UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

The Duffy antigen system and Plasmodium vivax: Immunological and epidemiological approaches to unconventional malaria in Sub-Saharan Africa

Permalink

https://escholarship.org/uc/item/1pw0d632

Author

Bradley, Lauren

Publication Date

2023

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

The Duffy antigen system and *Plasmodium vivax*: Immunological and epidemiological approaches to unconventional malaria in Sub-Saharan Africa

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Lauren Elizabeth Bradley

Dissertation Committee:
Professor Guiyun Yan, Chair
Assistant Professor Grace Lee
Assistant Professor Daniel Parker

2024

Chapter 1 © 2023 American Society of Tropical Medicine and Hygiene
All other materials © 2024 Lauren Elizabeth Bradley

DEDICATION

To

my family, friends and mentors

for their unwavering love and support

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my mentors and committee members. Their contributions to my learning, curiosity, scientific acumen, and grit are deeply appreciated.

I offer my deepest thanks to the research teams in Ethiopia who kept me safe, housed me, fed me, and educated me during my time conducting field work with them. Without their help the research conducted here would have been impossible.

Chapter 1 of this dissertation is a reprint of the material as it appears in Bradley L. et al. Determination of *Plasmodium vivax* and *Plasmodium falciparum* Malaria Exposure in Two Ethiopian Communities and Its Relationship to Duffy Expression. *The American Journal of Tropical Medicine and Hygiene*. 2023;109(5):1028-1035. doi:10.4269/ajtmh.22-0644, used with permission from the American Society of Tropical Medicine and Hygiene. The co-authors listed in this publication are Delenasaw Yewhalaw, Elizabeth Hemming-Schroeder, Paula Embury, Ming-Chieh Lee, Endalew Zemene, Teshome Degefa, Christopher King, James Kazura, Guiyun Yan, and Arlene Dent.

Thank you to the American Society for Tropical Medicine and Hygiene for inviting me to present on my Chapter 1 research during their 2021 annual meeting, and to the National Institutes' of Health for awarding me the F31 Ruth L. Kirschstein Predoctoral Individual National Research Service Award (5 F31 AI161887) which funded the presented dissertation research.

VITA

Lauren Elizabeth Bradley

2016	B.A. in Ecology and Evolutionary Biology
2016-2018	Research Assistant,
	Infectious Disease Diagnostics at LABS Inc.
2019	Field Associate - Ethiopia
2022-2023	Visiting Research Scholar,
	Colorado State University
2018-2020	Teaching Assistant, School of Biological Sciences,
	University of California, Irvine
2021	M.S. University of California, Irvine
2024	Ph.D. in Biological Sciences,
	University of California, Irvine

ABSTRACT OF THE DISSERTATION

The Duffy antigen system and *Plasmodium vivax*:
Immunological and epidemiological approaches to
unconventional malaria in Sub-Saharan Africa

by

Lauren Elizabeth Bradley
Doctor of Biological Sciences
University of California, Irvine, 2024
Professor Guiyun Yan, Chair

In the last two decades' tremendous progress has been made towards global goals of malaria control and elimination, yet there were still an estimated 227 million cases in 2019[1]. While the vast majority of these cases occur via the malaria species *Plasmodium falciparum* on the African continent, *Plasmodium vivax* remains the most geographically widespread. It is endemic to large swaths of Asia and South America, and remains among the most challenging of tropical diseases to control and eradicate due to its unique biology and global presence[2]. Additionally, while *Plasmodium vivax* was once considered to be a benign infection, it is now recognized as a significant global health threat[3]. Latent liver-stage hypnozoites can cause relapse

weeks or months after initial infection, and difficulties in hypnoizoite treatment related to G6PD deficiency based primaquine toxicity[4-6] significantly hinder control and intervention efforts.

In spite of its global dispersal, *P. vivax* is relatively absent in Sub-Saharan Africa. The lack of *P. vivax* endemicity in Africa has long been attributed to the near fixation of the Duffy negative phenotype on the continent[7]. The Duffy antigen receptor for chemokines (DARC) was established as the required binding antigen for *P. vivax* merozoites over four decades ago[8], yet recent work has highlighted the occurrence of *P. vivax* in confirmed Duffy negatives in several countries throughout Africa[9-13]. New attention has been given to these rare cases that challenge historical dogma and epidemiology; questioning how these infections are occurring, if there are alternate erythrocyte invasion mechanisms at play, if these infections are novel or merely historically undetected, and importantly what public health significance do they pose to Duffy negative individuals throughout Africa. The presented dissertation explores these infections through multiple lenses; investigating immunology, epidemiology and population genetics of unconventional *Plasmodium vivax* infections in Duffy negative people in Ethiopia.

In our first chapter we assess the impact Duffy genotype has on naturally acquired immunity. We calculated seroreactivity to a suite of *Plasmodium falciparum* and *P. vivax* target antigens and sequenced two sections of the DARC gene to interpret Duffy genotype in humans. We found that Duffy negative individuals developed significantly reduced immune responses to *P. vivax* antigens compared to *P. falciparum* antigens across almost all antigenic targets. Additionally, we assessed seroprevalence and found that overall immune responses to *P. vivax* were greatly diminished in Duffy negative individuals highlighting the need for consideration of Duffy heterogeneity in control efforts and vaccine development. In chapter 2 we first optimized a pooling strategy in order to process copious numbers of Dried Blood Spots (DBS), then assessed

malaria prevalence and parasitemia in individuals of different Duffy phenotypes. Out of over 15 thousand samples a mere 15 were both *P. vivax* positive and Duffy negative, highlighting the rarity of these infections in Ethiopia. Interestingly we found that these infections were most commonly found as mixed co-infections with *Plasmodium falciparum*, and that parasitemia did not vary between Duffy phenotype. Our third chapter is a literature review which sought to explore the best population genetic approaches currently available for malaria epidemiology. We found that despite microsatellite, SNP-barcode, and whole genome sequencing (WGS) all being commonly used with great success, each individually falls short in some aspect, making intensive population genetic and genomic analyses of malaria epidemiology challenging. Highly multiplexed amplicon deep sequencing for microhaplotypes is a novel genetic approach that poses great promise for garnering high-resolution genetic data and in this chapter we present recent developments of this technology, highlight its novel use in *P. falciparum* genomic exploration, and argue for its superiority as a population genetics approach and tool for malaria and specifically *P. vivax* malaria.

References

1. WHO: **World Malaria Report 2021.** *Geneva, Switzerland* 2021.
2. Rosenthal PJ: **Malaria in 2022: Challenges and Progress.** *Am J Trop Med Hyg* 2022, **106:**1565-1567.
3. Phyo AP, Dahal P, Mayxay M, Ashley EA: **Clinical impact of vivax malaria: A collection review.** *PLoS Med* 2022, **19:**e1003890.
4. Imwong M, Snounou G, Pukrittayakamee S, Tanomsing N, Kim JR, Nandy A, Guthmann JP, Nosten F, Carlton J, Looareesuwan S, et al: **Relapses of Plasmodium vivax infection usually result from activation of heterologous hypnozoites.** *J Infect Dis* 2007, **195:**927-933.
5. Baird K: **Origins and implications of neglect of G6PD deficiency and primaquine toxicity in Plasmodium vivax malaria.** *Pathog Glob Health* 2015, **109:**93-106.
6. Wells TN, Burrows JN, Baird JK: **Targeting the hypnozoite reservoir of Plasmodium vivax: the hidden obstacle to malaria elimination.** *Trends Parasitol* 2010, **26:**145-151.
7. Howes RE, Patil AP, Piel FB, Nyangiri OA, Kabaria CW, Gething PW, Zimmerman PA, Barnadas C, Beall CM, Gebremedhin A, et al: **The global distribution of the Duffy blood group.** *Nat Commun* 2011, **2:**266.
8. Louis H. Miller SJM, David F. Clyde, Mary H. McGinnis: **The resistance factor to Plasmodium vivax in Blacks: The Duffy Blood Group Genotype, FyFy.** *The New England Journal of Medicine* 1976, **295**.
9. Menard D, Barnadas C, Bouchier C, Henry-Halldin C, Gray LR, Ratsimbasoa A, Thonier V, Carod JF, Domarle O, Colin Y, et al: **Plasmodium vivax clinical malaria is**

commonly observed in Duffy-negative Malagasy people. *Proc Natl Acad Sci U S A* 2010, **107:**5967-5971.

10. Zimmerman PA: **Plasmodium vivax Infection in Duffy-Negative People in Africa.** *Am J Trop Med Hyg* 2017, **97:**636-638.

11. Wurtz NL, K. M.; Bogreau, H.; Pradines, B.; Rogier, C.; Mohamed Salem Boukhary, A. O.; Hafid, J. E.; Ould Ahmedou Salem, M. S.; Trape, J; Basco, L. K.; Briolant, S.: **Vivax malaria in Maritania includes infection of a Duffy-negative individual.** *Malaria Journal* 2011, **10**.

12. Lo E, Yewhalaw D, Zhong D, Zemene E, Degefa T, Tushune K, Ha M, Lee MC, James AA, Yan G: **Molecular epidemiology of Plasmodium vivax and Plasmodium falciparum malaria among Duffy-positive and Duffy-negative populations in Ethiopia.** *Malar J* 2015, **14:**84.

13. Woldearegai TG, Kremsner PG, Kun JF, Mordmuller B: **Plasmodium vivax malaria in Duffy-negative individuals from Ethiopia.** *Trans R Soc Trop Med Hyg* 2013, **107:**328-331.

CHAPTER 1

DETERMINATION OF *PLASMODIUM VIVAX* AND *PLASMODIUM FALCIPARUM* MALARIA EXPOSURE IN TWO ETHIOPIAN COMMUNITIES AND ITS RELATIONSHIP TO DUFFY EXPRESSION

Published in ***The American Journal of Tropical Medicine and Hygiene***

2023 109(5):1028-1035. DOI: 10.4269/ajtmh.22-0644. PMID: 37918005

Lauren Bradley[1], Delenasaw Yewhalaw[2,3], Elizabeth Hemming-Schroeder[4], Paula Embury[5], Ming-Chieh Lee[6], Endalew Zemene[3], Teshome Degefa[7], Christopher King[5], James Kazura[5], Guiyun Yan[1,6], Arlene Dent[5*]

[1]Department of Ecology and Evolutionary Biology, School of Biological Sciences, University of California Irvine, Irvine CA 92697, USA

[2]Department of Medical Laboratory Sciences and Pathology, College of Health Sciences, Jimma University, Jimma 5195, Ethiopia

[3] Tropical and Infectious Diseases Research Center, Jimma University, Jimma, Ethiopia

[4]Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523, USA

[5]Center for Global Health and Disease, Case Western Reserve University, Cleveland, OH 44106, USA

[6]Program in Public Health, College of Health Sciences, University of California at Irvine, Irvine CA 92697, USA

[7]School of Medical Laboratory Sciences, Faculty of Health Sciences, Jimma University, Jimma, Ethiopia

*Corresponding author: Arlene Dent, aed9@case.edu, (216) 368-6005

Word Counts: Abstract 206, Text 3570 Figures: 4, Tables: 3, Supplementary Tables: 1, Supplemental Figure 1

Running Title: Malaria exposure and Duffy expression in Ethiopia

Abstract

Despite historical dogma that Duffy blood group negativity of human erythrocytes confers resistance to *Plasmodium vivax* blood stage infection, cases of *P. vivax* malaria and asymptomatic blood stage infection (subclinical malaria) have recently been well documented in Duffy negative individuals throughout Africa. However, the impact of Duffy negativity on the development of naturally acquired immunity to *P. vivax* remains poorly understood. We examined antibody reactivity to *P. vivax* and *P. falciparum* antigens at two field sites in Ethiopia and assessed Duffy gene expression by PCR amplification and sequencing of the GATA-1 transcription factor-binding site of the Duffy antigen receptor for chemokines (DARC) gene promotor region that is associated with silencing of erythroid cell transcription and absent protein expression. Antibodies to three of the four *P. vivax* blood stage antigens examined, RBP2b, EBP2 and DBPIISal-1, were significantly lower (P-value <0.001) in Duffy negative individuals relative to Duffy positive individuals. In stark contrast no clear pattern was found across Duffy negative and positive genotypes for *P. falciparum* antibodies. We conclude that lack of erythroid Duffy expression is associated with reduced serologic responses, indicative of less naturally acquired immunity and less cumulative exposure to blood stage *P. vivax* parasites relative to Duffy positive individuals living in the same communities.

Keywords: malaria, *Plasmodium vivax*, Duffy blood group, serology, naturally acquired immunity

Background

Plasmodium vivax is globally the most widely distributed human malaria species. Blood stage infection and clinical vivax malaria-infection occur throughout tropical, subtropical and temperate regions of the world and it is the predominant *Plasmodium* species in Asia and South America[1]. One reason for this high prevalence is that dormant liver stage *P. vivax* hypnozoites, maintained in host hepatocytes for months to years, can lead to relapses of blood stage infection without continuous exposure to infective mosquitos[2,3]. Another unique feature of *P. vivax* biology is that merozoites appear to require expression of the Duffy antigen receptor for chemokines (DARC), here referred to as the Duffy blood group, to invade reticulocytes and immature erythroid lineage cells in bone marrow. It has long been accepted that the lack of Duffy expression, confers resistance to *P. vivax* blood stage infection[4-6]. Indeed, erythroid Duffy negativity is nearly fixed in populations indigenous to sub-Saharan Africa and has historically been credited with the lack of endemic *P. vivax* on the continent. Notably, *P. vivax* blood stage infection of Duffy positive and Duffy negative individuals also occurs in Madagascar, an Indian Ocean island country located 400 km from the southeastern coast of the African land mass[7]. However, recent observations indicate that *P. vivax* is endemic in many areas and regions of the continent, with the highest prevalence (approximately 8 percent) in the horn of Africa[8-10]. As many areas of sub-Saharan Africa experience reduced malaria burden and transition towards elimination, traditional methods for surveillance of blood stage infection such as blood smear microscopy, PCR and rapid diagnostic tests may be underestimating true malaria prevalence as they tend to miss low-density asymptomatic infections (subclinical malaria)[11,12]. This is especially true for *P. vivax* as traditional testing primarily accounts for active blood stage

infections, which can represent only a small portion of all *P. vivax* prevalence as they are unable to detect latent liver stage infections[13].

Serological tools that detect antibodies to *P. vivax* antigens are a potentially robust and sensitive measure for assessing *P. vivax* exposure in low-transmission settings[14-16]. Serology has the ability to capture information about not only concurrent blood stage infections, but also past infections that occurred at different times regardless of anti-malarial drug treatment[17,18]. There is currently little information on the relationship between erythroid Duffy blood group negativity and antibody responses to *P. vivax* in African populations.

We previously collected dried blood spots and whole blood for qPCR diagnosis of *P. vivax* and *P. falciparum* infection from healthy community residents of two sites in Southwestern Ethiopia. Plasma IgG antibodies to four *P. vivax* antigens and six *P. falciparum* antigens were quantified and genomic DNA was isolated for Duffy blood group sequencing. The aims of the study were to answer the following questions: (1) What is the prevalence of *P. vivax* and *P. falciparum* blood stage infection according to Duffy expression; (2) Do antibody prevalence and level differ based on Duffy expression; (3) Is there spatial clustering of *P. vivax* and *P. falciparum* infection and serology according to Duffy expression? These questions are of key epidemiological importance and will offer valuable insight into the role Duffy expression plays immunologically.

Methods

Ethical approval

Ethical approval was obtained from the Institutional Review Boards (IRB) at the University of California at Irvine, Case Western Reserve University, USA, and the Tropical and Infectious Diseases Research Center at Jimma University, Ethiopia. All residents willing to participate in the study were included; adults provided signed consent for themselves and assent for minors under 18 years of age after explanation of the study objectives and methodologies.

Study site

Figure 1 shows the location of villages of study participants within the two study sites, Arjo in the eastern Wollega Zone of the Oromia Regional State and Gambella district in the Gambella Regional State. These sites were chosen because both locations were previously found to have high DARC gene polymorphisms and different *P. vivax* malaria endemicity[19,20]. Arjo is an irrigation site for a government owned commercial sugar plantation, and Gambella is an irrigation site for a commercially owned large-scale rice farm. Malaria endemicity in Arjo is low (infection prevalence <5%), and *P. vivax* is the predominant malaria species[21]. Residents from 15 villages were enrolled (**Figure 1**). Malaria endemicity in Gambella is moderate (infection prevalence 10-20%), and *P. falciparum* is the predominant species[22]. Residents from 6 villages were studied.

Blood sample collection and processing

Dried blood spots containing 50ul of capillary blood collected by finger prick were collected from healthy study participants during a house-to-house Mass Blood Survey (MBS) in October 2018. Age and sex of each participant were recorded at the time of collection. Dry blood spot samples were transported to the University of California, Irvine and stored at -20^{o}C, and plasma

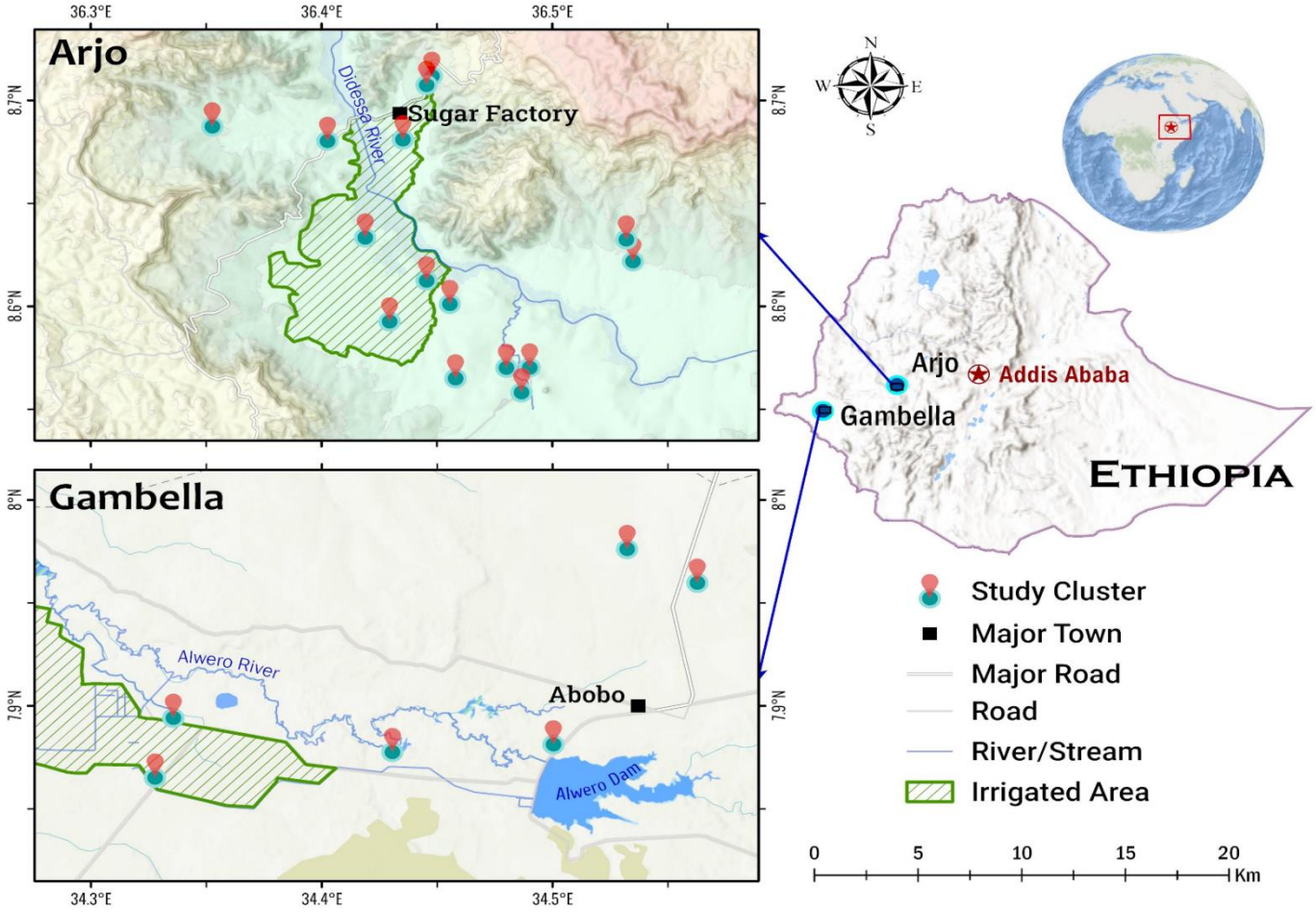


Figure 1 Map of the two selected Ethiopian study sites showing study resident locations and irrigation boundaries. Top: Arjo, Bottom: Gambella

was isolated from peripheral blood collected via venipuncture and stored at -80ºC until processing described below.

DNA was extracted from dried blood spots using established protocols[23] following standardized saponin/chelex procedures. DNA was eluted into ~170ul of molecular grade water, and either stored short term at 4ºC or at -20ºC for long term storage. *Plasmodium* species DNA determination and quantification of both *P. vivax* and *P. falciparum* were conducted using published protocols[24,25] with some modifications. qPCR amplification was conducted on a ThermoFisher QuantStudio 3 at a total volume of 20ul containing 6ul TaqMan Fast Advanced Master Mix, 0.5ul *P. vivax*-probe (VIC), 0.5ul *P. falciparum*-probe (FAM), 0.4ul forward and reverse primers and 2ul genomic DNA. Primers were prepared at 10uM and probes at 2uM concentrations.

Determination of Duffy genotypes

Duffy expression for each sample was conducted using established protocols and primers[9,26]. An approximately ~1,100bp region of the human DARC gene encompassing the 33rd nucleotide position was amplified and sequenced. Reactions contained 10ul DreamTaq Green Master Mix (ThermoFisher), 0.3ul forward and reverse primers and 2ul genomic DNA. Thermocycling was performed with the conditions of 94ºC for 2 min, 35 cycles of 94ºC for 30sec, 61ºC for 30sec, and 65ºC for 40sec, followed by a 2 min extension at 65ºC. Six microliters of amplified DNA were run on a 2.0% agarose gel to check for amplification. Successfully amplified samples were submitted to Retrogen Inc for Sanger sequencing. Sequencing result chromatograms were visually analyzed for five SNPs informing Duffy Fy genotypes; T33C, G125A, G145A, C265T, and G298A (see Duffy blood group nomenclature[26]).

