UNDERSTANDING MECHANISMS OF SUSTAINED MALARIA TRANSMISSION IN ZAMBIA THROUGH *PLASMODIUM FALCIPARUM* GENETICS

By
Julia C. Pringle

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
December, 2018

Abstract

Malaria remains an enormous public health burden, particularly in sub-Saharan Africa where it is among the leading causes of childhood mortality. Recently, renewed global commitment to malaria elimination has laid the groundwork for 35 nations to declare elimination targets and implement elimination programs. Zambia, one such country, is located in southern Africa where malaria control efforts have been challenging. Although Zambia has achieved significant progress towards its targeted elimination deadline of 2021, much work remains. Zambia faces a heterogenous transmission landscape, with regions of low prevalence in the south and regions of high transmission in the northwest, along the international border with the Democratic Republic of the Congo (DRC). To eliminate malaria, Zambia must address the obstacles to malaria control in regions where interventions have been ineffective as well as the barriers to elimination in regions where unknown mechanisms continue to sustain transmission. We examine the barriers to malaria control and elimination in two epidemiologically distinct regions in Zambia. We focus on the utility of *P. falciparum* genetic methods to draw inferences regarding the mechanisms that continue to sustain malaria transmission in these settings.

In southern Zambia, we use microsatellite genotyping to demonstrate that local malaria transmission contributes to the burden of malaria in spite of control measures including reactive case detection. We further identify a region in our study site as a local transmission hotspot, and suggest that this may be an area to prioritize for additional vector control measures.

Along the border between Nchelenge District, Zambia and Haut-Katanga Province, DRC, we use amplicon deep sequencing at two diverse *P. falciparum* loci, *Pfama1* and *Pfcsp,* to infer that cross-border malaria transmission may contribute to the high burden of malaria in this region. We characterize the extent to which this malaria-parasite population is genetically similar to the strain included in the vaccine, RTS,S/AS01. Our analysis indicates that only 5.2% of parasites in this region match the vaccine strain in the C-terminal *Pfcsp* locus, suggesting that the vaccine may have reduced efficacy in this region.

This dissertation demonstrates the value of incorporating parasite genetic analyses into malaria control and elimination efforts.

Thesis Advisory Committee

Thesis Advisory Committee Thesis Advisor:	Douglas E. Norris, Ph.D., M.S.
Thesis Committee Chair:	William J. Moss, M.D., M.P.H.
Thesis Committee Members:	Monica Mugnier, Ph.D.
	Margaret Taub, Ph.D.
Thesis Committee Alternates:	
	Melissa A. Hewett-Marx, Ph.D.
	Alan L. Scott, Ph.D.
	Frank C. Curriero, Ph.D.

Acknowledgements

I am forever grateful to my family and friends for their love and support. My parents, Cathy and Rick Pringle have always encouraged me and are consistently my number one fans. My sisters, Allison and Claire, have always been my two biggest role models and oldest friends. They and their families, Justin, Rahul, Colby, Maddie, Maya, Juniper, and Isabel have been a constant source of joy, and I am so appreciative to have each of them in my life. I'm especially grateful to Kyle Bucher for transplanting his life from North Carolina to Baltimore so that we could continue to be together these past few years. I'd also like to thank Eve Carlson, Emily Rector, Suzy Robert, Jenine Chami, Anna Rayne, Sarah Andrade, Emily Dutrow, Karen Pereira, Debbie Bozzi, Cathy Bachur, Hilary Szpara, Molly Deutsch-Feldman, and Katie Bradley for their friendship.

I feel incredibly lucky to work with such a wonderful and genuine group of people. I attribute much of the enjoyment I've experienced during my time at Hopkins to my advisers, Doug Norris, Bill Moss, and Melissa Hewett-Marx, for creating such a positive learning and work environment. I owe a lot of my great experience to each of your kindness and mentorship. I am especially appreciative of Doug, who is a phenomenal mentor and an overall great person. Thank you for challenging me to be a better scientist, but also providing support, encouragement, and so much good advice. I'd also like to thank Karen Charron and Yvonne Tam in the Department of International Health for being so supportive of my decision to pursue an MSPH degree. I am so thankful for how warmly I was welcomed into GDEC and how much these individuals did to accommodate my unique situation. Additionally, I have valued the mentorship and

feedback from the members of my thesis advisory committee: Doug Norris, Bill Moss, Monica Mugnier, Margaret Taub, Alan Scott, Frank Curriero, and Melissa Hewett-Marx.

The ICEMR team at Hopkins has been instrumental in making my experience here both productive and fun. Similarly, I can't imagine a better group of Norris lab members, past and present: Smita Das, Maureen Kessler, Giovanna Carpi, Christine Jones, Ilinca Ciubotariu, Jordan Hoffman, Mary Gebhardt, and honorary member Hannah MacLeod. Thank you all for helping make our work environment fun, positive, collaborative, and most importantly, full of chocolate. I'm grateful to all of the collaborators who have welcomed me into the labs and contributed so enormously to my learning experience; I'd like to especially thank the Macha Research Trust (MRT) laboratory in Zambia, the Greenhouse lab, Meshnick and Juliano labs, and the Bailey lab, along with our collaborators Jacob Almagro-Garcia and Joe Zhu. Additionally, I would have accomplished nothing were it not for the incredible administrative support from Gail O'Connor, Genevieve Williams, Trish Ward, Thom Hitzelberger, and Karla McCarthy in the MMI and IH departments at Johns Hopkins. Finally, I have been so fortunate to have a supportive cohort: Hannah, Elizabeth, Kristin, Jane, Zach, Brendan, and Phil.

Anne Jedlicka and Amanda Dziedzic on behalf of the Department of Molecular Microbiology and Immunology's Genomics Core Facility provided enormous support and guidance to me, and were especially instrumental in contributing to the amplicon deep sequencing library preparations that generated data for Chapters 3 and 4 of this dissertation. Thank you both for your generosity, mentorship, and countless contributions to these projects.

This research has been supported by the Molecular and Cellular Basis of Infectious Diseases Training Grant to the Department of Molecular Microbiology and Immunology (5T32AI007417-22) the Global Health Established Field Placements travel award from the Johns Hopkins Center for Global Health, the Johns Hopkins Malaria Research Institute, and from the NIH-funded Southern and Central Africa International Centers of Excellence in Malaria Research (2U19AI089680).

I would like to thank the Zambian study teams at the Tropical Diseases Research Centre, Ndola and at Macha Research Trust, Choma. I am grateful to the communities of Nchelenge and Choma Districts for their willingness to participate in this research.

Table of Contents

Section **Page**

List of Tables

Table **Page**

List of Figures

Figure **Page**

Chapter 1

Introduction

Human malaria is a disease caused by Apicomplexan parasites in the genus *Plasmodium*, and spread by female *Anopheles* mosquitoes. In 2016, there were an estimated 216 million cases of malaria worldwide, resulting in 445,000 deaths [1]. The majority of these cases and deaths occur among children living in sub-Saharan Africa who have not yet acquired immunity to clinical disease [2]. Of the *Plasmodium* species known to cause human disease—*P. falciparum, P. vivax, P. ovale, P. knowlesi, P. simian, and P. malariae—P. falciparum* is associated with the highest mortality, partially due to its unique ability to sequester in the blood vessels of its human host during infection, which can lead to cerebral complications including coma and death [3]. Given its clinical and epidemiological relevance, particularly in sub-Saharan Africa, *P. falciparum* malaria will be the primary focus of this dissertation and will often be referred to interchangeably with "malaria"[2,4].

A BRIEF HISTORY OF MALARIA CONTROL EFFORTS: FROM ERADICATION TO ELIMINATION

Malaria has plagued humans throughout our history, and it is thought that no other pathogen has shaped human evolution to such an extent [5,6]. Human genomes are rife with signatures of this historical selective pressure including sickle cell trait, glucose-6-phosphate deficiency, and thalassemia that confer varying degrees of protection against malaria to human hosts [6]. Despite the undeniable toll malaria has taken on human

populations throughout history, it was only relatively recently that we came to understand how the disease is transmitted. In fact, malaria, derived from the Italian "*mal'aria*," literally "bad air," reflects the prevailing theory throughout much of history that noxious air associated with poor hygiene practices and swamps was responsible for causing the disease. It wasn't until 1880 that Charles Louis Alphonse Laveran, a French physician working in Algeria, discovered parasites in the blood of his malarious patients [5]. It was later still, in 1897, when Ronald Ross first demonstrated that non-*falciparum* malaria could be transmitted between mosquitoes and birds [5].

By the 20th century, these key scientific discoveries into malaria pathology and transmission paved the way towards novel anti-malarial technologies capable of effectively disrupting transmission and treating infections. Notably, the discoveries of dichloro-diphenyl-trichloroethane (DDT) as an insecticide in 1939 and chloroquine as a safe and effective anti-malarial drug in 1934 contributed to the mounting enthusiasm for and confidence in the prospect of eventual global malaria eradication.

Interest in malaria eradication culminated in the World Health Organization's (WHO) proposal in 1955 of the Global Malaria Eradication Program (GMEP). The GMEP focused extensively on spraying the walls and eaves of households with DDT, a method referred to as indoor residual spraying (IRS) [7]. The *Anopheles* species implicated most in human malaria transmission tend to feed on humans indoors during the night and subsequently rest on walls and eaves of households to digest their blood meal. IRS targets these indoor, resting mosquitoes that have just fed, thus disrupting transmission by preventing potentially infected mosquitoes from becoming infectious. In addition to IRS, the GMEP relied heavily on treating infected individuals with

chloroquine, despite early reports of chloroquine resistance during the 1950's in Colombia and at the Cambodia-Thailand border [7–9]. Other established vector control practices including mosquito breeding habitat reduction were abandoned under the GMEP in favor of adopting the standardized eradication guidelines focused on DDT and chloroquine for all settings regardless of nuances in regional transmission epidemiology [7]. At the time the GMEP was conceived, it was thought that malaria could be eradicated within seven years, a prospect which is seen today as overly optimistic [7].

By the 1960's, chloroquine resistance had spread in *Plasmodium* populations throughout South America, Southeast Asia, and India, finally reaching Africa in the 1970's [8]. DDT resistance was first reported in Greece in 1951 and continued to spread among *Anopheles* vector populations thereafter [5,7,10]. As both DDT and chloroquine became less effective, the observed initial declines in malaria were quickly replaced with resurgence leading to a subsequent rise in the global malaria burden [7,11]. In fact, it is estimated that the risk of malaria death among children ages 0-9 years in Africa increased between 2.1 and 5.5-fold following the arrival of chloroquine resistance to the continent [12].

The GMEP was eventually abandoned in 1969 when it became clear that the goal of eradication was untenable [5,7,11,13]. Despite its failures, the GMEP achieved several important successes. Notably, it is estimated that 79 of the 178 countries endemic for malaria at the start of the 20th century achieved elimination between 1945-2010, with substantial gains driven by the GMEP agenda [13]. Malaria elimination successes during the GMEP were mainly restricted to geographically temperate regions, where the climactic suitability for transmission is less stable than in the tropics [13].

By the 1990's, the burden of malaria, particularly in Africa, had reached alarming numbers [14]. In 1998, the World Health Organization (WHO), the United Nations Children's Fund (UNICEF), the United Nations Development Programme (UNDP) and the World Bank responded to the rising number of malaria cases and jointly launched the Roll Back Malaria Partnership (RBM) with the goal of global malaria control. RBM's strategy focused on scaling up malaria control interventions with the goal of universal coverage, improving health systems, and decreasing malaria morbidity and mortality [15]. The lessons learned during the GMEP undoubtedly influenced the RBM goals. Importantly, RBM focused on reducing the malaria burden rather than eradication, and countries in sub-Saharan Africa were this time a major focus.

THE CURRENT STATE OF MALARIA CONTROL: SUCCESSES & CHALLENGES

During the early 2000's, RBM, along with the establishment of the US President's Malaria Initiative (PMI) in 2005 and increased funding towards malaria control through the Global Fund to Fight AIDS, Tuberculosis, and Malaria in 2002, contributed to a decline in global malaria transmission that continues to this day [16,17]. IRS, case management with improved drugs including artemisinin-combination therapies (ACTs), and insecticide treated bednets (ITNs) have been credited with the observed 40% decline in global malaria between 2000-2015 [17]. Currently, over 100 countries have successfully eliminated malaria, and the prospect of elimination is considered feasible by a growing number of additional nations [16]. In fact, 35 countries had established elimination goals and were actively implementing elimination programs as of September,

2015 [16]. Together, these 35 nations substantially reduced their collective malaria burden between 2000 and 2013 by 90% [16].

Although the past two decades have seen declines in malaria transmission worldwide, emerging threats to control and elimination necessitate constant monitoring and preparedness. In 2009, resistance to the frontline antimalarial drug, artemisinin, was first reported in Cambodia and subsequently reported throughout Southeast Asia [18–20]. Similarly, resistance to pyrethroids, the only class of insecticides currently used to impregnate ITNs, has become widespread among malaria vector populations [21]. Even when vectors have acquired pyrethroid resistance, ITNs continue to protect individual users by functioning as a physical barrier to host-seeking mosquitoes [22]. However, pyrethroid resistance reduces the ability of ITNs to decrease the lifespan of adult mosquitoes and thereby decrease community-level transmission [21]. Finally, the standard diagnostic tests for *P. falciparum* infections, or rapid diagnostic tests (RDTs), rely on detection of the parasite antigen, histidine-rich protein 2 (PfHRP2). Recently, a growing number of studies have reported parasites in Africa with *Pfhrp2/3* deleted from the genome, enabling parasites to remain undetected and parasite infections unconfirmed and untreated [23–26]. As the scale-up towards malaria elimination continues in a growing number of nations, it will be important to monitor these emerging threats, as well as support the development of novel drugs, insecticides, and diagnostics.

In addition to contending with emerging threats to malaria control, programs face unique challenges when transitioning from control to elimination. While malaria control may effectively be accomplished through scale-up of known interventions, eliminating malaria necessitates identifying and disrupting lingering mechanisms of transmission,

preventing reestablishment of transmission following imported cases, and reliably diagnosing low-parasitemia infections. While efficacy and effectiveness of malaria control interventions including IRS, ITNs, and ACTs have been extensively researched in the context of control, much remains unknown regarding the effectiveness of elimination strategies. Continuing to evaluate approaches to eliminate malaria in a variety of epidemiologically distinct settings will be valuable.

Furthermore, successful scale-up of vector control reduces anthropophilic vector populations but does not target mosquitoes that forage or rest outdoors. Thus, as control programs continue to decrease transmission, secondary mosquito vectors that are less well described come to play larger roles in sustaining this residual transmission [27,28]. Additional research into regionally specific secondary vectors will be an important component of the transition from malaria control to elimination.

Malaria control efforts beginning in the early 2000's have been highly successful, and the Millennium Development Goal (Goal #6C) to halt and reverse the trend of increasing malaria incidence was achieved [17,29]. However, with 35 nations currently implementing elimination programs, and numerous others scaling-up control efforts, addressing the aforementioned threats to control and elimination is critical.

CONTRIBUTIONS OF PARASITE GENOMICS TO MALARIA CONTROL & ELIMINATION

Since the early 2000's, Next Generation Sequencing (NGS) technologies have become more cost-effective and scalable. In 2002, Gardner et al. sequenced the first full

P. falciparum genome, and the number of full genomes sequenced since has been rapidly increasing [30]. Open access genomic databases, like the MalariaGEN Consortium's Pf3K Community Project [31] and PlasmoDB [32], encourage researchers to archive and share *Plasmodium* genomes to aid in the development of improved genomic marker selection for relevant epidemiological questions and to allow researchers to conduct global population genomic analyses [33].

Population genetic inferences have been used to evaluate the efficacy of malaria control strategies [34–36], monitor changes in transmission intensity [37,38], and identify source populations of on-going transmission [39]. Although parasite genotyping is not necessarily a replacement for monitoring prevalence and incidence data, it may be a complimentary tool offering additional nuanced insight into the routes of parasite spread. As the barriers to generating genetic and genomic data from parasite samples continue to decrease, it is possible to imagine that such assays may be incorporated into control and elimination operations. Thus, understanding the utilities and limitations of parasite genomic data for control and elimination programs is essential.

Documenting Transmission Intensity

Human malaria infections are commonly comprised of multiple genetically distinct parasite clones of the same species, hereby referred to as polyclonal infections [40]. In regions of high malaria transmission it is estimated that up to 80% of malaria infections are polyclonal [40,41]. The number of distinct genotypes detected in an individual infection is referred to as the multiplicity of infection, or MOI [42]. MOI has

been correlated with the development of clinical immunity to malaria [43–46]. In fact, high MOI is associated with protection from clinical disease in high transmission regions and with risk of severe disease in lower transmission settings [47–50].

MOI and malaria parasite population genetic diversity are considered indicators of malaria transmission intensity [33,51]. *Plasmodium* parasites spend the majority of their lifecycle in a haploid form, with a brief diploid stage in the mosquito midgut during which sexual recombination, or meiosis, occurs between male and female gametocytes [52–55]. During recombination, genes on different chromosomes may reassort, thus generating novel parasite clones when meiosis occurs between genetically distinct gametocytes [41,52,54,56]. In regions of high malaria transmission where polyclonal infections are common, the likelihood of a mosquito ingesting genetically distinct male and female gametocytes from a single blood meal is theoretically high, and therefore the rate of recombination and subsequent generation of novel strains is also high in holoendemic regions [51,55,56]. Conversely, in areas of low transmission where clonal malaria infections are more typical, a mosquito will more likely ingest genetically identical male and female gametocytes which produce more of the same clonal offspring during sexual reproduction, thus not contributing to novel genetic diversity [51,55]. Because MOI is a function of the frequency at which human hosts acquire genetically novel infections through infectious bites, the rate at which mosquitoes ingest genetically distinct clones during a blood meal, and the rate of recombination between genetically distinct strains, MOI is thought to be correlated with transmission intensity [51]. Indeed, MOI has been shown to decrease along with declining transmission in the Republic of the Congo [34], Djibouti [57], Tanzania [49], and The Gambia [58]. Thus, population genetic

diversity and MOI have been used to monitor changes in transmission intensity and evaluate whether programs or interventions have altered transmission patterns [34,36–38,49,59–61].

A "molecular barcode" comprised of 24 neutral, genome-wide, unlinked single nucleotide polymorphisms (SNPs) was developed in 2008 with the goal of providing researchers a common set of loci to answer malaria epidemiological questions [62]. In Thies, Senegal, 24-SNP barcoding revealed that an increase in malaria control interventions was followed by a shift towards decreased population genetic diversity and an increased proportion of infections containing clonal parasites [38]. A subsequent study in this setting revealed that following the initial decline in malaria transmission between 2006-2010, transmission rebounded between 2010-2013, an observation supported both by malaria incidence data from the region and genetic signatures of the parasites [37]. These studies highlight the utility of *P. falciparum* genetics to track transmission intensity and evaluate control programs.

Identifying Cross-Border and Imported Malaria

Malaria control efforts are constantly threatened by reintroduction of parasites through travel or cross-border movement of people and vectors [63]. In fact, malaria resurgence has been attributed to imported malaria in Greece [64] and Zanzibar [65], and imported cases were shown to contribute to the burden of *P. vivax* malaria in Venezuela [66]. Cross-border and imported malaria have been extensively researched in several Asian counties [39,67–70] where transmission intensity is low and efforts to eliminate

malaria have spanned multiple years. In sub-Saharan Africa, the focus has predominantly been on scaling up interventions, although the threat of cross-border malaria will inevitably become more pronounced as transmission is further reduced. While regional partnerships like Elimination 8 (E8) have established collaborative initiatives to eliminate malaria regionally, reducing the threat of cross-border transmission, only eight nations are currently included in this group. Notably, the Democratic Republic of the Congo (DRC) is not included in the E8 and potentially threatens the success of the E8 nations it borders in the absence of coordinated cross-border malaria control measures.

For nations in the pre-elimination stages, assessing the degree to which cross-border transmission threatens malaria elimination is valuable, and parasite genotyping studies offer a means to draw such estimates through measures of gene flow and population structure [33]. Microsatellite genotyping of *P. falciparum* infections collected in Yemen and Saudi Arabia, two regions of the Arabian Peninsula where control efforts have been less effective, revealed a contiguous parasite population suggestive of gene flow and cross-border transmission [71]. These findings were used to recommend additional control measures targeting imported malaria be put into place [71]. Similarly, at the China/Myanmar border, genetic approaches were instrumental in determining that cross-border transmission of *P. vivax* between the two nations was frequent [67].

The WHO defines malaria elimination as the abolishment of autochthonous, or locally sourced, malaria cases [72]. Therefore, documenting malaria elimination and maintaining elimination status necessitate distinguishing between autochthonous and imported cases. Genetic methods can be and have been used for this purpose. After maintaining successful malaria control in Panama for three decades, the nation

experienced a malaria epidemic, whereby cases increased 6-fold [73]. Genotyping using a panel of single nucleotide polymorphisms (SNPs) revealed that the infections in Panama were the result of an importation event from Colombia followed by clonal expansion during local transmission in Panama [73]. Similarly, microsatellite genotyping was used to confirm that an outbreak of malaria in Guatemala among United Nations peacekeeping soldiers had originated during their recent travels to the DRC [74]. These studies demonstrate the power of genotyping to track the spread of malaria parasites and identify outbreak sources.

Developing and Evaluating Malaria Vaccine Candidates

Developing a malaria vaccine capable of reducing malaria morbidity and mortality has been a research priority for decades since the observation that bites from infectious, irradiated mosquitoes conferred protection against malaria challenge in an animal model [75]. The only currently licensed malaria vaccine, RTS,S/AS01 (GlaxoSmithKline, 2015) includes a portion of the *P. falciparum* gene, *circumsporozoite protein* (CSP) fused with hepatitis-B surface antigen [76,77]. During Phase III clinical trials across eleven field sites in Africa, RTS,S/AS01 was shown to have a vaccine efficacy of 36.3% against clinical malaria from month zero to study end of among children aged 5–17 months who received 3 primary doses of RTS,S plus a booster at 20 months [78]. Further, previous research has demonstrated that efficacy of various malaria vaccine candidates, including RTS,S/AS01, varies depending upon the genetic composition of a given infection [79–82]. In fact, for parasites that deviated by even one amino acid in the C-terminal region of CSP included in the vaccine construct compared

to the vaccine strain, vaccine efficacy is reduced [80]. This observation highlights the threat that population genetic diversity may pose to malaria vaccine efficacy, and underlies the importance of conducting population genetic studies when designing future vaccine constructs [83]. To this end, researchers have sequenced *P. falciparum* loci contained in candidate vaccines to characterize global parasite population diversity with respect to vaccine targets [82,84] as well as monitored clinical trials for signals of strain-specific vaccine-induced immunity [85]. Thus, the era of NGS may benefit the future development of genetically-informed malaria vaccines.

Identifying Reservoirs of Transmission

As the malaria burden decreases, the transmission landscape becomes fragmented and heterogeneous [86]. Under these circumstances, it is most cost-effective to target control and elimination interventions to higher risk geographic hotspots or populations [86–88]. Thus, the identification of spatial regions or populations of individuals to whom interventions should be targeted is an important consideration in low transmission settings [87]. Current research efforts involve characterizing the transmissibility of the so-called asymptomatic reservoir comprised of individuals who may sustain transmission despite remaining free from symptoms that would cause them to seek treatment [89]. In this context, *P. falciparum* genetics can be useful. Interestingly, Searle et al. used the 24-SNP molecular barcode to compare parasites infecting asymptomatic and symptomatic individuals in Choma District, Zambia and found that the parasite populations infecting these groups of individuals were genetically distinct, suggesting that asymptomatically infected individuals are not contributing substantially to on-going transmission in this

region [90]. Similarly, the 24-SNP barcode was used to identify a region of Madagascar that was a hotpot for malaria transmission [91]. Approaches like these are valuable, as they allow control program leaders to more appropriately divert resources to the reservoirs and geographic foci of malaria transmission.

METHODS AND CHALLENGES IN MALARIA GENOMICS

Parasite genetic and genomic approaches are a valuable addition to malaria control and elimination efforts. However, using *P. falciparum* genetic information to answer epidemiological questions is not without its challenges including low concentrations of template DNA, deconvoluting genotypes in the context of polyclonal infections, and selecting appropriate genetic markers and bioinformatic methods tailored to the specific questions of interest.

Technical Challenges

The 23 Mb *P. falciparum* nuclear genome is substantially smaller than that of humans [30]. *P. falciparum* samples are typically collected from human blood and may contain up to 1,000 times more human than parasite DNA, leading to challenges of specifically sequencing the parasites [33,92,93]. Human leukocyte depletion and filtration have been used to remove human DNA from blood samples prior to preparing samples for sequencing with the goal of increasing the proportion of total DNA from *Plasmodium* [92,94]. However, leukocyte depletion assays require that whole blood be collected from study participants, an option that is not easily done in field settings where finger prick

blood samples can more readily be collected, stored, and transported on filter paper in the form of dried blood spots (DBS). Thus, parasite target enrichment methods from DBS have been used to selectively increase parasite DNA concentrations relative to that of the host. Two commonly used techniques include hybridization assays and polymerase chain reaction (PCR) based methods. In hybridization assays, oligonucleotide probes designed to tile the parasite genome are used to selectively capture the genome of interest prior to sequencing [93,95,96]. Hybrid capture methods have been shown to substantially increase the proportion of sequencing reads that map to the target genome [93], but custom hybridization probes remain costly, reducing the current scalability of the method. PCR-based enrichment methods, like selective whole genome amplification (sWGA) use primers that bind DNA motifs occurring more frequently in the parasite genome than in the human genome along with a highly processive DNA polymerase to selectively amplify the parasite genome [97,98]. While sWGA remains a cost-effective and scalable option for target enrichment, amplification-based methods may introduce bias if certain genetic variants preferentially amplify [Juliano, unpublished], although this was not observed in a recent *P. vivax* WGS analysis conducted from Peruvian isolates [99].

Sequencing low density *Plasmodium* isolates is further complicated due to the parasite's highly AT-rich genome [30]. The base composition of the *P. falciparum* genome is 81% AT on average, with some sub-telomeric and telomeric regions closer to 97% AT [30,92]. AT-rich regions of the genome tend to amplify less well than regions where the AT/GC content is more balanced which can lead to biased representation of the clones present in a sample if the genotyping method relies on a PCR step, as is often the

case for low density samples [100,101]. While DNA library preparations can be optimized to reduce the bias associated with AT-rich templates, [97] it is important to note the possible introduction of amplification bias during PCR. Of course, choosing a DNA library preparation strategy to minimize bias is challenging, as bypassing PCR may reduce detection of minor clones, whereas including a PCR step may introduce artifacts from amplification bias. Ultimately, establishing reliable methods to enrich *Plasmodium* genomes while controlling for amplification bias is desired. In fact, a novel molecular inversion probe (MIP) method for SNPs and microsatellites may provide an ideal path forward. MIP genotyping combines a dual hybrid capture probe and molecular barcoding system with subsequent DNA amplification, offering both the ability to genotype low density parasite DNA and a method to identify and remove amplification bias (Figure 1) [102]. Although not yet applicable for WGS, MIPs have the potential to improve upon current SNP and microsatellite genotyping assays while keeping costs low.

Markers and Methods

P. falciparum genetic and genomic analyses commonly rely on nuclear SNPs and microsatellites [103]. SNPs are biparentally inherited, individual nucleotides in the genome that are variant, most typically bi-allelic [104]. SNP loci can be genotyped through a variety of methods including but not limited to PCR followed by a restriction enzyme digest as in PCR-restriction fragment length polymorphism (PCR-RFLP), WGS, amplicon sequencing, MIPs, and TaqMan fluorescent probe high resolution melt assays (HRM) where wild-type and variant allele probes denature from template DNA at different temperatures [105,106]. PCR-RFLP and HRM assays often require a

preliminary PCR amplification step, but can be easily implemented in most laboratory settings equipped with basic PCR (PCR-RFLP) or light cycler (HRM) technology [106]. While more expensive and requiring specialized equipment, WGS can be done in a way that omits a PCR step, thus minimizing the potential for amplification bias if sufficient template DNA is available [106].

SNP selection involves identifying neutral, unlinked loci that are sufficiently heterozygous within the target population to be informative. However, selecting SNPs on such criteria has been criticized of favoring the inclusion of older mutations that have been diverging for longer, thereby resulting in overestimates of population structure and divergence [103]. Because SNPs in neutral locations of the genome are typically analyzed, SNP markers are appropriate for phylogeographical questions and detection of historical patterns of population differentiation and gene flow [103]. In fact, the 24-SNP *P. falciparum* molecular barcode comprised of globally informative neutral, unlinked SNPs is useful for assigning barcoded parasites to broad geographical regions of origin [105]. Finally, SNP genotyping is easily standardized across laboratories, and thus SNP data can be deposited into public databases, facilitating the direct comparison of genotypes across studies and geographical regions [103].

Microsatellites are biparentally inherited, variable length tandem repeats. Microsatellite assays, like SNP assays, generally require a PCR step to amplify microsatellite-containing regions prior to being analyzed for fragment length using capillary electrophoresis. [103]. Like SNPs, microsatellite loci can also be captured by WGS and amplicon deep sequencing, although repetitive regions of the *P. falciparum* genome are challenging to sequence, and therefore NGS approaches for microsatellite

analysis are less commonly used. Microsatellite genotyping assays have been challenging to standardize both within and across laboratories, and their comparability across studies is therefore limited [103,107]. Unlike bi-allelic SNPs which evolve slowly, microsatellites evolve rapidly and have numerous possible alleles at each locus [33,103,108]. Thus, more SNP than microsatellite loci are typically required to achieve the same power to identify population structure, discern genetic relatedness, and assign population of origin [103]. Although microsatellites have more diversity at each locus, SNPs tend to be more evenly spread across genomes, providing a more representative sample of genomic diversity [103,107].

In amplicon deep sequencing, 35-300bp regions of the *Plasmodium* genome are amplified using PCR and sequenced on an NGS platform, generating millions of sequencing reads per sample at the amplicon of interest. However, despite the depth obtained through amplicon deep sequencing, any SNPs or microsatellites contained within the targeted amplicon will be linked, due to their inherent proximity in the genome, which may preclude certain downstream population genetic analyses that require SNPs be unlinked. Additionally, amplicon sequencing, like SNP and microsatellites assays, requires a PCR step where amplification bias may be introduced. Amplicon deep sequencing has been used to characterize natural *Plasmodium* populations with respect to vaccine candidate genes, monitor drug resistance, and distinguish recurrent malaria episodes due to reinfection with new genetic clones versus recrudescence possibly due to treatment failures [109–111].

WGS is perhaps the least biased method for obtaining genetic information and performing population genetic analysis, particularly if the template DNA is sufficiently

concentrated to bypass target enrichment steps. WGS captures nearly the entire genome, so SNP and microsatellite regions need not be determined *a priori,* and regionally specific markers may be additionally identified. Importantly, WGS does not suffer from ascertainment bias as all loci are genotyped rather than a predetermined set meeting certain minor allele frequency thresholds [103]. However, although WGS costs continue to decline, it is not yet feasible for the majority of researchers to satisfy all of their genomic questions through WGS alone.

As different genetic markers are subject to differing evolutionary pressures and constraints, different markers are suitable to answer different epidemiological questions. Microsatellites, which are known to evolve rapidly [108] can detect recent population structuring events, whereas more slowly evolving SNPs will detect older patterns of gene flow [112]. Further, assays like HRM SNP genotyping and microsatellites may be more easily scalable in malaria endemic countries than methods relying on NGS platforms like amplicon deep sequencing and WGS. Considerations of the template DNA source, epidemiological questions, and available resources should guide the choice of genetic or genomic assay conducted.

