frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 12 June 2025
DOI 10.3389/fgene.2025.1526049

Check for updates

OPEN ACCESS

EDITED BY
Jing Hua Zhao,
University of Cambridge, United Kingdom

REVIEWED BY
Prashant Kumar Mallick,
National Institute of Malaria Research (ICMR), India
Jonathan Shortt,
University of Colorado Anschutz Medical Campus, United States

*CORRESPONDENCE
Karamoko Niaré,
karamoko_niare@brown.edu
Jeffrey A. Bailey,
jeffrey_bailey@brown.edu

RECEIVED 11 November 2024
ACCEPTED 14 May 2025
PUBLISHED 12 June 2025

CITATION
Niaré K, Crudale R, Fola AA, Wernsman Young N, Asua V, Conrad MD, Gashema P, Ghansah A, Hangi S, Ishengoma DS, Mazarati J-B, Zeleke AJ, Rosenthal PJ, Djimdé AA, Juliano JJ and Bailey JA (2025) Highly multiplexed molecular inversion probe panel in *Plasmodium falciparum* targeting common SNPs approximates whole-genome sequencing assessments for selection and relatedness. *Front. Genet.* 16:1526049. doi: 10.3389/fgene.2025.1526049

COPYRIGHT
© 2025 Niaré, Crudale, Fola, Wernsman Young, Asua, Conrad, Gashema, Ghansah, Hangi, Ishengoma, Mazarati, Zeleke, Rosenthal, Djimdé, Juliano and Bailey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Highly multiplexed molecular inversion probe panel in *Plasmodium falciparum* targeting common SNPs approximates whole-genome sequencing assessments for selection and relatedness

Karamoko Niaré[1,2]*, Rebecca Crudale[1], Abebe A. Fola[1,2], Neeva Wernsman Young[2], Victor Asua[3,4], Melissa D. Conrad[5], Pierre Gashema[6], Anita Ghansah[7], Stan Hangi[8], Deus S. Ishengoma[9,10], Jean-Baptiste Mazarati[6], Ayalew Jejaw Zeleke[11], Philip J. Rosenthal[5], Abdoulaye A. Djimdé[12,13], Jonathan J. Juliano[14,15,16] and Jeffrey A. Bailey[1,2]*

[1]Department of Pathology and Laboratory Medicine, Brown University, Providence, RI, United States, [2]Center for Computational Molecular Biology, Brown University, Providence, RI, United States, [3]Infectious Diseases Research Collaboration, Kampala, Uganda, [4]Institute for Tropical Medicine, University of Tubingen, Tubingen, Germany, [5]Department of Medicine, University of California, San Francisco, San Francisco, CA, United States, [6]Center for Genomic Biology, Institut d'Enseignement Supérieur de Ruhengeri, Ruhengeri, Rwanda, [7]Noguchi Memorial Institute for Medical Research, University of Ghana, Accra, Ghana, [8]Department of Pediatrics, HEAL Africa, Goma, Democratic Republic of Congo, [9]National Institute for Medical Research, Dar es Salaam, Tanzania, [10]Department of Biochemistry, Kampala International University in Tanzania, Dar es Salaam, Tanzania, [11]Department of Medical Parasitology, School of Biomedical and Laboratory Science, University of Gondar, Gondar, Ethiopia, [12]Pathogens genomics Diversity Network Africa, Bamako, Mali, [13]Malaria Research and Training Center, University of Science, Techniques and Technologies of Bamako, Bamako, Mali, [14]Institute for Global Health and Infectious Diseases, University of North Carolina, Chapel Hill, NC, United States, [15]Division of Infectious Diseases, School of Medicine, University of North Carolina, Chapel Hill, NC, United States, [16]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, United States

Introduction: The use of next-generation sequencing technologies (NGS) to study parasite populations and their response and evolution to interventions is important to support malaria control and elimination efforts. While whole-genome sequencing (WGS) is optimal in terms of assessing the entire genome, it is costly for numerous samples. Targeted approaches selectively enriching for the sequence of interest are more affordable and have higher throughput but sometimes lack adequate information content for key analyses.

Methods: We have developed a highly multiplexed molecular inversion probe (MIP) panel (IBC2FULL) targeting 4,264 single-nucleotide polymorphisms (SNPs) with ≥5% minor allele frequency (MAF) in Sub-Saharan African regions from publicly available *Plasmodium falciparum* WGS (n = 3,693). We optimized the panel alone and in combination with antimalarial drug resistance MIPs in laboratory *P. falciparum* strains at different parasitemias and validated it by

sequencing field isolates from the Democratic Republic of Congo, Ethiopia, Ghana, Mali, Rwanda, Tanzania, and Uganda and evaluating the population structure, identity-by-descent (IBD), signals of selection, and complexity of infection (COI).

Results: The new panel IBC2FULL consisted of 2,128 MIPs (containing 4,264 common SNPs) spaced by 5.1–18.4 kb across the entire genome. While these microhaplotypes were developed based on variations from Sub-Saharan African WGS data, 59.3% (2,529) of SNPs were also common in Southeast Asia. The MIPs were balanced to produce more a uniform and higher depth of coverage at low parasitemia (100 parasites/μL) along with MIPs targeting antimalarial drug resistance genes. Comparing targeted regions extracted from public WGS, we observed that IBC2FULL provided a higher resolution of the local population structure in Sub-Saharan Africa than current PCR-based targeted sequencing panels. For sequencing field samples (n = 140), IBC2FULL approximated WGS measures of relatedness, population structure, and COI. Interestingly, genome-wide analysis of extended haplotype homozygosity detected the same major peaks of selection as WGS. We also chose a subset of 305 high-performing MIPs to create a core panel (IBC2CORE) that produced high-quality data for basic population genomic analysis and accurate estimation of COI.

Discussion: IBC2FULL and IBC2CORE panels have been designed to provide an improved platform for malaria genomic epidemiology and biology that can approximate WGS for many applications and is deployable for malaria molecular surveillance in resource-limited settings.

KEYWORDS

malaria, *Plasmodium falciparum*, molecular inversion probe, molecular surveillance, genomic epidemiology, targeted sequencing

Introduction

Malaria remains a significant public health problem with unacceptably high morbidity and mortality rates (World Health Organization, 2023). Despite substantial effort over the past decades, malaria control and elimination interventions are threatened by multiple factors. Currently, drug and diagnostic resistance (Ariey et al., 2014; Ashley et al., 2014; Dondorp et al., 2009; Uwimana et al., 2020; World Health Organization, 2023) threaten longstanding testing and treatment strategies in Sub-Saharan Africa (SSA), where the vast majority of malarial deaths occur, primarily due to *Plasmodium falciparum* (World Health Organization, 2023). Artemisinin partial resistance has emerged in multiple locations in Eastern Africa and is rapidly spreading (Conrad et al., 2023; Fola et al., 2023; Ishengoma et al., 2024; Juliano et al., 2023; Mihreteab et al., 2023; Uwimana et al., 2020; Young et al., 2024). Furthermore, there is a lack of effective vaccines inducing long-acting sterilizing immunity (RTSS Clinical Trials Partnership, 2014; RTSS Clinical Trials Partnership, 2015), insecticide resistance continues to evolve, and the mosquito species *Anopheles stephensi* has spread in SSA from Asia, threatening urban environments (Balkew et al., 2020; Ranson and Lissenden, 2016; Seyfarth et al., 2019). To support malaria elimination efforts by overcoming these issues, multipurpose genomic tools that can be deployed to achieve the molecular surveillance of the disease in a cost-effective manner are currently needed.