Genotype information from sequence data was used to infer Duffy expression and red blood cell (RBS) phenotype (**Supplementary Table 1**). Dual peaks were determined to be heterozygous.

Plasma collection and quantification of IgG antibodies to Pf and Pv proteins

MAGPIX (Luminex) carboxylated microspheres were coupled to target four *P. vivax* and six *P. falciparum* recombinant protein antigens (**Table 1**), according to manufacturer's protocols with previously described modifications[27]. In order to determine an optimal antibody signal varying concentrations of each target protein were coupled to beads and tested with plasma pooled from 20 healthy Kenyan adult residents of Kisumu County, Kenya who donated blood samples in 2000. Quantification of IgG antibodies followed previous published methods[27,28]. Plasma pools were 4-fold serially diluted from ratios of 1:50 to 1:51,200, yielding a 6-point standard curve. Standard curves were run on each plate and used to determine the linear range of antibody binding for each bead. We quantified antibody magnitude using the dilution that fell within the linear range of the positive control pool. R-Phycoerythrin AfniPure F(ab')2 fragment goat anti-human IgG F(ab')2 (Jackson ImmunoResearch Laboratories) was used as the secondary antibody, and mean fluorescent intensity (MFI) values obtained during the assay were divided by the average MFI of malaria naïve negative controls on the same plate in order to normalize results and obviate variations across plates. Results are presented as the fold-increase of participant sample MFI over malaria naïve negative controls. Samples with an MFI greater than the mean plus 3 standard deviations (SD) of negative controls were considered positive, MFI values below this cutoff were set equivalent to malaria naïve controls at 1.

Table 1 Selected immune biomarkers used to test participant plasma and assess malaria exposure in Ethiopian communities.

Species	Abbreviation	Antigen	Parasite Stage
Plasmodium vivax	PvMSP1-19	Merozoite surface protein	Merozoite
	PvRBP2b	Reticulocyte binding protein	Merozoite
	PvEBP2	Erythrocyte binding protein	Merozoite
	PvDBPsal1	Duffy binding protein	Merozoite
Plasmodium falciparum	PfRh2a.b	Reticulocyte binding protein homolog	Merozoite
	PfMSP1-42	Merozoite surface protein	Merozoite
	PfAMA1	Apical membrane antigen1	Merozoite
	PfCSP	Circumsporozoite protein	Pre-erythrocytic
	PfEBA175	Erythrocyte binding antigen	Merozoite
	PfMSP3	Merozoite surface protein	Merozoite

Data analysis.

Malaria prevalence was calculated for Duffy-negative and Duffy-positive individuals for *Plasmodium vivax* and *P. falciparum* separately. Statistical significance between the two Duffy groups was tested using the Fisher's exact test. Mean fluorescent intensity (MFI) was compared across Duffy phenotypes by Kruskall-Wallis test. Significance was adjusted using the Bonferroni correction for multiple comparisons for four *P. vivax* antigens and six *P. falciparum* antigens. Raw MFI data was converted to binary data to assess overall seroprevalence. Given the proportional nature of the converted data, seroprevalence data was arcsine transformed and compared across Duffy expression and phenotypes for all *P. vivax* and *P. falciparum* antigens. Statistical significance was assessed via two-sample t-test for Duffy expression and Kruskall-Wallis test for Duffy phenotypes. Clustering of antibody profiles was assessed via principal component analysis (PCA) for *P. vivax* and *P. falciparum* antigens. Non-overlapping confidence ellipses representing a 95% confidence interval indicate significant variation between Duffy negative and Duffy positive individuals.

Results

Malaria burden and Duffy expression

A total of 527 dried blood spots were collected from 15 villages in Arjo and 6 villages in Gambella, and genotyped for T-33C mutation of the DARC gene. The prevalence of *P. falciparum* blood stage infection was approximately twice as high as *P. vivax* at both study sites. In total, 17 (3.2%) samples were positive for *P. falciparum*; *P. falciparum* prevalence was 0.7% in Duffy positives and 6.2% in Duffy negative. Overall, eight (1.5%) *P. vivax* infections were identified, with 2.1% prevalence among Duffy positives and 0.8% prevalence among Duffy negatives. Analyses of prevalence and Duffy expression showed a significant difference in *P. vivax* and *P. falciparum* infection distribution between Duffy negatives and Duffy positives ($P < 0.01$, **Table 2**), where *P. vivax* infections occurred more rarely in Duffy negative individuals. Based on qPCR none of the malaria positive samples exhibited a mixed infection with both *P. falciparum* and *P. vivax*.

Among the 527 samples genotyped for T-33C mutation, 395 randomly selected samples were successfully genotyped at four additional DARC SNPs (G125A, G145A, C265T and G298A). A detailed breakdown of number of samples in each Duffy phenotype, relative to the 13 different observed genotypes, is presented in **Supplementary Table 1**. Overall, 179 (45.3%) were homozygous Duffy negative Fy(a-b-), while the remaining 216 samples were Duffy positive. Duffy positive individuals were comprised of four distinct red blood cell (RBS) phenotypes inferred from genotype information; 30 (7.6%) Fy(a+b+), 63 (15.9%) Fy(a+b-), 1 (0.3%) Fy(a+b^{weak}), and 122 (30.9%) Fy(a+b-). Distribution of Duffy expression was similar in both study sites. In Arjo 93/204 individuals (45.6%) were homozygous Duffy negative Fy(a-b-).

In Gambella distribution was very similar with 86/191 (45.0%) of individuals exhibiting homozygous negativity Fy(a-b-).

Seroprevalence, Serologic reactivity and clustering effects of Duffy expression on serology

Due to limited number of children under 5 years old (n=1) and children ages 5-15 (n=32), only plasma samples from individuals over 15 years of age (n=362) were included in serologic reactivity analyses (**Table 3**). We observed distinctly different trends in *P. vivax* and *P. falciparum* antibodies among the different Duffy phenotypes. *P. vivax* antibodies exhibited significantly reduced seroprevalence in Duffy negative individuals compared to Duffy positive people ($P < 0.001$; **Figure 2**). Conversely, *P. falciparum* antibodies showed no significant variation between Duffy phenotypes ($P > 0.05$). Assessment of seroprevalence across Duffy phenotypes was also analyzed for all phenotypes with n>10. We found greater significance between Duffy phenotypes for *P. vivax* antibodies than for *P. falciparum* antibodies (**Supplemental Figure 1**).

Table 2 Prevalence of *Plasmodium vivax* and *P. falciparum* in Duffy negative and Duffy positive individuals from pooled Arjo and Gambella dried blood spots.

Duffy expression	n	*Plasmodium vivax*	*Plasmodium falciparum*
Positive	284	6 (2.11%)	2 (0.70%)
Negative	243	2 (0.82%)	15 (6.17%)
Total	527	8 (1.52%)	17 (3.23%)

Supplementary Table 1 Sample distribution of Duffy phenotypes and expression relative to 13 observed genotypes from Duffy Fy alleles from 395 subjects of all ages (*FyA, FyB*, *FyB, *FyX2*, *FyA^{ES}* and *FYB^{ES}*).**

RBC phenotype	Genotype	Antigen	Expression*	Nt 33	Nt 125	Nt 145	Nt 265	Nt 298	Total Genotype (n)	Total Phenotype (n)
(a+b+)	*FyA/FyB*	Fya^{+}Fyb^{+}	Positive	T/T	G/A	G/G	C/C	G/G	28	30
	*FyA/FyB**			T/T	G/A	G/G	C/C	G/A	2	
(a+b-)	*FyA/FyA*	Fya^{+}	Positive	T/T	G/G	G/G	C/C	G/G	6	63
	FyA/FyB^{ES}			T/C	G/A	G/G	C/C	G/G	56	
	FyA/FyA^{ES}			T/C	G/G	G/G	C/C	G/G	1	
(a+b^{weak})	*FyA/FyX2*	Fya^{+}Fyb$_{weak}$	Positive	T/T	G/A	G/G	C/T	G/A	1	1
(a-b+)	*FyB/FyB^{ES}*	Fyb^{+}	Positive	T/C	A/A	G/G	C/C	G/G or G/A	93	122
	FyB/FyB			T/T	A/A	G/G	C/C	G/G	20	
	*FyB/FyB**			T/T	A/A	G/G	C/C or C/T	A/G	7	
	FyB*/FyX2			T/T	A/A	G/G	C/T	A/A	1	
	FyA^{ES}/FyB*			C/T	G/A	G/G	C/C	G/A	1	
(a-b-)	FyB^{ES}/FyB^{ES}	no antigen	Negative	C/C	A/A	G/G or G/A	C/C	G/G	176	179
	FyA^{ES}/FyB^{ES}			C/C	G/A	G/G	C/C	G/G	3	

* Duffy expression is determined based on T-33C mutation in the promote region of the DARC gene.

Table 3 Distribution Duffy Phenotypes in adult Plasma samples collected from Arjo and Gambella, Ethiopia

Duffy Expression*	Duffy Phenotype	Arjo	Gambella	Total
Positive	a+b+	12 (6.7%)	18 (9.8%)	30 (8.3%)
	a+b-	36 (20.1%)	21 (11.5%)	57 (15.7%)
	$a+b^{weak}$	0	1 (0.5%)	1 (0.3%)
Negative	a-b+	52 (29.1%)	60 (32.8%)	112 (30.9%)
	a-b-	79 (44.1%)	83 (45.4%)	162 (44.8%)
Total		179	183	362

* Duffy expression is determined based on T-33C mutation in the promote region of the DARC gene.

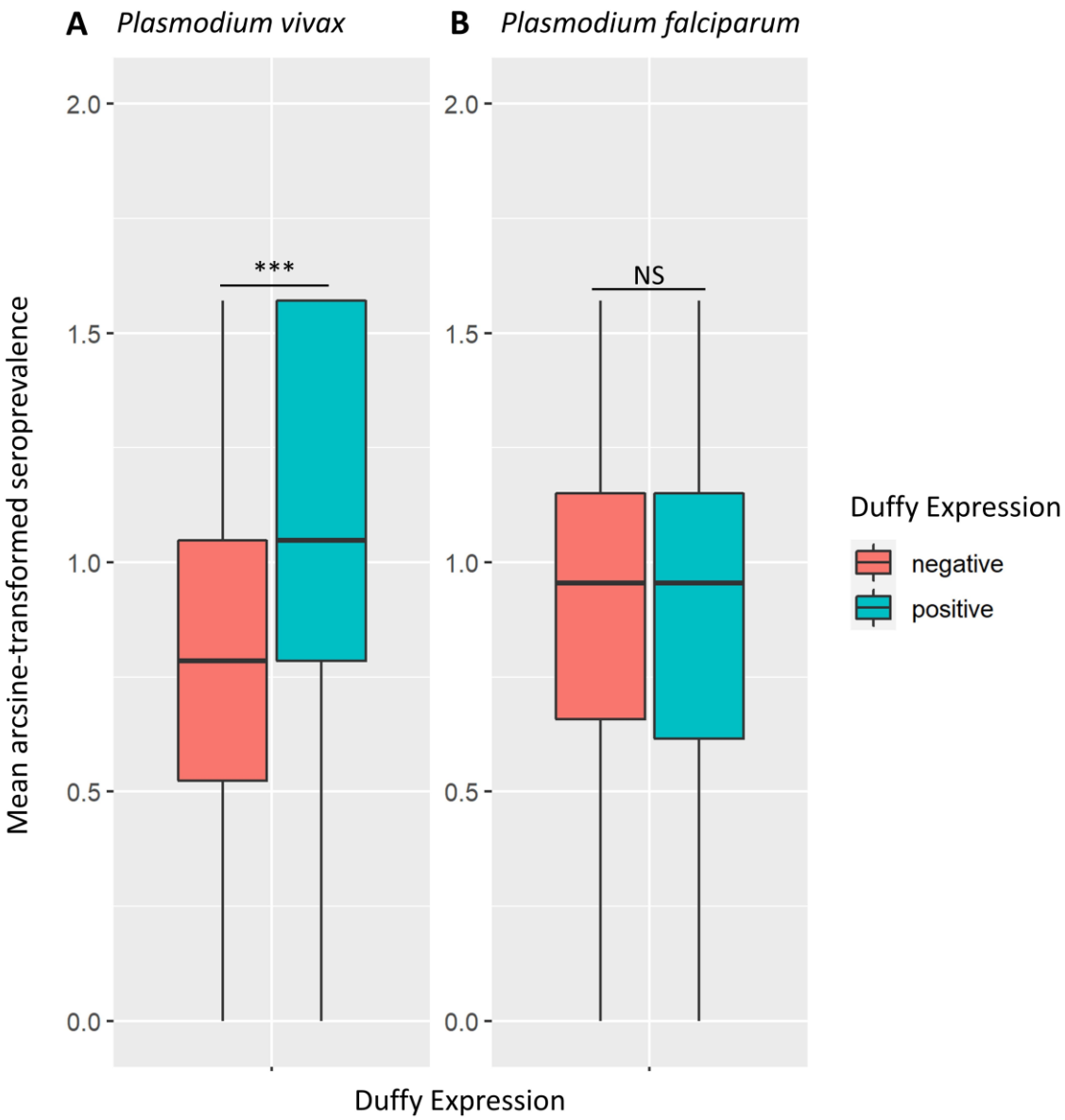
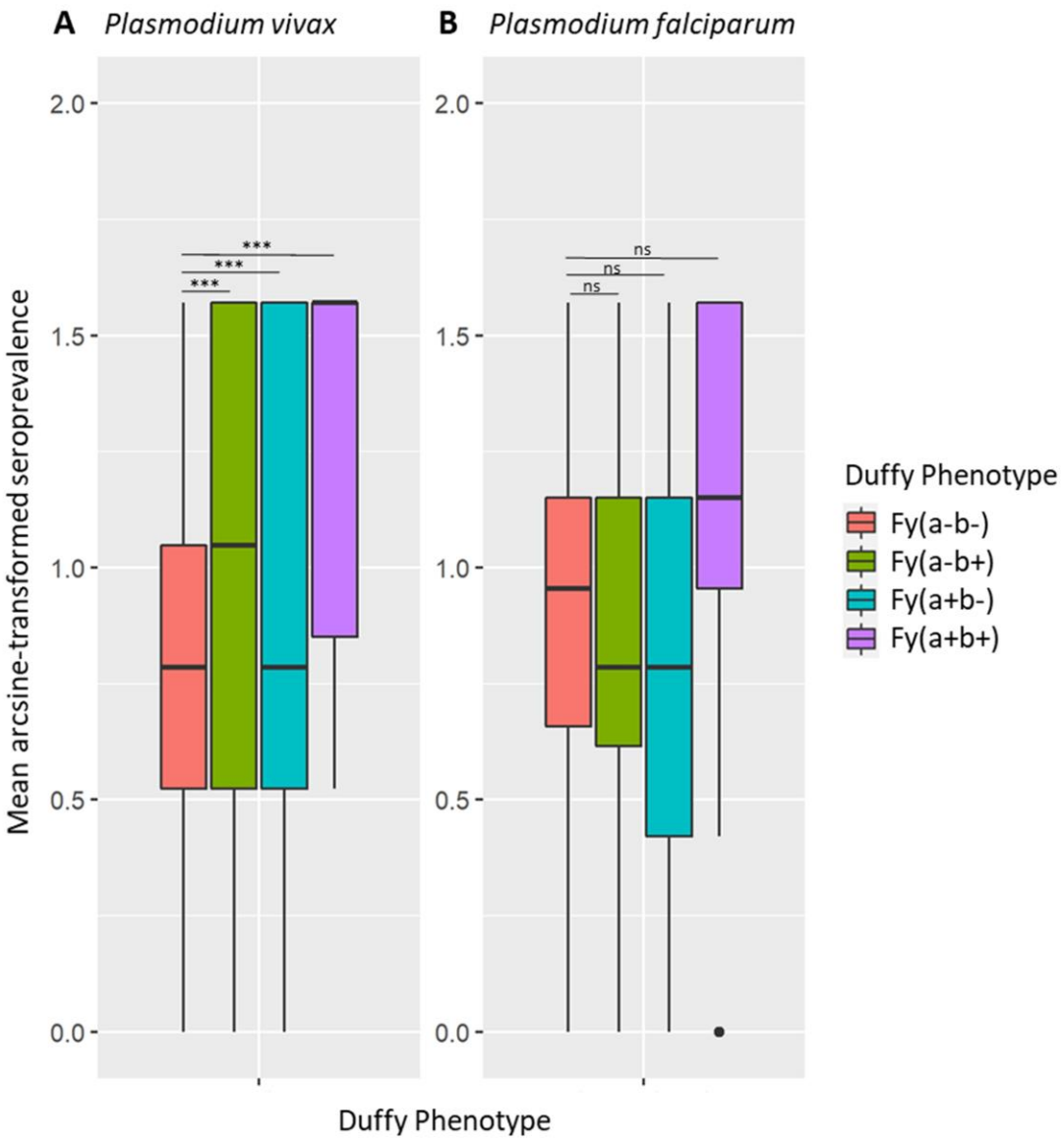


Figure 2 Seroprevalence box-plots of Duffy negative and Duffy positive individuals across 4 *Plasmodium vivax* antigens (A) and 6 *P. falciparum* antigens (B). Significance was obtained via two-sample T-test. NS, non-significant; ***, $P < 0.001$



Supplemental Figure 1 Seroprevalence box-plots of Duffy phenotypes across 4 *Plasmodium vivax* antigens (A) and 6 *P. falciparum* antigens (B). Phenotypes are classified as homozygous negative Fy(a-b-), heterozygous positive Fy(a-b+), Fy(a+b-), and homozygous positive Fy(a+b+). Significance was obtained via Kruskall-Wallis test. ns, non-significant; ***, $P < 0.001$

Serologic responses of *P. vivax* antigens were distinctly different in most cases depending on Duffy phenotype. Homozygous Duffy negative phenotype generally exhibited significantly lower MFI than Duffy positive phenotypes against PvRBP2b, PvEBP2, and PvDBPSal1 antigens, but not for PvMSP1-19 (**Fig. 3**). Conversely, serologic responses did not vary significantly for six *P. falciparum* antigens across Duffy phenotypes. No significant variation was found between Duffy negatives Fy(a-b-) and any Duffy positives Fy(a-b+), Fy(a+b-) and Fy(a+b+) for PfCSP, PfEBA175, and PfMSP3. Interestingly, homozygous negative Fy(a-b-) phenotype was found to show a significantly higher MFI than Fy(a+b-) phenotype for PfRh2a.b, PfMSP1 and PfAMA1 ($P < 0.05$).

The principal component analyses (PCA) showed distinct difference in clustering patterns between Duffy phenotype for *P. vivax* and *P. falciparum* (**Fig. 4**). For *P. vivax,* serologic response non-overlapping confidence ellipses provide evidence that antibody profiles differ distinctly between Duffy negative and Duffy positive individuals whereas *P. falciparum* antibody profiles did not show significant overlap.

