

Conference Info

- American Society of Tropical Medicine and Hygiene
- 2023 Annual Meeting
- October 18-22, 2023
- Hyatt Regency Chicago
- Chicago, Illinois, United States

Contact

- Name: **Bing Guo**
- Email: bing.guo@som.umaryland.edu
- GitHub: <https://github.com/bguo068>

Poster Info

- Abstract Presentation Number: **6155**
- In Poster Session **B (12-1:45pm, Friday Oct 20)**

6155

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

Bing Guo^{1,2}, Michele D. Spring³, Mariusz Wojnarski³, Brian A. Vesely³, Joana C. Silva¹, Norman C. Waters³, Shannon Takala-Harrison^{2,4}, and Timothy D. O'Connor^{1,5}

¹Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA; ²Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD USA; ³Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand; ⁴Co-senior authors

Introduction

- Multiple IBD calling methods used for *Pf*
- Comprehensive evaluation of IBD segment quality and its applicability in downstream analyses is lacking
- High recombination rate, low SNP density in *Pf* vs human necessitate IBD caller parameter optimization

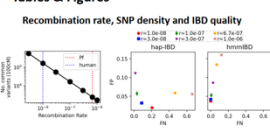
Methods

- Population-based simulation
- Tree-sequence-based True IBD
- Simulation under different demographic models to mimic low and high transmission settings
- Benchmarking metrics: IBD false positive/negative (FP/FN) rates, pairwise total IBD, and downstream estimates
- Parameter optimization via grid search
- Validation in empirical data (monoclonal, unrelated samples from in-house database and MalariaGEN PF7)

Results

- IBD callers designed for human genomes, such as phased-IBD, refined-IBD and hap-IBD perform well with human-like recombination rates but poorly with *Pf*-like rates
- IBD called using default parameters tends to have high FP/FN rates for *Pf*. Grid search of parameters, such as marker counts per segment, can largely improve IBD quality
- Using optimized parameters, most callers capture known population structure and selection signal, but only IBD from hmmIBD can accurately estimate effective population size (IBDNe).

Tables & Figures



Identity-By-Descent (IBD) callers are highly sensitive to low SNP density due to the high recombination rate in *Pf*

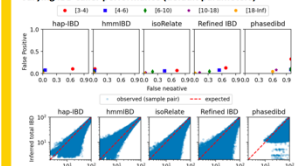
IBD caller parameter optimization can improve the quality of detected IBD segments

Not all IBD callers are suitable for downstream analyses sensitive to IBD quality (e.g., N_e estimation), even with optimization

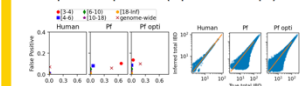
hmmIBD performs well for *Pf* genomes when given deconvoluted genotype data



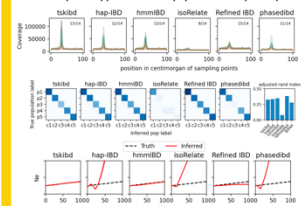
Varying IBD caller performance (default parameters)



IBD call parameter optimization (hap-IBD as an example)



Post-optimization benchmarking of downstream inferences, including selection detection and estimation of population structure (infomap) and effective population size (IBDNe)



Conclusions

- High recombination rate affects IBD quality
- Human-oriented IBD callers need parameter optimization for application to *Pf*
- IBD estimated using hmmIBD is relatively unbiased and is recommended for sensitive downstream analysis (such as N_e estimation)

Acknowledgment & Disclosure

We thank the Armed Forces Research Institute of Medical Sciences (AFRIMS) field and molecular teams and study participants. This work is supported by NIH (1R01AI145852) and US Department of Defense Global Emerging Infections Surveillance (GEIS) and Defense Medical Assistance Programs (DMAP).
Disclosure: Material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the author, and are not to be construed as official, or as reflecting true views of the Department of the Army or the Department of Defense. The investigators have adhered to the policies for protection of human subjects as prescribed in AR 70-25.

