(1):128-131. doi:10.4269/ajtmh.15-0442
6. Phelan JE, Turkiewicz A, Manko E, et al. Rapid profiling of *Plasmodium* parasites from genome sequences to assist malaria control. *Genome Med*. 2023;15(1):96. doi:10.1186/s13073-023-01247-7
7. Vanheer LN, Mahamar A, Manko E, et al. Genome-wide genetic variation and molecular surveillance of drug resistance in *Plasmodium falciparum*

- isolates from asymptomatic individuals in Ouélessébougou, Mali. *Sci Rep.* 2023;13(1):9522. doi:10.1038/s41598-023-36002-w
8. Ibrahim A, Diez Benavente E, Nolder D, et al. Selective whole genome amplification of *Plasmodium malariae* DNA from clinical samples reveals insights into population structure. *Sci Rep.* 2020;10(1):10832. doi:10.1038/s41598-020-67568-4
  9. Brokhattingen N, Matambisso G, Da Silva C, et al. Genomic malaria surveillance of antenatal care users detects reduced transmission following elimination interventions in Mozambique. *Nat Commun.* 2024;15(1):2402. doi:10.1038/s41467-024-46535-x
  10. Chin AZ, Maluda MCM, Jelip J, Jeffree MSB, Culleton R, Ahmed K. Malaria elimination in Malaysia and the rising threat of *Plasmodium knowlesi*. *J Physiol Anthropol.* 2020;39(1):36. doi:10.1186/s40101-020-00247-5
  11. Hussin N, Lim YAL, Goh PP, William T, Jelip J, Mudin RN. Updates on malaria incidence and profile in Malaysia from 2013 to 2017. *Malar J.* 2020;19(1):55. doi:10.1186/s12936-020-3135-x
  12. Dian ND, Mohd Salleh AF, Rahim MAFA, et al. Malaria Cases in a Tertiary Hospital in Kuala Lumpur, Malaysia: A 16-Year (2005–2020) Retrospective Review. *TropicalMed.* 2021;6(4):177. doi:10.3390/tropicalmed6040177
  13. Jacob CG, Thuy-Nhien N, Mayxay M, et al. Genetic surveillance in the Greater Mekong subregion and South Asia to support malaria control and elimination. *eLife.* 2021;10:e62997. doi:10.7554/eLife.62997

14. Dondorp AM, Nosten F, Yi P, et al. Artemisinin Resistance in *Plasmodium falciparum* Malaria. *N Engl J Med*. 2009;361(5):455-467.  
doi:10.1056/NEJMoa0808859
15. Fornace KM, Abidin TR, Alexander N, et al. Association between Landscape Factors and Spatial Patterns of *Plasmodium knowlesi* Infections in Sabah, Malaysia. *Emerg Infect Dis*. 2016;22(2):201-209.  
doi:10.3201/eid2202.150656
16. Abdul Rahim FA, Md Yatim MF, Abdul Mutalip MH, Mahmud MAF. Assessing the population's correct knowledge of malaria in Malaysia: a vital component for malaria elimination certification. *Malar J*. 2023;22(1):267.  
doi:10.1186/s12936-023-04704-1
17. Kader Maideen SF, Rashid A, Ahmad NI, Zahari SNA, Hamat RA. Sero-prevalence of malaria and the knowledge, attitudes and practices relating to the prevention of malaria among indigenous people living in the central forest spine in Peninsular Malaysia: a mixed-methods study. *Malar J*. 2022;21(1):281. doi:10.1186/s12936-022-04293-5
18. Checkley AM, Smith A, Smith V, et al. Risk factors for mortality from imported *falciparum* malaria in the United Kingdom over 20 years: an observational study. *BMJ*. 2012;344(mar27 3):e2116-e2116.  
doi:10.1136/bmj.e2116
19. Moss S, Mańko E, Vasileva H, et al. Population dynamics and drug resistance mutations in *Plasmodium falciparum* on the Bijagós Archipelago, Guinea-Bissau. *Sci Rep*. 2023;13(1). doi:10.1038/s41598-023-33176-1

