



OPEN

## Rapid detection of G6PD deficiency SNPs using a novel amplicon-based minion sequencing assay

Costanza Tacoli<sup>1,2,12</sup>, Mariana Kleinecke<sup>3,12</sup>, Kian Soon Hoon<sup>3</sup>, Hidayat Trimarsanto<sup>3,4</sup>, Muhammad Nadeem<sup>1</sup>, Rotha Eam<sup>1</sup>, Nimol Khim<sup>1</sup>, Angela Rumaseb<sup>3</sup>, Benedikt Ley<sup>3</sup>, Dean Sayre<sup>5</sup>, Sochea Phok<sup>6</sup>, Dysoley Lek<sup>7</sup>, Benoit Witkowski<sup>1</sup>, Jimee Hwang<sup>5</sup>, Sarah Auburn<sup>3,8,9</sup>, Ivo Mueller<sup>2,10</sup>, Ric N. Price<sup>3,8,9</sup> & Jean Popovici<sup>1,11</sup>✉

*Plasmodium vivax* malaria remains a significant global health challenge, complicated by the parasite's ability to form dormant liver stages (hypnozoites) that cause relapses. Radical cure of *P. vivax* malaria requires administration of a hypnozoitocidal drug, such as primaquine or tafenoquine. However, these drugs can cause severe haemolysis in individuals with glucose-6-phosphate dehydrogenase (G6PD) deficiency. G6PD deficiency is caused by more than 230 different variants at the gene level that confer different degrees of deficiency phenotypically. Understanding the distribution of different G6PD variants in affected populations is essential to inform safer antimalarial treatment strategies. This study aimed to develop a cost-effective sequencing assay targeting key regions of the G6PD gene, suitable for field deployment. A novel assay based on Nanopore technology was designed to amplify two amplicons covering exon 3 to exon 13, focusing on known variants associated with enzyme deficiency. A total of 79 samples from individuals in Cambodia, Vietnam, Afghanistan, and China were sequenced, and a bioinformatics pipeline was created for the targeted variant calling of 192 G6PD SNP mutations. The assay demonstrated reliable detection of known variants, with high concordance between runs, within runs, and with Sanger sequencing. The Nanopore MinION long-amplicon sequencing assay offers a robust and portable solution for large-scale G6PD genotyping in low-resource settings, that will improve malaria control and elimination strategies by enabling safer antimalarial treatment.

*Plasmodium vivax* remains an important public health threat in 41 endemic countries, causing 4–7 million cases of malaria each year<sup>1</sup>. Over the past two decades, intense malaria control efforts have reduced the global burden of disease considerably; however, in many areas where *P. vivax* and *P. falciparum* are co-endemic, *P. vivax* has become the predominant cause of malaria<sup>1–3</sup>. The primary reason for this shift in burden is the ability of *P. vivax* to cause recurrent episodes of malaria known as relapses. *P. vivax* and *P. ovale* are the only malaria parasites in humans capable of forming dormant liver stages (hypnozoites), which can reactivate weeks to months after an initial infection, causing recurrent parasitaemia. Up to 85% of recurrent *P. vivax* infections are attributable to relapses, and these contribute to significant morbidity and ongoing transmission of the parasite, which undermines elimination efforts<sup>4,5</sup>. Antimalarial treatment of *P. vivax* and *P. ovale* requires administration of a combination of schizonticidal and hypnozoitocidal drugs to kill both the asexual parasites, that cause acute disease, and the liver stages that result in relapses; this combination is referred to as 'radical cure'. The only licensed drugs with hypnozoitocidal activity are primaquine and tafenoquine. Whilst these 8-aminoquinoline compounds are well tolerated in most patients, they can cause severe haemolysis in individuals with glucose-

<sup>1</sup>Malaria Research Unit Institut Pasteur du Cambodge, Phnom Penh, Cambodia. <sup>2</sup>Walter and Eliza Hall Institute, Melbourne, Australia. <sup>3</sup>Global and Tropical Health Division, Menzies School of Health Research, Charles Darwin University, Darwin, NT, Australia. <sup>4</sup>Eijkman Research Center for Molecular Biology, National Research and Innovation Agency, Jakarta, Indonesia. <sup>5</sup>U.S. President's Malaria Initiative, Malaria Branch, Centers for Disease Control and Prevention, Atlanta, USA. <sup>6</sup>PMI Impact Malaria, Population Services International, Phnom Penh, Cambodia. <sup>7</sup>National Center for Parasitology, Entomology and Malaria Control, Phnom Penh, Cambodia. <sup>8</sup>Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok, Thailand. <sup>9</sup>Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK. <sup>10</sup>University of Melbourne, Melbourne, Australia. <sup>11</sup>Infectious Disease Epidemiology and Analytics G5 Unit, Institut Pasteur, Paris, France. <sup>12</sup>Costanza Tacoli and Mariana Kleinecke contributed equally to this work. ✉email: jpopovici@pasteur-kh.org

6-phosphate dehydrogenase (G6PD) deficiency (G6PDd)<sup>6</sup>. G6PDd is among the most common inherited enzymopathies worldwide, present in up to 35% of people living in malaria-endemic countries<sup>7</sup>.

The *G6PD* gene has a size of ~ 18.5 kilobases (kb) and is located on the X chromosome (Xq28). It consists of 13 exons (the first exon is non-coding) and 12 introns, encoding a product of 1,545 base pairs (bp). Over 230 variants have been associated with reduced enzyme activity<sup>8</sup>. Most of these variants are single nucleotide polymorphisms (SNPs) in exon 3–13<sup>6,8,9</sup>. Whilst males are either hemizygous-deficient or normal (wild-type), females can exhibit a more varied level of enzyme activity owing to the presence of two X chromosomes, one of which is randomly inactivated at the cellular level during early embryonic development through a process called lyonisation<sup>6</sup>. The degree of drug-induced haemolysis is related to the dose of drug administered and the underlying G6PD enzyme activity; the latter varies with the *G6PD* variant and the degree of lyonisation. In view of the risk of drug-induced haemolysis, the World Health Organization (WHO) recommends routine testing for G6PDd prior to treatment with either primaquine or tafenoquine. However, in practice, routine G6PD testing, despite the recent WHO pre-qualification of the SD Biosensor G6PD test, is rarely available in poor-resourced communities, where the greatest burden of *P. vivax* resides<sup>10</sup>. Geospatial mapping of different *G6PD* variants and their prevalence in *P. vivax* endemic countries can inform policy makers of the inherent risks of haemolysis if G6PD testing is not available, however such efforts have been confounded by the diversity of genotyping methods used<sup>11</sup>.