CVD·GLOBAL HEALTH

CENTER FOR VACCINE DEVELOPMENT AND GLOBAL HEALTH

[Lightning Talk Info](#)

- In Scientific Session **154**: Malaria - Genomics: Sharpening Tools to Guide Interventions and Uncover Biology
- Session Time: **8-9:45 AM, Sunday, Oct 22**
- Title: Benchmarking and Optimization of Identity-By-Descent Detection Methods for *Plasmodium falciparum*

[Extended Data](#)

Slides from 2023 FUTURE OF MALARIA RESEARCH SYMPOSIUM

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

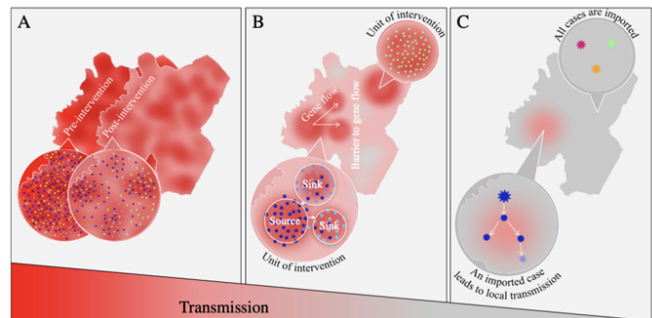
Bing Guo

Oct 13th, 2023

University of Maryland School of Medicine

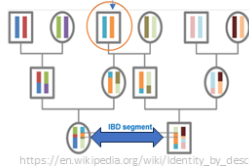
Background Parasite population genetics in malaria surveillance

- Surveillance is a core intervention for malaria reduction
- Population genetics is valuable in **all** transmission settings
 - High:
 - Genetic diversity
 - Effective population size
 - Intermediate:
 - Population structure
 - Migration
 - Low:
 - Relatedness network
 - Source of importation



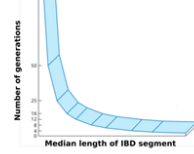
Background IBD: a powerful tool in *Pf* population genetics

Identity-by-descent (IBD) segment
(isolate pair, start pos, end pos)



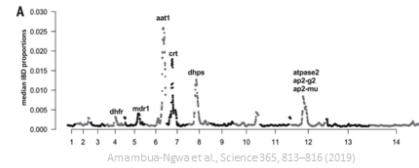
https://en.wikipedia.org/wiki/Identity_by_descent

Decay over generations
(Length indicates age)

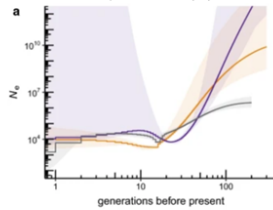


Browning, et al Annual Review of Genetics 46, 617–633 (2012)

Positional Enrichment (selection/drug resistance)



Population total IBD
(N_e , effective population size)



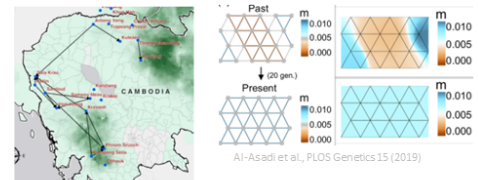
Morgan et al, Malaria Journal 19, 1–14 (2020)

Pair-wise total IBD
(relatedness network, community detection)



Henden et al, PLoS genetics 14 (2018)

Geo-indexed IBD
(migration direction, intensity)



Shetty et al, Nature Communications 10, 1–11, (2019)

3

Knowledge gap Potential IBD quality issues

- **Relative low SNP density** due to high recombination rate
 - Recombination rate is ~ 70 times higher in *Pf* than in humans
 - Human: 3000 cM – 2000k common SNPs – 66.7k SNPs per Morgan
 - *Pf*: 1500 cM – 20k common SNPs – **1.3k** SNPs per Morgan
- **Lack of a comprehensive evaluation**
 - Existing evaluation of *Pf*-oriented IBD callers have caveats
 - Pedigree-based simulation (focuses on close relatives with limited number of generations- from parent-offspring to 25 generations)
 - or less stringent definition of IBD accuracy
 - Applicability of human oriented IBD callers in *Pf* is unknown