20. Deelder W, Manko E, Phelan JE, Campino S, Palla L, Clark TG. Geographical classification of malaria parasites through applying machine learning to whole genome sequence data. *Sci Rep.* 2022;12(1):21150. doi:10.1038/s41598-022-25568-6
21. Ahmed AM, Pinheiro MM, Divis PC, et al. Disease Progression in *Plasmodium knowlesi* Malaria Is Linked to Variation in Invasion Gene Family Members. Hirayama K, ed. *PLoS Negl Trop Dis.* 2014;8(8):e3086. doi:10.1371/journal.pntd.0003086
22. Sundararaman SA, Plenderleith LJ, Liu W, et al. Genomes of cryptic chimpanzee Plasmodium species reveal key evolutionary events leading to human malaria. *Nat Commun.* 2016;7(1):11078. doi:10.1038/ncomms11078
23. Aurrecochea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, et al. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Research* [Internet]. 2009 Jan 1 [cited 2025 Dec 9];37(Database):D539-43. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkn814>
24. Li H. Minimap2: pairwise alignment for nucleotide sequences. Birol I, editor. *Bioinformatics* [Internet]. 2018 Sept 15 [cited 2024 Mar 26];34(18):3094-100. Available from: <https://academic.oup.com/bioinformatics/article/34/18/3094/4994778>
25. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Berger B, ed. *Bioinformatics.* 2018;34(15):2666-2669. doi:10.1093/bioinformatics/bty149

26. Otto TD, Böhme U, Sanders MJ, Reid AJ, Bruske EI, Duffy CW, et al. Long read assemblies of geographically dispersed *Plasmodium falciparum* isolates reveal highly structured subtelomeres. *Wellcome Open Res* [Internet]. 2018 May 3 [cited 2025 Dec 19];3:52. Available from: <https://wellcomeopenresearch.org/articles/3-52/v1>
27. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Hancock J, ed. *Bioinformatics*. 2018;34(5):867-868. doi:10.1093/bioinformatics/btx699
28. Zheng Z, Li S, Su J, Leung AWS, Lam TW, Luo R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci*. 2022;2(12):797-803. doi:10.1038/s43588-022-00387-x
29. MalariaGEN, Abdel Hamid MM, Abdelraheem MH, et al. Pf7: an open dataset of *Plasmodium falciparum* genome variation in 20,000 worldwide samples. *Wellcome Open Res*. 2023;8:22. doi:10.12688/wellcomeopenres.18681.1
30. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* [Internet]. 2010 Sept [cited 2025 Dec 18];20(9):1297-303. Available from: <http://genome.cshlp.org/lookup/doi/10.1101/gr.107524.110>
31. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993. doi:10.1093/bioinformatics/btr509

32. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575. doi:10.1086/519795
33. Lee S, Harrison A, Tessier N, et al. Assessing clonality in malaria parasites using massively parallel sequencing data. Published online 2016.
34. Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. hmmIBD: software to infer pairwise identity by descent between haploid genotypes. *Malar J.* 2018;17(1):196. doi:10.1186/s12936-018-2349-7
35. Newby G, Bennett A, Larson E, et al. The path to eradication: a progress report on the malaria-eliminating countries. *The Lancet.* 2016;387(10029):1775-1784. doi:10.1016/S0140-6736(16)00230-0
36. Jeffree SM, Ahmed K, Safian N, et al. *Falciparum* Malaria Outbreak in Sabah Linked to an Immigrant Rubber Tapper. *The American Journal of Tropical Medicine and Hygiene.* 2018;98(1):45-50. doi:10.4269/ajtmh.17-0081
37. Kounnavong S, Gopinath D, Hongvanthong B, Khamkong C, Sichanthongthip O. Malaria elimination in Lao PDR: the challenges associated with population mobility. *Infect Dis Poverty.* 2017;6(1):81. doi:10.1186/s40249-017-0283-5
38. Ahmed S, Reithinger R, Kaptoge SK, Ngondi JM. Travel Is a Key Risk Factor for Malaria Transmission in Pre-Elimination Settings in Sub-Saharan Africa: A Review of the Literature and Meta-Analysis. *The American Journal of Tropical Medicine and Hygiene.* 2020;103(4):1380-1387. doi:10.4269/ajtmh.18-0456