The Challenge of Polyclonality

As previously stated, the majority of malaria infections in high transmission regions are comprised of multiple genetically distinct parasite clones of the same species [41]. Polyclonal infections introduce the analytical challenge of phasing SNPs and microsatellites into complete haplotypes. In *P. falciparum* genetics, it has been common

practice to exclusively analyze monoclonal infections [38,113,114] or disregard loci where two or more alleles are present in polyclonal infections [91,115]. Alternatively, it is possible to use genetic distance metrics that account for multiple alleles at each locus [91; Liu and Tessema et al., in preparation]. Although these approaches may be suitable in some contexts, none are capable of characterizing individual clones within an infection as would be critical if the goal was discriminating between parasite reinfection and recrudescence following drug treatment or determining whether a vaccine candidate elicits parasite haplotype-specific immunity.

Like microsatellite and SNP assays, WGS necessitates grappling with the challenge of polyclonality. WGS requires fragmentation of template DNA during library preparation and subsequent mapping of sequenced reads to a reference genome. In the context of polyclonal infections, it is challenging to assign sequenced reads to the appropriate genome. Computational approaches including DEploid [117] have been successful at deconvoluting up to two *P. falciparum* genomes from polyclonal WGS data, but more research and validation is required before infections with greater than two clones can be reliably parsed bioinformatically [Carpi, unpublished].

In contrast to SNP and microsatellite assays that generate unphased genetic data as pools of alleles at the level of infected individuals, amplicon deep sequencing keeps parasite haplotypes intact and permits genetic analysis at the level of individual clones. Unlike WGS, amplicon libraries are not fragmented prior to sequencing, and therefore do not encounter the problem of deconvoluting sequencing reads to multiple reference genomes. Thus, amplicon deep sequencing is the method of choice for high transmission settings where polyclonal infections predominate.

In summary, genotyping and sequencing malaria parasites requires overcoming technical challenges associated with low density template DNA, an AT-rich genome, polyclonal infections, and field samples typically collected as DBS. Selecting genomic markers for study necessitates consideration of the extent of polyclonality in the region of interest, cost, density, volume, and form of template DNA, and the relevant epidemiological questions. Table 1 summarizes the information presented here regarding advantages and disadvantages of the markers and methods discussed.

MALARIA CONTROL IN ZAMBIA

Zambia is a malaria-endemic country in sub-Saharan Africa with an estimated 4.18 million cases and 7,000 deaths from *P. falciparum* malaria in 2016 [118,119]. Malaria is responsible for more childhood deaths in Zambia than any other disease, with 16% of under-five mortality attributable to malaria [120,121]. In Zambia, IRS and ITNs were introduced from 2000-2008 and, in combination with RDTs and ACTs, contributed significantly to a 66% decline in hospitalized cases and deaths during this period [122]. However, these gains in malaria control have been unevenly distributed throughout the country. While some regions of Zambia, most notably Southern Province, are now considered in a pre-elimination phase, transmission still remains high in multiple zones, particularly in districts bordering the DRC in the northern and western regions of the country [123].

Along with several regional neighbors, Zambia is committed to eliminating malaria in every district by 2021, an enormous undertaking given the highly

heterogeneous transmission landscape and the failure of current control methods to eliminate malaria in every district. Zambia now faces the dual challenge of transitioning from control to elimination in low transmission regions while continuing to work towards achieving control in others. In both settings, identifying mechanisms of sustained malaria transmission will enable the development of control and elimination strategies targeted to the transmission drivers in different epidemiological settings.

SOUTHERN AND CENTRAL AFRICA ICEMR

The International Centers of Excellence in Malaria Research (ICEMR) are seven-year research grants awarded by the National Institutes of Health (NIH) to investigate malaria transmission epidemiology, vector biology, and parasite genomics in epidemiologically diverse regions across the globe. The Southern and Central Africa ICEMR maintains four field sites in Southern and Central Africa: two in Zambia, one in Zimbabwe, and one located across the border from the northern Zambia site in Haut-Katanga Province, DRC. These four sites represent different epidemiological settings, and will each require that interventions be specifically tailored to address the distinct mechanisms that continue to sustain transmission in their respective regions. The focus of the research presented here is on malaria transmission in Macha, southern Zambia and along the international border separating Nchelenge Zambia from Kilwa and Kashowbe, DRC.

Macha Research Trust (MRT) is the site of the ICEMR field site in Macha, located in Choma District of Southern Province, Zambia. Southern Province, including Macha,

experiences very low levels of seasonal *P. falciparum* transmission. In fact, RDT prevalence among actively detected individuals in Macha declined from 9.2% in 2008 to less than 1% in 2013, and Southern Province is now considered to be a pre-elimination setting [124]. However, despite success in attaining low levels of malaria transmission, the incidence of pediatric malaria hospitalizations in Macha has plateaued since 2013 [Thuma, unpublished]. It is clear that the current elimination strategy is not sufficient to detect and disrupt the remaining avenues sustaining transmission in this setting. In this context, parasite genetic analyses offer a tool with which to track malaria transmission and draw inferences regarding its mode. In particular, parasite genetic analysis can distinguish locally transmitted from imported parasites and possibly identify local transmission hotspots for intervention prioritization.

Nchelenge District, Zambia is located in northern Zambia along the border with the DRC, and is characterized by intense, year-round transmission. Nchelenge District is located within Luapula Province, which has the highest malaria prevalence in children under five years in Zambia at 56% by RDT despite almost a decade of malaria control interventions including ITNs and IRS [125,126]. Located across the international border from Nchelenge District in Haut-Katanga Province, DRC, Kilwa and Kashobwe experience similar year-round, holoendemic malaria transmission. However, in Kilwa and Kashobwe, vector control initiatives have historically been deployed to a lesser extent than in neighboring Nchelenge. In fact, a cross-sectional household survey conducted in Kilwa and Kashobwe during February 2016 revealed that 31% of the sampled individuals reported sleeping under a bednet the previous night [127]. Following this survey, however, a mass ITN distribution campaign took place in Kilwa and

Kashobwe between September and October of 2016, and a more recent cross-sectional survey in August, 2017 demonstrated that 88% of surveyed individuals reported sleeping under a bednet [127]. However, despite the progress in ITN coverage in Kilwa and Kashobwe, IRS coverage remains low. In fact, only 0.3% of participants reported that their home had received IRS during the cross-sectional survey in February, 2016, and 0.21% in August, 2017 [127]. Furthermore, mitochondrial genomic analysis of the malaria vector, *Anopheles funestus* suggest no population level differentiation between mosquitoes collected in Nchelenge District and Haut-Katanga Province, DRC [128]. Given the historical lack of vector control initiatives in Haut-Katanga, DRC, and the observation of a contiguous mosquito population across this region, it is plausible that cross-border malaria has contributed to sustained transmission in spite of control efforts in Nchelenge District. Parasite population genetic analysis provides a method to test this hypothesis and draw inferences regarding the extent to which cross-border gene flow is on-going.

Finally, as Zambia continues to scale-up malaria control interventions with the goal of decreasing malaria morbidity and mortality, assessing whether the malaria vaccine, RTS,S/AS01, could be effective in this region is merited. Recent data has demonstrated that RTS,S/AS01 vaccine efficacy declines for parasites that differ in amino acid sequence at the C-terminal PfCSP locus from the vaccine strain, 3D7 [80]. Therefore, it is important to understand the proportion of parasites in this region that match 3D7 at the PfCSP locus. Amplicon deep sequencing enables the enumeration of individual parasite haplotypes within polyclonal infections and is therefore a suitable method for assessing the frequency of vaccine strain mismatch. Characterizing natural

parasite population diversity with respect to the RTS,S/AS01 vaccine enables evidence-based policy making should Zambia consider incorporating this vaccine into their current vaccination recommendations.

In summary, interrogating the *P. falciparum* genome can provide insights into epidemiological questions relevant to malaria control and elimination. The specific aims of this dissertation are to:

1. Use microsatellite genotyping to identify and characterize spatial and temporal patterns of local malaria transmission in a pre-elimination setting in Southern Province, Zambia.
2. Assess the extent of parasite gene flow between Nchelenge District, Zambia and Kilwa and Kashobwe, DRC using amplicon deep sequencing at two hypervariable genetic loci.
3. Characterize the *P. falciparum* population genetic diversity in relation to the malaria vaccine, RTS,S/AS01 in a high transmission setting in northern Zambia.

Table 1

	Cost	Ascertainment Bias	PCR Bias	Unlinked Loci	Challenging with Polyclonal Samples	Diversity per Locus	Comparability Across Laboratories
Markers							
SNPs	Low	Yes	Yes	Yes	Yes	Usually 2 alleles	Yes
Microsatellites	Low	Yes	Yes	Yes	Yes	High	No
Methods							
WGS	High	No	No (unless sWGA)	Yes	Yes	NA	Yes
Amplicon sequencing	Medium	Yes	Yes	No	No	User selected	Yes
MIPs	Medium	Yes	Minimized	Yes	Yes	User selected	TBD

Table 1: Summarizes the advantages and disadvantages of commonly used genetic markers and methods.

Figure 1:

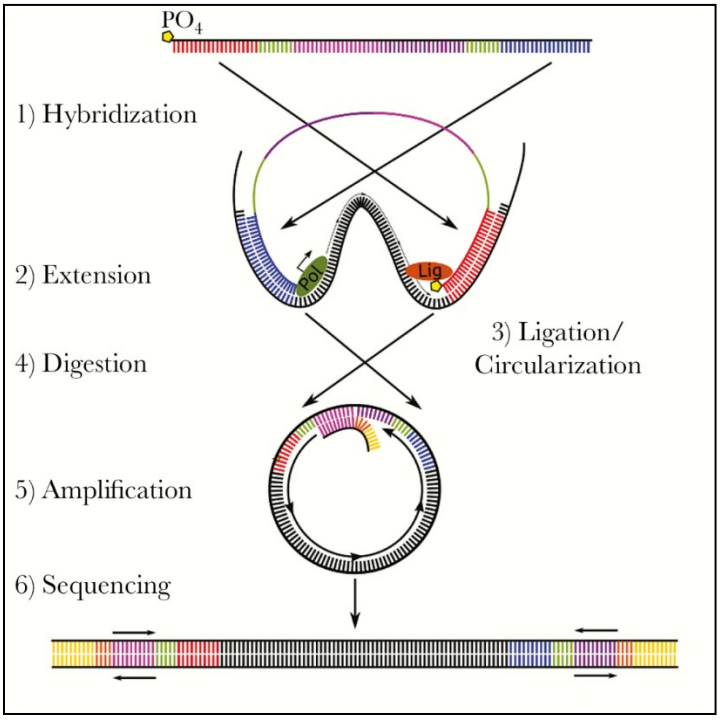

Figure 1: From: Aydmir et al. JID. **2018;** 218(6):946-955. A molecular inversion probe (MIP) is designed to hybridize with sequences flanking a region of interest (blue and red "probe" sequences). MIP molecules also contain unique molecular identification (UMI) sequences. Once captured, a DNA polymerase moves from one probe sequence (blue) across the DNA fragment of interest, extending the double stranded sequence until reaching the other hybridized, double stranded (red) probe sequence. A DNA ligase ligates the double stranded molecule together, forming a circle containing the MIP probes along with the captured DNA fragment of interest and UMI. Circularized molecules can be digested to generate linear double stranded fragments that serve as a PCR template. Because each fragment is uniquely tagged prior to PCR amplification, PCR errors can be detected and removed in downstream analysis.

REFERENCES

1. World Health Organization. World Malaria Report. **2017**.

2. World Health Organization. Malaria. **2015**. Available from: http://www.who.int/mediacentre/factsheets/fs094/en/

3. Bartoloni A, Zammarchi L. Clinical aspects of uncomplicated and severe malaria. Mediterr J Hematol Infect Dis. **2012**; 4(1).

4. Bruce MC, Macheso A, McConnachie A, Molyneux ME. Comparative population structure of Plasmodium malariae and Plasmodium falciparum under different transmission settings in Malawi. Malar J. **2011**; 10(1):38.

5. Centers for Disease Control and Prevention. The History of Malaria, an Ancient Disease. **2017**. Available from: https://www.cdc.gov/malaria/about/history/index.html

6. Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. Am J Hum Genet. **2005**; 77(2):171–92.

7. Ná Jera JA, Gonzá Lez-Silva M, Alonso PL. Some Lessons for the Future from the Global Malaria Eradication Programme (1955-1969). PLoS Med. **2011**; 8(1):e1000412.

8. Wellems TE, Plowe CV. Chloroquine-Resistant Malaria. J Infect Dis. **2001**; 184(6):770–776.

9. Payne D. Spread of chloroquine resistance in Plasmodium falciparum. Parasitol Today. **1987**; 3(8):241–246.

10. Livadas GA, Georgopoulos G. Development of resistance to DDT by Anopheles sacharovi in Greece. Bull World Health Organ. **1953**; 8(4):497–511.

11. Cohen JM, Smith DL, Cotter C, et al. Malaria resurgence: a systematic review and assessment of its causes. Malar J. **2012**; 11:122.

12. Trape J-F, Pison G, Preziosi M-P, et al. Impact of chloroquine resistance on malaria mortality. C R Acad Sci III. **1998**; 321(8):689-97.

13. Feachem RGA, Phillips AA, Hwang J, et al. Malaria Elimination 1 Shrinking the malaria map: progress and prospects. Lancet. **2010**; 376:1566–78.

14. Snow RW. Global malaria eradication and the importance of Plasmodium falciparum epidemiology in Africa. BMC Med. **2015**; 13:23.

15. Roll Back Malaria Partnership to End Malaria. Overview. **2018**. Available from: https://endmalaria.org/about-us/overview

16. Newby G, Bennett A, Larson E, et al. The path to eradication: a progress report on the malaria-eliminating countries. Lancet. **2016**; 387(10029):1775-84.

17. Bhatt S, Weiss DJ, Cameron E, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. **2015**; 526(7572):207-211.

18. Miotto O, Almagro-Garcia J, Manske M, et al. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. Nat Genet. **2013**; 45(6):648–55.

19. Dondorp AM, Nosten F, Yi P, et al. Artemisinin Resistance in Plasmodium falciparum Malaria. N Engl J Med. **2009**; 361(5):455–467.

20. Takala-Harrison S, Jacob CG, Arze C, et al. Independent Emergence of Artemisinin Resistance Mutations Among Plasmodium falciparum in Southeast Asia. J Infect Dis. **2015**; 211(5):670–679.

21. Toé KH, Jones CM, N'Fale S, Ismail HM, Dabiré RK, Ranson H. Increased Pyrethroid Resistance in Malaria Vectors and Decreased Bed Net Effectiveness, Burkina Faso. Emerg Infect Dis. **2014**; 20(10).

22. Kleinschmidt I, Bradley J, Knox TB, et al. Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: a WHO-coordinated, prospective, international, observational cohort study. Lancet Infect Dis. **2018**; 18(6):640–649.

23. Koita OA, Doumbo OK, Ouattara A, et al. False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. Am J Trop Med Hyg. **2012**; 86(2):194–8.

24. Parr JB, Verity R, Doctor SM, et al. Pfhrp2-deleted Plasmodium falciparum parasites in the Democratic Republic of Congo: A national cross-sectional survey. J Infect Dis. **2016**; 216(1):36-44.

25. Beshir KB, Sepúlveda N, Bharmal J, et al. Plasmodium falciparum parasites with histidine-rich protein 2 (pfhrp2) and pfhrp3 gene deletions in two endemic regions of Kenya. Sci Rep. **2017**; 7(1):14718.

26. Wurtz N, Fall B, Bui K, et al. Pfhrp2 and pfhrp3 polymorphisms in Plasmodium falciparum isolates from Dakar, Senegal: impact on rapid malaria diagnostic tests. Malar J. **2013**; 12(1):34.

27. Stevenson JC, Simubali L, Mbambara S, et al. Detection of Plasmodium falciparum Infection in Anopheles squamosus (Diptera: Culicidae) in an Area Targeted for Malaria Elimination, Southern Zambia. J Med Entomol. **2016**; 53(6):1482-1487.

28. Stevenson JC, Norris DE. Implicating Cryptic and Novel Anophelines as Malaria Vectors in Africa. Insects. **2016**; 8(1).

29. Cibulskis RE, Alonso P, Aponte J, et al. Malaria: Global progress 2000 - 2015 and future challenges. Infect Dis poverty. **2016**; 5(1):61.

30. Gardner M, Hall N, Fung E, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature. **2002**; 419:498–511.

31. The Pf3K Project (2016): pilot data release 5. **2016**. Available from: www.malariagen.net/data/pf3k-5

32. Aurrecoechea C, Brestelli J, Brunk BP, et al. PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res. **2009**; 37:D539–D543.

33. Neafsey DE, Volkman SK. Malaria Genomics in the Era of Eradication. Cold Spring Harb Perspect Med. **2017**; 7(8):a025544.

34. Ibara-Okabande R, Koukouikila-Koussounda F, Ndounga M, et al. Reduction of multiplicity of infections but no change in msp2 genetic diversity in Plasmodium falciparum isolates from Congolese children after introduction of artemisinin-combination therapy. Malar J. **2012**; 11(1):410.

35. Buchholz U, Kobbe R, Danquah I, et al. Multiplicity of Plasmodium falciparum

infection following intermittent preventive treatment in infants. Malar J. **2010**; 9:244.

36. Ofosu-Okyere a, Mackinnon MJ, Sowa MP, et al. Novel Plasmodium falciparum clones and rising clone multiplicities are associated with the increase in malaria morbidity in Ghanaian children during the transition into the high transmission season. Parasitology. **2001**; 123(Pt 2):113–123.

37. Daniels RF, Schaffner SF, Wenger E a., et al. Modeling malaria genomics reveals transmission decline and rebound in Senegal. Proc Natl Acad Sci. **2015**; 112(22):201505691.

38. Daniels R, Chang HH, Séne PD, et al. Genetic Surveillance Detects Both Clonal and Epidemic Transmission of Malaria following Enhanced Intervention in Senegal. PLoS One. **2013**; 8(4):4–10.

39. Lo E, Zhou G, Oo W, et al. Molecular inference of sources and spreading patterns of Plasmodium falciparum malaria parasites in internally displaced persons settlements in Myanmar–China border area. Infect Genet Evol. **2015**; 33:189–196.

40. Bell AS, Roode JC de, Sim D, Read AF. Within-host competition in genetically diverse malaria infections: parasite virulence and competitive success. Evolution. **2006**; 60(7):1358–1371.

41. Babiker H a, Ranford-Cartwright LC, Walliker D. The epidemiology of multiple Plasmodium falciparum infections - 3. Genetic structure and dynamics of Plasmodium falciparum infections in the Kilombero region of Tanzania. Trans R Soc Trop Med Hyg. **1999**; 93:S11–S14.

42. Mayengue PI, Luty AJF, Rogier C, Baragatti M, Kremsner PG, Ntoumi F. The multiplicity of Plasmodium falciparum infections is associated with acquired immunity to asexual blood stage antigens. Microbes Infect. **2009**; 11(1):108–114.

43. Reyburn H, Mbatia R, Drakeley C, et al. Association of transmission intensity and age with clinical manifestations and case fatality of severe Plasmodium falciparum malaria. JAMA. **2005**; 293(12):1461–1470.

44. Reyburn H, Griffin JT, Drakeley CJ, et al. Gradual acquisition of immunity to severe malaria with increasing exposure. Proc Biol Soc. **2015**; 282(1801):20142657.

45. Snow RW, Omumbo J a., Lowe B, et al. Relation between severe malaria morbidity in children and level of Plasmodium falciparum transmission in Africa. Lancet. **1997**; 349(9066):1650–1654.

46. Marsh K, Snow RW. Host-parasite interaction and morbidity in malaria endemic areas. Philos Trans R Soc Lond B Biol Sci. **1997**; 352(1359):1385–1394.

47. Al-Yaman F, Genton B, Reeder JC, Mokela D, Anders RE, Alpers MP. Humoral response to defined plasmodium falciparum antigens in cerebral and uncomplicated malaria and their relationship to parasite genotype. Am J Trop Med Hyg. **1997**; 56(4):430–435.

48. Engelbrecht, Fred, Felger, Ingrid, Genton, Blaise, Alpers, Michael, Beck H-P. Plasmodium falciparum: Malaria Morbidity is Associated with Specific Merozoite Surface Antigen 2 Genotypes. Exp Parasitol. **1995**; 81:90–96.

49. Beck HP, Felger I, Huber W, et al. Analysis of multiple Plasmodium falciparum infections in Tanzanian children during the phase III trial of the malaria vaccine SPf66. J Infect Dis. **1997**; 175(4):921–926.

50. Shigidi MMT, Hashim R a, Idris MN a, Mukhtar MM. Parasite Diversity in Adult Patients With Cerebral Malaria. Am J Trop Med Hyg. **2004**; 71(6):754–757.

51. Volkman SK, Neafsey DE, Schaffner SF, Park DJ, Wirth DF. Harnessing genomics and genome biology to understand malaria biology. Nat Rev Genet. **2012**; 13(5):315–328.

52. Babiker H a, Walliker D. Current views on the population structure of plasmodium falciparum: Implications for control. Parasitol Today. **1997**; 13(7):262–267.

53. McKenzie, F. Ellis, Ferreira, Marcelo U., Baird, Kevin J., Snounou, Georges, Bossert WH. Meiotic Recombination, Cross-Reactivity, and Persistence in Plasmodium falciparum. Evolution. **2001**; 55(7):1299–1307.

54. Kerr PJ, Ranford-Cartwright LC, Walliker D. Proof of intragenic recombination in Plasmodium falciparum. Mol Biochem Parasitol. **1994**; 66(2):241–248.

55. Mita T, Jombart T. Patterns and dynamics of genetic diversity in Plasmodium falciparum: What past human migrations tell us about malaria. Parasitol Int. **2014**; 64(3):238–243.

56. Talman AM, Domarle O, McKenzie FE, Ariey F, Robert V. Gametocytogenesis: the puberty of Plasmodium falciparum. Malar J. **2004**; 3:24.

57. Khaireh BA, Assefa A, Guessod HH, et al. Population genetics analysis during the

elimination process of Plasmodium falciparum in Djibouti. Malar J. **2013**; 12:201.

58. Haywood M, Conway DJ, Weiss H, et al. Reduction in the mean number of Plasmodium falciparum genotypes in Gambian children immunized with the malaria vaccine SPf66. Trans R Soc Trop Med Hyg. **1999**; 93 Suppl 1:65–68.

59. Gatei W, Kariuki S, Hawley W, et al. Effects of transmission reduction by insecticide-treated bed nets (ITNs) on parasite genetics population structure: I. The genetic diversity of Plasmodium falciparum parasites by microsatellite markers in western Kenya. Malar J. **2010**; 9(1):353.

60. Escalante AA, Ferreira MU, Vinetz JM, et al. Malaria Molecular Epidemiology: Lessons from the International Centers of Excellence for Malaria Research Network. Am J Trop Med Hyg. **2015**; 93(3 Suppl):79–86.

61. Baliraine FN, Afrane YA, Amenya DA, et al. A cohort study of Plasmodium falciparum infection dynamics in Western Kenya Highlands. BMC Infect Dis. **2010**; 10:283.

62. Daniels R, Volkman SK, Milner DA, et al. A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking. Malar J. **2008**; 7(1):223.

63. Sturrock HJW, Roberts KW, Wegbreit J, Ohrt C, Gosling RD. Tackling Imported Malaria: An Elimination Endgame. Am J Trop Med Hyg. **2015**; 93(1):139–144.

64. Danis K, Baka A, Lenglet A, et al. Autochthonous Plasmodium vivax malaria in Greece, 2011. Eurosurveillance. **2011**; 16(42):19993.

65. Menach A Le, Tatem AJ, Cohen JM, et al. Travel risk, malaria importation and malaria transmission in Zanzibar. Sci Rep. **2011**; 1:93.

66. Rodríguez-Morales AJ, Delgado L, Martínez N, Franco-Paredes C. Impact of Imported Malaria on the Burden of Disease in Northeastern Venezuela. J Travel Med. **2006**; 13(1):15–20.

67. Lo E, Lam N, Hemming-Schroeder E, et al. Frequent Spread of Plasmodium vivax Malaria Maintains High Genetic Diversity at the Myanmar-China Border, Without Distance and Landscape Barriers. J Infect Dis. **2017**; 216(10):1254-1263.

68. Liu H, Wang J-Z, Zhang Y, Guo X-R, Xu J-W. Risk Factors for Border Malaria in a Malaria Elimination Setting: A Retrospective Case-Control Study in Yunnan, China. Am J Trop Med Hyg. **2015**; 92(3):546–551.

69. Parker DM, Matthews SA, Yan G, et al. Microgeography and molecular epidemiology of malaria at the Thailand-Myanmar border in the malaria pre-elimination phase. Malar J. **2015**; 14(1):198.

70. Richards AK, Banek K, Mullany LC, et al. Cross-border malaria control for internally displaced persons: observational results from a pilot programme in eastern Burma/Myanmar. Trop Med Int Heal. **2009**; 14(5):512–521.

71. Al-Hamidhi S, Mahdy MAK, Idris MA, et al. The prospect of malaria elimination in the Arabian Peninsula: A population genetic approach. Infect Genet Evol. **2014**; 27:25–31.

72. World Health Organization. Overview of malaria elimination. **2018**. Available

from: http://www.who.int/malaria/areas/elimination/overview/en/

73. Obaldia N, Baro NK, Calzada JE, et al. Clonal outbreak of Plasmodium falciparum infection in eastern Panama. J Infect Dis. **2015**; 211(7):1087–96.

74. Patel JC, Taylor SM, Juliao PC, et al. Genetic Evidence of Drug-Resistant Malarial Strain from the Democratic Republic of the Congo Imported to Guatemala. Emerg Infect Dis. **2014**; 20(6):932–940.

75. Zavala F, Cochrane AH, Nardin EH, Nussenzweig RS, Nussenzweig V. Circumsporozoite proteins of malaria parasites contain a single immunodominant region with two or more identical epitopes. J Exp Med. **1983**; 157(6):1947–57.

76. Gordon DM, McGovern TW, Krzych U, et al. Safety, immunogenicity, and efficacy of a recombinantly produced Plasmodium falciparum circumsporozoite protein-hepatitis B surface antigen subunit vaccine. J Infect Dis. **1995**; 171(6):1576–85.

77. Casares S, Brumeanu T-D, Richie TL. The RTS,S malaria vaccine. Vaccine. **2010**; 28:4880–4894.

78. Clinical Trials Partnership. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. Lancet. **2015**; 386(4).

79. Takala SL, Plowe C V. Genetic diversity and malaria vaccine design, testing and efficacy: Preventing and overcoming “vaccine resistant malaria.” Parasite Immunol. **2009**; 31(9):560–573.

80. Neafsey DE, Juraska M, Bedford T, et al. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. N Engl J Med. **2015**; 373(21):2025–2037.

81. Takala SL, Coulibaly D, Thera MA, et al. Extreme polymorphism in a vaccine antigen and risk of clinical malaria: implications for vaccine development. Sci Transl Med. **2009**; 1(2):2ra5.

82. Takala SL, Coulibaly D, Thera MA, et al. Dynamics of Polymorphism in a Malaria Vaccine Antigen at a Vaccine-Testing Site in Mali. PLoS Med. **2007**; 4(3):e93.

83. Ouattara A, Barry AE, Dutta S, Remarque EJ, Beeson JG, Plowe C V. Designing malaria vaccines to circumvent antigen variability. Vaccine. **2015**; 33(52):7506–12.

84. Barry AE, Schultz L, Buckee CO, Reeder JC. Contrasting Population Structures of the Genes Encoding Ten Leading Vaccine-Candidate Antigens of the Human Malaria Parasite, Plasmodium falciparum. PLoS One. **2009**; 4(12):e8497.

85. Alloueche A, Milligan P, Conway DJ, et al. Protective Efficacy of the RTS,S/AS01 Plasmodium falciparum malaria vaccine is not strain specific. Am J Trop Med Hyg. **2003**; 68(1):97-101.

86. Bousema T, Griffin JT, Sauerwein RW, et al. Hitting Hotspots: Spatial Targeting of Malaria for Control and Elimination. PLoS Med. **2012**; 9(1):e1001165.

87. Sturrock HJW, Hsiang MS, Cohen JM, et al. Targeting Asymptomatic Malaria Infections: Active Surveillance in Control and Elimination. PLoS

Med. **2013**;10(6):e1001467.

88. Alegana VA, Atkinson PM, Lourenço C, et al. Advances in mapping malaria for elimination: fine resolution modelling of Plasmodium falciparum incidence. Sci Rep. **2016**; 6:29628.

89. Gonçalves BP, Kapulu MC, Sawa P, et al. Examining the human infectious reservoir for Plasmodium falciparum malaria in areas of differing transmission intensity. Nat Commun. **2017**; 8(1):1133.

90. Searle KM, Katowa B, Kobayashi T, et al. Distinct parasite populations infect individuals identified through passive and active case detection in a region of declining malaria transmission in southern Zambia. Malar J. **2017**; 16(1):154.

91. Rice BL, Golden CD, Anjaranirina EJG, Botelho CM, Volkman SK, Hartl DL. Genetic evidence that the Makira region in northeastern Madagascar is a hotspot of malaria transmission. Malar J. **2016**; 15(1):596.

92. Carlton JM, Perkins SL, Deitsch KW. Plasmodium Genomics and the Art of Sequencing Malaria Parasite Genomes. Malar Parasites Comp Genomics, Evol Mol Biol. Norfolk: Caister Academic Press; **2013.** p. 35–58.

93. Bright AT, Tewhey R, Abeles S, et al. Whole genome sequencing analysis of Plasmodium vivax using whole genome capture. BMC Genomics. **2012**; 13:262.

94. Sriprawat K, Kaewpongsri S, Suwanarusk R, et al. Effective and cheap removal of leukocytes and platelets from Plasmodium vivax infected blood. Malar J. **2009**; 8:115.

95. Carpi G, Walter KS, Bent SJ, Hoen AG, Diuk-Wasser M, Caccone A. Whole genome capture of vector-borne pathogens from mixed DNA samples: a case study of Borrelia burgdorferi. BMC Genomics. **2015**; 16(1):434.

96. Carpi G, Walter KS, Mamoun C Ben, et al. Babesia microti from humans and ticks hold a genomic signature of strong population structure in the United States. BMC Genomics. **2016**; 17(1):888.

97. Oyola SO, Manske M, Campino S, et al. Optimized whole-genome amplification strategy for extremely AT-biased template. DNA Res. **2014**; 21(6):661–71.

98. Oyola SO, Ariani C V, Hamilton WL, et al. Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. Malar J. **2016**; 15(1):597.

99. Cowell AN, Loy DE, Sundararaman SA, et al. Selective Whole-Genome Amplification Is a Robust Method That Enables Scalable Whole-Genome Sequencing of Plasmodium vivax from Unprocessed Clinical Samples. American Society for Microbiology; **2017**; 8(1):e02257-16.

100. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat Methods. **2009**; 6(4):291–295.

101. Aird D, Ross MG, Chen W-S, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. **2011**; 12(2):R18.

102. Aydemir O, Janko M, Hathaway NJ, et al. Drug-Resistance and Population Structure of Plasmodium falciparum Across the Democratic Republic of Congo Using High-Throughput Molecular Inversion Probes. J Infect Dis. **2018**; 218(6):946-955.

103. Morin PA, Luikart G, Wayne RK, Snp Workshop Group T. SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution. **2004**; 19(4):208-216.

104. Vignal A, Milan D, SanCristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol. **2002**; 34(3):275–305.

105. Daniels R, Volkman SK, Milner D a, et al. A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking. Malar J. **2008**; 7:223.

106. Koepfli C, Mueller I. Malaria Epidemiology at the Clone Level. Trends Parasitol. **2017**; 33(12):974–985.

107. Coates BS, Sumerford D V, Miller NJ, et al. Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. J Hered. **2009**; 100(5):556–564.

108. Payseur BA, Jing P, Haasl RJ. A genomic portrait of human microsatellite variation. Mol Biol Evol. **2011**; 28(1):303–12.

109. Lin JT, Hathaway NJ, Saunders DL, et al. Using Amplicon Deep Sequencing to

Detect Genetic Signatures of Plasmodium vivax Relapse. J Infect Dis. **2015**; 212:1–10.