The recent progress in next-generation sequencing technologies and analytical computational tools has paved the way for more effective malaria molecular surveillance to detect drug resistance markers, track malaria transmission, determine importation, and evaluate the impacts of interventions (Aydemir et al., 2018; Chang et al., 2017; Hathaway et al., 2018; Neafsey et al., 2021; Ngondi et al., 2017; Paschalidis et al., 2022; Rao et al., 2016; Schaffner et al., 2018; Volkman et al., 2012). Presently, single-nucleotide polymorphisms (SNPs) across the genome are the most robust input for the molecular surveillance of malaria, readily allowing for population genetic measures and statistics as well as directly underlying nonsynonymous changes leading to drug resistance. Beyond direct assessment of drug resistance markers, current population genomic analyses of malaria parasite populations aim mainly but not only to i) understand the origins, spatiotemporal spread and mechanisms of drugs, and diagnostic and vaccine resistances; ii) monitor parasite connectivity and importation into regions targeted for malaria elimination; and iii) estimate the complexity of infection (COI) to evaluate the transmission intensity. To power these analyses, large numbers of variant sites from across the genome are optimal. Targeted sequencing techniques, which consist of sequencing the genome at specific loci, are generally inexpensive tools that can be used for large-scale molecular surveillance of malaria with numerous samples. Unfortunately, increasing the number of targets to be amplified together in a single tube for more robust analysis is currently a major challenge for methods such as multiplex PCR-based assays (Campino et al., 2011; LaVerriere et al., 2022; Tessema et al., 2022). As a result, existing affordable targeted sequencing tools cover a small proportion of variations across the genome and cannot approach the whole-genome

sequencing (WGS) level of genetic information that is useful for detecting selection and detailing identity-by-descent (IBD).

Molecular inversion probe (MIP) capture (Cantsilieris et al., 2017) coupled with next-generation deep sequencing was first applied in 2018 (Aydemir et al., 2018) and has now been used for malaria molecular surveillance, particularly for large-scale countrywide surveys of drug resistance and population structure (Conrad et al., 2023; Juliano et al., 2023; Moser et al., 2020; Verity et al., 2020; Young et al., 2024). This method is based on the utilization of short oligonucleotide probes to hybridize to and capture as single-stranded circular regions of the genome (Cantsilieris et al., 2017). The extension and ligation arms within an MIP (analogous to PCR primers) are homologous to sequences flanking the specific target. Unique molecular identifiers (UMIs) allow for error correction and accurate capture counts. After hybridization of the probes to the template, the targeted sequence between the probe arms is captured by polymerase extension, followed by end ligation. The closed circular DNA formed in this step is preserved during exonuclease digestion and is thereby enriched. Later, PCR amplification adds sample barcodes to create a sequenceable Illumina library. MIPs represent an ultra-multiplexed targeted sequencing technique that can be used to sequence thousands of loci in hundreds to thousands of samples on a single Illumina run. Beyond SNPs, MIPs also provide microhaplotypes that are more informative for relatedness and population structure analysis (da Silva et al., 2023; Tessema et al., 2022).

We sought to develop a new MIP panel that could better approximate the information content of WGS. This new panel (IBC2FULL) captures microhaplotype information for 4,264 common SNPs from SSA that are densely distributed across the 14 chromosomes. We optimized the MIP panel and tested the resulting SNPs to analyze the IBD, population structure, COI, and selection metrics in field samples in SSA. We obtained an improved information content with estimates of the population structure compared to existing PCR-based targeted sequencing tools as well as measures of IBD and selection signatures comparable to WGS measures.

Materials and methods

Identification of informative SNPs across the genome for probe design

To select SNPs for the design of our MIP panel, we harnessed WGS read data from the Pf6 release (MalariaGEN et al., 2021) that was reanalyzed using an optimized GATK4 (v4.3.0.0) variant calling pipeline (Niaré et al., 2023). We selected SNPs detected from all field isolates collected in SSA that were separated into West Africa, which had the majority of samples (N = 2,164), and other African regions (N = 1,529). Other regions included East Africa (N = 539), Malawi (N = 349), Central West Africa (N = 275), and Central Africa (N = 366). For the two groups, BCFtools was used to retain SNPs with variant quality score log-odds >0 and minor allele frequency (MAF) ≥5% within the core regions (excluding subtelomeric and hypervariable regions) (Miles et al., 2016) of the 14 chromosomes after removing loci and samples with >10% and 20% missing genotypes, respectively, and pruning based on linkage disequilibrium (LD) score $r^2 \geq 0.20$.

Panel design, optimization, and data analysis pipeline

We used MIPTools (https://github.com/bailey-lab/MIPTools, version v0.4.0) to design probes to capture the targeted SNPs as previously described (Aydemir et al., 2018). Probes have two hybridization arms that are homologous to the flanks of the target and a nucleotide backbone linking them that includes priming sites for universal amplification to add sample indexes. The designed capture regions measure ~100 nucleotides in length. The MalariaGEN Pf3k variant dataset (MalariaGEN et al., 2021) was provided to the algorithm to exclude probes that were sitting on SNPs and indels with MAF ≥5%. Paralogous regions with ≥98% sequence identity were excluded from the design, as well as sequences with ≥98% identity with any region of the human genome to prevent nonspecific capture.

The entire panel design, also called Infectious Disease Epidemiology and Ecology Lab Big SNP barCode 2 Full (termed IBC2FULL for short), was tested by pooling equal volumes of all MIPs (Figure 1A). After phosphorylation, this panel was used to capture laboratory strains (7G8, DD2, and HB3) in which human blood was spiked in before Chelex extraction. Different probe pool concentrations (1 μM, 2 μM, 4 μM, and 8 μM) and parasitemias (100 parasites/μL, 1,000 parasites/μL, 4,000 parasites/μL, and 10,000 parasites/μL) were used in this optimization. The MIP captures were transformed into a sequenceable library by PCR using Illumina specific primers inserting sample-specific barcodes. The NextSeq 500 platform (Illumina, San Diego, CA, United States) was used at 2 × 150 base pair read length to sequence libraries for both lab strains and field samples at Brown University. MIPwrangler within the MIPTools suite built in the Jupyter Notebook (https://github.com/bailey-lab/MIPTools) was used for read processing and initial UMI-based PCR error correction and microhaplotype evaluation (Hathaway et al., 2018). The final processed microhaplotype data were formatted as FASTQ files and mapped onto the reference genome using BWA, followed by variant calling into VCF using FreeBayes (v1.3.5) (Garrison and Marth, 2012) from within MIPTools (https://github.com/Karaniare/IBC2FULL_miptools_analysis).

The sum rank of per probe UMI counts across samples at 1,000 parasites/uL parasitemia was used to estimate probe performance, which allowed us to calculate the new volumes of probes that are to be pooled to create a final rebalanced panel. Probes that failed to provide at least 10X UMI depth per sample were not added in this final optimized panel.

For IBC2CORE inclusion, we selected a subset of 305 high-performing probes spaced on average at 100 kb across the genome.

Source of clinical samples used for MIP panel validation

We used archived dried blood spot and whole blood samples collected between 2020 and 2023 by seven collaborative studies in

