Benchmarking and Optimization of Methods for the Detection of Identity-By-Descent in *Plasmodium falciparum*

Bing Guo[1, 2], Shannon Takala-Harrison[1, *], and Timothy D. O'Connor[2, *]

[1]Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD USA

[2]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA

[*]These authors jointly supervised this work

Abstract

Genomic surveillance is crucial for identifying at-risk populations for targeted malaria control and elimination. Identity-by-descent (IBD) is being used in *Plasmodium* population genomics to estimate genetic relatedness, effective population size (N_e), population structure, and positive selection. However, a comprehensive evaluation of IBD segment detection tools is lacking for species with high rates of recombination. Here, we employ genetic simulations reflecting *P. falciparum*'s high recombination rate and decreasing N_e to benchmark IBD callers, including probabilistic (`hmmIBD`, `isoRelate`), identity-by-state-based (`hap-IBD`, `phased IBD`) and others (`Refined IBD`), using genealogy-based true IBD and downstream inference of population characteristics. Our findings reveal that low marker density per genetic unit, related to high recombination rates relative to mutation rates, significantly affects the quality of detected IBD segments. Most IBD callers suffer from high false negative rates, which can be improved with parameter optimization. Optimized parameters allow for more accurate capture of selection signals and population structure, but `hmmIBD` is unique in providing less biased estimates of N_e. Empirical data subsampled from the MalariaGEN *Pf*7 database, representing different transmission

[1]Corresponding authors: Shannon Takala-Harrison, STakala@som.umaryland.edu, https://orcid.org/0000-0003-4674-8500; Timothy D. O'Connor, timothydoconnor@gmail.com, https://orcid.org/0000-0002-0276-1896.

settings, confirmed these patterns. We conclude that the detection of IBD in high-recombining species requires context-specific evaluation and parameter optimization and recommend that `hmmIBD` be used for quality-sensitive analysis, such as estimation of N_e in these species.

Keywords: Benchmarking, Optimization, Identity-By-Descent, Recombination rate, Marker density, *Plasmodium falciparum*

Introduction

Malaria is a mosquito-borne disease that poses a significant public health challenge globally, with an estimated 249 million clinical cases and 608,000 deaths occurring in 2022 (World Health Organization, 2023). Surveillance remains an important component of malaria control and elimination efforts. Advances in sequencing technologies and the scale of resequencing studies now allow for parasite genomic surveillance, which can provide insights into the efficacy of malaria interventions and guide the design of targeted elimination strategies in different transmission settings (Neafsey et al., 2021; Shetty et al., 2019; Wesolowski et al., 2018).

Identity-by-descent (IBD) is an essential tool in population genomics that has been used to estimate genetic relatedness (Gerlovina et al., 2022; Henden et al., 2018; Schaffner et al., 2018; Taylor et al., 2019), positive selection (S. R. Browning & Browning, 2020; Guo et al., 2024; Henden et al., 2018; Schaffner et al., 2023), effective population size (N_e) (Guo et al., 2024; Morgan et al., 2020), fine-scale population structure (Guo et al., 2024; Nait Saada et al., 2020; Shetty et al., 2019) and migration patterns (Al-Asadi et al., 2019; Shetty et al., 2019). However, the reliability of the IBD-based analysis is highly dependent on the accuracy of the detected IBD segments. Insufficient density of genetic markers, on a local or genome-wide scale, probably contributes to high error rates in the identification of IBD segments (S. R. Browning & Browning, 2020; Freyman et al., 2021; Zhou et al., 2020) and reduced accuracy of IBD-based estimates of population demography (S. R. Browning & Browning, 2015; Taylor et al., 2019). Many IBD detection methods have been designed for human genomes, where the demographic history and evolutionary parameters, including the recombination rate, differ considerably from *Plasmodium falciparum* (*Pf*). The effective population size of humans has increased rapidly in recent history (Gutenkunst et al., 2009), while that of *Pf* is decreasing, particularly in regions such as Southeast Asia and South America (Joy et al., 2003;

World Health Organization, 2023), due to enhanced malaria elimination efforts. More importantly, *Pf* genomes recombine about 70 times more frequently per unit of physical distance (Amambua-Ngwa et al., 2019; Gardner et al., 2002; Jiang et al., 2011; Miles et al., 2016; Su et al., 1999) than the human genome (Kong et al., 2002), while sharing a similar mutation rate (Bopp et al., 2013; Camponovo et al., 2023; Churcher et al., 2014; Hamilton et al., 2017; Huber et al., 2016; McDew-White et al., 2019; Neafsey et al., 2021) as human genomes (Campbell et al., 2012) on the order of 10^{-8} per base pair per generation. The decreasing population size (Guo et al., 2024) and the high recombination rate in *Pf* result in a reduced number of common variants, such as single nucleotide polymorphisms, per unit of genetic distance. If the marker density is too low, there may be insufficient genetic data to infer IBD. From a large human whole genome sequencing data set, we typically obtain millions of common biallelic SNP variants (Taliun et al., 2021), while only tens of thousands of common biallelic SNP markers with a minor allele frequency not less than 1%) are identified in the *Pf* genomes (MalariaGEN et al., 2023). Given that the human genome has a size of approximately 3,000 centimorgans (cM) and *Pf* with a length of 1,500 cM, the per-cM SNP density can be two orders of magnitude lower than in humans. Thus, it is critical to understand whether IBD detection methods can still generate accurate IBD segments under low SNP density conditions, considering the specific evolutionary parameters of the *Pf* genome.

Evaluating the quality of the detected IBD segments requires benchmarking with the known ground truth through simulation studies (Freyman et al., 2021; Shemirani et al., 2021; Tang et al., 2022; Zhou et al., 2020). The performance of IBD detection tools developed for use in the human context is typically measured using simulated genomes reflecting demographic and evolutionary parameters (B. L. Browning & Browning, 2011; Tang et al., 2022; Zhou et al., 2020), linkage disequilibrium patterns, and site frequency spectrum of human genomes (Freyman et al., 2021; Shemirani et al., 2021). Given the differences between human and *Pf* genomes, the quality assessment and parameter optimization derived from human studies likely do not apply directly to *Pf*. For tools explicitly designed for malaria parasites, such as `isoRelate` and `hmmIBD`, the evaluation of the quality of IBD was based on parent-offspring (Schaffner et al., 2018) or pedigree-based simulations (up to 25 generations) (Henden et al., 2018) that focused primarily on close relatives, which more likely mirrors low malaria transmission settings than high transmission settings. Furthermore, benchmarking methods and definitions of IBD accuracy are inconsistent across studies (Freyman

et al., 2021; Henden et al., 2018; Schaffner et al., 2018; Zhou et al., 2020), making the results of the quality evaluation of IBD difficult to compare. Considering the limitations of existing evaluations of IBD detection methods for *Pf* genomes, a unified benchmarking framework specifically designed for high recombining *Pf* genomes from low- and high-transmission settings is needed. Such a framework will assist researchers in comparing and prioritizing different IBD detection methods for intended downstream analysis.

In the present study, we developed a unified IBD benchmarking framework that reflects the demographic and evolutionary parameters of *Pf*. We evaluated how different recombination rates and marker densities affect the quality of detected IBD segments, performed IBD caller-specific parameter optimization, and benchmarked different IBD detection methods with their optimized parameters at both the IBD segment and downstream inference levels. Furthermore, we validated our findings from simulation data with empirical data sets constructed from subsets of samples from the publicly available whole genome sequencing database MalaraiGEN *Pf*7. Our findings indicate that a high recombination rate (given the same mutation rate) is associated with a low SNP density (per genetic unit), which substantially affects the accuracy of the detected IBD segments. IBD segments called by many of the benchmarked methods generally suffer from a high false negative rate (fraction of true IBD not detected) with the default parameters. IBD caller-specific parameter optimization can reduce errors in detected IBD segments and the resulting bias in IBD-based relatedness estimates. With optimized parameters, IBD inferred by most evaluated IBD callers can capture the main patterns of positive selection and population structure; however, only `hmmIBD` provides an accurate estimate of the trajectory of effective population size. To obtain optimal results, we recommend optimizing the parameters of human-oriented IBD callers before applying to *Pf* genomes, and using `hmmIBD` for IBD quality-sensitive analysis when genetic data from haploid genomes are available.

Results

Low SNP density due to high recombination rate affects the accuracy of IBD calls

Ideal IBD segments are typically defined as shared genomic regions descending unbroken by recombination from a common ancestor (S. R. Browning & Browning, 2012; Thompson, 2013), which can be derived from the ancestral recombination graph (ARG) with little bias (S. R. Browning & Browning, 2015; Guo et al., 2024; Kelleher et al., 2018; Tang et al., 2022). When ARG is unknown, IBD segments are generally inferred from mutational information, such as phased or unphased genotype data as a profile of the underlying ARG (Kelleher et al., 2019; Speidel et al., 2019). Thus, the accuracy of inferred IBD segments is highly dependent on the abundance of segregating sites available for IBD inference. However, high-recombining species like *Pf* may have a low density of mutation information in the genetic distance space.

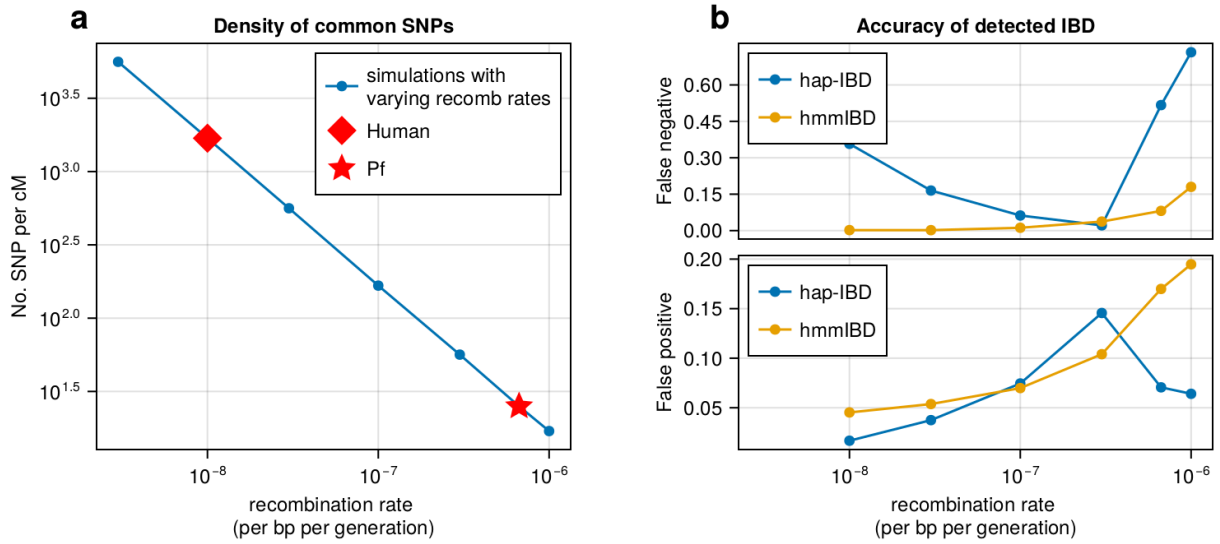

Figure 1: High recombination rates reduce genetic marker density and affect the quality of detected IBD segments. a, The number of common single nucleotide polymorphisms (SNPs) (minor allele frequency $\geq$ 0.01) per genetic unit (centimorgan, cM) in simulated genomes with different recombination rates. In these simulations (blue line), the mutation rates are fixed; the recombination rates vary widely to include the rate for both humans (red diamond) and *Pf* (red star). b, Accuracy of IBD segments detected from genomes simulated with different recombination rates. The accuracy of IBD segments is measured by the false negative rates (top panel) and false positive rates (bottom panel). The plotted error rates reflect the genome-wide rates (defined in Methods) of IBD segments called with default IBD caller parameters unless otherwise specified (see Supplementary Table 1). Only error rates for two IBD detection methods, `hmmIBD`, and `hap-IBD`, are included in (b) for simplicity. The error rates for other IBD callers are provided in Supplementary Figure 2. For both (a) and (b), the genomes were simulated under the single-population model (see Methods).