39. Ooi CH, Phang WK, Liew JWK, Atroosh WM, Lau YL. Profiling the imported human malaria in Sarawak, Malaysia in 2011–2019. *Tropical Med Int Health*. 2023;28(6):486-500. doi:10.1111/tmi.13875
40. Schaffner SF, Badiane A, Khorgade A, et al. Malaria surveillance reveals parasite relatedness, signatures of selection, and correlates of transmission across Senegal. *Nat Commun*. 2023;14(1):7268. doi:10.1038/s41467-023-43087-4
41. Mita T, Tanabe K, Kita K. Spread and evolution of *Plasmodium falciparum* drug resistance. *Parasitology International*. 2009;58(3):201-209. doi:10.1016/j.parint.2009.04.004
42. Amato R, Pearson RD, Almagro-Garcia J, et al. Origins of the current outbreak of multidrug-resistant malaria in southeast Asia: a retrospective genetic study. *Lancet Infect Dis*. 2018;18(3):337-345. doi:10.1016/S1473-3099(18)30068-9
43. Nair S, Li X, Arya GA, et al. Fitness Costs and the Rapid Spread of *kelch13*-C580Y Substitutions Conferring Artemisinin Resistance. *Antimicrob Agents Chemother*. 2018;62(9):e00605-18. doi:10.1128/AAC.00605-18

#### ADDITIONAL INFORMATION

Additional file 1: Supplementary\_Information

File format: DOCX

**Figure S1.** DNA concentration (A) and yield per sample (B) obtained from DNeasy® Blood and Tissue Kit (Qiagen, UK) and Monarch® Genomic DNA Purification Kit (NEB, UK). 12 frozen WBS were extracted with both kits and compared against each other with the Monarch® Genomic DNA Purification Kit

(NEB, UK) showing better extraction yields both in DNA concentration and total DNA yield. The remaining nine whole blood samples (WBS) from this study were extracted with this kit.

**Figure S2.** Estimated multiplicity of infection (MOI) using the within-host diversity metric  $F_{WS}^*$ . × denotes Malaysian isolates with their designated source region; \* the score of 1 dictates monoclonality and 0, polyclonal infection

**Table S1.** DNA concentration and yield obtained from both extraction kits, DNeasy® Blood and Tissue Kit (Qiagen, UK) and Monarch® Genomic DNA Purification Kit (New England Biolabs, UK)

**Table S2.** ENA Accession numbers for the project PRJEB90161

**Table S3.** Number of nuclear chromosome variants remaining following each filtration step

**Table S4.** Sequencing yields, assembly metrics and genomic coverage

**Table S5.** Regional frequencies of drug resistance genotypes and Malaysian isolates with respective genotypes

Additional file 2: Table S6

File format: XLS

**Table S6.** List of accession numbers and details of curated dataset used

**Table 1.** Patient characteristics linked to their respective *Plasmodium* isolates (n = 21)\*

Sample ID	Age (years)	Community Group	Hospital	Collection year	Parasitaemia (%)	Travel history
Pf_01	32	Chinese	Sibu	2008	1.66	Not reported
Pf_02	38	Dayak indigenous	Sibu	2008	0.97	PNG (Oceania)
Pf_03	33	Filipino	Sibu	2008	0.62	Not reported
Pf_04	44	Chinese	Sibu	2008	8.80	WNG (Oceania)
Pf_05	49	Dayak indigenous	Sibu	2008	1.60	Africa
Pf_06	37	Dayak indigenous	Sibu	2008	2.56	No travel
Pf_07	34	Dayak indigenous	Sibu	2008	0.12	Africa
Pf_08	24	Dayak indigenous	Sarikei	2008	2.40	No travel
Pf_09	19	Dayak indigenous	Sarikei	2009	0.14	No travel
Pf_10	43	Dayak indigenous	Sibu	2008	0.01	Ghana (Africa)
Pf_11	29	Chinese	Sibu	2008	1.05	No travel
Pf_12	43	Dayak indigenous	Sibu	2008	3.58	PNG (Oceania)
Pf_13	22	Malay	Sarikei	2009	4.19	No travel