Molecular methods such as polymerase chain reaction (PCR) followed by Sanger sequencing have been used to identify *G6PD* polymorphisms associated with enzyme deficiency. However, the application of these methods is usually tailored to identify variants known to be locally prevalent. While Sanger sequencing is accurate and useful for sequencing small genomic regions, it is less suitable for longer fragments due to its relatively high cost, slow processing time, and short read length compared to modern next-generation sequencing (NGS) technologies<sup>12</sup>. NGS platforms like Illumina are high-throughput and cost-effective but typically limited to 600 bp fragment lengths. In contrast, Nanopore sequencing stands out as one of the most advanced third-generation sequencing technologies, capable of generating sequence reads of varying lengths, from 20 kb to over 4 Mb<sup>13</sup>, enabling detection of variants across the entire gene with fewer amplicons. The MinION sequencer from Oxford Nanopore Technologies is compact, portable, and relatively inexpensive compared to other sequencing platforms.

The aim of this study was to develop a novel MinION long amplicon-based sequencing assay spanning critical regions of the *G6PD* gene designed for deployment in resource-limited settings. Specifically, we sequenced 10.8 kb of the *G6PD* gene in two amplicons, covering from exons 3 to 13 and their flanking intronic regions. We validated the assay by confirming the presence of known *G6PD* mutations in patients from Cambodia, Afghanistan, China and Vietnam, using Sanger sequencing. To streamline data analysis, we developed an open-access bioinformatics pipeline to identify the 192 *G6PD* SNP variants associated with reduced enzyme activity, as listed by Luzzatto et al.<sup>6</sup>. Finally, we estimated the economic advantage of using the MinION platform for sequencing the *G6PD* gene in 96 samples as compared to Sanger sequencing by breaking down the costs associated with each method taking into account both the per-sample costs and the overall cost of sequencing.

## Materials and methods

### Sample collection and selection

Blood samples were collected from 79 individuals (43 females and 36 males) enrolled in cross-sectional surveys or antimalarial clinical trials conducted in Cambodia, Vietnam and Afghanistan, as well as a patient of Chinese ethnicity presenting at an Australian hospital (Supplementary Data 1). Individuals provided 250–400  $\mu$ l of capillary blood or 4 ml of venous blood for screening of G6PD deficiency using either a qualitative test (Fluorescent Spot Test, FST) or a quantitative test (spectrophotometry or SD Biosensor). Female participants were categorized phenotypically as having G6PD normal enzyme activity (> 70% of the adjusted male median activity, AMM), intermediate activity (30–70% of the AMM), or deficiency (< 30% of the AMM). Males were categorized phenotypically as either normal (> 30% of the AMM) or deficient (< 30% of the AMM)<sup>14</sup>. Individuals tested with the FST were classified as G6PD-deficient if no fluorescence was observed after 30 min. For those tested using the SD Biosensor or spectrophotometry, G6PD activity below 4 IU/g Hb was classified as deficient, while activity between 4 and 6 IU/g Hb was considered intermediate<sup>15</sup>. The study was conducted following the declaration of Helsinki and all participants or their legal guardian provided informed consent.

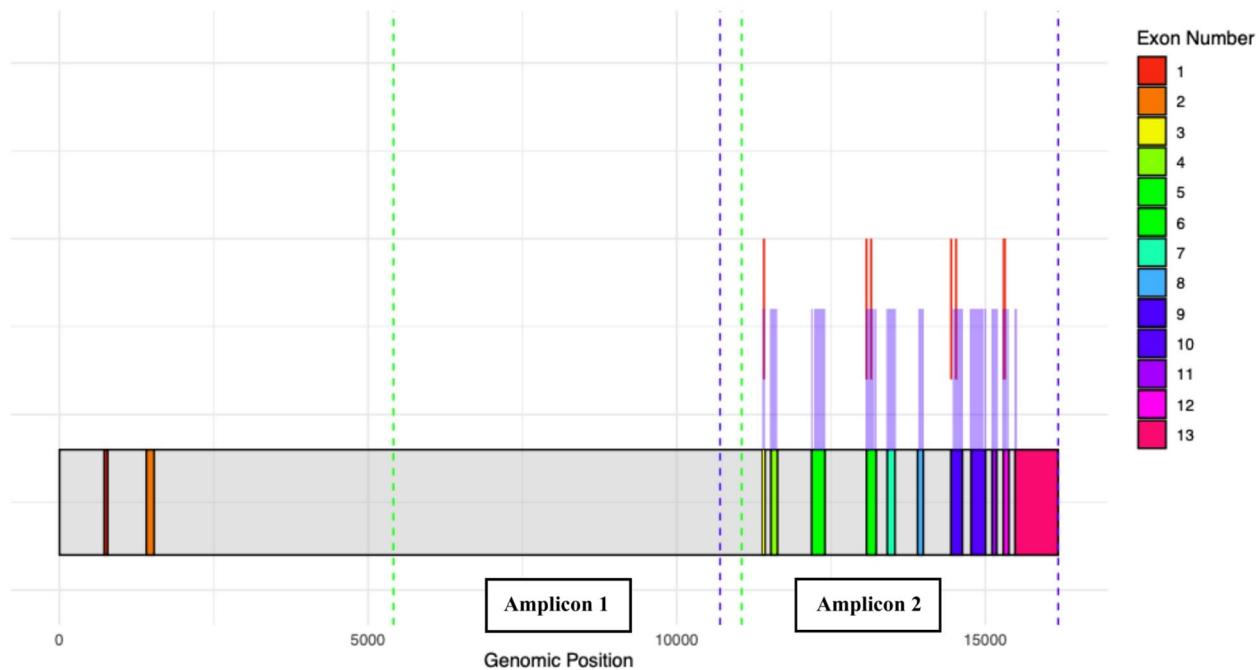
### DNA extraction and Sanger sequencing

For 11 individuals enrolled in Afghanistan, China and Vietnam, DNA was extracted from whole blood using column kits (Favorgen Biotech, Taiwan). For 68 participants from Cambodia, DNA was extracted from whole blood using QIAamp DNA mini kit (Qiagen, Germany).