110. Rao PN, Uplekar S, Kayal S, et al. A method for amplicon deep sequencing of drug resistance genes in *Plasmodium falciparum* clinical isolates from India. J Clin Microbiol. **2016**; 54(6):1500-1511.

111. Miller RH, Hathaway NJ, Kharabora O, et al. A deep sequencing approach to estimate Plasmodium falciparum complexity of infection (COI) and explore apical membrane antigen 1 diversity. Malar J. **2017**; 16:12936–17.

112. Chenet SM, Schneider K a, Villegas L, Escalante A a. Local population structure of Plasmodium: impact on malaria control and elimination. Malar J. **2012**; 11(1):412.

113. Jalloh A, Jalloh M, Matsuoka H. T-cell epitope polymorphisms of the Plasmodium falciparum circumsporozoite protein among field isolates from Sierra Leone: age-dependent haplotype distribution? Malar J. **2009**; 8:120.

114. Chenet SM, Taylor JE, Blair S, Zuluaga L, Escalante AA. Longitudinal analysis of Plasmodium falciparum genetic variation in Turbo, Colombia: implications for malaria control and elimination. Malar J. **2015**; 14:363.

115. Ochola-Oyier LI, Okombo J, Wagatua N, et al. Comparison of allele frequencies of Plasmodium falciparum merozoite antigens in malaria infections sampled in different years in a Kenyan population. Malar J. **2016**; 15(1):261.

116. 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):113–25.

117. Zhu SJ, Almagro-Garcia J, McVean G. Deconvolution of multiple infections in Plasmodium falciparum from high throughput sequencing data. Bioinformatics. **2018**; 34(1):9-15.

118. WHO. Zambia: Malaria Report. **2016.**

119. Kamuliwo M, Chanda E, Haque U, et al. The changing burden of malaria and association with vector control interventions in Zambia using district-level surveillance data, 2006–2011. Malar J. **2013**; 12(1):437.

120. World Health Organization. Zambia: WHO Statistical Profile. Available from: http://www.who.int/gho/countries/zmb.pdf?ua=1

121. UNICEF. Zambia: Malaria. **2015**. Available from: http://www.unicef.org/zambia/5109_8454.html

122. Masaninga F, Chanda E, Chanda-Kapata P, et al. Review of the malaria epidemiology and trends in Zambia. Asian Pac J Trop Biomed. **2013**; 3(2):89–94.

123. Malaria Atlas Project. Malaria Atlas Project: Zambia. **2010**. Available from: http://www.map.ox.ac.uk/explore/countries/zmb/

124. Laban NM, Kobayashi T, Hamapumbu H, et al. Comparison of a PfHRP2-based rapid diagnostic test and PCR for malaria in a low prevalence setting in rural southern Zambia: implications for elimination. Malar J. **2015**; 14(1):25.

125. Moss WJ, Dorsey G, Mueller I, et al. Malaria Epidemiology and Control within

the International Centers of Excellence for Malaria Research. Am J Trop Med Hyg. **2015**; 93:5–15.

126. Mukonka VM, Chanda E, Haque U, et al. High burden of malaria following scale-up of control interventions in Nchelenge District, Luapula Province, Zambia. Malar J. **2014**; 13(1):153.

127. Kobayashi T, Bobanga T, Umesumbu S, et al. The impact of mass bed net distribution on vector species and malaria prevalence in Kilwa and Kashobwe, Haut-Katanga Province, Democratic Republic of the Congo. Proc 67th Annu Am Soc Trop Med Hygeine. New Orleans; **2018**.

128. Jones CM, Lee Y, Kitchen A, et al. Complete Anopheles funestus mitogenomes reveal an ancient history of mitochondrial lineages and their distribution in southern and central Africa. Sci Rep. **2018**; 8(1):9054.

Chapter 2

Genetic evidence of focal *Plasmodium falciparum* transmission in a pre-elimination setting in Southern Province, Zambia

ABSTRACT

Background

Southern Province, Zambia, has experienced a dramatic decline in *Plasmodium falciparum* malaria transmission in the past decade and is targeted for elimination. Zambia's National Malaria Elimination Program recommends reactive case detection (RCD) within 140m of index households to enhance surveillance and eliminate remaining transmission foci.

Methods

To evaluate whether RCD captures local transmission, we genotyped 26 microsatellites from 106 samples collected from index (n=27) and secondary (n=79) cases detected through RCD in the Macha Hospital catchment area between January 2015 and April 2016.

Results

Participants from the same RCD event harbored more genetically related parasites than those from different RCD events, suggesting that RCD captures, at least in part, infections related through local transmission. Related parasites clustered in space and time, up to at least 250m from index households. Spatial analysis identified a putative focal transmission hotspot.

Conclusions

The current RCD strategy detects focal transmission events, although programmatic guidelines to screen within 140m of index households may fail to capture all secondary cases. This study highlights the utility of parasite genetic data in assessing programmatic interventions, and similar approaches may be useful to malaria elimination programs seeking to tailor intervention strategies to the underlying transmission epidemiology.

INTRODUCTION

The global decline in malaria between 2000 and 2015 has reinvigorated commitment to elimination [1]. Many countries have malaria elimination targets before 2030, including eight in southern Africa. Although interventions such as long lasting insecticide treated bed nets (LLINs), indoor residual spraying (IRS), and case management have been and remain important malaria control tools [1], transitioning from control to elimination will necessitate incorporating new strategies capable of permanently interrupting remaining avenues for local transmission [2].

Current elimination strategies recommend variations of active case detection (ACD) in which community health workers (CHWs) screen community members for malaria using rapid diagnostic tests (RDTs) to identify asymptomatically infected individuals who may contribute to ongoing transmission [3,4]. Despite consensus that reducing the asymptomatic reservoir is theoretically important to elimination by preventing reestablishment of local transmission, the empirical evidence that screen-and-treat initiatives reduce the burden of malaria in practice is inconsistent [5–8]. While simulation analyses and a study from Zambia reported decreases in malaria infection following screen-and-treat efforts [5,9], two cluster-randomized trials in Kenya and

Burkina Faso found no difference in the change in malaria burden between villages receiving longitudinal screen-and-treat and villages receiving no intervention [7,8].

Beyond conflicting evidence of effectiveness, ACD is costly and resource-intensive as an elimination strategy, particularly in low transmission regions where few individuals have levels of parasitemia above the RDT limit of detection [10]. Therefore, reactive case detection (RCD), a specific type of ACD, has been adopted by numerous control programs in low transmission settings, focusing screening efforts to geographic transmission "hotspots" rather than the entire community [3]. RCD, initiated after passive detection of an index case at a health facility, involves screening and treating individuals within the index household and in households a prescribed distance from the index household [3,4]. Multiple studies have reported geographic clustering of malaria cases in such hotspots and noted that individuals living in index case households have a higher likelihood of being infected, supporting RCD as a strategy to disrupt transmission in ecological niches that support malaria transmission [11–19].

In Zambia, RCD was incorporated into the National Malaria Elimination Program (NMEP) in 2011 for regions reporting less than 1% RDT positive prevalence, including Southern Province [20,21]. In 2014, a ground-truth evaluation of RCD in Southern Province revealed that across 26 RCD events, only 32% of the households screened were actually located within the targeted 140m radius, highlighting the challenges CHWs face in identifying eligible households [22]. Although RCD is theoretically appealing, the inconsistent data regarding ACD's ability to decrease community-level transmission coupled with the logistical challenges of RCD necessitates validating that RCD is appropriate for the malaria epidemiology of a given region. In particular, it is essential to

determine whether clusters of malaria cases are sustained through local transmission or other mechanisms. Whereas local transmission may be effectively targeted through interventions like RCD and focal vector control, eliminating case clusters that arose from household travel or occupational exposure would require different approaches. Distinguishing these scenarios empowers evidence-based prioritization of appropriate interventions tailored to the underlying transmission epidemiology.

Spatial and temporal patterns of malaria cases, sometimes in conjunction with travel data, have been commonly used to draw inferences regarding the epidemiology of local malaria transmission [23–25]. While an important aspect of epidemiological monitoring, spatial and temporal patterns, even when coupled with travel data, cannot reliably distinguish local transmission chains from clusters of unrelated infections. Understanding the contribution of local transmission requires comparing parasite genotypes from infected individuals [26,27]. To evaluate the ability of RCD to capture local malaria transmission in a region of Southern Province, Zambia, we performed microsatellite genotyping on dried blood spot (DBS) samples collected from index and secondary cases detected through RCD. We characterized the temporal and spatial scales on which genetically related parasites clustered, a proxy for local transmission, and identified a putative focal transmission hotspot.

METHODS

Study region

This study was conducted in the rural catchment area of Macha Hospital, an approximately 2,000km^2 region located in Choma District, Southern Province, Zambia with an estimated 56,000 residents, primarily subsistence farmers. This setting is characterized by a tropical savannah climate and experiences peak *Plasmodium falciparum* malaria during the single rainy season between November and April [28]. *Anopheles arabiensis* and *An. squamosus* are the primary and secondary malaria vectors in this region, respectively [28–30]. PfHRP2-based RDT prevalence among actively detected individuals in this region declined from 9.2% in 2008 to less than 1% in 2013 [31]. The introduction of artemisinin combination therapy with artemether-lumefantrine in 2004 along with LLIN distributions in 2007, 2012, and 2014 may have contributed to the decline in malaria in this setting. LLIN ownership in this region was estimated to be 83% in 2013 [32]. The Zambian NMEP aims to eliminate malaria from the country by 2021, and Southern Province, including the study area, is considered a pre-elimination setting. For regions like this with low level malaria transmission, the NMEP recommends RCD whereby CHWs screen using PfHRP2-based RDTs and treat positive individuals living within 140m of index case households.

Informed consent

This research was approved by the Institutional Review Board at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, USA and the Ethics Review Committee of the Tropical Diseases Research Centre in Ndola, Zambia. All adult

participants gave informed consent. All children had parental consent, and children ages 12-15 provided assent.

Enhanced reactive case detection

Following passive detection of PfHRP2-RDT positive index cases at fourteen rural health centers (RHCs) serving the catchment area of Macha Hospital, study team members at Macha Research Trust (MRT) were alerted by text message. Within one week of the index case's presentation at the RHC, study team members accompanied a CHW to screen all individuals living within an expanded RCD radius of 250m from the index case household using RDTs. RDT positive cases were offered treatment with artemether-lumefantrine. For all individuals screened, study team members collected a finger-prick blood sample as DBS on filter paper (Whatman 903™ Protein Saver Card), administered an epidemiological and demographic survey, and collected global positioning system (GPS) coordinates for each household. Households within 250m of the index household were pre-identified using satellite imagery to guide the field team in enrolling eligible households. DNA from the DBS samples was extracted using Chelex and screened for *P. falciparum* using quantitative PCR (qPCR) targeting mitochondrial *cytochrome-b* at the MRT laboratory in Macha, Zambia [31].

Passively detected cases at RHCs

Five of the 14 RHCs closest to Macha Hospital collected finger prick blood samples as DBS from a convenience sample of passively detected index cases as previously

described [33]. DBS were transported from the RHCs to MRT. Index cases were linked by name and visit date recorded at the RHC to the epidemiological and demographic survey questionnaire administered to the same individual up to one week later during RCD follow-up.

Sample selection

We selected all RCD events between January 1, 2015 and April 18, 2016 for which at least two participants, including the index case, were malaria positive by qPCR. One hundred and six samples (27 index, 79 secondary) across 49 RCD events met our inclusion criteria and DBS were transported to the University of California in San Francisco.

Laboratory procedures

DBS were stored with desiccant at -20° C until transport and processing. Six millimeter hole-punches from filter cards were extracted using the saponin Chelex method [34] and quantified using *var*-ATS ultra-sensitive qPCR [35]. Samples containing less than 10 parasites/μL of blood (45 of 106 samples) underwent selective whole genome amplification (sWGA) to selectively enrich for parasite DNA as previously described with some modifications described in Supplemental Table S1 [36,37]. For quality control, we assessed the concordance of alleles and determined the error rate introduced during sWGA by genotyping control samples with and without sWGA. Samples with >10 parasites/μL and post-sWGA samples were amplified at 26 microsatellite loci [36,37, Liu & Tessema et al., in preparation]. PCR products were diluted and sized by denaturing capillary electrophoresis on an Thermo Fisher 3730XL analyzer with GeneScan™

400HD ROX™ size standard (Thermo Fisher Scientific, Waltham, MA). The resulting electropherograms were analyzed using microSPAT [Murphy et al., in preparation]. Samples with < 10 successfully genotyped microsatellites and loci that did not amplify in at least 50% of the samples were excluded from the analysis. For each genotyped sample, we calculated the multiplicity of infection (MOI), or the number of genetically distinct parasite clones, as the second highest number of alleles present at any locus, accounting for the possibility of false positive allele calls.

Characterizing relatedness captured during RCD

For each pair of samples genotyped, pairwise genetic relatedness was determined using a modified identity by state (IBS) metric [39, Liu and Tessema et al., in preparation]. Pairwise IBS was calculated from allele similarity between isolates, allowing estimation of genetic relatedness from monoclonal as well as polyclonal samples. For each locus, IBS was estimated by the sum of the product of within host allele frequency (defined as 1/number of alleles detected) of shared alleles. The overall pairwise IBS was determined as the average of locus-specific estimates under the assumption of independent loci [Liu and Tessema et al., in preparation]. We calculated the physical distance between all sample pairs as Euclidean distance between households using R package geosphere [41]. Elapsed time between sample pairs was determined from the absolute value of the difference in sample collection dates. Sample pairs were considered related if their pairwise genetic relatedness was greater than or equal to 0.5, the 96^{th} percentile of pairwise relatedness across all sample pairs. We characterized mean pairwise genetic relatedness as well as the proportion of related samples across bins of increasing physical

distance. For each RCD event, we calculated the mean pairwise relatedness between all sample pairs from that event.

Characterizing the temporal and spatial boundaries of malaria transmission

We investigated the optimal temporal and spatial scales on which to detect related parasites in the study area. Given that the sampling framework inherently oversampled individuals clustered in time and space, we removed all pairwise comparisons between samples from the same RCD event. Using this subset of data corrected for the non-random sampling strategy, we binned the remaining pairwise comparisons into nine space-time categories and calculated the mean genetic relatedness among pairwise comparisons across increasing time and distance.

Identifying focal transmission hotspots

We developed a method to identify regions that sustain local transmission at higher than expected rates, referred to as focal transmission hotspots. We used R package sp [42–44] to draw lines between all pairs of individuals harboring related parasites after removing pairwise comparisons between samples from the same RCD event. We created a grid with 81 equally sized cells across the study region. For each cell, we counted the number of related lines that passed through, generating a count of observed related connections. To calculate an expected number of related connections for each cell under the assumption that connections were random, we generated 1,000 sampling replicates of size equal to the total number of related lines, each time sampling from all possible lines connecting all pairs of individuals regardless of relatedness. For each replicate, we

summed the number of lines that passed through each cell, generating an expected line count and 95% confidence interval for each cell. We calculated the ratio of observed to expected lines for each cell and exported this file for use in ArcMap (ESRI, version 10.5.1)[45].

RESULTS

Genotyping

Of the 26 microsatellite loci amplified, all but locus PolyA had genotyping success rates greater than 50% (Supplemental Figure S2). Locus PolyA was therefore excluded from the analysis. Eighty-five of the 106 extracted samples were successfully genotyped for at least 10 of the 25 microsatellite loci and were included in the analysis. Insufficient coverage was significantly associated with negative RDT result, being a secondary case, and low parasitemia (Supplemental Figure S3). Among the 21 samples with insufficient coverage, 19 underwent sWGA. The relationship between microsatellite coverage and parasitemia is shown in Supplemental Figure S3. Samples excluded from the analysis did not differ significantly from those included in terms of sex, age, self-reported overnight travel, or self-reported bednet use. In a quality control experiment, there was over 99% allele concordance in control samples genotyped pre- and post-sWGA (Supplemental Table S2).

The 85 successfully genotyped samples came from 47 different RCD events. Index (n=27) and secondary case individuals (n=58) did not differ significantly with respect to sex, age, self-reported bednet use, or self-reported overnight travel (Supplemental Table

S3). Index cases had a higher proportion of RDT positive individuals and a lower proportion of sWGA samples compared to secondary cases, as expected given the higher parasite densities generally seen in symptomatic infections (Supplemental Table S3). Among the samples included, the median MOI was two. MOI did not differ between index and secondary cases (Supplemental Figure S4) or between sWGA and non-sWGA samples.

Relatedness by physical distance

There was no significant difference in mean genetic relatedness between samples collected from within the same household and those 141-250m apart (Figure 1). Samples collected 1-140m apart were significantly less related than those within the same household (Student's t-test: $p=0.004$) and those 141-250m apart (Student's t-test: $p=0.05$), but this difference may reflect the small sample size of 10 comparisons in the 1-140m distance bin. When considered together, samples collected from participants located within 250m had significantly higher mean genetic relatedness (mean relatedness: 0.71) than participants more than 250m apart (mean relatedness: 0.31, Welch's t-test: $p<0.001$), indicating that screening with a 250m radius captures genetically related parasites. Among samples from the same RCD event, 78.6% were related, whereas only 2.9% of samples from different RCD events were related (test for difference in proportions, $p<0.001$). The distributions of relatedness among sample pairs from the same RCD event and among those from different RCD events are shown in Supplemental Figure S1. Among the related pairwise comparisons >250m apart, 60% occurred within 90 days, suggesting that relatedness among physically distant samples is nevertheless clustered temporally. In summary, RCD, both at the 140m and 250m radii, captured

genetically related cases, suggesting that case clustering is due to local transmission rather than household-level risk factors.

Relatedness within RCD events

Among the 56 sample pairs from the same RCD event, mean relatedness was 0.69. There was no significant difference in mean relatedness comparing two secondary cases, two index cases, or an index and secondary case from the same RCD event (ANOVA, p=0.62) (Figure 2). Among the seven RCD events with comparisons across multiple categories (index-index, secondary-secondary, or index-secondary), there was no difference in relatedness across categories, possibly reflecting the small sample size.

Relatedness across space and time

Related comparisons clustered focally in space and time (Figure 3), and pairwise genetic relatedness declined with increasing distance and elapsed time. Ninety-six percent of the 52 most highly related comparisons (related > 0.80) were within 10km and 92% occurred within 90 days (Figure 3), indicating that transmission occurred focally in space and time. Next, to verify that these relationships were not driven by the focal sampling framework, we removed all pairwise comparisons between samples from the same RCD event. Figure 4 illustrates the decline in genetic relatedness across increasing distance and elapsed time after removing these comparisons. After excluding pairs of samples from the same RCD event, most related comparisons occurred within 10km and 60 days, although the small sample size of comparisons within 1,000m precludes our ability to infer definitive boundaries between 0-1,000m. Marginal distributions of relatedness by physical distance and elapsed time are included in Supplemental Figure S5.

Identifying focal transmission hotspots

We identified a region in the northwest quadrant of our study area as a putative focal transmission hotspot (Figure 5). After removing pairs of samples from the same RCD event and accounting for the spatial distribution and density of households in the analysis, this region had a significantly higher than expected number of related connections. Conversely, two regions located in the west-central and north-central of our study site had significantly fewer than expected related connections. Parasite genetic data enabled the detection of finer-scale transmission heterogeneities in our study area than could have been detected through epidemiological and spatial patterns alone.

DISCUSSION

Microsatellite genotyping from RCD participants demonstrated that focal malaria transmission was ongoing between January 2015 and April 2016 in a region of Southern Province, Zambia targeted for elimination. Although spatial clustering of malaria cases has been previously described, [11–14], this is, to our knowledge, the first study to genotype *P. falciparum* identified during RCD in southern Africa to attribute case clustering to local transmission. Using genetic methods to clarify the mechanism through which clusters of malaria cases arise has important implications for programmatic planning, enabling evidence-based decisions regarding effective control tools suited to a region's unique malaria epidemiology. This study demonstrates that RCD captures genetically related infections and underscores the added value of genetic data to

understanding the underlying malaria transmission epidemiology of a given region beyond what can be inferred through spatial and temporal patterns alone.

Our data, implicating local transmission as a contributor to the sustained malaria burden, highlights the importance of addressing the underlying transmission biology and ecology that perpetuates malaria in this region. These results suggest that local transmission in this region extends up to 250m from the index household, and possibly beyond. Therefore, the Zambian NMEP's recommended 140m screening radius at one timepoint likely fails to capture infected individuals who may be capable of sustaining transmission after RCD. However, performing RCD across the entire landscape of local transmission may not be feasible or efficient, as previously suggested [46,47]. Exploring additional options, such as repeated visits or coupling focal vector control with RCD may be useful to interrupt local transmission in this and similar settings. With the goal of identifying regions where transmission-disrupting interventions like focal vector control could have the largest impact, we describe a novel approach for detecting focal transmission hotspots. Focal transmission hotspots may be ideal locations to implement additional transmission-disrupting interventions in conjunction with RCD. This proposed method for identifying focal transmission hotspots has not yet been rigorously assessed and will require expanded evaluation before being considered a reliable indicator for transmission foci.

The RCD sampling framework in this study is not the ideal dataset to delineate the spatial and temporal scales of local transmission. We did not genotype every index and secondary case between January 2015 and April 2016, but rather selected samples that came from RCD events where at least two individuals tested positive for malaria by

RDT or qPCR. Given the non-random inclusion criteria and somewhat sparse sampling of this study, the spatial and temporal boundaries of local transmission we can infer—10km and 60-90 days—require additional validation. Despite our inability to unequivocally identify the spatial and temporal boundaries of local transmission, our data strongly suggest that local transmission occurs focally in space and time. Temporal and spatial clustering was observed regardless of whether sample pairs from the same RCD event were included or excluded, indicating that our sampling framework did not solely determine the results. Our observations are consistent with simulation analyses from the region that suggested a 500m RCD radius would identify 77% of additional cases [47]. Similarly, our observation that transmission occurs temporally for up to 60-90 days is corroborated by 24-SNP molecular barcode data suggesting ongoing transmission by highly related subsequent infections within index households 30 and 90 days after the initial RCD intervention [48]. Defining exact boundaries of local transmission would necessitate sampling longitudinally from a spatially representative framework.

We inferred local transmission from pairwise genetic relatedness. Although we are unable to confidently assert that any sample pairs harboring related parasites represented direct transmission from one to another, it is improbable that we would see genetic relatedness values above 0.5 unless the individuals were connected in some manner through transmission. Our 0.5 threshold represented the 96th percentile of all pairwise relatedness values in the study, indicating how rare these related events are. Despite not having the granularity of connecting individual transmission events, the observed related infections likely indicate direct or indirect local transmission.

The ability to obtain genotypes from asymptomatic infections, the diverse, multiallelic loci used, and the use of a similarity metric capable of comparing infections with varying MOI contributed to the success of this study in a setting of low density, polyclonal infections. However, there is room for further improvement of these methods. For example, we were unable to assign meaning to categories of genetic relatedness values below 0.5, given the low precision of characterizing how relatedness across transmission generations decays for these markers in the context of polyclonal infections.

Characterizing the contribution of local transmission to sustained malaria is critical to programs transitioning from control to elimination. Knowledge of whether cases arise through travel, occupational exposure, or locally acquired transmission can guide intervention priorities. We discerned signatures of local malaria transmission using 106 DBS samples and a relatively low-tech genotyping method. Similar approaches could be valuable to other researchers and program leaders, particularly in resource-limited settings, interested in characterizing local transmission to guide malaria elimination.

Figure 1

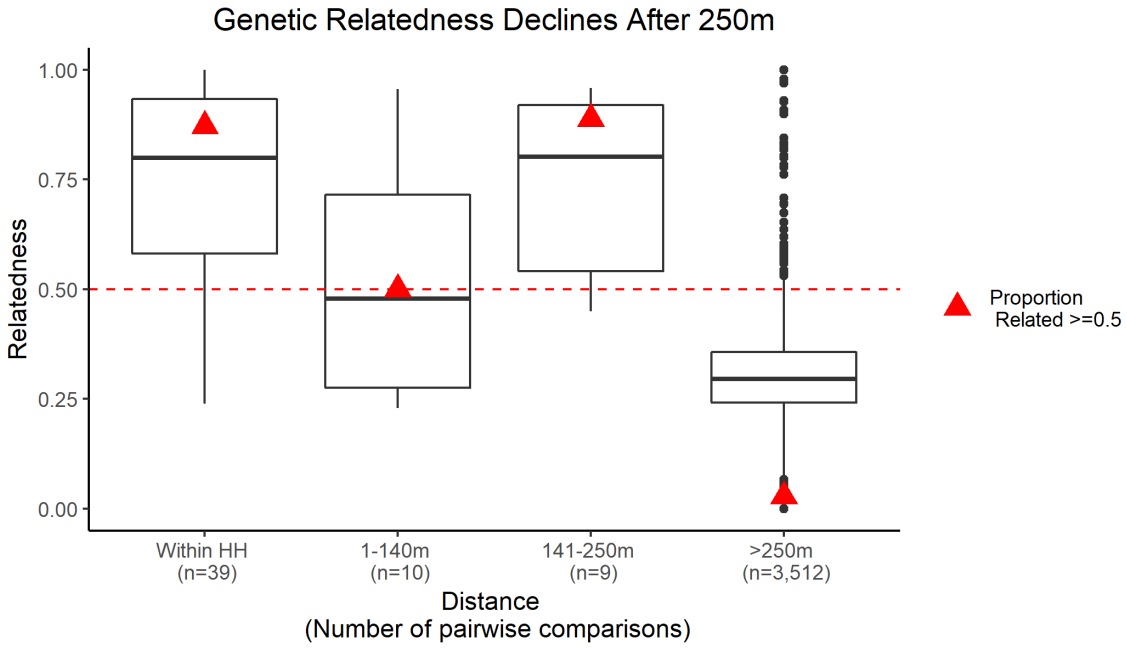

Figure 1: Genetic relatedness is shown for pairwise comparisons binned by increasing physical distance between sample pairs (within the same household (HH), samples 1-140m apart, samples 141-250m apart, and samples greater than 250m apart). The number of pairwise comparisons in each physical distance bin is listed below the bin label in parenthesis. Red triangles indicate the proportion of comparisons in each category that are related above the 0.5 genetic relatedness threshold demarcated by the red dashed line. Among sample pairs >250m apart, the mean physical distance was 20.2km (range: 297m-46.1km).

Figure 2

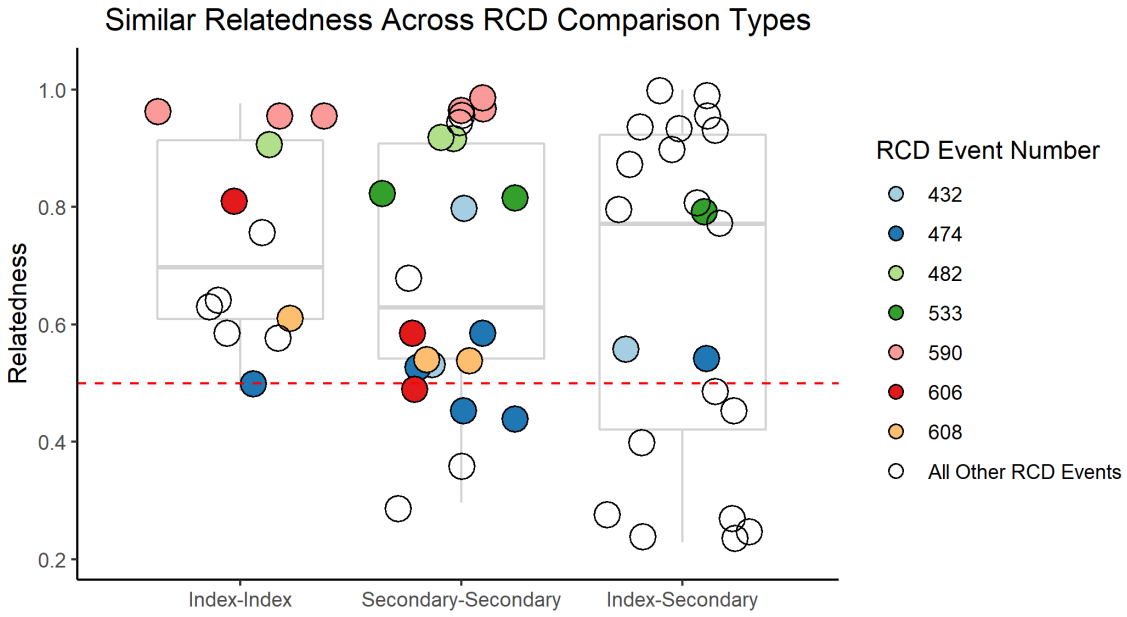

Figure 2: There were 56 pairwise comparisons between samples from the same RCD event, representing 23 unique RCD events. Each of the 56 comparisons were binned according to whether the comparison was between two index cases ("Index-Index"; n=12 pairwise comparisons), between two secondary cases ("Secondary-Secondary"; n=22 pairwise comparisons), or between an index case and a secondary case ("Index-Secondary"; n=22 pairwise comparisons) from the same RCD event. Genetic relatedness is shown for each of the 56 comparisons, represented as circles. For seven RCD events that had comparisons in more than one category (Index-Index, Secondary-Secondary, or Index-Secondary), comparisons are colored identically. Pairwise comparisons from RCD events for which only one category is represented are shown with hollow circles. The distributions of all pairwise comparisons for each of the three categories are shown by boxplots in light grey.

Figure 3

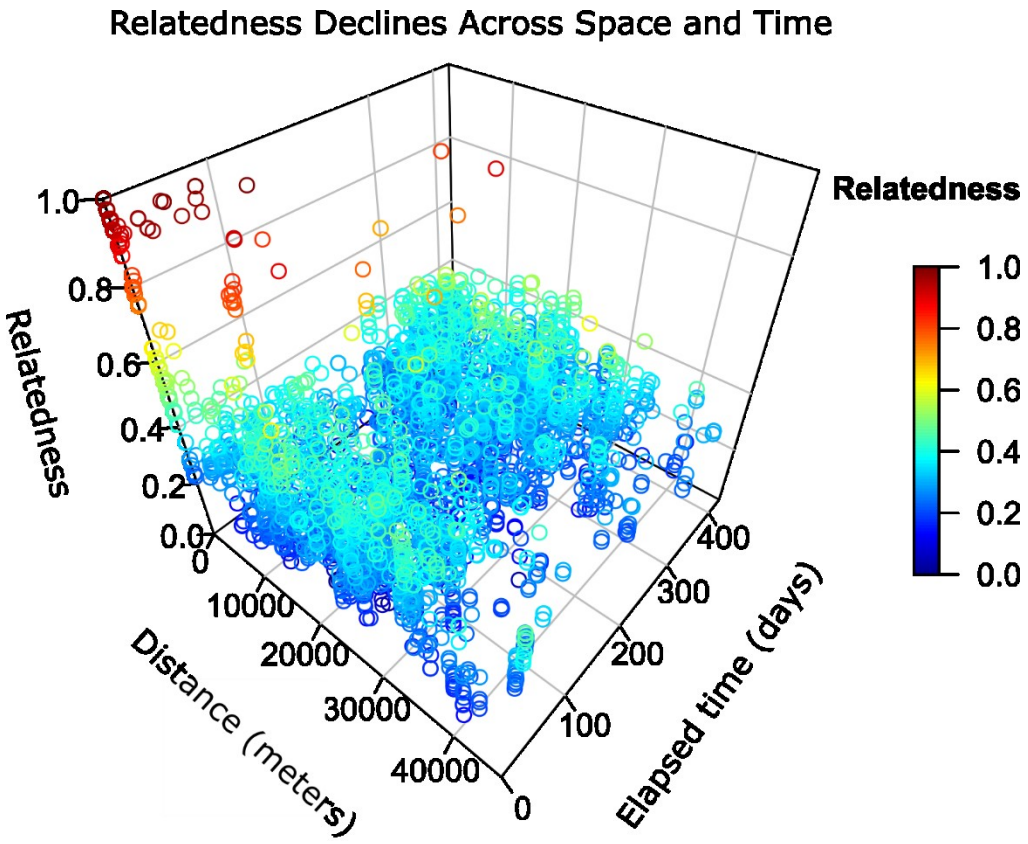

Figure 3: Each hollow circle represents a unique pairwise comparison between two samples. Circles are colored according to the genetic relatedness between the samples compared. Red indicates that the samples are highly related, and dark blue indicates that the samples are unrelated. Physical distance and elapsed time between sample pairs are shown on the horizontal axes.

