## Conference Info

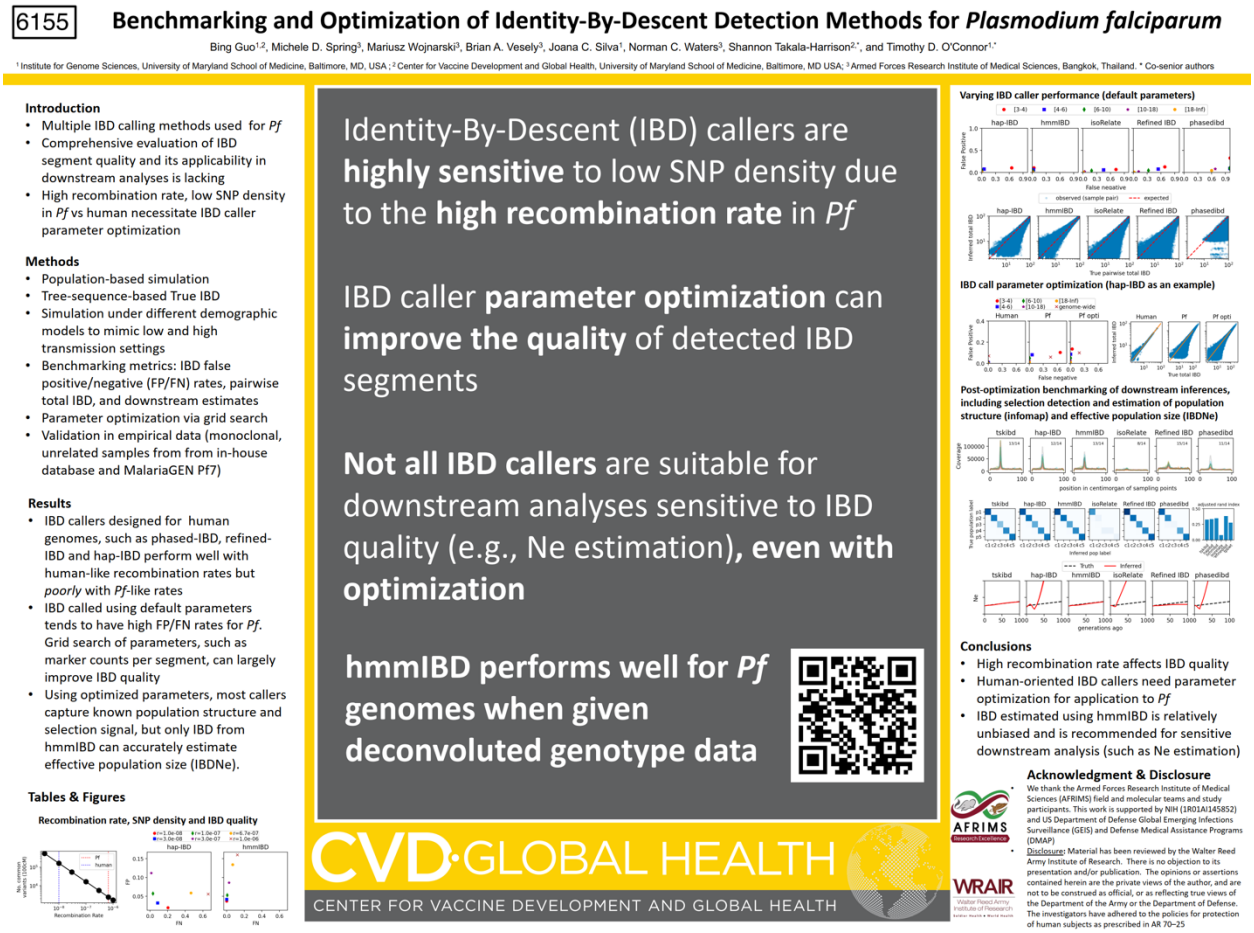
- American Society of Tropical Medicine and Hygiene
- 2023 Annual Meeting
- October 18-22, 2023
- Hyatt Regency Chicago
- Chicago, Illinois, United States

