UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Genetic diversity of human malaria parasites associated with continued malaria transmission in pre-elimination settings in South Africa

by

Hazel Beverly Gwarinda

Submitted in partial fulfilment of the requirements for the degree
Philosophiae Doctor in Biochemistry

In the Faculty of Natural and Agricultural Sciences
Department of Biochemistry, Genetics and Microbiology

UNIVERSITY OF PRETORIA
Pretoria, South Africa

October 2021

DECLARATION OF ORIGINALITY

I, **Hazel Beverly Gwarinda (nee Makowa)**, declare that the thesis, which I hereby submit for the degree ***Philosophiae Doctor*** in the Department of Biochemistry, Genetics and Microbiology at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

ETHICS STATEMENT

The author, whose name appears on the title page of this thesis, has obtained, for the research described in this work, the applicable research ethics approval.

The author declares that she has observed the ethical standards required in terms of the University of Pretoria's Code of Ethics for Researchers and the Policy guidelines for responsible research.

Signature: 

Student name: Hazel Beverly Gwarinda

Month Year: October 2021

PLAGIARISM DECLARATION

UNIVERSITY OF PRETORIA
FACULTY OF NATURAL AND AGRICULTURAL SCIENCES
DEPARTMENT OF BIOCHEMISTRY, GENETICS AND MICROBIOLOGY

Full name: **Hazel Beverly Gwarinda** Student number: **u16346476**

Title of work: **Genetic diversity of human malaria parasites associated with continued malaria transmission in pre-elimination settings in South Africa**

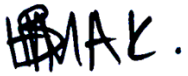
Declaration

1. I understand what plagiarism entails and am aware of the University's policy in this regard.

2. I declare that this thesis is my own, original work. Where someone else`s work has been used (whether from a printed source, the internet, or any other source) due acknowledgement was given and reference was made according to the departmental requirements.

3. I did not make use of another student's previous work and submit it as my own.

4. I did not allow and will not allow anyone to copy my work with the intention of presenting it as his own or her work.

SIGNATURE:  **DATE: OCTOBER 2021**

ACKNOWLEDGEMENTS

Most importantly, I would like to start by thanking God Almighty for giving me the strength and courage to endure this journey on both the good and bad days. His continued love, support and divine intervention made waking up each morning to pursue my dream all the easier.

I am indebted to my supervisor Prof. Lyn-Marie Birkholtz who has supported me throughout my PhD journey and altered the course of my life in many positive ways. She is my academic parent, and over the course of my PhD journey has become my personal role model. From the onset of my application to join her research group, she understood the challenging journey I had decided to partake as a female in science and accepted me despite also knowing that I was a wife and mother to a then 1 year old boy. None-the-less, she supported me unreservedly and provided me with equal opportunities to other group members, whilst constantly reminding me when the chips were down, to not be afraid of failure and that if at all I did fail at some point, to learn from the failure and continue to grow. I constantly had to remind myself of this during the Thesis writing process. Through her, I saw that with a good support structure around you, you can still be a good mother, a good wife and still chase after your dreams and be the best version of yourself. This inspiration made me keep pushing myself to the best of my ability.

Prof. Birkholtz afforded me the opportunity to visit the University of California, San Francisco in United States of America where I had the privilege of being mentored by infectious diseases expert Prof. Bryan Greenhouse and Dr. Sofonias Tessema for a month and learnt laboratory and data analysis techniques which I then used in my PhD project. I also had the opportunity to meet other world prominent leaders in science whilst there. For this exposure, I am entirely grateful to Prof. Greenhouse who took on the task of hosting and mentoring me there. I would also like to thank and acknowledge my co-supervisor Dr. Jaishree Raman for her invaluable input in facilitating access to study samples and communications with the Limpopo Department of Health. Her guidance and support during this project, right through to seminars, conferences and completion of this thesis was very instrumental.

I thank my family - my loving husband Tafara for the continued support and patience, as well as my loving son, Tomuvonga, for allowing me to take this journey with him even without asking for his permission. A huge thank you again goes to my loving mother Gloria, brothers and in-laws for their prayers, support and being my biggest cheer leaders.

It was a great pleasure and joy working with my wonderful colleagues and lab mates at the University of Pretoria, Malaria Parasite Molecular Laboratory who created a friendly, fun working environment, and always gave valuable feedback during presentations and in meetings. Special mention to Suzan, Masego, Phanankosi, Dikeledi, Jessica, Janie, Mariska, Sindisiwe and Meta who also became very good friends even outside the lab.

I would like to express my gratitude to the study population in Vhembe District, Limpopo Province, as well as the Folovhodwe, Madimbo, Manenzhe, Masisi, Mulala, Makuya and Tshipise clinic and laboratory staff. I would like to acknowledge Mr Eric Mabunda and health information officers from the Limpopo Department of Health, for facilitating access to patient information through the Limpopo Malaria Case (LMC) database. The University of Pretoria Institute for Sustainable Malaria Control acknowledges the South African Medical Research Council as Collaborating Centre for Malaria Research.

This project was supported by a bursary from the National Research Foundation (NRF) – The World Academy of Science (TWAS). The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily attributed to the NRF.

SUMMARY

South Africa is targeting malaria elimination (halting malaria transmission within the country's border) by 2023. However, hotspots of stable residual malaria transmission pose a threat to this goal. Whilst several factors associated with continued residual malaria transmission have been investigated, the contribution of parasite genetic diversity in these hotspots and across the southern African region in sustaining transmission has not been critically evaluated. Approaches to eliminate malaria in the country therefore require adoption of novel tools and strategies such as parasite population genetics tools to support better surveillance of the disease in aid of reaching elimination. These tools have successfully been used in eradication programs of other diseases such as polio where it has been shown that genetic diversity influences transmission of these diseases, and their use is gaining momentum in understanding malaria transmission dynamics. Thus, this PhD project aimed to use such tools to evaluate the genetic diversity of *Plasmodium falciparum* parasites from malaria hotspots in South Africa compared to regionally transmitted parasites to identify contributing factors associated with sustained transmission.

To address the existing knowledge gaps around *P. falciparum* population genetics in South Africa, this thesis characterised the within-host and population level diversity of *P. falciparum* parasite populations from the Vhembe District, Limpopo Province, at different spatial and temporal scales (Chapter 2). Consistent with the Vhembe District's classification as a stable 'high' transmission setting within South Africa, *P. falciparum* diversity was moderate to high and complex, with less pronounced population structure - all indicative of constant transmission at relatively moderate levels. Whilst this level of genetic diversity in Limpopo was attributed mostly to residual local transmission, in Kwa-Zulu Natal (KZN) Province, a significantly lower transmission setting, similar levels of parasite genetic diversity were attributed to imported infections as indicated in the comparative analysis in Chapter 3. This heterogeneity in transmission settings and parasite populations will therefore have significant implications for malaria control strategies in South Africa and will require differentiation in our approaches to reach our malaria elimination goals. This outcome therefore necessitated comparison of the South African parasite population to regional parasite genotypes to understand *P. falciparum* transmission dynamics in the southern African region.

As such, a meta-analysis on southern African parasite populations (Chapter 4) was conducted. A total of 5314 samples were analysed, with microsatellite genotyping data

collected from studies conducted in Namibia, Eswatini, South Africa and Mozambique. Noteworthy, this thesis revealed that the parasite population was structured and that there was an endogenous circulation of parasites in each of these countries. From an elimination perspective, in a low-transmission elimination setting such as Eswatini, for example, this shows that although imported infections play a significant role in continued transmission, there is also a contribution of locally acquired/generated parasites (parasites of local/internal origin) circulating within the individual countries preventing those countries from "getting to zero" (local) infections. Furthermore, there was some evidence of genetic connectivity of parasites between the different endemic countries, likely due to human migration. This suggests that malaria transmission may be sustained in the region, thus hindering success of control measures and progress towards elimination.

Taken collectively, this thesis addresses relevant knowledge gaps in our understanding of the parasite's contribution to malaria transmission dynamics in South Africa and the southern Africa region on a population genetics level. The data presented here may contribute to better decision making on a national level and for southern Africa as a regional block on what interventions should be put in place, and concentrated in which areas, with the goal of achieving national and regional malaria elimination.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACTs	arteminisinin-based combination therapies
AL	artemether-lumefantrine
AMOVA	analysis of molecular variance
ANOVA	analysis of variance
BIC	Bayesian Information Criterion
CHWs	community health care workers
CI	confidence interval
COVID-19	corona virus disease 2019
crt	chloroquine resistance transporter
CSP	circumsporozoite surface protein
DA	discriminant analysis
DAPC	discriminant analysis of principal components
DBS	dried blood spot
DDT	dichloro-diphenyl-trichloroethane
dhfr	dihydrofolate reductase
dhps	dihydropteroate synthase
DNA	deoxyribonucleic acid
E8	Elimination 8 Initiative
EHPs	environmental health practitioners
EIR	entomological inoculation rate
F_{ST}	Wright's F-statistics
F_{WS}	within-host fixation index
G6PD	gloucose-6-phosphate dehydrogenase
GAC	genotype accumulation curve
GLURP	glutamine rich protein
GMEP	Global Malaria Eradication Programme
He	heterozygosity

IAM	infinite alleles model
IBS	identity by state
IDC	intra-erythrocytic developmental cycle
IRS	indoor residual spraying
ISA	standardised index of association
ITNs	insecticide treated bed nets
LD	linkage disequilibrium
LD1/2	linear discriminant 1/2
LLITNs	long lasting insecticide treated bed nets
LMC	Limpopo malaria case
KZN	KwaZulu-Natal
MCMC	Markov chain Monte Carlo
mdr	multidrug resistance drug-protein 1
MIPs	molecular inversion probes
MLGs	multi-locus genotypes
MMS	malaria molecular surveillance
MOI	multiplicity of infection
MOSASWA	Mozambique, South Africa and Eswatini
MSP-1/2	merozoite surface proteins 1/2
N_e	effective population size
NGS	next-generation sequencing
NICD	National Institute for Communicable Disease
NMCPs	national malaria control programmes
PBS	phosphate buffered saline
PC	principal component
PCA	principal component analysis
PCR	polymerase chain reaction
pfhrp2/3	*Plasmodium falciparum* histidine-rich proteins 2/3

qPCR	quantitative PCR
RDTs	rapid diagnostic tests
RMSE	root mean squared error
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SMM	stepwise mutation model
SNPs	single nucleotide polymorphisms
SP	sulfadoxine/pyrimethamine
TARE-2	Telomere associated repetitive element -2
varATS	*var* gene acidic terminal sequence
WGS	whole-genome sequencing
WHO	World Health Organization

CHAPTER 1

LITERATURE REVIEW

1.1 Global burden of malaria and progress towards elimination

Globally, malaria remains one of the most life-threatening and highest-burden infectious diseases, with an estimated 229 million cases and 409 000 deaths reported in 2019 according to the World Health Organization (WHO) 2020 Malaria Report [1]. Approximately 85 % of all malaria-associated mortalities occur in sub-Saharan Africa, with children under the age of 5 years and pregnant women most at risk [2].

A century ago, malaria was found on every continent except for the arctic regions. However, by 1970 the disease had been eliminated (transmission within the defined area stopped) from most of the world's developed countries [3]. Today, 107 countries are malaria-free, and of the 91 remaining countries with on-going malaria transmission [1], 25 were identified by the WHO as having the potential to eliminate by 2025, 3 of which (South Africa, Botswana and Eswatini) belong to the southern African region [3-5]. A global reduction of up to 23 million malaria cases was reported in 2018 since 2010, attributable to sustained scale-up of control interventions [2]. However, this decline has not been uniform within and between countries, and in spite of the significant headway made, progress towards malaria elimination has stalled, with resurgence in cases experienced in most parts of the world during a global outbreak in 2017 [2].

Malaria is caused by protozoa of the genus *Plasmodium* and is transmitted by specific species of female *Anopheles* mosquitoes [6]. Amongst the six Plasmodium species (*Plasmodium falciparum, Plasmodium vivax*, *Plasmodium ovale curtisi, Plasmodium ovale wallikeri, Plasmodium malariae* and *Plasmodium knowlesi* [7]) that cause malaria in humans, *P. falciparum* is the most dominant and virulent species [8]. This species which will be the focus of this thesis is also associated with most of the malaria-related morbidity and mortality in southern Africa [8, 9], while in most countries outside of sub-Saharan Africa, *P. vivax* is the dominant malaria parasite [8]..

1.2 The *Plasmodium falciparum* life cycle

Human infections with the *P. falciparum* parasite start with a single bite of an infected female Anopheline mosquito. As summarised in Figure 1.1, during the mosquito's blood meal, forms of the parasite called sporozoites are introduced into the human host from the mosquito's salivary glands [10, 11]. When they enter the human host's bloodstream, they invade the liver cells known as hepatocytes within 30 to 60 min post-introduction, and they replicate and divide to produce tens of thousands of daughter merozoites [12], which are released into the bloodstream to invade uninfected erythrocytes where they begin to reproduce asexually. While sporozoite invasion and the liver stages last up to approximately 6 days, they do not cause disease. All clinical symptoms are caused by the parasite during its intra-erythrocytic, asexual developmental cycle (IDC). During the IDC, the parasite develops into three main stages: rings, trophozoites and multinucleated schizonts (Figure 1.1) [13]. Asexual schizogony causes erythrocytes to rupture, releasing newly formed daughter merozoites, which, again then infect uninfected erythrocytes and start another cycle of maturation and replication. This process happens several times leading to an increase in the number of parasites circulating in the infected human host. This cycle takes up to 48 hours and it is during this time that the clinical symptoms of malaria, which are associated with the synchronised/timely rupture of infected erythrocytes manifest. Symptoms of uncomplicated malaria caused by infection with the parasite are characterised by cyclic patterns of fevers and chills, while symptoms of severe disease are variable, ranging from impaired consciousness, shock, acute renal failure, and severe anaemia just to name a few [1, 14, 15].

A small proportion of asexual parasites transform/differentiate into sexual forms known as gametocytes, with five different stages (stages I-V) observed for *P. falciparum*. Once mature, stage V gametocytes are taken up during another of the mosquito's blood meals, and the gametocytes differentiate to male and female gametes. These gametes then recombine to form zygotes that develop into ookinetes that enter the mid gut wall and develop into oocytes that mature into sporozoites that migrate to the mosquitos salivary glands and start the parasite's life cycle again. This stage of the cycle takes between 8 to 15 days in the mosquito [16, 17].

Malaria is a completely preventable disease where either the mosquitoes that transmit the disease to humans are targeted and killed, or the pathogenic parasites are killed.

However, it is a very complex disease because while it is "caused" by the *Plasmodium* parasite, emergences that lead to it being epidemic or pandemic are also significantly determined by other factors related to the host (human and/or mosquito) and to host-environmental interactions.

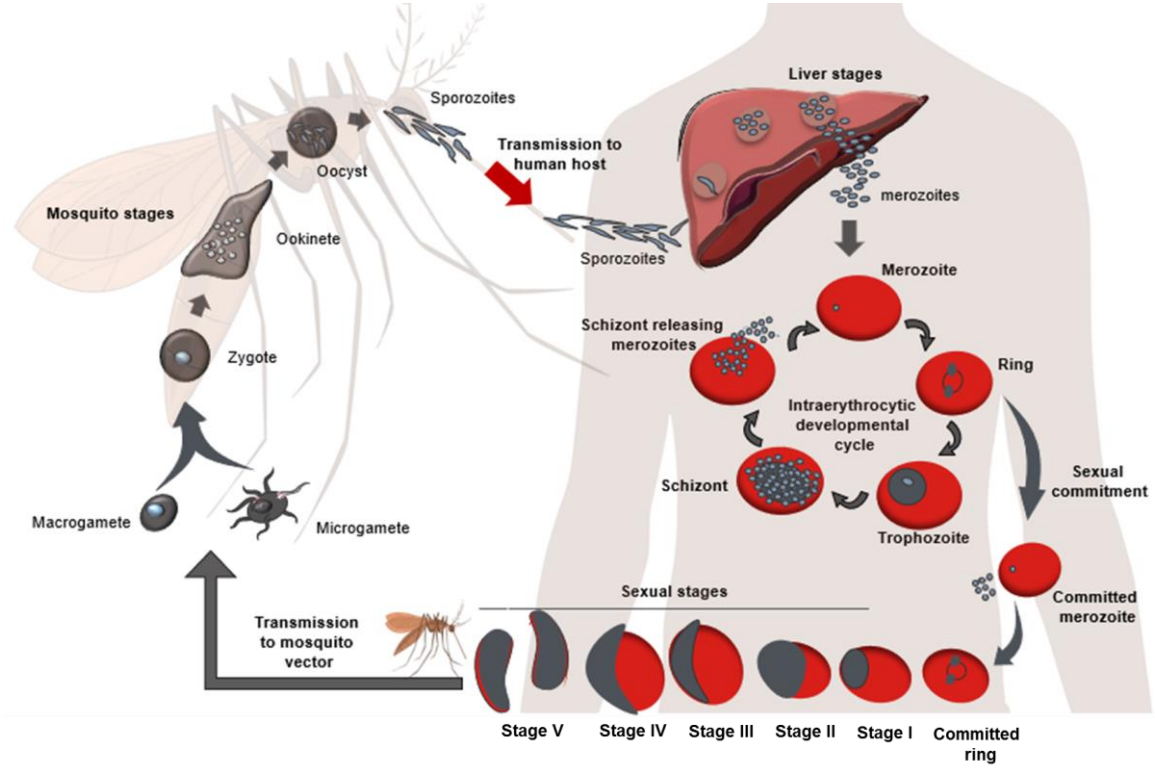


Figure 1.1 The *Plasmodium falciparum* life cycle. This figure describes the parasite's life cycle inside the human host where it undergoes asexual development in the liver and erythrocytes after sporozoites from a female Anopheles mosquito are introduced into the human host during a blood meal by the mosquito. A portion of the parasites then differentiates into sexual forms before being taken up again by the mosquito to undergo recombination and sexual development thus re-initiating the cycle.

1.3 Factors that affect malaria transmission

A variety of human, vector and environmental factors influence the distribution and intensity of malaria transmission. A change in any factor has the potential to impact the risk of infection profoundly. The intensity of malaria transmission in different countries and regions around the world varies based on local factors such as rainfall patterns as well as vector species and their proximity to breeding sites just to name a few [18, 19]. As such, the WHO guidelines on malaria elimination (halting malaria transmission within the country's border) acknowledge that malaria transmission intensity ranges from high to low

[20]. According to the WHO, transmission intensities therefore vary from very low (parasite prevalence >0 but <1 % of population or an annual parasite incidence of <100 cases per 1000 population at risk); to low (1-10 % parasite prevalence or an annual parasite incidence of 100–250 cases per 1000 population at risk); to moderate transmission (10-35 % parasite prevalence or an annual parasite incidence of 250–450 cases per 1000 population at risk); to extremely high-transmission settings (≥35 % parasite prevalence or an annual parasite incidence of ≥450 cases per 1,000 population at risk) [20, 21].

In this section, the influence of the environment, the human host, the mosquito vector, and the pathogenic parasite on transmission efficiency will be discussed.

1.3.1 Environmental factors

The geographic distribution and seasonality of malaria is mainly influenced by climatic conditions [19]. Environmental factors such as rainfall, temperature and humidity facilitate the presence of malaria in a region [18, 19]. When it rains, pools of stagnant water where mosquitoes can breed are created. Female mosquitoes lay their eggs in these pools of water, which develop in larvae and transform into adult mosquitoes if suitable temperature and humidity conditions are present [18]. For the parasite to be transmitted, the mosquito should stay alive long enough for the parasite to complete its multiplication cycle in the mosquito, which takes anywhere from 9 to 21 days at 25 ºC [18, 22, 23]. The cycle is shortened at higher temperatures thus increasing the risk of transmission [24, 25]. Higher temperatures increase the frequency of blood meals taken and the number of eggs laid by the mosquitoes [26]. This in-turn increases the number of mosquitoes in each area. At temperatures lower than 19 ºC, the multiplication cycle is not completed for *P. falciparum* and malaria cannot be transmitted, which aids in explaining why malaria is more widespread in warmer regions of the world [24, 25]. In areas that are drier or that receive little rainfall, the rain may indirectly affect malaria transmission through its effect on humidity [24, 25]. The relative humidity of the environment increases because of vegetation cover that increases after it rains. Warm weather may also influence human behaviour for example increased outdoor activities (e.g., farming) at times when Anopheline mosquitoes are most active. This may increase the chances of contact with the mosquitoes thus increasing the chances of an individual being infected [24, 25].

1.3.2 Human factors

Certain genetic and/or biological factors such as acquired immunity [27], expression of the Duffy antigen in the local human populations [28], sickle cell anaemia and G6PD (gloucose-6-phosphate dehydrogenase) deficiency [29] play an important role in malaria transmission. In areas of very high-transmission, the majority of individuals in the population develop immunity to the disease due to frequent exposure to infection [30]. As transmission declines to low intensities, individuals are faced with less frequent exposure to infection and therefore have decreased chances of acquired immunity in adults and elder children [30]. Irrespective of transmission intensity, children under the age of 5 years and pregnant women are at a high risk of infection because of their weaker immune systems. Naturally acquired immunity plays an important role in determining parasite kinetics and transmission as it affects asexual parasite densities and thus indirectly affects the development of gametocytes and transmission intensity [27].

Based on human genetic factors, some population groups are more prone to being infected by some *Plasmodium* species but not others. While *P. falciparum* is most prevalent in sub-Saharan Africa, *P. vivax* has a wider geographic distribution and is predominantly found in countries such as Ethiopia, India and Pakistan, as well as some parts of sub-Saharan Africa [1]. This is possibly due to the ability of *P. vivax* to develop at lower temperatures in the vector [31, 32], and its ability to survive for long periods as hypnozoites (dormant liver stage). These hypnozoites form a possible reservoir of the infection due to relapsing events and are a major obstacle towards vivax malaria eradication [1, 33-35]. *P. vivax* infection has however, in some studies been linked to high frequencies of the Duffy antigen in human populations [36]. In West/Central Africa, where infection with *P. vivax* is known to be rare, this has been linked to the lack of expression of the Duffy antigen in the local human populations [28]. This Duffy antigen negativity prevents the parasite from exploiting the entry mechanism on the erythrocyte surface [28].

Other genetic factors such as sickle cell anaemia and G6PD (gloucose-6-phosphate dehydrogenase) deficiency can also be involved in the process, severity and non-severity of malaria [29]. When G6PD, an enzyme that helps to maintain balance within human cells in response to oxidative stress is deficient in humans, this causes the rupture or destruction of red blood cells (haemolysis) in the G6PD-deficient individuals [37]. This impacts efforts to prevent of *P. vivax* malaria relapses since primaquine or tafenoquine, which are antimalarial drugs recommended to be given as treatment in combination with

chloroquine or an artemisinin-based combination therapy (ACT) to prevent *P. vivax* relapses cause haemolysis in G6PD-deficient individuals, which may lead to severe anaemia and potentially the need for a blood transfusion [38]. G6PD deficiency is however believed to provide a protective effect against malaria although available data to support this is conflicting [37, 38]. On the other hand, individuals with sickle cell disease, an autosomal recessive disorder that is characterized by chronic haemolytic anaemia, have an increased risk of malaria, mostly due to impaired splenic function [39]. Sickle cell disease is most common in children from sub-Saharan Africa, and if not diagnosed early and children born with the disease do not receive appropriate care, this may lead to death in infants [39, 40].

Human movement and migration is another very important factor that shapes transmission opportunities as it aids in the movement of parasites across a wide range of geographic scales [36]. Most population movements in Africa usually involve economic migrants moving from areas high transmission. This facilitates spread of, for example, more virulent or drug resistant parasite strains [36]. Additionally depending on how far the human is travelling, this may facilitate either local and/or imported transmission in different transmission settings. Other human activities, behaviour and characteristics that influence contact with mosquitoes and hence transmission opportunities, include spending long periods of time outdoors [41] and differences in mosquitoes being attracted to the person that is influenced by for example body size [42] and odour profiles [43].

1.3.3 Mosquito vector factors

One of the most essential determinants to malaria transmission success is the mosquito vector. Malaria transmission intensity may be influenced by the type of *Anopheles* mosquito present in an area at a given time [44]. Vector species that prefer human blood and support parasite development are the most dangerous [44] as they are endowed with high vectorial capacity (the total number of potentially infectious bites that would eventually arise from all the mosquitoes biting a single perfectly infectious (i.e. all mosquito bites result in infection) human on a single day) [45]. Mosquito vectors in Africa are anthropomorphic, and have a high preference to feed on humans [46]. *Anopheles gambiae* is the most prevalent species complex that efficiently infects humans with sporozoites [47] and it is predominantly found in Africa [46]. High numbers of this vector species increase the probability of mosquito bites thus increasing the chances of disease

transmission when an individual is frequently bitten in an area where parasites are circulating [42]. This therefore increases transmission intensity in the area [44]. To kill the vectors, insecticides are used [48, 49]. However, if the mosquitoes develop resistance to the insecticides, this means that they can no longer be killed by the insecticides and, therefore, large numbers of mosquitoes will survive in communities [50]. Apart from the genetic basis of insecticide resistance, there are also behavioural adaptations such as feeding and resting patterns adopted by the mosquito to avoid insecticides [51-53]. Most *Anopheles* mosquitoes are active either at dusk, dawn, or at night, and prefer to feed and rest indoors [51-53]. To avoid exposure to insecticides through insecticide-treated bed nets (ITNs) or indoor residual spraying of insecticides, some of these mosquitoes then adopt outdoor feeding behaviour thereby increasing the risk of malaria infections [51-53].

The degree to which different human host populations are targeted by mosquitoes differs and it may play a pivotal role in determining sources of onward transmission [42]. Onward transmission is also dependent on the effective contact rates between humans and mosquitoes [42]. This makes the abundance of vectors, competent species and the frequency in which they take infected blood meals from individuals the most important factors in determining transmission success [24, 25]. The chances of successful infection can also be influenced by the ability of ookinetes to escape the mosquito's immune system, the mosquito's feeding behaviour and length/duration of the parasites' multiplication in the mosquito [54, 55].

1.3.4 Parasite factors

There are also several parasite factors that influence transmission dynamics. Some of these include the total parasite density [56, 57], gametocyte density [27] and gametocyte sex ratio [58]. The presence of competing parasites strains, the parasites' response to host immunity and environmental stress factors affecting gametocytogenesis influence the gametocyte sex ratio [58, 59]. The gametocyte sex ratio is typically female biased, with 3 to 5 times more female than male gametocytes [58]. When gametocyte densities are low, this usually results in lower rates of mosquito infections and even more reduced quantities of oocysts in infected mosquitoes. However, even with low oocyst densities, mosquitoes are infectious and capable of continuing the transmission cycle to the next human host [60]. Low-density asexual parasites in asymptomatic parasite carriers who are immune to infection, may also form an important source of onward transmission to mosquitoes and

should be considered in elimination efforts together with gametocytes that can be carried for months [61]. Furthermore, parasite genetic factors such as the overall genetic diversity of the parasite population and the number of distinct parasite clones circulating within an individual infection [62-64] can be linked to transmission intensity and will be discussed in more detail in the following sections describing parasite genetic diversity.

1.4 Malaria control and elimination

Eradicating malaria is a process which first goes through different stages of malaria control and elimination. When controlling malaria, disease incidence, prevalence, morbidity or mortality are reduced to a locally acceptable level as a result of deliberate intervention efforts [21]. Therefore, in order to sustain control, the continued implementation of interventions is required [21]. Once local transmission is interrupted (reduction to zero incidence of indigenous cases) for all human parasites as a result of these deliberate intervention efforts, certification of malaria elimination in a country can then be achieved [21]. Prior to country certification, there could be sub-national certification as is the case in the KwaZulu-Natal Province of South Africa [65, 66] with parasite genetic diversity being one of the factors that drive low level transmission in the area as described in Chapter 3 of this thesis. However, continued measures to prevent the re-establishment of transmission would still be required [21]. Only when there is a permanent reduction to zero of the worldwide incidence of infection caused by human malaria parasites as a result of the deliberate activities, can malaria then be eradicated [21]. Once eradication is achieved, interventions will no longer be required [21]. The different stages in the malaria elimination continuum are defined as; malaria control (incidence >5 malaria cases per 1000 population at risk), pre-elimination (incidence between >0 but <5 malaria cases per 1000 population at risk), elimination (incidence <1 malaria case per 1,000 population at risk) and prevention of reintroduction (0 malaria cases per 1,000 population at risk) per malaria season [20].

1.4.1 Vector control strategies and challenges for malaria elimination

Several different tools and strategies to control, eliminate and eradicate malaria by targeting the mosquito vector have been employed. Vector control strategies aim to reduce or prevent contact between the mosquito vector and the human host [67] and include the use of insecticides, long lasting insecticide treated bed nets (LLITNs),

larviciding and insecticide-based indoor residual spraying (IRS). The first large scale attempt to eradicate malaria was the Global Malaria Eradication Programme (GMEP), which ran from 1955 until 1969 [67]. Despite being called a global programme, Africa was not included in the programme because transmission intensity in sub-Saharan Africa was just too high for the intervention to lead to malaria elimination in the region in the short term. This programme was a partial success and led to malaria elimination in some regions of the world using dichloro-diphenyl-trichloroethane (DDT) as a long-lasting insecticide to kill the mosquitoes [67]. Some regions that were nearing elimination then started to experience resurgence in malaria cases because in addition to other challenges such as social and operational issues, the mosquito vectors also developed resistance to DDT thus hampering malaria elimination efforts [67]. This led to the GMEP being abandoned and post GMEP, sustained implementation of vector control interventions became the focus.

These vector control interventions were promoted through the Roll Back Malaria initiative in areas where elimination was not yet feasible. This initiative saw the use of insecticide treated bed nets (ITNs), IRS and larviciding to successfully control the mosquito vector and reduce the mortality rate in children [48, 49, 68-75]. However, these strategies are only partially effective and although they managed to eliminate malaria in countries such as China, Sri Lanka and Paraguay [76, 77], they do not eliminate malaria in high-transmission settings of Africa where coverage of interventions may be poor due to financial constraints [74, 75]; outdoor feeding and resting behaviour of some *Anopheles* vectors, which may render bed nets ineffective [68]; and resistance to available insecticides being developed by the mosquito vectors. Alternative measures such as removing the breeding grounds of the mosquito and aquatic habitats of the larval stages of the vector also exist, however they are too expensive to execute when there are many water bodies [68]. Instead of relying on a single control measure, cost-effective strategies employing integrated vector management (a decision-making process for the management of vector populations, so as to reduce or interrupt transmission of vector-borne diseases) are promoted as a better way of controlling malaria [68].

The success of a vector control program can be evaluated using different metrics such as the entomological inoculation rate (EIR), vectorial capacity and sporozoite rate [78]. The EIR, which has long been considered the gold standard metric for measuring transmission, evaluates the number of infectious mosquito bites received by an individual per year [78]. This metric can be used in various transmission settings, however it lacks

precision since distribution of mosquitoes is highly heterogeneous and the mosquito biting rates are difficult to estimate [79, 80]. The vectorial capacity assesses how well local mosquitoes transmit malaria parasites to humans and reduction in this metric was the basis for the GMEP recommended by the WHO in the 1950s. The sporozoite rate is a measure of the number of collected mosquitoes that contain sporozoites [78]. All these methods however need to be adapted to the various transmission settings. Despite being very effective, vector control interventions alone are insufficient to achieve elimination. As cases decrease, parasite detection and treatment become more important in a county's quest to achieve elimination. The results discussed in the upcoming chapters of this thesis can therefore be used to make a case for integrated parasite control which would be a decision-making process for the management of parasite populations, so as to reduce or interrupt transmission of malaria.

1.4.2 Strategies targeting the parasite and challenges for malaria elimination

In addition to vector control strategies, interventions that specifically target the parasite are essential for malaria control and elimination. Amongst these strategies, the deployment of effective antimalarial drug therapies is of paramount importance. Following its discovery in the early 1940s, chloroquine became the preferred antimalarial and was widely used to treat malaria during the GMEP [67] until resistance emerged in the 1980s [67, 81]. Chloroquine was replaced in many countries with sulfadoxine/pyrimethamine (SP). Unfortunately parasites' resistance to SP rapidly evolved [81], resulting in SP being replaced in many countries by arteminisinin-based combination therapies (ACTs). ACTs combine artemisinin derivatives with other long-lasting partner drugs such as mefloquine, lumefantrine and piperaquine with a different mode of action to reduce the chances of the parasite developing resistance [82]. Currently ACTs are the recommended WHO treatments and are the most widely employed first-line treatments against uncomplicated falciparum malaria. Although ACTs have been highly effective in reducing malaria associated cases and deaths parasites resistant to ACTs are present in the greater Mekong region in southeast Asian countries of Thailand, Vietnam, Myanmar and Cambodia [83], with some reports of artemisinin resistant parasites in Africa emerging [84-86].

Due to the complex transmission patterns in Africa, it has now become apparent that it may not be possible to eradicate malaria with the current tools alone, and novel

interventions including vaccines, are going to be required. While current vaccines cannot eliminate malaria since they can only be used in specific categories of the populations and require multiple does, they could form part of an integrated parasite control programme. The vaccine candidates currently undergoing clinical trials, target the different life cycle stages of the *P. falciparum* parasite [87]. A blood stage vaccine based on the merozoite surface antigen apical membrane antigen 1 (*AMA1*) showed high protection (64.3 %) against malaria caused by parasites homologous to the vaccine strain, but much lower protection (17.4 %) against heterologous strains in Malian children [88]. RTS,S, which is the pre-erythrocytic stage vaccine based on the circumsporozoite protein of *P. falciparum* showed an efficacy of 36.3 % in children aged 5 – 17 months old who received four doses of the vaccine [89], although its efficacy declined over time [90]. Although these current vaccines have great potential, they only provide short-term protection against malaria infection and would need to be given continuously to all members of the community if malaria control by vaccination is to be achieved, which was until recently deemed not feasible [91]. The RTS,S vaccine, however, received a positive scientific opinion from the European Medicines Agency and was recommended by the World Health Organization for malaria vaccine pilot programmes in Ghana, Malawi and Kenya [92]. Furthermore, in a recent historic and much awaited decision, the WHO on the 6th of October 2021 recommended the wide rollout of the malaria vaccine to protect children in Africa [91]. This gives countries the opportunity to decide how to use the vaccine as part of their malaria control programs [91]. Already, evidence from a study conducted in Mali and Burkina Faso revealed how the vaccine was used in combination with regular doses of antimalarial drugs just before the rainy season to lower risk of clinical and severe malaria by up to 60 % and 70 %, respectively, compared with children who only received one or the other interventions [93]. This highlights the importance of complementing new and existing tools in the fight against malaria in moderate- to high-transmission settings [93].

None-the-less, these interventions meant to kill the parasite and save lives are therefore more likely to result in a reduction in transmission intensity, but not in elimination or eradication of malaria [94, 95]. As it is evident that it is challenging to target the parasite, interventions targeting the parasite need to be continuously monitored for their effectiveness to see the effect they have on the reduction of transmission intensity.

The most widely collected metric to classify geographical regions based on malaria endemicity [74, 96, 97] is to measure the proportion of individuals who are parasite positive in a specific population, at a given time point (parasite positivity rate) [78].

Measurement of the parasite positivity rate for the assessment of transmission intensity has been shown to saturate at high-transmission intensities due to acquired immunity and multiple infections [78]. On the other hand, in low-transmission settings, where assessments of effectiveness of control measures are needed more [98, 99], this method lacks accuracy. Other commonly used methods currently rely on routine surveillance reports of active malaria cases from health care facilities. However, estimates derived from routine surveillance data are usually lower than those derived from surveys of parasite prevalence due to incomplete case detection and/or incorrect diagnosis at public and private health services [100]. Also, since traditional case detection depends on several factors such as antimalarial chemotherapeutics use, antimalarial drug resistance, acquired immunity and parasite seasonal variation, the number of circulating parasites in the bloodstream may be affected. As a result, this could underestimate transmission intensity [101].

1.5 Malaria foci and *Plasmodium* hotspots

Following the sustained implementation of effective interventions, transmission intensity decreases with malaria transmission becoming more focal and concentrated in a small proportion of individuals. This uneven distribution of malaria across the population leads to the formation of malaria foci and hotspots. This reduces the efficacy of control strategies and profiling of cases becomes important to determine the kind of intervention required [102].

Foci of malaria transmission are well defined locations containing the epidemiological and climatic factors necessary for malaria transmission to occur [103]. Within these foci, there are malaria hotspot areas, which are defined specific geographic areas or specific groups/clusters of people where transmission intensity is much higher due to highly favourable conditions [104]. Individuals within the focus usually have higher chances being infected with *P. falciparum* than the population outside the hotspot [105, 106]. Hotspots differ in size spatially, and the extent over which they can be identified and targeted differs from entire countries, to small geographical areas less than 1 km^2 [104, 107]. Once such areas are identified, it is critical that individuals be appropriately diagnosed, with special emphasis placed on determining the level of parasitaemia, and gametocytaemia as well as *Plasmodium* species present, to ensure appropriate treatment

and interruption of the transmission cycle. Untreated infections can last for several months and cause recurring clinical episodes with asymptomatic periods [108].

Microscopy and/or rapid diagnostic tests (RDTs), which are the backbone of malaria diagnosis, can give information on the number of people that have been infected at any given time. However, they both have difficulty in fully identifying mixed infections, infections with low parasitaemia, and infections that lack detectable concentrations of *Plasmodium falciparum* histidine-rich proteins 2 and/or 3 (pfhrp2/3) antigens [109, 110]. *Pfhrp2/3* gene diversity and deletion can affect the performance of pfhrp-based RDTs used in malaria-endemic regions to test for malaria thus impacting malaria control and elimination efforts [111]. Other molecular diagnostic tools such as loop-mediated isothermal amplification method (LAMP) and lateral immuno-assays are currently under development and may potentially be used for routine diagnosis and surveillance [112-114]. The WHO also recommends focal mass drug administration (defined as targeted drug administration to smaller populations of people who are at higher risk of malaria) at low transmission for the purposes of elimination precisely because it becomes difficult to diagnose and treat all cases [115-117]. Current surveillance data that inform decision making are unfortunately sometimes incomplete, of a low quality, and are unable to provide critical information that would aid in giving a better picture of the type of infection and clearer patterns of transmission.

Malaria control programs in low-transmission settings therefore require additional innovative molecular tools (genotyping) to the currently used surveillance tools, to accurately track changes in *Plasmodium* parasite populations in relation to changing transmission intensity in order to better target and tailor interventions to malaria foci and *Plasmodium* hotspots based on local epidemiology, so that the impact and response to changing transmission dynamics can be maximised [62, 118-120]. Use of these additional genotyping tools may also allow associations between malaria infections and the individuals driving transmission or so called hot-pops to be evaluated [121]. These tools would also be essential in accurately mapping the routes of parasite dispersal by human carriers as these would enable targeting of interventions to both source (where imported infections originate from) and sink (recipient areas where transmission is substantially contributed to) populations [122, 123]. Additionally, imported cases, which are a contributing factor to residual transmission in low-transmission areas, can also be accurately distinguished from local cases [62, 64, 119, 120].

Understanding the parasite genetic diversity and gene flow dynamics within and between populations through population genetics tools may therefore be an effective way for implementing these measures [62, 64, 120, 122-124]. However, in order to have successful malaria surveillance from such tools and evaluation of malaria elimination programs, data from both patient demographic records and molecular epidemiology should be merged to facilitate the identification of individuals carrying specific *Plasmodium* lineages associated with malaria hotspots [62, 64, 120, 122-124]. It is therefore important to have a good estimation of malaria incidence in countries to be able to assess the success of malaria control programs as well as to help to identify foci of residual transmission over different spatial and temporal scales [62, 64, 120, 122-124].

In order to accelerate progress towards elimination, the need to use molecular epidemiology for surveillance is addressed in one of the pillars of the WHO global technical strategy for malaria 2018-2030, which focuses on the need to transform surveillance into a core intervention [125]. This has successfully been implemented in eradication programs of other diseases by assessing the genetic epidemiology and disease surveillance for elimination of polio, outbreaks of Ebola, on-going transmission of tuberculosis [126] and most recently, in supporting the timely control of corona virus disease 2019 (COVID-19) outbreaks [127-133].

1.6 Malaria and COVID-19

At the time of writing this thesis, a new coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), had been identified. The virus, associated with severe pneumonia was first identified in Wuhan, China in December 2019 [134] and rapidly spread across the globe [134]. By March 11, 2020, the COVID-19 epidemic was declared a global pandemic by the WHO [134]. The current prevalence of the disease in Africa is much lower than expected, with some suggestion that malaria offers some protection against COVID-19 [134]. As of August 2020, out of the roughly 17 million global COVID-19 confirmed cases and close to 700 000 deaths [135], 54 countries from the African region had a cumulative total of 736 288 cases, which accounts for roughly 4 % of all global cases, and reported 15 418 deaths [134, 135]. By May 2021 however, 1.9 million COVID-19 cases and 61,649 deaths were reported in the southern Africa Elimination 8 Initiative (E8) region alone [136], and South Africa is currently the most affected as it carries the highest burden of COVID-19 cases compared to the rest of Africa [137].

The spread of COVID-19 was mainly curbed through interventions such as wearing of masks, social distancing, curfews, border closures and travel bans that prevented human movement beyond local, regional and international borders [138, 139]. In Africa, countries were forced to prioritise COVID-19 containment and mitigation measures often at the cost of other potentially deadly infectious diseases such as malaria [136]. These measures resulted in the readjustment of resources that were initially allocated towards routine public health programs such as immunisation and surveillance [138]. However, joint advocacy efforts from the E8 Initiative were instrumental in ensuring that malaria health budgets in the E8 region were not significantly changed and that the national malaria control programmes (NMCPs) received the appropriate resources to continue facilitating routine vector control and community case management during the COVID-19 pandemic [140]. None-the-less, reports from the E8 Initiative showed that the region, as a whole, experienced an increase in the total annual malaria cases and deaths in 2020 compared to 2019 [136]. A study on the impact COVID-19 has had on malaria burden and services from five countries (Angola, Eswatini, Mozambique, Namibia and South Africa) in the E8 region revealed that malaria case management and surveillance were negatively impacted by the pandemic, with people less likely to seek healthcare for fear of being infected with the virus [136]. The deployment of vector control interventions such as IRS and distribution of LLINs were affected to a lesser extent [136]. The malaria workforce was redeployed for COVID-19 duties, however experienced limited movement due to travel restrictions and staff self-isolating due to illness. Additionally, other interruptions in the flow of case management and IRS activities were mainly as a result of challenges experienced with procurement and supply management of malaria commodities [136]. A separate study in Zimbabwe also revealed that there was an increase in the cases and deaths by malaria compared to the same period between 2017 to 2019 before the COVID-19 pandemic [141]. While it was indicated that further studies were required to explore possible explanations for these observed trends [141], another possible explanation for some of the observed upwards trends in malaria cases in the region may have also been as a result of the overlapping of symptoms between malaria and COVID-19 such as fever, fatigue, breathing difficulties, and acute onset headache just to name a few [134]. This may have led to misdiagnosis of malaria for COVID-19 and *vice versa*, particularly when clinicians relied mainly on symptoms for diagnosis [134]. Targets for eliminating malaria in the E8 region by 2030 may therefore now be in further jeopardy due to disruptions in health services caused by the COVID-19 pandemic.

One positive from the border closures and restrictions on cross-border travel has been the reduction in the importation of malaria infections though formal borders. Additionally, the COVID-19 pandemic emphasised the importance of collaborative pathogen genomics for effective disease surveillance and outbreak management in Africa [127]. This is because genomics of SARS-CoV-2, which was mainly done through collaborative work showed that genomics was instrumental in the timely development of diagnostics and vaccines, monitoring for viral evolution that affects diagnosis, transmissibility and virulence, and in understanding COVID-19 transmission dynamics, thus supporting timely control of COVID-19 outbreaks [127-133]. COVID-19 surveillance using genomics was also found to be useful in the overall assessment of the effectiveness of infection prevention and control measures [127, 133] and to track the spread of new concerning variants of SARS-CoV-2, which have high-transmission rates and the potential to affect COVID-19 medical interventions [127, 142].

Since there has generally been a reluctance to use regional hubs of collaboration within the southern African region [140], hopefully the lessons learnt from the benefits of the use of genomics in COVID-19 through international collaborations, and funding for molecular surveillance activities from institutions such as the Bill and Melinda Gates Foundation this will encourage uptake of malaria molecular surveillance (MMS) collaborations on the malaria parasite and mosquito vector development and survival [143]. Hopefully, malaria control and elimination programmes can also be encouraged to efficiently share data for molecular surveillance of the malaria parasite between countries in the E8 region through taking advantage of accessible regional platforms such as the KwaZulu-Natal Research Innovation and Sequencing Platform [127]. As the E8 region continues to address the COVID-19 pandemic, it is also important that malaria remains a priority on the political agenda and that NMCP adapt and learn from the COVID-19 response to ensure the timely attainment of the elimination targets.

Therefore, in this thesis, the focus will be on how parasite population genetics can be used as a value-added tool to improve surveillance in pre-elimination settings of southern Africa.

1.7 Malaria parasite population genetics

Population genetics is the study of the genomic composition of populations and investigates allele frequency variation over space and time, in response to different evolutionary processes such as natural selection, genetic drift, mutation and gene flow [144]. Malaria population genetics has been adopted for use by several countries around the world to understand malaria transmission patterns and the impact on malaria interventions; as well as in the early detection of anti-malarial drug or insecticide resistance and the early detection of diagnostic resistance (pfhrp2/3 deletions) [62, 64, 118, 122, 126, 145-160].

Prior evidence from malaria parasite population genetics data generated from majority malaria endemic countries in Africa and other parts of the world suggests that genetic complexity and diversity of parasites is associated with transmission intensity [62, 64]. People in high-transmission settings often carry many parasite genomes within them due to multiple mosquito vectors carrying multiple distinct parasite clones (Figure 1.2). When an uninfected individual is bitten by a single mosquito carrying multiple parasite clones that may be related, this is known as co-transmission [161]. However, also as likely, or more likely to contribute to a higher number of distinct parasite clones circulating in an individual in high transmission settings is superinfection, wherein the same individual is infected multiple times with unrelated parasite clones before clearing parasites (either spontaneously or via antimalarial drugs) [11, 158, 162]. It is generally believed that as malaria transmission declines, infected individuals go down to fewer parasite genomes per person and finally to a very few parasites that are actually present and circulating in the population and thus reflect the level of transmission intensity [120]. Therefore the number of distinct parasite clones circulating within an individual infection, known as multiplicity of infection (MOI) decreases, and subsequently results in a less diverse parasite population with strong linkage disequilibrium (LD) (described further in section 2.2.7), thereby leading to a spatially fragmented parasite population [120].

