New approaches for measuring fitness of *Plasmodium falciparum* mutations implicated in drug resistance
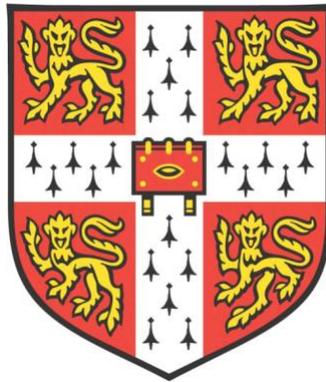
Manuela Carrasquilla
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

St. Edmunds College

January, 2019

Declaration

I hereby declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically stated in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit of 60,000 words, as required by the School of Biological Sciences.

Manuela Carrasquilla

mc898@cam.ac.uk

September, 2018

A mis hermanos

Acknowledgements

I am grateful with my uncle Alejandro, who is an anthropologist based in Leticia, Colombia, for showing me the huge social impact of tropical diseases, and the need for multidimensional solutions in the developing world in order to effectively tackle this problem. I´ll be forever grateful for my time there, and look forward to going back.

I would like to begin thanking my supervisors. Dr. Marcus Lee, for accepting me to be his PhD student while starting his own lab. He has been a great mentor who gave me enough trust and freedom to move around the different aspects of my project, as he provided me with the guidance and skills to start asking questions of my own. Also Dr. Julian Rayner for always being supportive, for guiding me through the PhD process, for his input and ideas and encouraging me to build connections in Colombia. For allowing me to prepare a workshop in bioinformatics in Bogotá which was one of the most rewarding teaching experiences I´ve had. I also want to thank my thesis committee members, Dr. Oliver Billker (Sanger) and Dr. Pete Bull (University of Cambridge) for their input, and Annabel Smith and Christina Hedberg-Delouka from the Graduate office for all their help and support.

I am incredibly grateful to all members of the Lee and Rayner groups at the Sanger Institute. In particular my friends and colleagues Sophie Adjalley and Chuan Cao for their incredible support, and for showing me how scientists can easily be kind and respectful. Alejandro Marin, Silvia Kariuki, and Arthur Talman for their friendship, scientific guidance and all their support through these last few years. Theo Sanderson for many bioinformatic sessions and for making me think on how to go around ways of analyzing data. Liam Prestwood for making things work in the lab and for his patience. Also, many former members of the malaria programme for their support, guidance and friendship. I don't think there is a better scientific environment than the malaria programme at the Sanger Institute. Also, Mandy Sanders, Wiesia Johnson, Abrar Ahmed and Katrina Robinson for all the support with sequencing and all other administrative help.

Finally, I want to express my gratitude to my friends and family. Thanks to my parents, my brother and my sister for accompanying me throughout the whole process of becoming a scientist. For all the support and for always understanding my career decisions and the fact that

I had to leave Colombia for the PhD, and the fact that I always want to find ways to apply my research there. To my partner, Marco, I will be eternally grateful for his never-ending support, for his patience, for his bioinformatics lessons, and for helping me become a better scientist and a better person. I can be difficult under stress and I´m glad he´s there to tell me when I´m being irrational. I think he´s suffered every word written on this thesis and I will always admire his positivity and kindness.

Abstract

The repeated emergence of drug resistance in *Plasmodium falciparum* underscores the importance of understanding the genetic architecture of current resistance pathways, as well as any associated fitness costs. Why resistance emerges in particular regions of the world has been linked to particular genetic backgrounds that better tolerate resistance-associated polymorphisms; this is likely to play a key role in driving the epidemiology of drug resistance, however is infrequently studied at a large scale in a laboratory setting. The first results chapter establishes a barcoding approach for *P. falciparum* with the aim of tracking parasite growth *in vitro*. The strategy used was adapted for *P. falciparum* by using a pseudogene (*PfRh3*) as a safe harbour to insert unique molecular barcodes. These libraries of barcoded *P. falciparum* vectors were also used as a readout of transfection efficiency. The second chapter establishes a proof of principle for phenotyping by barcode sequencing, using a panel of barcoded parasites generated in different genetic backgrounds that comprise sufficient genetic diversity to pilot the method. These were grown in the presence and absence of antimalarial compounds, and growth phenotypes were measured in parallel using BarSeq. The third results chapter studies the contribution of mutations in *Pfkelch13,* a molecular marker of artemisinin resistance, to parasite fitness. Combining CRISPR/Cas9-based genome editing and high throughput sequencing, the impact of *Pfkelch13* alleles on fitness in the context of particular strain backgrounds is revealed. In particular, the impact of genetic background in the emergence and spread of drug-resistant lineages (referred to as KEL1) in Southeast Asia carrying a Y580 *Pfkelch13* allele. Overall, given the current pace of genome sequencing of pathogenic organisms such as *P. falciparum*, it will be important to increase the scale of experimental genetics, in order to tackle in real-time natural variation that might be under constant selection from drugs, thus anticipating the emergence of drug resistance in changing parasite populations. Through this work, tools were developed to facilitate parallel phenotyping by measuring in vitro growth using high-throughput sequencing. The work also develops novel approaches to address the importance of genetic background and a potential role for positive epistasis in a lineage responsible for the recent outbreak of drug-resistant malaria in Southeast Asia.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACTs: Artemisinin combination therapies

AT: Adenine-thymine

BarSeq: Barcode Sequencing

Bp: Base-pair

CRISPR/Cas9: Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9

dAMP: Adenine Nucleotides

ddNTPs: Dideoxynucleotides triphosphate

DHA: Dihydroartemisinin

DNA: Deoxyribonucleic acid

dNTPs: deoxyribonucleotide triphosphat

DSB: double strand break

DV: Digestive Vacuole

GC: Guanosine-cytosine

gRNA/sgRNA: Single guide RNA

GWAS: Genome-Wide Association Study (GWAS)

Hb: Haemoglobin

hDHFR: human dihydrofolate reductase

HDR: Homology-directed repair

HRI/HRII: Homology region I/II

HTS: High-throughput sequencing

Hz: Hemozoin crystals

IC$_{50}$: Inhibitory concentration assay

IDC: Intraerythrocytic developmental cycle

iRBC: parasite-infected Red Blood Cells

Kb: Kilobase-pair

MDR: Multidrug resistance gene

Ne: Effective population size

NGG: Any nucleobase and two guanosine PAM motif.

NGS: Next Generation Sequencing

NHEJ: Non-Homologous End Joining

PAM: Protospacer Adjacent Motif

PCR: Polymerase Chain Reaction

Pfcrt/CRT: Plasmodium falciparum chloroquine resistance transporter

Pfkelch13/**K13:** Kelch 13 protein (PF3D7_1343700)

PfRh3/**Rh3:** Reticulocyte binding protein homologue 3

PlasmoGEM: Plasmodium Genetic Modification Project

RBC: Red blood cell

RPMI: Roswell Park Memorial Institute Media

RSA$_{(0-3h)}$: Ring Stage Survival Assay (0-3h)

SNP: Single nucleotide polymorphism

ug: Microgram

uL: Microliter.

mL: Mililiter.

uRBC: uninfected Red Blood Cells

yDHODH: Yeast dihydroorotate dehydrogenase

yFCU: Yeast cytosine deaminase and uridyl phosphoribosyl transferase.

ZFN: Zinc Finger Nucleases

Chapter 1: Introduction

1.1 Biology and epidemiology of malaria

1.1.1 The global burden of malaria

Malaria is a vector-borne disease that has been an enormous burden on humanity for thousands of years (Harper and Armelagos 2010). It currently puts at risk 3.4 billion people in endemic countries; accounting for nearly half of the human population according to the World Health Organization (WHO) (World Health Organization 2016). Populations affected are mainly in low-income countries where control efforts remain challenging, and where burden persists given socioeconomic factors that have led to changes in surveillance and policies, and to an increase in prevalence (Pigott *et al.* 2012). Historically malaria was even more widespread, and it has taken lives in many geographical locations and ecological conditions, ranging from the south of the United States to Southeast Asia, Southern Europe, Africa, Asia, and Central and South America. The outcomes of malaria varied widely between these various places, mainly because of differences in economic resources and implementation of control strategies, because global technologies for malaria control have not been efficiently utilised in all locations (Institute of Medicine, 1991).

In the past decade, renewed efforts towards malaria elimination have been initiated, including the United Nations Millennium Development Goals of 2015, the Roll Back Malaria Strategy, or the Medicines for Malaria Ventures (MMV) initiatives all aiming to significantly reduce the burden of malaria in endemic countries, and eventually to eliminate it (Hemingway *et al.* 2016; Van Voorhis *et al.* 2016). These efforts have clearly had an impact, with the World Health Organization estimating in 2015 that for the past 15 years global burden has decreased by 41%. All these world efforts have therefore drastically reduced mortality, primarily in Sub-Saharan Africa. The reduction in mortality has been driven by improvements in prevention, including extensive use of insecticide-treated bed nets in houses, better diagnostic tools, improvement in policy for usage of particular antimalarials, and overall better health-care

policies worldwide (Moonen *et al.* 2010). Despite this generally positive trend, there were still 214 million estimated cases and a total of 438,000 deaths in 2015 (World Health Organization 2016). Even more troubling, according to the most recent report of the WHO, there has been little improvement over the past two years, with Africa having an unchanged number of deaths reported in 2015 and 2016 (World Health Organization 2017).

Over this same time period, there has been a significant increase in our knowledge regarding the basic biology of the *Plasmodium* species that cause malaria. A notable example is the huge impact the completion of the reference genome of the human malaria parasite *Plasmodium falciparu*m has had for the field (Gardner *et al.* 2002), which has led to significant contributions in the areas of drug and vaccine development (Le Roch, Chung, and Ponts 2012).

Despite improvements in control and basic biology in the last decade, the global burden of malaria continues to be enormous. Currently, the population that is most affected by this parasitic disease is young children (0-10 years old), and it has had a dramatic impact on the development of many countries; from an economic and public health point of view. In addition, there are still challenges to the collective efforts towards elimination: the current lack of an efficient vaccine, the appearance of drug-resistant parasites to almost all antimalarial drugs that have been introduced into clinical practice, and the lack of effective prevention strategies that are spread worldwide rather than in individual countries.

1.1.2 *Plasmodium* parasites

To better understand the long-standing evolutionary arms race between *Plasmodium* parasites and their hosts it is important to do so in the context of its origin and natural history. Parasite species belonging to the Phylum Apicomplexa are unicellular eukaryotes that can infect various vertebrate hosts, ranging from reptiles to mammals and birds (Faust and Dobson 2015).

In total there are six *Plasmodium* species that routinely infect humans; *Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae, Plasmodium knowlesi, Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri*. The two that cause the overwhelming majority of malaria mortality and complications are *P. falciparum* and *P. vivax* (Sutherland *et al.* 2010). Malaria symptoms for

both *P. falciparum* and *P. vivax* are very similar; both include high fever, chills, vomiting and headaches, usually accompanying the replicating asexual cycles of the parasite within the red blood cells of the host. After these symptoms start in a patient, it is essential to start treatment as soon as possible, as a delay can lead to potentially life-threatening consequences (Miller *et al.* 2002; Mackintosh, Beeson, and Marsh 2004).

The full cycle of the malaria parasite will be discussed later in the chapter. However, it is important to note that different *Plasmodium* species have different lengths of asexual replication cycles, *P. knowlesi*, for example, has a 24-hour cycle, whereas *P. falciparum, P. vivax* and *P. ovale* have a 48-hour cycle. Finally, *P. malariae* has the longest asexual life cycle, lasting approximately 72 hours (Ott 1967).

Plasmodium parasites infecting humans seem to have originated on multiple occasions, and as a consequence harbour large genetic diversity. Initially, *P. falciparum* was thought to be closely related to rodent and avian parasites. Waters *et al.* (1991) used small-subunit ribosomal RNA (SSU rRNA) to show the phylogenetic relationships between these parasites, and *P. falciparum* was forming a monophyletic clade with the avian malaria parasite *P. gallinaceum*. These findings suggested lateral transfer given the spread and adaptation of the vector with rapidly developing populations in the area (Waters, Higgins, and McCutchan 1991). A few years later, Escalante AA & Ayala FJ (1994) also utilised SSU rRNA of different species inside the genus *Plasmodium* to try to test the earlier hypothesis, however including other species such as *Plasmodium reichenowi,* which infects chimpanzees, showing that it was closer phylogenetically to *P. falciparum* than was previously thought. The latter was then positioned inside a clade of ape malaria, more distantly related to both avian parasites and other human-infecting parasites like *P. vivax*. This supported the hypothesis that the two species diverged when humans and chimpanzees diverged from their common ancestor 5 million years ago (Escalante and Ayala 1994).

Amongst those infecting humans - including the species accounting for the highest mortality, *P. falciparum* - there has been a debate over the highly similar morphological features with species infecting both chimpanzees (*Pan troglodytes*) and gorillas (*Gorilla gorilla*).

In 2010 it was shown that all of these parasites fall within the subgenus Laverania, and that the most closely related species to *P. falciparum* originated in a gorilla and was named *Plasmodium praefalciparum* (Sundararaman *et al.* 2016; Liu *et al.* 2010).

P. vivax causes significantly less mortality than *P. falciparum* but has a much broader geographical distribution. It occurs at very low frequency in Africa, due to the high prevalence of the Duffy negative genetic trait in Sub-Saharan Africa. *P. vivax* relies on the Duffy antigen for the essential process of red-blood cell invasion. The widespread prevalence of this genetic trait in Africa means that the geographical distribution of *P. vivax* is almost sympatric to that of *P. falciparum*. *P. vivax* primarily occurs in temperate regions of Asia and the Americas, as opposed to *P. falciparum* that occurs predominantly in sub-Saharan Africa (Baird 2013), although the latter is also found throughout Asia and the Americas at low endemicity.

Fig. 1.1 shows the distribution of *P. falciparum,* the species of study throughout this project. Distribution is measured by the prevalence of malaria caused by *P. falciparum* in children 2-10 years old, which includes those younger than five years old that are the most vulnerable population as immunity has not been acquired yet (Malhotra *et al.* 2009).
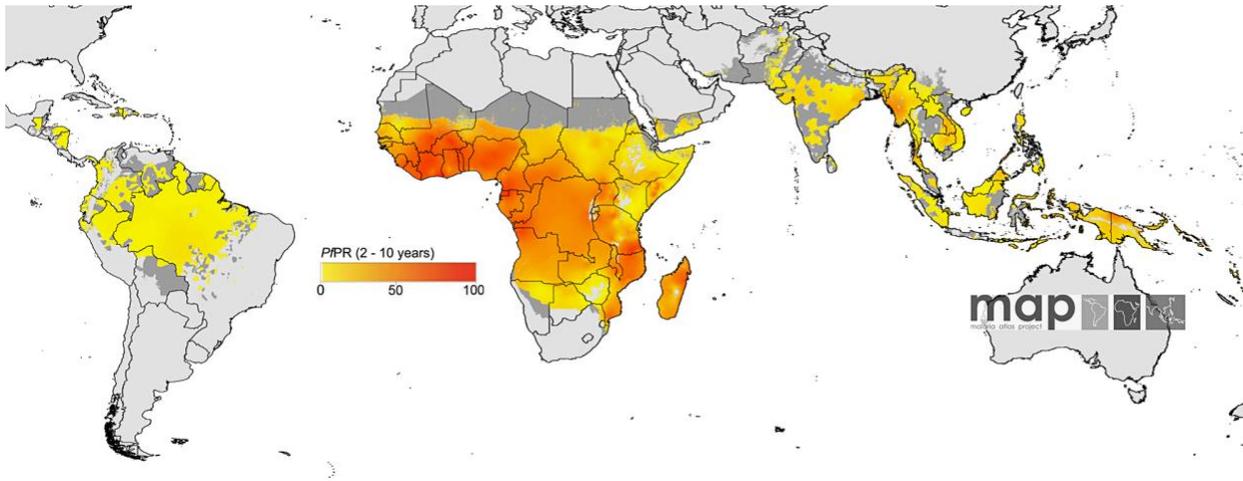
Figure 1.1 Map showing the spatial distribution of malaria caused by *P. falciparum* worldwide. Prevalence of malaria in children between 2 and 10 years of age as a predictive of malaria prevalence. The lower end of the colour spectrum (white) indicates absent or very low risk of malaria, whereas prevalence is marked in orange and red. Map taken from the Malaria Atlas Project from the University of Oxford (https://map.ox.ac.uk/).

1.1.3 The human malaria parasite *Plasmodium falciparum*

Besides being the parasite species accounting for the highest mortality worldwide, *P. falciparum* has also been considered the most virulent, having evolved complex mechanisms to avoid being cleared by the immune system of the host, including the ability to cytoadhere to endothelial surfaces using the protein products of a variable gene family (Miller *et al.* 2002). As seen in Fig. 1.1, it is primarily present in sub-Saharan Africa but also has an important impact in South America and Southeast Asia. It is important to note that both the epidemiology and burden of disease in the different geographical locations vary greatly. Studying the population structure of parasites around the globe therefore would allow a more thorough understanding of the genetic architecture that makes specific parasite populations more successful under certain ecological conditions or selective pressures than others (Amato *et al.* 2018). This idea will be expanded in more detail in this same chapter.

The life cycle of *P. falciparum*, like that of other species in the genus, is complex, alternating between a human host and a mosquito vector (Bannister and Mitchell 2003). Fig. 1.2 shows a representation of the life cycle of *P. falciparum*.

The cycle starts with the infectious bite of a female *Anopheles sp*. mosquito, which as it feeds on a blood meal from the host will inject a variable amount (usually a median of 50-100) of the haploid, motile form called the sporozoite, which wait inside the salivary glands of the mosquito until blood feeding occurs (step 1) (Kebaier, Voza, and Vanderberg 2009). Sporozoites enter the skin and travel through the blood vessels until they reach the liver, where they invade liver cells (step 2-4). Once the liver cells (hepatocytes) are infected, parasites continue to replicate until becoming a multinucleated liver schizont that later ruptures releasing merozoites into the bloodstream, which subsequently infect erythrocytes (step 5) (Tavares *et al.* 2013; Aly, Vaughan, and Kappe 2009).

Once parasites enter the red blood cells, they initiate the Intraerythrocytic Developmental Cycle (IDC), which is accompanied by the appearance of the typical physiological symptoms of malaria. In *P. falciparum*, the IDC usually takes 48 hours, although there is some evidence of subtle variations in the length in life cycle of different strains cultured in the laboratory (Rovira-Graells *et al.* 2016). During this stage of the life cycle, parasites, for the most part, undergo asexual replication, with systematic changes in different morphological and developmental stages shown in Fig. 1.2. This cycle initiates with a merozoite infecting a red blood cell (RBC), and shortly after entering a ring-like stage. This ring stage last for up to the first 24 hours, and is characterised by the development of membrane structures within the infected RBC. The parasites then enter the trophozoite stage, regarded as being a the most metabolically active developmental stage of the IDC (Bozdech *et al.* 2003). At this stage, the digestive vacuole (DV) appears and can be observed by light microscopy due to the formation of hemozoin crystals derived from hemoglobin digestion. Trophozoites take approximately 12 hours to reach maturation, a process that ultimately leads to a transition from the G1/G0 cell cycle stage to the S phase, where a very fast DNA replication of the haploid genome starts to take place. The last stage in the asexual cycle, the schizont, can be easily distinguished as

multiple nuclei appear within the infected red blood cells. These nuclei are ultimately packaged into new daughter cells, called merozoites, and can number up to 36 depending on parasite strain. At maturity, the schizont ruptures and the merozoites are released into the bloodstream. Repeated rounds of this process leads to symptoms in malaria-infected patients such as high fevers and anaemia, peaking consistently with the length of the IDC (Lee *et al.* 2014; Bousema *et al.* 2014). Fig. 1.3 shows in detail the different developmental stages that the parasite undergoes throughout the IDC. This is also the phase of the life cycle that is the easiest to maintain *in vitro,* and is the main part of the life cycle that will be discussed throughout this project (step 5-6).
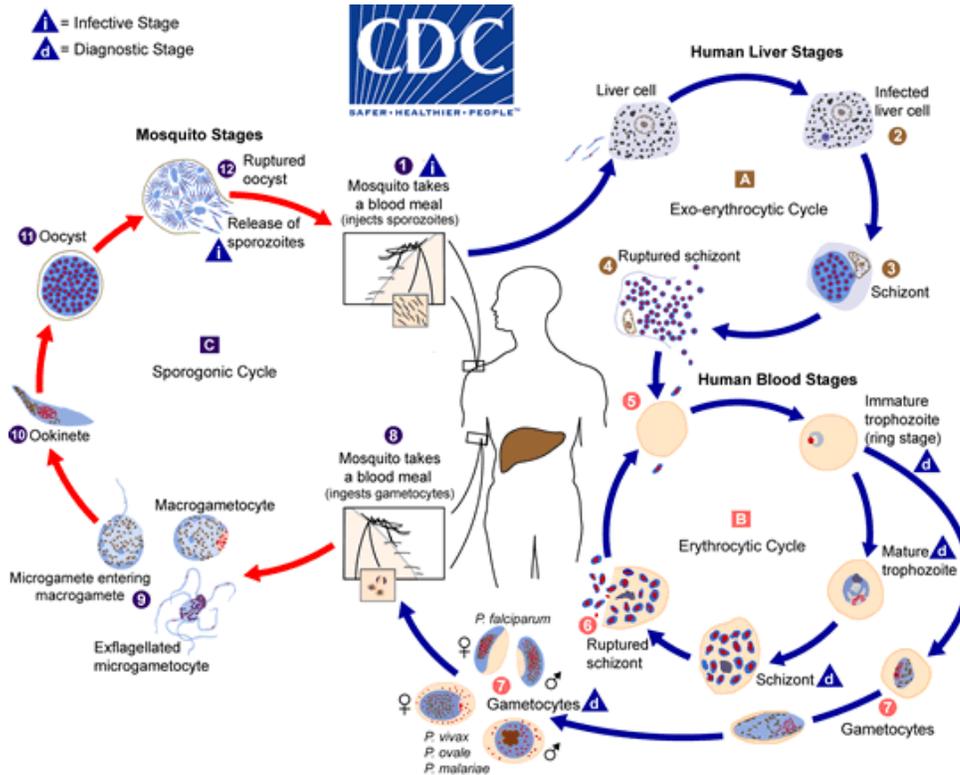
Figure 1.2 The life cycle of the human malaria parasite *P. falciparum.* It divides into three main developmental stages: the liver, intraerythrocytic and mosquito stages. Injected sporozoites migrate to the liver, where they generate liver-stage merozoites (A), a process which takes between 12-16 days. After this, parasites migrate into the bloodstream where most of the typical symptoms of malaria occur (B). Sexual differentiation takes place during the blood stages, with a later migration to the peripheral blood for uptake by a female *Anopheles* mosquito, within which fertilisation occurs thus completing the life cycle (C). Figure reproduced from the Centers for Diseases Control and Prevention (http://www.cdc.gov/malaria/about/biology).
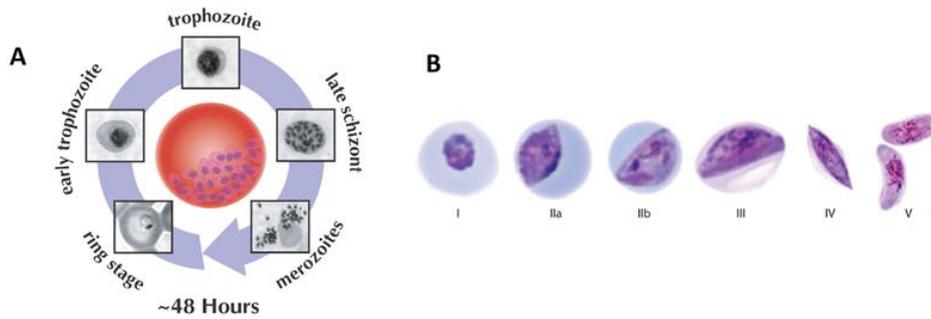
Figure 1.3 Intraerythrocytic Developmental Cycle. A. Figure showing the Intraerythrocytic Developmental Cycle (IDC) of *P. falciparum*. Modified from (Bozdech *et al.* 2003). **B.** Development of sexual forms (gametocytes) into five distinct morphological stages. Only the last stage of differentiation of female and male gametocytes is transmissible to female *Anopheles* mosquitoes. Adapted from (Bousema *et al.* 2014).

Once parasites are established in the bloodstream, they must enter a developmental stage capable of being transmitted to the mosquito host in order to continue their life cycle. Extensive evidence has been produced addressing the underlying molecular mechanisms of this developmental change, often referred to as "switching", when a small fraction of the parasite population becomes a transmissible sexual form. Specific transcription factors of the ApiAP2 family (De Silva *et al.* 2008) have a critical role in switching, with one member of the family, AP2-G, validated as essential for the primary developmental switch in both cultured *P. falciparum* parasites and the *in vivo* rodent malaria model *Plasmodium berghei* (Kafsack *et al.* 2014; Sinha *et al.* 2014).

Figure 1.3B shows the morphology of the two forms of *P. falciparum* gametocytes (7). Of note, parasites at the early stages of gametocyte development are very similar morphologically to asexual trophozoites entering the S phase; therefore, it is difficult to distinguish them by microscopy. Once gametocytes have finalised the five distinct stages of maturation shown in Fig. 1.3A, they go to the peripheral bloodstream where they can be taken up by a mosquito. Once they are taken up in a blood meal by the mosquito they develop either into the male microgamete (1n), which rapidly undergoes three rounds of DNA replication producing eight haploid genomes, or a female macrogamete (1n), leading to fertilisation (2n). Previous work by Billker *et al.* showed that the process of gamete activation is influenced by two main factors,

driven by a drastic change in the host environment. The considerable drop in temperature from 37℃ to 18℃ following the transition between the mammalian and mosquito hosts, and the presence of xanthurenic acid, a molecule produced by the mosquito, are sufficient changes to trigger the differentiation of these two forms into gametes (Billker *et al.* 2004). Once this has taken place, the gametes mate inside the mosquito forming a zygote (2n), which later undergoes a round of replication followed by meiosis to form a tetraploid ookinete (Fig. 1.2, steps 9-10) (Bousema *et al.* 2014). Ookinetes then cross the midgut epithelium of the mosquito to form oocysts. The latter undergo sporogony, within which several rounds of replication take place, resulting in haploid sporozoites that migrate to the salivary glands of the mosquito. By being injected into a new host, sporozoites re-start the life cycle of *P. falciparum* (11-12) (Bannister and Mitchell 2003).

1.2 Next Generation Sequencing Technologies

Until recently, DNA sequencing relied on capillary sequencing, a method developed by Fred Sanger in 1973 (Sanger *et al.* 1973). By using deoxynucleotides (ddNTPs) to prime the DNA polymerase, and by incorporating radioactive nucleotides, the order of bases in a sequence could be determined during the process of DNA replication (Bentley *et al.* 2008).

More recently, there have been multiple platforms developed that fall into the category referred to as Next Generation Sequencing. These platforms have been developed in order to reach a higher number of more reliable sequence 'reads', and most importantly at a lower cost and considerably higher throughput than was previously accomplished with Sanger sequencing (Sanger *et al.* 1973).

Next Generation Sequencing (NGS) has accelerated the field of genomics in various ways. Access to higher quality sequences has allowed expansion to other areas of genomics such as the accurate discovery of polymorphisms in DNA or transcriptomics, both of which have been shown to have important biological and clinical implications. The most well-known NGS platform comes from the biotechnology company Illumina, characterised by generating short-reads from DNA sequencing. More recently, other companies such as Pacific Biosciences or

Oxford Nanopore have developed long-read sequencing, but will not be discussed here as all the next generation sequencing used in this study was performed using the Illumina platform.

1.2.1 Illumina sequencing

Illumina sequencing technology utilises a cycle of reversible termination chemistry, which was initially described by Bentley *et al.* 2008. It involves three major steps in sample preparation and processing: 1) library preparation, 2) generation of clusters, and 3) sequencing. Illumina sequencing libraries are generated by shearing DNA into short length fragments. There are different ways by which the shearing process can take place, including sonication (with a Covaris machine or equivalent) or by using enzymatic digestion. Fig. 1.4A shows the fragments generated after shearing, which are then end-repaired to make blunt ends. After this, adenine nucleotides (dAMP) are added to generate a dA-tail to the 3' end of the blunt DNA fragments. The latter can be used as a template to ligate Illumina sequencing adaptors, which contain a thymine nucleotide overhang. Lastly, the reaction is catalysed using a T4 DNA ligase. Upon completion of the library, the fragments are washed to remove free adaptors.

Appropriately sized fragments that contain the adaptors are then selected and bound into flow cells (Fig. 1.4B). The flow cell is a glass-based surface coated with oligonucleotides that are complementary to the sequencing adaptors attached to the DNA fragments (Bentley *et al.* 2008).
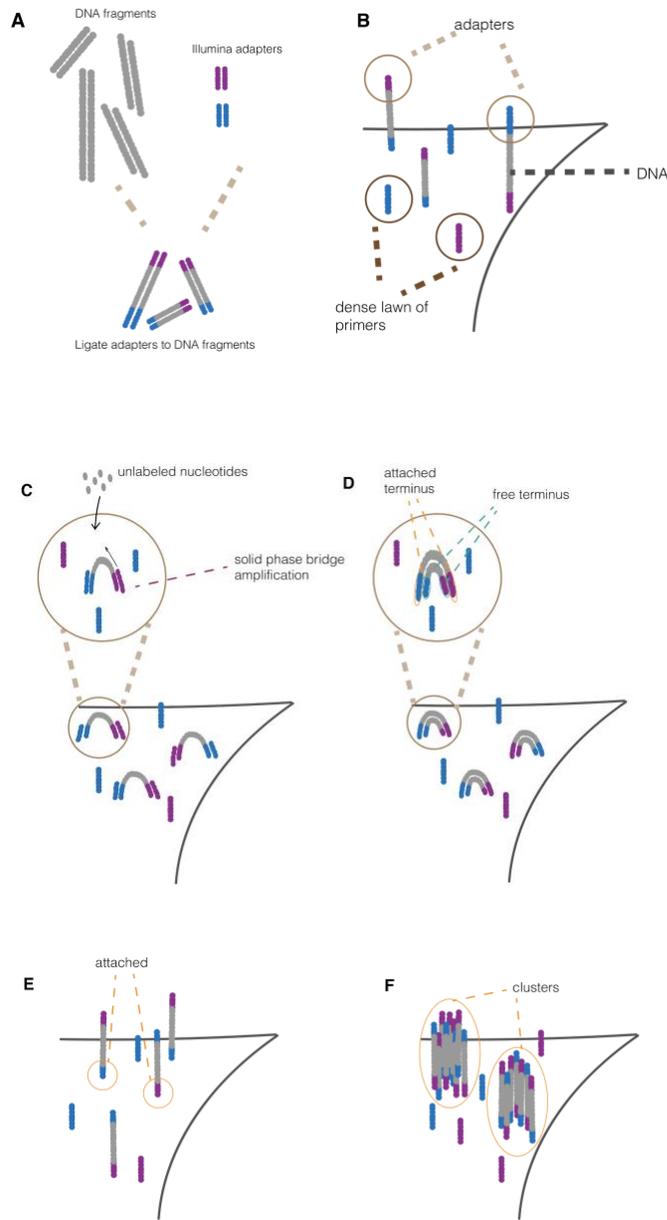
Figure 1.4 Generation of a DNA sequencing library for the Illumina sequencing platform. DNA is sheared using sonication or enzymatic digestion. **A.** Illumina sequencing adaptors are ligated to blunt-ended fragments. **B.** Sheared DNA fragments are immobilised after size quantification. **C.** Unlabeled nucleotides are incorporated for solid-phase bridge amplification. **D-E.** Paired-end reads are generated after synthesis of the complementary strand. **F.** Generation of clusters after single DNA molecules attached to the flow cell are clonally amplified.

The flow cell contains complementary primers to the sequencing adaptors described previously, allowing single-stranded DNA molecules from the library to hybridise (Fig. 1.4B),

followed by a bridge amplification step in which the DNA molecules bound to the flow cell bind to another complementary oligonucleotide from the flow cell with its free adaptor (Fig. 1.4C). The generation of paired-end reads follows this process after acting as a template for synthesis of the complementary strand (Fig. 1.4D-E), which continues until clusters are generated that result from clonal amplification of single molecules of DNA, with thousands of templates per cluster being generated (Fig. 1.4F).

The next step in Illumina sequencing relies on the use of single modified nucleotides (dNTPs) that contain a terminator sequence that prevents incorporation of a subsequent nucleotide. These dNTPs are also fluorescently labelled, with a fluorophore that is specific to each DNA base; meaning that they can be excited in different channels by using a laser (Fig. 1.5). Imaging allows the intensity of each fluorophore to be accurately measured, and the process of base determination takes place. After this process, the terminator moiety is removed to allow for another cycle to proceed. The base calling process for the Illumina platform uses a software (Bustard) that determines the order of base incorporation in each cycle (Erlich *et al.* 2008).
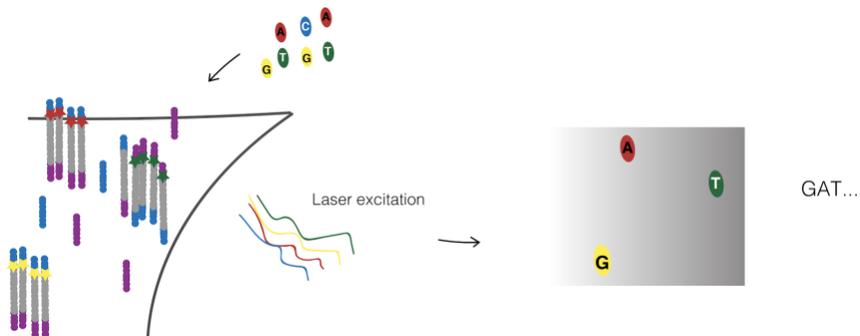
Figure 1.5 DNA sequence determination. The addition of labelled reversible terminator nucleotides, primers and a polymerase to the flow cell allow for the initiation of the first chemistry cycle where the first base is identified after laser excitation and imaging. Multiple cycles take place and allow for determination of the sequence order of the bases identified in a particular DNA molecule.

Once the base calling has been performed, there are several bioinformatic tools in place to analyse short reads coming from the Illumina sequencing platform. Algorithms have been developed, such as the Burrows-Wheeler transform method (BWA) to allow for short reads to be efficiently aligned to a reference genome (Li and Durbin 2009).

1.3 Genetics and genomics of *Plasmodium falciparum*

1.3.1 Efficient adaptation of *P. falciparum* to *in vitro* culture

One of the major accomplishments of the last 50 years of malaria research was the successful adaptation of parasites to a stable *in vitro* culture system, thanks to the work of Trager and Jensen in 1976, who were able to successfully co-culture patient isolates with uninfected red blood cells (Trager and Jensen 1976). The use of supplemented media, the capacity to keep the culture at 37℃, and maintaining conditions similar to those in the bloodstream of a patient meant keeping the culture at elevated concentrations of carbon dioxide. Since this critical accomplishment, a large amount of research has come out from the establishment of the system, including one of the most significant achievements, the sequencing of the whole genome of *P. falciparum*. In their pioneering work, Gardner *et al.* assembled the first full genome, obtained from a clone named 3D7 of an adapted parasite strain NF54 that was isolated from a traveller that came back to the Netherlands (Gardner *et al.* 2002).

1.3.2 Sequencing and assembly of the whole genome of *P. falciparum*

The genome of most *Plasmodium spp.* is composed of 14 chromosomes, and is around 23 million bases which harbour approximately 5400 genes. One of the major findings of the grand effort to sequence the genome was the fact that *P. falciparum* had a surprisingly high adenine (A) and thymine (T) content (~80% A-T, and up to ~90% in intergenic regions of the genome). As well as being interesting evolutionarily, this is also one of the key reasons why validation of gene function in *P. falciparum* has been challenging. Even though the genome sequence revealed the sequence of many genes of interest, the high A-T content makes the generation of transfection constructs in *E. coli* challenging, as large *P. falciparum* DNA fragments are unstable in bacteria (Rug and Maier 2013).

Analysis of the genes present in the *P. falciparum* genome revealed some surprises, including the lack of one of the two major DNA repair mechanisms. The absence of non-

homologous end joining (NHEJ) was observed in various bioinformatics analyses performed since the assembly of its genome in 2002 by Gardner and collaborators (Gardner *et al.* 2002). Proteins such as Ku70, Ku80 and DNA ligase IV that act to repair stochastic double-strand breaks of DNA, and are key for NHEJ in other organisms are absent in *P. falciparum* (Lee *et al.*, 2014; Straimer, Lee, *et al.*, 2012). This has significant implications for CRISPR/Cas9 genome editing, discussed further below.

It took almost two decades after *in vitro* culture systems were developed to understand how to efficiently deliver DNA across the four membranes surrounding the nucleus of the parasite. Between 1995-96 the research groups of T.E. Wellems and A. Cowman published transfection by electroporation of *P. falciparum* ring stages (Wu, Kirkman, and Wellems 1996; Wu *et al.* 1995; Crabb *et al.* 1997). With this crucial event, molecular genetics in *P. falciparum* seemed possible. However, modification of the genome still required very long periods of time (up to six months of continuous culture) for homologous recombination to occur. Depending on the design of the constructs, transfections can either lead to stable maintenance of circularised plasmids (episomes), or chromosomal integration of the whole plasmid with single-crossover recombination driven by homologous recombination to repair a stochastic DNA damage event at the target locus (Duraisingh *et al.* 2002).

1.3.3 Using genomic data to understand the diversity of *P. falciparum* populations worldwide

Once a reference genome had been sequenced and assembled, understanding the genetic diversity of *P. falciparum* worldwide became a priority. Over the last decade, the next generation sequencing technologies described earlier were successfully adapted with the aim of performing large-scale genomic studies of parasites directly isolated from patients in malaria-endemic regions.

As well as being of specific interest in understanding epidemiological features of malaria, genomics of *P. falciparum* also serves as a good system to study some of the general concepts in population genetics, for example how the effective population size (N_e) can

determine how specific variants spread in a given population. The latter can start from the notion that there have been evolutionary bottlenecks that have shaped genetic diversity of parasites in the different regions of the world, and that some variants that might be deleterious are more easily lost to natural selection in areas with a high N_e.