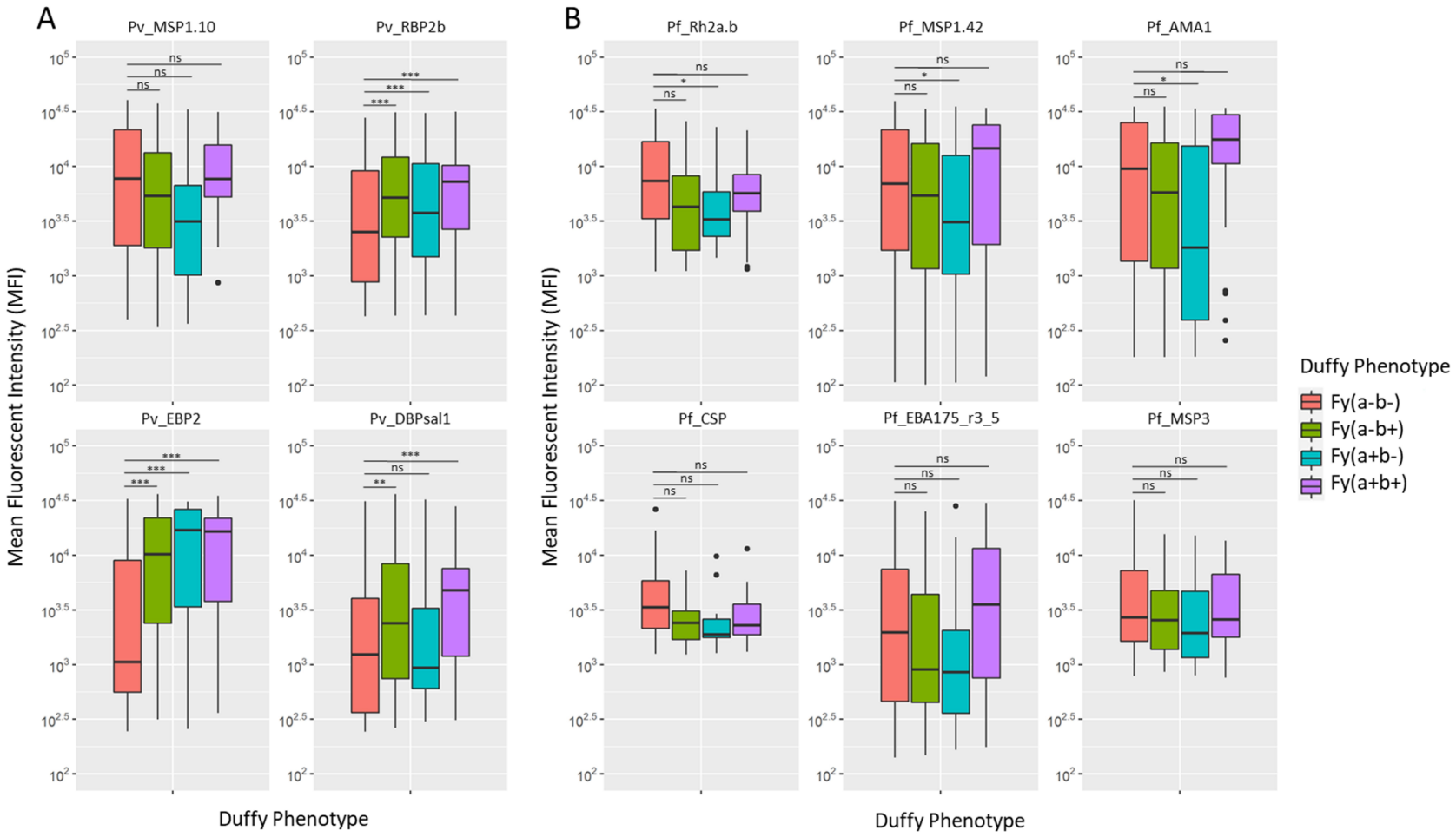


Figure 3 Mean fluorescent intensity box-plot of Plasmodium vivax (A) and P. falciparum antigens (B) across different Duffy phenotypes. Phenotypes are classified as homozygous negative Fy(a-b-), heterozygous positive Fy(a-b+) and Fy(a+b-), and homozygous positive Fy(a+b+). Central box plot represents interquartile range with a median center line. P-values were obtained via Kruskall-Wallis test with Bonferroni correction. ns, non-significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$

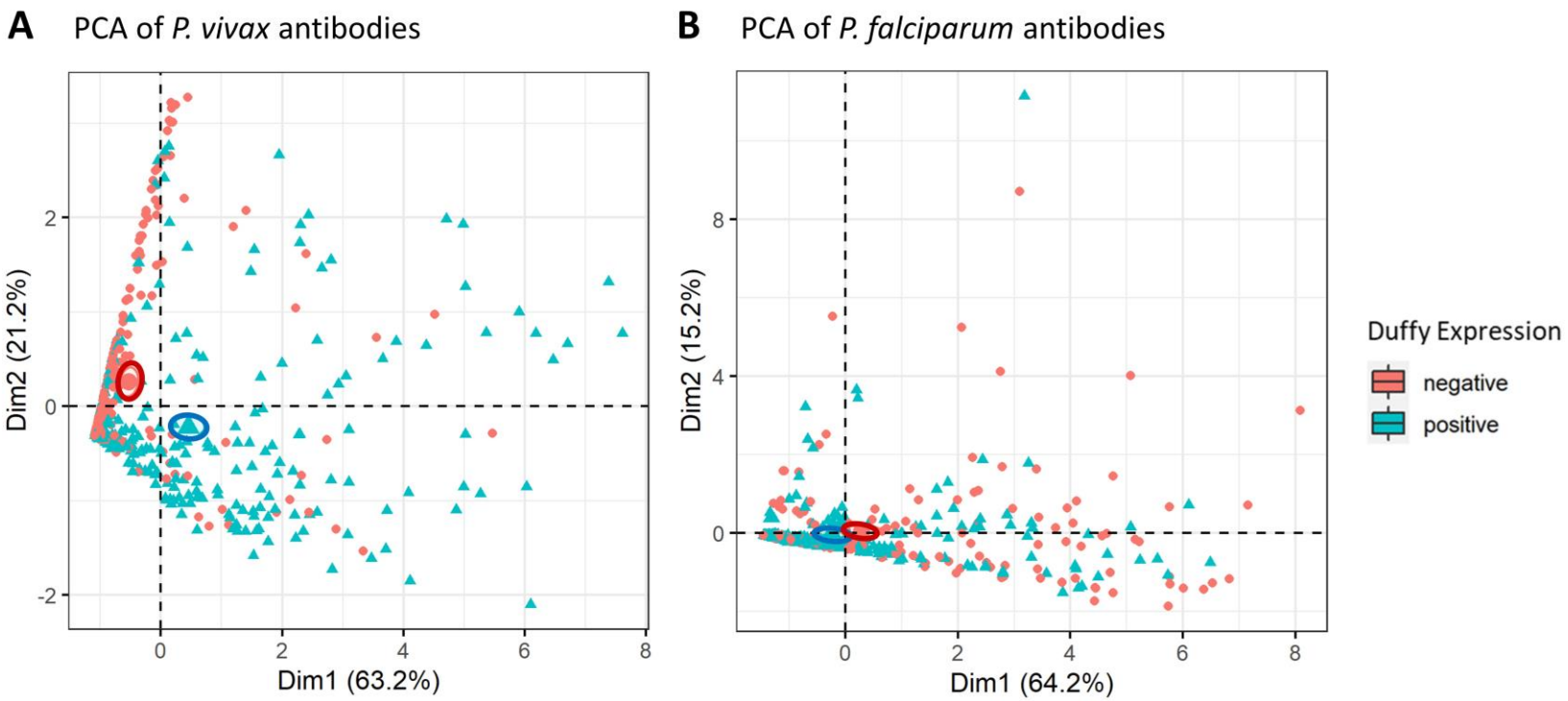


Figure 4 PCA on antibody profiles of Duffy negative and Duffy positive individuals for combined *P. vivax* and *P. falciparum* serologic responses.

Discussion

Overall, our study sites exhibited high levels of Duffy heterogeneity with approximately 45.3% (179/395) individuals being homozygous negative for the Duffy antigen. These levels are notably lower than other areas of the continent, with west and central Africa generally having nearly fixed allele levels of Duffy negativity (>97%)[29]; yet are higher than previous studies on Duffy expression in Ethiopia which recorded levels of Duffy negativity to range from 20-30% in areas around Harar and Jimma[9,30]. Cases of *P. vivax* were found predominantly in Duffy positive individuals, yet did not exhibit significant variation from Duffy negative infection rates. Conversely and interestingly *P. falciparum* infections were mainly found in Duffy negative individuals and exhibited significant variations in infection levels across Duffy phenotypes. Given that these results are derived from very small sample sizes of positive malaria infections caution should be exercised in their interpretation. Antibody levels were generally significantly reduced in homozygous Duffy negative peoples Fy(a-b-), when compared to the three Duffy positive phenotypes for all *P. vivax* target antigens except PvMSP1-19. No significant difference in antibody levels was observed between Duffy negatives and positives for *P. falciparum* antigens. This phenomenon was more clearly observed in seroprevalence analyses where *P. vivax* seroprevalence was significantly lower in Duffy negative individuals, while *P. falciparum* seroprevalence did not differ between Duffy expression. Additionally, PCA visualizations depict clear clustering of Duffy expression for *P. vivax* antigens, yet yield no observable pattern for *P. falciparum*; corroborating the pattern that Duffy negative peoples experience greatly reduced development of *P. vivax* antibodies.

Despite historical evidence that Duffy negativity inhibits infection by *P. vivax*[4-6,31], our study adds to the growing body of research showing that cases of active *P. vivax* infection in

Duffy negative peoples occur in Africa[28,32-35]. We also observed *P. falciparum* infections occurred more frequently in Duffy negative individuals. This pattern could simply be resultant of small number of *P. falciparum* infections in a low-transmission zone. Some studies, however, have linked Duffy binding-like (DBL) domains to the susceptibility of *P. falciparum* infected red blood cell (RBCs) by natural killer (NK) cells[36-38]. Natural killer (NK) cell killing of *P. falciparum* infected RBCs requires specific contact between NK natural cytotoxicity receptors (NCRs) and DBL-1α domain expressed on the *P. falciparum* erythrocyte membrane protein (PfEMP1)[38]. It is possible that there exists an interaction between phenotypic Duffy positivity, and NK innate recognition of DBL domains on *P. falciparum* infected RBCs. The pattern observed in our study is highly novel and interesting; warranting further study and investigation, yet we acknowledge that it arises from very small sample sizes which warrants caution in interpretation and presents a limiting factor in this aspect of our study. Any such conclusions will require further large-scale studies of *P. falciparum* infections across a wide range of Duffy positive and negative individuals. Future studies should seek to determine if there is a relationship between Duffy phenotype and immune response by NK cells to *P. falciparum* infection.

The role that Duffy expression plays in the development of natural immunity is not well understood, and there has been little work done exploring the relationship between Duffy negativity and immune response development given exposure to *P. vivax* infection[39,40]. Our data shows distinct trends of immune response across Duffy expression for *P. vivax* antibodies, but not for *P. falciparum* antibodies. Duffy negative individuals had notably and significantly lower immune responses to *P. vivax* antigens than did Duffy positives, suggesting that successful blood stage infection plays a role in broad immune response, despite the fact that only one of our target

antigens was directly related to the Duffy antigen system (PvDBPSal1). The notable exception to the observed pattern in *P. vivax* antigens was the Pv-merozoite surface protein (PvMSP1-19), which did not vary significantly between Duffy phenotypes, but was also the only class of antigen to be analyzed for both *P. falciparum* and *P. vivax* (PvMSP1-19/PfMSP1-42/PfMSP3). Kano et al. (2018)[41] found that Duffy-negative individuals had a greater chance to acquire high levels of anti-PvMSP1-19 antibodies than Fy(a+/b+) individuals in the Northeast of Amazonas State of Brazil, but we did not find such a phenomenon. The finding on the relationship between Duffy phenotypes and antibody response to *P. vivax* antigens should be interpreted with caution because both the frequency and levels of the antibody response to PvMSP1-19 was low. Previous work has shown that certain immunities established against one *Plasmodium* species are able to cross react against other *Plasmodium* species through commonly shared epitopes, and do indeed exhibit high levels of cross reactivity between *P. vivax* and *P. falciparum* antigens[42]. It is possible, that the lack of antibody variation for PvMSP1-19 between Duffy phenotypes could be attributed to cross-species reactivity of the MSP antigen class, however, several studies have also shown MSP antibodies to be highly species specific[40,43] and therefore further study is required to determine why PvMSP1-19 development does not appear to be influenced by Duffy expression[44,45].

Vivax malaria, though far less pervasive and causal of much less mortality than *P. falciparum* in Africa, presents a notable problem for management and elimination and is increasingly attributed to significant morbidity in endemic areas[46]. Vaccination development efforts for *P. vivax* have been hampered by a multitude of species-specific challenges. These include lifecycle complexities such as relapses and latent-stage hypnozoites, technical challenges in in-vitro culturing for pre-clinical experimentation, and difficulties in both understanding and

establishing functionally-conserved epitopes from antigens at various life stages of infection[47]. Our work clearly shows that Duffy expression is now an additional complexity to be considered in *P. vivax* vaccine development. Areas of high Duffy negativity and indeed nearly all of Sub-Saharan Africa may not respond to *P. vivax* vaccine efforts as predicted if Duffy negative individuals experience significantly reduced development of *P. vivax* antibodies. This information is vitally important to further understanding the epidemiology of *P. vivax* on the continent, and establishing future elimination and management strategies.

Due to the rarity of *P. vivax* infection in Duffy negative people, our study was limited in that very few active cases of *P. vivax* were found, and that even fewer of these occurred in Duffy negative peoples. Our study also primarily examined merozoite blood-stage antigens for *P. vivax* and therefore could be presenting an incomplete picture of antibody profile and development across different lifecycle stages. Future studies would do well to closely examine antibody profiles for individuals both with and without active *P. vivax* infection, and include more pre-erythrocytic serologic markers to assess if reduction in antibody response is occurring only at the erythrocytic stage where Duffy negativity inhibits infection. Inclusion of salivary gland antibodies[48] to known local vectors would also be highly beneficial as they could potentially control for exposure differences and ensure that variations in antibody profiles are not simply resultant of reduced exposure.

Our study adds to the growing body of literature showing that Duffy expression potentially plays a significant role in *P. vivax* naturally acquired immunity, and that Duffy negative people appear to develop a greatly reduced immune response to *P. vivax* compared to Duffy positive individuals. In addition to vaccine developments and implementation, this work is of vital importance as studies and projects continue to use serological tools as a surveillance

method. The use of serologic tools to inform malaria transmission is often highly beneficial and can give a higher resolution of transmission dynamics[14-18]. They are able to account for asymptomatic and sub-patent infections (subclinical malaria) missed by traditional microscopy, RDT and qPCR; and are also able to provide insight into the timeline of both infections and prevalence's by accounting for individual antibody half-life's. Thus, studies aiming to use serologic assays as a surveillance tool in areas of high Duffy admixture should strongly consider the effect of Duffy expression levels in the area of study on their results. Finally, our work adds to the growing evidence that Duffy negativity may no longer offer complete resistance to infection by *P. vivax*[28,31-35,49]. These findings, combined, leave us concerned that as *P. vivax* endemicity in Duffy negative populations continues, these people could not only be more susceptible to infection, but also lack naturally acquired immunity; leaving them more susceptible to greater levels of infection prevalence and morbidity.

Contributors

LB, GY, JK and AD conceived this paper; LB drafted it with contributions from EHS and MCL. LB, DY and PE provided field and antibody data; LB and PE conducted laboratory analyses. Findings were reviewed and interpreted by LB and GY with contributions from EHS, CK, DY, JK and AD.

Acknowledgements

The authors thank the field teams and staff at Jimma University in Ethiopia who were involved in this project. We would also like to acknowledge the local healthcare facilities and thank the communities, villages, and individuals for their willingness to participate. We are grateful to the two anonymous reviewers for their constructive suggestions and comments.

Financial support: This study is funded by the National Institutes of Health (U19 AI129326 and D43 TW001505). The funders did not have a role in study design or development, data collection and analysis, publication process or manuscript preparation.

Potential conflicts of interest: All authors: No reported conflicts of interest.

This work was presented in part at the American Society for Tropical Medicine and Hygiene Annual Meeting 2021

References

1. World Health Organization. World Malaria Report 2021. Geneva, WHO, 2021. https://apps.who.int/iris/rest/bitstreams/1398397/retrieve. Accessed 1 May 2022.
2. Howes RE, Battle KE, Mendis KN, Smith DL, Cibulskis RE, Baird JK, Hay SI, 2016. Global Epidemiology of *Plasmodium vivax.* Am J Trop Med Hyg 95: 15-34.
3. Battle KE, Baird JK, 2021. The global burden of *Plasmodium vivax* malaria is obscure and insidious. PLoS Med 18: e1003799.
4. Miller LH, Mason, SJ, Clyde DF, McGinnis MH, 1976. The resistance factor to *Plasmodium vivax* in Blacks: The Duffy Blood Group Genotype, FyFy. N Engl J Med 295:302-304.
5. Barnwell JW, Nichols ME, Rubinstein R, 1989. In vitro evaluation of the role of the Duffy Blood Group in Erythrocyte Invasion by *Plasmodium vivax*. J Exp Med 169 (5): 1795–1802.
6. Langhi DM, Jr., Bordin JO, 2006. Duffy blood group and malaria. Hematology 11: 389-98.
7. Menard D et al, 2010. *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. Proc Natl Acad Sci USA 107: 5967-71.
8. Howes RE et al, 2015. *Plasmodium vivax* Transmission in Africa. PLoS Negl Trop Dis 9(11): e0004222.
9. Lo E, Hostetler JB, Yewhalaw D, Pearson RD, Hamid MMA, et al 2019. Frequent expansion of *Plasmodium vivax* Duffy Binding Protein in Ethiopia and its epidemiological significance. PLoS Negl Trop Dis 13(9): e0007222.

10. Ketema T, Bacha K, Getahun K, Portillo HAD, Bassat Q, 2021. *Plasmodium vivax* epidemiology in Ethiopia 2000-2020: A systematic review and meta-analysis. PLoS Negl Trop Dis 15: e0009781.
11. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM, 2007. Vivax malaria: neglected and not benign. Am J Trop Med Hyg. 2007 Dec; 77(6 Suppl): 79–87.
12. Baum E et al, 2015. Submicroscopic and asymptomatic *Plasmodium falciparum* and *Plasmodium vivax* infections are common in western Thailand - molecular and serological evidence. Malar J 14: 95.
13. Baird JK, 2021. Basic Research of *Plasmodium vivax* Biology Enabling Its Management as a Clinical and Public Health Problem. Front Cell Infect Microbiol 11: 696598.
14. Corran P, Coleman P, Riley E, Drakeley C, 2007. Serology: a robust indicator of malaria transmission intensity? Trends Parasitol 23: 575-82.
15. Cook J, Reid H, Iavro J, Kuwahata M, Taleo G, Clements A, McCarthy J, Vallely A, Drakeley C, 2010. Using serological measures to monitor changes in malaria transmission in Vanuatu. Malar J 9: 169.
16. Yman V et al, 2016. Antibody acquisition models: A new tool for serological surveillance of malaria transmission intensity. Sci Rep 6: 19472.
17. van den Hoogen LL et al, 2015. Serology describes a profile of declining malaria transmission in Farafenni, The Gambia. Malar J 14: 416.
18. Assefa A et al, 2019. Multiplex serology demonstrate cumulative prevalence and spatial distribution of malaria in Ethiopia. Malar J 18: 246.

19. Ntumngia FB, Schloegel J, Barnes SJ, McHenry AM, Singh S, King CL, Adams JH, 2012. Conserved and variant epitopes of *Plasmodium vivax* Duffy binding protein as targets of inhibitory monoclonal antibodies. Infect Immun 80: 1203-8.
20. Payne RO et al, 2017. Human vaccination against *Plasmodium vivax* Duffy-binding protein induces strain-transcending antibodies. JCI Insight 2(12): e93683
21. Hawaria D, Getachew H, Zhou G, Demissew A, Habitamu K, Raya B, Lee MC, Yewhalaw D, Yan G, 2019. Ten years malaria trend at Arjo-Didessa sugar development site and its vicinity, Southwest Ethiopia: a retrospective study. Malar J 18: 145.
22. Taffese HS, Hemming-Schroeder E, Koepfli C, Tesfaye G, Lee MC, Kazura J, Yan G, Zhou G, 2018. Malaria epidemiology and interventions in Ethiopia from 2001 to 2016. Infec Dis Poverty 7: 6.
23. Bereczy SM, A.; Gil, J. P.; Farnert, A, 2005. Short Report: Rapid DNA Extraction from archive blood spots on filter paper for genotyping *Plasmodium falciparum*. Am J Trop Med Hyg 72: 249-251.
24. Veron V, Simon S, Carme B, 2009. Multiplex real-time PCR detection of *P. falciparum, P. vivax* and *P. malariae* in human blood samples. Exp Parasitol 121: 346-51.
25. Shokoples SE, Ndao M, Kowalewska-Grochowska K, Yanow SK, 2009. Multiplexed real-time PCR assay for discrimination of Plasmodium species with improved sensitivity for mixed infections. J Clin Microbiol 47: 975-80.
26. Wurtz NL et al, 2011. Vivax malaria in Maritania includes infection of a Duffy-negative individual. Malar J 10: 336.

27. Dent AE et al, 2016. Contrasting Patterns of Serologic and Functional Antibody Dynamics to *Plasmodium falciparum* Antigens in a Kenyan Birth Cohort. Clin Vaccine Immunol 23: 104-16.

28. Weber GE et al, 2017. Sero-catalytic and Antibody Acquisition Models to Estimate Differing Malaria Transmission Intensities in Western Kenya. Sci Rep 7: 16821.

29. Howes RE et al, 2011. The global distribution of the Duffy blood group. Nat Commun 2: 266.

30. Woldearegai TG, Kremsner PG, Kun JF, Mordmuller B, 2013. *Plasmodium vivax* malaria in Duffy-negative individuals from Ethiopia. Trans R Soc Trop Med Hyg 107: 328-31.

31. Wertheimer SP, Barnwell JW, 1989. *Plasmodium vivax* interaction with the Human Duffy Blood Group Glycoprotein: Identification of a parasite receptor-like protein. Exp Parasitol 69: 340-350.

32. Zimmerman PA, 2017. *Plasmodium vivax* Infection in Duffy-Negative People in Africa. Am J Trop Med Hyg 97: 636-638.

33. Mendes C, Dias F, Figueiredo J, Mora VG, Cano J, de Sousa B, do Rosario VE, Benito A, Berzosa P, Arez AP, 2011. Duffy negative antigen is no longer a barrier to *Plasmodium vivax*--molecular evidences from the African West Coast (Angola and Equatorial Guinea). PLoS Negl Trop Dis 5: e1192.

34. Ngassa Mbenda HG, Das A, 2014. Molecular evidence of *Plasmodium vivax* mono and mixed malaria parasite infections in Duffy-negative native Cameroonians. PLoS One 9: e103262.

35. Kepple D et al, 2021. *Plasmodium vivax* From Duffy-Negative and Duffy-Positive Individuals Share Similar Gene Pools in East Africa. J Infect Dis 224: 1422-1431.

36. Oguariri RM, Borrmann S, Klinkert MQ, Kremsner PG, Kun JF, 2001. High prevalence of human antibodies to recombinant Duffy binding-like alpha domains of the *Plasmodium falciparum*-infected erythrocyte membrane protein 1 in semi-immune adults compared to that in nonimmune children. Infect Immun 69: 7603-9.
37. Howell DP, Levin EA, Springer AL, Kraemer SM, Phippard DJ, Schief WR, Smith JD, 2008. Mapping a common interaction site used by *Plasmodium falciparum* Duffy binding-like domains to bind diverse host receptors. Mol Microbiol 67: 78-87.
38. Mavoungou E, Held J, Mewono L, Kremsner PG, 2007. A Duffy binding-like domain is involved in the NKp30-mediated recognition of *Plasmodium falciparum*-parasitized erythrocytes by natural killer cells. J Infect Dis 195: 1521-31.
39. Wang R et al, 2005. Immune responses to *Plasmodium vivax* pre-erythrocytic stage antigens in naturally exposed Duffy-negative humans: a potential model for identification of liver-stage antigens. Eur. J. Immunol 35: 1859-1868.
40. Wang et al, 2016. Naturally Acquired Antibody Responses to *Plasmodium vivax* and *Plasmodium falciparum* Merozoite Surface Protein 1 (MSP1) C-Terminal 19 kDa Domains in an Area of Unstable Malaria Transmission in Southeast Asia. PLoS One 11: e0151900.
41. Kano FS et al, 2018. Susceptibility to *Plasmodium vivax* malaria associated with DARC (Duffy antigen) polymorphisms is influenced by the time of exposure to malaria. Sci Rep 8: 1-14.
42. Muh F et al, 2020. Cross-species reactivity of antibodies against *Plasmodium vivax* blood-stage antigens to *Plasmodium knowlesi*. PLoS Negl Trop Dis 14: e0008323.

