Biology Methods and Protocols, 2026, **11**(1), bpag008
https://doi.org/10.1093/biomethods/bpag008
Published: 5 February 2026
Methods Article

BIOLOGY Methods & Protocols
OXFORD

Resolving haplotypes of the glucose-6-phosphate dehydrogenase gene using long-range polymerase chain reaction and Oxford Nanopore sequencing

Kamonwan Chamchoy[1], Beatriz Aira C. Jacob[2], and Usa Boonyuen[2,*]

[1]Princess Srisavangavadhana Faculty of Medicine, Chulabhorn Royal Academy, Bangkok 10210,Thailand
[2]Department of Molecular Tropical Medicine and Genetics, Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand

*Corresponding author. Department of Molecular Tropical Medicine and Genetics, Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand. E-mail: usa.boo@mahidol.ac.th

Abstract

Glucose-6-phosphate dehydrogenase (G6PD) deficiency is the most common human enzymopathy and poses a major concern on safe administration of oxidative drugs, including antimalarials such as primaquine and tafenoquine. Common diagnostic approaches, such as enzyme assays, polymerase chain reaction (PCR)-based methods, and short-read genotyping, often fail to identify heterozygous female carriers and are unable to determine the phase of compound heterozygous mutations. To address these limitations, a workflow combining long-range PCR with Oxford Nanopore Technologies (ONT) sequencing was developed, enabling comprehensive analysis of the entire *G6PD* locus with direct haplotype resolution. The accuracy of the developed method was independently validated by Sanger sequencing for exonic variant detection and by adaptive sampling-based ONT sequencing for phasing accuracy. A total of 24 samples (20 females, 4 males) were analyzed using two long-range amplicons (∼12 and ∼14 kb) with an 8.2 kb overlap spanning both variant-rich and variant-sparse regions. ONT sequencing revealed 36 distinct variants across exonic, intronic, and regulatory regions. The design consistently captured multiple informative heterozygous sites, markedly improving haplotype reconstruction in females. Nanopore sequencing generated long reads (N50 ∼11 kb) with deep coverage (>700-fold), supporting accurate variant detection and phasing. These findings demonstrate the feasibility and robustness of a nanopore-based long-read approach for comprehensive *G6PD* haplotyping, integrating variant detection and phasing within a single analytical workflow, and providing a foundation for future studies on carrier detection and other X-linked genes.

Keywords G6PD deficiency, haplotype phasing, Oxford Nanopore sequencing, long-range PCR, X-linked disorders

Introduction

Glucose-6-phosphate dehydrogenase (G6PD), a key enzyme in the pentose phosphate pathway, generates the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH), an essential molecule for cellular protection against oxidative stress. Because red blood cells depend almost exclusively on G6PD for NADPH production, individuals with G6PD deficiency are highly vulnerable to acute hemolysis when exposed to oxidative triggers such as fava beans, certain drugs, and infections [1, 2]. This concern is especially relevant for the administration of antimalarial agents like primaquine and tafenoquine in malaria-endemic regions, where G6PD deficiency is highly prevalent [1–3]. Moreover, neonates with G6PD deficiency are at increased risk of developing severe hyperbilirubinemia, which can progress to kernicterus, a potentially fatal neurological disorder [4, 5]. The severity of drug-induced hemolysis in G6PD-deficient individuals ranges from mild to life-threatening and is influenced by both the extent of oxidative stress and the patient's G6PD status. As an X-linked disorder, G6PD deficiency manifests predominantly in hemizygous males and homozygous females. In heterozygous females, however, a wide spectrum of G6PD activity can be observed due to X-chromosome inactivation, ranging from deficient, similar to hemizygous males, to normal levels. As a consequence, heterozygous females may also experience acute hemolytic episodes when exposed to oxidative stress [1, 2, 6].

Diagnosis of G6PD deficiency is primarily based on phenotypic assays. These include spectrophotometric enzyme activity measurement, the fluorescent spot test (FST), and rapid diagnostic tests

Received: 05 November 2025; **Revised:** 20 January 2026; **Accepted:** 29 January 2026

designed for field use. While these methods can accurately identify hemizygous males and homozygous females who present with markedly reduced enzyme activity, their performance is less reliable in individuals with intermediate activity [7–10]. As a result, a considerable number of heterozygous carriers are overlooked, with published estimates indicating that 30%–35% are missed [11, 12]. These diagnostic limitations highlight the need for complementary genetic approaches to improve identification of heterozygous carriers, particularly those who may have intermediate enzyme activity that cannot be accurately detected by phenotypic assays.

The *G6PD* gene (NG_009015.2) spans approximately 16,197 bp and comprises 13 exons and 12 introns. It transcribes a 1548-base pair (bp) mRNA that encodes a 515-amino-acid G6PD enzyme. Notably, exon 1 is a noncoding region, with the translation start codon (ATG) located in exon 2. The *G6PD* promoter region features a TATA-like sequence (TTAAAT) located within a CpG island upstream of exon 2 [13, 14]. To date, over 200 *G6PD* variants have been reported worldwide, including both coding and noncoding variants with variable or unknown clinical significance, while most reported variants are single-nucleotide substitutions within the coding region of the *G6PD* gene. These missense mutations cause a change in amino acid that compromise the structural stability and enzymatic function of G6PD [2]. Compound mutations, involving multiple missense variants have also been described [15–17]. Several noncoding variants of the *G6PD* gene, including those located in the 5′ and 3′ untranslated regions, have been frequently reported across multiple populations. These variants include c.486-34delT, c.1311C>T, c.1365-13T>C, and c.*357G>A [16, 18]. However, their functional significance remains unclear and warrants targeted investigation to determine whether they influence G6PD expression through alterations in transcriptional regulation, pre-mRNA splicing, or mRNA stability [18–20].

Extending genetic analysis beyond coding regions to encompass noncoding and regulatory sequences may provide additional insights into the genetic architecture of *G6PD*. This approach is particularly relevant in heterozygous females, in whom random X-chromosome inactivation results in variable enzyme activity and complicates phenotype prediction. Although the functional consequences of noncoding variants have yet to be fully established, their inclusion in comprehensive analyses may support future efforts to improve carrier detection and enhance understanding of genotype–phenotype relationships.

However, current approaches to *G6PD* genotyping remain insufficient. Previously, genotyping of the *G6PD* gene has relied on methods such as polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP), allele-specific polymerase chain reaction (AS-PCR), or high-resolution melting (HRM) analysis, which primarily target known variants, limiting its ability to detect novel variants. Although Sanger sequencing is widely regarded as the gold standard for genotyping, it is not well suited for comprehensive analysis of the full genomic region of *G6PD*. Sequencing across the entire locus typically requires multiple overlapping PCR reactions, which is labor intensive and limits throughput. In addition, the short-read length of Sanger sequencing makes it difficult to interpret complex variation and does not reliably resolve whether coexisting variants are in *cis* or in *trans*. As a result, phasing multiple variants across the gene, particularly in heterozygous females with compound heterozygosity, remains challenging without additional information such as family-based data.

Advances in DNA sequencing technologies have led to the development of third-generation sequencing, enabling the direct analysis of long nucleotide sequences. Among these, the long-read sequencing platform developed by Oxford Nanopore Technologies (ONT) presents a promising approach to overcome the limitations of *G6PD* analysis. ONT offers real-time sequencing with lower costs, a compact desktop-sized device, and scalable throughput options. ONT reports that nanopore sequencing can generate reads of approximately 10–100 kb [21, 22]. In the current long-range PCR-based workflow, these read lengths enable comprehensive characterization of sequence variation across the entire *G6PD* gene, including coding, intronic, and untranslated regions, while also supporting accurate haplotype phasing.

In this study, an integrated sequencing and analysis pipeline was developed (Fig. 1a), combining long-range PCR with ONT, to achieve full-length variant profiling and direct haplotype phasing in heterozygous females, an outcome not attainable with conventional short-read sequencing.

Methods

Ethics approval and study samples

The study procedures were approved by the Human Ethics Committee, Faculty of Tropical Medicine, Mahidol University (MUTM 2021-075-05) and the Human Ethics Committee, Chulabhorn Royal Academy (approval number 074/2566). Twenty-four archived diagnostic specimens were randomly selected based on DNA availability and meeting minimum quality/quantity requirements for long-range PCR and Oxford Nanopore sequencing. Samples were excluded if DNA was unavailable or insufficient for long-range PCR, or if long-range PCR failed. No selection was made based on clinical characteristics or known *G6PD* genotypes. The data were fully anonymized, and the authors had no access to information to identify individual participants.

DNA extraction

Genomic DNA (gDNA) was extracted from 24 blood samples (20 females and 4 males) using the QIAamp DNA Blood Mini Kits (QIAGEN, Germany), following the manufacturer's instructions. DNA concentration was measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA).