*G6PD* genotypes were determined by Sanger sequencing exon 3–13 following seven distinct PCR reactions using a protocol adapted from Kim et al.<sup>16</sup>. Chromatograms were visualized and analyzed with the Qiagen CLC Genomics Workbench (Qiagen, Germany). SNP mutations were identified using Mega-X software v.11 by comparison with the NCBI genomic RefSeqGene sequence NG\_009015.2:g.

### Long-amplicon amplification, library preparation and minion sequencing

The protocol, including primer sequences, PCR cycling conditions, and library preparation steps, can be found in Supplementary Information. Briefly, PCR amplifications were performed on two fragments of approximately 5.6 kb and 5.7 kb each, encompassing exons 3–13 and their flanking intronic regions (Fig. 1). The reverse primer for the second fragment was positioned approximately 800 bp downstream of exon 13 to include its 3' untranslated region. Amplification was performed using High Fidelity PrimeSTAR GXL DNA Polymerase (Takara Bio, Japan). The first fragment was obtained through two repeated PCRs utilizing the same primers. PCR products were run on 1% agarose gel for 1–1.5 h at 100 V.



**Fig. 1.** The *G6PD* gene, highlighting all exons from 1 to 13 (Transcript variant 3 from RefSeq NM\_001360016.2:c, GenCode: ENST00000393562.10). The regions covered by Amplicon 1 and Amplicon 2 are indicated with green and dark purple dotted vertical lines, respectively. The position of variants that were identified in this study are marked with vertical red lines, while others included in the pipeline are shown in light purple.

A detailed list of the consumables and reagents used for the library preparation is provided by Oxford Nanopore at <https://nanoporetech.com/document/native-barcoding-amplicons> and summarised in Supplementary Information. Agencourt AMPure XP beads (Beckman Coulter, US) were used for high-throughput purification of PCR amplicons following the manufacturer's instructions, maintaining a 1:1 ratio of beads to each PCR product during all purification steps. The Qubit Fluorometer 2.0 and Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, US) were used to quantify PCR amplicons at each purification stage, tracking the DNA quantity to ensure it met the required levels for library preparation, including final pooling and loading onto the MinION flow cell.

Library preparation was carried out according to Oxford Nanopore Technologies instructions. Briefly, PCR amplicons of the two fragments from the same sample were quantified and pooled in equimolar amounts after purification, aiming for an ideal final concentration of 200–400 ng (120–240 fmol) per sample. DNA was then end-repaired, 5' phosphorylated, and prepared for ligation with the NEBNext Ultra II End Repair/dA-Tailing Module. To pool multiple samples onto one flow cell, the amplicons of each sample were labeled with unique barcodes using NEB Blunt/TA Ligase Master Mix. For the amplicon libraries of the first two runs performed (namely, run 1 and run 2) and the latter one (run 4), the ligation kit SQK-LSK109 with the amplicons-native barcoding expansion kit EXP-NBD104 was used, this was replaced by the expansion kit EXP-NBD196 for the 96-samples format of run 3. NEBNext Quick T4 DNA Ligase was used for the ligation of Nanopore adapters (Adapter Mix II) to the barcoded fragments, forming a complete library molecule. Before loading the prepared library, free oligonucleotides and small fragments were removed using AMPure XP beads. Lastly, priming of the flow cell, and loading of the DNA library was achieved using the flow cell priming kit (EXP-FLP002).

All runs were performed until exhaustion of the flow cell (up to 28 h). For the pool libraries of run 1 and run 2, total reads were estimated at multiple times between 16 and 24 h for comparison in data yield. Since the coverage was highly satisfactory, when working in a 96-well plate, the clean-up step following end-prep was omitted to speed up the library preparation process, and an extended run time was included as a compensatory measure. Omitting the clean-up step and running until flow cell exhaustion allowed the initial volume and concentrations of pooled amplicons to be reduced (Supplementary Information). This method successfully processed final libraries with a concentration as low as 120 ng (below the recommended 200 ng) without any significant negative effects on sequencing quality, accuracy, or data yield. For run 2, 'high accuracy basecalling' was performed during the run while for the other runs including for multiplexing 96 samples, it was switched to 'fast basecalling' to improve computing time.

### Data analysis pipeline

A pipeline for targeted variant calling is available at [https://github.com/vivaxgen/G6PD\\_MinION](https://github.com/vivaxgen/G6PD_MinION). A detailed breakdown of the pipeline steps can be found in Supplementary Information. After MinION sequencing, raw FAST5 reads were basecalled and demultiplexed using Guppy version 6.5.7 as a plugin in Genious Prime 2024.

During the installation process, the reference sequence was automatically indexed (samtools version 1.20)<sup>17</sup>. The pipeline analysed the base-called sequencing data in five steps: (i) trimming (chopper version 0.8.0)<sup>18</sup>, (ii) mapping (minimap2 version 2.28-r1209)<sup>19</sup>, (iii) variant calling (freebayes version 1.3.6 or clair3 version 1.0.9)<sup>20,21</sup>, and, (iv) a Python script that generates a variant report after applying quality-based filtering to remove low-confidence variant calls. Specifically, variants were filtered based on quality metrics (e.g., QQUAL score) produced during variant calling, and empirical thresholds applied within the pipeline to minimize background noise while balancing sensitivity and specificity ( $Q \geq 10$ , minimum depth of 20). These thresholds are not fixed globally, but rather are adapted based on sample characteristics, and are consistent with practices established in previous Nanopore-based variant calling workflows<sup>22</sup>. For the variant calling the pipeline was designed to take a list of known variants, to report the presence/absence of the variant as well as to characterize hemi-, homo-, and heterozygosity. The variants currently included can be found in Supplementary Data 2. The pipeline also has a variant discovery mode, which is still under development; this mode applies similar quality filtering criteria to candidate novel variants, but additional post hoc curation and validation steps are needed to mitigate the effects of noise and background.