Plasmodium falciparum parasites from different transmission settings vary genotypically, with a strong correlation between malaria transmission intensity, MOI, and population structure of parasites (Figure 1.2) [162, 163]. Sub-Saharan Africa is defined as a high transmission malaria area with multiple-clone *P. falciparum* infections and a high genetic diversity amongst the parasite populations [164]. The incidence of highly diverse *P.*

falciparum parasites in high-transmission settings occurs due to the high levels of outbreeding between multiple distinct parasite genotypes during sexual recombination in the mosquito. Multiple-clone infections are evaluated as MOI in humans (Figure 1.2) and in these high-transmission settings, they usually come about as a result of multiple bites from different mosquitoes carrying genetically distinct *P. falciparum* parasites (superinfection) [165]. Parasite populations are also highly mixed across different geographic scales in high transmission settings [145]. However as malaria control interventions are employed, the level of genetic diversity in the parasite population starts to decrease and parasites start to separate into geographic clusters [124]. By contrast, patients infected with malaria in low-transmission areas such as South American and some parts in Africa, countries have infections caused by a single clone of *P. falciparum* parasites with limited genetic diversity [124, 164]. The parasite population is clonal with identical genotypes that persist over time and across different geographic areas (Figure 1.2) [124]. These identical genotypes self-replicate leading to high levels of parasite relatedness that is evaluated as the LD [145, 166]. As a result, there is a higher prevalence of infections that are genetically identical [124, 145, 163, 166, 167]. Individuals in these areas are said to have a low MOI and there is usually a less effective parasite population size [163]. As low-transmission countries move towards malaria elimination assessing the level of genetic diversity therefore aids in distinguishing between local and imported cases.

In the mosquito, a measure of exposure to infectious mosquitoes is evaluated as the EIR (Figure 1.2), which is interpreted as the number of *P. falciparum* infective bites received by an individual during a season or annually [78]. This metric can be used in various transmission settings, however it lacks precision since distribution of mosquitoes is highly heterogeneous and the mosquito biting rates are difficult to estimate [79, 80].

Therefore, this illustration (Figure 1.2) summaries how genomics allows us to assess transmission by acting as a value-added tool for surveillance. In this case the reduction in the number of parasite genomes in an infected person and the emergence of clonal parasites upon employing interventions is primarily being measured. These metrics can be collected by sequencing or genotyping highly polymorphic loci in the parasite genome.

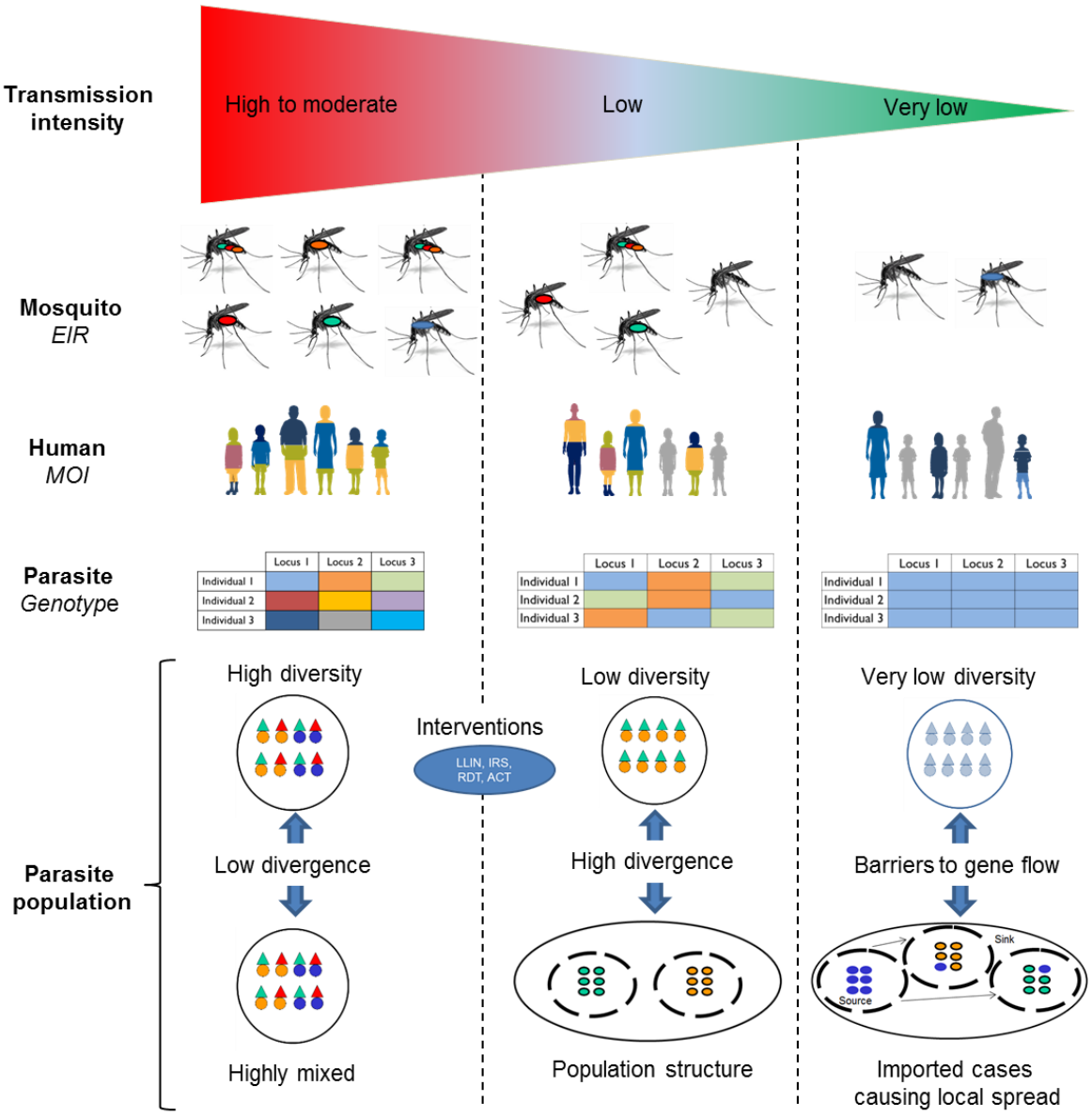


Figure 1.2 Parasite transmission dynamics assessed on an individual and population level. In the mosquito, this is evaluated as the entomological inoculation rate (EIR). In humans, multiplicity of infection (MOI) is measured and the number of different clones that are carried in each individual represented by the different colours. The parasite genome of each individual will therefore differ based on the variability of infecting clones. On a population level, numerous individuals each carry different or similar alleles across their genomes which then leads to either high or low genetic diversity. Diversity refers to the amount of allelic variation among individuals in a population, whereas divergence refers to the amount of allelic variation between different populations. Under balancing selection, high diversity would be expected at a locus under selection, but low divergence would be expected between populations.

1.7.1 *P. falciparum* genotyping strategies

The *P. falciparum* parasite's genome, first published in 2002 [168] spans 23 MB over 14 chromosomes, encoding more than 5000 genes [168]. The parasite also has a 6 kb

mitochondrial genome and a 35 kb plastid genome [169]. This makes it one of the smallest genomes compared to that of other pathogens. Curated genome variation data on 7,000 *P. falciparum* samples from 28 malaria-endemic countries was recently released and will contribute towards the facilitation of research into the evolutionary processes affecting malaria control and to the accelerated development of the surveillance toolkits required for malaria elimination [170].

A major source of genetic diversity in the parasite arises through the recombination of alleles thereby introducing new allele combinations into the genome [171, 172]. In the *P. falciparum* genome, recombination occurs at relatively high rates but varies between and within parasite populations [171-173]. The formation of new alleles through recombination occasionally enables the parasite to evade the human host's immune response [172] or effects of antimalarials. When mosquitoes feed on an individual infected with multiple parasite genotypes or on different individuals carrying different parasite genotypes [171, 172], these genotypes recombine during the sexual stage of the parasite's life cycle thereby making the new variants.

Genotyping tools allow determination of genetic diversity across the parasite's genome by typically assessing thousands of molecular markers. Genetic diversity across the *P. falciparum* genome manifests in the form of single nucleotide polymorphisms (SNPs), short insertions and deletions, inversions, non-coding variable number tandem repeats, translocations, microsatellites and gene copy number variations [174]. Modern approaches of genotyping and assessing the genetic variation in the parasite population include small SNP genotyping panels [175-178], large SNP genotyping arrays [179, 180], targeted sequencing of amplicons [181-186] or molecular inversion probes (MIPs) [187] and shotgun whole-genome sequencing (WGS). However, the most applied genotyping techniques for *P. falciparum* include microsatellites [122, 145, 152, 153, 188, 189] and allele-specific polymerase chain reaction (PCR) [190] with agarose gel profiling of PCR amplicon size polymorphisms [191], due to their simplicity and low cost. The most common genotyping strategies used to establish the genetic diversity in different *P. falciparum* populations are discussed in more detail in the below sections.

1.7.1.1 Size polymorphic antigenic markers

Size polymorphism has been used to genotype of *Plasmodium* parasites and is mostly based on characterising size differences in surface proteins, some of which are

associated with antigenic makers: merozoite surface proteins 1 (MSP-1) and 2 (MSP-2), the glutamine rich protein (GLURP) expressed throughout the asexual life cycle and the circumsporozoite surface protein (CSP), associated with sporozoites [192, 193].

This genotyping strategy amplifies the polymorphic regions of the parasite genome to identify haplotypes. The antigenic markers are amplified by PCR and the size of the amplicon can either be determined by gel or capillary electrophoresis [194]. The strategy is useful for polygenomic infections, where more than one parasite genome contributes to the infection, which allows the estimation of the MOI level across patients [195]. The mean number of genetically distinct parasite lineages infecting a person can be estimated using this genotyping method to distinguish between high and low transmission levels. However, since size polymorphisms are dependent on characterising a limited set of genes, the level of detail obtained is obviously constrained [193]. Further drawbacks associated with this technique include that the interpretation of agarose gels shows significantly variable results when repeatedly performed on the same individual [196]. Genotyping accuracy can be improved with increasing the numbers of molecular markers assessed. Polymorphic markers with larger repeats such as the CSP and GLURP protein have been used, however they are less diverse compared to MSP-1 and MSP-2 [145, 196]. Other challenges are that loci are under strong immune selection will affect the distribution of alleles within a population. Further, there are non-random combinations of MSP-1 alleles in individuals with multiple infections, leading to underestimation of genetic diversity. Although these methods are easy to perform and are inexpensive, they provide lower resolution than sequencing methods.

1.7.1.2 High-throughput single nucleotide polymorphism genotyping

There are several different technologies that exist for genotyping SNP loci. SNPs, which were previously discovered through WGS are defined as deoxyribonucleic acid (DNA) sequence variations, which occur when either one of the four nucleotide bases (adenine, cytosine, guanine, or thymine) in the genome differs between parasites or paired chromosomes in an individual. Approaches such as Sequenom/Agena, Taqman and high-resolution melting curves have been employed in the malaria genomics field for genotyping single biallelic SNPs in large numbers of parasite or vector samples [120]. The Taqman approach for example led to the development of a SNP molecular barcode, which is comprised of different combinations of SNPs that together express a unique pattern of variation of sequences, and these patterns are then used to describe and distinguish

parasite clones from each other [146]. The technique can be easily applied to a variety of field isolates with a more than 99 % success of distinguishing different parasite genotypes [146].

SNP-based barcodes are able to distinguish malaria parasites that originate from different geographic regions, hence they are used to identify sources of infections in pre-elimination settings [197, 198]. SNP barcodes have also been used in *P. vivax*, where a panel of 42-SNP barcode was developed using clinical samples from parasite populations originating from Africa, Asia and South America [198]. In *P. falciparum*, a panel of 24 SNPs were selected from over 112000 SNPs from 18 parasite genomes [146]. The selected SNPs segregate independently and they are broadly distributed across the genome, exhibiting a 35 % minor allele frequency, which is defined as the frequency at which the second most common allele occurs within a population [146]. The panel is the most widely used in different malaria endemic regions and has been previously used to study the vector and human interactions by tracking parasite genotypes over time with the aim of understanding parasite populations, to investigate the impact of various intervention strategies in high transmission areas such as Zambia and Zimbabwe, as well as to investigate clonal outbreaks of *P. falciparum* in Panama [146, 199, 200]. This genotyping technique is sensitive and requires small amounts of genomic DNA for amplification of genome fragments containing the SNP sites [201]. The technique is also inexpensive at scale although it requires specialized instrumentation [118].

SNP-based barcodes have been described as more informative compared to the size polymorphism genotyping, as they have the capacity to differentiate between strains due to the variable number of alleles that exist among the 24 markers [202]. Compared to microsatellites, SNPs are more amenable and they overpower the limitations of traditional genotyping based on length polymorphisms [203, 204]. However, a drawback of DNA based barcodes (24 *P. falciparum* and 42 *P. vivax* SNPs) is that they are located in the nuclear genomes, which makes them easily subject to recombination, thus possibly disrupting the associations among the panel SNPs. As a result, the correct genotyping of the parasites could be impaired [205]. Additionally, due to high effective population sizes of the nuclear genomes in comparison with the uniparentally evolving organelle genome such as the mitochondria [206], the rate of mutation in the nuclear genome is expected to be higher. Therefore, together, the higher recombination and mutation rates, as well as large effective size in the nuclear genome, this is expected to lead to early disruption of the SNP panel. Technical and biological constraints however limit the scalability and

discriminatory power of SNP-based methods especially in sub-Saharan African settings where infected individuals carry multiple-strain (polyclonal, MOI > 1) infections even when they are in low-transmission areas where mostly single strain (monoclonal, MOI =1) infections would have been expected as low density minority strains become harder to differentiate [122, 123, 188, 207].

Taking these factors into consideration, next-generation sequencing (NGS) of short, highly variable regions of the genome with numerous alleles (microhaplotypes) of 100–300 bp length and mostly composed of 3 or more SNPs is gaining momentum in the field of malaria genomics [208, 209]. This is because genotyping numerous genetically diverse loci offers the potential for higher resolution comparisons of infections at a population level [209]. This type of amplicon sequencing approach has provided insight into complex infections [183], vaccine efficacy [210], and has also been applied for surveillance of known and novel alleles at drug resistance loci [211, 212]. Most recently, an amplicon sequencing panel was designed for *P. falciparum* to characterise highly heterozygous microhaplotypes so as to infer relatedness of infections and could have great utility in low transmission settings of southern Africa [186].

1.7.1.3 Polymorphic microsatellite genotyping

Microsatellite genotyping in *P. falciparum* is used to investigate parasite genetics, population structure and dynamics of clinical isolates [122, 145, 152, 153, 188, 189, 213]. Other notable uses of microsatellites have been to identify the spread of *P. falciparum* antimalarial drug resistance alleles across continents, from Asia to Africa [214] and more recently, across the Eastern Greater Mekong subregion [215, 216]. Microsatellites are defined as short tandem nucleotide repeats, generally limited to between 1 and 6 nucleotides each, found in specific genomic localisations across the genome. In the A+T-rich genome of *P. falciparum* [199], microsatellites are typically characterised as $(TA)_n$ and $(TAA)_n$ motifs, with the variation in the numbers of tandem repeats among strains [217]. These remain important to understand the genetic diversity, relatedness of parasite populations and understanding transmission dynamics.

With microsatellite genotyping, genotypes are obtained accurately from several microsatellite regions (over many loci), providing higher resolution data than size polymorphisms described above. Microsatellites can be detected from limited amounts of genomic DNA and the microsatellite loci are amplified using multiplex PCR [218]. The

sizes of microsatellites can be highly reproducibly measured with capillary electrophoresis, which has a high resolution and precision to detect one nucleotide difference between microsatellites, that allows the diversity of loci to be fully detected [196]. Microsatellites therefore describe allelic diversity at large number of loci, which span across the genome and can type either individuals of populations [219]. Unlike antigenic markers, microsatellites are not under immune selection and the size of alleles have predictable lengths that are easily used to compare between various samples [196].

Microsatellite genotyping is limited due to variable mutation rates [220, 221] and the number of copies of the repeat, thus, the higher the repeat count, the higher the mutation rate, which may result in the detection of false positive alleles [222]. High mutation rates are due to the interplay between strand slippage events and mismatch repair, which counteracts DNA slippage during replication. As a result, the mutation rates allow the microsatellites to be used to understand the emergence of recent and local population genetic patterns [220, 221, 223].

In most studies, *P. falciparum* parasites are characterised using 10 commonly used microsatellites, which are from seven chromosomes [145, 194]. The different loci were selected at random with respect to gene function [219]. Some of the loci are situated in genes with full sequences available, with majority of the loci uncharacterised [145]. This makes it challenging to reveal homology with other organisms, further making it difficult to determine the functions of the sequences [224]. Nonetheless, loci are situated in coding regions and are attributed to continuous reading frames with conserved sequences that flank the repeat regions [224].

In the present study, a panel of 26 microsatellite loci were used to measure genetic diversity in clinical isolates. The 26 microsatellite loci includes a panel of 16 newly identified loci that flank the 10 (PolyA, Ta81, TA87, TA1, TA109, TA40, ARA2, PfG377, PfPK2 and TA60) microsatellite loci previously described [225]. The 16 were selected and identified microsatellite loci in the up- and down-stream flanking regions of the 10 microsatellite loci, to allow the evaluation of short-range haplotypes. The 26 loci were carefully evaluated for their utility as genotyping markers and they were found to be neutral and polymorphic [225]. In previous studies, the panel was used to determine the genetic diversity of southern African parasites to infer local and cross border infections, to use genetic diversity as a surveillance tool to understand the dynamics of malaria

parasites and to evaluate temporal spatial clustering of parasites from Zambia [218, 226, 227].

Considering the evaluation time frame of three years to certify an area as malaria free, the most suitable markers would be those that quickly evolve such as microsatellite markers. For low to medium income countries in Africa, microsatellite genotyping is also a cost effective approach that gives good genetic resolution for the characterisation of community, country or region level genetic diversity and when distinguishing between local and imported *P. falciparum* strains [122, 145, 189, 224]. Therefore, for these reasons, in this thesis it was appropriate to use microsatellite markers to analyse samples from different transmission settings within South Africa. . Additionally this technique will allow for the direct comparison of regional genetic data from southern African countries working towards malaria elimination by 2023 which will be described further in Chapter 4.

1.8 Malaria in southern Africa and challenges for elimination

Malaria-endemic countries from the southern Africa region, excluding Mozambique, accounted for < 11 % of the 229 million cases reported from the WHO African region in 2020 [228]. This relatively low case load in southern Africa compared to the rest of the continent resulted in the region being earmarked for elimination by the WHO [228, 229].

Currently, as part of the Elimination 8 Initiative (E8), 8 countries in southern Africa are leading the efforts to malaria elimination on the continent [140]. These include the four low-transmission front-line countries (South Africa, Namibia, Eswatini and Botswana), which aim to eliminate by 2023 and pave way for another 4 higher transmission, second-line countries (Mozambique, Zimbabwe, Zambia and Angola) to eliminate by 2030 [140, 229] (Figure 1.3). The E8 was created by Health Ministers from all eight countries to coordinate the implementation of a regional malaria elimination strategy as it was seen that malaria elimination in any one southern African country would not be possible without regional cooperation and collaboration [230, 231] due to the high levels of interconnectedness of southern African countries [122, 232]. This interconnectedness is caused mainly by highly mobile migrant human populations who facilitate the constant movement of malaria parasites across country-borders as they travel across country borders, mostly from high-transmission countries to low-transmission countries in the region [122, 232]. The implementation of five cross-border malaria control initiatives and

the deployment of malaria health units at strategic points along shared borders in the southern African region was supported by the E8, such that these units led to a 30 % and 46 % reduction in malaria incidence and mortality, respectively, in the E8 border regions [140]. These cross-border malaria initiatives included BOMOZISA (Botswana, Mozambique, Zimbabwe and South Africa), MOSASWA (Mozambique, South Africa and Eswatini), Trans-Kunene (TKM - Angola and Namibia), Trans-Zambezi (TZMI - Namibia and Zambia) and ZIM-ZAM (Zambia and Zimbabwe) [140].

A novel data sharing platform known as The Situation Room, also allowed for the early detection of and prompt response to malaria outbreaks in the E8 [140]. However, despite advanced and well controlled vector control strategies, to which the E8 countries' low malaria numbers can be ascribed, South Africa, and other countries in the region have not been able to reach their elimination goals at the proposed target dates [1, 76]. The decline in malaria cases and deaths has not been uniform within and between low-transmission countries in the E8 region (Figure 1.3) from the year 2000 to 2019. Cases are higher in second line high-transmission countries compared to frontline low-transmission countries. Low-transmission countries however face a unique challenge in that these front-line countries share porous borders with areas of higher transmission (Figure 1.3), associated with both human migration and mosquito movement [122, 188]. For instance, malaria in low-transmission settings such as South Africa and Namibia is heterogeneous and the highest malaria incidence in these countries occurs at international border regions. In low-transmission countries in sub-Saharan Africa, the character of disease transmission is believed to be similar to that in high-transmission countries thus hampering malaria elimination efforts [73]. It is therefore important to try and understand why progress to elimination has stalled in the southern African region by using new tools such as parasite population genetics to understand regional transmission dynamics.

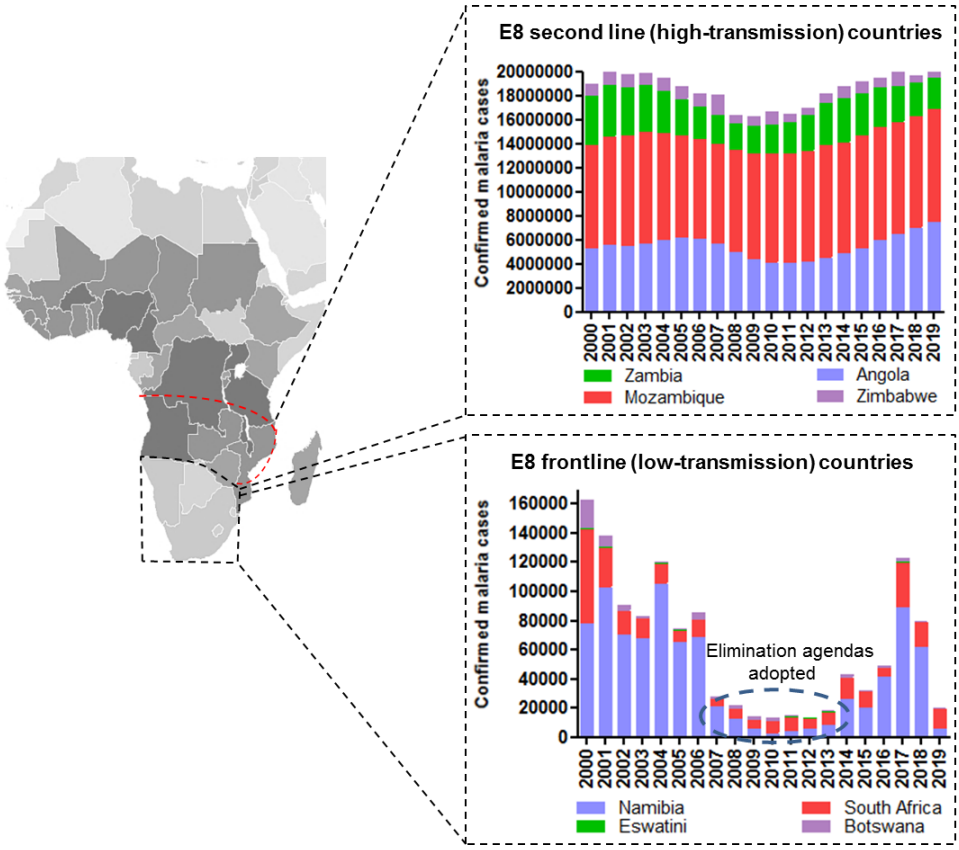


Figure 1.3 Distribution of malaria cases in southern Africa E8 region. The map shows the varied distribution of malaria incidence in the African continent with dark colours representing regions of high-transmission intensity and the lighter colours representing regions of low-transmission intensity. Focus is on the southern African E8 region relevant to this study The data used to generate the graphs was obtained from the World Health Organization 2020 malaria report [1].

1.9 Malaria in South Africa

Malaria transmission in South Africa is seasonal, occurring mainly during the hot, rainy season between September and May, with majority of the malaria cases and malaria-related deaths due to *P. falciparum* infections. Although endemic malaria transmission is limited to border regions of three provinces, namely KwaZulu-Natal (KZN), Mpumalanga and Limpopo, transmission intensity in these three provinces is very heterogeneous [66, 233], with focal areas of sustained local transmission occurring in certain areas [233] (Figure 1.4A).

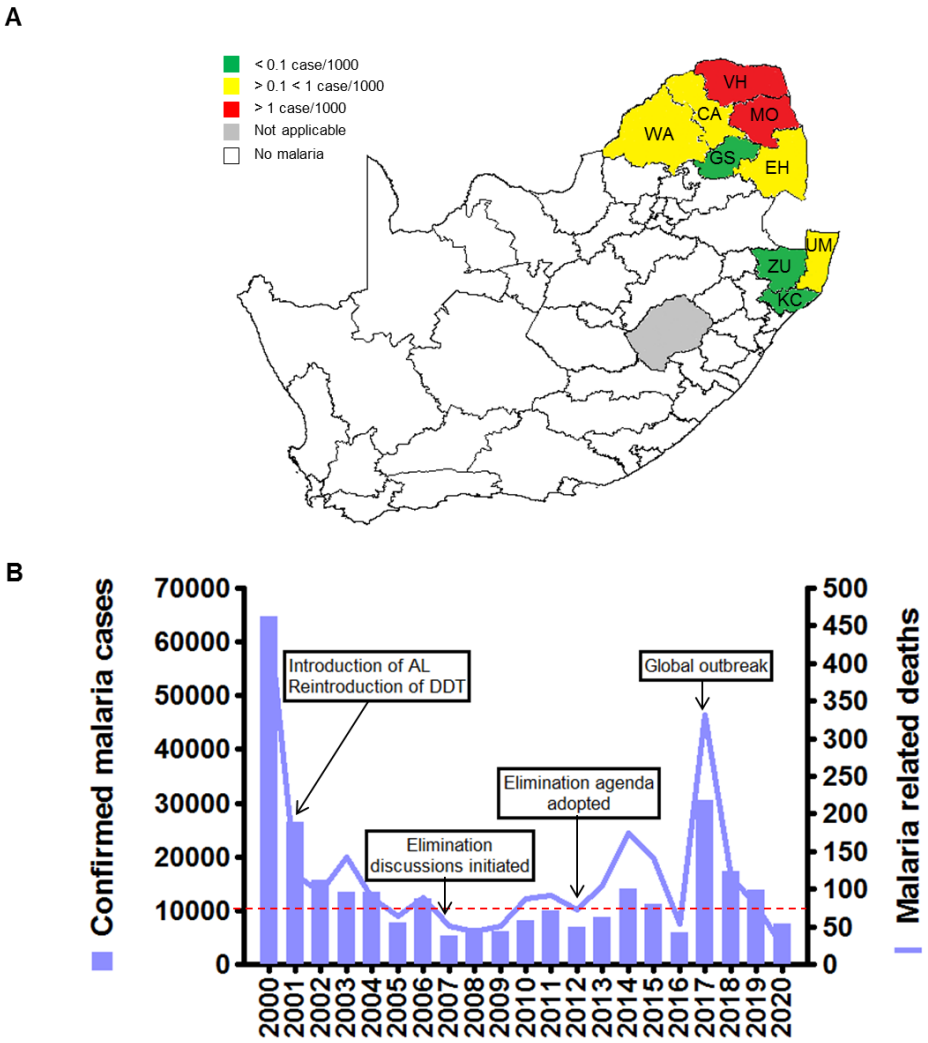


Figure 1.4 Malaria situation in South Africa. (A) Map showing hotspots of local transmission in South Africa indicated by district incidence rates in the remaining malaria endemic provinces of South Africa staged along the elimination continuum, 2018. The following districts are found in the Limpopo Province: VH = Vhembe, MO = Mopani, CA = Capricorn, WA = Waterberg and GS = Greater Sekhukhune; in the Mpumalanga Province: EH = Ehlanzeni; and in the KwaZulu-Natal Province: UM = Umkhanyakude, KC = King Cetshwayo and ZU = Zululand. The districts are colour coded based on incidence rates as indicated in the key. Cases represent local, unknown, and untraceable cases for 2018. Figures were compiled with data from the World Health Organisation Malaria Report and South Africa Malaria Elimination Strategy 2019-2023 respectively. (B) Total malaria cases and deaths in South Africa from 2000 to 2020 showing that residual malaria hampers malaria elimination efforts in South Africa. The bars represent the confirmed malaria cases per year and the solid blue line malaria related deaths. The red line represents the average annual malaria case numbers that have averaged at around 10 000 for the past two decades thus representing the residual malaria.

Despite sustained malaria intervention efforts and prompt, effective case management, the north-eastern border regions of the country (Limpopo and Mpumalanga Provinces in particular) remain stable hotspots for malaria transmission. While South Africa's overall malaria burden is low (0.1-<1 local cases per 1000 population at risk) to very low transmission (≤0.1 local cases per 1000 population at risk), there are certain areas within

the Limpopo Province (the highest transmission area in the country) that have higher transmission intensities (≥1 local cases per 1000 population at risk) [66] (Figure 1.4A). In this province, the Vhembe District has the greatest burden of disease and is classified as a moderate transmission area with 3.79 local cases/1000 population at risk reported in 2018 [66]. Mostly local transmission occurs in the Vhembe District and is part of approximately 40 sub-districts in South Africa where local transmission occurs. The cause and dynamics of local transmission in these areas are poorly understood. However, in the KZN Province, which has the least transmission intensity in the country, very few locally transmitted cases (<0.1 cases/1000 population at risk) are reported (Figure 1.4A), with importation rates of more than 80 % [66]. Overall, imported malaria cases in the country account for close to 60 % of the total cases reported [66], mostly from Mozambique and Zimbabwe [233, 234]. Therefore, South Africa's latest strategic malaria elimination plan proposes a phased approach to achieve zero local transmission by targeting interventions to initially clear foci of transmission and ensuring systems are established to support elimination interventions [66].

The use of the insecticide, DDT, for IRS has been the mainstay of vector control in South Africa and has resulted in a notable decline in cases and deaths after its re-introduction in 2000 [233] (Figure 1.4B). Although tremendous progress has been made towards elimination since the 1999/2000 malaria outbreak, a resurgence of cases resulted in outbreak in 2017 [66], with South Africa reporting more than 30 000 cases, [66, 143, 233] (Figure 1.4B). Annual malaria case numbers in the country have also consistently averaged at around 10 000 for the past two decades (Figure 1.4B) in spite of the vector control and case management interventions employed, thus reflecting residual malaria transmission, which hampers elimination efforts in the country [229]. While there has been no drug resistance since the year 2000, and it is not really sustaining transmission currently, some of the other causative factors that have been shown to be driving residual malaria in South Africa include entomological influence with the predominant malaria vectors belonging to the *An. gambiae* and *An. funestus* complexes [235-240], and climate change, which have been shown to impose distinct biological constraints.

The authorisation of community health care workers (CHWs) in moderate malaria transmission areas of the Limpopo Province to test for malaria, and for environmental health practitioners (EHPs) to test for malaria and treat on the spot with artemether-lumefantrine (AL) [241] came as a welcome development to the malaria control programme in the Limpopo Province. Use of CHWs and EHPs in testing and treatment

aids in reaching especially asymptomatic cases and this ultimately leads to less burden on health facilities [241]. Reduced secondary transmission has been reported from an index case as a result of this practice, and all contact cases are also tested and treated on the spot [241]. This provides a good opportunity to administer primaquine in these areas to target and clear mature gametocytes, which are the transmissible forms of the *P. falciparum* parasite [241] as registered nurses are authorised to test, treat and administer primaquine based on national recommendations [242]. This work that is done by CHWs is also instrumental even in the capturing and accurate classification of imported cases in particularly mobile migrant populations from different countries since the CHWs stay in the same communities in which they work and they interact with community members on a daily basis [241]. Where illegal residents from neighbouring countries may be staying and working in these communities (for example farming or mining communities), or the community is too far and members are unable to access health care facilities, they can also be tested and treated on the spot by these CHWs and EHPs [241].

In 2001 the KZN Province became one of the first places outside the Greater Mekong region to use an artemisinin-based combination therapy, as first-line treatment for uncomplicated falciparum malaria [243]. This was after it became apparent that previous malaria outbreaks experienced in 1986 and 2000 were mainly driven by chloroquine resistant and SP resistant *P. falciparum* parasites in circulation respectively [244, 245], together with pyrethroid resistance in the *An. funestus* mosquito populations present in South Africa [246]. The re-introduction of DDT [247] and introduction of AL treatment policies reduced malaria cases and deaths since then, and led to a drug resistance surveillance programme as part of the NMCP from 2000 in which known drug resistance markers were monitored [248]. Therefore, when the 2017 outbreak was experienced, it was known that the outbreak was definitely not due to drug resistance but rather due to system health failures [66]. This kind of molecular surveillance was, however, not done in the rest of the E8 region and was therefore not taken up well by programmes since MMS has not been a priority. Although malaria elimination in the KZN province has not yet been achieved, this kind of drug resistance molecular surveillance programme highlighted the importance of MMS in the fight against malaria in South Africa and the region. The KZN Province has now for the past two decades been able to maintain very low levels of local transmission partly because of the introduction of AL and is targeted first in line for malaria elimination a in South Africa by 2023 [66, 248]. These timelines may however be affected due to recent world events caused by the new global health pandemic, COVID-19, that

affected malaria control and elimination programs not only in South Africa but other countries in the E8 region [141, 249].

1.10 Rationale of project

Passive and active surveillance systems have been strengthened by a number of countries, including South Africa, to meet the surveillance needs for elimination [233]. Although these surveillance systems may identify the location where both symptomatic and asymptomatic infections occur, they are limited in their ability to accurately determine the geographic origin of infections given that they depend on reported travel histories from patients, mainly based on human recall, which may at times result in incorrect information [123].

Malaria knows no borders, and the southern African region is no exception [122, 188, 250]. This has to do mainly with human and mosquito mobility for transmitting malaria. Malaria transmission in border areas between some low- and high-transmission countries is similar [73]. This presents a challenge for malaria elimination, partly because there are several gaps in knowledge of malaria infection risk factors at low-transmission in the region [140]. It remains unclear exactly how much malaria is being imported across the borders, which and how cases are spreading locally - a clear distinction that is required to achieve elimination. This therefore makes it imperative, in the South African context, to understand the genetic contribution of locally transmitted parasites compared to 'imported' parasites to overall parasite loads and to continued malaria transmission in low-transmission settings and how this influences the path to malaria elimination in the southern African region.

To achieve elimination, highly sensitive, accurate and innovative tools such as molecular genotyping of *P. falciparum* infections are therefore required as additional tools to the currently available surveillance systems to support better surveillance of malaria in aid of reaching elimination [119]. Most of the studies on parasite population genetics studies done in Africa have however been conducted in areas of high-transmission [64, 145]. Nonetheless, there is evidence already in the southern African region that the genetic diversity in low-transmission settings can be determined with microsatellite genotyping [122, 188, 227]. Therefore, these microsatellite studies provided support that the same microsatellite genotyping strategy could be used to study the South African parasite

population. There is currently no data describing the *P. falciparum* parasite population's genetic diversity in South Africa at different spatial and temporal scales and how it relates to the rest of the southern African region (Figure 1.5). As such, it is unclear if the genetic diversity of the parasite population is a driving factor that contributes to and fuels the sustained residual malaria seen in specific hotspots in the country. Therefore, this study focussed on evaluating the diversity of *P. falciparum* parasites in South Africa and associate this with temporal and geospatial variances to identify contributing factors associated with sustained residual transmission as the country and region work towards malaria elimination. The contribution of parasites carried by asymptomatic migrant populations to malaria transmission in South Africa, and the connectivity of parasite population within South Africa and between South Africa and neighbouring countries was also assessed to be able to inform targeting and/or selection of interventions and surveillance.

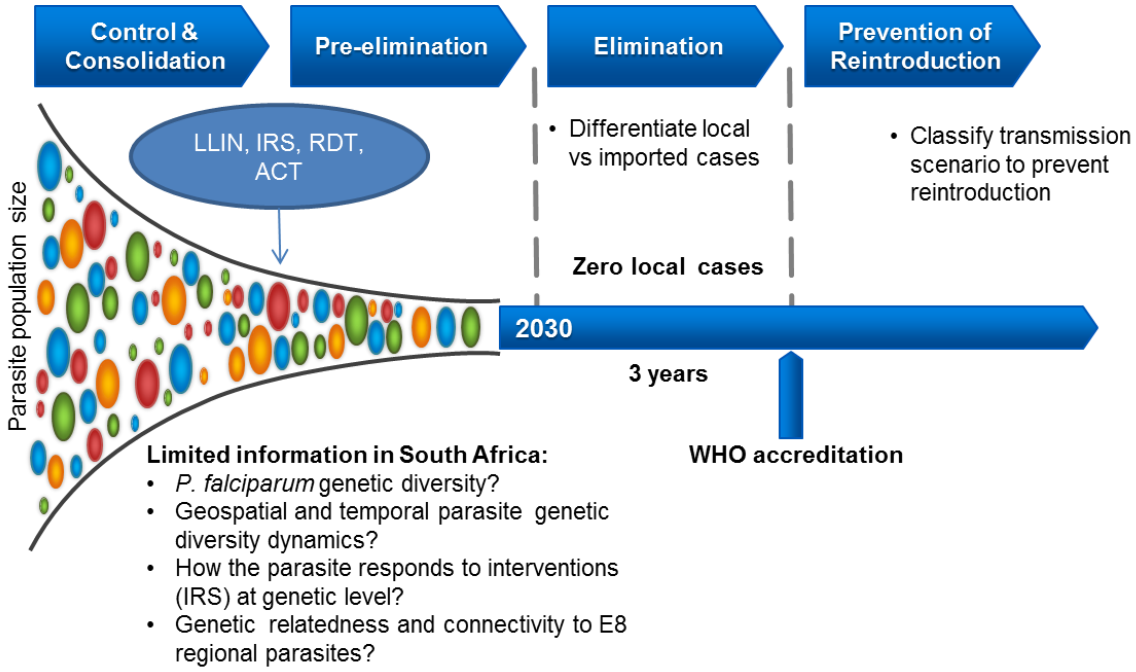


Figure1.5 South Africa malaria elimination plan. Knowledge gaps of *P. falciparum* population genetics in South Africa presented alongside the malaria elimination continuum. South Africa is in the pre-elimination phase. The specific research questions highlighted here are addressed in this thesis as the country works towards malaria elimination**.**

1.11 Project aim and objectives

HYPOTHESIS

Parasite diversity is a contributing factor to sustained malaria transmission dynamics in southern Africa.

AIM

This study aimed to assess the genetic diversity in *Plasmodium falciparum* parasites from malaria hotspots in South Africa and malaria endemic neighbouring countries to enable the identification of factors contributing to sustained transmission in South African malaria hotspots.

OBJECTIVES

1. To determine *P. falciparum* within-host and population level genetic diversity in the Vhembe District of South Africa using microsatellite genotyping (Chapter 2).
2. To evaluate genetic diversity of parasites from the KZN Province (lowest burden of cases) in South Africa and compare them to those from the Vhembe District, Limpopo Province (highest burden of cases) (Chapter 3).
3. To conduct a meta-analysis on southern African parasite populations to understand regional transmission dynamics (Chapter 4).

RESEARCH OUTPUTS

Research findings generated in this study resulted in the following manuscripts:

Journal articles

Published manuscripts

1. Gwarinda, H.B., Tessema, S.K., Raman, J., Greenhouse, B., and Birkholtz, L. Parasite genetic diversity reflects continued residual malaria transmission in Vhembe District, a hotspot in the Limpopo Province of South Africa (2021). Malaria Journal 20 (1):96. doi: https://doi.org/10.1186/s12936-021-03635-z.
2. Raman, J., Gast, L., Balawanth, R., Tessema, S., Brooke, B., Maharaj, R., Munhenga, G., Tshikae, P., Lakan, V., Mwamba, T., Makowa, H., Sangweni, L., Mkhabela, M., Zondo, N., Mohulatsi, E., Nyawo, Z., Ngxongo, S., Msimang, S., Dagata, N.,

Greenhouse, B., Birkholtz, L., Shirreff, G., Graffy, R., Qwabe, B. and Moonasar, M. High levels of imported asymptomatic malaria but limited local transmission in KwaZulu-Natal, a South African malaria-endemic province nearing malaria elimination (2020). Malaria Journal, 19:152. doi: https://doi.org/10.1186/s12936-020-03227-3

Manuscripts in preparation

1. Niemand, J., Maboane, S., Gwarinda, H.B., van Heerden, A., van de Watt, M. and Birkholtz, L. Genetic diversity in southern African *Plasmodium* isolates displays differential sensitivity to gametocytocidal compounds.

Conference presentations

Research findings generated in this study were presented at the following conferences:

Oral presentations

1. Gwarinda, H.B., Tessema, S.K., Raman, J., Greenhouse, B., and Birkholtz, L. *Plasmodium falciparum* genetic diversity associated with continued malaria transmission in pre-elimination settings in South Africa. Short Talk (virtual; Wellcome Connecting Science Conferences). Genomic Epidemiology of Malaria Conference (07-09 June 2021).
2. Gwarinda, H.B., Tessema, S.K., Raman, J., Greenhouse, B., and Birkholtz, L. Genetic diversity of *Plasmodium falciparum* in Vhembe District, a hotspot in the Limpopo Province of South Africa. Short Talk (virtual). 6th Annual Johns Hopkins Malaria Research Institute Future of Malaria Research Symposium (13 November 2020).

Poster presentations

1. Gwarinda, H.B., Tessema, S.K., Raman, J., Greenhouse, B., and Birkholtz, L. Genetic diversity of *Plasmodium falciparum* in Vhembe District of Limpopo Province, South Africa: Insights from microsatellite markers. Poster presentation. 5th MRC Office of Malaria Research Conference, University of Pretoria, South Africa (30 July – 1 August 2019).
2. Gwarinda, H.B., Tessema, S.K., Raman, J., Greenhouse, B., and Birkholtz, L. Genetic-based malaria parasite drug resistance monitoring for malaria elimination agenda. Poster presentation. 3rd Medical Research Council Office of Malaria Research Conference, Johannesburg, South Africa (August 2017).

CHAPTER 2

PARASITE GENETIC DIVERSITY REFLECTS CONTINUED RESIDUAL MALARIA TRANSMISSION IN VHEMBE DISTRICT, A HOTSPOT IN THE LIMPOPO PROVINCE OF SOUTH AFRICA

This data chapter has been published as:
Gwarinda, H.B., Tessema, S.K., Raman, J., Greenhouse, B., and Birkholtz, L. 2021. "*Parasite genetic diversity reflects continued residual malaria transmission in Vhembe District, a hotspot in the Limpopo Province of South Africa*." *Malaria Journal* 20 (1):96. doi: https://doi.org/10.1186/s12936-021-03635-z.

2.1 Introduction

Malaria remains a global health problem, with ~241 million cases reported worldwide in 2020, 95 % of which occurred in the World Health Organization (WHO) African Region [228]. As the southern Africa region, excluding high-transmission countries like Mozambique, accounted for < 10 % of the 213 million cases reported in the WHO African region [229], three southern African countries (South Africa, Botswana and Eswatini) have been earmarked for malaria elimination by 2025 [5] guided by the WHO Global Technical Strategy for Malaria [251]. Unfortunately, like a number of other regions in the world, southern Africa experienced a resurgence in malaria cases and deaths during the 2017/2018 season [229]. This resulted in South Africa reporting more than 30 000 cases, a surge in numbers previously only experienced during the 1999/2000 drug and insecticide resistance outbreak [66, 143, 233].

Plasmodium falciparum is the predominant species and accounts for the majority of cases and fatalities in the South Africa [66], with *An. arabiensis* (of the *An. gambiae* complex) the main vector species associated with transmission [66]. Historically, *An. funestus* and members of the *An. gambiae* complex were the main vectors from the 1930s-1950s, however, after the 1950s, *An. funestus* was been associated with outbreaks and epidemics, specifically those recorded during the period 1996 to 2001 [66]. Malaria transmission in South Africa mainly occurs in the hot and rainy season between September and May [66]. While South Africa has made significant progress in the reduction in malaria cases since the 1999/2000 outbreak through the implementation of sustained vector control and case management interventions, progress has recently

stalled [66, 143, 233, 252-255]. Transmission intensity is very heterogeneous within and between the three endemic provinces: KZN, Mpumalanga and Limpopo [66, 233]. The Vhembe District in the Limpopo Province carries the greatest burden of disease in South Africa, accounting for over 60 % of the country's malaria burden and with 3.79 local cases/1000 population at risk in 2018, is classified as a moderate transmission area. In contrast, the endemic districts within KZN are classified as very low transmission areas with transmission intensities of <0.1 cases/1000 population at risk and very few locally-acquired cases [66].

Vhembe District is situated in the north-eastern border region of the country bordered by Mozambique to the southeast, Zimbabwe to the north and Botswana to the northwest. This district experiences sustained, seasonal malaria transmission [233]. While importation of cases, mostly from Mozambique and Zimbabwe [233], has been implicated in on-going transmission in South Africa, the majority (63 %) of the cases notified from 1998 to 2017 in the district were classified as locally-acquired based on travel histories obtained from the patients [233, 234]. Other causative factors for the persistent residual transmission observed in the Vhembe District, include antimalarial drug resistance, insecticide resistance and vector species variance between *An. gambiae* and *An. funestus* [235, 236, 238, 240, 248, 253].