Targeted amplification of *G6PD* gene

The reference sequence for *G6PD* (accession ID NG_009015.2) retrieved from The National Institute of Health's National Center for Biotechnology Information (NCBI) GenBank (http://www.ncbi.nlm.nih.gov/gene/) was used as a target for PCR amplification. PCR primers were designed using Primer3 (https://primer3.ut.ee/) and subsequently analyzed for secondary structure and potential dimer formation using OligoEvaluator (https://www.oligoevaluator.com/). Primer locations and amplicon coverage across the *G6PD* locus are shown in Fig. 1b. The primers used are listed in Table 1. Each 50 μL PCR reaction contained 1× PrimeSTAR GXL Buffer, 200 μM of each dNTP, 0.3 μM of each primer, 2.5 μL gDNA, sterile water and 1.25 U PrimeSTAR GXL DNA Polymerase (Takara Bio, USA). The PCR

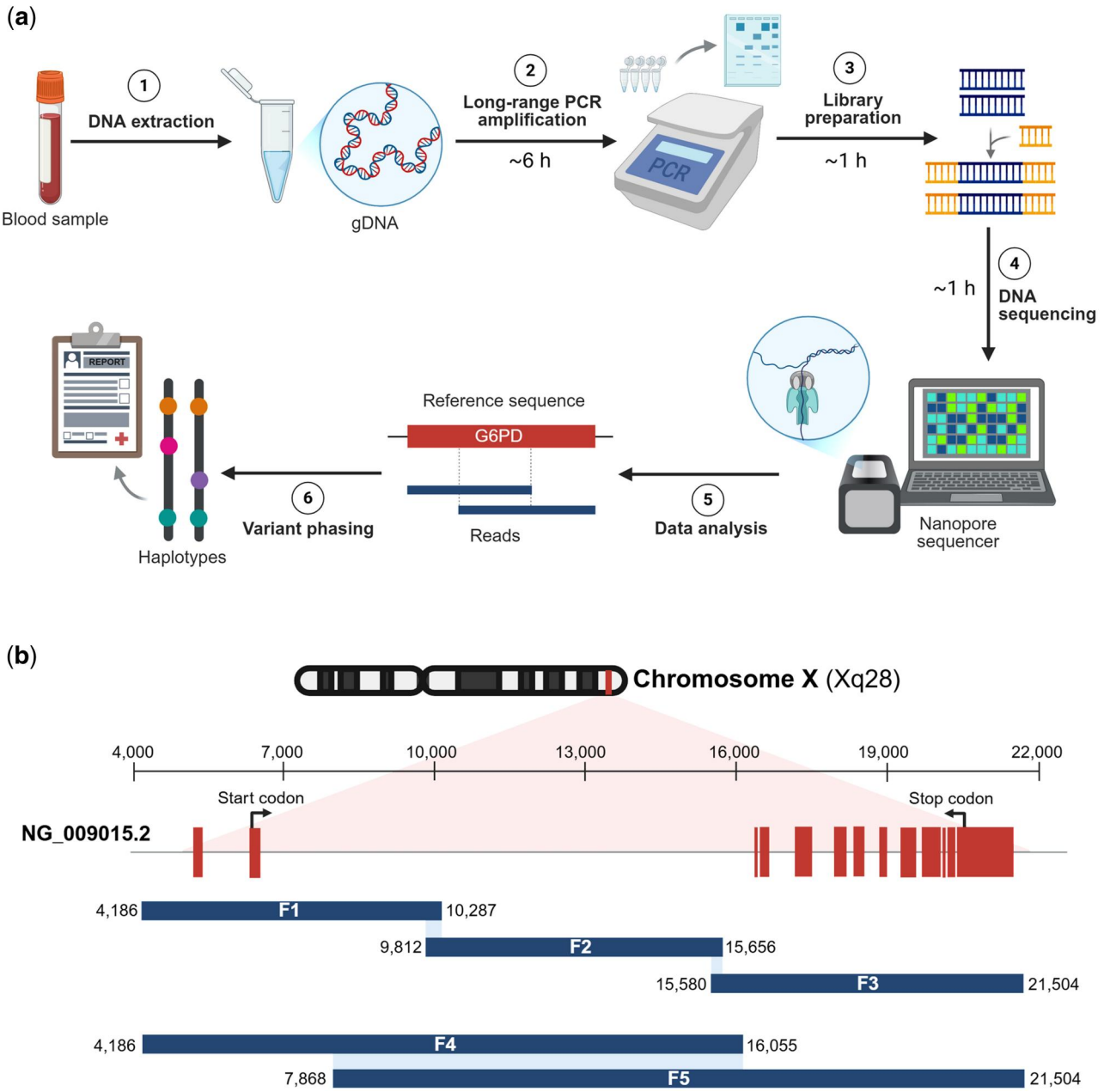

Figure 1 Study workflow and amplicon design. (a) Overview of the long-range PCR and Oxford Nanopore sequencing workflow, followed by basecalling, alignment, variant calling, and read-based phasing to generate haplotype-resolved variants. (b) Primer locations and amplicon coverage across the *G6PD* locus (NG_009015.2) on chromosome X (Xq28); overlapping regions between adjacent amplicons are indicated.

conditions were as follows: 1 cycle of 98°C for 5 minutes, followed by 30 cycles of 98°C for 10 seconds and 68°C for 7 minutes per cycle for fragments 1–3 or 13 minutes 30 seconds per cycle for fragments 4–5, with a final extension at 68°C for 5 minutes. For amplicons ≥10 kb in length (fragments 4–5), primers were used at a final concentration of 0.25 μM each. Thereafter, PCR products were analyzed by 1% agarose gel electrophoresis.

PCR products were pooled and subjected to purification using PrimeWay Gel Extraction/PCR Purification Kit (1st Base, Malaysia), following the manufacturer's protocols. The concentration of purified PCR products was determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA).

Library preparation for Oxford Nanopore sequencing

Sequencing of the full-length *G6PD* gene was performed on 24 unique samples using the Ligation Sequencing Amplicons-Native Barcoding Kit 96 V14 (SQK-NBD114.96), following the manufacturer's protocols. This included 18 samples analyzed via fragments 1–3 and 14 via fragments 4–5, with eight samples overlapping both designs to resolve ambiguous zygosity. In brief, the amplicons were subjected to end-repair treatment reaction consisting of 1.75 μL Ultra II end prep reaction buffer (New England Biolabs [NEB]), 1 μL diluted DNA control sample (DCS), 0.75 μL Ultra II end prep enzyme mix (NEB), and 11.5 μL equimolar PCR products from each sample (200 fmol). The

Table 1 Primers used for amplification of *G6PD* gene.

Product	Primer	Sequence (5′–3′)	Primer (bp)	Primer T_m (°C)	Product size (bp)
Fragment 1	F1_F	TAAGACGTAATTCCGATCACATTCC	25	59.02	6101
	F1_R	GGAAACTAAGGCCCAGAGAAGT	22	59.61	
Fragment 2	F2_F	CACAAACACAGAAGGAAGAGCAAATG	26	61.44	5845
	F2_R	TAAAAGCTAACCCAAAGAAAAGATGAAC	28	59.13	
Fragment 3	F3_F	GGGGTTGTCTGTTTCTATTCGTC	23	59.32	5925
	F3_R	CCATCCCCTCTCCTTCCTAAATAC	24	59.71	
Fragment 4	F1_F	TAAGACGTAATTCCGATCACATTCC	25	59.02	11,869
	F4_R	GGAGCTAAGCATTCCACTTAAGAAG	25	59.70	
Fragment 5	F5_F	CTGACTCCCCATAAGCCTGAG	21	59.58	13,636
	F3_R	CCATCCCCTCTCCTTCCTAAATAC	24	59.71	

reaction was incubated at 20°C for 5 minutes, followed by 65°C for 5 minutes.

Next, native barcodes were ligated to the end-repaired PCR products. The ligation reaction included 3.75 μL of PCR products, 1.25 μL of native barcode (ONT) and 5 μL of Blunt/TA ligation master mix (NEB). The mixture was incubated at room temperature for 20 minutes and stopped by adding 2 μL EDTA (ONT). Barcoded samples were pooled and mixed with 0.4× AMPure XP beads. Clean-up was performed with 80% ethanol washes and the barcoded DNA library was eluted in 35 μL nuclease-free water.

Adapter ligation was then performed by combining 30 μL pooled barcoded samples, 5 μL native adaptor (ONT), 10 μL NEBNext quick ligation reaction buffer, and 5 μL Quick T4 DNA ligase (NEB). This reaction was incubated at room temperature for 30 minutes, followed by clean-up using 20 μL AMPure XP beads and washing with Long Fragment Buffer (ONT). The final library was eluted in 25 μL of Elution Buffer (ONT) after a 10-minute incubation at 37°C. The library was loaded onto R10.4.1 PromethION flow cells and sequenced for 2 hours.

Validation of variant calls and phasing accuracy

To validate variant calling, PCR amplification covering all 13 exons of the *G6PD* gene was performed on all 24 samples, according to protocols described in a previous report [23]. The resulting PCR products were purified and subsequently subjected to DNA sequencing (1st Base, Malaysia). Sequencing chromatograms were visualized using SnapGene Viewer version 8.2.1.

Furthermore, phasing accuracy was validated using adaptive sampling performed on eight samples (two males and six females), following ONT pharmacogenomic (PGx) sequencing workflow, with four samples sequenced per flow cell. Male samples, which are hemizygous for the X-linked *G6PD* gene, were included as internal controls to confirm accurate variant detection and the absence of spurious phasing, while female samples enabled direct assessment of haplotype resolution. In brief, 2 μg of extracted gDNA was subjected to DNA repair and end-prep treatment in a 60 μL reaction containing Ultra II end prep reaction buffer (3.5 μL), Ultra II end prep enzyme (3 μL), NEBNext FFPE DNA repair buffer (3.5 μL), NEBNext FFPE DNA repair mix (3.5 μL), and 48 μL of gDNA. The reaction was carried out at 20°C for 15 minutes before heat inactivation at 65°C for 5 minutes. The reaction was purified with AMPure XP beads and eluted in 20 μL nuclease-free water. Thereafter, 1.6 μg end-prepped DNA was subjected to native barcode ligation at room temperature for 25 minutes and AMPure XP beads purification with a final elution in 32 μL nuclease-free water. Adaptor ligation was carried out for 30 minutes at room temperature using 3.2 μg DNA in 50 μL reaction, followed by AMPure XP beads purification with a final elution in 97 μL elution buffer. The 500 ng DNA library was prepared and loaded onto R10.4.1 PromethION flow cells and sequenced for 72 hours with intermittent wash and reload every 24 hours.