43. Priest JW, Plucinski MM, Huber CS, Rogier E, Mao B, Gregory CJ, Candrinho B, Colborn J, Barnwell JW, 2018. Specificity of the IgG antibody response to *Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae,* and *Plasmodium ovale* MSP119 subunit proteins in multiplexed serologic assays. Malar J 17: 417.

44. Herrera SV et al, 2005. Antibody Response to *Plasmodium vivax* Antigens in Fy-Negative Individuals From the Colombian Pacific Coast. Am J Trop Med Hyg 73: 44-49

45. Maestre A et al, 2010. Acquired Antibody Responses against *Plasmodium vivax* Infection Vary with Host Genotype for Duffy Antigen Receptor for Chemokines (DARC). PLoS One 5: e11437

46. Cui L, Brashear A, Menezes L, Adams J. Elimination of *Plasmodium vivax* malaria: problems and solutions. In: Rodriguez-Morales AJ, editor. Current Topics and Emerging Issues in Malaria Elimination. IntechOpen; 2021:1–27

47. De SL, May S, Shah K, Slawinski M, Changrob S, Xu S, Barnes SJ, Chootong P, Ntumngia FB, Adams JH, 2021. Variable immunogenicity of a vivax malaria blood-stage vaccine candidate. Vaccine 39: 2668-2675.

48. Kearney EA, Agius PA, Chaumeau V, Cutts JC, Simpson JA, Fowkes FJI, 2021. Anopheles salivary antigens as serological biomarkers of vector exposure and malaria transmission: A systematic review with multilevel modelling. eLife 10:e73080

49. Wilairatana P, Masangkay FR, Kotepui KU, De Jesus Milanez G, Kotepui M, 2022. Prevalence and risk of *Plasmodium vivax* infection among Duffy-negative individuals: a systematic review and meta-analysis. Sci Rep 12: 3998.

Figure Captions

Figure 1. Map of the two selected Ethiopian study sites showing study resident locations and irrigation boundaries.

Figure 2: Box-plots of seroprevalence for Duffy negative and Duffy positive individuals across 4 Plasmodium vivax antigens (A) and 6 *P. falciparum* antigens (B). Significance was obtained via two-sample t-test. NS, non-significant; ***, $P < 0.001$.

Figure 3: Mean fluorescent intensity box-plot of *Plasmodium vivax* (A) and *P. falciparum* antigens (B) across different Duffy phenotypes. Phenotypes are classified as homozygous negative Fy(a-b-), heterozygous positive Fy(a-b+) and Fy(a+b-), and homozygous positive Fy(a+b+). Central box plot represents interquartile range with a median center line. P-values were obtained via Kruskall-Wallis test with Bonferroni correction for multiple comparisons. NS, non-significant: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Figure 4: PCA on antibody profiles of Duffy negative and Duffy positive individuals for combined *P. vivax* and *P. falciparum* serologic responses.

Supplementary Materials

Supplemental Figure 1: Seroprevalence box-plots of Duffy phenotypes across 4 *Plasmodium vivax* antigens (A) and 6 *P. falciparum* antigens (B). Phenotypes are classified as homozygous negative Fy(a-b-), heterozygous positive Fy(a-b+), Fy(a+b-), and homozygous positive Fy(a+b+). Significance was obtained via Kruskall-Wallis test. NS, non-significant; ***, $P < 0.001$

Supplementary Table 1: Sample distribution of Duffy phenotypes and expression relative to 13 observed genotypes from Duffy Fy alleles from 395 subjects of all ages (FyA, FyB, FyB*, FyX2, FyA^{ES} and FYB^{ES}).

CHAPTER 2

EPIDEMIOLOGY OF *PLASMODIUM VIVAX* IN DUFFY NEGATIVES AND DUFFY POSITIVES FROM COMMUNITY AND HEALTH CENTER COLLECTIONS IN ETHIOPIA

*Accepted to **Malaria Journal** with Minor Revision*

Resubmitted with Revision on November 20th 2023

Lauren Bradley[1], Delenasaw Yewhalaw[2,3], Elizabeth Hemming-Schroeder[4], Brook Jeang[1], Ming-Chieh Lee[1], Endalew Zemene[3], Teshome Degefa[5], Eugenia Lo[6], Christopher King[7], James Kazura[7], Guiyun Yan*[1,8]

[1] Department of Ecology and Evolutionary Biology, School of Biological Sciences, University of California Irvine, Irvine, CA 92697, USA

[2] Department of Medical Laboratory Sciences and Pathology, College of Health Sciences, Jimma University, Jimma 5195, Ethiopia

[3] Tropical and Infectious Diseases Research Center, Jimma University, Jimma, Ethiopia

[4] Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523, USA

[5] School of Medical Laboratory Sciences, Faculty of Health Sciences, Jimma University, Jimma, Ethiopia

[6]Department of Microbiology and Immunology, Drexel University, Philadelphia, PA 19104, USA

[7]Center for Global Health and Disease, Case Western Reserve University, Cleveland, OH 44106, USA

[8]Program in Public Health, College of Health Sciences, University of California at Irvine, Irvine, CA 92697, USA

*Corresponding author: Guiyun Yan, guiyuny@hs.uci.edu. (949) 413-0935

Word Counts: Abstract 299, Text 3,435 Figures: 3, Tables: 3

Running Title: Malaria epidemiology and Duffy expression in Ethiopia

Abstract

Background Malaria remains a significant cause of morbidity and mortality in Ethiopia with an estimated 3.8 million cases in 2021 and 61% of the population living in areas at risk of malaria transmission. Throughout the country *Plasmodium vivax* and *P. falciparum* are co-endemic, and Duffy expression is highly heterogeneous. The public health significance of Duffy negativity in relation to *P. vivax* malaria in Ethiopia, however, remains unclear. This study seeks to explore the prevalence and rates of *P. vivax* malaria infection across Duffy phenotypes in clinical and community settings.

Methods A total of 9,580 and 4,667 subjects from community and health facilities from a malaria endemic site and an epidemic-prone site in western Ethiopia were enrolled and examined for *P. vivax* infection and Duffy expression from February 2018 to April 2021. Association between Duffy expression, *P. vivax* and *P. falciparum* infections were examined for samples collected from asymptomatic community volunteers and symptomatic subjects from health centers.

Results Infection rate of *P. vivax* among Duffy positives was 2-22 fold higher than Duffy negatives in asymptomatic volunteers from the community. Parasite positivity rate was 10-50 fold higher in Duffy positives than Duffy negatives among samples collected from febrile patients attending health centers and mixed *P. vivax* and *P. falciparum* infections were significantly more common than *P. vivax* mono infections among Duffy negative individuals. *P. vivax* parasitemia measured by 18sRNA parasite gene copy number was similar between Duffy positives and Duffy negatives.

Conclusions Duffy negativity does not offer complete protection against infection by *P. vivax*, and cases of *P. vivax* in Duffy negatives are widespread in Ethiopia, being found in

asymptomatic volunteers from communities and in febrile patients from health centers. These findings offer evidence for consideration when developing control and intervention strategies in areas of endemic *P. vivax* and Duffy heterogeneity.

Keywords Malaria, Duffy blood group, *Plasmodium vivax*, qPCR, gene copy number

Background

In spite of significant progress towards malaria control in the past two decades, malaria remains a major cause of mortality and morbidity in Africa[1]. According to the World Health Organization, *Plasmodium vivax* and *P. falciparum* contributed to approximately 700 thousand and 230 million cases, respectively, in Africa in 2021[2]. In Ethiopia there were an estimated 3.8 million cases in 2021 and 61% of the population resides in areas with endemic transmission[2, 3]. *Plasmodium vivax* and *P. falciparum* account for approximately 33% and 67% of all malaria cases, respectively, and it is one of only a few countries in Africa where *P. vivax* remains consistently endemic[4].

Current endemicity of *P. vivax* in Africa correlates with areas of high heterogeneity in Duffy expression[5,4]. The Duffy antigen receptor for chemokines (DARC), often referred to as the Fy glycoprotein, is a silent heptahelical chemokine receptor located on chromosome 1 and expressed on the surface of erythrocytes. DARC has been recognized as the binding antigen of *P. vivax,* and a single point mutation located in the GATA-1 transcription factor binding site of the DARC gene promoter (-67T>C) causes this receptor to not be expressed, resulting in a Duffy negative phenotype[6,7]. The absence of this receptor on red blood cells has been shown to confer resistance to blood-stage infection by *P. vivax*[5, 8, 9]. This negative phenotype is nearly fixed in sub-Saharan Africa, correlating with the general lack of endemic *P. vivax* on the continent. Despite this established dogma, cases of *P. vivax* are being found in confirmed Duffy negative individuals throughout different African countries[10-14]. In addition to Duffy blood group, other population-level factors that influence *P. vivax* epidemiology include climatic conditions, age, socioeconomic status, access to healthcare, malaria control measures[15, 16]. Whether discovery of increasing number of *P. vivax* infections in Duffy negatives results from more recent research on

P. vivax in Africa or from new *P. vivax* genetic variants, the data suggest that Duffy negativity no longer confers complete resistance to blood-stage *P. vivax* infection[17, 18]. Furthermore, Duffy negatives generally develop reduced natural immunity to *P. vivax* blood-stage antigens[19-21].

There remains little information on the public health significance of *P. vivax* infection in individuals lacking the Duffy antigen in Africa. For example, how frequent are Duffy negative individuals infected with *P. vivax* compared to Duffy positive individuals from the same communities? How frequently does *P. vivax* contribute to clinical malaria among Duffy negatives compared to Duffy positives from areas of same endemicities? This population-based study aimed to address these questions in two locations with varying malaria endemicities in southwestern Ethiopia, using samples from communities and health centers

Methods

Study Sites

Samples were collected from two study sites, Arjo-Didessa and Gambella, both located in western Ethiopia (**Figure 1**) with a rainy season lasting from May to October. The Arjo-Didessa sugarcane plantation is located within the Oromia Regional State 395 km west of the Ethiopian capital Addis Ababa and the area covers most of the Arjo-Didessa sugarcane irrigation scheme. It is located at an elevation ranging from 1200m to 1500m above sea level, and comprises 15 villages in 3 districts (Jimma Arjo, Bedele District, and Dado Hana District). It contains 1 health center, 3 health posts, and 9 command posts which are smaller scale health posts located within the temporary residential areas formed by migrant workers. The sugarcane plantation was formerly the Didessa Wildlife Sanctuary before 2006 when the state owned sugarcane plantation was developed to supply the proximal sugarcane factory. It is one of the biggest sugarcane developments in the country, currently covering 5000 hectares with plans to expand to 80,000 hectares[22-24]. Gambella is located in the Abobo District in the Gambella Regional State, 811 km west of Addis Ababa. The area's elevation ranges from 400-600m above sea level and as of 2019 had a population of 20,080. The main socio-economic activity in the area is farming of cotton, maize and sorghum, or working fruit plantations to produce mango, papaya and banana. Additionally, the Alwero Dam provides fishing opportunities and employs approximately 2000 people at a large-scale rice irrigation scheme that currently spans 3,000 hectares with plans to expand to 10,000 hectares. The district comprises 19 villages containing 4 health centers and 16 health posts[25, 26]. The populations at both locations primarily consisted of local villagers and migrant workers with long-term residency. These sites were chosen for the study as both areas have high levels of Duffy admixture, and continuous *P. vivax* endemicity[12, 27].

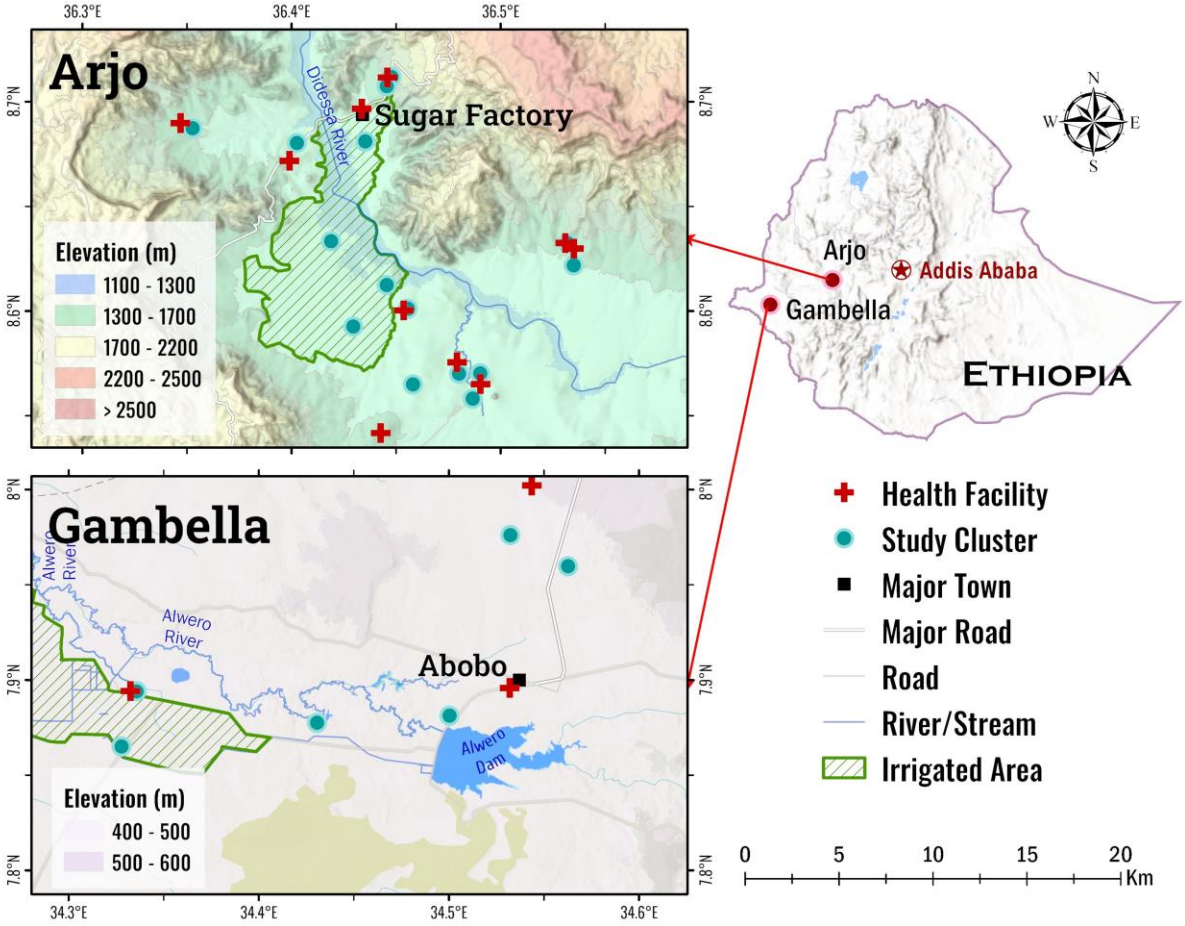


Figure 1 Figure 1 Map showing location of both study sites; Arjo and Gambella, in western Ethiopia. Includes locations of study clusters, health facilities, and major towns in the regions. The map was created with Esri ArcGIS Pro 3.1 with data sources from our field survey, and elevation data from NASA SRTM v3 (https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003).

Blood sample collection

Finger prick blood samples were collected throughout both study sites from community members who were asymptomatic during cross-sectional surveys, and from febrile volunteers attending health centers from February 2018 to April 2021. From each individual, a total of 3 blood spots, equaling ~50ul, was pressed to Whatman 3MM filter paper for storage and transportation. For community collections all residents willing to participate were included in the study and provided signed informed consent and/or assent for minors under 18 years old. At the time of sample collection, for both clinical and community samples, the age and sex of participants were recorded when possible. Dried blood spots were transported to the University of California Irvine and stored at 4°C.

DNA Extraction and qPCR of *Plasmodium* species

Parasite DNA was extracted from dried blood spots (DBS) using a standardized saponin/chelex method[28]. DNA was eluted to ~200ul molecular grade water stored at 4°C in the short term or -20°C for long term storage. *Plasmodium* species-specific primers and probes were used to amplify the 18sRNA gene using a previously described protocol with modification[29]. Real time PCR was conducted at a total volume of 12ul containing; 6µl ThermoFisher FastAdvanced MM (2X), 0.5µl of each species-specific probe, 0.4µl of each forward and reverse species-specific primer, and 2µl parasite DNA. Reaction conditions were set as follows: 50°C for 2 min, and 45 cycles of 95°C for 2 min, 95°C for 3 seconds, 60°C for 30 seconds and run on a QuantStudio 3 Real-Time PCR System.

Duffy Sequencing

An approximately ~600-bp fragment of the human DARC gene encompassing the -33rd nucleotide position located in the GATA-1 box of the promoter region was amplified sequenced

following established protocols to assess Duffy expression[27, 30, 31]. Specifically, the total volume for amplification was a 20ul reaction mixture containing; 10µl DreamTaq Green PCR MM (2X), 0.3µl of each forward and reverse primer, and 2µl genomic DNA. Thermocycling conditions were set at; 94°C for 2 minutes, 35 cycles of 94°C for 30 seconds, 61 for 30 seconds, and 65 for 40 seconds followed by a 2-minute extension at 65°C. Five microliters of PCR product were run on a 1.5% agarose gel to confirm amplification. PCR product which had successful amplification was cleaned enzymatically to remove remaining primers and dNTPs; 2µl SAP and 0.2µl XO1 per was added to PCR product and cleaned via the following thermocycling conditions; 37°C for 15 minutes, 80°C for 15 minutes, and then held at a 4C extension. Sanger sequencing was conducted by Retrogen Inc. using forward primers and chromatogram results were visually analyzed via Chromas for a T→C mutation at the 33rd nucleotide position indicating Duffy negativity. Only samples positive for *P. vivax* mono and mixed infections were sequenced for Duffy expression.

Data Analysis

Malaria prevalence was calculated for both study settings at each study site separately. Overall prevalence of both *Plasmodium* species we compared between study sites for community and clinical collections via the Chi-Square test for independence. Given that only *P. vivax* positive samples were sequenced for Duffy expression, rates of Duffy negativity in the population was not directly assessed for this study, the overall rate of *P. vivax* in Duffy negative and Duffy positive individuals was calculated by dividing the number of *P. vivax* infections by the expected number of Duffy negative and positive individuals at each site. Expected Duffy negative and positive populations were calculated by multiplying the total number of samples by the Duffy negativity rate in Arjo (43.6%) and Gambella (45.9%) as determined in a previously

published study[21]. The ratio of mixed (Pv +Pf) to mono (Pv) infections was determined for each study setting, community and health facility, for both Duffy negatives and Duffy positives. Comparisons of the rate of *P. vivax* in Duffy negatives to Duffy positives, and the ratio of mixed to mono infections for Duffy negatives and positives were made via Fisher's Exact test for both community and health facility collected samples. Parasite Gene Copy Number (GCN) was calculated from qPCR Cycle threshold (Ct) values via standard curve to estimate parasite density. Log_{10} transformed GCN was compared between community and health facility settings for both *P. vivax* and *P. falciparum* via two-sample t-test, and between Duffy negatives and Duffy positives for both settings via Fisher's Exact test.

Results

Prevalence of *P. vivax* Across Study Sites, Collection Method and Duffy Expression

A total of 14,247 dried blood spots were collected from two study sites in southwestern Ethiopia (**Figure 1**) from February 2018 to December 2021. Asymptomatic community collections were made via cross-sectional surveys conducted during the spring and late-fall of each year and making up 9,580 of the total dried blood spots. The remaining 4,667 samples were from symptomatic infections collected from health clinics and facilities in the regions via passive case detection (PCD). In total 344 DBS were positive for only *P. vivax*, 937 for only *P. falciparum* and 35 samples exhibited a mixed infection being positive for both *P. vivax* and *P. falciparum* (**Table 1**). A total of 7,519 of these DBS were collected from Arjo; 5,454 from cross-sectional surveys and 2,065 from passive case detection. In Gambella 6,728 samples were collected in total; 4,126 were collected from the community during cross-sectional surveys and 2,602 via passive case detection (**Table 1**). Overall, *P. vivax* and *P. falciparum* infection rate was significantly higher in Gambella than in Arjo ($P < 0.001$ for both species).

Table 1 PCR Prevalence of *Plasmodium vivax* (Pv) and *P. falciparum* (Pf) infections among community-based asymptomatic sampling and sample positivity among febrile patients detected by passive case surveillance from health centers in two sites in Ethiopia.

Settings	Site	Samples (n)	Total *Plasmodium* infections	Mixed Pv and Pf infections	Pv mono infections	Pf mono infections	P-value*
Community	Arjo	5454	19 (0.35%)	0	3 (0.06%)	16 (0.29%)	< 0.05
	Gambella	4126	424 (10.28%)	8 (0.19%)	133 (3.22%)	283 (6.86%)	< 0.001
	Total	9580	443 (4.62%)	8 (0.08%)	136 (1.42%)	299 (3.12%)	
Health Center	Arjo	2065	313 (15.16%)	9 (4.36%)	114 (5.52%)	190 (9.20%)	< 0.001
	Gambella	2602	560 (21.52%)	18 (0.69%)	94 (3.61%)	448 (17.22%)	< 0.001
	Total	4667	873 (18.79%)	27 (0.57%)	208 (4.46%)	638 (13.67%)	

*Fishers exact test comparison between Pv and Pf mono infection rate

Duffy genotyping was performed only on *P. vivax* mono infections and mixed *P. vivax* and *P. falciparum* infections across all study sites and collection methods (**Table 2**). Of the 379 *P. vivax* positive and mixed-species infections, 345 were successfully sequenced at the T33C promoter of the GATA-1 transcription factor. Among the community-based cross-sectional samples, infection rate of *P. vivax* among the Duffy negatives and positives was low and similar in Arjo, but significantly higher infection rate was found in Gambella among Duffy positives than Duffy negatives (5.6% vs. 0.26%, $P < 0.001$; **Table 2**). Similarly, sample positivity rate was more than 10-50 fold higher in Duffy positive than Duffy negatives in both sites among samples collected from the health center settings (**Table 2**), suggesting a much reduced *P. vivax* burden among Duffy negative people in febrile patients.