The addition of malaria parasite population genetic data to the standard surveillance data collected by malaria control programmes has assisted in understanding malaria transmission dynamics and also allowed for spatio-temporal inferences to be made from e.g. local *vs*. imported malaria cases [62, 64, 124, 145, 150, 152-154, 159]. These data reflect that typically, parasite genetic complexity decreases with a decline in malaria transmission. This holds true for low-transmission settings such as Southeast Asia and Latin America where parasite complexity of infections within individual human hosts is low compared to high-transmission settings of sub-Saharan Africa where parasite within-host complexity of infection is high. However, the limited data available from microsatellite genotyping of parasites from low-transmission settings in sub-Saharan Africa (from the KZN Province in South Africa, Eswatini, and Namibia [122, 188, 255]), suggests the opposite. In these study sites, parasites were highly complex and diverse, attributed to either frequent importations from neighbouring high-transmission settings or on-going local transmissions.

To better understand the malaria parasite-associated factors that contribute to residual malaria transmission in the Vhembe District, microsatellite genotyping was applied to describe the population of *P. falciparum* parasites in South Africa at different spatial and temporal scales. Additionally, the parasite load and distribution of anti-malarial drug resistance genes in genetically diverse *P. falciparum* parasites were also examined. This work identifies the parasite genetic factors associated with sustained residual transmission in the Vhembe District in South Africa to help with decision making around the selection and/or targeting of interventions as the country works towards malaria elimination.

2.2 Materials and methods

2.2.1 Ethical approval

Ethical approval for the study was obtained from the University of Pretoria, Faculty of Health Sciences Research Ethics Committee (Ethics Reference No. 406-2014) and Limpopo Department of Health (Ref: LP_201906_011). Patient information was extracted from the Limpopo Malaria Case (LMC) database and limited to the clinics from which malaria rapid diagnostic test (RDT) samples were collected.

2.2.2 Study site & study design

Study samples in the form of RDTs were randomly collected from seven primary health clinics within five known source health districts (Nzhelele/Tshipise, Thohoyandou, Mutale, Elim, and Levubu/Shingwedzi) within the Vhembe District, Limpopo Province, South Africa (Figure 2.4). Out of a total of 1892 confirmed cases that were reported during the study period at the 7 study clinics, a total of 1153 (61 %) RDT samples from symptomatic patients that tested positive for malaria on *P. falciparum*-specific RDTs (First Response® Malaria Antigen *P. falciparum* card test HRP2, Premier Medical Corporation, India) were randomly collected from the seven different clinics in the wet and dry seasons from January to December of the years 2016, 2017 and 2018. Samples were transported and stored at room temperature in sealed bags with desiccant.

2.2.3 DNA isolation and quantification

DNA was extracted from the nitrocellulose strip of First Response™ RDTs in accordance with the Worldwide Antimalarial Resistance Network Guidelines [256]. Briefly, the RDT cassettes were opened using sterile forceps and the nitrocellulose strips containing the parasite DNA removed. Using a clean pair of scissors sterilised with 95 % ethanol, three to four pieces of about 3 mm x 3 mm of the nitrocellulose strip were cut out from the DNA concentrated region approximately halfway between the blood application site and the result lines of the RDT test strip. The Saponin-Chelex method [257] was then used to extract DNA from the nitrocellulose strip first by incubating the strips in 1 mL of 0.5 % saponin prepared in 1X phosphate buffered saline (PBS; Ca^{2+}, Mg^{2+} free, pH 7.4) at 4 ºC overnight to lyse the erythrocytes. The tubes containing the mixture were then centrifuged the following day at 10 000 x*g* for 5 s before all the PBS/saponin was aspirated. Fresh 1 ml of PBS with no saponin was then added, vortexed briefly, and incubated at 4 ºC for 15-30 min before being centrifuged and the solution aspirated as before. A solution of 150 µL of 10 % chelex stored at room temperature was then transferred into the tubes and parasite DNA was extracted by incubating the tubes for 10 min in a 95 ºC heat-block, vigorously vortexing each sample every 2 min throughout the incubation. After incubation, the tubes were centrifuged for 5 min at 10 000 x*g* and the solution was transferred to a clean tube before a final round of centrifugation for 10 min again at 10 000 x*g* and transfer of the final supernatant with DNA. The extracted genomic DNA was then stored at -20 ºC until used.

To detect parasite densities in the very low blood volume (5µL) RDT samples, two ultra-sensitive *Plasmodium* specific quantitative PCR (qPCR) methods: *var* gene acidic terminal sequence (varATS) [258] and telomere associated repetitive element-2 (TARE-2) [258, 259], were used as described before. Starting from a 2 % parasitaemia (100 000 parasites/µL of blood) stock solution of 3D7 ring stage parasites, standard curves were generated from duplicate serial ten-fold dilutions of known density controls (100 000 parasites/µL, 10 000 parasites/µL, 1 000 parasites/µL, 100 parasites/µL, 10 parasites/µL, 1 parasites/µL, 0.1 parasites/µL, and 0.01 parasites/µL) of Saponin-Chelex extracted genomic DNA from 3D7 ring-stage parasites spotted with a 5 µL volume on filter paper and unused First Response™ RDTs. Calculations on the estimates of parasite density (number of parasites/µL of blood) were made based on WHO malaria treatment guidelines, assuming that 2 % parasitaemia of *P. falciparum* equates to approximately 100 000 parasites/µL of blood [260]. The experimental run included a negative (no template,

uninfected whole blood) and positive (ring stage 3D7 genomic DNA template) controls done in duplicate.

To analyse the relationship between the parasite densities and C_T values, linear regression analysis was done using Prism 5 (GraphPad Software, San Diego, CA, USA). The parts of the standard curve that adhere to 95 % confidence intervals were used to derive the linear equation for calculation of sample concentrations. The limit of detection (LOD) was defined as the lowest detectable parasite density of the control samples within the range on the standard curve, above the background.

Isolates (test samples) were then screened for *P. falciparum* using ultra-sensitive varATS [258] and TARE-2 [258, 259] quantitative PCR (qPCR) as described before. Parasite density was quantified on a random subset of 323 high volume samples from all years. It was not possible to quantify all samples due to a low DNA volume for a proportion of samples. The genotyping threshold was set at a parasite concentration of ≥ 10 parasites/µL of blood.

2.2.4 Microsatellite genotyping

A total of 1153 samples were genotyped using a previously described 26 microsatellite marker panel protocol [122, 188, 189]. Briefly, the 26 microsatellite loci were amplified using a semi-nested PCR protocol. The primary PCR was performed in 4 groups of multiplex reactions, and 1 µL of the primary amplified product used as a template for the secondary individual PCR for each marker. To determine repeat length sizes, the labelled PCR products were diluted and sized by denaturing capillary electrophoresis on an ABI 3730XL analyser using GeneScan™ 400HD ROX™ Size Standard (Thermo Fisher Scientific). MicroSPAT software (https://github.com/Greenhouse-Lab/MicroSPAT/releases/tag/v2.0.3) was then used to automate identification of true alleles and differentiate real peaks from artefacts of the resulting electropherograms using a classifier algorithm based on the location and size of locus-specific patterns relative to a primary peak as done in studies conducted in Eswatini [188], Namibia [122], the KZN Province of South Africa [255] and China [189] that used similar experimental conditions as those used in his study. Multiple alleles per locus were scored if minor peaks were at least a third of the height of the major peak. Collated genotyping data from all samples was processed with similar microSPAT software settings in which a semi-supervised naïve Bayes classifier was used [188] to avoid variability in allele calling. All samples that

met the genotyping threshold of ≥ 10 parasites/μL of blood were genotyped and had to have a genotyping coverage ≥ 60 % (alleles detected on at least 15 or more loci) to be included in the downstream population genetics analysis.

A genotype accumulation curve (GAC) was produced to assess whether the panel of 26 microsatellite markers had sufficient discriminatory power to distinguish between unique parasite haplotypes in different individuals in the Vhembe District sample set. In order to establish the distribution of the number of unique parasite haplotypes that can be identified in the sample set per different number of loci, loci were randomly sampled 1000 times without replacement and the number of multilocus genotypes observed per number of loci counted. This analysis also confirmed the number of loci necessary to discriminate individuals.

2.2.5 Characterisation of within-host genetic diversity

The within-host diversity of infections was determined using the within-host fixation index (F_{WS}) and multiplicity of infection (MOI) which identifies genetically distinct *P. falciparum* clones. To reduce the probability false positive alleles influencing MOI, the MOI in an individual sample was defined as the second highest number of alleles detected at any of the 26 microsatellite loci genotyped. Since the *P. falciparum* parasite is haploid in the human host, multiple peaks or alleles correspond to an infection with multiple genotypes or strains (a polygenomic or polyclonal infection).

The F_{WS} index is a measure of diversity in an individual infection relative to the population level genetic diversity and was determined as previously described [63, 122, 188]. A low F_{WS} indicates high within-host diversity relative to the population thereby suggesting higher chances of outbreeding. F_{WS} was calculated for each sample using the equation: $Fws = 1 - Hw/Hs$, where *Hw* is the allele frequency of each unique allele found at a particular locus for each individual and *Hs* is the heterozygosity of the local parasite population. Outbreeding is reported as 1-F_{WS} with a 0-value indicative of a perfect clone and therefore low within-host diversity. Previously described thresholds of Fws ≥ 0.95 (1-Fws ≤ 0.05) were used to identify samples containing a single genotype (or "clonal" infections) in spite of additional genotypes that may be present at relatively low proportions; and Fws ≤ 0.70 (1-Fws ≥ 0.30) to describe samples with highly diverse infections respectively [63, 149, 188, 261].

2.2.6 Characterisation of population level genetic diversity

Population level genetic diversity was estimated using expected heterozygosity (*He*) which is defined as the probability of randomly drawing a pair of different alleles from the allele pool. Heterozygosity was therefore calculated on each locus using the equation: $He = [\left(\frac{n}{n-1}\right) * (1 - \sum p_{i^2})]$, where n is the number of genotyped samples and p_i is the frequency of the i^{th} allele in the population [122, 188]. Values for *He* range from 0 indicating no diversity, to 1 indicating that all alleles are different and therefore there is maximum diversity. The mean *He* was calculated by taking the average *He* across all loci. The number of haplotypes (unique multilocus genotypes) as well as the unique alleles detected per locus (A - allelic richness), were also determined. The effective population size (N_e) was estimated using the infinite alleles model (IAM) and the step-wise mutation model (SMM) as described before [145]. Differences in the N_e were then evaluated across space (study sites – source health districts) and time (year of infection) and evidence of a parasite population bottleneck tested for using a simulation approach in BOTTLENECK software version 1.2.02 [262].

2.2.7 Linkage disequilibrium

To assess whether alleles from different loci were associated with each other, the multilocus linkage disequilibrium (LD) was determined as previously described [263] using the *Poppr* package in R software [264]. LD was determined for the whole dataset which includes monoclonal and polyclonal infections, with LD for monoclonal infections alone determined as a precaution against the bias that may result from presence of any false dominant haplotypes [145]. The Monte Carlo method was used to test the significance of LD in the complete data set of each population stratified by geographic origin of the infection (source health district). In this study, 10 000 permutations were completed. LD values range from 0 (no loci in LD) to 1 (all loci in LD). Pairwise standardised index of association (ISA) over all loci was assessed to determine whether the observed pattern of LD is due to a single or multiple pairs of loci.

2.2.8 Geospatial population substructure and genetic differentiation

To determine the influence of geographic origin of infections on genetic diversity, the analysis of variance (ANOVA) pairwise t-test was used to compare MOI, 1-F_{WS} and *He*

between the parasite populations stratified by source health district. Population substructure between the geographic areas was investigated by measuring Wright's F-statistics (F_{ST}), using the *adegenet* package [265] in R. Hendrick's G_{ST} and Jost's D, were calculated using the mmod package [266] in R. Genetic differentiation between populations ranges from 0 to 1 representing absence of to complete differentiation respectively. The Monte Carlo method was used to test the significance of pairwise F_{ST} between source health districts. In this study, 999 permutations were completed. Additionally, discriminant analysis of principal components (DAPC) using the *adegenet* package in R software was used to confirm signatures of population structure [265, 267]. DAPC infers population structure based on whether haplotypes (estimated from multi-locus genotypes generated from all major and minor allele data) clustered into distinct genetic populations. Unlike the traditional principal component analysis (PCA) which identifies linear axes that explain the most variability in all groups together, DAPC seeks to detect the linear axes which explain the most between-group variability in data [265]. K-means clustering was used to detect the number of inferred genetic clusters in the parasite population, and the best number of clusters chosen was that with the lowest associated Bayesian Information Criterion (BIC). To prevent over fitting of clusters, the optimal number of principal components (PC) to be retained was confirmed by cross validation of the DAPC. Data was divided into a training set (90 % of data), and a validation set (10 % of data), and members of each of the identified clusters were stratified by random sampling to ensure that at least one member of each cluster is represented in both training and validation sets. DAPC was then performed on the training set with variable numbers of PCs retained. The extent to which the analysis was able to accurately predict group memberships of individuals in the validation set was used to identify the optimal number of PCs to be retained. Sampling and DAPC procedures were repeated 1000 times at each level of PC retention, and the optimal number of PCs retained was associated with the lowest root mean square error. The resultant clusters were then plotted in a scatterplot of the first and second linear discriminants of DAPC.

2.2.9 Characterising pairwise genetic relatedness between infections

To determine the genetic connectivity/relatedness of pairs of infections including all alleles detected in both monoclonal and polyclonal infections of successfully genotyped samples, a modified identity by state (IBS) metric was used [122]. Overall, pairwise IBS was calculated as the average of locus specific estimates under the assumption of

independent loci. This metric was calculated using the formula: $IBS = \frac{1}{n}\sum_{i=1}^{n}\frac{Si}{XiYi}$ where *n* is the number of genotyped loci; S_i is the total number of shared alleles at locus *i* between samples *X*; and Y_iX_i is the number of alleles in sample *X* at locus *I;* and Y_i is the number of alleles in sample *Y* at locus *i*. A total of 85078 pairs of infection within the Vhembe District dataset were analysed and highly related infection pairs above the cut-off of IBS ≥ 0.5 [122] identified. Pairwise comparisons of relatedness of parasite pairs were then grouped into two categories depending on whether they occurred between two parasites isolated from individuals who acquired infections in the same (within) source health district or between two parasites from individuals who acquired infections from different (between) source health districts respectively. ANOVA pairwise t-test was used to compare differences in the proportions of highly related infections within and between the groups.

2.2.10 Influence of temporal variation on genetic diversity

To determine the influence of temporal changes in transmission on genetic diversity, the ANOVA pairwise t-test was used to compare MOI, 1-F_{WS} and *He* between the parasite populations stratified by year of transmission. Finer scale stratification by month of infection was also assessed to determine how MOI changes in the dry and wet transmission seasons and pairwise t-tests were used to compare MOI between the different seasons.

2.2.11 Assessing the impact of control interventions on the complexity of infections

The impact of the main strategy for control, indoor residual spraying (IRS) on the within-host diversity of *P. falciparum* was also assessed based on the number of unique parasite genotypes (MOI) identified in individuals from sprayed *vs.* unsprayed households. In the Limpopo Province, like the rest of South Africa, IRS typically takes place at the start of each malaria season (wet season between September and May), with the number of households to be sprayed determined by the provincial malaria control programme based on factors such as the number of housing structures within the malaria endemic area, availability of insecticide and insecticide resistance data, as well as malaria prevalence data for the previous year [233].

2.2.12 Assessing the association between polyclonality and anti-malarial drug resistance mutation prevalence

To determine whether there was any association between polyclonality (determined using microsatellite genotyping) and anti-malarial drug resistance mutation prevalence, a random subset of *P. falciparum* positive samples from the Vhembe District for which parasite microsatellite genetic diversity data was available were subjected to further molecular characterisation by genotyping drug resistance markers: chloroquine resistance transporter (*crt76*), multidrug resistance drug-protein 1 (*mdr86)*, dihydrofolate reductase (*dhfr51, dhfr59, dhfr108* and *dhfr164*) and dihydropteroate synthase (*dhps436 dhps437, dhps581,* and *dhps540*) as described before [268]. The drug resistance genotyping had been performed by Dr. Shehu Shagari Awandu, during his PhD studies. The proportion of individuals carrying infections with anti-malarial drug resistance mutations was compared between individuals with monoclonal (MOI = 1) infections and those with polyclonal (MOI > 1) infections. An ANOVA test was performed to compare the differences in the number of anti-malarial drug resistance mutations carried by individuals with monoclonal versus polyclonal infections. The prevalence of individuals with polyclonal and monoclonal infections carrying specific anti-malarial drug resistance mutations was also assessed.

2.3 Results

2.3.1 Detection of malaria parasites using ultra-sensitive varATS and TARE-2 qPCR

To detect parasite density in RDTs to establish whether there would be sufficient genomic DNA quantities to conduct microsatellite genotyping, varATS and TARE-2 qPCR were performed on control samples. Both the varATS and TARE-2 methods had good qPCR efficiencies of 98.6 % and 94.5 % respectively as determined on the *P. falciparum* 3D7 culture control, which were comparable to previously published values within the desired efficiency range of 90 % -105 % [258]. Both methods could detect 0.1 parasites/µL of blood based on the standard curves obtained (Figure 2.1A) and could therefore both be used for quantification of parasite loads in the RDT samples. Due to a low DNA volume for a proportion of samples, it was not possible to quantify all samples. Each of the two qPCR assays detected a unique set of samples/infections. Therefore, there was no direct comparison done between the assays. Of the 109 samples quantified by the varATS

qPCR method, 98 (90 %) had parasite densities above the genotyping threshold of 10 parasites/µL, while 88 % (188/214) of samples quantified by the TARE-2 qPCR method had parasite densities above the genotyping threshold which suggests that most samples could therefore be genotyped. Median parasite density was 5160 parasites/µL of blood for the samples detected by varATS qPCR (n = 109) and 382.7 parasites/µL of blood for the TARE-2 qPCR (n = 214) (Figure 2.1B). The significant difference (*P*<0.0001, Mann Whitney test) in these parasite densities may be attributed to higher density samples being randomly selected for the varATS method compared to the TARE-2 method. Since the mean parasite densities of the subset of randomly selected quantified samples assayed using both methods were above the genotyping threshold of ≥ 10 parasites/µL of blood, this was an indication that most of the study samples collected would have sufficient quantities of DNA for microsatellite genotyping to be performed on them. Therefore all 1153 samples collected in the Vhembe District between 2016 and 2018 were therefore subjected to microsatellite genotyping.

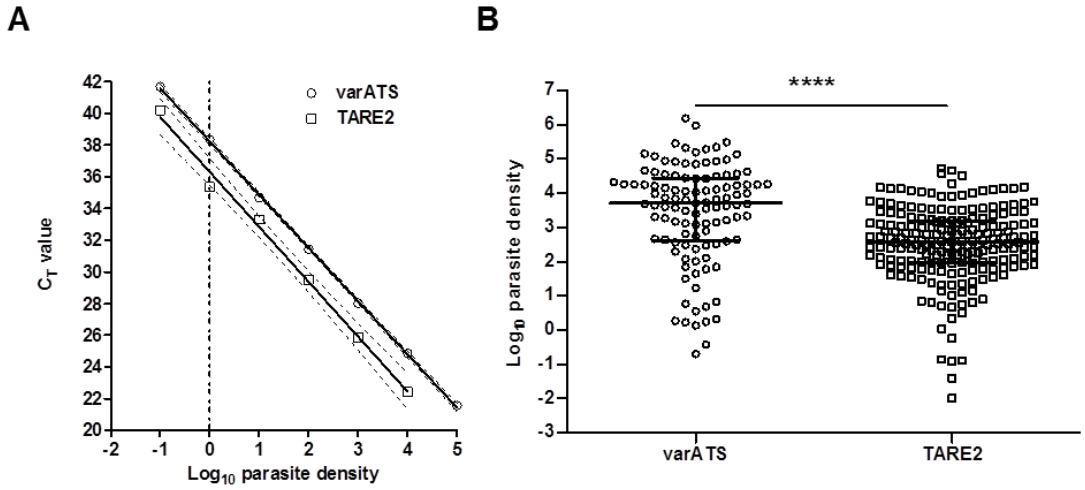


Figure 2.1 Outcomes of the varATS and TARE-2 qPCR methods. (A) Standard curves for both methods showing dilution series of known parasite densities of genomic DNA from *P. falciparum* 3D7 cultures plotted against the corresponding C_T values. Parasite density measurements are represented on a Log_{10} scale. The dashed lines parallel to the best fit lines indicate the 95 % confidence intervals of the best fit line, while the vertical dashed line shows from where the intercept values were derived. (B). Distribution of parasite densities of randomly collected samples quantified using the two qPCR methods. Each dot represents an individual sample and the corresponding parasite densities in Log_{10} scale. Summary lines are displayed with the middle line indicating median and the top and bottom lines indicating interquartile ranges of 75 % and 25 % respectively. Two-way *P* values (Mann Whitney test) are indicated in the plot where, ns: *P*>0.05; *: *P*≤0.05; **: *P*≤0.01; ***: *P*≤0.001; ****: *P*≤0.0001.

2.3.2 Study population and quality control of genotyped samples

Of the total 1153 *P. falciparum* RDT positive samples genotyped, 65 % (747/1153) had a good genotyping coverage with alleles detected at a minimum of 15 of the 26 microsatellite loci evaluated and were therefore included in the final sample set for population genetics analysis. The genotype accumulation curve (Figure 2.2) showed that there was an expected plateau in the number of multi-locus genotypes (MLGs) (unique haplotypes identified in individuals in the population) identified per number of loci, which confirmed that the panel of 26 markers used was informative and had sufficient discriminatory power to carry out genetic diversity analysis in the Vhembe District parasite population. All unique MLGs were detected using at least 17 markers as indicated by the 100 % mark (Figure 2.2).

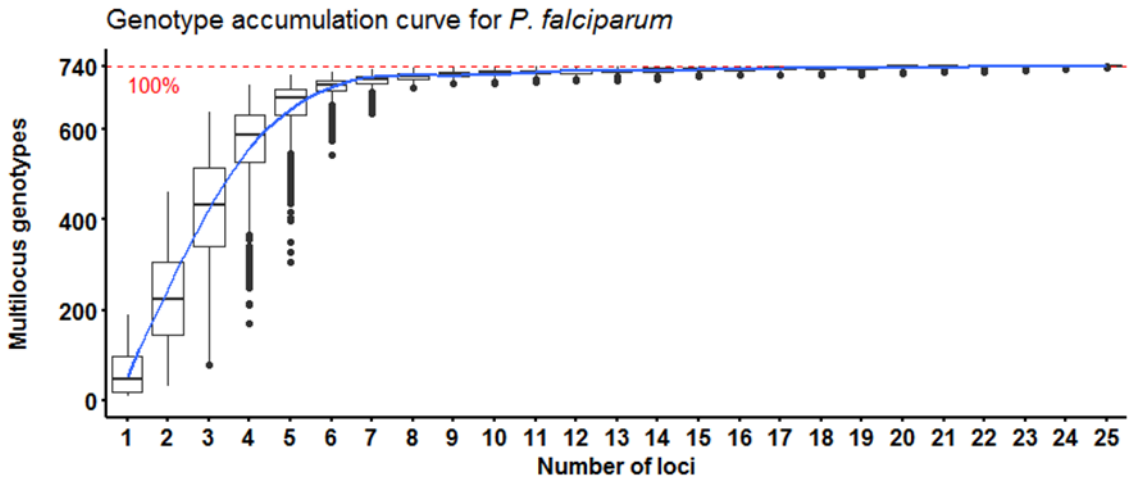


Figure 2.2 Genotype accumulation curve for the *P. falciparum* population in the Vhembe District. The number of loci was randomly sampled 1000 times without replacement up to n − 1 loci. The red dashed line represents 100 % of the multilocus genotypes in the parasite population.

2.3.3 Sample demographics

The highest proportion (56 %, 418/747) of samples successfully genotyped were collected in 2017 (Figure 2.3), predominately from Mulala Clinic (41 %, 306/747). After linking genotyped data with patient information from the LCM database, samples were stratified based on the source of infection health district (Figure 2.3). This is because it was established that patients that acquired their infections from similar source areas did not necessarily visit the same clinics for testing and treatment therefore genetic differentiation/separation of parasites using clinic locations would not be clear due to parasites mixed from different source locations.

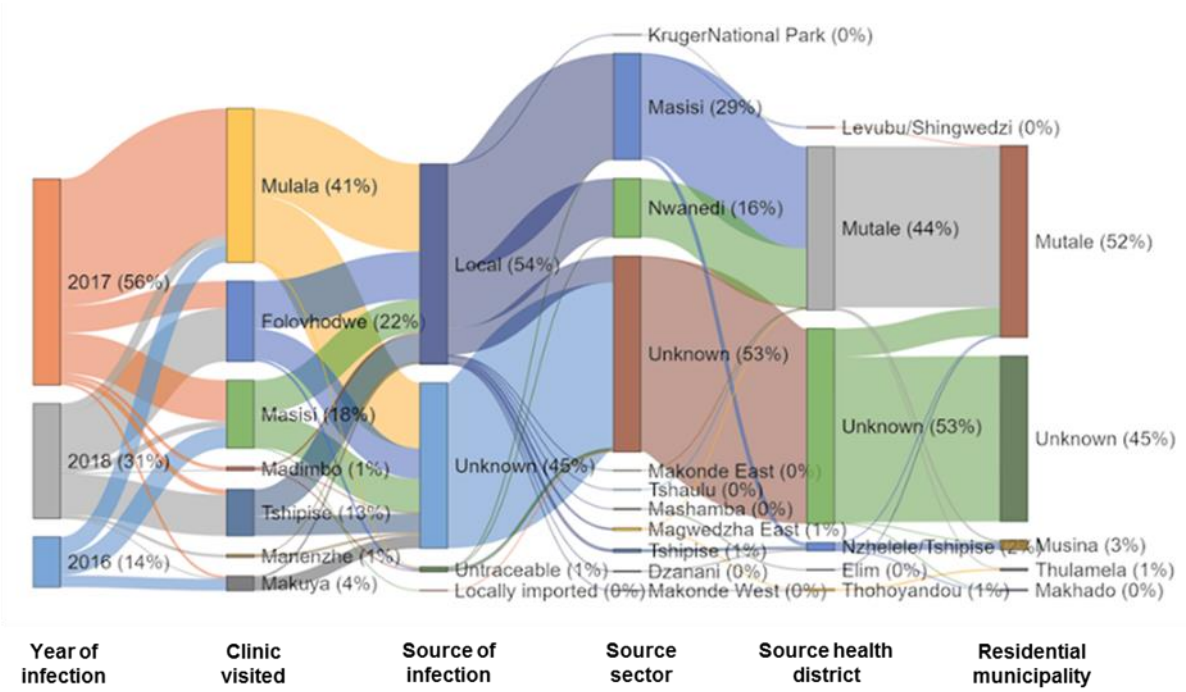


Figure 2.3 Sankey diagram linking the sampling period to the geographic locations where *P. falciparum* infections were acquired in the Vhembe District, Limpopo Province, South Africa. This illustration represents the distribution of the proportion of infections from the year of infection to the clinic that patients visited and from where RDT samples were then collected. Information on the source of infection is derived from linking RDT sample identities to patient unique identifiers on the LCM database. Samples whose patient unique identifiers could not be linked to the RDT identity were classified as "Unknown" for the source of infection, together with source sectors, source health districts and patient residential municipalities where information was unknown. The thickness of the linking nodes is proportional to the number of samples in each section.

The level of parasite mixing between samples collected from the different clinics was supported by the diversity statistics (Table 2.1) which showed no significant differences (P>0.05, ANOVA pairwise t-tests) in within-host (MOI and 1-Fws) and population level (He and LD) diversity between the clinics. This suggests mixed parasite populations. It was therefore, concluded that source health districts may be more informative for geospatial analysis than clinics and would therefore be used for onward geospatial analyses in this Chapter.

Table 2.1 *P. falciparum* genetic diversity statistics of RDT samples collected from selected clinics during different years in the Vhembe District.

Clinic	Year	n	MLG	MOI	1-Fws	He	LD
Folovhodwe		**160**	**160**	**2.17**	**0.233**	**0.74**	**0.144**
	2017	2	2	2.00	0.240	0.76	NA
	2018	158	158	2.33	0.225	0.74	0.145
Madimbo		**7**	**7**	**1.75**	**0.123**	**0.73**	**0.072**
	2017	7	7	1.75	0.123	0.73	0.072
Makuya		**27**	**27**	**1.75**	**0.142**	**0.71**	**0.195**
	2016	25	25	2.00	0.156	0.68	0.209
	2017	2	2	1.50	0.128	0.88	NA
Manenzhe		**9**	**9**	**2.00**	**0.212**	**0.75**	**0.169**
	2018	9	9	2.00	0.212	0.75	0.169
Masisi		**138**	**135**	**1.89**	**0.184**	**0.70**	**0.154**
	2016	91	87	1.93	0.194	0.73	0.165
	2017	36	36	1.83	0.159	0.72	0.108
	2018	12	12	1.92	0.200	0.73	0.101
Mulala		**311**	**308**	**2.25**	**0.210**	**0.74**	**0.138**
	2016	87	85	2.45	0.219	0.73	0.184
	2017	224	222	2.05	0.200	0.73	0.131
Tshipise		**95**	**94**	**1.93**	**0.197**	**0.72**	**0.152**
	2017	8	8	1.75	0.161	0.67	0.047
	2018	86	83	2.10	0.233	0.72	0.165
Total		**747**	**742**	**2.13**	**0.220**	**0.73**	**0.144**

n = number of isolates genotyped; MLG = number of unique haplotypes (or multilocus genotypes); He = heterozygosity; MOI = multiplicity of infection; 1-Fws = outbreeding; LD = linkage disequilibrium.

Patient demographic and travel history information revealed that 54 % (403/747) of the infections in the sample set were locally acquired (Figure 2.3), predominately from Masisi (29 %; 217/747) and Nwanedi (16 %; 120/747) source sectors, both situated within the Mutale health district. This correlates to residential status, with 52 % (388/747) of the individuals residing in Mutale. Forty five percent of the genotyped samples, however, could not be linked to source health districts, due to missing or illegible unique identifiers and were therefore classified as "unknown". It should be noted that during the initial phase of this study, no patient information was collected, with only year of infection and healthcare facility name recorded. The geographical distribution of the source health districts with the proportion of samples collected per year is represented in Figure 2.4. The highest number of RDT samples (200/747) was collected in 2017 from the Mutale source health district, with the lowest number (1/747) collected in 2017 from Thohoyandou. A

slightly higher proportion (54 *vs.* 46 %) of females compared to males was sampled during the study with a median age of 23 years, ranging between 19 and 65 years.

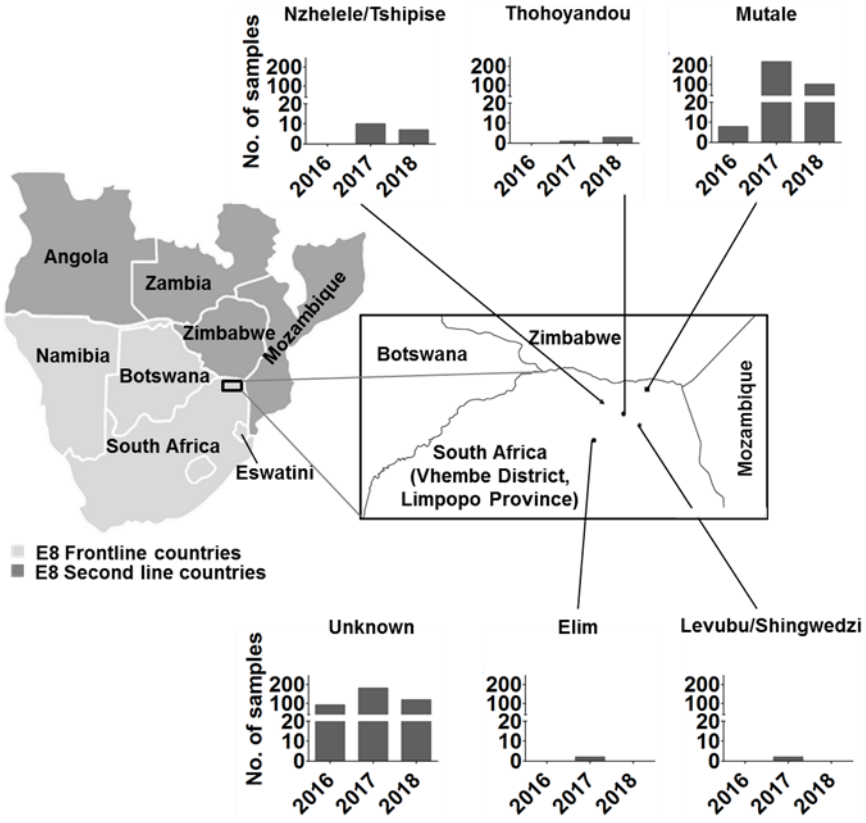


Figure 2.4 Study area and design. Map of the study site in the Vhembe District, Limpopo Province, South Africa. Bar graphs show the proportion of samples collected from the different source health districts and the years of transmission. The study district shares borders with Zimbabwe, Botswana, and Mozambique. South Africa is shown alongside other malaria endemic countries in the E8 region.

2.3.4 Microsatellite genotyping indicates parasite complexity

To characterise the level of genetic diversity within individual infections, MOI and outbreeding were assessed. Overall, a high proportion (66 %; 493/747) of polyclonal infections with a moderate mean MOI = 2.13 was observed in the genotyped samples (Figure 2.5A). This suggests a moderate complexity of infection within individual samples and thus moderate to high within-host diversity in the parasite population. Mean MOI did not differ significantly (P = 0.73, ANOVA, n = 353 excluding unknowns) between male (mean MOI = 2.11) and female participants (mean MOI = 2.16). The different age groups also exhibited similar mean MOIs (from 2 to 2.17) which were not significantly different (P = 0.94, ANOVA, n = 353 excluding unknowns). Parasite density and MOI were poorly

correlated (Pearson's r = 0.07 [95 % CI: -0.09-0.23], *P* = 0.4042, t-test, n = 313), which suggests that MOI is influenced by other factors other than parasite density (Figure 2.5B).

Similarly, to the 34 % (254/747) of monoclonal infections observed, only 40 % (299/747) of samples had 1-F_{WS} < 0.05 (Fws ≥ 0.95, indicating effectively clonal infections, Figure 2.5C). Mean outbreeding (1-F_{WS}) was low at 0.22, with only 33 % of samples with a stringent 1-Fws value of ≥ 0.30 (Fws ≤ 0.70), suggesting that these were the most highly diverse infections (Figure 2.5C). A significant positive relationship (Pearson's r = 0.85 [95 % CI: 0.83-0.87], *P* < 0.001, t-test, n = 747) was seen between outbreeding (1-F_{WS}) and MOI (Figure 2.5D), which suggests that both metrics support the presence of within-host diversity.

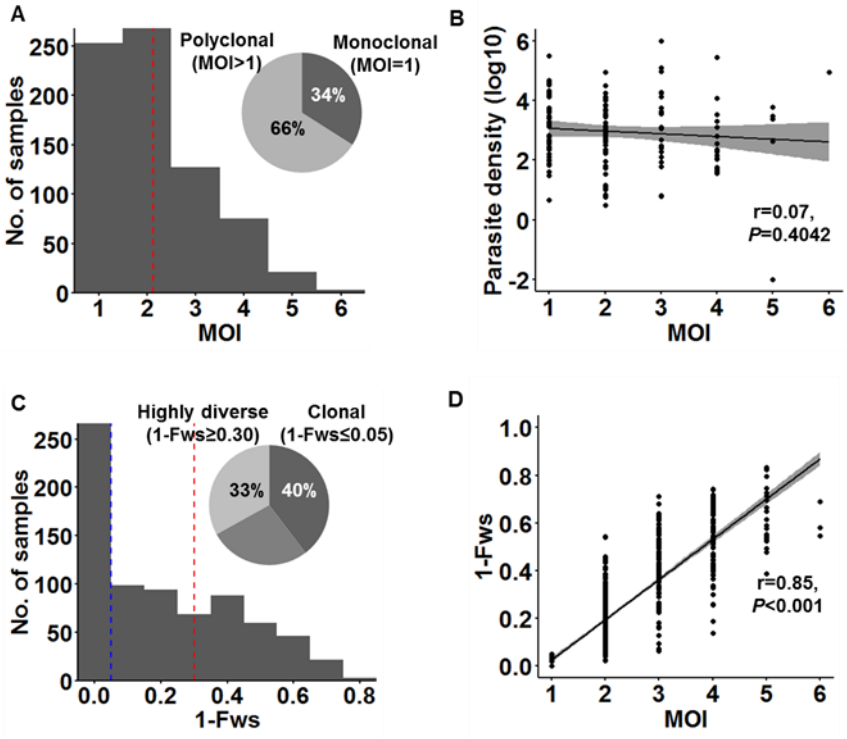


Figure 2.5 Within-host diversity of *P. falciparum* parasites in Vhembe District, Limpopo Province, South Africa. (A) Multiplicity of infection (MOI) measured as the second highest number of alleles detected at any of the 26 loci. The dashed red line indicates the mean MOI = 2.13. The inserted pie chart shows proportion of single (34 %) to multiple (66 %) infections. (B) Distribution between MOI and parasite density. The linear regression curve (95 % CI) shows dependence between parasite density and MOI, measured by Pearson's correlation. (C) Within-host diversity index (1-Fws). 1-Fws value shows outbreeding and a value of 0 indicates a perfect clone. The dashed blue line indicates the cut-off of 1-Fws ≤ 0.05 which represents samples with clonal infections. The dashed red line indicates the cut-off of 1-Fws ≥ 0.30 which represents samples with high genetic diversity. The inserted pie chart shows proportion of clonal (40 %) to highly diverse (33 %) infections. (D) Distribution of MOI vs 1-Fws. The linear regression curve (95 %CI) shows the linear dependence between 1-Fws and MOI, measured by Pearson's correlation.

2.3.5 Parasite population reflects moderate to high genetic diversity

To characterise parasite genetic diversity on a population level, the number of unique haplotypes in the population, allelic richness and the level of heterozygosity were analysed. Almost all samples (99 %, 742/747) analysed had unique haplotypes, indicative of high levels of outcrossing and a highly diverse parasite population. This was supported by a high allelic richness of an average of 12.2 across all 26 loci (Figure 2.6A), and a moderate to high mean *He* of 0.74 (Figure 2.6B) that indicates frequent recombination of different parasite clones. This suggests a larger *P. falciparum* population than that expected in a low-transmission setting, but is consistent with the Vhembe District being the highest transmission setting in South Africa. Locus PfPK2 was the most diverse marker (*He* of 0.91) and contributed to the high genotypic diversity whereas locus AS1 was the least diverse (He = 0.44). Overall genotypic evenness was high at 0.74 (equally abundant genotypes per locus yield a value of 1 and where a single genotype is dominant at a locus, the value is closer to 0). Locus Ara2 had the highest proportion of equally abundant genotypes (genotypic evenness of 0.86) and therefore contributed to the most genotypic evenness while locus AS7 had the least genotypic evenness at 0.58. Overall, the parasite population exhibited a moderate to high genotypic richness, evenness, and diversity.

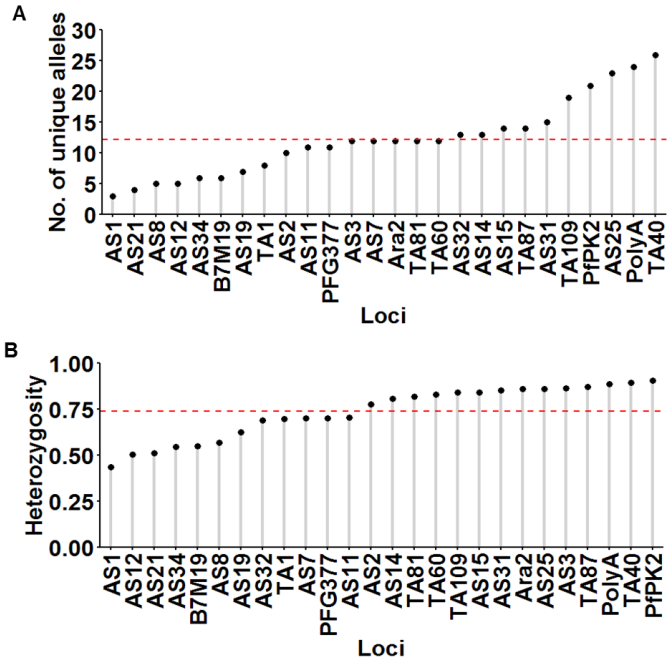


Figure 2.6 Population level diversity of *P. falciparum* parasites in the Vhembe District. (A) Distribution of number of unique alleles (allelic richness) detected in 26 microsatellite loci. The dashed red line indicates the mean = 12.2. (B) Population level genetic diversity measured as the distribution of heterozygosity in 26 microsatellite loci. The dashed red line indicates the mean heterozygosity = 0.74.

Additionally, a low LD (standardised index of association, ISA = 0.08) was observed between alleles of the *P. falciparum* haplotypes (Table 2.2) and this was not due to a single locus. The observed ISA value fell outside of the re-sampled distribution expected under no linkage when compared to histograms showing results of 10 000 permutations. The Monte Carlo method was used to test the significance of LD in the complete data set and alleles of monoclonal infections (n = 253) were linked across loci with *P* = 0.0001. For all infections including polyclonal infection, LD was also low at 0.138 but significant (*P* = 0.0001, n = 747). This was emphasized by a small proportion (0.26 %, 221/85078) of pairwise infections in the sample set being highly related (IBS ≥ 0.5). The significantly low LD therefore indicates high recombination of distinct parasite clones which therefore supports the moderate to high within-host diversity observed in Vhembe and is consistent with the presence of some degree of local transmission.

Table 2.2 Multilocus linkage disequilibrium in *P. falciparum* populations of Vhembe district.

Population	All infections		Single clones	
	n	ISA (*P* value)	n	ISA (*P* value)
Mutale	329	0.14 (0.0001)	111	0.03 (0.2633)
Unknown	393	0.14 (0.0001)	134	0.12 (0.0001)
Levubu/Shingwedzi	2	NA	1	NA
Thohoyandou	4	0.08 (0.0319)	3	-0.11 (1)
Nzhelele/Tshipise	17	0.14 (0.0001)	4	0.05 (0.3665)
Elim	2	NA	0	NA
TOTAL	**747**	**0.14 (0.0001)**	**253**	**0.08 (0.0001)**

n = number of isolates; ISA = standardised index of association; NA = not applicable
P value = The Monte Carlo method was used to test the significance of LD.

2.3.6 Parasites are fragmented based on their level of within-host diversity

To further evaluate the genetic relatedness between the parasite genotypes and explore the structure of parasite populations, k-means clustering was employed based on individual multi-locus genotype discrimination. Eight genetic clusters were identified in the parasite population with the parasite populations in genetic clusters 1, 2, 4, 7 and 8 observed using DAPC (Figure 2.7A) separated by linear discriminant 1 (LD1) from parasites in clusters 5 and 6. Linear discriminant 2 (LD2) further separated clusters 2, 4 and 8 from clusters 1 and 7. The proportion of correct assignment of the haplotypes to each inferred cluster ranged between 0.95 and 1.

The DAPC analysis could be performed with missing data in place with missing data which was randomly distributed in the data set basically replaced by the mean allele

frequency in the case of multi-locus genotype (MLG) calculations. MLGs were generated from all major and minor allele data. Although samples with missing genotypes/data were included in the DAPC analysis, samples with 83 % and 90 % of data were missing for the PfPK2 and TA1 loci, respectively, were excluded.

Specific infection source districts were not necessarily linked to particular genetic clusters observed from the DAPC analysis, as infections from the different areas were represented/distributed in the eight different inferred genetic clusters (Figure 2.7B), implying free gene flow and possible parasite mixing between these sites. The spatial genetic variation was underpowered due to the high proportion of samples whose source health district was unknown.

Interestingly however, clusters 5 and 6 contained the majority (98 %, 249/253) of monoclonal (MOI = 1) infections (Figure 2.7C). This was consistent with the mixture of mostly clonal (1-Fws < 0.05) and less diverse (medium, 1-Fws > 0.05 but < 0.30) infections in clusters 5 and 6 (Figure 2.7D). Clusters 1 and 2, and to a lesser extent cluster 7, appeared fragmented and contained only highly diverse infections (1-Fws > 0.30, Figure 2.7D). This indicates that although most of the parasite population is indeed structured/fragmented, the most highly diverse infections result in separation of the parasite population. The moderate genetic diversity and MOI, and less pronounced population structure are all indicative of constant transmission at overall relatively moderate levels.