Data analysis

For amplicon-based sequencing, raw DNA sequencing data were obtained in the POD5 format using the MinKNOW version 24.06.16 software (https://nanoporetech.com). Reads were basecalled, trimmed, and demultiplexed using Dorado version 7.4.14 with a super accuracy (SUP) model running at 400 bases per second (bps) (dna_r10.4.1_e8.2_400bps_sup@v4.3.0). The processed data was generated in FASTQ format. While ONT recommends a quality score above 7, a stricter threshold of Q10 was applied to pass reads for downstream genotyping to increase stringency and decrease the global error rate [24, 25]. For the first set of fragments (F1, F2, and F3), filtered reads were obtained via Filtlong version 0.2.1 (https://github.com/rrwick/Filtlong) at a minimum length of 5000 bp. The minimum read length was 10,000 bp for the second set of fragments (F4 and F5). Filtered FASTQ reads were aligned using minimap2 version 2.24 [26, 27] to the *G6PD* reference sequence (NG_009015.2) and the GRCh38 whole-genome reference; the latter was used to validate results against adaptive sampling data. Sequence metrics were obtained using Nanoplot version 1.46.0 [28] and NanoStat version 1.6.0 [29]. File manipulation steps were performed with samtools version 1.17 [30], including conversion from sequence alignment map (SAM) to the sorted binary alignment map (BAM) format with a minimum mapped read length of 5000 bp, as well as depth and coverage calculations. Variants were called using Clair3 version 1.0.10 with thresholds of 0.25 for SNPs and 0.3 for INDELs. The higher threshold for INDELs was utilized to account for the inherently higher error rate associated with ONT for these variants [31, 32]. Phased variant call format (VCF) files were acquired using WhatsHap version 1.7 [33]. Aligned and phased reads were visualized using Integrative Genomics Viewer (IGV) version 2.19.1 [34, 35].

For adaptive sampling, live basecalling and mapping (hg38) were performed using the MinKNOW version 25.09.16 and Dorado version 7.11.2 with the high accuracy model at 400 bps (dna_r10.4.1_e8.2_400bps_hac@v5.2.0), and a minimum *Q* score of 9. Genetic variants in

the *G6PD* gene were identified and haplotyped in the epi2me-labs suite through the pharmacogenomics workflow (wf-pgx version 0.1.7).

Results

Initial development of long-read sequencing of *G6PD* gene

Based on the NG_009015.2 reference sequence, the *G6PD* gene spans 16,197 bp and is located within the genomic coordinates 4987–21,183. Long-range PCR followed by Oxford Nanopore sequencing was developed to enable haplotype-resolved variant detection across the full locus (Fig. 1a). In the absence of prior information about variant-rich regions within the *G6PD* gene, primers were designed to amplify the full-length gene in three overlapping fragments, each approximately 6 kb in size (Fig. 1b). All three amplicons were successfully generated using a total PCR run time of 3 hours 40 minutes. The amplicon sizes were 6102, 5845, and 5925 bp for fragments 1, 2, and 3, respectively (Fig. 2a). All fragments produced bands of the expected sizes on agarose gel electrophoresis.

Using three targeted amplicons (fragments 1–3), *G6PD* variants were successfully identified in male samples by ONT long-read sequencing. In females, phasing was limited: among 14 cases, only three (IDs: F1, F3, and F19) could be resolved with these initial fragments (Fig. 3). Notably, all variants detected in one sample (ID: F3) were homozygous across the *G6PD* gene (Fig. S1). In the remaining cases, haplotype determination was not achieved. The limitation was the short overlap regions, in which heterozygous variants were either absent or present in only a single region, thereby preventing haplotype linkage across fragments (Fig. S1).

To overcome this limitation, variant distribution across the female samples was assessed using IGV. The overall pattern of variant distribution across the *G6PD* locus was broadly comparable among individuals, although the number of heterozygous sites differed considerably. Some samples carried only a limited set of heterozygous variants, whereas others had more extensive sets; however, in all cases the distribution was uneven. Clusters of heterozygous sites were consistently observed within genomic coordinates 9000–10,000 and 20,000–22,000. Extended intervals including genomic coordinates 13,000–18,000, contained few heterozygous positions (Fig. 4). Consequently, informative variants were often absent from fragments 1 to 3 overlaps, hindering haplotype linkage despite sufficient read length and sequencing depth.

Extended long-range PCR design for improved haplotype resolution of *G6PD*

Based on the observed distribution of heterozygous variants across the *G6PD* gene, a new strategy for long-range PCR amplification was developed. The existing primer sets were adapted to generate two amplicons with substantially increased overlap. These newly designed fragments, designated as fragment 4 and fragment 5, spanned genomic coordinates 4186–16,055 (11,869 bp) and 7868–21,504 (13,636 bp), respectively, on the *G6PD* reference sequence (Fig. 1b). Amplification of fragments 4–5 was successfully achieved using a total PCR runtime of approximately 6 hours 50 minutes (Fig. 2b). The two new amplicons produced an 8187 bp overlap (NG_009015.2:7,818–16,055), which consistently contained multiple heterozygous variants (Fig. 4). This larger overlap region ensured that informative sites were included, thereby providing the necessary anchors for haplotype linkage across the gene. In addition, the design of fragments 4 and 5 deliberately encompassed both variant-rich intervals (NG_009015.2:9000–10,000) and variant-sparse intervals (NG_009015.2:13,000–18,000). By simultaneously capturing regions with high heterozygous density and bridging across low-density stretches, the continuity of haplotypes was preserved. As a result, haplotype reconstruction was markedly improved in the previously unresolved female samples (Fig. 5).

Sequencing performance metrics

Nanopore sequencing of the long-range PCR products generated high-quality reads for both fragment sets (Table 2). For fragments 1–3, a total yield of 244,314,606 bp was obtained from 18 barcoded samples, corresponding to 73,785 reads, of which 73,771 passed the quality threshold (Q-score > 10). The N50 was 5847 bp, and 26,350 reads were successfully mapped to the *G6PD* reference sequence. The average per-base coverage depth across the locus was 500-fold (range 252–1201-fold). The mean *Q*-score was 15.1, corresponding to an estimated sequencing accuracy of 96.9%, while high-depth consensus

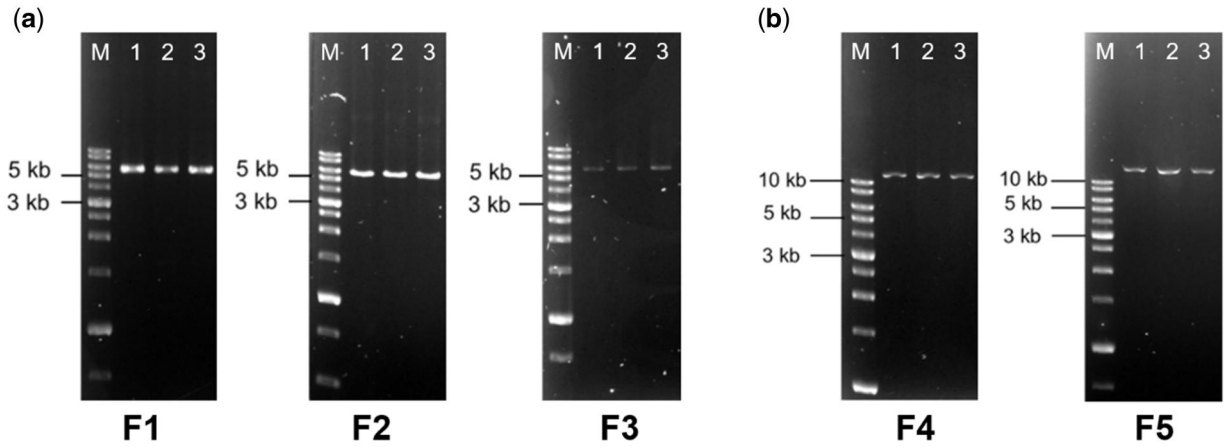

Figure 2 Agarose gel electrophoresis of long-range PCR amplicons. (a) Initial development using fragments 1–3. (b) Extended long-range PCR using fragments 4–5. Lanes 1–3 show representative samples; M, DNA size marker.

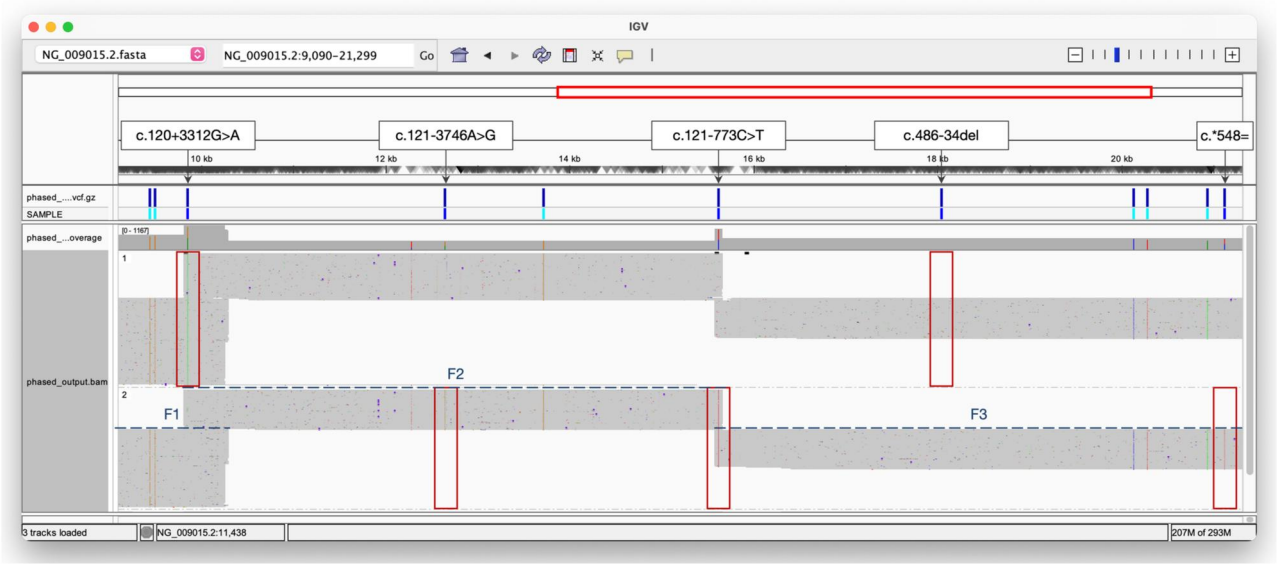

Figure 3 Representative phasing of *G6PD* variants in a heterozygous female sample (ID: F19). Phasing was successfully achieved using fragments 1–3, owing to the presence of heterozygous variants (c.120 + 3312G > A and c.121-773C > T) within the overlapping regions. Heterozygous variants are indicated by red boxes.