## 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)**



## 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

UNIVERSITY of MARYLAND SCHOOL OF MEDICINE
CENTER FOR VACCINE DEVELOPMENT AND GLOBAL HEALTH

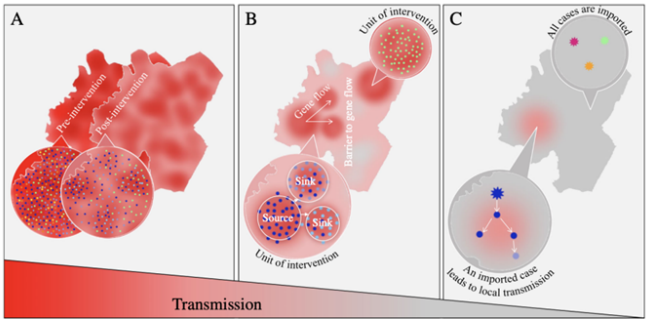
# Benchmarking and Optimization of Identity-By-Descent Detection Methods for *Plasmodium falciparum*

Bing Guo
Oct 13th, 2023
University of Maryland School of Medicine

UNIVERSITY of MARYLAND SCHOOL OF MEDICINE
CENTER FOR VACCINE DEVELOPMENT AND GLOBAL HEALTH

## 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



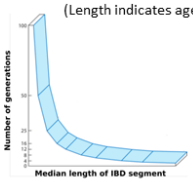

Wesolowski et al., BMC Med 16, 190 (2018)

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

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

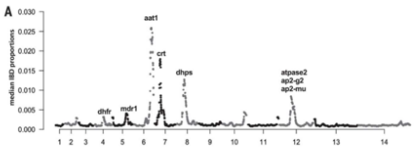

https://en.wikipedia.org/wiki/Identity_by_descent

Decay over generations
(Length indicates age)



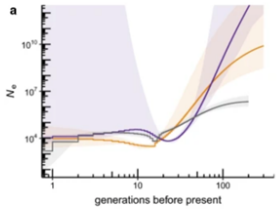

Browning, et al Annual Review of Genetics 46, 617–633 (2012)

Positional Enrichment (selection/drug resistance)




Amambua-Ngwa et al., Science 365, 813–816 (2019)

Population total IBD
(Ne, effective population size)

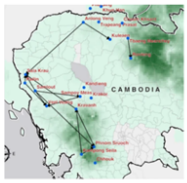


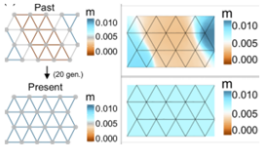

Morgan et al., Malaria Journal 19, 1–14 (2020)

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

Henden et al., PLoS genetics 14 (2018)

Geo-indexed IBD
(migration direction, intensity)




Al-Asadi et al., PLOS Genetics 15 (2019)

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





# 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

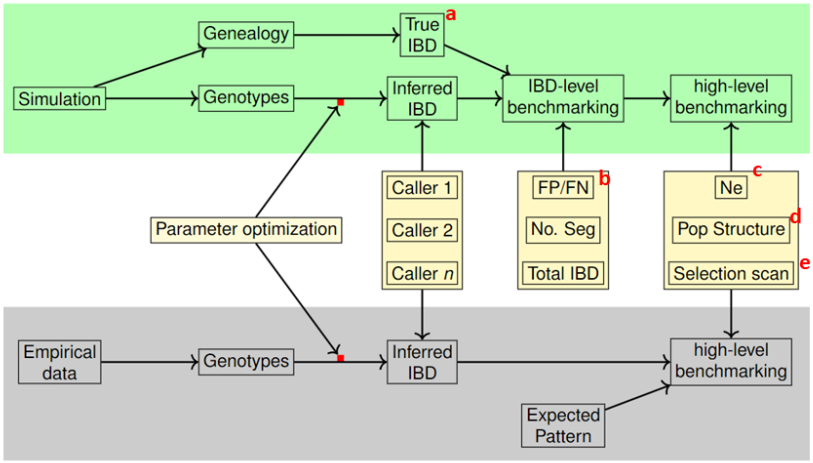
# Hypotheses

- High recombination rate and rapid decline in Ne (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