An important study that utilised NGS methods was undertaken in 2012, in which a catalogue of single nucleotide polymorphisms (SNPs) was generated in order to characterise *P. falciparum* populations (Manske *et al.* 2012). The aim was to address how parasites are related across different regions of the world, and how this could be linked to ecological and epidemiological variables in the different regions. In this study, the global population structure of field parasites, and the extent to which these polymorphisms determined population structure was established for the first time. The result of the analysis of these polymorphisms by Principal Component Analysis (PCA) used the difference in the mean fixation index (F_{ST}) - a measurement typically used to differentiate populations - and a neighbour-joining phylogenetic tree to show a high degree of genetic structure in the global parasite populations. Clear differences between regions were also identified, with, for example, Thailand and Cambodia in Southeast Asia forming distinct clusters (green and blue respectively in Fig. 1.6), whereas Mali and Burkina Faso in West Africa (yellow and brown, respectively) showed a higher level of admixture. The latter is a feature generally consistent with more stable transmission, which is the case for West Africa, as opposed to low or unconstant transmission commonly observed in Southeast Asia. This finding can also be explained by the high levels of inbreeding observed in Southeast Asia (Manske *et al.* 2012).
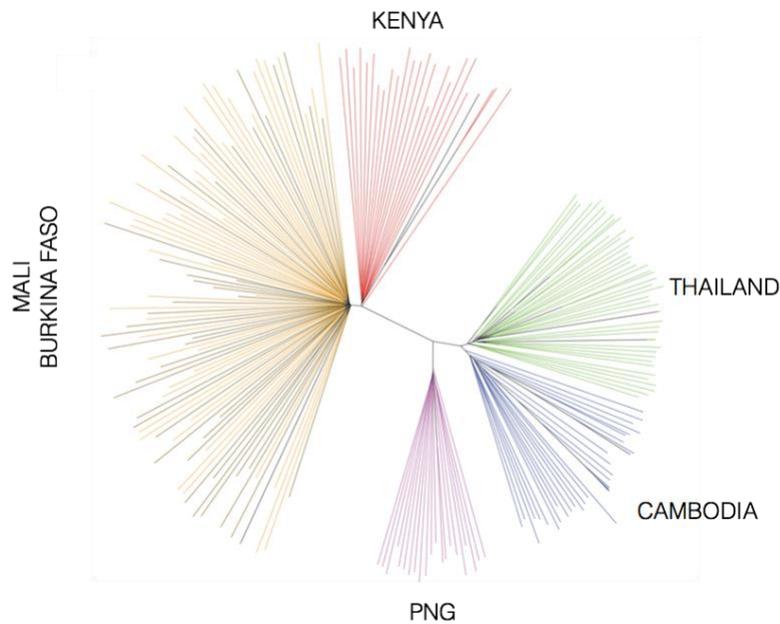
Figure 1.6 Genomic diversity of *P. falciparum* isolates. Neighbor-joining tree of 227 world samples included in this study showing high degree of population structure in Southeast Asia as opposed to West Africa (Burkina Faso and Mali). Lines indicate an individual sample and the length of the branches refers to the genetic distance between samples, showing highly structured populations in SEA. Modified from (Manske *et al.* 2012).

The findings from this study are relevant for our understanding of antimalarial drug resistance, as overall for populations with low genetic diversity and high levels of inbreeding (e.g. Southeast Asia or South America), mutations that appear to be beneficial under a certain selective pressure might increase in frequency; an example that has been seen with antimalarials such as chloroquine and more recently artemisinin (Payne 1987; Ashley *et al.* 2014). This would be consistent with differences in N_e, and will be addressed later in the dissertation from the perspective of some of the variants involved in antimalarial drug resistance.

1.4 Antimalarial drug resistance

1.4.1 A history of antimalarial drug resistance

The constant process of adaptation between pathogens and hosts has been extensively studied and is often referred to by Leigh Van Valen's New Evolutionary Law ("Red Queen Hypothesis") (Valen, 1973). The latter usually implies that this co-adaptation between pathogens and hosts occurs in cycles, indicating that fitness lost through adaptive compromise is then restored via compensatory mechanisms. Some of these fluctuations can come not from host immunity, but rather from externally applied selective pressures such as, in the case of *Plasmodium* parasites, an antimalarial compound, and ultimately how the process of adaptation takes places will depend on the population and evolutionary history of the parasite and host. Given this, understanding the emergence and spread of antimalarial drug resistance not only has to be understood in the context of the history of different compounds being utilised and approved for the clinic, but also in the context of the various drug policies that exist worldwide, with all their socioeconomic implications for public health.

The mechanisms by which resistance mutations emerge, persist and spread within parasite populations depends heavily on biological fitness, or the ability of a parasite to persist in a given population, which is in turn shaped by external factors such as host or local environment, meaning that resistance-conferring alleles can be passed to the next generation at different rates (Moura de Sousa *et al.* 2017).

Generally, parasites have evolved mechanisms which allow them to optimize biochemical processes in order to cope with such selective pressures, which might incur an associated fitness cost. An example of drugs impacting negatively fitness are mutations in cytochrome b that result in decreased susceptibility to the antimalarial atovaquone but negatively impact fitness (Peters *et al.* 2002), or copy number amplifications in the multidrug resistance gene *pfmdr1* leading to increased survival upon exposure to the antimalarial drug mefloquine but reduced fitness (Preechapornkul *et al.* 2009). Another line of evidence that can be extrapolated from the field is the reversal of resistance to chloroquine in Malawi after the

drug was removed from use, supporting the idea that mutations in the chloroquine resistance gene *pfcrt* had a fitness cost in these African parasites (Kublin *et al.* 2003). Although fitness landscapes are shaped by various ecological and evolutionary forces, measuring parasite fitness in laboratory settings has proven to be useful in order to study the impact of drug resistance on parasite fitness, specifically determined by the *in vitro* growth of individual parasite strains over time. However, due to the limitations in transfection efficiency for *P. falciparum*, a general feature of most assays that are aimed at measuring parasite fitness is that they rely on head-to-head competition comparing at most two lines, and have not involved parallel phenotyping at a large scale, limiting the understanding of the *in vitro* fitness of cultures *P. falciparum* strains.

Identifying markers of drug resistance allows accurate tracking of the origin and spread of resistance mutations, and interrogation of their fitness effects. Some regions in the world are clearly more prone to the emergence of resistance alleles. Specifically, Southeast Asia has historically been a "hot-spot" of drug resistance as will be described with some examples later in this chapter. Some of the explanations for this very localised epidemiology of resistance have a lot to do with both environmental factors and natural history, an example being the study of the worldwide genetic diversity of *Plasmodium falciparum* populations performed by Manske *et al.* 2012 (Manske *et al.* 2012).

Various studies have shown that Southeast Asia has overall low mosquito transmission rates as compared to West Africa (Dalrymple, Mappin, and Gething 2015). Infections in a low transmission setting are usually characterised by a high proportion of patients carrying symptomatic infections; this is thought to result in high drug treatment rates and low levels of immunity, in contrast to areas of high transmission, which have a high prevalence of asymptomatic infections due to the more routine development of effective clinical immunity. As such, the epidemiology of drug resistance in Southeast Asia is potentially driven in part by widespread drug exposure of largely monoclonal parasite infections (Blasco *et al.* 2017).

Today the current pipeline of drugs for treatment is being compromised as it relies on a limited number of compounds to which resistance is becoming widespread (Flannery *et al.* 2013). Table 1.1 shows the major antimalarial compounds that have been introduced into the

clinic, the year in which they were introduced, the year when resistance was first reported and some references to potential mechanisms of action (Paloque *et al.* 2016).

Table 1.1 Summary of main antimalarials that have been introduced into the field and for which resistance has been reported. Adapted from Paloque *et al.* (2016).

Antimalarial	Year Introduced	First resistance reported	Reference to main resistance mechanisms
Quinine	1632	1910	Disruption of drug accumulation inside the food vacuole by reduced propensity of the drug transporter PfMDR1 to transport the anti-malarial (Sanchez *et al.* 2010; Koenderink *et al.* 2010).
Chloroquine	1945	1957	Drug extrusion from the digestive vacuole by mutated drug transporter Pfcrt due to negativity of the transporter allowing efflux of protonated chloroquine (Ibraheem *et al.* 2014; Summers, Nash, and Martin 2012; Setthaudom *et al.* 2011).
Proguanil	1948	1949	Modification of the drug target: reduced inhibition of enzymatic activity by the drug (Ménard *et al.* 2008).
Sulfadoxine-pyrimethamine	1967	1967	
Mefloquine	1977	1982	Reduction of parasite susceptibility to mefloquine by amplification of *pfmdr1* copy number (Sidhu *et al.* 2006).
Atovaquone	1996	1996	Modification of the drug target by inhibition of the ubiquinol oxidation pocket of cytochrome bc1 complex (Kessl, Meshnick, and Trumpower 2007).
Artemisinins	1980s	2006	Delayed progression of ring stages, up-regulation of unfolded protein response, quiescence (Witkowski *et al.* 2010; Ariey *et al.* 2014; Mok *et al.* 2015).

1.4.2 Antimalarials and their mechanisms of action

1.4.2.1 The Digestive Vacuole

There have been many compounds derived or synthesized that have antimalarial properties and for which the mechanism of action seems to be associated with the blockage of the heme detoxification pathway undertaken by the parasite inside the DV.

The DV is a highly specialized organelle that is formed by endocytosis of RBC cytosol at structures known as cytostomes, which are formed by invagation of the parasite plasma membrane and parasitophorous vacuolar membrane (Dluzewski *et al.* 2008; Aikawa *et al.* 1966). Abu Bakar *et al.* 2010 used live imaging to further characterize the formation of this organelle, using fluorescence dextran probes to demonstrate a change in the acidity of the compartment after its formation (Bakar *et al.* 2010). As a result, the mature DV of *Plasmodium* parasites is highly acidic (pH 5.2-5.5) and is maintained by a proton gradient through an ATPase pump (Klonis *et al.* 2007; Dhingra *et al.* 2017). During most of the intraerythrocytic developmental cycle this compartment plays an essential role in the process of degradation of host hemoglobin (Hb), primarily carried out by aspartic and cysteine proteases including plasmepsins 2 and 3 (PM2 and PM3) and falcipains. The breakdown products include short peptides and amino acids, as well as heme that is ultimately detoxified into inert hemozoin crystals (Hz) (Chugh *et al.* 2013). It has been found that antimalarials belonging to the quinoline class interfere with the heme detoxification process. In Fig. 1.7 some of the quinolines are labelled as QN (Quinine), CQ (Chloroquine), Amodiaquine (AQ), and Piperaquine (PPQ) (Blasco *et al.* 2017).
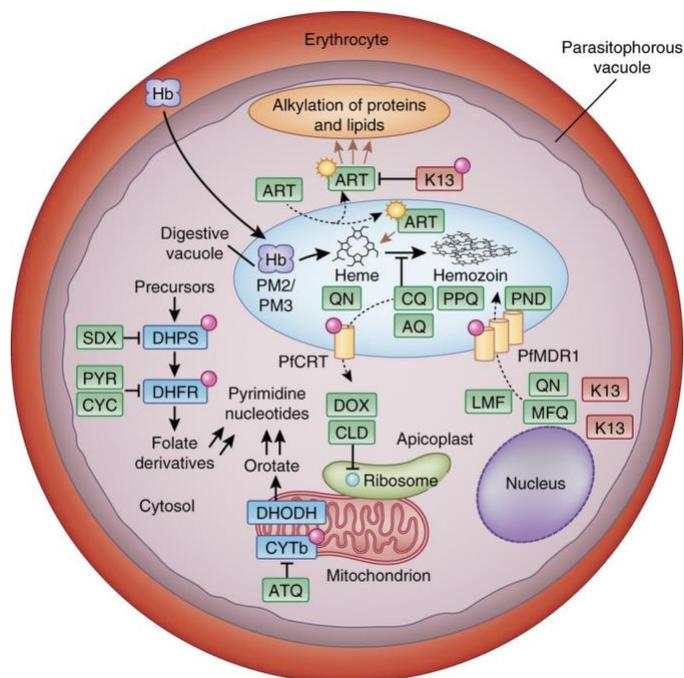
Figure 1.7 Diagram of the digestive vacuole of *P. falciparum* and various antimalarials with their suggested mechanism of action. The blue oval corresponds to the DV of the parasite where multiple antimalarials have been shown to act. In particular, 4-aminoquinolines have an ability to interfere with the process of detoxification of heme coming from hemeoglobin degradation (quinine (QN), chloroquine (CQ), piperaquine (PPQ), amodiaquine (AQ), mefloquine (MFQ) and lumefantrine (LMF)). Additional known and validated pathways being targeted by other antimalarials are related to the pyrimidine biosynthesis (sulfadoxine (SDX) inhibiting dihydro dihydropteroate synthase (DHPS) and pyrimethamine (PYR)/cycloguanil (CYC) inhibiting dihydrofolate reductase (DHFR). Electron transport chain in the mitochondria (atovaquone (ATQ) inhibiting cytochrome B (CYTb) and dihydroorotate dehydrogenase (DHODH). Protein translation in the apicoplast (clindamycin (CLD) and doxycycline (DOX)), and artemisinin (ART), for which variant forms of Pfkelch13 (K13) increase survival measured by RSA$_{(0-3h)}$. Figure taken from (Blasco *et al.* 2017)**.**

1.4.2.2 Quinolines

Resistance to quinolines is generally thought to be related to processes undertaken in and involving the DV. In particular, it has been associated with specific polymorphisms in two main genes; the chloroquine resistance transporter gene (*Pfcrt*) and the multidrug resistance 1 gene (*Pfmdr1*) (Fidock *et al.* 2000). Historically, some quinolines have been used as a monotherapy, however with the emergence and spread of drug resistance, they are mostly now being used in combination with the frontline antimalarial artemisinin, which will be

described further on this chapter. An exception to this is quinine, which has been recently used as monotherapy for the treatment of severe malaria, however there has been some contesting evidence against it, showing that artesunate as a monotherapy might be safer for this purpose (Noubiap 2014). Some of the quinolines that have been used in the field include quinine, chloroquine, mefloquine, piperaquine, primaquine, and lumefantrine, shown inside the DV in Fig. 1.7. Quinine was the first compound to be utilised in a widespread manner for the treatment of malaria, and it dates to the beginning of the 17th century. It was isolated from the bark of the "quina-quina" tree, native to South America. Although there are some legends that indicate that it had been used by the indigenous populations for the treatment of fevers, the evidence is not very strong, at least as it is written in the post-colonial history of the Americas. The trigger for widespread use is alleged to be when a Spanish countess was cured by indigenous people in Peru with the bark of this tree, resulting in the transport of the plant to Europe in 1638 (Achan *et al.* 2011).

Although quinine has been in use for a very long time (Table 1), issues related to toxicity and adverse side effects were always present (Goldenberg and Wexler 1988), emphasising the case for a search for a safer and effective substitute. Although there have been some reports of quinine treatment failure thought to be related to resistance, primarily in South America in the 1960s, other hypothesis as to why it may not be fully effective are related to its pharmacological properties, including poor tolerability by the different age groups treated, however full-scale resistance to quinine is not thought to be present today (Achan *et al.* 2011)**.**

1.4.2.3 Chloroquine

As an alternative to quinine, chloroquine (CQ) was synthesized in the 1940s and its main purpose was to treat malaria. CQ, another 4-aminoquinoline compound, is a weak base that accumulates in the DV because it becomes protonated there due to the low pH environment. Once accumulated in the DV, CQ can block the heme detoxification process, leading to toxicity to the parasite (Lewis *et al.* 2014).

Chloroquine was extensively used after its introduction, as it was highly efficacious and also had pharmacological properties that made it suitable for patients of all ages. It was central

to the WHO global malaria eradication campaign in the mid 1950s, which combined extensive administration of chloroquine with vector control strategies (primarily through DDT), leading to some lower transmission geographical regions being cleared of malaria (Wellems and Plowe 2001). However, its use came with the almost unavoidable consequence of drug resistance, and its success was compromised when there started to be reports of independent emergence of resistant parasites in Southeast Asia and South America. After 20 years, CQ resistance spread to Sub-Saharan Africa, with devastating effects in morbidity and mortality, mostly in paediatric populations (Fidock *et al.* 2000). Figure 1.8 shows the likely trajectory taken by chloroquine-resistant parasites emerging in Southeast Asia, entering Africa through Zanzibar (Schwartz *et al.* 1983).

Figure 1.8 World map showing the emergence of chloroquine resistance. Resistant parasites emerged in the late 1950s in Southeast Asia (marked with a red circle) and South America and subsequently spread 20 years later to areas with high endemicity in Africa. Map adapted from the Malaria Atlas Project from the University of Oxford. The first major study that aimed to understand the mechanism of resistance to chloroquine was published by Wellems *et al.* in 1990. The aim of this study came from the observation that in resistant parasites, chloroquine was being expelled from the DV, therefore limiting efficacy. However, these strains could become sensitive again by the addition of a compound called veramapil (VP), which counteracted the resistance mechanism (Martiney *et al.* 1995). In mammalian tumour cells, the reversal of resistance by VP had been observed in a similar manner, and was mediated by a multidrug resistance gene (*mdr*), an ATP-driven P-glycoprotein transporter. Given the ability to reverse the effect in other biological systems, the study aimed at understanding whether MDR proteins in *P. falciparum* would also be involved in a mechanism of resistance to chloroquine. To isolate the resistance determinants, the authors performed a genetic cross in a chimpanzee between the chloroquine-sensitive *P. falciparum* strain HB3 (CQS) - originally isolated from Honduras - and the chloroquine resistant strain Dd2 (CQR), originating from Laos. The purpose of the cross would be to dissect the genetic basis of the phenotype by studying the segregation of that trait in the progeny. After completing the cross and obtaining the progeny, the two candidate multidrug resistance genes (*Pfmdr1* and *Pfmdr2*) did not seem to be involved in resistance, suggesting the phenotype was linked to different genetic loci. Notably, the progeny of the cross was split into two clear groups:

23

sensitive and resistant, with no intermediate phenotypes (Wellems *et al.* 1990). Restriction Fragment Length Polymorphisms (RFLPs) were used on all the progeny to study the inheritance of all 14 chromosomes, finding a 400 kb region on chromosome 7 to be linked to modulating chloroquine resistance (Wellems, Walker-Jonah, and Panton 1991). This region was later narrowed down to 36 kb by Su *et al.* in 1997 (Su *et al.* 1997).

Later in 2000, Fidock *et al.* looked for transcripts arising from this 36 kb region on chromosome 7, and found previously uncharacterised transcripts coming from a gene they named *Pfcrt,* which was defined as the Chloroquine Resistance Transporter. One finding from these studies was the observation that the resistance phenotype was associated with the presence of at least four independent mutations in the *Pfcrt* gene: K76T, A220S, N326D and I356L. The chloroquine resistant strain Dd2 also had an additional four mutations, which were speculated to compensate for any loss of fitness caused by the resistance mutations, a mechanism that often occurs in pathogens during their process of co-evolving within hosts (Fidock *et al.* 2000). The importance of these mutations has also been supported by genome editing studies that will be described further (Straimer, Lee, *et al.* 2012).

1.4.3 Artemisinin: The current frontline antimalarial

Artemisinin is currently the frontline antimalarial in use. It, or one of its derivatives, has been used as monotherapy or more recently in combination with other antimalarials as Artemisinin Combination Therapies (or ACTs), and has contributed significantly to the recent reductions in the global burden of malaria (World Health Organization 2016).

The origin of the use of artemisinin dates to the year 340 AD. This record comes from a book published in this year by Ge Hong called "Handbook of prescriptions for emergencies". In the text, a plant called *Artemisia annua* L. ("qinghaosu" in Chinese), commonly referred to as 'wormwood' (shown in Figure 1.9A), was extracted in water in order to treat symptoms that are characteristic of a person infected with malaria (Guo 2016).

This plant was rediscovered after the emergence of resistance to chloroquine in the 1950s. With the concern of not having an effective treatment for malaria, a call was made

among research groups in China to examine their traditional medicines for antimalarial properties. With this effort, the creation of Project 523 took place under the leadership of Youyou Tu, who later received the Nobel Prize for Physiology and Medicine in 2015. The project aimed at isolating possible components from all the available collections of herbal materials in order to identify extracts with possible pharmacological properties against *Plasmodium* parasites. Extracts of *Artemisia annua* L. (Figure 1.9B shows the structure of the active extract) showed outstanding properties against different *Plasmodium* species, and this led researchers to extend their work to clinical trials in order to treat *P. falciparum*, the species that at the time was becoming resistant to all other antimalarials introduced in the region (Tu 2011; Schlitzer 2007).

A   **B**
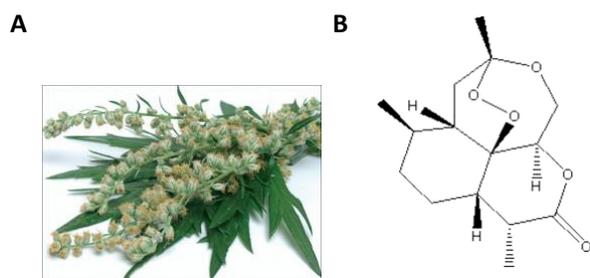
Figure 1.9 *Artemisia annua.*

A. Picture taken from (Nosten 2010). **B.** Structure of artemisinin taken from (Shandilya et al. 2013)

By 2001, the WHO was recommending every country to implement in their malaria healthcare policies a combination of the extract from *Artemisia annua L.* together with another antimalarial (partner drug) (Price and Douglas 2009). The reasoning behind the decision to include a partner drug was that despite the rapid effect of artemisinin on parasite clearance, the drug also has a very short half-life in the plasma. Thus the longer lasting partner drug would act to eliminate any parasites remaining after artemisinin has disappeared from circulation (Eastman and Fidock 2009). Additionally, the principle of using a combination therapy approach for any pathogenic disease is based on the fact that there is a theoretically lower probability of acquiring resistance mutations to multiple, unrelated compounds, than to a single compound, therefore reducing the risk of treatment failure (Rybak and McGrath 1996).

1.4.4 Artemisinin resistance

The WHO recommendation of 2001 regarding use of artemisinin combination therapies (ACTs) was widely taken up and proved effective, but only seven years later Noedl and collaborators showed in clinical trials that *P. falciparum* parasites in Cambodia displayed some level of resistance to artemisinin (Noedl *et al.* 2008).

The study by Noedl *et al.* used monotherapy to investigate the presence of resistance to artemisinin, even though artemisinin was not meant to be used alone for treatment in endemic countries. In their study in Western Cambodia, they enrolled volunteers that were infected with *P. falciparum* but without significant complications. They treated groups of patients with either artemisinin (artesunate) or with quinine and tetracycline (previously described). Artemisinin-resistance was defined by using molecular, *in vivo*, *in vitro*, and pharmacokinetic criteria to distinguish delayed parasite clearance from those patients that cleared parasites at the effective rate. To be defined as failing treatment, patients had to have parasites for more than seven days after the start of the treatment or experienced reemergence of parasites within 28 days of the start of treatment. From the patients included in the study, four out of the 60 patients had reemergence of parasites, and were classified as treatment failure, although two of these were due to drug inefficacy, and only two due to drug resistance (Figure 1.10A). Clinical artemisinin resistance was subsequently defined as the time it would take the drug to clear half of the total parasitaemia in the circulating blood, measured by thick blood smear, combined with a higher IC_{50} to the compound. Figure 1.10B provides additional evidence to their statement by showing a higher inhibitory concentration (IC_{50}) for dihydroartemisinin from the two slow-clearing patients, distinguishing them with the other two patients that had treatment failure but no artemisinin-resistant parasites (Noedl *et al.* 2008).
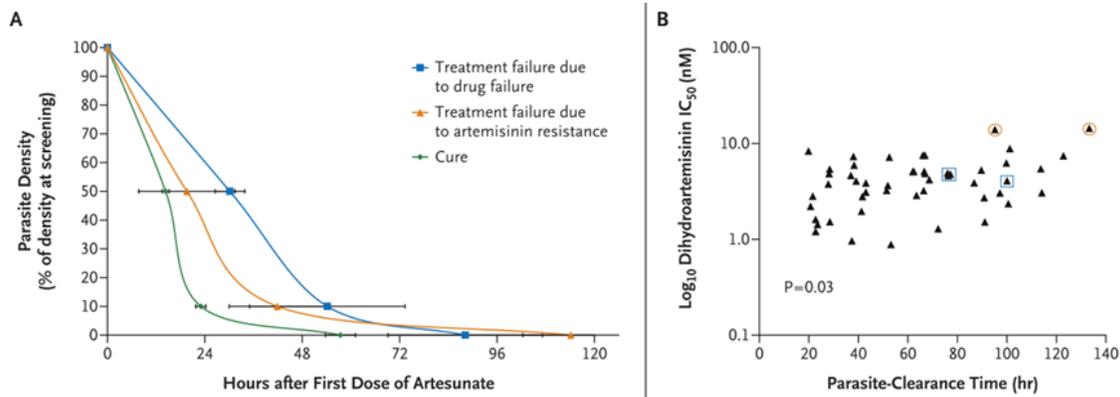
Figure 1.10 Resistance to artemisinin monotherapy. A. Curves showing the reduction of parasitaemia over time for the 60 patients involved in the study by Noedl *et al.* 2008. **B.** Inhibitory concentrations for all patients involved in the study with those circled in red showing patients that had slow-clearance because of resistance and those circled in blue having slow-clearance because of treatment failure. Reproduced from (Noedl *et al.* 2008).

1.4.5 Genetics and epidemiology of artemisinin resistance

1.4.5.1 Population genetics of artemisinin resistance

Following additional studies showing that there was indeed evidence of resistance to artemisinin, various research groups started to analyse whole genomes of parasites isolated from distinct geographical regions to unravel the genetic architecture of what seemed to be an alarming phenotype (Straimer *et al.* 2015).

In 2013, Miotto and collaborators aimed to address the population structure of drug-resistant *P. falciparum* and - given it was a phenotype so far only seen within Southeast Asia - tried to tackle the question of how resistance emerged within these populations and whether it could also emerge in areas of high mortality such as West Africa.

Consistent with the results by Manske *et al.* (Manske *et al.* 2012), analysing the genome sequences of 290 global samples by Principal Component Analysis identified clear population structure within Southeast Asia. This finding was consistent with the presence of low levels of genetic diversity in the area as a consequence of the low transmission rates. Figure 1.11A shows a neighbour-joining tree of all the samples included in the study, showing evidence that

there is a surprising level of structure in Southeast Asia even within geographically close locations. Samples from within Cambodia were analysed to identify their ancestry, with the finding that some parasites (genotype group KH1 comprising Northeastern Cambodia, Thailand and Vietnam) were all more closely related amongst them. However, three other groups of samples were genetically more distant than samples from the region (genotypes KH2, KH3, KH4 from Western Cambodia), and an additional set of examples that accounted for the population with the highest level of admixture (genotype KHA) (Figure 1.11B) (Miotto *et al.* 2013).
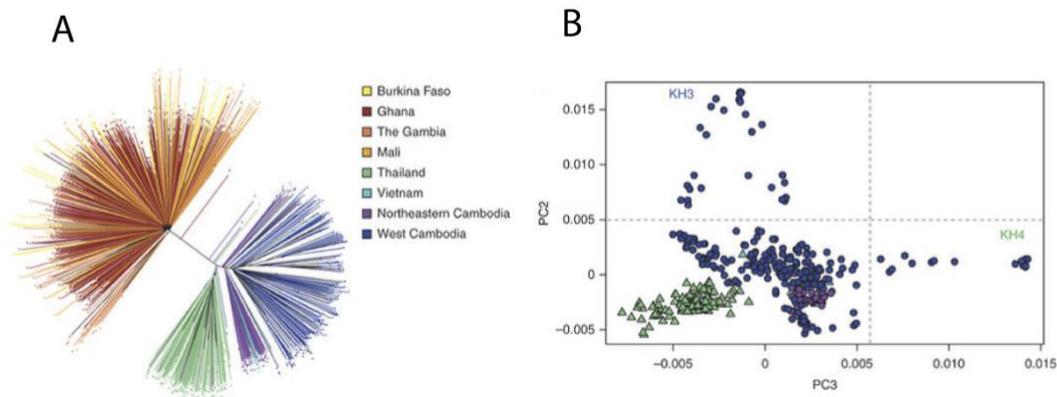
Figure 1.11 Population structure of *P. falciparum*. A. Neighbour-joining tree consisting of 825 samples isolated from malaria patients based on a pairwise distance matrix. The distance between samples represented by mismatches is defined by the length of the branches shows two highly differentiated set of samples corresponding to West Africa and Southeast Asia. A high degree of structure is observed within Southeast Asian samples explained by the distinct clusters in the tree, whereas a less structured population was observed for West Africa. **B.** A principal component analysis showing PC2 and PC3 for the Southeast Asian samples only, showing population structure within samples from Cambodia. Reproduced from (Miotto *et al.* 2013)**.**

This analysis raised the clear question whether these high levels of population structure had anything to do with the different phenotypes seen in the region, in particular, slower parasite clearance after treatment with artemisinin. To answer this question, Miotto *et al.* correlated the values that were available for clearance time of parasitaemia between the different genotype groups (KH1, KH2, KH3, KH4 and KHA). With this analysis, they found that in particular, KH2 had a much higher clearance time than KH1, which they characterised as having a "wild-type" phenotype to artemisinin tolerance. Interestingly, KHA showed a very variable

phenotype, suggesting that this was an admixed population and that perhaps those genomic regions which were responsible for the higher tolerance to artemisinin had been ancestrally acquired.

Amongst the most interesting aspects that were discussed by the authors is the possibility that some of the genotypes (e.g. KH2) had accumulated a particular set of mutations that allowed them to be resistant but also to compensate for a potential loss of fitness as a consequence of drug pressure. These genotypes seemed to be spreading in the area and had a slower clearance rate when treated with artemisinin. If this was the case, there was an urgent need to understand whether specific genetic backgrounds exposed to the same selective pressure could harbour additional mutations leading to the phenotypes observed in the field, and ultimately to lead to a spread of resistance to artemisinins (Miotto *et al.* 2013).

This study led other groups to try to identify genetic markers of artemisinin resistance, and with access to whole genome sequences of the parasite sub-populations identified by Miotto and collaborators, some of these were also experimentally validated.

1.4.5.2 A search for molecular markers of artemisinin resistance

The previous findings were worrying as at the time the whole world was relying on artemisinin for the treatment of malaria. With the improvement of sequencing technologies and genome analysis tools, researchers around the globe attempted to characterise the genetic basis and epidemiology of resistance, now threatening the lives of millions of people. In 2014, a research study used an *in vitro* drug sensitivity assay that correlates well with the clinical measurement of artemisinin resistance, described earlier. The latter defines the time it takes the drug to clear 50% of the circulating parasites in a patient. Together with genomic data, the aim of the study was to identify a molecular marker of artemisinin resistance (Ariey *et al.* 2014). The same group had previously published the Ring Stage Survival Assay (0-3h), or $RSA_{(0-3h)}$. It consists of treating early (0 to 3 hour) ring stages with a concentration that correlates with the concentration used to treat malaria in patients: 700nM of the primary metabolite from artemisinin, dihydroartemisinin (DHA). The usefulness of this protocol was primarily because in artemisinin-resistant isolates from patients in a region in Cambodia, the early rings stages were

resilient to the treatment, resulting in the reemergence of parasites as a consequence of a delayed developmental stage that could be coming from quiescence, or dormancy (Witkowski *et al.* 2010).

In the study of 2014, Ariey *et al.* used a lab-adapted, artemisinin-sensitive strain isolated from Tanzania (F32-TEM) and cultured it with increasing concentrations of artemisinin for five consecutive years. In parallel, a clone was cultured for the same amount of time in the absence of artemisinin pressure as a control. Whole genome sequencing of the two clones was performed before and after the phenotype of the selected line (F32-ART5) reached a 10% increase in the clearance half-life as measured by the RSA$_{(0-3h)}$. The acquisition of mutations in seven genes at around 40 cycles after drug selection was consistent with an increase in resistance measured by RSA of 8-10%. Among the mutations acquired in this period, only those that were present in the drug-selected clone and not in the untreated or in 3D7 were used for phenotype validation. With this, the authors validated their importance in the resistance phenotype of parasites isolated from patients in the field, primarily in regions of the world in which at the time treatment failure was occurring at an increasing rate. Given all the 49 sequenced parasite strains had been culture adapted, they performed *in vitro* RSA$_{(0-3h)}$ on all of them, finding that some of the candidate genes that were observed in their 125-cycle selection as being important for acquisition of resistance also contained polymorphisms in these parasites isolated from patients. In particular, various mutations in PF3D7_1343700 (*Pfkelch13*) were associated with a high-resistance phenotype measured with RSA$_{(0-3h)}$, as seen in Figure 1.12 (Ariey *et al.* 2014). Notably, some mutations had a higher frequency and a slightly lower RSA$_{(0-3h)}$ (e.g. C580Y); whereas others had a higher RSA but their global prevalence was lower (e.g. I543T), at least within the strains that were included in this study, using the wild type strain 3D7 as a control.
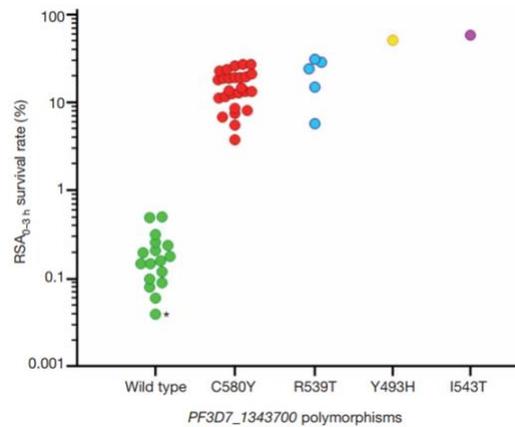
Figure 1.12 A molecular marker of artemisinin resistance. Ring-Stage Survival assay (RSA) performed on 0-3 hour ring stages of culture-adapted strains that had reported a delayed parasite clearance phenotype *in vivo,* and contained mutations in *Pfkelch13*. Reproduced from (Ariey *et al.* 2014)**.**

Overall, the authors confirmed that *Pfkelch13* is an essential molecular marker of artemisinin resistance, and established the phenotype obtained after a long selection experiment with samples isolated from the field. Additionally, the authors showed that the mutation C580Y is the most prevalent. However, there was one *Pfkelch13* mutation from the selection experiment that conferred a higher level of resistance, but is not common as C580Y, arguing that perhaps these mutations were associated with a higher fitness cost, leading to their lower prevalence.

1.4.5.3 *Pfkelch13* and different artemisinin resistant populations

With access to whole genome sequences from many regions of the world, and with the recent establishment of *Pfkelch13* as a molecular marker of artemisinin resistance (Ariey *et al.* 2014), Miotto *et al.* published a study in 2014 aimed at understanding which other genes were involved in resistance pathways in those genotypes that were spreading in Southeast Asia. Partly the importance of this study was to understand the reason behind some mutations in *Pfkelch13* being absent from regions of high transmission intensity such as West Africa. The authors performed a Genome-Wide Association Study (GWAS) in which they examined 1612 samples with phenotypic data *(ex vivo* and *in vitro* RSA$_{(0-3h)}$) coming from Cambodia, Thailand, Myanmar, Laos and Bangladesh. Similar to the results from Ariey *et al.* 2014, they found that

the only mutation in the propeller domain of *Pfkelch13* that had a strong association with the observed phenotype was C580Y, and in regions with some level of resistance, it had almost reached levels of genetic fixation (F_{st} of 1). Interestingly, they were able to identify phenotypic boundaries between subpopulations with different levels of resistance (low vs. high), all of which contained C580Y, and used these subpopulations to dissect some of the components of the genetic background important in the differentiation between medium and high levels of artemisinin resistance, as measured by parasite clearance half-life and $RSA_{(0-3h)}$. The principle of compensatory mechanisms, or stepwise acquisition of resistance-associated polymorphisms has been studied for other candidate molecular markers such as *Pfcrt* (Gabryszewski *et al.* 2016) or *Pfdhfr* (Lozovsky *et al.* 2009*)*.

1.4.5.4 Link of *Pfkelch13* to artemisinin resistance

It is still not clear how mutations in the kelch propeller domain encoded by *Pfkelch13* confer artemisinin resistance. Some studies have linked these polymorphisms to an increase in the levels of the phosphatidylinositol-3-phosphate kinase (PfPI3K), and its lipid product PI3P in ring stages, increasing tolerance to artemisinin (Kahrstrom 2015; Mbengue *et al.* 2015), but studies of populations of resistant parasites suggest that the mechanism of resistance is more complex.

In mammalian cells, kelch proteins have the capacity of acting as cytosolic adaptors of substrates targeted to the ubiquitination pathways as a response to oxidative stress. Keap1 - a kelch-like protein in humans - has been shown to bind to Nrf2 (nuclear factor erythroid 2-related factor 2), which acts as a repressor of Keap1 leading it to the ubiquitination pathway and subsequent degradation by the proteasome. When a cell is in normal homeostatic conditions, Nrf2 is continuously being ubiquitinated, however under stress conditions such as oxidative stress, some residues of the kelch-like protein Keap1 have been seen to be modified by ubiquitin leading to a decline in the activity of one of the ligase subunits which form part of the ubiquitination process, in turn leading to an accumulation of Nrf2 (McMahon *et al.* 2003). This has been shown to have some level of effect on transcription as it would control the

expression of genes related to cell stress response, specifically related to ubiquitination and degradation of proteins by the proteasome (Taguchi *et al.* 2011).

The link between mutations in *Pfkelch13* and resistance to artemisinin has also been studied in the context of how these kelch-like proteins act and affect transcriptional response in human cells. A study published in 2015 by Mok *et al.* showed strong differences in the transcriptional response of artemisinin tolerant and sensitive parasites. In particular, they observed the activation of the unfolded protein response in the resistant backgrounds, potentially related to the decline in activity of the ubiquitin-proteasome system given the mutations present in *Pfkelch13*, similar to what had been observed for kelch proteins in mammalian cells (Mok *et al.* 2015).