Figure 4

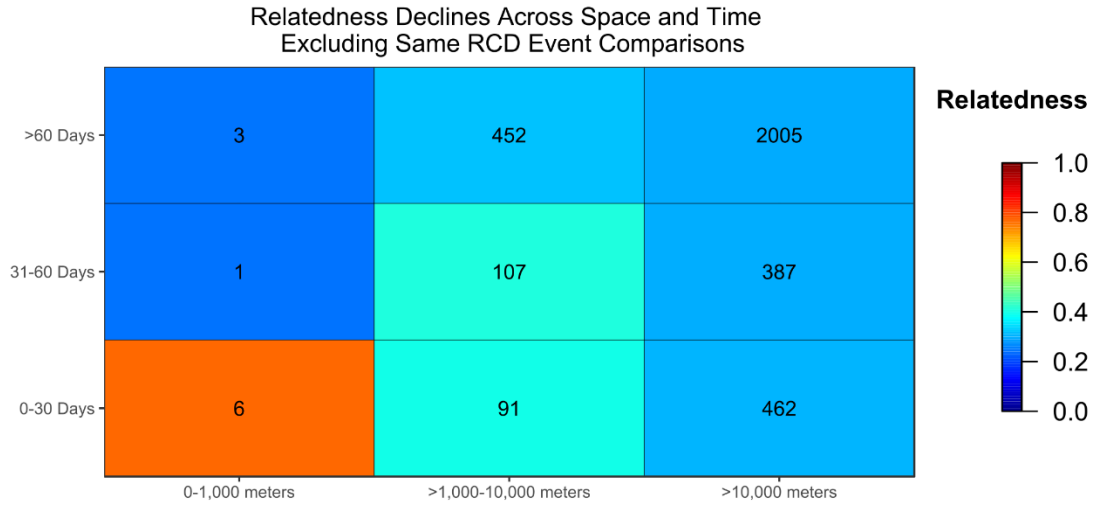

Figure 4: Pairwise comparisons (excluding those from the same reactive case detection event) were binned into nine space-time categories represented by the nine boxes in the matrix. The number in each bin denotes the number of pairwise comparisons in each category. The mean relatedness of the pairwise comparisons in each bin is shown by the color gradient, where red is highly related and dark blue is less related.

Figure 5

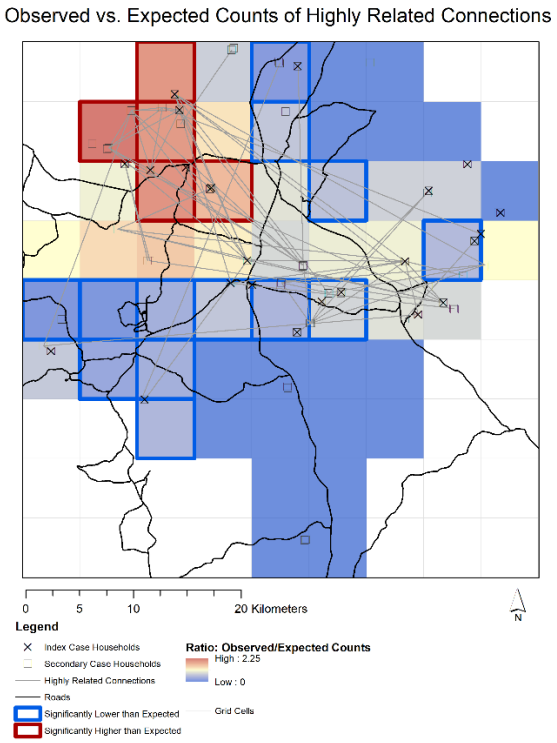

Figure 5: A map of the study area is overlaid with 81 equally sized, approximately 5km grid cells. For each grid cell, the ratio comparing the number of related connections (pairwise comparisons related at or above 0.5 after excluding comparisons from the same RCD event) to the number of connections expected is depicted by a color gradient. Red indicates that the cell has more related connections than expected; blue indicates that the cell has fewer related connections than expected. Grid cells with significantly more related connections than expected are outlined by a thick red border. Grid cells with significantly fewer than expected related connections are outlined by a thick blue border. We observed similar MOI comparing individuals living within and outside of focal transmission hotspots (mean MOI among individuals living in focal transmission hotspots= 1.62; mean MOI among individuals living outside of focal transmission hotspots=1.92; Student's T-test p=0.28).

Figure S1

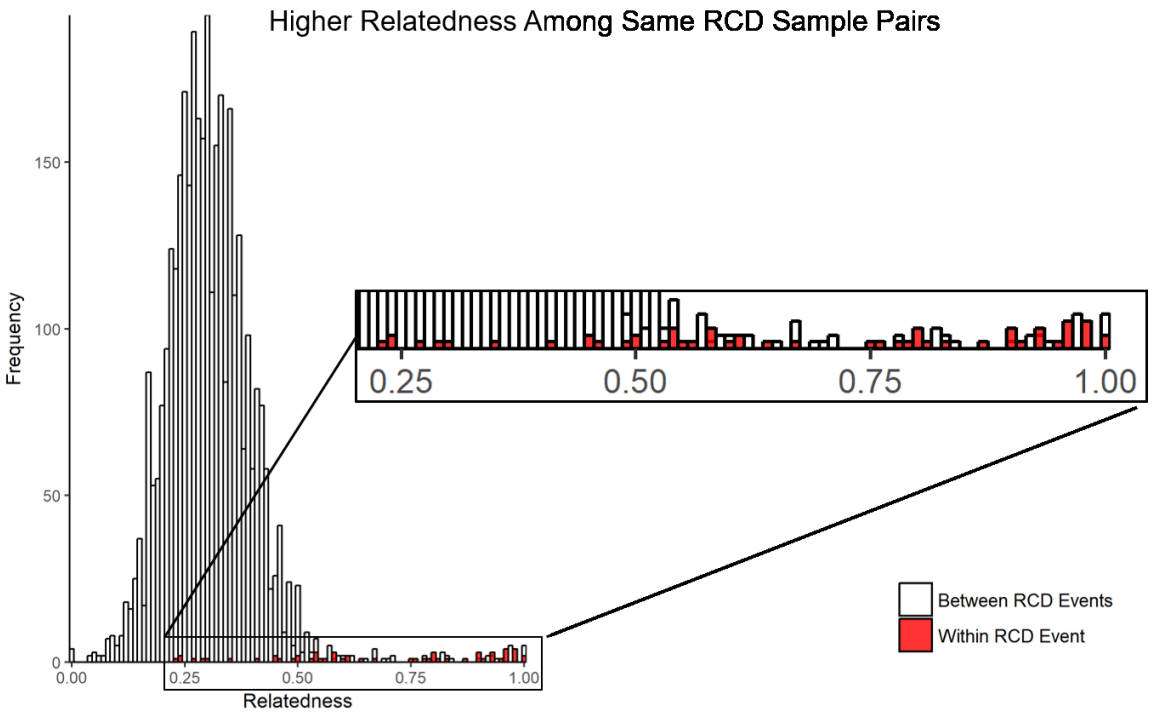

Figure S1: The histogram shows the distribution of genetic relatedness values across all pairwise comparisons in the study. White bars denote relatedness values for samples from different RCD events, and red bars indicate values for samples from the same RCD event. The boxed region of the histogram is shown magnified to the right of the histogram.

Figure S2

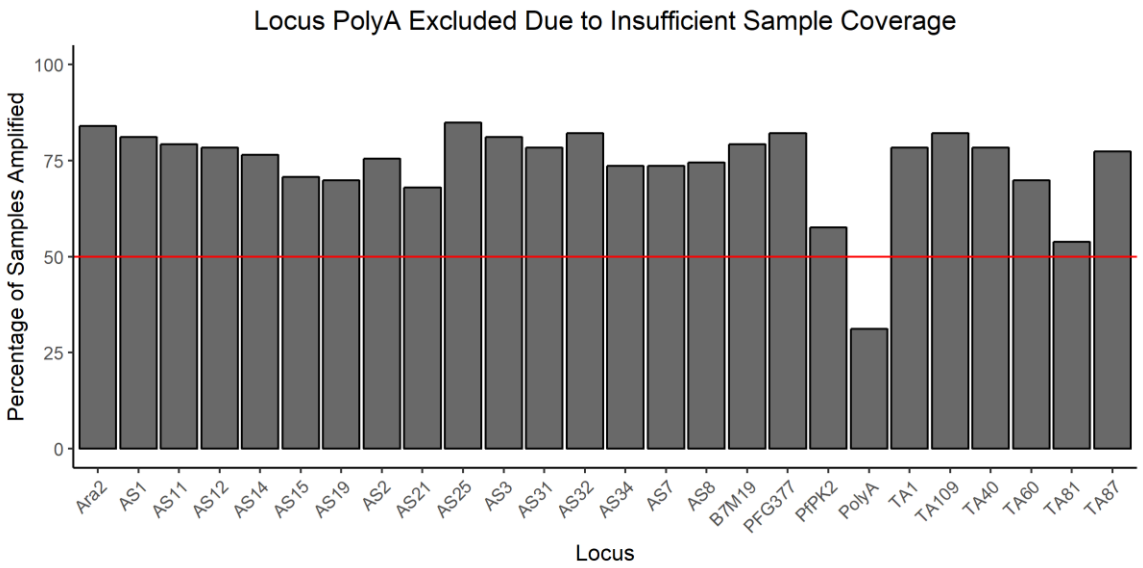

Figure S2: The percentage of samples successfully genotyped at each locus is shown by grey bars. The red horizontal line marks the inclusion threshold above which loci were required to perform for inclusion in subsequent analyses. Locus PolyA was not included in the analysis**.** Among the 25 microsatellite loci included in this study, the mean number of alleles detected per locus across all samples was 10.64 (range: 3-26).

Figure S3

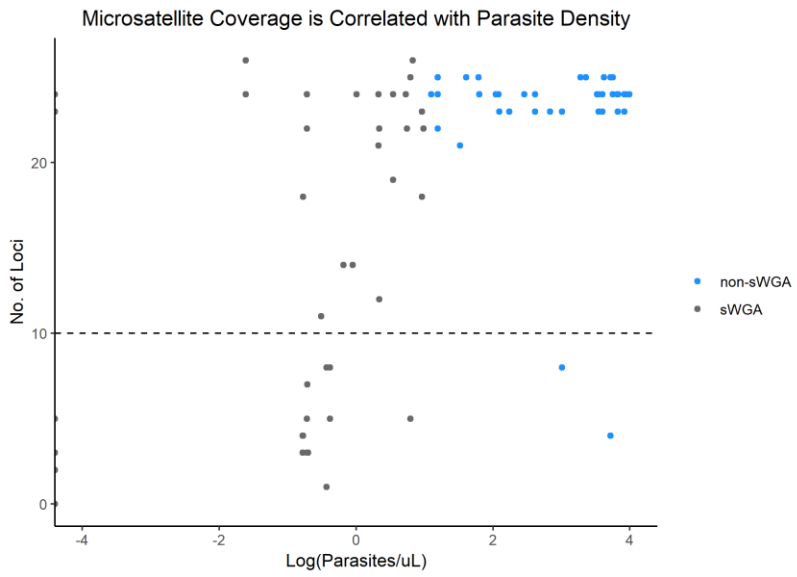

Figure S3: Parasite density (parasites/µL) and the number of microsatellite loci successfully amplified (after removing loci for which <50% of samples amplified) are shown for each sample. The dashed horizontal line marks the coverage cutoff above which samples were required to fall for inclusion in the analysis. Points in blue denote samples that did not undergo selective whole genome amplification; points in grey indicate those that did. Points lying directly on the Y-axis denote negative infinity log values, obtained for samples with 0 parasites/µL. Eighty-one percent of insufficient coverage samples tested negative by RDT compared to 40% of sufficient coverage samples (Fisher's exact test of equal proportions p=0.001). Among insufficient coverage samples, all were secondary cases; among sufficient coverage samples 68% were secondary cases (Fisher's exact test of equal proportions p=0.001). The median parasitemia among insufficient coverage samples was 0.19 parasites/µL compared to 287 parasites/µL among samples with sufficient coverage (Student's t-test p<0.001).

Figure S4

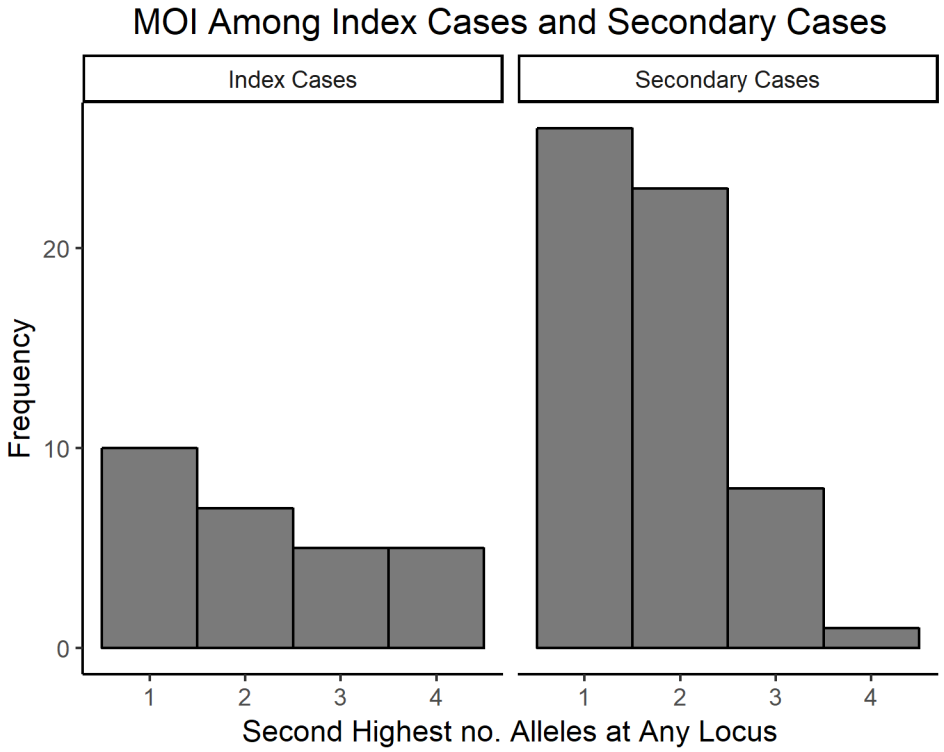

Figure S4: Multiplicity of infection (MOI) was defined as the second highest number of alleles observed at any locus for each sample. The distribution of MOI is shown for index (left) and secondary (right) cases above.

Figure S5

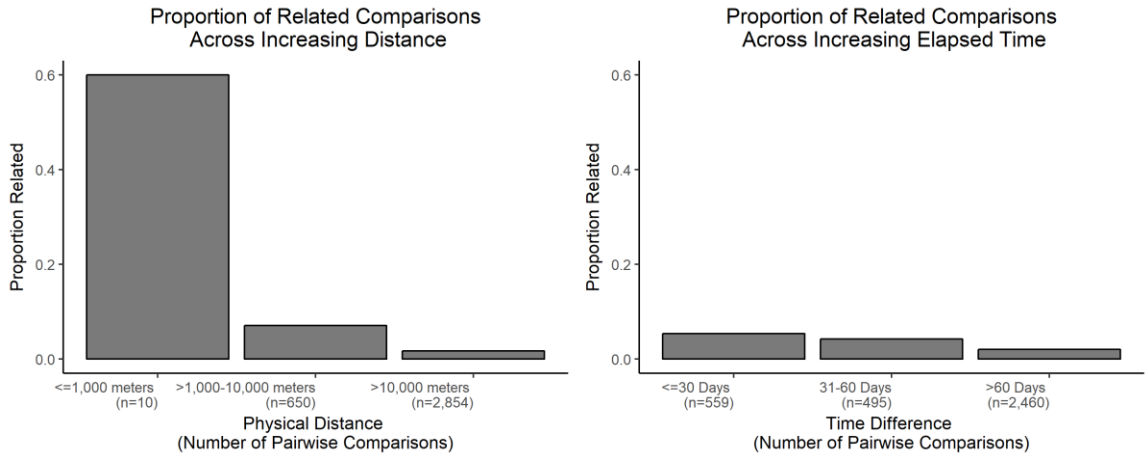

Figure S5: After excluding pairwise comparisons between samples from the same RCD event, the proportion of comparisons related at or above 0.5 was plotted against bins of increasing physical distance (left) and bins of increasing elapsed time (right). The number of comparisons in each bin is given in parenthesis below the X-axis category label.

Table S1:

Reagents	Final volume - mL	Reaction condition
10X phi29 Reaction Buffer	5 mL	35^0C for 5 min 34^0C for 10 min 33^0C for 15 min 32^0C for 20 min 31^0C for 30 min 30^0C for 16 hrs 65°C for 15 min 10°C hold
100X Bovine Serum Albumin (BSA)	0.5 mL	
100 µM Primer set 6A[a]	1.25 mL	
250 µM Primer set 10A[b]	0.5 mL	
5 mM dNTP mix (70AT:30GC ratio)	20 mL	
1000 U/mL phi29 DNA Polymerase	3 mL	
Master mix per well	30 mL	
Template DNA	20 mL	
Total reaction volume	50 mL	

Table S1: PCR reagents and reaction condition for selective whole genome amplification of low density *P. falciparum* samples. Primer sets as described in [a]Sundararaman et al., 2016 and [b]Oyola et al., 2016 were used.

Table S2:

Loci	Pre-sWGA			Post-sWGA (duplicates)					
	1000 p/mL	100 p/mL	10 p/mL	1000 p/mL		100 p/mL		10 p/mL	
AS1	176	176	176	176	176	176	176	176	176
AS2	195	195	195	195	195	195	195	195	195
AS3	170	170	170	170	170	170	170	NA	170
AS32	242	242	258	242	242	242	242	242	242
AS7	182	182	182	182	182	182	182	182	182
AS8	198	198	NA	198	198	198	198	198	198
AS11	162	162	162	162	162	162	162	162	162
AS12	165	165	165	165	165	165	165	165	165
AS14	197	197	197	197	197	197	197	197	197
AS15	128	128	NA	128	128	128	128	128	128
AS19	179	179	179	179	179	179	179	179	176
AS21	157	157	NA	157	157	157	157	157	157
AS34	187	187	187	187	187	187	187	187	187
AS25	104	104	104	104	104	104	104	104	104
B7M19	155	155	155	155	155	155	155	155	155
TA109	156	156	156	156	156	156	156	156	156
AS31	197	197	197	197	197	197	197	197	197
Ara2	146	146	NA	146	146	146	146	146	146

PfPK2	182	182	182	182	182	182	182	182	182
TA1	171	171	171	171	171	171	171	171	171
TA87	107	107	107	107	107	107	107	107	107
TA81	121	121	121	121	121	121	121	121	121
TA60	215	215	215	215	215	215	NA	215	215
PFG377	155	NA	155	155	155	155	155	155	155
PolyA	104	104	104	104	104	104	104	NA	104
TA40	271	271	271	271	271	271	271	271	NA
Coverage (%)	100	96	84.6	100	100	100	96	92.3	96

Table S2: Comparisons of microsatellite alleles in 26 loci pre- and post- sWGA of control samples with parasite densities of 1000 p/µL, 100 p/µL and 10 p/µL. Alleles detected erroneously following sWGA are indicated in red. "NA" denotes that no alleles were detected.

Table S3:

	Index Cases (%) N=27	Secondary Cases (%) N=58	Missing Data	p-value*
Sex				
Male	16 (59.3)	30 (51.7)	0	0.64
Female	11 (40.7)	28 (48.3)		
sWGA				
No	25 (92.6)	34 (58.6)	0	**0.0019**
Yes	2 (7.4)	24 (41.4)		
RDT Result				
Negative	0 (0)	34 (58.6)	0	**<0.001**
Positive	27 (100)	24 (41.4)		
Age Category				
<=5	5 (18.5)	14 (24.1)	0	0.18
5-16	10 (37.1)	30 (51.8)		
>16	12 (44.4)	14 (24.1)		
Sleep Under Bednet (usually)	12 (48.0)	26 (46.4)	4	1.0
No	13 (52.0)	30 (53.6)		
Yes				
Overnight Travel in Past Month	20 (74.0)	51 (87.9)	0	0.13

No **Yes**	7 (26.0)	7 (12.1)		
Mean Age (years) (range)	20.4 (2.2-76)	15.2 (1.6-75)	0	0.25

*Fisher's exact test for categorical variables and Student's T-test for continuous variables. All statistical tests implemented in R (version 3.4.4) with α=0.05.

Table S3: Demographic and epidemiological characteristics are compared between index and secondary case samples included in the analysis. P-values are bold for variables that differed significantly between index and secondary cases.

REFERENCES

1. Bhatt S, Weiss DJ, Cameron E, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature **2015**; 526:207-211.
2. Moonen B, Cohen JM, Snow RW, et al. Operational strategies to achieve and maintain malaria elimination. Lancet **2010**; 376(9752):1592–603.
3. Sturrock HJW, Novotny JM, Kunene S, et al. Reactive Case Detection for Malaria Elimination: Real-Life Experience from an Ongoing Program in Swaziland. PLoS One **2013**; 8(5):e63830.
4. World Health Organization. A framework for malaria elimination. **2017.**
5. Kern SE, Tiono AB, Makanga M, et al. Community screening and treatment of asymptomatic carriers of Plasmodium falciparum with artemether-lumefantrine to reduce malaria disease burden: a modelling and simulation analysis. Malar J **2011**; 10:210.
6. Sutcliffe CG, Kobayashi T, Hamapumbu H, et al. Reduced Risk of Malaria Parasitemia Following Household Screening and Treatment: A Cross-Sectional and Longitudinal Cohort Study. PLoS One **2012**;7(2):e31396.
7. Tiono AB, Ouédraogo A, Ogutu B, et al. A controlled, parallel, cluster-randomized trial of community-wide screening and treatment of asymptomatic carriers of Plasmodium falciparum in Burkina Faso. Malar J **2013**; 12:79.
8. Halliday KE, Okello G, Turner EL, et al. Impact of Intermittent Screening and Treatment for Malaria among School Children in Kenya: A Cluster Randomised Trial. PLoS Med **2014**; 11(1):e1001594.

9. Larsen DA, Bennett A, Silumbe K, et al. Population-Wide Malaria Testing and Treatment with Rapid Diagnostic Tests and Artemether-Lumefantrine in Southern Zambia: A community Randomized Step-Wedge Control Trial Design. Am J Trop Med Hyg **2015**; 92(5):913–921.

10. Rogawski ET, Congpuong K, Sudathip P, et al. Short Report: Active Case Detection with Pooled Real-Time PCR to Eliminate Malaria in Trat Province, Thailand. Am J Trop Med Hyg **2012**; 86(5):789–791.

11. Bejon P, Williams TN, Liljander A, et al. Stable and Unstable Malaria Hotspots in Longitudinal Cohort Studies in Kenya.PLoS Med **2010**; 7(7):e1000304.

12. Bousema T, Drakeley C, Gesase S, et al. Identification of Hot Spots of Malaria Transmission for Targeted Malaria Control. J Infect Dis **2010**; 201(20111):1764–1774.

13. Pinchoff J, Henostroza G, Carter BS, et al. Spatial patterns of incident malaria cases and their household contacts in a single clinic catchment area of Chongwe District, Zambia. Malar J **2015**; 14:305.

14. Björkman A, Cook J, Sturrock H, et al. Spatial Distribution of Falciparum Malaria Infections in Zanzibar: Implications for Focal Drug Administration Strategies Targeting Asymptomatic Parasite Carriers. Clin Infect Dis **2017**; 64(9):1236–43.

15. Maria Van Eijk A, Ramanathapuram L, Sutton PL, et al. What is the value of reactive case detection in malaria control? A case-study in India and a systematic review. Malar J **2016**; 15:67.

16. Stresman GH, Kamanga A, Moono P, et al. A method of active case detection to target reservoirs of asymptomatic malaria and gametocyte carriers in a rural area in Southern Province, Zambia. Malar J **2010**; 9:265.

17. Littrell M, Sow GD, Ngom A, et al. Case investigation and reactive case detection for malaria elimination in northern Senegal. Malar J **2013**; 12:331.

18. Stresman GH, Baidjoe AY, Stevenson J, et al. Focal Screening to Identify the Subpatent Parasite Reservoir in an Area of Low and Heterogeneous Transmission in the Kenya Highlands. J Infect Dis **2015;** 212(11):1768-77.

19. Aidoo EK, Afrane YA, Machani MG, et al. Reactive case detection of Plasmodium falciparum in western Kenya highlands: effective in identifying additional cases, yet limited effect on transmission. Malar J **2018**; 17(1):111.

20. Zambia National Malaria Elimination Centre. National Malaria Elimination Strategic Plan 2017-2021 Moving from accelerated burden reduction to malaria elimination in Zambia. **2017**.

21. Larsen DA, Chisha Z, Winters B, et al. Malaria surveillance in low-transmission areas of Zambia using reactive case detection. Malar J **2015**; 14(1):465.

22. Searle KM, Hamapumbu H, Lubinda J, et al. Evaluation of the operational challenges in implementing reactive screen-and-treat and implications of reactive case detection strategies for malaria elimination in a region of low transmission in southern Zambia. Malar J **2016**; 15(1):412.

23. Sturrock HJW, Roberts KW, Wegbreit J, Ohrt C, Gosling RD. Tackling Imported Malaria: An Elimination Endgame. Am J Trop Med Hyg **2015**; 93(1):139–144.

24. Ruktanonchai NW, DeLeenheer P, Tatem AJ, et al. Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data. PLoS Comput Biol **2016**; 12(4):e1004846.

25. Reiner RC, Menach A Le, Kunene S, et al. Mapping residual transmission for malaria elimination. Elife **2015**; 4:e09520.

26. Mi Choi K, Kyoung Choi Y, Kang YA, et al. Study of the Genetic Discrimination Between Imported and Autochthonous Cases of Malaria in South Korea. J Travel Med **2011**; 18(1):63–66.

27. Rice BL, Golden CD, Anjaranirina EJG, Botelho CM, Volkman SK, Hartl DL. Genetic evidence that the Makira region in northeastern Madagascar is a hotspot of malaria transmission. Malar J **2016**; 15(1):596.

28. Kent RJ, Thuma PE, Mharakurwa S, Norris DE. Seasonality, blood feeding behavior, and transmission of Plasmodium falciparum by Anopheles arabiensis after an extended drought in Southern Zambia. Am J Trop Med Hyg **2007**; 76(2):267-74.

29. Stevenson JC, Pinchoff J, Muleba M, et al. Spatio-temporal heterogeneity of malaria vectors in northern Zambia: implications for vector control. Parasit Vectors **2016**; 9(1):510.

30. Fornadel CM, Norris LC, Glass GE, Norris DE. Analysis of Anopheles arabiensis blood feeding behavior in southern Zambia during the two years

after introduction of insecticide-treated bed nets. Am J Trop Med Hyg **2010**; 83(4):848-53.

31. Laban NM, Kobayashi T, Hamapumbu H, et al. Comparison of a PfHRP2-based rapid diagnostic test and PCR for malaria in a low prevalence setting in rural southern Zambia: implications for elimination. Malar J **2015**; 14(1):25.

32. Pinchoff J, Hamapumbu H, Kobayashi T, et al. Factors Associated with Sustained Use of Long-Lasting Insecticide-Treated Nets Following a Reduction in Malaria Transmission in Southern Zambia. Am J Trop Med Hyg **2015**; 93(5):954–960.

33. Searle KM, Katowa B, Kobayashi T, et al. Distinct parasite populations infect individuals identified through passive and active case detection in a region of declining malaria transmission in southern Zambia. Malar J **2017**; 16(1):154.

34. Plowe C V, Djimde A, Bouare M, Doumbo O, Wellems TE. Pyrimethamine and proguanil resistance-conferring mutations in Plasmodium falciparum dihydrofolate reductase: Polymerase chain reaction methods for surveillance in Africa. Am J Trop Med Hyg **1995**; 52(6):565–568.

35. Hofmann N, Mwingira F, Shekalaghe S, Robinson LJ, Mueller I, Felger I. Ultra-Sensitive Detection of Plasmodium falciparum by Amplification of Multi-Copy Subtelomeric Targets. Seidlein L von, editor. PLOS Med **2015**; 12(3):e1001788.

36. Oyola SO, Ariani C V, Hamilton WL, et al. Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. Malar J **2016**; 15:597.

37. Sundararaman SA, Plenderleith LJ, Liu W, et al. Genomes of cryptic chimpanzee Plasmodium species reveal key evolutionary events leading to human malaria. Nat Commun **2016;** 7:11078.

38. Anderson TJ, Haubold B, Williams JT, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Mol Biol Evol **2000**; 17(10):1467–1482.

39. 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):113–25.

40. Jacquard A, Charlesworth B, Charlesworth D. The Genetic Structure of Populations. Berlin: Springer, **2012**.

41. Hijmans RJ. geosphere: Spherical Trigonometry. **2017**. Available at: https://cran.r-project.org/package=geosphere

42. Bivand RS, Pebesma E, Gómez-Rubio V. Applied Spatial Data Analysis with R. New York: Springer, **2013**.

43. Pebesma E, Bivand RS. Classes and Methods for Spatial Data: the sp Package. **2005**; Available at: https://cran.r-project.org/web/packages/sp/vignettes/intro_sp.pdf

44. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; **2018**. Available at: https://www.r-project.org/

45. ESRI. ArcGIS Desktop: Release 10.5.1. Redlands, CA: Environmental Systems Research Institute; **2017**.

46. Deutsch-Feldman M, Hamapumbu H, Lubinda J, et al. Efficiency of a Malaria Reactive Test-and-Treat Program in Southern Zambia: A Prospective, Observational Study. Am J Trop Med Hyg **2018**; 98(5):1382–1388.

47. Searle KM, Shields T, Hamapumbu H, et al. Efficiency of Household Reactive Case Detection for Malaria in Rural Southern Zambia: Simulations Based on Cross-Sectional Surveys from Two Epidemiological Settings. PLoS One **2013**; 8(8):e70972.

48. Searle KM, Pringle JC, Hamapumbu H, et al. Evaluating the Efficiency of Reactive Case Detection to Achieve Malaria Elimination in Rural Southern Zambia using Follow-up Household Visit and Parasite Genotyping [abstract 1326]. In: Program and abstracts of the Proc 66th Annu Meet Am Soc Trop Med Hygeine. Baltimore, MD: **2017**.

Chapter 3

Amplicon deep sequencing reveals a contiguous *Plasmodium falciparum* population along the international border between Zambia and the Democratic Republic of the Congo

ABSTRACT

The Zambian National Malaria Elimination Program (NMEP) aims to eliminate malaria by 2021, a challenging goal given the high transmission regions in the north of the country along the international border with the Democratic Republic of the Congo (DRC). Cross-border malaria transmission in this region may threaten the feasibility and sustainability of Zambia's elimination targets. In order to assess whether cross-border malaria transmission is on-going between Nchelenge District, Luapula Province in north Zambia and two villages, Kilwa and Kashobwe, across the border from Nchelenge District in Haut-Katanga Province, DRC, we compared *Plasmodium falciparum* populations on either side of the border to detect signatures of population genetic structure. Randomly selected households in Nchelenge, Kilwa, and Kashobwe were visited in June and July 2016 and consenting individuals provided a dried blood spot (DBS) for detection of *P. falciparum* by PCR and additional population genetic analysis. Among the DBS which tested positive for *P. falciparum* by PCR, 41 samples from Nchelenge and 39 from the DRC were extracted and prepared for amplicon deep sequencing at the *Pfama1* and *Pfcsp* loci. We compared within-country and between-country pairwise parasite relatedness, tested for genetic differentiation between the parasite populations in the two countries, and visualized population structure using discriminatory analysis of principal components and a haplotype network. We detected

no population structure or differentiation between *P. falciparum* isolates from Nchelenge District, Zambia and Haut-Katanga Province, DRC, supporting the idea that cross-border transmission is on-going in this region. Additional studies aimed at clarifying the spatial resolution of *Pfama1* and *Pfcsp* to detect population structure are merited. Continuing to monitor this trend as DRC scales up vector control measures will provide valuable information regarding the effect of interventions on reducing cross-border transmission.