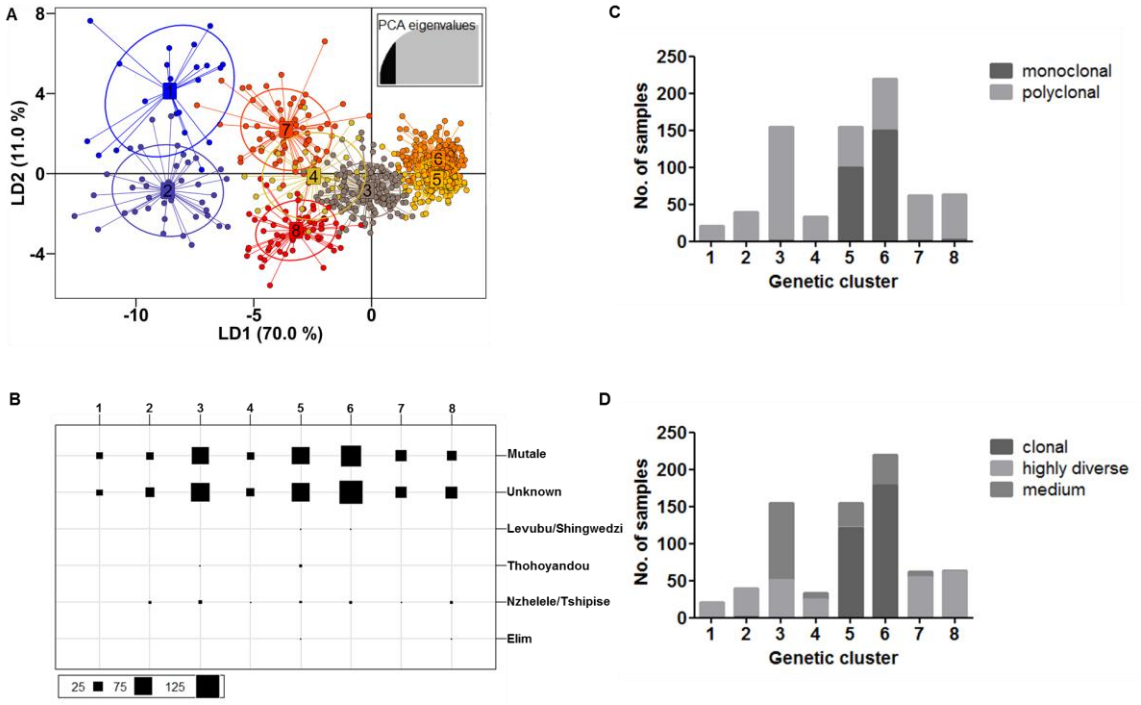


Figure 2.7 Fragmentation of parasites based on inferred genetic clusters. (A) Scatterplot of the discriminant analysis of principal components (DAPC) based on discrimination of the 8 inferred genetic clusters of *P. falciparum* populations. Samples whose source of infection is unknown are included. Individual multi-locus genotypes appear as dots. Colours and lines represent population membership. Analysis is based on retention of 50 principal components (top right insert). (B) Assignment of infection source health districts to the inferred genetic clusters. Columns correspond to the inferred genetic clusters, and rows correspond to the infection source health districts. The solid black squares indicate the number of individuals in each category with the size of each square being directly proportional to the number of individuals. (C) Proportion of monoclonal (MOI = 1) to polyclonal (MOI > 1) infections in the fragmented genetic clusters. (D) Proportion of clonal (1-Fws ≤ 0.05) to highly diverse (1-Fws ≥ 0.30) infections. Medium represents infections with 1-Fws > 0.05 but < 0.30.

2.3.7 No geospatial correlation exists to explain genetic diversity

To determine whether there was a correlation between the level of genetic diversity and the geographic origin of infections (source health districts), geospatial analyses were performed. Overall, the levels of within-host and population level diversity were not influenced by where infections were reported since the mean MOI (global ANOVA P = 0.36, n = 747, Figure 2.8A); the level of outbreeding (global ANOVA, P = 0.27, n = 747, Figure 2.8B) and heterozygosity (global ANOVA P=0.23, n = 747, Figure 2.8C) between the different source health districts did not significantly differ. Only differences in the MOI from Elim and Thohoyandou were observed compared to the overall mean MOI (pairwise t-test, n = 747, P≤0.01 and P ≤ 0.05, respectively), which may be due to small sample size from these sites.

The limited spatial structure observed in the DAPC analysis was confirmed by the very low F_{ST} values (Figure 2.8D), with only the Elim and Levubu/Shingwedzi districts sharing significant F_{ST} with two other districts each, implying some degree of separation of the parasites based on geographic origin but again associated with small sample size from these areas. A Hendrick's G_{ST} at -0.0394 and Jost's D at -0.0122 supports general parasite mixing within Vhembe District. This overall lack of separation of parasites based on geospatial data agrees with patient demographic and travel history information, with 99 % of the infections locally acquired of which 95 % are within the Mutale health district. This also correlated to residential status, with 93 % of the individuals residing in Mutale. Additionally, 1-IBS analysis showed that some infections were genetically connected within and between source areas, with highly related infections significantly higher (P < 0.0001, ANOVA, n = 221) within source areas (mean 1-IBS = 0.76 ± 0.022) compared to between source areas (mean 1-IBS = 0.56 ± 0.007) which confirms some level of local transmission.

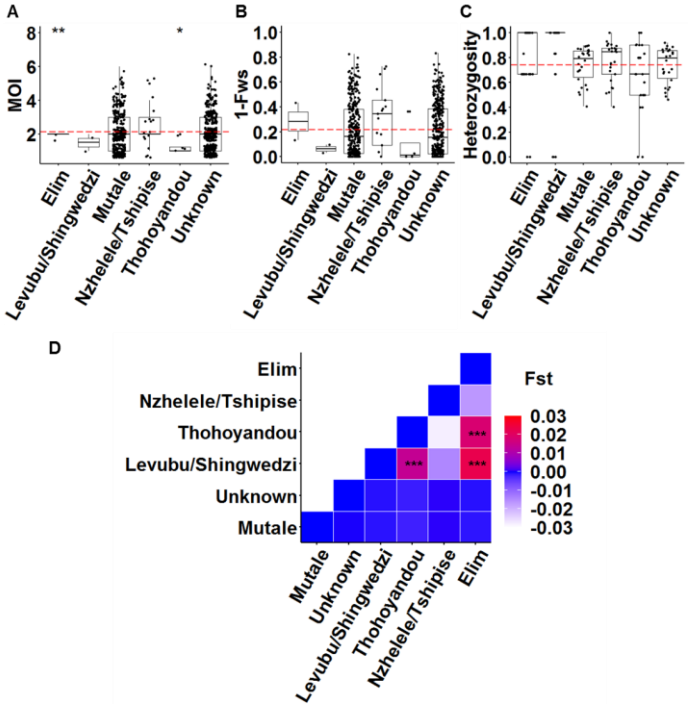


Figure 2.8 Geospatial variation of within-host and population level genetic diversity of *P. falciparum* parasites in the source health districts in Vhembe. (A) Multiplicity of infection (MOI) measured as the second highest number of alleles detected at any of the 26 loci across the different source health districts. The dashed red line indicates the mean MOI = 2.13. (B) Outbreeding, (1-F_{WS}) measured across the different source health districts. The dashed red line indicates the mean 1-F_{WS} = 0.22. (C) Population level genetic diversity measured as the distribution of heterozygosity in 26 microsatellite loci across the different source health districts. The dashed red line indicates the mean heterozygosity = 0.74. (D) Heatmap showing matrix of between source health district pairwise F_{ST}. The range of F_{ST} values for the pairwise comparisons is shown in the legend and significance was tested using the Monte Carlo method. Pairwise P values (t-test) are indicated where, not significant: P>0.05; *: P≤0.05; **: P≤0.01; ***: P≤0.001; ****: P≤0.0001.

2.3.8 Transmission dynamics was not temporally influenced

To determine the influence of seasonality on the level of parasite genetic diversity, correlations between the two were made at different temporal scales. The level of genetic diversity did not differ between the years of transmission as demonstrated by global ANOVA *P* values (n = 747) for MOI (Figure 2.9A), outbreeding (Figure 2.9B) and heterozygosity (Figure 2.9C) of 0.90, 0.77 and 0.07, respectively, providing no evidence for differing transmission intensity during that period. No genetic differentiation (F_{ST} range -0.00005 to 0.0003) between parasite populations from the different years was observed, which suggests that the malaria outbreak experienced in 2017 was not due to an introduction of a completely distinct/new parasite population to Vhembe.

Stratification of MOI distribution by transmission season (month of infection) (Figure 2.9D) showed that infections were persistently complex throughout the year. Despite an approximately 15-fold reduction in the number of cases and notable decrease in rainfall levels between the wet (high-transmission) and dry (low-transmission) seasons (Figure 2.9D), mean MOI was also not significantly different (*P* = 0.42, Bonferroni *P* adjustment method) between the two seasons (wet season mean MOI = 2.14 ± 0.056, and mean MOI = 1.96 ± 0.178 in the dry season). Infections were similarly complex (mean MOI ± SE = 2.15 ± 0.075 and 2.14 ± 0.101; *P* = 1, Bonferroni *P* adjustment method) between infections from sprayed or unsprayed households respectively suggesting the maintenance of similarly complex infections despite vector control implementation during wet seasons, therefore suggesting possible co-transmission of strains. These data may indicate that the mean MOI may be seasonally stable thus reflecting the continued residual transmission in the Vhembe District. Out of the 220 patients who knew which insecticides were sprayed in their households, 53 % (116/220), 44 % (97/220) and 3 % (7/220) reported use of DDT, Fendona® and K-Othrine® respectively. This data was verified from the LCM database. Mean MOI of infections in the houses sprayed with DDT, Fendona® and K-Othrine® were 2.19 ± 0.109, 2.13 ± 0.106 and 1.86 ± 0.404 respectively and they did not differ significantly (P > 0.05, ANOVA). Therefore, the complexity of infections and genetic diversity of the parasite population was stable - neither influenced by transmission season nor which insecticide was used for IRS.

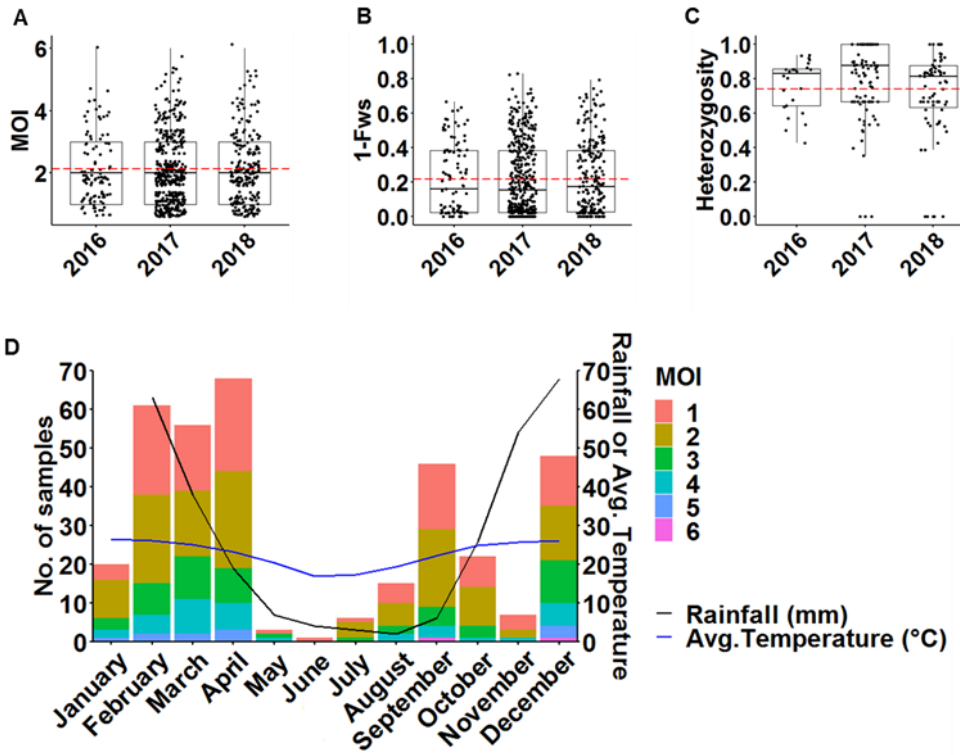


Figure 2.9 Temporal variation of within-host and population level genetic diversity of *P. falciparum* parasites in Vhembe district. (A) Multiplicity of infection (MOI) measured as the second highest number of alleles detected at any of the 26 loci across the different years. The dashed red line indicates the mean MOI = 2.13. (B) Within-host diversity index (outbreeding, 1-F_{WS}) measured across the different years. The dashed red line indicates the mean 1-Fws = 0.22. (C) Population level genetic diversity measured as the distribution of heterozygosity in 26 microsatellite loci across the different years from when the samples were collected. The dashed red line indicates the mean heterozygosity = 0.74. Pairwise *P* values (t-test) are indicated in all plots where, not significant ns: *P*>0.05; *: *P*≤0.05; **: *P*≤0.01; ***: *P*≤0.001; ****: *P*≤0.0001 (D) MOI distribution of collated samples throughout all the months of the year against different rainfall and temperature patterns.

In support of the presence of a stable parasite population in the Vhembe District, estimated population size, N_e, was 4475 (95 % confidence interval [CI], 1923–10194) and 10843 (95 % CI, 4660–24701), under IAM and SMM models respectively. There was no decline in the N_e over space and time (Table 2.3) which indicates a relatively stable parasite population. Additionally, under the SMM, there was no significant deviation of *He* based on the observed number of alleles (Wilcoxon test for heterozygosity excess, *P*-value = 1) revealed, suggesting no recent history of parasite population bottlenecks.

Table 2.3 Effective population size, N_e, across source health districts and year of infection. Estimates were calculated based on both the infinite-allele (IAM) and the stepwise (SMM) mutation models at 95 % confidence intervals (CI).

	IAM - N_e	**95 % CI**	**SMM - N_e**	**95 % CI**
ALL	4475	(1923, 10194)	10843	(4660, 24701)
Year of infection				
2016	4475	(1923, 10194)	10843	(4660, 24701)
2017	4251	(1826, 9684)	9998	(4296, 22775)
2018	4251	(1826, 9684)	9998	(4296, 22775)
Study site				
Mutale	4475	(1923, 10194)	10843	(4660, 24701)
Unknown	4475	(1923, 10194)	10843	(4660, 24701)
Levubu/Shingwedzi	4251	(1826, 9684)	9998	(4296, 22775)
Thohoyandou	3052	(1312, 6953)	6015	(2585, 13701)
Nzhelele/Tshipise	4251	(1826, 9684)	9998	(4296, 22775)
Elim	6289	(2703, 14327)	18868	(8108, 42980)

2.3.9 Correlation between polyclonality and anti-malarial drug resistance mutation prevalence

To assess the probability of individuals with polyclonal (MOI > 1) infections carrying infections with mutant drug resistance genes, a subset of samples was assessed. A total of 24 samples had complete data for both the microsatellite parasite genetic diversity and drug resistance molecular markers (*crt76*, *mdr86*, *dhfr51*, *dhfr59*, *dhfr108*, *dhfr164* and *dhps436*, *dhps437*, *dhps540* and *dhps581*). The proportion of mutant (59 %, 80/136), and wildtype/mutant (57 %, 8/14) genotypes was higher in individuals carrying polyclonal infections compared to those with monoclonal infections. The number of drug resistance genes with mutations per individual infection was also not influenced (P = 0.79, ANOVA) by whether an individual had a monoclonal or polyclonal infection (Figure 2.10A). It was observed that all polyclonal and all monoclonal infections harboured *dfhr108*, *dhps540* and *dphs580* drug resistance gene mutations (Figure 2.10B). None of the infections harboured *dhfr164* and *mdr86* mutations. The proportions of polyclonal infections that harboured *crt76* (29 %, 4/14) and *dhps436* (29 %, 4/14) mutations was, however, lower than those of monoclonal infections (50 %, 5/10 for *crt76* and 40 %, 4/10 for *dhps436*) respectively. This may suggest that individuals with polyclonal infections were less likely

to harbour chloroquine resistance mutations than individuals with monoclonal infections. On the other hand, higher proportions of 100 % (14/14), 86 % (12/14) and 86 % (12/14) of polyclonal infections harboured *dhps437*, *dhfr51* and *dhfr59* mutations compared to 90 % (9/10), 70 % (7/10) and 70 % (7/10) in monoclonal infections respectively. This may show that individuals with polyclonal infections were more likely to harbour SP resistance mutations than individuals with monoclonal infections. Therefore, while the prevalence of drug resistant gene mutations in individual infections may not be influenced by having polyclonal infections, the class of drug resistance genes that are more prone to carrying mutant genes are associated with polyclonal infections.

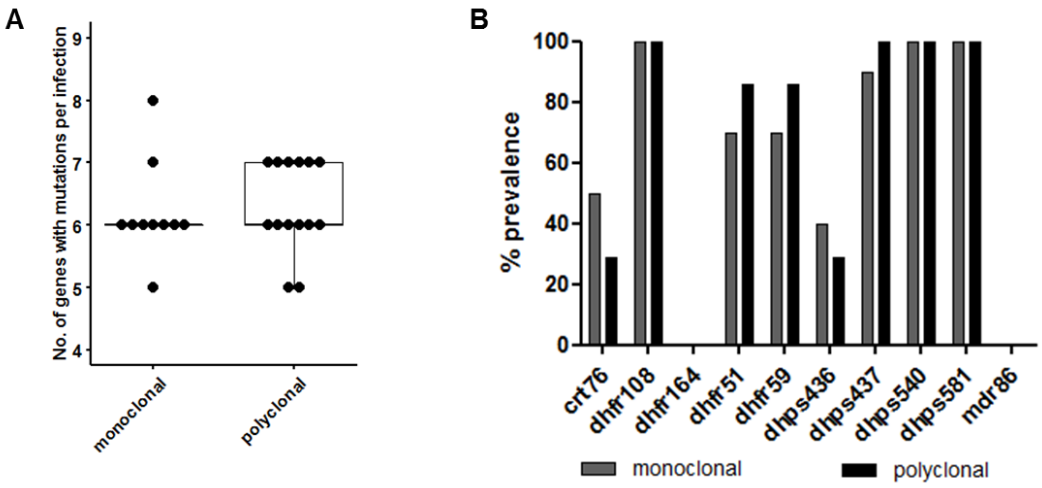


Figure 2.10 Correlation between polyclonality and mutation in anti-malarial drug resistance genes. (A) Comparison of the number of drug resistance genes with mutations per individual in polyclonal vs monoclonal infections. *P* = *P* values (ANOVA) are indicated in the plot where, not significant ns: *P*>0.05 is not shown; *: *P*≤0.05; **: *P*≤0.01; ***: *P*≤0.001; ****: *P*≤0.0001. In this case there is no significant difference between polyclonal and monoclonal infections. (B) Prevalence of polyclonal and monoclonal infections carrying specific drug resistance genes with mutations.

2.4 Discussion

The impact of *P. falciparum* parasite genetic diversity on malaria transmission intensity is only now being clarified in South Africa. This is the first study in South Africa to use multilocus genotyping to demonstrate that the continued residual malaria transmission in a malaria hotspot within South Africa is associated with parasite diversity.

Although South Africa is overall a low-transmission setting, the Vhembe District is a transmission hotspot. Data from this study show that the parasite population within this hotspot behaves genetically like those typically seen in high-transmission settings (based on incidence data) such as Guinea, Mali, and The Gambia [145, 269-271] with a high

degree of complexity and diversity. Such parasite complexity is typically associated with high levels of gene flow between areas of different transmission intensities that serve to increase allelic richness by introducing new alleles into the population, thereby increasing the level of heterozygosity in the parasite population [188, 255]. Although the Vhembe District is located along the border with Mozambique and Zimbabwe, very limited parasite genetic evidence exists for imported malaria cases [268] and rather, the high level of heterozygosity, with strong gene flow and a clear indication of gene mixing observed in this study suggest local transmission in a hotspot area where relatively higher transmission occurs compared to other areas in the country. This is supported by a marked level of gene flow and parasite mixing between parasite populations from the different source areas within the district, resulting in a low but significant LD, with frequent and random association between alleles and a panmictic parasite population [272]. Geospatial and temporal variances had little effect on within-host and population diversity in the Vhembe District suggesting that genetic diversity was stable over space and time. This observation was supported by an absence of recent parasite population bottlenecks, suggestive that the diversity of the parasite population is shaped by other factors other than transmission intensity. The fact that mean MOI remained seasonally stable may therefore reflect the contribution of complex infections to continued residual transmission in the Vhembe District.

A limitation of this study was that many samples could not be linked to their district of origin due to missing patient information, thereby leaving the geospatial analyses underpowered in most study districts. Additionally, the sampling plan that targeted few selected clinics in a known hotspot area based on case incidence data, may have introduced a sampling bias as most of the samples came from the Mutale district. It was therefore not possible to get genetic representation of parasites from the other source districts and those seeded through importation from neighbouring high-transmission settings reported in the province if any and what their contribution to the local transmission dynamics is.

Hotspot areas in other low-transmission settings such as Malaysia have demonstrated opposite trends where parasite complexity and heterozygosity are low, and LD is high using both microsatellite markers and merozoite surface protein variants for genotyping [153, 273]. In Madagascar, however, high levels of genotypic diversity and a high proportion of polyclonal infections were associated with a transmission hotspot using single nucleotide polymorphism genotyping [274]. These differences may have been due

to differences in technology platforms used for genotyping in comparison to this study or due to the challenge of non-standardisation of study designs in molecular epidemiology studies. Additionally, since “low” in Africa is very different than “low” in Asia (which is generally much lower based on for example entomological inoculation rates and parasite prevalence rates) it could have also been due to differences in underlying epidemiology of parasite transmission between Malaysia and Madagascar. In other low-transmission areas of southern Africa [122, 188, 255], where a similar panel of microsatellite markers was used, the level of genetic diversity was as high as that in the Vhembe District. However, in Eswatini and in the KZN province of South Africa, the high levels of within-host and population diversity and lack of parasite population structure could be explained by high levels of importation from neighbouring high-transmission areas. On the other hand, in north-eastern Namibia, while infections were genetically diverse, transmission was mostly as a result of locally acquired infections and fine-scale parasite fragmentation based on the geographic origin of parasites was observed [122]. Detectable genetic clusters of different genotypes mean if studies are strategically designed and patient parasite genetic and demographic data accurately linked in all remaining endemic areas in South Africa and neighbouring endemic countries, particularly at the border areas, parasites from different countries can easily be detected or traced back to their origins. This information can then be used to make better decisions on a national level and as a regional block on what interventions should be put in place and concentrated in which areas. Therefore, regional comparison of parasite genotypes will be informative to understand parasite mixing in the context of malaria transmission.

Some clones in an infection may exist at lower proportions than others due to either competitive suppression by other genotypes or possibly host specific selection due to immunity or receptor polymorphisms. In *P. chabaudi*, minor clones in mixed/multiple infections may produce as many, or more, oocysts than they would have as a single clone infection [275], highlighting that competitive stress may increase transmission of certain clones. Multiple distinct parasite clones are implicated in high gametocyte production and emergence of highly virulent and drug resistant parasite strains due to intense within-host competition [276-279]. While the relationship between parasite density and MOI is complex, the fact that the total parasite density did not increase with an increase in distinct parasite clones may be suggestive of within-host competition [280]. Alternatively, since both MOI and acquired immunity increase with exposure, this observation may suggest that as more strains are acquired, immunity becomes more effective at reducing parasite density [280]. This current study showed that a higher proportion of individuals with

polyclonal infections generally harbour mutant anti-malarial drug resistant genes compared those with monoclonal infections, however, the association for more virulent infections being caused due to this could not be ascertained as patient follow-up was not a part of the study design. Patterns of genetic diversity observed in individuals with anti-malaria drug resistance gene mutations could therefore not be correlated with treatment response or clinical outcome. Another limitation of this aspect of the study was that the sample size was also relatively small with limited geospatial information such as longitude and latitude coordinates of the individual samples making it difficult to do a spatial assessment of the drug resistance genes and was therefore not included in the final published manuscript. None-the-less, more clarity may be achieved by conducting a more important and detailed longitudinal study of the contribution of specific clones to parasite transmissibility and virulence in the Vhembe District.

The results generated in this study are a useful addition to the growing resource of *P. falciparum* genetic data in the southern African E8 region which will facilitate more detailed evaluations of cross-border parasite spread in future studies. The genetic data may also be useful to control programmes with decision making for malaria elimination as it can be used as a reference point for the assessment of the effectiveness of on-going interventions over time, selection and/or targeting of interventions. Furthermore, the genetic data may be used for the identification of imported cases and/or outbreaks, as well as monitoring for the potential spread of antimalarial drug resistance and potentially revealing infection patterns.

2.5 Conclusion

In this study, the impact of *P. falciparum* genetic diversity on continued malaria transmission intensity in the Vhembe District is demonstrated. The *P. falciparum* population is moderate to highly diverse and genetically complex, which is key and advantageous to the parasite's evolution and survival. This data could be informative as a reference point in evaluating the efficacy of strategic control interventions over time, aimed at eliminating residual malaria transmission in malaria transmission ‘hotspots’ in South Africa. Furthermore, this data can be used to identify imported cases and/or outbreaks, as well as monitor for the potential spread of antimalarial drug resistance. Linking data on *P. falciparum* genetic diversity from Vhembe/South Africa to that from neighbouring sub-Saharan countries in the E8 regional initiative would need to be investigated as the transmission dynamics in the region is not fully understood.

CHAPTER 3

GENETICALLY DIVERSE AND COMPLEX ASYMPTOMATIC IMPORTED *PLASMODIUM FALCIPARUM* CASES MAY SEED LOCAL TRANSMISSION IN PRE-ELIMINATION SETTINGS OF SOUTH AFRICA

Part of the data in this chapter was published as:
Raman, J., Gast, L., Balawanth, R., Tessema, S., Brooke, B., Maharaj, R., Munhenga, G., Tshikae, P., Lakan, V., Mwamba, T., **Makowa, H.**, Sangweni, L., Mkhabela, M., Zondo, N., Mohulatsi, E., Nyawo, Z., Ngxongo, S., Msimang, S., Dagata, N., Greenhouse, B., Birkholtz, L., Shirreff, G., Graffy, R., Qwabe, B. and Moonasar, M. 2020. "High levels of imported asymptomatic malaria but limited local transmission in KwaZulu-Natal, a South African malaria-endemic province nearing malaria elimination." Malaria Journal 19:152. doi: https://doi.org/10.1186/s12936-020-03227-3.

3.1 Introduction

The KwaZulu-Natal (KZN) Province of South Africa has made substantial progress in reducing its malaria burden and is on the verge of eliminating the disease [65]. Using a phased district-level approach to malaria elimination [66] guided by the World Health Organization's Global Technical Strategy for Malaria 2016–2030 [251], all three remaining malaria endemic districts (uMkhanyakhude, Zululand and King Cetshwayo; Figure 3.1) within the KZN Province have consistently reported less than 1 local malaria case per 1000 population at risk from 2008 to date, and are therefore targeted for elimination [65, 66, 255].

Low level focal transmission, however, persists in the province particularly in the uMkhanyakude district [255], which shares borders with Eswatini, and the southern region Mozambique. Despite regional spikes in cases caused by the 2017 upsurge, cases in the KZN Province remain low [281, 282]. However, the continued implementation of crucial vector control [247], case management interventions, and treatment with effective artemisinin-based combination therapy [243, 255], has not led to the attainment of malaria elimination in the province.

Mozambique contributes to roughly over 70 % of imported cases annually in South Africa [233, 283], and contributed to 4 % of the world's malaria cases in 2019 [1]. Although the

South African government through its health department has applied cross-border strategies to strengthen prevention and provide treatment services in the high risk areas with its neighbouring higher transmission countries exporting malaria [252, 284], frequent, continued and sometimes illegal cross-border migration from the neighbouring high-transmission countries such as Mozambique and Zimbabwe mostly due to economic migrants (mobile migrant populations) persist and thus pose a risk to local spread and a challenge to malaria elimination in South Africa [65, 255]. Some of these imported cases are sometimes missed since some illegal migrants for example, may not be able to access government health facilities due to a lack of valid identity/immigration documents and therefore such cases may not be captured through the national and/or provincial surveillance database. Therefore, this necessitates understanding the contribution/dynamics of imported infections carried by mobile migrant populations to sustained transmission. In this case, an imported malaria case is defined one whose source of infection can be traced back to an area outside of South Africa where the patient has recently travelled to or from [255].

Individuals from high-transmission acquired partial immunity to malaria due the frequent exposure to malaria infections, and often become asymptomatic carriers of the disease [10, 54]. This presents another challenge for malaria elimination in low-transmission settings where all malaria carriers should be detected and treated. A recent study in very low prevalence settings, found that low density asymptomatic infections make up to 70-80 % of the total number of reported cases and are responsible for 20-50 % of all infections from humans to mosquitoes [285]. Although the absolute density of infection necessary to transmit is unknown, sub-patent infections are known to transmit malaria [286, 287]. Asymptomatic carriers, with either sub-patent or patent infections are therefore an important source of malaria transmission and a major hindrance to elimination. Therefore, accurate characterisation of the numeric distribution of parasite densities in particularly migrant cases coming into elimination settings is essential for planning treatment, regional control, and elimination strategies.

Although several factors have been associated with residual/focal transmission in the southern African region, it is unclear to what extent the parasite genetic diversity of asymptomatic imported infections in very low-transmission areas of southern Africa drives local transmission. It is therefore of paramount importance to enhance our understanding of *P. falciparum* parasite biology and the contribution of genetic diversity in imported infections to sustaining transmission in KZN Province. In addition, parasite population

information could potentially assistance with the more accurate classification of local and imported cases, a requirement for elimination certification. Currently case classification is achieved using patient travel information collected via traditional surveillance methods which are highly dependent on human recall. Information gathered on source areas of infection is therefore often incomplete and/or inaccurate [123, 288].

Combining patient demographic data with parasite population genetic analysis allows us to see that parasites with high genetic relatedness/similarities cluster together in spite of their geographic origin making it easy to identify imported infections in low-transmission settings. Individuals from high-transmission settings often carry many parasite genomes/strains within them due to multiple mosquito bites from many different mosquito vectors [64, 120, 145]. As transmission intensity reduces so does the number of parasite genomes/strains circulating in the population [64, 120, 145]. This association between transmission intensity and genetic diversity is therefore a useful method to measure changes in malaria transmission [62, 120, 169]. However, in low-transmission African settings this relationship is not always clear due to the constant importation of cases from neighbouring higher transmission countries [188, 250]. While studies on the genetic diversity of the parasite have been initiated in southern Africa [122, 188, 227, 289], studies that assess the parasite's diversity in migrant asymptomatic carriers in the region are limited [188] with little information on such characterisation done in South Africa [255, 259, 268].

Therefore, this study firstly sought to quantify parasite densities in asymptomatic imported cases from highly mobile migrant populations in the KZN Province and determine how parasite density in this group is influenced by age of the infected individuals and within-host parasite diversity. Secondly, this study aimed to evaluate to what extent parasite genotypes of imported infections contribute to those of local parasites represented by those in the Limpopo Province of South Africa (characterised in Chapter 2) by resolving genetic differences and similarities within and between *P. falciparum* parasites identified in asymptomatic imported cases from the KZN Province. This data will provide information on the factors driving and sustaining low-level malaria transmission in the KZN Province and could serve as a national reference point for planning diagnosis, treatment, and surveillance of imported asymptomatic cases including mobile migrant populations in South Africa in order to achieve elimination.

3.2 Materials and methods

3.2.1 Ethical approval

Ethical approval for the Limpopo samples was obtained from the University of Pretoria, Faculty of Health Sciences Research Ethics Committee (Ethics Reference No. 406-2014) as described in Chapter 2.2.1. The National Institute for Communicable Disease (NICD) holds ethical approval from the KwaZulu-Natal Provincial Department of Health, the Health Ethics Review Committee of the University of Witwatersrand (M170869), and Advarra Research Compliance Solutions (Maryland, USA). Invited study participants in the KZN Province included adults and children between the ages of 2 to 18 years for whom written consent was obtained from their parents, guardians, and caregivers.

3.2.2 Study site & sample collection

The study participants from KZN Province were chosen from individuals that visited the informal KwaPhuza border market, located along the border between uMhlabuyalingana municipality in the uMkhanyakude District of the KZN Province of South Africa and the Maputo Province in Mozambique. Since the number of people that cross from Mozambique into South Africa through that border is unpredictable and unknown, the number of individuals selected for the study was based on a random sampling method with individuals entering the South African side of the border market every Wednesday over a 6-week period invited to participate in the study. Dried blood spot (DBS) samples were collected from all consenting individuals that tested positive for malaria by standard RDT (First Response™ Malaria Ag *P. falciparum* HRP2 Detection Rapid Card Test, Premier Medical Corporation Ltd, India) and highly sensitive RDT (Alere™ Malaria AG P.F. Ultra-sensitive, Abbott, USA). Munktell TFN cards (Munktell, Germany) labelled with unique patient identifiers similar to those of the participants' demographic data were used to collect the blood spots samples. The DBS samples were air-dried and individually packed into zip-lock packets containing desiccant until further processing at the NICD in Johannesburg. Individuals that tested positive for malaria by RDTs were immediately treated with artemether–lumefantrine by registered nurses based on the South African malaria treatment guidelines [290]. The recent travel history of these individuals was also captured by survey personnel. Limpopo Province study participants were selected as described in Chapter 2.2.2.

3.2.3 Genomic DNA extraction

DNA was extracted from DBS samples by first punching out three circles of approximately 3 mm diameter each and placing the punched circles into 1.5 ml wells of a 96-well plate. The Saponin-Chelex method [257] was then used to extract the DNA first by incubating the punched DBS in 1ml of 0.5 % saponin prepared in 1X PBS (calcium and magnesium free, pH 7.4) at 4 ºC overnight to lyse the erythrocytes. The 96-well plate containing the mixtures in the different wells was then centrifuged the following day at 10 000 x*g* for 5 s before all the PBS/saponin was aspirated. Fresh 1 ml of PBS with no saponin was then added to each well, vortexed briefly, and incubated at 4 ºC for 15-30 min before being centrifuged and the solution aspirated as before. A solution of 150 µL of 10 % Chelex stored at room temperature was then transferred into the tubes and parasite DNA was extracted by incubating the tubes for 10 min in a 95 ºC heat-block, vigorously vortexing each sample every 2 min throughout the incubation. After incubation, the tubes were centrifuged for 5 min at 10 000 x*g* and the solutions were transferred to a clean tube before a final round of centrifugation for 10 min again at 10 000 x*g* and transfer of the final supernatant with DNA. Tips were changed between tubes throughout the extraction process. The extracted genomic DNA was then stored at -20 ºCuntil used.

3.2.4 Ultra-sensitive DNA quantification

Isolates were quantified for *P. falciparum* the using the ultra-sensitive varATS [258] qPCR method as described before in Chapter 2.2.3 using a QuantStudio™12K Flex Real-Time PCR system (Thermo Fisher Scientific) and parasite density determined. Genomic DNA isolated from 3D7 ring-stage parasites spotted from a 20 µL volume of infected blood onto filter paper were used to generate the standard curve for quantification of samples. Based on the available demographic data of study participants, the influence of age and geographic origin of infections on parasite densities of the study cohort was assessed, including statistical (ANOVA test) comparisons of parasite densities between the Limpopo and KZN Provinces. The genotyping threshold was set at a parasite density of ≥ 10 parasites/µL of blood.

3.2.5 Microsatellite genotyping of *P. falciparum* and characterisation of parasite within-host and population level diversity

DBS samples with parasite densities that met the genotyping threshold were genotyped using the same panel of 26 microsatellite markers used in Chapter 2.2.4 [122, 188, 189]. Genotyping data for samples from both the Limpopo and KZN Provinces was processed together with similar microSPAT software settings [188] to avoid variability in allele calling, and the genotyping coverage threshold was maintained at ≥ 60 % as before for inclusion in the downstream population genetics analysis.

As described in Chapter 2.2.5, the within-host diversity of infections was assessed using multiplicity of infection (MOI) and the within-host fixation index (F_{WS}). The MOI was determined in the overall sample set, and assessments of changes in MOI made based on differences in parasite density, age, and by origin of infection.

Like Chapter 2.2.6, genetic diversity of the parasite population was evaluated by heterozygosity (*He*) and the number of unique alleles detected across all loci. Mean genetic diversity was then determined by origin of infection (i.e. provinces of Mozambique from where the surveyed individuals were coming from). This also included samples where the origin of infection was unknown or not provided. The number of haplotypes defined as the multi-locus genotypes across all genotyped loci was also determined.

Differences in parasite within-host and population level diversity were compared between isolates from the Vhembe District, Limpopo Province, and those from the KZN Province using ANOVA pairwise t-test.

LD was also assessed like in Chapter 2.27, however, a total of 999 permutations were completed for this study. To ensure that the pattern of linkage disequilibrium seen is not due to a single pair of loci, LD was calculated over all pairs of loci.

To visualise the parasite genetic diversity within and between infections across the genotyped loci, the haplotype rainbows version v1.0.0 package (https://github.com/nickjhathaway/HaplotypeRainbows) was used in R.

3.2.6 Geospatial population substructure and genetic differentiation

To determine the influence of geographic origin of infections on genetic diversity, the ANOVA pairwise t-test was used to compare MOI, 1-F_{WS} and *He* between the parasite populations stratified by origin of infection. Population substructure between the geographic areas was investigated by measuring Gst, Gst' and Jost's D using the *adegenet* package [265] in R. The Monte Carlo method was used to test the significance of pairwise F_{ST} between source health districts. In this study, 999 permutations were completed. Additionally, DAPC analysis using the *adegenet* package in R software was used to detect signatures of population structure between parasites from the different source areas from where the individuals were coming from [265, 267]. K-means clustering was used to detect the number of inferred genetic clusters in the parasite population, and the best number of clusters chosen was that with the lowest associated Bayesian Information Criterion (BIC). The resultant clusters were then plotted in a scatterplot of the first and second linear discriminants of DAPC. Haplotypes were estimated from all allele data. Analysis of Molecular Variance (AMOVA) was also used to assess parasite population differentiation in the source areas of infection by evaluating where the most variation exists in a hierarchical population structure [291]. To determine the genetic connectivity/relatedness of pairs of infections, a modified identity by state (IBS) metric [122] was used as described in Chapter 2.2.9.

3.2.7 Comparing parasite genetic diversity of infections from the KZN Province to those in Limpopo Province

To determine the differences in parasite genetic diversity in two South African endemic provinces, namely, Vhembe District, Limpopo Province and uMkhanyakude District, KZN Province, parasite within-host and population level diversity were compared using ANOVA pairwise t-test. To determine the extent to which imported genotypes from the KZN populations contribute to those in the Limpopo Province where mostly local transmission occurs, DAPC analysis was repeated including isolates from both the Limpopo Province and the KZN Province.

3.3 Results

3.3.1 Characteristics of KZN study participants

From a total of 885 consenting individuals that were approached to take part in the study at the border market in the KZN Province, 65 (7 %) tested positive for malaria by standard falciparum-specific malaria RDT, which suggests that malaria prevalence was low. Of these 65 individuals, 51 (78 %) were confirmed positive for *P. falciparum* infection by PCR through amplification of the 18s rRNA gene and therefore DBS samples of these individuals were then used for subsequent downstream analysis. All cases were asymptomatic, and the ages of the study participants ranged from 5 to 45 years with the median age being 23 years. Based on travel history data, all 51 cases were classified as imported from neighbouring Mozambique, specifically Gaza (2 %), Inhambane (41 %), Maputo (32 %) districts and some unknown (25 %) because of lack of information of where in Mozambique the individuals were coming from (Figure 3.1).

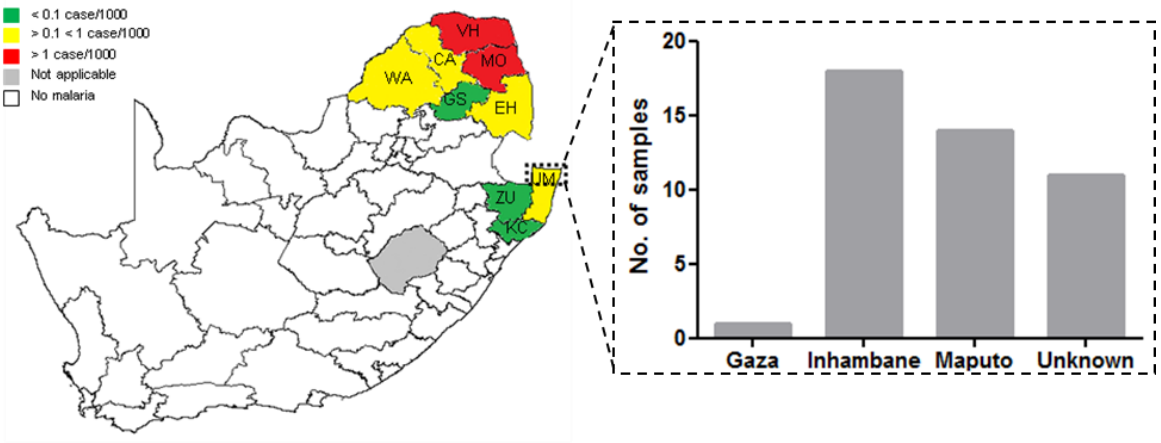


Figure 3.1 Study area and design. Map showing the study site in the KZN Province of South Africa where imported cases from Mozambique were identified. The KZN province is shown alongside other malaria hotspots of local transmission in South Africa indicated by district incidence rates reported in 2018. The following districts are found in the Limpopo Province: VH = Vhembe, MO = Mopani, CA = Capricorn, WA = Waterberg and GS = Greater Sekhukhune; in the Mpumalanga Province: EH = Ehlanzeni; and in the KwaZulu-Natal Province: UM = uMkhanyakude, KC = King Cetshwayo and ZU = Zululand. The districts are colour coded based on incidence rates as indicated in the key. Bar graphs show the proportion of samples coming from the different source districts in Mozambique. The "Unknown" group represents samples with no information of where in Mozambique the individuals were coming from.

3.3.2 High parasite density in asymptomatic imported cases

To determine the level of parasite density in the imported asymptomatic infections, ultra-sensitive quantification was performed. A high proportion of 59 % (30/51) of the

asymptomatic cases had above sub-patent (100 parasites/µL of blood) parasite levels with a mean parasite density of 5568 parasites/µL of blood and the highest parasite density detected in an individual being 183323 parasites/µL of blood (Figure 3.2A). Parasite density was positively correlated with age of study participants (Pearson's r = 0.33 [95 % CI: 0.02 to 0.56]; *P* = 0.038) (Figure 3.3B). This suggests that older asymptomatic migrants may have increased levels of parasite density compared to their younger counterparts. While no significant difference (*P* > 0.05, ANOVA) in parasite density was observed between infections from the Gaza, Inhambane and Maputo Provinces, parasite density of infections from Inhambane was significantly higher (*P* = 0.043 ANOVA, n = 51) than that from unknown areas within Mozambique (Figure 3.3C). Overall, this may suggest that these asymptomatic migrants are coming from areas of similar transmission intensity within Mozambique.

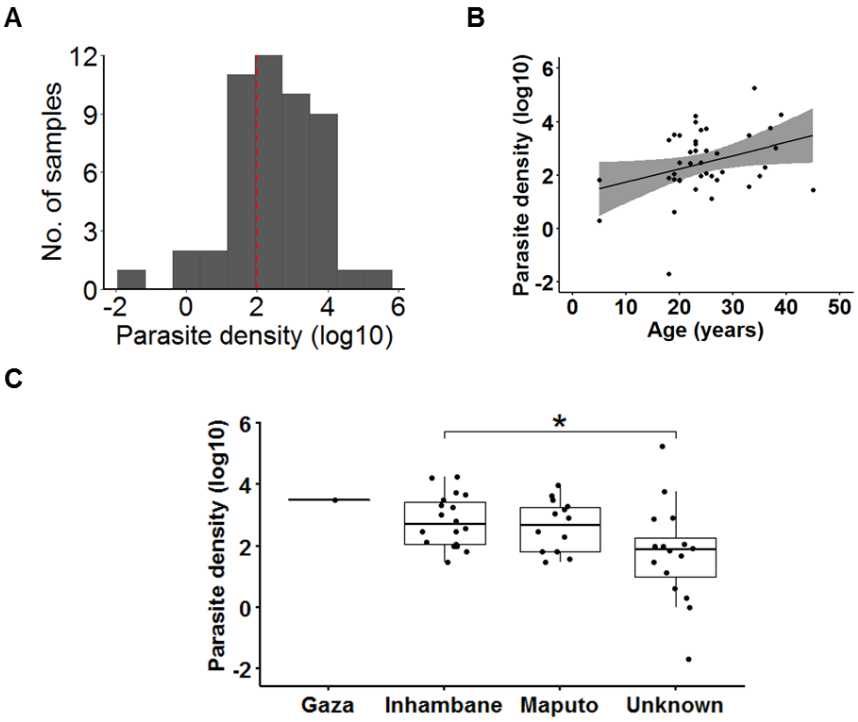


Figure 3.2 Parasite density of asymptomatic, imported *Plasmodium falciparum* infections in the KwaZulu-Natal Province of South Africa. (A) Distribution of parasite density measured as the number of parasites per µL of blood in the quantified samples. The dashed red line indicates sub-patent threshold of 100 parasites/µL of blood. (B) Correlation between parasite density and age of study participants measured by Pearson's correlation at 95 % confidence interval. (C) Comparison of parasite density from the different source health districts within Mozambique.

As expected, mean parasite density of the asymptomatic cases from the KZN Province (mean = 5568 parasites/µL blood, n=51) was significantly lower (*P* = 5.3e-05, ANOVA pairwise t-test) than that of symptomatic cases from the Limpopo Province (mean = 51527 parasites/µL blood, n = 109) (Figure 3.3A). However, interestingly, as evidenced by the

interleaved distribution of parasite densities (Figure 3.3B) these "low" parasite densities of the asymptomatic KZN cases were within detectable ranges found in local symptomatic cases in South Africa. This suggests relatively "high" parasite density in the asymptomatic imported cases in relation to locally acquired infections, which may be indicative of them originating from a high-transmission setting. Therefore, without treatment, these "high" parasite density infections in the asymptomatic imported cases may sustain malaria transmission in low transmission settings in South Africa.