There have been other studies aimed at studying artemisinin response in both sensitive and resistant parasites isolated in the field and adapted to culture. A study by Dogovski *et al.* 2015 showed that when wild-type parasites were treated with artemisinin, there was an accumulation of ubiquitinated proteins, again linking the mechanisms of action of this drug to cell-stress response mechanisms. When the authors made the same approach to study the response of resistant parasites to artemisinin they observed that the accumulation of ubiquitination declined, potentially a mechanism that would allow a delayed cell death; an observation that had been described before. To validate their approach, they included in their study a compound that was known to inhibit the proteasome. Interestingly what they saw was that this compound would synergise the effect of artemisinin, suggesting that the mechanism of action of artemisinin directly interferes with the cell stress response pathways and those resistant parasites seem to have enhanced proteostasis mechanisms (Dogovski *et al.* 2015).

Although these studies do not definitively establish how the resistance phenotypes correlate with mutations in *Pfkelch13*, they provide some critical knowledge on the relevance of this protein and the necessity to perform more thorough studies on other candidates on the genetic background of drug-resistant parasites. To understand what these mechanisms might be, some of the available information on similar proteins in human cells and their effect in regulating transcription of genes involved in cell-stress response mechanisms such as the

ubiquitin-proteasome can be studied in various ways. An example can be the work by Miotto *et al.* (O. Miotto *et al.* 2015) which can be a starting point at trying to understand these different mechanisms.

Overall, studying the effect of these haplotypes seen in the field and their effect on parasite fitness, understanding the genetic architecture of drug resistance, and identifying new antimalarial compounds that avoid current resistance pathways are all urgent priorities for the field.

1.4.5.5 Biological fitness and antimalarial drug resistance

The underlying genetic architecture and epidemiological features that explain why resistance emerges in some areas of the world and not in others has been studied through various approaches. Whether mutations in *Pfkelch13* had a direct impact on resistance was experimentally validated by Straimer *et al.* and by Ghorbal *et al.*, using genome editing technologies to introduce polymorphisms previously reported by Ariey *et al.*, and observed in parasite populations in the field. Some of the lines included by Straimer *et al.* would contain some of these polymorphisms in K13 previously shown to be associated with the high tolerance phenotype after artemisinin pressure. Figure 1.12 shows the different *Pfkelch13* mutations that gave high RSA$_{(0-3h)}$ values such as I543T, Y493H and R539T: all of these have a lower prevalence when compared to C580Y, which is widespread throughout Southeast Asia. To study the contribution of each one of these haplotypes, the mutations were removed from parasite isolates from Cambodia which were artemisinin-tolerant. These mutations were also introduced into standard laboratory strains and artemisinin-sensitive field isolates. Fig. 1.13 shows the different backgrounds in which mutations in *Pfkelch13* were introduced. Fig. 1.13A shows the effect of mutations observed in the field on parasite survival using the RSA method. It was clear from this experiment that mutations that had the highest RSA values were present at very low frequencies in the field, as opposed to the most common mutation C580Y. Figure 1.13B shows a different experiment aimed at understanding the contribution of genetic background to resistance phenotypes. Because of its high prevalence, mutation C580Y was introduced into different genetic backgrounds, and RSA$_{(0-3h)}$ values were obtained. Remarkably,

when this mutation was introduced in Cambodian isolates Cam2, Cam3.II and CamWT, and even a lab-adapted strain originally from Laos, Dd2, the survival rate was significantly higher when compared with FCB, a lab-adapted strain originally from Colombia (Straimer *et al.* 2015).
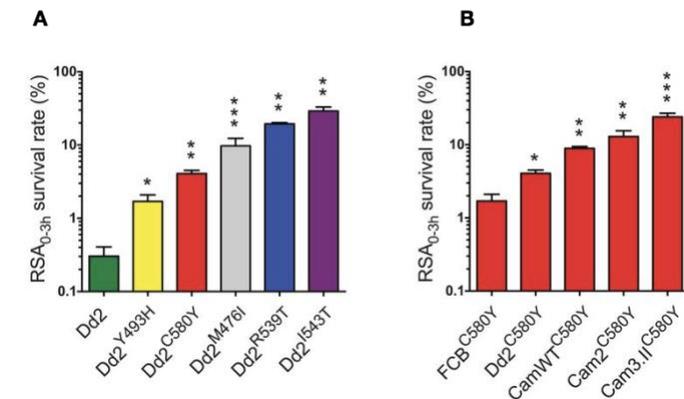
Figure 1.13 ZFN-based editing of *Pfkelch13*. ZFN-based genome editing to address epidemiologically relevant SNPs and their effect on artemisinin resistance. **A.** *Pfkelch13* mutations seen in the field and their effect on artemisinin tolerance in the lab adapted strain Dd2, measured by RSA$_{(0-3h)}$. **B.** The most prevalent mutation seen in the field for *Pfkelch13,* C580Y introduced into different genetic backgrounds and the effect on artemisinin tolerance measured by RSA$_{(0-3h)}$. Reproduced from (Straimer *et al.* 2015).

Evidence from this work supports the idea that the genetic background has an important effect on artemisinin resistance. Understanding the different levels by which mutations in *Pfkelch13* impact the biological fitness of parasite isolates is therefore key to improve surveillance strategies.

1.4.5.6 Mode of action of artemisinin

The mechanism of action of artemisinin is still not fully understood, and is likely to be more complicated than it was previously thought.

Artemisinin is a sesquiterpene lactone with an endoperoxide bridge (Figure 1.9B). There have been studies in the past showing that the breakage of the endoperoxide bridge leads to the release of free radicals which were in part responsible for its antimalarial properties. In 1991, Meshnick and collaborators suggested that artemisinin would get activated into toxic free radicals for the parasite by entering into contact with the heme inside the parasite, which is produced as a result of the digestion of haemoglobin inside red blood cells (Meshnick 2002). This leads to a mechanism of action whereby release of free radicals interferes with the heme detoxification performed by the parasite inside its DV, leading to membrane destabilisation (Antoine *et al.* 2014). Other studies have shown slightly contradictory results, as there were observations of artemisinin accumulating in the parasite membrane rather than the DV, and explicitly showing artemisinins could kill early ring stages which lack heme, therefore in principle the compound could not be activated by the iron present in the process of haemoglobin degradation leading to the production of toxic free radicals. In 2004, Eckstein-Ludwig *et al.* suggested that artemisinins acted by inhibiting the sarco-endoplasmic reticulum $Ca2+$-ATPase (SERCA) ortholog in *Plasmodium falciparum, PfATP6* (Eckstein-Ludwig *et al.* 2003).

Additionally, other studies have suggested artemisinins are activated by the electron transport chain in the mitochondria, leading to its membrane being depolarised and ultimately damaged as a consequence of the toxic free radicals being released upon artemisinin activation (W. Li *et al.* 2005).

Another mechanism that has been described for the mode of action of artemisinin comes from the observation that a small proportion of parasites would enter a state of dormancy, resuming the infection by having a recrudescence after completion of the treatment. Less evidence exists as to how artemisinin enhances this biological state, however, this putative mechanism provides yet another reason why artemisinin should be used in

combination with compounds for which no phenotype of quiescence has been observed (Witkowski *et al.* 2010).

Overall, the mechanism of action of artemisinin is complex, and different studies have shown evidence of it having a pleiotropic effect on the cell, targeting multiple proteins. Given the challenges in validating some of these targets, it is essential to direct efforts towards policies for the use of combination therapies and to invest in large-scale screening of compound libraries to identify new candidates to either use with it, or to replace it.

1.4.6 Artemisinin combination therapies (ACTs)

Artemisinin combination therapies (ACTs) have been used, ever since resistance to artesunate monotherapy became a threat, as the front-line antimalarial. According to the World Health Organisation, it has been one of the significant contributors to the reduction in malaria mortality in areas of high endemicity, attributed primarily to low ACT failure rates (less than 10%) (World Health Organization 2016). Table 1.2 shows the different ACTs that have been used thus far to treat malaria in different regions of the world. Ideally, countries should have availability of as many combinations as possible, given that there have been reports of resistance to the partner drugs currently available, therefore having a more extensive chemical space would be advantageous for any health care policy. Unfortunately, this is not the case, and neighbouring countries have different health care policies and surveillance strategies, which comes with great consequences for control and eventually spread of resistant parasites (Eziefula 2016).

Table 1. 2 Six Artemisinin Combination Therapies are recommended by the World Health Organisation.

Artemisinin derivative	Partner compound
Artesunate	Mefloquine
Artesunate	Sulphadoxine/Pyrimethamine
Arthemeter	Lumefantrine
Artesunate	Amodiaquine
Dihydroartemisinin	Piperaquine
Artesunate	Pyronaridine

1.4.6.1 Resistance to ACTs

Ever since the different partner-drug combinations with artemisinin have been adopted to the different countries, there have also been various reports showing some level of resistance to the partner drug, specifically in the context of artemisinin-resistant parasites from Southeast Asia (Amaratunga *et al.* 2016; Amato *et al.* 2017).

The first combination to be introduced into clinical practice was artesunate and mefloquine in the mid-1990s; resistance to the latter had already been reported in the past, thought to be associated with amplification of a multidrug resistance locus (*Pfmdr1*) (Price *et al.* 2004). Previous studies had seen an increase in mefloquine resistance when used as a monotherapy or in combination with sulfadoxine-pyrimethamine, leading to high treatment failure in Thailand and Cambodia. When mefloquine was used in combination with artesunate, however, a decrease of treatment failure was observed after application of a 3-day dose of artesunate was performed. As a consequence of the effectiveness of the new treatment regime with artesunate, which would be enough for two parasite life cycles, the rate of resistance to mefloquine decreased, meaning that sensitivity of parasites to the drug as measured by the inhibitory concentration (IC_{50}) had increased (Nosten *et al.* 2000; Woodrow and White 2016).

Similar strategies were then used to treat *P. falciparum* malaria worldwide. More recently, there have been reports of treatment failure to one of the most widely used ACTs in the Greater Mekong Subregion (Western Cambodia): dihydroartemisinin-piperaquine. In 2017,

there were two independent studies published which showed an association of specific genetic markers and the treatment failure to this combination therapy. What both studies found after performing a Genome-Wide Association Study (GWAS) was that there was an amplification of the aspartic proteases plasmepsin 2 and 3, localised in chromosome 14 in piperaquine resistance lines. Interestingly, they also found that an amplification in *Pfmdr1* that had been observed and possibly attributed to mefloquine pressure in the past was absent from samples having the plasmepsin 2 and 3 amplification, suggesting that perhaps they represent opposing mechanisms of resistance. The latter is currently being explored in the region and use of triple combination therapies is being encouraged in areas of high resistance (Witkowski *et al.* 2017; Amato *et al.* 2017).

In another study, Dhingra *et al.* had showed the results on a line originating from Indochina (Dd2) which had been selected in their earlier work (Eastman *et al.* 2011) using continuous piperaquine pressure. Comparative studies using Whole Genome Sequencing revealed a novel haplotype in Dd2 for the chloroquine resistance transporter PfCRT, carrying the mutation C101F with a possible association to the phenotype, and a de-amplification of *Pfmdr1* (previously associated with both mefloquine and chloroquine resistance). With this finding, the authors were able to use genome editing tools to validate the resistance phenotype observed. Among their findings, was the fact that if the mutation was inserted into a Dd2 background, it led to both a decrease in susceptibility to piperaquine and a decrease in replication rates, possibly because of a fitness cost associated with that particular haplotype. Additionally, they found that *Pfmdr1* had no impact on piperaquine susceptibility as it has been previously reported, but rather that the de-amplification observed in the field was a result of the removal of mefloquine from use and an expansion of a piperaquine resistance-conferring haplotype occurring in the same genetic background (single copy of *Pfmdr1*) (Dhingra *et al.* 2017).

Given these findings, it is essential to understand the waves of drug resistance that have taken place worldwide by employing large-scale population studies. However, it is equally important to perform validation of polymorphisms identified in the field, specifically in the context of fitness in different genetic backgrounds. Different genome editing tools have been used for this

purpose. This will help to accumulate knowledge into the different pathways and mechanisms that *P. falciparum* uses in the process of acquiring resistance, and the implications for its biological fitness.

1.5 Genome engineering

For many pathogenic organisms, genome engineering has been crucial for understanding mechanisms of drug resistance, identifying essential targets for drug discovery, validation of potential molecular markers of resistance observed in field studies, among other applications. Together with the recent advancements in Next-Generation Sequencing technologies, the field of genome editing has also been growing rapidly, leading for example to genetic screens at the whole-genome level being carried out, and to the generation of large numbers of genetically modified pathogens. These extensive collections have led to subsequent validation of gene essentiality, leading to the identification of many exciting targets. Some recent studies exploring at the genome-scale level the function of genes in *Plasmodium spp.* and *Toxoplasma gondii* have both shed light into the biology of these parasites and also discovered their high degree of essentiality as a potential tool for vaccine development or drug discovery (Sidik *et al.* 2016; Bushell *et al.* 2017; Zhang *et al.* 2018). The case of *P. falciparum* has perhaps benefited most from these technologies, even though still many gene functions remain unknown given the challenges of genetic modification specific to this species.

1.5.1 DNA double-strand breaks and different repair mechanisms

The principle for all genome engineering approaches is to take advantage of the DNA repair pathways of cells that are brought into action as a consequence of double-strand breaks in the DNA. Figure 1.14 shows the two main DNA repair pathways that are activated upon DNA cleavage by a site-specific nuclease.

Induced double strand break by site-specific nuclease

Homology-directed repair (HDR)

Non-homologous end joining (NHEJ)

Deletion

Insertion

Donor sequence

HDR

Integration of region of donor sequence
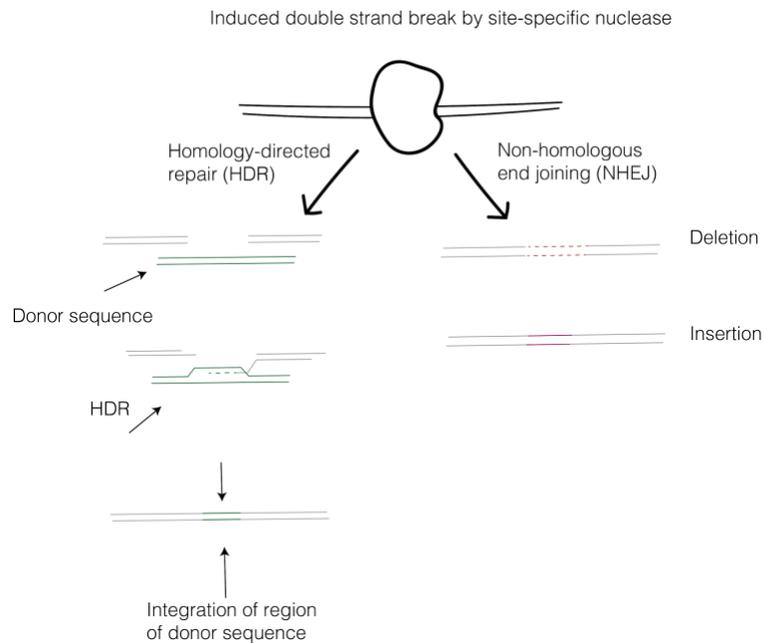
Figure 1.14 Main DNA-repair mechanisms. Two main DNA repair mechanisms that can be undertaken by the cell after a nuclease-induced double strand break. Homology-directed repair (HDR) requires a sequence of homology to be provided (donor sequence) for introducing modifications into the genome. The NHEJ pathway has been referred to as the "error-prone" pathway as it can generate disruptions to the genomic sequence by introducing insertions or deletions as errors generated by the repair machinery. In *Plasmodium sp*., the NHEJ pathway is likely to be absent based on various bioinformatic analysis (Kirkman *et al.* 2014).

Double strand breaks can be extremely damaging for cells and it is therefore essential that cells have mechanisms in place to repair them, in order to allow the cell to copy their genome during DNA replication in an accurate matter. The latter has been shown to include chromatin modifications that allow for enzymes involved in the repair process to access the DNA after the damage has occurred (Dinant *et al.* 2008; Gupta *et al.* 2016). In eukaryotes, Double Strand Breaks (DSB) can be repaired typically by two main DNA repair pathways. The first one is considered as being "error-free" homologous recombination (homology-directed repair or HDR). The other "error-prone" mechanism of repair is by end joining (Non-homologous end joining or NHEJ) (Krogh and Symington 2004).

The HDR DNA repair pathway relies on having a template to repair a DSB in an "error-free" manner. This template can come from a sister chromatid, a segment of exogenous DNA such as a donor plasmid, or another homologous chromosome for the case of diploid organisms

(Davis and Maizels 2014). In 1981, Orr-Weaver and collaborators published a study in the budding yeast *Saccharomyces cerevisiae*, which showed that inducing double-strand breaks in DNA increases the rate of repair by initiating the homologous recombination DNA repair pathway. This study showed that plasmid DNA could be integrated into chromosomal DNA of yeast via this mechanism (Orr-Weaver *et al.* 1981). This finding would later on apply to many eukaryotic organisms, including *P. falciparum*, and HDR has formed the basis of all genetic modification approaches in this organism*.*

By contrast, the canonical NHEJ pathway has been referred to as "error-prone" as it introduces mutations in the genetic code, thereby promoting genetic variability. NHEJ depends on the presence of the proteins Ku70/Ku80 and the DNA ligase IV, in a process which joins the two ends of a DSB together, often introducing or removing small numbers of bases as it does so (Pinder *et al.* 2015). It is important to note that one of the features that became apparent with the sequencing of the *P. falciparum* genome was that it lacked one of the two major DNA repair mechanisms. Several bioinformatic analyses performed during its assembly in 2002 by Gardner *et al.* (Gardner *et al.* 2002) showed that such proteins (Ku70, Ku80 and DNA ligase IV) that act to repair stochastic double-strand breaks of DNA, and are key for NHEJ in other organisms, were absent in *P. falciparum* (Straimer, Lee *et al.* 2012; Lee *et al.* 2014). There are, however, other apicomplexan parasites such as *Toxoplasma gondii* that do have NHEJ, indicating that *Plasmodium* parasites have lost the machinery at some stage during their evolution (Kirkman *et al.* 2014).

The way cells sense a DBS and initiate the repair mechanisms is quite complex and Fig. 1.15 shows the multiple steps undertaken by the machinery of the cell, which has been extensively studied in the budding yeast model. The first steps are of sensing the double strand break and the two fragments are tethered together with the MRX-Sae2 complex (Gnügge and Symington 2017). Once the two ends are close there are two proteins involved in resectioning the 5' to 3' strands. These form part of the STR-Dna2 complex which commits the cell to repair via homologous recombination and essentially protects the cell from the "error-prone" repair pathway. Tails of single-strand DNA (ssDNA) are generated on the 5'-3' direction on both

strands after exonucleolytic processing (by Exo 1). Rad51 is a nucleoprotein that protect the ssDNA tails generated during the process of resectioning described, leading to a loop formation necessary for the process of strand invasion once it recognises the complementary strand; in haploid organisms this is primarily the sister chromatid generated during DNA replication. The process of D-loop formation and strand invasion can take place as a single cross-over, double-crossover or by break-induced replication which is the result of the distal chromosome arm being lost and the rest of the strand being synthesized by the replication machinery (Lee *et al.* 2014; Mimitou and Symington 2011).
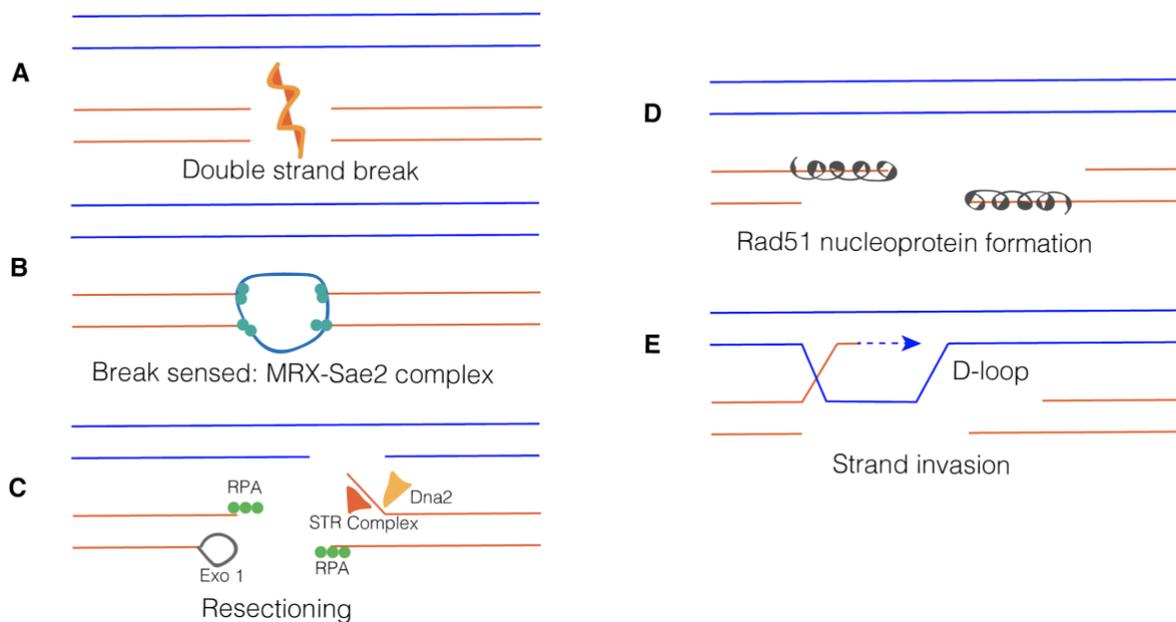
Figure 1.15 DNA repair by homology-directed repair. A. The first step when there is a chromosomal double-strand break, in a typical diploid organism in which the other chromosome is used as a template for repair. **B.** Representation of the sensing step performed by the MRX-Sae2 complex, holding together the two fragments of DNA. **C.** Resectioning of the strands is carried out by Exo1 and the STR complex. The latter step exposes the 3' end and allows for binding of RPA. **D**. RPA is replaced by the Rad51 protein which enables the formation of a nucleoprotein filament (in grey), allowing for the formation of the D-loop and strand invasion to take place (**D-E**). Adapted from Lee *et al.* 2014).

In *P. falciparum* there are not many studies showing the overall mechanisms used for repairing double-strand breaks; rather most studies have focused towards the understanding of the expansion of antigenic variation, acquisition of drug-resistant alleles and overall genetic diversity and evolution. *P. falciparum* chromosomes are arranged by having genes relevant for cell-cycle progression and DNA replication in the centre (core genome), whereas genes important for antigenic variation (clonally variant gene families), which in *P. falciparum* include var, rifin and stevor gene families, are located in the sub-telomeric regions. The latter has been extensively studied given that a large proportion of malaria-associated virulence factors, which allow the parasite to evade the host's adaptive immune system and lead to chronic infections, are due to these gene families. The fact that this chromosomal organisation is maintained despite *P. falciparum* lacking major DNA repair pathways remains puzzling, specifically given the mechanisms by which these antigenic gene families get propagated through recombination

(Claessens *et al.* 2014), and the highly repetitive nature and extremely high AT content of the genome overall. DSBs in the sub-telomeric regions of *P. falciparum* may be repaired through the process of extending telomeres by a telomerase (referred to as "telomere healing") (Calhoun *et al.* 2017), but other questions about genome integrity remain.

Another important aspect that has been widely studied in the context of *P. falciparum* DNA repair is related to the evolution of drug resistance. As mentioned earlier, it has been shown that certain parasite genotypes are more prone to become resistant to antimalarials, whereas others take a longer time to become resistant (e.g. to chloroquine), potentially because of a fitness cost in particular environmental conditions and given their natural history. One hypothesis that has been explored, given the fact that at some point all the resistance phenotypes have been linked in some way to DNA polymorphisms, is that some genotypes are more capable of mutating their DNA (referred to as Accelerated Resistance to Multiple Drugs, ARMD) (Hegreness *et al.* 2008). It has been proposed that these parasites might both become resistant faster and alleviate any fitness cost linked with resistance-conferring polymorphisms. ARMD has been studied in detail in other biological systems such as yeast and mammalian systems, where they seem to be associated with defects in the DNA repair machinery, leading to a higher mutation rate and the likelihood of resistance-associated mutations to propagate more rapidly (Rathod *et al.* 1997).

ARMD was studied in the context of *P. falciparum* in 2016. The transcriptional response after different DNA damaging agents were applied to cultured parasites leading to the observation that when one of these damaging agents such as methyl methanesulphonate (MMS) was applied, the transcriptional response would vary between strains that had been reported to have ARMD (Dd2 and W2) in an earlier study by Rathod *et al.* 1997 (Rathod *et al.* 1997). The ARMD strains were proposed to contain the genetic background to tolerate de novo mutations associated with drug resistance in higher frequencies than the two non-ARMD strains included in the study, 3D7 and D6, which can remain susceptible to compounds for longer periods of time. What the authors found was that ARMD strains had a failure in response to DNA damaging agents such as MMS, as they were not upregulating genes involved

in repair pathways that were seen in the non-ARMD, as confirmed by transcriptomics analysis using microarray technologies. The authors collected genomic DNA at the same time points as they isolated RNA, and after performing Whole Genome Sequencing they found that 18 genes involved in DNA repair had non-synonymous mutations. This helped to confirm the idea that parasites becoming resistant to antimalarials can harbour polymorphisms in genes involved in DNA repair mechanisms, increasing the likelihood of *de novo* mutations to emerge and be maintained given specific genetic backgrounds, as in the case of the multidrug-resistant W2 and Dd2 (Gupta *et al.* 2016).

Even though this and other studies provided some evidence of higher mutation rates in certain *P. falciparum* backgrounds, it does not exactly extrapolate to parasites directly isolated from the field and therefore it does not necessarily explain how resistant parasites emerge and spread globally.

1.5.2 Genome editing technologies

Understanding of the function of genes crucial for essential biological functions can be achieved by both forward and reverse genetics. In addition to this, introducing single polymorphisms identified from large-scale genomics studies as potentially having an important role in a specific biological function is an important tool. Editing technologies are also powerful as they allow the study of genes known to be essential - as it has been shown that in many cases polymorphisms might be enough to explain a phenotype, without inactivating the genes (Winzeler *et al.* 1999).

Forward genetics is when a phenotype of interest is chosen, and the genetic underpinning of that phenotype is explored. The use of genetic crosses in *P. falciparum,* used as described above to uncover the genetic basis of chloroquine resistance, is one example of forward genetics. Another is the use of random mutagenesis, which has been carried out in *P. falciparum* by using transposable elements. Zhang *et al.* 2018 took the approach of using the transposable element piggyBac, that inserts at the nucleotide target site TTAA, which, taking into account the high AT content of *P. falciparum*, allows the theoretical possibility of multiple

insertions into every gene in the genome. Many mutants were generated. These approaches end up being highly useful, especially in an organism that is not very genetically tractable such as *P. falciparum* (Zhang *et al.* 2018).

Reverse genetics takes the opposite approach by studying the function of specific genes in order to identify the possible phenotypes that might occur after these genes are disrupted. Generally, it requires prior knowledge of the genomic sequence of the organism of interest to design specific targeting vectors with sufficient homology to the target of interest. For this reason, the field of reverse genetics has been growing exponentially since access to sequencing technology has become more affordable and ubiquitous in many research environments around the world. As discussed earlier, in *Plasmodium* parasites there have been challenges that have made the technologies lag behind other organisms, even when compared to other Apicomplexan parasites such as *Toxoplasma gondii* (Sidik *et al.* 2016). Stable transfection technologies were developed in the 1990s, and the genome was sequenced in the early 2000s, but the functions of the majority of *P. falciparum* genes remain unknown. The discovery of gene functions by reverse genetics has been slow, with less than 10% of the total genes examined (Sanderson *et al.* 2017), and a generally slow trajectory of new genes investigated every year, with the only exceptions being two significant spikes in reported phenotypes in 2008 when a knockout screen was performed on exported proteins in the asexual cycle (A. G. Maier *et al.* 2008), and a knockout screen of the *P. falciparum* kinome in 2011 (Solyakov *et al.* 2011).

1.5.2.1 The *Plasmodium* Modification Project (*Plasmo*GEM)

Plasmodium berghei has been a more genetically tractable organism than the human malaria parasites largely because it has a much higher transfection efficiency than *P. falciparum*, and as a consequence of this more gene functions have been studied in this species. A significant contribution has come as a result of the establishment of the *Plasmodium* Modification Project (*Plasmo*GEM), which has attempted to overcome the limitations of transfections for scalable approaches to reverse genetics.

The *Plasmo*GEM resource is comprised of a library of targeting vectors available for use for reverse genetics screens in *P. berghei*. They were generated by using the low copy linearised

plasmid pJAZZ-OK from (Godiska *et al.* 2010) to insert *P. berghei* genomic DNA libraries. After end-sequencing to identify the fragments contained within any given clone, recombineering technology from bacteriophage N15 was used to convert these pJAZZ-OK clones into targeting vectors. The significant advantage of these vectors is the fact that they exist inside *E.coli* as a low-copy linear prophage with covalently closed hairpin ends, which allows large DNA fragments, including higher % A+T sequences and repetitive sequences, to be maintained inside the bacterial cell (Ooi *et al.* 2008).

Generating vectors compatible with *Plasmodium* was performed initially in three steps. The first step was to generate the size-selected gDNA library in pJAZZ-OK linear vectors derived from bacteriophage N15, with an average insert size of 9kb, and covering almost the full genome of *P. berghei*. This library was then used for generation of knock-out vectors using homologous recombination in *E. coli*, expressing the temperature-sensitive plasmid pSC101gbaA-tet: encoding the bacteriophage lambda Red-ET. The first step was to replace a sequence of interest by the positive/negative selectable cassette zeo-pheS using short homology arms as overhangs in the primers used to amplify the target of interest, and flanked by gateway sites. This is used as an intermediate vector for the gateway reaction for generating the final vector containing the positive/negative selectable markers hDHFR/yFCU, shown to work for *Plasmodium* parasites (Pfander *et al.* 2011). Using this efficient recombineering technology, Gomes *et al.* 2015 published a study in which an 11bp-long unique barcode was introduced to the final vector to track individually mutated parasite clones over time using whole genome sequencing approaches (Gomes *et al.* 2015). Figure 1.16 shows the generation of the knockout targeting vectors. The initial linear vector from genomic DNA library previously described, flanked by NotI restriction sites, was replaced with zeo-pheS cassette flanked by sites attR2 and attR1 for gateway recombineering. Finally, a unique barcode was used for the second step in a gateway reaction to introduce the positive and negative selectable marker for selection of mutant clones of *P. berghei.*
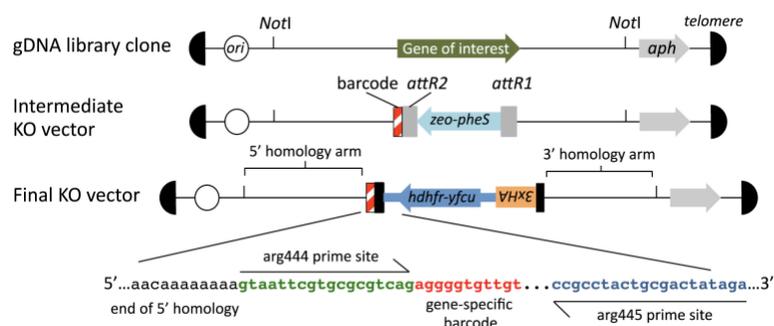
Figure 1.16 PlasmoGEM vector construction from gDNA library. Schematic representation of the different steps in the generation of the final knockout vector. The gDNA library contains a region of homology to gene of interest flanked by NotI sites. Intermediate vector contains the zeo-pheS cassette flanked by attR1/attR2 for gateway recombineering. The final vector contains two homology arms to gene of interest for replacement with positive/negative selection cassettes, a unique barcode and constant primer binding sites arg444/arg445 flanking amplicon for NGS. Reproduced from (Gomes *et al.* 2015).

1.5.2.2 Barcode sequencing

The use of barcodes in these *P. berghei* modification vectors is linked to the development of a new reverse genetics approach referred to as Barcode Analysis by Sequencing (BarSeq). This technique was originally developed for the study of gene function in yeast. In this work, the authors selected mutants for two well-characterised drug targets to grow each strain in a pool under drug pressure. As each strain was individually barcoded, they were able to amplify the 20-mer barcode with primers containing the required sequence for Illumina NGS, and fitness of particular mutants was confirmed, outperforming their previous method using barcode microarray assay, for this specific drug target (Giaever and Nislow 2014; Smith *et al.* 2009).

The *Plasmo*GEM team adapted BarSeq technology to establish a scalable pipeline for doing experimental genetics in *P. berghei. Plasmo*GEM vectors carry a unique 11bp barcode which does not exist anywhere else in the genome, next to annealing sites for oligonucleotides containing standard Illumina sequences for BarSeq. These unique barcodes allow multiple barcoded mutants to be generated as pools in a single transfection, with BarSeq being used to track the relative growth rate of each mutant within the pool. As a proof of principle, protein

kinases were studied and those that had been previously shown to be essential were confirmed by this approach (Tewari *et al.* 2010; Gomes *et al.* 2015).

The application of BarSeq in *P. berghei* was extended to half the genes in the genome in 2017 by Bushell and collaborators. In this pioneering study, barcode sequencing was used to determine growth of 2578 individual mutants. This study found that in this parasite species, the proportion of essential genes was surprisingly high, as compared to other members across the evolutionary tree of Apicomplexan parasites, and the evolution of parasitism. Additionally, it shed light on a number of potential antimalarial drug targets, given their high degree of essentiality (Bushell *et al.* 2017). This study showed the power of these kinds of approaches for studying gene functions, and the value of using genetic modification in a large-scale matter.

1.5.2.3 Nuclease-mediated genome editing

So far, large-scale reverse genetics screens in *P. falciparum* have not been performed. It is not clear why there exists such a big difference in transfection efficiency across *Plasmodium* species, with *P. berghei* having a much high efficiency than *P. falciparum*, leading to considerably fewer genes being targeted and phenotypes studied in the human parasite (Bushell *et al.* 2017).

Technologies such as nuclease-mediated genome editing have accelerated genome modification in many organisms and have helped in the identification of functions of genes that were previously unknown. These site-specific nucleases work by inducing a DNA double-strand break at a target locus and thus engaging the DNA repair machinery, as discussed previously.

Different site-specific nucleases have been identified such as Zinc Finger Nucleases (ZFNs), TALENs and more recently CRISPR-Cas. Both ZFN and CRISPR-Cas have been used in *P. falciparum* and are illustrated in Fig. 1.17, together with the conventional allelic exchange which was introduced in the 1990's and has been used most extensively (Crabb *et al.* 1997; Wu *et al.* 1995).
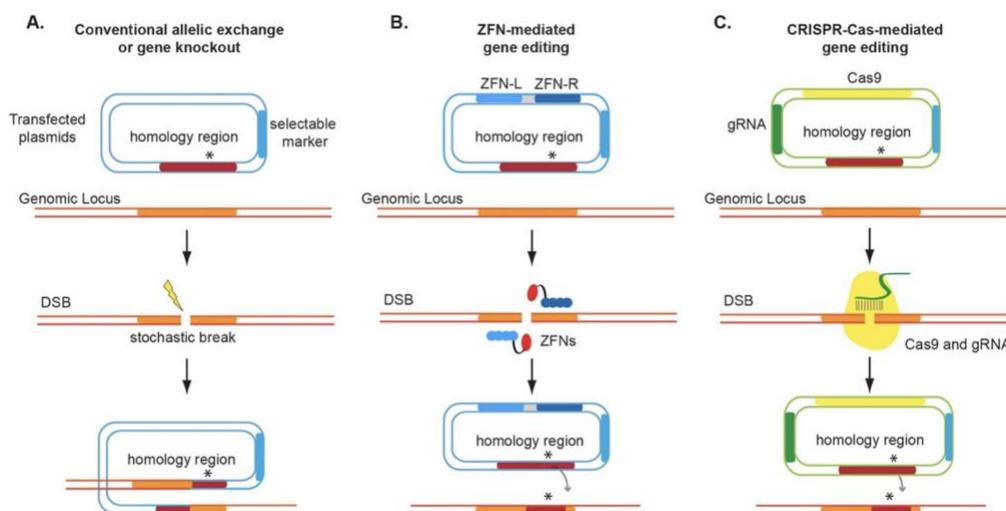
Figure 1.17 Genome editing in *P. falciparum*. Schematic representation of some of the different strategies that have been used to modify the genome of *P. falciparum.* **A.** Conventional approach that was used since the 1990s, which would likely be triggered by a stochastic DNA break near the site of interest (marked with an asterisk), thus forcing the repair machinery to use a homology-directed repair. Plasmid integration was stimulated by providing a sequence of homology on the plasmid. Given the stochasticity of the process, it could take up to six months of continuous culture under drug selection to recover mutant parasites. **B-C.** Two different approaches of genome editing by using directed site-specific nucleases to target a locus of interest for subsequent double-strand break and repair. As in **1.17A,** a construct with homology to the target is provided. Zinc Finger Nuclease-based genome editing relies on the heterodimerization of two zinc finger proteins (ZFN-L and ZFN-R) fused to the FokI restriction enzyme (shown in red), whereas the CRISPR-Cas system consists of the nuclease Cas9 bound to a single gRNA that has homology to a region of the genome thus forming an RNA-DNA structure for double strand break. Reproduced from (Lee and Fidock 2014).

ZFNs consist of binding domains from transcription factors and a FokI restriction enzyme with nuclease activity. The binding domains provide a unique specificity to a DNA sequence, generating a heterodimer, and then inducing a double-strand break through the FokI activity (Urnov *et al.* 2010). By providing a donor sequence with enough homology, homologous recombination takes place at the site of the ZFN induced DSB, leading to gene editing or disruption. While ZFNs can be very efficient, the design of the constructs is challenging, and the costs are high as a unique ZFN needs to be designed for every gene. This limits the potential of ZFNs to become scalable to a genome-wide approach (Porteus and Carroll 2005).