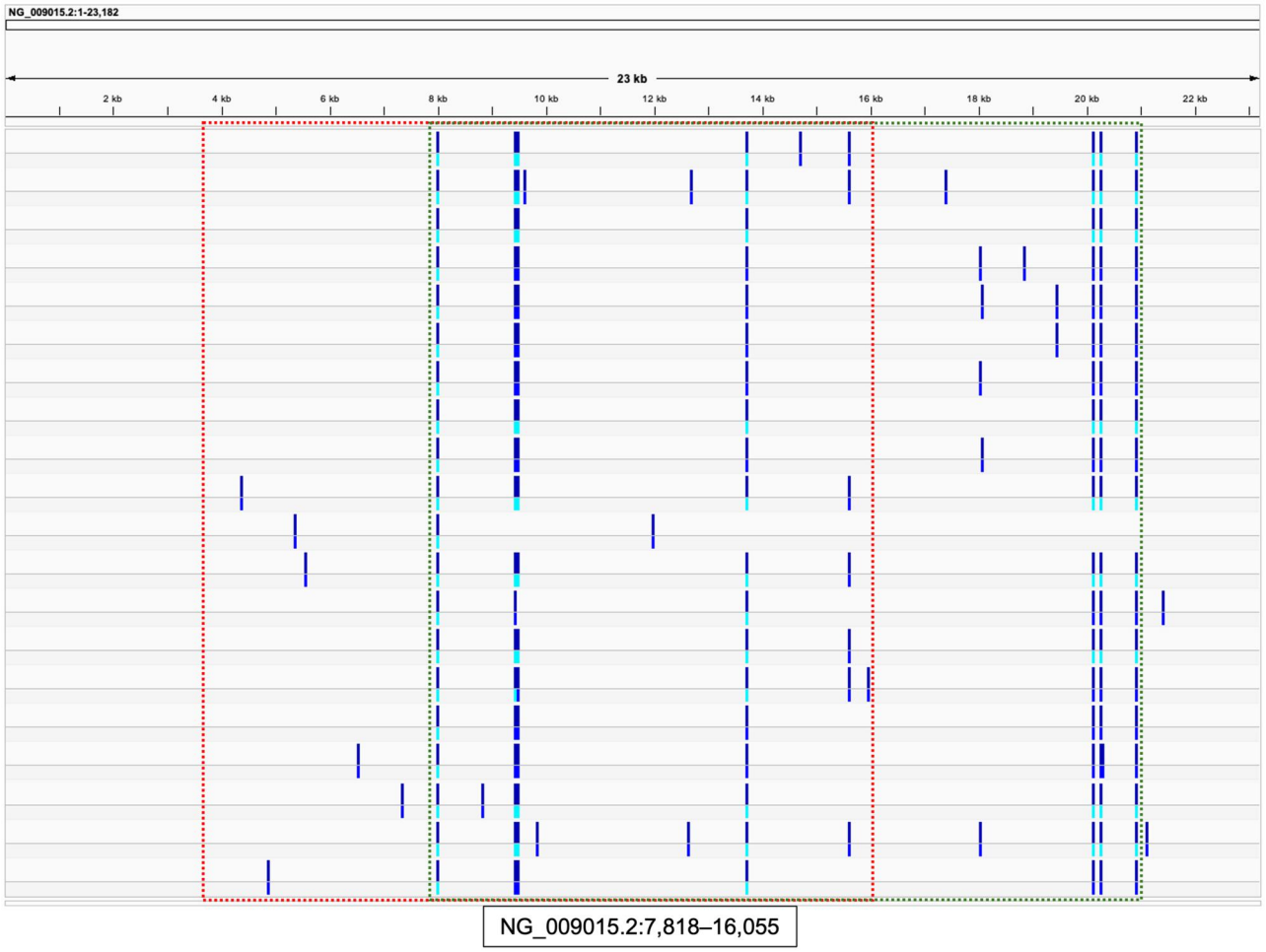

Figure 4 Distribution of variants across the *G6PD* gene in female samples visualized by IGV. Variants are shown as cyan bars for homozygous positions and blue bars for heterozygous positions.

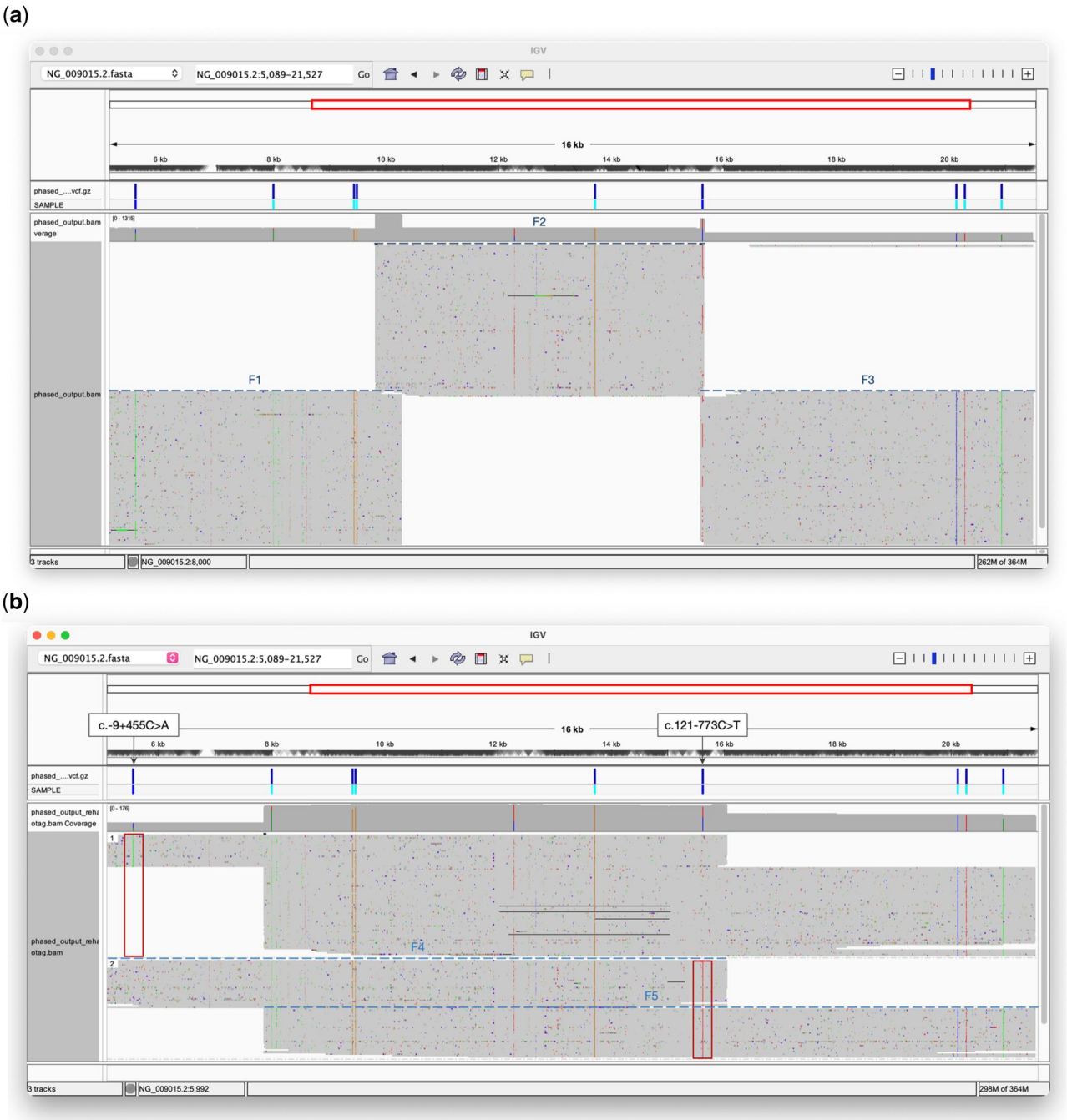

Figure 5 Representative phasing of *G6PD* variants in a heterozygous female sample (ID: F12). (a) Phasing was not successful using fragments 1–3. (b) Phasing with fragments 4 and 5 enabled accurate haplotype reconstruction, demonstrating that the heterozygous variants c.-9 + 455C > A and c.121-773C > T are located in *trans* on different haplotypes. Heterozygous variants are indicated by red boxes.

building resulted in a lower observed post-alignment error rate of 1.18%.