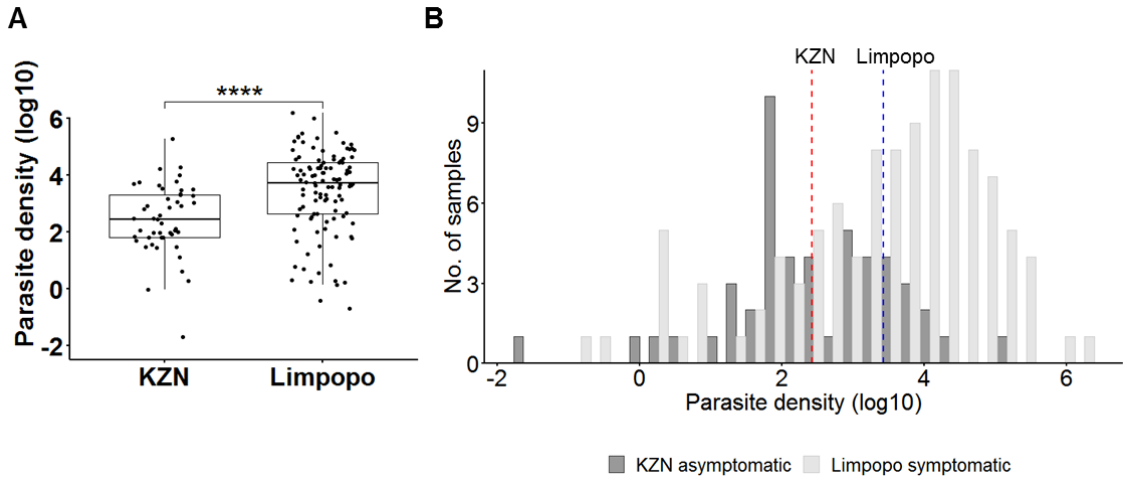


Figure 3.3 Comparison of parasite density between asymptomatic imported *Plasmodium falciparum* infections from the KwaZulu-Natal Province and symptomatic local cases from the Limpopo Province of South Africa. (A) Box plots showing full distribution of the data at different parasite densities associated with the province in South Africa from where the samples were collected. The box plots show summary statistics (i.e., the median and box indicating the 25th and 75th percentile interquartile ranges. (B) Interleaved histograms showing distribution of parasite density in asymptomatic cases from the KZN Province in relation to symptomatic cases from the Limpopo Province. The dashed red line indicates mean parasite density of quantified KZN asymptomatic cases (log10 = 2.43; i.e., 5568 parasites/µL blood). The dashed blue line indicates mean parasite density of quantified Limpopo symptomatic cases (log10 = 3.43; i.e., 51527 parasites/µL blood). The qPCR for both KZN and Limpopo Province samples was run on the same plate, under the same experimental conditions using the varATS method.

Since the median parasite density of quantified KZN samples (n = 51) was 241 parasites/µL of blood, which was above the genotyping threshold of ≥ 10 parasites/µL of blood, therefore all 51 samples were genotyped. Of the 51 samples, 44 (86 %) had a genotyping coverage ≥ 60 % and therefore had good genotyping quality data, allowing them to be included in downstream analysis.

3.3.3 Imported, asymptomatic parasites in KZN are complex and diverse

To determine the level of genetically distinct parasites circulating in the asymptomatic imported infections, parasite within-host and population-level diversity was then assessed.

Alleles were successfully detected at 25 of the 26 microsatellite loci in the sample set and no alleles observed at all at locus PfPK2 which was the most diverse (He = 0.91) locus in the Limpopo sample set as described in Chapter 2.3.5. At locus TA1, only one allele was detected in the asymptomatic imported infections thereby making this locus the least diverse (He = 0, Figure 3.4A). In comparison to the Limpopo sample set in Chapter 2.3.5, locus TA1 had up to 8 unique alleles identified and was highly diverse with a heterozygosity of 0.70. Parasite population level genetic diversity revealed that overall, the parasite isolates were genetically diverse (mean H_e = 0.68; range = 0.47-0.90) (Figure 3.4A). This observation was supported by unique alleles ranging from 1 to 16 (mean number of unique alleles = 7.56) (Figure 3.4B), detected across all loci, which indicated the presence of allelic richness within the population. The allelic richness may be attributed to frequent and random mixing of parasite populations.

A high proportion of 82 % (36/44) of multiple infections (MOI > 1) was observed in the samples, with a mean MOI of 2.41 (range 1-5, (Figure 3.4C) indicating complex infections within individual samples and thus moderate to high within-host diversity in the parasite population. This observation was supported by high within-host diversity as described by F_{WS} (Figure 3.4D). A mean F_{WS} in the sample population of 0.70 (range: 0.22-1.00) was observed, with only 18 % of the samples with a F_{WS} value >0.95, indicating clonal expansion. Fifty-five percent (55 %) of the samples had an $F_{WS} \leq 0.70$, which suggests that these infections exhibited the most genetic diversity [63, 149]. Moreover, F_{WS} was strongly negatively correlated (Pearson's r = -0.83 [95 % CI: -0.71 to -0.90]; p<0.001) with MOI which suggests that both metrics agree in their interpretation of within-host diversity.

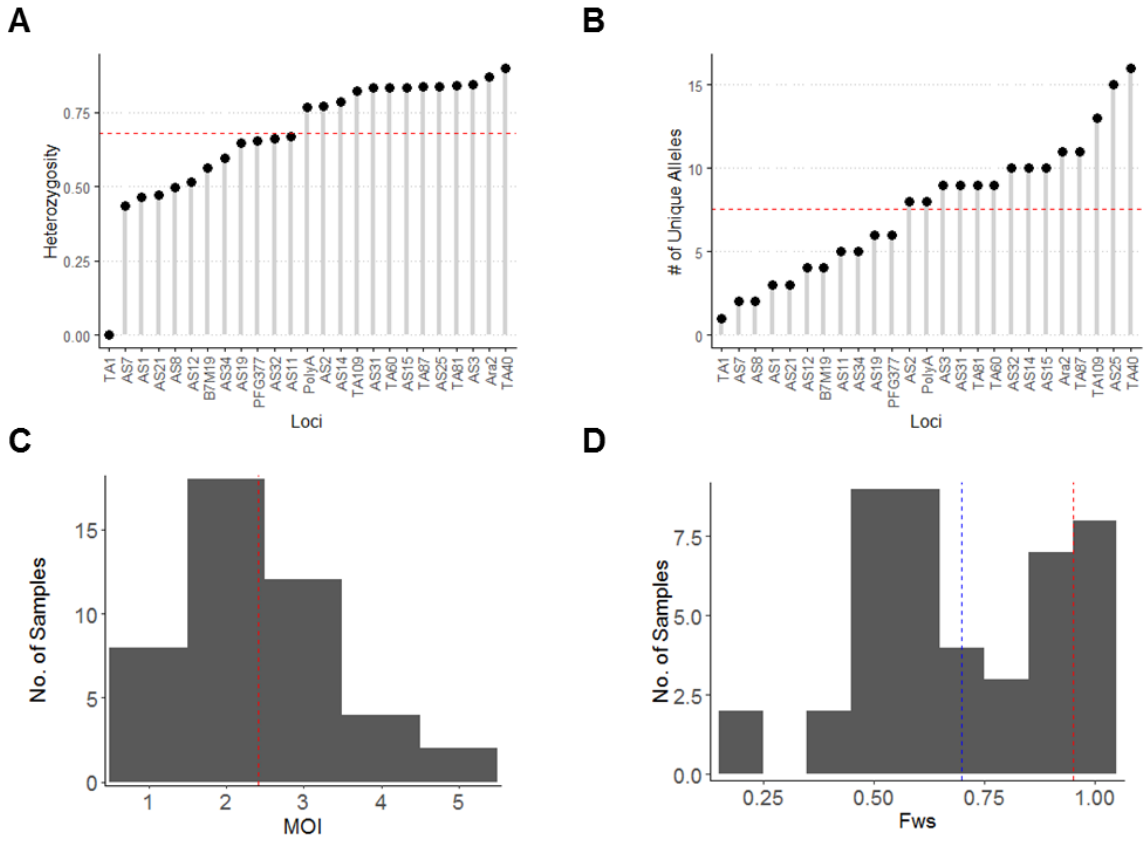


Figure 3.4 Within-host and population diversity of malaria parasites in asymptomatic imported cases in KZN. (A) Population level genetic diversity measured as the distribution of heterozygosity in 26 microsatellites. The dashed red line indicates the mean hetorozygosity = 0.68. Values for H_e range from 0-1, with 0 representing no diversity and 1 representing 100 % of alleles being different. (B) Allelic richness measured as the number of unique across all loci. The dashed red line indicates the mean number of unique alleles = 7.56. (C) Within-host diversity: Multiplicity of infection (MOI) measured as the second highest number of alleles detected at any of the 26 loci. The dashed red line indicates the mean MOI = 2.41 (D) Within-host diversity: Within-host fixation index (F_{WS}) measures the relationship between the genetic diversity of an individual infection relative to that of the parasite population. The red dashed line indicates ≥0.95, a cut-off point used to indicate clonal expansion, and the blue dashed line indicates ≤0.70, a threshold for infections exhibiting high within-host diversity.

A poor correlation was observed between parasite density and MOI (Pearson's r = -0.06 [95 % CI: -0.24 to 0.35]; p=0.6857) (Figure 3.5A), as well as between age and MOI (Pearson's r = 0.05 [95 % CI: -0.27 to 0.37]; p=0.7511) (Figure 3.5B) which suggests that in this sample set, the number of distinct parasite clones circulating within an individual infection (MOI) is influenced by other factors other than parasite density and age.

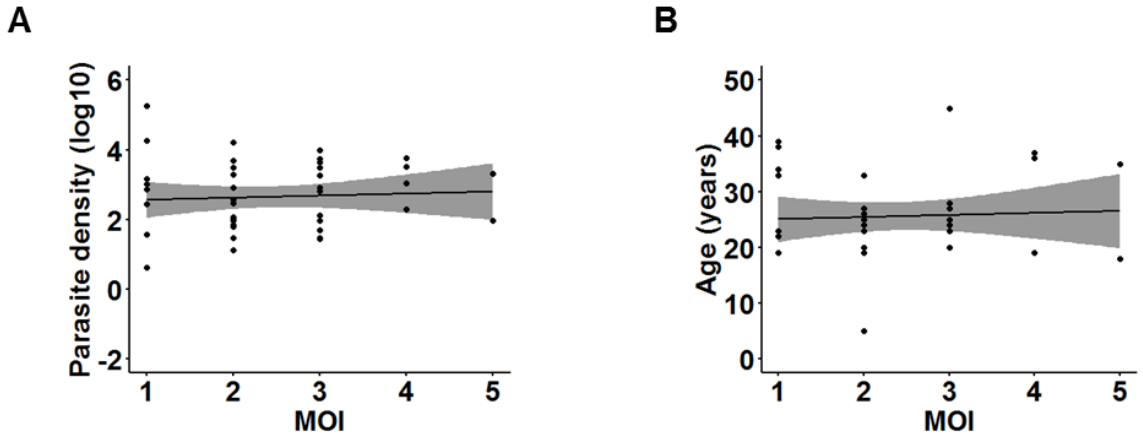


Figure 3.5 Relationship between parasite density, age, and MOI of asymptomatic, imported *Plasmodium falciparum* infections in the KwaZulu-Natal Province of South Africa. (A) Correlation between parasite density and MOI measured by Pearson's correlation at 95 % confidence interval (B) Correlation between age of study participants and MOI measured by Pearson's correlation at 95 % confidence interval. The black lines represent the linear relationships fitted to the MOI values for both parasite density and age of the infected individuals represented by the black dots.

3.3.4 Alleles observed at different loci recombine freely into new genotypes during sexual reproduction

To evaluate whether alleles from different loci were genetically related an analysis of LD was performed. As shown in the heatmap (Figure 3.6A), alleles in the parasite population were related between different pairs of loci, however, this was low as indicated by the low standardised index of association (rd) values (majority blue squares). However, where they were highly related (red squares), it was not due to a single pair of loci which suggests that the overall observed LD would not be biased by a single pair of loci. Since there was only 1 unique allele detected at locus TA1, the index of association on this locus got an undefined value as indicated by the grey squares in heatmap (Figure 3.6A). Overall, low multilocus LD was observed in the sample set with a standardised index of association (rd) of 0.108 (Figure 3.6B). The observed value of 108 fell outside of the re-sampled distribution expected under no linkage (Figure 3.6B) when compared to histograms showing results of 999 permutations which therefore supports sexual reproduction instead of clonal expansion. This suggests that inbreeding between parasite strains is low (therefore high outcrossing) and that fewer related copies of genes are inherited resulting in low relatedness in the parasite population. We, however, find significant (p = 0.001) support for the hypothesis that alleles are linked across loci. As a result of outcrossing between distinct parasite clones, there is a high genotypic diversity observed per locus in the parasite population. Haplotype "rainbows" (Figure 3.6C) are

observed which represent the myriad input of multiple alleles/genotypes per locus per individual in the sample set as opposed to SNP barcodes [292] which only have binary input. Allele frequencies within each sample represent the different haplotypes in each sample. The same colour within a column/locus represents the same major haplotype in different samples, and the different colours represent the minority haplotypes. Colours across columns do not relate to each other which end up creating a repeating "rainbow" across all samples. Where white spaces are seen, no alleles were detected. Although in locus PolyA for example, up to 8 unique alleles were detected in the whole sample set as described in Chapter 3.3.3, this allelic richness was contributed to by only a few individuals (15/44) as compared to for example locus AS2, which had the same number of unique alleles in the sample set but the genetic diversity was contributed to by all (44/44) of the individuals in the sample set. This demonstrates the moderate to high level of parasite genetic diversity within and between these samples and shows how certain individuals may contribute to the observed genetic diversity of the parasite population.

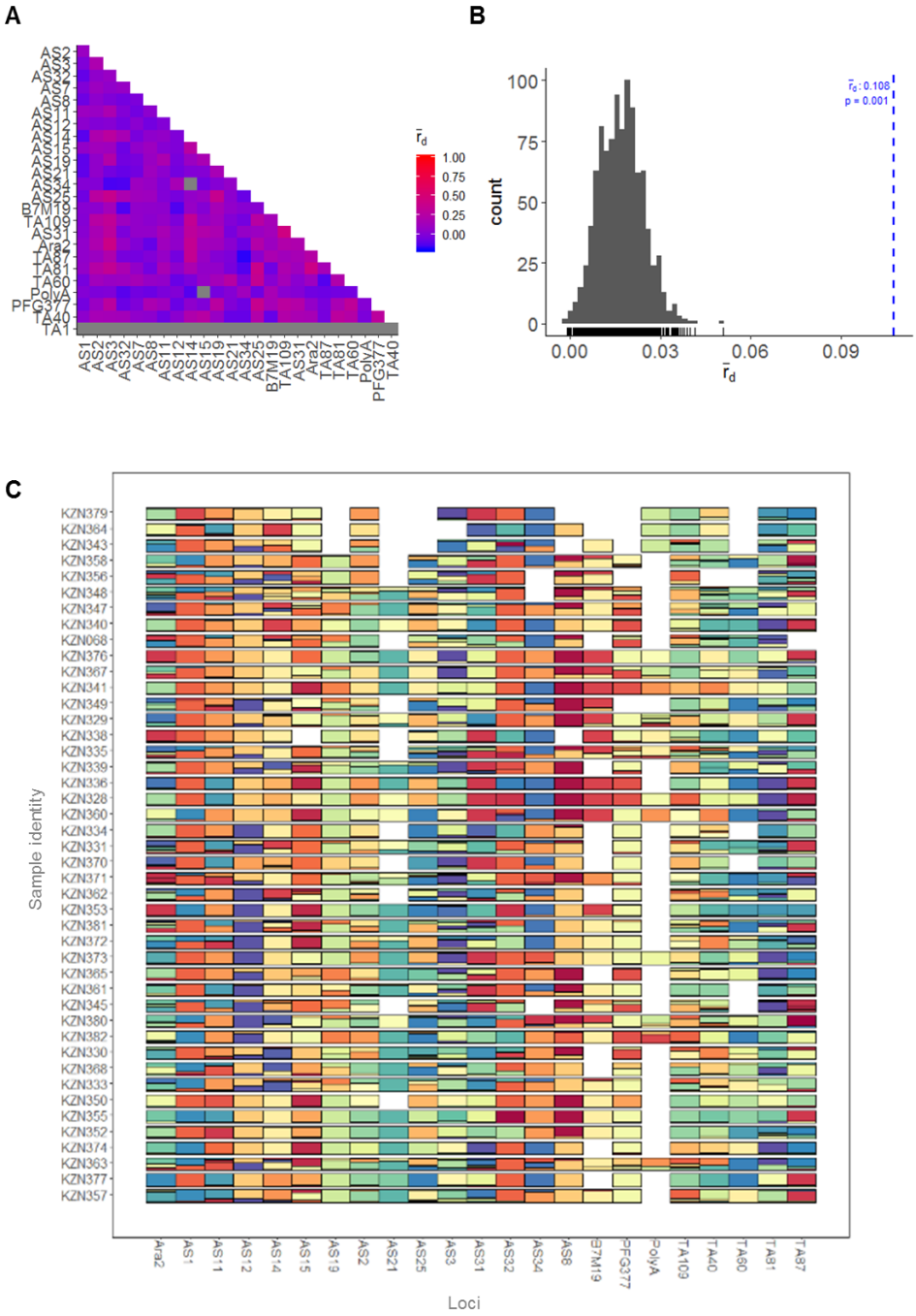


Figure 3.6 Random associations of alleles across loci. (A) Pattern of linkage disequilibrium described by pairwise association/relatedness of alleles across the different loci. LD ranges from 0 (no linkage) to 1 (linkage between alleles) and is colour coded blue and red respectively. (B) Low but significant multilocus linkage disequilibrium. The dashed blue line indicates that the observed $\bar{r}_d$=0.108 falls outside of the re-sampled distribution expected under no linkage and that that alleles are linked across loci with p<0.001. (C) Haplotype "rainbow" showing the level of genetic diversity within and between individual samples. Samples are ordered according to haplotypes that are most like each other and therefore appear next to each other. Samples are shown on the y-axis and loci on the x-axis. Allele frequencies within each sample are considered when adjusting the bars which represent the different haplotypes in each sample. The same colour within a column/loci represents the same major haplotype in different samples, with different colours that end up creating a repeating "rainbow" across all samples. Colours across columns do not relate to each other. Where white spaces are seen, no alleles were detected.

3.3.5 Genetic diversity of imported cases was not influenced by geospatial variation

To establish the differences and/or similarities of the genetically diverse infections over different geospatial scales, parasite within-host and population level was assessed within and between infections from the different source areas within Mozambique. Overall, the levels of within-host diversity were not influenced by where infections were reported since the mean MOI (global ANOVA $P = 0.3$, n = 51, Figure 3.7A) and the level of outbreeding (global ANOVA, $P = 0.68$, n = 51, Figure 3.7B) between the different source health districts did not significantly differ. Heterozygosity was, however, significantly different ($P \leq 0.0001$, n = 51, ANOVA) between Gaza and all other sites which may have been because of differences in sample size (Gaza 1 sample) (Figure 3.7C). A lack of geographic clusters as observed by the DAPC suggests parasite mixing between the source areas (Figure 3.7D). This observation was supported by pairwise genetic relatedness between samples calculated using the identity by state (IBS) metric which identified only a single pair of highly related infections (IBS > 0.6) suggesting limited direct transmission between these Mozambican individuals. Other estimates of parasite population differentiation statistics also showed no differentiation of the parasite populations based on their source areas of infection with Gst = -0.1051, G'st = -0.2404 and Jost's D = -0.08306. This was also supported by AMOVA which showed that there was no significant differentiation ($P = 0.6364$, randtest) between source areas of infection. Additionally, the source areas of infection exhibited extremely small variance components (-0.326) compared to Error (36.7) (i.e., the variation from individuals within populations) which suggests that most of the variance was experienced within individuals than between geographic populations supporting that the parasite population is panmictic. Therefore, the imported infections from Mozambique were mixed and could not be differentiated by their source of origin.

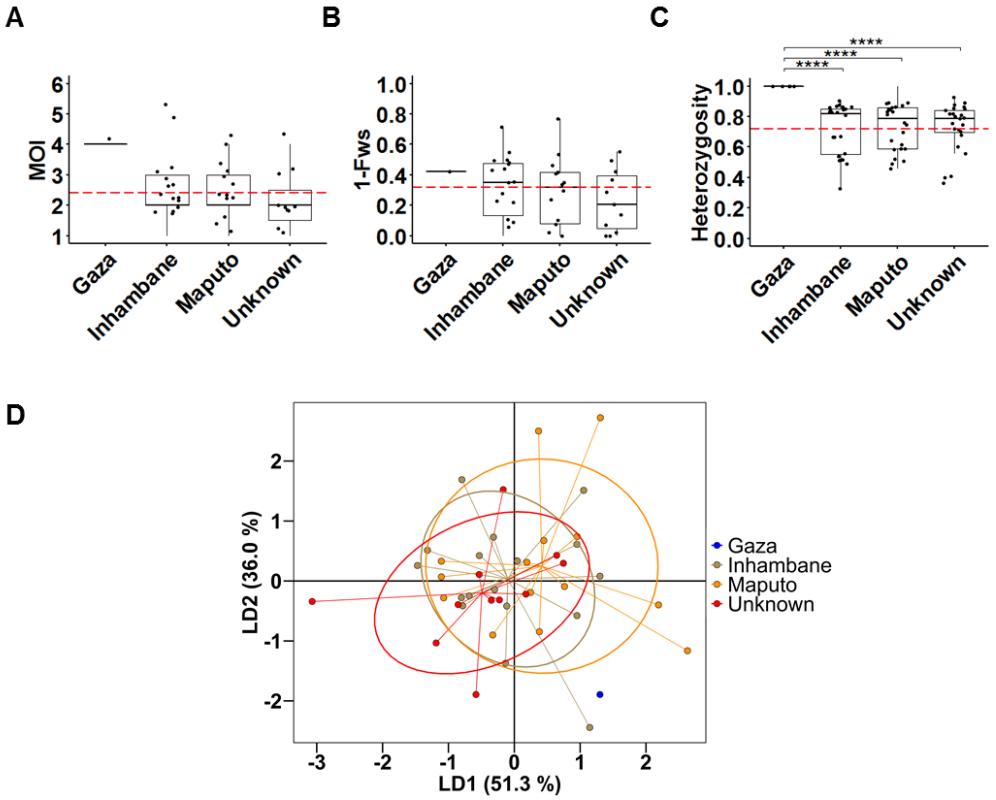


Figure 3.7 Geospatial variation of within-host and population level genetic diversity of *P. falciparum* parasites in the source health districts in Mozambique. (A) Multiplicity of infection (MOI) measured as the second highest number of alleles detected at any of the 26 loci. The dashed red line indicates the mean MOI = 2.41. (B) Outbreeding, ($1\text{-}F_{WS}$) was measured across the different source health districts. The dashed red line indicates the mean $1\text{-}F_{WS}$ = 0.38. (C) Population level genetic diversity measured the level of heterozygosity across the different source health districts. The dashed red line indicates the mean heterozygosity = 0.72. Pairwise *P* values (t-test) are indicated in all plots where, ns: $P>0.05$; *: $P\leq0.05$; **: $P\leq0.01$; ***: $P\leq0.001$; ****: $P\leq0.0001$. (D) *Plasmodium falciparum* population structure of imported cases described by discriminant analysis of principal components (DAPC). Scatterplot based on DAPC discrimination of the four *P. falciparum* populations from Gaza, Inhambane, Maputo and Unknown. All samples were identified in the KZN Province of South Africa and were identified as being imported from Mozambique. Individual multi-locus genotypes appear as dots. Colours and lines represent population membership. Analysis is based on retention of 5 principal components.

3.3.6 Imported infections in the KZN Province are more complex and outbred than locally acquired infections in the Limpopo Province

To compare the differences in parasite genetic diversity of imported infections identified in the KZN Province (low-transmission setting) to the majority of locally acquired infections identified in the Limpopo Province (moderate-transmission setting), pairwise comparisons between the two provinces were performed. In spite of a smaller sample size, the mean MOI (2.41 ± 0.16) and proportion of polyclonal infections (82 %) of samples from the KZN Province (n = 44) was higher but not significantly different (P = 0.1, ANOVA pairwise t-

test) from that of samples from the Vhembe District in the Limpopo Province with a mean MOI of 2.13 ± 0.04 and proportion of polyclonal infections of 66 % (n = 747) (Figures 3.8A and B). The level of outbreeding was significantly higher (*P* = 0.01, ANOVA pairwise t-test) in KZN samples than in Limpopo samples (Figure 3.8C) suggesting possible differences in transmission in the two settings with possible superinfection in Mozambique due to higher vector numbers and transmission intensity compared to the Limpopo Province. The population level diversity as described by heterozygosity was also not significantly different (*P* = 0.64, ANOVA pairwise t-test) between the two provinces (Figure 3.8D). Therefore, the KZN sample set representing imported, asymptomatic cases revealed a higher within host diversity compared to Limpopo samples representing a majority of locally acquired symptomatic cases.

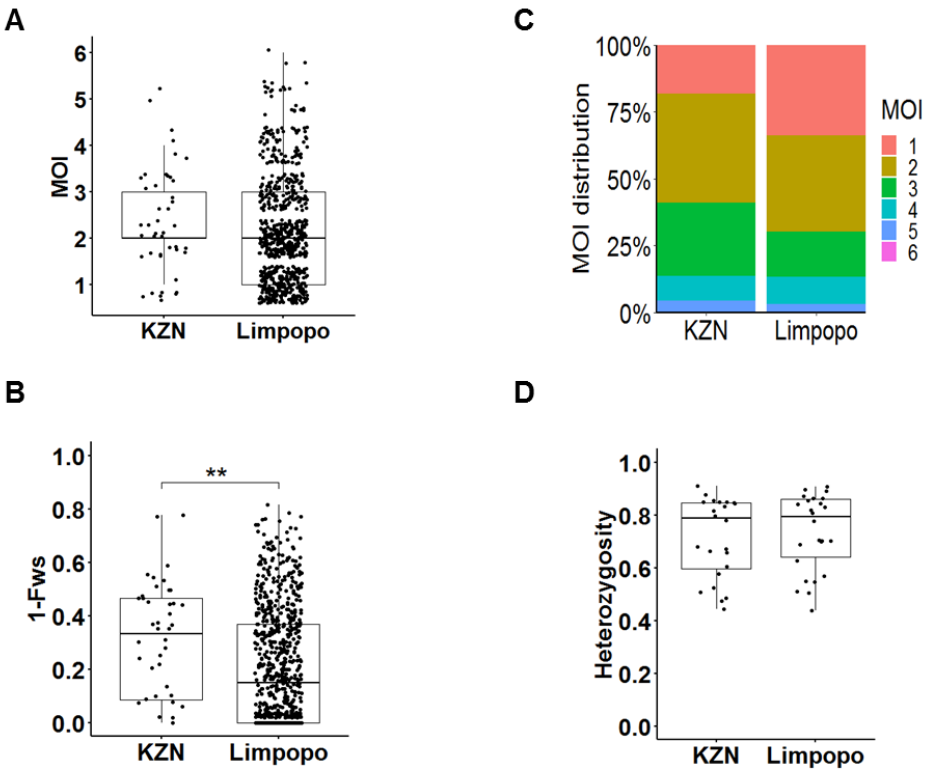


Figure 3.8 Provincial variation of within-host and population level genetic diversity of *P. falciparum* parasites in two of South Africa's malaria endemic provinces. (A) Multiplicity of infection (MOI) measured as the second highest number of alleles detected at any of the 26 loci between the KZN and Limpopo Provinces. (B) Proportion of monoclonal infections (MOI =1) to polyclonal infections relative to each population size between samples collected from the two provinces (n = 44, KZN; n = 747, Limpopo). (C) Within-host diversity index (outbreeding, 1-F_{WS}) measured between the KZN and Limpopo Provinces. (D) Population level genetic diversity measured as the distribution of heterozygosity in 26 microsatellite loci between the KZN and Limpopo Provinces from when the samples were collected. Pairwise *P* values (ANOVA t-test) are indicated in all plots where, ns: *P*>0.05; *: *P*≤0.05; **: *P*≤0.01; ***: *P*≤0.001; ****: *P*≤0.0001.

3.3.7 Diverse imported infections from Mozambique may seed local transmission in South Africa

To investigate how parasite haplotypes of the imported infections would compare to the majority of locally acquired infections, cluster analyses were performed. Eight genetic clusters were inferred from the parasite populations comprising infections from both the Limpopo and the KZN Provinces (Figure 3.9A). These parasites are fragmented based on their level of within-host diversity as described by MOI (Figure 3.9B) and outbreeding (Figure 3.9C). Genetic clusters 1, 2, 3, 4 and 7 which had the majority of polyclonal and the most highly diverse infections were associated with KZN samples originally from Mozambique. Therefore, we conclude that highly diverse imported infections from Mozambique may seed local transmission in South Africa. It is however unknown whether the imported cases from Mozambique fuel secondary transmission in KZN and Limpopo equally.

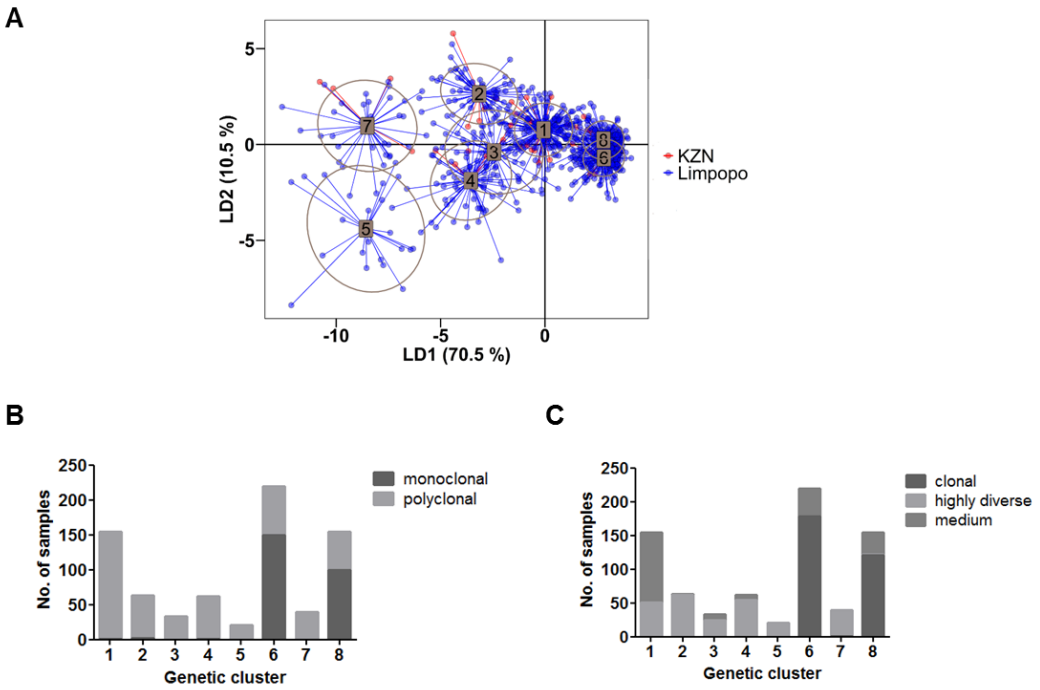


Figure 3.9 Fragmentation of parasites from two endemic provinces in South Africa based on inferred genetic clusters. (A) Scatterplot of the discriminant analysis of principal components (DAPC) based on discrimination of the 8 inferred genetic clusters of *P. falciparum* populations. Samples collected from the Vhembe District in the Limpopo Province and those collected from the KZN Province were included in the analysis. Individual multi-locus genotypes appear as dots. Colours and lines represent population membership. Analysis is based on retention of 50 principal components. Inferred genetic clusters are numbered 1 to 8 and parasite populations are colour coded red and blue based on province from where infections were identified. (B) Proportion of monoclonal (MOI = 1) to polyclonal (MOI > 1) infections in the fragmented genetic clusters. (C) Proportion of clonal (1-Fws ≤ 0.05) to highly diverse (1-Fws ≥ 0.30) infections. Medium represents infections with 1-Fws > 0.05 but < 0.30.

3.4 Discussion

This present study was able to establish that *P. falciparum* parasite density in asymptomatic imported cases of migrants from Mozambique was within the range of that found in local symptomatic cases in South Africa. The parasites circulating in these asymptomatic migrant individuals were also characterised as being complex and genetically diverse. We provide the first piece of genetic evidence suggesting that these genetically diverse imported parasites may be seeding local transmission in elimination settings in South Africa.

It was not surprising that the parasite densities of the migrant asymptomatic individuals detected in this study were above patent levels since it has been shown that immunity to blood stage parasites is greater in higher-transmission settings [293]. Asymptomatic individuals in high-transmission settings therefore have on average higher parasite densities which often extend up to 10 000 parasites/µL [293] compared to asymptomatic individuals in lower transmission settings [294]. This is a great cause for concern because imported cases of migrants carrying patent asexual parasite densities as observed in this current study may contribute to the persistent low level focal transmission in low transmission settings in the KZN Province if they go undetected and untreated. Cross-sectional surveys conducted in other low-transmission areas along border regions between eastern Myanmar and north-western Thailand and in western Cambodia have revealed geometric mean parasite densities of 5158 parasites/mL of blood in *P. falciparum* infections [293] which were much lower than those detected in this study. This may be attributed to different transmission intensities which are higher in sub-Saharan Africa than in Southeast Asia.

It was surprising that the parasite density of the current study cohort increased with an increase in age of study participants. This is because in high-transmission settings, individuals are expected to have acquired immunity due to frequent exposure as they grow [10, 54, 295]. Therefore the correlation between age and parasite density in high-transmission settings is expected to be inversely proportional [54]. Cross-sectional surveys in sub-Saharan Africa have demonstrated an extremely greater infection burden among school going children [296-299]. Our results may therefore suggest that other factors other than acquired immunity contribute to the relationship between age and parasite density in migrant individuals. Since most of the migrants are economic migrants, it may be that these individuals travel more often and for longer distances than their

younger counterparts to and from other higher transmission areas within Mozambique itself or even outside Mozambique where they then acquire additional strains to which they are not immune. This data therefore highlights the complexity in the association between age, parasite load and measures of *P. falciparum* transmission intensity [296-304].

Given that the individuals in this study were identified as originating from Mozambique, the observed moderate to high genetic diversity and complexity of infections of the parasites was not surprising as other studies in the southern African region had detected similar levels in some individuals from Mozambique [188]. In high-transmission countries such as Mozambique, mosquito vector numbers are high and people are usually infected with multiple distinct parasite clones as a result of multiple different mosquito bites resulting in multiple distinct clones circulating in an individual infection and therefore high levels of outcrossing of clones during recombination in the mosquito [188]. Also, when outbreeding is high, mating of unrelated individuals can alter gene frequency in the population as different copies of genes are inherited resulting in an increase in heterozygosity at any given locus. Cross-sectional surveys in pre-elimination settings of South Africa (Vhembe District) also revealed that MOI in individuals with ultra-low density infections ranged from 1 to 3 (mean MOI = 1.8) in local subjects and 1 to 5 (mean MOI = 2.8) in migrant subjects [259] which shows that even with low density infections, *P. falciparum* parasites imported by migrant individuals are more likely to be genetically complex and diverse than those of locally acquired infections. It was interesting to see how the most genetically diverse clusters of infections from the Limpopo Province where mostly local transmission occurs were associated with the highly diverse imported infections which gave the first insights of genetic data supporting that imported infections may seed local transmission in eliminating settings in South Africa. The Gaza Province in Mozambique is a major source of infections to the KZN sink [65]. This is of practical value in designing control interventions as South Africa is funding IRS operations in Gaza and Inhambane Provinces to try and get rid of the source affecting KZN [65].

Findings on the factors that drive *P. falciparum* MOI in high-transmission areas also remain poorly understood and present contrasting views even in terms of correlations with the hosts' age and parasite density. Our study results showed that MOI and host age were poorly correlated which is in support of other studies where a negative correlation is observed [276, 305-308]. This contrasts with other studies which report positive correlations [309-311]. The poor correlation is, however, attributed to acquired anti-

parasite immunity in older individuals living in high-transmission settings [276, 305-308]. The exact age at which efficient immunity is acquired varies based on transmission intensity and could differ between different malaria endemic areas [301]. Individuals may also develop immunity to infection from some, but not all strains to which they are exposed [312] and may not be protected against further infection by the same strains for which immunity was developed. This has implications for malaria vaccine efficacy in high transmission areas and may lend credence to the fact that vaccine rollout will not take place in the Limpopo Province where South Africa experiences the highest transmission. Our observations may also be an indication of within-host competition among genetically distinct co-infecting parasite strains.

As the KZN Province moves towards elimination, the introduction of the transmission-blocking drug primaquine, together with routine active detection at known informal border crossings, and strengthening of cross border initiatives are some of the strategies that have been employed to address imported and introduced malaria [255].

3.5 Conclusion

The findings presented in this Chapter show that *P. falciparum* genetic diversity data in an eliminating setting such as KZN with high connectivity to high-transmission countries like Mozambique strongly points to imported cases contributing to continued transmission likely due to the contribution of highly diverse parasite populations that asymptomatic migrants carry across the border. Therefore, this genetic data has been an informative tool in capturing infections not observed through the national malaria surveillance database and assessing the contribution of these imported asymptomatic cases to local transmission in South Africa. The data generated here could serve as a national reference point for planning diagnosis, treatment, and surveillance of imported asymptomatic cases including mobile migrant populations in South Africa to achieve elimination.

CHAPTER 4

POPULATION STRUCTURE AND GENETIC CONNECTIVITY OF *PLASMODIUM FALCIPARUM* IN PRE-ELIMINATION SETTINGS OF SOUTHERN AFRICA

4.1 Introduction

One of the core objectives of the regional malaria elimination initiative, the Elimination 8 (E8), is preventing cross-border malaria transmission [140], in order to accelerate the region towards malaria elimination by 2030. Therefore, the E8 strategic priorities include understanding regional connectivity, identifying source and sink populations in the region and cross border blocks of high parasite connectivity to inform intervention harmonization and synchronization. The potential impact of imported infections on local transmission is an important consideration [140] for eliminating countries that share porous borders with areas of higher transmission, as importation can play a significant role in sustaining or re-establishing local transmission [140]. Identifying blocks of high parasite connectivity within and across the country's borders and coordinating elimination strategies accordingly would therefore be critically essential for success towards elimination on both a national and regional level. This is important not just at borders but also in country as determining local transmission routes can also contribute to better decision making about appropriate interventions.

The parasite population in Africa had been portrayed as a single contiguous population [145, 261, 269, 271] until recently [36] when signatures of parasite population structure were observed/reported suggesting that the parasite population is in fact fragmented. This fragmentation which clustered *P. falciparum* parasite populations in sub-Saharan Africa into major western, central, and eastern regional subgroups as well as a highly divergent Ethiopian subpopulation has been attributed to the parasite's ancestry associated to the respective regional blocks which corresponded with both the parasite's origin and with historical human population movement; as well as the use of interventions that may also drive the selection of for example drug resistant parasite strains [36]. The southern African parasite population was, however, inadequately represented in sub-Saharan African parasite population studies with only Zimbabwe and Malawi [36, 145] from the southern African region included.

Limited *P. falciparum* population genetics studies that have been conducted separately in low-transmission settings of southern Africa include Namibia where moderate to high parasite genetic diversity, fine-scale parasite population structure and cross-border parasite genetic connectivity with neighbouring high-transmission countries (Zambia and Angola) [122] was identified. Similarly, Zambia identified cross-border parasite genetic connectivity with neighbouring higher transmission country of the Democratic Republic of Congo [250], and Eswatini with neighbouring Mozambique linked through travel history data [188]. Although these studies provided baseline evidence to try and understand the level of parasite genetic connectedness, gene flow and population structure in the southern Africa region, it is important to cover a wider regional geographic sampling scale in a single study so as to directly compare how parasite populations from the different countries relate to each other and understand southern African regional transmission dynamics collectively as collated genetic data is used from as many countries as possible.

Some of the studies showed that the level of parasite genetic diversity does not necessarily reflect the transmission intensity in the country possibly due to importation of malaria from higher-transmission countries [188]. This contrasted with high-transmission countries studied extensively in sub-Saharan Africa where the high transmission intensity has been reflected by high levels of parasite genetic diversity [64, 145, 149, 269, 271]. If measures of parasite genetic diversity are to have utility in assessing the level of transmission intensity in low-transmission settings in southern Africa, then parasites in this region should therefore be assessed together to identify the best parasite genetic diversity metrics to be used as indicators of transmission intensity in the region. Even more fundamental is to define the extent to which these metrics can be used at all, in which settings, and with which caveats, as they may not be informative in all settings, at least without proper context. This study therefore presents an opportunity to identify the parasite genetic diversity metrics that would be the best indicators of transmission intensity within the southern African region.

Thus, as a pilot study to a future regional study that will include all countries within the E8 region, this study set out to do a meta-analysis by comparing parasite populations recently genotyped from South Africa [255, 289] as described in Chapters 2 and 3, to those from neighbouring countries (Namibia, Eswatini and Mozambique) [122, 188] where similar technology was used, to understand how the South African parasite population compares to that of other parasite populations in the southern African region as the country and the region work towards malaria elimination. This Chapter aimed to evaluate the population

structure, genetic connectivity, and gene flow patterns between different *P. falciparum* populations from the selected countries in the southern African region. Additionally, parasite genetic metrics that would be the most useful to assess transmission intensity based on the collective genetic diversity represented by parasites in the southern African region were evaluated.

4.2 Materials and methods

4.2.1 Ethical approval

Ethical approval for the study was obtained from the University of Pretoria, Faculty of Health Sciences Research Ethics Committee (Ethics Reference No. 406-2014) and the Limpopo Department of Health (Ref: LP_201906_011). The National Institute for Communicable Disease holds ethical approval for analysis of the KwaZulu-Natal (KZN) samples from the KwaZulu-Natal Provincial Department of Health, the Health Ethics Review Committee of the University of Witwatersrand (M170869), and Advarra Research Compliance Solutions (Maryland, USA). The other collated genetic data was obtained based on ethical approval and protocols of the local National/Institutional ethical review committees of Namibia [122], Eswatini [188] and Mozambique.

4.2.2 Datasets used

A meta-analysis was conducted based on the collated data from publications of parasite population genetics studies conducted in Namibia [122], Eswatini [188], South Africa [289] and data generated from Mozambique provided by Professor Bryan Greenhouse after personal communication with him. The genotyping data was comparable in terms of the same microsatellite genotyping technology and equipment used to generate it and similar microSPAT software settings used for allele calling for all samples collectively. The microsatellite genotyping protocol is as described in Chapter 2.2.4.

4.2.3 *P. falciparum* population level diversity in southern Africa

On a population level, the heterozygosity, number of unique alleles per locus (allelic richness) and multilocus linkage disequilibrium (using both monoclonal and polyclonal data) for each population of isolates defined by geographical location of sample collection sites (countries) was calculated in R using the poppr package as described in Chapter

2.2.6. ANOVA pairwise t-tests were used to compare differences in population level (*He*) diversity between the 4 countries/populations and assess whether parasite population level genetic diversity reflects the transmission intensity observed in the four countries.

Since allelic richness is biased by sample size, to assess the distribution of alleles across the populations and the number of alleles private to each population with a standardised sample size, the ADZE software [313] was used. To compensate for differences in sample sizes, a rarefaction approach that considers the maximum equal-sized sub-samples from each population was considered. Principal component analysis (PCA) was also performed as a quality control measure using the *factoextra* package in R to ensure that the microsatellite markers conform to expected distribution. Genetic bottlenecks of the parasite populations and LD were determined as described in Chapter 2.2.6 and 2.2.7 respectively and compared between the different parasite populations per country.

4.2.4 *P. falciparum* within-host diversity in southern Africa

For the within-host diversity, the MOI and Fws index were calculated as previously described in Chapter 2.2.5. ANOVA pairwise t-tests were used to compare differences in within-host (MOI and 1-F_{WS}) diversity between parasite populations from the four countries and assess whether within-host diversity reflects the transmission intensity observed in the four countries.

4.2.5 Population structure and differentiation within southern Africa

To determine signatures of population structure between the four countries, Discriminant Analysis of Principle Components (DAPC) was performed using the *adegenet* package in R software [265, 267] with countries used as *priori* groups. A scatterplot of the first and second linear discriminants of DAPC was then plotted. To prevent over fitting of clusters, the optimal number of principal components (PC) to be retained was confirmed by cross validation of the DAPC. Cross-validation provides an objective optimisation procedure for identifying the 'goldilocks point' in the trade-off between retaining too few and too many PCs in the model. Data was divided into a training set (90 % of data), and a validation set (10 % of data), and members of each of the identified clusters were stratified by random sampling to ensure that at least one member of each group or population in the original data is represented in both training and validation sets. DAPC was then performed on the training set with variable numbers of PCs retained. The extent to which the analysis was

able to accurately predict group memberships of individuals in the validation set was used to identify the optimal number of PCs to be retained. Sampling and DAPC procedures were repeated 30 times at each level of PC retention, and the optimal number of PCs retained was associated with the lowest root mean square error. Population differentiation between the geographic areas was also determined by measuring pairwise measuring Wright's F-statistics (F_{ST}), using the *adegenet* package [265] in R. Hendrick's G_{ST} and Jost's D, were calculated using the mmod package [266] in R. The Monte Carlo method was used to test the significance of pairwise F_{ST} between the countries by completing 999 permutations.

An isolation by distance approach which correlates genetic distance to geographic distance was used to test the significance of population structure. The Monte Carlo method was used to test the significance and was based on 999 replicates.

4.2.6 Parasite genetic connectivity and gene flow within southern Africa

To examine genetic connectivity of parasite genotypes across countries, the number and proportional distribution of multi-locus genotypes (MLG) genotypes as well as genotypes per locus shared across populations was assessed. The extent and direction of parasite gene flow within and between countries was then determined to infer parasite migration patterns among populations and identify cross border blocks of high parasite connectivity to identify the major source and sink areas in the region. Estimates of historical gene flow patterns were made using dominant allele data and divMigrate online software [314] (https://popgen.shinyapps.io/divMigrate-online/). Asymmetric bidirectional gene flow was assumed. Relative migration and gene flow was determined based on Wright's equation $FST=1/4Nem+1$, where *Ne* = population size and *m* = gene flow [315]. The migration patterns between different parasite populations were tested for different levels of gene flow calculated using *Jost's D* method [316].

To estimate more recent migration or gene flow patterns across countries the BayesAss v3.0.4 program [317] was used, again with dominant allele data. Bayesian inference with Markov chain Monte Carlo (MCMC) simulations was used to estimate the fraction of immigrants per population. Mixing parameters for migration rates, allele frequencies and inbreeding coefficients were optimised to ensure that the acceptance rates for each parameter was between the recommended target ranges of 20 % to 60 % [317]. To

calculate the average of gene flow estimates, MCMC simulations were performed using 10^7 iterations, with a burnin of 10^6 and a sampling interval of 100.