INTRODUCTION

Significant progress has been made in reducing global malaria transmission since the early 2000's [1,2]. In fact, it is estimated that the global malaria burden declined by 40% between 2000 and 2015 [1]. Consequently, 35 countries established malaria elimination targets and programs as of September 2015 [3]. With a growing number of countries pursuing malaria elimination goals, it is critical to address the threat of cross-border malaria transmission, which complicates national malaria control efforts and threatens the sustainability of successful elimination campaigns.

The World Health Organization (WHO) defines malaria elimination as the reduction to zero of all indigenous malaria cases [4]. Therefore, documenting successful elimination necessitates distinguishing between locally transmitted and imported cases. Once elimination is achieved, continuing to monitor and avert cross-border and imported malaria is critical to preventing resurgence. Imported malaria contributed to the failure of the Global Malaria Eradication Program (GMEP) of the 1950's and 1960's [5] and continues to contribute to the burden of malaria in Venezuela [6], Greece [7], and

Zanzibar [8]. Similarly, sugar-cane workers traveling from Mozambique into Swaziland led to malaria resurgence in the 1970's [9], and human population movement continues to threaten progress towards malaria elimination in Swaziland today [10]. Sri Lanka, a country that successfully eliminated malaria in 2012, is similarly threatened with malaria resurgence through regular importation events [11].

Controlling cross-border malaria transmission is particularly relevant to regions located along international borders. Since malaria transmission in border regions is influenced by the policies of two or more countries, achieving malaria elimination in this context can be challenging in the absence of formal international partnerships and programs [12,13]. Further, international borders are especially vulnerable to reintroduction of parasites due to regular influxes of migrants and travelers in these settings [12,13].

Assessing the risk that imported malaria poses to malaria-eliminating countries typically relies on estimating human population movement [14], often accomplished through travel surveys among passively detected malaria cases [5]. Determining whether cases arose through importation based on recent travel history may be reliable in settings where malaria has been eliminated or when transmission is very low. However, travel histories may not be a reliable tool to measure cross-border malaria in higher transmission settings, or along international borders where risk is similarly high on both sides. Particularly in these settings, more nuanced methods to track the routes of parasite dissemination across international borders and identify reservoirs of ongoing cross-border malaria will be necessary to achieve national and regional elimination goals [5,15]. Parasite genotyping may allow more refined discrimination of parasite locality of

origin and enables gene flow between populations to be estimated, although the precise spatial resolution at which structure may be detected remains to be established. In fact, microsatellite genotyping at eleven loci in the *P. vivax* genome from isolates collected longitudinally in internally displaced person (IDP) settlements and surrounding villages along the Myanmar-China international border revealed bidirectional and frequent malaria transmission between IDP settlements and the community [15]. Similarly, microsatellite genotyping revealed parasite gene-flow between Yemen and Saudi Arabia, highlighting the importance of addressing cross-border transmission during malaria elimination in the Arabian Peninsula [16].

Zambia, a malaria-endemic country in southern Africa, aims to eliminate malaria by 2021. While southern Zambia has progressed substantially towards this goal, northern Zambia, particularly the regions bordering the Democratic Republic of the Congo (DRC), continues to experience holoendemic transmission. In fact, Luapula Province, located in northern Zambia, experiences the highest malaria prevalence among children under five years in Zambia at 56% by rapid diagnostic test (RDT) [17]. In Nchelenge District, located in Luapula Province directly along the international border with the DRC, malaria control interventions including long-lasting insecticide treated bednets (LLINs) and indoor residual spraying (IRS) have been used for over a decade, with only modest impact [18].

Two villages, Kilwa and Kashobwe, are located in the DRC directly across the international border from Nchelenge, Zambia. A cross-sectional survey conducted in Kilwa and Kashobwe during July of 2016 revealed that only 2.5% of individuals reported having received IRS in their home [19]. Given the unequal scale-up of vector

control on either side of the border, it is plausible that cross-border movement of people and vectors between Zambia and DRC contributes to the failure of malaria control interventions in Nchelenge District. Comparative analyses of the *Plasmodium falciparum* parasite populations on either side of the border will enable the detection of signatures of population genetic structure or lack thereof, corollaries of cross-border transmission.

METHODS

Sample Collection

A detailed description of sample collection, DNA extraction, and *P. falciparum* quantification has been previously described [20]. Briefly, dried blood spot (DBS) samples were collected (Whatman 903™ Protein Saver Card) from participants living in randomly selected households in Nchelenge, Zambia, and Kilwa and Kashobwe, DRC between June and July 2016. DBS samples were transported to the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, USA. DNA was extracted from the DBS samples using 20% Chelex and the concentration of *P. falciparum* specific DNA was quantified using qPCR for the *Pfldh* gene [21].

Amplicon Library Preparation

The 103 *P. falciparum* positive DNA extracts were amplified at the *Pfama1* and *Pfcsp* loci. *Pfcsp* amplicon generation was previously described [20]. Amplicons of *Pfama1* were similarly generated. In brief, primers (forward primer: CCAACAAAACCTCTTATGTCACCA; reverse primer: TTAGGTTGATCCGAAGCACTCA) were fused with Illumina adapter sequences and were used to amplify a 454 bp region of *Pfama1* [22]. *Pfama1* amplification reaction

components included 9.25 µL of the extracted template DNA, 12.5 µL KAPA Hifi HotStart ReadyMix (Kapa Biosystems, Wilmington, Massachusetts), 0.25 µL of the 20 µM reverse primer, 1.0 µL of the 5 µM forward primer (forward and reverse primers contained Illumina adapter sequences), and 2 µL of 25 µM magnesium chloride. Amplicon generation cycling conditions were: 95 °C for 5 minutes, then 32 cycles of 98 °C for 30 seconds, 58 °C for 1 minute, and 72 °C for 1 minute, followed by 72 °C for 5 minutes, and 4 °C thereafter. Amplicon sizes were confirmed using TapeStation (Agilent 4200, Santa Clara) and purified using AMPure beads (Beckman Coulter, Brea, California). Unique Nextera (Illumina, San Diego, California) indexes were added to amplicons in a subsequent PCR reaction described by Illumina [22]. Indexed amplicons underwent an additional TapeStation verification and AMPure bead purification. Confirmed amplicons were quantified using PicoGreen (ThermoFisher Scientific, Waltham, Massachusetts), diluted, and combined in equal concentrations into a single pool for 300 bp paired-end sequencing on the Illumina MiSeq platform at the Johns Hopkins School of Medicine Sequencing and Synthesis Facility.

Bioinformatic Processing

Forward and reverse read pairs were stitched using FLASH [23], sorted by primer sequence, and trimmed for quality (sliding window = 50 bp, step size = 5 bp, quality threshold = 20). Identical reads within individuals were clustered and consensus parasite haplotypes generated using SeekDeep [24]. Additional details regarding the bioinformatic pipeline are provided in Pringle et al. [20].

Within vs. Between Population Genetic Distance

Fasta files containing consensus sequences for each parasite clone were aligned using Geneious (version 9.1.5) (https://www.geneious.com). Software DnaSP (version 6.10.01) was used to calculate the number of unique *Pfama1* and *Pfcsp* haplotypes in each country, the nucleotide diversity (π) within each country and among all samples, and the average number of nucleotide differences between samples from within and between countries. Geneious was used to calculate pairwise genetic distance between parasite pairs as percent nucleotide identity. Pairwise comparisons were then grouped into three categories depending on whether they occurred between two parasites isolated from individuals both living in the DRC, parasites within two individuals from Zambia, or between parasites within individuals living in different countries. We compared the mean within-country percent nucleotide identity between parasite pairs to the mean between-country relatedness between parasite pairs.

Population Structure and Differentiation

To detect signatures of population structure between parasites from Zambia and the DRC, R package, adegenet [25] was used to conduct a discriminant analysis of principal components (DAPC) [26]. DAPC seeks to detect the linear axes which explain the most between-group variability in the data, unlike classical principal components analysis (PCA) which identifies the linear axes that explain the most variability overall [26]. We performed DAPC using sampling locations as group priors. Samples from Nchelenge District were classified as "inland" or "lakeside" samples, denoting the two ecologically distinct sub-regions of this study area. DRC samples were classified by their village of origin, Kilwa or Kashobwe. In addition to DAPC analysis, Templeton, Crandall, and Sing (TCS) haplotype networks [27] were constructed for each amplicon and visualized using

tcsBU [28]. We used the software package, STRUCTURE (version 2.3.4) [29] without reference sequences, to estimate the number of populations from which our samples originated and then assign each parasite to their respective population of origin. Finally, software DnaSP was used to test for population differentiation between *P. falciparum* isolates from Zambia and the DRC at the *Pfama1* and *Pfcsp* loci.

RESULTS

Amplicon Generation and Sequencing

One hundred and three DBS were amplified at the *Pfama1* and *Pfcsp* loci. Of these 103 samples, *Pfama1* successfully amplified across 74 samples (71.8%). The median parasite density of samples with successful *Pfama1* amplicon generation was 2,814 parasites/μL (p/μL). Among samples that failed to amplify at the *Pfama1* locus, median parasitemia was 6 p/μL, suggesting that low parasitemia was the cause of amplicon generation failure. Of the 74 unique samples with successful *Pfama1* amplicon generation, 70 had sequencing reads that passed all quality filters described in the methods section. Among these 70 individuals, 248 parasite clones were identified. Of the 248 *Pfama1* clones, 134 were from DRC (Kilwa= 49; Kashobwe= 85), and 114 were from Zambia (Nchelenge Inland= 38; Nchelenge Lakeside= 78).

Amplicon generation and sequencing results for *Pfcsp* are described in detail in Pringle et al. [20]. Briefly, 77 of the 103 samples generated successful *Pfcsp* amplicons, and 53 of these 77 resulted in sequencing reads passing quality control filtering. In total, 193 *P. falciparum* clones were characterized at the *Pfcsp* locus. Of the 193 clones, 86 were

from DRC (Kilwa= 24; Kashobwe= 62), and 108 were from Zambia (Nchelenge Inland= 30, Nchelenge Lakeside= 78).

Data Availability Statement

The unique *Pfcsp* DNA sequences generated in this study have been deposited in GenBank (accession numbers MG715504-MG715555).

Within vs. Between Population Genetic Distance

There was no significant difference in the mean pairwise percent nucleotide identity comparing parasites from within the same country to parasites from different countries for either *Pfama1* or *Pfcsp* (Figure 1). Similarly, the mean number of nucleotide differences between samples from within the same country was similar to the mean number of nucleotide differences among all samples (Table 1). These observations suggest that the parasites from Zambia and DRC originate from a single population.

Population Structure and Differentiation

DAPC analysis did not detect population structure between parasites from Zambia and the DRC at either the *Pfama1* or *Pfcsp* loci (Figure 2). Interestingly, the DAPC analysis suggests that some population differentiation may exist between the inland and lakeside regions within Nchelenge District. For both *Pfama1* and *Pfcsp*, STRUCTURE estimated the number of populations from which samples were derived to be one, and that all samples came from the same contiguous population. Further, TCS haplotype networks for both *Pfama1* and *Pfcsp* revealed that the majority of haplotypes were shared between both countries in similar proportions (Figure 3). Finally, DnaSP detected no statistically

significant population differentiation between parasites from Zambia and DRC at either the *Pfama1* (p=0.053) or *Pfcsp* (p=0.67) loci (Table 1).

DISCUSSION

Amplicon deep sequencing at two highly variable *P. falciparum* loci, *Pfama1* and *Pfcsp,* indicates that parasites from Nchelenge District, Zambia and Haut-Katanga Province, the DRC belong to a single genetic population, supporting the hypothesis that cross-border malaria is on-going in this region. Following sample collection for this study, a mass bednet distribution campaign was carried out in Kilwa and Kashobwe during October, 2016 [19]. Conducting a second cross-sectional population genetic analysis of parasites from Nchelenge District and Haut-Katanga Province would provide insight into whether the recent scale-up of vector control in the DRC impacted parasite genetic diversity and multiplicity of infection, indicators of transmission intensity. Similarly, additional samples from within Nchelenge District would be necessary to further investigate whether parasite population structure exists at a fine geographic scale and whether it is driven by ecological differences between the inland and lakeside regions.

While clear signatures of population differentiation are easily interpretable, it is more challenging to attribute cause for observations with no discernable population structure. A lack of population structure could either reflect the true underlying biology of an admixed population or may be an artifact of genetic markers incapable of detecting structure on a small spatial scale in high transmission zones [30,31]. The same *Pfcsp* amplicon that was examined in this study revealed parasite population genetic structure

on a continental scale [20]. Additional research is merited to assess the utility of our *Pfama1* and *Pfcsp* amplicons to detect population structure on smaller geographic scales. While our data suggest *P. falciparum* parasites from Nchelenge, Zambia and Kilwa and Kashobwe, DRC exist as a single panmictic population, it is possible that increasing either the number of SNPs characterized (*Pfama1* 38 SNPs, *Pfcsp* 25 SNPs) or the number of isolates sequenced would reveal finer scale population structuring.

It is typical to analyze neutral, unlinked SNPs in population genomic analysis. Here, we used amplicon deep sequencing to characterize two highly variable *P. falciparum* genes, *Pfama1* (38 SNPs) and *Pfcsp* (25 SNPs) which are both thought to be under balancing selective pressure [32]. The SNPs within each of these two amplicons are linked due to their close proximity in the *P. falciparum* genome. It is possible that the use of non-neutral, linked loci has biased our analysis such that we have failed to detect true population differentiation between Nchelenge District, Zambia and Haut-Katanga Province, DRC. However, in regions where the majority of infections are comprised of multiple, genetically distinct parasite clones, amplicon deep sequencing is the only method capable of preserving parasite haplotypes. Prior genetic studies in Nchelenge District demonstrated that 78% of samples from this area comprised polyclonal infections [33]. Amplicon deep sequencing is required to avoid invoking biased haplotype reconstruction methods or discarding polyclonal infections prior to analysis. As methods for handling polyclonal genetic data continue to improve, it may be possible to select unlinked, neutral loci for additional analyses assessing the contribution of cross-border transmission to the malaria burden in this region.

Despite the use of non-neutral, linked SNPs, our data suggesting a contiguous *P. falciparum* population are consistent with whole genome sequencing analyses from *Anopheles funestus* mosquitoes that did not detect population structure of vectors between Nchelenge, Zambia and Haut-Katanga Province, DRC [33, Lee, unpublished]. These data supporting one vector population suggest a possible mechanism that might drive the regular genetic crossing and lack of population differentiation among *P. falciparum* isolates from either side of the border.

Cross-border malaria transmission along the international border between Zambia and the DRC highlights the importance of expanding existing regional partnerships [5] like Elimination 8 (E8) in southern Africa, or the Asia Pacific Elimination Network in Asia. While the DRC remains unaffiliated with the E8, cross-border malaria transmission may likely threaten the sustainability of elimination efforts in the countries it borders, including Zambia, which aims to eliminate malaria by 2021. A study which looked at how frequently the Global Fund funded malaria projects aiming to establish multi-national control efforts [13] found that these proposals are rarely funded, and that there is little guidance for what makes these projects successful. Developing new strategies to guide, fund, and support regional initiatives that encourage international cooperation towards malaria elimination may enhance current and future efforts. As efforts to eliminate malaria across the globe continue to expand, addressing the threat posed by cross-border transmission is essential. Parasite population genetic analysis may be of value to malaria control programs in assessing the risk of cross-border transmission in a variety of settings.

Table 1:

	π: *ama1*	**π: *csp***	**# *ama1* haplotypes**	**# *csp* haplotypes**	**Average # nucleotide differences *ama1***	**Average # nucleotide differences *csp***	**Genetic differentiation *ama1***	**Genetic differentiation *csp***
Within Country:								
Zambia	0.023	0.026	64	42	10.32	6.5		
DRC	0.024	0.025	55	37	11.04	6.2		
Overall	0.024	0.025	92	52	10.76	6.3	p=0.053	p=0.67

Table 1: DnaSP (version 6.10.01) was used to calculate nucleotide diversity (π), the number of unique DNA haplotypes, the average number of nucleotide differences between *P. falciparum* clones, and population differentiation between parasites from Zambia and DRC.

Figure 1:

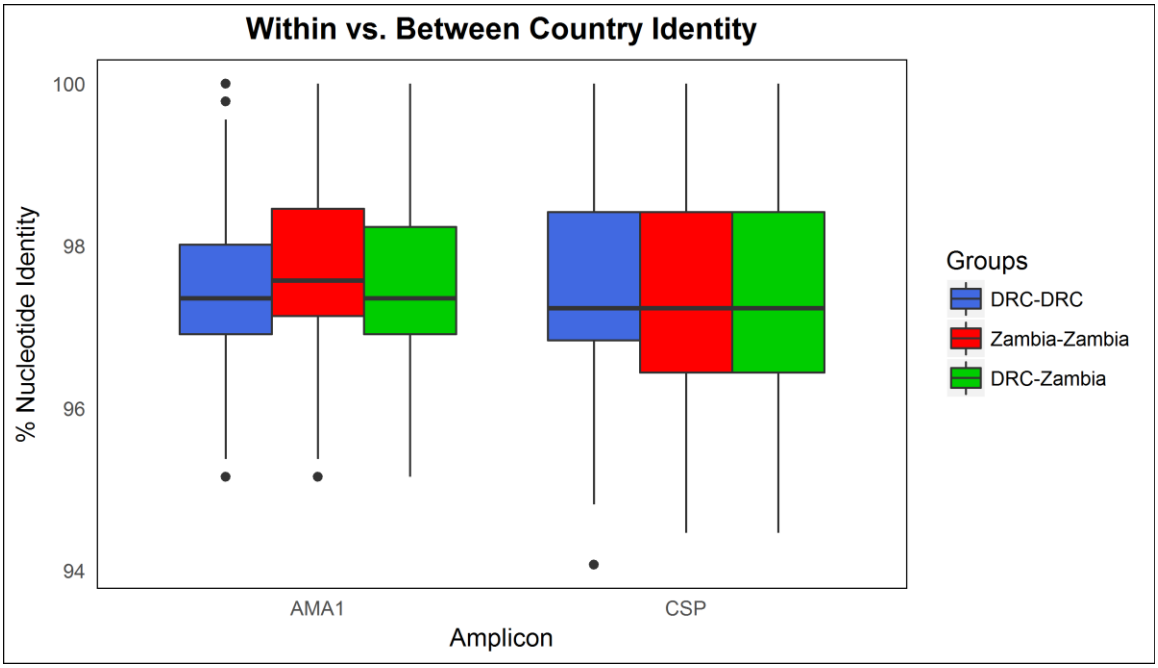

Figure 1: Pairwise percent nucleotide identity between all pairs of samples is plotted for comparisons between parasite isolates from DRC (blue), comparisons between parasite isolates from Zambia (red), and comparisons between parasite isolates from different countries (green).

Figure 2:

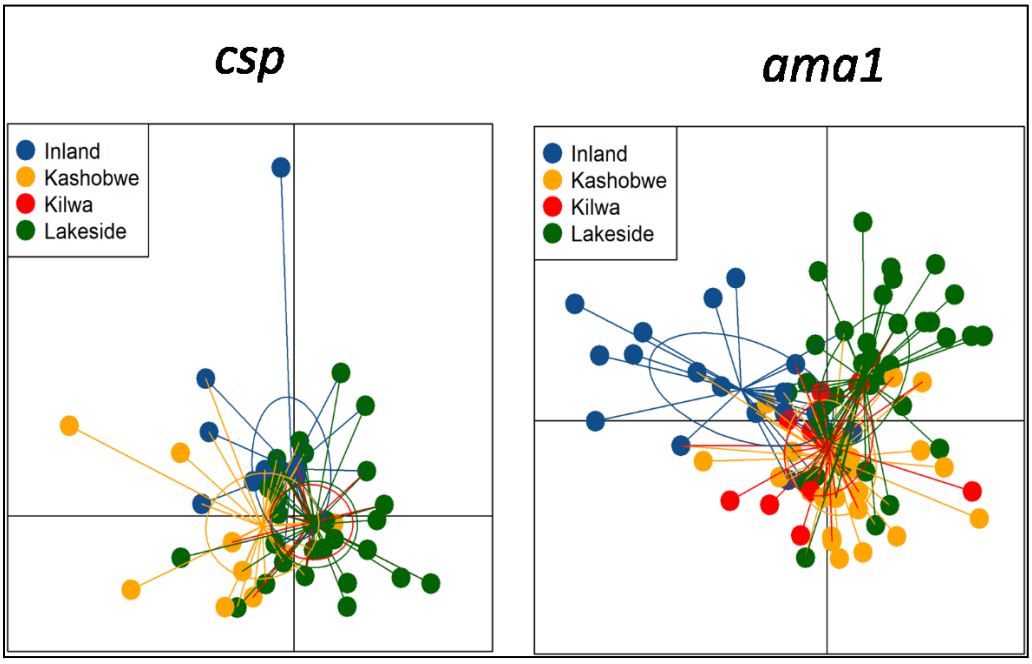

Figure 2: Discriminatory analysis of principal components (DAPC) seeks to maximize between-group genetic distance to detect signatures of population structure. Samples were grouped according to the village in which they were collected (inland and lakeside samples come from two ecologically distinct regions of Nchelenge District, Zambia; Kilwa and Kashobwe are located in the DRC).

Figure 3:

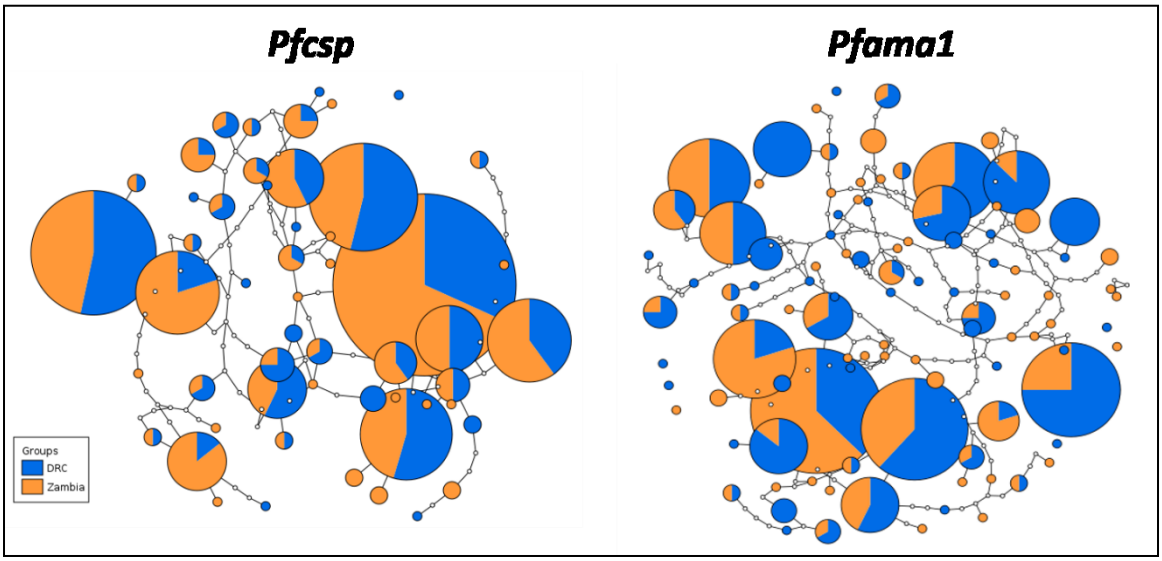

Figure 3: TCS haplotype networks depict each unique DNA haplotype as a circle. Circles are sized based on the number of samples with that haplotype. Haplotypes are connected based on genetic distance (number of nucleotide differences) between sequences. Colors denote whether the sequences came from individuals living in Zambia (orange) or the DRC (blue).

REFERENCES

1. Bhatt S, Weiss DJ, Cameron E, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. **2015**; 526(7572):207-211.

2. Cibulskis RE, Alonso P, Aponte J, et al. Malaria: Global progress 2000 - 2015 and future challenges. Infect Dis poverty. **2016**; 5(1):61.

3. Newby G, Bennett A, Larson E, et al. The path to eradication: a progress report on the malaria-eliminating countries. Lancet. **2016**; 387(10029):1775-84.

4. World Health Organization. Overview of malaria elimination. **2018**. Available from: http://www.who.int/malaria/areas/elimination/overview/en/

5. Sturrock HJW, Roberts KW, Wegbreit J, Ohrt C, Gosling RD. Tackling Imported Malaria: An Elimination Endgame. Am J Trop Med Hyg. **2015**; 93(1):139–144.

6. Rodríguez-Morales AJ, Delgado L, Martínez N, Franco-Paredes C. Impact of Imported Malaria on the Burden of Disease in Northeastern Venezuela. J Travel Med. **2006**; 13(1):15–20.

7. Danis K, Baka A, Lenglet A, et al. Autochthonous Plasmodium vivax malaria in Greece, 2011. Eurosurveillance. **2011**; 16(42):19993.

8. Menach A Le, Tatem AJ, Cohen JM, et al. Travel risk, malaria importation and malaria transmission in Zanzibar. Sci Rep. **2011**; 1:93.

9. Packard RM. Agricultural development, migrant labor and the resurgence of malaria in Swaziland. Soc Sci Med. **1986**; 22(8):861–7.

10. Koita K, Novotny J, Kunene S, et al. Targeting imported malaria through social networks: a potential strategy for malaria elimination in Swaziland. Malar J. **2013**; 12(1):219.

11. Wickramage K, Premaratne RG, Peiris SL, Mosca D. High attack rate for malaria through irregular migration routes to a country on verge of elimination. Malar J. **2013**; 12:276.

12. Wangdi K, Gatton ML, Kelly GC, Clements AC. Cross-Border Malaria: A Major Obstacle for Malaria Elimination. Adv Parasitol. **2015**; 79–107.

13. Gueye C, Teng A, Kinyua K, Wafula F, Gosling R, McCoy D. Parasites and vectors carry no passport: how to fund cross-border and regional efforts to achieve malaria elimination. Malar J.; **2012**; 11(1):344.

14. Pindolia DK, Garcia AJ, Wesolowski A, et al. Human movement data for malaria control and elimination strategic planning. Malar J.; **2012**; 11(1):205.

15. Lo E, Lam N, Hemming-Schroeder E, et al. Frequent Spread of Plasmodium vivax Malaria Maintains High Genetic Diversity at the Myanmar-China Border, Without Distance and Landscape Barriers. J Infect Dis. **2017**; 216(10):1254-1263.

16. Al-Hamidhi S, Mahdy MAK, Idris MA, et al. The prospect of malaria elimination in the Arabian Peninsula: A population genetic approach. Infect Genet Evol. **2014**; 27:25–31.

17. Moss WJ, Dorsey G, Mueller I, et al. Malaria Epidemiology and Control within

the International Centers of Excellence for Malaria Research. Am J Trop Med Hyg. **2015**; 93:5–15.

18. Mukonka VM, Chanda E, Haque U, et al. High burden of malaria following scale-up of control interventions in Nchelenge District, Luapula Province, Zambia. Malar J. **2014**; 13(1):153.

19. Kobayashi T, Bobanga T, Umesumbu S, et al. The impact of mass bed net distribution on vector species and malaria prevalence in Kilwa and Kashobwe, Haut-Katanga Province, Democratic Republic of the Congo. Proc 67th Annu Am Soc Trop Med Hygeine. New Orleans; **2018**.

20. Pringle JC, Carpi G, Almagro-Garcia J, et al. RTS,S/AS01 malaria vaccine mismatch observed among Plasmodium falciparum isolates from southern and central Africa and globally. Sci Rep.; **2018**; 8(1):6622.

21. Laban NM, Kobayashi T, Hamapumbu H, et al. Comparison of a PfHRP2-based rapid diagnostic test and PCR for malaria in a low prevalence setting in rural southern Zambia: implications for elimination. Malar J.; **2015**; 14(1):25.

22. Illumina. 16S Metagenomic Sequencing Library Preparation. Illumina.com. **2013**; 1–28.

23. Mag T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. **2011**; 27(21):2957–296310.

24. Hathaway NJ, Parobek CM, Juliano JJ, Bailey JA. SeekDeep: single-base

resolution de novo clustering for amplicon deep sequencing. Nucleic Acids Res. **2018**; 46(4):e21

25. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. **2008**; 24(11):1403–1405.

26. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. **2010**; 11(1):94.

27. Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. Mol Ecol. **2000**; 9(10):1657–1659.

28. Múrias dos Santos A, Cabezas MP, Tavares AI, Xavier R, Branco M. tcsBU: a tool to extend TCS network layout and visualization. Bioinformatics. **2016**; 32(4):627–628.

29. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. Genetics. **2000**; 155(2):945-59.

30. Arnaud-Haond S, Alberto F, Teixeira S, Procaccini G, Serr EA, Duarte CM. Assessing Genetic Diversity in Clonal Organisms: Low Diversity or Low Resolution? Combining Power and Cost Efficiency in Selecting Markers. J Hered. **2005**; 96(4):434-40.

31. Reusch TB. New markers- old questions: population genetics of seagrasses. Mar Ecol Prog Ser. **2001**; 211:261–274.

32. Miller RH, Hathaway NJ, Kharabora O, et al. A deep sequencing approach to estimate Plasmodium falciparum complexity of infection (COI) and explore apical membrane antigen 1 diversity. Malar J. BioMed Central; **2017**; 16(1):490.

33. Pringle JC, Kobayashi T, Carpi G, Doctor SM, Juliano J, Meshnick S, Mulenga M, Chaponda M, Muleba M, Bobanga T, Norris DE, Moss WJ. Parasite genetics approach to evaluate malaria transmission dynamics in Zambia. Am Soc Trop Med Hyg 65th Annu Meet. Atlanta, GA; **2016**.

34. Jones CM, Lee Y, Kitchen A, et al. Complete Anopheles funestus mitogenomes reveal an ancient history of mitochondrial lineages and their distribution in southern and central Africa. Sci Rep. Nature Publishing Group; **2018**; 8(1):9054.

Chapter 4

RTS,S/AS01 malaria vaccine mismatch observed among *Plasmodium falciparum* isolates from southern and central Africa and globally

ABSTRACT

The RTS,S/AS01 malaria vaccine encompasses the central repeats and C-terminal of *Plasmodium falciparum* circumsporozoite protein (PfCSP). Although no Phase II clinical trial studies observed evidence of strain-specific immunity, recent studies show a decrease in vaccine efficacy against non-vaccine strain parasites. In light of goals to reduce malaria morbidity, anticipating the effectiveness of RTS,S/AS01 is critical to planning widespread vaccine introduction. We deep sequenced C-terminal *Pfcsp* from 77 individuals living along the international border in Luapula Province, Zambia and Haut-Katanga Province, the Democratic Republic of the Congo (DRC) and compared translated amino acid haplotypes to the 3D7 vaccine strain. Only 5.2% of the 193 PfCSP sequences from the Zambia-DRC border region matched 3D7 at all 84 amino acids. To further contextualize the genetic diversity sampled in this study with global PfCSP diversity, we analyzed an additional 3,809 *Pfcsp* sequences from the Pf3k database and constructed a haplotype network representing 15 countries from Africa and Asia. The diversity observed in our samples was similar to the diversity observed in the global haplotype network. These observations underscore the need for additional research assessing genetic diversity in *P. falciparum* and the impact of PfCSP diversity on RTS,S/AS01 efficacy.