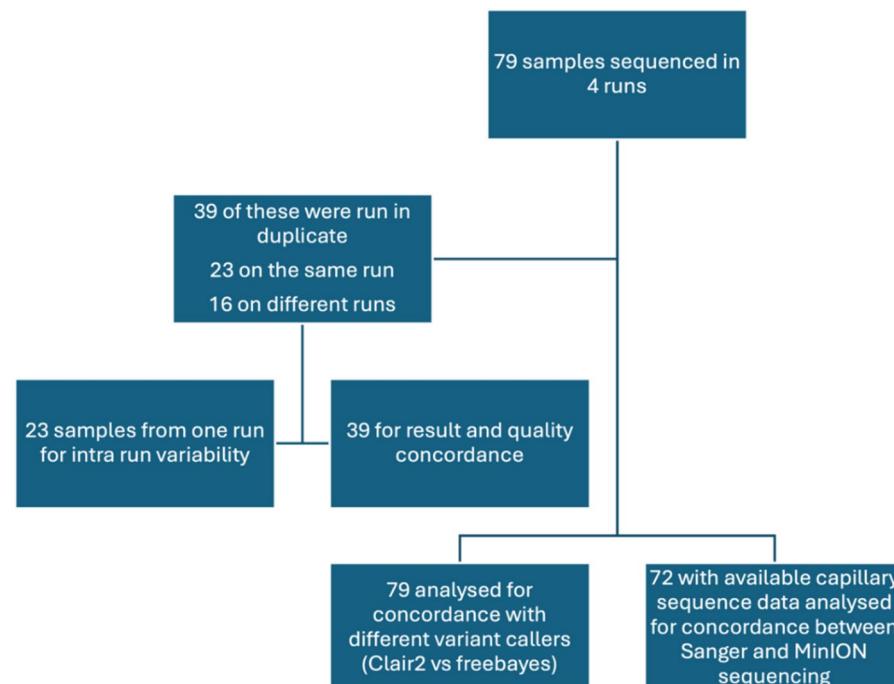
## Results

### Assay development and samples sequenced

A total of 79 individuals from four different countries were sequenced in four different runs (Fig. 2). In run 1, 3 samples were investigated, and ~ 7000 reads per sample were obtained after 16 h of sequencing, with an increase of ~ 6000 passed reads at the 24-hour time-point. Similar numbers were observed when working with an increased number of samples, with over 10,000 reads per amplicon reached at 20 h in run 2 (11 samples, data not shown). In run 3, to optimize MinION library preparation in a 96-well plate format and assess the impact of protocol modifications, samples from 23 individuals (2 females and 21 males) with known G6PD genotypes were sequenced in duplicates within the plate. Of these, one sample of Cambodian origin could not achieve concordant results within the run duplicate due to low read depth. Finally, run 4 (8 samples) was performed to provide a final check of the developed protocol. A total of 16 samples were run at multiple times across runs to assess reproducibility. All remaining samples were concordant between and within runs and showed agreement with Sanger sequencing results (Supplementary Data 1).

### Sequencing performance

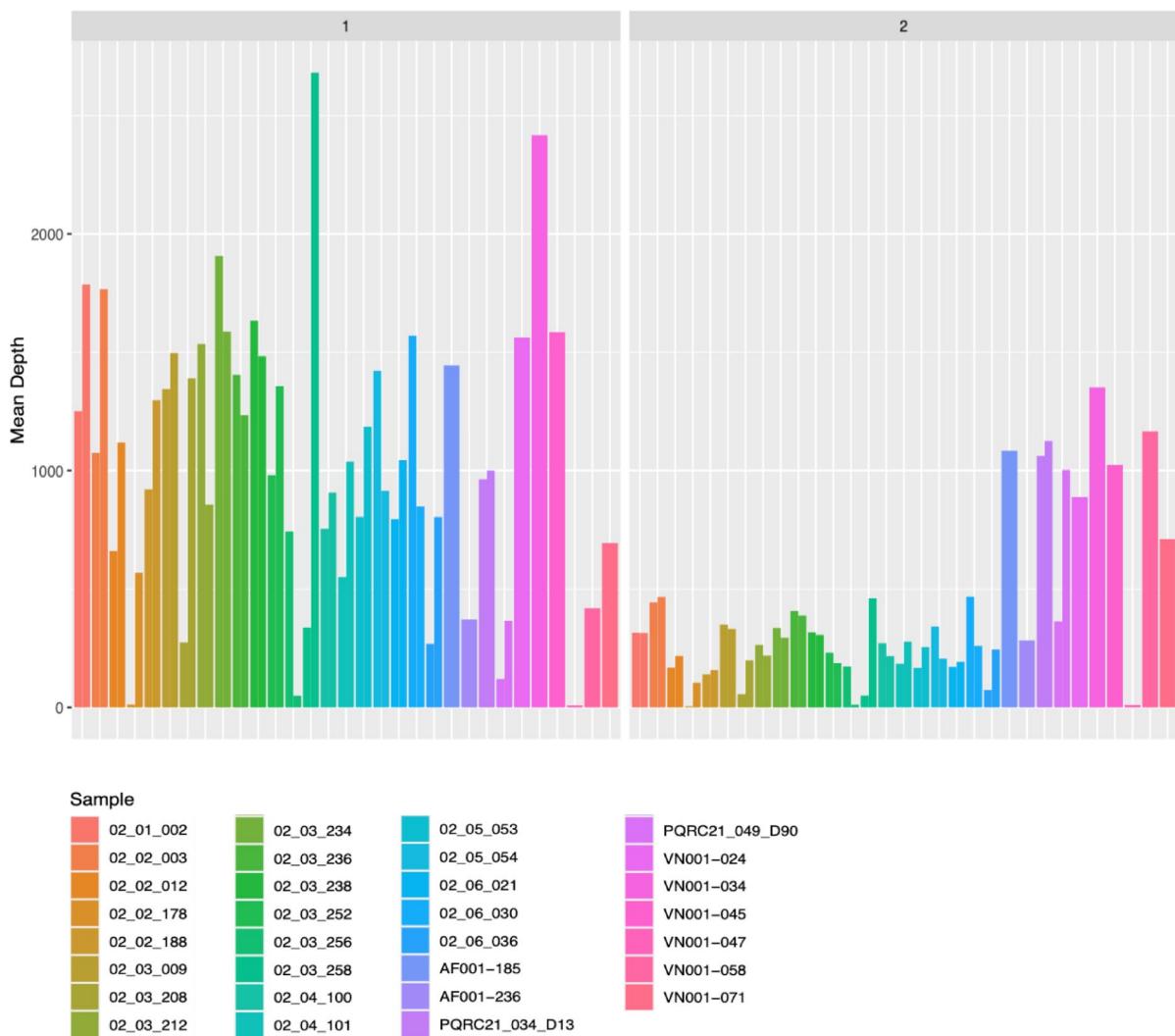
A summary of the data acquired at each run is given in Table 1. The mean number of reads per amplicon per sample varied from 686 to 94,641, resulting in a depth variation from 498 to 52,778 (Supplementary Data 3). These results were influenced by factors including running time, sample input concentration, available sequencing pores and sample multiplexing. For example, in run 2, an ultra-high depth of 52,000 was observed. Despite this variability, coverage and mapping quality were consistently high across the four runs, ranging from 97 to 100% for coverage and a score of 59–60 for mapping quality. In the sequencing run with 96 samples, 2.77 million reads and approximately 5.14 Gb of bases were generated. Following analysis, the average depth was of ~ 950 for Amplicon 1 and of ~ 500 for Amplicon 2 (Fig. 3). Four samples (VN001-047, 02\_02\_178, 02\_02\_092