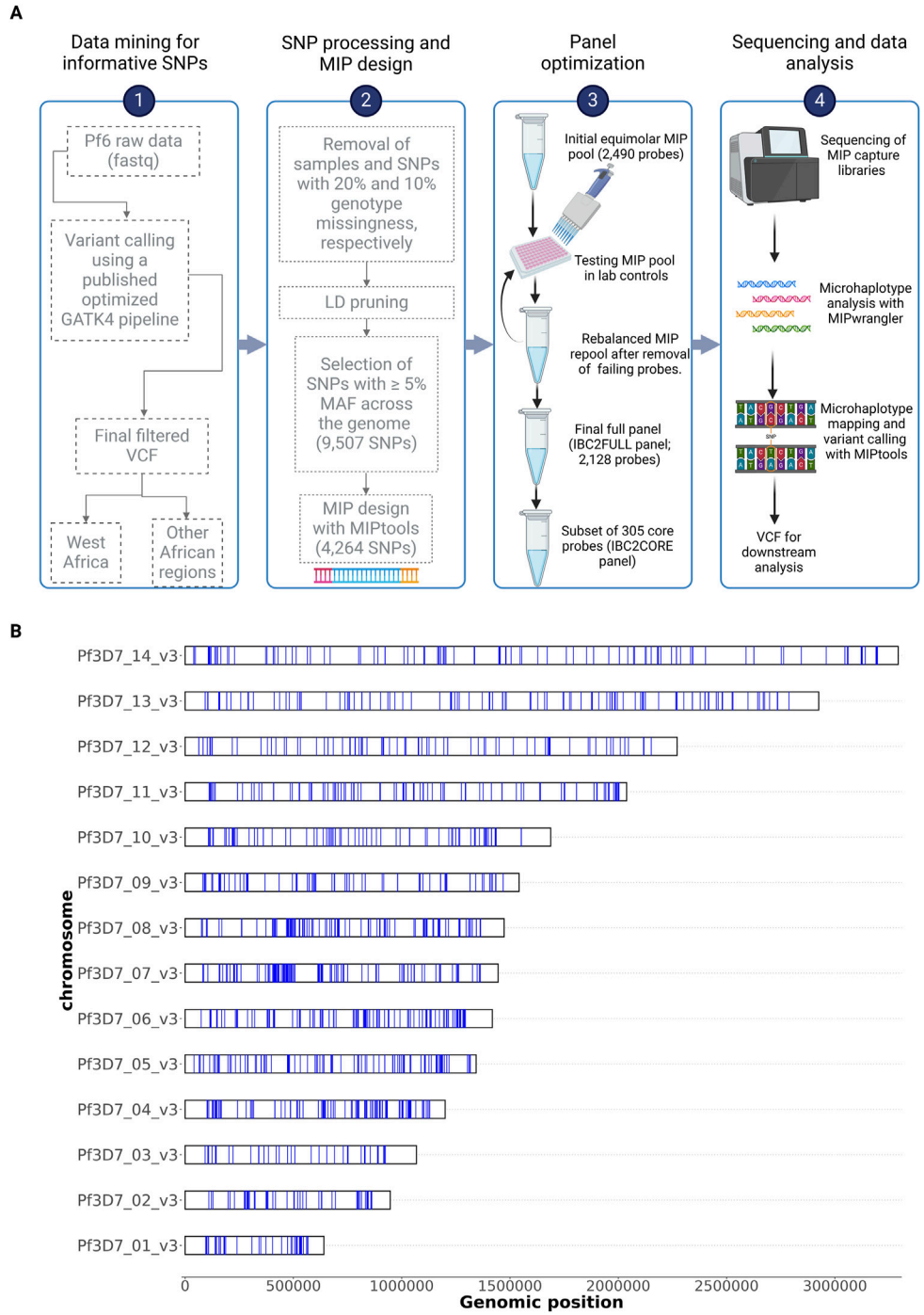


FIGURE 1
IBC2FULL panel design. **(A)** Summary of the IBC2FULL panel development pipeline. **(B)** Genome-wide distribution of microhaplotype loci targeted by the IBC2FULL panel. Each blue band represents a microhaplotype captured by IBC2FULL. LD, linkage disequilibrium; MAF, minor allele frequency; VCF, variant call format; Pf6, MalariaGEN's Pf6 release (MalariaGEN et al., 2021).

SSA, namely, the Democratic Republic of Congo (DRC), Ethiopia, Ghana, Mali, Rwanda, Tanzania, and Uganda. Ethical approval was obtained by each participating study with agreements for further use of the samples. Twenty high-quality samples lacking validated or candidate artemisinin partial resistance mutations were selected by country and subject to DNA extraction using Tween–Chelex or MagBead extraction.

Population genomic analyses

Microhaplotype heterozygosity analysis

MIPs provide an accurate phasing between SNPs located within the same captured region. Genomic loci captured by the MIPs were considered microhaplotypes, and the total numbers of their unique sequences in our various datasets were estimated using custom

scripts. All SNP positions with missing genotypes were removed before concatenating nucleotides from polymorphic sites to construct haplotype sequences. Microhaplotype heterozygosity was computed using the formula $1-\sum_{i=1}^{k}\frac{p_i^2}{N^2}$, where p_i is the individual microhaplotype count and N is the sum of all microhaplotype counts found at a given capture location.

Estimation of complexity of infection using MIP data

For VCFs generated from IBC2FULL, we only used targeted SNP positions for COI calculation. For WGS VCF, we used SNPs with MAF ≥5% that remained after variant recalibration and removal of sites and samples with ≥10% and 20% of genotype missingness. *THE REAL McCOIL* package (v2.0) (Chang et al., 2017) was used to compute COI. The total numbers of Markov chain Monte Carlo and burn-in iterations were set to 2,000 and 500, respectively.

Population structure analysis

To analyze the population structure, we pruned targeted SNP data based on minor allele frequency ≥1% (Amambua-Ngwa et al., 2023), LD < 0.20, and sample missingness rate <20%. PLINK (v1.90b6.27) was used to compute the pairwise variance–standardized genetic relationship matrix. We used this matrix to perform and visualize principal component analysis using R packages PCA and factoextra, respectively. For validation, IBC2FULL was compared with MAD^{4}HatTeR and SpotMalaria panels using their targeted loci extracted from Pf6 data.

Identity-by-descent analysis

Filtered MIP and WGS SNP data were used for pairwise IBD analysis using hmmIBD (v3.0) (Schaffner et al., 2018) with a maximum number of fit iterations of 5 and assuming 0.1% genotype error rate. Sample pairs with at least 60 informative sites were selected for clustering and visualization using the igraph R package.

Analysis of selection signatures

The *rehh* R package was used to compute the allele-specific integrated haplotype homozygosity score (iHS) across the entire core genome using filtered VCFs from IBC2FULL and WGS. To test whether an allele is under selection with a significant increase in the iHS, the *P* value was calculated using formula $P = -\log_{10}(2\Phi(-|iHS|))$, where Φ (x) is the normal distribution function (Gautier and Vitalis, 2012).

Results

IBC2FULL is an optimized panel approximating the information content of the whole genome

After a stepwise optimization (Supplementary Methods, Supplementary Tables S1, S2; Supplementary Figures S1, S2), we successfully developed a large barcode panel (IBC2FULL) of 2,128 MIPs for 4,264 SNPs representing 44.9% of common (≥5% MAF) SNPs found in SSA across the entire genome excluding *var* genes and hypervariable regions (Miles et al., 2016) (Figures 1A,B). However, 59.3% (2,529) of these informative SNPs were also found in Southeast Asian parasite populations. The panel targeted diverse regions with a mean microhaplotype heterozygosity score of 0.4 and 95% CI: 0.39–0.41 (Figure 2A).

Read depth improved significantly with parasitemia (Figure 2B), but even lower-parasitemia samples performed well with more than 10 UMIs per MIP on average at 100 parasites/μL. Despite this performance, there was an undesirable variability in the read depth across probes (standard deviation of per probe UMI count = 441), which was effectively addressed by re-pooling rebalanced volumes of individual MIPs (standard deviation = 77; Figure 2C). To ensure that IBC2FULL can be combined with other MIP panels without affecting the performance of both, we spiked in the panel capturing antimalarial drug resistance genes that was previously used in field samples from Africa (Aydemir et al., 2018) and captured lab controls at 1,000 parasites/μL. After sequencing, no difference was found in the read depth between IBC2FULL with and without the 11 genes tested, with the median UMI count >100 (Figure 2D). To provide an even more targeted cost-effective panel (Supplementary Table S3), we selected 305 MIPs corresponding to 1–2 top performing probes based on UMI coverage in every 100-kb window across the genome. This panel targeted 976 informative SNPs with high microhaplotype heterozygosity (0.435 on average), which represented 22.9% of the IBC2FULL loci and 10.3% of the genome.

SNPs captured by IBC2FULL display high-resolution population structure in Sub-Saharan Africa

To evaluate the information content of IBC2FULL and IBC2CORE in terms of resolving local population structures, we performed principal component analysis with their targeted SNPs extracted from the Pf6 WGS from Sub-Saharan African samples and compared them to the full WGS core-genome SNP complement as well as two current targeted sequencing tools MAD^{4}HatTeR (Aranda-Díaz et al., 2024) and SpotMalaria (Jacob et al., 2021). The first dimension explained up to 37.2% of variation between isolates and showed separation of West (Benin, Burkina Faso, Ghana, The Gambia, Guinea, Ivory Coast, Mali, Mauritania, and Senegal) and Central West (Cameroon and Nigeria) African parasites from the rest of the populations with all tools, except SpotMalaria. However, IBC2FULL displayed better discrimination resolution than other targeted tools and approximated the full SNP complement of WGS (Figures 3A–E). Together, dimensions 1 and 2 further discriminated Central Africa (DRC) from East Africa (Ethiopia, Kenya, Madagascar, Tanzania, and Uganda) and Southeast Africa (Malawi) with IBC2FULL and WGS only. West African and Central West African parasites were also closely related, but WGS and IBC2FULL showed the greatest population definition, followed by MAD^{4}HatTeR and IBC2CORE.