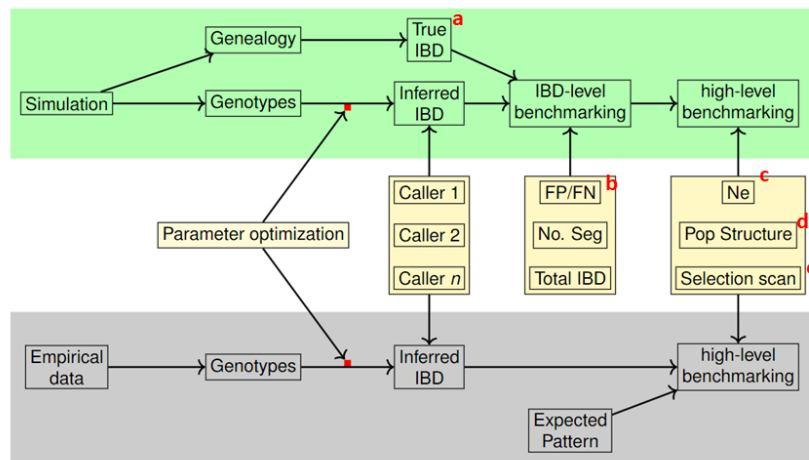
4

Hypotheses

- High recombination rate and rapid decline in N_e (due to reduced transmission) is associated with low SNP density in *Pf* genomic data and thus affects the performance of IBD calling algorithms
- Optimizing and prioritizing IBD callers via an evaluation system designed for *Pf* can improve the quality of detected IBD segments and results of downstream analyses

5

Methods Overall strategy



a tskibd

Guo et al., bioRxiv (2023) DOI: 10.1101/2023.07.14.549114

b FP/FN definition

Zhou et al., American Journal of Human Genetics 106, 426–437 (2020)

c IBDNe

Browning, American Journal of Human Genetics 97, 404–418 (2015)

d Infomap

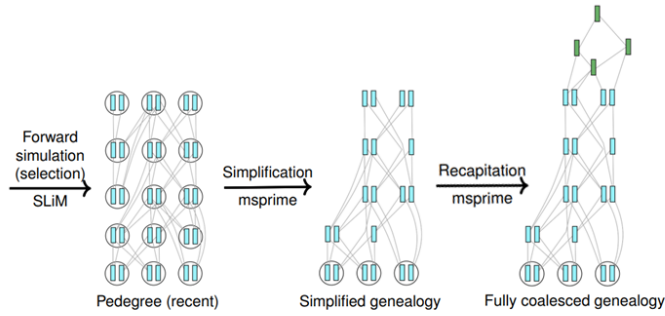
M. Rosvall et al., Eur. Phys. J. Spec. Top. 178, 13–23, (2009)

e Selection stats ($X_{iR,s}$)

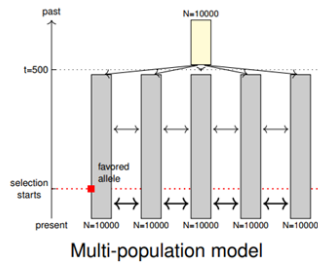
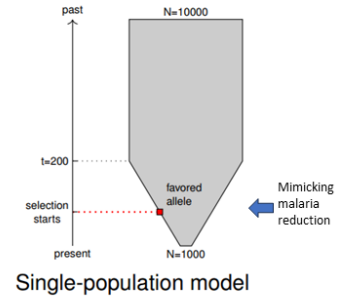
Henden et al., PLoS genetics 14, (2018)