**Fig. 2.** Overview of the samples selected for sequencing and their corresponding analyses.

Run	Amplicon	Nr. reads	Coverage	Mean depth	Mean base quality	Mean map quality
1 (n=3)	1	1,140.33	100.00	1,032.78	24.73	60.00
	2	2,518.00	100.00	2,114.96	21.70	59.97
2 (n=11)	1	94,641.82	100.00	52,778.65	24.28	59.99
	2	51,155.73	97.83	43,459.04	21.48	60.00
3 (n=96)	1	1,280.92	100.00	957.67	19.42	59.98
	2	686.47	99.99	498.90	18.81	59.97
4 (n=8)	1	6,021.5	100.00	5,300.46	19.13	59.98
	2	4,704.38	100.00	27,452.97	18.50	59.99

**Table 1.** Summary of sequencing metrics per run and amplicon, including the average number of reads, mean depth, mean base quality, mapping quality and coverage. Mean base quality reflects the confidence of each base call, while mean mapping quality reflects the confidence that reads are correctly aligned to the reference. Coverage represents the proportion of the genome sequenced, and depth indicates the average number of times each base is read. 'n' represents the number of samples processed at each run. data shown were collected at flow cell exhaustion. Detailed metrics for each sample and run are available in supplementary data 3.



**Fig. 3.** Mean sequencing depth for each sample. Each sample is represented by two bars of the same color, indicating duplicate sequencing runs associated with different barcodes. The left panel depicts the depth for Amplicon 1, while the right panel displays the depth for Amplicon 2.

and 02\_03\_256) showed low sequencing depth (< 50x) in one or both of the amplicons in one of the runs but could successfully be recovered in a subsequent run (Fig. 3, Supplementary Data 3).

### Assay validation and SNPs detection

The MinION pipeline was tested on samples from all 79 individuals, identifying seven known G6PD variants in 58 (73.4%). According to MinION results, of the 36 males included in the study, 58.3% (21/36) were hemizygous for G6PD variants, while the remaining 15 (41.7%) had no major known G6PD polymorphisms. Out of the 43 females sequenced, five (5/43, 11.6%) were homozygous for the G6PD Viangchan variant (Val291Met), and 28 (65.1%) were identified as heterozygous. Of these 28 heterozygous females, 26 carried the Viangchan variant, 1 had both the Viangchan and the Aureus (Ile48Thr) variants, and 1 had the Mahidol (Gly163Ser) variant.

Sanger sequencing was performed for 53 individuals carrying mutations according to MinION results (Table 2), and for 19 wild-type individuals to confirm absence of known polymorphisms. Seven samples could not be Sanger sequenced due to low DNA amounts or low quality. A final subset of 72 individuals had data from both Sanger and MinION sequencing: 61 from Cambodia, 3 from Afghanistan, 7 from Vietnam and, 1 from China (Supplementary Data 1). Five samples of Cambodian origins (2 males and 3 females) were not sequenced by Sanger due to low quantity of extracted DNA. Among the individuals carrying G6PD mutations, 43 (43/53, 81.1%) had exclusively Viangchan, 1 (1.9%) carried both Viangchan and Aureus variants, 3 (5.7%) had the Mediterranean variant (Ser188Phe), 3 (5.7%) had the Kaiping (Arg463His) and 1 each (1.9%) had the Canton (Arg459Leu), the Kerala-Kalyan (Glu317Lys), and the Mahidol variants. Complete concordance between Sanger and MinION was observed for all except one sample (Supplementary Data 1).

### Cost comparison of Sanger and MinION sequencing of the G6PD gene

To highlight the potential cost savings that could be achieved through the adoption of the MinION platform, we evaluated the cost of reagents and consumables (including MinION flow cells and sequencer) required to sequence the G6PD gene for 96 and for 960 samples (Table 3). Note that this is not a thorough cost efficiency comparison, but a crude comparison between the two approaches which does not capture all programmatic expenses, including hardware maintenance, service contracts, or personnel time.