INTRODUCTION

Although indoor residual spraying (IRS) and insecticide treated bednets (ITNs) have dramatically decreased malaria transmission, the global impact of malaria remains high with an estimated 216 million cases reported in 2016 [1,2]. Sub-Saharan Africa experiences a disproportionately high burden of *Plasmodium falciparum* malaria, even in regions with high coverage of IRS and ITNs [1,3]. Recent World Health Organization (WHO) goals aim to reduce both malaria mortality and case incidence by 90% of 2015 levels by 2030 [1]. Given the inadequacy of IRS and ITNs to eliminate malaria in all transmission settings, additional tools are necessary [3]. Of particular interest is an effective vaccine which might enhance control efforts and reduce malaria associated morbidity and mortality, particularly in regions refractory to current interventions.

Circumsporozoite protein (CSP), the dominant surface protein coating infectious stage sporozoites, has been a focus of vaccine development since the observation that bites from irradiated, infectious mosquitoes induce protective immune responses [4] *P. falciparum* CSP (PfCSP) has three distinct regions: the conserved amino (N)-terminal region, the central repeat region (CRR) comprised of 37–42 NANP repeats, and a polymorphic carboxyl (C) -terminal containing two sub-regions of high diversity known as Th2R and Th3R that elicit T-cell responses [5,6]. The CRR contains the most immunogenic B-cell sporozoite epitopes and anti-CSP antibodies induced by exposure to irradiated, live sporozoites prevent infection by binding at the CRR in animal models [7].

Malaria vaccine development efforts have spanned multiple decades and recently culminated in the licensure of the RTS,S/AS01 vaccine by GlaxoSmithKline (GSK) in 2015. The RTS,S/AS01 vaccine is a recombinant protein vaccine containing a portion of the NANP repeats (B-cell epitopes) and the C-terminal region (B-cell and T-cell epitopes) of the PfCSP fused with hepatitis B surface antigen (HBsAg) and is administered with a novel adjuvant, AS01 [8,9]. The vaccine construct is based on the *P. falciparum* 3D7 clone, which was derived from the NF54 strain isolated from a patient living near Schipol Airport in Amsterdam [8,10]. Phase III clinical trials carried out at 11 sites across seven countries in sub-Saharan Africa demonstrated an overall vaccine efficacy (estimated using negative binomial regression) against clinical malaria from month zero to study end (children: median 48 months until study end, infants: median 38 months until study end) of 36.3% in children aged 5–17 months who received 3 primary doses of RTS,S plus a booster at 20 months [11]. In 2015, the European Medicines Agent (EMA) approved the use of RTS,S/AS01 [12]. The Malaria Vaccine Implementation Programme (MVIP) led by WHO will begin RTS,S/AS01 implementation in three high transmission regions of Ghana, Kenya, and Malawi in 2018 with the goals of continued evaluation of the vaccine's impact on mortality, evaluating the feasibility of deploying the four dose vaccine series, and continued monitoring of vaccine safety [12].

Because the gene encoding *P. falciparum* CSP (*Pfcsp*) is globally diverse [13-15], multiple studies were conducted during the RTS,S/AS01 Phase II clinical trials to monitor *Pfcsp* haplotypes from vaccine and placebo recipients for signals of allele-specific vaccine-induced immunity. Four studies conducted in The Gambia, Kenya, and Mozambique found no evidence of allele-specific vaccine-induced immunity [16-19].

These genetic surveillance analyses relied either on Sanger sequencing [16,18] or oligonucleotide hybridization assays to assign genotypes to *P. falciparum* isolates [17,19]. While state of the art assays at the time, the advent of affordable and scalable next generation sequencing technologies with the capacity to rapidly analyze larger sample sets has rendered both methods outdated.

Following the Phase III clinical trials, researchers used Illumina MiSeq and PacBio next generation sequencing technologies respectively to sequence both the C-terminal and CRR regions of *Pfcsp* from parasites collected from individuals vaccinated with RTS,S/AS01 or placebo at 11 Phase III trial sites [20]. Cumulative vaccine efficacy was reduced from 50.3% for parasites with a perfect *Pfcsp* C-terminal sequence match to only 33.4% for parasites with any amino acid mismatch in this region [20]. Although previous studies did not provide evidence to support the allele dependent nature of RTS,S/AS01 vaccine efficacy [16-19], this recent analysis using technologically advanced methods and a larger sample size suggests allele-specific immunity is important in eliciting protection [20]. Further, this observation offers a potential explanation into the wide range of RTS,S/AS01 efficacies observed during Phase III clinical trials across the 11 sites (range: 22.0–74.6% against clinical malaria from month zero until the end of follow-up among children receiving three primary doses of RTS,S/AS01 plus a booster at 20 months; vaccine efficacy estimated through negative binomial regression) [11].

In Zambia, malaria risk is heterogeneous with regions targeted for malaria elimination in the south and districts in which prevalence is greater than 50% throughout the year in the north [21]. Nchelenge District is located in northern Zambia in Luapula Province, the province with the highest malaria prevalence in children younger than five

years of age (>50% by malaria rapid diagnostic test in 2014) [21]. Despite a decade of malaria control interventions in Nchelenge District, including implementation of ITNs and IRS, malaria transmission remains holoendemic, with prevalence greater than 50% through 2017 (unpublished data) [3,22]. While Zambia scales up to meet its 2021 malaria elimination goal, it is important to consider the utility of introducing RTS,S/AS01 into the current arsenal of tools to reduce malaria morbidity and mortality in this region. Towards this goal, the genetic diversity of the C-terminal *Pfcsp* was characterized with respect to the vaccine strain in parasites collected from the border of northern Zambia and the Democratic Republic of the Congo (DRC). Further, the genetic diversity of the C-terminal *Pfcsp* sequences in this study was contextualized within global *Pfcsp* diversity from the Pf3k database.

METHODS

Sample Collection, DNA Extraction, and Quantification

Dried blood spot (DBS) samples were collected at one time point from consenting individuals living in randomly selected households during June and July 2016 in Nchelenge District, Zambia as well as two villages, Kilwa and Kashobwe, located directly across the Zambian border in the DRC. One hundred and three DBS samples from unique individuals (64 Zambian participants and 39 DRC participants) ranging in age from 8 months to 72 years (mean 17.7 years) and identified to be *P. falciparum*positive through qPCR screening of the DBS in Zambia were shipped to the

Johns Hopkins Bloomberg School of Public Health, extracted using 20% Chelex, and quantified using *P. falciparum lactose dehydrogenase* (*Pfldh*) qPCR [23].

Amplicon Generation and Sequencing

A 300-bp amplicon containing the C-terminal *Pfcsp* (839-1,139-bp, clone 3D7 0304600.1, PlasmoDB [24]) was amplified from *P. falciparum* positive samples using the forward primer GACAAGGTCACAATATGCCAAA and reverse primer ACATTAAACACACTGGAACATTTTTC fused with Illumina MiSeq adapter sequences for library indexing during PCR [25]. PCR amplification reaction components included: 10 μL DNA template, 12.5 μL KAPA Hifi HotStart ReadyMix (Kapa Biosystems, Wilmington, Massachusetts), 0.25 μL each of 20 μM forward and reverse primers containing Illumina adapters, and 2 μL of 25 μM magnesium chloride. PCR cycling conditions were 95 °C for 5 minutes followed by 30 cycles of 98 °C for 20 seconds, 61 °C for 30 seconds, 72 °C for 1 minute, 72 °C for 5 minutes, and a holding step at 4 °C. Amplicon size (300-bp) was verified using TapeStation (Agilent 4200, Santa Clara). Seventy-five percent of the 103 samples successfully generated 300-bp *Pfcsp*amplicons. Among samples with >50 p/μL by qPCR, the amplicon generation success rate was 100%; for samples with <50 p/μL by qPCR, the success rate was 7%. The arithmetic mean of parasite copy number for samples that failed to generate amplicons was 11.1 p/μL (range: 0.5 p/μL–44.5 p/μL) compared with arithmetic mean 8,666 p/μL (range: 4.0 p/μL–81,196.4 p/μL) for samples for which amplicon generation was successful.

Amplicons were uniquely barcoded in a subsequent PCR reaction containing Nextera (Illumina, San Diego, California) indexes as described by Illumina [25]. Indexed amplicon sizes were verified using TapeStation, purified using AMPure beads (Beckman Coulter, Brea, California), and quantified using PicoGreen (ThermoFisher Scientific, Waltham, Massachusetts). The *Pfcsp* amplicons were normalized and combined into a single pool for 300-bp paired end sequencing on a MiSeq at the Sequencing and Synthesis Facility at the Johns Hopkins School of Medicine.

Bioinformatic Processing and Analysis

Forward and reverse reads were merged using FLASH [26], trimmed for quality (sliding window = 50-bp, step size = 5-bp, quality threshold = 20) and collapsed by haplotype using SeekDeep's [27] default Illumina settings and allowing for one high quality mismatch within individuals. Samples included in the final analysis were supported by high read coverage, with an average of 29,439 reads. Haplotypes found to represent at least 1% of a sample were considered in the final analysis in order to minimize the inclusion of false positive haplotypes. The number of genetically distinct parasite haplotypes per individual, or complexity of infection (COI), was determined by the number of unique haplotype clusters per individual, as estimated by SeekDeep.

The *Pfcsp* sequences were translated to amino acid sequences and aligned to the 3D7 vaccine reference strain (3D7 0304600.1, PlasmoDB) [24] in Geneious (version 9.1.5). The number of amino acid differences was calculated for each sequence and the 3D7 reference sequence for the 84 amino acids in the C-terminal amplicon (amino acids 288–371) of PfCSP (total length = 397 amino acids).

Data Availability Statement

The unique DNA sequences obtained from this study have been deposited in GenBank (accession numbers MG715504-MG715555).

Pf3k Sequence Acquisition and Global Diversity Analysis

Global *Pfcsp* diversity was examined by mining the MalariaGEN Pf3k Project (release 5) [28] which includes 2,512 *P. falciparum* full genomes from 14 countries worldwide. We retrieved all genetic variants on chromosome 3 available in Pf3k in variant call format (VCF) from release 5.1. Variant calls were made using GATK best practice haplotypeCaller [29,30]. Variants in the C-terminal *Pfcsp* region (nucleotides 866–1,113) were extracted for the 1,147 monoclonal infections from Africa and Asia. The individual *Pfcsp* haplotype sequences for the 1,365 multi-clonal samples were reconstructed using DEploid [31] with appropriate reference panels of mono-clonal samples from the Pf3k dataset [29]. We constructed a network based on the method by Templeton, Crandall, and Sing (TCS) using PopArt [32] to assess genealogical relationships between the global *Pfcsp* haplotypes found in Pf3k and the haplotypes from Zambia and the DRC [33,34]. Genetic diversity metrics were calculated using DnaSP (version 6.10.01). We compared sequences between African and Asian countries in terms of diversity and for evidence of population differentiation by calculating F_{ST}. Similarly, we compared samples from east and west African countries and calculated F_{ST} for signatures of population structure. For the purposes of comparing east and west Africa, we grouped samples in the DRC with east African samples.

Ethics Approval and Informed Consent

This research was approved by the Institutional Review Board at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, USA, the Ethics Review Committee for the Tropical Diseases Research Centre in Ndola, Zambia, and by Le Comité d'éthique de l'Université Protestante au Congo in Kinshasa, the Democratic Republic of the Congo. All studies were conducted in accordance with the ethical guidelines set forth by the aforementioned review boards. All adults who participated in these studies gave informed consent. All child participants gave assent and had parental consent for participation.

RESULTS

Of the 103 DBS samples extracted from our collections in Zambia and the DRC, 77 yielded suitable *Pfcsp* amplicons for sequencing. Fifty five of the 77 samples sequenced passed quality filtering steps implemented by FLASH and SeekDeep. Two samples were excluded from the analysis for lacking full length *Pfcsp* sequences. Overall, 193 PfCSP haplotype sequences from 53 individuals were characterized, corresponding to 52 unique haplotypes (Table 1) ranging in population frequency from 1 to 22 (mean 3.7) observations across the 53 individuals. The 53 individuals were infected with a mean of 3.6 genetically distinct parasite haplotypes (range: 1–10). Of the 193 parasite sequences characterized from the 53 human samples, only ten matched the 3D7 haplotype at all 84 amino acids (5.2%). The median number of amino acid differences

across all 193 parasite haplotypes (Zambia: n=108, DRC: n= 86, Kilwa= 24, Kashobwe= 62, Nchelenge Inland= 30, Nchelenge Lakeside= 78) in comparison to 3D7 was seven. Of the 10 3D7-type parasites, eight were found in individuals from Zambia and two were found in individuals from the DRC. The proportion of individuals harboring a 3D7-type parasite were similar between countries (Zambia: 25.8%, DRC: 9.1%, Pearson's chi-squared test of equal proportions: $p = 0.24$) and between sampling sites (Kilwa: 0%, Kashobwe: 13.3%, Nchelenge lakeside: 31.8% Nchelenge inland: 11.1%, Pearson's chi-squared test of equal proportions: $p = 0.20$). The frequency of 3D7 (vaccine) matched parasites was lower than previously reported in other African countries [13,17]. Furthermore, all 10 3D7-type parasites were found in the context of polyclonal infections (range 3–7 haplotypes). The 3D7-type was not the major haplotype in nine of these ten (90%) polyclonal infections (median relative abundance of 3D7-type haplotypes in an individual = 5%; range = 1–24%). Divergent regions of the PfCSP amplicon from 3D7 were visualized by plotting the percentage of samples sharing the 3D7 reference amino acid at each of the 84 amino acid positions (Figure 1). Twenty-two polymorphic amino acid positions were identified, eight of which were within the Th2R region (amino acids 311–327) and six were contained in Th3R (amino acids 352–363).

The *Pfcsp* nucleotide diversity observed within Nchelenge District, Zambia, and Kilwa and Kashobwe in the DRC appeared to be representative of African *Pfcsp* in the TCS haplotype network constructed using samples from this study in addition to 3,809 sequences from the Pf3k database, as the proportion of sequences that matched 3D7 was similar comparing our study area (5.2%) to all of the African isolates (5.3%). In total, we

identified 393 unique *Pfcsp* haplotypes, of which seven account for 51.3% of all the 4,002 sequences analyzed in this study.

No clear population structure was identified between east and west African isolates ($F_{ST} = 0.008$), although signatures of moderate population differentiation were observed between African and Asian samples (Figure 2), with $F_{ST} = 0.163$. We observed higher nucleotide diversity among African isolates than in Asian isolates, but no difference between east and west African isolates (Table 2). Among all African isolates analyzed, the 3D7 haplotype represented only 5.3% of the African 2,635 sequences. Including both Asian and African isolates, the 3D7 haplotype represented only 3.6% of 4,002 sequences. Among the Asian isolates, only three sequences matched the 3D7 haplotype (0.2%), all reported from Bangladesh.

DISCUSSION

RTS,S vaccine efficacy has been shown to decline from 50.3% for vaccine matched parasite CSP haplotypes to 33.4% for unmatched haplotypes [20]. In our study sites, only 10 parasite sequences of the 193 recovered (5.2%) were an exact match to the amino acid sequence in the C-terminal CSP region of the vaccine strain. The proportion of individuals harboring 3D7-type parasites was similar between Zambia and the DRC. The frequency of vaccine matched parasites observed in this study is lower than previously reported in other African countries [13,17].

It was previously demonstrated that RTS,S vaccine efficacy declined substantially for parasites not matching 3D7 in the C-terminal region of *Pfcsp* [20], and that vaccine

efficacy declines as the number of amino acid differences increases [20]. Therefore, the implication of parasites along the Zambia-DRC border differing from 3D7 at a median of seven amino acids is potentially significant. All 10 3D7-type parasites identified in our study occurred in the context of polyclonal infections (range 3–7 haplotypes). While Neafsey *et al*. evaluated the proportion of infections containing a 3D7 matched parasite haplotype as a function of COI [20], how vaccine efficacy differs between monoclonal 3D7-type infections, infections where 3D7 is the major of multiple haplotypes, and infections where 3D7 is the minor of multiple haplotypes has not yet been studied. Additional studies aimed at clarifying the effect of polyclonal infections on vaccine efficacy are warranted.

RTS,S/AS01, the only currently licensed malaria vaccine, is based on the sequence of just one parasite clone, 3D7, of African origin [35]. Given that the vaccine is based on an African parasite clone, considering to what extent circulating parasites from Asia differ relative to those in Africa can provide us with insight into how well RTS,S/AS01 may perform if implemented in Asian countries. In this study, we observe moderate population differentiation between Asian and African isolates ($F_{ST} = 0.163$). Previous studies that aimed to assess global *Pfcsp* diversity identified population structure between isolates from Africa and Asia, consistent with our observations [14,15]. Barry *et al*. used *Pfcsp* sequences from GenBank (n = 604) and characterized global diversity in an approximately similar C-terminal *csp* region (nucleotides 909–1140) to this study (nucleotides 866–1113) [15]. Although the *Pfcsp* network created by Barry *et al*. included only five of the sixteen countries represented in this study, the overall pattern of high global *Pfcsp* diversity was consistent across the two studies, strengthening the

conclusions presented in both analyses. Notably, previous analyses of global *Pfcsp*diversity have not included isolates from multiple east African countries [14,15,36]. Here, we include 910 *Pfcsp* sequences from three east African countries, providing, to our knowledge, the first large scale characterization of *P. falciparum* genetic diversity in relation to the RTS,S/AS01 vaccine across multiple countries in this historically understudied area. Further, previous research has focused on describing *Pfcsp* diversity from monoclonal malaria infections [13] which may underestimate true population genetic diversity. This study characterizes *Pfcsp* haplotypes from multiple clones present in polyclonal infections, providing a more complete analysis of population diversity.

Ideally, an effective malaria vaccine would provide protection against the majority of circulating parasites across multiple geographic regions. Our data provide evidence that *Pfcsp* exhibits high genetic diversity both locally and globally. Interestingly, the prevalence of the 3D7-type parasite strains in our study area (5.2%) is the same as that across all of the African countries included in the Pf3k dataset (5.3%, n = 140/2635). Among Asian isolates, the 3D7 haplotype is even less frequently observed (n = 3/1367) at only 0.2% prevalence. In fact, Bangladesh is the only Asian country in the Pf3k dataset in which the 3D7 haplotype was observed. These data support previous observations that C-terminal *Pfcsp* is diverse globally, and that 3D7-type parasites are more frequently found in African countries than Asian counties [14,15]. The high degree of global genetic *Pfcsp* diversity may potentially reduce RTS,S/AS01 vaccine effectiveness, particularly in Asian countries where 3D7 was not or only rarely observed.

Monitoring differential vaccine efficacy by PfCSP haplotype during possible future RTS,S/AS01 implementation programs will be valuable.

Publically available sequence databases provide unparalleled opportunities to understand global pathogen population genetics. However, it is important to acknowledge the limitations of drawing inferences from non-randomly sampled sequences. Notably, countries, as well as regions within countries, have unequal rates of sample deposition into databases, leading to a geographically biased set of sequences which over-represent genotypes from a small number of geographic foci while under-representing large swaths of the globe. Further, the conclusions drawn from sequences obtained from any given sequence repository are subject to change as sample sizes and geographic distributions are continually updated and expanded. Finally, the sequences obtained from Pf3k represent a multitude of sampling strategies, time periods, and sequencing technologies, which prevent samples from various regions from being optimally comparable. These samples come from patients across a spectrum of ages rather than specifically from children who are the recipients of the RTS,S/AS01 vaccine. However, despite these inherent limitations, sequence databases are powerful tools capable of elucidating global patterns in pathogen population genetic diversity. These resources, coupled with functional laboratory studies as well as observational field research, have a critical role to play in vaccine development efforts against a repertoire of global pathogens, including malaria. In the context of this study and with these limitations, we recognize that we have likely underestimated the true global diversity of *Pfcsp*.

The data presented here highlight the diversity of C-terminal *Pfcsp*, particularly in sub-Saharan Africa which is a key target region for malaria vaccination programs. This

study underscores the importance of incorporating population genetic studies into future malaria vaccine design, laboratory and clinical evaluation. Most importantly, assessing the diversity of C-terminal *Pfcsp* should be a component of the RTS,S/AS01 Malaria Vaccine Implementation Programme and in further refining our understanding of how genetic diversity affects RTS,S/AS01 efficacy.

Table 1: Amino Acid Haplotype Frequencies

Rank order of unique haplotypes	Frequency	Number of AA Differences
1	22	8
2	15	5
3	13	8
4	11	7
5	10	0
6	10	6
7	8	7
8	7	7
9	7	8
10	7	8
11	5	7
12-15	4	5-7
16-21	3	4-8
22-33	2	2-9
34-52	1	5-10

Table 1: Summarizes the frequency distribution of the observed 52 unique C-terminal haplotypes from Zambia and DRC. Haplotypes are listed in descending order of frequency, with Rank 1 representing the most common haplotype. The number of amino acid (AA) differences was calculated against 3D7 reference 0304600.1 (PlasmoDB).

Table 2: *Pfcsp* Global Diversity Statistics

Continent	Region	Country	n	h	S	K	Hd	π
Africa			2,635	370	33	4.11	0.948	0.117
	East Africa		910	181	26	4.05	0.951	0.116
		DRC	268	64	22	4.01	0.950	0.114
		Malawi	534	150	25	3.99	0.951	0.114
		Zambia	108	31	19	4.45	0.933	0.127
	West Africa		1725	264	31	4.11	0.945	0.117
		The Gambia	105	23	20	4.19	0.905	0.120
		Ghana	1109	232	29	4.17	0.948	0.119
		Guinea	175	49	21	3.94	0.939	0.112
		Mali	166	53	24	3.84	0.935	0.110
		Nigeria	7	4	6	2.95	0.857	0.084
		Senegal	163	35	22	4.04	0.928	0.116
Asia			1,367	44	20	2.58	0.780	0.074
		Bangladesh	80	20	15	2.54	0.859	0.072
		Cambodia	752	23	17	2.66	0.807	0.076
		Laos	129	11	13	2.68	0.799	0.077
		Myanmar	81	6	9	0.65	0.349	0.019
		Thailand	197	14	12	2.09	0.589	0.060
		Vietnam	128	15	15	3.07	0.759	0.088

Table 2: Summarizes the samples included in the *Pfcsp* network analysis, including 193 samples sequenced in this study and 3,809 from Pf3k database. n= number of sequences, h= number of unique haplotypes, S= number of segregating sites (out of total possible 35), K= average number of pairwise nucleotide differences, Hd= haplotype diversity, π= nucleotide diversity.

Figure 1

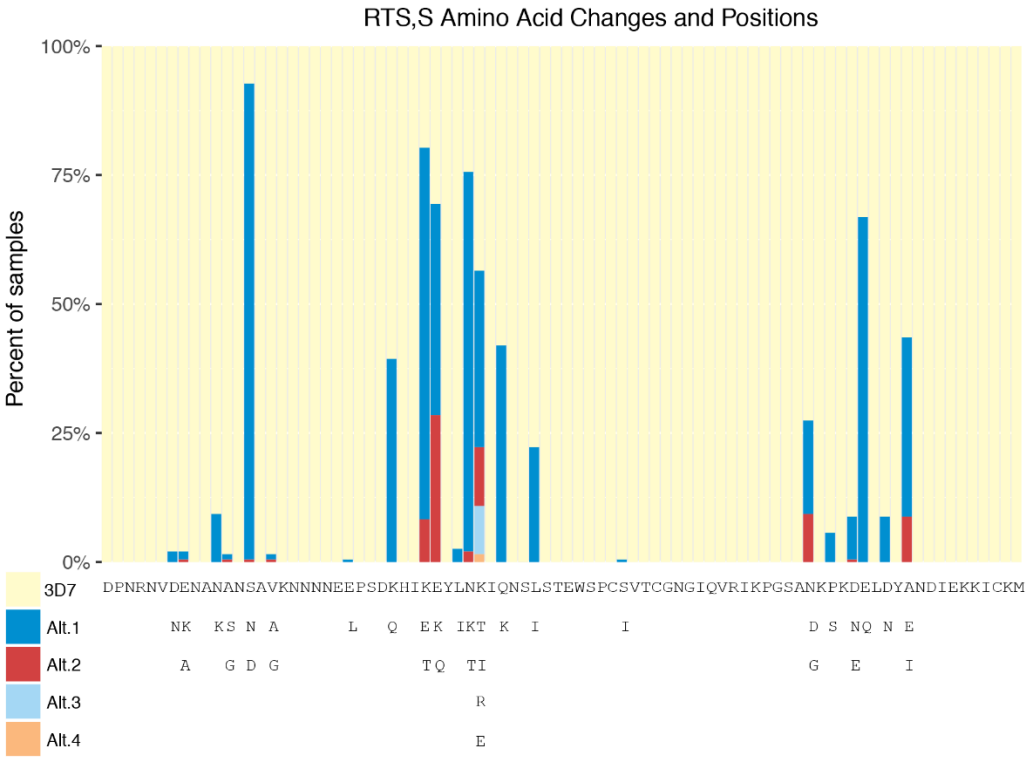

Figure 1: RTS,S Amino Acid Changes and Positions. The 84 amino acids (positions 288–371) comprising the C-terminal amplicon are represented by columns in the bar-chart. The percentage of samples sharing the 3D7 amino acid are represented in pale yellow. Non-3D7 amino acid alternatives are represented in descending order of frequency in dark blue, red, light blue, or orange. Below the bar-chart, the 3D7 amino acid sequence is shown, with positions corresponding to the coordinates above. The substitutions at each of the 84 positions are enumerated below the 3D7 sequence.

Figure 2

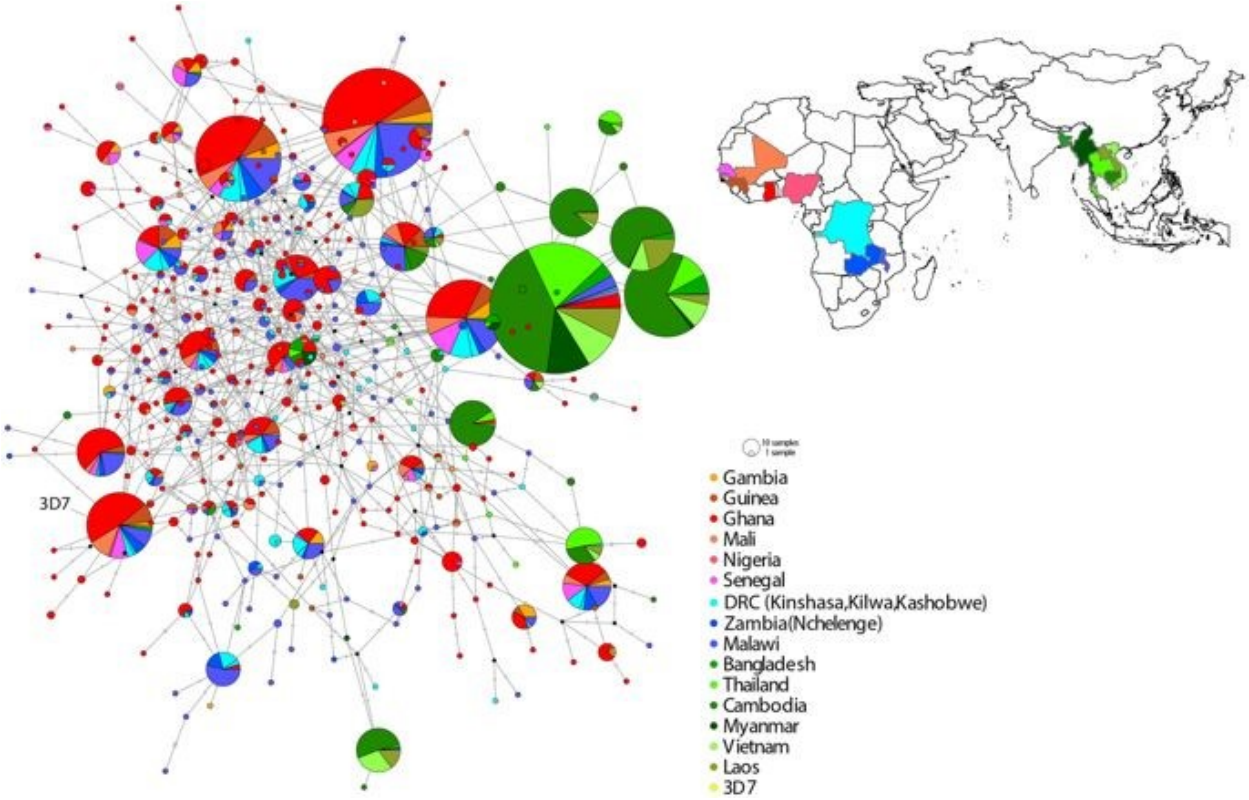

Figure 2: African and Asian *Pfcsp* Haplotype Network. Templeton, Crandall, and Sing (TCS) network summarizing the global diversity of the C-terminal *Pfcsp* from 4,002 sequences. Circles represent unique nucleotide haplotypes, and circles are scaled according to the frequency which the haplotype was observed. Vaccine strain 3D7 (0304600.1, PlasmoDB) is included for reference. Haplotype colors match the geographic origin of the samples depicted on the map.

REFERENCES

1. World Health Organization. World Malaria Report (2017).

2. Bhatt, S. *et al.* The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*. **526**, 207-211 (2015).

3. Mukonka, V.M. *et al*. High burden of malaria following scale-up of control interventions in Nchelenge District, Luapula Province, Zambia. *Malar J*. **13**, 153 (2014).

4. Clyde DF, Most H, McCarthy VC, Vanderberg JP. Immunization of man against sporozite-induced falciparum malaria. *Am J Med Sci*. **266**, 169-177 (1973).

5. Malik A, Egant JE, Houghtent RA, Sadofft JC, Hoffman SL. Human cytotoxic T lymphocytes against the *Plasmodium falciparum* circumsporozoite protein. *Med Sci*. **88**, 3300-3304 (1991).

6. Good, M.F. *et al.* Human T-cell recognition of the circumsporozoite protein of *Plasmodium falciparum*: Immunodominant T-cell domains map to the polymorphic regions of the molecule. *Immunology*. **85**, 1199-1203 (1988).

7. Zavala F, Cochrane AH, Nardin EH, Nussenzweig RS, Nussenzweig V. Circumsporozoite proteins of malaria parasites contain a single immunodominant region with two or more identical epitopes. *J Exp Med*. **157**, 1947-1957 (1983).

8. Casares S, Brumeanu T-D, Richie TL. The RTS,S malaria vaccine. *Vaccine*. **28**, 4880-4894 (2010).

9. Gordon, D.M. *et al*. Safety, immunogenicity, and efficacy of a recombinantly produced *Plasmodium falciparum* circumsporozoite protein-hepatitis B surface antigen subunit vaccine. *J Infect Dis*. **171**, 1576-1585 (1995).

10. Walliker, D. *et al*. Genetic Analysis of the Human Malaria Parasite *Plasmodium falciparum*. *Source Sci New Ser*. **236**, 1661-1666 (1987).

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

12. World Health Organization. Q&A on the malaria vaccine implementation programme (MVIP). http://www.who.int/malaria/media/malaria-vaccine-implementation-qa/en/ (2018).

13. Jalloh A, Jalloh M, Matsuoka H. T-cell epitope polymorphisms of the *Plasmodium falciparum* circumsporozoite protein among field isolates from Sierra Leone: age-dependent haplotype distribution? *Malar J.* **8**, 120 (2009).

14. Zeeshan, M. *et al*. Genetic Variation in the *Plasmodium falciparum* Circumsporozoite Protein in India and Its Relevance to RTS,S Malaria Vaccine. Gruner AC, ed. *PLoS One*. **7**, e43430 (2012).

15. Barry AE, Schultz L, Buckee CO, Reeder JC. Contrasting Population Structures of the Genes Encoding Ten Leading Vaccine-Candidate Antigens of the Human Malaria Parasite, *Plasmodium falciparum*. Rénia L, ed. *PLoS One*. **4**, e8497 (2009).