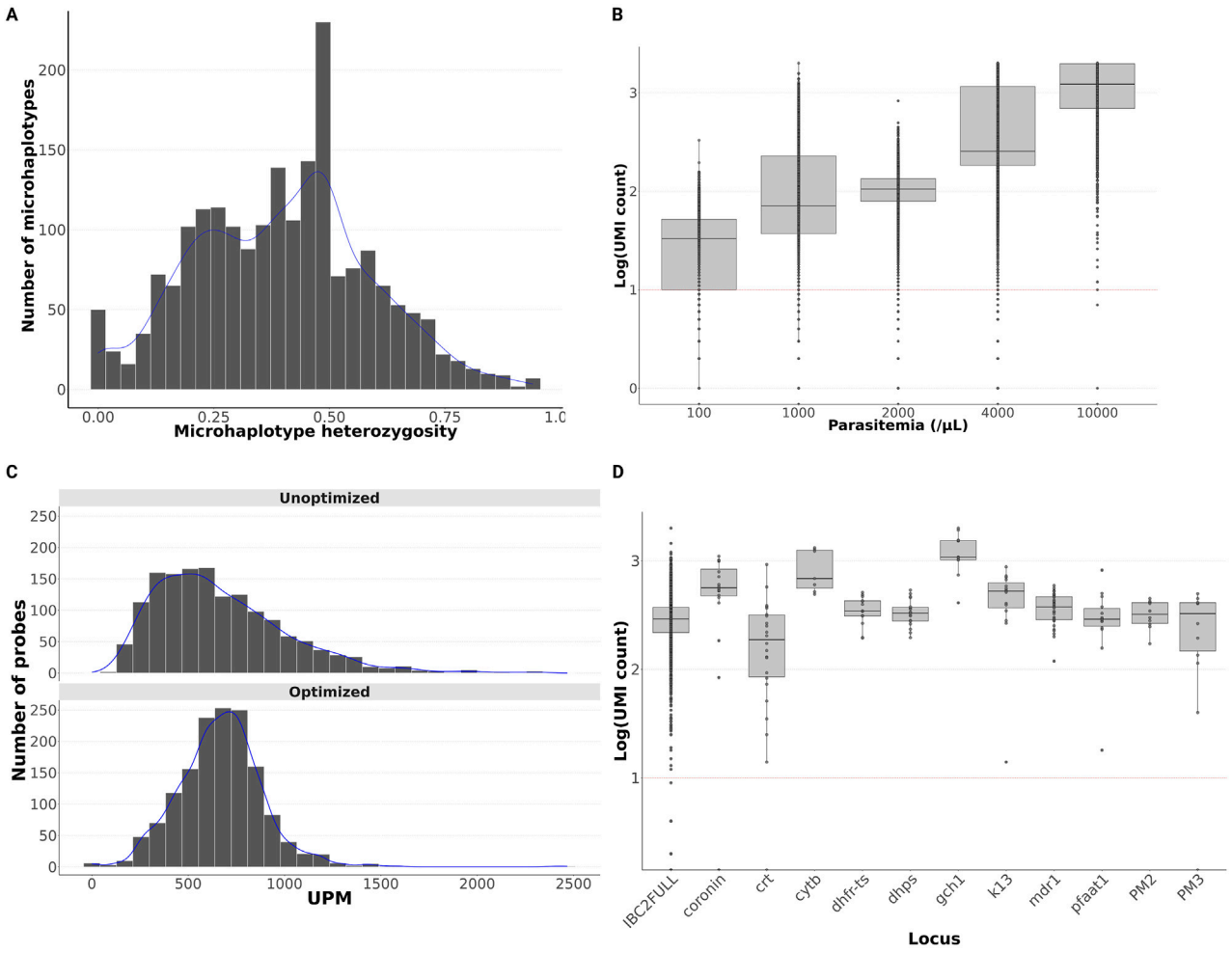


FIGURE 2
IBC2FULL panel optimization. **(A)** Distribution of microhaplotype heterozygosity by probe. Heterozygosity was calculated from the Pf6 dataset in African field samples. **(B)** Sequencing depth variation by parasitemia represented by $\log_{10}$ transformation of the unique molecular index (UMI) count across samples. Each dot indicates the UMI count of one probe. **(C)** UMI depth distribution of the unfiltered panel containing 2,490 probes after one and two NextSeq runs, and one run of the filtered and rebalanced panel with 2,128 probes remaining. Second NextSeq was performed to verify if the low read depth of some probes is due to their intrinsic performance rather than insufficient sequencing. After this additional sequencing, 362 probes still showed low read depth and were removed from the final panel. UMI depth normalized per million (UPM). **(D)** IBC2FULL and main antimalarial drug resistance genes showing similar sequencing depth (log10 transformation of UMI) after spiking with drug resistance MIPs. The test was conducted in lab controls at 1,000 parasites/µL.

IBC2FULL and whole-genome sequencing equally detect relatedness and strong signatures of positive selection

To assess the ability of our IBC2FULL to accurately determine parasite relatedness, we used it to sequence field samples (n = 52) collected in Tanzania in 2022, for which WGS data were available (Ishengoma et al., 2024), and performed pairwise IBD analysis. We found a significant correlation (R = 0.81, $p < 10^{-16}$) with the IBD sharing scores between IBC2FULL and WGS (Figure 4A), and their direct comparison showed no statistical difference (Figure 4B), although the latter is based on the comprehensive analysis of informative sites with ≥5% MAF. After classifying IBD sharing into short intervals of 0.1, there was a similar distribution of sample pairs across them with both tools, with the majority of pairs below 0.1 IBD (Figure 4C). Additionally, there was a strong concordance of both methods in clustering field samples at IBD >0.5 (Figure 4D).

Because of the high SNP density of IBC2FULL, we were able to scan the entire genome for signatures of positive selection using the same data generated with field isolates from Tanzania. To test the null hypothesis of no selection, we estimated the *p*-values around the iHS based on the Gaussian cumulative distribution for both WGS and IBC2FULL panels. The estimated SNP-level iHS score was equally normally distributed for both tools (Supplementary Figure S3), suggesting that IBC2FULL can be used to detect strong signatures of positive selection accurately. Interestingly, IBC2FULL was able to detect all the strongest signals of selection ($p < 10^{-4}$) found with WGS, including *ama1, trap,* PF3D7_0711500, and PF3D7_1035100 (Figure 4E). Overall, we detected 24.3% of the WGS signals of selective sweep at *p*-value < 0.05, including mainly weaker selection (Supplementary Table S4).