Pairwise genetic relatedness of highly related parasites (all allele data) was then determined using IBS as described before in Chapter 2.2.9. Dr Sofonias Tessema ran the R script for generation of the IBS results, however, the resultant data was analysed by the PhD candidate. The geographic distance between parasites was determined based on the longitude and latitude coordinates of the clinics in the different districts of Mozambique (since residential longitude and latitude coordinates of each patient were not available) from where the cases of infection were identified, and the national coordinates used for Eswatini.

4.3 Results

4.3.1 Meta-analysis data population

To ensure that a good representation in sample size across the 4 selected countries would be achieved, all the data from the selected studies was used in this meta-analysis. A total of 5314 samples of curated genotype data was collected from publications from Namibia (n = 2585) [122], Eswatini (n = 835) [188] and South Africa (n = 747) [289]. Data from Mozambique (n = 1147) is still not yet published, however, this was obtained from Professor Bryan Greenhouse after personal communication with him. Additionally, 46 of the samples from the Mozambique data set was obtained from published data [255] also presented in Chapter 3. The study sites representing the countries from where the samples were collected are shown in Figure 4.1A. A genotyping coverage threshold of ≥ 60 % where alleles in each sample had to be detected at a minimum of at least 15 loci was maintained for the downstream population genetics analysis (Figure 4.1B). Therefore, allelic data from as many good quality genotyped samples as could be accessed was available for each country.

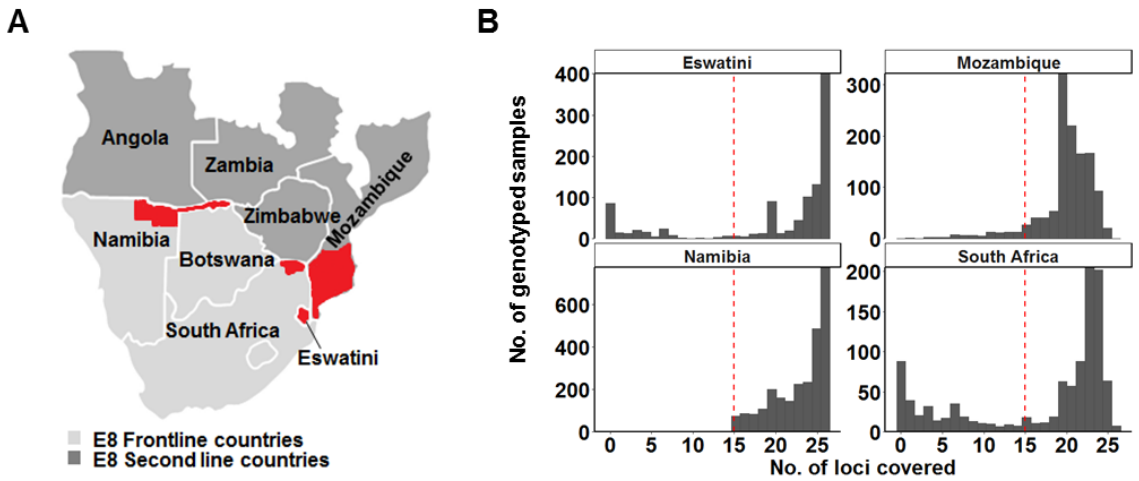


Figure 4.1 Study area and design. (A) Map of the study site showing the countries from where genotyping data was collected. The specific areas where samples were collected are coloured in red: the north-eastern border region of Namibia (the Kavango and Zambezi regions), the north-eastern border region of South Africa (the Vhembe District in the Limpopo Province), the whole of Eswatini and parts of southern Mozambique. Study sites represent countries in the southern Africa Elimination-8 region. (B) Only samples that met the 60 % genotyping coverage threshold were analysed. The red dashed line indicates the threshold which represents alleles being detected on at least 15 of the 26 loci.

4.3.2 Allelic patterns across southern African parasite populations

To determine patterns of allelic richness in the 4 countries in comparison to each other, frequency distribution of the number of unique alleles across all loci and pairwise ANOVA analyses were performed. Allelic distribution across loci generally showed a similar trend in the number of unique alleles identified at each locus per country (Figure 4.2A). Exceptions were locus TA1 which showed much lower number of unique alleles in South Africa compared to the other 3 countries; and loci PfPK2, AS32, AS34 and B7M19 which had much higher number of unique alleles identified in Namibia compared to the other countries. The overall number of unique alleles was significantly higher ($P \leq 0.01$, pairwise t-test) in Namibia (mean A = 17.3 ± 1.46) compared to South Africa (mean A = 12.2 ± 1.22) and Eswatini (mean A = 13.3 ± 1.27) ($P \leq 0.05$, pairwise t-test), and was not significantly ($P > 0.05$, ANOVA pairwise t-test) different between the other country pairs including Mozambique which had a mean A = 14.0 ± 1.38 (Figure 4.2B). This was based on both uncorrected and adjusted/standardised sample sizes. Allelic richness patterns therefore did not reflect transmission intensity in the region. The greatest allelic richness was observed in Namibia which was unexpected since Mozambique is the highest transmission country of the 4 study countries. However, this observation may be explained by the fact that the southern region of Mozambique where most study samples may have been collected from has a much lower transmission intensity for example in

Maputo city than other areas in Mozambique where transmission intensity is much higher such as Cabo Delgado Province [318]. Transmission intensity is, however, still much higher in Mozambique than Namibia, and the observed differences could have been due to other aspects such as sample size. Another alternative is that maybe allelic diversity is not only affected by transmission intensity. Population movement in the study area as well as settlement factors such as how clustered homesteads and villages are plays a role. The closer they are, with lower population movement, can lead to circulation of a few clones. Another consideration is that samples from Namibia show greater diversity due to importation from Angola rather than because of transmission intensity.

PCA showed that all loci were positively correlated as they pointed to the same side of the plot (Figure 4.2C), which suggests that the microsatellite markers conform to expected distribution. However, loci TA1, PfPK2, B7M19, PolyA, AS3 and TA40 were the most variable and contributed between 5 % to over 15 % variance (Figure 4.2C). Therefore, this confirms the observed distribution per loci shown in Figure 4.2A and suggests that the level of unique allele richness at some loci is influenced by the country of origin which may be a reflection of different transmission dynamics between the countries or technical differences in sampling.

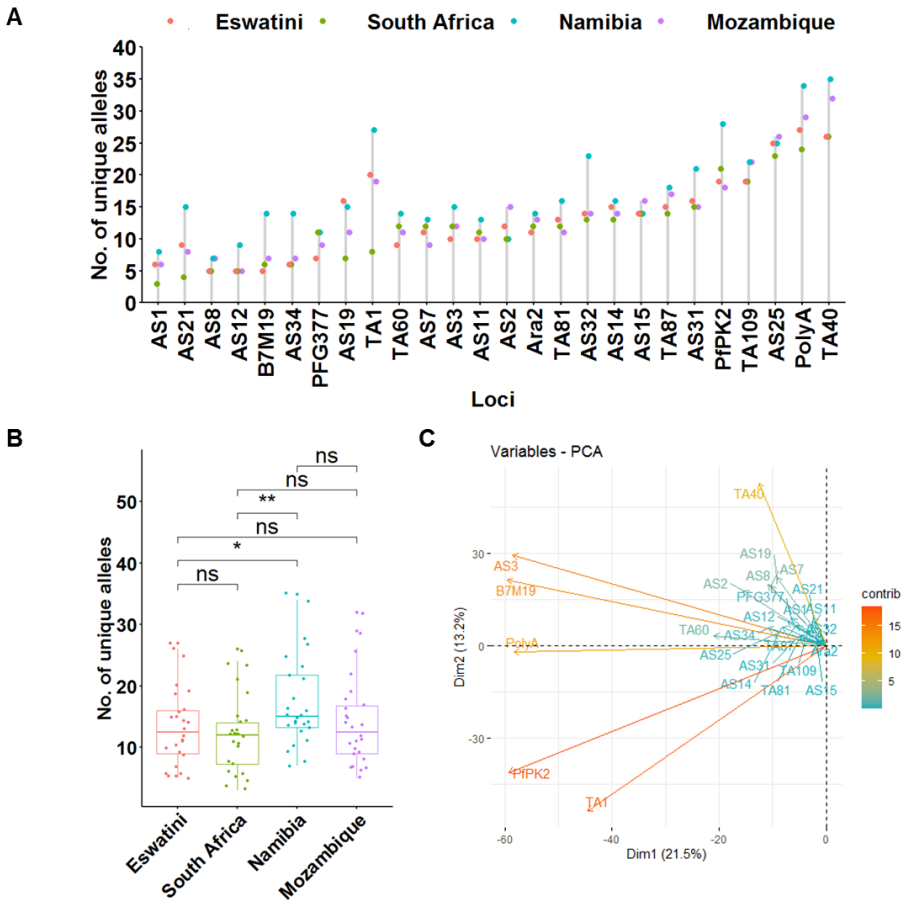


Figure 4.2 Allelic patterns across southern African parasite populations. (A) Distribution of the number of unique alleles per locus across countries. (B) Comparison of the allelic richness between countries. The corresponding box plots show summary statistics with jitters indicating distribution of individual data points. Pairwise *P* values (t-test) of the number of unique alleles compared between countries are indicated where, not significant (ns): *P*>0.05; *: *P*≤0.05; **: *P*≤0.01. (C) Principal Component Analysis (PCA) showing variability of loci with data from all countries combined. Positive correlated variables (loci) point to the same side of the plot and negative correlated variables point to opposite sides of the graph. The colour scale represents contribution of variables with the highest variability represented by a deep orange and least variability represented in turquoise.

4.3.3 Parasite population level diversity is high and stable and does not reflect transmission intensity in southern African

To determine which metrics of parasite population level diversity were good indicators to reflect transmission intensity in the region, heterozygosity and the numbers of unique haplotypes were assessed within and between countries in the region. Evidence of parasite population bottlenecks was also assessed. The level of heterozygosity was high (mean He = 0.75) but did not differ significantly between all countries (*P* > 0.05, ANOVA pairwise t-test) (Table 4.1). Therefore, this may suggest that the level of heterozygosity

does not reflect transmission intensity and may not necessarily be a good indicator of transmission intensity in the southern African regional parasite populations. Eswatini had the highest proportion of identical haplotypes (2.2 %, 18/835) in its population, followed by South Africa (0.9 %, 7/747), Mozambique (0.7 %, 8/1147) and Namibia (0.3 %, 9/2585) (Table 4.1). These identical haplotypes were calculated based on multilocus genotypes from all allele data. The low proportion of identical haplotypes in all countries confirms the high level of genetic diversity of the parasite population mostly because of sexual recombination of distinct parasite clones. Although the proportions of identical haplotypes are relatively low, the differences observed between the four countries may suggest a higher likelihood of some parasites undergoing clonal expansion or inbreeding in Eswatini as the lowest transmission country. Additionally, the sampling was quite different between countries – in Eswatini for example, there was denser sampling which included reactive case detection around cases. The more densely you sample, the more likely you are going to find closely related infections. The lowest number of identical haplotypes identified in Namibia may have therefore been influenced by the wider geographical scale at which sampling was done compared to other countries. Other factors such as the differences in sampling parameters i.e. the site where samples were collected, period of sample collection, sample sizes and area of collection are also important when considering the comparisons in this data. Sampling done over a period of peak transmission may yield different results from sampling from an entire malaria season. Additionally, data collected from the different countries was sub-national data with the exception of Eswatini where national data was collected. Therefore this data represented a defined geographical area and was not representative of national data. There is great heterogeneity in malaria transmission in these countries.

Significant linkage disequilibrium (LD) ($P \leq 0.001$, Monte Carlo test, 1000 permutations) was detected in all four countries which suggests some level of local transmission (Table 4.1). Multilocus LD was also detected in the overall southern African population (LD = 0.18; $P \leq 0.001$, Monte Carlo test) which may suggest some level of population structure. Surprisingly, the highest LD was observed in Namibia (LD = 0.212) although it had the least number of identical haplotypes. This inferred relatedness may indicate genetically related clusters due to rapid reduction in transmission and effective parasite population size influenced by different processes.

Table 4.1 Genetic diversity of *Plasmodium falciparum* parasite populations from pre-elimination settings in southern African.

Population	n	h	A ± SE	He ± SE	MOI ± SE	1-Fws ± SE	LD
Eswatini	835	817	13.3±1.3	0.75±0.03	2.32±0.04	0.22 ± 0.01	0.17***
South Africa	747	740	12.2±1.2	0.74±0.03	2.13±0.04	0.21 ± 0.01	0.14***
Namibia	2585	2576	17.3±1.45	0.75±0.03	2.74±0.03	0.22 ± 0.01	0.212***
Mozambique	1147	1139	14.0±1.38	0.74±0.03	2.14±0.03	0.24 ± 0.01	0.119***
TOTAL	**5314**	5272	**14.2 ± 1.3**	**0.75±0.03**	**2.33±0.03**	**0.22 ± 0.01**	**0.18*** **

n = number of isolates genotyped; h = number of haplotypes (or multilocus genotypes); A = mean number of alleles per locus; He = heterozygosity; MOI = multiplicity of infection; 1-Fws = outbreeding, LD = linkage diseqilibrium. ***: $P \leq 0.001$ (Monte Carlo test, 1000 permutations).

Allele frequency distribution showed L-shaped mode shift graphs (Figure 4.3) for parasites from all the four countries which suggests that there was no evidence for recent parasite genetic bottlenecks in any of the parasite populations as expected under the assumption of mutation drift equilibrium. This lack of bottlenecks can be attributed to the importation of malaria parasites especially from high-transmission countries. Most alleles were found in the rarest class with allele frequencies ≤ 0.10 as is expected from neutral evolution. This therefore suggests that the higher proportion of shared alleles observed in the low-transmission countries (Eswatini, South Africa and Namibia) as observed by LD may not have been due to recent intervention but rather had been sustained from previous years before sampling was done. Estimates of effective population size (N_e) were highest in Namibia (3148), followed by Mozambique (1206), Eswatini (1027) and South Africa (847) which correlated to the different sample sizes. This indicates high parasite population diversity. Overall, this data shows that the genetic diversity of the parasite population in these countries is high and stable. Therefore, metrics of parasite population genetics may not be the best at reflecting transmission intensity in the region. Limitations that arise from sampling parameters discussed earlier also apply here.

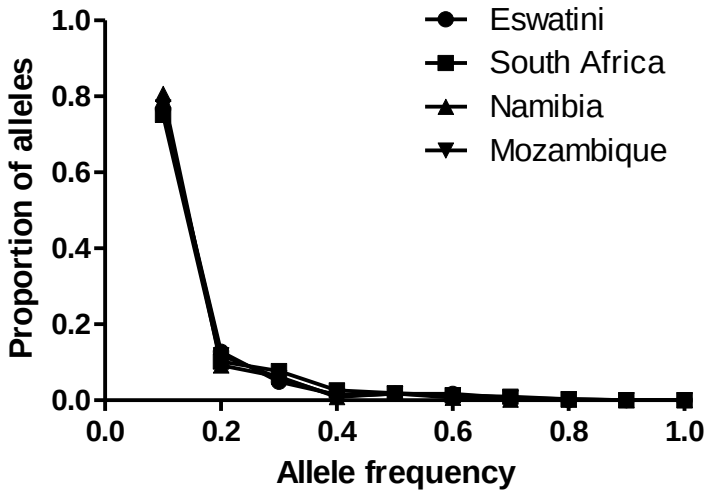


Figure 4.3 Allele frequency distribution for *Plasmodium falciparum* parasite populations in southern African countries. Genetic population bottleneck analysis was performed using the BOTTLENECK software. The resultant overlaid mode-shift graphs for each country showing an L-shaped distribution that indicates lack of recent genetic bottlenecks are represented.

4.3.4 Parasite within-host diversity does not reflect transmission intensity in southern Africa

To determine whether metrics of within-host diversity were better indicators to reflect transmission intensity in the region, MOI distribution and level of outbreeding were evaluated within and between countries. The overall southern Africa population showed that the MOI in the region is moderate with a mean MOI of (2.33) (Figure 4.4A). Evaluating the relationship between transmission intensity and MOI between the four countries revealed that MOI is significantly higher in Namibia (MOI maximum = 10) than all countries, confirming that infections were the most complex in this country (Figure 4.4B). Interestingly, this MOI was higher than that of Mozambique (MOI maximum = 7) which is a higher transmission country. At an MOI of 1, the data shows that the proportion of monoclonal infections relative to the total number of samples in each country is instead highest in Mozambique, as shown by the spread of MOI in each country (Figure 4.4B). The proportion of polyclonal infections with MOI > 1 were therefore: 77 % for Namibia, 70 % for Eswatini, 66 % for South Africa and 64 % for Mozambique suggesting that the complexity of infections in southern Africa does not reflect transmission intensity. South Africa and Eswatini are expected to have similar MOI distribution patterns to Mozambique as that is where they import malaria from due to close geographic proximity. Namibia on the other hand imports cases from Angola which has even higher transmission than

southern Mozambique. This highlights the importance to contextualize sampling patterns for the studies while interpreting results.

Although the mean MOI between South Africa and Mozambique was not significantly different ($P > 0.05$, pairwise t-test), the level of outbreeding was significantly higher ($P \leq 0.01$, pairwise t-test) in Mozambique (Figure 4.4C) compared to South Africa, which may suggest different biological scenarios of the parasite in the different countries (i.e. co-transmission which may result in more related coinfecting clones vs superinfection where the different clones in an individual infection may not be related). When taken up by a mosquito during a blood meal, these unrelated clones are more likely to out-cross thus generating more diverse parasites in the population. South Africa had the highest proportion (40 %, 297/747) of clonal parasites as described by 1-Fws (1-Fws < 0.05) (Figure 4.4D) which suggests higher levels of inbreeding in this country compared to the other three. This was also reflected in the smallest proportion of highly diverse (1-Fws > 0.30) parasites in South Africa (28 %, 207/747) compared to Eswatini (34 %, 281/835), Namibia (32 %, 818/2585) and Mozambique (37 %, 425/1147) (Figure 4.4D). Given that the Fws metric is more granular than MOI, the differences in this metric observed between South Africa versus the other 3 countries which appear more like each other may be due to factors such as South Africa having been the first of the four countries to have introduced the use an artemisinin-based combination therapy as first-line treatment for uncomplicated falciparum malaria. This may have had an influence in getting rid of more diverse clones in the parasite population possibly leading to fewer distinct clones circulating in the parasite population, leading to lower outbreeding of fewer distinct clones. Therefore, both the level of outbreeding and MOI are, however, also not good indicators of transmission intensity in the region. This highlights other underlying factors happening to the parasite in the region.

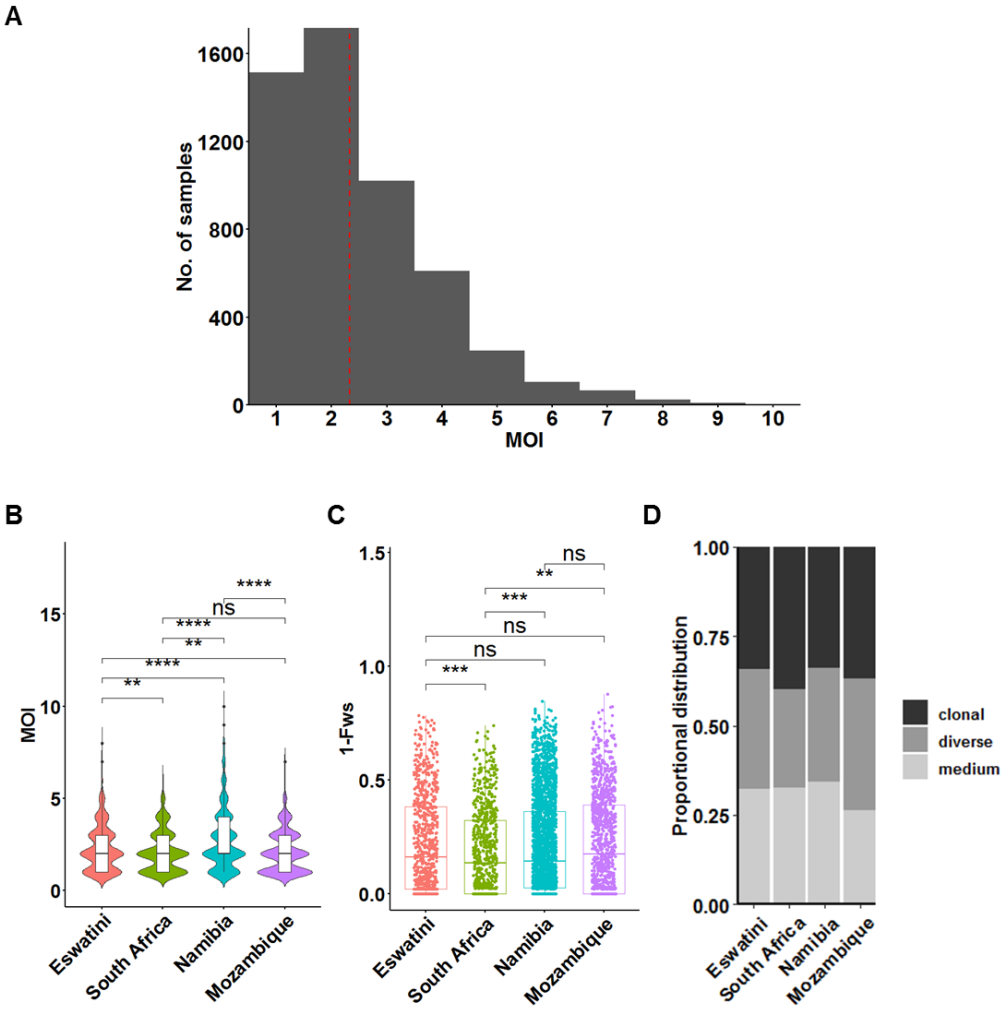


Figure 4.4 Within-host diversity described in relation to transmission intensity across the different endemic countries. (A) MOI was measured as the second highest number of alleles detected at any of the 26 loci. Distribution in the southern African population is shown. The dashed red line indicates the mean MOI = 2.33. (B) The violin plots showing full distribution of the data at different MOI values are colour coded to correspond to the associated country. The corresponding box plots show summary statistics (i.e., the median and box indicating the 25th and 75th percentile interquartile ranges; with dots representing potential outliers). (C) Within-host diversity index (1-Fws) describing outbreeding. The corresponding box plots show summary statistics with jitters indicating distribution of individual data points. Pairwise *P* values (t-test) of MOI and 1-Fws compared between countries are indicated where, not significant (ns): P>0.05; *: P≤0.05; **: P≤0.01; ***: P≤0.001; ****: P≤0.0001. (D) Proportional distribution of clonal (1-Fws ≤ 0.05), moderately diverse (medium; 0.30 ≥ 1-Fws ≥ 0.05) and highly diverse (1-Fws ≥ 0.30) infections as described by the within-host diversity index (1-Fws).

4.3.5 Significant *P. falciparum* population structure in southern Africa

To establish whether parasite populations from the different countries in the southern African region were genetically similar based on their geographic origins, population structure analysis and genetic differentiation of parasites was performed. Genetic

differentiation between pairs of parasite populations was very low as described by Nei's G_{ST}, and ranged from -0.00008 to 0.002 (Figure 4.5A). Jost's D ranged from -0.00003 to 0.00113 (Figure 4.5B) which suggests that up to 0.1 % of alleles were unique between the most genetically distant parasite populations (i.e. Namibia and South Africa). The distribution of private/unique alleles across the four countries showed that Namibia contained the majority of (161/214) private alleles, followed by South Africa (37/214), Eswatini (14/214) and Mozambique (2/214). Similarly, the rarefaction approach which accounted for differences in sample size confirmed that Namibia has both the highest allelic richness and the highest number of private alleles compared to the other countries. The smallest values in both categories of the mean number of distinct alleles per locus and the mean number of private alleles per locus occur in South Africa and Mozambique respectively. This private allelic richness therefore suggests that there is an endogenous circulation of parasites in each of the countries, or possibly random chance due to sampling (since overall frequencies are quite similar) or potentially technical differences. Genetic distance (*Jost's D*) between parasites from the different countries was significantly positively correlated (r^2 = 0.72, *P* = 0.001, Mantel test of matrix correlation) to geographic distance (Figure 4.5C) thus supporting that parasites were isolated based on the geographic distance between them. It is important to note that these values are all very small indicating little if any meaningful difference based on this metric.

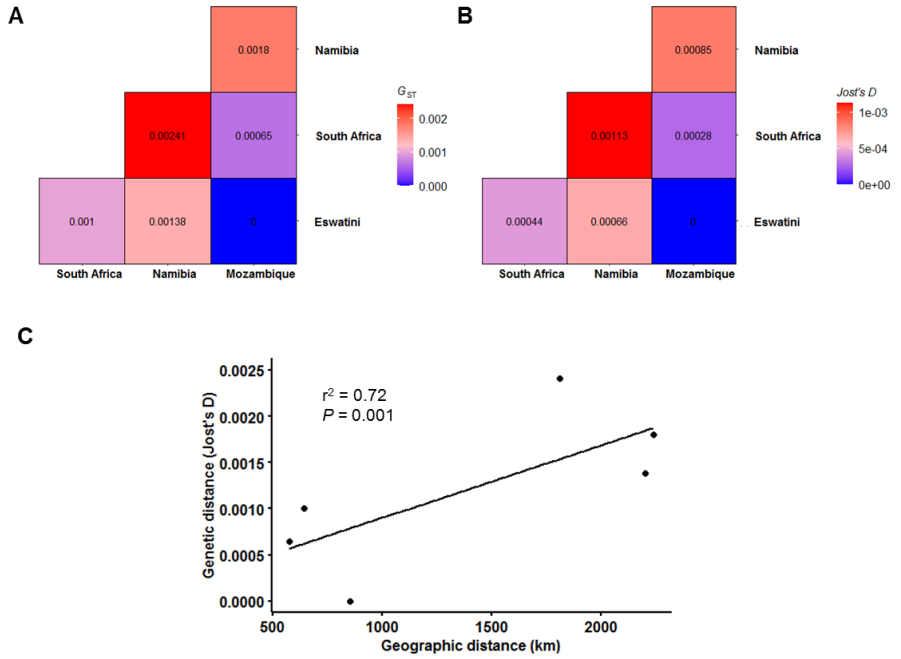


Figure 4.5 Population differentiation of *Plasmodium falciparum* parasites in southern Africa. (A) Pairwise genetic differentiation between parasites from the different countries described by Nei's G_{ST} values. (B) Pairwise genetic differentiation as described by *Jost's D*. (C) Relationship between genetic distance (*Jost's D*) and geographical distance (km). The isolation-by-distance model was tested using the Mantel test.

Using a different analysis tool, clear separation of the Namibian parasite population by LD1 (67.2 % genetic variance explained by DA eigenvalues – top left insert of Figure 4.6A) from those in the MOSASWA (Mozambique, South Africa and Eswatini) region was observed in the DAPC analysis (Figure 4.6A) which suggests that the parasites in Namibia are genetically distinct from those in the MOSASWA block. This suggests interrupted clades of gene flow and provides evidence that some lineages/haplotypes may be unique to the different geographical locations (countries). Another possible explanation for this observation is that the MOSASWA block is a contiguous area whilst Namibia is isolated. LD2 (29.1 % genetic variance explained by DA eigenvalues – bottom left insert of Figure 4.6A) separated the South African parasite population from the Mozambique and Eswatini parasite populations. Where haplotypes from the different countries overlapped, this suggests transmission connectivity supported by geographic proximity. The lack of population differentiation between Eswatini and Mozambique suggests strong gene flow between both populations. While some haplotypes from Mozambique and South Africa also overlap, this is to a lesser extent than those from Eswatini and Mozambique which suggests a less strong gene flow between Mozambique and South Africa.

In order to establish the optimum number of PCs which maximizes the probability of assigning new individuals to their actual group, cross-validation of the DAPC was performed. The scatter plot (Figure 4.6B) shows an almost flat line with mean successful assignment ranging from 87.2 % to 89.2 % for 150 to 550 PCs which suggests that group membership is well-predicted by any of these models. The number of PCs that achieved the highest (0.892) mean successful prediction of subsamples was 500; and that the number of PCs that achieved the lowest (0.11) root mean squared error (RMSE) was 450, also with a high mean successful prediction of subsamples of 0.892. Therefore because of the low RMSE, 450 PCs were selected as the number of PCs that, when kept, allow for the generation of a DAPC with the most informative and generalisable results. The mean successful assignment suggests that after the DAPC was run with 450 PCs using only 90 % of the samples in the training set, the DAPC was able to correctly place the left-out 10 % of the data in the validation set (or build a model that accurately placed individuals) in the correct group 89.2 % of the time. This model therefore has a mean success rate that is 64.2 % higher than the mean success for a random chance approach which suggests that the model to discriminate between groups is not just happening by chance. The mean and CI for random chance was 25.0 % (23.9 %, 26.2 %) represented as horizontal solid (mean) and dashed (CI) lines respectively on the plot generated for cross-validation. However, to prevent possible over fitting that arises from retaining a high number of PCs,

a more conservative number of PCs (150) (top left insert in Figure 4.6A) with a high mean successful prediction of subsamples of 0.872 and low RMSE of 0.129 (Figure 4.6B) was retained for the DAPC analysis for the full data set. The cross-validated predictions that show a similar pattern, therefore make the DAPC findings on the significant population structure observed in Figure 4.6A more reliable.

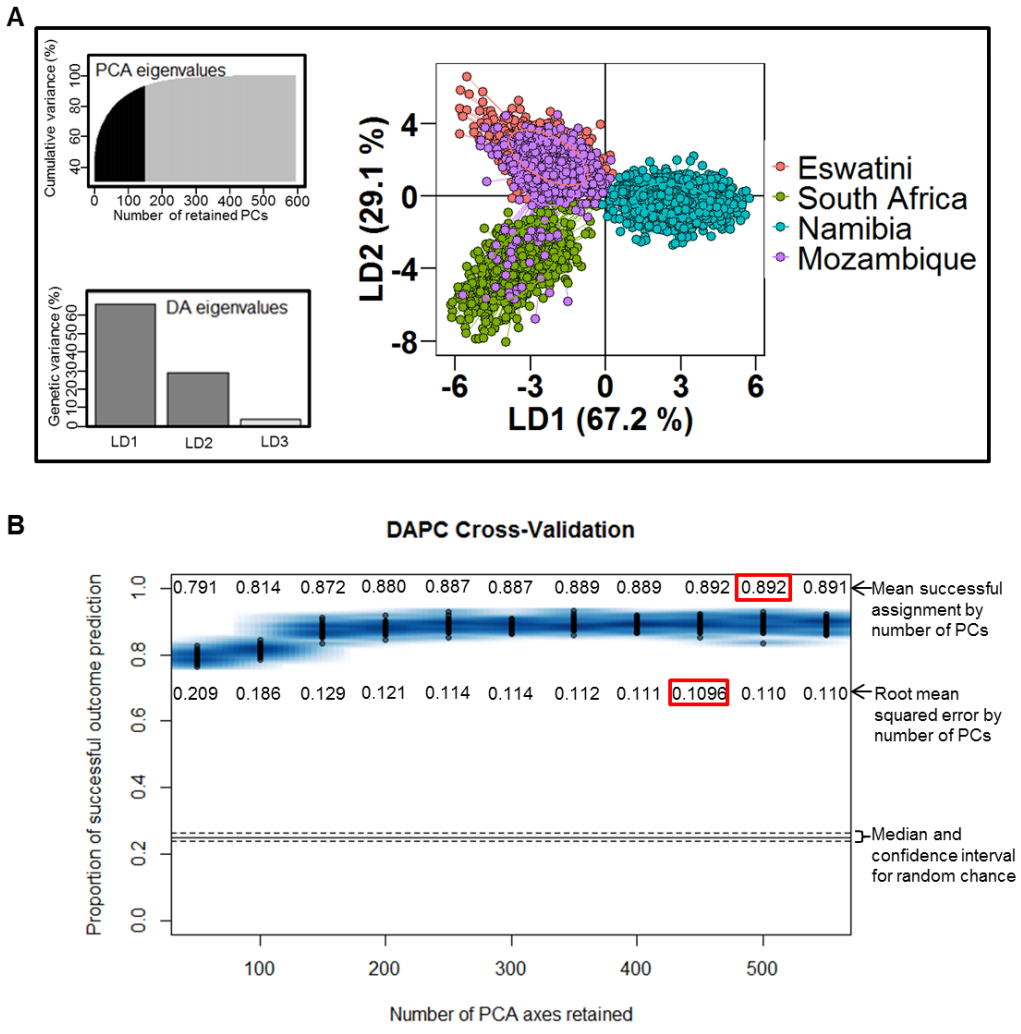


Figure 4.6 Population structure of *Plasmodium falciparum* populations in southern Africa as described by DAPC. (A) Scatterplot of the discriminant analysis of principal components (DAPC) based on discrimination of *P. falciparum* populations stratified by country of origin of infections. Individual multi-locus genotypes appear as dots. Colours and lines represent population membership. Analysis is based on retention of 150 principal components as seen in the top left insert of the figure. (B) Scatterplot of the DAPC cross-validation. The number of PCs retained in each DAPC varies along the x-axis, and the proportion of successful outcome prediction varies along the y-axis. Individual replicates appear as points, and the density of those points in different regions of the plot is displayed in blue. DAPC is carried out on the training set with variable numbers of PCs retained, and the degree to which the analysis is able to accurately predict the group membership of excluded individuals (those in the validation set) is used to identify the optimal number of PCs to retain. At each level of PC retention, the sampling and DAPC procedures were repeated 30 times. The mean and CI for random chance are represented as horizontal solid (mean) and dashed (CI) lines on the plot generated for cross-validation.

An assessment of the alleles that most differentiated the Namibian cluster from those in the MOSASWA block identified PFG377, AS8, AS21, AS7, B7M19 and PolyA as the major loci that contributed alleles that led to this differentiation as shown in the loading plot (Figure 4.8A). On the other hand, distinction of the South African from the Mozambique and Eswatini parasite populations identified AS8, AS3 and PolyA (Figure 4.7B) as the major loci contributing to the observed population structure. This shows how certain haplotypes can lead to the distinction/separation of parasite populations.

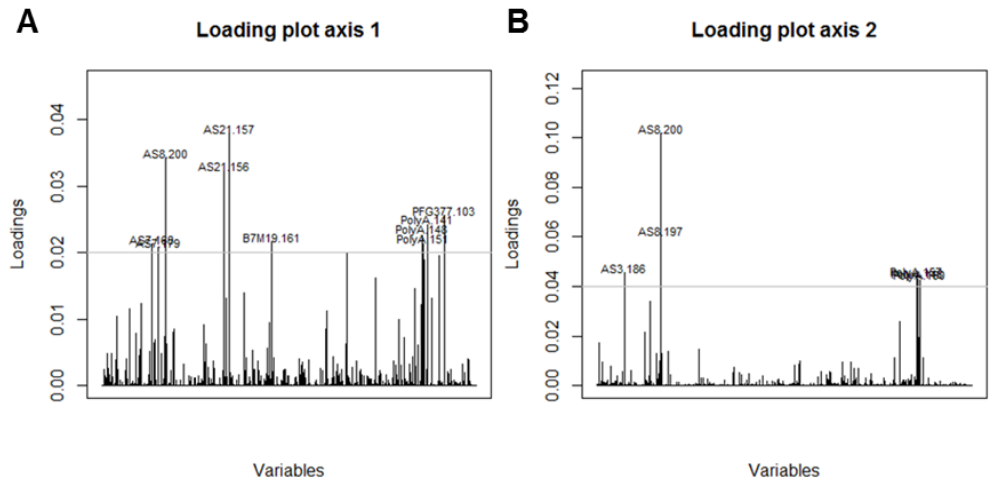


Figure 4.7 Loading plot from DAPC of *P. falciparum* parasite populations showing the contribution of alleles to (A) the first DAPC eigenvalue separating Namibia isolates from all other countries and (B) the second DAPC eigenvalue separating isolates from the MOSASWA block.

4.3.6 Evidence of cross-border parasite genetic connectivity in southern Africa

To determine the level of interconnectivity of parasites in the region, genetic relatedness, and connectivity between parasites from the different countries was assessed. Out of the 5314 samples studied, there were no identical multi-locus genotypes shared across the different parasite populations from the different countries. This was expected given the high level of parasite genetic diversity in the region described in sections 4.3.2 and 4.3.3 and high variability in microsatellite markers. Parasite genotypes were shared across countries at different loci (Figure 4.9), which suggests inferred interconnectivity between parasite populations. The number of genotypes per locus shared across the 4 populations ranged from 15 (locus AS12) to 168 (locus AS25) (Figure 4.9). There were abundant genotypes (shown in red) in most loci and that certain parasite genotypes were shared across all the 4 countries with the most prevalent genotypes found in markers such as TA81. Although other loci such as AS2 and PolyA had several shared genotypes across

all populations up to 54 and 100 respectively, these genotypes were all found at low frequencies. What is also clear across loci for example AS25 is that while some genotypes may be shared between all countries, some are only found in specifically two or three countries which may be an indication of parasite gene flow between the countries, adaptation of parasite strains to certain areas, or just due to sampling issues with rare alleles.

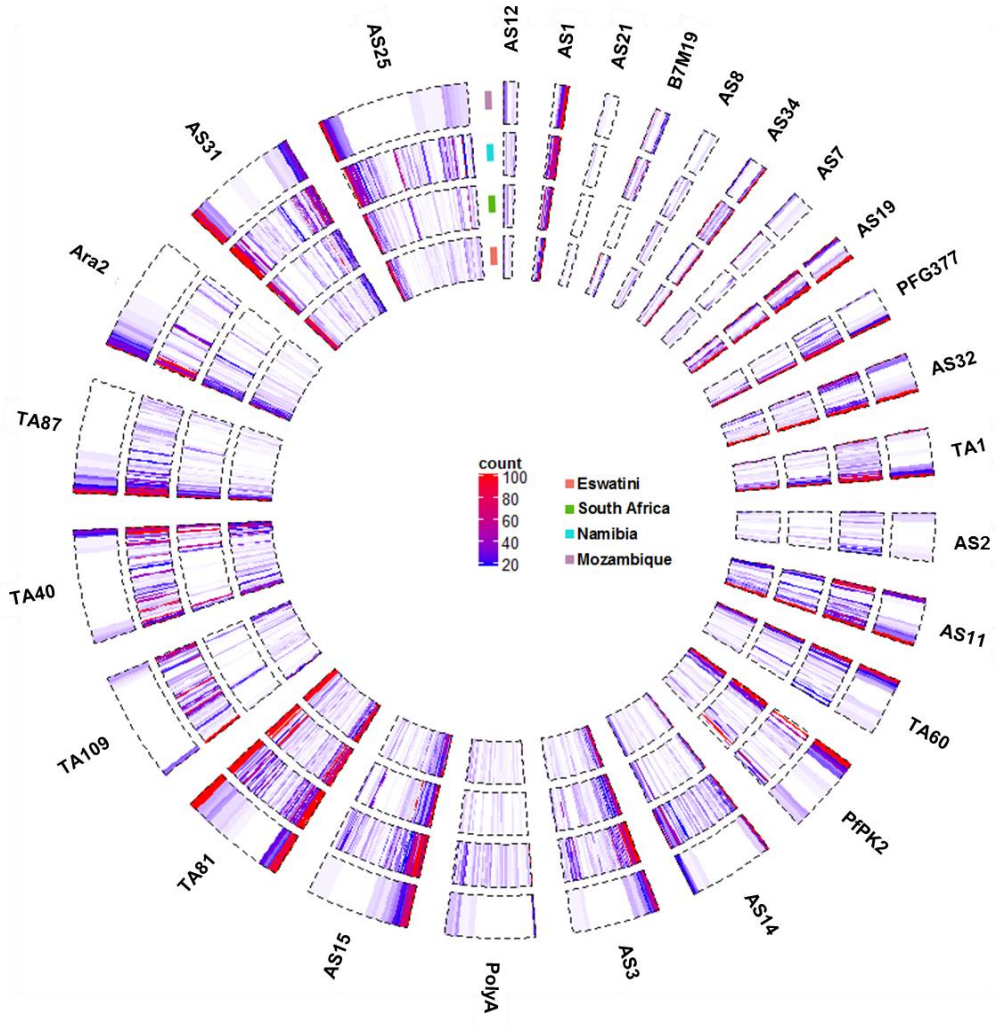


Figure 4.8 Number of shared genotypes per locus across parasite populations from the four countries in the southern African Elimination-8 region. The circular heatmap shows the number of genotypes per locus shared across the four populations in clockwise ascending order. The colour scale shown represents white for 0 or 1 shared genotypes, blue scale for low frequency while the red indicates high frequency of ≥100. The tracks from outwards going towards the inner part of the circle represent the countries in order of decreasing transmission intensity (i.e., Mozambique (outermost), Namibia, South Africa, Eswatini (innermost)).

These results were supported by evidence of geneflow between parasites from the different countries. The greatest historical geneflow was observed between Mozambique and Eswatini as shown by the thickness of the cords in the circos plot (Figure 4.10A).

Mozambique received 77 % of migrants from Eswatini and Eswatini received 100 % of its migrants from Mozambique as indicated by relative migration of the parasites between the population pair determined by *divMigrate* online (Figure 4.10A). While in practice, it is very unlikely that Mozambique received many if any migrants from Eswatini given the differences in transmission intensity, it is interesting that this analysis indicates the expected differential in the direction. The least relative migration was observed from Namibia to Eswatini (2 %), Namibia to Mozambique (2 %) and South Africa to Namibia (2 %). Although South Africa is relatively geographically close to Eswatini and Mozambique, parasite migration to and from these countries (Eswatini and Mozambique) was relatively low and ranged between 3 % and 7 %. The migration patterns displayed in this plot support the population structure observations made in the DAPC analysis in Figure 4.6A. These results suggest that Mozambique is the major source area and Eswatini the major sink area of parasite populations between these 4 countries in the region.

Patterns of recent parasite gene flow estimates based on BayesAss migration data confirmed that most of the gene flow is between Eswatini and Mozambique. Mozambique had the majority of (parasite) emigrants (21 % to Eswatini, 20 % to South Africa, and 19 % to Namibia) (Figure 4.10B) compared to any other country which supports that it is the major source country. Eswatini on the other hand received most of its migrants from Mozambique (18 %) compared to 0.1 % from South Africa and 0.02 % from Namibia. Namibia and South Africa both received the least migrants of less than 1 % from each of all the other countries. The ‘humps’ in Figure 4.10B represent gene flow originating from within the same country which is indicated in all 4 countries. This supports endogenous circulation of parasites revealed by private allelic data in each of the countries. Overall, these analyses showed gene flow between *P. falciparum* parasite populations in all four countries which implies interconnectivity between the parasite populations. Additionally, there appeared to have been considerably more gene flow between the countries historically than recently which may have been influenced by implementation of control interventions and cross border interventions.

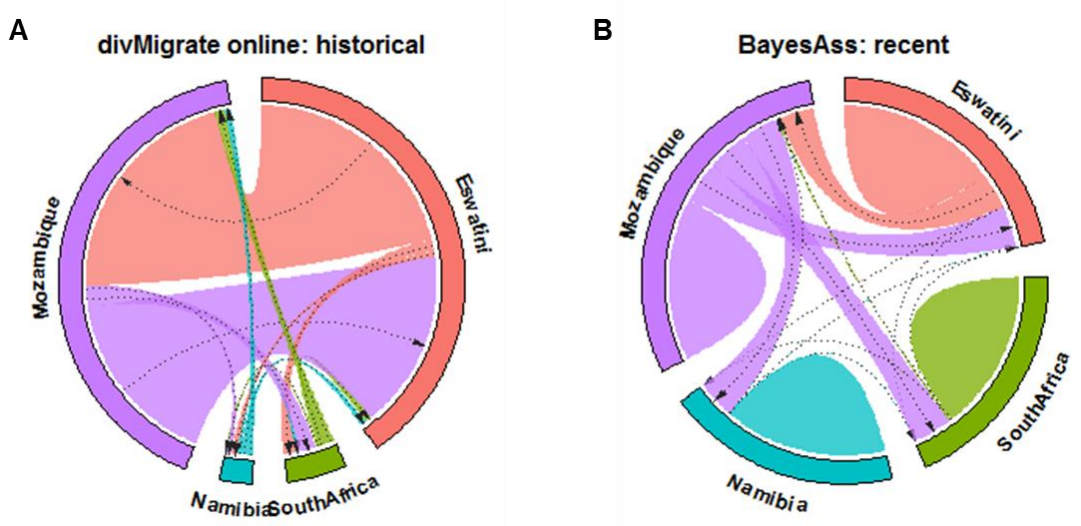


Figure 4.9 Gene flow diagrams of historic and recent inferred genetic connectivity between *P. falciparum* parasites among four countries in the southern African Elimination-8 region. (A) Historical gene flow estimates were derived from divMigrate online. The relative migration values were derived from *Jost's D*. (B) Recent gene flow estimates were calculated using BayesAss. In both plots, arrows indicate direction of gene flow from one country to another. Arrow thickness reflects the strength of gene flow between different parasite populations. Patterns for each diagram are independent. Similar widths of arrows do not represent the same amount of gene flow across each of the two diagrams.

Since the greatest gene flow was observed between Mozambique and Eswatini, the level of pairwise relatedness (IBS) between parasites from these 2 countries was assessed. Out of 11476 pairwise comparisons, 9841 (86 %) of the infections were found to be highly related (IBS >0.6). Of these highly related parasites, there was no significant difference ($P > 0.05$, ANOVA) in the level of relatedness between infections from Mozambique (mean IBS = 0.71 ± 0.07) and Eswatini (mean IBS = 0.70 ± 0.07) (Figure 4.11A). There was a poor correlation (Pearson's r = 0.0013, $P = 0.89$) between the genetically related infections and geographic distance between them (Figure 4.11B) which confirms the lack of parasite population structure and parasite mixing between the 2 countries. Infections more than 600 km apart could be linked through their genetic similarity which suggests interconnectivity between parasites from the different countries most likely due to human movement since mosquitoes cannot fly such long distances.