Table 2 Rate of *Plasmodium vivax* (Pv) infections among Duffy negative and Duffy positive individuals in both community-based asymptomatic and passive case surveillance from health centers at two study sites in Ethiopia.

Setting	Site	Duffy Negatives	Duffy Positives	Rate of Pv in Duffy Negatives	Rate of Pv in Duffy Positives	P-value*
Community	Arjo	2525	2929	0.04% (1/2525)	0.07% (2/2929)	> 0.05
	Gambella	1894	2232	0.26% (5/1894)	5.60% (125/2232)	< 0.001
Health Center	Arjo	956	1109	0.21% (2/956)	10.64% (118/1109)	< 0.001
	Gambella	1194	1408	0.59% (7/1194)	6.04% (85/1408)	< 0.001

*P-value calculated via Fishers Exact Test for comparing Rate of Pv in Duffy Negatives to Duffy Positives

Interestingly, we found that a considerably large proportion of malaria infections were mixed species infection. Among the 133 successfully sequenced community-based samples, eight out of 133 (6.0%) malaria infections were mixed species and 26 out of 212 (12.3%) samples were mixed infections from the health center settings (**Table 3**). Among the Duffy negatives, *P. vivax* was found more frequently found in the form of mixed-species infection than mono infections, whereas mono *P. vivax* infections were far more common in Duffy positives. In the community asymptomatic samples, the ratio of mixed species infection to *P. vivax* mono infection was 0.5 among Duffy negatives, but this ratio was reduced to 0.05 in Duffy positives ($P < 0.05$; **Table 3**). In febrile samples from health centers the ratio of mixed species infection to *P. vivax* mono infection was 3.5 among Duffy negatives, far greater than the ratio observed in Duffy positive (0.10; $P < 0.001$). This data strongly suggests that in Duffy negative individuals *P. vivax* is more frequently found in mixed infections compared to *P. vivax* only mono infections.

Table 3 Distribution of Duffy phenotypes across *Plasmodium vivax* (Pv) and Mixed (Pv and Pf) infections in both community-based asymptomatic and passive case surveillance via health centers from two study sites in Ethiopia

Setting	Infection	n	Duffy negative	Duffy positive	Ratio of mixed species infection to *P. vivax* only infections		Fisher's exact test
					Duffy negative	Duffy positive	
Community	Pv	125	4	121	0.5	0.05	$P < 0.05$
	Pv + Pf	8	2	6			
Health Center	Pv	186	2	184	3.5	0.10	$P < 0.001$
	Pv + Pf	26	7	19			

P. vivax Parasitemia in Community and Clinical Samples and Across Duffy Expressions

Analyses of qPCR data revealed significant differences in the parasitemia between cross-sectional samples without clinical symptoms and clinical samples collected during passive case detection from health centers for both *P. vivax* and *P. falciparum*. In both *P. vivax* and *P. falciparum* infections parasitemia was significantly higher in samples collected via passive case detection than via cross-sectional survey ($P < 0.001$, **Figure 2**). Symptomatic *P. vivax* infections showed a geometric mean gene copy number (GCN) of 2.03 parasites/µl, which was significantly higher than the asymptomatic *P. vivax* infections, which had a geometric mean of 0.94 parasites/µl ($P < 0.001$, **Figure 2**). Similarly, symptomatic *P. falciparum* infections exhibited a geometric mean of 1.67 parasites/µl, which was significantly higher than the asymptomatic *P. falciparum* infections which had a mean of 0.90 parasites/µl ($P < 0.001$). Community Duffy-negative and Duffy-positive samples exhibited a similar parasitemia, with a GCN of 1.28 and 0.93 parasites/µl respectively ($P > 0.05$, **Figure 3**). Similarly, PCD Duffy-negative and Duffy positive samples showed a mean GCN of 1.93 and 2.07 parasites/µl respectively ($P > 0.05$, **Figure 3**). These data do not include four Duffy negative samples as their gene copy numbers fell just outside of our standard curve based cutoff range. Given the substantial differences in sample sizes between Duffy-negatives and Duffy-positives it is possible that the lack of significance observed here is indeed due to a small samples size of Duffy negatives.

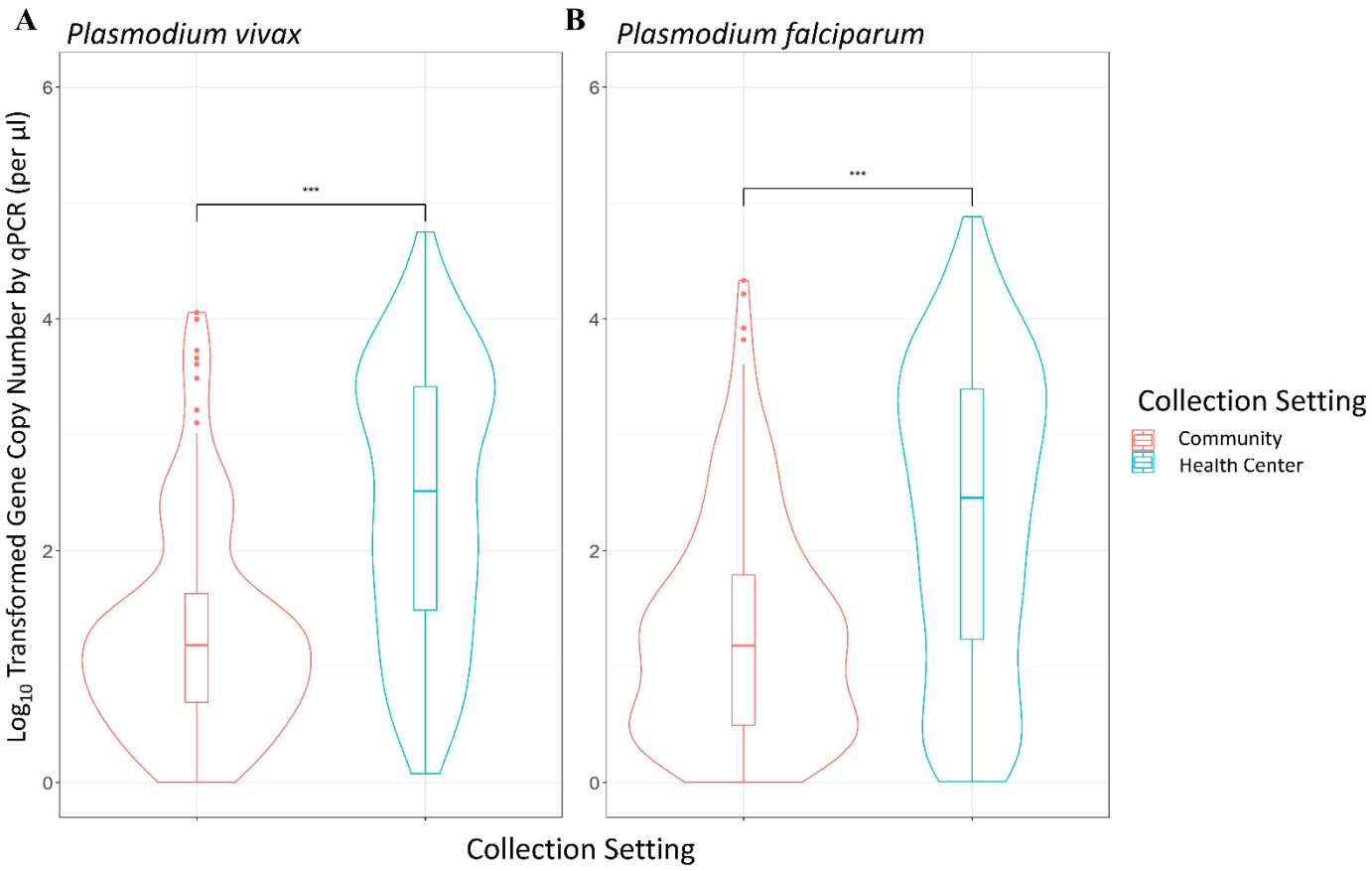


Figure 2 Log-transformed parasite gene copy number of community and clinical samples. Violin plots of the log-transformed malaria parasite gene copy number of samples collected from asymptomatic communities and febrile patients of all ages from health centers in Ethiopia. (A) *Plasmodium vivax*, and (B) *P. falciparum* by qPCR for individuals of all ages. The central box represents the interquartile range with the median shown as the center line in the box. ***, $P < 0.001$ based on Fisher's exact test.

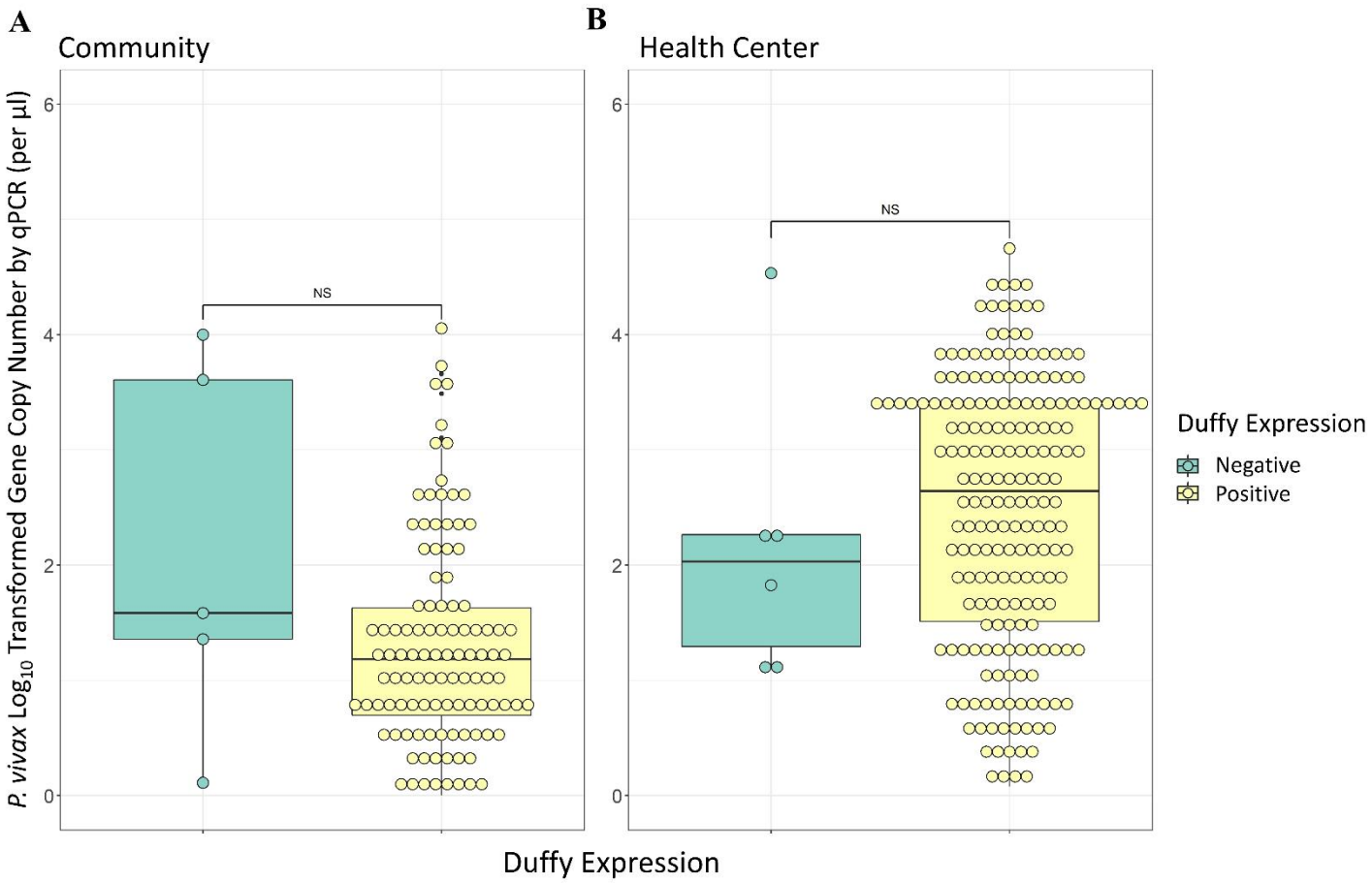


Figure 3 Box plots of the log-transformed parasite gene copy number for Duffy negative and Duffy positive individuals. Box plots of the log-transformed Plasmodium vivax parasite gene copy number for Duffy negative and Duffy positive individuals of all ages. A) asymptomatic community samples; and B) febrile malaria samples from health centers. Box plots represent the interquartile range with the median expressed as the center line. NS, non-significant based on Fisher's exact test.

Discussion

This study sought to examine *P. vivax* malaria burden in Duffy negative individuals at two field sites with similar proportion of Duffy negativity, but different malaria endemicities in southwest Ethiopia. We found, firstly, that *P. vivax* posed a significant health burden at both sites, but was far more prevalent in the community in Gambella than in Arjo where infection prevalence was over 50 times higher. In febrile patients *P. vivax* was found more often in Arjo than in Gambella; however, this difference was much less drastic than in the community with Arjo exhibiting only 1.5 times more *P. vivax* clinical infections than Gambella. Across both sites and collection settings *P. vivax* was found far less frequently in Duffy negatives than Duffy positives. In the community Duffy positives had approximately 2 and 22-fold greater infection rate of *P. vivax* than Duffy negatives at Arjo and Gambella, respectively. In febrile patients and samples collected from health facilities this trend was even more apparent; in Arjo and Gambella Duffy positives exhibited a 51 and 10-fold greater positivity rate of *P. vivax* infections, respectively, than Duffy negatives. The variations in rate of infection were highly significant for samples from health centers at both sites, but only significant for community samples from Gambella. The lack of significance in Arjo community samples could potentially be due to the small sample size as only three *P. vivax* infections were found in the community in Arjo. These strongly suggest that *P. vivax* infections, though commonly found in Duffy negative individuals, are still predominantly occurring in Duffy positive people. Despite the significant variations in rate of *P. vivax* infection between Duffy expressions, we did not observe significant differences in parasitemia between Duffy negatives and Duffy positives in either the community or health centers. Perhaps most interestingly this study highlights a pattern of mixed versus mono infections related to Duffy negativity. We found the ratio of mixed to mono *P. vivax* infections

among Duffy negative and positive individuals we found that Duffy negatives exhibited a 10 and 35-fold greater ratio of mixed to mono infections than Duffy positives in both the community and clinical settings, respectively. Therefore, for Duffy negatives, *P. vivax* is predominantly found in mixed infections more than mono infections.

Since the level of *P. vivax* exposure remained consistent among both Duffy positive and Duffy negative individuals across our distinct study locations, the observed diminished burden of *P. vivax* in Duffy negative individuals underscores that while Duffy negativity does not confer absolute resistance to *P. vivax* infection, it does exert a significant inhibitory effect on infection establishment. The mechanism behind *P. vivax* infections of Duffy negatives remains highly elusive, however, several studies have highlighted potential invasion mechanism adaptations of *P. vivax* that may circumvent Duffy-based infection inhibition and allow for infection on a lesser scale[17, 32]. One of the most well studied of these potential adaptations is the *P. vivax* Duffy binding protein 1 (PvDBPI) copy number expansion. Several different studies have clearly shown that PvDBP gene amplification both facilitated binding to alternative lower affinity receptors in Duffy negatives, and also suggested that the binding affinity of DARC with high copies of PvDBP could be much higher than with single-copy PvDBP parasites[17, 33-35], providing a potential selective pressure towards gene duplication and thus increased infectivity. Two additional ligands, *P. vivax* glycosylphosphatidylinositol-anchored micronemal antigen (PvGAMA) and *P. vivax* merozoite surface protein-1 paralog (PvMSP1P), were recently found capable of binding to both Duffy positive and negative red blood cells, suggesting possible involvement in Duffy-independent invasion pathways[36].

Collectively our findings build on previous work documenting *P. vivax* infections in Duffy negative individuals in numerous African countries[37, 38] including Cameroon[39],

Madagascar[10], Angola and Equatorial Guinea[40], Kenya[41], Ethiopia[4]. These studies are consistent with our findings and support the conclusion that Duffy negative individuals are not completely resistant to infection by *P. vivax*, yet still have a greatly reduced prevalence of *P. vivax* infections compared to Duffy positive individuals. Our data show that regardless of exhibiting no significant variation in parasitemia between Duffy positives and Duffy negatives, several *P. vivax* infections from Duffy negatives exhibited relatively high levels of parasitemia potentially implying that these parasites readily infect and adapt to Duffy negativity, allowing for greater erythrocyte invasion. Despite this, several studies have ample evidence that parasitemia of *P. vivax* is greatly reduced in Duffy negatives, supporting the hypothesis that parasite infectivity to the human erythrocyte, though not completely inhibited, is indeed reduced in the absence of the Duffy antigen[42, 43]. Several prior studies have also observed that *P. vivax* infections within Duffy negative individuals are frequently mixed infections, yet these data are limited in that they are predominantly descriptive and do not explore these mixed infections in detail nor compare their prevalence between Duffy negatives and positives[10, 44, 45]. Thus this current study stands out in its efforts to systematically evaluate the prevalence of mixed infections in individuals with Duffy negative status as compared to those with Duffy positive status. Our findings thus shed light on the noteworthy phenomenon that *P. vivax* infections in Duffy negatives frequently encompass mixed-species infections, especially when compared to Duffy positives.

It warrants mention that in the present study is limited in that Duffy expression (negative vs. positive) was inferred based on genotype data of the T33C point mutation in the promoter region of the GATA-1 transcription factor binding site of the Duffy antigen receptor for chemokines (DARC) gene, which is known to alter erythroid expression and eliminate Duffy antigen expression on the red blood cell surface[30, 31, 46]. However, the direct antigen expression

(phenotype) was not assessed. It is therefore possible for a genotypically categorized Duffy negative individual to potentially express Duffy receptors in some quantity, and the *P. vivax* strains infecting Duffy negatives in this study may be utilizing such an expression in invasion, despite genotypic negativity. Additionally, we did not assess the prevalence or burden of *P. falciparum* across Duffy negatives and positives as Duffy expression is not known to be associated with *P. falciparum* infection.

Understanding the distribution of *P. vivax* in Africa and exploring the significance of Duffy expression continues to be a challenging and intricate endeavor. Given the low parasitemia often associated with *P. vivax* infections of Duffy negative individuals, microscopy and RDTs are often not sensitive enough to detect infection, hindering their diagnosis and study in the field. Indeed, corresponding microscopy data from this area accounted for only approximately 70%, of all qPCR confirmed *P. vivax* positive infections[22], highlighting the need for more sensitive molecular detection tools in the field. This has significant implications for malaria elimination on the continent as a high proportion of *P. vivax* cases are likely being overlooked by traditional diagnostic methods. This work shows that not only does *P. vivax* transmission remain widespread in Ethiopia, but these asymptomatic community infections make up a significant portion of *P. vivax* cases resulting in a large undetected parasite reservoir that may greatly complicate and hinder interventions and elimination efforts. Finally, it is clear through the current study that Duffy negativity is not a definitive barrier to infection, and *P. vivax* infections were detected in Duffy negative individuals, encompassing both asymptomatic and febrile malaria instances, frequently occurring in mixed infections. Importantly, these trends persist across both study sites representing high and low endemic settings. This information is vital to

informing control and elimination strategies in areas of Sub-Saharan Africa with variable *P. vivax* endemicity and of high Duffy heterogeneity.

Declarations

Ethical Approval Ethical and scientific approval and clearance was obtained from the institutional scientific and ethical review boards of the University of California, Irvine, USA and Jimma University, Ethiopia. Written informed consent/assent for participation in the study was obtained from all participants and/or parents/guardians (for minors under the age of 18).

Consent for Publication Not applicable.

Availability of Data and Materials The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests The authors declare that they have no competing interests.

Funding This study is funded by the National Institutes of Health (5 F31 AI161887, U19 AI129326 and D43 TW001505). These funders were not involved in study design or development, data collection and analysis, publication process or manuscript preparation.

Authors' contributions LB, and GY conceived this paper; LB drafted it with contributions from MCL. LB, EZ, TD and DY provided field data; LB, EHS and BJ conducted laboratory analyses. Findings were reviewed and interpreted by LB and GY with contributions from EL, EHS and DY; EL, CK, JK, GY edited and revised the manuscript.

Acknowledgements The authors sincerely thank the field team at Jimma University in Ethiopia and volunteers for their contribution. We thank the two anonymous reviewers for their valuable comments.

References

1. Girum T, Shumbej T, Shewangizaw M: **Burden of malaria in Ethiopia, 2000-2016: findings from the Global Health Estimates 2016.** *Trop Dis Travel Med Vaccines* 2019, **5:**11.
2. **World Malaria Report.** *World Health Organization* 2022, **Geneva**
3. Dawit G Ayele TTZHGM: **Prevalence and risk factors of malaria in Ethiopia.** *Malaria Journal* 2012, **11**.
4. Ketema T, Bacha K, Getahun K, Portillo HAD, Bassat Q: **Plasmodium vivax epidemiology in Ethiopia 2000-2020: A systematic review and meta-analysis.** *PLoS Negl Trop Dis* 2021, **15:**e0009781.
5. Twohig KA, Pfeffer DA, Baird JK, Price RN, Zimmerman PA, Hay SI, Gething PW, Battle KE, Howes RE: **Growing evidence of Plasmodium vivax across malaria-endemic Africa.** *PLoS Negl Trop Dis* 2019, **13:**e0007140.
6. Langhi DM, Jr., Bordin JO: **Duffy blood group and malaria.** *Hematology* 2006, **11:**389-398.
7. Hoher G, Fiegenbaum M, Almeida S: **Molecular basis of the Duffy blood group system.** *Blood Transfus* 2018, **16:**93-100.
8. Louis H. Miller SJM, David F. Clyde, Mary H. McGinnis: **The resistance factor to Plasmodium vivax in Blacks: The Duffy Blood Group Genotype, FyFy.** *The New England Journal of Medicine* 1976, **295**.
9. Howes RE, Patil AP, Piel FB, Nyangiri OA, Kabaria CW, Gething PW, Zimmerman PA, Barnadas C, Beall CM, Gebremedhin A, et al: **The global distribution of the Duffy blood group.** *Nat Commun* 2011, **2:**266.