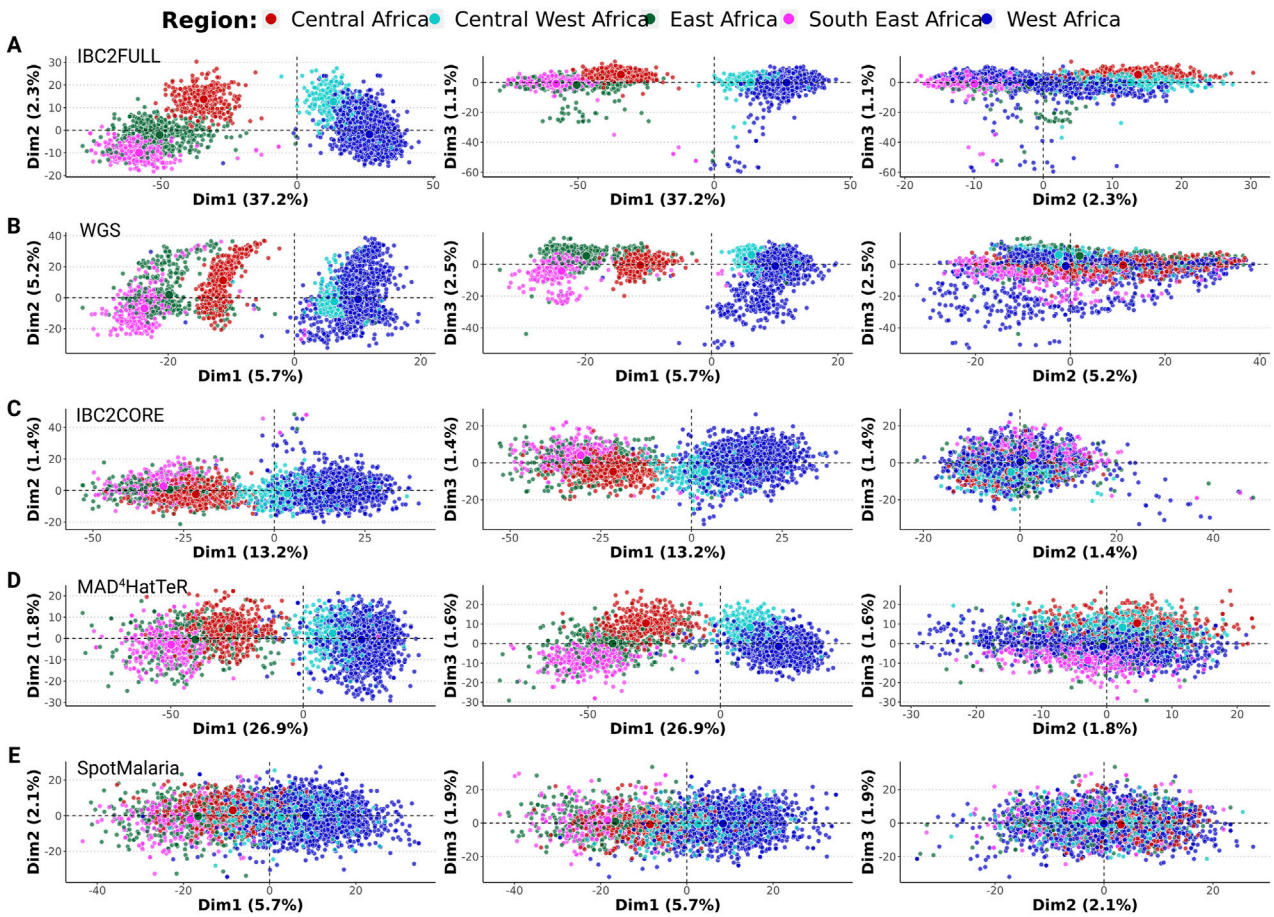


FIGURE 3
Local population structure in Sub-Saharan Africa based on SNPs extracted from the Pf6 dataset. First three dimensions were displayed for **(A)** IBC2FULL; **(B)** whole-genome sequencing (WGS) after selecting SNPs with ≥1% minor allele frequency; **(C)** IBC2CORE; **(D)** MAD^{4}HatTeR; **(E)** SpotMalaria panels (n = 3,693). The first three dimensions (Dim 1, 2, and 3) of the PCA with the proportions of variance explained are shown. Each dot represents a sample colored by region defined in Sub-Saharan Africa.

IBC2FULL and IBC2CORE are effective tools for measuring COI

We measured a strong correlation in the estimated COI between IBC2CORE and WGS and IBC2FULL (Figure 5A, *R = 0.94* versus WGS, *R = 0.97* versus IBC2FULL). IBC2FULL and WGS were also correlated (*R = 0.91*). For further validation, we conducted MIP sequencing on recent archived field samples (n = 140) from high-transmission sites in West Africa (Ghana and Mali) and Central Africa (DRC) as well as lower-transmission settings in East Africa (Ethiopia, Rwanda, Tanzania, and Uganda) to assess the COI. The prevalence of polygenomic infections ranged between 0% in Rwanda and 90% in Mali and was in keeping with the intensity of malaria transmission in these areas (Figure 5B). Having theoretically demonstrated that IBC2FULL loci clustered parasite populations by regions defined in SSA with the same resolution as WGS using the Pf6 dataset (Figure 3), we confirmed these findings with our actual MIP data from recent samples (Figures 5C,D). Interestingly, both IBC2FULL and IBC2CORE displayed high-resolution population structure by country based on the first three principal components combined (Figures 5C,D).

Discussion

While WGS allows for comprehensive analysis of variants across the genome, only a subset of these are commonly used after filtering for quality and informativeness, especially those with high frequency in the circulating parasite populations. We successfully developed a new targeted sequencing panel IBC2FULL that incorporates ~50% of the most common SNPs in African parasites. Comparison with WGS showed that we obtained the same levels of information for key population genomic metrics that are employed to inform malaria control and elimination efforts. Interestingly, we approximated the WGS in detecting strong signals of positive selection across the genome, which has not been previously evaluated for targeted panels.

The potential contributions of these new panels to malaria control and elimination and for studying the biology of the parasites are numerous. IBC2FULL could refine measures of parasite connectivity within countries and across borders, thus aiding in tracking the origins and evolution of drug or diagnosis resistances in regions where malaria parasite populations show minimal local population structure, such as much of SSA