To assess how recombination rates affect the density of segregating sites per genetic unit, we simulated genomes with varying recombination rates but a fixed mutation rate. For this analysis, we used the single-population demographic model for simplicity (see Methods). We found that the density per centimorgan (cM) of common biallelic SNPs (minor allele frequency ≥ 0.01), in selectively neutral scenarios, is inversely correlated with recombination rates ranging from 3×10^{-9} to 10^{-6} per base pair per generation (Figure 1 a). For example, the SNP density of *Pf*-like genomes is 25 SNPs per cM, which is approximately 1/67 of that of the human-like genomes (1,660 SNPs per cM). We further assessed how the low density of SNPs associated with a high recombination rate affects the accuracy of the detected IBD segments by comparing them with the true IBD based on simulated genealogical trees (via `tskibd` (Guo et al., 2024)). We evaluated the accuracy using two IBD segment-level metrics, including the false negative rate (FN), which represents the proportion of a true IBD segment not covered by inferred IBD segments of the same genome pairs, and the false positive rate (FP), which indicates the faction of a true IBD segment not covered by inferred IBD segments (see Methods for detailed definitions and Supplementary Figure 1 for an overview of the methods). Our analysis showed that as the recombination rate increases and the SNP density decreases, both the genome-wide FN (Figure 1 b top) and FP (Figure 1 b bottom) increase for IBD inferred from `hmmIBD`. The patterns vary in the other four IBD detection methods, and all suffer elevated FNs and/or FPs as the recombination rate increases (Figure 1 b and Supplementary Figure 2), except for `isoRelate`, of which the IBD quality tends to be better for input data with lower marker density. The results suggest that a low SNP density per genetic unit can dramatically affect the reliability of detected IBD segments.

Varying quality of IBD inferred from simulated *Pf* genomes via different IBD callers

Multiple IBD callers have been used for *Pf*, including *Pf*-oriented, Hidden Markov Model-based methods, such as `hmmIBD` (Schaffner et al., 2018) and `isoRelate` (Henden et al., 2018), and those originally designed for human genomes, such as `Refined IBD` (Morgan et al., 2020) and `Beagle` (version 4.1) (Shetty et al., 2019). For *Pf*-oriented callers, the performance is largely unexplored in population samples, where most genomes are not closely related and share mainly short IBD segments. For human-oriented callers, applying to *Pf* directly without further optimization could

be problematic, given the significant differences in recombination rates and demographic history between humans and *Pf*. We analyzed the `hap-IBD` (Zhou et al., 2020) and `phased IBD` (Freyman et al., 2021) IBD callers in addition to `hmmIBD`, `isoRelate` and `Refined IBD`, since the former represents two recent key advancements in the development of IBD detection methods that scale well to large sample sizes and genome sizes.

To evaluate the applicability and accuracy of these IBD detection methods in analyzing *Pf* genomes, we performed benchmarking analyses in simulated genomes (Supplementary Figure 1, top panel), mimicking the high recombination rate and the decreasing population size of *Pf* populations. Our analyses include three sets of comparisons: (1) baseline benchmarking, where we mainly used the default parameter values for each IBD caller and compared the performance in *Pf* genomes at the level of IBD segment and their simple statistics; (2) post-optimization benchmarking, where we used parameter values optimized specifically for each IBD caller so that our low- and high-level comparisons are based on the optimal performance of each method; (3) human-like genome benchmarking, where detected IBD segments are expected to have low error rates for human-oriented IBD callers and thus were used as an internal control to validate our benchmarking pipeline (see Supplementary Table 1 for the IBD caller parameters and Methods for demographic models).

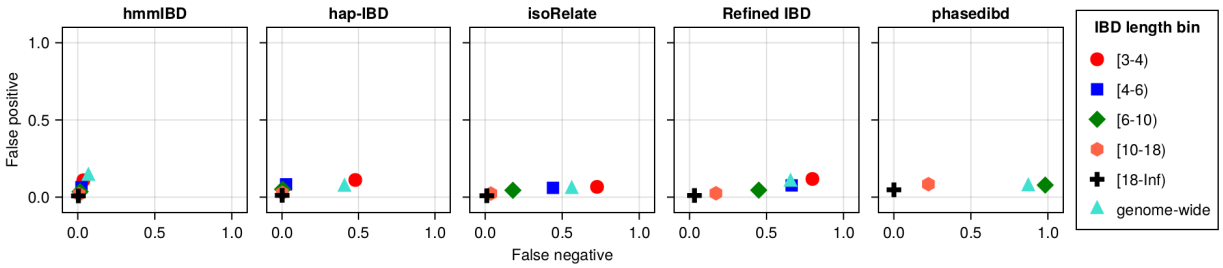

Figure 2: Accuracy of IBD segments detected from *Pf* genomes varies across IBD callers. IBD segments were inferred from genomes simulated under the single-population model with a recombination rate compatible with *Pf*. The accuracy of IBD was evaluated using the calculated false positive rate (y axis) and false negative rate (x axis). The rates were calculated for different length bins in centimorgans, including [3-4), [4-6), [6-10), [10-18), [18, inf) centimorgan and at the genome-wide level (defined by overlapping analysis between true IBD segments and inferred IBD segments from each genome pair). The IBD callers analyzed here, from left to right, include `hap-IBD`, `hmmIBD`, `isoRelate`, `Refined IBD`, and `phased IBD`. The results of the simulations under the multiple population model are provided as Supplementary Data.

From baseline benchmarking, we found that, under the single population model with shrinking population sizes, all callers except `hmmIBD`, including human-oriented IBD callers (`hap-IBD`, `Refined IBD`, and `phased IBD`) and *Pf*-oriented `isoRelate`, suffer from high FNs, especially for short IBD

segments (genome-wide FN/FP rates tend to reflect shorter IBD segment as most segments are short) (Figure 2). Among them, `hap-IBD` only suffers high FNs for IBD segments of shorter length ([3-4) cM) but not longer IBD segments. Similarly, genetic relatedness metrics based on pairwise total IBD are largely underestimated for these four callers (Supplementary Figure 3). For *Pf*-oriented IBD callers(`hmmIBD` and `isoRelate`), we found that by default `hmmIBD` has relatively low FN/FP error rates (Figure 2) and is less biased for relatedness estimates (Supplementary Figure 3). In contrast, `isoRelate` has high false negative rates despite the low FP rates (Figure 2). Thus, both *Pf*- and human-oriented IBD callers can suffer high error rates when detecting IBD from genomes with a high recombination rate and a low marker density, with the exent depending on factors including underlying assumptions and methodologies.

IBD-caller-specific parameter optimization for *Pf* improves IBD accuracy

Given that these IBD callers are designed independently and optimized for different species or genotype datasets, we hypothesized that optimization of IBD caller-specific parameters under a unified framework (including a demographic model and a recombination rate consistent with *Pf* genomes) can potentially improve the performance of these IBD callers. Considering that each IBD caller provides multiple parameters to control its behavior and searching the entire parameter space is inefficient, our optimization focused mainly on parameters potentially affected by or needing adjustment according to differences in marker density between the new target (*Pf*) and previously tested populations (humans). For example, the `min-marker` parameter of `hap-IBD` sets the minimum number of markers in an IBD segment (including the seed and extension segments) to be reported; the default value recommended for the human genome (100) can be too high for *Pf* genomes, which have a much lower marker density, causing a high rejection rate of candidate segments and thus high false negative rates in the inferred IBD tracts, as shown above (Figure 2 and Supplementary Figure 3). We therefore iterated over a list of discrete values for the `min-marker` parameter or did a grid search over two or more parameters and looked for the value(s) that allow the generation of inferred IBD with low and balanced error rates. For IBD callers that do not explicitly have marker density-related parameters, such as `hmmIBD`, we explored other parameters that likely affect IBD quality. We summarized all the parameters explored and their corresponding values in Supplementary Table 2, with detailed results provided as Supplementary Data.

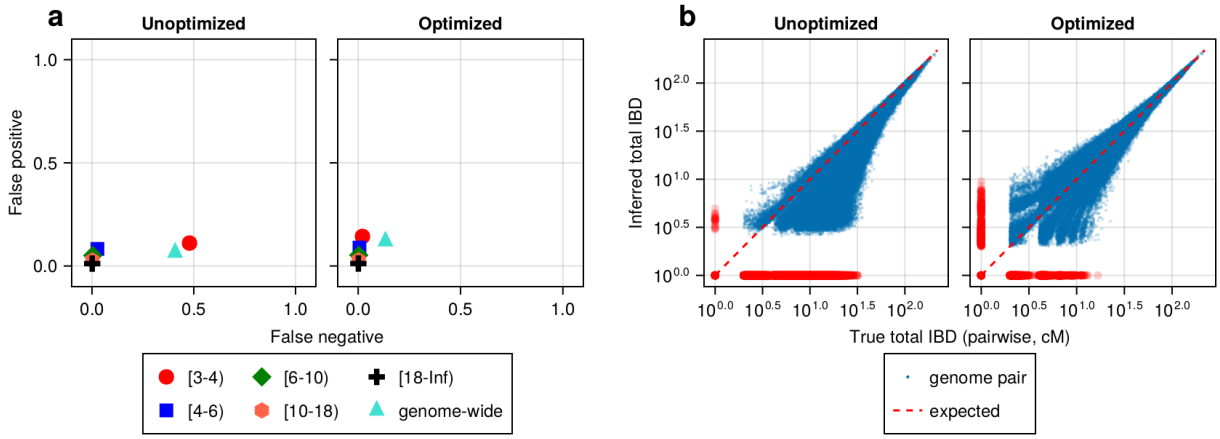

Figure 3: IBD caller-specific parameter optimization can improve the quality of IBD segments inferred from simulated *Pf* genomes (using `hap-IBD` as an example). a, Quality of detected IBD measured by false positive and false negative rates before (left column) and after (right column) `hap-IBD`-specific parameter optimization. As indicated in the axis legend, the error rates were calculated for different length ranges (in centimorgans), including [3-4), [4-6), [6-10), [10-18), [18, inf) and at the genome-wide level. b, Quality of detected IBD measured by total genome pairwise IBD, an estimate of genetic relatedness, before (left column) and after (right column) `hap-IBD` parameter optimization. Each dot represents a pair of genomes with the coordinates x and y being true and inferred total IBD. Note: both the x and y axes in (b) use log scales. In (b), the blue dots are the pairs with nonzero true and inferred total IBD while red dots are pairs with either true total IBD or inferred total IBD being 0; zero-valued total IBD was replaced with 1.0 cM for visualization purposes. The red dotted line of $y = x$ indicates the expected pattern, that is, true total IBD equal to inferred total IBD if the inferred IBD was 100% accurate.

We found that most IBD callers have a key parameter that can dramatically affect their FN/FP rates (last column in Supplementary Table 2). For example, the FN rates of IBD called from `hap-IBD` change substantially when the `min-marker` parameter varies (see Supplementary Data). With a value of 70, the FN rate for short IBD segments dramatically decreases so that the FN and FP error rate tends to be balanced at the cost of a slight increase in the FP rates (Figure 3 a). When the inferred pairwise total IBD is compared with the true total IBD (obtained via `tskibd` (Guo et al., 2024)), these relatedness estimates change from being highly biased (underestimated) before parameter optimization (Figure 3 b, left column) to being more balanced after optimization (Figure 3 b, right column). Similarly, parameter optimization dramatically reduces FN rates for `Refined IBD` and `phased IBD` with relatively small increases in FP rates (Supplementary Figure 4 a). With these changes, the relatedness estimates are less biased (Supplementary Figure 4 b). In contrast, the quality metrics remained largely unchanged during parameter optimization attempts for `hmmIBD` and `isoRelate`, with `hmmIBD` being more accurate and unbiased and `isoRelate` suffering from high false negative rates and an underestimated relatedness (Supplementary Figure 4). The observation

that human-oriented IBD callers are more optimizable for the simulated *Pf* genome compared to *Pf*-oriented IBD callers confirms that IBD callers originally targeting a non-*Pf* population need adjustment before they can be used for *Pf*.

The human-oriented IBD callers, when not optimized for *Pf*, performed worse than `hmmIBD`, especially for `Refined IBD` and `phased IBD` (Supplementary Figure 4 a). To exclude potential problems in our benchmarking pipeline, we simulated genomes with human recombination rates and the demographic history of the human population in the UK (Supplementary Figure 5 a, left column, and b, left column; also see Methods). Then we evaluated human-oriented IBD callers with the same benchmarking pipeline *without* parameters optimized for *Pf*. The results of the analysis showed that these human-oriented IBD callers indeed perform much better for human genomes, showing consistently lower FN/FP error rates (Supplementary Figure 5 a, left column) and less biased total IBD-based relatedness estimates (Supplementary Figure 5 b, left column), when compared with results in *Pf*-like genomes (Supplementary Figure 4 a-b, left columns). These results support the robustness of our benchmarking pipeline (Supplementary Figure 5 a-b, left columns) and demonstrate that directly applying human-oriented IBD callers to *Pf* may present challenges. Furthermore, we found that the performance of these IBD callers is sensitive to various demographic configurations (single-population demographic model in Supplementary Figure 4 a-b, left columns, versus UK human demographic model in Supplementary Figure 5 a-b, right columns) even with the *same* (*Pf*) recombination/mutation rates and default IBD caller parameters, suggesting that IBD caller parameter optimization may also be demography dependent (more details are provided as Supplementary Data).