ZFNs have been used in *P. falciparum* to validate candidate markers associated with drug resistance. To date, multiple loci have been targeted using this technology, including

relevant genes involved in mechanisms of drug resistance, including *Pfcrt* (Straimer, Lee, *et al.* 2012), *phosphatidylinositol-4-OH kinase* (McNamara *et al.* 2013), and more recently a molecular marker associated with increased tolerance to artemisinin, *Pfkelch13* (Straimer *et al.* 2015). Despite being a useful technology for the implementation of nuclease-driven homologous recombination, the rate of successful recombination relies on the cleavage activity of ZFNs, the distance to the site where the polymorphism is being incorporated, and the length of the homology of the donor sequence. From the previous studies by using different nuclease-induced strategies, it has been established that longer arms of homology correlate with the efficiency of recombination, and given the difficulty of successful cloning of long stretches of an AT-rich genome such as *Plasmodium*, into *E. coli,* this becomes challenging.

Like ZFNs, Tal effector-like nucleases (TALENs) have also been shown to be a powerful tool for genome editing in many organisms. TALEN design is more straightforward than ZFNs and could potentially increase throughput, however perhaps because it contains a highly repetitive binding domain of 34-amino acids, and the large size of the nuclease, there is no published evidence of this method being used for *P. falciparum* (Lee and Fidock 2014).

1.5.2.4 CRISPR/Cas9-based genome editing

Recently CRISPR/Cas9 has emerged as an easily programmable and affordable genetic modification tool for many organisms, and is now by far the most widespread genome editing tool in general use (Cong and Zhang 2015). In the last few years, it has also been shown to work in *P. falciparum*, which is very promising as reverse genetics now seems feasible and scalable for tackling different biological questions (Ghorbal *et al.* 2014; Wagner *et al.* 2014).

This platform is based on a bacterial CRISPR (Clustered regularly-interspaced short palindromic repeats) and an associated protein Cas9 (Figure 1.18). As part of the immune mechanisms used by many bacterial species to protect themselves from foreign DNA, they incorporate foreign nucleic acids between CRISPR arrays within their genome, containing a unique sequence for recognition by the gRNA-Cas9 complex. The subsequently transcribed

CRISPR RNAs will contain "protospacers" that are complementary to the foreign sequences. This transcript will then couple with a second RNA (transactivating RNA), and after processing by an RNase, will form a complex with the nuclease (Cas9). The Cas9 nuclease will target and cleave the foreign DNA, upstream of a Protospacer Adjacent Motif (PAM) (of variable sequence depending on the CRISPR system). The way this natural mechanism has been adapted to genome editing in other organisms is by using a fusion of the CRISPR transcript with the protospacer or sequence of homology to the genome, coupled with part of the transactivating RNA. The resulting complex comprises a 20-nucleotide single guide RNA (gRNA), which directs the nuclease to the target gene (Sander and Joung 2014).

Figure 1.18 Cas9/gRNA-binding.

Figure illustrating the binding of the gRNA/Cas9 complex, upstream of the recognition motif (PAM). The first 20 nt of the gRNA is complementary to the target DNA sequence (in red), directing the Cas9 nuclease to the target site and after unwinding of the target and the formation of a gRNA-DNA hybrid, leading finally to a double strand break. Reproduced from (Sander and Joung 2014).
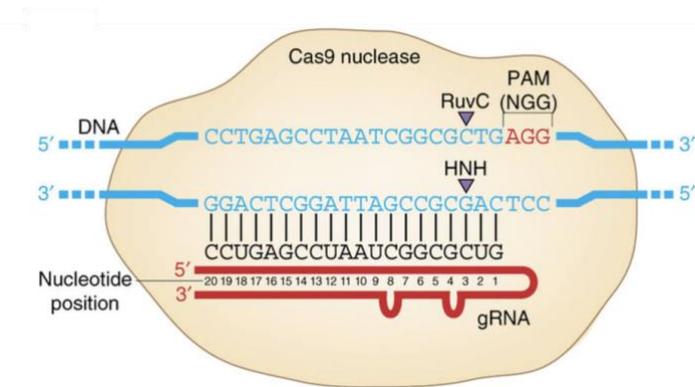
This technology has now been adapted to *P. falciparum* with various approaches, the general method being to provide the parasite with plasmids bearing the Cas9 and gRNA under the control of specific promoters. Two different RNA polymerase III-based promoters have been used to transcribe gRNAs, which rely on a defined transcription start site to produce an accurate gRNA. In the study by Ghorbal *et. al* 2014, a U6 upstream region was used, whereas T7 RNA polymerase, derived from bacteriophage T7, was adapted in a study by Wagner *et al.* in the same year (Wagner *et al.* 2014).

Ghorbal *et al.* used an approach in which two plasmids were co-transfected. The first plasmid carried the Cas9 along with the selectable marker yDHODH, and the second plasmid

carried the gRNA and the donor sequence of homology to the target site for repair by homologous recombination. The U6 small nuclear RNA (snRNA) requires the RNA polymerase III in order to get transcribed, initiating with a guanosine nucleotide as was established in the original CRISPR-Cas9 paper by Ran *et al.* (Ran *et al.* 2013). However, Ghorbal *et al.* were able to achieve expression of the gRNA without the initial G, which for the genome of *P. falciparum* would be advantageous given the low G-C content (less than ~20% in the genome), and the fact that the *Streptococcus pyogenes* Cas9 has an NGG PAM requirement. Fig. 1.19 shows the plasmid design used by Ghorbal *et al.* which consisted of the pUF1 plasmid driving the expression of Cas9, flanked by two nuclear localization signals (NLS), by *Plasmodium*-specific promoters and with the drug-selectable marker y*dhodh.* The second plasmid contains the sgRNA expression cassette and the donor sequence for generating a knockout, with homology arms flanking the positive drug-selectable cassette h*dhfr*. This vector also contains the negative selectable marker y*FCU*. In the same study, the strategy was used to generate single mutations in the genome by providing a region of homology that contains the mutations of interest, and in this particular case the system was used to validate the importance of the *Pfkelch13* mutation C580Y, increasing tolerance to artemisinin in the lab strain NF54.
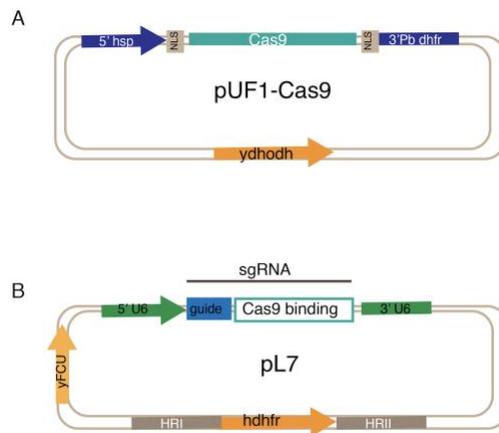
Figure 1.19 Two plasmid CRISPR/Cas9 approach for *P. falciparum*. A. Schematic representation of the Cas9-expressing plasmid, flanked by nuclear-localisation signals and a 5' *hsp* promoter and 3' *Pbdhfr* terminator sequence. A positive selectable marker yDHODH is also present in this plasmid. **B.** Plasmid containing the sgRNA expression cassette using the U6 promoter for RNA Polymerase III. Homology regions shown are flanking a positive selection cassette hDHFR for generating gene knockouts and also a negative selectable marker yFCU is present. Reproduced from (Ghorbal *et al.* 2014)**.**

Overall, recent advances in genome sequencing technologies and the implementation of genome editing technologies in disease-causing pathogens such as *P. falciparum* have added new layers to the study of infectious diseases. There is now the need to use these tools at scale and understand, almost at real time, what the changes in the genome of this parasite mean and what the implications are for surveillance and for containment of alleles that might be modulating drug resistance to current or future antimalarial compounds, and what their implications are for parasite fitness in the different populations worldwide.

1.6 Project Aims

My project will entail developing a method to perform parallel phenotyping in *P. falciparum* using BarSeq and Next Generation Sequencing in order to explore epidemiologically relevant resistance associated alleles and their effects on parasite fitness. For this, the following specific aims were established, which are organised into three independent chapters:

1. The first chapter will aim at establishing a barcode sequencing approach for *P. falciparum* in order to track parasite growth *in vitro*, as previously developed for *P. berghei*. In addition to this, transfection efficiency will be quantified with BarSeq by using pools of barcoded vectors, with the limitations for *P. falciparum* being discussed in the context of a growing field of genome-scale CRISPR/Cas9 screens.

   i.   First, a complex pool of 96 uniquely barcoded vectors will be generated aimed at performing experimental genetics at scale in *P. falciparum*
   ii.   Pools of vectors will be used as readout of transfection efficiency in *P. falciparum* by comparing the barcode complexity in the input pool and output post-transfection
   iii. A constant genomic region (*PfRh3*) will be evaluated as a safe harbor for barcode insertion by using CRISPR/Cas9-based genome editing

2. The second results chapter aims at adapting BarSeq to *P. falciparum* to study both the inherent fitness and drug response of different antimalarial compounds of a panel of strains harbouring sufficient genetic diversity and drug-resistance profiles.

   i.   Libraries of uniquely barcoded parasites will be generated using the pools described in chapter 3
   ii.   Different genetic backgrounds will be used in order to cover sufficient genetic and phenotypic diversity for developing a proof-of-principle of BarSeq in *P. falciparum*
   iii. Both inherent fitness and drug response of the various strains will be measured and compared with standard inhibitory concentration assays

3. The last results chapter will be centered around the contribution of specific *Pfkelch13* alleles generated with CRISPR/Cas9 have to parasite fitness and artemisinin resistance. In this chapter, similar approaches as in other results chapters will be used for measuring competitive parasite growth in vitro, and the epidemiological relevance of these results will be discussed.

i. First, CRISPR/Cas9 editing vectors will be built in order to efficiently edit *Pfkelch13* in a representative strain belonging to the KEL1 recently described in Amato *et al.* 2018. Additional lab-adapted strain 3D7 will be used as a non-KEL1 strain.

ii. Edited lines will be used to perform competition assays to measure the impact of variant C580Y in *Pfkelch13* has on parasite fitness measured as *in vitro* growth.

iii. CRISPR/Cas9 approaches will be used to mimic a double strand break at position 580 in *Pfkelch13*, and by providing a pool for homologous-directed repair containing a donor sequence with homology to the locus and at position 580 any codon to change to, the experimental outcome of which codons were tolerated will be studied in the context of different genetic backgrounds.

Chapter 2: General Methods

2.1 *In vitro* cultures of *Plasmodium falciparum*

2.1.1 Parasite culture maintenance and synchronization

For routine culture of parasites adapted to culture, various parasite strains were used (Table 2.1), and the protocol established by Trager and Jensen in 1976 was followed (Trager and Jensen 1976). The different parasite strains were routinely cultured in O+ RBCs provided by anonymous healthy donors from the National Health Services (NHS), and kept in incubators at 37°C and routinely gassed with a mix with concentrations of 1% O_2, 3% CO_2 and 96% N_2. The blood was previously depleted of leucocytes, washed twice in Roswell Park Memorial Institute Media (RPMI), and resuspended at 50% haematocrit in a medium provided by Sigma-Aldrich containing RPMI-1640 supplied with HEPES, Albumax II, gentamicin, glutamax and supplementation with 20mM HEPES. Unless otherwise stated, this was the method for media preparation. Certain strains required a modification to the culturing protocol and the medium was supplemented with human serum, or certain drugs were added for various purposes, mainly for selecting transfections or performing drug assays. The routine measurements relevant for the maintenance of the cultures were parasitaemia and haematocrit which are described below.

Parasitaemia refers to the number of parasite-infected Red Blood Cells (iRBC) with respect to the total parasite population. The total parasite count can be from a blood smear taken from a malaria-infected patient (Warhurst and Williams 1996), or it can be from a routine laboratory *in vitro* culture. For the experiments throughout this work, counts for culture maintenance were performed in at least five different fields and the parasitaemia obtained by dividing the number of iRBCs over the total count of uninfected Red Blood Cells (uRBC).

Haemeatocrit refers to the percentage of red blood cells that are in the total volume of culture. haematocrits can vary but normally they are in the range of 1-5% for standard *in vitro* cultures.

2.1.1.1 Culture maintenance

In order to maintain viable and healthy parasite cultures, parasitaemia was maintained at less than 5% and hematocrit throughout all the experiments performed was kept between 3-5%. The latter is a standard parasite culture media that has not been overly modified since first described by Trager and Jensen (Trager and Jensen 1976). Depending on the parasite strain (which will be specified accordingly), cultures were diluted when the parasitaemia would reach these percentages. For field isolates the range of parasitaemia was between 1 and 5% in order to avoid parasites under high stress which has been shown to lead to an abnormal production of gametocytes (Carter *et al.* 2013). Other parasite strains that have been long-adapted to laboratory conditions were kept at parasitaemia below or above the mentioned range used for field isolates.

Parasitaemia was measured by resuspending parasite cultures at ~50% haematocrit for homogenization and to avoid aggregation of cells. A drop of the suspension of approximately 2uL was used to perform the blood smear that was later stained with 10% Giemsa solution (Sigma-Aldrich) for 10 minutes and observation at 100X magnification with a light microscope was performed. For a more precise measurement of parasitaemia, flow cytometry was used as in section 2.4.2.1.

2.1.1.2 Parasite synchronization

Parasite synchronization is needed in order to overcome the challenges of the asynchrony reached by parasites in culture, in which only a few life cycles are required before losing synchrony, as reported by Trager and Jensen (Radfar *et al.* 2009; Trager and Jensen 1976). Several methods have been described that allow parasite synchronization.

Sorbitol at 5% in water and sterile filtered was used mainly for enriching for ring stages and reducing the time window of parasites in the culture. According to a modified version of the standard protocol developed by Lambros and Vanderberg in 1979 (Lambros and

Vanderberg 1979). Parasites are centrifuged at 1100g for 5 minutes on either 15 or 50mL Falcon tubes in order to remove the media. The remaining parasite pellets were resuspended in 7 volumes of 5% sorbitol and left at 37°C inside the incubator for 5 minutes. After incubating the falcon tubes were centrifuged at a lower speed (800g for 5 minutes), sorbitol was removed with a vacuum aspirator and a washing step was performed with complete media. The washed pellet was then resuspended at the appropriate haemeatocrit.

A different synchronization protocol was performed by using a Percoll gradient, attempting to isolate both late stages, segmenting schizonts or rings whilst avoiding cellular stress from sorbitol treatment. This was particularly relevant for highly sensitive strains such as recently adapted field isolates to enrich for later stages, usually late trophozoites or schizonts (Saliba and Jacobs-Lorena 2013), and for narrowing the stage window required for particular experiments. Percoll purification used throughout this work was a 63% solution prepared in 10X PBS and homogenized with RPMI-1640 as previously described in (Blackman *et al.* 1990). Parasite pellet was corresponding to a 50mL culture flask at 3% haematocrit was resuspended in a maximum volume of 5mL of complete media and loaded above the percoll solution, avoiding mixing of the two phases created with the loading. The percoll gradient at this stage was centrifuged 1300g for 11 minutes with no brake in order to get the gradient and isolate the required stages. Figure 2.1 shows the graphical representation of the gradient described, and the layers that get separated at this concentration.
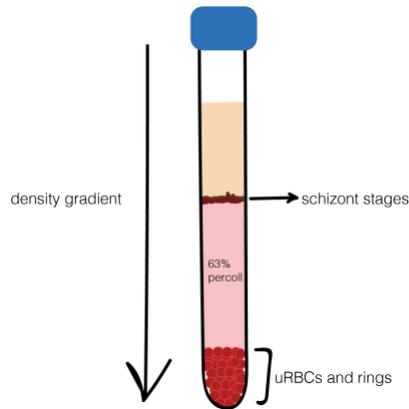
Figure 2.1 Density gradient using a 63% v/v of Percoll solution. Used for separation of late stages of *P. falciparum* based on density. Uninfected red blood cells (uRBCs) will settle at the bottom of the 15mL falcon tube, layered on top the percoll solution and the media from the culture. In between these two phases, a thin layer (brown) will contain the late stages.

Additional synchronization steps were performed using a magnetic cell separation (MACS). The latter was useful for enrichment of late stages (mid-trophozoites and schizonts). Cultures were centrifuged for 5 minutes at 1100g and resuspended in complete media before loading them through the magnetic columns from Miltenyi Biotec. After cultures have passed through the column, the late stages were retained with the magnet and the uninfected red blood cells went through together with the rings. Once this was completed, the column was removed from the magnet and the parasites captured on the matrix of the column based on their hemozoin content can be diluted with standard RPMI 1640.

2.1.2 Thawing and freezing parasites

Table 2.1 shows all the parasite strains used throughout this dissertation. Frozen parasite stocks were kept in vials either in a -80 freezer or in liquid nitrogen. The vials were thawed at room temperature and transferred to a 50mL falcon tube. 1/10th of 12% NaCl (prepared from stock from Sigma-Aldrich) per volume of pellet was added to the thawed cells while shaking the falcon tube gently. After adding this solution, the parasite pellet was left at

room temperature for two minutes. Ten times the pellet volume of 1.6% NaCl was added slowly to the cells while agitating them slowly as with the previous solution. Once this solution has been added, the falcon tubes were centrifuged at 800g for 5 minutes and the supernatant was removed using a vacuum or a strip pipette. After removing the media, the parasites were resuspended dropwise in the third solution which was 1.2% NaCl and 2% glucose, also by agitating slowly the tubes. Another step of centrifugation with the previous settings was conducted and a posterior washing step using standard culture media were performed before finally resuspending the pellet in 10mL of culture media and adding 50% RBCs to reach a final haematocrit of 3%. All solutions used for the thawing process were pre-warmed inside one of the 37°C incubator and at the point of preparation they were sterile filtered using a 0.22 uM membrane filters.

Ideally, parasites were frozen down once a 2% parasitaemia was reached and the majority of the parasite population was at the ring stage. For freezing down cultured parasites, a glycerolyte solution was used using 28% glycerol, 3% sorbitol, 0.65% NaCl and was sterile filtered through a 0.22uM membrane filter. Parasite cultures were centrifuged at 1100g for 5 minutes and pellet was resuspended in 2 times their volume of the glycerolyte solution described. This protocol was modified from an original described by Radfar *et al.* 2009 (Radfar *et al.* 2009).

Table 2.1 Parasite strains used, origin and year when they were adapted into culture. Table includes the references to the strain in the fourth column (Walliker *et al.* 1987; Udeinya *et al.* 1983; Guinet *et al.* 1996; Lee and Fidock 2016; Hadley *et al.* 1987).

Strain	Origin	Year (adapted)	Reference
3D7	Netherlands (I)	1970s	Walliker et al. 1987
V1/S	Vietnam	1980	Udeinya et al. 1983
PH0212-c (CAM)	Pursat, Cambodia	2010	Lee and Fidock 2016
Dd2	SEA-Laos (II)	1982	Guinet et al. 1996
7G8	SA-Brazil (III)	1980	Hadley et al. 1987
3D7Goldberg	Netherlands (IV)	1970s	Walliker et al. 1987

(I) probably West Africa, clone of NF54 isolated in the Netherlands

(II) clone of W2

(III) clone of IMTM22

(IV) 3D7 from Goldberg lab, 38 hour life cycle

2.1.3 Cloning of parasites by limiting dilution

In order to generate isogenic parasite lines either from culture strains or from transfections for various experimental procedures, parasites were cloned using a limiting dilution protocol. The method was modified from an already described protocol (Rosario 1981) that uses a 96-well plate in which parasites are diluted at 0.5-0.8 parasites per well at 1.8% haematocrit. The expectation was to obtain approximately 50% of clonal parasites distributed across the plate. For more recently adapted field strains the protocol used was slightly modified in order to allow parasites to grow better. Calculations for the later were performed for 1 parasite per well rather than less than that. The times taken for parasites to become visible via microscopy ranged from 17 to 21 days in well adapted strains and more than 25 days for the recently adapted field isolates. Detection of positive wells was first performed using DNA stain Sybr Green as in (Lyko *et al.* 2012) in order to have an estimate of the amount of DNA present in each well and discard either empty well defined by an absorbance and emission equal to the negative control wells, or much higher than the average wells indicating possibly more than one parasite per well. Positive wells were later validated using light microscopy.

2.2 Transfection of *Plasmodium falciparum*

Genetic manipulation of *P. falciparum* was performed via standard transfections previously described (Deitsch, Driskill, and Wellems 2001; Wu *et al.* 1995). For all the electroporation methods used in this project, cells were monitored for different time ranges until parasites were visible via light microscopy. Times for this observation varied from 14 days in a very efficient transfection to 40 days in the most inefficient ones. Monitoring was performed by smearing and staining with 10% Giemsa for 100x magnification over weeks at least twice a week as described earlier. Drug selections were performed from day 1 post transfection and for a course of six days for CRISPR/Cas9 and for the whole time for episomal transfections, in order to select for parasites stably replicating episomes. During the project different drug selectable markers were used. Following these regimes, the most common plasmid that was used encoded the human dihydrofolate reductase gene (h*dhfr*), and parasites

were selected 24 hours after transfection with WR99210 compound, and for a total of six days for integration or until parasites were recovered for episomal transfections. The other drug selectable marker that was used throughout this project was the yeast dihydroorotate dehydrogenase (y*DHODH*), for DSM1 selection.

The different methods for transfection used over the course of this project are described below. Parasites ready for transfection were resuspended in cytomix for washing (120mM KCl, 0.15 mM $CaCl_2$, 2mM EGTA, 5mM $MgCl_2$, 10mM K_2HPO_4, 25mM HEPES, pH 7.6) and either cytomix or Primary Cell Solution (P3) solution from Amaxa for resuspension of DNA:

2.2.1 Direct electroporation of infected red blood cells (iRBCs)

2.2.1.1 Lonza 4D Nucleofactor

Cultured parasites with a hematocrit of 3% were split and centrifuged at 1100g in order to obtain a final pellet volume of 100uL, at a parasitemia close to 5%, with a high prevalence of early rings. The pellet was washed in cold complete cytomix with recipe described earlier. After washing the parasites were resuspended in 100uL of solution P3 containing 50-100ug of plasmid DNA and were loaded to an Amaxa cuvette. The program utilised for transfection of infected erythrocytes was the equivalent for the cell solution used to dissolve the plasmid DNA. The electroporated cells were rapidly resuspended in warm complete parasite culture medium and left to recover from one to two hours before posterior centrifugation. The pellets were then resuspended at the desired hematocrit in a 5mL culture well of a 6-well plate.

2.2.1.2 BioRad Gene Pulser Xcell Electroporator

Conditions of cultures and parasitaemia to use for transfection remain mostly unchanged for this electroporation method, except more RBCs are loaded. BioRad cuvettes can harbour a larger volume and are considerably more affordable than the Lonza cuvettes, yielding similar transfection efficiencies, therefore they were the electroporator of choice for most experiments during the course of this project. For the 0.2-cm cuvettes, 150uL of packed iRBCs at >5% parasitaemia were homogenised with 300uL of of plasmid DNA (between 50 and 100ug) previously resuspended in complete cytomix. The cuvettes were then electroporated using the

following settings: 950 microfarad (μF) at a voltage of 0.31 kilovolts (Kv) as in Deitsch *et al.* (Deitsch *et al.* 2001).

2.2.1.3 Electroporation of uRBCs preloaded with DNA

Preloading of RBCs was performed following the protocol developed by Deitsch *et al.* (Deitsch *et al.* 2001) in 2001. For this approach, 300uL of RBCs were homogenized with a lower concentration of plasmid DNA resuspended in cytomix. Concentrations varied from 10-50ug throughout the course of this project. Once the RBCs were electroporated, they were mixed with parasites in culture, allowing for reinvasion to occur and providing other preloaded cells to increase the efficiency of parasites taking up the plasmid. Unless otherwise specified, settings remained the same as previously described. For the case of the field isolates used in this study, this method was particularly helpful and some specific optimizations to the protocol were performed to increase the efficiency.

2.3 Molecular cloning

2.3.1 Saponin lysis and genomic DNA extraction from parasite cultures

Preparation of genomic DNA (gDNA) was performed from culture volumes of not less than 5mL at 3% haematocrit and more than 2% parasitaemia to assure sufficient yield for molecular experiments. Parasite cultures was centrifuged and pellet resuspended in 0.15% Saponin in PBS and left at room temperature for five minutes in order for cellular lysis to take place. Saponin-lysed pellets were then washed twice with PBS to remove all remains of blood and media. Pellets were then directly used for gDNA extraction. Briefly, extraction of genomic DNA was performed using the DNeasy Blood Tissue Kit from Qiagen, using 20uL of Proteinase K. For lysis, 200uL of lysis buffer (AL) were added and the samples were incubated at 56°C for 10-15 minutes and precipitated using 200uL of 100% Ethanol. The samples were then eluted in Elution Buffer (Qiagen) at a final volume of 60uL and concentration was measured using a either a Qubit or a Nanodrop, depending on experiment. Extracted DNA was kept at -20C and used for different experiments, ranging from PCR for to whole genome amplification.

2.3.2 Molecular cloning

2.3.2.1 Oligonucleoetide design

All the primers that were used throughout this study were generated using software LaserGene from DNA Star or with the free online platform Benchling (https://benchling.com/). For the first approach, secondary structures and melting temperatures were validated using Olicocalc (http://biotools.nubic.northwestern.edu/OligoCalc.html). Once the sequence was obtained and validated it was ordered through the Agresso purchasing system from the Wellcome Sanger Institute, using mainly Sigma-Aldrich or IDB providers. A full list of primers is provided in the appendix.

DNA synthesis was performed for certain sequences in order to facilitate the cloning process. The design was performed using the software specified and the synthetic DNA was ordered using gBlocks from IDT (https://sg.idtdna.com/site/order/gblockentry), or alternatively if the construct was challenging to synthetize, GeneArt from Thermo Fisher was used (https://www.thermofisher.com/th/en/home/life-science/cloning/gene-synthesis/geneart-gene-synthesis.html).

2.3.2.2 Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction or PCR is normally used to increase the number of copies of a particular DNA sequence, which can be later confirmed by methods such as capillary sequencing or Next Generation Sequencing. The way the reaction works is by cycling through different temperatures using a thermostable polymerase. The latter can either be a proofreading enzyme to avoid random incorporation of nucleotides, or non-proofreading used primarily to confirm by agarose gel rather than sequencing (KAPA from KAPA biosystems or GoTaq from Promega, respectively). Reactions for the thermocycler depended on the manufacturer instructions of the thermostable polymerases used, but generally it started with DNA denaturation at 95°C, followed by a drop-in temperature dependent on the melting temperature of the primer and thus an accurate annealing temperature. For all the oligonucleotides designed throughout this project, the temperature was adjusted in ordered to improve annealing of primer and DNA template, usually ranging between 50°C and 60°C. After

denaturation at 95°C, given the high AT content of the genome of *P. falciparum*, the extension temperature required was reduced from 72°C to 62°C in most reactions performed, unless specified otherwise. The extension time would also depend on the length of the fragment being amplified. The reaction was followed by multiple cycles of denaturation-annealing-elongation, varying from 5 to 35 cycles, depending on experimental setup and whether it was a nested PCR reaction, adjusted to increase specificity in a secondary reaction.

BarSeq-specific reactions for library amplification for NGS required a 2-step nested PCR reaction. For library amplification from a plasmid pool, DpnI treatment was used to remove any remaining vector template. For this, a final reaction volume of 25uL was used with a total of 80ng of plasmid DNA (PCR 1). Alternative, a PCR product from endogenous genes was used as template to amplify integrated barcodes. For measuring integrated barcodes, specific requirements and further reactions are listed in section 2.5.

2.3.2.3 PCR for bacterial colonies

PCR was also performed as a screening strategy for bacterial colonies of the plasmids transformed throughout this project. For these reactions, GoTaq enzyme was used according to manufacturer's settings and was performed directly from the agar plate. The reaction volume was of 10uL for this reaction, and the conditions for PCR slightly varied for screening colonies.

2.3.2.4 Confirmation of PCR products and purification

In order to confirm the correct size of the PCR reactions performed during these studies, or to confirm plasmids after performing restriction digestions, agarose gels were made at a concentration of 0.8% agarose in tris-acetate-ethylenademiaminetetraacetic acid (EDTA), TAE. After agarose was completely dissolved in TAE, either ethidium bromide (Sigma) or Sybr Safe at a final concentration of 0.5ug/ml were used to visualise DNA bands. Agarose solution was let at room temperature in the tray until solid and the DNA samples were loaded with 1X of loading buffer. 5uL of DNA ladder (Hyperladder from Bioline) were added to one well of the gel in order to define the fragment size accurately. PCR bands were observed using UV light and were excised, and a QIAgen kit was used to purify the PCR products from the gel band, following

manufacturer's instructions. The latter was only performed for certain reactions, specifically for Gibson assemblies. Otherwise, a PCR purification was performed and is described later.

For purifying PCR products, a Macherey and Nagel PCR purification kit was used following the instructions of the manufacturer.

2.3.2.5 Restriction digestions, annealing of oligonucleotides and DNA ligation

All the conditions required for the restriction endonucleases used in this study were established using the New England Biolabs (NEB) website for optimization of digestion (https://www.neb.com/tools-and-resources/usage-guidelines/optimizing-restriction-endonuclease-reactions). Most reactions were performed at an incubation temperature of 37°C, unless specified otherwise for using particular enzymes. The volume for a restriction digestion was of 10uL and consisted of 1uL of each enzyme, 2uL of CutSmart buffer also from Promega (unless specified otherwise), plasmid (0.5-1ug total), and nuclease free water.

Annealing of oligonucleotides was performed by adding equimolar amounts of complementary primers (at 100uM concentration), followed by Buffer number 2 from New England Biolabs, for a total volume of 10uL. The thermocycling program for annealing (5-minute incubation at 95°C and gradual reduction until room temperature was achieved), was performed on the PCR tubes, and the annealed oligonucleotides were diluted 1:100 in 5mM Tris (pH 8.0) for following T4 ligation.

For vector cloning, after plasmid digestion was performed, DNA fragments were ligated using the standard T4 DNA ligase from Promega in a total volume of 10uL. The proportion of plasmid to insert used was following the manufacturer instructions, and the reaction was incubated at 16°C for 1-4 hours or at room temperature overnight, a modification of the protocol of the manufacturer (New England Biolabs).

2.3.2.6 Gibson Assemblies

Gibson cloning was used to amplify homology regions from *P. falciparum* from different genes for the various experiments performed. PCR products of the homology regions would have a sequence of overlap to one of the vectors used defined by the oligonucleotides used. A 2x master mix of Gibson Assembly Master Mix from New England Biolabs (NEB) was used, and a

ratio of insert (region of homology or two homology arms) to cut the vector as described earlier of 2:1 was used for a two or three-fragment assembly at 50°C for 25 minutes. Colonies were screened for integration of the homology region and grown in LB medium with 1:1000 dilution of 100mg/mL of ampicillin, as all vector backbones contain this a cassette for ampicillin resistance.

2.3.2.7 Bacterial transformation

The ligated plasmids were transformed into chemically ultracompetent cells XL-10 (Agilent Technologies). For all the transformations, cells in a total volume of 30-50uL were defrosted on ice for 30 minutes before adding the DNA. The total amount of DNA transformed ranged between 0.1-50ng following manufacturer's instructions, and was gently added directly to the tube containing the cells. After adding them they were incubated on ice for another 30 minutes. The cell suspension was heat shocked for 30 seconds and rapidly moved to ice for two minutes, following addition of 900uL SOC recovery medium. The suspension was incubated in a shaking heat block pre-warmed at 37°C for one hour and then plated into Luria-Bertani (LB) agar plates, supplemented with 100ug/uL of ampicillin (Sigma-Aldrich), and incubated overnight at 37°C. Colonies were grown in mini preps from Macherey and Nagel and stocks were made at a final concentration of 25% glycerol for storing at -80°C.

2.4 Parasite phenotypic assay

2.4.1 Drug sensitivity assays

Standard drug sensitivity assays were performed following the protocol by Smilkstein *et al.* 2004 (Smilkstein *et al.* 2004) which was modified slightly to accommodate the availability of equipment in the laboratory. The principle of the protocol is to take advantage of the absence of DNA in enucleated Red Blood Cells in order to stain with SYBR Green I only DNA coming from parasitised RBCs, therefore measuring growth by the amount of DNA present in a given well, in presence of a particular compound.

Serial dilutions of the compound of interest were performed in duplicate as shown in Fig 2.2, obtaining a final volume of media in all the wells of 100uL. The last two columns were left as a control (no drug was added to those wells), and were used for the calculations of parasite growth and inhibitory concentration. A total of 100uL of parasitized culture, ideally at the ring stages were added to all the wells. The final parasitaemia would be between 0.5 and 1%, depending on growth rates of particular strains, and a final haematocrit of 1.8%. A total of 200uL of culture and drug suspension were incubated at the optimal gas concentrations described earlier for a total of 72 hours.

After the 72-hour incubation, a final concentration of 3X of SYBR Green I (stock at 10,000x) was diluted in lysis buffer (20mM Tris pH 7.5, 5mM EDTA, 0.008% saponin, 0.08% Triton X-100). From the lysis solution, 15uL of the solution were loaded onto a 96 well plate. The incubated parasites were resuspended with a multichannel pipette and 30uL were taken to load on top of the SYBR Green solution. The plate was covered in foil and left at 37°C for 30 minutes and then loaded on a FLUOstar Omega Microplate reader. All the results obtained from the plate reader were interpreted as parasite growth and the half inhibitory concentration (IC$_{50}$). Compounds used throughout the project are listed in table 2.2, and the analysis and statistical tests was performed using an R script using the package drc (for generating dose-response curves) (Ritz *et al.* 2015). The latter takes into account different parameters for analysis: the lower limit or parasite activity in the absence of drug, the upper limit or concentration value at which parasites reach 100% inhibition, the slope of the curve that indicates the kinetics of the reaction and the shift when the different strains reach 50% inhibition (Nguyen *et al.* 2014).

Table 2.2 List of compounds used for inhibitory concentration assays and the range of concentrations used for setting up the serial dilution plates for BarSeq experiments. The concentration for chloroquine was overall higher for all given the high IC50 of some of the strains used such as V1/S and CAM.

Compound	Concentration (range in nM)
Chloroquine (C6628 SIGMA)	[0-1000]
Piperaquine (C7874 SIGMA)	[0-500]
Mefloquine (C2319 SIGMA)	[0-200]
DHA*(A6979 LKT Labs)	[0-200]
Lumefantrine (L5420 SIGMA)	[0-200]
Halofantrine (H9414 SIGMA)	[0-200]
Atovaquone (A7986 SIGMA)	[0-200]

* Dihydroartemisinin
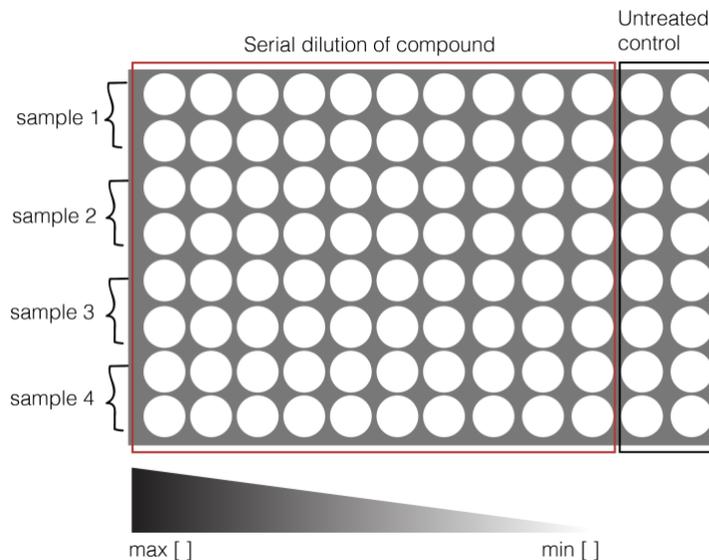
Figure 2.2 Plate layout for inhibitory concentration assays. Each sample is loaded in duplicate (e.g. rows A and B for sample 1). The selected drug is loaded at the maximum concentration on column 1 and a serial dilution of 1:2 per column is performed until column 10. Columns 11-12 are the negative controls which are used for the analysis of the dose response curve.

2.4.2 Ring-stage Survival Assay (RSA)

2.4.2.1 RSA measured by thin smear and light microscopy:

A modified version of the protocol by (Witkowski *et al.* 2013) was used for pulsing early ring stages with dihydroartemisinin (DHA). Briefly, late-stage parasites (schizonts) were purified using the 63% percoll solution described earlier in section 2.1.1.2. The layer with the schizonts

was carefully removed and washed in 50mL of warm complete medium. Pellet was resuspended in ⅕ of the original volume of the culture: for a 50mL original culture, the late stages were transferred after purification to a 10mL flask for posterior re-invasion. The later would take place within three hours to make sure parasites were in the 0-3-hour window required for the assay. A final concentration of 700nM DHA (unless otherwise specified for specific experiments) was added to the exposed (treated) parasites. The assay was performed in triplicate and a control well using the same concentration of DMSO present in the DHA stock was used in order to measure final survival of treated parasites by comparing to the final parasitaemia of the untreated. After the six-hour treatment, the media with the drug was washed twice with 15mL of complete medium, and the parasites were left to recover until they reached 72 hours since start of treatment. In order to measure the percentage survival, the following formula was used:

$$Growth\ rate = NE/INI$$
$$\%\ Survival = (DHA/NE)*100$$

* NE=non-exposed, INI=initial parasitaemia and DHA=exposed.

The variables used for the measurement were calculated from thin blood smears and observed under the light microscope. A total of 10,000 red blood cells were counted in order to have an accurate measurement of survival.

2.4.2.2 Two-color flow cytometry approach to measure labelled parasite populations and for performing RSA$_{(0-3h)}$

The steps for parasite preparation were the same as with the thin blood smear approach described above, however after 72 hours a specific treatment for measuring parasite survival using flow cytometry was used.