10.3389/fgene.2025.1526049

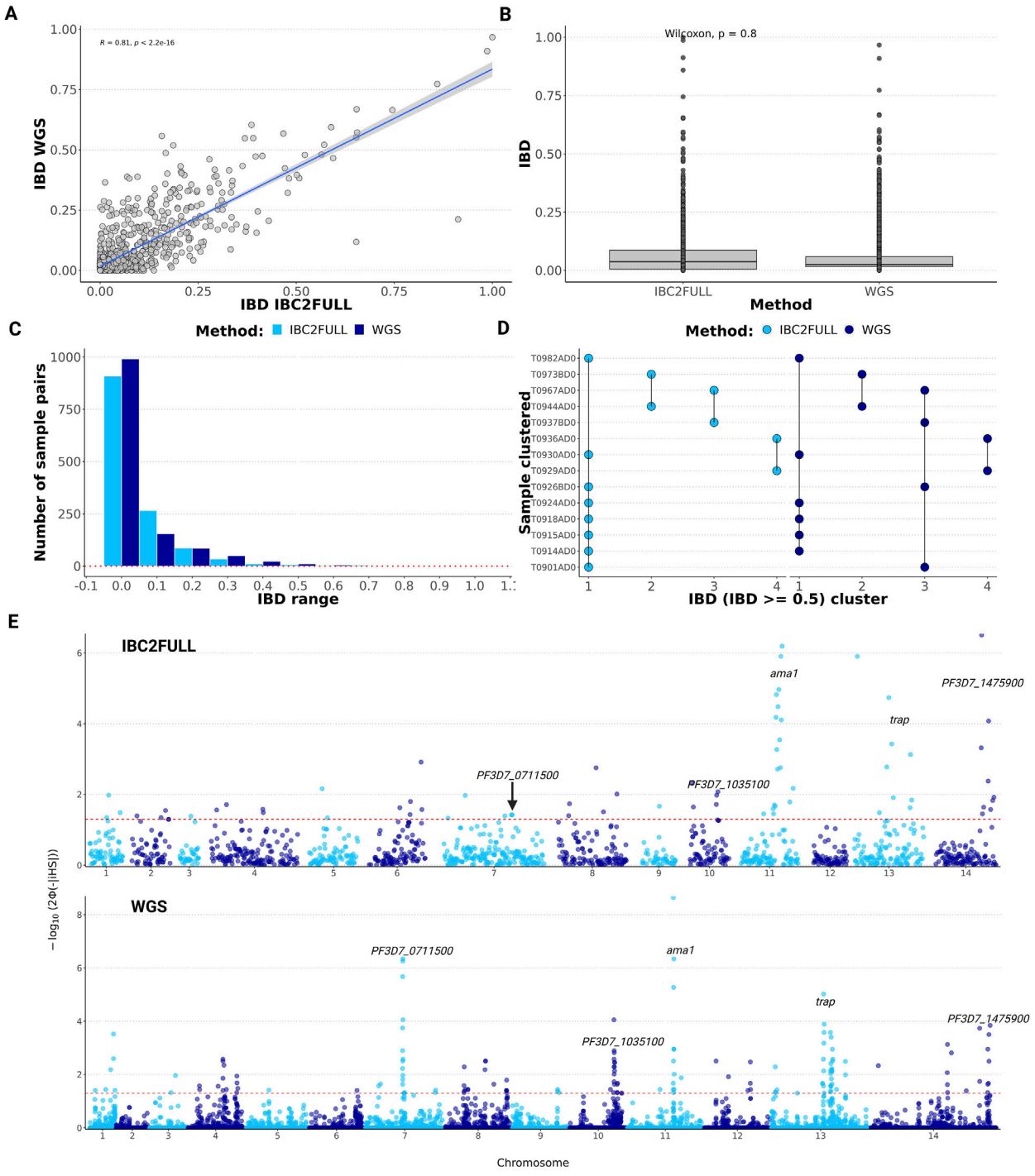


FIGURE 4
Measures of relatedness and signatures of positive selection across the genome in recent field isolates from Tanzania. **(A)** Correlation of pairwise identity-by-descent (IBD) sharing between IBC2FULL and whole-genome sequencing (WGS) (n = 52). **(B)** Direct comparison of IBC2FULL and WGS pairwise IBD. **(C)** Histogram showing sample pair distribution by IBD range for IBC2FULL versus WGS. **(D)** Concordance between IBD-based sample clustering by IBC2FULL and WGS data. Four clusters were found in total by each of the tools at IBD ≥0.5. Each dot represents a sample assigned to a cluster name on the x-axis, and the vertical lines link samples that belong to the same cluster. **(E)** Manhattan plot showing *p*-values (formula is shown on the y-axis where Φ represents the Gaussian distribution function) of the integrated haplotype homozygosity score (iHS) across the 14 chromosomes of *Plasmodium falciparum* for IBC2FULL versus WGS (n = 52). Major signals detected by both tools were similar and are displayed, including *ama1*, *trap*, PF3D7_0711500 (putative regulator of chromosome condensation), PF3D7_1035100 (unknown function), and PF3D7_1475900 (KELT protein). Each dot represents an SNP.

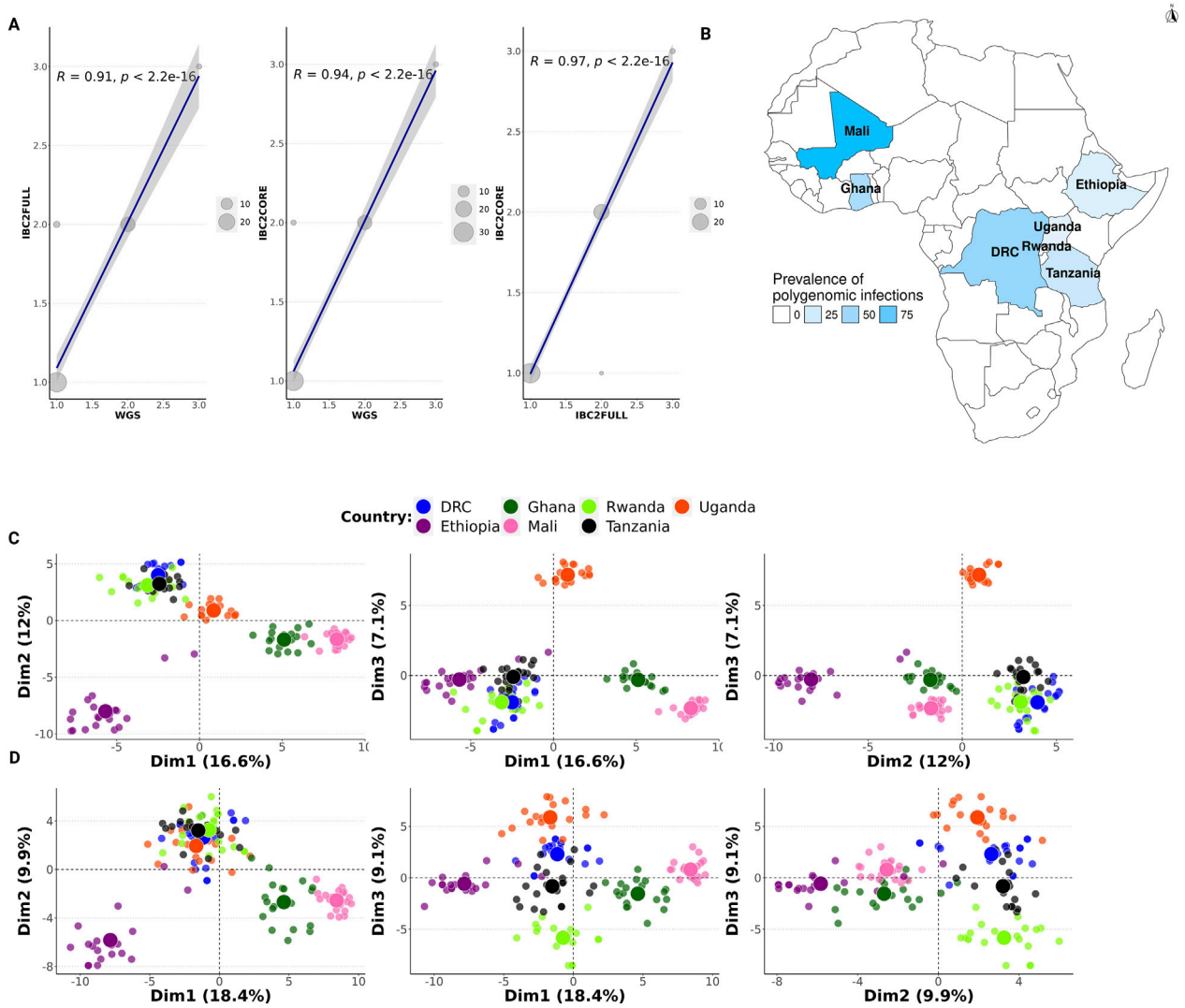


FIGURE 5
Complexity of infection and population structure in Africa based on molecular inversion probe (MIP) sequencing. **(A)** Comparison of the complexity of infection (COI) between IBC2FULL, IBC2CORE, and WGS. Both x- and y-axes show COI values but for different panels. Dots represent field isolates from Tanzania (n = 52), and the dot size is proportional to the number of samples with the same COI. **(B)** Prevalence of polygenomic infections (n = 140) in African countries studied using IBC2CORE. **(C** and **D)** Population structure in Africa based on MIP sequencing using recent field samples (n = 140) for IBC2FULL and IBC2CORE, respectively. First three dimensions are shown. Each dot represents a sample colored by country.

(MalariaGEN et al., 2021; Manske et al., 2012). Specifically, the high-resolution data generated with our panel could serve to resolve fundamental questions regarding the local emergence, importation, and clonal expansion of resistant parasites. It can also be used to monitor any parasite importation or disease outbreak in regions at the pre-elimination stage with very low prevalence of malaria. To assess the impact of control interventions, both IBC2FULL and its subset IBC2CORE, could be deployed to measure the transmission intensity over time and across zones through the evaluation of the prevalence of polygenomic infections. With the possibility of scanning IBD and haplotype sharing across the genome, IBC2FULL can be leveraged to examine the extent and speed of the selective sweep due to drug resistance by analyzing SNPs that are located in the vicinity of molecular markers. Furthermore, selected WGS sequencing could be leveraged in combination with IBC2FULL to provide more powerful and defined resolution of selection.

Our findings showed that IBC2FULL provides a higher resolution in population structure compared to existing amplicon sequencing tools such as SpotMalaria and the microhaplotype panel MAD^{4}HatTeR (Aranda-Díaz et al., 2024; Jacob et al., 2021). The improvement in resolution with IBC2FULL could be explained by the multitude of loci with increased microhaplotype heterozygosity covered, its SNP density across the genome, its representativity of the geography, and good data quality with error correction. Our panels were validated using recent archived field samples from different regions in SSA, and the relatedness and COI reflect the geographical locations and transmission intensities of the sampling sites, respectively (World Health Organization, 2023). With IBC2FULL, we reproduced the same results obtained from WGS in

performing genome-wide scanning for signatures of positive selection, and some of our hits with high iHS signals such as *ama1* and *trap* genes have been previously reported in Africa (Amambua-Ngwa et al., 2012; Coulibaly et al., 2022; Mu et al., 2010). Currently, the IBC2FULL panel requires only the Illumina NextSeq platform, and its total cost per sample was 20 times cheaper than that of WGS coupled with selective whole-genome amplification using Illumina NovaSeq X Plus (Supplementary Table S3). However, there is a minimum upfront investment of $19,675 in purchasing 10 μM of probes, but this can be used to process 336,900 samples costing $3 each, including capture, library preparation, and sequencing. For obligatory upfront investment, the IBC2CORE panel costs only $2,820 and can be used to sequence 36,872 samples at $1.54 per sample.