Pf_14	28	Malay	Sibu	2009	0.08	DRC (Africa)
Pf_15	19	Dayak indigenous	Sibu	2009	0.02	No travel
Pf_16	25	Chinese	Sibu	2009	0.38	DRC (Africa)
Pf_17	47	Chinese	Sibu	2009	1.30	PNG (Oceania)
Pf_18	40	Dayak indigenous	Sibu	2009	2.72	PNG (Oceania)
Pf_19	50	Dayak indigenous	Sibu	2010	0.01	Africa
Pf_20	52	Dayak indigenous	Sarikei	2010	0.11	South Africa (Africa)
Pf_21	31	Dayak indigenous	Sibu	2010	2.43	Africa

\* all confirmed with *P. falciparum* infections using PCR and microscopy; PNG:

Papua New Guinea; DRC: Democratic Republic of the Congo; WNG: Western  
New Guinea

**Table 2.** Genotypic drug resistance profiles and geographical classification produced by *Malaria-Profiler* and multiplicity of infection of the Malaysian isolates. Isolates were considered monoclonal if their  $F_{ws}$  score was  $>0.95$ .

Sample ID	Drug resistance	Gene (mutations)	Geographical classification (Highest predictive probabilities)	$F_{ws}$ score*
Pf_02	Chloroquine	<i>pfcr</i> (A220S, C72S, K76T) <i>pfmdr1</i> (N1042D, Y184F)	SEA (0.50) and Oceania (0.28)	0.93
	Pyrimethamine	<i>pfdhfr-ts</i> (C59R, S108N)		
Pf_03	Chloroquine	<i>pfcr</i> (A220S, C72S, K76T)	Oceania (1.00) and SEA (0.001)	0.92
Pf_05	Chloroquine	<i>pfcr</i> (A220S)	SEA (0.61) and South Asia (0.36)	0.96
Pf_06	Chloroquine	<i>pfcr</i> (A220S, R371I, Q271E, I356T, K76T, M74I)	Central Africa (0.91) and West Africa (0.09)	0.92
		<i>pfmdr1</i> (N86Y)		
	Pyrimethamine	<i>pfdhfr-ts</i> (S108N)		
	Sulfadoxine	<i>pfpppk-dhps</i> (A437G, I431V, S436A)		
Pf_07	Chloroquine	<i>pfcr</i> (R371I, Q271E, I356T, K76T, M74I)	Central Africa (0.91) and West Africa (0.09)	0.97