A two-color flow cytometric approach was also used in order to measure survival of parasites treated with DHA *in vitro.* The protocol used for this approach was adapted and slightly modified from Amaratunga *et al.* 2014 in (Amaratunga, *et al.* 2014). The two dyes used

for this assay were SYBR Green I Nucleic Acid Stain (Sigma-Aldrich) and MitoTracker Deep Red (Thermo Fisher), with 497/520 nm and 644/665 Excitation/Emission, respectively. The concentrations were modified as it was easier to perform the gating than with the concentrations published. A total concentration of 2xSYBR-green, 160 nM mitotracker deep red in complete media were used for the staining. Briefly, 100uL of culture were loaded into a 96-well plate, washed once with PBS and resuspended in 100uL of the two-dye mix. The parasites were left at 37°C inside the incubator for 30 minutes, covered. After staining, the plates were centrifuged and two washes with PBS were performed before resuspending in a total volume of 200uL of PBS. This was then diluted 1 in 10 for loading to a benchtop Cytoflex Flow Cytometer from Beckman Coulter. The gating strategies were performed following the published protocol and are shown as follows, after being exported from the flow cytometer and loaded for posterior analysis using the software FlowJo Version 10, and are shown in Fig. 2.3.
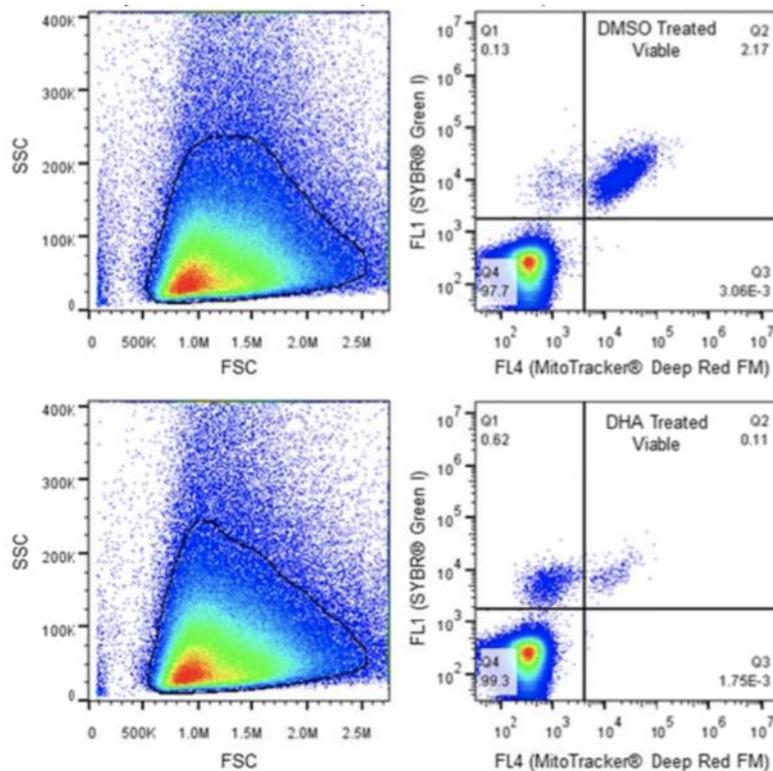
Figure 2.3 Gating for two-color flow cytometry approach. Two left panels represents the total population of cells in the control wells. Top right panel represent the live cells in the non-exposed parasites (NE). The bottom right panel are the exposed parasites (DHA). For both gating panels, the right gate (Q2) represents the mitotracker positive (staining only mitochondria thus only live parasites will be on this gate) and SYBR green positive cells which are the surviving cells after the assay. The top left gate (Q1) includes both live cells but mitotracker negative and dead cells as SYBR green is a DNA dye. Measurement of survival will be by using the formula described for the microcopy approach. Modified from (Amaratunga, *et al.* 2014).

2.5 Barcode Sequencing

2.5.1 Competition assays using Next Generation Sequencing (NGS)

For measuring the relative growth of individual parasites in a pool, parasites were carefully quantified in order to achieve a 1:1 ratio of each strain measured in the competition. Depending on the experimental designed, gDNA was collected and extracted following the protocols described earlier in this chapter. Each experiment was performed in duplicate or triplicate, which will be specified accordingly in the following chapters.

2.5.1.1 Library preparation and Illumina sequencing (BarSeq)

Because *Plasmodium falciparum* has the ability to stably maintain replicating episomes, a first round of PCR amplification was performed in order to discard plasmid, and the strategy will be described in the following chapter. Briefly, the first PCR amplifying from outside the homology regions was performed using the proofreading DNA polymerase KAPA Hot Start Ready Mix 2x from Kapa Biosystems. The conditions were optimized to work for *Plasmodium falciparum*: 30 cycles of 95°C for 3 minutes (initial denaturation), 98°C for 20 seconds (denaturation), 57°C for 20 seconds (annealing), 62C for 2 minutes (extension), and a final extension at 68°C for 2 minutes. All the PCR products were verified on a 1% agarose gel for size confirmation, and purified using the protocol described earlier, with their concentration accurately measured using nanodrop.

PCR products from outside the homology region that were purified were used as a template for a nested PCR, with a total of 100ng loaded. Two PCRs were performed using a strategy of adaptor ligation as described in the overview of Illumina Sequencing in the previous chapter. First PCR was performed in duplicate for analysis purposes, using primers 91 and 97 listed in appendix, the later amplified the region of interest containing the barcode, with PCR settings as described previously in Gomes *et al.* 2015 (Gomes *et al.* 2015). These primers bind the annealing sites annotated as arg444/arg445 in Figure 1.16 of introductory chapter, flanking each barcode used in most of the experiments in this project, unless specified otherwise. This generated an amplicon of 167bp which was later used in a second PCR reaction. The later one was performed in order to anneal the Illumina adapters on the 5' end, together with a unique index tag to be able to identify each sample in a BarSeq pool. Throughout the length of the project, multiplexing never exceeded 96 different samples in a single lane of MiSeq. This final library was amplified using the universal primer PE1.0 and any of the 96 synthesized unique oligonucleotides. All of the primers are listed in appendix. All the PCR reactions performed in the process of getting the library ready for NGS were ran on a 1% agarose gel for size validation, then PCR purification was performed as described earlier. The concentration of the samples was measured using a dsDNA Qubit HS Kit from Thermo Fisher and pooled together in

equimolar amounts, for a final dilution of 4nM that was used for Illumina MiSeq sequencer service at the Sanger Institute.

The diluted libraries were loaded at 50% PhiX concentration and at a low cluster density (<400k cluster density; $4x10^5$ cluster/mm2) given their low complexity. The reads were paired-end with a length of 150bp. Illumina MiSeq v300 kits were provided for each run for the Illumina Walk-Up Sequencing service available at the Wellcome Trust Sanger Institute, or were used from the Bespoke team at the Sanger Institute.

2.5.1.2 Fitness calculation for competition assays

Once the reads were obtained from the MiSeq sequencer, a Python script (see https://gist.github.com/mgalardini/4bc16dc0101aaacce675d719579cee66) was developed by Marco Galardini in order to find individual barcodes in a complex pool. A fasta file containing a list of sequences to look for was provided only exact matches were counted. Sequences with low quality base calling or non-exact matches were excluded for posterior analysis. The output file would be composed of the haplotypes/barcodes, day, condition and the read counts, which were later analysed using R.

A comprehensive R script was developed in order to start from the input file described earlier and to work out the relative abundance of each parasite in the pool. Briefly, this was measured as the proportion of each barcode over time, and the relative growth rate was measured as the log2 of the change in barcode proportion over time. All of the analysis that led to all the figures presented throughout this work were performed using R packages ggplot2 and tidyverse (https://www.tidyverse.org/), which combines other data managing packages such as "plyr", "dplyr" and "broom".

Chapter 3: Establishing a safe harbour for barcode insertion in the *Plasmodium falciparum* genome and using a pool of barcoded plasmids as a readout of transfection efficiency.

3.1 Summary:

The aim of this chapter was to provide a proof of principle for barcode sequencing in *P. falciparum* parasites. I first adapted the barcode cassette from the *Plasmo*GEM linear KO vectors (Bushell *et al.* 2017) to a circular *P. falciparum* donor plasmid, pCC1 (adapted from Maier *et al.* 2006). I then established the pseudogene (*Pfrh3*) as a feasible safe harbour to insert barcodes by co-transfecting the barcode donor plasmid with a Cas9 nuclease and a single guide RNA targeting *Pfrh3*. I then combined these two strategies, and performed pooled episomal transfections to understand transfection efficiencies of *P. falciparum* and to address the limitations for generating barcoded parasites at a large scale.

3.2 Background:

3.2.1 Barcode Sequencing for measuring fitness

Next generation sequencing (NGS) has pushed the scale of our understanding of biological systems in a number of different ways beyond simply enabling the production of more whole genome sequences. A key example is the application of NGS to quantify the relative growth rate of numerous strains in parallel, for example in chemical-genetic interactions screens, otherwise known as chemogenomics. Before NGS, competitive chemogenomic assays in yeast were performed using barcode microarrays as a readout of fitness, with each individual strain having a different barcode. The advantage of NGS technologies allowed the group of Corey Nislow to establish Barcode Analysis by Sequencing (BarSeq) technology, applying it to a panel of yeast deletion strains (Giaever and Nislow 2014). As each strain was individually barcoded, the authors were able to amplify all 20-mer barcodes

with conserved primers containing the required sequence for Illumina NGS. The quantities of the barcodes could then be recorded using NGS and the fitness of particular mutants measured. This approach outperformed their previous method using microarrays, particularly for non-essential genes (Smith *et al.* 2009). More recently, the Billker and Rayner groups at the Wellcome Sanger Institute formed the *Plasmodium* Genetic Modification Project (PlasmoGEM), and developed a library of uniquely barcoded knockout vectors targeting, at the latest count, more than half of the genome of the rodent malaria parasite *P. berghei*.

3.2.2 CRISPR-Cas9 based editing in *Plasmodium falciparum*

As outlined in Chapter 1, reverse genetics has not yet been implemented at a large scale in *P. falciparum*. This is in large part because there are some clear bottlenecks that make transfection efficiency particularly low in this species, as compared to other species like *P. berghei* or *P. knowlesi* (Bushell *et al.* 2017; Moon *et al.* 2013), and even lower when compared to other Apicomplexans such as *Toxoplasma gondii* (Sidik *et al.* 2016). However, factors affecting the efficiency of both transfection and homologous recombination still remain to be fully investigated in *P. falciparum* and other *Plasmodium* species.

The success of CRISPR/Cas9-based genome editing in *P. falciparum* has been a promising development as reverse genetics now seems feasible and scalable for tackling different biological questions. However, due to the absence of NHEJ, even simple gene disruptions require the provision of a donor sequence when transfecting, adding complexity to vector design and scale. However, the absence of NHEJ has the side benefit that off-target effects are likely to be minimal, as unwanted double-strand breaks would likely lead to cell death by the inability of the parasite to repair these lesions (Ghorbal *et al.* 2014; Wagner *et al.* 2014).

Another challenge in the application of CRISPR/Cas9 to *P. falciparum* is design of the gRNAs required to target Cas9 to the correct site. The targeting region of gRNAs are conventionally 20 nucleotides long, and require a downstream Protospacer Adjacent Motif, or PAM, at the target site that is critical for the nuclease to induce cleavage. While multiple

variants of Cas9 are now available, in the case of the conventional *Streptococcus pyogenes* Cas9 that has been published in multiple studies in *P. falciparum* (Ghorbal *et al.* 2014; Wagner *et al.* 2014), the PAM requirement is NGG. Given the low GC content of the *P. falciparum* genome, there can be challenges in designing gRNAs near to the target site that meet this requirement. Another consideration that is relevant for the design are the "off-target" scores, which refer to the "uniqueness" of the target sequence and the likelihood that the nuclease will cleave elsewhere in the genome, although as noted above this is most problematic for organisms relying on NHEJ for DNA repair. This score, incorporated in platforms like Benchling, was used to identify gRNAs that had at least three "mismatches" to other regions of the genome. The proximity of repair to the cut site has also been shown to impact the editing efficiency, with efficiency decreasing as the distance to the cut site increases (Sander and Joung 2014). Lastly, the "on-target" score, which measures the activity of the nuclease based on predictive models, has been developed using mammalian systems and yeast models. Because it is unclear whether these predictions are also true for *P. falciparum* on-target scores were not be taken into consideration throughout this project (Doench *et al.* 2014).

3.2.3 Using a pseudogene for barcode insertion

The first step in establishing Barcode Analysis by Sequencing (BarSeq) for *P. falciparum* was to establish a safe harbour in the genome in which barcodes could be inserted without any effect on fitness. *Pfrh3* (PF3D7_1252400), or reticulocyte binding protein homologue 3, was selected as the target locus because it has been characterised as a pseudogene (Duraisingh *et al.* 2002) and can be disrupted without any effects on parasite growth (Taylor *et al.* 2001). Located on chromosome 12, this gene was first identified in a study by Taylor *et al.* 2001 which set out to identify *P. falciparum* homologues of *P. yoelii* and *P. vivax* proteins involved in invasion of erythrocytes. Two key invasion proteins had come out of other studies as having sequence similarity; *PfRh2a* and *PfRh2b* (Rayner *et al.* 2000). These two proteins are almost identical except in their C-terminal amino acid region, and are both non-essential for invasion. *Pfrh3* was identified as part of the same family based on sequence homology to other species. Unlike *PfRh2a* and *2b* however, the 5' end of the gene contains two reading frame shifts. While

the gene is transcribed in the schizont stages, it would not be predicted to produce a functional protein, confirmed by using polyclonal antibodies raised against different fragments of *Pfrh3* (Taylor *et al.* 2001). With this evidence in favour of *Pfrh3* being a pseudogene and not essential for parasite growth, it was selected for barcode insertion.

3.3 Methods and Results

3.3.1 *Pfrh3* (PF3D7_1252400) as a safe harbour for barcode insertion

3.3.1.1 Assembly of barcoding vector with homology arms for targeting Pfrh3

In order to amplify homology arms of approximately 1kb for *Pfrh3*, genomic DNA of 3D7 was extracted using the protocol described in section 2.3.1. This region was also sequenced from other strains used later in this project to confirm that the gRNA target sequences were conserved, so the same designs could be used across multiple strains. Primers designed to amplify the homology regions contained sequences of overlap with each other and included restriction sites NcoI and NheI for subsequent ligation of barcodes (p204 and p205). They also contained a sequence of overlap to the plasmid backbone of pCC1 (Fig. 3.1A), to allow a three-fragment Gibson assembly to be performed. The first homology arm, HRI (928bp), was amplified using primer p203/p204 and HRII (1200bp) was amplified using p205/p206, for the final assembly shown in Fig. 3.2A. Because of the high AT content in the genome of *P. falciparum* (Pollack *et al.* 1982), a lower extension temperature was used than that predicted for the primers. The cycling settings for this reaction for a Phusion High Fidelity DNA polymerase were: 95℃ 2m (initial denaturation), followed by 34 cycles of 95℃ 20s (denaturation), 52℃ 20s (annealing), 62℃ 1m (extension) and finalising at 62℃ for 5m (final extension). Lanes 2 and 3 in Fig 3.1C show amplification of the two homology arms, which were later purified using a Qiagen MinElute PCR purification kit.
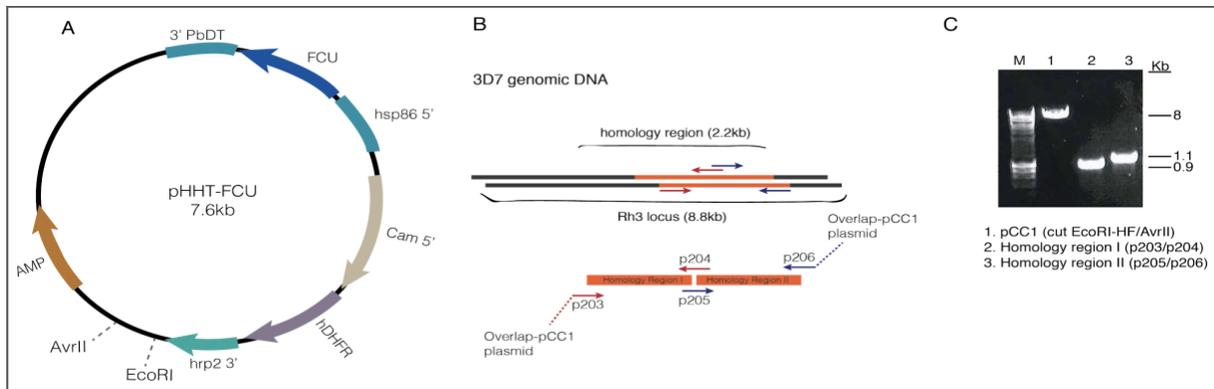
Figure 3.1 Amplification of homology arms of *Pfrh3* and cloning into pHHT-FCU backbone. A. Figure of pHHT-FCU vector containing an Ampicillin (AMP) cassette. Initially published by Meier *et al.* 2006. The first selectable marker encodes the h*dhfr* (human dihydrofolate reductase) gene, driven by the calmodulin promoter (Cam 5'), with a histidine-rich protein 2 3' UTR as a terminator sequence (hrp2 3'). The second selectable marker, yFCU, is for negative selection with compound 5-FC (FCU). **B.** Design of homology arms for amplification and cloning into pHHT-FCU (A). All primers used for this reaction are labelled in the second panel, and the dashed lines indicate the 20bp nucleotide overlap for Gibson assembly into the linearised vector. **C.** PCR-amplified fragments used for Gibson assembly into the pCC1 plasmid backbone. Lane 1 shows pCC1 plasmid linearised using restriction enzymes EcoRI and AvrII for Gibson assembly of homology regions of *Pfrh3*. Lanes 2 and 3 are the two homology regions amplified from gDNA, with overlapping regions to linearised pCC1 plasmid (p203 and p206) and also overlap between the two regions (p204/p205) for a three-way assembly. M stands for DNA marker (Hyper Ladder).

Assembly of the homology arms with the pCC1 plasmid was performed by Gibson assembly as described in section 2.3.2.6, and confirmed by both Sanger sequencing and with diagnostic digestion using NheI/SacI (6.5 and 3kb). Fig. 3.2A shows the two homology regions and the NcoI/NheI restriction sites for subsequent insertion of the barcode. The genomic region in *Pfrh3* in between the homology arms encodes the guide RNA sequences shown with red arrows, upstream of the NGG Protospacer Adjacent Motif (PAM) in blue (Fig. 3.2A).
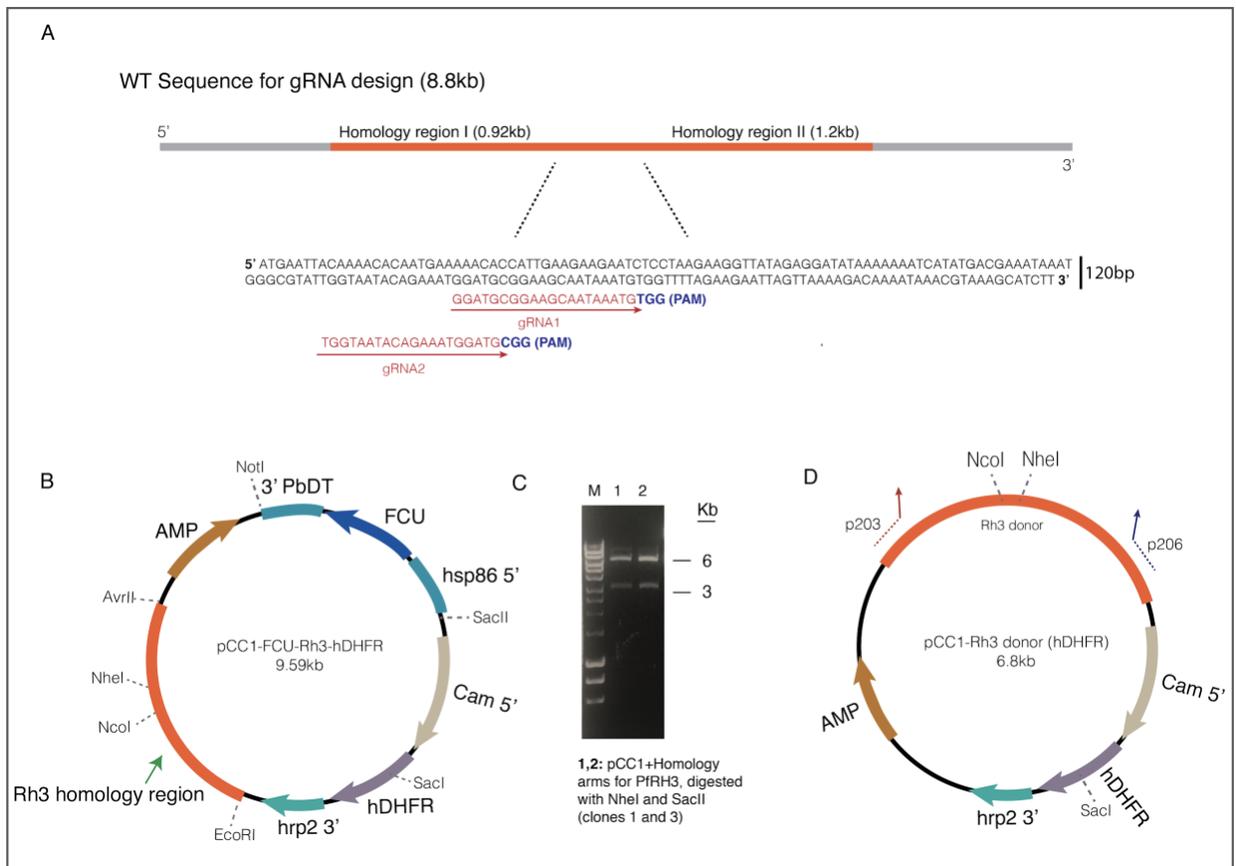
Figure 3.2 Final pCC1 vector with donor sequence for *Pfrh3*. A. Wild type locus of *Pfrh3* (8.8kb) showing the homology regions designed for cloning into the pCC1 backbone. The 120bp region between HRI and HRII contains the target sequences of the guide RNAs. **B.** Plasmid containing homology regions assembled using Gibson cloning, with restriction sites NcoI and NheI for barcode ligation. **C.** A confirmation digest using restriction enzymes SacII and NheI to verify insertion of the homology region for Rh3 after Gibson assembly (two independent clones shown in lanes 1-2). M stands for DNA marker (Hyper Ladder). **D.** Final construct after assembly of homology arms for Rh3 and removing the negative selectable marker. Primers p203/p206 were used for the Gibson reaction into pCC1 backbone (B).

Fig. 3.2B shows the assembled pCC1 vector similar to the one published by Meier *et al.* for *P. falciparum* (Maier, Rug *et al.* 2008), confirmed using restriction enzymes NheI and SacII (approximately 6.5 and 3.1kb bands in Fig. 3.2C). pCC1 utilises the 5'UTR sequence of the *P. falciparum* calmodulin gene (CAM) to drive expression of the human dihydrofolate reductase gene-selectable marker (h*dhfr*), with the aim of conferring resistance to the drug WR99210. The original pCC1 includes the yeast negative selectable marker cytosine deaminase (FCU), to select double crossover recombinants by using compound 5-Flurocytosine (5-FC). Once the homology regions were assembled, multiple attempts were performed to ligate the barcodes between these regions. However, the plasmid became unstable, and despite screening multiple colonies

it was clear that ligating the barcodes was problematic. For this reason, the negative selectable marker was removed to decrease plasmid size, using a restriction digestion with NotI and SacII, shown in Fig. 3.2B, and free ends were repaired using the Fast DNA End Repair Kit. A finalised version of the construct without the negative selectable marker is shown in Fig. 3.2D and it was this version that was used for barcode insertion and subsequent transfection.

3.3.1.2 Design of 30bp DNA barcodes for insertion into pCC1-Rh3 donor backbone

In order to find unique barcodes not found across the genome of *P. falciparum*, the first approach was to design 30bp barcodes with a high GC content (50%) which would immediately increase the chances of producing sequences that do not exist in *P. falciparum* given its strikingly low GC content. In order to generate a large number of 30bp sequences, a website designed by the Maduro Lab at the University of California, Riverside (http://www.faculty.ucr.edu/~mmaduro/random.htm) was used. Once this strategy was established, Frank Schwach at the Sanger Institute wrote a script for generating random sequences and then searching for them across the whole genome of the reference strain 3D7 of *P. falciparum*. All sequences produced were assessed for their potential to form internal secondary structures using predicting software as described in section 2.3.2.1. Ultimately, two independent barcodes as a proof-of-principle were cloned after ordering them as oligonucleotides.

Fig 3.3A-B shows the two sequences that were used for the pilot experiment to understand whether the *Pfrh3* locus could function as a safe harbour for barcode insertion, and how they were annealed into a digested pCC1 backbone. Oligonucleotides p171/p172 were used for Barcode 1 and p173/p174 for Barcode 2 (Fig. 3.3B). Annealed primers were ligated in between the homology regions for *Pfrh3* using restriction sites NcoI and NheI, generated with primers p205 and p206, marked in blue and red, respectively in Fig. 3.1A. Fig. 3.3C shows Sanger sequencing of barcode 2 inserted into the donor vector.
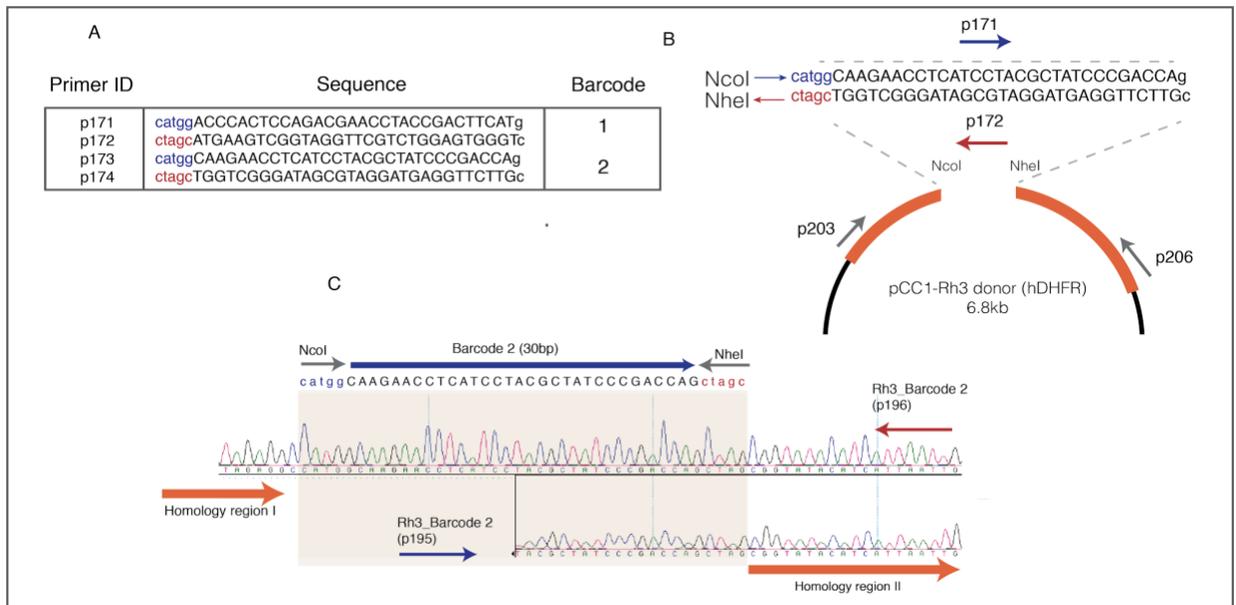
Figure 3.3 Insertion of a 30bp DNA barcode into *Pfrh3*. A. 30bp sequences for two barcodes designed containing at least 50% GC content, flanked by NcoI/NheI restriction sites marked in blue and red, respectively. **B.** Ligation of barcode into pCC1-Rh3 donor (hDHFR) by standard T4 ligation. **C.** Confirmation of barcode insertion into pCC1 backbone by Sanger sequencing, using sequencing primers p191/p196.

3.3.1.3 Design of gRNAs targeting Rh3 and assembly into pDC2-Cas9 expressing vector

gRNA expression vectors for targeting *Pfrh3* were generated as a modified version of the pDC2 vector developed in the lab of David Fidock (Fidock *et al.* 2000). This vector contains the sequence for expressing a nuclease Cas9 together with a sgRNA under the control of a U6 promoter as in (Ghorbal *et al.* 2014). After transfection, the plasmid can be selected for with DSM-1, which targets dihydroorotate dehydrogenase (DHODH), and for which the yeast enzyme provides resistance (Ganesan *et al.* 2011). The gRNAs were ordered as standard oligonucleotides, and after annealing were ligated into the pDC2 backbone, previously digested using restriction enzyme BsbI, shown in Fig. 3.4.

Taking into consideration the criteria for selection of gRNAs described in section 3.2.2, two independent guide sequences were identified (Fig. 3.4A), taking into account the total number of mismatches and their position relative to the PAM (Hsu *et al.* 2013). The positive sign on the left refers to the strand orientation, in the case of the two guide RNAs selected (in green), they

were both on the positive strand (+). Fig. 3.4B shows the oligos used to generate the two selected guide RNAs. A map for the BbsI-digested vector is shown in Fig. 3.4C, in which the desired guide sequences were ligated. Confirmation of successful ligation was obtained by PCR and direct Sanger sequencing (Fig. 3.4D and E).
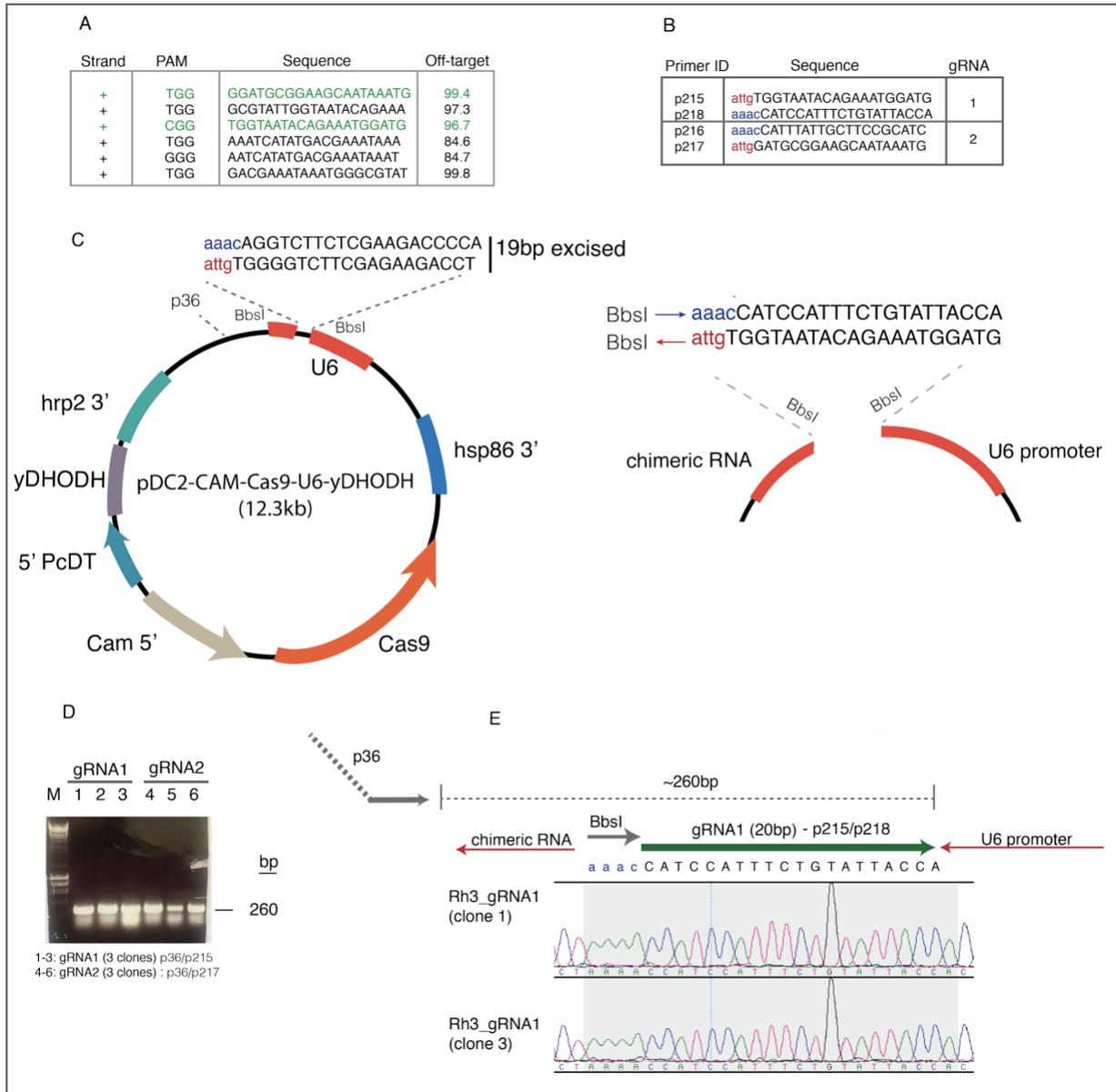
Figure 3.4 gRNA design strategy for targeting *Pfrh3*. A-B. Guide RNAs designed using software Benchling. Sequences in green were selected for cloning into pDC2-CAM-Cas9-U6-yDHODH (**A**) and the oligonucleotides ordered for annealing and ligation (**B**). **C.** Sequence flanked by BbsI sites that is replaced by gRNAs shown in Fig. 3A-B. **D.** Confirmation of gRNA1 and gRNA2 after ligation using the gRNA sequence as a reverse primer (p215 and p217), and p36 as forward primer resulting in a 262bp PCR product. **E.** Sanger sequencing confirming the insertion of the gRNA1 in between BbsI sites by using primer p36. In blue and red (**B, C, D**) are the sites for annealing into the pDC2 backbone previously digested with BbsI.

3.3.2 *Pfrh3* can be used as a safe harbour to insert DNA barcodes

Once the correct insertion of the barcode sequences into pCC1 were confirmed (Fig. 3.3C-D), parasites were transfected. These first attempts were for optimization and in order to improve the overall low efficiency of transfection of *P. falciparum* (Hasenkamp *et al.* 2012), and to validate whether *Pfrh3* could be used as a harbour to drop in the 30 nucleotide fragment encoding a single barcode as a proof-of-principle.

Both the Cas9-gRNA and donor-barcode vectors were co-transfected into the standard laboratory reference strain, 3D7. In this case, transfections were performed by electroporation of 50 µg of DNA per plasmid. An AMAXA 4D electroporator system was used, and the parasite stage selected based on previous studies were rings (Wu *et al.* 1996).

Parasites were obtained within three weeks after electroporation, with continuous drug selection using compound WR99210. A set of genotyping reactions were designed to validate integration of the 30bp barcode sequence into the non-essential gene *Pfrh3* by differentiating between the WT and edited locus (Fig. 3.5A-B). In both cases, the forward primer p191 anneals outside the *Pfrh3* homology arms, so a positive PCR product could not be generated from non-integrated episomes remaining in the transfected parasites (Crabb *et al.* 1997). Reverse primers recognised either the WT genomic sequence (p215) or edited barcode sequence (p174), as shown in Fig. 3.5A-B. The presence and size of PCR products was evaluated by gel electrophoresis, as shown in Fig. 3.5C. The first three lanes correspond to the wild type 3D7, which was used as a control, and as expected only yielded a band for the unedited locus (lane 3). In contrast, the transfected line was positive for the barcode insertion (lane 5), but no unedited wild type product was visible (lane 6), indicating the efficiency for the editing reaction was 100% as per this genotyping method.
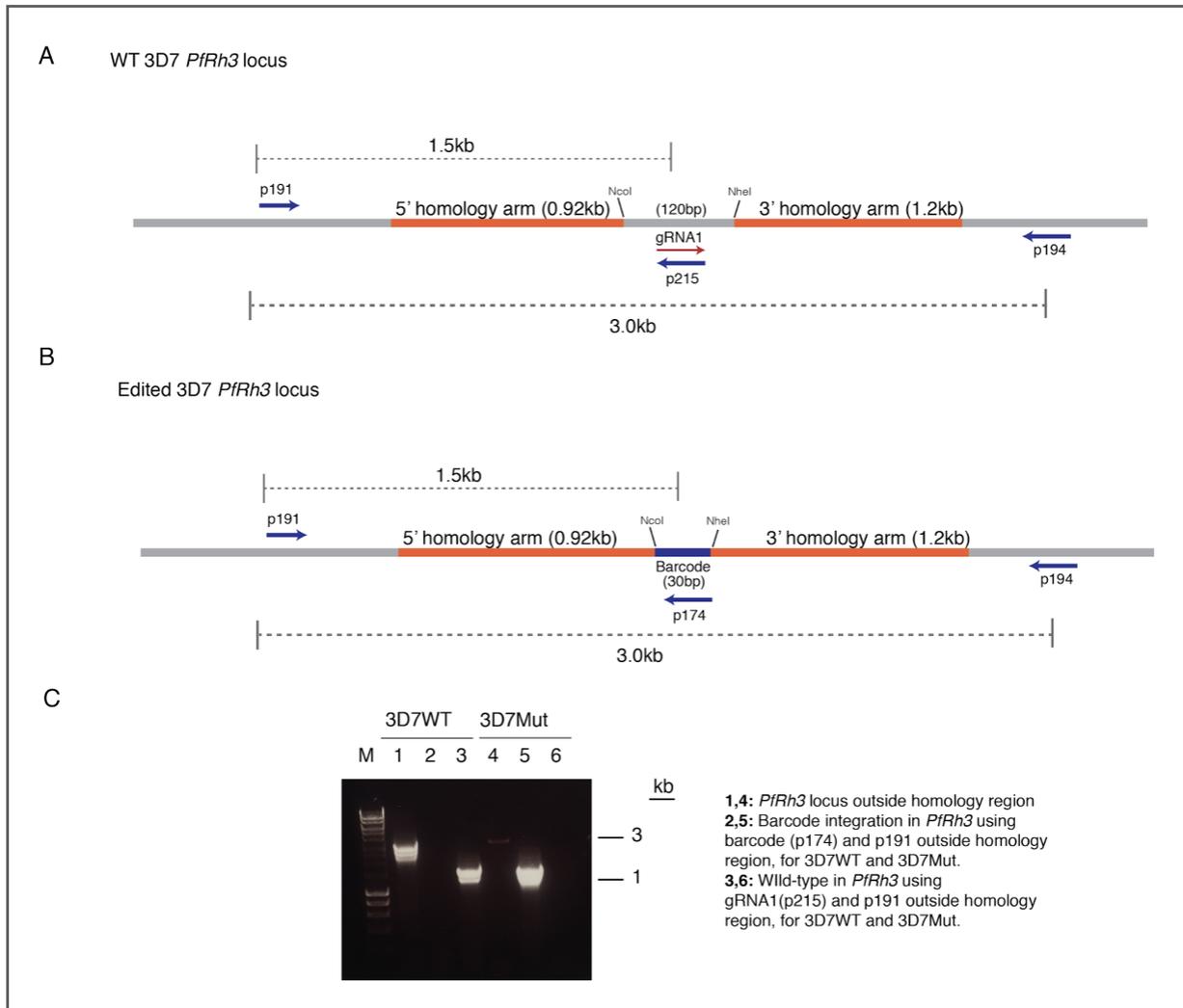
Figure 3.5 Strategy for genotyping of *Pfrh3* mutants. A-B. Genotyping strategy for determining WT and edited locus within the transfected population of parasites to measure the efficiency of integration. The WT locus was amplified by using p191 and p215 (A). The edited locus also uses p191 outside the homology region and p174 within the barcode as a reverse primer (B). **C.** Gel electrophoresis showing the genotyping reactions for both wild type and edited 3D7. Lanes 1,4 show the amplification of a 3kb PCR product (faintly visible in lane 4) that was used for confirmation using Sanger sequencing. Lanes 2,5 report from the edited population and lanes 3,6 amplify the WT locus.