6

Methods -- Simulations

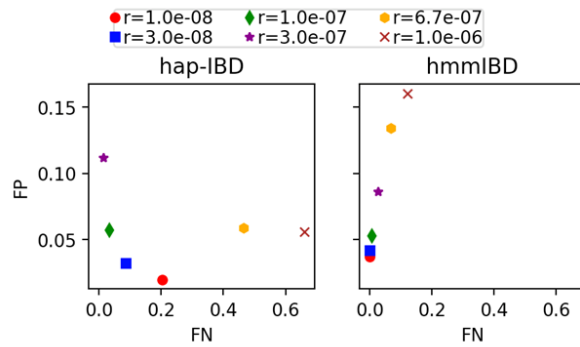
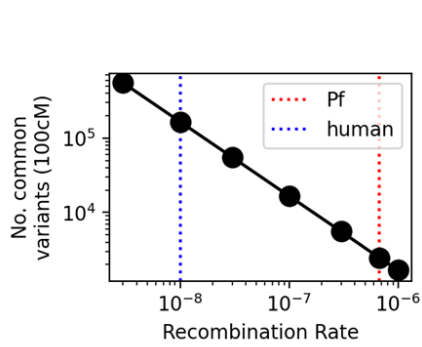


Haller et al., Molecular Ecology Resources 19, 552–566(2019)
Baumdicker et al., Genetics 220, iyab229 (2022)



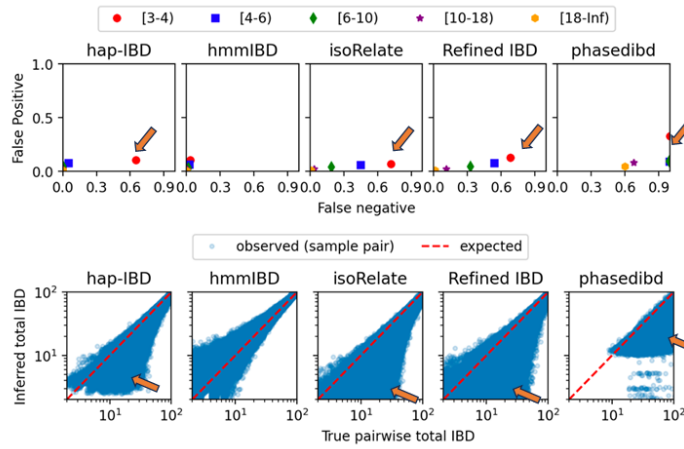
7

Results Low SNP density due to high recombination rate can reduce IBD accuracy



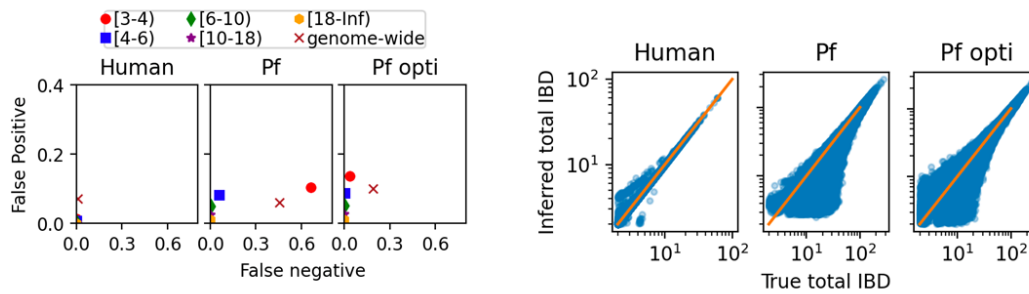
8

Results Quality of detected IBD varies across IBD callers with *Pf* genomes



9

Results IBD-caller specific parameters optimization for *Pf* genomes improves IBD quality