Our full panel could be an effective alternative to WGS (Neafsey et al., 2008; Van Tyne et al., 2011) in terms of cost-effectiveness and throughput. Whereas WGS still remains mainly for fundamental research with a limited number of samples, the MIP technique has already been applied to monitor the spread of drug resistance in the context of molecular surveillance using thousands of field isolates (Juliano et al., 2023; Moser et al., 2020; Verity et al., 2020; Young et al., 2024). Our new panel is similar to these previous MIP sequencing panels in terms of laboratory protocols and computational data analysis and can be combined with them to analyze the prevalence of drug resistance markers and the genetic background of resistant parasites simultaneously to address multiple questions of interest. We demonstrated this by spiking independently designed probes for drug resistance (Aydemir et al., 2018) into IBC2FULL to extend the application of the panel without affecting the overall sequencing depth. As perspectives, the new development would be dedicated to increasing the resolution of EHH around drug resistance markers by densifying the panel with probes covering the main drug resistance genes and flanking regions.

As a general limitation, MIP sequencing is currently considered less sensitive compared to PCR amplicon sequencing methods (Aranda-Díaz et al., 2024; Aydemir et al., 2018; LaVerriere et al., 2022; Sadler et al., 2024; Tessema et al., 2022), which could hinder the use of our panels in low-parasitemia samples below 100 parasites/μL. Currently, to mitigate this issue, one of the best approaches is to perform selective whole-genome amplification in low-parasitemia samples prior to running IBC2FULL.

In conclusion, we present a new cost-effective targeted sequencing tool that approximates WGS for many genomic analyses and fits well for malaria molecular surveillance using numerous field samples even in resource-limited settings.

Data availability statement

Publicly available datasets were analyzed in this study. These data can be found here: https://malariagen.github.io/parasite-data/pf7/Data_access.html#.

Author contributions

KN: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing. RC: investigation, Writing – review and editing. AF: Resources, Writing – review and editing. NW: Resources, Writing – review and editing. VA: Resources, Writing – review and editing. MC: Resources, Writing – review and editing. PG: Resources, Writing – review and editing. AG: Resources, Writing – review and editing. SH: Resources, Writing – review and editing. DI: Resources, Writing – review and editing. J-BM: Resources, Writing – review and editing. AZ: Resources, Writing – review and editing. PR: Resources, Writing – review and editing. AD: Resources, Writing – review and editing. JJJ: Resources, Supervision, Validation, Writing – review and editing. JAB: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by Brown Institutional Research Funding and NIH R01AI137410, R01AI139520, and R01AI139179.

Acknowledgments

The authors would like to thank all researchers and participants involved in the studies sharing data or samples.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2025.1526049/full#supplementary-material

References

Amambua-Ngwa, A., Button-Simons, K. A., Li, X., Kumar, S., Brenneman, K. V., Ferrari, M., et al. (2023). Chloroquine resistance evolution in Plasmodium falciparum is mediated by the putative amino acid transporter AAT1. *Nat. Microbiol.* 8 (7), 1213–1226. doi:10.1038/s41564-023-01377-z

Amambua-Ngwa, A., Park, D. J., Volkman, S. K., Barnes, K. G., Bei, A. K., Lukens, A. K., et al. (2012). SNP genotyping identifies new signatures of selection in a deep sample of West African Plasmodium falciparum malaria parasites. *Mol. Biol. Evol.* 29 (11), 3249–3253. doi:10.1093/molbev/mss151

Aranda-Díaz, A., Vickers, E. N., Murie, K., Palmer, B., Hathaway, N., Gerlovina, I., et al. (2024). *Sensitive and modular amplicon sequencing of Plasmodium falciparum diversity and resistance for research and public health*. bioRxiv.org: The Preprint Server for Biology. doi:10.1101/2024.08.22.609145

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.-C., Khim, N., et al. (2014). A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature* 505 (7481), 50–55. doi:10.1038/nature12876

Ashley, E. a., Dhorda, M., Fairhurst, R. M., Amaratunga, C., Lim, P., Suon, S., et al. (2014). Spread of artemisinin resistance in Plasmodium falciparum malaria. *N. Engl. J. Med.* 371 (5), 411–423. doi:10.1056/NEJMoa1314981

Aydemir, O., Janko, M., Hathaway, N. J., Verity, R., Mwandagalirwa, M. K., Tshefu, A. K., et al. (2018). Drug-resistance and population structure of Plasmodium falciparum across the democratic republic of Congo using high-throughput molecular inversion probes. *J. Infect. Dis.* 218 (6), 946–955. doi:10.1093/infdis/jiy223

Balkew, M., Mumba, P., Dengela, D., Yohannes, G., Getachew, D., Yared, S., et al. (2020). Geographical distribution of *Anopheles stephensi* in eastern Ethiopia. *Parasites and Vectors* 13 (1), 35. doi:10.1186/s13071-020-3904-y

Cantsilieris, S., Stessman, H. A., Shendure, J., and Eichler, E. E. (2017). Targeted capture and high-throughput sequencing using molecular inversion probes (MIPs). *Methods Mol. Biol. Clift. N.J.* 1492, 95–106. doi:10.1007/978-1-4939-6442-0_6

Campino, S., Auburn, S., Kivinen, K., Zongo, I., Ouedraogo, J. B., Mangano, V., et al. (2011). Population genetic analysis of Plasmodium falciparum parasites using a customized Illumina GoldenGate genotyping assay. *PLoS One* 6 (6), e20251. doi:10.1371/journal.pone.0020251

Chang, H.-H., Worby, C. J., Yeka, A., Nankabirwa, J., Kamya, M. R., Staedke, S. G., et al. (2017). The real McCOIL: a method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites. *PLoS Comput. Biol.* 13 (1), e1005348. doi:10.1371/journal.pcbi.1005348

Conrad, M. D., Asua, V., Garg, S., Giesbrecht, D., Niaré, K., Smith, S., et al. (2023). Evolution of partial resistance to artemisinins in malaria parasites in Uganda. *N. Engl. J. Med.* 389 (8), 722–732. doi:10.1056/NEJMoa2211803

Coulibaly, A., Diop, M. F., Kone, A., Dara, A., Ouattara, A., Mulder, N., et al. (2022). Genome-wide SNP analysis of Plasmodium falciparum shows differentiation at drug-resistance-associated loci among malaria transmission settings in southern Mali. *Front. Genet.* 13, 943445. doi:10.3389/fgene.2022.943445

da Silva, C., Boene, S., Datta, D., Rovira-Vallbona, E., Aranda-Díaz, A., Cisteró, P., et al. (2023). Targeted and whole-genome sequencing reveal a north-south divide in P. falciparum drug resistance markers and genetic structure in Mozambique. *Commun. Biol.* 6 (1), 619. doi:10.1038/s42003-023-04997-7

Dondorp, A. M., Nosten, F., Yi, P., Das, D., Phyo, A. P., Tarning, J., et al. (2009). Artemisinin resistance in Plasmodium falciparum malaria. *N. Engl. J. Med.* 361 (5), 455–467. doi:10.1056/NEJMoa0808859

Fola, A. A., Feleke, S. M., Mohammed, H., Brhane, B. G., Hennelly, C. M., Assefa, A., et al. (2023). Plasmodium falciparum resistant to artemisinin and diagnostics have emerged in Ethiopia. *Nat. Microbiol.* 8 (10), 1911–1919. doi:10.1038/s41564-023-01461-4

Garrison, E., and Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing*. arXiv. arXiv [q-bio.GN].