For Sanger sequencing, the entire coding exon regions is amplified using seven separate PCR reactions per sample, as described elsewhere<sup>16,23</sup>. Each of these PCR products was sent for purification and sequencing at the cost of \$2.20 per purification and \$4.95 per sequencing reaction (forward and reverse) to Macrogen, Korea, totaling \$7.90 per PCR product. Therefore, the total cost for sequencing a single sample using Sanger was of \$84.70 (7 purifications × \$2.20 + 14 sequencing reactions × \$4.95). However, with an estimated 10–15% discount commonly offered by sequencing providers, the Sanger sequencing cost per sample would decrease to \$76.23–\$72.00, respectively. For 96 samples, this would have amounted to \$8,131 without discount, or \$6,915–\$7,318 after a 10–15% discount, and for 960 samples, the cost would scale linearly to \$81,312 without discount or \$69,150–\$70,272 after discount.

In comparison, the MinION Mk1C Starter Pack was \$4,900 and included a few kits, six flow cells and the MinION sequencer. Once purchased, the consumable costs for 96 samples (including flow cells, barcoding and adapter kits, ligation and end-prep enzymes, and wash kits), was about \$1,180, that is, \$12.29 per sample. The cost of AMPure XP beads (Beckman Coulter, US) purification was approximately \$11.03 per sample (\$1058 total for 96 samples). Thus, the total MinION cost including beads was approximately \$7,138 for 96 samples (sequencer + consumables + beads), or \$2,238 for consumables and beads alone (no sequencer). For 960 samples (10 plates), consumables and beads alone would therefore cost approximately \$22,380, and including the one-time sequencer purchase the total would be about \$27,280. Even with the slightly higher price of the new MinION Mk1D Starter Pack (\$4,950 including five flow cells), the cost advantage over Sanger sequencing remains substantial, especially as sample numbers grow.

Name	ClinVar ID	ClinVar Interpretation	Genomic mutation	mRNA mutation	Protein change	Molecular Consequence	Exon
Aureus	10,402	Pathogenic/Likely pathogenic	16417T>C	143T>C	p.Ile48Thr	Missense	3
Mahidol	10,367	Pathogenic/Likely pathogenic	18078G>A	487G>A	Gly163Ser	Missense	6
Mediterranean	100,057	Pathogenic	18,154 C>T	563 C>T	Ser188Phe	Missense	6
Viangchan	10,386	Pathogenic/Likely Pathogenic	19451G>A	871G>A	Val291Met	Missense	9
Kerala-Kalyan	10,401	Pathogenic/Likely Pathogenic	19529G>A	949G>A	Glu317Lys	Missense	9
Canton	100,058	Pathogenic	20304G>T	1376G>T	Arg459Leu	Missense	12
Kaiping	100,059	Pathogenic	20316G>A	1388G>A	Arg463His	Missense	12

**Table 2.** Seven G6PD variants identified in 53 sequenced samples based on minion results confirmed by Sanger sequencing. The table provides detailed information on each variant, including the corresponding clinvar ID, and clinvar interpretation. The reference sequences used were NCBI: NG\_009015.2:g for genomic mutations, NM\_001360016.2:c for mRNA mutations, and NP\_001346945.1p for protein mutations (Isoform B).

Category	Description	Cost per sample (USD)	Cost for 96 samples (USD)	Cost for 960 samples (USD)
Sanger Sequencing	PCR purification + sequencing (7 fragments $\times$ Fw & Rv)	\$7.90	\$8,131 (no discount)	\$81,312 (no discount)
	Estimated discount (10–15%)	\$72.00–\$76.23	<b>\$6,915 – \$7,318</b>	<b>69,120–\$73,180</b>
MinION Sequencing	Consumables (flowcells, kits, barcoding, ligation, wash kits)	\$12.29	\$1,180	\$11,805
	AMPure XP beads purification (additional cost)	\$11.03	\$1,058	\$10,588
	One-time MinION Mk1C Starter Pack		\$4,900	\$4,900
Total MinION cost	First run (consumables + beads + sequencer)		\$7,138	\$27,293
	Subsequent runs (consumables only, excluding beads)		\$1,180	\$11,805
	Subsequent runs (including beads)		<b>\$2,238</b>	<b>\$22,393</b>

**Table 3.** Cost comparison of Sanger and minion sequencing for 96 and 960 samples of the G6PD gene. For Sanger, the price per sample includes seven PCR product purifications and 14 sequencing reactions. MinION consumables include flow cells, barcoding/adapters, end-prep/ligation enzymes, and wash kits, while ampure XP beads are listed separately. The total minion cost also includes the one-time purchase of the Mk1C sequencer.

Finally, in terms of turnaround time, MinION library preparation is done within a workday and results are delivered within 2–3 days for 96 samples on a same flow cell. On the other hand, PCR of 96 samples can be done in a single workday, as can be Sanger sequencing. However, as most laboratories do not have Sanger sequencing capacity which is quite expensive, PCR products have to be shipped to a sequencing centre, whether a commercial provider or academic collaborator, before being analysed and this can take weeks in malaria endemic areas.

## Discussion

We describe a novel sequencing assay and bioinformatics pipeline that reliably detects seven known G6PD variants associated with deficient enzyme activity, with the potential to identify up to 192 known variants (Supplementary Data 2). The MinION assay accurately genotyped 72 samples, as confirmed by Sanger sequencing. For these samples, we observed high concordance in sequencing results, even when a sample was run multiple times or in duplicate within the same run (Supplementary Data 1). Variant calls were also highly concordant when using two different callers, clair3 and freebayes (Supplementary Data 4) for both MinION and Sanger sequencing data, suggesting that our pipeline is robust across different bioinformatics tools. The only two discordant samples between runs identified during the study were obtained from two Cambodian male and female participants, for which Sanger sequencing was not performed (Supplementary Data 1). Of these, one sample (02\_02\_178) was run in duplicate during run 3 and was identified as homozygous for the Viangchan variant in one replicate and as wild-type in its within-plate duplicate, but the wild-type duplicate had very low read depth. The second sample, (02\_02\_092) was run twice (in run 3 and run 4). While in run 3 it showed low read depth consistent with a wild-type genotype, in run 4 it gave discordant results according to the two variant callers. Specifically, it was identified as heterozygous for the Viangchan variant by clair 3 and as homozygous for the same variant by freebayes. This discrepancy may have arisen from a variety of factors, including differences in variant calling thresholds. Variations in filtering parameters could influence allele frequency detection, potentially leading to false positive homozygous calls or missed heterozygous variants. Further investigation is needed to refine filtering criteria and improve the accuracy of genotype classification.