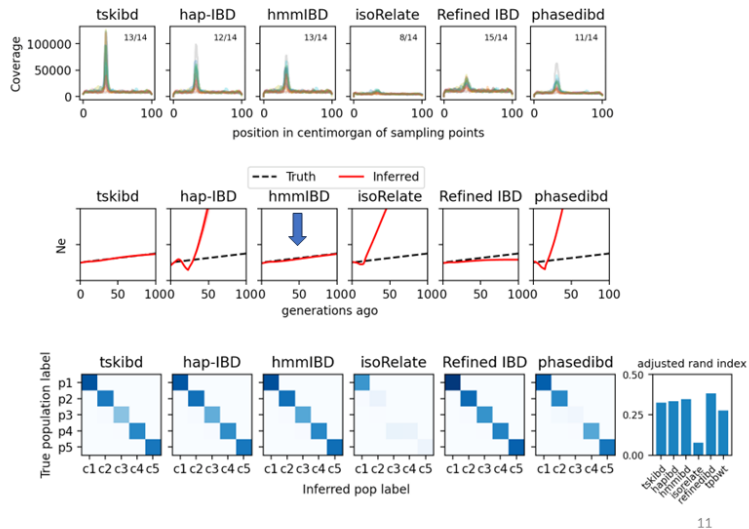


* Use hapIBD as an example

10

Results Post-optimization benchmarking via downstream inferences

- Most IBD callers can capture main pattern of simulated selection signal and population structure
- N_e estimates are very sensitive to IBD quality. Only hmmIBD generated realistic estimates



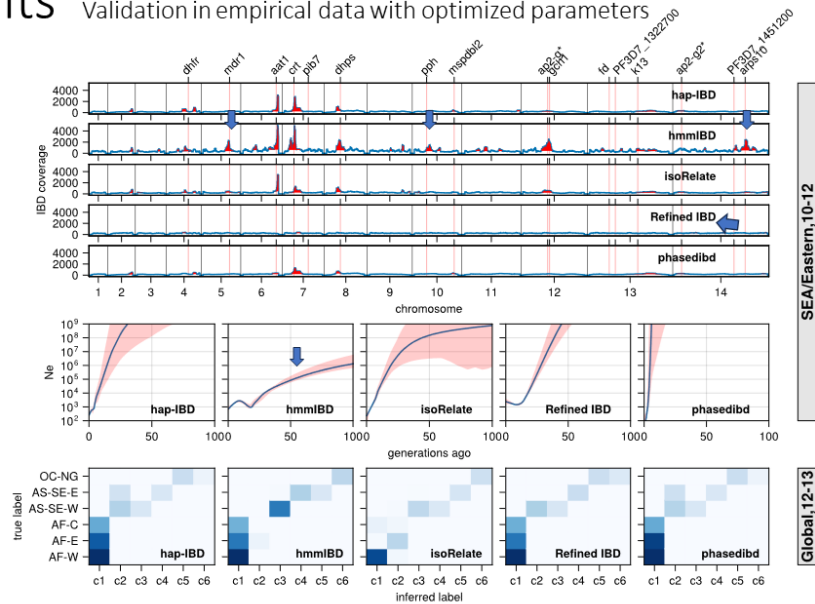
11

Methods -- Validation in Empirical data

- Processing MalariaGEN Pf7 data
 - Haploid genomes:
 - $F_{ws} > 0.95$; dominant allele ($AD \geq 5$, Fraction ≥ 0.9); missingness < 0.1
 - $MAF > 0.01$, biallelic site-only
 - Imputation: beagle 5.1 (no panel)
 - Unrelated isolates only (total IBD < 0.5 genome size)
- Constructing Datasets
 - "Single" population datasets: Eastern Southeast Asia and West Africa
 - For analyzing N_e and positive selection signal
 - Cross-continental dataset (structured)
 - For inferring population structure via IBD network and community detection algorithm

12

Results Validation in empirical data with optimized parameters



13

Conclusions and Next Steps

- Summary
 - High recombination rate affects IBD quality
 - Human-oriented IBD callers need parameter optimization for application to *Pf*
 - IBD estimated using hmmlBD is relatively unbiased and is recommended for sensitive downstream analysis (such as N_e estimation)
- Moving forward
 - Improve the computational efficiency of hmmlBD for large data sets
 - Benchmark IBD callers in other simulation models that incorporate
 - High inbreeding
 - Polyclonal infection
 - Different genotyping error rates

14

Acknowledgements

- Dissertation Committee members
 - Dr. Timothy O'Connor (Co-mentor)
 - Dr. Shannon Takala-Harrison (Co-mentor)
 - Dr. Joana Carneiro da Silva
 - Dr. Maciej F. Boni
 - Dr. David Serre
 - Dr. Michelle D. Shardell
- AFRIMS Department of Bacterial and Parasitic Diseases
- This work is supported by
 - NIH 1R01AI145852 (Takala-Harrison, O'Connor)