10. Menard D, Barnadas C, Bouchier C, Henry-Halldin C, Gray LR, Ratsimbasoa A, Thonier V, Carod JF, Domarle O, Colin Y, et al: **Plasmodium vivax clinical malaria is commonly observed in Duffy-negative Malagasy people.** *Proc Natl Acad Sci U S A* 2010, **107:**5967-5971.

11. Wurtz NL, K. M.; Bogreau, H.; Pradines, B.; Rogier, C.; Mohamed Salem Boukhary, A. O.; Hafid, J. E.; Ould Ahmedou Salem, M. S.; Trape, J; Basco, L. K.; Briolant, S.: **Vivax malaria in Maritania includes infection of a Duffy-negative individual.** *Malaria Journal* 2011, **10**.

12. Woldearegai TG, Kremsner PG, Kun JF, Mordmuller B: **Plasmodium vivax malaria in Duffy-negative individuals from Ethiopia.** *Trans R Soc Trop Med Hyg* 2013, **107:**328-331.

13. Abdelraheem MH, Albsheer MM, Mohamed HS, Amin M, Mahdi Abdel Hamid M: **Transmission of Plasmodium vivax in Duffy-negative individuals in central Sudan.** *Trans R Soc Trop Med Hyg* 2016, **110:**258-260.

14. Russo G, Faggioni G, Paganotti GM, Djeunang Dongho GB, Pomponi A, De Santis R, Tebano G, Mbida M, Sanou Sobze M, Vullo V, et al: **Molecular evidence of Plasmodium vivax infection in Duffy negative symptomatic individuals from Dschang, West Cameroon.** *Malar J* 2017, **16:**74.

15. Habtamu K, Petros B, Yan G: **Plasmodium vivax: the potential obstacles it presents to malaria elimination and eradication.** *Trop Dis Travel Med Vaccines* 2022, **8:**27.

16. Auburn S, Cheng Q, Marfurt J, Price RN: **The changing epidemiology of Plasmodium vivax: Insights from conventional and novel surveillance tools.** *PLoS Med* 2021, **18:**e1003560.

17. Gunalan K, Niangaly A, Thera MA, Doumbo OK, Miller LH: **Plasmodium vivax Infections of Duffy-Negative Erythrocytes: Historically Undetected or a Recent Adaptation?** *Trends Parasitol* 2018, **34:**420-429.

18. Wilairatana P, Masangkay FR, Kotepui KU, De Jesus Milanez G, Kotepui M: **Prevalence and risk of Plasmodium vivax infection among Duffy-negative individuals: a systematic review and meta-analysis.** *Sci Rep* 2022, **12:**3998.

19. Kano FS, de Souza AM, de Menezes Torres L, Costa MA, Souza-Silva FA, Sanchez BAM, Fontes CJF, Soares IS, de Brito CFA, Carvalho LH, Sousa TN: **Susceptibility to Plasmodium vivax malaria associated with DARC (Duffy antigen) polymorphisms is influenced by the time of exposure to malaria.** *Sci Rep* 2018, **8:**13851.

20. Maestre A, Muskus C, Duque V, Agudelo O, Liu P, Takagi A, Ntumngia FB, Adams JH, Sim KL, Hoffman SL, et al: **Acquired antibody responses against Plasmodium vivax infection vary with host genotype for duffy antigen receptor for chemokines (DARC).** *PLoS One* 2010, **5:**e11437.

21. Bradley: **DETERMINATION OF PLASMODIUM VIVAX AND PLASMODIUM FALCIPARUM MALARIA EXPOSURE IN TWO ETHIOPIAN COMMUNITIES AND ITS RELATIONSHIP TO DUFFY EXPRESSION** *American Journal of Tropical Medicine and Hygiene* 2023, **In Press**.

22. Hawaria D, Getachew H, Zhou G, Demissew A, Habitamu K, Raya B, Lee MC, Yewhalaw D, Yan G: **Ten years malaria trend at Arjo-Didessa sugar development site and its vicinity, Southwest Ethiopia: a retrospective study.** *Malar J* 2019, **18:**145.

23. Jiang A-L, Lee M-C, Zhou G, Zhong D, Hawaria D, Kibret S, Yewhalaw D, Sanders BF, Yan G, Hsu K: **Predicting distribution of malaria vector larval habitats in Ethiopia**

by integrating distributed hydrologic modeling with remotely sensed data. *Scientific Reports* 2021, **11**.

24. Demissew A, Hawaria D, Kibret S, Animut A, Tsegaye A, Lee MC, Yan G, Yewhalaw D: **Impact of sugarcane irrigation on malaria vector Anopheles mosquito fauna, abundance and seasonality in Arjo-Didessa, Ethiopia.** *Malar J* 2020, **19:**344.

25. Taffese HSH-S, E.; Koepfli, C.; Tesfaye, G.; Lee, M. C.; Kazura, J.; Yan, G.; Zhou, G.: **Malaria epidemiology and interventions in Ethiopia from 2001 to 2016.** *Infectious Diseases of Poverty* 2018, **7**.

26. Haileselassie W, Parker DM, Taye B, David RE, Zemene E, Lee MC, Zhong D, Zhou G, Alemu T, Tadele G, et al: **Burden of malaria, impact of interventions and climate variability in Western Ethiopia: an area with large irrigation based farming.** *BMC Public Health* 2022, **22:**196.

27. Lo E, Yewhalaw D, Zhong D, Zemene E, Degefa T, Tushune K, Ha M, Lee MC, James AA, Yan G: **Molecular epidemiology of Plasmodium vivax and Plasmodium falciparum malaria among Duffy-positive and Duffy-negative populations in Ethiopia.** *Malar J* 2015, **14:**84.

28. Bereczy SM, A.; Gil, J. P.; Farnert, A: **Short Report: Rapid DNA Extraction from archive blood spots on filter paper for genotyping plasmodium falciparum.** *Am J Trop Med Hyg* 2005, **72:**249-251.

29. Veron V, Simon S, Carme B: **Multiplex real-time PCR detection of P. falciparum, P. vivax and P. malariae in human blood samples.** *Exp Parasitol* 2009, **121:**346-351.

30. King CL, Adams JH, Xianli J, Grimberg BT, McHenry AM, Greenberg LJ, Siddiqui A, Howes RE, da Silva-Nunes M, Ferreira MU, Zimmerman PA: **Fy(a)/Fy(b) antigen**

polymorphism in human erythrocyte Duffy antigen affects susceptibility to Plasmodium vivax malaria. *Proc Natl Acad Sci U S A* 2011, **108:**20113-20118.

31. Lo E, Hostetler JB, Yewhalaw D, Pearson RD, Hamid MMA, Gunalan K, Kepple D, Ford A, Janies DA, Rayner JC, et al: **Frequent expansion of Plasmodium vivax Duffy Binding Protein in Ethiopia and its epidemiological significance.** *PLoS Negl Trop Dis* 2019, **13:**e0007222.

32. Menard D, Chan ER, Benedet C, Ratsimbasoa A, Kim S, Chim P, Do C, Witkowski B, Durand R, Thellier M, et al: **Whole genome sequencing of field isolates reveals a common duplication of the Duffy binding protein gene in Malagasy Plasmodium vivax strains.** *PLoS Negl Trop Dis* 2013, **7:**e2489.

33. Gunalan K, Lo E, Hostetler JB, Yewhalaw D, Mu J, Neafsey DE, Yan G, Miller LH: **Role of Plasmodium vivax Duffy-binding protein 1 in invasion of Duffy-null Africans.** *Proc Natl Acad Sci U S A* 2016, **113:**6271-6276.

34. Pearson RD, Amato R, Auburn S, Miotto O, Almagro-Garcia J, Amaratunga C, Suon S, Mao S, Noviyanti R, Trimarsanto H, et al: **Genomic analysis of local variation and recent evolution in Plasmodium vivax.** *Nat Genet* 2016, **48:**959-964.

35. Hostetler JB, Lo E, Kanjee U, Amaratunga C, Suon S, Sreng S, Mao S, Yewhalaw D, Mascarenhas A, Kwiatkowski DP, et al: **Independent Origin and Global Distribution of Distinct Plasmodium vivax Duffy Binding Protein Gene Duplications.** *PLoS Negl Trop Dis* 2016, **10:**e0005091.

36. Popovici J, Roesch C, Rougeron V: **The enigmatic mechanisms by which Plasmodium vivax infects Duffy-negative individuals.** *PLoS Pathog* 2020, **16:**e1008258.

37. Zimmerman PA: **Plasmodium vivax Infection in Duffy-Negative People in Africa.** *Am J Trop Med Hyg* 2017, **97:**636-638.

38. Lo E, Russo G, Pestana K, Kepple D, Abagero BR, Dongho GBD, Gunalan K, Miller LH, Hamid MMA, Yewhalaw D, Paganotti GM: **Contrasting epidemiology and genetic variation of Plasmodium vivax infecting Duffy-negative individuals across Africa.** *Int J Infect Dis* 2021, **108:**63-71.

39. Ngassa Mbenda HG, Das A: **Molecular evidence of Plasmodium vivax mono and mixed malaria parasite infections in Duffy-negative native Cameroonians.** *PLoS One* 2014, **9:**e103262.

40. Mendes C, Dias F, Figueiredo J, Mora VG, Cano J, de Sousa B, do Rosario VE, Benito A, Berzosa P, Arez AP: **Duffy negative antigen is no longer a barrier to Plasmodium vivax--molecular evidences from the African West Coast (Angola and Equatorial Guinea).** *PLoS Negl Trop Dis* 2011, **5:**e1192.

41. Ryan: **Evidence for Transmission of Plasmodium vivax amond a Duffy Antigen Negative Population in Western Kenya.** *American Journal of Tropical Medicine and Hygiene* 2006, **74**.

42. Albsheer MMA, Pestana K, Ahmed S, Elfaki M, Gamil E, Ahmed SM, Ibrahim ME, Musa AM, Lo E, Hamid MMA: **Distribution of Duffy Phenotypes among Plasmodium vivax Infections in Sudan.** *Genes (Basel)* 2019, **10**.

43. Abate A, Bouyssou I, Mabilotte S, Doderer-Lang C, Dembele L, Menard D, Golassa L: **Vivax malaria in Duffy-negative patients shows invariably low asexual parasitaemia: implication towards malaria control in Ethiopia.** *Malar J* 2022, **21:**230.

44. Oboh MA, Badiane AS, Ntadom G, Ndiaye YD, Diongue K, Diallo MA, Ndiaye D: **Molecular identification of Plasmodium species responsible for malaria reveals Plasmodium vivax isolates in Duffy negative individuals from southwestern Nigeria.** *Malar J* 2018, **17:**439.

45. Oboh MA, Singh US, Ndiaye D, Badiane AS, Ali NA, Bharti PK, Das A: **Presence of additional Plasmodium vivax malaria in Duffy negative individuals from Southwestern Nigeria.** *Malar J* 2020, **19:**229.

46. Tournamille C, Colin Y, Cartron JP, Van Kim C: **Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy–negative individuals.** *Nature Genetics* 1995, **10:**224-228.

Figure Captions

Figure 1. Figure 1 Map showing location of both study sites; Arjo and Gambella, in western Ethiopia. Includes locations of study clusters, health facilities, and major towns in the regions. The map was created with Esri ArcGIS Pro 3.1 with data sources from our field survey, and elevation data from NASA SRTM v3 (https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003).

Figure 2 Violin plots of the log-transformed malaria parasite gene copy number of samples collected from asymptomatic communities and febrile patients of all ages from health centers in Ethiopia. (A) *Plasmodium vivax,* and (B) *P. falciparum* by qPCR for individuals of all ages. The central box represents the interquartile range with the median shown as the center line in the box. ***, $P < 0.001$ based on Fisher's exact test.

Figure 3 Box plots of the log-transformed *Plasmodium vivax* parasite gene copy number for Duffy negative and Duffy positive individuals of all ages. A) asymptomatic community samples; and B) febrile malaria samples from health centers. Box plots represent the interquartile range with the median expressed as the center line. NS, non-significant based on Fisher's exact test.

CHAPTER 3

MICROHAPLOTYPES: A POWERFUL TOOL FOR MALARIA POPULATION GENETICS

For: ***Trends in Parasitology*** *– Review*

Lauren Bradley[1], Guiyun Yan[1,2], Elizabeth Hemming-Schroeder[3]

[1] Department of Ecology and Evolutionary Biology, School of Biological Sciences, University of California Irvine, Irvine CA 92697, USA

[2]Program in Public Health, College of Health Sciences, University of California at Irvine, Irvine CA 92697, USA

[3]Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523, USA

Abstract As malaria interventions continue to cause decline in global prevalence and certain regions approach pre-elimination status, the changing landscape of malaria epidemiology must be met with powerful and high-resolution genomic approaches to discerning population diversity and transmission dynamics. While microsatellites, single nucleotide polymorphisms, and whole genome sequencing have all been used with great success, each possesses limiting constraints. Highly multiplexed amplicon deep sequencing of microhaplotypes is a powerful, novel genotyping approach capable of surmounting previous constraints. Microhaplotypes are able to; garner high-resolution population data, assess point mutations in genes of interest, detect minor clones, determine multiplicity of infection and is overall less expensive than previous approaches making it an ideal assay for high-throughput genetic analyses. Microhaplotypes should therefore be strongly considered in future genetic epidemiology and population genetic studies in malaria.

Keywords Malaria, population genetics, microhaplotypes, genomics

Introduction

In the past two decades' extensive public health campaigns have resulted in drastic reductions in malaria mortality and morbidity with cases declining by 27% globally from 2000 to 2015, and deaths in Africa down 44% from 2000 to 2019[1]. Advances in targeted interventions, genetic epidemiology and surveillance have all contributed greatly to this global reduction, yet over recent years' progress has notably stalled[2-4]. While disruptions in malaria treatment and interventions due to the COVID-19 pandemic no doubt contributed to a contraction in recent progress[1, 5], other biologic factors have also played important roles in this interference of advancement towards malaria elimination.

As cases of malaria decrease, recent work has shown more unstable transmission patterns, increases in outbreaks and areas of hot-spots, rises in non-falciparum malarias previously non-endemic areas, and expansion in anti-malaria drug resistance[6-8]. Recent successes in intervention and control have also correlated with the discovery of more expansive asymptomatic reservoirs than were previously thought to exist[9, 10], and transmission stemming from these subclinical infections is a key obstacle to malaria control. Asymptomatic reservoirs may be both undetected and long-lasting; predictive of future periods of patent infection leading to local hot-spots and a key factor in determining routes of connectivity[11]. Importantly, in all areas but especially those approaching pre-elimination settings, detection of minor clones and resolving haplotypes in polyclonal infections are both vital to understanding fine-scale transmission dynamics[12]. This is especially true for *Plasmodium vivax* as it remains far more widespread and relatively understudied. Additionally relapses from latent stage hypnozoites enhance transmission resulting in higher prevalence of polyclonal infections, thus *Plasmodium*

vivax often retains higher levels of genetic diversity even at low transmission settings compared to *P. falciparum* malaria[13, 14].

These ongoing complexities surrounding malaria elimination must be met with accurate, timely, and detailed information on parasite diversity, transmission dynamics, and mechanisms of adaptation to interventional pressures including genetic markers of anti-malarial resistance. The importance of population genetics and genomics for surveillance in the context of control and elimination therefore cannot be overstated[3, 15, 16]. Genetic epidemiology must empower policymakers and health authorities to make informed decisions regarding targeted interventions, resource allocation, treatment policy and surveillance strategies. An ideal genotyping method is capable of: 1) discriminating between infections in endemic areas where polyclonality is high 2) maintain quality coverage in low-parasitemia or asymptomatic infections 3) offer a high power of relationship inference for population genetic analyses 4) garner sequence data on specific genes or points of interest 5) be standardized and thus able to be directly compared across multiple studies; yet common methodologies fail to meet all of these requirements.

Highly multiplexed amplicon deep sequencing for microhaplotypes is a relatively novel sequencing methodology that garners detailed, in-depth genomic data and has immense potential as a tool for population genetics in malaria surveillance[17]. In this paper we overview current common molecular approaches used to acquire genomic data including microsatellites, SNP barcodes and whole genome sequencing (WGS), and discuss the limiting factors of each. Limitations in common genotyping methods turn highlights the need for a novel approach to population genetics that is high-throughput, high-resolution, cost-effective and practical for mitigating complexities in the current landscape of malaria genetics.

Traditional approaches to malaria population genetics

Microsatellites

Microsatellites are highly polymorphic short tandem repeats of DNA, usually consisting of 1-10 base pair long nucleotide repeats[18]. These length-type polymorphisms are common and widely distributed throughout the genome; with the AT richness of the *Plasmodium falciparum* genome there are approximately 132 thousand potential repeated motifs that can be used for microsatellite analyses in the 3D7 reference genome[19]. Given this abundance and their relative ease of use, microsatellites have been used in hundreds of studies defining transmission patterns of *P. falciparum* and *P. vivax* for over two decades; predating other genetic approaches to population genetic analyses. A consensus panel of 12 microsatellite loci was first developed for *P. falciparum* in 1999[20], and has been heavily referenced in the majority of population genetics studies on the species in many studies over the past twenty years[19, 21-23]. While less accord has been reached on non-falciparum malaria microsatellite panels, two panels established in 2007 and 2008[24, 25] have been commonly applied to several studies with great success. In Thailand 10 *Plasmodium vivax* markers were used to genotype infections along the country borders which harbor 95% of the remaining malaria, the center of the country being now nearly malaria-free. They found that despite this reduction in prevalence and low transmission, genetic diversity throughout the country remained high[26]. In Papua New Guinea the same microsatellite panel was used along with 10 *P. falciparum* markers pre- and post-intervention and showcased that, for *Plasmodium vivax* in particular, genetically diverse and intermixing parasite populations remained common despite substantial prevalence reduction[22].

Despite their extensive use in population genetics the major limitation in microsatellite genotyping is its' limited discrimination of alleles of similar sizes[27]. Assays are often unable to

discriminate allele size differences less than 20bp, and stutter peaks and non-specific amplification from sub-optimal PCR conditions commonly confound fine-scale differentiation[28]. Minority clones common in low-transmission areas also present difficulties for microsatellite genotyping (**Figure 1**). Minor, mixed or contaminated parasite sub-populations can be missed by microsatellite typing when they represent less than 20% of the overall parasite population. Common filters that disregard minor peaks less than one-quarter to one-third the size of major peaks are often used in microsatellite protocols and potentially give incorrect impressions of clonality with mixed parasite populations[29]. Microsatellites are also significantly limited for population genetic analyses as they cannot genotype point mutations in specific genes of interest and thus must often be used in conjunction with other genotyping methods to garner more detailed genetic information.

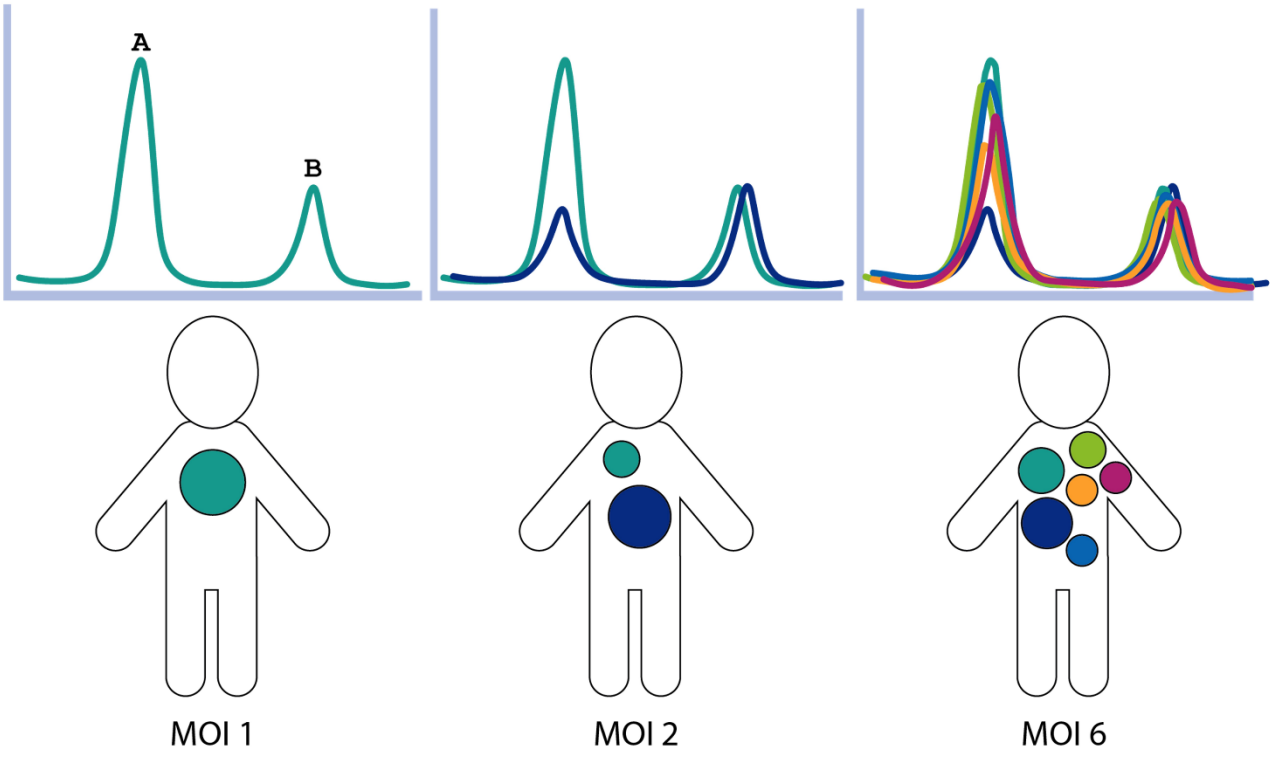


Figure 1 Visual depiction of microsatellite data representing two microsatellite markers of varying length (A and B) and how differentiation between both peaks and infections is complicated as multiplicity of infection (MOI) increases resulting in an inability to detect minor clones.