Sequencing depth varied between samples and runs, and was dependent upon the sequencing running time, fragments size and/or number of samples multiplexed. Sequence depth also varied for duplicates, and while DNA concentrations were quantified and equimolarly pooled at each run, the cause of this variability is unclear. Despite the depth variation, coverage and mapping quality remained consistently high across all runs, with coverage exceeding 97% and average mapping quality scores between 59 and 60. However, this variation may explain why a sample from a Cambodian female was identified differently as homozygous or heterozygous by the two sequencing methods.

The MinION assay identified seven known variants, with one of them having two different heterozygous variants. A total of 19 samples had no known variants detected according to both sequencing methods, of which 3 samples came from female individuals with intermediate G6PD activity (although one of these was close to normal with 68% of AMM activity) and 4 from males phenotypically diagnosed as deficient (Supplementary Data 1). Of these, 1 had G6PD activity determined by both spectrophotometry and by SD Biosensor, and the two tests led discrepant results, while for the other 3 enzyme activity was assessed using the FST method. This could indicate the influence on enzyme activity of novel variants that have yet to be described in the literature

or an error associated to the testing. Further work is now underway to adapt the pipeline for variant discovery and laboratory confirmation.

Geographical mapping of the prevalence of different G6PD variants is complicated by the variety of methods used in different surveys and countries<sup>11</sup>. In our Cambodian dataset, all 68 samples were collected from individuals in Kampong Speu province, located in Western Cambodia. MinION sequencing identified the Viangchan mutation as the most common variant, present in 47 samples (69.1%) (Supplementary Data 1). In contrast, samples from Afghanistan carried the Mediterranean variant, while samples from Vietnam showed Canton, Kaiping and Viangchan mutations. However, the small sample sizes and non-representative sampling in these countries limit the ability to draw broad conclusions about variant distribution and prevalence at population level.

Our study has several limitations. The assay was designed to detect only variants within exon 3–13 excluding the gene promoter region, the first non-coding exon, and exon 2. Additionally, the assay was also assessed in a limited number of geographical areas and known variants, using venous blood samples, so further investigation of other populations and sampling types, such as dried blood spots, is warranted. Nevertheless, the application of this standardized assay in larger populations has significant potential to enhance our understanding of G6PD genetic variants in *P. vivax*-endemic regions, including association of different variants with enzyme deficiency and associated haemolytic risk in patients treated with primaquine or tafenoquine. The assays will also generate valuable insights into G6PD mutations in both coding and non-coding regions of the gene, improving our understanding of genotype-phenotype correlations and their implications for the diagnostics of G6PD deficiency.

In terms of cost-effectiveness, MinION sequencing offers a far more economical approach for sequencing G6PD gene than traditional Sanger sequencing given 96 samples per flow cells are run at a time, emphasizing its suitability for large-scale studies. Additionally, MinION significantly reduced turnaround time, with results delivered within 2–3 days, whereas Sanger sequencing can take weeks depending on the sequencing provider's location.

## Conclusions

The choice of sequencing technology for gene analysis can significantly impact both the financial and logistical aspects of a study, especially when dealing with large sample sizes. We present a novel long-amplicon sequencing assay that reliably detected G6PD mutations in exons 3–13 in Asian populations from Cambodia, Afghanistan, Vietnam and China. Our results demonstrate that MinION sequencing is a cost-effective and time-saving alternative to Sanger sequencing. Additionally, this study highlights the reliability of MinION sequencing, making it a recommended choice for G6PD genotyping in large-scale studies such as cross-sectional surveys and clinical trials. These advantages are particularly valuable in resource-limited settings, where efficiency and affordability are essential.

## Data availability

The data supporting the findings of this study are available in the European Nucleotide Archive (ENA) under the following primary accession number: PRJEB88995.

Received: 9 June 2025; Accepted: 5 December 2025

Published online: 21 December 2025

## References

1. World & Health Organization (WHO). *World Malaria Report 2023*. 2023 Nov 30. Available from: <https://www.who.int/publications/item/9789240086173>
2. Battle, K. E. et al. Mapping the global endemicity and clinical burden of *Plasmodium vivax*, 2000–17: A Spatial and Temporal modelling study. *Lancet* **394** (10195), 332–343. [https://doi.org/10.1016/S0140-6736\(19\)31096-7](https://doi.org/10.1016/S0140-6736(19)31096-7) (2019).
3. Price, R. N., Commons, R. J., Battle, K. E., Thriemer, K. & Mendis, K. *Plasmodium Vivax* in the era of the shrinking *P. falciparum* map. *Trends Parasitol.* **36** (6), 560–570. <https://doi.org/10.1016/j.pt.2020.03.009> (2020).
4. Obadia, T. et al. Developing sero-diagnostic tests to facilitate *Plasmodium Vivax* serological test-and-treat approaches: Modeling the balance between public health impact and overtreatment. *BMC Med.* **20** (1), 98. <https://doi.org/10.1186/s12916-022-02285-5> (2022).
5. Commons, R. J., Simpson, J. A., Watson, J., White, N. J. & Price, R. N. Estimating the proportion of *Plasmodium Vivax* recurrences caused by relapse: A systematic review and Meta-Analysis. *Am. J. Trop. Med. Hyg.* **103** (3), 1094–1099. <https://doi.org/10.4269/ajtmh.20-0186> (2020).
6. Luzzatto, L., Ally, M. & Notaro, R. Glucose-6-phosphate dehydrogenase deficiency. *Blood* **136** (11), 1225–1240. <https://doi.org/10.1182/blood.2019000944>. (2020).
7. Howes, R. E. et al. G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: A Geostatistical model-based map. *PLoS Med.* **9** (11), e1001339. <https://doi.org/10.1371/journal.pmed.1001339> (2012).
8. Pfeffer, D. A. et al. Genetic variants of Glucose-6-Phosphate dehydrogenase and their associated enzyme activity: A systematic review and Meta-Analysis. *Pathogens* **11** (9), 1045. <https://doi.org/10.3390/pathogens11091045> (2022).
9. Minucci, A. et al. Glucose-6-phosphate dehydrogenase (G6PD) mutations database: Review of the old and update of the new mutations. *Blood Cells Mol. Dis.* **48** (3), 154–165 (2012).
10. Sadhewa, A. et al. A review of the current status of G6PD deficiency testing to guide radical cure treatment for Vivax malaria. *Pathogens* **12** (5), 650. <https://doi.org/10.3390/pathogens12050650> (2023). PMID: 37242320; PMCID: PMC10220632.
11. Howes, R. E., Battle, K. E., Satyagraha, A. W., Baird, J. K. & Hay, S. I. G6PD deficiency: Global distribution, genetic variants and primaquine therapy. *Adv. Parasitol.* **81**, 133–201. <https://doi.org/10.1016/B978-0-12-407826-0.00004-7>. (2013)
12. Shendure, J. et al. DNA sequencing at 40: Past, present and future. *Nature* **550** (7676), 345–353 (2017). <https://doi.org/10.1038/nature24286>. PMID: 29045398.
13. Jain, M. et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **36** (4), 338–345. <https://doi.org/10.1038/nbt.4060> (2018).