For the extended design (fragments 4–5), sequencing of 14 barcoded samples yielded 321,330,978 bp with 48,971 total reads, of which 48,968 passed the *Q*-score filter. The read length N50 increased to 10,948 bp, and 14,404 reads were aligned to the reference. Coverage was deeper than the initial fragment set, with an average per-base depth of 763-fold (range 35–1648-fold). The mean *Q*-score improved slightly to 15.6 (~97.2% raw read accuracy), with an observed post-alignment error rate of 1.79%. Despite a wider range of per-sample depths, high uniformity was maintained, with 97.0% and 87.6% of targeted bases achieving at least 30- and 50-fold coverage, respectively. While the lower proportion of mapped reads relative to total pass reads reflects stringent alignment and demultiplexing criteria, the extended amplicons produced longer reads and more robust coverage breadth. This improved data quality directly facilitated superior haplotype phasing, successfully phasing 100.0% of heterozygous sites and yielding a median phase-block

Table 2 Sequencing metrics of *G6PD* long-range PCR amplicons obtained from two designs.

	Metrics	
Parameters	**Fragments 1–3**	**Fragments 4–5**
Pre-alignment		
Number of barcodes	18	14[a]
Total yield	244,314,606 bp	321,330,978 bp
Read length (mean ± SD)	3469.8 ± 2,166.0 bp	6000.9 ± 4,592.3 bp
Read length (N50)	5847 bp	10,948 bp
Total number of reads	73,785 reads	48,971 reads
Pass (*Q*-score > 10)	73,771 reads	48,968 reads
Q-score (mean ± SD)	15.1 ± 0.2	15.6 ± 0.2
Post-alignment		
Reads mapped to the reference (proportion of mapped reads)	26,350 reads (35.7%)	14,404 reads (29.4%)
Average per-base coverage depth, fold (range)	500.2 (252.1–1201.0)	763.4 (34.8–1647.9)
Percentage of bases above 30-fold	100.0	97.0
Percentage of bases above 50-fold	100.0	87.6
Error rate (%)	1.18	1.79
Phasing among samples with heterozygous variants		
Proportion of heterozygous sites phased (%)	71.6	100.0
Median phase-block length (range)	1493 bp (0–11,272 bp)	11,493 bp (6615–16,054 bp)
Proportion of gene covered by longest phase block (%)	69.6	99.1

[a] Eight unresolved samples from the initial design (fragments 1–3) were reanalyzed using fragments 4–5.

length of 11,493 bp, and the longest phase block covering 99.1% of the *G6PD* gene. Detailed per-sample metrics are present in Table S1.

Detected variants and haplotype configuration of *G6PD* in the studied samples

Across the 24 DNA samples analyzed (20 females and 4 males), a total of 36 distinct variants were identified within the *G6PD* locus, comprising 22 intronic, 6 exonic, 2 in the 3′ UTR, 2 upstream, and 1 downstream (Table 3). Seven variants were highly recurrent in this study, each with an allele frequency >0.7, including c.120 + 1469T>A (rs112987151), c.120 + 2900T>G (rs2472394), c.120 + 2955A>G (rs2472393), c.121 – 2679C>G (rs2515685), c.1311T>C (rs2230037), c.1365 – 13C>T (rs2071429), and c.*357G>A (rs1050757). Most of them are intronic variants in intron 2, except for one synonymous exonic change (c.1311T>C) and one 3′UTR variant (*c.357G>A). Despite their high frequency, only c.1311T>C has been assigned a clinical classification by World Health Organization (WHO) 2022 (Class C; no hemolytic risk). Consistently, ClinVar annotations indicated minimal clinical relevance overall, reporting four variants as benign, one as of uncertain significance (c.*357G>A), and two without a formal assertion of pathogenicity (Table S2) [36].

Five missense substitutions of clinical relevance were detected. According to WHO 2022 classifications, *G6PD* Mahidol (c.487G>A, rs137852314), *G6PD* Viangchan (c.871G>A, rs137852327), *G6PD* Canton (c.1376G>T, rs72554665), and *G6PD* Valladolid (c.406C>T, rs979416826) are assigned to Class B, associated with acute hemolytic anemia. Each of these variants is consistently classified as pathogenic in ClinVar. The fifth missense variant, *G6PD* Gond (c.477G>C, rs370918918), remains unclassified by WHO and carries conflicting interpretations in ClinVar [36, 37]. Notably, all five missense variants were detected at relatively low allele frequencies in this study (0.023–0.068), consistent with their classification as pathogenic but uncommon alleles. The observed low frequencies may also be influenced by the limited sample size and therefore should be interpreted with caution. Allele frequency analysis additionally revealed distinctive distribution patterns in this study.

Among the two individuals carrying the Viangchan mutation (c.871G>A), both also harbored c.1311T, c.1365-13C, and c.*357G, consistent with the well-documented linkage disequilibrium pattern of these variants in Asian populations [16, 18, 20, 38, 39]. Notably, the same combination of c.1311T, c.1365 – 13C, and c.*357G was also observed in several individuals without the Viangchan allele, indicating that this haplotype occurs independently of the coding mutation (Tables S2 and S3). Moreover, a long-range *cis* haplotype defined by six common variants (c.120 + 2900G, c.120 + 2955G, c.121 – 2679G, c.1311C, c.1365 – 13T, and c.*357A) was detected, which was present on 31 of 44 X chromosomes (70.5%) (Tables S1 and S2). Finally, allele frequency analysis revealed that the A allele at c.120 + 1469 was fixed in this study across all 44 X chromosomes examined (allele frequency =1.00), with complete absence of the reference T allele (Table 3).

All variants in the four male samples were detected in the hemizygous state, consistent with the X-linked inheritance of *G6PD*. As G6PD is an X-linked gene, male samples inherently carry a single haplotype and therefore do not require phasing. Consequently, haplotype reconstruction was applicable only to female samples. Among the 20 females, the number of heterozygous variants ranged from 0 to 9 per individual (median 5.5), with uneven distribution across the locus (Tables S2 and S3). With the initial three-amplicon design (fragments 1–3), haplotype phasing was successful in only two females (F1 and F19), which carried two and five heterozygous variants, respectively (Table S3). In both cases, informative heterozygous sites were located within the overlap regions of fragments 1–3, allowing for haplotype reconstruction. In contrast, several other females carried six to nine heterozygous variants but could not be phased, as most heterozygous positions were located outside these short overlaps. Clusters of heterozygous variants were most frequently observed in intron 2 and the 3′ UTR region (Table 3 and Table S3). The extended design generated an 8187 bp overlap between fragments 4 and 5, which encompassed the cluster within intron 2 as well as mid-gene regions that

Table 3 Summary of distinct *G6PD* variants detected across the study samples.

Genomic position (NG_009015.2)	mRNA position (NM_001042351.1)	Location, Variant	Hemizygous male (X*Y)	Homozygous female (X*X*)	Heterozygous female (X*X)	Cohort allele frequency
g.4387C>T	N/A	2 kb upstream	0	0	1	0.023
g.4876T>C	N/A	2 kb upstream	0	1	1	0.068
g.5367C>T	c.-9 + 262C>T	Intron 1	0	0	1	0.023
g.5560C>A	c.-9 + 455C>A	Intron 1	0	0	1	0.023
g.6544A>C	c.120 + 7A>G	Intron 2	0	0	1	0.023
g.6781C>T	c.120 + 244C>T	Intron 2	0	0	1	0.023
g.7352G>A	c.120 + 815G>A	Intron 2	0	0	1	0.023
g.7498G>A	c.120 + 961G>A	Intron 2	1	0	0	0.023
g.8006T>A	c.120 + 1469T>A	Intron 2	4	20	0	1.000
g.8845G>T	c.120 + 2308G>T	Intron 2	0	0	2	0.045
g.9437T>G	c.120 + 2900T>G	Intron 2	3	10	9	0.727
g.9492A>G	c.120 + 2955A>G	Intron 2	3	9	10	0.705
g.9849G>A	c.120 + 3312G>A	Intron 2	0	0	1	0.023
g.9618G>A	c.120 + 3081G>A	Intron 2	0	0	1	0.023
g.11982A>G	c.121-4413A>G	Intron 2	0	0	1	0.023
g.12649A>G	c.121-3746A>G	Intron 2	0	0	1	0.023
g.12693C>T	c.121-3702C>T	Intron 2	0	0	1	0.023
g.13716C>G	c.121-2679C>G	Intron 2	3	12	7	0.773
g.13801G>C	c.121-2594G>C	Intron 2	0	0	1	0.023
g.14031A>T	c.121-2364A>T	Intron 2	0	0	2	0.045
g.14708G>A	c.121-1687G>A	Intron 2	0	0	1	0.023
g.14746G>A	c.121-1649G>A	Intron 2	0	0	1	0.023
g.15622C>T	c.121-773C>T	Intron 2	0	0	8	0.182
g.15964G>T	c.121-431G>T	Intron 2	0	0	1	0.023
g.17326C>T	c.406C>T	Exon 5, Valladolid	1	0	0	0.023
g.17397G>C	c.477G>C	Exon 5, Gond	0	0	1	0.023
g.18043del	c.486-34del	Intron 5	0	0	3	0.068
g.18078G>A	c.487G>A	Exon 6, Mahidol	1	0	2	0.068
g.18865C>T	c.771-39C>T	Intron 7	0	0	1	0.023
g.19451G>A	c.871G>A	Exon 9, Viangchan	0	0	2	0.045
g.20134T>C	c.1311T>C	Exon 11, Synonymous	3	9	10	0.705
g.20280C>T	c.1365-13C>T	Intron 11	3	9	10	0.705
g.20304G>T	c.1376G>T	Exon 12, Canton	1	0	1	0.045
g.20930G>A	c.*357G>A	3′ UTR	3	9	10	0.705
g.21121C>T	c.*548C>T	3′ UTR	0	0	1	0.023
g.21416G>A	N/A	500 bp downstream	0	0	1	0.023

Allele frequencies reported here were calculated within this technical-validation cohort (20 females, 4 males; total alleles = 44 for the X-linked G6PD locus) and are not intended to represent population-level estimates.

were largely variant-sparse (Fig. 1b). This broader overlap provided sufficient informative sites for phasing and enabled haplotype reconstruction in additional female samples that had previously failed with fragments 1–3.