Given that low SNP densities tend to be associated with high error rates (Figure 1 b), we tested whether increasing SNP density by including rarer variants could improve IBD segment quality. We found that including rare biallelic SNPs in genotype data used for `hmmIBD` IBD detection does not dramatically change FN/FP error rates, with only a slight increase in FP rate and a decrease in FP rates for short IBD segments ([3-4) cM) for different demographic models (Supplementary Data).

Post-optimization benchmarking via downstream inferences

IBD-based downstream analyses, such as estimation of N_e, selection signals, and population structure, are key applications of IBD segments in population genetics, which often rely on the high

quality of input IBD segments. With optimized IBD caller-specific parameters tailored for *Pf*-like genomes, we can expect IBD callers to perform at their best, which allows high-level benchmarking by comparing IBD-based downstream estimates.

For the detection of selection signals based on IBD, we simulated positive selection (selection coefficient $s = 0.2$) with a sweep on each of the 14 chromosomes via the single-population model, used inferred IBD from different callers to identify IBD peaks, and validated the peaks using statistics based on the integrated haplotype score (iHS) X_{iHS} (see Methods and our recent work (Guo et al., 2024)). The validated peaks were inferred as regions under selection; we treated each validated peak as a true signal if the peak contains the selected site as specified in simulations. We found that the IBD segments from most callers capture a majority of the simulated positive selection signals except `Refined IBD`, which is less sensitive and only detects 3 of 14 true selection signals (Supplementary Figure 6 a). In addition, `isoRelate` tends to suffer from an increased level of false positives or a relatively low signal-to-noise ratio, as shown in the IBD coverage curve (Supplementary Figure 6 a).

For IBD-based population structure inference, we simulated *Pf*-like genomes with the multiple-population demographic model under a selectively neutral condition (see Methods). We used IBD network community detection algorithm InfoMap (Csardi & Nepusz, 2006; Rosvall et al., 2009) to obtain subpopulation assignments. Similarly to IBD-based selection signal detection, we found that IBD inferred from most callers can recapitulate the simulated population structure with accuracy comparable to that of true IBD (`tskibd` in Supplementary Figure 6 b). The exception is that `isoRelate` tends to generate a larger number of smaller groups, which is likely due to the high FN rates for short IBD segments, thus misses the connectivity among genomes that are distantly related and tends to capture "family" structure or closely related, small subgroups.

For IBD-based N_e inference (via `IBDNe` (S. R. Browning & Browning, 2015)), we simulated neutral *Pf* genomes using the single-population model, since `IBDNe` were designed to work with a homogeneous population (S. R. Browning & Browning, 2015). The tool `IBDNe` estimates the trajectory of N_e for the last 100-200 generations based on population total amount of IBD of different length bins, with short IBD segments informative for the more distant past and longer for more recent time frames. Since this method relies on the IBD segment length to estimate the age of shared ancestors, we expect that IBD-based estimates of N_e would be highly sensitive to quality

issues of IBD inputs. Our results showed that most of the compared IBD callers suffer from a wild oscillation, which has previously been observed (S. R. Browning & Browning, 2015), and tends to deviate from the truth for older generations(Figure 4), consistent with the general pattern of high false negative or positive rates in shorter IBD length bins for these IBD callers (Supplementary Figure 4 a, right column). Meanwhile, the IBD inferred by `hmmIBD` generated highly accurate estimates comparable to the estimates based on true IBD (Figure 4). We explored the mechanisms underlying bias in N_e estimates by comparing the total amount of inferred IBD from the entire population for specific length ranges with the truth (`tskibd`). We found that the strong bias in N_e estimates for older generations is likely associated with underestimation of population-level total IBD for short IBD segments (Supplementary Figure 7). The bias is mostly obvious for the shorter bins in `hap-IBD`, `isoRelate`, and `phased IBD`, followed by `Refined IBD`. For `hmmIBD`, both the estimates of N_e (Figure 4) and population total IBD are relatively unbiased (Supplementary Figure 7 column 2), consistent with the relatively low and balanced FP/FN rates of IBD inferred via this caller (Figure 2, leftmost panel). Together, these results suggest that IBD-based N_e estimates are highly sensitive to the quality of input IBD segments, and that `hmmIBD` is recommended for this analysis.

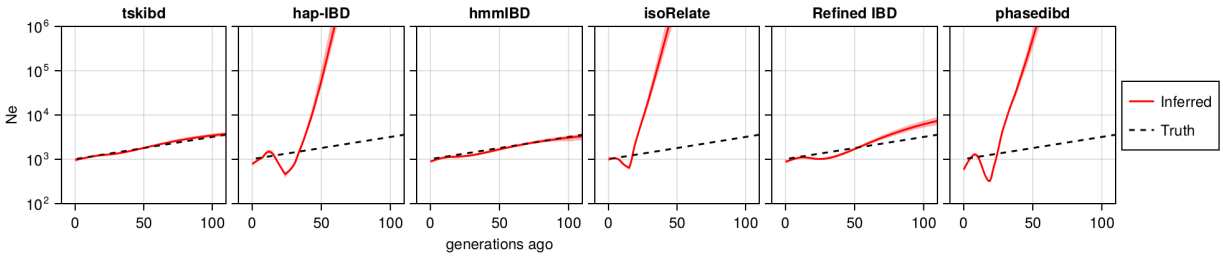

Figure 4: Post-optimization benchmarking of different IBD callers by comparing downstream estimates N_e. With parameters optimized for each IBD caller, the performance of IBD callers was evaluated by comparing the N_e trajectory for the recent 100 generations estimated via `IBDNe` based on true (black dashed line) IBD versus inferred IBD (red solid line). True IBD was calculated from simulated genealogical trees via `tskibd` inferred IBD includes those inferred from `hap-IBD`, `hmmIBD`, `isoRelate`, `Refined IBD`, and, `phased IBD`, with their N_e estimates shown from left to right. The shading areas surrounding the red lines indicate 95% confidence intervals as determined by `IBDNe`. See Supplementary Figure 10 for pre-optimization results.

To provide more context on the application of the quality-sensitive N_e estimation in *Pf*, we further explored whether excluding short IBD segments ($<$ 4cM), which often suffer from higher error rates, could improve the quality of N_e estimates for non-`hmmIBD` callers. The result showed that N_e estimates based on only longer IBD segments ($\geq$ 4 cM) indeed exhibit reduced oscillation

of the trajectory for some scenarios, such as estimates based on IBD from for `hap-IBD` in the simulated data (Supplementary Figure 8 a), but are often associated with wide confidence intervals or underestimation of N_e in older generations (Supplementary Figure 8 a for the simulated data set and b for empirical datasets). We then explored the potential reasons underlying the oscillation commonly observed in the estimated N_e trajectories in recent time frames (B. L. Browning & Browning, 2011; Harris et al., 2020; Morgan et al., 2020). We hypothesized that the small oscillation or drop around 20 generations ago for all IBD callers (Supplementary Figure 8 a, second row and b, both rows) is partially due to the inclusion of IBD segments with TMRCA < 1.5 generations ago in `IBDNe` input, which likely affects N_e estimation (S. R. Browning & Browning, 2015). We found that the removal of these IBD segments (when TMRCA information is available, for example, from simulation) can greatly mitigate this problem (Supplementary Figure 8 a), especially for `hmmIBD` and `Refined IBD`. The findings suggest that special attention should be paid when interpreting (1) a recent drop in an estimated N_e trajectory in an empirical data set where TMRCA-based filtering of IBD segments is less practical and (2) the extremely large estimates in older generations partially stemming from high error rates for short IBD segments.

To confirm that downstream inference benefits from parameter optimization (based on IBD-level metrics), we compared post-optimization results (Supplementary Figure 6 and Figure 4) with pre-optimization estimates (Supplementary Figure 9 and Supplementary Figure 10). We found that parameter optimization indeed increased the power in selection detection (in `isoRelate`, `Refined IBD` and `phased IBD`, Supplementary Figure 6 a versus Supplementary Figure 9 a), improved population structure inference (`phased IBD` and `hap-IBD`, Supplementary Figure 6 b versus Supplementary Figure 9 b), and reduced oscillation on the N_e trajectory (`Refined IBD`, Figure 4 versus Supplementary Figure 10).

Validation in empirical data set

Although evolutionary parameters reflecting *Pf* were used, simulated populations are still a much simplified approximation of empirical data from real-world parasite populations. However, benchmarking IBD callers with empirical data (population samples) at the IBD segment level is challenging, since true IBD segments are not available. We focused on high-level analyses by evaluating whether IBD-based downstream estimates are consistent with expected patterns. To perform anal-

yses similar to simulation analysis, we constructed "single" or "multiple" population data sets by filtering and subsetting samples from the MalariaGEN *Pf*7 dataset (MalariaGEN et al., 2023) (see Methods for details). We compared the performance of different IBD callers with the empirical data set by performing downstream analysis, including N_e estimation and selection signal detection with "single" population data sets and InfoMap population structure inference with the "multiple" population data set.

With optimized parameters (Supplementary Table 1), IBD estimates from most callers, except `Refined IBD`, tend to capture most of the selection signals from the single population data set focusing on Southeast Asia, including known selective sweeps associated with loci involved in anti-malarial drug resistance and sexual commitment, including dihydrofolate reductase (*dhfr*) (Miotto et al., 2013), multidrug resistance protein 1 (*pfmdr1*) (Koenderink et al., 2010), amino acid transporter 1 (*pfaat1*) (Amambua-Ngwa et al., 2019), chloroquine resistance transporter (*pfcrt*) (Martin & Kirk, 2004), dihydropteroate synthase (*dhps*) (Brooks et al., 1994), Apicomplexan-specific ApiAP2-g(*ap2-g*) (Early et al., 2022) and apicoplast ribosomal protein S10 (*arps10*) (Miotto et al., 2015) (Figure 5 a). Although `hmmIBD` can detect more peaks beyond shared ones, it appears to suffer from a noisy background, likely due to the relatively high FP to FN ratios for short (<4 cM) IBD segments (Figure 1 and Figure 2).

Similar to the simulation analysis, IBD detected using most callers resulted in N_e estimates with unrealistic oscillations for the Eastern SEA data set, including small estimates at generation 0 and extremely large estimates in the more distant past (> 20 generations ago) (Figure 5 b). The problems are much less severe with the estimates from `hmmIBD`, consistent with a high quality of the inferred IBD segments from `hmmIBD`. The estimated N_e pattern (from `hmmIBD`) mirrored the expected reduction in malaria in this region due to the intense efforts to eliminate malaria in recent decades (World Health Organization, 2023).

InfoMap-based community detection using the global, multiple-population *Pf* data set shows that IBD from different callers can capture a similar expected pattern of continental population structure, with African *Pf* parasites being less structured, and Southeast Asian parasites being more structured and separate from Oceanian parasites (Figure 5 c), consistent with previous characterization using non-IBD-based methods (MalariaGEN et al., 2023). Similarly to simulation analysis, `isoRelate` fails to capture the true underlying population structure: IBD estimates from `isoRelate`