14. Ley, B. et al. Performance of the access Bio/CareStart rapid diagnostic test for the detection of glucose-6-phosphate dehydrogenase deficiency: A systematic review and meta-analysis. *PLoS Med.* **16** (12), e1002992. <https://doi.org/10.1371/journal.pmed.1002992> (2019). PMID: 31841583; PMCID: PMC6909027.
15. Roca-Feltrer, A. et al. Field trial evaluation of the performances of point-of-care tests for screening G6PD deficiency in Cambodia. *PLoS One.* **9** (12), e116143. <https://doi.org/10.1371/journal.pone.0116143> (2014).
16. Kim, S. et al. Performance of the CareStart™ G6PD deficiency screening test, a point-of-care diagnostic for primaquine therapy screening. *PLoS One.* **6** (12), e28357. <https://doi.org/10.1371/journal.pone.0028357> (2011). Epub 2011 Dec 2. PMID: 22164279; PMCID: PMC3229584.
17. Danecek, P. et al. Twelve years of samtools and BCFtools. *Gigascience* **10** (2), giab008. <https://doi.org/10.1093/gigascience/giab008> (2021). PMID: 33590861; PMCID: PMC7931819.
18. De Coster, W. & Rademakers, R. NanoPack2: Population-scale evaluation of long-read sequencing data. *Bioinformatics* **39** (5), btad311. <https://doi.org/10.1093/bioinformatics/btad311> (2023).
19. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34** (18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191> (2018).
20. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *ArXiv Preprint.* [arXiv:1207.3907](https://arxiv.org/abs/1207.3907) (2012).
21. Zheng, Z. et al. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat. Comput. Sci.* **2** (12), 797–803. <https://doi.org/10.1038/s43588-022-00387-x> (2022). Epub 2022 Dec 19. PMID: 38177392.
22. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford nanopore sequencing. *Genome Biol.* **20** (1), 129. <https://doi.org/10.1186/s13059-019-1727-y> (2019). PMID: 31226931; PMCID: PMC6588364.
23. Matsuoka, H. et al. Glucose-6-phosphate dehydrogenase (G6PD) mutations in Cambodia: G6PD Viangchan (871G>A) is the most common variant in the Cambodian population. *J. Hum. Genet.* **50** (9), 468–472. <https://doi.org/10.1007/s10038-005-0284-y> (2005).

## Author contributions

Ivo Mueller, Ric Price and Jean Popovici conceived the study. Costanza Tacoli and Mariana Kleinecke designed the study. Costanza Tacoli designed the laboratory assay. Costanza Tacoli, Ratha Eam, Muhammad Nadeem, Nimol Khim and Angela Rumaseb contributed to sequencing data generation and laboratory data analysis. Costanza Tacoli, Mariana Kleinecke, Hidayat Trimarsanto, Kian Soon Hoon, and Muhammad Nadeem developed the bioinformatic pipeline and conducted the bioinformatic data analysis. Benedikt Ley, Dean Sayre, Sochea Pok, Dysoley Lek, Jimee Hwang, Benoit Witkowski, Sarah Auburn, Ivo Mueller, Ric Price and Jean Popovici contributed to essential field-based sample collections and metadata, and guidance on the study design and interpretation. Costanza Tacoli, Mariana Kleinecke, Sarah Auburn, Ric Price and Jean Popovici wrote the manuscript. All authors approved the final version of the manuscript.

## Disclaimer

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centres for Disease Control and Prevention or the United States Agency for International Development.

## Funding

The collection of patient samples was funded by the U.S. President's Malaria Initiative, USA, the UK Department for International Development, UK Medical Research Council, UK National Institute for Health Research, and the Wellcome Trust through the Joint Global Health Trials Scheme (MR/K007424/1) and the Bill & Melinda Gates Foundation [INV-024389]. M.N. was supported by the IDEX International Master's Scholarship from Université Paris-Saclay, France. R.N.P. and S.A. are funded by an Australian National Health and Medical Research (NHMRC) Leadership Investigator Grants (2008501 and 2025377). J.P. is supported by the NIH/NIAID (R01AI173171, R01AI175134 and R61AI187100). J.P. and BW are supported by the Pasteur International Unit PyESMEE.

## Declarations

### Competing interests

The authors declare no competing interests.

### Ethical approval

Ethics approval was obtained from the following national and local committees and authorities: The Human Research Ethics Committee of the Northern Territory Department of Health (HREC), Australia, the Islamic Republic of Afghanistan, Ministry of Public Health, Institutional Review Board, Afghanistan, the Oxford Tropical Research Ethics Committee (OxTREC), UK, the National Ethics Committee for Health Research, Cambodia, the U.S. Centers for Disease Control and Prevention, and the Ministry of Health Evaluation Committee on Ethics in Biomedical Research, Vietnam.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-31822-4>.

**Correspondence** and requests for materials should be addressed to J.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025