16. Waitumbi, J.N. *et al*. Impact of RTS,S/AS02A and RTS,S/AS01B on genotypes of *P. falciparum* in adults participating in a malaria vaccine clinical trial. *PLoS One*. **4,** 1-8 (2009).

17. Alloueche, A. *et al.* Protective Efficacy of the RTS,S/AS02 *Plasmodium falciparum* Malaria Vaccine is not Strain Specific. *Am J Trop Med Hyg*. **68**, 97-101 (2003).

18. Enosse, S. *et al*. RTS,S/AS02A malaria vaccine does not induce parasite CSP T cell epitope selection and reduces multiplicity of infection. *PLoS Clin Trials*. **1**, e5 (2006).

19. Bojang, K.A. *et al*. Efficacy of RTS, S/AS02 malaria vaccine against *Plasmodium falciparum* infection in semi-immune adult men in The Gambia: a randomised trial. *Lancet*. **358**, 1927-1934 (2001).

20. Neafsey, D.E. *et al*. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N Engl J Med*. **373,** 2025-2037 (2015).

21. Moss, W.J. *et al.* JW. Malaria Epidemiology and Control within the International Centers of Excellence for Malaria Research. *Am J Trop Med Hyg*. **93**, 5-15 (2015).

22. Kamuliwo, M. *et al.* The changing burden of malaria and association with vector control interventions in Zambia using district-level surveillance data, 2006–2011. *Malar J*. **12**, 437 (2013).

23. Parr, J.B. *et al.* Estimation of *Plasmodium falciparum* Transmission Intensity in Lilongwe, Malawi, by Microscopy, Rapid Diagnostic Testing, and Nucleic Acid Detection. *Am J Trop Med Hyg*. **95**, 373-377 (2016).

24. Aurrecoechea, C. *et al.* PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res*. **37**, D539-D543 (2009).

25. Illumina. 16S Metagenomic Sequencing Library Preparation. *Illumina.com*. 1-28 (2013).

26. Mag T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* **27**, 2957-2963 (2011).

27. Hathaway NJ, Parobek CM, Juliano JJ, Bailey JA. SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing. *Nucleic Acids Res*. November 2017.

28. The Pf3K Project (2016): pilot data release 5. www.malariagen.net/data/pf3k-5.

29. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. **20**, 1297-1303 (2010).

30. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. **43**, 491-498 (2011).

31. Zhu SJ, Almagro-Garcia J, McVean G. Deconvolution of multiple infections in *Plasmodium falciparum* from high throughput sequencing data. *Bioinformatics*. **34**, 9-15 (2017).

32. Leigh JW, Bryant D. popart : full-feature software for haplotype network construction. Nakagawa S, ed. *Methods Ecol Evol*. **6**, 1110-1116 (2015).

33. Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. *Mol Ecol*. **9**, 1657-1659 (2000).

34. Templeton, A.R., Crandall, K.A., Sing C.F. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*. **132**, 619-33 (1992).

35. Preston M.D. *et al.* A barcode of organellar genome polymorphisms identifies the geographic origin of Plasmodium falciparum strains. *Nat Commun*. **5**, 1-7 (2014).

36. Bailey, J.A. *et al.* Use of Massively Parallel Pyrosequencing to Evaluate the Diversity of and Selection on *Plasmodium falciparum* csp T-Cell Epitopes in Lilongwe, Malawi. *J Infect Dis*. **206**, 580-587 (2012).

Chapter 5

Conclusions

Zambia, along with numerous global partners and regional governments, is committed to malaria elimination by 2021 and elimination throughout southern Africa by 2030. Zambia has made great strides towards this goal in the past two decades, decreasing malaria incidence substantially. Although malaria control has been highly successful in some regions of the country, transmission has been refractory to control interventions in others. Given the uncertainty surrounding the mechanisms of sustained transmission in spite of control measures, new research is urgently needed to guide evidence-based strategic planning. The research presented in this dissertation aims to fill a critical knowledge gap by identifying mechanisms through which malaria transmission continues despite current control measures in two epidemiologically distinct regions in Zambia.

Plasmodium falciparum genetic and genomic analysis provides a means with which to track how malaria spreads and identify barriers to control and elimination. In fact, parasite genetics have proven useful in addressing a variety of epidemiological questions including identifying outbreak sources, monitoring the spread of drug resistance, and evaluating whether control interventions have successfully bottlenecked parasite populations, an indicator of decreased transmission. In the research presented in this dissertation, we describe the utility of *P. falciparum* genetic analyses to enhance our understanding of malaria transmission dynamics in Zambia.

Choma District, located in Southern Province, Zambia experiences low-level, seasonal malaria transmission and Southern Province is currently considered to be a pre-elimination setting. The Zambian National Malaria Elimination Program (NEMP) recommends reactive case detection (RCD), whereby all individuals living within 140m of index case households are screened for malaria using rapid diagnostic tests (RDTs) to identify and disrupt the remaining pockets of transmission in low transmission zones, including Choma District. Because RCD is a resource-intensive strategy, it is important to validate its ability to capture local transmission events. Prior to this dissertation, the spatial distribution of malaria cases was the primary tool used to assess whether RCD was effective at capturing secondary transmission events [1]. In fact, prior research in Macha demonstrated that the majority of secondary malaria cases detected during RCD-screening events lived within the index case household, possibly implicating local transmission as a source of new infections [2]. However, the spatial patterns of cases alone do not distinguish between local malaria transmission and other mechanisms, like household travel, a common occupational exposure, or clustering of malaria risk factors. Parasite genotyping enables a finer resolution understanding of how low-level malaria transmission is sustained in this region and whether RCD adequately captures local transmission events. We used microsatellites to genotype *P. falciparum* isolates captured during an expanded RCD sampling framework in Choma District to determine that parasite genetic relatedness is high on focal temporal and spatial scales, and that it declines both with increasing physical distance and increasing elapsed time. Our data suggest that local malaria transmission in on-going in Choma District, Southern Province, Zambia, and that it extends up to at least 250m from index case households. The current

NMEP guidelines of screening within 140m of index case households during RCD may miss infections that could contribute to the sustained malaria burden in this region. Although it may be operationally impractical to extend the RCD screening radius beyond 140m, these data suggest that interventions like focal vector control could be coupled with RCD to more effectively disrupt local transmission.

As malaria control programs transition from scaling up malaria control to focusing on elimination strategies, it is important to consider targeting interventions to high-risk geographical hotspots or high-risk demographic groups. In our study region in Choma District, we developed a method for identifying putative focal malaria transmission hotspots to which targeted vector control interventions could be directed. By detecting geographical zones across which parasite pairs were more related than would be expected by chance alone, we identified a region in northern Macha as a focal transmission hotspot. Although this focal transmission hotspot method requires additional validation, this approach may be of use in other low prevalence regions aiming to develop evidence-based methods to direct malaria control interventions. Importantly, we used a relatively low-tech, low-cost genotyping method to discern important signals of local transmission patterns in a pre-elimination setting, exemplifying how parasite genetic approaches could potentially be translatable to control programs, even in resource limited settings.

Nchelenge District, in Luapula Province, northern Zambia continues to experience holoendemic transmission despite scale-up of malaria control interventions including ITNs and IRS for over a decade. Identifying how malaria transmission circumvents control measures in this region would enhance operational planning. In order

to test the hypothesis that cross-border transmission from neighboring DRC contributes to sustained transmission, we performed amplicon deep sequencing at two highly polymorphic *P. falciparum* loci, *Pfama1* and *Pfcsp*, from individuals living in randomly selected households in both Zambia and DRC between June and July, 2016. We compared pairwise percent nucleotide identity between parasites from the same country and from those between different countries in order to determine whether within-country parasite pairs were more closely related than between-country pairs, an indicator of country-level parasite population genetic structure and gene flow. Our results indicate that between-country relatedness was similar to within-country parasite relatedness, suggesting an absence of population structure between *P. falciparum* from Zambia and the DRC. Additionally, we observed that the majority of *Pfama1* and *Pfcsp* DNA haplotypes were found in similar proportions in parasites from Zambia and the DRC. Furthermore, discriminatory analysis of principal components (DAPC) suggested no population partitioning between parasites from our sampling sites in these countries. However, DAPC revealed possible signatures of population structure between parasites from the two ecologically distinct regions, the lakeside and the inland, within Nchelenge District in Zambia. These two regions have unique ecotypes, one bordering the large Lake Mweru and the other, a marshier area with a single large stream. Thus, it is plausible that this difference in focal ecology has driven parasite population structure within Nchelenge District. Our small sample size precludes us from drawing strong conclusions regarding this suggested population structure, and additional sequencing of parasite samples from within Nchelenge District are merited to further test this hypothesis.

Power simulation experiments estimate that genotyping 10 samples per population at 20 SNPs had complete power to detect F_{ST} of 0.2, which would indicate strong genetic differentiation between populations [3]. To detect moderate population differentiation (F_{ST} = 0.01), 20 SNPs in 60 samples per population is associated with power of 0.8 [3]. Further, Morin et al. demonstrated that 100 samples per country and 75 SNPs is required to detect low F_{ST} of 0.0025 with power of 0.8 [3]. Of course, these simulations were done for unlinked, neutral, SNPs, and additional simulations may be required to fully assess the effect of using linked SNPs on such power calculations. From Morin et al.'s power simulations, we estimate that we had sufficient power to detect strong signatures of population structure, but insufficient SNPs and/or sample size to detect very low signatures of population differentiation. We recommend sequencing additional samples to achieve population sizes of at least 100 samples per country in order to further test for low level population structuring between samples from Zambia and the DRC. Unfortunately, amplicon deep sequencing is currently the best method to use in regions like Nchelenge District/Haut-Katanga Province where polyclonal infections are prevalent, meaning that it is unlikely that higher numbers of SNPs could be characterized than the numbers used here (*Pfama1*: 38 SNPs; *Pfcsp*: 25 SNPs) unless long-read sequencing technology like Pacific Biosystems was used. However, as computational methods for phasing SNPs and microsatellites from polyclonal infections improve, there may be increased opportunities to conduct additional population genetic analyses using large numbers of unlinked, neutral loci, thereby providing additional insights into the degree to which parasites on either side of the border are panmictic, and whether cross-border transmission is on-going in this region. In summary, our results

suggest that *P. falciparum* parasites on either side of the Zambia-DRC border are part of a large, panmictic parasite population that does not exhibit strong population differentiation. Additional sequencing is required to test whether there is low level population differentiation between parasites from these two countries.

Although our data represent small sample sizes and rely on linked, non-neutral SNPs, our results are corroborated by *Anopheles funestus* genomic data suggesting that malaria vector populations on either side of the Zambia-DRC border also represent a single interbreeding population. Taken together, these observations indicate that cross-border malaria transmission may be on-going in this region. A cross-sectional survey among randomly sampled households in Kilwa and Kashobwe, DRC in 2016 revealed that IRS has been minimally deployed in this region at the time of sample collection. Given our data which suggest gene flow between parasites on either side of the Zambia-DRC border, it is possible that cross-border transmission from the DRC contributes to the sustained burden of malaria in spite of control interventions in Nchelenge District. Additional simulation and modeling experiments are merited to further explore the temporal and spatial resolution on which the SNPs used in this analysis can detect population structuring events.

These results highlight the importance of international partnerships to coordinate malaria control programs, particularly along international borders where transmission on either or both side of the border may be high. Although the Elimination 8 (E8) partnership in Southern Africa aims to do precisely this, the DRC is notably absent from this organization. Continuing to build regional capacity to address cross-border malaria,

particularly with the DRC, will be critical to malaria elimination efforts in Southern Africa.

In 2015, GlaxoSmithKline (GSK) licensed the first malaria vaccine, RTS,S/AS01, based on a portion of the C-terminal amplicon of the circumsporozoite protein (CSP) from just one *P. falciparum* clone, 3D7. Research from Phase III Clinical Trials demonstrated that any amino acid deviations in the C-terminal portion of CSP included in the vaccine resulted in decreased vaccine efficacy [4]. Given that RTS,S/AS01 has only a 36.3% vaccine efficacy against severe malaria to begin with, further decreases in vaccine efficacy due to population genetic variation are concerning. As countries consider whether to incorporate RTS,S/AS01 into their national control programs, it is critical to understand the potential impact of parasite population diversity on vaccine efficacy. Towards this goal we sequenced *P. falciparum* isolates collected from individuals living along the Zambia-DRC border between Nchelenge District, Zambia and Haut-Katanga Province, DRC, a region of year-round, holoendemic malaria transmission. We found that only 10% of the parasite clones from this study area matched the 3D7 clone in the C-terminal region of CSP at every amino acid. Further, we used the open database, Pf3k, to compare sequences from our study area to those of global isolates. We found that 5.3% of African parasite sequences and only 0.2% of Asian parasite sequences perfectly matched the 3D7 vaccine strain in the C-terminal amplicon of CSP. Ideally, an effective malaria vaccine would provide protection against the majority of circulating parasites across multiple geographic regions. Our data provide evidence that *Pfcsp* exhibits high genetic diversity both locally and globally. It is possible that such diversity could undermine RTS,S/AS01 vaccine efficacy.

The data presented here highlight the diversity of C-terminal *Pfcsp*, particularly in sub-Saharan Africa which is a key target region for malaria vaccination programs. This study underscores the importance of incorporating population genetic studies into future malaria vaccine design, laboratory and clinical evaluation. Most importantly, assessing the diversity of C-terminal *Pfcsp* should be a component of the RTS,S/AS01 Malaria Vaccine Implementation Programme and in further refining our understanding of how genetic diversity affects RTS,S/AS01 efficacy.

In conclusion, *P. falciparum* genetic and genomic approaches are valuable for answering epidemiologically relevant questions for malaria control and elimination. With more countries aiming to eliminate malaria within the coming decades, it will be important to identify and address the mechanisms that continue to sustain transmission in spite of control measures. Parasite genetic approaches will be a critical contributor to our understanding of the barriers to malaria control and elimination in a variety of epidemiological settings.

REFERENCES

1. Searle KM, Hamapumbu H, Lubinda J, et al. Evaluation of the operational challenges in implementing reactive screen-and-treat and implications of reactive case detection strategies for malaria elimination in a region of low transmission in southern Zambia. Malar J. **2016**; 15(1):412.

2. Deutsch-Feldman M, Hamapumbu H, Lubinda J, et al. Efficiency of a Malaria Reactive Test-and-Treat Program in Southern Zambia: A Prospective, Observational Study. Am J Trop Med Hyg. **2018**; 98(5):1382–1388.

3. Morin PA, Martien KK, Taylor BL.Assessing statistical power of SNPs for population structure and conservation studies. Mol Ecol Resour. **2009**; 9(1):66-73.

4. Neafsey DE, Juraska M, Bedford T, et al. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. N Engl J Med. Massachusetts Medical Society; **2015**; 373(21):2025–2037.

Appendix A: Chelex DNA extraction and *Pfldh* quantitative PCR protocol

SOP #:2014-1; Steve Meshnick Laboratory; University of North Carolina at Chapel Hill

TITLE: Chelex gDNA Extraction and quantitative PCR of PfLDH

Version: 1

Author(s): Stephanie Doctor, Steve Taylor, Jaymin Patel (modified by Julia Pringle)

1. Introduction

This SOP describes the procedures necessary to efficiently and reliably extract genomic DNA (gDNA) from these blood spots using Chelex resin, and use real-time PCR (qPCR) to detect *Plasmodium falciparum* in each sample.

The blood spots will be on filter paper and will be deposited in Eppendorf tubes in which they will undergo a two-day procedure to extract gDNA using Chelex 100 (Bio-Rad, Richmond, CA), a chelating ion-exchange resin. This procedure requires approximately 15 minutes on day one and three hours on day two with an overnight incubation in between. Finally, the extracted gDNA will be used to set up a real-time PCR assay to detect *Plasmodium falciparum*.

2. Materials Needed

2.1. 10% saponin in water (Sigma Aldrich #47036-50G-F)
2.2. Vortex
2.3. Water bath
2.4. Molecular grade water
2.5. Eppendorf tubes (1.5mL)
2.6. 20% Chelex in water (Bio-Rad #143-2832)
2.7. Ice bucket
2.8. SYBR Green Light Scanner Master Mix
2.9. PfLDH forward primer at 100 uM (ACG ATT TGG CTC GAG CAG AT)
2.10. PfLDH reverse primer at 100 uM (TCT CTA TTC CAT TTT GTC ACT CTT TC)
2.11. 5 mL tube
2.12. 1-50 uL filtered tips
2.13. DNA standards
2.14. qPCR plate adhesive covers
2.15. 96-well qPCR plates

3. Procedure

3.1. Chelex – Initial incubation

3.1.1. Put gloves on. Wipe down work area with 70% ethanol.

3.1.2. Add 1ml of 1x PBS and 50 uL of 10% saponin (1g saponin in 10 mL water) to **each tube containing dried blood spots**. Vortex samples for 20-30 seconds.

3.1.3. Store the plate overnight at 4°C. Store extra saponin at -20C.

3.2. Chelex – DNA extraction

3.2.1. Centrifuge samples for 1 minute at 1000 rpm. Aspirate and discard the supernatant (reddish brown PBS/saponin mix) from each well of the plate.

3.2.2. Add 1ml of PBS (without saponin) to each tube containing a dried blood spot. Vortex 20-30 seconds.

3.2.3. Incubate samples at 4°C for 30 minutes.

3.2.4. Set the water bath at this time to 95-99°C

3.2.5. Centrifuge tubes at 1000 rpm for 1 minute. Aspirate and discard as much fluid as possible from each well. Press the tip down on the filter paper into the lower end of the tube without packing it excessively.

3.2.6. Add 100 uL of sterile water to each tube containing a dried blood spot.

3.2.7. Ensure Chelex beads are suspended in the water and they don't settle at the bottom of the tube by vortexing or inverting tube containing the Chelex solution.

3.2.8. Add 50 uL of 20% Chelex (5g Chelex in 25 mL water) to each sample, being certain that Chelex beads remain suspended in the solution prior to adding to the samples. Cut the ends of the pipette tips off to ensure Chelex beads are picked up.

3.2.9. Incubate samples in a 95-99°C water bath for 12 minutes. Vortex vigorously every 2-3 minutes.

3.2.10. After incubation, centrifuge at 1400 rpm for 5 minutes.

3.2.11. Transfer as much solution as possible from the spun plate into new tubes, not worrying if Chelex beads have been carried over. **It is essential that: 1) the source and destination tubes are the same, 2) you change tips after every dispense, 3) you transfer as much liquid as possible from the source to the destination tube, and 4) you don't transfer the blood spots themselves.**

3.2.12. Spin the new plate for 10 minutes at 1400 rpm.

3.2.13. Transfer the final, white to yellowish supernatant into a new real-time PCR plate, avoiding the pelleted Chelex at the bottom of the plate. **It is essential that each tube with no sample is empty.**

3.2.14. Label the new plate indicating that it contains chelex extracted DNA with sample ID. Apply a new fplate cover to the plate and store at -20°C.

3.2.15. Wipe down the bench area and pipettes used in the procedure.

3.3. qPCR setup

3.3.1. Take your sample plates out of the -20C freezer to thaw.

3.3.2. Make qPCR master mix

1. Fill a bucket with ice and gather SYBR Green qPCR Mix, PfLDH forward and reverse primers, and molecular grade water. Vortex each and spin briefly before using.
2. In the PCR hood, create the qPCR master mix in a 5 mL tube as follows:

Master Mix Component	**Volume per reaction**
SYBR Green Master qPCR Mix	12.5 µL
PfLDH F primer (20uM)	0.75 µL
PfLDH R primer (20uM)	0.75 µL
Water	7.0 µL
DNA	4 µL

Total Volume	25.0 µL

3.3.3. Return all materials to the freezer

3.3.4. Apply a Bio-Rad qPCR adhesive cover and use the plastic blade to seal.

3.3.5. Cover each sample plate with a new foil cover and return to the -20C freezer.

3.4. PfLDH qPCR assay

Step	**Temperature**	**Time**	**Number of cycles**
Holding	50°C	2 minutes	
Holding	95°C	10 minutes	
Cycling	95°C 60°C	15 seconds 1 minutes	40

Appendix B: Selective whole genome amplification

From the Greenhouse Laboratory; University of California, San Francisco

Summary

- This protocol describes a method for selective whole genome amplification (sWGA) of *Plasmodium falciparum* from samples that contain a high proportion of human DNA (Oyola, Ariani et al., 2016).
- This method is designed for use on samples where the human contamination is too high to allow efficient or successful downstream processing of the parasite DNA, for example where parasitemias are very low.
- Because the protocol is designed using the phi29 polymerase it is equivalent to the Qiagen Repli-G MDA kit or the GE Healthcare Genomiphi multiple-displacement amplification (MDA) kit.
- This reaction should produce long amplicons of several kb to 10's of kb in length and should be suitable for a range of procedures particularly whole-genome-sequencing.

Materials:

- *phi29* Enzyme kit (NEB #M0269L)
- dNTPs (10mM pool - 7mM AT and 3mM GC*)
- sWGA primer pool (10A=250uM and 6A100uM)
- Nuclease-free water
- 15 ml Falcon tube
- PCR plates skirted type for running the sWGA reactions (AB-1400-L)
- PCR plate seal [Building 100]

*dNTPs are adjusted 70:30 ratio of AT and GC (since Pf genome is AT reach)

Methods summary:

1. In the aster mix room: A master-mix of enzyme, buffer, dNTPs and primers is made up sufficient for 100 wells.
2. Add 40uL of gDNA (We will be able to adjust this for high parasitaemia samples in due course).
3. Seal plate using Aluminum foil seals – NOT heat seal
4. Put it on ice and take the plate to building 100.
5. Run the step-down isothermal temperature profile for 16 hours (note that shorter times were tested and 16hrs was better in terms of coverage).
6. The reactions are then heat-treated to inactivate the phi29 enzyme.

7. Typical QC of the material measures total DNA produced and includes some form of qPCR/genotyping to ensure the gDNA of interest has been amplified and allow some idea of the quality of the material.

Master-mix recipe for 50µL/100µL sWGA reaction volume:

sWGA PCR	6A+10AD - 50uL		6A+10AD - 100uL	
	1x - 50uL	100X	1x- 100uL	100X
H20			2.5	250
10X NEB Phi29 Buffer	5	500	10	1000
10mg/ml BSA	0.5	50	1	100
100uM Primer set 6A	1.25	125	2.5	250
250uM Primer set 10A	0.5	50	1	100
5mM dNTP	20	2000	40	4000
Phi29 Enzyme	3	300	3	300
Total	**30.25**	**3025**	**60**	**6000**
DNA	20		40	
Total	**50**		**100**	

• Cycling condition

- **Step-down protocol**
  - 35^0C for 5 min,
  - 34^0C for 10 min,
  - 33^0C for 15 min,
  - 32^0C for 20 min,
  - 31^0C for 30 min,
  - 30^0C for 16 hrs
- **Heat-inactivation of the phi29**
  - 65°C for 15 min
- **Cool until ready to remove**
  - 10°C forever (only use 4°C for short periods to prevent the PCR machine overworking.

Appendix C: Microsatellite Loci Amplification

From the Greenhouse Laboratory; University of California, San Francisco (version 001, March 1, 2016):

1. **Overview:** This SOP describes the procedure of microsatellite genotyping of P. falciparum malaria parasites.

2. **Safety:** Adhere to local safety regulations. Wear appropriate personal protective equipment.

3. **Materials:**

   a. *Equipment:*

Item
Range of single and multichannel pipettes
Barrier Tips for pipettes
Troughs
96 well PCR plates
384 well PCR plates
1.5mL Eppendorf tubes
96 and 384 well PCR machines
CE Plates (96-well)
Centrifuge
Vortex
latex or nitrile gloves
PCR plate seals
PCR plate sealer

   b. *Reagents: All reagents should be stored according to the instructions supplied with them and disposed of at the expiry date recorded on the product.*

Item	Company	Product Number	Storage
2x Type-It Multiplex Master Mix	Qiagen	1053044	-20 freezer
Microsatellite marker primers, tagged and untagged	IDT		-20 freezer
TE Buffer (10x)	Teknova	T3457	Room temp
Molecular-grade water			Room temp
dNTPs	Omega Bio-Tek	101414-958	-20 freezer
Phusion II Hot Start	Thermo Fisher	F-518	-20 freezer
5X HF Phusion Buffer	Thermo Fisher	F-549L	-20 freezer

DBS extracted DNA (see Dual Extraction Protocol)			-20 freezer
ROX			refrigerator
Formamide			-20 freezer

4. Preparation of Reagents:

Prepare dNTPs and 10X Primer Mixture for Groups 1-4 (see below) from stock solution to appropriate concentration using TE buffer.

Group 1-10x Primer Mixture (100uL)				
Primer	**Concentration (µM)**	**Volume (µL)**	**Multiple**	
AS1	1.25	1.25		12.5
AS11	1.25	1.25		12.5
AS12	2.5	2.5		25
AS20	2	2	10	20
AS3	1.5	1.5		15
AS31	1.25	1.25		12.5
AS32	2	2		20
AS34	2	2		20
TA1	3	3		30
TA109	3	3		30
TE Buffer		60.5		605

Group 2-10x Primer Mixture (100uL)				
Primer	**Concentration (µM)**	**Volume (µL)**	**Multiple**	
TA60RevNewII + TA60NewForT	3	3		30
AS25	1.5	1.5		15
PFG377	2	2		20
AS19	3	3	10	30
Ara2New	3	3		30
AS21	1.25	1.25		12.5
AS8	2	2		20
TE Buffer		65.5		625

Group 3- 10x Primer Mixture (100uL)				
Primer	**Concentration (µM)**	**Volume (µL)**	**Multiple**	
TA81	3	3		30
AS7	2	2		20
AS15	1.25	1.25	10	12.5
TA40	3	3		30
PolyA New	3	3		30
PfPk2 New	3	3		30
TE Buffer		65.5		675

Group 4-10x Primer Mixture (100uL)				
Primer	**Concentration (µM)**	**Volume (µL)**	**Multiple**	
AS14	2	2		20
TA87	3	3		30
B7M19	2	2		20
AS2	2	2	10	20
TE Buffer		82		820

5. Methods:

1. Thaw reagents and template DNA samples
2. For each 96-well rack of template DNA samples, prepare 4 multiplex 96-well PCR reaction plates (Groups 1-4) as follows:

Master Mix	Volume (µL)	Multiple	Final Volume (µL)
2x Type-It Multiplex Master Mix	12.5	110	1375
10x Primer Mix	2.5		275
Water	5		550
Template DNA	5		
Reaction Volume			25

3. Run Groups 1-4 in 96-well PCR machine

a. PCR cycling conditions for Groups 1 & 4:
    1. 95°C 5 minutes
    2. 95°C 30 sec
    3. 60°C 1:30 min
    4. 72°C 30 seconds
    5. X30 (from step 2)
    6. Extension @ 60°C for 30 min
    7. Infinite hold @ 4°C
b. PCR cycling conditions for Groups 2 & 3:
    1. 95°C 5 minutes
    2. 95°C 30 sec
    3. 56°C 1:30 min
    4. 72°C 30 seconds
    5. X30 (from step 2)
    6. Extension @ 60°C for 30 min
    7. Infinite hold @ 4°C

4. Prepare Plates 1-10 of individual marker PCR reactions as follows using previously prepared Groups 1-4 (from Step 3) as template DNA:

Microsatellite Master Mix 5µl scale (1µl temp) Phusion PCR				
	Final Conc.	**Volume (µl)**	**Multiple**	**Total (µl)**
H2O		1.95		214.5
P1 10uM	500nM	0.25		27.5
P2 10uM	500nM	0.25		27.5
5X HF Phusion Buffer	1X	1	110	110
dNTP 2mM	200uM	0.5		55
Phusion II Hot Start (1U/uL)	0.4u (0.04u/µl)	0.05		5.5
		4		Mix / Well
THEN ADD template DNA		1		

Plate	**Marker**	**Group**	**Location**	**Cycling Condition**
1	AS1	1	Q1	Phu3
	AS32	1	Q2	
	TA1	1	Q3	
	AS11	1	Q4	
2	AS34	1	Q1	
	TA109	1	Q2	
	AS31	1	Q3	

	PfPk2New	3	Q4	
3	B7M19	4	Q1	
4	AS12	1	Q1	Phu 7
	AS19	2	Q2	
	AS21	2	Q3	
5	AS15	3	Q1	
	AS7	3	Q2	
6	AS3	1	Q1	Phu4
	AS8	2	Q2	
	AS2	4	Q3	
7	TA81	3	Q1	Phu5
	TA87	4	Q2	
8	PFG377	2	Q1	Phu1
	TA40	3	Q2	
9	AS25	2	Q1	Phu10
	PolyA	3	Q2	
10	TA60	2	Q1	Phu13
	Ara2	2	Q2	
	AS14	4	Q3	

5. Prepare 12 dilution plates as follows:

Dye	Locus	CE Plate	Dilution Ratio	Dilution Vol. (uL)	Water Vol. (uL)	Indiv. Rxn Plate	Indiv. Rxn Quad
FAM	AS1	1	1:500	0.8	400	1	Q1
NED	AS11		1:400	1		1	Q4
HEX	AS31		1:400	1		2	Q3
FAM	AS32	2	1:400	1	400	1	Q2
HEX	TA1		1:400	1		1	Q3
NED	AS34		1:200	2		2	Q1
FAM	TA109	3	1:400	1	400	2	Q2
NED	TA87		1:350	1.144		7	Q2
FAM	AS2	4	1:500	0.8	400	6	Q3
NED	AS14		1:300	1.32		10	Q3
HEX	TA81	5	1:500	0.8	400	7	Q1
FAM	AS12		1:500	0.8		4	Q1
NED	TA60	6	1:50	2	100	10	Q1
FAM	PFG377		1:100	1		8	Q1

HEX	AS19	7	1:500	0.8	400	4	Q2
FAM	Ara2		1:500	0.8		10	Q2
NED	AS21		1:300	1.32		4	Q3
FAM	AS8	8	1:150	1.32	400	6	Q2
NED	AS15		1:500	0.8		5	Q1
FAM	B7M19	9	1:500	0.8	400	3	Q1
NED	AS3		1:250	1.6		6	Q1
HEX	AS25		1:400	1		9	Q1
FAM	AS7	10	1:400	1	400	5	Q2
HEX	PfPK2		1:450	0.88		2	Q4
FAM	TA40	11	1:100	4	400	8	Q2
FAM	PolyA	12	1:200	2	400	9	Q2

a. Seal dilution plates and vortex
b. Spin plates down in centrifuge (short spin) to remove any liquid from seal

6. Prepare final CE plates as follows:

CE Plate	**Volume (µL) per well**	**Multiple**	**Final Volume (µL)**
Formamide	9.9	1,176	11,642.4
ROX	0.1		117.6
Diluted DNA	2		
Master Mix per well			**10**
Reaction Volume			12

7. Submit CE order to Quintara Biosciences via their online form (copy and paste from excel file)

Appendix D: AMA1 Amplicon Generation (Illumina Adapters Included)

JP Updated: October 22, 2018

Primers (includes Illumina adapter sequences)

AMA1_553_For: TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCA ACA AAA CCT CTT ATG TCA CCA

AMA1_1490_Rev: GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GTT AGG TTG ATC CGA AGC ACT CA

Hydrate lyophilized primers in 1x TE (10mM Tris +1mM EDTA)

Master Mis

Reagent	Amount Per Reaction (µL)
KAPA HiFi Taq Master Mix	12.5
Nuclease Free Water	0
Forward Primer (5uM stocks)	1.0
Reverse Primer (20uM stocks)	0.25
Magnesium chloride (25µM)	2.0
Template DNA	9.25
Total Volume	25

Cycling Conditions

95°- 5 min

33 cycles of:

- 95°- 30 seconds
- 58°- 1 minute
- 72°- 1 minute

72°- 5 min

4°-forever

Expected Product Size: ~550bp

Appendix E: CSP Amplicon Generation (Illumina Adapters Included)

JP Updated: October 22, 2018

Primers (Illumina adapter sequences included)

CSP_1190_For: TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GAC AAG GTC ACA ATA TGC CAA A

CSP_1490_Rev: GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GAC ATT AAA CAC ACT GGA ACA TTT TTC

Hydrate lyophilized primers in 1x TE (10mM Tris +1mM EDTA)

Master Mis

Reagent	Amount Per Reaction (µL)
KAPA HiFi Taq Master Mix	12.5
Nuclease Free Water	0
Forward Primer (20uM stocks)	0.25
Reverse Primer (20uM stocks)	0.25
Magnesium chloride (25µM)	2.0
Template DNA	10
Total Volume	25

Cycling Conditions

95°- 2 min

33 cycles of:

95°- 30 seconds

62°- 30 seconds

60°- 45 seconds

72°- 5 min

4°-forever

Expected Product Size: ~350bp

JULIA C. PRINGLE

Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe St.
Baltimore, MD 21205, USA
Mobile: (610)-329-0915
Email: julia.pringle108@gmail.com
Birthdate: October 8, 1989
Place of birth: Baltimore, MD, USA

EDUCATION

PhD: Department of Molecular Microbiology and Immunology 09/2014-Present
Johns Hopkins Bloomberg School of Public Health
Area of concentration: Molecular epidemiology and genetics
Dissertation: "Understanding mechanisms of sustained malaria transmission in Zambia through *Plasmodium falciparum* genetics"
GPA: 3.88 of a maximum 4.0
Expected data of completion: 11/2018

MSPH: Department of International Health 09/2016-Present
Johns Hopkins Bloomberg School of Public Health
Area of concentration: Global Disease Epidemiology and Control
Thesis: "High prevalence of RTS,S/AS01 vaccine mismatch among *Plasmodium falciparum* isolates sampled from Southern and Central Africa and globally"
GPA: 3.88 of a maximum 4.0
Expected data of completion: 11/2018

B.A.: Biology 09/2008-05/2012
Haverford College
Area of concentration: Molecular biology
Thesis: *Wolbachia* Interact with Different Components of Their Host's Cytoskeleton in Various Species of *Drosophila*
GPA: 3.74 of a maximum 4.0
Graduation Honors: *Cum laude*

WORK EXPERIENCE

PhD Candidate: Laboratory of Dr. Douglas Norris 11/2014-Present
Department of Molecular Microbiology and Immunology
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe St.
Baltimore, MD 21205, USA
Hours per week: 40
Duties, Accomplishments, and Related Skills:

- Lead a team of nine scientists to characterize the temporal and spatial patterns of malaria transmission hot-spots in Macha, Zambia using microsatellite genotyping, transmission network modeling, and geospatial analysis.
- Demonstrated that the extent of focal malaria transmission in Macha, Zambia is larger than would be captured by the currently implemented reactive case detection screen-and-treat strategy where individuals living within 140 meters of an index case are tested for malaria at one time point. These results were shared during a talk at the Genetic Epidemiology of Malaria Conference and were recently submitted as a manuscript to The Journal of Infectious Diseases.
- Established and oversee an on-going collaboration with eleven researchers across three institutions aimed at quantifying the magnitude of cross-border malaria transmission between northern Zambia and southeastern Democratic Republic of the Congo using amplicon deep sequencing and parasite population genetic analyses.
- Implemented a new genetic sequencing pipeline in the laboratory to determine that only 10% of malaria parasites in northern Zambia match the malaria vaccine strain, suggesting that vaccine efficacy may be attenuated if deployed in this region; this work was recently published by Scientific Reports.
- Process and analyze routinely collected mosquito specimens from two field sites in Zambia in order to characterize the behavior and ecology of malaria vectors in regions targeted for malaria elimination.
- Contribute to grant writing and Institutional Review Board (IRB) application preparation on behalf of an international research team.
- Contribute to the design and implementation of a new longitudinal cohort study in Southern Province, Zambia aimed at capturing and genotyping all incident malaria infections within a focal geographic region to distinguish between locally acquired and travel-related cases.