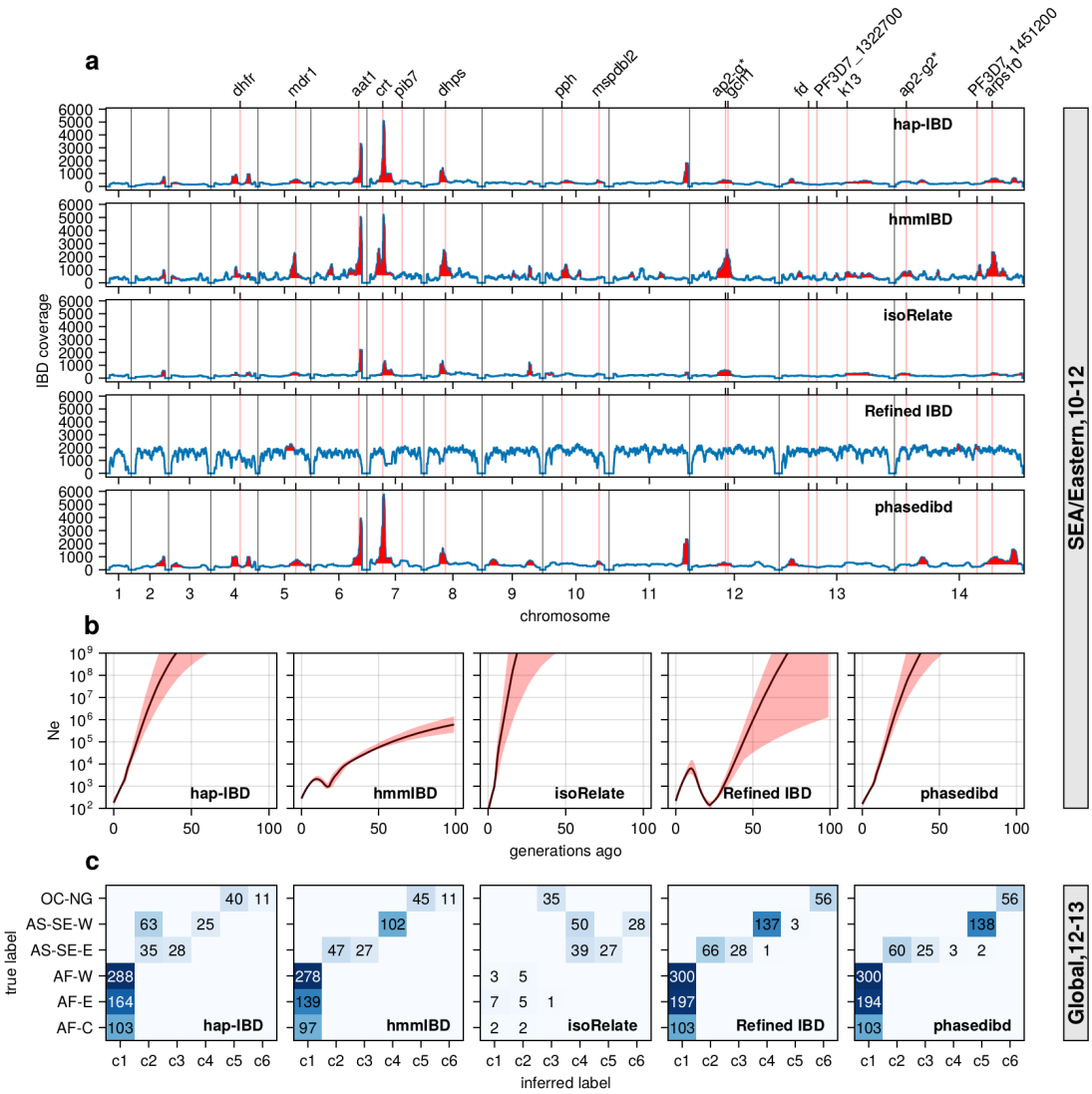

Figure 5: Validation of the performance of IBD callers in empirical data sets by comparing IBD-based downstream analyses. a, IBD coverage and detected selection signals in the SEA data set using different IBD callers (rows 1 to 5). Annotations and corresponding vertical dotted lines at the top indicate the center of known and putative drug resistance genes and genes related to sexual commitment; red shading indicates regions that are inferred to be under positive selection (see Methods for definitions). b, N_e estimates of the SEA data set based on IBD inferred from different callers. Line plots are point estimates; the shading areas around the line plots indicate confidence intervals based on bootstrapping (generated by `IBDNe`). c, Inference of the population structure of the structured data set by the InfoMap community detection algorithm using the IBD inferred from different IBD callers. The rows of the heatmap are geographic regions of isolates, and the columns are the largest, inferred communities, labeled as c1 to c6. The heat map color represents the number of isolates in each block with the given row and column labels. The columns are rearranged so that the diagonal blocks tend to have the largest values per row for better visualization.

tend to generate IBD communities with a small number of isolates (small groups) potentially due to high false negative rates for short IBD segments. This issue is most obvious in high-transmission

settings such as Africa where parasites have low relatedness on average and mainly share short IBD segments (Figure 5 c).

To further confirm that IBD caller parameter optimization improves IBD-based downstream estimates, we performed similar analyses except using IBD detected with nonoptimized parameters (see Supplementary Table 1). We found that the height and number of detected IBD peaks for `hap-IBD` and `phased IBD` before parameter optimization (Supplementary Figure 11 a) decreased dramatically compared to the results after optimization (Figure 5 a). Pre-optimization N_e estimates also suffered more extreme oscillations for human-oriented IBD callers (Supplementary Figure 11 b), although their post-optimization results were still highly biased (Figure 5 b). Consistent with high false negative rates for shorter IBD segments called with non-optimized parameters, pre-optimization IBD estimates from `hap-IBD` and `phased IBD` fail to reveal the expected population structure, with the detected communities being very small in African parasite populations (Supplementary Figure 11 c versus Figure 5 c). The difference between pre- and post-optimization in empirical studies further demonstrated the importance of IBD caller parameter optimization for *Pf*, especially for IBD callers not validated for *Pf*.

Computation efficiency comparison and improvement

As the cost of whole genome sequencing decreased and the availability of whole genome sequencing samples improved rapidly, as exemplified by MalariaGEN *Pf*7 (n = 20,864) (MalariaGEN et al., 2023), it is also important to prioritize IBD callers that scale well for large sample sizes. Therefore, we compared the IBD inference time and maximum memory usage for different IBD callers. Given that not all IBD callers can take advantage of multithreading to accelerate the inference process, we performed the comparison with or without enabling parallelization. When using a single thread, IBD callers involving probabilistic inference algorithms, including `isoRelate`, `Refined IBD`, and `hmmIBD`, are much slower for those based on identity-by-state-based or positional Burrows-Wheeler transform (PBWT) based algorithms, such as `hap-IBD` and `phased IBD`, with the runtime difference of about two orders of magnitude (Figure 6 a). When using a single thread, the maximum memory consumption is highest in `Refined IBD`, followed by `isoRelate`, `phased IBD` and `hap-IBD`, with `hmmIBD` being around ten times more efficient (Supplementary Figure 12 a). With multithreading, the patterns are similar to single-thread comparison as most allow parallelization (Figure 6 b and

Supplementary Figure 12 b). The exception is `hmmIBD`, as it currently only supports a single-thread computation. Notably, with sample sizes smaller than n = 1,000, the runtime of the single-threaded `hmmIBD` was still shorter than that of multiple-threaded (10 threads) `isoRelate` , which is generally in agreement with a previous evaluation (Schaffner et al., 2018), suggesting that the algorithm underlying or its implementation of `hmmIBD` is more efficient or simple compared to `isoRelate`. However, while `hmmIBD` produces more accurate IBD segments from *Pf* genomes than IBS-based methods, it remains significantly slower, highlighting the need for further enhancements to efficiently analyze large WGS data sets like MalariaGEN *Pf*7.

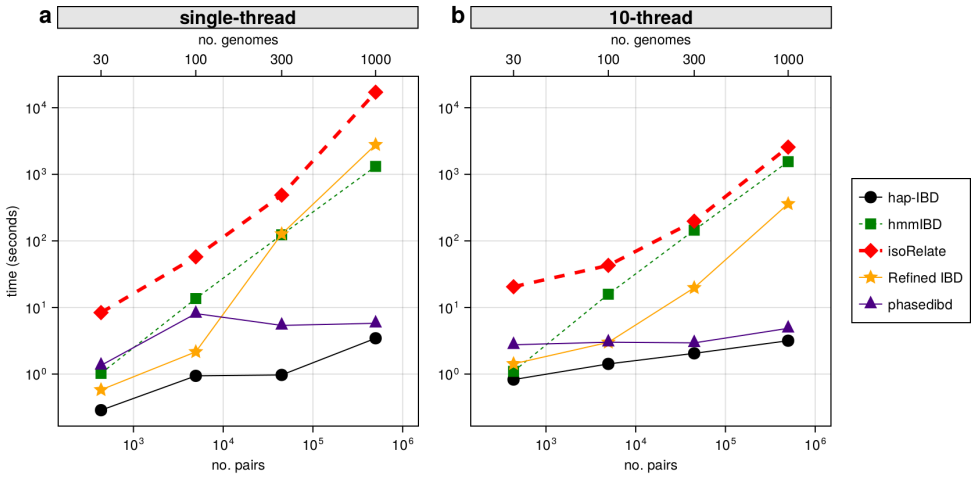

Figure 6: Comparison of computational runtime for IBD calling process for different callers. a, Runtime for different IBD callers to detect IBD from genomes of different sample sizes in single-thread mode. The comparison is based on *Pf* genomes of size of 100 cM simulated under the single population model. The x-axis tick labels include the number of pairs of genomes analyzed (below the plot, on a linear scale) and the sample size (number of haploid genomes, above the plot, on a nonlinear scale) analyzed. The line styles and markers for different callers/tools are provided in the legend box on the far right of the figure, which is shared across the three subplots. b, Runtime in multithreading mode. (b) is organized similarly to (a), except that the IBD calling processes were run in multithreading mode with 10 threads. Also, see Supplementary Figure 12 for the maximum memory usage for different callers.

Discussion

Our study sought to address important questions regarding the application of identity-by-descent (IBD) in highly recombining species, such as *Plasmodium falciparum*, focusing on the reliability of IBD segment detection and the prioritization of detection methods for accurate IBD-based analyses. We observed that: (1) Low marker density per genetic distance, resulting from high recombination

rates relative to mutation rates, significantly affects the accuracy of IBD detection; (2) Optimizing IBD caller-specific parameters can improve the performance of IBD callers for *Pf* genomes, especially for callers designed for human genomes; (3) After parameter optimization, most IBD callers effectively capture expected positive selection signals and population structure patterns; however, only `hmmIBD` can provide plausible N_e estimates for simulated genomes mimicking the *Pf* population; (4) Probabilistic (HMM-based) IBD detection methods are significantly less computationally efficient than IBS-based methods; `hmmIBD` will need enhancement through parallelization and memory optimization, to enable scalability to larger data sets. Consequently, we recommend optimizing human-oriented IBD callers before their application in malaria parasite research and endorse the use of `hmmIBD` for quality-sensitive IBD-based inferences, for example, the estimation of N_e.

Comparing the performance of multiple IBD detection methods often requires a unified framework, which should include a uniform definition of accuracy and a simulated ground truth that mimics *Pf* genomes. Our benchmarking framework incorporates several novel features. First, it utilizes a consistent definition of IBD length-specific accuracy based on the overlap of IBD segment lengths, closely aligned with metrics used in `hap-IBD` (Zhou et al., 2020), `phased IBD` (Freyman et al., 2021), and `Refined IBD` (B. L. Browning & Browning, 2013), as detailed in our Methods section. The approach differs from the original evaluation of `hmmIBD`, which assesses the accuracy of IBD based on the fraction of SNPs that share IBD, which could overlook the precise accuracy of IBD length of the detected segments (Schaffner et al., 2018). In particular, the original study of `isoRelate` defined the accuracy (true positive rate) using a less stringent overlap-by-count criterion where a segment is counted as accurate when at least 50% of a detected IBD segment is overlapped by true segments (Henden et al., 2018). Second, we generated *Pf*-like genomes via population genetic simulation that reflect a realistic distribution of IBD segment lengths. This contrasts with previous studies, where methods such as `hap-IBD` and `Refined IBD` used human-like data from population genetic simulations (B. L. Browning & Browning, 2013; Zhou et al., 2020), whereas `hmmIBD`, `isoRelate`, and phased-ibd relied on simulations based on artificial recombination or pedigree models (Henden et al., 2018; Schaffner et al., 2018). These models (no-population-based genetic simulation) often produce long shared IBD segments typical of close relatives, failing to capture the IBD length distribution in population samples predominantly comprising distant relatives. Third, our benchmarking extends beyond the segment-level evaluation of IBD callers and

includes downstream inferences of population structure and effective population size, providing a more thorough assessment of their application in real-world analyses. This comprehensive approach seeks to complement and broaden the scope of quality evaluations that are predominantly focused on the IBD segment level in original studies, thus improving our understanding of the performance of IBD detection methods in the context of *Pf* genomes.

The density of markers per genetic unit, as a measure of information enrichment for coancestry inference, plays a crucial role in determining whether and how a candidate IBD segment is accepted in the final IBD calls. IBS-based methods, such as `hap-IBD` and `phased IBD`, first identify long IBS segments ($\geq$ 2 centimorgan) as candidate IBD segments, and subsequently merge short ones separated with small gaps (Zhou et al., 2020), allow a certain number of discordant markers to account for phasing errors (Freyman et al., 2021), or eliminate false positive segments by removing candidate segments supported by only a small number of markers (Zhou et al., 2020). Similarly, `Refined IBD`, which combines an IBS-based method with an HMM probabilistic model, uses a LOD score to decide whether a candidate IBD segment should be rejected or accepted (B. L. Browning & Browning, 2013). In these studies, default values for threshold parameters related to marker number/density were shown to be effective for genomes like humans. Although the use of different parameter values has been considered for these algorithms to account for marker density differences between sequencing data versus microarray genotype data (Zhou et al., 2020), per-genetic-unit marker density in human data, even for microarray genotype data, is still significantly higher than in *Pf* sequencing data. We evaluated how different levels of per-genetic-unit marker density affect the detection of IBD segments by varying recombination rates in simulations. With simulated *Pf*-like genomes of low marker density and high recombination rate, we found that human-oriented IBD callers suffer high false negative rates. One potential explanation is that the thresholds optimized for human data are too stringent for *Pf* causing excessive rejection of candidate segments. The effect of marker-density on IBD detection is further confirmed by our findings that adjusting the values of marker-density-related parameters using a grid-search approach could significantly reduce IBD error rates and generate more accurate IBD-based downstream estimates. Even though IBD accuracy can be improved by parameter optimization, we found that error rates of the detected IBD segments are still higher in *Pf* genomes than those of human genomes even after IBD caller parameters are optimized. There are several possible explanations for the high error rates of IBD

segments detected from low marker density data: (1) Detected IBD segments can only start and end at the genotype marker site, bypassing any non-genotyped sites; (2) A lower marker density is linked with greater uncertainty in the distribution of IBD endpoints (S. R. Browning & Browning, 2020); (3) Ancestral relationships including IBD are too difficult to be reliability inferred given limited mutational information (Ishigohoka & Liedvogel, 2024; Kelleher et al., 2019; Mehra et al., 2024; Speidel et al., 2019).