SNP-based genotyping

Following the sequencing and assembly of the West African 3D7 clone of *P. falciparum*, rapid characterization of genomic variation and profiling of geographically diverse isolates via Sanger sequencing initiated the era of malaria genomics[30]. This template was then repeated and applied to *P. vivax*, and identified ample variants allowing targeted single nucleotide polymorphism (SNP) genotyping of parasites, promoting more extensive analyses of demographic variation and selection[31]. A 24 SNP barcode developed for *P. falciparum*[32] and a 42 SNP barcode was subsequently developed for *P. vivax*[33], these barcodes and others derived from them have been utilized extensively as globally applicable SNP panels to fingerprint infections. They have been successfully applied in a plethora of locals including Nigeria[34], Malawi[35], Kenya[36], Senegal[37, 38], Zambia[39], Papua New Guinea[40], Sri Lanka[41], and China[42] by assessing population diversity, structure and dynamics. Recent work expanded on the initially developed *P. vivax* barcode, and found this updated panel to far outperform microsatellite genotyping in Papua New Guinea (PNG)[40]. Despite their extensive use SNP-barcodes still lack the discriminatory power and resolution to detect minor clones, assess multiplicity of infection and reconstruct haplotypes in areas of high transmission and polyclonality, and identify subtle but important local transmission dynamics in high resolution[27, 43-45]. Rigorous assessment of 24- and 96-SNP molecular barcodes for *P. falciparum* highlighted difficulties in assessment of minor clones and fine-scale population genetic dynamics[46]. Additionally, SNP barcoding often requires large amount of DNA template and often struggles with low grade infections common in asymptomatic reservoirs[27]

Whole Genome Sequencing

Whole genome sequencing undoubtedly provides the most comprehensive view of a pathogen genome and its advent revolutionized molecular epidemiology[47]. In recent years WGS data availability from field collected samples has greatly expanded with continually developing technologies. Until recently studies required vast amounts of venous blood and leukocyte depletion at the time of collection, challenging for most field operations[48]. Yet recent developments in pre-enrichment methods including selective Whole Genome Amplification (sWGA) allow for WGS on easily collected and stored dried blood spots (DBS) expanding the use of WGS for field studies[49, 50]. Regardless of sample enrichment, WGS has been extensively used in the characterization of genetic diversity and population dynamics in multiple *Plasmodium* species[51-55]. In Ethiopia WGS data was used to investigate adaptive molecular changes in *P. vivax* populations and compare them to existing genomic data from Southeast Asia *P. vivax* isolates[54]. From this study they were able to assess that pressure from anti-malarial drugs appeared weaker in Ethiopia than Asia, and found an absence of MDR1 (multi-drug resistant) amplification suggesting that *P. vivax* populations in Ethiopia remain susceptible to mefloquine treatment. In Cape Verde off the West African coast and in pre-elimination phases of malaria control WGS data on *P. falciparum* was used to investigate parasite population structuring during an outbreak and found that not only did parasites carry antimalarial mutations in *pfcrt* and *pfhdr* genes, but samples were highly genetically clustered and genetically similar, suggesting the outbreak resulted from a clonal expansion of local parasites[55].

Despite the high-resolution genetic data garnered by WGS, generally several issues present themselves that warrant consideration. Firstly, the low ratio of parasite to human DNA

present in even high-grade infections can dramatically hinder successful sequencing. As human DNA gets amplified and sequenced along with parasite DNA, human genetic material can contaminate reactions and hinder sequencing efficacy, especially in low parasitemia or asymptomatic samples[27, 53, 56, 57]. In some instances this has been addressed by the addition of a pre-enrichment step such as sWGA of parasite DNA, yet this presents its own challenges as sWGA may preferentially amplify the majority clone in multi-clonal samples thus exaggerating the estimated proportion of dominant clones[49]. While WGS may be able to provide the most high resolution data, it remains cost-prohibitive for high throughput studies, especially those where assessment of MOI is the primary objective[27, 58]. For certain cases with polyclonal infections WGS may be able to discern parasite strains, yet as MOI increases and minority clones under 10% in frequency are more common, coverage of each individual clone is reduced and overall sequence resolution is lost[27]. These limitations are of particular importance in *P. vivax* studies as vivax parasites preferentially invade young reticulocytes leading to generally low-density infections, early gametocyte emergence enhances transmission, relapse from hypnozoites complicates clonal differentiation, and the difficulties in establishing a *P. vivax* reference genome limit data interpretation[59]. Finally, WGS approaches garner massive amounts of data, requiring complex bioinformatics pipelines and expertise which may be prohibitive to certain labs or locations[60, 61].

Highly multiplexed amplicon deep sequencing of Microhaplotypes

Thus far we have discussed three of the most commonly used genotyping methods for malaria genomics, how each of them have been used with great success, and where they individually fall short. Due to the inherent limitations of these approaches, many in depth and large-scale studies often utilize multiple molecular approaches in order to garner a comprehensive understanding of malaria dynamics and genomics[58, 62-64]. This allows complete informational coverage where alternately using only one of the discussed approaches may leave gaps in the knowledge and result in an incomplete epidemiological view. While these methodologies all capture valuable genomic information each subject to technical and biological constraints; particularly in polyclonal infections which represent the vast majority of cases in Sub-Saharan Africa. Microsatellite sequencing falls short in its ability to detect minor clones, which is of vital importance in low-grade polyclonal infections, and its limitations in calling and reporting on specific alleles and point mutations of interest. SNP barcoding methodologies were developed in part to combat the difficulties in standardization and throughput associated with microsatellites, and are well adept at assessing point mutations and analyzing population structure and diversity in exploration of transmission dynamics. However, SNP's lack the discriminatory power to resolve multilocus haplotypes in complex infections which represent the vast majority of infections in Sub-Saharan Africa, even in areas of low-endemicity. Whole genome sequencing provides the highest resolution of genomic data, yet remains technically challenging and costly, placing limitations on high-throughput studies. Additionally, in low-level infections the abundance of human DNA compared to parasite genetic material presents laboratory complexities and often results in a substantial loss of both read coverage and resolution. An ideal genotyping method would be capable of: 1) discriminating between

infections in endemic areas where polyclonality is high 2) maintain quality coverage in low-parasitemia or asymptomatic infections 3) offer a high power of relationship inference for population genetic analyses and 4) garner sequence data on specific genes or points of interest. Recent developments of a novel assay may present a solution to the historic problems in malaria genomics, and may meet all the requirements of one such ideal assay (**Table 1**). Highly multiplexed amplicon deep sequencing, also known by microhaplotypes, Amp-seq, and other colloquial names is a novel genotyping assay that we believe will revolutionize malaria genomics; offering a high-resolution cost effective solution.

Microhaplotypes are targeted deep sequence data of short, highly variable sections of interest in the genome, generally consisting of several closely linked SNPs (**Figure 2**). This recent development in next generation sequencing (NGS) allows for incredibly detailed genomic characterization by utilizing simple library preparation and massive efficiency of scale to greatly increase efficiency and decrease the cost of genotyping hundreds of genetically variable targets[65]. Additionally, microhaplotypes, unlike traditional SNPs, are multiallelic, with targeted regions generally consisting of 3 or more traditional SNPs; thus yielding significantly higher discriminatory power in complex infections. They have been shown across multiple biologic fields and disciplines to provide drastically increased power for relationship inference[65-69], therefore proving themselves as a potentially powerful tool for malaria genomics and epidemiology. This tool is especially useful in areas of high polyclonality, with minor clones and a need for high-resolution data on fine-scale transmission dynamics where other common methodologies fall short; SNPs in their inability to resolve multilocus haplotypes and granularity of population genetic dynamics, and WGS in coverage limitations in low-grade infections, or infections with high clonality (**Figure 3**).

Table 1 Comparison of traditional genotyping methods and their ideal applications

	Pros	Cons	Ideal Application
Microsatellites	Affordable Ease of data interpretation and analysis	Difficulties in detecting minor clones Does not sample significant or coding portion of genome Cannot genotype point mutations in genes of interest Lack of standardization of selected markers between studies limits data comparison	Population Structure
SNP Barcoding	Cost effective Can genotype point mutations in genes of interest	Unable to resolve multilocus haplotypes in complex infections Does not have sensitivity to detect minor clones Does not sample significant portion of genome	Population Structure Assessment of point mutations
Whole Genome Sequencing	Samples significant portion of the genome providing ideal breadth of coverage Can genotype point mutations in genes of interest	Difficulties in both resolving multilocus haplotypes in complex infections and sensitivity to detect minor clones Contamination of DNA by human genome Expensive Labor intensive data analysis	Population Structure Assessment of point mutations
Microhaplotypes	Resolves multilocus haplotypes in complex infections Detection of minor clones Genotypes points of interest Allows for study comparison through panel standardization Overall less expensive for high-throughput studies once panel is developed	Pipeline and panel development expensive Analysis is complex and requires intensive bioinformatics	Population structure Assessment of point mutations Resolve complex infections Detect minor clones High resolution sequence data on targeted genes of interest

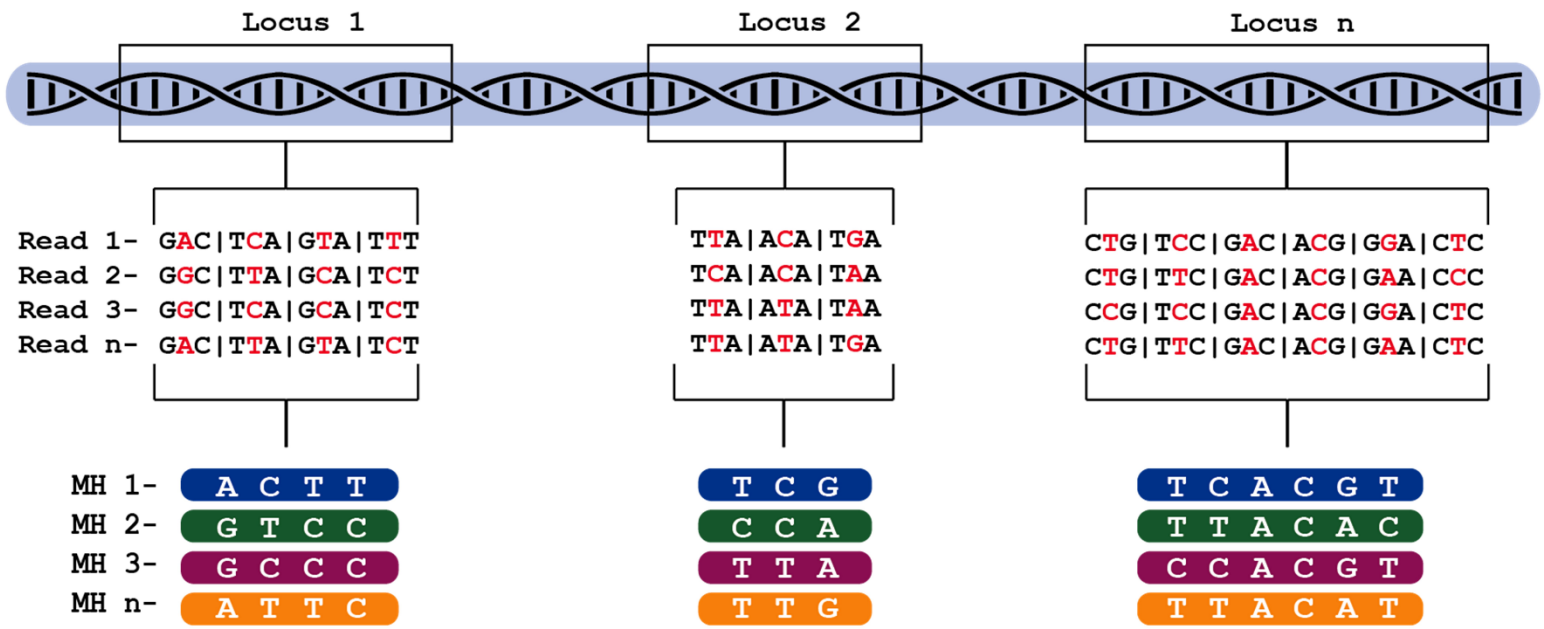


Figure 2 The genomic structure of a microhaplotype panel. A Microhaplotype panel is made up of n loci with each loci being a short highly polymorphic region or a targeted gene of interest with single nucleotide polymorphisms (SNPs). In depth read coverage allows for high resolution genomic data throughout the genome yielding multiple distinguishable microhaplotypes (MH); multiallelic consensus sequences representing different parasite clones within a polyclonal infection.

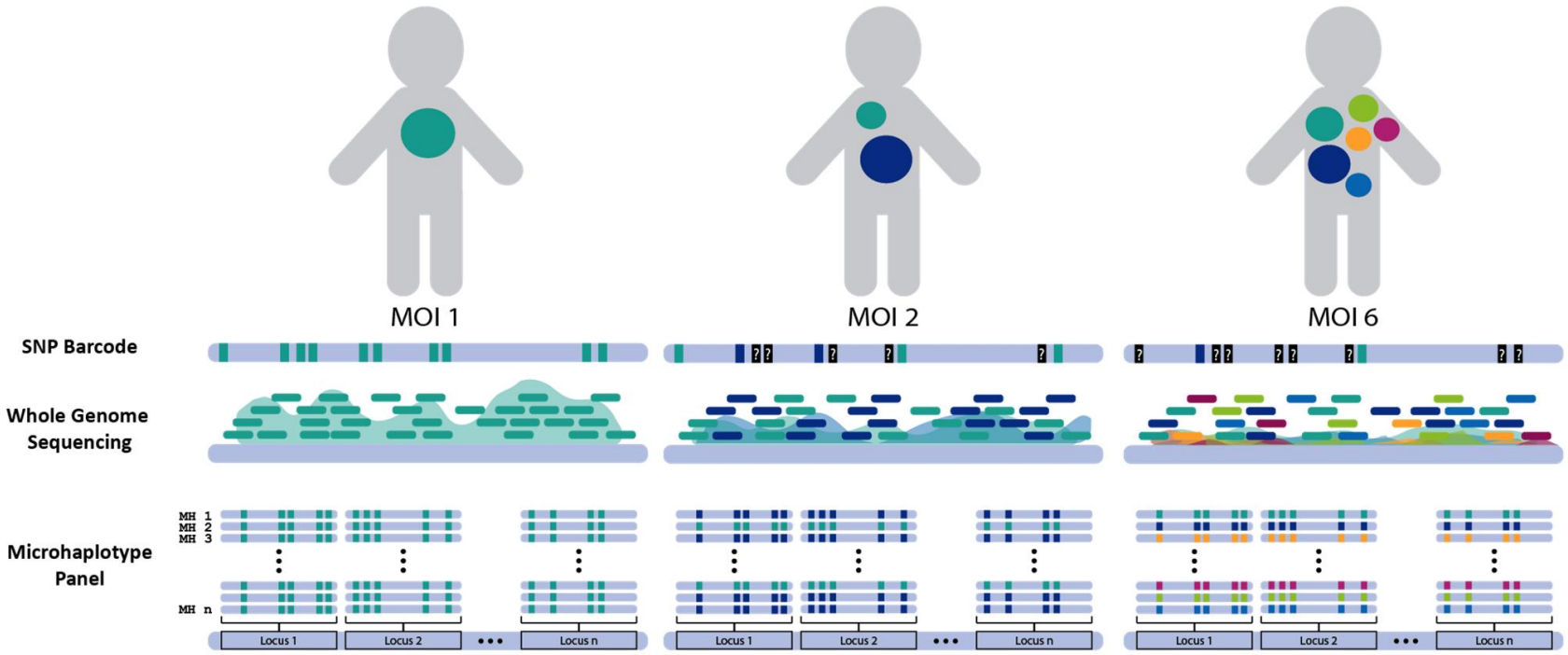


Figure 3 Schematic of different commonplace sequencing approaches used for malaria population genetic studies. Highlights that as multiplicity of infection (MOI) increases, SNP barcoding and Whole Genome Sequencing underperform in their ability to resolve multilocus haplotypes, experience reduced coverage resolution, and lose sensitivity to detect minor clones. Microhaplotypes offer deep sequence data on multiple loci throughout the genome.

While initial applications of this methodology focused on one or a few targets of interest[68], new work has expanded upon this approach and taken influence from SNP-barcoding to develop a microhaplotype panel consisting of numerous genetically diverse loci and genomic areas of interest. Bioinformatic pipelines allow for the bespoke selection of these targets across different species and settings and a workflow highlighting panel development and target selection is highlighted in **Figure 4**. A recent study on *Plasmodium falciparum* developed one such panel; selecting 93 discrete microhaplotype targets with high diversity (median expected heterozygosity = 0.7) and 7 drug resistance loci, out of the potential 4,465 publicly available high-diversity genomic regions[17]. They found that the novel sequencing method achieved very high coverage (median 99%), specificity (99.8%), and sensitivity (90%). Furthermore, the microhaplotype panel far outperformed biallelic SNP-barcodes in its ability to discriminate related from unrelated polyclonal infections, even across a wide range of parasite densities. A malaria case study conducted an evaluation of new *P. falciparum* microhaplotype panels one consisting of solely 4 highly diverse antigens, and a second consisting of 129 drug resistance markers, highly diverse loci, and a locus to detect *P. vivax* co-infection. They found that the two panels performed comparatively, were both capable of evaluating the number of distinct parasite strains within samples (MOI), and showed tremendously strong performance across a wide range of parasitemia levels without the need for DNA pre-amplification[69].

This methodology is therefore clearly capable of; distinguishing individual parasite strains in complex infections, providing insight into targets of interest including potential co-infections and drug resistance markers, and garnering information on population structure at both a fine and broad geographic scale across a wide range of parasite densities including low-grade infections. Additionally, it is an immensely efficient and cost effective assay, with simultaneous

deep-sequencing of numerous and readily customizable loci allowing for high-throughput and bespoke approaches to malaria epidemiology and surveillance. Highly multiplexed microhaplotype panels are therefore a powerful new tool for integrated approaches to malaria genomics, allowing for more routine generation of valuable genetic data that could inform real time data on malaria intervention, control and elimination efforts.

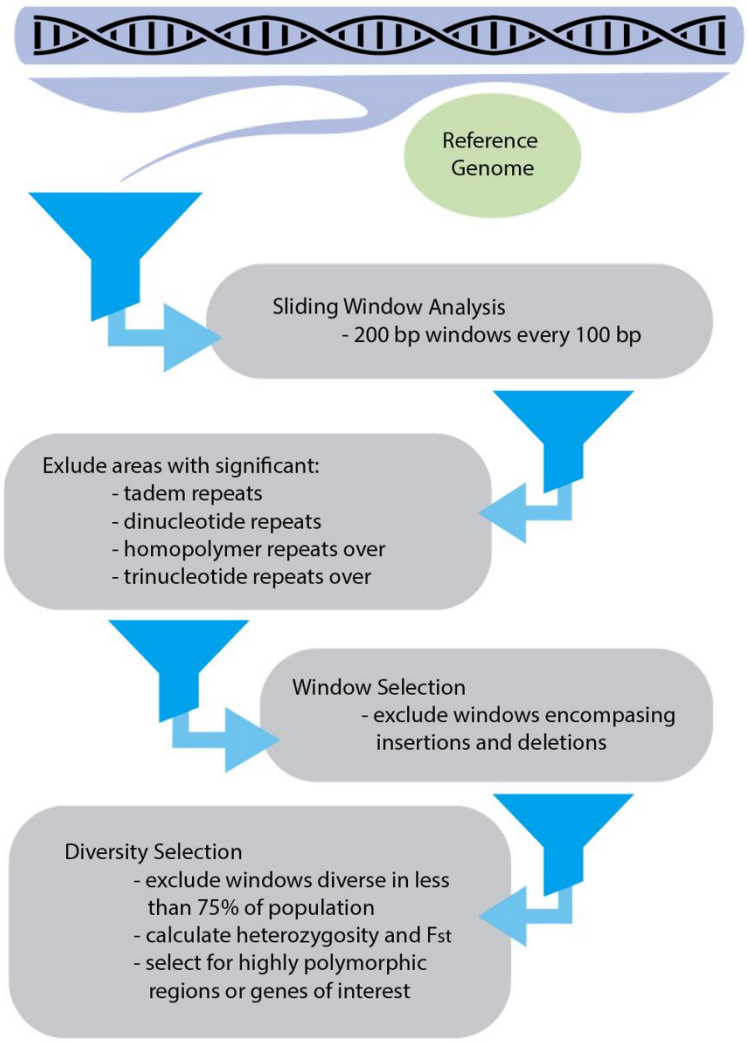


Figure 4 Flowchart depicting bioinformatics workflow for the identification of microhapotype panel targets. Final target selection would ideally include combination of highly polymorphic regions and target genes of interest.