While the editing reaction seemed very efficient, dilution cloning was carried out in order to isolate parasite clones that had no remaining wild type population. Bulk culture was diluted such that 0.5-0.8 parasites would be present in each well in a 96-well plate, so that any parasites that emerge are likely to be clonal. After 17 days, an aliquot from each well was transferred to a Sybr Green I lysis plate to measure DNA concentration and thus identify

positive wells, as described in section 2.1.3. Fig. 3.6A shows two of the clones (A8 and C9) that were expanded and subsequently genotyped. Integration of barcode 1, flanked by restriction sites NcoI and NheI, was confirmed by Sanger sequencing (Fig. 3.6B).
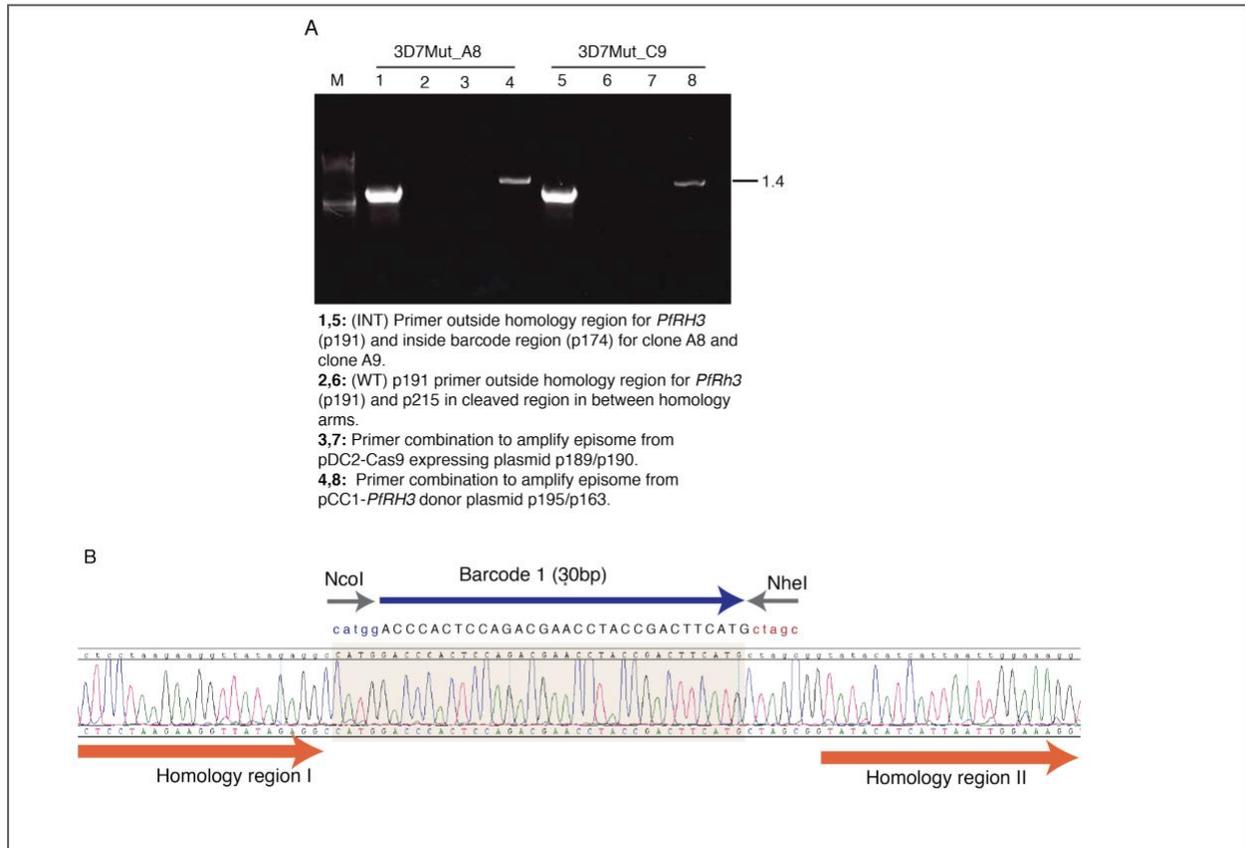
Figure 3.6 Genotyping of *Pfrh3* mutant clones. A. Genotyping strategy performed on two independent clones isolated after limiting dilution cloning. PCR amplification of the barcoded locus (lanes 1,5) and the wild type locus (lanes 2,6) indicate both clones are barcoded. Additional reactions were performed to detect the presence of episomes for the pDC2-Cas9 plasmid (lanes 3,7) and the donor pCC1 plasmid (lanes 4,8). **B.** Confirmation of integrated barcode in clone A8 by Sanger sequencing. The sequencing was performed on a PCR product that would amplify the genomic *Pfrh3* locus but not the episome, using primer combination p191/p194 shown in Fig. 3.5A-B.

Overall, the integration efficiency at *Pfrh3* was high, as shown from PCR genotyping performed on the bulk transfection before cloning, where there was no amplification using the primer combination that would identify the presence of the wild type *Pfrh3* locus. Two sets of reactions were able to determine the presence of the episomes within the transfected

population, and confirmed that that pCC1 plasmid remains replicating after drug pressure was removed (transfected parasites were selected for only six days with WR99210). By contrast, no amplification for pDC2-yDHODH, the plasmid containing Cas9 and the gRNA, could be observed, potentially because the parasites were not selected with DSM-1. The persistence of pCC1 was not surprising, as it has been proposed that episomes can start aggregating and rather than replicate as circular vectors, form concatemers that can persist within the population (Williamson *et al.* 2002). This might explain the presence of a PCR product even after dilution cloning, suggesting that extended drug-free culturing does not produce parasites that lack episomes.

3.3.3 Adapting the BarSeq approach to *Plasmodium falciparum* by generating pools of barcoded plasmids

The work above establishes a CRISPR-based approach for tagging parasite lines at the *Pfrh3* locus with barcodes, but making individual barcoded mutants one at a time is laborious. In *P. berghei*, dozens of different barcoded parasite mutants can be generated with a single transfection, and this has allowed fitness measurements in the blood stages at a large scale (Bushell *et al.* 2017; Gomes *et al.* 2015). I therefore decided to test whether pooled transfection approaches could be applied to *P. falciparum,* by generating pools of barcoded vectors that could be transfected either in the absence of the Cas9 cleavage event such that they are maintained as episomes, or in the presence of *Pfrh3* cleavage to insert the barcodes into the genome. Initial experiments using episomal barcodes were performed in order to understand: 1. transfection efficiency and 2. the number of unique plasmids taken up by a single parasite, for the purpose of having an estimate of the potential limitations in performing large-scale genetic screens in this species.

In order to generate pools of barcodes, the backbone vector from the *P. berghei Plasmo*GEM resource, pGEM, (Gomes *et al.* 2015) was used as template to amplify barcode cassettes. The linear knockout vector design is shown in Fig. 3.7A, which contains constant primer binding sites for Illumina sequencing (in green), flanked by a negative/positive selectable marker (h*DHFR*/y*FCU*), and generic homology regions targeting a gene of interest in *P. berghei*.

The availability of over 2000 barcode vectors coming from this useful resource was more than sufficient for the experiments performed in this study. A plate of 96 different *P. berghei* gene-targeting vectors was obtained as bacterial glycerol stocks (plate ID PbSTM_21). All 96 different bacteria carrying the unique pGEM vectors were cultured individually (to avoid competition) in deep 96-well plates overnight at 30℃, in TB medium with 30ug/mL of kanamycin. After each individual bacteria grew to an appropriate optical density (OD), and assuming similar growth rates across wells, they were pooled and a single library containing all 96 vectors was extracted using a Macherey-Nagel midi prep kit as described in section 2.3.2.7.
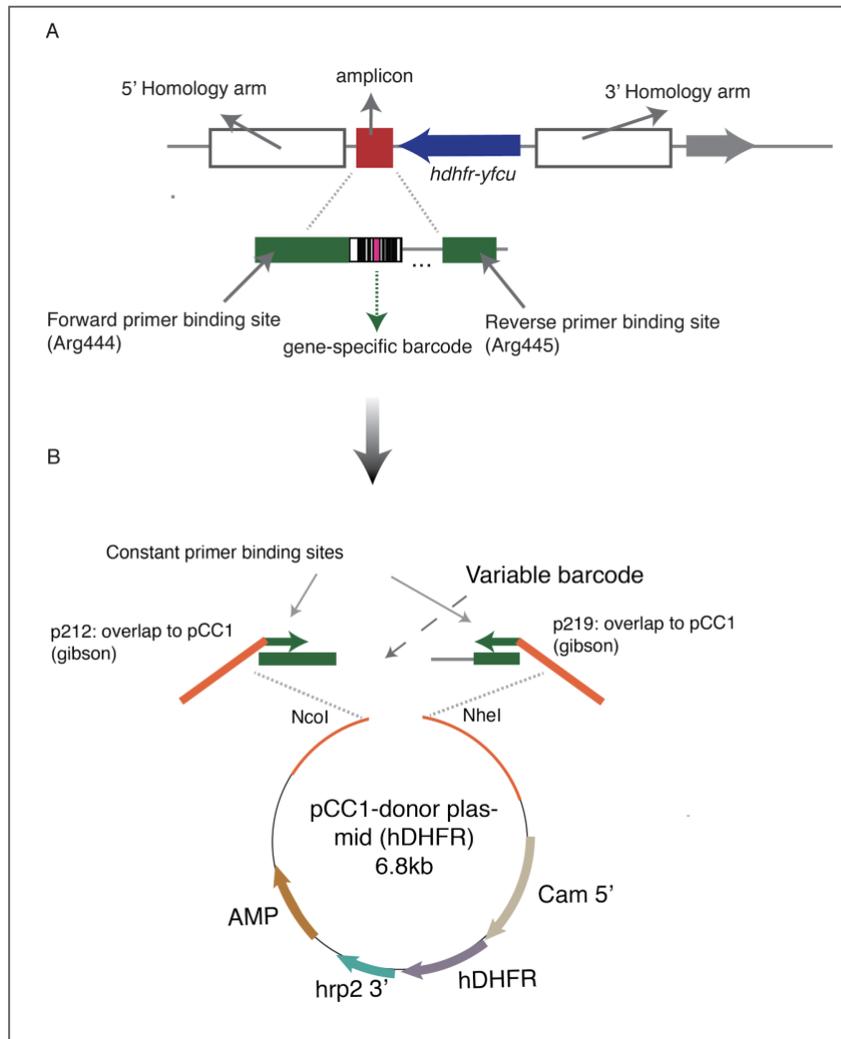
Figure 3.7 Adapting the PlasmoGEM approach to *P. falciparum*. A. Amplicon from PlasmoGEM vectors containing a unique barcode, which is 10-11bp long and not found in the genome of *P. falciparum*. Primers used to amplify the barcode with the annealing site for NGS are shown in green (Arg444/Arg445). Flanking the amplicon are the positive/negative selectable markers and the homology region targeting any gene of the *P. berghei* genome. **B.** Amplicon shown in 3.7B is amplified with p212/p219 which contain 20nt overlap (orange) to pCC1 for Gibson assembly.

Barcode cassettes from this pool of 96 different vectors were amplified using primers designed to amplify the 100 bp amplicon used for generation of *P. berghei* BarSeq sequencing libraries, but with the addition of sequences of overlap to the *P. falciparum* vector pCC1 described above (Fig. 3.2D) to permit Gibson assembly. As before, the Rh3 donor plasmid had been linearised

with NheII and NcoI, as marked in Fig. 3.3B. The one-step cloning reaction is shown in Fig. 3.7B, in which amplification of pooled barcodes from Fig. 3.7A was performed using primers containing a 20bp nucleotide overlap to pCC1 for Gibson assembly. The final assembled vector which was used for parasite transfection was transformed into One-Shot TOP10 competent *E. coli* cells. Several hundred colonies were obtained after the transformation, which were pooled together and plasmids extracted in a large midi-prep to obtain enough material for transfection. Multiple glycerol stocks were made from this prep, which were routinely used throughout the project.

The complexity of the plasmid prep was confirmed by Sanger sequencing. Fig. 3.8 shows the chromatogram of the plasmid pool, using sequencing primers p195 and p196. As expected there is a high-complexity region corresponding to the barcode region (brown highlight), indicating that multiple different pCC1 vectors, each containing a unique barcode, were present within the midiprep.
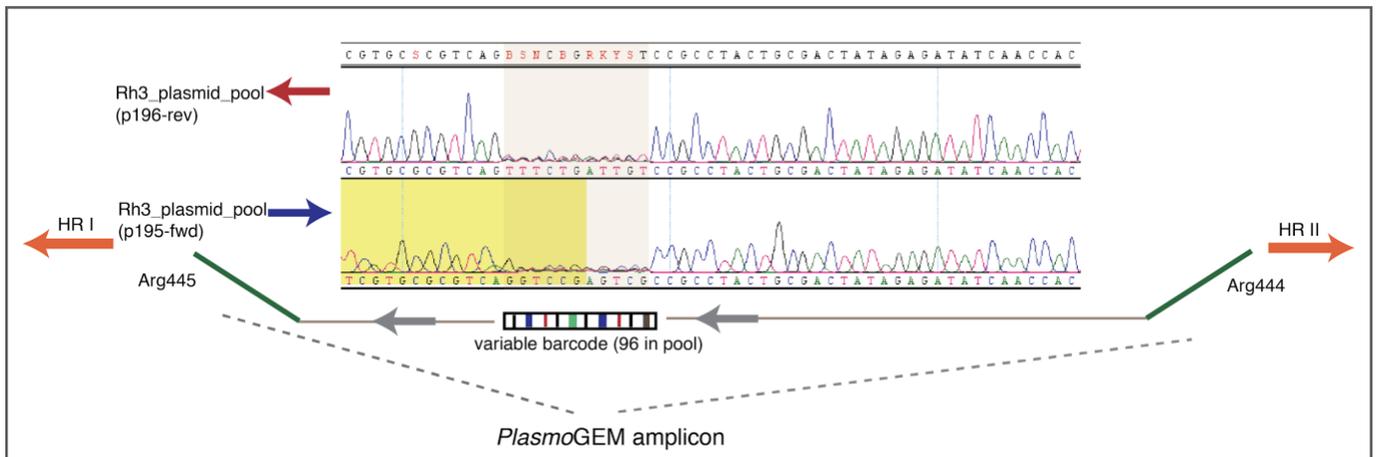
Figure 3.8 Genotyping of pool of barcoded vectors used for transfection. Sanger sequencing of barcode region ligated in between homology arms for *Pfrh3* in plasmid backbone pCC1. The high complexity of the pool is highlighted (light brown). Primer combination used was p196/p195.

To assess the relative abundance of different barcodes within the pCC1 pool, a direct amplification strategy was used, which consists of two PCR steps and a single round of purification of PCR products, to prepare libraries for Next Generation Sequencing (Cruaud *et al.*

2017). With the purified products, a nested PCR reaction was performed as shown in Fig. 3.9. The first set of primers, Arg444 and Arg445, are long oligonucleotides that amplify directly from the constant binding site and thus will amplify any barcode within those flanking regions. This product (PCR1 in Fig. 3.9) is approximately 100 bp long and contains binding sites for the second PCR that introduces the Illumina adapters, labelled in red. The latter was performed by using plasmid DNA from the pool, with PCR settings modified from (Gomes *et al.* 2015), as per requirements for HiFi KAPA polymerase: 95℃ for 2min, followed by 35 cycles of 98℃ for 20 seconds, 58℃ for 20 seconds and 72℃ for 30 seconds, and a final extension at 72℃ was performed for 1 minute. The number of cycles had been previously optimised to work from infected blood of *P. berghei* (Gomes *et al.* 2015), and appears to be sufficient to reach barcode saturation, and thus was selected for this approach.
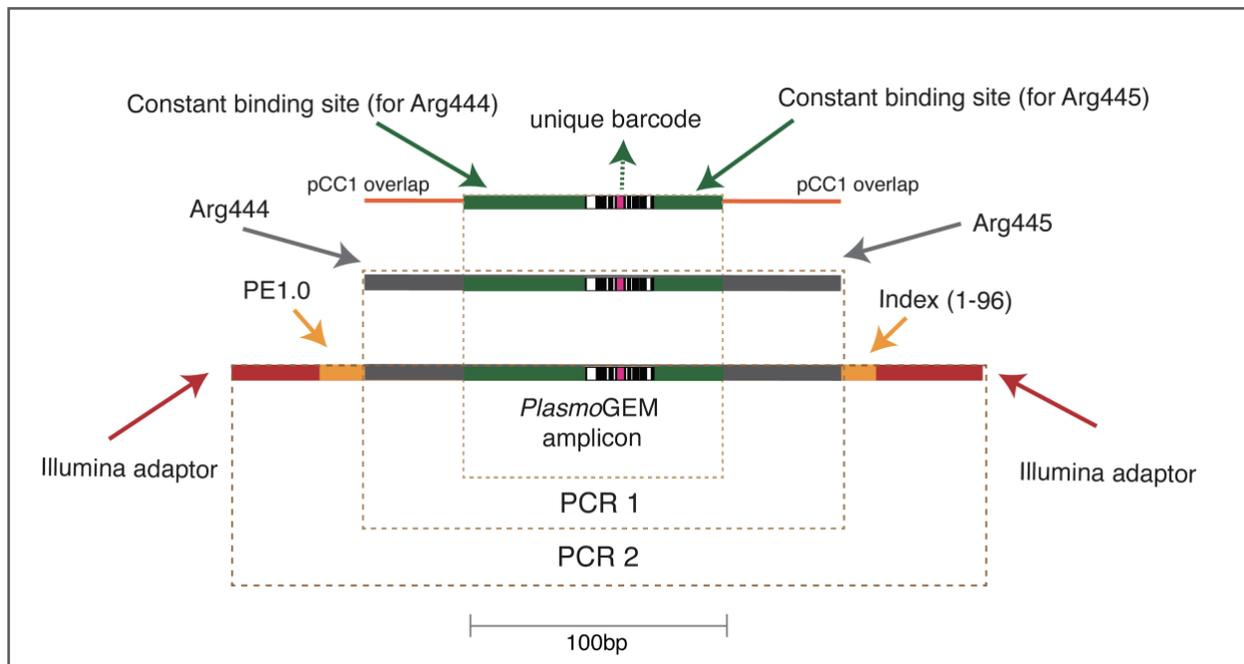
Figure 3.9 Library preparation for BarSeq. Method for generation of libraries compatible with Illumina platform MiSeq. A nested PCR approach uses primers Arg444/Arg445 to amplify from the constant binding sites (green). These primers have overhangs compatible with Illumina index tags PE1.0 and index 1-96 which anneal the Illumina adapters that bind to the flow cell (Gomes *et al.* 2015).

Because of the low complexity of the samples, due to the flanking regions of the barcode being the same for every vector, the order in which each base is read will essentially be the same for every cluster formed in the flow cell until the barcode is reached. This can produce errors during Illumina sequencing, so it is important to increase the diversity in the clusters as much as possible to avoid introducing errors in base calling during the imaging process. To overcome this issue, genomic DNA from the bacteriophage PhiX is added to the sequencing libraries, known as a spike-in. Increasing the concentration of PhiX in the sample mix increases the diversity of the clusters. This had been previously optimised for the *P. berghei* work, and the settings were left constant throughout this dissertation. A normal sequencing run for MiSeq, specifically for whole genome sequencing, will have up to 5% PhiX, but given the low complexity of the BarSeq samples 40-50% PhiX was routinely spiked into each sample. Another method for dealing with the low complexity was to use a low cluster density. Normally, for whole genome sequencing, a cluster density of approximately 8×10^5 clusters/mm^2 is used, however for the barcode sequencing runs, the final cluster density for these amplicons was 4×10^5 clusters/mm^2.

The raw fastq reads for two input pools that were sequenced were obtained directly from the sequencer and analysed. Both pools originated from the same midi prep but were sequenced on different days and using different multiplexing strategies, thus providing total read count libraries of different length. Fig. 3.10A shows the distribution of the raw counts after being counted from the fastq files. The green arrow shows that most barcodes for both pools are above one thousand reads, however the second input shows overall lower abundance for all the barcodes. In order to understand how the abundance would affect the overall complexity of barcodes, the proportion of each barcode in the experiment is shown in Fig. 3.10B for both pools, showing that even though the abundance of input pool 2 is lower in terms of absolute counts, the two proportions are consistent within sequencing replicates, with an R^2 of 0.98. Because all subsequent analysis is performed by assessing changes in the relative proportion of different barcodes in different experiments, variation in the overall read counts are not problematic. Given the sensitivity of the approach even barcodes at very low

abundance (10% of the most abundant barcode in the pool) have been shown to be detected and are above the sequencing threshold defined in Gomes *et al.* 2015 (Gomes *et al.* 2015).
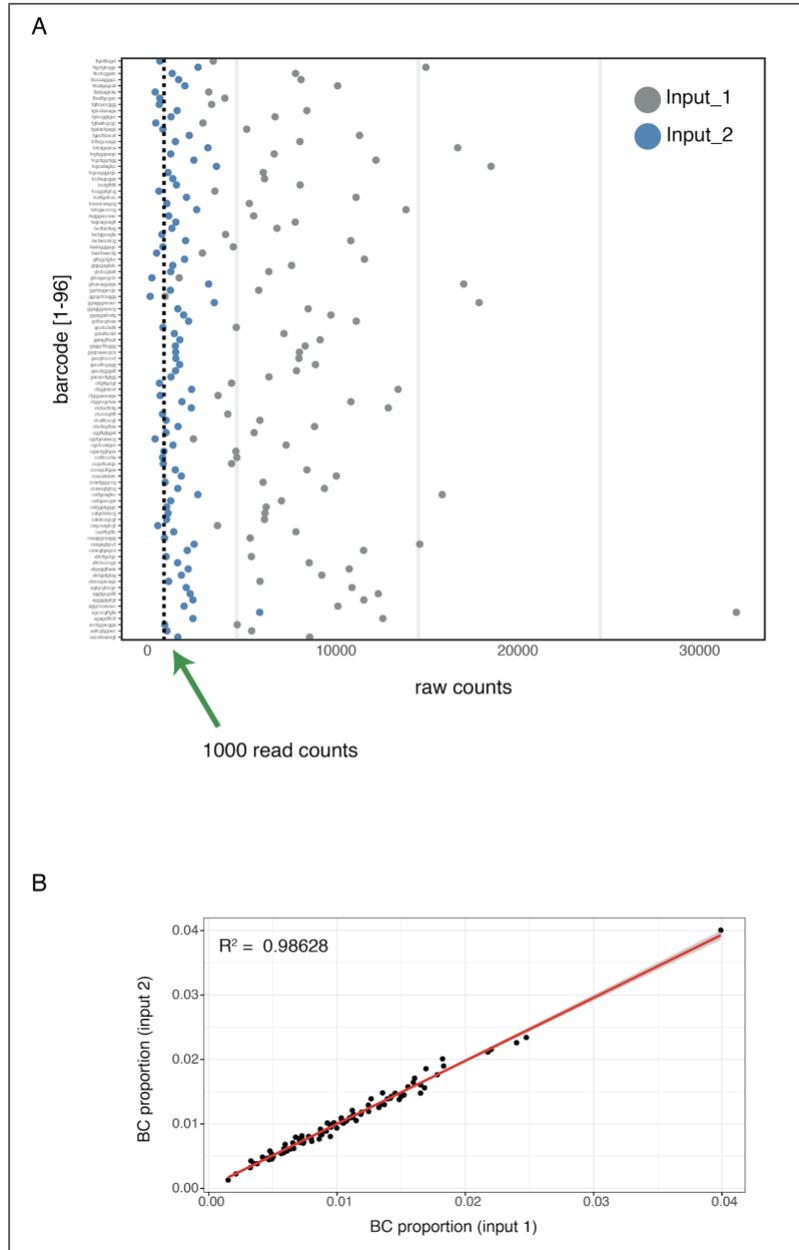
Figure 3.10 Plasmid pool sequencing and comparison. **A.** Raw counts for two pools sequenced as part of different multiplexed runs but originally from the same plasmid preparation. The y-axis are the barcode sequences present in the pools and x-axis indicates the raw counts. **B.** Linear correlation of the proportions between the two barcode (BC) input pools, showing an R^2 of 0.98.

3.3.4 A limited number of plasmids are taken up by *P. falciparum* when transfected with a complex vector pool

The efficiency of transfection in *P. falciparum* has been previously assessed using a luciferase-based approach, where a plasmid expressing a luciferase reporter was electroporated into infected erythrocytes. By performing a time course of luciferase activity, efficiency of transfection was assessed in a more sensitive manner than other methods, such as the time it takes parasites to become visible by microscopy after transfection, or to reach a certain parasitaemia (Skinner-Adams *et al.* 2003). Using the luciferase approach in combination with the best transfection protocol, the efficiency can reach up to 1.08×10^{-5}, a measurement equivalent to the ratio of parasites expressing luciferase at different time points post transfection until parasiteamia reaches 1%, to the input of parasites that got electroporated initially (Hasenkamp *et al.* 2012). However, these studies have the limitation that they do not provide any information about how plasmids get distributed throughout the parasite population - a few parasites could be taking up many plasmids, or each parasite could be only taking up one, and the luciferase activity would be unchanged.

I therefore used my pool of barcoded vectors, coupled with a high throughput sequencing approach, to gain a better understanding of the dynamics of DNA uptake in *P. falciparum*, and assess the "multiplicity of transfection", i.e. the number of unique plasmids taken up by each parasite. Barcode vectors were used to examine both the proportion of these vectors taken up by parasites, and how these uptake events were distributed over the parasite population.

Transfections were performed in the laboratory strains 3D7 (Transfections 1 and 2) and Dd2 (Transfections 3-5). 50ug of the precipitated plasmid donor pool described earlier were loaded into each transfection, which were performed using the same conditions used as in section 2.2.1.1. Once resistant parasites were detectable by microscopy after continuous selection with WR99210, cultures were expanded to 10mL. In order to get an accurate representation of the parasite population, the parasitaemia and haematocrit were kept at 3% and >5%, respectively. Genomic DNA was isolated, and the barcodes were amplified by the nested PCR reaction described (Fig. 3.9). The workflow for all barcoding experiments is shown

in Fig. 3.11, and consists of 3 steps: 1. transfection of plasmid pools, 2. recovery of transfected parasites, and 3. library amplification for high throughput sequencing.
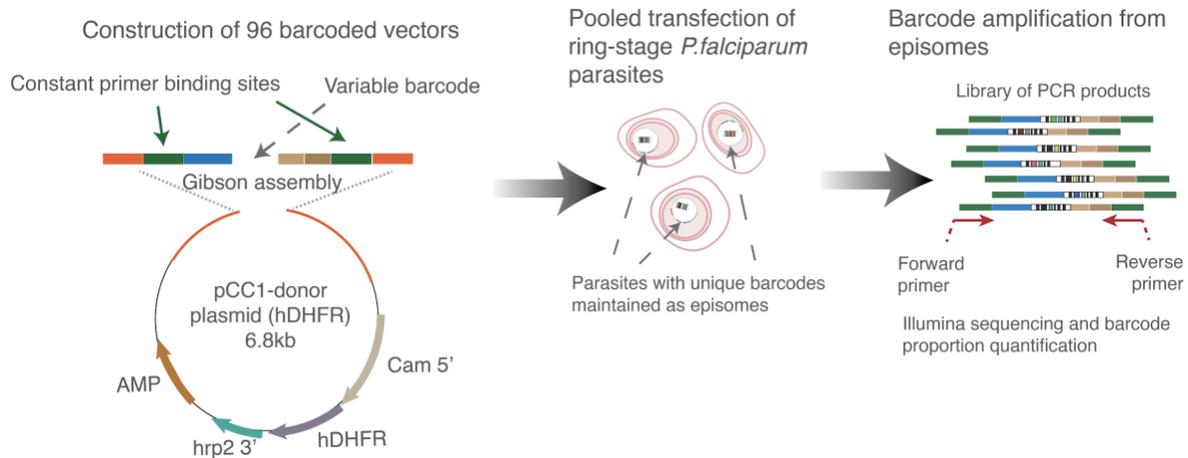
Figure 3.11 Graphical representation of a BarSeq approach for measuring transfection efficiency using pools of amplicons containing 96 unique barcodes. The three steps in the process are in between the gray arrows and are 1. transfection of pools of 96 barcoded vectors, 2. recovery of transfected parasites with continuous selection, and 3. library amplification for high throughput sequencing.

The first approximation for understanding the barcode complexity was to design a genotyping PCR amplifying genomic DNA extracted after parasites were visible by microscopy. These PCR products were then examined by Sanger sequencing, as shown in Fig. 3.12 which shows the chromatogram across the barcode region of the recovered episomes. A region of high complexity was observed in the barcode region, consistent with the recovery of multiple different barcoded plasmids post-transfection (Fig. 3.10).
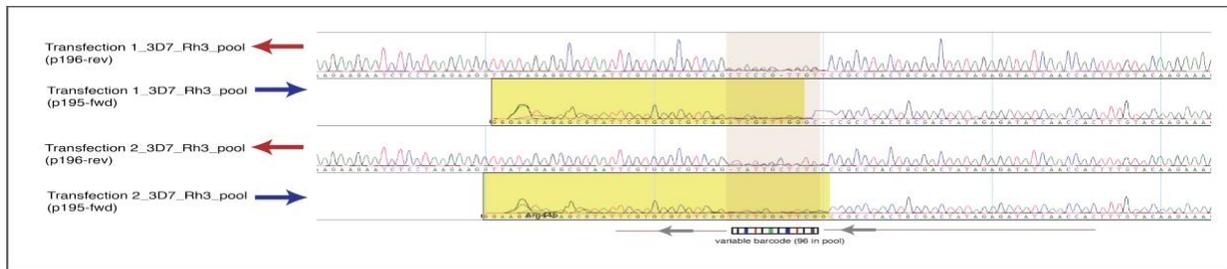
Figure 3.12 Genotyping of two independent transfections (Tx1 and Tx2). Chromatogram for the two transfections showing the barcode region inside the light brown box as being the only region with high complexity. Forward (p196) and reverse (p195) primers used for Sanger sequencing from episomes, and the yellow region reflects poor sequence quality at the start of the p195-primed reactions.

To more accurately measure the distribution of vectors maintained episomally, library preparation of PCR amplicons was generated as in Fig. 3.9. Analysis of the raw fastq reads was performed to measure the relative proportion of each barcode in each transfection. Figure 3.13A shows the barcode proportions in the five independent transfections, with the y-axis representing the proportion of the total counts for each barcode, sorted on the x-axis by their abundance in the input pool.

From this figure, it cannot be determined whether the most abundant barcodes in the input were the ones that got into the parasites most effectively to be stably maintained, as no clear correlation can be determined between the input abundance and recovery. Rather, these pilot studies were aimed at understanding the dynamics of transfection in *P. falciparum*. Fig. 3.13B-C shows the distribution of barcodes across the different transfections, which indicates the level of stochasticity between transfections. Fig. 3.13B shows the distribution of the log ratio of barcodes recovered in the different transfections, and the distribution of barcodes in the input pool (gray). To further investigate the high degree of stochasticity observed in the distributions, the difference between the log ratio in the input pool and the output (different transfections) was measured (Fig. 3.13C). From the results in Fig. 3.13C, the distribution of the difference between input and output suggests that the source of the stochasticity may result from the specific complement of plasmids taken up by the parasites from each independent transfection, not necessarily being the most highly abundant barcodes in the input the ones that get taken up. Alternatively, stochasticity may be introduced by different growth rates that individual parasite clones might have, and that competition might have occurred in the time

taken for the culture to reach a parasitaemia high enough for genomic DNA to be extracted. However, this latter possibility seems less likely given that all barcodes should have equal fitness. Regardless of the explanation, it can be concluded that transfections 3, 4 and 5 seem to have a distribution in which only a few barcodes comprise most of the DNA pool and are on the more positive side of the distribution, and that there is a higher amount of noise observed from the lower barcode proportions that represent less than 0.01% of the population.
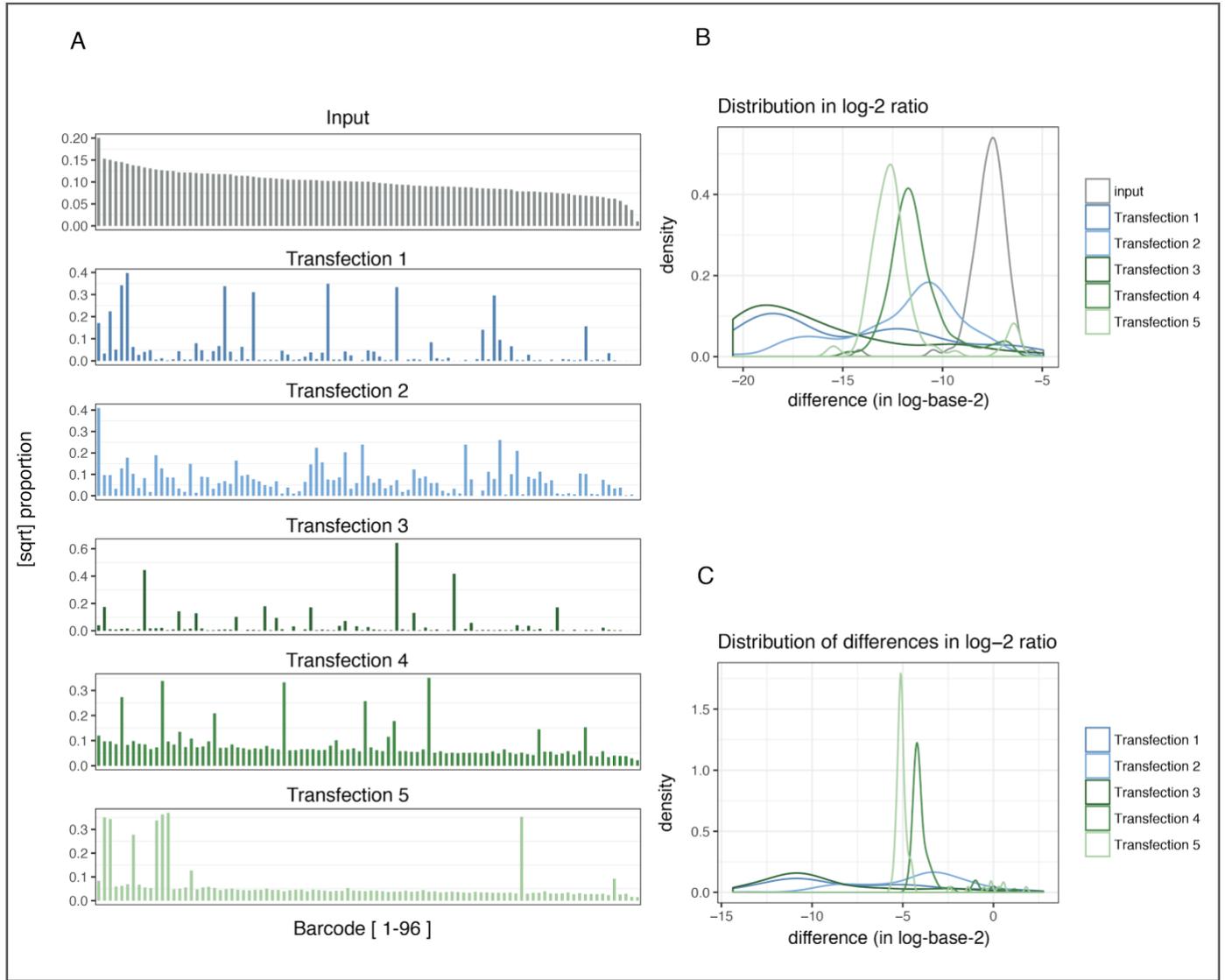
Figure 3.13 Barcode abundance and distribution measured with NGS. Transfections 1 and 2 were performed in 3D7 and 3-5 in Dd2. **A.** Transfection of a pool of 96 uniquely barcoded vectors. **A.** The data is sorted by the most abundant barcodes present in the input pool (first panel, gray). The y axis represents the proportion of each barcode in the total pool, as measured with Next Generation Sequencing. **B.** Distribution of the log2 of the barcode proportion across 5 transfections and input (gray) **C.** Density plot showing the difference between the log2 ratio of the input and the output (transfections 1-5). Transfections 1 and 2 were performed in 3D7 and 3-5 in Dd2.

To understand the relationship between abundance of barcodes in the input and the output across transfections, each of their proportion values were plotted against the proportion of the barcodes in the input. Fig. 3.14A shows the linear correlation of the different transfections to the input. From this analysis, generally the lower counts for the output are widely distributed across the x-axis (input), and this would suggest that the most abundant vectors were not necessarily those that were present at the endpoint of the transfection. The highest linear correlation between abundance in the input and the output was observed in transfection 2, (R^2=0.214), but this is a relatively weak positive correlation, and in some cases, there was essentially random uptake of barcodes, with R^2=0.0034 and R^2=0.017 for transfections 4 and 5, respectively.