Supervisor: Dr. Douglas Norris (4106142710); Dr. William Moss (4105021165)
Okay to contact this Supervisor: Yes

Teaching Assistant: Vector Biology and Vector Borne Diseases 3rd Term: 2016, 2017, 2018
Department of Molecular Microbiology and Immunology
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe St.
Baltimore, MD 21205, USA
Hours per week: 6
Duties, Accomplishments and Related Skills:

- Wrote and graded student exams covering topics in arthropod biology, medical entomology, tick-borne diseases, arboviruses, and vector-borne parasitic infections.
- Organized and held review sessions for a class with an average size of 40 students.
- Mentored students individually to understand core concepts in vector biology, ecology, and disease transmission cycles.

- Served as a Teaching Assistant for this course during the third term of academic years 2015-2016, 2016-2017, and 2017-2018.

Supervisor: Douglas Norris (4106142710)
Okay to contact this Supervisor: Yes

Teaching Assistant: Spatial Analysis I: ArcGIS 1st Term: 2018
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe St.
Baltimore, MD 21205, USA
Hours per week: 6
Duties, Accomplishments and Related Skills:

- Provide technical software assistance to students learning the geographical information systems (GIS) software, ArcGIS.
- Hold office hours twice a week to provide support for course assignments and reinforce GIS concepts.
- Help students navigate GIS assignments including delineating food deserts in Baltimore City, identifying the source of an *E. coli* outbreak in California, and creating sexually transmitted infection risk maps.
- Contribute to grading of GIS assignments for a class with 108 students.
- Served as a Teaching Assistant for this course during the first term of academic year 2017-2018 and will continue to serve as a Teaching Assistant for this course during the first term of academic year 2018-2019.

Supervisor: Timothy Shields (4105029077)
Okay to contact this Supervisor: Yes

Teaching Assistant: Spatial Analysis II: Spatial Data and Technologies 2nd Term: 2018
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe St.
Baltimore, MD 21205, USA
Hours per week: 6
Duties, Accomplishments and Related Skills:

- Provide technical support for mapping and spatial analysis software including ArcGIS, QGIS, and Google Earth Pro to 52 students enrolled in the course.
- Hold weekly office hours to mentor students through concepts and processes in spatial data collection, generation, and analysis including mobile GPS technologies and 3D mapping.
- Graded course assignments.
- Served as a Teaching Assistant for this course during the second term of academic year 2017-2018 and will continue to serve as a Teaching Assistant for this course during the second term of academic year 2018-2019.

Supervisor: Timothy Shields (4105029077)
Okay to contact this Supervisor: Yes

Laboratory Technician- Laboratory of Dr. William Snider 08/2012-07/2014
Neuroscience Research Center
The University of North Carolina at Chapel Hill School of Medicine
Neuroscience Research Building
115 Mason Farm Rd.
Chapel Hill, NC 27599, USA
Hours per week: 40
Duties, Accomplishments and Related Skills:

- Conducted research to understand the roles of mitogen-activated protein kinase (MAPK) and glycogen synthase kinase (GSK) signaling pathways in the developing brain using mouse models.
- Demonstrated that perturbations in GSK signaling result in neuronal migration defects in the mouse cortex; these data were published in the journal Elife.
- Maintained a colony of numerous transgenic mouse lines totaling over 300 cages.
- Oversaw laboratory compliance with the Institutional Animal Care and Use Committee (IACUC) guidelines for the ethical treatment of animals.
- Assisted in laboratory management tasks including purchasing and preparation of reagents.
- Trained and supervised two undergraduate work-study students conducting part-time research in the laboratory.

Supervisor: Dr. William Snider (9198438623)
Okay to contact this Supervisor: Yes

Undergraduate Thesis Researcher-Laboratory of Dr. Rachel Hoang 08/2011-05/2012
Department of Biology
Haverford College
Marian E. Koshland Integrated Natural Sciences Center, L207
370 Lancaster Ave.
Haverford, PA 19041, USA
Hours per week: 40
Duties, Accomplishments and Related Skills:

- Pioneered new research in the laboratory aimed at characterizing the host-pathogen dynamics between *Drosophila* and their endosymbiont, *Wolbachia*, a bacterium previously implicated in the pathogenicity of *Onchocerca volvulus* to cause river blindness.
- Captured and analyzed confocal microscopy images to deduce patterns of host-pathogen interactions between *Wolbachia* and *Drosophila* embryos.
- Developed and implemented novel experimental protocols to ascertain the cellular compartments of *Drosophila* embryos commandeered by *Wolbachia* during infection.
- Determined that *Wolbachia* are trafficked within Golgi vesicles of *Drosophila melanogaster* embryos.

Supervisor: Dr. Rachel Hoang (6108961000)
Okay to contact this Supervisor: Yes

Teaching Assistant: Biochemistry and Molecular Biology Laboratory 1/2012-5/2012
Department of Biology
Haverford College
Marian E. Koshland Integrated Natural Sciences Center, L207
370 Lancaster Ave.
Haverford, PA 19041, USA
Hours per week: 6
Duties, Accomplishments and Related Skills:

- Assisted approximately 30 students with technical and analytical aspects of laboratory experiments.
- Resolved issues that arose during laboratory sessions including equipment malfunctions and protocol errors.
- Prepared materials for laboratory sessions two times each week.

Supervisor: Dr. Rachel Hoang (6108961000); Dr. Judy Owen (6108961192)
Okay to contact this Supervisor: Yes

Intern: MOMobile at Riverside Correctional Facility 05/2011-07/2011
Maternity Care Coalition
Riverside Correctional Facility
8151 State Rd
Philadelphia, PA 19136, USA
Hours per week: 40
Duties, Accomplishments and Related Skills:

- Applied for and received funding through the Center for Peace and Global Citizenship at Haverford College to self-design and execute a ten-week summer internship in 2011.
- Designed a ten-week internship to work with the non-profit organization, Maternity Care Coalition, at their site inside Philadelphia's Riverside Correctional Facility, providing prenatal services to pregnant inmates.
- Developed curriculum and co-instructed a prenatal workshop covering topics in nutrition, child development, sexually transmitted infections, HIV/AIDS, and post-partum depression.
- Conducted home visits to new mothers recently released from prison to facilitate access to early childhood services.
- Accompanied incarcerated women to the hospital to provide support during their deliveries.

Supervisor: Danyell Williams (215-680-8853)
Okay to contact this Supervisor: Yes

Research Assistant- Laboratory of Dr. Philip Meneely 06/2010-08/2010
Department of Biology
Haverford College
Marian E. Koshland Integrated Natural Sciences Center, L207

370 Lancaster Ave.
Haverford, PA 19041, USA
Hours per week: 40
Duties, Accomplishments and Related Skills:

- Conducted research using *C. elegans* to interrogate the role of meiotic complex protein, HIM-8's, role in homologous chromosome segregation.
- Maintained a colony of *C. elegans* with higher than normal incidence of improper chromosomal segregation.
- Identified the point mutation in the gene encoding him-8 that is responsible for the abnormal chromosome segregation phenotype using PCR and DNA sequencing.
- Utilized confocal microscopy to visualize the chromosomal abnormalities of the him-8 mutant colony of *C. elegans.*

Supervisor: Dr. Philip Meneely (6108964996)
Okay to contact this Supervisor: Yes

Science Library Staff 09/2008-05/2012
White Science Library
Haverford College
Marian E. Koshland Integrated Natural Sciences Center, L207
370 Lancaster Ave.
Haverford, PA 19041, USA
Hours per week: 10
Duties, Accomplishments and Related Skills:

- Operated the library's front desk and ensured effective patron service.
- Assisted patrons in locating and accessing library materials.
- Resolved issues with library accounts, printers, and room reservations.

Supervisor: Dora Wong (6108961416)
Okay to contact this Supervisor: Yes

PROFESSIONAL PUBLICATIONS

Peer Reviewed Publications

1. **Pringle JC,** Carpi G, Almagro-Garcia J, Zhu SJ, Kobayashi T, Mulenga M, Bobanga T, Chaponda M, Moss WJ, Norris DE. High prevalence of RTS,S/AS01 vaccine mismatch among *Plasmodium falciparum* isolates sampled from Southern and Central Africa and globally. Scientific Reports. 2018 Apr 26;8:6622. PubMed Central PMCID: PMC5920075.

2. Jones CM, Lee Y, Collier T, **Pringle JC**, Kitchen A, Stevenson JC, Coetzee M, Muleba M, Yamasaki Y, Cornel AJ, Norris DE, Carpi G. Mitochondrial genomics

of *Anopheles funestus* in Southern and Central Africa. Scientific Reports. Accepted: 2018 May 30.

3. Das S, Muleba M, Stevenson JC, **Pringle JC**, Norris DE. Beyond the entomological inoculation rate: characterizing multiple blood feeding behavior and *Plasmodium falciparum* multiplicity of infection in *Anopheles* mosquitoes in northern Zambia. Parasites & Vectors. 2017 Jan 26;10(1):45. PubMed Central PMCID: PMC5267472.

4. Stevenson JC, Pinchoff J, Muleba M, Lupiya J, Chilusu H, Mwelwa I, Mbewe D, Simubali L, Jones CM, Chaponda M, Coetzee M, Mulenga M, **Pringle JC**, Shields T, Curriero FC, Norris DE. Spatio-temporal heterogeneity of malaria vectors in northern Zambia: implications for vector control. Parasites & Vectors. 2016 Sep 21;9(1):510. PubMed Central PMCID: PMC5031275.

5. Stevenson JC, Simubali L, Mbambara S, Musonda M, Mweetwa S, Mudenda T, **Pringle JC**, Jones CM, Norris DE. Detection of *Plasmodium falciparum* infection in *Anopheles squamosus* (Diptera: Culicidae) in an area targeted for malaria elimination, Southern Zambia. Journal of Medical Entomology. 2016 Nov;53(6):1482 – 1487. PubMed Central PMCID: PMC5106822.

6. Morgan-Smith M, Wu Y, Zhu X, **Pringle J**, Snider WD. GSK-3 signaling in developing cortical neurons is essential for radial migration and dendritic orientation. Elife. 2014 Jul 29;3:e02663. PubMed Central PMCID: PMC4109311.

Manuscripts Submitted and in Preparation

1. **Pringle JC,** Tessema S, Chen A, Carpi G, Wesolowski A, Murphy M, Shields TM, Hamapumbu H, Searle KM, Kobayashi T, Katowa B, Musonda M, Stevenson JC, Thuma PE, Greenhouse B, Moss WJ, Norris DE. Genetic evidence of focal *Plasmodium falciparum* transmission in a pre-elimination setting in Southern Province, Zambia. *Submitted: Journal of Infectious Diseases, August 8, 2018.*

2. Jones CM, Lee Y, Carpi G, Collier TC, Hanemaaijer MJ, Lanzaro GC, **Pringle JC,** Braak L, Cornel AJ, Norris DE. Genomic divergence of *Anopheles funestus* supports sweep-before-differentiation model. *In preparation.*

Works in Progress

1. **Pringle JC**, Carpi G, Kobayashi T, Mulenga M, Muleba M, Chaponda M, Bobanga T, Juliano J, Meshnick S, Moss WJ, Norris DE. Documenting malaria transmission across the Zambia-Democratic Republic of the Congo border through *Plasmodium falciparum* genotyping.

PROFESSIONAL PRESENTATIONS

1. **Julia C. Pringle**, Sofonias Tessema, Anna Chen, Giovanna Carpi, Timothy M. Shields, Harry Hamapumbu, Kelly M. Searle, Tamaki Kobayashi, Ben Katowa, Michael Musonda, Jennifer C. Stevenson, Philip E. Thuma, Bryan Greenhouse, William J. Moss, Douglas E. Norris. Microsatellite genotyping reveals focal *Plasmodium falciparum* transmission patterns in a pre-elimination setting in Southern Province, Zambia. *Genetic Epidemiology of Malaria Conference,* 2018; Hinxton, UK. (Oral presentation)

2. **Julia C Pringle**, Giovanna Carpi, Jacob Almagro-Garcia, Sha Joe Zhu, Tamaki Kobayashi, Modest Mulenga, Thierry Bobanga, Mike Chaponda, William J Moss, and Douglas E Norris. Malaria parasite population genetic diversity: a threat to RTS,S/AS01 vaccine efficacy. *Johns Hopkins Vaccine Initiative: Vaccine Day Poster Session*, 2018; Baltimore, MD. (Poster presentation)

3. **Julia C. Pringle**, Tamaki Kobayashi, Giovanna Carpi, Steven Meshnick, Jonathan Juliano, Modest Mulenga, Mike Chaponda, Mbanga Muleba, Thierry Bobanga, William J Moss, Douglas E Norris. Evaluating Cross-Border Malaria Transmission between Zambia and the Democratic Republic of Congo: A Parasite Genetics Approach. *66th Annual Meeting of the American Society of Tropical Medicine and Hygiene,* 2017; Baltimore, MD. (Poster presentation)

4. **Julia C. Pringle**, Tamaki Kobayashi, Giovanna Carpi, Steven Meshnick, Jonathan Juliano, Modest Mulenga, Mike Chaponda, Mbanga Muleba, Thierry Bobanga, William J. Moss, Douglas E. Norris. Amplicon Deep Sequencing Suggests Cross-Border Malaria Transmission between Zambia and the Democratic Republic of Congo *Young Malaria Investigators Conference*, 2017; Baltimore, MD. (Poster presentation)

5. **Julia C. Pringle**, Tamaki Kobayashi, Steven Meshnick, Stephanie Doctor, Douglas Norris, William J. Moss. A Parasite Genetics Approach to Evaluate Malaria Transmission Dynamics in Zambia. *65th Annual Meeting of the American Society of Tropical Medicine and Hygiene*, 2016; Atlanta, GA. (Poster presentation)

6. **Julia C. Pringle**, Kelly M. Searle, Giovanna Carpi, Christine M. Jones, Tamaki Kobayashi, Douglas E. Norris, William J. Moss, for the Southern African International Center for Excellence in Malaria Research. A genomics approach to track malaria transmission in different epidemiologic settings in Zambia. *Maryland Genetics, Epidemiology and Medicine Training Program: Genetics Research Day Poster Session*, 2017; Baltimore, MD. (Poster presentation)

7. **Julia C. Pringle**, Giovanna Carpi, Douglas E. Norris, William J. Moss. *Plasmodium falciparum* Population Genetics on a Fine Spatial Scale. *Young Malaria Investigators Conference*, 2016; Rockville, MD. (Poster presentation)

8. **Julia C. Pringle**, Ben Katowa, Kelly Searle Tamaki Kobayashi, Mwiche Siame, Mike Chaponda, Modest Mulenga, Mbanga Muleba, Douglas E. Norris, William J. Moss. Evaluating *Plasmodium falciparum* Genetic Diversity in Nchelenge District, Zambia. *Johns Hopkins Center for Global Health: Global Health Day Poster Session*, 2016; Baltimore, MD. (Poster presentation)

9. **Julia C. Pringle**, Smita Das, Tamaki Kobayashi, Kelly Searle, Ben Katowa, Mwiche Siame, Mike Chaponda, Modest Mulenga, Mbanga Muleba, Douglas E. Norris, William J. Moss. *Plasmodium falciparum* Genetic Diversity Pre and Post Vector Control Campaigns in Nchelenge District, Zambia. *64th Annual Meeting of the American Society of Tropical Medicine and Hygiene*, 2015; Philadelphia, PA. (Poster presentation)

10. **Julia C. Pringle**, Smita Das, Tamaki Kobayashi, Kelly Searle, Ben Katowa, Mwiche Siame, Mike Chaponda, Modest Mulenga, Mbanga Muleba, Douglas E. Norris, William J. Moss. *Plasmodium falciparum* Genetic Diversity Pre and Post Vector Control Campaigns in Nchelenge District, Zambia. *Young Malaria Investigators Conference,* 2015; Rockville, MD. (Poster presentation)

ABSTRACTS

1. Jones, CM, Y Lee, T Collier, **JC Pringle**, JC Stevenson, M Coetzee, M Muleba, Y Yamasaki, AJ Cornel, DE Norris and G Carpi for the Southern Africa International Centers of Excellence for Malaria Research. 2017. *Anopheles funestus* in central and southern Africa: analysis of mitochondrial diversity. *66th Annual Meeting of the American Society of Tropical Medicine and Hygiene*, 2017; Baltimore, MD. (Poster presentation)

2. Carpi, G, **JC Pringle**, M Muleba, JC Stevenson, M Chaponda, M Mulenga, WJ Moss and DE Norris for the Southern Africa International Centers of Excellence for Malaria Research. 2017. Within-vector parasite diversity: insights from *Plasmodium falciparum* deep whole-genome sequencing from field-caught mosquitoes in northern Zambia. *66th Annual Meeting of the American Society of Tropical Medicine and Hygiene*, 2017; Baltimore, MD. (Poster presentation)

3. Searle, KM, **JC Pringle**, H Hamapumbu, M Musonda, B Katowa, T Kobayashi, JC Stevenson, DE Norris, PE Thuma and WJ Moss for the Southern Africa International Centers of Excellence for Malaria Research. 2017. Evaluating the efficiency of reactive case detection to achieve malaria elimination in rural

southern Zambia using follow-up household visits and parasite genotyping. *66th Annual Meeting of the American Society of Tropical Medicine and Hygiene*, 2017; Baltimore, MD. (Poster presentation)

4. Jones, CM, G Carpi, Y Lee, T Collier, **JC Pringle**, JC Stevenson, M Coetzee, M Muleba, Y Yamasaki, AJ Cornel, and DE Norris. 2016. Mitogenomics of *Anopheles funestus* in Southern Africa. *Young Malaria Investigators Conference*, 2017; Rockville, MD. (Poster presentation)

5. Giovanna Carpi, Christine M. Jones, **Julia C. Pringle**, Jennifer Stevenson, Douglas E. Norris. Vector Biology, Ecology and Genetics – Malaria Transmission and Control. *World Malaria Day*, 2016; Baltimore, MD. (Poster presentation)

HONORS AND AWARDS

1st Place: Poster Competition 04/27/2018
Johns Hopkins Vaccine Initiative: 10th Annual Vaccine Day
Johns Hopkins Bloomberg School of Public Health- Baltimore, MD

Martin Frobisher Fellowship Recipient 03/23/2018
Department of Molecular Microbiology and Immunology
Johns Hopkins Bloomberg School of Public Health- Baltimore, MD

Pre-Doctoral Fellowship Recipient 11/16/2017
Johns Hopkins Malaria Research Institute
Johns Hopkins Bloomberg School of Public Health- Baltimore, MD

Molecular and Cellular Basis of Infectious Disease Training Grant Recipient 2016, 2017, 2018
Department of Molecular Microbiology and Immunology
Johns Hopkins Bloomberg School of Public Health- Baltimore, MD

Summer Institute in Statistical Genetics Travel Scholarship 03/21/2017
Department of Biostatistics
University of Washington School of Public Health- Seattle, WA

Graduate Research Fellowship: Honorable Mention 04/2016
National Science Foundation

Dr. Lloyd and Mae Rozeboom Scholarship 2015, 2016, 2017
Department of Molecular Microbiology and Immunology
Johns Hopkins Bloomberg School of Public Health- Baltimore, MD

Global Health Established Field Placements Travel Award 03/23/2015
Center for Global Health

Johns Hopkins Bloomberg School of Public Health- Baltimore, MD

Self-Designed Summer Internship Funding Award 04/2011
Center for Peace and Global Citizenship
Haverford College- Haverford, PA

Magill-Rhoads Scholarship 08/2008-05/2012
Haverford College- Haverford, PA

JOB RELATED TRAINING

Workshop: Summer Institute in Statistical Genetics 7/12/17-7/21/17
Location: University of Washington School of Public Health: Department of Biostatistics, Seattle, Washington
Description: I received a travel scholarship to attend three intensive courses at the Summer Institute in Statistical Genetics. The courses covered topics in population genetics, molecular epidemiology, and phylogenetics. I participated in analytical workshops covering BEAST (Bayesian evolutionary analysis by sampling trees) software implementation for phylogenetic analyses, with an emphasis on viral pathogens. Similarly, I attended population genetic analysis workshops using R statistical packages.

Malaria Single Nucleotide Polymorphism SNP Barcode Workshop 06/11/2015-06/20/2015
Location: Harvard T. H. Chan School of Public Health, Boston, Massachusetts
Description: I applied for and was accepted as a participant in the inaugural Malaria Single Nucleotide Polymorphism (SNP) barcode workshop hosted by the Broad Institute and the Harvard T.H. Chan School of Public Health. During the workshop, I processed human blood samples in the laboratory in order to genotype malaria parasites detected in the blood using a novel method developed at the Broad Institute. I learned how to analyze SNP genotyping data from malaria parasites to draw inferences about malaria transmission dynamics. I attended lectures on population genetic theory and the utility of genetics to track and monitor the spread of pathogens with the goal of informing public health decision making.

ACADEMIC AND COMMUNITY SERVICE

Conference Co-Organizer-4th Annual Future of Malaria Research Symposium 05/2018-Present
Johns Hopkins Malaria Research Institute
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe St.
Baltimore, MD 21205, USA
Hours per week: 3
Duties, Accomplishments and Related Skills:

- Selected as a co-organizer of the 4th Annual Future of Malaria Research Symposium in Rockville, Maryland on October 26, 2018.
- Oversee logistical operations for a conference with 150 expected participants.
- Coordinate keynote speaker invitations and abstract selections.

Supervisor's name: Genevieve Williams (genevieve.williams@jhu.edu)
Okay to contact this supervisor: Yes

Haverford College Alumni Admissions Volunteer- Baltimore, MD 08/2012-Present
Haverford College
370 Lancaster Ave.
Haverford, PA 19041, USA
Hours per week: As needed
Duties, Accomplishments and Related Skills:

- Interview and evaluate prospective students on behalf of Haverford College.
- Generate reports about interviewees' ambitions and attributes that would make them well-suited for Haverford College.
- Represent Haverford College at local college fairs.

Supervisor's name: Amy Abolafia (aabolafi@haverford.edu)
Okay to contact this supervisor: Yes

Next Scholars Mentor 03/2017-03/2018
New York Academy of Sciences
7 World Trade Center
250 Greenwich Street, 40th floor
New York, NY 10007-2157, USA
Hours per week: 1
Duties, Accomplishments and Related Skills:

- Communicated regularly with a mentee in order to promote and support college-aged women pursuing careers in science, technology, engineering, and mathematics (STEM).
- Monitored a mentee's progress through the Next Scholars Program training modules.
- Advised a mentee about graduate school applications, networking, and job searching in STEM fields.

Supervisor's name: Julie Nadel (Jnadel@nyas.org)
Okay to contact this supervisor: Yes

Books Through Bars Volunteer Coordinator 08/2011-05/2012
4722 Baltimore Ave
Philadelphia, PA 19143, USA
Hours per week: 3
Duties, Accomplishments and Related Skills:

- Led the Books Through Bars Volunteer program at Haverford College.

- Organized and coordinated biweekly volunteer trips of Haverford students to the Books Through Bars headquarters in Philadelphia.
- Responded to inmates' requests for reading material by sending book packages compiled from donated materials.

Supervisor: Marilou Allen
Okay to contact this supervisor: No; deceased.

Mentoring and Student Teaching (MAST) High School Science Tutor 01/2010-05/2010
Haverford College
370 Lancaster Ave.
Haverford, PA 19041, USA
Hours per week: 5
Duties, Accomplishments and Related Skills:

- Supervised a group of three high school students with the execution of weekly science experiments including aspirin synthesis, a shark dissection, and rocket launching.
- Mentored students to understand scientific concepts underlying their weekly project.
- Supported students as they prepared presentations to showcase their work at the end of the program.

Supervisor: Kate Heston (6108961369)
Okay to contact this supervisor: Yes

SKILLS AND RELEVANT COURSEWORK

Programming and Statistical Languages: R, Stata, Linux Command Line Interface

Mapping and Spatial Analysis Software: ArcGIS, QGIS, SaTScan

Genetic Analysis Software: Geneious, Structure, DnaSP, BEAST, MEGA

Model Organisms: *Plasmodium falciparum, Aedes aegypti, Anopheles gambiae, Mus musculus, Drosophila melanogaster, Escherichia coli, Caenorhabditis elegans*

Laboratory Skills:
DNA:
Extraction
Purification
Quantification
PCR
Quantitative PCR (qPCR)
Genotyping methods:
Taqman
High resolution melt SNP calling
Capillary electrophoresis microsatellite analysis

Sanger sequencing
Next Generation Sequencing:
DNA library preparations
Amplicon deep sequencing
Whole genome sequencing
Hybrid capture target enrichment
Selective whole genome amplification
In vitro techniques
Tissue culture
Immunohistochemistry
Western blot
Microscopy
Immunofluorescence microscopy
Confocal microscopy
In vivo techniques
Profusions (mice)
Necropsy (mice)
Tissue sectioning
Infectious blood-feeding (mosquitoes)

Relevant Coursework

YEAR	ANALYTICAL COURSEWORK	INSTITUTION
2012	Statistical Methods with R	Haverford College
2010	Epidemiology and Global Health	Haverford College
2015	Epidemiological Methods I	Johns Hopkins Bloomberg School of Public Health
2015	Epidemiological Methods II	Johns Hopkins Bloomberg School of Public Health
2016	Epidemiological Methods III	Johns Hopkins Bloomberg School of Public Health
2015	Statistical Methods in Public Health I	Johns Hopkins Bloomberg School of Public Health
2015	Statistical Methods in Public Health II	Johns Hopkins Bloomberg School of Public Health
2016	Statistical Methods in Public Health III	Johns Hopkins Bloomberg School of Public Health
2017	Statistical Methods in Public Health IV	Johns Hopkins Bloomberg School of Public Health
2016	Stata Programming	Johns Hopkins Bloomberg School of Public Health
2016	Spatial Analysis I: ArcGIS	Johns Hopkins Bloomberg School of Public Health
2017	Introduction to Risk Sciences	Johns Hopkins Bloomberg School of Public Health

2017	Population Genetic Data Analysis	Summer Institute in Statistical Genetics: University of Washington
2017	Advanced Population Genetics	Summer Institute in Statistical Genetics: University of Washington
2017	Evolutionary Dynamics and Molecular Epidemiology of Viruses	Summer Institute in Statistical Genetics: University of Washington
2017	Statistical Machine Learning: Methods, Theory, and Applications	Johns Hopkins Bloomberg School of Public Health
	IMMUNOLOGY & INFECTIOUS DISEASE	
2010	Immunology	Haverford College
2011	Bacterial Pathogenesis	Haverford College
2014	Principles of Immunology I	Johns Hopkins Bloomberg School of Public Health
2014	Principles of Immunology II	Johns Hopkins Bloomberg School of Public Health
2014	Fundamentals of Virology	Johns Hopkins Bloomberg School of Public Health
2014	Biology of Parasitism	Johns Hopkins Bloomberg School of Public Health
2015	Pathogenesis of Bacterial Infections	Johns Hopkins Bloomberg School of Public Health
2015	Vector Biology and Vector Borne Diseases	Johns Hopkins Bloomberg School of Public Health
2015	Infectious Diseases and Child Survival	Johns Hopkins Bloomberg School of Public Health
2016	Vaccine Development and Application	Johns Hopkins Bloomberg School of Public Health
	PUBLIC HEALTH	
2012	Violence and Public Health	Haverford College
2012	Epidemics, Plagues, and Diseases in History	Haverford College
2015	Introduction to Emergency Preparedness	Johns Hopkins Bloomberg School of Public Health
2015	Introduction to International Health	Johns Hopkins Bloomberg School of Public Health
2015	Implementation and Sustainability of Community Based Health Programs	Johns Hopkins Bloomberg School of Public Health
2018	Global Disease Control Programs and Policies	Johns Hopkins Bloomberg School of Public Health
	CELL AND MOLECULAR BIOLOGY	

2010	Advanced Genetic Analysis	Haverford College
2014	Molecular Biology and Genomics	Johns Hopkins Bloomberg School of Public Health
2015	Cell Structure and Dynamics	Johns Hopkins Bloomberg School of Public Health