Analytical Limitations

One potential issue with microhaplotype panels as a population genetic and epidemiological tool is that with the generation of massive amounts of data comes a need for vastly complex bioinformatics. Realizing the full potential of data generated via this methodology will therefore undoubtedly require significant improvement in and development of downstream analyzes and software's. Currently few programs exist allowing for the sequence discrimination required to discern different parasite strains and assess MOI, with the central bioinformatic challenge being able to accurately resolve true biologic variation, from that of errors introduced via PCR amplification and sequencing. Recent work has focused on developing software suites to handle just this. Seekdeep is one such open-suite software designed for *de novo* amplicon analysis; specifically with the intention of being able to consistently differentiate true biologic single base pair differences across a wide range of variations and applications whilst yielding improved accuracy and sensitivity to operational taxonomic units (OTUs)[70]. It was found to far outperform previously used software suites including DADA2[71] and MED[72], both of which were considered best-in-class programs also aiming for single-base resolution. When compared to OTU based clustering programs including USEARCH (aka UCLUST/UPARSE)[73], and Swarm[74], neither of which are capable of resolving at the single-base level, Seekdeep was also found superior and showed better accuracy and precision resulting in drastically improved haplotype recovery and far fewer false haplotypes. More recently the development of an identity-by-descent (IDB) framework seeks to accurately calculate genetic distance between polyclonal infections, while explicitly accounting for complexity of infection and population allele frequencies in order to provide a reliable inference. Dcifer's (Distance for complex infections: fast estimation of relatedness) IBD-based framework allows for just that[75]

and has been utilized with great success on *Plasmodium* microhaplotype data. Despite these innovative developments in analytic resources, difficulties remain in the complexities surrounding their use and implementation. As developments in NGS continue, and microhaplotype panels gain more traction in malarial genomics, use of these software suites will likely continue to expand. With this expansion we hope to see a subsequent continued expansion both in analytic options, and resources to assist with ease of use and understanding of the options presented here. With their ability for rapid computation and power of inference these tools will prove invaluable for obtaining epidemiological insight from pathogenic data.

Concluding Remarks

In summation, this review has discussed the need for a novel population genetic and genomic sequencing approach for malaria, as changing landscapes and continued efforts towards elimination highlight the need for a more cost-effective, high-throughput and high-resolution genotyping approach. The advent of highly-multiplexed deep sequencing for microhaplotypes offers a solution that is ideal for numerous applications in malaria genetic epidemiology. Microhaplotype panels are overall less expensive once pipelines and panels have been developed, they allow for genotyping specific points of interest including anti-malarial markers of vital importance with increased intervention efforts. Perhaps most importantly they are capable of detecting minor clones and resolve multilocus haplotypes to accurately discern multiplicity of infection with greater power of inference than previous methods have allowed. For these reasons the use of microhaplotypes in malaria genetic epidemiology should be greatly expended and employed, detailed genetic information on population and transmission dynamics is of vital importance to malaria elimination efforts, and high-resolution data will empower

researchers, policy makers, and health officials to continue intervention and elimination efforts most strategically and effectively.

References

1. WHO: **World Malaria Report 2021.** *Geneva, Switzerland* 2021.
2. Musoke D, Atusingwize E, Namata C, Ndejjo R, Wanyenze RK, Kamya MR: **Integrated malaria prevention in low- and middle-income countries: a systematic review.** *Malar J* 2023, **22:**79.
3. Nsanzabana C: **Strengthening Surveillance Systems for Malaria Elimination by Integrating Molecular and Genomic Data.** *Trop Med Infect Dis* 2019, **4**.
4. Rosenthal PJ: **Malaria in 2022: Challenges and Progress.** *Am J Trop Med Hyg* 2022, **106:**1565-1567.
5. Park J, Kang S, Seok D, Baek YJ, An SY, Lee J, Jun A, Kim SY: **Barriers against and strategies for malaria control during the COVID-19 pandemic in low- and middle-income countries: a systematic review.** *Malar J* 2023, **22:**41.
6. Mogeni P, Omedo I, Nyundo C, Kamau A, Noor A, Bejon P, Hotspot Group A: **Effect of transmission intensity on hotspots and micro-epidemiology of malaria in sub-Saharan Africa.** *BMC Med* 2017, **15:**121.
7. Oboh MA, Oyebola KM, Idowu ET, Badiane AS, Otubanjo OA, Ndiaye D: **Rising report of Plasmodium vivax in sub-Saharan Africa: Implications for malaria elimination agenda.** *Scientific African* 2020, **10**.
8. Conrad MD, Rosenthal PJ: **Antimalarial drug resistance in Africa: the calm before the storm?** *Lancet Infect Dis* 2019, **19:**e338-e351.
9. Bousema T, Okell L, Felger I, Drakeley C: **Asymptomatic malaria infections: detectability, transmissibility and public health relevance.** *Nat Rev Microbiol* 2014, **12:**833-840.

10. Whittaker C, Slater H, Nash R, Bousema T, Drakeley C, Ghani AC, Okell LC: **Global patterns of submicroscopic Plasmodium falciparum malaria infection: insights from a systematic review and meta-analysis of population surveys.** *Lancet Microbe* 2021, **2:**e366-e374.

11. Cheaveau J, Mogollon DC, Mohon MAN, Golassa L, Yewhalaw D, Pillai DR: **Asymptomatic malaria in the clinical and public health context.** *Expert Rev Anti Infect Ther* 2019, **17:**997-1010.

12. Eldh M, Hammar U, Arnot D, Beck HP, Garcia A, Liljander A, Mercereau-Puijalon O, Migot-Nabias F, Mueller I, Ntoumi F, et al: **Multiplicity of Asymptomatic Plasmodium falciparum Infections and Risk of Clinical Malaria: A Systematic Review and Pooled Analysis of Individual Participant Data.** *J Infect Dis* 2020, **221:**775-785.

13. Barry AE: **Complex infections in vivax malaria: the more you look, the more you find.** *Trends Parasitol* 2021, **37:**1022-1023.

14. Arnot A, Alyssa B, John R: **Understanding the population genetics of Plasmodium vivax is essential for Malaria Control and Elimination.** *Malaria Journal* 2012, **11**.

15. Neafsey DE, Taylor AR, MacInnis BL: **Advances and opportunities in malaria population genomics.** *Nat Rev Genet* 2021, **22:**502-517.

16. Garrido-Cardenas JA, Gonzalez-Ceron L, Manzano-Agugliaro F, Mesa-Valle C: **Plasmodium genomics: an approach for learning about and ending human malaria.** *Parasitol Res* 2019, **118:**1-27.

17. Tessema SK, Hathaway NJ, Teyssier NB, Murphy M, Chen A, Aydemir O, Duarte EM, Simone W, Colborn J, Saute F, et al: **Sensitive, Highly Multiplexed Sequencing of**

Microhaplotypes From the Plasmodium falciparum Heterozygome. *J Infect Dis* 2020, **225:**1227-1237.

18. Mathema VB, Nakeesathit S, White NJ, Dondorp AM, Imwong M: **Genome-wide microsatellite characteristics of five human Plasmodium species, focusing on Plasmodium malariae and P. ovale curtisi.** *Parasite* 2020, **27:**34.

19. McDew-White M, Li X, Nkhoma SC, Nair S, Cheeseman I, Anderson TJC: **Mode and Tempo of Microsatellite Length Change in a Malaria Parasite Mutation Accumulation Experiment.** *Genome Biol Evol* 2019, **11:**1971-1985.

20. Anderson TJ, Su XZ, Bockarie M, Lagog M, Day KP: **Twelve microsatellite markers for characterization of Plasmodium falciparum from finger-prick blood samples.** *Parasitology* 1999, **119 (Pt 2):**113-125.

21. Ajogbasile FV, Kayode AT, Oluniyi PE, Akano KO, Uwanibe JN, Adegboyega BB, Philip C, John OG, Eromon PJ, Emechebe G, et al: **Genetic diversity and population structure of Plasmodium falciparum in Nigeria: insights from microsatellite loci analysis.** *Malar J* 2021, **20:**236.

22. Kattenberg JH, Razook Z, Keo R, Koepfli C, Jennison C, Lautu-Gumal D, Fola AA, Ome-Kaius M, Barnadas C, Siba P, et al: **Monitoring Plasmodium falciparum and Plasmodium vivax using microsatellite markers indicates limited changes in population structure after substantial transmission decline in Papua New Guinea.** *Mol Ecol* 2020, **29:**4525-4541.

23. Imwong M, Suwannasin K, Kunasol C, Sutawong K, Mayxay M, Rekol H, Smithuis FM, Hlaing TM, Tun KM, van der Pluijm RW, et al: **The spread of artemisinin-resistant**

Plasmodium falciparum in the Greater Mekong subregion: a molecular epidemiology observational study. *Lancet Infect Dis* 2017, **17:**491-497.

24. Imwong M, Nair S, Pukrittayakamee S, Sudimack D, Williams JT, Mayxay M, Newton PN, Kim JR, Nandy A, Osorio L, et al: **Contrasting genetic structure in Plasmodium vivax populations from Asia and South America.** *Int J Parasitol* 2007, **37:**1013-1022.

25. Karunaweera ND, Ferreira MU, Munasinghe A, Barnwell JW, Collins WE, King CL, Kawamoto F, Hartl DL, Wirth DF: **Extensive microsatellite diversity in the human malaria parasite Plasmodium vivax.** *Gene* 2008, **410:**105-112.

26. Kittichai V, Koepfli C, Nguitragool W, Sattabongkot J, Cui L: **Substantial population structure of Plasmodium vivax in Thailand facilitates identification of the sources of residual transmission.** *PLoS Negl Trop Dis* 2017, **11:**e0005930.

27. Zhong D, Koepfli C, Cui L, Yan G: **Molecular approaches to determine the multiplicity of Plasmodium infections.** *Malar J* 2018, **17:**172.

28. Manrique P, Hoshi M, Fasabi M, Nolasco O, Yori P, Calderon M, Gilman RH, Kosek MN, Vinetz JM, Gamboa D: **Assessment of an automated capillary system for Plasmodium vivax microsatellite genotyping.** *Malar J* 2015, **14:**326.

29. Figan CE, Sa JM, Mu J, Melendez-Muniz VA, Liu CH, Wellems TE: **A set of microsatellite markers to differentiate Plasmodium falciparum progeny of four genetic crosses.** *Malar J* 2018, **17:**60.

30. M.J. G, Bart B, al. e: **Genome sequence of the human malaria parasite Plasmodium falciparum.** *Nature* 2002, **419**.

31. Neafsey DE, Galinsky K, Jiang RH, Young L, Sykes SM, Saif S, Gujja S, Goldberg JM, Young S, Zeng Q, et al: **The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum.** *Nat Genet* 2012, **44:**1046-1050.

32. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, Park DJ, Rosen D, Angelino E, Sabeti PC, Wirth DF, Wiegand RC: **A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking.** *Malar J* 2008, **7:**223.

33. Baniecki ML, Faust AL, Schaffner SF, Park DJ, Galinsky K, Daniels RF, Hamilton E, Ferreira MU, Karunaweera ND, Serre D, et al: **Development of a single nucleotide polymorphism barcode to genotype Plasmodium vivax infections.** *PLoS Negl Trop Dis* 2015, **9:**e0003539.

34. Bankole BE, Kayode AT, Nosamiefan IO, Eromon P, Baniecki ML, Daniels RF, Hamilton EJ, Durfee K, MacInnis B, Okafor H, et al: **Characterization of Plasmodium falciparum structure in Nigeria with malaria SNPs barcode.** *Malar J* 2018, **17:**472.

35. Fulakeza J, McNitt S, Vareta J, Saidi A, Mvula G, Taylor T, Mathanga DP, Small DS, Skarbinski J, Gutman JR, Seydel K: **Comparison of msp genotyping and a 24 SNP molecular assay for differentiating Plasmodium falciparum recrudescence from reinfection.** *Malar J* 2019, **18:**84.

36. Mitchell RM, Zhou Z, Sheth M, Sergent S, Frace M, Nayak V, Hu B, Gimnig J, Ter Kuile F, Lindblade K, et al: **Development of a new barcode-based, multiplex-PCR, next-generation-sequencing assay and data processing and analytical pipeline for multiplicity of infection detection of Plasmodium falciparum.** *Malar J* 2021, **20:**92.

37. Sy M, Badiane AS, Deme AB, Gaye A, Ndiaye T, Fall FB, Siddle KJ, Dieye B, Ndiaye YD, Diallo MA, et al: **Genomic investigation of atypical malaria cases in Kanel, northern Senegal.** *Malar J* 2021, **20:**103.

38. Sy M, Deme AB, Warren JL, Early A, Schaffner S, Daniels RF, Dieye B, Ndiaye IM, Diedhiou Y, Mbaye AM, et al: **Plasmodium falciparum genomic surveillance reveals spatial and temporal trends, association of genetic and physical distance, and household clustering.** *Sci Rep* 2022, **12:**938.

39. Searle KM, Katowa B, Musonda M, Pringle JC, Hamapumbu H, Matoba J, Lubinda M, Shields T, Kobayashi T, Stevenson JC, et al: **Sustained Malaria Transmission despite Reactive Screen-and-Treat in a Low-Transmission Area of Southern Zambia.** *Am J Trop Med Hyg* 2020, **104:**671-679.

40. Fola AA, Kattenberg E, Razook Z, Lautu-Gumal D, Lee S, Mehra S, Bahlo M, Kazura J, Robinson LJ, Laman M, et al: **SNP barcodes provide higher resolution than microsatellite markers to measure Plasmodium vivax population genetics.** *Malar J* 2020, **19:**375.

41. Dewasurendra RL, Baniecki ML, Schaffner S, Siriwardena Y, Moon J, Doshi R, Gunawardena S, Daniels RF, Neafsey D, Volkman S, et al: **Use of a Plasmodium vivax genetic barcode for genomic surveillance and parasite tracking in Sri Lanka.** *Malar J* 2020, **19:**342.

42. Liu P, Shen L, Wang S, Qin P, Si Y, Pan M, Zeng W, Qin Y, Chen X, Zhang Y, et al: **Increasing proportions of relapsing parasite species among imported malaria in China's Guangxi Province from Western and Central Africa.** *Travel Med Infect Dis* 2021, **43:**102130.

43. Auburn S, Barry AE: **Dissecting malaria biology and epidemiology using population genetics and genomics.** *Int J Parasitol* 2017, **47:**77-85.

44. Apinjoh TO, Ouattara A, Titanji VPK, Djimde A, Amambua-Ngwa A: **Genetic diversity and drug resistance surveillance of Plasmodium falciparum for malaria elimination: is there an ideal tool for resource-limited sub-Saharan Africa?** *Malar J* 2019, **18:**217.

45. Chang HH, Worby CJ, Yeka A, Nankabirwa J, Kamya MR, Staedke SG, Dorsey G, Murphy M, Neafsey DE, Jeffreys AE, et al: **THE REAL McCOIL: A method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites.** *PLoS Comput Biol* 2017, **13:**e1005348.

46. Argyropoulos DC, Tan MH, Adobor C, Mensah B, Labbe F, Tiedje KE, Koram KA, Ghansah A, Day KP: **Performance of SNP barcodes to determine genetic diversity and population structure of Plasmodium falciparum in Africa.** *Front Genet* 2023, **14:**1071896.

47. Tang P, Croxen MA, Hasan MR, Hsiao WWL, Hoang LM: **Infection control in the new age of genomic epidemiology.** *Am J Infect Control* 2017, **45:**170-179.

48. Auburn S, Campino S, Clark TG, Djimde AA, Zongo I, Pinches R, Manske M, Mangano V, Alcock D, Anastasi E, et al: **An effective method to purify Plasmodium falciparum DNA directly from clinical blood samples for whole genome high-throughput sequencing.** *PLoS One* 2011, **6:**e22213.

49. Cowell AN, Loy DE, Sundararaman SA, Valdivia H, Fisch K, Lescano AG, Baldeviano GC, Durand S, Gerbasi V, Sutherland CJ, et al: **Selective Whole-Genome Amplification Is a Robust Method That Enables Scalable Whole-Genome Sequencing of Plasmodium vivax from Unprocessed Clinical Samples.** *mBio* 2017, **8**.

50. Oyola SO, Ariani CV, Hamilton WL, Kekre M, Amenga-Etego LN, Ghansah A, Rutledge GG, Redmond S, Manske M, Jyothi D, et al: **Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification.** *Malar J* 2016, **15:**597.

51. Winter DJ, Pacheco MA, Vallejo AF, Schwartz RS, Arevalo-Herrera M, Herrera S, Cartwright RA, Escalante AA: **Whole Genome Sequencing of Field Isolates Reveals Extensive Genetic Diversity in Plasmodium vivax from Colombia.** *PLoS Negl Trop Dis* 2015, **9:**e0004252.

52. Shen H, CHen J: **Whole-genome sequencing and analysis of plasmodium falciparum isolates from China-Myanmar border area.** *Infectious Diseases of Poverty* 2018, **7**.

53. de Oliveira TC, Corder RM, Early A, Rodrigues PT, Ladeia-Andrade S, Alves JMP, Neafsey DE, Ferreira MU: **Population genomics reveals the expansion of highly inbred Plasmodium vivax lineages in the main malaria hotspot of Brazil.** *PLoS Negl Trop Dis* 2020, **14:**e0008808.

54. Auburn S, Getachew S, Pearson RD, Amato R, Miotto O, Trimarsanto H, Zhu SJ, Rumaseb A, Marfurt J, Noviyanti R, et al: **Genomic Analysis of Plasmodium vivax in Southern Ethiopia Reveals Selective Pressures in Multiple Parasite Mechanisms.** *J Infect Dis* 2019, **220:**1738-1749.

55. Da Veiga Leal S, Ward D, Campino S, Benavente ED, Ibrahim A, Claret T, Isaias V, Monteiro D, Clark TG, Goncalves L, et al: **Drug resistance profile and clonality of Plasmodium falciparum parasites in Cape Verde: the 2017 malaria outbreak.** *Malar J* 2021, **20:**172.

56. Nag S, Kofoed PE, Ursing J, Lemvigh CK, Allesoe RL, Rodrigues A, Svendsen CA, Jensen JD, Alifrangis M, Lund O, Aarestrup FM: **Direct whole-genome sequencing of Plasmodium falciparum specimens from dried erythrocyte spots.** *Malar J* 2018, **17:**91.

57. Teyssier NB, Chen A, Duarte EM, Sit R, Greenhouse B, Tessema SK: **Optimization of whole-genome sequencing of Plasmodium falciparum from low-density dried blood spot samples.** *Malar J* 2021, **20:**116.

58. Kanai M, Sachel M: **Comparative Analysis of Plasmodium falciparum Genotyping via SNP Detection, Microsatellite Profiling, and Whole-Genome Sequencing.** *Antimicrobial Agents and Chemotherapy* 2022, **66**.

59. Brashear AM, Cui L: **Population genomics in neglected malaria parasites.** *Front Microbiol* 2022, **13:**984394.

60. Akoniyon OP, Adewumi TS, Maharaj L, Oyegoke OO, Roux A, Adeleke MA, Maharaj R, Okpeku M: **Whole Genome Sequencing Contributions and Challenges in Disease Reduction Focused on Malaria.** *Biology (Basel)* 2022, **11**.

61. Oakeson KF, Wagner JM, Mendenhall M, Rohrwasser A, Atkinson-Dunn R: **Bioinformatic Analyses of Whole-Genome Sequence Data in a Public Health Laboratory.** *Emerg Infect Dis* 2017, **23:**1441-1445.

62. Knudson A, Gonzalez-Casabianca F, Feged-Rivadeneira A, Pedreros MF, Aponte S, Olaya A, Castillo CF, Mancilla E, Piamba-Dorado A, Sanchez-Pedraza R, et al: **Spatio-temporal dynamics of Plasmodium falciparum transmission within a spatial unit on the Colombian Pacific Coast.** *Sci Rep* 2020, **10:**3756.

63. Rovira-Vallbona E, Rosanas-Urgell A: **High Proportion of Genome-Wide Homology and Increased Pretreatment pvcrt Levels in Plasmodium vivax Late Recurrences: a Chloroquine Therapeutic Efficacy Study.** *Antimicrobial Agents and Chemotherapy* 2021, **6**.

64. Oriero EC, Demba MA, Diop MF, Ishengoma DS, Amenga-Etego LN, Ghansah A, Apinjoh T, Issiaka S, Djimde A, D'Alessandro U, et al: **Plasmodium malariae structure and genetic diversity in sub-Saharan Africa determined from microsatellite variants and linked SNPs in orthologues of antimalarial resistance genes.** *Sci Rep* 2022, **12:**21881.

65. Campbell NR, Harmon SA, Narum SR: **Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing.** *Mol Ecol Resour* 2015, **15:**855-867.

66. Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC: **Microhaplotypes provide increased power from short-read DNA sequences for relationship inference.** *Mol Ecol Resour* 2018, **18:**296-305.

67. Pakstis AJ, Gandotra N, Speed WC, Murtha M, Scharfe C, Kidd KK: **The population genetics characteristics of a 90 locus panel of microhaplotypes.** *Hum Genet* 2021, **140:**1753-1773.

68. Lerch A, Koepfli C, Hofmann NE, Kattenberg JH, Rosanas-Urgell A, Betuela I, Mueller I, Felger I: **Longitudinal tracking and quantification of individual Plasmodium falciparum clones in complex infections.** *Sci Rep* 2019, **9:**3333.

69. LaVerriere E, Schwabl P, Carrasquilla M, Taylor AR, Johnson ZM, Shieh M, Panchal R, Straub TJ, Kuzma R, Watson S, et al: **Design and implementation of multiplexed**

amplicon sequencing panels to serve genomic epidemiology of infectious disease: A malaria case study. *Mol Ecol Resour* 2022, **22:**2285-2303.

70. Hathaway NJ, Parobek CM, Juliano JJ, Bailey JA: **SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing.** *Nucleic Acids Res* 2018, **46:**e21.

71. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP: **DADA2: High-resolution sample inference from Illumina amplicon data.** *Nat Methods* 2016, **13:**581-583.

72. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML: **Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences.** *ISME J* 2015, **9:**968-979.

73. Edgar RC: **UPARSE: highly accurate OTU sequences from microbial amplicon reads.** *Nat Methods* 2013, **10:**996-998.

74. Mahe F, Rognes T, Quince C, de Vargas C, Dunthorn M: **Swarm: robust and fast clustering method for amplicon-based studies.** *PeerJ* 2014, **2:**e593.

75. Gerlovina I, Gerlovin B, Rodriguez-Barraquer I, Greenhouse B: **Dcifer: an IBD-based method to calculate genetic distance between polyclonal infections.** *Genetics* 2022, **222**.