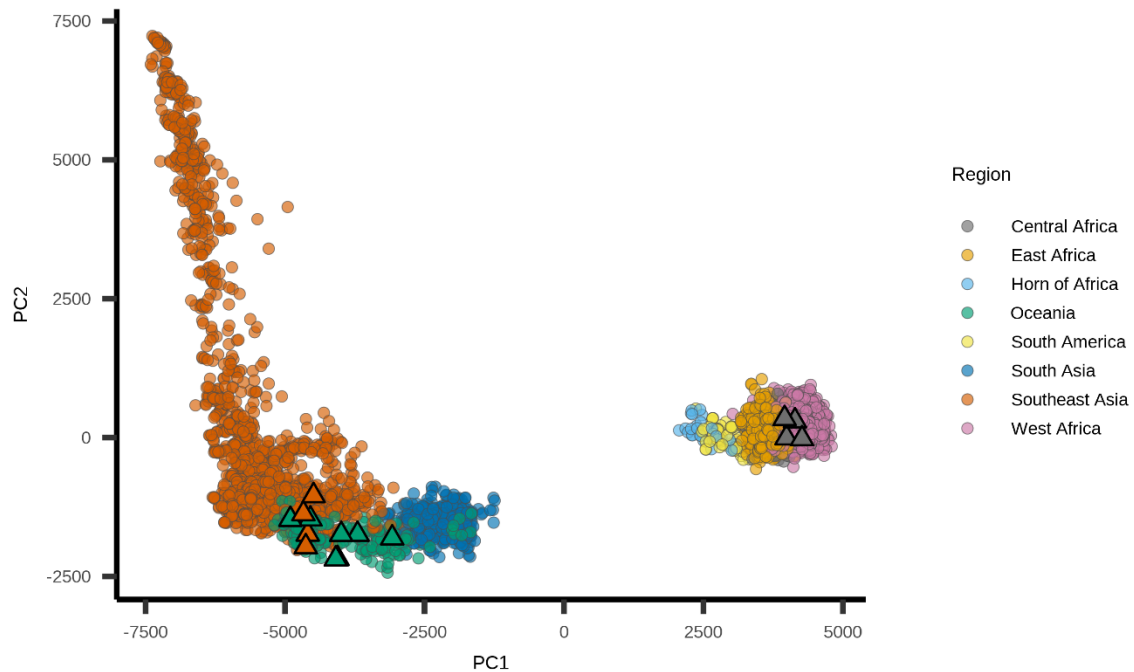
		<i>pfmdr1</i> (N86Y, Y184F)		
	Pyrimethamine	<i>pfdhfr-ts</i> (N51I, C59R, S108N)		
	Sulfadoxine	<i>pfpppk-dhps</i> (A437G, I431V, S436A)		
Pf_08	Chloroquine	<i>pfcr1</i> (A220S, C72S, K76T)	SEA (0.55) and South Asia (0.36)	0.96
Pf_09	Sensitive	N/A	Oceania (0.93) and SEA (0.06)	0.90
Pf_11	Chloroquine	<i>pfcr1</i> (C72S, K76T)	Oceania (0.98) and SEA (0.01)	0.94
		<i>pfmdr1</i> (N86Y)		
	Pyrimethamine	<i>pfdhfr-ts</i> (I164L)		
	Sulfadoxine	<i>pfpppk-dhps</i> (A437Gn, A581G)		
Pf_12	Chloroquine	<i>pfcr1</i> (A220S, C72S, K76T)	Oceania (1.0) and Horn of Africa (0.0006)	0.96
		Pyrimethamine		
	Sulfadoxine	<i>pfpppk-dhps</i> (A437G)		
Pf_13	Chloroquine	<i>pfcr1</i> (A220S)	SEA (0.46) and South Asia (0.39)	0.96
	Pyrimethamine	<i>pfdhfr-ts</i> (S108N)		
Pf_15	Sensitive	N/A	SEA (0.80) and South Asia (0.09)	0.97
Pf_17	Chloroquine	<i>pfcr1</i> (A220S, C72S, K76T)	Oceania (0.98) and SEA (0.02)	0.79
		<i>pfmdr1</i> (N86Y)		

	Pyrimethamine	<i>pfdhfr-ts</i> (C59R, S108N)		
	Sulfadoxine	<i>pfpppk-dhps</i> (A437G)		
Pf_18	Sulfadoxine	<i>pfpppk-dhps</i> (A437G)	Oceania (0.99) and SEA (0.004)	0.81
Pf_20	Sensitive	N/A	Central Africa (0.64) and West Africa (0.36)	0.70
Pf_21	Sulfadoxine	<i>pfpppk-dhps</i> (A437G)	Central Africa (0.92) and West Africa (0.06)	0.92