In our benchmark analysis, we used only common biallelic SNPs as markers for inferring IBD, excluding rare variants and indels. The use of this additional information can potentially provide more dense and richer genotype information, thus enhancing our understanding of the population's ancestral relationships, a key aspect on which the inference of the IBD segment depends. For instance, large-scale whole genome sequencing studies reveal that rare variants account for the majority of all segregating sites (MalariaGEN et al., 2021; Taliun et al., 2021), which contain crucial information for deciphering recent evolutionary history. However, rare variants are typically not utilized for two main reasons: (1) rare genotype calls are more prone to genotyping or phasing errors; (2) rare genotypes are very sparsely distributed across many sites and are less informative per site (S. R. Browning & Browning, 2020; Schaffner et al., 2018). As a result, including rare variation may increase IBD inference time due to data sparsity and cause reduced accuracy in detected IBD segments due to genotyping/phasing errors. However, with increasing use of high-depth sequencing and single-cell sequencing, the quality issues associated with rare variants are likely to diminish. To address the computational burden of rare variants, our ongoing work aims to improve the computational efficiency of the `hmmIBD` algorithm and implement a highly parallelizable alternative *hmmibd-rs* to reduce runtime. In addition, we are in the process of developing a tabular encoding strategy for rare variant genotypes and related algorithms for efficient rare allele-sharing analyses. Although our simulation analysis indicates that including rare variants seems detrimental (by increasing error rates) for certain IBD callers, this observation could be skewed due to the small sample size simulated and limited number of cut-off values tested, underlining the need for further analysis with a large sample size. Indels are another significant source of underutilized genetic variations in the *Pf* genome, with abundance on par with that of biallelic SNPs in the MalariaGEN *Pf*7 data set (MalariaGEN et al., 2023). These indels are, in part, the result of microplasticity related to the high AT content (up to 90% in non-coding regions) in *Pf* genomes (Hamilton et

al., 2017). Using these variants could also increase the marker density to infer IBD segments. However, additional research is necessary to determine whether the inclusion of these variants can reduce uncertainty in the inference of IBD for *Pf* or introduce more bias due to challenges such as sequence-read mapping.

Genotyping of highly diverse multiallelic loci, including microhaplotypes, is facilitated by amplicon sequencing panels such as AMPLseq (LaVerriere et al., 2022), Paragon (Tessema et al., 2022) and SpotMalaria V2 (Jacob et al., 2021), shown to be informative in inferring IBD-based relatedness parameters (Taylor et al., 2019). In comparison to the whole genome phased haploid genotype data we analyzed here, amplicon panels are usually comprised of 100 - 200 loci approximately 100 - 400 bp in length, and represent only a small fraction ($< 0.5\%$) of the genome. We expect that genotype data over these limited genomic intervals might not provide sufficient data to accurately infer with high confidence the distribution of IBD segment lengths and the positional enrichment. As a result, amplicon sequencing data might not be suitable for a comprehensive scan for selection signals or quality-sensitive IBD-based analysis, such as estimating effective population size. Since these panels usually include highly diverse regions, antigen genes and drug resistance genes, IBD segment estimates and downstream analysis based on these panels can potentially be biased from the effects of non-neutral evolution. This includes the impact of positive selection, which has been investigated in our previous study (Guo et al., 2024), as well as balancing selection, the effects of which are currently under scrutiny as part of our ongoing research efforts. However, amplicon sequencing benefits from lower parasite DNA requirements, and, depending on the panel, read depth necessary to estimate multiplicity of infection, and can provide important information on genetic relatedness via tools like `Dcifer` (Gerlovina et al., 2022)), especially when WGS data or reliable deconvolution are infeasible.

Although a substantial portion of this study concentrated on *Pf*, the main findings and methodologies may be relevant to high-recombining species beyond *Pf*. For instance, in regions with intermediate and low malaria transmission, the incidence of *Pf* has markedly decreased, allowing other species, such as *Plasmodium vivax*, to become predominant (World Health Organization, 2023). Clinical malaria caused by simian *Plasmodium* species, e.g., *Plasmodium knowlesi*, have also increased in some geographic areas where human *Plasmodium* species have declined (Amir et al., 2018). We expect that the performance of IBD callers will be similar in other *Plasmodium* species,

given their likely comparable high recombination rates (Bright et al., 2014; Ibrahim et al., 2023); however, generalization would need further exploration as part of future work, considering variations in evolutionary histories and parasite biology (Escalante et al., 2004; Lee et al., 2011; Loy et al., 2017). For other species that exhibit high recombination, but have markedly distinct evolutionary parameters from *Plasmodium*, the benchmarking framework established in this study can easily be tailored to prioritize and evaluate IBD detection methods in a context-specific manner.

While we have conducted numerous simulation analyses complemented by carefully designed validation studies, our work is subject to a few caveats: (1) Our simulations did not explicitly incorporate inbreeding within the complicated life cycle of the parasite (Anderson et al., 2000), except for the increased inbreeding potential due to the reduced population size in the single population model. Inbreeding can be pervasive, especially in low transmission settings, leading to a change in the length distribution of IBD toward longer segments and a potentially reduced marker density. (2) Our optimization is based on simple accuracy metrics and only focuses on a subset of parameters, to allow faster iteration over different values. Investigating a larger parameter space, including genotyping error rate and higher-level accuracy metrics may generate different optimal values and further improve IBD-based downstream estimates. (3) We assume a constant recombination rate and mutation rate as static genomic/population parameters, rather than traits capable of evolving over time. If this assumption proves to be inaccurate, such as with recombination rates that vary between individuals and populations (Smukowski & Noor, 2011), a more complex benchmarking framework will be required.

In this study, we evaluated the performance of existing IBD segment detection methods in analyzing genomic data of malaria parasite *Pf*, which is characterized with high recombination rates and low marker densities. Our findings underscore that a high recombination rate, relative to the mutation rate, can compromise the accuracy of detected IBD segments when using methods originally calibrated for the human genome, characterized by a significantly lower recombination rate and a higher marker density (per genetic unit). The accuracy of IBD detection can be improved by parameter optimization via grid search techniques. We advocate for a context-specific evaluation of IBD detection methods when applying them to untested species. Specifically for *Pf*, our research indicates that hidden Markov model-based probabilistic methods, such as `hmmIBD` produce less biased IBD estimates leading to more accurate downstream inferences. This is especially important

for analyses that heavily rely on the accuracy of detected IBD segments, such as N_e inference.

Methods

Simulation overview

We used population genetic simulations to allow the generation of (1) ground truth, including true IBD segments, true sites under positive selection, true trajectory of population size, and true subpopulation assignments (population structure), and (2) inferred patterns, including IBD inferred from phased genotype data via different IBD callers and IBD-based downstream inferences of N_e, positive selection, and population structure. By comparing inferred patterns with ground truth, we calculated metrics at the IBD segment level and the IBD-based downstream estimate level (high level) for benchmarking and optimizing various IBD detection methods for *Pf* genomes. As described in our accompanying work (Guo et al., 2024), we combined the flexible forward simulator `SLiM` (Haller & Messer, 2019; Haller et al., 2019) and the efficient coalescent simulator `msprime` (Baumdicker et al., 2022) to simulate genomes similar to *Pf* reflecting the high recombination rate, strong positive selection, and population size shrinkage due to malaria reduction (Supplementary Figure 1).

Each simulation consists of four main steps: (1) forward simulation via `SLiM`. In this phase, we used `SLiM` to explicitly model each individual of hermaphrodite diploids in the entire population, which allows the simulation of complex positive selection (of different strength, duration, and starting allele frequency) as well as migration (population structure) and inbreeding. With tree sequence recording toggled on, the forward simulation generates a partially coalesced genealogy up to the first generation of the forward simulation time window (base population), which contains information on each individual of each involved generation. To save computation memory and time, neutral mutations were not simulated in this stage, and only the non-neutral sites under positive selection were generated and recorded within the tree sequence. (2) Genealogy simplification via *pyslim* (Haller et al., 2019). The tree sequence information from the previous phase contains redundant information because many ancestral and present-day individuals unrelated to finally sampled present-day individuals were present, the latter of which is usually much smaller than the population size. The simplification step removes this redundant information and prepares a minimal tree

sequence for the next stage. (3) Coalescent simulation via `msprime`. The genealogical trees from the previous step may not have already fully coalesced, as the grand common ancestor for each marginal tree may not yet included. The step starts from the "root" nodes of the trees and completes them backward in time, generating fully coalesced genealogical trees carrying the non-neutral mutation information. The fully coalesced tree sequence can be used to generate true IBD (described below) and allows for the addition of neutral mutations. (4) Simulation of neutral mutations via `msprime`. The above-mentioned non-neutral mutation only occurs at a few sites, and the majority of genotype data needs to be generated from neutral mutation simulations. `msprime` can efficiently add neutral mutations to existing genealogical trees mainly with encoded ancestral information. The resulting mutation information (neutral and non-neutral) can be easily converted to phased genotype data in variant calling format (VCF). A detailed implementation of the simulation can be found in a dedicated GitHub repository (https://github.com/bguo068/bmibdcaller_simulations).

In these simulations, we assumed constant recombination rates over the genome, such as 6.67×10^{-7} per base pair per generation for *Pf* (Amambua-Ngwa et al., 2019; Conway et al., 1999) and 1.0×10^{-8} for human (Kong et al., 2002), and a mutation rate of 1.0×10^{-8} for both *Pf* (Bopp et al., 2013; Camponovo et al., 2023) and human unless otherwise specified. The values of parameters used to model effective/census population size changes, population structure, and positive selection are described in the following section or in our related publication (Guo et al., 2024).

Simulation demographic models

We used three different demographic models in the simulations, including the single-population model, the multiple-population model, and the UK European human population models (Zhou et al., 2020).

Single- and multiple-population models have been described in our accompanying work (Guo et al., 2024). Briefly, the single population model describes a population with a shrinking population size from 10,000 200 generations ago to 1,000 in the present-day generation in an exponentially decreasing pattern, to mimic malaria reduction in settings where transmission was previously high and now is low, such as Southeast Asia. This model was used to benchmark IBD detection methods at the IBD segment level and at the (high) level of downstream estimation, including selection signal detection and N_e estimation. The multiple-population model describes a structured population with

accelerated migration in recent times, in which an ancestral population with a size of 10,000 splits into five contemporary subpopulations each with a constant size of 10,000. This model is mainly used to benchmark IBD calling methods via IBD network-based community detection. More details of the two models were provided in our related work (Guo et al., 2024) and the GitHub repository (see Code availability).

The demographic model of the human population in the United Kingdom (UK human model) is similar to the simulation model used in (Zhou et al., 2020) to evaluate the performance of the IBD caller `hap-IBD` (Zhou et al., 2020). With the UK human demographic model, the population goes through a bottleneck event from a constant size of 10,000 to 3,000 that occurred 5,000 generations ago and then grows at a rate of 1.4% per generation beginning 300 generations ago, and 25% beginning 10 generations ago (Zhou et al., 2020). We simulated 14 chromosomes with a size of 60 cM for each genome so that the total genome size is smaller than the actual human genome size to reduce simulation time. We used this demographic model as a control to detect IBD segments with human-oriented callers, which can help validate our IBD accuracy evaluation pipeline. We also used this model to test whether the demographic model plays a key role in the performance of IBD callers by replacing the human recombination rate with that of *Pf*.

To separate the effects of positive selection from that of demographic models and recombination rates, we mostly simulated neutral genomes by setting the selection coefficient s to 0 for the above models, except in the case where we need to benchmark IBD callers for detecting positive selection signals.