To define how well barcode abundance correlated across transfections and input, a Pearson correlation matrix was built and its values are plotted in Fig. 3.14B. Transfection 2, as predicted, had the highest correlation to the input (0.47), as opposed to the lower ones from transfections 3 to 5 with correlation values between 0.02 and 0.016). Lastly, with the different replicates available, it is possible to infer that replicates of 3D7 have an overall lower stochasticity in plasmid uptake than Dd2, having a correlation to the input pool of over three times higher, with an average across transfection replicates of 0.36, and only 0.1 for Dd2.
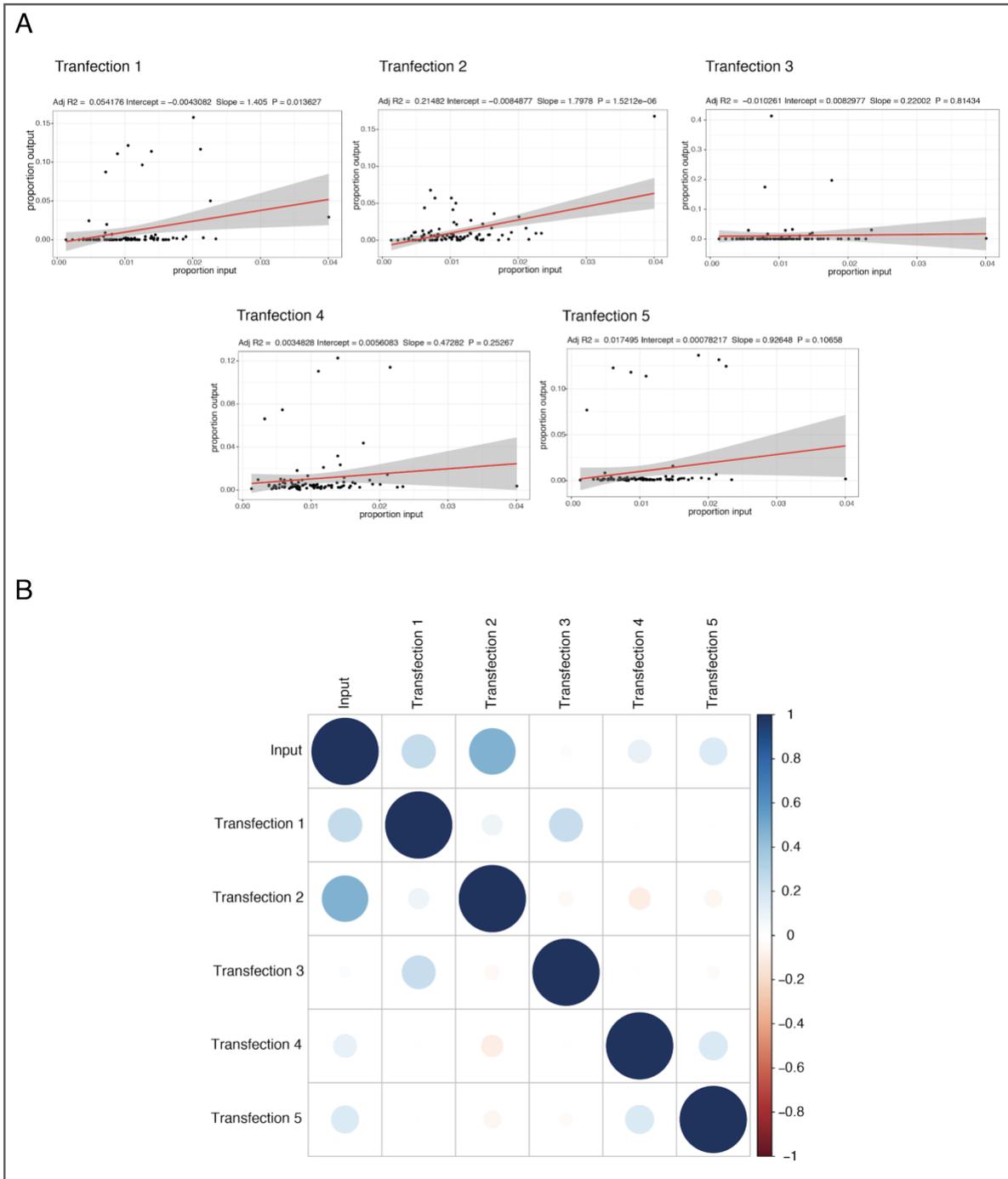
Figure 3.14 Correlation of barcode proportion across transfections and input. A. Scatter plot showing the independent variable (input), and the dependant variable (output) for each of the five transfections. **B.** Plot of values of correlation matrix of all transfections to the input pool and to each other.

Assuming the distributions of the DNA plasmids follow a bimodal distribution, Theo Sanderson designed an analysis in which these distributions are represented as a superposition of two normal distributions: the more positive distribution representing barcodes successfully taken up by the parasites and the more negative representing barcodes that did not make it in the transfection. For each transfection, a mixture of the two distributions was used to calculate the number of unique barcodes that were present in each transfection, and are shown for each transfection in Table 3.1. Additionally, a random sampling of barcodes was simulated taking the relative proportions in the input, and the numbers of DNA molecules that this translates to, which are also shown in Fig. 3.15.
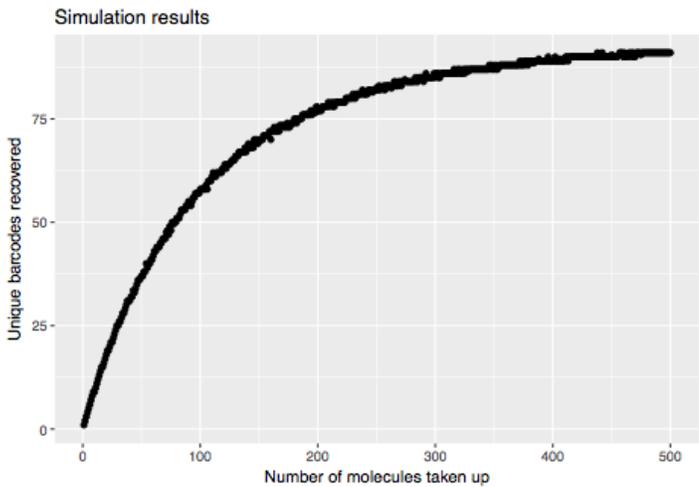
Transfection	Unique Barcodes	Molecules of DNA taken up
1	46	66
2	64	123
3	26	31
4	19	21
5	9	9

Figure 3.15 Simulation of uptake of DNA molecules. Simulation performed by Theo Sanderson from the Rayner group. Taking random sampling of 96 barcodes, assuming the bimodal distribution represented as two normal distributions that indicate whether the barcodes were successfully maintained in the transfection (the more positive distribution) or barcodes that never made it to the transfection (more negative values of the distributions)**.**

Table 3. 1 Results of the simulation from barcode distribution Table shows molecules that the number of barcodes correspond to, considering the molecular mass of a standard plasmid.

3.3.5 Co-transfection results in high multiplicity of transfection for those parasites successfully transfected

The fact that large numbers of unique plasmids, albeit many at relatively low abundance could be recovered from each transfection might suggest a potential for multiplexed screening. However, in *P. falciparum*, it is known that a single parasite is capable of taking up at least two plasmids when co-transfected, as has been shown for GFP/RFP colocalisation studies (M. C. S. Lee *et al.* 2008; Struck *et al.* 2008), or more recently two-plasmid CRISPR/Cas9 approaches (Ghorbal *et al.* 2014; Wagner *et al.* 2014). To establish whether the majority of parasites were taking up multiple plasmids, limiting dilution cloning was performed from the bulk parasites recovered in transfection 5 to examine the distribution of plasmids in individual parasites. This transfection was selected because it was the least complex in terms of barcode numbers. Barcode sequencing of cloned parasites (Fig. 3.16) showed that each clone contained multiple barcodes. In every one of the eight clones sequenced, 4 to 8 barcodes were detected, suggesting that all parasites had in fact taken up multiple plasmids. In fact, only two different combinations of barcodes were observed among the clones (clones 1 and 5 represent the two types), with most of the clones having the same barcode content as the sequenced bulk population. Even though the transfection efficiency is very low, there is clearly a non-random distribution of plasmid uptake, and a strong preference for individual parasites to take up multiple plasmids. This suggests that alternate approaches will be required for scaling reverse genetics in *P. falciparum* using a pooled vector strategy, as the pooled vectors will not be distributed randomly across the parasite population, and individual parasites will receive vectors targeting multiple genes.
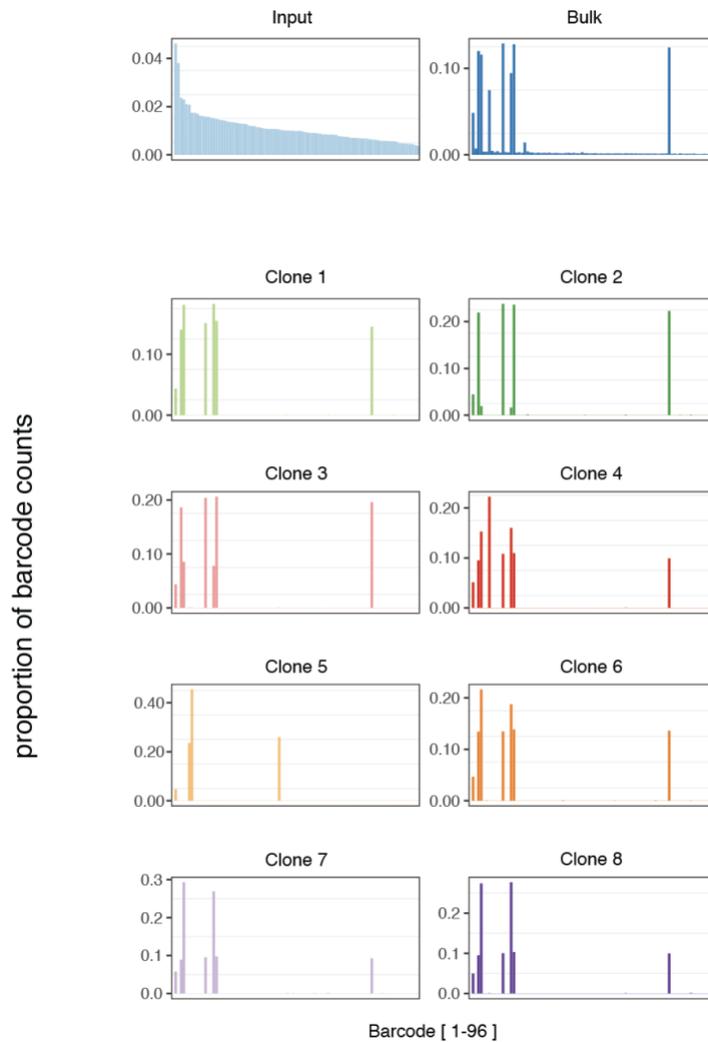
Figure 3.16 Distribution of barcodes across individual clones. Measurement of the complexity of episomes in individual parasite clones obtained with limiting dilution, showing two clone-types, clone 5 representing one clone type and the rest another type. The y axis is the proportion of each barcode in the total read counts.

3.4 Discussion

A key element of the planned approach to monitor parasite fitness in competitive growth assays is the ability to tag individual parasites with a unique barcode. Here, I established and targeted a non-essential locus (*Pfrh3*) as a safe harbour in which to insert a single 30bp

DNA barcode using CRISPR/Cas9 expressing vectors (Fig. 3.6A-B). This strategy was effective, and was extended in order to have a better understanding of transfection efficiency, as well as testing whether the barcoding strategy can be made scalable by simultaneous transfection of multiple barcodes, and then subsequently cloning out single barcoded parasites by limiting dilution. The results show that the complexity of stably maintained episomes can be used as a readout of transfection efficiency, by Sanger sequencing and by a high throughput sequencing strategy (Figs. 3.12, 3.13A). Remarkably, from the results, *P. falciparum*, as *P. berghei*, is able to take up multiple vectors (potentially a few dozen) when transfected in a ring-stage culture. However, the results and analyses also raise questions on whether it is possible to perform reverse gene targeting at the whole-genome scale, such as in other species. The results of using multiple transfections in two different laboratory strains (Dd2 and 3D7) and measuring the complexity after transfection by analysing next generation sequencing reads, indicate that individual parasites routinely take up multiple plasmids. These results have implications for future CRISPR-based approaches that rely on the introduction of libraries of gRNA plasmids, such as employed by Sidik *et al.* 2016 in *T. gondii* (Sidik *et al.* 2016), and indicate that further work will be required to understand transfection parameters in *P. falciparum.*

Even though there has been more than a decade after the first transfections were performed in *P. falciparum*, many factors related to the limitations in the efficiency at which DNA gets delivered to the parasite remain unknown. From this data, from the complex pool of barcoded vectors, an estimation could be made that for each transfection (one cuvette), approximately 60 million parasites were electroporated, and between them they took up only 9-120 molecules of DNA, as shown by the simulations given the distributions shown in Fig. 3.13B. Given this result, I was also able to show that independence in the uptake events is not possible. If plasmid uptake was completely independent, it would be expected that each of these molecules would be found in a different parasite, which translates to a transfection efficiency of approximately $\sim 1 \times 10^{-6,}$, consistent with previous observations (Hasenkamp, *et al.* 2012; Skinner-Adams *et al.* 2003). With these results on efficiency, the chances of a single parasite taking up 2 molecules would then be approximately 1×10^{-36}, and given the limited

number of parasites that were electroporated in the cuvette (60 million), that would be too low for any transfected parasites to be observed. Multiple plasmid uptake must therefore be the norm, and in fact, two-plasmid transfection approaches that rely on dual-selection markers are routinely used for co-localization experiments or delivery of Cas9 and donor plasmids (Ghorbal *et al.* 2014; Wagner *et al.* 2014). The clonal dilution performed show that parasites are routinely transfected with up to 5 molecules, suggesting that plasmid-uptake events are far from statistically independent.

There are a number of possible hypothetical bottlenecks in *Plasmodium* transfection. The data presented suggests that availability of DNA is not likely to be a crucial limitation, and it is possible that lower concentrations of DNA might even achieve superior results. It is also seems likely that the use of "two plasmid" systems, for instance for supplying a donor and a Cas9 vector, are unlikely to lead to significantly lower efficiencies than one plasmid systems with current transfection conditions. Further understanding these dynamics will be particularly important for the development of multiplexed CRISPR-screens, where low multiplicities of transfection will be crucial. This knowledge will be fundamental to the exploration of the parasite's genome and the unravelling of gene function as well as host-parasite interactions.

Chapter 4: Using a panel of genetically diverse barcoded parasites to assess fitness and antimalarial resistance: a proof of principle of BarSeq in *Plasmodium falciparum*.

4.1 Summary

The aim of this chapter was to establish a proof of principle of the application BarSeq to assess *P. falciparum* fitness and response to antimalarial drugs. First, I generated a panel of barcoded parasites using a scalable approach with the plasmid pools described in chapter 3. Second, I used this panel of barcoded parasites, which included strains from different parts of the world and with different sensitivities to antimalarial compounds, in pooled growth competition assays with and without drug pressure. Changes in the relative proportion of each barcode over time were quantitated using BarSeq high throughput sequencing, and relative changes in parasite growth within the panel in the presence or absence of antimalarial compounds was assessed.

4.2 Background:

4.2.1 Barcode sequencing for measuring growth at a high resolution

In the previous chapter, a method was developed for generating barcoded parasites. We showed that *Pfrh3* can be used as a safe harbour to insert these barcodes, and that plasmid pools can be transfected and recovered episomally. Once this was optimised, a first attempt to do competitive assays and measure fitness (as parasite growth rate *in vitro*) using pools was developed and will be expanded in this chapter.

Tracking growth rates using Barcode Sequencing is relatively new to apicomplexan parasites. The *Plasmo*GEM team at the Sanger Institute used this approach for high throughput growth phenotyping in *P. berghei,* in one of the first attempts in any *Plasmodium* species to use systematic reverse genetic approaches to generate knockout mutants and insert barcodes for

tracking their fitness (Bushell *et al.* 2017). In a related study undertaken by Sidik *et al.* 2016, a CRISPR/Cas9 based genome editing approach was used to target all genes across the genome of *Toxoplasma gondii.* In their study, which took advantage of the presence of the NHEJ pathway in *Toxoplasma*, a Cas9 nuclease was constitutively expressed from a plasmid, and a library of 10 sgRNAs per predicted protein-coding gene was provided in order to establish gene essentiality at the whole genome scale (Sidik *et al.* 2016). The sgRNAs themselves acted as the barcodes, with the absence of particular sgRNAs from the bulk culture indicative of gene essentiality. These two studies, therefore, used versions of Barcode Sequencing to track parasite growth, but with the purpose to understand gene essentiality, not parasite fitness or drug response.

By contrast in other systems, particularly the yeast model *Saccharomyces cerevisiae*, barcode sequencing has been used to track with high resolution the evolution of clonal lineages in order to quantify the process of adaptive evolution. Recently, Levy *et al.* 2015 use this approach to track the adaptive mutations that were acquired through asexual evolution, which were essential in determining or shifting the overall fitness landscapes of the population. By using a highly efficient barcode insertion strategy to generate thousands of barcoded lines, the appearance and expansion of subpopulations that have a fitness advantage could be tracked. The authors developed a system in which they could insert barcodes into a neutral region of the yeast genome using the Cre-*lox*P recombination system, which the authors referred to as a "landing-pad". Once the libraries of barcoded yeast coming from a clonal lineage were generated, they were pooled together and tracked over 168 generations. By this approach, the acquisition of beneficial mutations over time due to adaptive evolution was measured, providing an overview of the concept that independent beneficial mutations that appear in an asexually evolving population can act collectively after an accumulation of stochastically acquired independent mutations, and have a very strong impact in driving the dynamics of the population. This study also provided evidence of the power of approaches to perform high throughput phenotyping in asexually replicating populations (Levy *et al.* 2015).

4.2.2 Measuring fitness in *Plasmodium falciparum*

Parasite fitness is an oft discussed concept, particularly in the context of drug resistance mutations, but has proven difficult to quantitate. Due to the limitations in transfection efficiency for *P. falciparum*, a general feature of most assays that are aimed at measuring parasite fitness is that they rely on head-to-head competition comparing at most two lines. One approach to measuring parasite fitness in this way was to use a GFP-competitor line co-cultured with a non-fluorescent query line, which allowed a readout using flow cytometry. In one example, a drug-resistant mutant in elongation factor 2 was mixed in a 1:1 ratio with a GFP-expressing wild-type Dd2 strain. By performing co-culturing experiments, the authors were able to validate that evolved resistant mutants had a cost to parasite fitness (Baragaña *et al.* 2015).

Another study assessing parasite fitness used a real time PCR approach to track different epidemiologically relevant *pfmdr1* alleles, and measure their contribution to parasite fitness and to modulating drug response to antimalarials such as chloroquine and mefloquine. In this approach, competition assays between isogenic lines with the various *pfmdr1* alleles introduced by gene-editing approaches were set up, and growth of each line quantified by designing primers which would detect the specific polymorphisms in the different strains. The concentration of each allele was then used as a measure of relative growth in competition (Hayward *et al.* 2005).

More recently, the group of David Fidock used pyrosequencing to determine the contribution of different *Pfkelch13* alleles to fitness in isogenic lines with epidemiologically relevant polymorphisms generated using zinc-finger nuclease mediated gene editing, described in section 1.5.2.3 (Straimer *et al.* 2017). This method requires four different enzymes which all work to convert inorganic pyrophosphate released after nucleotide incorporation performed by a DNA polymerase (enzyme 1), to ATP by an ATP-sulfurylase (enzyme 2), releasing necessary energy for a third enzyme, a luciferase to emit light with the oxidation of a luciferin molecule after incorporation of a specific nucleotide provided. A fourth enzyme (apyrase) is required to degrade the nucleotides that do not get incorporated, and thus allow for the order of

nucleotides to be determined by the light emission of a given nucleotide (Ahmadian *et al.* 2006).

Overall, these methods have highlighted the importance of measuring fitness as part of the overall understanding of the complexity of drug resistance in *P. falciparum* parasites. However, all these approaches have only allowed for phenotyping two lines in parallel, which makes scaling up difficult. Therefore, I undertook to develop an approach to combine barcode insertion by CRISPR editing with Barcode Sequencing to measure *in vitro* parasite growth of multiple parasite lines in parallel.

4.2.3 Drug sensitivity assay in *Plasmodium falciparum*

Reliable drug sensitivity assays are essential to monitor any infectious disease. The effectiveness of a compound with known or unknown mode of action can be typically measured with *ex vivo* or *in vitro* sensitivity assays. These have been particularly relevant in regions with high prevalence of drug-resistant *P. falciparum* parasites, like Southeast Asia, where a continuous surveillance of changes in drug susceptibility remains crucial. Because of the emergence and spread of parasites with an increased tolerance to the current frontline antimalarial artemisinin, the following are crucial for surveillance of resistance to both artemisinin and partner drugs: 1. monitoring molecular markers of resistance (if known), and 2. phenotypic validation using drug-sensitivity assays like pulse-based ring-stage assay (RSA$_{(0-3h)}$) or standard 48-72h inhibitory concentration assays to pinpoint any change in the tolerance levels of parasites to particular compounds over time (Lim *et al.* 2013).   The RSA$_{(0-3h)}$, developed by Witkowski *et al.* 2010 and described in section 1.4.5.2, allowed for characterisation of parasites originating from Cambodia and showing a higher survival to artemisinin treatment, highlighting the resistance mechanisms present in these parasites, at the early ring stages (Witkowski *et al.* 2010). An alternative assay was developed by the group of Colin Sutherland, and showed equivalent results to the standard RSA$_{(0-3h)}$. By pulsing with DHA for 6 hours with the different concentrations used in a standard 48-hour inhibitory concentration assay, a shift in survival was shown to be associated with carrying a mutant allele of *pfap2-mu*, another candidate of artemisinin resistance in Africa. This assay also provided the

advantage of not requiring a tight synchronization as the RSA$_{(0-3h)}$ does (Henriques *et al.* 2015). *Ex vivo* drug sensitivity assays rely on the short-term maturation of ring stages extracted from a patient infected with malaria and have been essential for determining the status of resistance in the field. However, throughout this dissertation, *in vitro* assays were performed by using both flow cytometry and microscopy from Giemsa stains, as described in section 2.4, which rely on culture-adapted parasite strains that are maintained (Russell *et al.* 2013).

In this chapter, approaches to integrate both fitness and drug response measures, using pools of barcoded parasites, were performed and the phenotypes were validated with standard 72-hour inhibitory concentration assays for measuring tolerance to partner drugs, or by RSA$_{(0-3h)}$ for artemisinin.

4.3 Methods and Results

4.3.1 Generating barcoded parasite clones from different genetic backgrounds

I used the barcoding approaches developed in the previous chapter as the starting point for generating parasite strains carrying a barcode into the *Pfrh3* locus, by transfecting the plasmid pools described. The issue of multiplicity of transfection, with multiple barcode plasmids being delivered to a single cell, will not impact our strategy as only one barcode can be inserted at the target site in a given line. To tag parasite lines, the plasmid shown in Fig. 3.4 expressing Cas9 and a guide RNA, and the plasmid pool shown in Fig. 3.7B bearing diverse barcodes, were co-transfected into different genetic backgrounds.

The rationale for transfecting multiple barcodes was to facilitate to scale up of library generation. Depending on the efficiency of integration, from a single transfection multiple clones bearing different barcodes could be isolated via dilution cloning, resulting in individually barcoded parasites. A total of 50ug of the mixed plasmid pool and the plasmid for Cas9 expression were transfected using the protocol in section 2.2.1.1. Parasites were selected for with WR99210 (donor plasmid) for six days, and on average it took 3-5 weeks for parasites to become visible by microscopy. At this point they were expanded and cloned by limiting dilution.

The time from the point of electroporation until parasites were visible was highly variable. Fig. 4.1 summarises 9 different successful transfections in the 3D7 background, showing the number of days until parasites were visible by microscopy on the y-axis. From these results, it is evident that transfection outcomes, at least with this transfection method, are remarkably variable and potential stochasticity may be unavoidable, as was previously observed with the episomal pools.
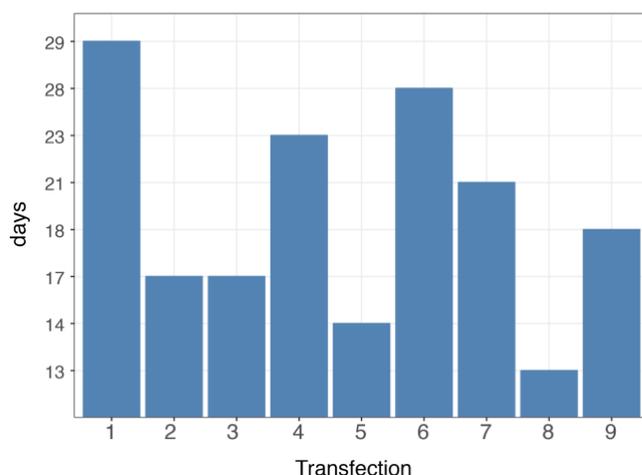
Figure 4.1 Length of transfections for editing of *Pfrh3*. Independent transfections were performed on a standard laboratory strain (3D7), with the number of days until parasites were visible by microscopy shown on the y-axis.

A set of genotyping reactions were designed to confirm the integration of the barcodes after transfection, including an additional reaction to ensure that the nested PCR would not produce any remaining episomes. To confirm their presence, and estimate the efficiency of barcode integration, the PCR reactions shown in Fig. 4.2A were performed. Fig. 4.2B shows an example of a bulk culture post-transfection, with barcode integration detected by amplifying from within the amplicon (p239) and a primer outside the homology region contained within the plasmid (p191). Little to no wild-type locus was typically observed (lane 2). Lanes 3 and 4 show reactions to detect the two plasmids that were co-transfected, pCC1 and pDC2 containing the donor region for homologous recombination and the Cas9 and gRNA, respectively. In order to confirm the integration efficiency and the complexity of barcodes, a first PCR was performed

using p191/p194 (marked with the blue arrows), represented with the dashed square in Fig. 4.2B. This 3kb product flanks the homology regions and thus does not capture episomal donors on pCC1. Fig. 4.2C shows the chromatogram of two independent transfections with the barcode pools integrated into the native locus, *Pfrh3*. From this figure, not only a high complexity can be predicted from the variable barcode region but it also shows high recombination efficiency. Both transfections shown were performed in a 3D7 background, and once parasites were visible, dilution cloning was performed in order to isolate individually barcoded parasite clones which could then be used in competition assays. Besides this reference strain, additional strains (listed in table 2.1) were selected based on drug-resistance profiles to be included in the pilot BarSeq experiment.
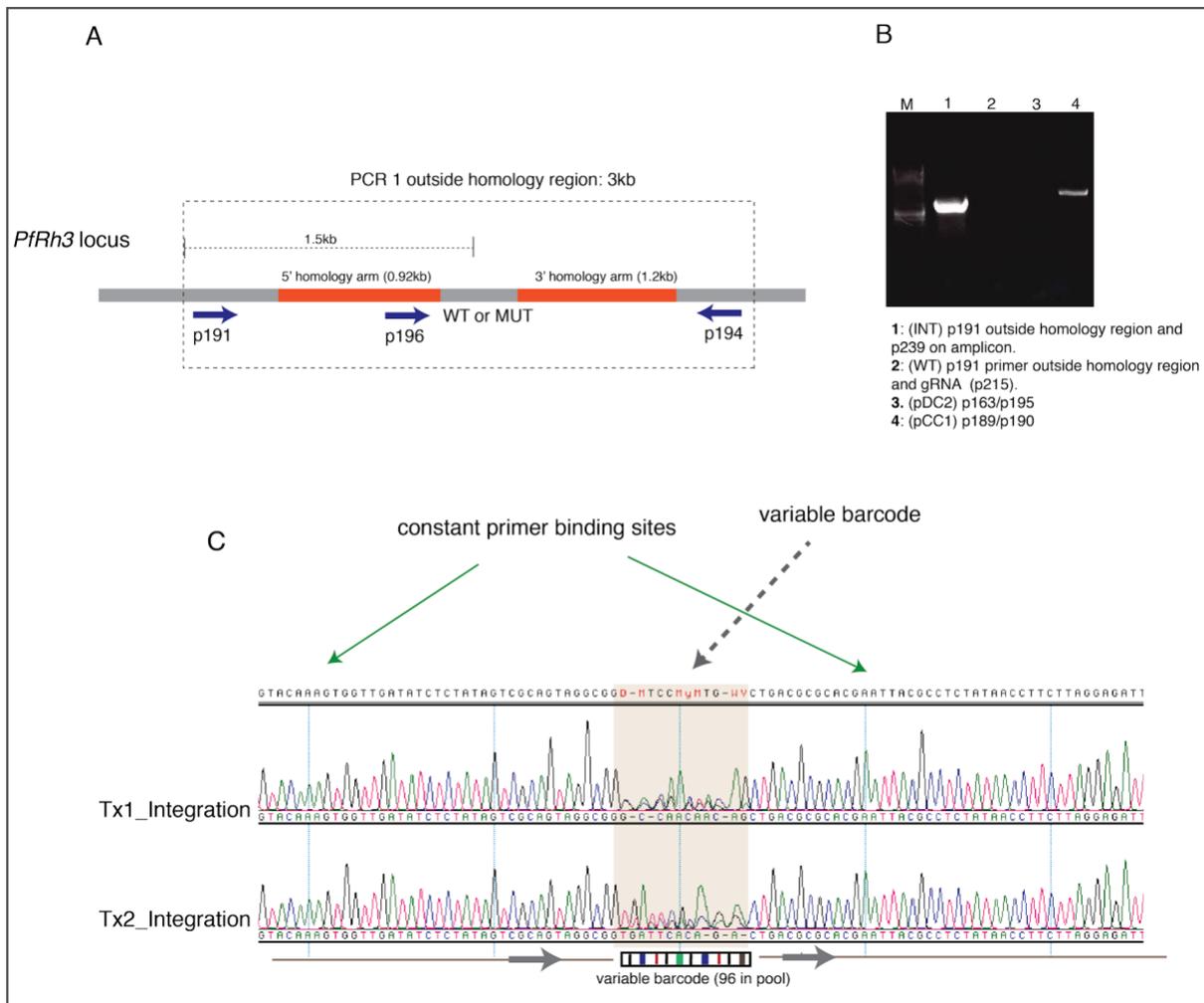
Figure 4.2 Genotyping strategy and integration efficiency of barcode pool. A-B. Genotyping strategy for integrated barcodes into *Pfrh3*. A long PCR (1) was performed by using primer combination p191/p194 (3kb) and used for Sanger sequencing shown in C. Forward primer p191 was used in combination with either p215 (WT locus) or p239 (edited barcode) for genotyping by PCR (1.5kb product). An example of a parasite line with near 100% editing is shown in (B). Additional PCR reactions were performed to detect the presence of the pDC2-Cas9 plasmid (p163/p195) or the pCC1 donor plasmid (p189/p190). **C.** Chromatogram showing a high complexity region indicating multiple barcodes are present in the bulk culture. This product is flanked by constant primer binding sites for nested PCR reactions described earlier to allow amplicon sequencing by NGS.

The next step was to generate the library for Illumina Sequencing using the nested PCR reaction described in Chapter 3 (Fig. 3.9). The reaction was performed by using as template the 3kb PCR product from outside the homology region for *Pfrh3*, as shown in 4.2A, again to ensure that only integrated barcodes were amplified and quantitated. After completion of the nested PCR, purification and dilution to 4µM were performed for Next Generation Sequencing. Fastq

files were analysed as before and the square root of the barcode proportions of the input pool and the two independent transfections are shown in Fig. 4.3. From these results, a range of 5 to 7 uniquely barcoded parasites were recovered, and in order to isolate them, the bulk population was cloned by limiting dilution (section 2.1.3), and clones containing unique barcodes were confirmed by Sanger sequencing.
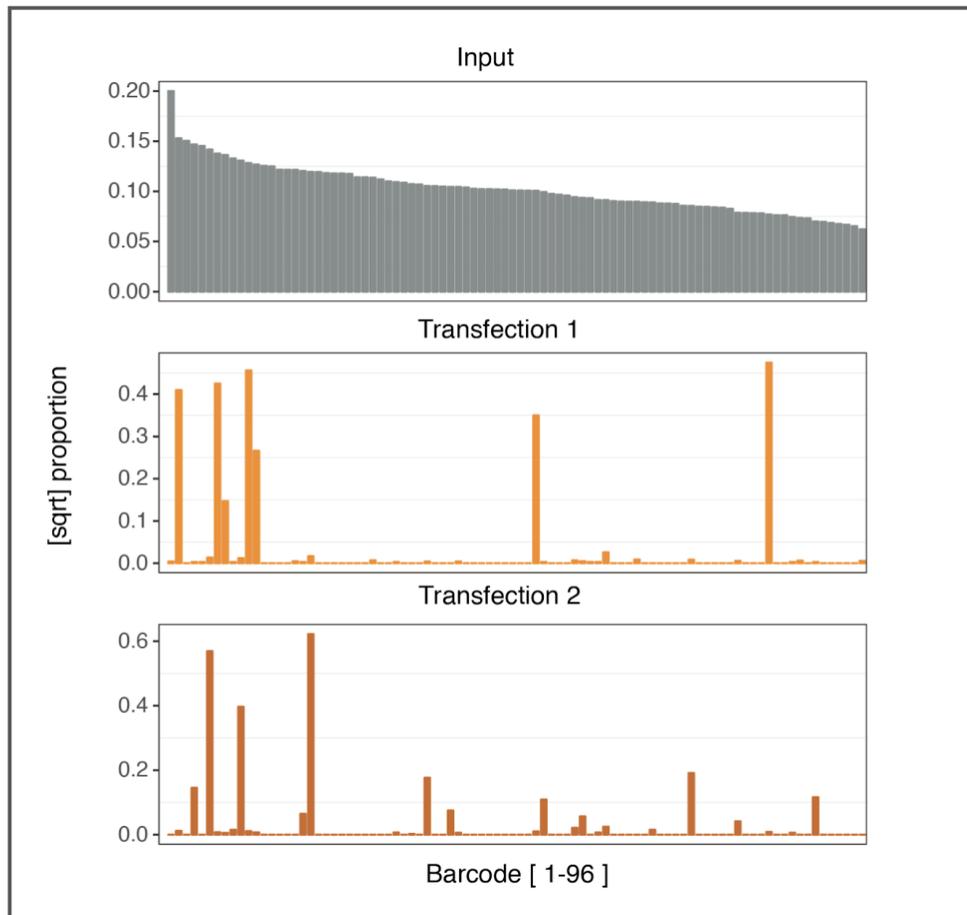
Figure 4.3 Integrated barcodes measured by NGS. Transfection of a pool of 96 uniquely barcoded vectors together with the pDC2 plasmid expressing Cas9. The data is sorted by the most abundant barcodes present in the input pool (first panel, gray). The y axis represents the square root of the proportion of each barcode in the total pool, per individual transfection as obtained with Next Generation Sequencing.

The same strategy was used to barcode other strains listed in Table 4.1. Overall, unique barcodes were introduced into three clones of 3D7 and two clones for V1/S. Unfortunately, the

efficiency of integration for the Cambodian parasite (CAM) was lower and even after screening multiple clones, only one barcoded line could be obtained.

Table 4.1 Barcoded parasite strains. A barcode ID was assigned based on the order of barcodes on the original PlasmoGEM PbSTM_21 plate from which they were amplified, and that was correctly matched with the integrated barcode from the pool.

Strain	Clone	Origin	Barcode ID	Barcode Sequence
3D7	A	Netherlands (I)	15	GTCTCCGTATT
3D7	B	Netherlands (I)	24	TAATAGGGAGC
3D7	C	Netherlands (I)	26	TAGCAGAAGTT
V1/S	A	Vietnam	1	TCTTCTGAATCA
V1/S	B	Vietnam	23	GACATGGGATT
PH0212-c (CAM)	A	Pursat, Cambodia	7	GTGAGAGTATC

(I) probably West Africa, clone of NF54 isolated in the Netherlands

4.3.2 A pilot study of BarSeq in *P. falciparum*

To establish the parameters for BarSeq, two experimental set-ups were tested. To examine the ability of BarSeq to detect inherent growth differences between strains, a long-term co-culture experiment was performed over 30 days (4.3.2.1). A shorter duration set of experiments were performed to test the impact of drug pressure on barcode abundance (4.3.2.2). The barcoded parasite clones listed in Table 4.1 were mixed together in equal proportions. In order to adjust their parasitaemia, flow-cytometry was used to obtain an accurate measurement of live parasites by co-staining with both Mitotracker Deep Red and Sybr Green I. The parasitaemia was adjusted so that each of the six clones would be initially present at 0.5% parasitemia out of the total 3% of the mix, which was then used as the initial time point after confirmation of total parasitemia by microscopy. All competition assays were performed in triplicate unless otherwise specified, and maintained for varying durations, ranging from hours to several weeks depending on experimental design. Figure 4.4 summarises the experimental design up to the library of PCR products for sequencing.
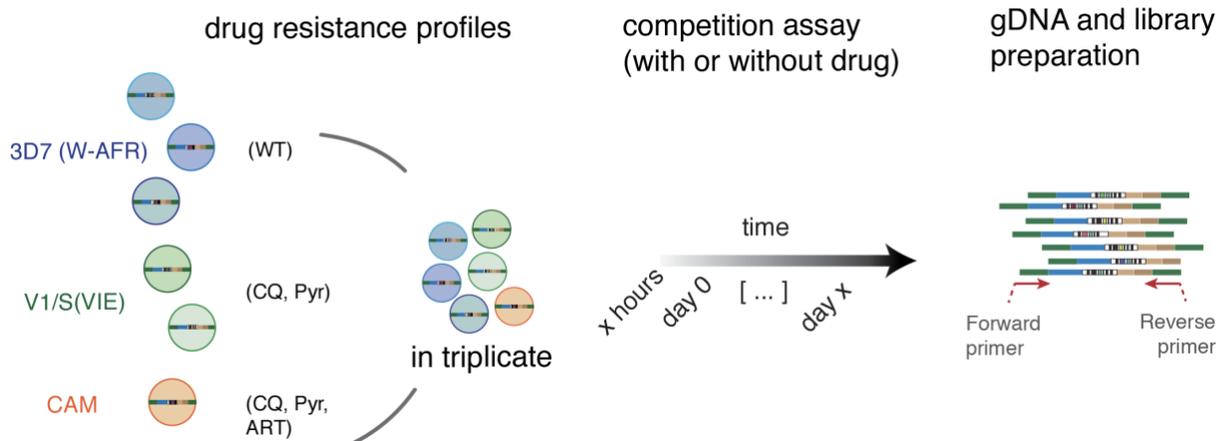
Figure 4.4 General experimental design for Barcode Sequencing experiments. Different parasite clones of genetically diverse strains with different drug sensitivity profiles were pooled together in triplicate. A competition assay was performed for varying lengths of time depending on the experiment. Genomic DNA was extracted at each timepoint and the nested PCR reactions described above were performed prior to loading onto a MiSeq sequencer. The CAM strain is the only *Pfkelch13* mutant, carrying the C580Y mutation and having an increased survival to artemisinin (Ariey *et al.* 2014).