Validation of variant calls and phasing accuracy

Variant detection was independently validated by PCR amplification followed by Sanger sequencing of clinically relevant exonic *G6PD* variants (Mahidol, Viangchan, Canton, Valladolid, and Gond), a synonymous exonic variant (C.1311T>C), and the intronic variant (c.1365-13C>T). The Sanger sequencing results showed complete concordance with variants identified by the long-read Oxford Nanopore sequencing approach, including correct identification of zygosity status, (Fig. 6). This corresponded to 100% agreement across all validated variants, thereby supporting the accuracy of variant calling (Tables S4 and S5).

Phasing accuracy was also evaluated using an Oxford Nanopore adaptive sampling–based PGx sequencing workflow, which enables direct observation of haplotype structures across multiple variants within an individual DNA molecule. In male samples, which are hemizygous for the X-linked *G6PD* gene, variant calls obtained by adaptive sampling were fully concordant with those from long-range PCR–based Oxford Nanopore sequencing, confirming consistent variant detection and the absence of spurious phasing. In female samples, adaptive sampling enabled direct haplotype resolution, and the phased haplotypes were fully consistent with those inferred from long-range PCR–based long-read sequencing, providing independent validation of phasing accuracy (Fig. 7 and Table S6).

Discussion

Comprehensive characterization of the *G6PD* gene benefits from the analysis of both coding and noncoding regions, as sequence variation outside exons may affect splicing, regulation, or gene expression. Although the functional relevance of many noncoding variants remains to be confirmed, their inclusion offers a more complete view of genetic diversity across the locus. This is particularly important for *G6PD*, an X-linked gene, where accurate phasing is critical for interpreting compound heterozygosity in females. Previous studies have suggested that compound mutations can have additive effects on enzyme activity, leading to more severe deficiency than single variants

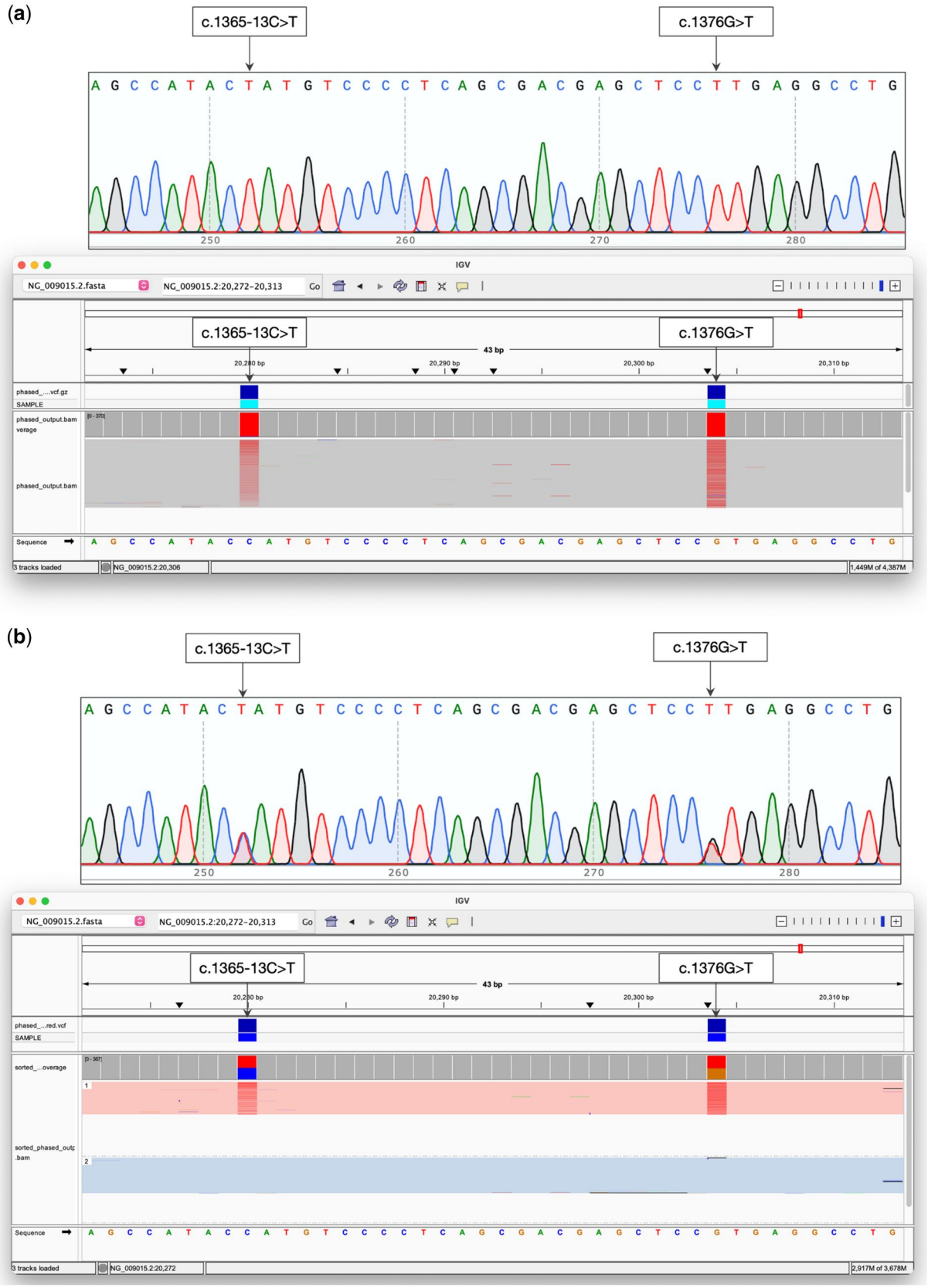

Figure 6 Basecalling and haplotyping comparison between sanger sequencing and the developed method. While both methods show concordant detection of c.1365-13C > T and c.1376G > T (*G6PD* Canton) in (a) a hemizygous male (ID: M1) and (b) a heterozygous female (ID: F17), direct haplotyping of these variants in the female sample was achieved only by the developed method.

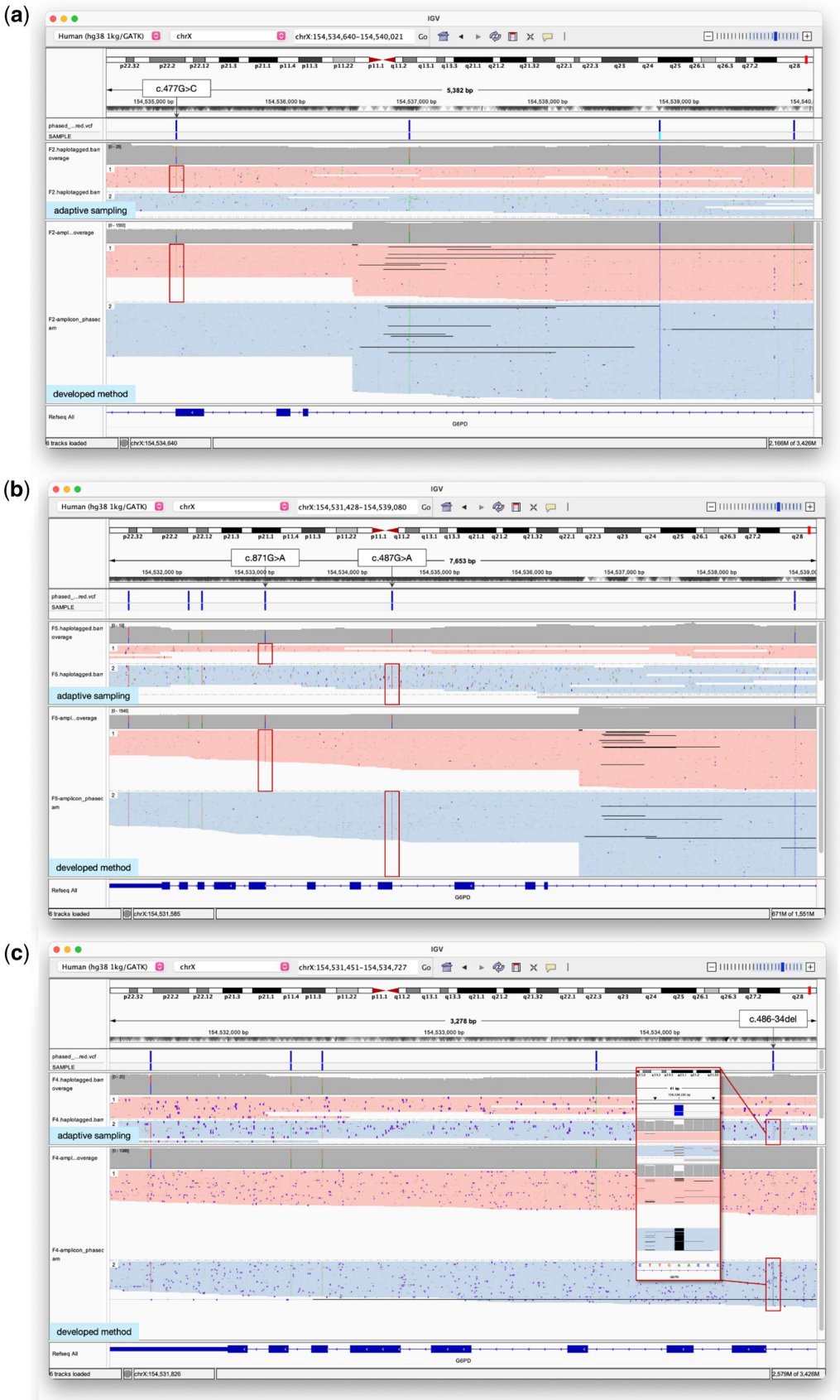

Figure 7 Comparison of aligned and phased reads between adaptive sampling and the developed method in heterozygous female samples. Both methods yielded concordant results, successfully detecting: (a) a single exon variant, c.477G>C, *G6PD* Gond (ID: F2); (b) two exon variants in *trans*, c.487G>A (*G6PD* Mahidol) and c.871G>A (*G6PD* Viangchan) (ID: F5), and (c) an intronic deletion, c.486-34delT (ID: F4).

alone [15, 40]. Distinguishing whether variants are in *cis* or in *trans* is therefore essential for accurate genotype interpretation and clinical risk assessment.