Positive selection simulation

To evaluate the performance of IBD callers via IBD-based selection signal detection, we simulated positive selection within the single-population model with selection coefficients s to 0.2 with a single origin starting 80 generations ago. 14 chromosomes with a size of 100 cM were simulated independently, each with a selected site at 33.3 centimorgan from their left ends. For selection simulation, we conditioned on the establishment of the selective sweeps, that is, the allele under selection should not be lost at the present-day generation. If lost, the simulation was rerun for a maximum of 100 times until the selective sweep is established.

IBD calling and default parameters

We generated true IBD from simulated genealogical trees encoded in the tree sequence format as described in the above Simulation overview subsection. The true IBD detection algorithm `tskibd` has been described and validated in our accompanying work (Guo et al., 2024). Briefly, we sampled the local/marginal trees from the tree sequence along a chromosome from left to right (smaller coordinates to larger coordinates) with a step size of 0.01 cM and tracked the changes in the most recent common ancestor (MRCA) for each pair of present-day nodes (leaf nodes). At each sampling point, for each pair of leave nodes, if MRCA changes, the shared ancestral segment breaks; if MRCA stays the same, the shared ancestral segment elongates. The ancestral segment generated from any of the breaks or those present when the right end of the chromosome is reached are candidates for the IBD segment. We report a shared ancestral segment as an IBD segment only if the length of the segment is not shorter than the threshold, such as 2 centimorgans.

For inferred IBD, we used phased genotype data as input. Only biallelic sites with a minor allele frequency no less than 0.01 (common SNPs) were included, unless otherwise stated. When needed, a genetic map is generated based on the constant recombination rate as specified in the corresponding simulations. For IBD detection methods designed for diploids, we converted each haploid to a pseudohomozygous diploid, for example, a genotype call for a site and sample "0" will be converted to "0|0". The resulting IBD segments of each pair of pseudo-homozygous diploids (A and B) will have redundant information due to the 100% runs of homozygosity. Each IBD segment will be shared 4 times due to different combinations of two haplotypes in A (A / hap1 and A / hap2) with those in B (B / hap1 and B / hap2). We only keep one pair A / hap1 - B / hap1 and remove other combinations, including A / hap2 - B / hap1, A / hap2 - B / hap2, and A / hap1 - B / hap2.

As each IBD detection method provides multiple tunable parameters from the command-line interface or the programming interface from a Python or R script, we provided the detailed values we used for them in Supplementary Table 1 for two scenarios – default and optimized. For the default scenarios, the parameters mostly follow the original documentation for each caller. The exceptions are parameters that need to be consistent across different IBD callers to allow benchmarking, including the minimum length of the reported IBD segment and the minimum minor allele frequency

of the SNPs included. Details of how the optimized parameter values were obtained are described below.

Benchmarking metrics at the IBD segment level

We used a few metrics to benchmark IBD methods by directly comparing the result at the IBD segment level or using their simple aggregates, including the false negative rate and false positive rate, pairwise total IBD and population-level total IBD per length bin.

We calculated the false positive rate and the false negative rate following the definition used in the work of Zhou *et al.* (Zhou et al., 2020). Rates were first calculated per segment via segment-overlapping analysis; then, they were averaged over all segments of the same length bin. (1) The false positive rate per segment is defined as the proportion of a given IBD segment of some genome pair (g_1, g_2) from the inferred set (for example, IBD called via `hmmIBD`) that is not covered by any IBD segment from the truth set (generated by `tskibd`), for the same genome pair (g_1, g_2). Similarly, the false negative rate per segment is defined as the proportion of a given IBD segment of some genome pair (g_1, g_2) from the truth set that is not covered by any IBD segment from the inferred set, for the same genome pair (g_1, g_2). For a given inferred set and truth set, each inferred IBD segment is associated with a per-segment false positive rate, whereas each true IBD segment is associated with a per-segment false negative rate. (2) The average false positive rate is calculated as the average per-segment false positive rates for all inferred IBD segments the length of which falls in a certain range (length bin); the average false negative rate is calculated as the average per-segment false positive rates for all true segments of a length bin. We calculated the average rates for the following length bins: [3-4), [4-6), [6-10), [10-18), [18, inf), similar to Zhou *et al.*'s method.

To capture genome-wide bias instead of per segment length bin, we also calculated per-(genome-)pair genome-wide rates and their averages across all genome pairs. The per-pair genome-wide false positive rate is the ratio of two sums: (1) the numerator sum is the total length of parts of all inferred IBD segments of a certain genome pair that are not covered by any true IBD segments of the same genome pair; (2) the denominator sum is the total length of all inferred IBD segments of a certain genome pair (pairwise total inferred IBD). The per-pair genome-wide false negative rate is defined in a similar way as the percentage of pairwise total true IBD that are not overlapped by any inferred segments of the same pair. We then obtain the aggregate metric by averaging these

rates for all genome pairs.

We calculated and compared pairwise total IBD from truth *versus* inferred set as it is a useful metric to estimate genetic relatedness and build IBD sharing networks for inference of population structure. It was calculated as the sum of the lengths of all inferred or true IBD segments of each genome pair.

Given that the N_e estimator `IBDNe` internally utilizes quantities of population-level total IBD of different length bins, we also calculated these quantities to better examine IBD accuracy that is closely relevant to N_e inference. We defined a list of non-overlapping length bins of a width of 0.05 centimorgan that cover all possible lengths. For instance, for the length bin [2.00 - 2.05), we sum up the length of all IBD segments, of which the segment's length falls into this length bin, from any genome pair of the whole population.

The calculation of these metrics has been implemented in the `ishare/ibdutils` tools (https://github.com/bguo068/ishare).

IBD caller parameter optimization

We optimized key IBD caller-specific parameters by iterating each parameter over the list of discrete values, or two or more parameters over a grid of discrete values. Optimized parameters were mainly those related to marker density for IBD callers `hap-IBD`, `phased IBD`, `Refined IBD`, and `isoRelate`. Other parameters were searched, if they can potentially have a great impact on the quality of the detected IBD. The optimal values for an explored parameter are determined by the length-bin-specific or genome-wide error rates (FN and FP) for detected IBD segments as defined above. The parameter value or combination of values that generate lower and generally balanced error rates were selected as optimal values. The parameter searched, the value lists explored and the optimal values selected are summarized in Supplementary Table 1 and Supplementary Table 2.

When the optimized values are different across different demographic models, we used the ones optimized for the single-population model for downstream analyses. We provided detailed heatmaps of error rates for all demographic models tested (see the Supplementary Data).

Benchmarking with IBD-based downstream analysis

At a higher level, we benchmarked different IBD callers by comparing downstream estimates based on true IBD sets *versus* inferred IBD sets. These IBD-based estimates include positive selection scan, N_e estimates, and population structure inference via the community detection algorithm InfoMap.

We scanned positive selection signals using the IBD-based thresholding method followed by validation with integrated haplotype score-based statistics X_{iHS} as previously described (Guo et al., 2024). Briefly, we first calculated the IBD coverage defined as the number of IBD segments shared by any genome pair that cover a given sampling point, which is evenly spaced along a chromosome with a step size of 0.1 centimorgan. For each chromosome, we defined the peak identification threshold as the 5% trimmed mean plus two standard deviations of IBD coverage. We defined a positive selection candidate genomic region as a core region and two extension regions. The core regions are regions where IBD coverage is above the threshold. For each core region, two extension regions were determined by extending the core region to both sides until the IBD coverage reaches the chromosome median. To help remove a false positive selection signal, we calculated X_{iHS} statistics for each SNP and designated a candidate region as a validated region if there is at least one significant SNP (X_{iHS} based p-value < 0.05) within the region. To calculate X_{iHS} statistics, we first calculated the raw integrated Haplotype Score (iHS) and then standardized them within their allele frequency bin using the *scikit-allel* Python package. We squared the standardized iHS to obtain the X_{iHS} statistics. A p-value is calculated for the X_{iHS} statistic of each SNP based on a chi-square distribution with one degree of freedom.

We inferred the trajectory of effective population size, N_e, for the last 100 generations using `IBDNe`. As this method uses IBD shared by diploid individuals as input, we converted each haploid genome to pseudo-homozygous diploid individuals. We infer the trajectory N_e using most of the default parameters except for setting the minregion parameter to 10 cM to allow the inclusion of short contigs in the analysis. The final estimates are scaled by 0.25 to compensate for the haploid-to-diploid conversion. By default, we used the value of 2 cM for the mincm parameter to only include IBD segments $\geq$ 2 cM for inferring N_e trajectory; as indicated in the Results section, we also set mincm to 4 to test whether excluding short IBD segments can improve the accuracy of N_e estimates. As the `IBDNe` algorithm does not work well when IBD segments shared by close relatives

are included in the input, we followed the procedure described in the original work by Browning *et al.* (S. R. Browning & Browning, 2015). For simulated data, we utilized the TRMCA information of the true IBD segments to filter the IBD segments before calling `IBDNe`. For true IBD segments (generated by `tskibd`), we excluded IBD segments with TMRCA < 1.5 ; for inferred IBD segments (called by `hmmIBD`, `hap-IBD`, `isoRelate`, `Refined IBD`, `phased IBD`), we remove (inferred) segments that overlap with any true IBD segment with TMRCA < 1.5 shared by the same genome pair. For empirical data where true IBD and TMRCA are not available, we pruned highly related isolates by iteratively removing the genome that has the highest number of close relatives defined by pairwise total IBD > 0.5 of genome size until no close relatives are present in the remaining subgroup (Guo et al., 2024).

For population structure inference, we first built the pairwise total IBD matrix, each element being the total IBD for a pair of genomes. The matrix was then squared and used as a weighted adjacency matrix to construct an IBD-sharing network. We then ran the InfoMap algorithm to infer community membership. Genomes assigned the same membership are inferred to be of the same subpopulation. We noted that the square transformation of the original total IBD matrix can help better capture the simulated, underlying population structure. For empirical data, we excluded IBD segments short than 4 cM when calculating the total IBD matrix to help reduce noise due to false positives and set each element with a value < 5 to zero in the unsquared IBD matrix to decrease the density of the IBD matrix.

Before all these high-level benchmarking analyses, we pruned highly related samples as mentioned above. As IBD-based estimates can be biased by strong positive selection in empirical data, we removed high IBD sharing peak regions as previously described (Guo et al., 2024) before any downstream analyses.

Processing empirical data sets

We constructed empirical data sets for validation using genotype data from whole genome sequencing samples from the MalariaGEN *Pf*7 database (MalariaGEN et al., 2023). We used the *malariagen_data* Python package to download high-quality monoclonal samples that pass QC. Monoclonal samples are determined by $F_{\mathrm{ws}} > 0.95$ (F_{ws} table available from the malariaGEN website). Quality control labels were extracted directly from the metadata (provided in the *malariagen_data* pack-

age). We also performed an initial site filtering before downloading the large genotype data by removing multiallelic sites, non-SNP variants, and SNPs with missingness > 0.8 or with minor allele count < 10 (ultra rare alleles).

We then generated haploid genomes (phased genotype data) using the dominant allele from genotype calls. The dominant allele of each genotype call is determined by the per-sample allele depth (AD) fields. For each sample and site, the allele supported by 90% of total reads (total AD values) in a genotype call with at least 5 total reads was used as the dominant allele. Genotype calls without dominant alleles were marked as missing; those with dominant alleles were replaced with a phased genotype homozygous for the dominant allele. The genotype data were further filtered by sample missingness and SNP minor allele frequency and missingness. The resulting genotype data had per-SNP and per-sample missingness <0.1 and minor allele frequency ≥ 0.01. Genotypes based on dominant alleles were further imputed without a panel using Beagle 5.1 (B. L. Browning et al., 2018). These processing steps generated phased, imputed, pseudo-homozygous diploid genotype data, ready for IBD detection.

We constructed two different data sets, that is, the "single" population data set and the "structured, multiple" population data set by subsampling the above haploid genomes according to sampling time and location. For each data set, we set the time window to 2-3 years to reduce the sample time heterogeneity, and then shifted the window within all possible sampling years and chose one that maximized the sample size. For the "single" population model, we further restricted the sample locations to a relatively small geographic region, such as eastern Southeast Asia, as the data set was used for N_e estimation which assumes a homogeneous population. For the "multiple" population data set, we included samples from different continental or subcontinental regions using "Population" labels from the meta-information table provided with the MalariaGEN *Pf*7 database. To make the sample size of each "population" more balanced, we set a maximum number of samples of 300. Populations with samples larger than 300 are subsampled to a size of 300; Populations with a size smaller than 100 were not included in the multiple population data set. The details of the sampling location and time information were summarized in Supplementary Table 3, Supplementary Table 4 and Supplementary Table 5.

The detail in the preparation and analysis of the empirical data sets can be found at (https://github.com/bguo068/bmibdcaller_empirical).