Gautier, M., and Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28 (8), 1176–1177. doi:10.1093/bioinformatics/bts115

Hathaway, N. J., Parobek, C. M., Juliano, J. J., and Bailey, J. A. (2018). SeekDeep: single-base resolution *de novo* clustering for amplicon deep sequencing. *Nucleic Acids Res.* 46 (4), e21. doi:10.1093/nar/gkx1201

Ishengoma, D. S., Mandara, C. I., Bakari, C., Fola, A. A., Madebe, R. A., Seth, M. D., et al. (2024). *Evidence of artemisinin partial resistance in North-western Tanzania: clinical and drug resistance markers study*. medRxiv: The Preprint Server for Health Sciences. doi:10.1101/2024.01.31.24301954

Jacob, C. G., Thuy-Nhien, N., Mayxay, M., Maude, R. J., Quang, H. H., Hongvanthong, B., et al. (2021). Genetic surveillance in the Greater Mekong subregion and South Asia to support malaria control and elimination. *eLife* 10, e62997. doi:10.7554/eLife.62997

Juliano, J. J., Giesbrecht, D. J., Simkin, A., Fola, A. A., Lyimo, B. M., Pereus, D., et al. (2023). *Country wide surveillance reveals prevalent artemisinin partial resistance mutations with evidence for multiple origins and expansion of high level sulfadoxine-pyrimethamine resistance mutations in northwest Tanzania*. medRxiv: The Preprint Server for Health Sciences. doi:10.1101/2023.11.07.23298207

LaVerriere, E., Schwabl, P., Carrasquilla, M., Taylor, A. R., Johnson, Z. M., Shieh, M., et al. (2022). Design and implementation of multiplexed amplicon sequencing panels to serve genomic epidemiology of infectious disease: a malaria case study. *Mol. Ecol. Resour.* 22 (6), 2285–2303. doi:10.1111/1755-0998.13622

MalariaGEN, Ahouidi, A., Ali, M., Almagro-Garcia, J., Amambua-Ngwa, A., Amaratunga, C., et al. (2021). An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide samples. *Wellcome Open Res.* 6, 42. doi:10.12688/wellcomeopenres.16168.2

Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., et al. (2012). Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* 487 (7407), 375–379. doi:10.1038/nature11174

Mihreteab, S., Platon, L., Berhane, A., Stokes, B. H., Warsame, M., Campagne, P., et al. (2023). Increasing prevalence of artemisinin-resistant HRP2-negative malaria in Eritrea. *N. Engl. J. Med.* 389 (13), 1191–1202. doi:10.1056/NEJMoa2210956

Miles, A., Iqbal, Z., Vauterin, P., Pearson, R., Campino, S., Theron, M., et al. (2016). Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum. *Genome Res.* 26 (9), 1288–1299. doi:10.1101/gr.203711.115

Moser, K. A., Madebe, R. A., Aydemir, O., Chiduo, M. G., Mandara, C. I., Rumisha, S. F., et al. (2020). Describing the current status of Plasmodium falciparum population structure and drug resistance within mainland Tanzania using molecular inversion probes. *Mol. Ecol.* 30, 100–113. doi:10.1111/mec.15706

Mu, J., Myers, R. A., Jiang, H., Liu, S., Ricklefs, S., Waisberg, M., et al. (2010). Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat. Genet.* 42 (3), 268–271. doi:10.1038/ng.528

Neafsey, D. E., Schaffner, S. F., Volkman, S. K., Park, D., Montgomery, P., Milner, D. A., Jr, et al. (2008). Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. *Genome Biol.* 9 (12), R171. doi:10.1186/gb-2008-9-12-r171

Neafsey, D. E., Taylor, A. R., and MacInnis, B. L. (2021). Advances and opportunities in malaria population genomics. *Nat. Rev. Genet.* 22, 502–517. doi:10.1038/s41576-021-00349-5

Ngondi, J. M., Ishengoma, D. S., Doctor, S. M., Thwai, K. L., Keeler, C., Mkude, S., et al. (2017). Surveillance for sulfadoxine-pyrimethamine resistant malaria parasites in the Lake and Southern Zones, Tanzania, using pooling and next-generation sequencing. *Malar. J.* 16 (1), 236. doi:10.1186/s12936-017-1886-9

Niaré, K., Greenhouse, B., and Bailey, J. A. (2023). An optimized GATK4 pipeline for Plasmodium falciparum whole genome sequencing variant calling and analysis. *Malar. J.* 22 (1), 207. doi:10.1186/s12936-023-04632-0

Paschalidis, A., Watson, O. J., Aydemir, O., Verity, R., and Bailey, J. A. (2022). *coiaf: directly estimating complexity of infection with allele frequencies*. bioRxiv. doi:10.1101/2022.05.26.493561

Ranson, H., and Lissenden, N. (2016). Insecticide resistance in african Anopheles mosquitoes: a worsening situation that needs urgent action to maintain malaria control. *Trends Parasitol.* 32 (3), 187–196. doi:10.1016/j.pt.2015.11.010

Rao, P. N., Uplekar, S., Kayal, S., Mallick, P. K., Bandyopadhyay, N., Kale, S., et al. (2016). A method for amplicon deep sequencing of drug resistance genes in Plasmodium falciparum clinical isolates from India. *J. Clin. Microbiol.* 54 (6), 1500–1511. doi:10.1128/JCM.00235-16

RTSS Clinical Trials Partnership (2014). Efficacy and safety of the rts,S/AS01 malaria vaccine during 18 Months after vaccination: a phase 3 randomized, controlled trial in children and Young infants at 11 african sites. *PLoS Med.* 11 (7), e1001685. doi:10.1371/journal.pmed.1001685

RTS,S Clinical Trials Partnership (2015). Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *Lancet* 386 (9988), 31–45. doi:10.1016/S0140-6736(15)60721-8

Sadler, J. M., Simkin, A., Tchuenkam, V. P. K., Gerdes Gyuricza, I., Fola, A. A., Wamae, K., et al. (2024). *Application of a new highly multiplexed amplicon sequencing tool to evaluatePlasmodium falciparumantimalarial resistance and relatedness in individual and pooled samples from Dschang, Cameroon*. medRxiv. doi:10.1101/2024.10.03.24314715

Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F., and Neafsey, D. E. (2018). hmmIBD: software to infer pairwise identity by descent between haploid genotypes. *Malar. J.* 17 (1), 196. doi:10.1186/s12936-018-2349-7

Seyfarth, M., Khaireh, B. A., Abdi, A. A., Bouh, S. M., and Faulde, M. K. (2019). Five years following first detection of *Anopheles stephensi* (Diptera: Culicidae) in Djibouti,

Horn of Africa: populations established-malaria emerging. *Parasitol. Res.* 118 (3), 725–732. doi:10.1007/s00436-019-06213-0

Tessema, S. K., Hathaway, N. J., Teyssier, N. B., Murphy, M., Chen, A., Aydemir, O., et al. (2022). Sensitive, highly multiplexed sequencing of microhaplotypes from the Plasmodium falciparum heterozygome. *J. Infect. Dis.* 225 (7), 1227–1237. doi:10.1093/infdis/jiaa527

Uwimana, A., Legrand, E., Stokes, B. H., Ndikumana, J.-L. M., Warsame, M., Umulisa, N., et al. (2020). Emergence and clonal expansion of *in vitro* artemisinin-resistant Plasmodium falciparum kelch13 R561H mutant parasites in Rwanda. *Nat. Med.* 26 (10), 1602–1608. doi:10.1038/s41591-020-1005-2

Van Tyne, D., Park, D. J., Schaffner, S. F., Neafsey, D. E., Angelino, E., Cortese, J. F., et al. (2011). Identification and functional validation of the novel antimalarial resistance locus PF10_0355 in Plasmodium falciparum. *PLoS Genet.* 7 (4), e1001383. doi:10.1371/journal.pgen.1001383

Verity, R., Aydemir, O., Brazeau, N. F., Watson, O. J., Hathaway, N. J., Mwandagalirwa, M. K., et al. (2020). The impact of antimalarial resistance on the genetic structure of Plasmodium falciparum in the DRC. *Nat. Commun.* 11 (1), 2107. doi:10.1038/s41467-020-15779-8

Volkman, S. K., Neafsey, D. E., Schaffner, S. F., Park, D. J., and Wirth, D. F. (2012). Harnessing genomics and genome biology to understand malaria biology. *Nat. Rev. Genet.* 13 (5), 315–328. doi:10.1038/nrg3187

World Health Organization (2023). *World malaria report 2023*. World Health Organization.

Young, N. W., Gashema, P., Giesbrecht, D., Munyaneza, T., Maisha, F., Mwebembezi, F., et al. (2024). *High frequency of artemisinin partial resistance mutations in the great lake region revealed through rapid pooled deep sequencing*. bioRxiv. doi:10.1101/2024.04.29.24306442