Despite its clinical importance, sequencing the entire *G6PD* gene has remained technically challenging due to its genomic length and the presence of large introns, particularly intron 2. Conventional genotyping methods, such as PCR-based genotyping, Sanger sequencing, or short-read next-generation sequencing, often fail to capture the full spectrum of genetic variation across this locus. Xia *et al.* reported that nearly 20% of pathogenic variants were missed by a targeted genotyping kit designed for 12 common *G6PD* mutations [41]. While exon-based sequencing can identify additional coding variants, it leaves intronic and regulatory regions largely unexplored. These limitations are particularly problematic in female carriers, where short-read data are insufficient to resolve haplotype phase and compound heterozygosity with confidence.

This study is the first to report and validate the application of long-range PCR coupled with Oxford Nanopore long-read sequencing for full-length haplotyping of the *G6PD* gene. Although the individual components of this workflow are well established, reliable haplotype reconstruction required gene-specific experimental optimization. In particular, redesign of the F4 and F5 amplicons to generate a variant-rich overlapping region was critical for accurate phasing across the full gene, an optimization that has not previously been described for *G6PD*. Rather than introducing a new sequencing platform or bioinformatic framework, this work highlights a practical yet nontrivial design consideration for locus-focused long-read haplotyping, especially for genes with population-specific variant architectures and clinically relevant compound heterozygosity. Given the importance of G6PD deficiency in pharmacogenetics and carrier screening, the establishment of a validated full-gene long-read haplotyping strategy represents a meaningful methodological and translational contribution.

This approach overcame the limitations of short-read sequencing by enabling comprehensive variant identification and direct phasing of heterozygous alleles within a single experiment. The workflow is operationally efficient, with PCR (∼7 hours), library preparation (∼1 hour), and sequencing (typically 1 hour to achieve sufficient depth), making it feasible within a standard laboratory schedule. Sequencing metrics further demonstrate the robustness of the approach, with 321 Mb of total output, an N50 read length of 10,948 bp, an average per-base coverage depth of 763-fold (range 35–1648-fold), and a mean *Q*-score of 15.6. The combination of long read lengths and high depth provided reliable consensus accuracy for variant calling and haplotype phasing [22].

Nevertheless, several technical limitations warrant consideration. The mean *Q*-score of 15.6 (corresponding to a 97.2% sequencing accuracy) remains lower than the accuracy typically achieved with short-read sequencing (Q30 or higher) [42], which may reduce confidence in variant calls at sites with low coverage. In addition, coverage depth varied substantially between samples, raising concern that low-depth libraries could fail to capture heterozygous variants and compromise haplotype resolution. Only a fraction of pass reads aligned to the reference, likely contributing to reduced coverage in some cases. These observations underscore the need for future optimization of library preparation, barcode normalization, and demultiplexing strategies to achieve more uniform sequencing depth and ensure reliable clinical implementation.

ONT sequencing is also characterized by higher INDEL error rates, particularly within homopolymeric regions where basecalling remains challenging [25, 43]. In this study, raw error rates were 1.18% for fragments 1–3 and 1.79% for fragments 4–5. Such artifacts can generate false-positive calls or obscure true variants, such as the c.486-34delT intronic deletion common in Thai populations [17, 44], which was also observed in three heterozygous females in this study. To mitigate these effects, basecalling was performed using the SUP model, and more stringent thresholds were applied during Clair3 variant calling. These measures resulted in high concordance between the variants detected in this study and those identified by short-read sequencing and adaptive sampling approaches.

Based on these findings, a minimum per-base coverage depth of 50-fold is recommended for high-confidence *G6PD* genotyping using Clair3, consistent with previous report [45]. In this study, mean coverage substantially exceeded this threshold for both fragments 1–3 (500-fold) and fragments 4–5 (763-fold). However, the wide coverage range observed for fragments 4–5 (35–1648-fold) suggests that coverage variability is not barcode-specific but more likely reflects inconsistencies in DNA extraction or library preparation. These technical factors should be carefully controlled in future applications.

Beyond depth, read length emerged as a critical determinant of haplotype phasing performance. To ensure contiguous phasing across the full *G6PD* locus, the read-length N50 should approximate the target amplicon size. fragments 4–5, with an N50 of 10,948 bp, achieved complete phasing of heterozygous sites and near-complete gene coverage by the longest phase block (99.1%). In contrast, fragments 1–3, with a shorter N50 of 5847 bp, achieved only 71.6% phasing efficiency and a markedly shorter median phase-block length of 1493 bp. While equimolar pooling of PCR amplicons is generally recommended for multiplexing, this strategy must account for the bias of ONT sequencing toward longer and lower-GC amplicons [46]. Adjustments to molar ratios may therefore be necessary to ensure balanced representation of shorter or GC-rich fragments.

Despite the modest sample size, the mutation spectrum observed in this pilot work highlights the substantial genetic heterogeneity of the *G6PD* gene. Because samples were randomly selected from archived diagnostic specimens based on DNA availability and suitability for long-range PCR/ONT sequencing, the cohort was assembled for technical validation and is not population-representative; therefore, observed variant frequencies and haplotype distributions should not be interpreted as population estimates. Across 24 samples, 36 distinct variants were identified, spanning exonic, intronic, UTR, and regulatory regions. Most female samples carried multiple heterozygous variants, frequently involving compound mutations across coding and noncoding sites. Pathogenic missense variants such as Mahidol, Viangchan, Canton, and Valladolid (WHO Class B) were identified, all of which have direct clinical relevance due to their association with drug-induced hemolysis, particularly following exposure to primaquine or tafenoquine [47–49]. In one female sample, long-read sequencing enabled direct phasing of Mahidol and Viangchan variants, demonstrating that they were located in *trans* configuration, consistent with results obtained using adaptive sampling. This finding demonstrates the unique utility of long-read sequencing for resolving compound heterozygosity in females, which cannot be reliably inferred from genotype data alone.

Established linkage disequilibrium patterns frequently reported in Asian populations, including the co-segregation of Viangchan with c.1311T, c.1365 – 13C, and c.*357G, and the association of c.1311T

and c.1365 – 13C with c.*357G, were also detected in this study [16, 18, 20, 38, 39], further supporting the reliability of the developed workflow. Importantly, full-gene analysis also revealed additional haplotypic configurations that require further investigation. A long-range haplotype comprising six recurrent variants (c.120 + 2900G, c.120 + 2955G, c.121 – 2679G, c.1311C, c.1365 – 13T, and c.*357A), was identified together with the fixation of c.120 + 1469A across all chromosomes. Several of these alleles are annotated as variants relative to the reference transcript (NM_001042351.1) but in fact represent predominant alleles in the Thai population. These findings highlight the limitations of reference-based annotation and underscore the importance of population-informed interpretation of *G6PD* variation.

These results emphasize that successful haplotype reconstruction depends not only on sequencing depth and read length, but also on the strategic distribution of informative heterozygous variants within the amplicon overlaps. The initial three-amplicon design (F1, F2, and F3) yielded limited haplotype resolution due to variant-poor overlaps, whereas the redesigned fragments (F4 and F5) spanning both variant-rich and variant-sparse regions markedly improved phasing performance. This suggests a key methodological principle that effective haplotype phasing requires not only long reads but also primer design that is informed by the underlying variant landscape of the target locus.

Finally, the long-read sequencing workflow described here provides a framework for comprehensive *G6PD* analysis, enabling full-gene variant detection and haplotype phasing within a single assay. Future studies incorporating larger and more diverse cohorts will be essential to refine population-specific haplotypes and to benchmark this approach against alternative long-read pharmacogenomic workflows. In addition, ONT's ability to detect native DNA modifications offers opportunities to integrate sequence and methylation data, potentially revealing regulatory mechanisms that influence G6PD expression [50]. More broadly, the methodological insights gained from this work may inform haplotype phasing strategies for other X-linked genes and complex loci where long-range phase information is critical for both research and translational genomics.

Conclusions

In summary, this study validates a long-range PCR and Oxford Nanopore long-read sequencing approach for full-gene haplotyping of *G6PD*. The optimized amplicon design enabled comprehensive variant detection and accurate phasing of heterozygous alleles, particularly in female carriers, overcoming key limitations of short-read methods. This workflow provides a practical framework for resolving clinically relevant *G6PD* haplotypes and supports future applications in pharmacogenetic screening and translational genomics.

Author contributions

Kamonwan Chamchoy (Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Resources [equal], Validation [equal], Visualization [equal], Writing—original draft [equal]), Beatriz A. C. Jacob (Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Validation [equal], Visualization [equal], Writing—original draft [equal]), and Usa Boonyuen (Conceptualization [lead], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [lead], Resources [lead], Supervision [lead], Validation [equal], Visualization [supporting], Writing—original draft [equal], Writing—review & editing [lead])

Supplementary material

Supplementary material is available at *Biology Methods and Protocols* online.

Conflicts of interest

The authors declare no competing interests.

Funding

This work was supported by the Chulabhorn Royal Academy to K.C. and by Mahidol University through the Fundamental Fund, fiscal year 2025, provided by the National Science Research and Innovation Fund (NSRF) to U.B. B.A.C.J. acknowledges support from the Department of Science and Technology–Philippine Council for Health Research and Development (DOST-PCHRD).