Measuring computational runtime and memory usage

The genomes were simulated with $N_0 = 1{,}000$ and s = 0.0 under the single population model. The runtime and maximum memory usage were measured using the GNU time 1.7 utility. To allow a more appropriate comparison, we ensured: (1) the time used for data pre- and post-processing was excluded; (2) memory resource allocation was capped at 30 gigabytes per IBD call; (3) input genotype data included only common SNPs with minor allele frequency ≥ 0.01; (4) the minimum reported IBD segment length was set to 2.0 cM.

Code availability

Custom tools or scripts were provided in the following GitHub repositories: (1) bmibdcaller_simulations: a Nextflow pipeline to benchmark different IBD detection methods and optimize IBD caller-specific parameters by simulating *Plasmodium falciparum*-like genomes and using true IBD (https://github.com/bguo068/bmibdcaller_simulations). (2) bmibdcaller_empirical: a Nextflow pipeline to benchmark IBD callers with empirical data by comparing IBD-based estimates with expected patterns (https://github.com/bguo068/bmibdcaller_empirical). (3) ishare/ibdutils: a Rust crate and command line tools designed to facilitate the analysis of rare-variant sharing and identity-by-descent (IBD) sharing (used here mainly for fast IBD segment overlapping analysis) (https://github.com/bguo068/ishare, command line tool `ibdutils`).

Data availability

All empirical data used were publicly available from MalariaGEN *Pf*7 (MalariaGEN et al., 2023).

Acknowledgement

This publication uses MalariaGEN data as described in "Pf7: an open dataset of Plasmodium falciparum genome variation in 20,000 worldwide samples' MalariaGEN et al, Wellcome Open Research 2023, 8:22 https://doi.org/10.12688/wellcomeopenres.18681.1". This work was supported by NIH 1R01AI145852 granted to ST-H and TDO by the U.S. National Institutes of Health.

References

Al-Asadi, H., Petkova, D., Stephens, M., & Novembre, J. (2019). Estimating recent migration and population-size surfaces (M. DeGiorgio, Ed.). *PLOS Genetics*, *15*(1), e1007908–e1007908. https://doi.org/10.1371/journal.pgen.1007908

Amambua-Ngwa, A., Amenga-Etego, L., Kamau, E., Amato, R., Ghansah, A., Golassa, L., Randrianarivelojosia, M., Ishengoma, D., Apinjoh, T., Maïga-Ascofaré, O., Andagalu, B., Yavo, W., Bouyou-Akotet, M., Kolapo, O., Mane, K., Worwui, A., Jeffries, D., Simpson, V., D'Alessandro, U., … Djimde, A. A. (2019). Major subpopulations of Plasmodium falciparum in sub-Saharan Africa. *Science (New York, N.Y.)*, *365*(6455), 813–816. https://doi.org/10.1126/science.aav5427

Amir, A., Cheong, F. W., de Silva, J. R., Liew, J. W. K., & Lau, Y. L. (2018). Plasmodium knowlesi malaria: Current research perspectives. *Infection and Drug Resistance*, *11*, 1145–1155. https://doi.org/10.2147/IDR.S148664

Anderson, T. J. C., Paul, R. E. L., Donnelly, C. A., & Day, K. P. (2000). Do malaria parasites mate non-randomly in the mosquito midgut? *Genetics Research*, *75*(3), 285–296. https://doi.org/10.1017/S0016672300004481

Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A. P., Tsambos, G., Zhu, S., Eldon, B., Ellerman, E. C., Galloway, J. G., Gladstein, A. L., Gorjanc, G., Guo, B., Jeffery, B., Kretzschumar, W. W., Lohse, K., Matschiner, M., Nelson, D., Pope, N. S., … Kelleher, J. (2022). Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, *220*(3), iyab229. https://doi.org/10.1093/genetics/iyab229

Bopp, S. E. R., Manary, M. J., Bright, A. T., Johnston, G. L., Dharia, N. V., Luna, F. L., McCormack, S., Plouffe, D., McNamara, C. W., Walker, J. R., Fidock, D. A., Denchi, E. L., & Winzeler, E. A. (2013). Mitotic Evolution of Plasmodium falciparum Shows a Stable Core Genome but Recombination in Antigen Families. *PLOS Genetics*, *9*(2), e1003293. https://doi.org/10.1371/journal.pgen.1003293

Bright, A. T., Manary, M. J., Tewhey, R., Arango, E. M., Wang, T., Schork, N. J., Yanow, S. K., & Winzeler, E. A. (2014). A high resolution case study of a patient with recurrent Plasmodium

vivax infections shows that relapses were caused by meiotic siblings. *PLoS neglected tropical diseases*, *8*(6), e2882. https://doi.org/10.1371/journal.pntd.0002882

Brooks, D. R., Wang, P., Read, M., Watkins, W. M., Sims, P. F., & Hyde, J. E. (1994). Sequence variation of the hydroxymethyldihydropterin pyrophosphokinase: Dihydropteroate synthase gene in lines of the human malaria parasite, Plasmodium falciparum, with differing resistance to sulfadoxine. *European journal of biochemistry*, *224*(2), 397–405. https://doi.org/10.1111/j.1432-1033.1994.00397.x

Browning, B. L., & Browning, S. R. (2011). A Fast, Powerful Method for Detecting Identity by Descent. *The American Journal of Human Genetics*, *88*(2), 173–182. https://doi.org/10.1016/j.ajhg.2011.01.010

Browning, B. L., & Browning, S. R. (2013). Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data. *Genetics*, *194*(2), 459–471. https://doi.org/10.1534/genetics.113.150029

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics*, *103*(3), 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Browning, S. R., & Browning, B. L. (2012). Identity by Descent Between Distant Relatives: Detection and Applications. *Annual Review of Genetics*, *46*(1), 617–633. https://doi.org/10.1146/annurev-genet-110711-155534

Browning, S. R., & Browning, B. L. (2015). Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. *American Journal of Human Genetics*, *97*(3), 404–418. https://doi.org/10.1016/j.ajhg.2015.07.012

Browning, S. R., & Browning, B. L. (2020). Probabilistic Estimation of Identity by Descent Segment Endpoints and Detection of Recent Selection. *The American Journal of Human Genetics*, *107*(5), 895–910. https://doi.org/10.1016/j.ajhg.2020.09.010

Campbell, C. D., Chong, J. X., Malig, M., Ko, A., Dumont, B. L., Han, L., Vives, L., O'Roak, B. J., Sudmant, P. H., Shendure, J., Abney, M., Ober, C., & Eichler, E. E. (2012). Estimating the human mutation rate using autozygosity in a founder population. *Nature Genetics*, *44*(11), 1277–1281. https://doi.org/10.1038/ng.2418

Camponovo, F., Buckee, C. O., & Taylor, A. R. (2023). Measurably recombining malaria parasites. *Trends in Parasitology*, *39*(1), 17–25. https://doi.org/10.1016/j.pt.2022.11.002

Churcher, T. S., Cohen, J. M., Novotny, J., Ntshalintshali, N., Kunene, S., & Cauchemez, S. (2014). Public health. Measuring the path toward malaria elimination. *Science (New York, N.Y.)*, *344*(6189), 1230–1232. https://doi.org/10.1126/science.1251449

Conway, D. J., Roper, C., Oduola, A. M. J., Arnot, D. E., Kremsner, P. G., Grobusch, M. P., Curtis, C. F., & Greenwood, B. M. (1999). High recombination rate in natural populations of Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(8), 4506–4511.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, complex systems*, *1695*(5), 1–9.

Early, A. M., Camponovo, F., Pelleau, S., Cerqueira, G. C., Lazrek, Y., Volney, B., Carrasquilla, M., de Thoisy, B., Buckee, C. O., Childs, L. M., Musset, L., & Neafsey, D. E. (2022). Declines in prevalence alter the optimal level of sexual investment for the malaria parasite Plasmodium falciparum. *Proceedings of the National Academy of Sciences*, *119*(30), e2122165119. https://doi.org/10.1073/pnas.2122165119

Escalante, A. A., Cornejo, O. E., Rojas, A., Udhayakumar, V., & Lal, A. A. (2004). Assessing the effect of natural selection in malaria parasites. *Trends in Parasitology*, *20*(8), 388–395. https://doi.org/10.1016/j.pt.2004.06.002

Freyman, W. A., McManus, K. F., Shringarpure, S. S., Jewett, E. M., Bryc, K., The 23 and Me Research Team, & Auton, A. (2021). Fast and Robust Identity-by-Descent Inference with the Templated Positional Burrows–Wheeler Transform. *Molecular Biology and Evolution*, *38*(5), 2131–2151. https://doi.org/10.1093/molbev/msaa328

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., … Barrell, B. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*. https://doi.org/10.1038/nature01097

Gerlovina, I., Gerlovin, B., Rodríguez-Barraquer, I., & Greenhouse, B. (2022). Dcifer: An IBD-based method to calculate genetic distance between polyclonal infections (S. Browning, Ed.). *Genetics*, *222*(2), iyac126. https://doi.org/10.1093/genetics/iyac126

Guo, B., Borda, V., Laboulaye, R., Spring, M. D., Wojnarski, M., Vesely, B. A., Silva, J. C., Waters, N. C., O'Connor, T. D., & Takala-Harrison, S. (2024). Strong positive selection biases identity-by-descent-based inferences of recent demography and population structure in Plasmodium falciparum. *Nature Communications*, *15*(1), 2499. https://doi.org/10.1038/s41467-024-46659-0

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genetics*, *5*(10), e1000695. https://doi.org/10.1371/journal.pgen.1000695

Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W., & Ralph, P. L. (2019). Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources*, *19*(2), 552–566. https://doi.org/10.1111/1755-0998.12968

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, *36*(3), 632–637. https://doi.org/10.1093/molbev/msy228

Hamilton, W. L., Claessens, A., Otto, T. D., Kekre, M., Fairhurst, R. M., Rayner, J. C., & Kwiatkowski, D. (2017). Extreme mutation bias and high AT content in Plasmodium falciparum. *Nucleic Acids Research*, *45*(4), 1889–1901. https://doi.org/10.1093/nar/gkw1259

Harris, D. N., Kessler, M. D., Shetty, A. C., Weeks, D. E., Minster, R. L., Browning, S., Cochrane, E. E., Deka, R., Hawley, N. L., Reupena, M. S., Naseri, T., Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed Population Genetics Working Group, McGarvey, S. T., & O'Connor, T. D. (2020). Evolutionary history of modern Samoans. *Proceedings of the National Academy of Sciences*, *117*(17), 9458–9465. https://doi.org/10.1073/pnas.1913157117

Henden, L., Lee, S., Mueller, I., Barry, A., & Bahlo, M. (2018). Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. *PLoS genetics*, *14*(5), e1007279–e1007279. https://doi.org/10.1371/journal.pgen.1007279

Huber, J. H., Johnston, G. L., Greenhouse, B., Smith, D. L., & Perkins, T. A. (2016). Quantitative, model-based estimates of variability in the generation and serial intervals of Plasmodium falciparum malaria. *Malaria Journal*, *15*(1), 490. https://doi.org/10.1186/s12936-016-1537-6

Ibrahim, A., Manko, E., Dombrowski, J. G., Campos, M., Benavente, E. D., Nolder, D., Sutherland, C. J., Nosten, F., Fernandez, D., Vélez-Tobón, G., Castaño, A. T., Aguiar, A. C. C., Pereira, D. B., Da Silva Santos, S., Suarez-Mutis, M., Di Santi, S. M., Regina De Souza Baptista, A., Dantas Machado, R. L., Marinho, C. R., … Campino, S. (2023). Population-based genomic study of Plasmodium vivax malaria in seven Brazilian states and across South America. *The Lancet Regional Health - Americas*, *18*, 100420. https://doi.org/10.1016/j.lana.2022.100420

Ishigohoka, J., & Liedvogel, M. (2024, February). High-recombining genomic regions affect demography inference. https://doi.org/10.1101/2024.02.05.579015

Jacob, C. G., Thuy-Nhien, N., Mayxay, M., Maude, R. J., Quang, H. H., Hongvanthong, B., Vanisaveth, V., Ngo Duc, T., Rekol, H., van der Pluijm, R., von Seidlein, L., Fairhurst, R., Nosten, F., Hossain, M. A., Park, N., Goodwin, S., Ringwald, P., Chindavongsa, K., Newton, P., … Miotto, O. (2021). Genetic surveillance in the Greater Mekong subregion and South Asia to support malaria control and elimination. *eLife*, *10*, e62997. https://doi.org/10.7554/eLife.62997

Jiang, H., Li, N., Gopalan, V., Zilversmit, M. M., Varma, S., Nagarajan, V., Li, J., Mu, J., Hayton, K., Henschen, B., Yi, M., Stephens, R., McVean, G., Awadalla, P., Wellems, T. E., & Su, X. Z. (2011). High recombination rates and hotspots in a Plasmodium falciparum genetic cross. *Genome Biology*. https://doi.org/10.1186/gb-2011-12-4-r33