# Methods Overall strategy




[a] tskibd
Guo et al., bioRxiv (2023) DOI: 10.1101/2023.07.14.549114

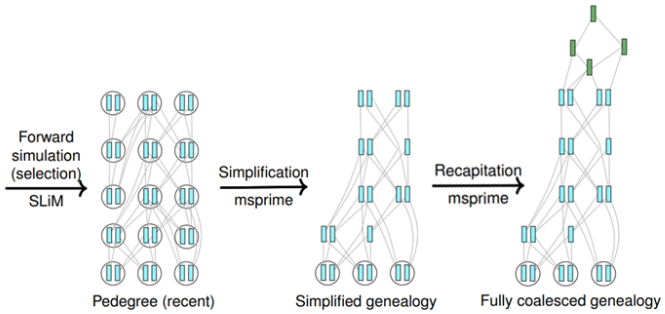
[b] FP/FN definition
Zhou et al., American Journal of Human Genetics 106, 426–437 (2020)

[c] IBDNe
Browning, American Journal of Human Genetics 97, 404–418 (2015)

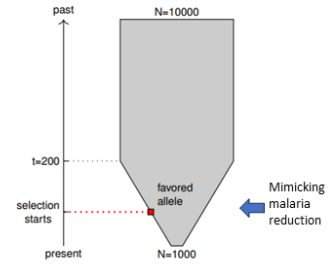
[d] Infomap
M. Rosvall *et al.*, *Eur. Phys. J. Spec. Top.* **178**, 13–23, (2009)

[e] Selection stats ($X_{iR,s}$)
Henden et al., PLoS genetics 14, (2018)

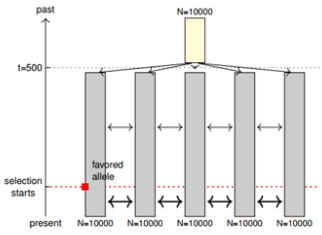
# Methods -- Simulations




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




Single-population model



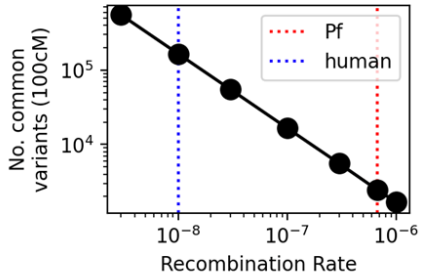

Multi-population model

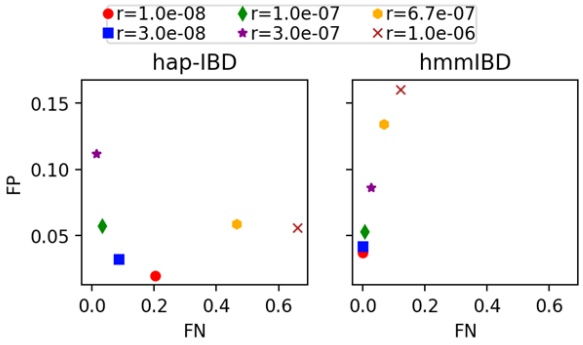




# Results

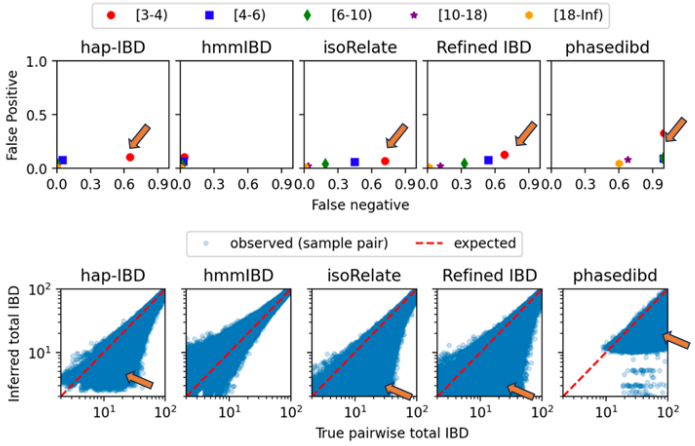
Low SNP density due to high recombination rate can reduce IBD accuracy