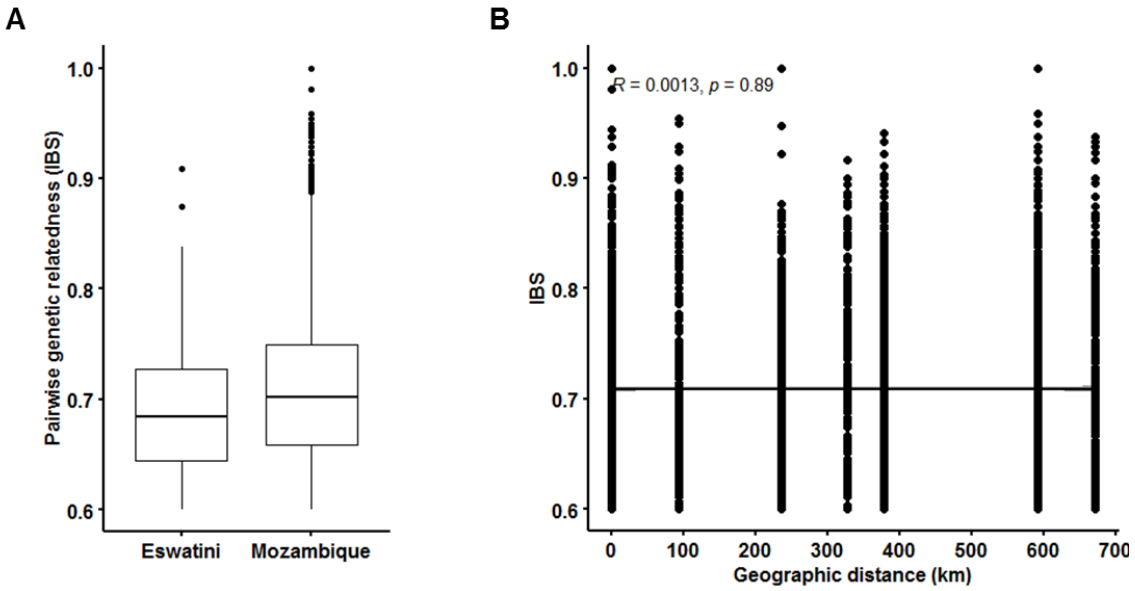


Figure 4.10 Pairwise genetic relatedness between parasite infections from Eswatini and Mozambique. (A) Box plots comparing the distribution and mean of highly related infections in the two countries. (B) Correlation of IBS and geographic distance of the infections.

4.4 Discussion

The potential use of *P. falciparum* population genetics tools in elimination efforts in southern Africa is now gaining momentum. Findings from this Chapter demonstrate strong evidence for parasite population structure in the southern African region and provide insight into patterns of parasite gene flow and connectivity within and across national borders of the selected endemic countries in the E8 region. This data demonstrates the potential of efficiently generating genetic information with similar shared criteria that can be used for regional malaria control and elimination efforts, as well as the potential utility of supplementing traditional surveillance data with molecular surveillance data generated by direct evaluation of the parasite population genetics.

Failure of both indices of parasite within-host and population level diversity to reflect the level of transmission intensity in the southern African parasite populations as shown in this study highlights the complexity of transmission dynamics in such highly interconnected areas in the region. This study has shown that the maintenance of high levels of parasite genetic diversity may be a critical barrier to malaria elimination in this region. Connectivity between endemic and pre-elimination areas in the region aided by human migration of infected individuals may offset the effect of geographic fragmentation in parasite

populations thus enabling re-introduction to areas where the disease may already have been controlled.

Genetic data confirmed the high level of gene flow particularly in the MOSASWA block, which is possibly linked to geographic proximity of the countries and human migration for economic reasons. Relative migration analysis of multilocus haplotypes allowed investigation of the strength and direction of parasite flow amongst the parasite populations from the four different countries. This showed more parasite migration between Mozambique and Eswatini, and relatively limited migration to and from South Africa and Namibia which was consistent with genetic differentiation, private allelic and clustering analyses. Similar trends of the lack of separation between parasites from countries with different malaria transmission intensities sharing borders in sub-Saharan Africa were also observed between Zambia (moderate-transmission) and the Democratic Republic of the Congo (high-transmission) [250]. This was also true for Namibia (low-transmission) and higher-transmission neighbours Angola and Zambia [122]. There was no evidence of parasite population structure observed between the countries involved suggesting contiguous transmission zones in those areas [122, 250]. This evidence was supported by the shared parasite genotypes at most of the loci thus inferring genetic connectivity between geographically distinct locations.

The wide dispersal of shared parasite genotypes across the different countries suggests that connectivity among the different endemic areas, likely caused by human migration (since mosquitoes cannot fly over such long distances), sustains disease transmission in the region hindering elimination efforts. The small proportion of infections from Mozambique that clustered with those from South Africa suggested limited genetic connectivity/relatedness and thus possibly less importation of parasites from Mozambique into South Africa. This observation may have been influenced more by the fact that the whole South African parasite population was not entirely represented since the sampling plan used to get the samples was more opportunistic and was not entirely designed for a national population genetics study. Therefore, only a small proportion of samples collected from a hotspot area in one of the provinces in the country was used which may have been more representative of that area only and not the entire country, a limitation that can be addressed by adjusting the sample collection process to include a wider representation of all or as many infections as possible in all remaining endemic areas in the country in future.

Additionally, the close geographic proximity of this South African hotspot area where the majority of samples were collected [289], to the Zimbabwean southern border region and the direct transport route that links the two countries raises a question of whether parasites from Zimbabwe instead may be more similar to those in the Limpopo Province of South Africa and may be the greater source of imported infections seeding local transmission in that area. Alternatively, it may also be possible that there is not a lot of importation in this area and that there is primarily sustained local transmission. Unfortunately, there were no samples genotyped using the same criteria as that of all other samples in this study that represented the Zimbabwean or other neighbouring country parasite populations at the time this analysis was done to enable us to make this assessment. However, in low transmission settings, importation from higher transmission settings leads to seeding events which then becomes local transmission.

Overall, these analyses showed gene flow between *P. falciparum* parasite populations in all four countries which implies interconnectivity between the parasite populations. The results, based on the analysed sample set suggest that Mozambique is the major source area and Eswatini the major sink area of parasite populations between these 4 countries in the region. This confirms the importation reported between these two countries based on patient travel history information. The limited gene flow to and from South Africa (Vhembe District, Limpopo Province) and Namibia also confirms the high level of local transmission as reported through patient travel data [122, 289]. Although the signals of gene flow to and from these areas were not very strong, they were present over a wide geographic range (distance), which indicates possible genetic connectivity of the parasite populations. This suggests that movement of malaria parasites by human reservoirs may connect geographically distinct malaria transmission areas in southern Africa.

While cross border genetic connectivity between certain countries in the region was already reported in a separate studies [122, 188, 250], collating this genetic data with that in the rest of the region would therefore make the regional analysis in the E8 region stronger as a wider geographic scale is covered. Genotype data was collated from 4 and not all the 8 countries in the E8 region for the study because of the unavailability of data from the remaining countries on the chosen microsatellite genotyping platform. A major constraint of studies that require direct comparisons from microsatellite markers is that they require the same experimental conditions and genotyping analysis parameters, and data processing settings to be able to be directly compared to each other, as any

inconsistences in data processing settings may potentially distinguish/differentiate between otherwise identical alleles/genotypes (difference of a single allele may be interpreted as a different allele/genotype). With additional microsatellite genetic data from the remaining 4 countries in the E8 region, or other types of data, which may be less prone to the limitations that arise from microsatellite markers, and may have higher resolution to detect signals of genetic relatedness (e.g. highly multiplexed amplicon sequencing), genetic data of a larger spatial scale can be shared within the region facilitating the identification of origins of imported infections so as to be able to put interventions in place where there is a higher risk of this occurring. A major strength of this current study was the sample size in which a good representation from affected areas was achieved.

The patterns of private allelic richness that suggest an endogenous circulation of parasites in the eliminating countries is an interesting finding which could further hamper elimination efforts in the respective countries. From an elimination perspective, in a low-transmission elimination setting like Eswatini or the KZN Province of South Africa, this shows that although imported infections play a significant role in continued transmission, there is also a contribution of locally acquired/generated parasites (parasites of local/internal origin) circulating within the individual countries preventing those countries from "getting to zero" (local) infections. This finding may therefore assist by giving further insights into the behaviour of the parasite for example in the MOSASWA malaria cross-border initiative works through harmonized collaborative efforts to achieve zero local transmission in Eswatini, South Africa and Maputo Province, Mozambique by 2023 and pre-elimination status in southern Mozambique (Maputo and Gaza Provinces) by 2025 [233, 319, 320].

Harmonising the time and place when appropriate interventions can be deployed is, however, governed by understanding the impact of imported malaria infections on local transmission. In the event of an endogenous circulation of parasites, local control measures such as vector control will be necessary. However, if there is a strong genetic connectivity of local and imported infections, then, interventions aimed at decreasing malaria in the source areas of infection or decreasing vulnerability to importation may be required [3].

The results generated in this study are a useful starting point to a future larger study which includes all 8 countries in the E8 region which will facilitate with decision making for malaria elimination efforts in the region by targeting of interventions effectively in both

source and sink areas on both a national and regional level thus preventing continued spread of infections in the region and hopefully achieve elimination. This data can also be used as an early warning system for preventing undetected spread of for example drug resistant and imported infections.

4.5 Conclusion

Studying the parasite population genetics in the selected four countries provided a preliminary understanding of parasite genetic connectivity in this area of the Elimination 8 region of Africa, as well as added to the knowledge of understanding local and cross-border malaria transmission dynamics in the region. Results in this study showed strong signals of parasite population structure and genetic connectivity between malaria parasite populations across national borders which calls for strengthening the harmonisation of malaria control and elimination efforts between the Elimination 8 countries. This data also proves its potential utility as an additional surveillance tool for malaria surveillance on both a national and regional level as countries work towards malaria elimination. Due to its retrospective nature, this study could not however optimize comparisons between countries due to the different sampling approaches used. It is possible that genetic metrics could be more useful with standardized sampling taking into account heterogeneity of malaria transmission. Stratification based on transmission intensity, risk of importation, hotspots and outbreaks can ensure that genetic analysis be more useful.

CHAPTER 5

GENERAL DISCUSSION AND CONCLUSIONS

In the absence of vaccines, malaria control and elimination programs rely on interventions aimed at reducing disease transmission. However, traditional methods for measuring malaria incidence have limited power in low-transmission settings. At low-transmission, residual transmission may be driven by clusters of cases that originate from local cases or from imported cases that seed local transmission. In order to accurately distinguish between these cases, the genetic variation of the parasite populations should be studied to better target interventions and track local transmission in elimination settings. The molecular tools used in this study to generate this data demonstrated the utility of microsatellite genotyping as a highly sensitive method for monitoring the parasite population and understanding transmission dynamics in South Africa and part of the E8 region particularly the MOSASWA block.

For the first time, it was shown that the *P. falciparum* population in the Limpopo Province, an area of moderate-transmission intensity in South Africa, was characterised by complex and moderate to high genetic diversity. The stability of this genetic diversity over different temporal and geospatial scales confirmed sustained local residual transmission which is hindering South Africa's progress towards malaria elimination. It was also shown for the first time in South Africa that asymptomatic mobile populations from neighbouring Mozambique may carry genetically diverse parasites that may seed local transmission in pre-elimination settings of South Africa. This again is another contributing factor that presents major challenges for malaria elimination efforts in South Africa. These results can be used by NMCPs to inform intervention selection, resource deployment, surveillance, and case investigation in the affected pre-elimination areas in South Africa. Amongst some of the limitations of the study however, were the small sample sizes in Elim and Thohoyandou, as well as the large number of samples with unknown origin that underpowered the geospatial analysis to some extent.

Parasite migration and gene flow patterns suggested that Mozambique is the major source of infection detected in Eswatini, the major sink country in the MOSASWA block correlated well with reported travel history data [233]. The confirmed interconnectivity of parasites from within and between countries in the E8 region reinforces the view that malaria elimination in any southern African country would not be possible without regional

cooperation and collaboration [140]. This is because continued movement of infected people from areas of high- to low-transmission allows for the continuous movement of the malaria parasite across borders making country elimination efforts difficult. These results can be used to inform program planning and resource deployment. Countries should therefore continue to work together through established cross-border initiatives that are coordinated and supported by the E8 Initiative for elimination to be achieved [140].

Malaria surveillance using population genetics tools as demonstrated in this study can be used by NMCPs as a value-added surveillance tool to distinguish malaria source and sink populations, distinguish between local and imported cases and guide decision-making around intervention selection and implementation. Where new interventions are then employed based on this genetic information, the level of genetic diversity can then be evaluated to assess whether the employed interventions have been effective or not. Given the high level of interconnectedness within the E8 region, country genetic data must be shared in a timely and consistent manner through the appropriate regional platforms for such information to have an impact on a regional level.

While novel data sharing platforms developed and supported by the E8 such as the Situation Room have been instrumental in the early detection of and prompt response to malaria outbreaks in the region [140], there remain persistent challenges that impede malaria elimination efforts in the region. These include the reluctance by NMCPs to share data on regional platforms and to adopt new technologies and techniques such as single-low-dose-primaquine [321] and the genomic surveillance described in this thesis and in other studies in the region [122]. This is in spite of it being widely acknowledged that such methods are required to achieve malaria elimination [140]. NMCPs must therefore be encouraged and supported to adopt and implement novel techniques with evidence of impact and training.

To perform robust population genetics studies, adequate sampling of *Plasmodium* infections in a locality is necessary. In low-transmission areas such as South Africa, Eswatini and Namibia, sampling from different localities or provinces within a country with different levels of transmission intensity, as well as from different time points, is necessary to characterise the dynamics of heterogeneous structured populations since they may need different control strategies at a national level. Such sampling should include not only infections from symptomatic patients but also asymptomatic people and people living in areas of difficult accessibility that are less likely to access health care.

In general, a *Plasmodium* population genetics study must start with a decision regarding appropriate genetic markers. Markers must have the correct sensitivity for the research question. It is possible to have too much information if markers are too different with nothing to link them, or too little information where no signal will be achieved [119]. Genetic polymorphisms such as SNPs and microsatellite markers are valuable to explore the dynamics of parasites with important phenotypes. Such genetic information is used as molecular barcodes to track the origin of infections [175, 176, 206, 322]. However, it is important to acknowledge how each of the different genetic markers could or could not provide precise information for control and elimination purposes in the South African and southern African regional context. For example, the fact that SNPs become less informative in polyclonal infections vs. multi allelic markers such as microsatellites or microhaplotypes may render them unsuitable for operational investigations [223].

Microsatellites and microhaplotypes, on the other hand, would be more suitable for this purpose due to their mutation rate [186, 220]. Although mutation patterns of microsatellites are complex [221], they are still valuable to provide information for epidemiological time scales. Microsatellites have been extensively used to investigate the population structure of *Plasmodium* [122, 145, 152, 153, 188, 189, 213, 223] and optimized protocols for microsatellite genotyping and analysis in *P. falciparum* and even *P. vivax* are currently available [155, 223, 322] as well as guidelines for data interpretation according to the specific research question. Considering the evaluation time frame of three years to certify an area as malaria free, the most suitable markers would be those evolving more quickly such as microsatellites. However, microsatellite genotyping poses several challenges with standardisation of methods. This makes analysing data that was not generated and processed together difficult.

High resolution genotyping which is more standardised across laboratories and can evaluate more markers than microsatellites can therefore be used as an alternative to determine genetic diversity and accurately assign parasites to their source of origin [186]. Genotyping several loci can also further help to simultaneously address the questions of unique challenges of malaria elimination in the E8 countries such as *Plasmodium* species identification, drug resistance monitoring and diagnostic resistance (pfhrp2/3 deletions). High resolution genotyping such as deep amplicon sequencing as described in Chapter 1.7.1.2, however, usually requires sophisticated infrastructure with adequate computing power and highly trained personnel for data analysis [119, 186, 323]. As it is not feasible for this sequencing capacity to be built in each country, the establishment of highly

sophisticated sub-regional laboratory networks where samples are processed collectively and data is shared as seen with COVID-19, should be considered [127]. Building and/or strengthening national capacity is also important to complement the regional lab as this will support NMCPs more effectively in real time. Use of Nanopore sequencing (MinION) [324] for example, which is a mobile sequencing platform used in the field and has reportedly been used to detect different sequences of the *P. falciparum* parasite genomes in the field can aid in such efforts. Therefore, if these approaches can be used, there can be both a national and regional genotyping database where information is generated and shared more timeously to put interventions in place where necessary. Such platforms can also be used to concurrently assess the emergence and trends of antimalarial drugs and insecticide resistance, as well as important changes in parasites or vectors such as deletion of pfhrp2/3 genes.

While such work is already being done and data is available [118], data collected is fragmented, with limited regional impact. A more inclusive sample set would therefore have a greater impact. For example, using longitudinal samples/data from a particular region to look at a particular population of parasites as artemisinin resistance emerges and find out the particular parasite strains/genes sweeping throughout the population. This will provide early warning signs for emerging resistance. Collection of data has already started along the Mymmar-Thailand border which is a hot zone for artemisinin resistance [325]. In Africa, in Thies, Senegal, routine genetic surveillance for drug resistance has been on-going. They have been tracking resistance to various antimalarials using molecular markers (*crt, mdr, dhfr, dhps*) for just over two decades since 2008 [326]. Key implications have been that genetic surveillance for drug resistance markers reveals dynamic trends, and that genetic surveillance for emerging drug resistance loci provides early warning system [326]. Also in Mali, a powerful antimalarial genetic group leads the genetic surveillance of Pf(K13) for potential emergence of ACT resistance across Africa [207, 327].

While there are many on-going malaria genomics studies, the next critical step is to integrate this information which is primarily in research projects into the NMCPs surveillance systems with the ultimate vision of using the data in the same way as the polio or HIV communities use their genomic databases [119]. The important thing, however, is how the information generated is going to be used to inform decision making. This is the challenge that may be faced for the next few of years to come. The way in which the malaria community is going to take this challenge of this big genetic data that is

being and will continue to be generated and incorporate it in such a way that is actually informative for the people making decisions daily about how the programmes are progressing and how to allocate resources remains to be seen.

There will be need for teams that will generate and evaluate the genetic data, technical validation as is handled with standard routine data; political validation in terms of getting governments and decision makers to decide that this is an important and valuable tool that they want to spend resources on; then the actual use of the data in the decision making process [119]. The biggest technical barriers currently (in addition to the very important political and coordination barriers highlighted above) are the development of analytical approaches to appropriately answer the questions of interest using available data and tailored to Plasmodium biology and epidemiology [119]. While there may be challenges along the way and difficulties may be encountered, it is time to start integration of genomic and genetic surveillance into NMCPs as it will go a long way in malaria eradication efforts as learnt from other diseases ([127].

Lessons learnt from feedback from malaria control programmes and researchers in the Asia-Pacific region on implementing parasite genotyping into national surveillance frameworks, which may also be applicable to South Africa and the southern African regional context as a whole, revealed that whilst the utility of multiple use cases for parasite genotyping was acknowledged, they agreed that the priority of each use case would vary with different endemic settings [328]. They also agreed that a one-size-fits-all approach to molecular surveillance would unlikely be applicable across the Asia–Pacific region, and that agreement on the spectrum of value-added activities would help support data sharing across national boundaries [328]. Other outcomes from this consultation included the fact that knowledge exchange was required to establish local expertise in different laboratory-based methodologies and bioinformatics processes [328]. Collaborative research involving local and international teams was also pointed out as a way to help maximize the impact of analytical outputs on the operational needs of NMCPs [328]. It was also noted that research was required to explore the cost-effectiveness of genetic epidemiology for different use cases to help to leverage funding for wide-scale implementation [328]. Engagement between NMCPs and local researchers would also be critical throughout this process [328].

In future studies, genetics may even allow the prediction of vectoral capacity particularly in elimination settings to answer the question of whether an imported parasite can be transmitted. This can be done by evaluating the parasite's sexual stage protein *Pfs*47,

which is already known to play an important role in the process in which ookinetes evade immunity of anophiline mosquitoes [55]. There is currently on-going work assessing *Pfs*47 and its polymorphisms, associated with different geographic origins as well as the different abilities of mosquitoes to actually become infectious/infected with the parasite [329]. Similarly, other researchers are investigating the potential for being able to predict vectoral capacity [330, 331]. While these are mostly research projects, by combining the needs of NMCP and research communities, these kinds of innovations may be fast tracked.

Incorporating genomic surveillance into routine malaria surveillance in the E8 region may be met with intellectual tension with some advocating for the continued use of traditional methods to eliminate malaria given that malaria has been eliminated in many places without these new tools; and also, that introducing new technologies may take away precious time from NMCPs. However, is important to ask critical questions which can be answered through genomics for malaria elimination to be achieved. DNA can bridge the gap between current surveillance and what is actually going on in communities by being able to get the granularities of what is actually going on in individual infections as it relates to transmission.

CHAPTER 6 REFERENCES

1. WHO. World Malaria Report [Internet]. World Health Organisation; 2020.
2. WHO. World Malaria Report [Internet] World Health Organisation; 2019.
3. WHO. Eliminating malaria. World Health Organisation; 2016.
4. WHO. Update on the E-2020 Initiative of 21 malaria-eliminating countries. 2018.
5. WHO. World Malaria Day: WHO launches effort to stamp out malaria in 25 more countries by 2025 [press release]. World Health Organisation, 21 April 2021 2021.
6. Kiszewski A, Mellinger A, Spielman A, Malaney P, Sachs SE, Sachs J. A global index representing the stability of malaria transmission. The American journal of tropical medicine and hygiene. 2004;70(5):486-98.
7. Singh B, Sung LK, Matusop A, Radhakrishnan A, Shamsul SS, Cox-Singh J, et al. A large focus of naturally acquired Plasmodium knowlesi infections in human beings. The Lancet. 2004;363(9414):1017-24.
8. WHO. World malaria report 2020: 20 years of global progress and challenges. World Health Organisation; 2020.
9. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, Hay SI, et al. The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. PLoS Med. 2008;5(2):e38.
10. Doolan DL, Dobaño C, Baird JK. Acquired immunity to malaria. Clinical microbiology reviews. 2009;22(1):13-36.
11. Nkhoma SC, Nair S, Cheeseman IH, Rohr-Allegrini C, Singlam S, Nosten F, et al. Close kinship within multiple-genotype malaria parasite infections. Proceedings of the Royal Society B: Biological Sciences. 2012;279(1738):2589-98.
12. Jones MK, Good MF. Malaria parasites up close. Nature medicine. 2006;12(2):170-1.
13. Greenwood BM, Fidock DA, Kyle DE, Kappe SH, Alonso PL, Collins FH, et al. Malaria: progress, perils, and prospects for eradication. The Journal of clinical investigation. 2008;118(4):1266-76.
14. Bartoloni A, Zammarchi L. Clinical aspects of uncomplicated and severe malaria. Mediterranean journal of hematology and infectious diseases. 2012;4(1).
15. Marsh K, Forster D, Waruiru C, Mwangi I, Winstanley M, Marsh V, et al. Indicators of life-threatening malaria in African children. New England journal of medicine. 1995;332(21):1399-404.
16. Carter R, Graves P. Gametocytes. Malaria: principles and practice of malariology Volume 1. 1988:233-305.
17. Eichner M, Diebner HH, Molineaux L, Collins WE, Jeffery GM, Dietz K. Genesis, sequestration and survival of *Plasmodium falciparum* gametocytes: parameter estimates from fitting a model to malariatherapy data. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2001;95(5):497-501.
18. Snow RW, Gilles HM. The epidemiology of malaria. Essential malariology: CRC Press; 2017. p. 85-106.
19. Stresman GH. Beyond temperature and precipitation: ecological risk factors that modify malaria transmission. Acta tropica. 2010;116(3):167-72.
20. WHO. A framework for malaria elimination: World Health Organization; 2017.
21. WHO. WHO malaria terminology. World Health Organization; 2016.
22. Cohen JM, Ernst KC, Lindblade KA, Vulule JM, John CC, Wilson ML. Topography-derived wetness indices are associated with household-level malaria risk in two communities in the western Kenyan highlands. Malaria journal. 2008;7(1):1-12.

23. Oesterholt M, Bousema J, Mwerinde O, Harris C, Lushino P, Masokoto A, et al. Spatial and temporal variation in malaria transmission in a low endemicity area in northern Tanzania. Malaria Journal. 2006;5(1):1-7.
24. CDC. Where Malaria Occurs [Available from: https://www.cdc.gov/malaria/about/distribution.html#.
25. OU. Communicable Diseases Module: 6. Factors that Affect Malaria Transmission 2021 [Available from: https://www.open.edu/openlearncreate/mod/oucontent/view.php?id=89&printable=1.
26. Machault V, Gadiaga L, Vignolles C, Jarjaval F, Bouzid S, Sokhna C, et al. Highly focused anopheline breeding sites and malaria transmission in Dakar. Malaria Journal. 2009;8(1):1-21.
27. Bousema T, Drakeley C. Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. Clinical microbiology reviews. 2011;24(2):377-410.
28. Russo G, Faggioni G, Paganotti GM, Dongho GBD, Pomponi A, De Santis R, et al. Molecular evidence of *Plasmodium vivax* infection in Duffy negative symptomatic individuals from Dschang, West Cameroon. Malaria journal. 2017;16(1):1-9.
29. Leslie T, Briceño M, Mayan I, Mohammed N, Klinkenberg E, Sibley CH, et al. The impact of phenotypic and genotypic G6PD deficiency on risk of *Plasmodium vivax* infection: a case-control study amongst Afghan refugees in Pakistan. PLoS medicine. 2010;7(5):e1000283.
30. Snow RW, Marsh K. The consequences of reducing transmission of *Plasmodium falciparum* in Africa. 2002.
31. Culleton R, Carter R. African *Plasmodium vivax*: distribution and origins. International journal for parasitology. 2012;42(12):1091-7.
32. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM. Vivax malaria: neglected and not benign. The American journal of tropical medicine and hygiene. 2007;77(6_Suppl):79-87.
33. Baird JK. Neglect of Plasmodium vivax malaria. Trends in parasitology. 2007;23(11):533-9.
34. Culleton R, Ndounga M, Zeyrek FY, Coban C, Casimiro PN, Takeo S, et al. Evidence for the transmission of *Plasmodium vivax* in the Republic of the Congo, West Central Africa. The Journal of infectious diseases. 2009;200(9):1465-9.
35. White NJ. Determinants of relapse periodicity in *Plasmodium vivax* malaria. Malaria journal. 2011;10(1):1-36.
36. Amambua-Ngwa A, Amenga-Etego L, Kamau E, Amato R, Ghansah A, Golassa L, et al. Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. Science. 2019;365(6455):813-6.
37. Cappellini MD, Fiorelli G. Glucose-6-phosphate dehydrogenase deficiency. The lancet. 2008;371(9606):64-74.
38. Francis RO, Jhang JS, Pham HP, Hod EA, Zimring JC, Spitalnik SL. Glucose-6-phosphate dehydrogenase deficiency in transfusion medicine: the unknown risks. Vox sanguinis. 2013;105(4):271-82.
39. Wastnedge E, Waters D, Patel S, Morrison K, Goh MY, Adeloye D, et al. The global burden of sickle cell disease in children under five years of age: a systematic review and meta-analysis. Journal of global health. 2018;8(2).
40. Piel FB, Steinberg MH, Rees DC. Sickle cell disease. New England Journal of Medicine. 2017;376(16):1561-73.
41. Russell TL, Govella NJ, Azizi S, Drakeley CJ, Kachur SP, Killeen GF. Increased proportions of outdoor feeding among residual malaria vector populations following

increased use of insecticide-treated nets in rural Tanzania. Malaria journal. 2011;10(1):1-10.
42. Gonçalves BP, Kapulu MC, Sawa P, Guelbéogo WM, Tiono AB, Grignard L, et al. Examining the human infectious reservoir for *Plasmodium falciparum* malaria in areas of differing transmission intensity. Nature communications. 2017;8(1):1-11.
43. Busula AO, Bousema T, Mweresa CK, Masiga D, Logan JG, Sauerwein RW, et al. Gametocytemia and attractiveness of *Plasmodium falciparum*–infected Kenyan children to Anopheles gambiae mosquitoes. The Journal of infectious diseases. 2017;216(3):291-5.
44. Cohuet A, Simard F, Wondji CS, Antonio-Nkondjio C, Awono-Ambene P, Fontenille D. High malaria transmission intensity due to *Anopheles funestus* (Diptera: Culicidae) in a village of savannah–forest transition area in Cameroon. Journal of medical entomology. 2004;41(5):901-5.
45. Brady OJ, Godfray HCJ, Tatem AJ, Gething PW, Cohen JM, McKenzie FE, et al. Vectorial capacity and vector control: reconsidering sensitivity to parameters for malaria elimination. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2016;110(2):107-17.
46. Cornel AJ, Lee Y, Almeida APG, Johnson T, Mouatcho J, Venter M, et al. Mosquito community composition in South Africa and some neighboring countries. Parasites & vectors. 2018;11(1):1-12.
47. Beier JC. Malaria parasite development in mosquitoes. Annual review of entomology. 1998;43(1):519-43.
48. Curtis CF, Mnzava AE. Comparison of house spraying and insecticide-treated nets for malaria control. Bulletin of the World Health Organization. 2000;78:1389-400.
49. Pluess B, Tanser FC, Lengeler C, Sharp BL. Indoor residual spraying for preventing malaria. Cochrane database of systematic reviews. 2010(4).
50. Hemingway J, Ranson H, Magill A, Kolaczinski J, Fornadel C, Gimnig J, et al. Averting a malaria disaster: will insecticide resistance derail malaria control? The Lancet. 2016;387(10029):1785-8.
51. Sanou A, Nelli L, Guelbéogo WM, Cissé F, Tapsoba M, Ouédraogo P, et al. Insecticide resistance and behavioural adaptation as a response to long-lasting insecticidal net deployment in malaria vectors in the Cascades region of Burkina Faso. Scientific reports. 2021;11(1):1-14.
52. Kreppel K, Viana M, Main B, Johnson P, Govella N, Lee Y, et al. Emergence of behavioural avoidance strategies of malaria vectors in areas of high LLIN coverage in Tanzania. Scientific reports. 2020;10(1):1-11.
53. Sougoufara S, Harry M, Doucouré S, Sembène P, Sokhna C. Shift in species composition in the Anopheles gambiae complex after implementation of long-lasting insecticidal nets in Dielmo, Senegal. Medical and veterinary entomology. 2016;30(3):365-8.
54. Bousema T, Okell L, Felger I, Drakeley C. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. Nature Reviews Microbiology. 2014;12(12):833-40.
55. Molina-Cruz A, Garver LS, Alabaster A, Bangiolo L, Haile A, Winikor J, et al. The human malaria parasite Pfs47 gene mediates evasion of the mosquito immune system. Science. 2013;340(6135):984-7.
56. Schneider P, Bousema JT, Gouagna LC, Otieno S, Van De Vegte-Bolmer M, Omar SA, et al. Submicroscopic Plasmodium falciparum gametocyte densities frequently result in mosquito infection. The American journal of tropical medicine and hygiene. 2007;76(3):470-4.

57. Ouédraogo AL, Bousema T, de Vlas SJ, Cuzin-Ouattara N, Verhave J-P, Drakeley C, et al. The plasticity of *Plasmodium falciparum* gametocytaemia in relation to age in Burkina Faso. Malaria journal. 2010;9(1):1-8.
58. Reece SE, Drew DR, Gardner A. Sex ratio adjustment and kin discrimination in malaria parasites. Nature. 2008;453(7195):609-14.
59. Paul RE, Brey PT, Robert V. Plasmodium sex determination and transmission to mosquitoes. Trends in parasitology. 2002;18(1):32-8.
60. Stone WJ, Eldering M, van Gemert G-J, Lanke KH, Grignard L, van de Vegte-Bolmer MG, et al. The relevance and applicability of oocyst prevalence as a read-out for mosquito feeding assays. Scientific reports. 2013;3(1):1-8.
61. Slater HC, Ross A, Ouédraogo AL, White LJ, Nguon C, Walker PG, et al. Assessing the impact of next-generation rapid diagnostic tests on *Plasmodium falciparum* malaria elimination strategies. Nature. 2015;528(7580):S94-S101.
62. Auburn S, Barry AE. Dissecting malaria biology and epidemiology using population genetics and genomics. International Journal for Parasitology. 2017;47(2-3):77-85.
63. Auburn S, Campino S, Miotto O, Djimde AA, Zongo I, Manske M, et al. Characterization of within-host *Plasmodium falciparum* diversity using next-generation sequence data. PloS one. 2012;7(2):e32891.
64. Escalante AA, Ferreira MU, Vinetz JM, Volkman SK, Cui L, Gamboa D, et al. Malaria molecular epidemiology: Lessons from the International Centers of Excellence for Malaria Research Network. The American Journal of Tropical Medicine and Hygiene. 2015;93(3_Suppl):79-86.
65. Maharaj R, Seocharan I, Qwabe B, Mkhabela M, Kissoon S, Lakan V. Decadal epidemiology of malaria in KwaZulu-Natal, a province in South Africa targeting elimination. Malaria journal. 2019;18(1):1-7.
66. Malaria Elimination Strategic Plan for South Africa 2019-2023 [Internet]. South African Department of Health. 2019.
67. Nájera JA, González-Silva M, Alonso PL. Some lessons for the future from the Global Malaria Eradication Programme (1955–1969). PLoS medicine. 2011;8(1):e1000412.
68. Control mCGoV. A research agenda for malaria eradication: vector control. PLoS medicine. 2011;8(1):e1000401.
69. Alonso PL, Lindsay SW, Armstrong J, De Francisco A, Shenton F, Greenwood B, et al. The effect of insecticide-treated bed nets on mortality of Gambian children. The Lancet. 1991;337(8756):1499-502.
70. D'Alessandro U, Olaleye B, Langerock P, Aikins M, Thomson M, Cham M, et al. Mortality and morbidity from malaria in Gambian children after introduction of an impregnated bednet programme. The Lancet. 1995;345(8948):479-83.
71. Habluetzel A, Diallo D, Esposito F, Lamizana L, Pagnonl F, Lengeler C, et al. Do insecticide-treated curtains reduce all-cause child mortality in Burkina Faso? Tropical medicine & international health. 1997;2(9):855-62.
72. Nevill C, Some E, Mung'Ala V, Muterni W, New L, Marsh K, et al. Insecticide-treated bednets reduce mortality and severe morbidity from malaria among children on the Kenyan coast. Tropical medicine & international health. 1996;1(2):139-46.
73. Bhatt S, Weiss D, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 2015;526(7572):207-11.
74. Noor AM, Kinyoki DK, Mundia CW, Kabaria CW, Mutua JW, Alegana VA, et al. The changing risk of Plasmodium falciparum malaria infection in Africa: 2000–10: a spatial and temporal analysis of transmission intensity. The Lancet. 2014;383(9930):1739-47.

75. Noor AM, Mutheu JJ, Tatem AJ, Hay SI, Snow RW. Insecticide-treated net coverage in Africa: mapping progress in 2000–07. The Lancet. 2009;373(9657):58-67.
76. WHO. WHO calls for reinvigorated action to fight malaria 2020 [Available from: https://www.who.int/news/item/30-11-2020-who-calls-for-reinvigorated-action-to-fight-malaria.
77. WHO. From 30 million cases to zero: China is certified malaria-free by WHO. 2021.
78. Tusting LS, Bousema T, Smith DL, Drakeley C. Measuring changes in Plasmodium falciparum transmission: precision, accuracy and costs of metrics. Advances in parasitology. 2014;84:151-208.
79. Drakeley C, Schellenberg D, Kihonda J, Sousa C, Arez A, Lopes D, et al. An estimation of the entomological inoculation rate for Ifakara: a semi-urban area in a region of intense malaria transmission in Tanzania. Tropical Medicine & International Health. 2003;8(9):767-74.
80. Smith DL, Dushoff J, McKenzie FE, Dobson AP. The risk of a mosquito-borne infectionin a heterogeneous environment. PLoS biology. 2004;2(11):e368.
81. Packard RM. The origins of antimalarial-drug resistance. New England Journal of Medicine. 2014;371(5):397-9.
82. Cui L, Su X-z. Discovery, mechanisms of action and combination therapy of artemisinin. Expert review of anti-infective therapy. 2009;7(8):999-1013.
83. Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, et al. Artemisinin resistance in Plasmodium falciparum malaria. New England journal of medicine. 2009;361(5):455-67.
84. Asua V, Conrad MD, Aydemir O, Duvalsaint M, Legac J, Duarte E, et al. Changing prevalence of potential mediators of aminoquinoline, antifolate, and artemisinin resistance across Uganda. The Journal of Infectious Diseases. 2021;223(6):985-94.
85. Uwimana A, Umulisa N, Venkatesan M, Svigel SS, Zhou Z, Munyaneza T, et al. Association of Plasmodium falciparum kelch13 R561H genotypes with delayed parasite clearance in Rwanda: an open-label, single-arm, multicentre, therapeutic efficacy study. The Lancet Infectious Diseases. 2021.
86. Uwimana A, Legrand E, Stokes BH, Ndikumana J-LM, Warsame M, Umulisa N, et al. Emergence and clonal expansion of in vitro artemisinin-resistant Plasmodium falciparum kelch13 R561H mutant parasites in Rwanda. Nature medicine. 2020;26(10):1602-8.
87. Hill AV. Vaccines against malaria. Philosophical Transactions of the Royal Society B: Biological Sciences. 2011;366(1579):2806-14.
88. Thera MA, Doumbo OK, Coulibaly D, Laurens MB, Ouattara A, Kone AK, et al. A field trial to assess a blood-stage malaria vaccine. New England Journal of Medicine. 2011;365(11):1004-13.
89. Rts S. 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. The Lancet. 2015;386(9988):31-45.
90. Olotu A, Fegan G, Wambua J, Nyangweso G, Leach A, Lievens M, et al. Seven-year efficacy of RTS, S/AS01 malaria vaccine among young African children. New England Journal of Medicine. 2016;374(26):2519-29.
91. Vogel G. WHO gives first malaria vaccine the green light. American Association for the Advancement of Science; 2021.
92. WHO. RTS,S malaria vaccine reaches more than 650 000 children in Ghana, Kenya and Malawi through groundbreaking pilot programme. 2021.

93. Chandramohan D, Zongo I, Sagara I, Cairns M, Yerbanga R-S, Diarra M, et al. Seasonal malaria vaccination with or without seasonal malaria chemoprevention. New England Journal of Medicine. 2021;385(11):1005-17.
94. Bhattarai A, Ali AS, Kachur SP, Mårtensson A, Abbas AK, Khatib R, et al. Impact of artemisinin-based combination therapy and insecticide-treated nets on malaria burden in Zanzibar. PLoS medicine. 2007;4(11):e309.
95. Fegan GW, Noor AM, Akhwale WS, Cousens S, Snow RW. Effect of expanded insecticide-treated bednet coverage on child survival in rural Kenya: a longitudinal study. The Lancet. 2007;370(9592):1035-9.
96. Gething PW, Casey DC, Weiss DJ, Bisanzio D, Bhatt S, Cameron E, et al. Mapping *Plasmodium falciparum* mortality in Africa between 1990 and 2015. New England Journal of Medicine. 2016;375(25):2435-45.
97. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, et al. A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS medicine. 2009;6(3):e1000048.
98. Daniels RF, Schaffner SF, Wenger EA, Proctor JL, Chang H-H, Wong W, et al. Modeling malaria genomics reveals transmission decline and rebound in Senegal. Proceedings of the National Academy of Sciences. 2015;112(22):7067-72.
99. Yukich J, Briët O, Bretscher MT, Bennett A, Lemma S, Berhane Y, et al. Estimating *Plasmodium falciparum* transmission rates in low-endemic settings using a combination of community prevalence and health facility data. 2012.
100. Cibulskis RE, Aregawi M, Williams R, Otten M, Dye C. Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods. PLoS medicine. 2011;8(12):e1001142.
101. Corran P, Coleman P, Riley E, Drakeley C. Serology: a robust indicator of malaria transmission intensity? Trends in parasitology. 2007;23(12):575-82.
102. Carter R, Mendis KN, Roberts D. Spatial targeting of interventions against malaria. Bulletin of the World Health Organization. 2000;78:1401-11.
103. WHO. Malaria elimination: a field manual for low and moderate endemic countries. Malaria elimination: a field manual for low and moderate endemic countries. World Health Organisation: 2007.
104. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. PLoS medicine. 2012;9(1):e1001165.
105. Bousema T, Stevenson J, Baidjoe A, Stresman G, Griffin JT, Kleinschmidt I, et al. The impact of hotspot-targeted interventions on malaria transmission: study protocol for a cluster-randomized controlled trial. Trials. 2013;14(1):1-12.
106. Bousema T, Stresman G, Baidjoe AY, Bradley J, Knight P, Stone W, et al. The impact of hotspot-targeted interventions on malaria transmission in Rachuonyo South District in the Western Kenyan Highlands: a cluster-randomized controlled trial. PLoS medicine. 2016;13(4):e1001993.
107. Bejon P, Williams TN, Nyundo C, Hay SI, Benz D, Gething PW, et al. A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots. elife. 2014;3:e02130.
108. Crowell V, Yukich JO, Briët OJ, Ross A, Smith TA. A novel approach for measuring the burden of uncomplicated *Plasmodium falciparum* malaria: application to data from Zambia. PLoS One. 2013;8(2):e57297.
109. Gamboa D, Ho M-F, Bendezu J, Torres K, Chiodini PL, Barnwell JW, et al. A large proportion of *P. falciparum* isolates in the Amazon region of Peru lack pfhrp2 and pfhrp3: implications for malaria rapid diagnostic tests. PloS one. 2010;5(1):e8091.

110. Koita OA, Doumbo OK, Ouattara A, Tall LK, Konaré A, Diakité M, et al. False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. The American journal of tropical medicine and hygiene. 2012;86(2):194.
111. Nderu D, Kimani F, Thiong'o K, Karanja E, Akinyi M, Too E, et al. *Plasmodium falciparum* histidine-rich protein (*Pf*HRP2 and 3) diversity in Western and Coastal Kenya. Scientific reports. 2019;9(1):1-9.
112. McCreesh P, Mumbengegwi D, Roberts K, Tambo M, Smith J, Whittemore B, et al. Subpatent malaria in a low transmission African setting: a cross-sectional study using rapid diagnostic testing (RDT) and loop-mediated isothermal amplification (LAMP) from Zambezi region, Namibia. Malaria journal. 2018;17(1):1-11.
113. Tambo M, Mwinga M, Mumbengegwi DR. Loop-mediated isothermal amplification (LAMP) and Polymerase Chain Reaction (PCR) as quality assurance tools for Rapid Diagnostic Test (RDT) malaria diagnosis in Northern Namibia. PloS one. 2018;13(12):e0206848.
114. Salminen T, Mehdi F, Rohila D, Kumar M, Talha SM, Prakash JAJ, et al. Ultrasensitive and robust point-of-care immunoassay for the detection of plasmodium falciparum malaria. Analytical Chemistry. 2020;92(24):15766-72.
115. Eisele TP. Mass drug administration can be a valuable addition to the malaria elimination toolbox. Malaria journal. 2019;18(1):1-5.
116. Shah MP, Hwang J, Choi L, Lindblade KA, Kachur SP, Desai M. Mass drug administration for malaria. Cochrane Database of Systematic Reviews. 2021;9:CD008846.
117. WHO. Meeting report of the WHO Evidence Review Group on mass drug administration for malaria: 11–13, Geneva, Switzerland. World Health Organization Geneva; 2018.
118. Neafsey DE, Taylor AR, MacInnis BL. Advances and opportunities in malaria population genomics. Nature Reviews Genetics. 2021:1-16.
119. Tessema SK, Raman J, Duffy CW, Ishengoma DS, Amambua-Ngwa A, Greenhouse B. Applying next-generation sequencing to track falciparum malaria in sub-Saharan Africa. Malaria journal. 2019;18(1):268.
120. Volkman SK, Neafsey DE, Schaffner SF, Park DJ, Wirth DF. Harnessing genomics and genome biology to understand malaria biology. Nature Reviews Genetics. 2012;13(5):315.
121. Cotter C, Sturrock HJ, Hsiang MS, Liu J, Phillips AA, Hwang J, et al. The changing epidemiology of malaria elimination: new strategies for new challenges. The Lancet. 2013;382(9895):900-11.
122. Tessema S, Wesolowski A, Chen A, Murphy M, Wilheim J, Mupiri A-R, et al. Using parasite genetic and human mobility data to infer local and cross-border malaria connectivity in Southern Africa. eLife. 2019;8:e43510.
123. Wesolowski A, Taylor AR, Chang H-H, Verity R, Tessema S, Bailey JA, et al. Mapping malaria by combining parasite genomic and epidemiologic data. BMC medicine. 2018;16(1):190.
124. Daniels R, Chang H-H, Séne PD, Park DC, Neafsey DE, Schaffner SF, et al. Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. PloS One. 2013;8(4).
125. WHO. Global technical strategy for malaria 2016-2030: World Health Organization; 2015.
126. WHO. Technical consultation on the role of parasite and anopheline genetics in malaria surveillance. World Health Organisation; 2019.