Joy, D. A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A. U., Ho, M., Wang, A., White, N. J., Suh, E., Beerli, P., & Su, X.-z. (2003). Early Origin and Recent Expansion of Plasmodium falciparum. *Science*, *300*(5617), 318–321. https://doi.org/10.1126/science.1081449

Kelleher, J., Thornton, K. R., Ashander, J., & Ralph, P. L. (2018). Efficient pedigree recording for fast population genetics simulation. *PLoS Computational Biology*, *14*(11), 1–21. https://doi.org/10.1371/journal.pcbi.1006581

Kelleher, J., Wong, Y., Wohns, A. W., Fadil, C., Albers, P. K., & McVean, G. (2019). Inferring whole-genome histories in large population datasets. *Nature Genetics*, *51*(9), 1330–1338. https://doi.org/10.1038/s41588-019-0483-y

Koenderink, J. B., Kavishe, R. A., Rijpma, S. R., & Russel, F. G. M. (2010). The ABCs of multidrug resistance in malaria. *Trends in Parasitology*, *26*(9), 440–446. https://doi.org/10.1016/j.pt.2010.05.002

Kong, A., Gudbjartsson, D. F., Sainz, J., Jonsdottir, G. M., Gudjonsson, S. A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., Shlien, A., Palsson, S. T., Frigge, M. L., Thorgeirsson, T. E., Gulcher, J. R., & Stefansson, K. (2002). A high-resolution recombination map of the human genome. *Nature genetics*, *31*(3), 241–247. https://doi.org/10.1038/ng917

LaVerriere, E., Schwabl, P., Carrasquilla, M., Taylor, A. R., Johnson, Z. M., Shieh, M., Panchal, R., Straub, T. J., Kuzma, R., Watson, S., Buckee, C. O., Andrade, C. M., Portugal, S., Crompton, P. D., Traore, B., Rayner, J. C., Corredor, V., James, K., Cox, H., … Neafsey, D. E. (2022). Design and implementation of multiplexed amplicon sequencing panels to serve genomic epidemiology of infectious disease: A malaria case study. *Molecular Ecology Resources*, *22*(6), 2285–2303. https://doi.org/10.1111/1755-0998.13622

Lee, K.-S., Divis, P. C. S., Zakaria, S. K., Matusop, A., Julin, R. A., Conway, D. J., Cox-Singh, J., & Singh, B. (2011). Plasmodium knowlesi: Reservoir Hosts and Tracking the Emergence in Humans and Macaques. *PLoS Pathogens*, *7*(4), e1002015. https://doi.org/10.1371/journal.ppat.1002015

Loy, D. E., Liu, W., Li, Y., Learn, G. H., Plenderleith, L. J., Sundararaman, S. A., Sharp, P. M., & Hahn, B. H. (2017). Out of Africa: Origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax. *International journal for parasitology*, *47*(2-3), 87–97. https://doi.org/10.1016/j.ijpara.2016.05.008

MalariaGEN, Abdel Hamid, M. M., Abdelraheem, M. H., Acheampong, D. O., Ahouidi, A., Ali, M., Almagro-Garcia, J., Amambua-Ngwa, A., Amaratunga, C., Amenga-Etego, L., Andagalu, B., Anderson, T., Andrianaranjaka, V., Aniebo, I., Aninagyei, E., Ansah, F., Ansah, P. O., Apinjoh, T., Arnaldo, P., … Van Der Pluijm, R. W. (2023). Pf7: An open dataset

of Plasmodium falciparum genome variation in 20,000 worldwide samples. *Wellcome Open Research*, *8*, 22. https://doi.org/10.12688/wellcomeopenres.18681.1

MalariaGEN, Ahouidi, A., Ali, M., Almagro-Garcia, J., Amambua-Ngwa, A., Amaratunga, C., Amato, R., Amenga-Etego, L., Andagalu, B., Anderson, T. J. C., Andrianaranjaka, V., Apinjoh, T., Ariani, C., Ashley, E. A., Auburn, S., Awandare, G. A., Ba, H., Baraka, V., Barry, A. E., … Ye, H. (2021). An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide samples. *Wellcome open research*, *6*, 42. https://doi.org/10.12688/wellcomeopenres.16168.2

Martin, R. E., & Kirk, K. (2004). The Malaria Parasite's Chloroquine Resistance Transporter is a Member of the Drug/Metabolite Transporter Superfamily. *Molecular Biology and Evolution*, *21*(10), 1938–1949. https://doi.org/10.1093/molbev/msh205

McDew-White, M., Li, X., Nkhoma, S. C., Nair, S., Cheeseman, I., & Anderson, T. J. C. (2019). Mode and Tempo of Microsatellite Length Change in a Malaria Parasite Mutation Accumulation Experiment (C. Baer, Ed.). *Genome Biology and Evolution*, *11*(7), 1971–1985. https://doi.org/10.1093/gbe/evz140

Mehra, S., Neafsey, D. E., White, M., & Taylor, A. R. (2024, April). Systematic bias in malaria parasite relatedness estimation. https://doi.org/10.1101/2024.04.16.588675

Miles, A., Iqbal, Z., Vauterin, P., Pearson, R., Campino, S., Theron, M., Gould, K., Mead, D., Drury, E., O'Brien, J., Rubio, V. R., MacInnis, B., Mwangi, J., Samarakoon, U., Ranford-Cartwright, L., Ferdig, M., Hayton, K., Su, X.-z., Wellems, T., … Kwiatkowski, D. (2016). Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum. *Genome Research*, *26*(9), 1288–1299. https://doi.org/10.1101/gr.203711.115

Miotto, O., Almagro-Garcia, J., Manske, M., MacInnis, B., Campino, S., Rockett, K. A., Amaratunga, C., Lim, P., Suon, S., Sreng, S., Anderson, J. M., Duong, S., Nguon, C., Chuor, C. M., Saunders, D., Se, Y., Lon, C., Fukuda, M. M., Amenga-Etego, L., … Kwiatkowski, D. P. (2013). Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. *Nature Genetics*, *45*(6), 648–655. https://doi.org/10.1038/ng.2624

Miotto, O., Amato, R., Ashley, E. A., MacInnis, B., Almagro-Garcia, J., Amaratunga, C., Lim, P., Mead, D., Oyola, S. O., Dhorda, M., Imwong, M., Woodrow, C., Manske, M., Stalker, J., Drury, E., Campino, S., Amenga-Etego, L., Thanh, T.-N. N., Tran, H. T., … Kwiatkowski,

D. P. (2015). Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nature Genetics*, *47*(3), 226–234. https://doi.org/10.1038/ng.3189

Morgan, A. P., Brazeau, N. F., Ngasala, B., Mhamilawa, L. E., Denton, M., Msellem, M., Morris, U., Filer, D. L., Aydemir, O., Bailey, J. A., Parr, J. B., Mårtensson, A., Bjorkman, A., & Juliano, J. J. (2020). Falciparum malaria from coastal Tanzania and Zanzibar remains highly connected despite effective control efforts on the archipelago. *Malaria Journal*, *19*(1), 1–14. https://doi.org/10.1186/s12936-020-3137-8

Nait Saada, J., Kalantzis, G., Shyr, D., Cooper, F., Robinson, M., Gusev, A., & Palamara, P. F. (2020). Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations. *Nature Communications*, *11*(1), 6130. https://doi.org/10.1038/s41467-020-19588-x

Neafsey, D. E., Taylor, A. R., & MacInnis, B. L. (2021). Advances and opportunities in malaria population genomics. *Nature Reviews Genetics*, 1–16. https://doi.org/10.1038/s41576-021-00349-5

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *European Physical Journal Special Topics*, *178*(1), 13–23. https://doi.org/10.1140/epjst/e2010-01179-1

Schaffner, S. F., Badiane, A., Khorgade, A., Ndiop, M., Gomis, J., Wong, W., Ndiaye, Y. D., Diedhiou, Y., Thwing, J., Seck, M. C., Early, A., Sy, M., Deme, A., Diallo, M. A., Sy, N., Sene, A., Ndiaye, T., Sow, D., Dieye, B., … Volkman, S. K. (2023). Malaria surveillance reveals parasite relatedness, signatures of selection, and correlates of transmission across Senegal. *Nature Communications*, *14*(1), 7268. https://doi.org/10.1038/s41467-023-43087-4

Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F., & Neafsey, D. E. (2018). HmmIBD: Software to infer pairwise identity by descent between haploid genotypes. *Malaria Journal*, *17*(1), 10–13. https://doi.org/10.1186/s12936-018-2349-7

Shemirani, R., Belbin, G. M., Avery, C. L., Kenny, E. E., Gignoux, C. R., & Ambite, J. L. (2021). Rapid detection of identity-by-descent tracts for mega-scale datasets. *Nature Communications*, *12*(1), 3546. https://doi.org/10.1038/s41467-021-22910-w

Bandiera_abtest: a

Cc_license_type: cc_by

Cg_type: Nature Research Journals

Primary_atype: Research
Subject_term: Genome informatics;Population genetics
Subject_term_id: genome-informatics;population-genetics.

Shetty, A. C., Jacob, C. G., Huang, F., Li, Y., Agrawal, S., Saunders, D. L., Lon, C., Fukuda, M. M., Ringwald, P., Ashley, E. A., Han, K. T., Hlaing, T. M., Nyunt, M. M., Silva, J. C., Stewart, K. E., Plowe, C. V., O'Connor, T. D., Takala-Harrison, S., Noedl, H., … Onyamboko, M. A. (2019). Genomic structure and diversity of Plasmodium falciparum in Southeast Asia reveal recent parasite migration patterns. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-10121-3

Smukowski, C. S., & Noor, M. A. F. (2011). Recombination rate variation in closely related species. *Heredity*, *107*(6), 496–508. https://doi.org/10.1038/hdy.2011.44

Speidel, L., Forest, M., Shi, S., & Myers, S. R. (2019). A method for genome-wide genealogy estimation for thousands of samples. *Nature Genetics*, *51*(9), 1321–1329. https://doi.org/10.1038/s41588-019-0484-x

Su, X.-z., Ferdig, M. T., Huang, Y., Huynh, C. Q., Liu, A., You, J., Wootton, J. C., & Wellems, T. E. (1999). A Genetic Map and Recombination Parameters of the Human Malaria Parasite Plasmodium falciparum. *Science (New York, N.Y.)*, *286*(5443), 1351–1353. https://doi.org/10.1126/science.286.5443.1351

Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., Taliun, S. A. G., Corvelo, A., Gogarten, S. M., Kang, H. M., Pitsillides, A. N., LeFaive, J., Lee, S.-b., Tian, X., Browning, B. L., Das, S., Emde, A.-K., Clarke, W. E., Loesch, D. P., … Abecasis, G. R. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, *590*(7845), 290–299. https://doi.org/10.1038/s41586-021-03205-y

Tang, K., Naseri, A., Wei, Y., Zhang, S., & Zhi, D. (2022). Open-source benchmarking of IBD segment detection methods for biobank-scale cohorts. *GigaScience*, *11*, giac111. https://doi.org/10.1093/gigascience/giac111

Taylor, A. R., Jacob, P. E., Neafsey, D. E., & Buckee, C. O. (2019). Estimating relatedness between malaria parasites. *Genetics*, *212*(4), 1337–1351. https://doi.org/10.1534/genetics.119.302120

Tessema, S. K., Hathaway, N. J., Teyssier, N. B., Murphy, M., Chen, A., Aydemir, O., Duarte, E. M., Simone, W., Colborn, J., Saute, F., Crawford, E., Aide, P., Bailey, J. A., & Greenhouse, B. (2022). Sensitive, Highly Multiplexed Sequencing of Microhaplotypes From the Plasmodium falciparum Heterozygome. *The Journal of Infectious Diseases*, *225*(7), 1227–1237. https://doi.org/10.1093/infdis/jiaa527

Thompson, E. A. (2013). Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. *Genetics*, *194*(2), 301–326. https://doi.org/10.1534/genetics.112.148825

Wesolowski, A., Taylor, A. R., Chang, H.-H., Verity, R., Tessema, S., Bailey, J. A., Alex Perkins, T., Neafsey, D. E., Greenhouse, B., & Buckee, C. O. (2018). Mapping malaria by combining parasite genomic and epidemiologic data. *BMC medicine*, *16*(1), 190–190. https://doi.org/10.1186/s12916-018-1181-9

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

Zhou, Y., Browning, S. R., & Browning, B. L. (2020). A Fast and Simple Method for Detecting Identity-by-Descent Segments in Large-Scale Data. *American Journal of Human Genetics*, *106*(4), 426–437. https://doi.org/10.1016/j.ajhg.2020.02.010