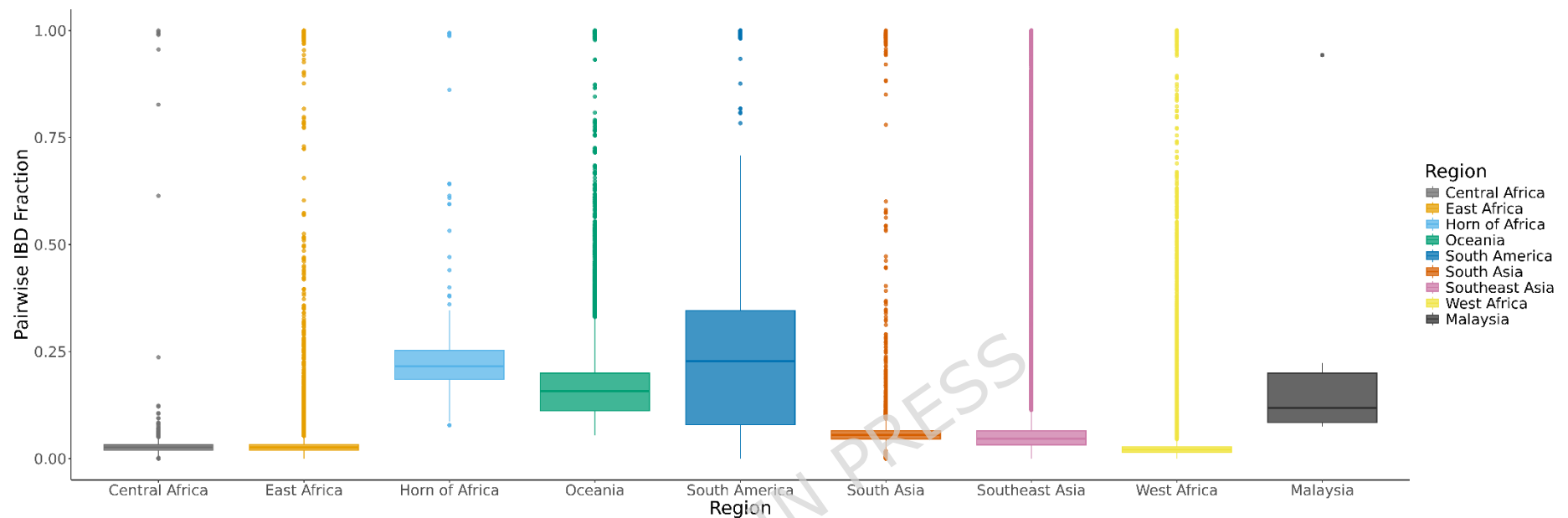
\* the score of 1 dictates monoclonality and 0, polyclonal infection; SEA:

Southeast Asia; N/A not applicable.

ARTICLE IN PRESS



**Figure 1.** Principal component analysis (PCA) combined with multidimensional scaling (MDS) demonstrated that Malaysian isolates cluster with a curated *P. falciparum* reference database. The global dataset comprised 9,518 *P. falciparum* isolates from 32 countries across eight geographical regions. The PCA plot is shown by geographical region, with Malaysian isolates denoted by a  $\Delta$  and coloured according to their predicted regions. These isolates clustered within the regions consistent with patients' reported travel histories.



**Figure 2.** Pairwise identity-by-descent (IBD) sharing between Malaysia and global regions.

Distribution of pairwise IBD fractions calculated between individuals from Malaysia ( $n = 6$ ) and individuals from eight global regions ( $n = 7,134$ ). The x-axis denotes the comparison region, and the y-axis shows the pairwise IBD fraction.

Colours indicate geographical regions: Central Africa ( $n = 212$ ), East Africa ( $n = 619$ ), Horn of Africa ( $n = 16$ ), Oceania ( $n = 271$ ), South Asia ( $n = 310$ ), Southeast Asia ( $n = 4,030$ ), West Africa ( $n = 1,615$ ), and South America ( $n = 61$ ). All samples were filtered to include only monoclonal infections ( $F_{WS} > 0.95$ ). Higher IBD fractions indicate more recent shared ancestry between Malaysian samples and individuals from the corresponding region.