Data availability

The data underlying this article are available in the article and in its online supplementary material.

References

1. Cappellini MD, Fiorelli G. Glucose-6-phosphate dehydrogenase deficiency. *Lancet* 2008;**371**:64–74.
2. Luzzatto L, Ally M, Notaro R. Glucose-6-phosphate dehydrogenase deficiency. *Blood* 2020;**136**:1225–40.
3. Howes RE, Piel FB, Patil AP *et al.* G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. *PLoS Med* 2012;**9**:e1001339.
4. Kaplan M, Hammerman C. Glucose-6-phosphate dehydrogenase deficiency: a hidden risk for kernicterus. *Semin Perinatol* 2004;**28**:356–64.
5. Olusanya BO, Emokpae AA, Zamora TG, Slusher TM. Addressing the burden of neonatal hyperbilirubinaemia in countries with significant glucose-6-phosphate dehydrogenase deficiency. *Acta Paediatr* 2014;**103**:1102–9.
6. Domingo GJ, Advani N, Satyagraha AW *et al.* Addressing the gender-knowledge gap in glucose-6-phosphate dehydrogenase deficiency: challenges and opportunities. *Int Health* 2019;**11**:7–14.
7. Roca-Feltrer A, Khim N, Kim S *et al.* Field trial evaluation of the performances of point-of-care tests for screening G6PD deficiency in Cambodia. *PLoS One* 2014;**9**:e116143.
8. Peters AL, Veldthuis M, van Leeuwen K *et al.* Comparison of spectrophotometry, chromate inhibition, and cytofluorometry versus gene sequencing for detection of heterozygously glucose-6-phosphate dehydrogenase-deficient females. *J Histochem Cytochem* 2017;**65**:627–36.
9. Henriques G, Phommasone K, Tripura R *et al.* Comparison of glucose-6 phosphate dehydrogenase status by fluorescent spot test and rapid diagnostic test in Lao PDR and Cambodia. *Malar J* 2018;**17**:243.
10. Halim SA, Bahar R, Abdullah WZ *et al.* Performance comparison between conventional fluorescent spot test and quantitative assay in detecting G6PD deficiency in neonates. *Oman Med J* 2023;**38**:e524.

11. Lyu RY, Chen XW, Zhang M *et al.* [Detection of gene mutation in glucose-6-phosphate dehydrogenase deficiency by RT-PCR sequencing]. *Zhongguo Dang Dai Er Ke Za Zhi* 2016;**18**:630–4.
12. Chen S, Gao J, Wu Q *et al.* Reduction of missed diagnosis of G6PD deficiency in heterozygous females by G6PD/6PGD ratio assay combined with amplification refractory mutation system PCR. *Hum Hered* 2023;**88**:1–7.
13. Laliotis GP, Bizelis I, Rogdakis E. Comparative approach of the de novo fatty acid synthesis (lipogenesis) between ruminant and non ruminant mammalian species: from biochemical level to the main regulatory lipogenic genes. *Curr Genomics* 2010;**11**:168–83.
14. Song J, Sun H, Zhang S, Shan C. The multiple roles of glucose-6-phosphate dehydrogenase in tumorigenesis and cancer chemoresistance. *Life (Basel)* 2022;**12**:271.
15. He Y, Zhang Y, Chen X *et al.* Glucose-6-phosphate dehydrogenase deficiency in the Han Chinese population: molecular characterization and genotype–phenotype association throughout an activity distribution. *Sci Rep* 2020;**10**:17106.
16. Boonyuen U, Jacob BAC, Wongwigkan J *et al.* Genetic analysis and molecular basis of G6PD deficiency among malaria patients in Thailand: implications for safe use of 8-aminoquinolines. *Malar J* 2024;**23**:38.
17. Pengsuk N, Chamchoy K, Duangdee C *et al.* Integrated phenotypic and genotypic approaches for accurate diagnosis of G6PD deficiency: implications for drug safety in Thailand. *Pharmacol Res Perspect* 2025;**13**:e70117.
18. Amini F, Ismail E. 3′-UTR variations and G6PD deficiency. *J Hum Genet* 2013;**58**:189–94.
19. Geck RC, Powell NR, Dunham MJ. Functional interpretation, cataloging, and analysis of 1,341 glucose-6-phosphate dehydrogenase variants. *Am J Hum Genet* 2023;**110**:228–39.
20. Chaowanathikhom M, Nuchnoi P, Palasuwan D. Significance of 3′UTR and pathogenic haplotype in glucose-6-phosphate deficiency. *Lab Med* 2017;**48**:73–88.
21. Oxford Nanopore Technologies. *Short, Long, or Ultra-Long: Which Read Length is Right for You?* https://nanoporetech.com/blog/short-long-or-ultra-long-which-read-length-is-right-for-you (25 October 2025, date last accessed).
22. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet* 2020;**21**:597–614.
23. Boonyuen U, Songdej D, Tanyaratsrisakul S *et al.* Glucose-6-phosphate dehydrogenase mutations in malaria endemic area of Thailand by multiplexed high-resolution melting curve analysis. *Malar J* 2021;**20**:194.
24. Jamshidi J, Rowntree C, Fadaee S *et al.* Long-range PCR and Nanopore sequencing for localisation and phasing variants: an end-to-end clinical application workflow. *BMC Med Genom* 2025;**18**:186.
25. Delahaye C, Nicolas J. Sequencing DNA with nanopores: troubles and biases. *PLoS One* 2021;**16**:e0257521.
26. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100.
27. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 2021;**37**:4572–4.
28. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 2023;**39**:btad311.
29. De Coster W, D'Hert S, Schultz DT *et al.* NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018;**34**:2666–9.
30. Danecek P, Bonfield JK, Liddle J *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* 2021;**10**:giab008.
31. McClinton B, Watson CM, Crinnion LA *et al.* Haplotyping using long-range PCR and Nanopore sequencing to phase variants: lessons learned from the ABCA4 locus. *Lab Invest* 2023;**103**:100160.
32. Liau Y, Maggo S, Miller AL *et al.* Nanopore sequencing of the pharmacogene CYP2D6 allows simultaneous haplotyping and detection of duplications. *Pharmacogenomics* 2019;**20**:1033–47.
33. Martin M, Patterson M, Garg S *et al.* WhatsHap: fast and accurate read-based phasing. *bioRxiv*. 2016:085050.
34. Robinson JT, Thorvaldsdóttir H, Wenger AM *et al.* Variant review with the integrative genomics viewer. *Cancer Res* 2017;**77**:e31–4.
35. Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6.
36. Geck RC, Wheelock MK, Powell RL *et al.* Functional evidence for G6PD variant classification from mutational scanning. *bioRxiv*. 2025.
37. Luzzatto L, Bancone G, Dugué P-A *et al.* New WHO classification of genetic variants causing G6PD deficiency. *Bull World Health Organ* 2024;**102**:615–7.
38. Chamchoy K, Sudsumrit S, Wongwigkan J *et al.* Molecular characterization of G6PD mutations identifies new mutations and a high frequency of intronic variants in Thai females. *PLoS One* 2023;**18**:e0294200.
39. Tantular IS, Kawamoto F. Distribution of G6PD deficiency genotypes among Southeast Asian populations. *Trop Med Health* 2021;**49**:97.
40. Yan T, Cai R, Mo OH *et al.* Incidence and complete molecular characterization of glucose-6-phosphate dehydrogenase deficiency in the Guangxi Zhuang autonomous region of southern China: description of four novel mutations. *Haematologica* 2006;**91**:1321–8.
41. Xia Z, Wang X, Ye H *et al.* Evaluation of strategies for identification of infants with pathogenic glucose-6-phosphate dehydrogenase variants in China. *Front Genet* 2022;**13**:844381.
42. Illumina. *View Run Metrics*. BaseSpace Sequence Hub Help, 2024. https://help.basespace.illumina.com/data/view-data/view-runs/run-metrics (1 November 2025, date last accessed).
43. Hall MB, Wick RR, Judd LM *et al.* Benchmarking reveals superiority of deep learning variant callers on bacterial Nanopore sequence data. *Elife* 2024;**13**:RP98300.
44. Boonyuen U, Jacob BAC, Chamchoy K *et al.* Improved genetic screening with zygosity detection through multiplex high-resolution melting curve analysis and biochemical characterisation for G6PD deficiency. *Trop Med Int Health* 2025;**30**:437–57.
45. Zheng Z, Li S, Su J *et al.* Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci* 2022;**2**:797–803.
46. Whitford W, Hawkins V, Moodley KS *et al.* Proof of concept for multiplex amplicon sequencing for mutation identification using the MinION nanopore sequencer. *Sci Rep* 2022;**12**:8572.
47. Chu CS, Bancone G, Moore KA *et al.* Haemolysis in G6PD heterozygous females treated with primaquine for *Plasmodium vivax* Malaria: a nested cohort in a trial of radical curative regimens. *PLoS Med* 2017;**14**:e1002224.
48. Rueangweerayut R, Bancone G, Harrell EJ *et al.* Hemolytic potential of tafenoquine in female volunteers heterozygous for glucose-6-phosphate dehydrogenase (G6PD) deficiency (G6PD Mahidol variant) versus G6PD-normal volunteers. *Am J Trop Med Hyg* 2017;**97**:702–11.
49. Kheng S, Muth S, Taylor WRJ *et al.* Tolerability and safety of weekly primaquine against relapse of *Plasmodium vivax* in Cambodians with glucose-6-phosphate dehydrogenase deficiency. *BMC Med* 2015;**13**:203.
50. Simpson JT, Workman RE, Zuzarte PC *et al.* Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* 2017;**14**:407–10.

Biology Methods and Protocols, 2026, 11, 1–14
https://doi.org/10.1093/biomethods/bpag008
Methods Article