127. Inzaule SC, Tessema SK, Kebede Y, Ouma AEO, Nkengasong JN. Genomic-informed pathogen surveillance in Africa: opportunities and challenges. The Lancet Infectious Diseases. 2021.
128. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance. 2020;25(3):2000045.
129. Vogels CB, Brito AF, Wyllie AL, Fauver JR, Ott IM, Kalinich CC, et al. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. Nature microbiology. 2020;5(10):1299-305.
130. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell. 2020;182(4):812-27. e19.
131. Young BE, Fong S-W, Chan Y-H, Mak T-M, Ang LW, Anderson DE, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. The Lancet. 2020;396(10251):603-11.
132. Giandhari J, Pillay S, Wilkinson E, Tegally H, Sinayskiy I, Schuld M, et al. Early transmission of SARS-CoV-2 in South Africa: An epidemiological and phylogenetic report. International Journal of Infectious Diseases. 2021;103:234-41.
133. Meredith LW, Hamilton WL, Warne B, Houldcroft CJ, Hosmillo M, Jahun AS, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. The Lancet infectious diseases. 2020;20(11):1263-72.
134. Hussein MIH, Albashir AAD, Elawad OAMA, Homeida A. Malaria and COVID-19: Unmasking their ties. Malaria Journal. 2020;19(1):1-10.
135. WHO. Coronavirus disease (COVID-19). World Health Organisation; 2020.
136. M Chisenga JTG, J Martins, Q Dlamini, B Candrinho, P Uusiku, P Moonasar, I Kleinschmidt, C Sikaala, J Chimumbwa The Impact of COVID-19 on Malaria in the E8 region. South Africa Malaria Research Conference2021.
137. Stiegler N, Bouchard J-P, editors. South Africa: Challenges and successes of the COVID-19 lockdown. Annales Médico-psychologiques, revue psychiatrique; 2020: Elsevier.
138. Bango ZA, Tawe L, Muthoga CW, Paganotti GM. ast and current biological factors affecting malaria in the low transmission setting of Botswana: A review. Infection, Genetics and Evolution. 2020:104458.
139. Ogunleye OO, Basu D, Mueller D, Sneddon J, Seaton RA, Yinka-Ogunleye AF, et al. Response to the novel corona virus (COVID-19) pandemic across Africa: successes, challenges, and implications for the future. Frontiers in pharmacology. 2020;11:1205.
140. Raman J, Fakudze P, Sikaala CH, Chimumbwa J, Moonasar D. Eliminating malaria from the margins of transmission in Southern Africa through the Elimination 8 Initiative. Transactions of the Royal Society of South Africa. 2021:1-9.
141. Gavi S, Tapera O, Mberikunashe J, Kanyangarara M. Malaria incidence and mortality in Zimbabwe during the COVID-19 pandemic: analysis of routine surveillance data. Malaria journal. 2021;20(1):1-9.
142. WHO. Tracking SARS-CoV-2 variants: World Health Organisation; [Available from: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/.
143. Maharaj R, Raman J, Morris N, Moonasar D, Durrheim D, Seocharan I, et al. Epidemiology of malaria in South Africa: From control to elimination. South African Medical Journal. 2013;103(10):779-83.
144. Amorim A. Population Genetics: Brenner's Encyclopedia of Genetics (Second Edition), Academic Press; 2013.

145. Anderson TJ, Haubold B, Williams JT, Estrada-Franco § JG, Richardson L, Mollinedo R, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Molecular Biology and Evolution. 2000;17(10):1467-82.
146. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, Park DJ, et al. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. Malaria Journal. 2008;7(1):223.
147. Ishengoma DS, Saidi Q, Sibley CH, Roper C, Alifrangis M. Deployment and utilization of next-generation sequencing of *Plasmodium falciparum* to guide anti-malarial drug policy decisions in sub-Saharan Africa: opportunities and challenges. Malaria journal. 2019;18(1):267.
148. Koepfli C, Mueller I. Malaria epidemiology at the clone level. Trends in parasitology. 2017;33(12):974-85.
149. Mobegi VA, Duffy CW, Amambua-Ngwa A, Loua KM, Laman E, Nwakanma DC, et al. Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. Molecular Biology and Evolution. 2014;31(6):1490-9.
150. Patel JC, Taylor SM, Juliao PC, Parobek CM, Janko M, Gonzalez LD, et al. Genetic evidence of importation of drug-resistant *Plasmodium falciparum* to Guatemala from the Democratic Republic of the Congo. Emerging Infectious Diseases. 2014;20(6):932.
151. Salgueiro P, Vicente JL, Figueiredo RC, Pinto J. Genetic diversity and population structure of *Plasmodium falciparum* over space and time in an African archipelago. Infection, Genetics and Evolution. 2016;43:252-60.
152. Carter TE, Malloy H, Existe A, Memnon G, Victor YS, Okech BA, et al. Genetic diversity of *Plasmodium falciparum* in Haiti: Insights from microsatellite markers. PloS One. 2015;10(10).
153. Razak MRMA, Sastu UR, Norahmad NA, Abdul-Karim A, Muhammad A, Muniandy PK, et al. Genetic diversity of *Plasmodium falciparum* populations in malaria declining areas of Sabah, East Malaysia. PloS One. 2016;11(3).
154. Bei AK, Niang M, Deme AB, Daniels RF, Sarr FD, Sokhna C, et al. Dramatic changes in malaria population genetic complexity in Dielmo and Ndiop, Senegal, revealed using genomic surveillance. The Journal of Infectious Diseases. 2018;217(4):622-7.
155. Fola AA, Harrison GA, Hazairin MH, Barnadas C, Hetzel MW, Iga J, et al. Higher complexity of infection and genetic diversity of *Plasmodium vivax* than *Plasmodium falciparum* across all malaria transmission zones of Papua New Guinea. The American journal of tropical medicine and hygiene. 2017;96(3):630-41.
156. Redmond SN, MacInnis BM, Bopp S, Bei AK, Ndiaye D, Hartl DL, et al. De novo mutations resolve disease transmission pathways in clonal malaria. Molecular biology and evolution. 2018;35(7):1678-89.
157. Waltmann A, Koepfli C, Tessier N, Karl S, Fola A, Darcy AW, et al. Increasingly inbred and fragmented populations of *Plasmodium vivax* associated with the eastward decline in malaria transmission across the Southwest Pacific. PLoS neglected tropical diseases. 2018;12(1):e0006146.
158. Nkhoma SC, Trevino SG, Gorena KM, Nair S, Khoswe S, Jett C, et al. Resolving within-host malaria parasite diversity using single-cell sequencing. bioRxiv. 2018:391268.
159. Amambua-Ngwa A, Jeffries D, Amato R, Worwui A, Karim M, Ceesay S, et al. Consistent signatures of selection from genomic analysis of pairs of temporal and spatial *Plasmodium falciparum* populations from The Gambia. Scientific Reports. 2018;8(1):9687.
160. Dalmat R, Naughton B, Kwan-Gett TS, Slyker J, Stuckey EM. Use cases for genetic epidemiology in malaria elimination. Malaria journal. 2019;18(1):1-11.

161. Nkhoma SC, Trevino SG, Gorena KM, Nair S, Khoswe S, Jett C, et al. Co-transmission of Related Malaria Parasite Lineages Shapes Within-Host Parasite Diversity. Cell Host & Microbe. 2020;27(1):93-103. e4.
162. Nkhoma SC, Nair S, Al-Saai S, Ashley E, McGready R, Phyo AP, et al. Population genetic correlates of declining transmission in a human pathogen. Molecular ecology. 2013;22(2):273-85.
163. Sisya TJ, Kamn'gona RM, Vareta JA, Fulakeza JM, Mukaka MFJ, Seydel KB, et al. Subtle changes in *Plasmodium falciparum* infection complexity following enhanced intervention in Malawi. Acta Tropica. 2015;142:108-14.
164. Echeverry DF, Nair S, Osorio L, Menon S, Murillo C, Anderson TJ. Long term persistence of clonal malaria parasite *Plasmodium falciparum* lineages in the Colombian Pacific region. BMC Genetics. 2013;14(1):2.
165. Volkman SK, Neafsey DE, Schaffner SF, Park DJ, Wirth DF. Harnessing genomics and genome biology to understand malaria biology. Nature Reviews Genetics. 2012;13(5):315-28.
166. Zhong D, Afrane Y, Githeko A, Yang Z, Cui L, Menge DM, et al. Plasmodium falciparum genetic diversity in western Kenya highlands. The American journal of tropical medicine and hygiene. 2007;77(6):1043-50.
167. Daniels R, Chang H-H, Séne PD, Park DC, Neafsey DE, Schaffner SF, et al. Genetic Surveillance Detects Both Clonal and Epidemic Transmission of Malaria following Enhanced Intervention in Senegal. PLOS ONE. 2013;8(4):e60780.
168. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature. 2002;419(6906):498-511.
169. Conway DJ. Molecular epidemiology of malaria. Clinical Microbiology Reviews. 2007;20(1):188-204.
170. MalariaGEN AA, Ali M, Almagro-Garcia J, Amambua-Ngwa A, Amaratunga C, Amato R, et al. An open dataset of *Plasmodium falciparum* genome variation in 7,000 worldwide samples. Wellcome Open Research. 2021;6.
171. Jiang H, Li N, Gopalan V, Zilversmit MM, Varma S, Nagarajan V, et al. High recombination rates and hotspots in a *Plasmodium falciparum* genetic cross. Genome biology. 2011;12(4):1-15.
172. Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, Waisberg M, et al. *Plasmodium falciparum* genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. Nature genetics. 2010;42(3):268-71.
173. Mu J, Awadalla P, Duan J, McGee KM, Joy DA, McVean GAT, et al. Recombination hotspots and population structure in *Plasmodium falciparum*. PLoS biology. 2005;3(10):e335.
174. Cheeseman IH, Gomez-Escobar N, Carret CK, Ivens A, Stewart LB, Tetteh KK, et al. Gene copy number variation throughout the *Plasmodium falciparum* genome. BMC genomics. 2009;10(1):1-11.
175. Campino S, Auburn S, Kivinen K, Zongo I, Ouedraogo J-B, Mangano V, et al. Population genetic analysis of *Plasmodium falciparum* parasites using a customized Illumina GoldenGate genotyping assay. PloS one. 2011;6(6):e20251.
176. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, Park DJ, et al. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. Malaria journal. 2008;7(1):1-11.
177. Lee Y, Marsden CD, Nieman C, Lanzaro GC. A new multiplex SNP genotyping assay for detecting hybridization and introgression between the M and S molecular forms of *Anopheles gambia*e. Molecular ecology resources. 2014;14(2):297-305.

178. Baniecki ML, Faust AL, Schaffner SF, Park DJ, Galinsky K, Daniels RF, et al. Development of a single nucleotide polymorphism barcode to genotype *Plasmodium vivax* infections. PLoS neglected tropical diseases. 2015;9(3):e0003539.
179. Neafsey D, Lawniczak M, Park D, Redmond S, Coulibaly M, Traoré S, et al. SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. Science. 2010;330(6003):514-7.
180. Van Tyne D, Park DJ, Schaffner SF, Neafsey DE, Angelino E, Cortese JF, et al. Identification and functional validation of the novel antimalarial resistance locus PF10_0355 in *Plasmodium falciparum*. PLoS genetics. 2011;7(4):e1001383.
181. Bailey JA, Mvalo T, Aragam N, Weiser M, Congdon S, Kamwendo D, et al. Use of massively parallel pyrosequencing to evaluate the diversity of and selection on *Plasmodium falciparum* csp T-cell epitopes in Lilongwe, Malawi. The Journal of infectious diseases. 2012;206(4):580-7.
182. Aragam NR, Thayer KM, Nge N, Hoffman I, Martinson F, Kamwendo D, et al. Diversity of T cell epitopes in *Plasmodium falciparum* circumsporozoite protein likely due to protein-protein interactions. PLoS One. 2013;8(5):e62427.
183. Juliano JJ, Porter K, Mwapasa V, Sem R, Rogers WO, Ariey F, et al. Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. Proceedings of the National Academy of Sciences. 2010;107(46):20138-43.
184. Gandhi K, Thera MA, Coulibaly D, Traore K, Guindo AB, Ouattara A, et al. Variation in the circumsporozoite protein of *Plasmodium falciparum*: vaccine development implications. PLoS One. 2014;9(7):e101783.
185. Early AM, Lievens M, MacInnis BL, Ockenhouse CF, Volkman SK, Adjei S, et al. Host-mediated selection impacts the diversity of *Plasmodium falciparum* antigens within infections. Nature communications. 2018;9(1):1-10.
186. Tessema SK, Hathaway NJ, Teyssier NB, Murphy M, Chen A, Aydemir O, et al. Sensitive, highly multiplexed sequencing of microhaplotypes from the *Plasmodium falciparum* heterozygome. J infect dis. 2020.
187. Aydemir O, Janko M, Hathaway NJ, Verity R, Mwandagalirwa MK, Tshefu AK, et al. Drug-resistance and population structure of *Plasmodium falciparum* across the Democratic Republic of Congo using high-throughput molecular inversion probes. The Journal of infectious diseases. 2018;218(6):946-55.
188. Roh M, Tessema S, Murphy M, Nhlabathi N, Mkhonta N, Vilakati S, et al. High genetic diversity of *Plasmodium falciparum* in the low transmission setting of the Kingdom of Eswatini. The Journal of Infectious Diseases. 2019.
189. Liu Y, Tessema SK, Murphy M, Xu S, Schwartz A, Wang W, et al. Confirmation of the absence of local transmission and geographic assignment of imported falciparum malaria cases to China using microsatellite panel. Malaria Journal 2020;19:244.
190. Reeder JC, Marshall VM. A simple method for typing Trypanosoma falciparum merozoite surface antigens 1 and 2 (MSA-1 and MSA-2) using a dimorphic-form specific polymerase chain reaction. Molecular and biochemical parasitology. 1994;68(2):329-32.
191. Snounou G. Genotyping of Plasmodium spp. Malaria Methods and Protocols: Springer; 2002. p. 103-16.
192. Robert F, Ntoumi F, Angel G, Candito D, Rogier C, Fandeur T, et al. Extensive genetic diversity of *Plasmodium falciparum* isolates collected from patients with severe malaria in Dakar, Senegal. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1996;90(6):704-11.
193. Coppi A, Natarajan R, Pradel G, Bennett BL, James ER, Roggero MA, et al. The malaria circumsporozoite protein has two functional domains, each with distinct roles as sporozoites journey from mosquito to mammalian host. J Exp Med. 2011;208(2):341-56.

194. Zhong D, Koepfli C, Cui L, Yan G. Molecular approaches to determine the multiplicity of *Plasmodium* infections. Malaria Journal. 2018;17(1):172.
195. Kolakovich KA, Ssengoba A, Wojcik K, Tsuboi T, Al-Yaman F, Alpers M, et al. *Plasmodium vivax*: favored gene frequencies of the merozoite surface protein-1 and the multiplicity of infection in a malaria endemic region. Experimental parasitology. 1996;83(1):11-8.
196. Greenhouse B, Myrick A, Dokomajilar C, Woo JM, Carlson EJ, Rosenthal PJ, et al. Validation of microsatellite markers for use in genotyping polyclonal *Plasmodium falciparum* infections. The American journal of Tropical Medicine and Hygiene. 2006;75(5):836-42.
197. Preston MD, Campino S, Assefa SA, Echeverry DF, Ocholla H, Amambua-Ngwa A, et al. A barcode of organellar genome polymorphisms identifies the geographic origin of *Plasmodium falciparum* strains. Nature Communications. 2014;5:4052.
198. Baniecki ML, Faust AL, Schaffner SF, Park DJ, Galinsky K, Daniels RF, et al. Development of a Single Nucleotide Polymorphism Barcode to Genotype *Plasmodium vivax* Infections. PLoS Neglected Tropical Diseases. 2015;9(3):e0003539.
199. Carlton JM, Volkman SK, Uplekar S, Hupalo DN, Pereira Alves JM, Cui L, et al. Population Genetics, Evolutionary Genomics, and Genome-Wide Studies of Malaria: A View Across the International Centers of Excellence for Malaria Research. The American journal of tropical medicine and hygiene. 2015;93(3 Suppl):87-98.
200. Obaldia N, 3rd, Baro NK, Calzada JE, Santamaria AM, Daniels R, Wong W, et al. Clonal outbreak of *Plasmodium falciparum* infection in eastern Panama. The Journal of Infectious Diseases. 2015;211(7):1087-96.
201. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, Park DJ, et al. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. Malaria Journal. 2008;7(1):223.
202. Messerli C, Hofmann NE, Beck H-P, Felger I. Critical Evaluation of Molecular Monitoring in Malaria Drug Efficacy Trials and Pitfalls of Length-Polymorphic Markers. Antimicrobial Agents and Chemotherapy. 2016;61(1):e01500-16.
203. Benavente ED, Ward Z, Chan W, Mohareb FR, Sutherland CJ, Roper C, et al. Genomic variation in *Plasmodium vivax* malaria reveals regions under selective pressure. PloS ONE. 2017;12(5):e0177134.
204. Tripathi M, Das A. Genotyping malaria parasites with DNA barcodes. 2015.
205. Tripathi M, Das A. Genotyping malaria parasites with DNA barcodes. Tropical Medicine & International Health. 2015;20(12):1636-8.
206. Preston MD, Campino S, Assefa SA, Echeverry DF, Ocholla H, Amambua-Ngwa A, et al. A barcode of organellar genome polymorphisms identifies the geographic origin of *Plasmodium falciparum* strains. Nature communications. 2014;5(1):1-7.
207. Apinjoh TO, Ouattara A, Titanji VP, Djimde A, Amambua-Ngwa A. Genetic diversity and drug resistance surveillance of *Plasmodium falciparum* for malaria elimination: is there an ideal tool for resource-limited sub-Saharan Africa? Malaria journal. 2019;18(1):1-12.
208. Kidd K, Pakstis A, Speed W, Lagace R, Chang J, Wootton S, et al. Microhaplotype loci are a powerful new type of forensic marker. Forensic Science International: Genetics Supplement Series. 2013;4(1):e123-e4.
209. Oldoni F, Kidd KK, Podini D. Microhaplotypes in forensic genetics. Forensic Science International: Genetics. 2019;38:54-69.
210. Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, et al. Genetic diversity and protective efficacy of the RTS, S/AS01 malaria vaccine. New England Journal of Medicine. 2015;373(21):2025-37.

211. Rao PN, Uplekar S, Kayal S, Mallick PK, Bandyopadhyay N, Kale S, et al. A method for amplicon deep sequencing of drug resistance genes in *Plasmodium falciparum* clinical isolates from India. Journal of clinical microbiology. 2016;54(6):1500-11.
212. Ngondi JM, Ishengoma DS, Doctor SM, Thwai KL, Keeler C, Mkude S, et al. Surveillance for sulfadoxine-pyrimethamine resistant malaria parasites in the Lake and Southern Zones, Tanzania, using pooling and next-generation sequencing. Malaria journal. 2017;16(1):1-7.
213. Greenhouse B, Myrick A, Dokomajilar C, Woo JM, Carlson EJ, Rosenthal PJ, et al. Validation of microsatellite markers for use in genotyping polyclonal *Plasmodium falciparum* infections. The American journal of tropical medicine and hygiene. 2006;75(5):836.
214. Roper C, Pearce R, Nair S, Sharp B, Nosten F, Anderson T. Intercontinental spread of pyrimethamine-resistant malaria. Science. 2004;305(5687):1124-.
215. Imwong M, Suwannasin K, Kunasol C, Sutawong K, Mayxay M, Rekol H, et al. The spread of artemisinin-resistant *Plasmodium falciparum* in the Greater Mekong subregion: a molecular epidemiology observational study. The Lancet Infectious Diseases. 2017;17(5):491-7.
216. Imwong M, Hien TT, Thuy-Nhien NT, Dondorp AM, White NJ. Spread of a single multidrug resistant malaria parasite lineage (PfPailin) to Vietnam. The Lancet Infectious Diseases. 2017;17(10):1022-3.
217. Su X-z, Wellems TE. Toward a High-Resolution *Plasmodium falciparum* Linkage Map: Polymorphic Markers from Hundreds of Simple Sequence Repeats. Genomics. 1996;33(3):430-44.
218. Tessema S, Wesolowski A, Chen A, Murphy M, Wilheim J, Mupiri A-R, et al. Using parasite genetic and human mobility data to infer local and cross-border malaria connectivity in Southern Africa. Elife. 2019;8:e43510.
219. Barker GC. Microsatellite DNA: a tool for population genetic analysis. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2002;96:S21-S4.
220. Schlötterer C. Evolutionary dynamics of microsatellite DNA. Chromosoma. 2000;109(6):365-71.
221. Wu C-H, Drummond AJ. Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov Chain Monte Carlo. Genetics. 2011;188(1):151-64.
222. Neafsey DE, Volkman SK. Malaria Genomics in the Era of Eradication. Cold Spring Harbor Perspectives in Medicine. 2017:a025544.
223. Chenet SM, Schneider KA, Villegas L, Escalante AA. Local population structure of Plasmodium: impact on malaria control and elimination. Malaria journal. 2012;11(1):412.
224. Anderson T, SU X-Z, Bockarie M, Lagog M, Day K. Twelve microsatellite markers for characterization of *Plasmodium falciparum* from finger-prick blood samples. Parasitology. 1999;119(2):113-25.
225. Liu Y, Tessema SK, Murphy M, Xu S, Schwartz A, Wang W, et al. Confirmation of the absence of local transmission and geographic assignment of imported falciparum malaria cases to China using microsatellite panel. Malaria Journal. 2020;19(1):244.
226. Roh ME, Tessema SK, Murphy M, Nhlabathi N, Mkhonta N, Vilakati S, et al. High genetic diversity of *Plasmodium falciparum* in the low transmission setting of the Kingdom of Eswatini. bioRxiv. 2019:522896.
227. Pringle JC, Tessema S, Wesolowski A, Chen A, Murphy M, Carpi G, et al. Genetic evidence of focal *Plasmodium falciparum* transmission in a pre-elimination setting in Southern Province, Zambia. The Journal of infectious diseases. 2019;219(8):1254-63.
228. Organization WH. World malaria report 2021. 2021.
229. World malaria report 2019 [Internet]. World Health Organization. 2019.

230. Gueye CS, 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. Malaria journal. 2012;11(1):1-10.
231. Wangdi K, Gatton ML, Kelly GC, Clements AC. Cross-border malaria: a major obstacle for malaria elimination. Advances in parasitology. 2015;89:79-107.
232. Ruktanonchai NW. Human Migration Movement among E8 Countries Malaria Atlas Project. University of Southampton. 2014.
233. Raman J, Morris N, Frean J, Brooke B, Blumberg L, Kruger P, et al. Reviewing South Africa's malaria elimination strategy (2012–2018): Progress, challenges and priorities. Malaria Journal. 2016;15(1):438.
234. Adeola A, Ncongwane K, Abiodun G, Makgoale T, Rautenbach H, Botai J, et al. Rainfall trends and malaria occurrences in Limpopo Province, South Africa. International Journal of Environmental Research and Public Health. 2019;16(24):5156.
235. Brooke B, Koekemoer I, Kruger P, Urbach J, Misiani E, Coetzee M. Malaria vector control in South Africa. South African Medical Journal. 2013;103(10):784-8.
236. Christian R, Dahan-Moss Y, Munhenga G, Lobb L, Erlank E, Dandalo L, et al. Malaria Vector Surveillance Report, South Africa, January–December, 2017. National Institute for Communicable Diseases-Bulletin, Johannesburg, South Africa. 2016.
237. Daygena TY, Massebo F, Lindtjørn B. Variation in species composition and infection rates of Anopheles mosquitoes at different altitudinal transects, and the risk of malaria in the highland of Dirashe Woreda, south Ethiopia. Parasites & vectors. 2017;10(1):343.
238. Kyalo D, Amratia P, Mundia CW, Mbogo CM, Coetzee M, Snow RW. A geo-coded inventory of anophelines in the Afrotropical Region south of the Sahara: 1898-2016. Wellcome open research. 2017;2.
239. Ogola EO, Fillinger U, Ondiba IM, Villinger J, Masiga DK, Torto B, et al. Insights into malaria transmission among *Anopheles funestus* mosquitoes, Kenya. Parasites & vectors. 2018;11(1):577.
240. Samuel M, Qwabe B, Dlamini D, Mabaso N, Manyawo Z, Zhikali J, et al. Malaria Vector Surveillance Report, South Africa, January–December 2018. 2018.
241. Eric M. Effects of targeted testing and treatment of malaria in moderate transmission regions in South Africa: Southern Africa Malaria Research Conference Research and Control: Untited Against Malaria; 2021 [Available from: http://malariaconference.mrc.ac.za/AbstractBook.pdf.
242. DOH. National guidelines for the treatment of malaria, South Africa 2019: South Africa Department of Health; 2019 [Available from: https://www.health.gov.za/wp-content/uploads/2020/11/national-guidelines-for-the-treatment-of-malaria-south-africa-2019.pdf.
243. Barnes KI, Durrheim DN, Little F, Jackson A, Mehta U, Allen E, et al. Effect of artemether-lumefantrine policy and improved vector control on malaria burden in KwaZulu–Natal, South Africa. PLoS Med. 2005;2(11):e330.
244. Bredenkamp B, Sharp BL, Mthembu SD, Durrheim DN, Barnes KI. Failure of sulphadoxine-pyrimethamine in treating *Plasmodium falciparum* malaria in KwaZulu-Natal. South African medical journal= Suid-Afrikaanse tydskrif vir geneeskunde. 2001;91(11):970, 2-, 2.
245. Roper C, Pearce R, Bredenkamp B, Gumede J, Drakeley C, Mosha F, et al. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. The Lancet. 2003;361(9364):1174-81.
246. Hargreaves K, Koekemoer L, Brooke B, Hunt R, Mthembu J, Coetzee M. Anopheles funestus resistant to pyrethroid insecticides in South Africa. Medical and veterinary entomology. 2000;14(2):181-9.

247. Maharaj R, Mthembu D, Sharp B. Impact of DDT re-introduction on malaria transmission in KwaZulu-Natal. South African Medical Journal. 2005;95(11):871-4.
248. Ukpe IS, Moonasar D, Raman J, Barnes K, Baker L, Blumberg L. Case management of malaria: treatment and chemoprophylaxis. South African Medical Journal. 2013;103(10):793-8.
249. Brooke B, Raman J, Frean J, Rundle K, Maartens F, Misiani E, et al. Implementing malaria control in South Africa, Eswatini and southern Mozambique during the COVID-19 pandemic. SAMJ: South African Medical Journal. 2020;110(11):1072-6.
250. Pringle JC, Wesolowski A, Berube S, Kobayashi T, Gebhardt ME, Mulenga M, et al. High *Plasmodium falciparum* genetic diversity and temporal stability despite control efforts in high transmission settings along the international border between Zambia and the Democratic Republic of the Congo. Malaria journal. 2019;18(1):1-13.
251. WHO. Global technical strategy for malaria 2016-2030 [Internet]. World Health Organization; 2015.
252. Moonasar D, Maharaj R, Kunene S, Candrinho B, Saute F, Ntshalintshali N, et al. Towards malaria elimination in the MOSASWA (Mozambique, South Africa and Swaziland) region. Malaria Journal. 2016;15(1):419.
253. Moonasar D, Nuthulaganti T, Kruger PS, Mabuza A, Rasiswi ES, Benson FG, et al. Malaria control in South Africa 2000–2010: beyond MDG6. Malaria Journal. 2012;11(1):294.
254. Khosa E, Kuonza LR, Kruger P, Maimela E. Towards the elimination of malaria in South Africa: a review of surveillance data in Mutale Municipality, Limpopo Province, 2005 to 2010. Malaria Journal. 2013;12(1):7.
255. Raman J, Gast L, Balawanth R, Tessema S, Brooke B, Maharaj R, et al. High levels of imported asymptomatic malaria but limited local transmission in KwaZulu-Natal, a South African malaria-endemic province nearing malaria elimination. Malaria Journal. 2020;19:152.
256. Module M. Preparation of Rapid Diagnostic Tests (RDTs) for DNA extraction v1.1. WWARN Procedure. 2011.
257. Plowe CV, 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. The American Journal of Tropical Medicine and Hygiene. 1995;52(6):565-8.
258. 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. PLoS Medicine. 2015;12(3):e1001788.
259. Awandu SS, Raman J, Bousema T, Birkholtz L-M. Ultralow-density *Plasmodium falciparum* Infections in African Settings. Clinical Infectious Diseases. 2019;69(8):1463-4.
260. Organization WH. Guidelines for the treatment of malaria: World Health Organization; 2015.
261. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. Nature. 2012;487(7407):375.
262. Piry S, Luikart G, Cornuet J-M. BOTTLENECK: a program for detecting recent effective population size reductions from allele data frequencies. Journal of Heredity. 1999;90:502-3.
263. Agapow PM, Burt A. Indices of multilocus linkage disequilibrium. Molecular Ecology Notes. 2001;1(1-2):101-2.
264. Kamvar ZN, Tabima JF, Grünwald NJ. *Poppr:* an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2014;2:e281.

265. Jombart T. *adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24(11):1403-5.
266. Winter DJ. MMOD: an R library for the calculation of population differentiation statistics. Molecular Ecology Resources. 2012;12(6):1158-60.
267. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics. 2010;11(1):94.
268. Awandu SS. Understanding the contribution of malaria parasites to continued residual malaria in a pre-elimination setting in South Africa: implications for elimination strategies: University of Pretoria; 2018.
269. Mobegi VA, Loua KM, Ahouidi AD, Satoguina J, Nwakanma DC, Amambua-Ngwa A, et al. Population genetic structure of *Plasmodium falciparum* across a region of diverse endemicity in West Africa. Malaria Journal. 2012;11(1):223.
270. Murray L, Mobegi VA, Duffy CW, Assefa SA, Kwiatkowski DP, Laman E, et al. Microsatellite genotyping and genome-wide single nucleotide polymorphism-based indices of *Plasmodium falciparum* diversity within clinical infections. Malaria journal. 2016;15(1):275.
271. Nabet C, Doumbo S, Jeddi F, Konaté S, Manciulli T, Fofana B, et al. Genetic diversity of *Plasmodium falciparum* in human malaria cases in Mali. Malaria journal. 2016;15(1):353.
272. Conway DJ, Roper C, Oduola AM, Arnot DE, Kremsner PG, Grobusch MP, et al. High recombination rate in natural populations of *Plasmodium falciparum*. Proceedings of the National Academy of Sciences. 1999;96(8):4506-11.
273. Anthony TG, Conway DJ, Cox-Singh J, Matusop A, Ratnam S, Shamsul S, et al. Fragmented population structure of *Plasmodium falciparum* in a region of declining endemicity. The Journal of Infectious Diseases. 2005;191(9):1558-64.
274. 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. Malaria Journal. 2016;15(1):1-12.
275. Taylor LH, Walliker D, Read AF. Mixed–genotype infections of malaria parasites: within–host dynamics and transmission success of competing clones. Proceedings of the Royal Society of London Series B: Biological Sciences. 1997;264(1383):927-35.
276. Sondo P, Derra K, Lefevre T, Diallo-Nakanabo S, Tarnagda Z, Zampa O, et al. Genetically diverse *Plasmodium falciparum* infections, within-host competition and symptomatic malaria in humans. Scientific reports. 2019;9(1):1-9.
277. Nassir E, Abdel-Muhsin A-MA, Suliaman S, Kenyon F, Kheir A, Geha H, et al. Impact of genetic complexity on longevity and gametocytogenesis of *Plasmodium falciparum* during the dry and transmission-free season of eastern Sudan. International Journal for Parasitology. 2005;35(1):49-55.
278. Bose J, Kloesener MH, Schulte RD. Multiple-genotype infections and their complex effect on virulence. Zoology. 2016;119(4):339-49.
279. Pollitt LC, Mideo N, Drew DR, Schneider P, Colegrave N, Reece SE. Competition and the evolution of reproductive restraint in malaria parasites. The American Naturalist. 2011;177(3):358-67.
280. Bushman M, Morton L, Duah N, Quashie N, Abuaku B, Koram KA, et al. Within-host competition and drug resistance in the human malaria parasite *Plasmodium falciparum*. Proceedings of the Royal Society B: Biological Sciences. 2016;283(1826):20153038.
281. Community SAD. SADC Malaria Status by 2014 Report2015.
282. Community SAD. SADC Malaria Status Report,. 2017.

283. Elimination-8. Country overview South Africa [Available from: https ://malariaelimination8.org/south-africa/.
284. Maharaj R, Moonasar D, Baltazar C, Kunene S, Morris N. Sustaining control: lessons from the Lubombo spatial development initiative in southern Africa. Malaria journal. 2016;15(1):1-9.
285. Okell LC, Bousema T, Griffin JT, Ouédraogo AL, Ghani AC, Drakeley CJ. Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. Nature communications. 2012;3(1):1-9.
286. Harris I, Sharrock WW, Bain LM, Gray K-A, Bobogare A, Boaz L, et al. A large proportion of asymptomatic Plasmodium infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. Malaria journal. 2010;9(1):1-8.
287. Manjurano A, Okell L, Lukindo T, Reyburn H, Olomi R, Roper C, et al. Association of sub-microscopic malaria parasite carriage with transmission intensity in north-eastern Tanzania. Malaria journal. 2011;10(1):1-8.
288. Wesolowski A, Stresman G, Eagle N, Stevenson J, Owaga C, Marube E, et al. Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. Scientific reports. 2014;4(1):1-7.
289. Gwarinda HB, Tessema SK, Raman J, Greenhouse B, Birkholtz L-M. Parasite genetic diversity reflects continued residual malaria transmission in Vhembe District, a hotspot in the Limpopo Province of South Africa. Malaria Journal. 2021;20(1):96.
290. South African National Department of Health. National guidelines for the treatment of malaria, South Africa. South African National Department of Health; 2018.
291. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;131(2):479-91.
292. Tessema SK, Hathaway NJ, Teyssier NB, Murphy M, Chen A, Aydemir O, et al. Sensitive, highly multiplexed sequencing of microhaplotypes from the *Plasmodium falciparum* heterozygome. The Journal of infectious diseases.jiaa527.
293. Imwong M, Stepniewska K, Tripura R, Peto TJ, Lwin KM, Vihokhern B, et al. Numerical distributions of parasite densities during asymptomatic malaria. The Journal of infectious diseases. 2016;213(8):1322-9.
294. Slater HC, Ross A, Felger I, Hofmann NE, Robinson L, Cook J, et al. The temporal dynamics and infectiousness of subpatent *Plasmodium falciparum* infections in relation to parasite density. Nature communications. 2019;10(1):1-16.
295. Trape J-F, Rogier C. Combating malaria morbidity and mortality by reducing transmission. Parasitology today. 1996;12(6):236-40.
296. Mawili-Mboumba DP, Akotet MKB, Kendjo E, Nzamba J, Medang MO, Mbina J-RM, et al. Increase in malaria prevalence and age of at risk population in different areas of Gabon. Malaria journal. 2013;12(1):1-7.
297. Pullan RL, Bukirwa H, Staedke SG, Snow RW, Brooker S. Plasmodium infection and its risk factors in eastern Uganda. Malaria journal. 2010;9(1):1-11.
298. Walldorf JA, Cohee LM, Coalson JE, Bauleni A, Nkanaunena K, Kapito-Tembo A, et al. School-age children are a reservoir of malaria infection in Malawi. PloS one. 2015;10(7):e0134061.
299. Wilson S, Booth M, Jones FM, Mwatha JK, Kimani G, Kariuki HC, et al. Age-adjusted *Plasmodium falciparum* antibody levels in school-aged children are a stable marker of microgeographical variations in exposure to Plasmodium infection. BMC infectious diseases. 2007;7(1):1-9.

300. Bretscher MT, Maire N, Felger I, Owusu-Agyei S, Smith T. Asymptomatic *Plasmodium falciparum* infections may not be shortened by acquired immunity. Malaria journal. 2015;14(1):1-10.
301. Vafa M, Troye-Blomberg M, Anchang J, Garcia A, Migot-Nabias F. Multiplicity of *Plasmodium falciparum* infection in asymptomatic children in Senegal: relation to transmission, age and erythrocyte variants. Malaria journal. 2008;7(1):17.
302. Bereczky S, Liljander A, Rooth I, Faraja L, Granath F, Montgomery SM, et al. Multiclonal asymptomatic *Plasmodium falciparum* infections predict a reduced risk of malaria disease in a Tanzanian population. Microbes and infection. 2007;9(1):103-10.
303. Sondén K, Doumbo S, Hammar U, Vafa Homann M, Ongoiba A, Traoré B, et al. Asymptomatic multiclonal *Plasmodium falciparum* infections carried through the dry season predict protection against subsequent clinical malaria. The Journal of infectious diseases. 2015;212(4):608-16.
304. Tran TM, Li S, Doumbo S, Doumtabe D, Huang C-Y, Dia S, et al. An intensive longitudinal cohort study of Malian children and adults reveals no evidence of acquired immunity to *Plasmodium falciparum* infection. Clinical Infectious Diseases. 2013;57(1):40-7.
305. Ntoumi F, Contamin H, Rogier C, Bonnefoy S, Trape J-F, Mercereau-Puijalon O. Age-dependent carriage of multiple *Plasmodium falciparum* merozoite surface antigen-2 alleles in asymptomatic malaria infections. The American journal of tropical medicine and hygiene. 1995;52(1):81-8.
306. Agyeman-Budu A, Brown C, Adjei G, Adams M, Dosoo D, Dery D, et al. Trends in multiplicity of *Plasmodium falciparum* infections among asymptomatic residents in the middle belt of Ghana. Malaria journal. 2013;12(1):1-6.
307. Konaté L, Zwetyenga J, Rogier C, Bischoff E, Fontenille D, Tall A, et al. 5. Variation of *Plasmodium falciparum* msp1 block 2 and msp2 allele prevalence and of infection complexity in two neighbouring Senegalese villages with different transmission conditions. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1999;93(Supplement_1):21-8.
308. Bendixen M, Msangeni H, Pedersen B, Shayo D, Bedker R. Diversity of *Plasmodium falciparum* populations and complexity of infections in relation to transmission intensity and host age: a study from the Usambara Mountains, Tanzania. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2001;95(2):143-8.
309. Peyerl-Hoffmann G, Jelinek T, Kilian A, Kabagambe G, Metzger W, Von Sonnenburg F. Genetic diversity of *Plasmodium falciparum* and its relationship to parasite density in an area with different malaria endemicities in West Uganda. Tropical medicine & international health. 2001;6(8):607-13.
310. Vafa M, Troye-Blomberg M, Anchang J, Garcia A, Migot-Nabias F. Multiplicity of *Plasmodium falciparum* infection in asymptomatic children in Senegal: relation to transmission, age and erythrocyte variants. Malaria journal. 2008;7(1):1-9.
311. Koukouikila-Koussounda F, Malonga V, Mayengue PI, Ndounga M, Vouvoungui CJ, Ntoumi F. Genetic polymorphism of merozoite surface protein 2 and prevalence of K76T pfcrt mutation in *Plasmodium falciparum* field isolates from Congolese children with asymptomatic infections. Malaria Journal. 2012;11(1):1-7.
312. GUPTA S, P. DAY K. A theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria. Parasite immunology. 1994;16(7):361-70.
313. Szpiech ZA, Jakobsson M, Rosenberg NA. ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics. 2008;24(21):2498-504.
314. Sundqvist L, Keenan K, Zackrisson M, Prodöhl P, Kleinhans D. Directional genetic differentiation and relative migration. Ecology and evolution. 2016;6(11):3461-75.

315. Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97.
316. Nei M. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences. 1973;70(12):3321-3.
317. Wilson GA, Rannala B. Bayesian inference of recent migration rates using multilocus genotypes. Genetics. 2003;163(3):1177-91.
318. Ejigu BA. Geostatistical analysis and mapping of malaria risk in children of Mozambique. PloS one. 2020;15(11):e0241680.
319. Sharp BL, Kleinschmidt I, Streat E, Maharaj R, Barnes KI, Durrheim DN, et al. Seven years of regional malaria control collaboration—Mozambique, South Africa, and Swaziland. The American journal of tropical medicine and hygiene. 2007;76(1):42-7.
320. Elimination-8. E8 strategic plan (2015-2020). Windhoek; 2015.
321. Chen I, Poirot E, Newman M, Kandula D, Shah R, Hwang J, et al. An assessment of the supply, programmatic use, and regulatory issues of single low-dose primaquine as a *Plasmodium falciparum* gametocytocide for sub-Saharan Africa. Malaria journal. 2015;14(1):1-10.
322. Fola AA, Kattenberg E, Razook Z, Lautu-Gumal D, Lee S, Mehra S, et al. SNP barcodes provide higher resolution than microsatellite markers to measure *Plasmodium vivax* population genetics. Malaria Journal. 2020;19(1):1-15.
323. Ghansah A, Kamau E, Amambua-Ngwa A, Ishengoma DS, Maiga-Ascofare O, Amenga-Etego L, et al. Targeted Next Generation Sequencing for malaria research in Africa: current status and outlook. BioMed Central; 2019.
324. Runtuwene LR, Tuda JS, Mongan AE, Makalowski W, Frith MC, Imwong M, et al. Nanopore sequencing of drug-resistance-associated genes in malaria parasites, *Plasmodium falciparum*. Scientific reports. 2018;8(1):1-13.
325. Cerqueira GC, Cheeseman IH, Schaffner SF, Nair S, McDew-White M, Phyo AP, et al. Longitudinal genomic surveillance of *Plasmodium falciparum* malaria parasites reveals complex genomic architecture of emerging artemisinin resistance. Genome biology. 2017;18(1):1-13.
326. Wong W, Griggs AD, Daniels RF, Schaffner SF, Ndiaye D, Bei AK, et al. Genetic relatedness analysis reveals the cotransmission of genetically related *Plasmodium falciparum* parasites in Thiès, Senegal. Genome Medicine. 2017;9(1):1-12.
327. Ndwiga L, Kimenyi KM, Wamae K, Osoti V, Akinyi M, Omedo I, et al. A review of the frequencies of indigenous *Plasmodium falciparum* Kelch 13 artemisinin resistance mutations in Africa. International Journal for Parasitology: Drugs and Drug Resistance. 2021.
328. Noviyanti R, Miotto O, Barry A, Marfurt J, Siegel S, Thuy-Nhien N, et al. Implementing parasite genotyping into national surveillance frameworks: feedback from control programmes and researchers in the Asia–Pacific region. Springer; 2020.
329. Molina-Cruz A, Canepa GE, e Silva TLA, Williams AE, Nagyal S, Yenkoidiok-Douti L, et al. *Plasmodium falciparum* evades immunity of anopheline mosquitoes by interacting with a Pfs47 midgut receptor. Proceedings of the National Academy of Sciences. 2020;117(5):2597-605.
330. Molina-Cruz A, Canepa GE, Kamath N, Pavlovic NV, Mu J, Ramphul UN, et al. Plasmodium evasion of mosquito immunity and global malaria transmission: The lock-and-key theory. Proceedings of the National Academy of Sciences. 2015;112(49):15178-83.
331. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. Science. 2015;347(6217).

CHAPTER 7 APPENDIX A

7.1 A.1 List of primers and PCR conditions for each microsatellite marker

Microsatellite Marker	Primer sequence (5'-3')	Primary PCR conditions	Secondary PCR conditions		
			Cycling Conditions	Ann. Temp (°C)	Ext. Temp (°C)
TA1	F: GTGTCTTTTTTATCTTCATCCCCAC and R: CCGTCATAAGTGCAGAGC	95 ºC5 min 95 ºC30 sec 60 ºC1:30 min 72 ºC30 sec 30 cycles 60 ºC30 min	98 ºC3 min 98 ºC10 sec 60 ºC1:30 min Ann.(°C) 30 secs 30 cycles Ext (°C) 30 secs Final extension (°C) 30 min	58	66
TA109	F: GTGTCTTAACGAGGTTAAATCAGGACAACA and R: GGGAAAATTATATCCCTTCTTGG				
AS1	F: CAAAGGAAAAAGACATGAACGA and R: GTGTCTTTTGTTGTTTAGATGGCTTAATTTTT				
AS11	F: GTGTCTTCCAATACCTTTAACCAGGCAAA and R: GGAGATACAGTGGTCGTTCCA				
AS31	F: GTGTCTTCTTGTGTAGGCACATCAGCA and R: CGAATGATACAACATTATGTCAATGAA				
AS32	F: GAAGGGTTCATTTTATTTAACTCTGTA and R: GTGTCTTAAATGGAACCATACAGGGAAAG				
AS34	F: GTGTCTTCAATTCTTATGTTCATAGGCAAGTTTA and R: CGTAAATTTTAATTATTCCCCTGGT				
B7M19	F: TGGTCTTGGTCTCGTTTTCA and R: GTGTCTTTTTCTTTTTCTTTCATTTCCAC				
AS3	F: TTCGAAGGGAAATACTGAAAGG and R: GTGTCTTCAGCTGTAATTTTCTTTGTTTTGTTTA			58	60
AS2	F: AACAATTTTGCAAACAATACAA and R: GTGTCTTTTTGCTAAGTTTTTATTTCCTGA				
TA87	F: AATGGCAACACCATTCAAC and R: GTGTCTTACATGTTCATATTACTCAC			57	63
AS12	F: GTGTCTTAAAAATGCAAAAATGGATAACAA and R: CACTATCTACTAAAGCATCATCATT			56	60
AS14	F: CATGAATAATACAAATGTTAAGTTGG and R: GTGTCTTAGATAAAACGGGTTCATTTCTT			56	66
ARA2	F: GTGTCTTCATATGAATAAACAAAGTATTGCTGA and R: TGGTCAAGTGGTACAGATCTTTTT	95 ºC5 min 95 ºC30 sec 56 ºC1:30 min 72 ºC30 sec 30 cycles 60 ºCfor 30 min			
TA60	F: GTGTCTTGAAAGCGATCCTCAAAGAGAAA and R: TTACAGGTAGTTATATCTTCACT				
AS21	F: GCCTTCATCTATACATAAACAGACA and R: GTGTCTTTGATGATATAAATAATGAACATTCTGA			56	60
AS19	F: GTGTCTTGGTTGATGTTTATCTATATCTTCTTT and R: TATGATATGAAACATACAAATGA				
AS15	F: TGAACAATAGTGATATAAACAGCA and R: GTGTCTTAGAACATAAATCAACATTACTACAA				
AS7	F: TTTTTAATAACACACACAATACAAAAG and R: GTGTCTTTCCTTTTTGGATAGATTTAATATGTTG				
AS25	F: GTGTCTTATATTTCCTTGGCCCCCTTT and R: CAAAAATAAACCACAATAATGA			50	60
Polyα	F: GTGTCTTGAAATTATAACTCTACCA and R: AAAATATAGACGAACAGA				
TA40	F: TTTTGGTTTCCAAGGGATT and R: GTGTCTTTTAAGGCCACGAGGAAATTG			54	66
PFG377	F: TTATCCCTACGATTAACA and R: GTGTCTTGATCTCAACGGAAATTAT				
PfPK2	F: GTGTCTTAAAGAAGGAACAAGCAGA and R: CTTTCATCGATACTACGA			58	66
AS8	F: TTTCTAGCATGGTCCAAAT and R: GTGTCTTTGGAAAATATTATGAGCATTTATGGA			58	60
TA81	F: TGGACAAATGGGAAAGGATA and R: GTGTCTTTTTCACACAACACAGGATT			57	63

*F - forward primer, R - reverse primer