# Results

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

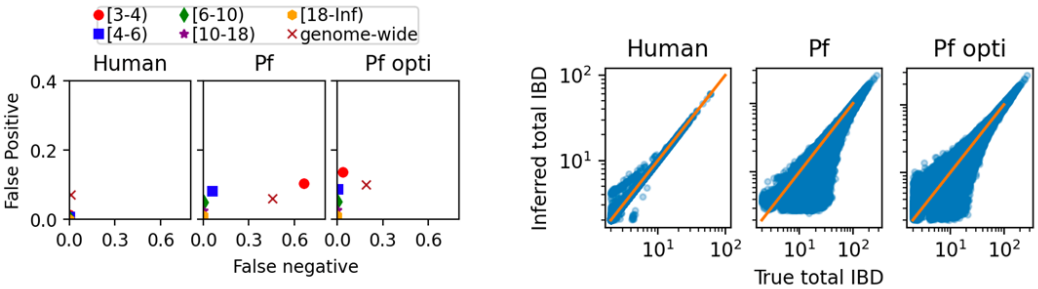







# Results

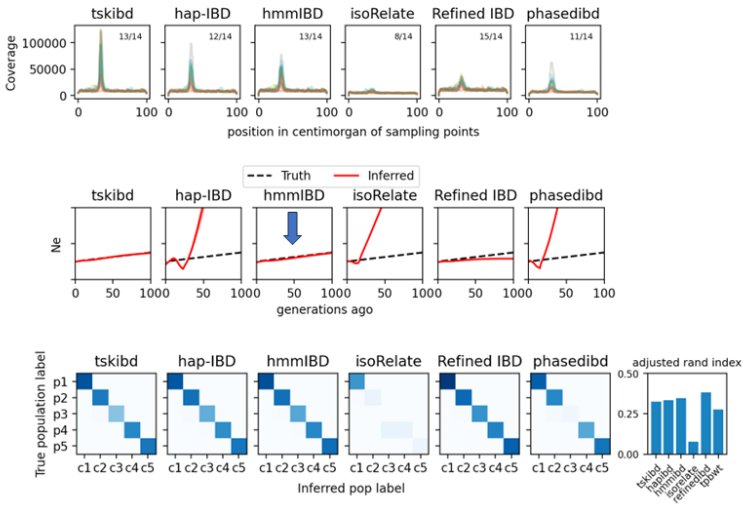
IBD-caller specific parameters optimization for *Pf* genomes improves IBD quality




* Use hapIBD as an exmaple

# Results Post-optimization benchmarking via downstream inferences

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



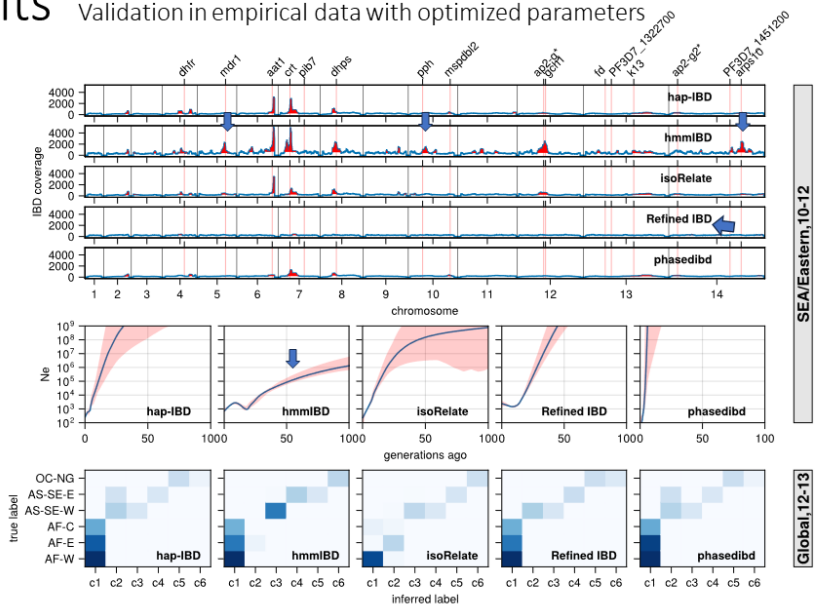



# Methods -- Validation in Empirical data

- Processing MalariaGEN Pf7 data
  - Haploid genomes:
    - $F_{ws}$ > 0.95; dominant allele (AD>=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 Ne and positive selection signal
  - Cross-continental dataset (structured)
    - For inferring population structure via IBD network and community detection algorithm

# Results

Validation in empirical data with optimized parameters






# 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 hmmIBD is relatively unbiased and is recommended for sensitive downstream analysis (such as Ne estimation)
- Moving forward
  - Improve the computational efficiency of hmmIBD for large data sets
  - Benchmark IBD callers in other simulation models that incorporate
    - High inbreeding
    - Polyclonal infection
    - Different genotyping error rates

# 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)