4.3.2.1 A 30-day competition assay shows inherent fitness differences across strains

In order to understand whether there were inherent differences in the *in vitro* fitness across strains, a competition assay was performed over a one-month period. The assay revealed that the non-3D7 strains, V1/S and CAM, were almost completely outcompeted by 3D7 after 29 days of co-culture (Fig. 4.5). By comparing the barcode proportion of the total counts, a significant change in proportion between 3D7 and non-3D7 strains was observed. By contrast, the growth rates across three independently barcoded clones of 3D7, presented as 3D7A-C, were similar across the one-month assay. As it was a first pilot for BarSeq, only PCR duplicates were measured, therefore all further assays included biological duplicates or triplicates. The last time point of the assay was on day 31, however for the analysis, data was only considered until day 25. This is based on previously optimised parameters for *P. berghei* in which any measurement with counts accounting for less than 10% of the most abundant barcode as part of the total library of sequencing reads were removed.
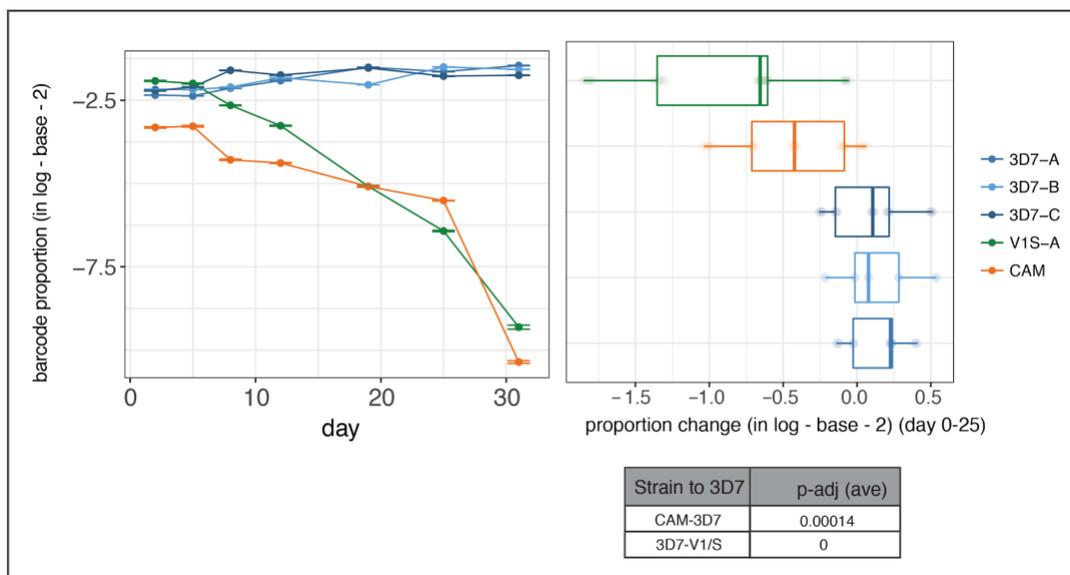
Figure 4.5 Competition assay in absence of drug over 31 days. Pilot of competition assay of three independent strains, with internal replicates for 3D7 as three uniquely barcoded clones in order to establish whether the internal replicates had similar growth rates. Second panel is the change in proportion from day 0 until day 25 and below the adjusted p-value after comparison of means from an analysis of variance (ANOVA) and TukeyHSD test. Technical duplicates were performed for this assay.

4.3.2.2 Change in barcode proportions reveal drug resistance phenotypes

To determine if BarSeq could be used in short-term assays to identify drug-sensitivities, triplicate cultures were initiated containing the barcoded clones described above, synchronised at early ring stages. The pooled cultures were grown in the presence or absence of different concentrations of chloroquine for 18 consecutive days, with samples isolated and extracted every four days. The assays were performed in 6-well plates at 3% hct, and at each timepoint 4mL at 5-8% parasitaemia was collected, lysed with 0.15% saponin, and used for DNA extraction and BarSeq. The remaining culture was monitored and adjusted such that the parasitaemia would be in the 1-5% range. BarSeq samples were prepared as in Fig. 4.4 and were multiplexed onto one MiSeq lane. To determine the appropriate chloroquine concentration, standard 72h drug-sensitivity assays were performed (described in section 2.4.1) and were compared with

previously published IC_{50} values, if available. Fig. 4.6 shows a standard inhibitory curve for the three strains used.

Based on the IC_{50} obtained from this analysis, and confirmation with the published data (Ref IC_{50} in Fig. 4.6), two concentrations were used for the BarSeq approach based on the IC_{50} of the chloroquine-sensitive strain, 3D7. For the strain from Cambodia (CAM), a reference IC_{50} was not available, however because this strain was used in further experiments, the IC_{50} value was confirmed by performing additional drug sensitivity assays, which will be relevant for the next chapter.
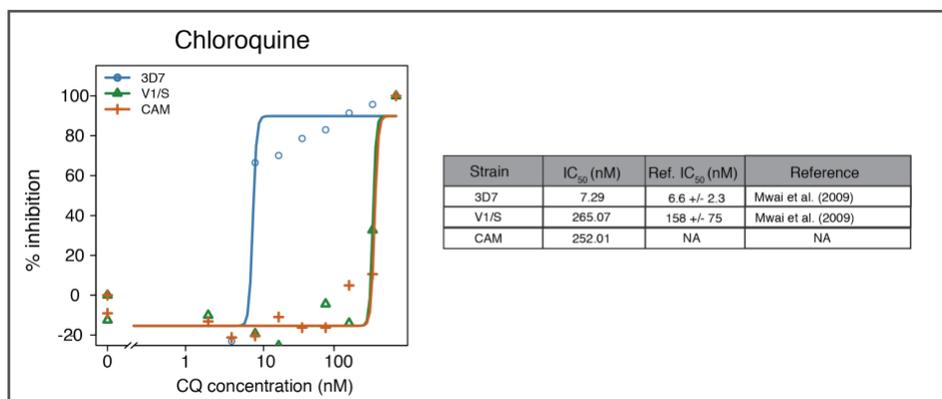
Figure 4.6 Chloroquine standard dose-response curve. Dose response curve for strains used in competition assays measured by using a dose response model (drc) (Ritz *et al.* 2015). Second panel showing the experimental values obtained and the published values (Ref. IC_{50}), both of which were considered for design of BarSeq experiments. Technical duplicates were performed for the assay, with three biological replicates showing the similar IC_{50} curves.

Figure 4.7 shows BarSeq analysis of relative growth of each barcoded line in the presence of $1\times$ and $3\times$ IC_{50} concentrations of chloroquine for the 3D7 strain. The upper panel shows the growth curves, with the y-axis representing the log of the proportion of each barcode in the total counts per replicate, per day. Overall, the internal replicates of 3D7 and V1/S, represented by unique barcode sequence per clone, showed no difference in growth as expected, whereas some differences were observed when comparing 3D7 to the resistant strains V1/S and CAM (p-value 0.02 and 0.03, respectively). The bottom panel in Fig. 4.7A shows the change in proportion over time, and an analysis of variance (ANOVA) coupled with a Tukey HSD test to compare the means across groups between days until day 13, with the

significance values listed in Fig. 4.7B. The final sample on day 18 was not taken into account as the counts for all 3D7 barcoded clones were below the threshold in the high chloroquine treated samples. As expected, the phenotypes observed by BarSeq were consistent with the differences in drug IC_{50}, showing a significant difference in the means between the chloroquine resistant (V1/S and CAM) and sensitive (3D7) strains at sub-lethal concentrations, with p-values < 0.01. This provides evidence that pressuring the mixed parasite pool at a relatively low chloroquine concentration, 3-fold over IC_{50} of the sensitive strain, was sufficient for clear phenotypic characterisation within two weeks of drug exposure.
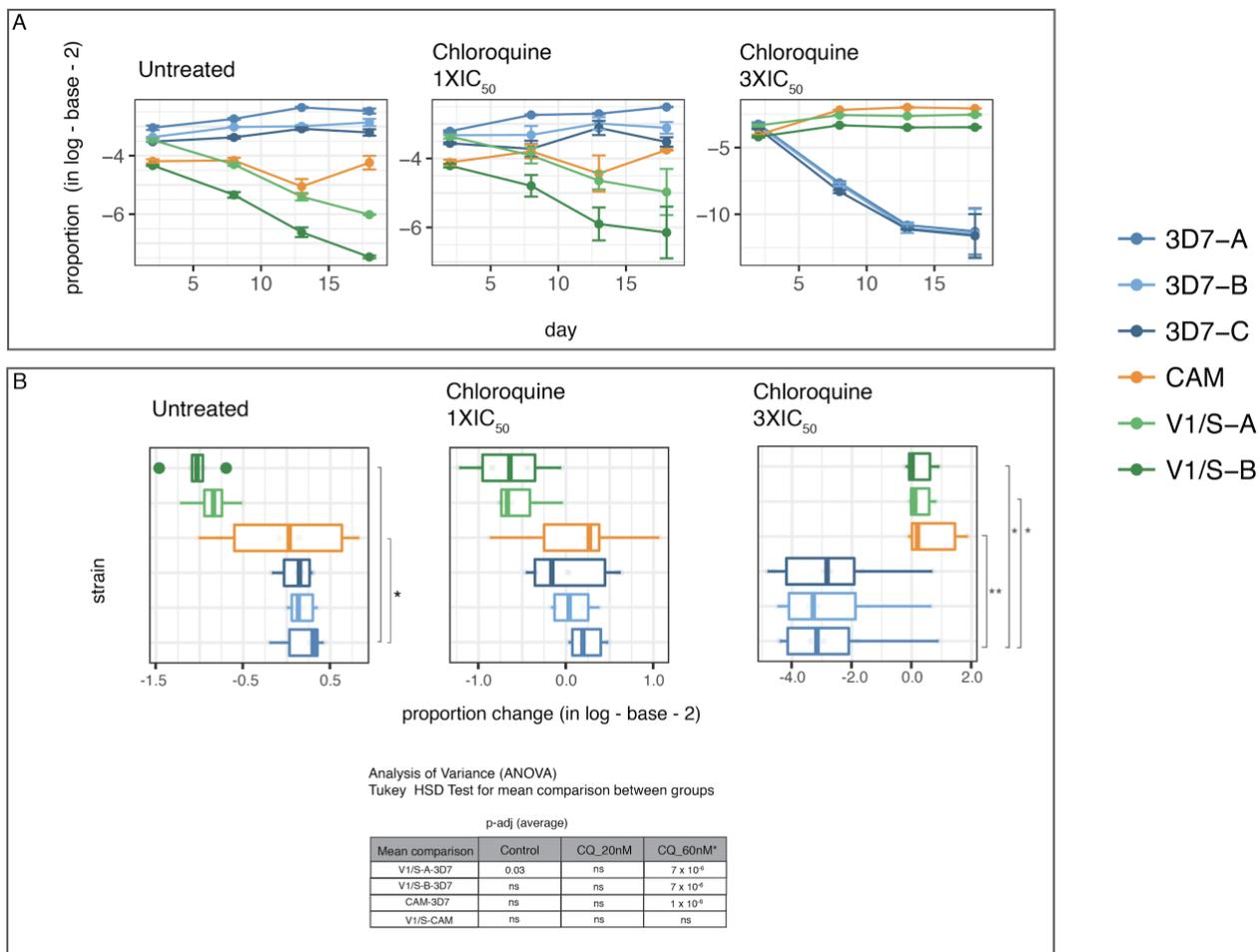
Figure 4.7 A. BarSeq of the pool of five clones, in the absence of drug and with increased chloroquine concentration.

The x-axis corresponds to time (in days), and the y-axis represents the log2 of the proportion of each barcode in the total read counts obtained. Left panel is the untreated control and middle and right panels show the increasing concentration of chloroquine based on the values obtained in Fig 4.6, for 1× and 3× the IC_{50} for 3D7. **B.** Change in proportion of each barcode over 18 days of competition. **C.** Analysis of variance across groups and mean comparison using a TukeyHSD for all timepoints except day 18, with significance values shown as the adjusted p-value. Technical triplicates were perfomed for this assay, and two biological replicates confirmed the same results.

In order to understand whether the method would be consistent in discriminating changes in response to higher dosages over shorter time frames, a linear regression was performed for each growth curve, and the slope of that regression at each concentration is shown in Fig. 4.8B. As expected, the slope of the sensitive strain, 3D7 decreases as the concentration increases, indicating the approach can also be used to determine the effect of

122

increasing concentrations on survival. The three concentrations used were equivalent to 2× to 10× the IC_{50} for 3D7. To estimate whether there was a difference in the rate at which the less tolerant strain would become outcompeted at different drug concentrations, a linear regression model using R was performed for each parasite clone, and the log value of the slope is shown in Fig. 4.8B. On the y-axis of Fig. 4.8B, the log of the value of the slope for all the time points is measured by a linear regression for the different curves. Because it is shown in log base 2, the positive curves corresponding to the chloroquine-tolerant strains have a slope greater than 1, in contrast with less than 1 for the sensitive clones. As expected, the difference in the slope between 2× and higher concentrations was significant, whereas between 5× and 10× IC_{50}, no significant difference could be determined.
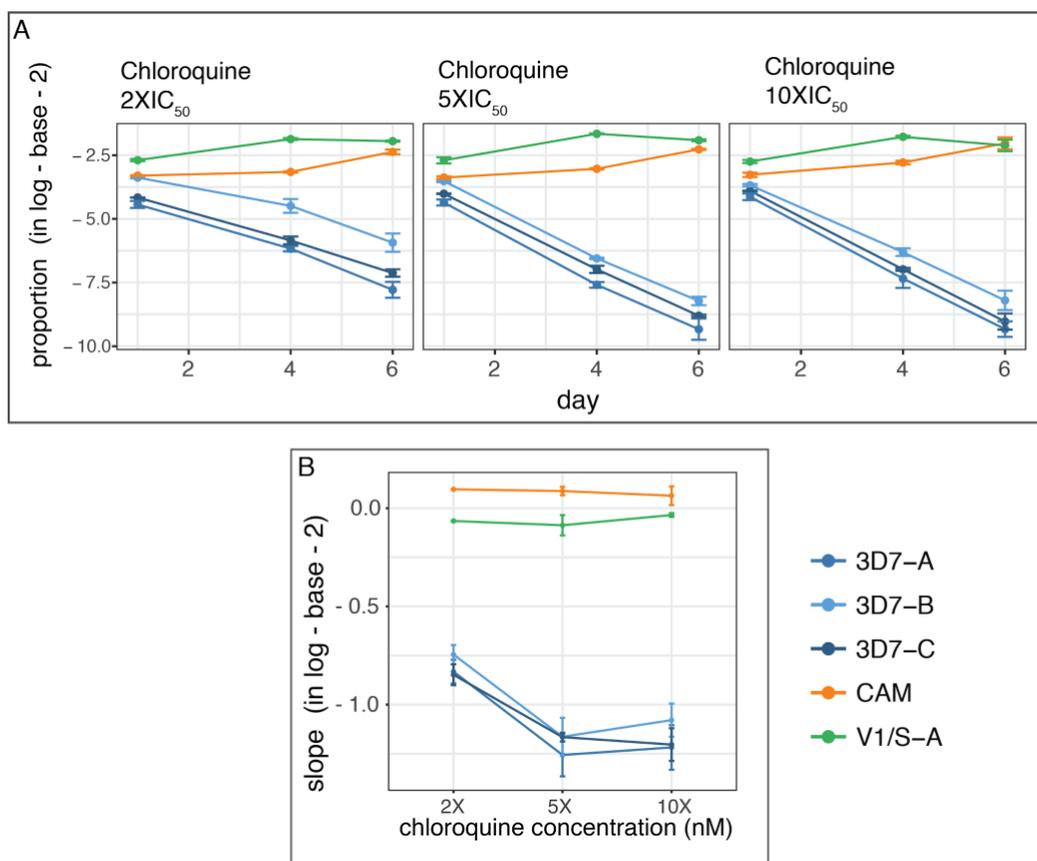
Figure 4.8 Response to increasing concentrations of chloroquine measure by BarSeq. A. Growth curves at different concentrations of chloroquine ranging from 2-10× the IC_{50} for the sensitive strain, 3D7 over a course of 6 days (x-axis) **B.** Average slope from triplicates at the three concentrations used, taken from performing a linear regression on the growth curves shown in (A). The y-axis represents the log2 of the average slope across all timepoints for each strain, and the x-axis represents the increasing concentration in nM. Technical triplicates were performed for this assay.

4.3.2.3 Change in barcode counts reflect mefloquine sensitivity, but not DHA or piperaquine

Given the efficacy of the protocol in revealing the expected phenotype for well-characterised drug sensitivities of the strain panel, other clinically relevant antimalarial compounds (particularly those which are used in combination with the frontline antimalarial, artemisinin) were applied to the experimental setup. Mefloquine has been used in combination with artesunate since rising resistance in Thailand and Cambodia rendered it ineffective as a monotherapy (Vijaykadga *et al.* 2006; Denis *et al.* 2006). Resistance to mefloquine has been

associated with a copy number amplification in the multidrug resistance gene *Pfmdr1* (Price *et al.* 2004). Recently malaria programs in Southeast Asia have been encouraged to use triple combinations of dihydroartemisinin-piperaquine-mefloquine, informed in part by genomic surveillance studies showing a rise in a molecular marker of piperaquine resistance (amplification in plasmepsin 2 and 3), and a deamplification in *Pfmdr1* (Amato *et al.* 2018). To establish the concentration conditions for these drugs for use in BarSeq experiments, the drug-response curves for mefloquine, piperaquine and DHA were obtained experimentally (Fig. 4.9A) and compared to those available in the literature (Duraisingh *et al.* 2000).

A

Mefloquine

Strain	IC$_{50}$ (nM)	Ref. IC$_{50}$ (nM)	Reference
3D7	29.26	21.9-26.1	Durasingh, M. et al. 2002
V1/S	8.75	25 *	Durasingh, M. et al. 2002
CAM	19.00	NA	NA

* The exact value for V1/S is not provided in the reference but rather its presented as a figure.

B

Piperaquine

Strain	IC$_{50}$ (nM)	Ref. IC$_{50}$ (nM)	Reference
3D7	23.11	27 +/- 17	Mwai. et al. 2009
V1/S	35.04	42 +/- 10	Mwai. et al. 2009
CAM	31.76	NA	NA

C

DHA

Strain	IC$_{50}$ (nM)	Ref. IC$_{50}$ (nM)	Reference
3D7	1.39	2 +/- 0.1	Mwai. et al. 2009
V1/S	1.24	2 +/- 1	Mwai. et al. 2009
CAM	2.29	NA	NA

Figure 4.9 Dose-response curves for mefloquine, piperaquine and DHA. A-C Dose response curves for strains used in competition assays measured by using a dose response model (drc) (Ritz *et al.* 2015), for mefloquine (A), piperaquine (B), and dihydroartemisinin (DHA) (C). Panels under the growth curves represent the respective experimental values obtained and the published values (Ref. IC$_{50}$), both of which were taken into account for the design of BarSeq experiments. The x-axis corresponds to the concentrations of the compound, on a log scale, and the y-axis to the percentage inhibition. Technical duplicates were performed for the assay.

Figure 4.10A-B shows a BarSeq experiment using the 3D7 IC$_{50}$ of either mefloquine or piperaquine as the selection pressure. As expected, for the case of mefloquine, even though the difference in IC$_{50}$ fold-change in concentration across strains is small (2 or 3-fold), by using the concentration of the wild-type strain 3D7, a sensitivity phenotype for V1/S and CAM was

validated at this low concentration. Interestingly, V1/S shows a hypersensitive phenotype when exposed to mefloquine, based on this approach. An analysis of variance and mean comparison were performed as with chloroquine and the statistical significance is shown (p-adj of 6.6x10-5). Mefloquine hypersensitivity in chloroquine resistant strains has been previously shown by Durasingh *et al.* (2000), and it has important implications for implementing policies for drug combinations, as the genetic architecture of hypersensitivity has not been dissected (Duraisingh *et al.* 2000). An additional assay used a dilution of mefloquine as with chloroquine, confirming a dose-dependent effect, with no significant fluctuations at the lower concentrations (between $2\times$ and $5\times$ IC_{50}). At 500nM mefloquine, all strains maintain the same proportion of the total reads, as the read counts drop over time, suggesting that in this short period of time during the competition assay, there is still DNA from dead parasites that is amplified; however, it does not represent changes in growth (Fig. 4.11). Giemsa smears were made at every time point and confirmed the absence of live parasites by microscopy at 500nM.
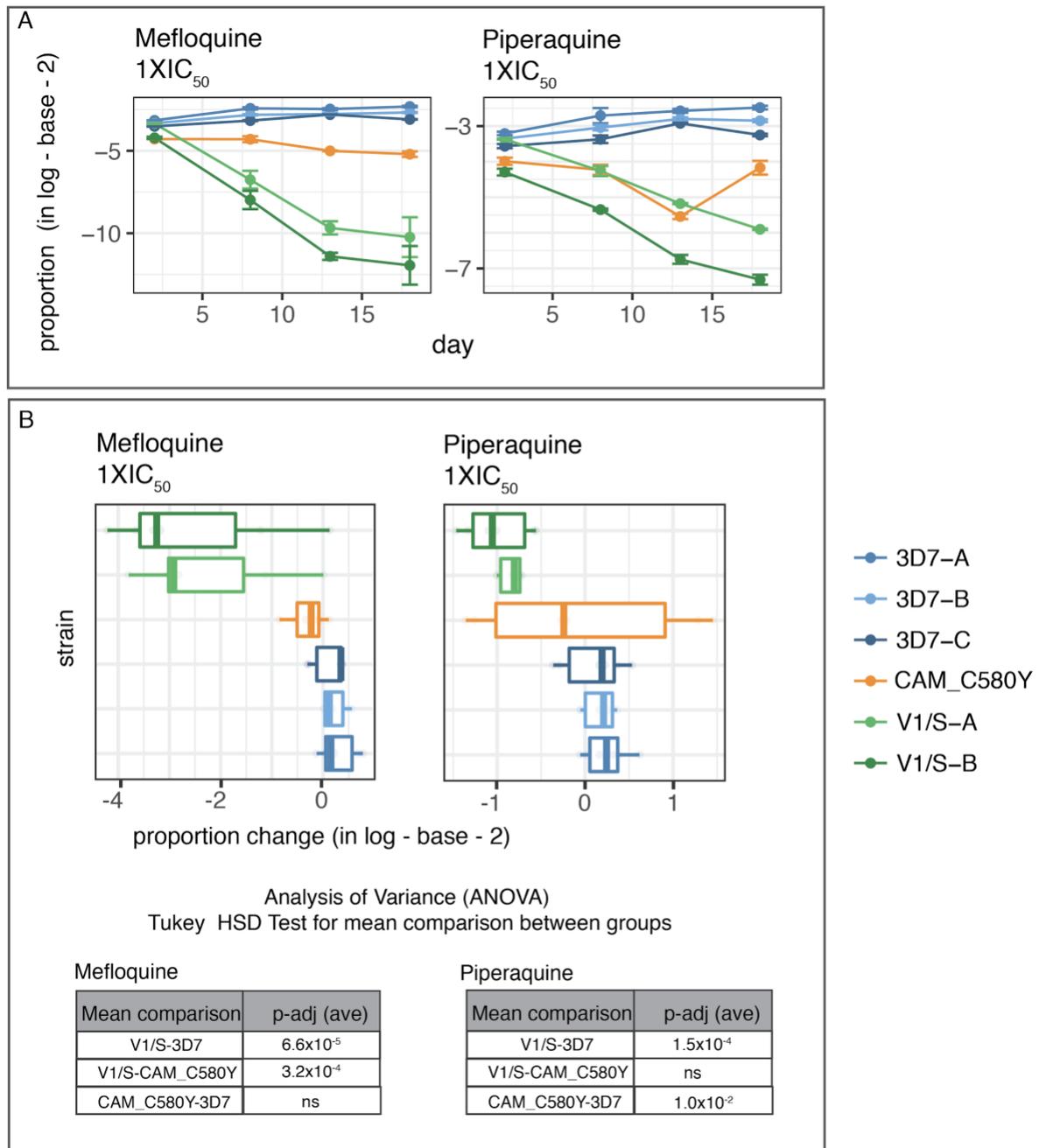
Figure 4.10 BarSeq of barcoded clones in presence of mefloquine and piperaquine. **A.** Competition assays of the barcoded pool in the presence of different antimalarial compounds. The y-axis represents the log2 of the proportion of each barcode in the total read counts obtained over time. **B.** Change in proportion until day 13 for mefloquine and until day 18 for piperaquine, panel below shows significance between the means of the different groups measured for both compounds by performing an analysis of variance (ANOVA) and TukeyHSD test for mean comparison. Technical triplicates were performed for this assay.

Piperaquine exposure at $1 \times IC_{50}$ revealed significant differences between 3D7 and V1/S (Fig. 4.10A-B), however, the magnitude of the change in proportion over time at this drug concentration is comparable to the result obtained for the untreated control (Fig. 4.7A), reflecting differences in the inherent fitness of the strains. To show that the standard assay at a sublethal concentration of piperaquine was ineffective for this specific panel of strains, a replicate experiment was performed (at the IC_{50}), together with an untreated control, and the change in proportion was measured. Fig. 4.12 shows no significant difference across experiments, evidence that the drug has no effect at this concentration on the strains used. Note that for this analysis, the proportion of the total reads corresponding the V1/S-B clone at the lowest concentration was too low for analysis, as the fluctuations would not be accurately representative of a change in the slope of a linear regression, or the rate at which the proportion decreases over time. In addition, the Cambodian strain has not been associated with piperaquine resistance, and does not carry any amplifications in plasmepsin 2 and 3.
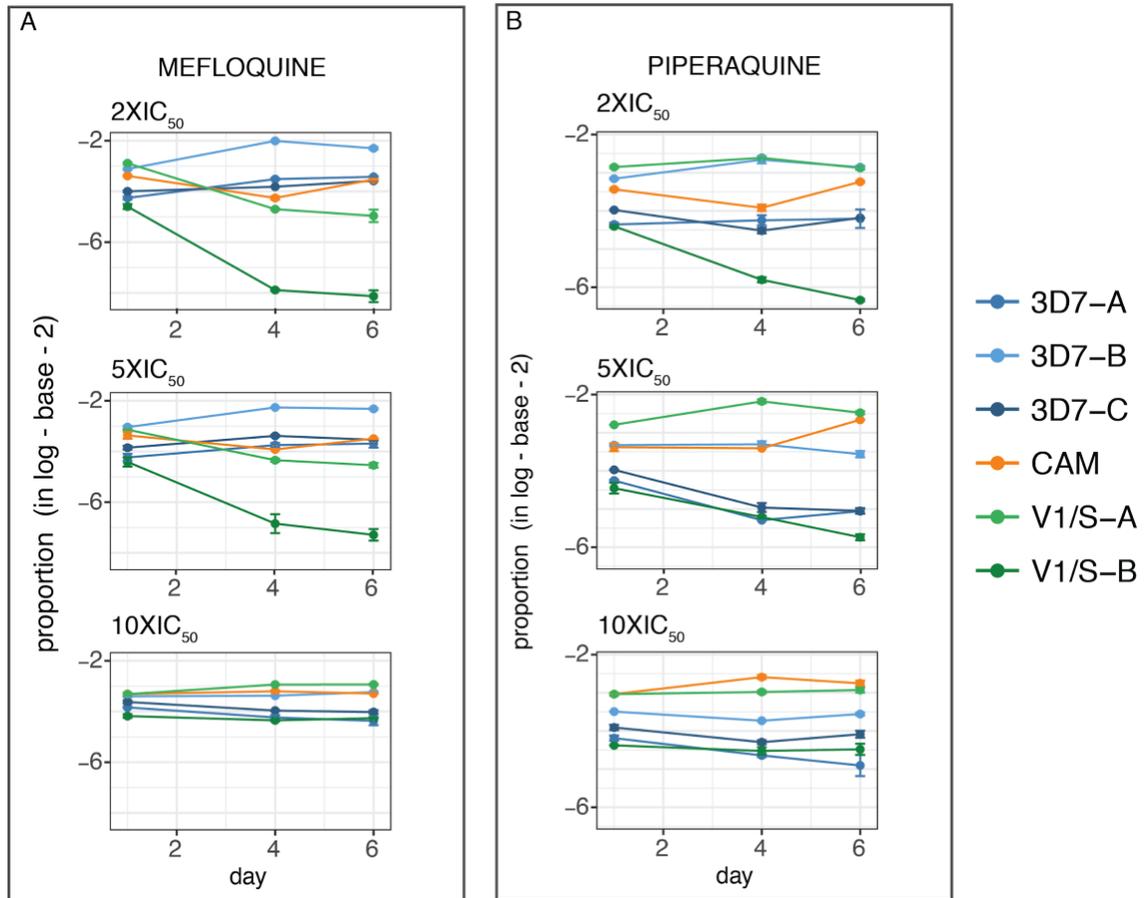
Figure 4.11 Response to increasing concentrations of mefloquine and piperaquine measured by BarSeq
Growth curves at different concentrations of mefloquine **(A)** and piperaquine **(B)**, ranging from 2-10× the IC_{50} for the strain used as the wild-type (3D7), over a course of 6 days (x-axis), y-axis representing the log2 of the proportion of the total reads in each time point. Technical triplicates were performed for this assay.
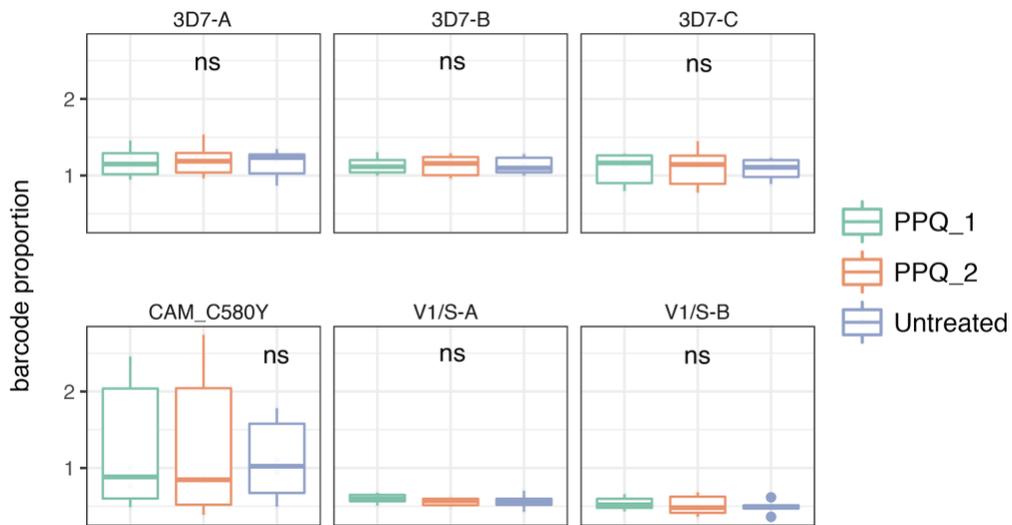
Figure 4.12 Unchanged barcode proportion with piperaquine exposure at low dose. Analysis of the change in barcode proportion over time for the pool of barcoded parasites over a course of 18 days. Two independent biological triplicates using piperaquine are shown together with the data from the untreated control from Fig. 4.7A. Each panel represent a different clone and an analysis of variance was performed and the mean was compared between groups showing no statistical differences for any of the strains used.

It has been shown by the group of Didier Ménard that an alternative assay to the standard 72-hour inhibition assay correlates better with treatment outcomes. A piperaquine survival assay (PSA) was developed, in which ring-stage parasites are pulsed with 200nM of piperaquine for 24 hours, then a washing step is performed and parasitaemia is measured at 72 hours from start of treatment. As with the $RSA_{(0-3h)}$ for artemisinin, a survival rate greater than 10% is considered resistance (Duru *et al.* 2015). A similar drug treatment regime to the PSA may thus be necessary in order to use the BarSeq approach to measure piperaquine sensitivity. We anticipated that a modified BarSeq approach might also be necessary for artemisinin, given the short half-life of the drug in vivo (2-4 hours) (Titulaer *et al.* 1991). Figure 4.13 shows a BarSeq experiment performed by using dihydroartemisinin (DHA). Two approaches were used as an attempt to detect the expected phenotype for the parasite background harboring the C580Y mutation in *Pfkelch13*. In the event that a standard competition assay at a sub-lethal dose would not reveal the phenotype, a parallel experiment was performed in which the pool of barcoded parasites was pressured at the IC_{50} of 3D7 every 24 hours, starting at ring stages. Surprisingly, both experimental approaches revealed that the drug was not having any effect at this concentration and amount of time (Fig. 4.13 A-B). A significant difference was observed between V1/S and both 3D7 and CAM in the pulsed experimental set up (Fig. 4.13 B), however when the average change in proportion was compared (Fig. 4.13 C), no significant differences could be detected between these two experimental designs, suggesting that shorter pulses at a more precise parasite stage, akin to the $RSA_{(0-3h)}$ would be necessary to distinguish the K13-mutant CAM line from 3D7 and V1/S.
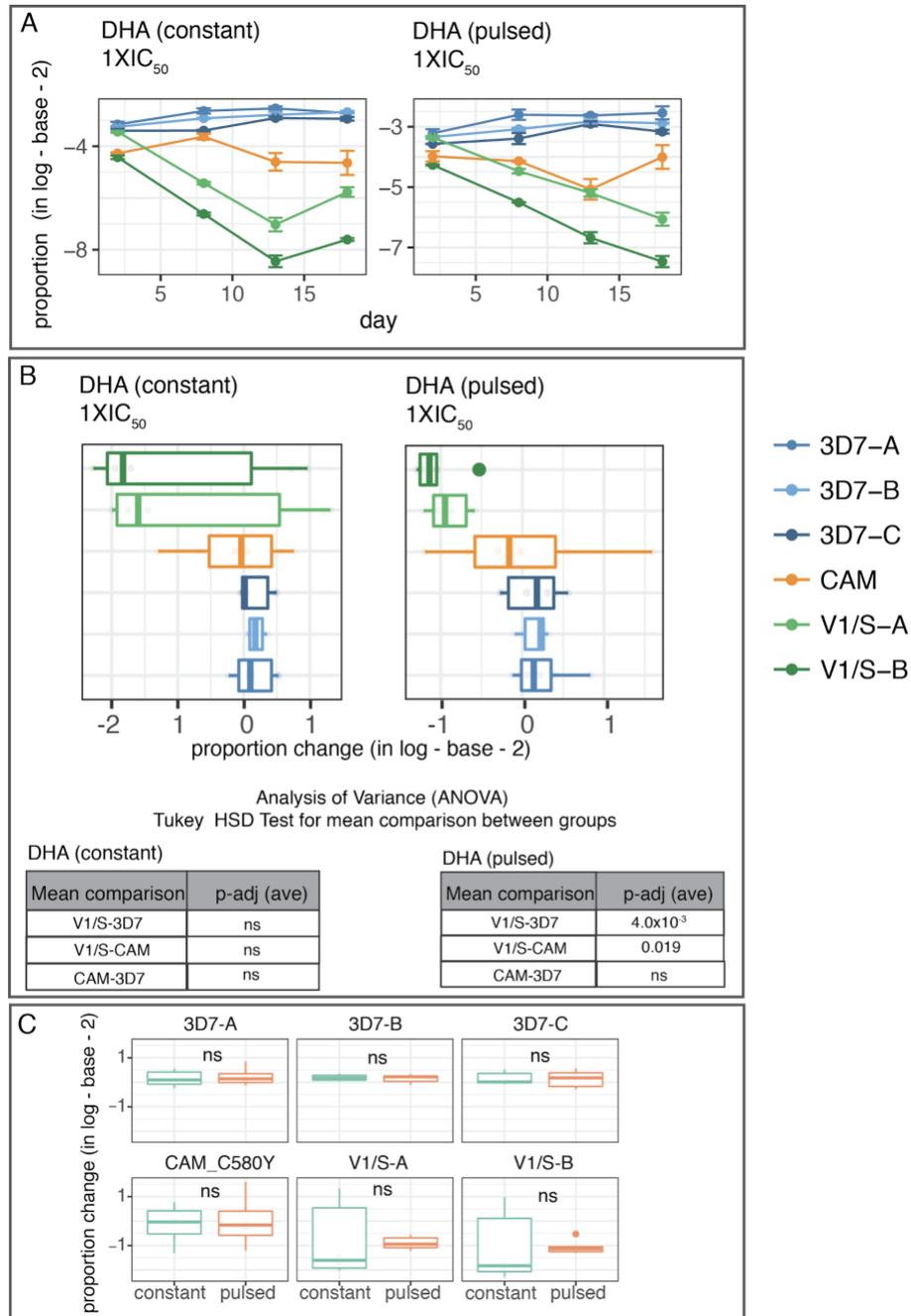
Figure 4.13 Response to constant pressure or pulsed DHA on panel of barcoded clones by BarSeq. Competition assay with the panel of six clones in the presence of the main active metabolite of artemisinin (DHA). To mimic the short half-life of the compound, two strategies were tested, constant pressure at $1\times IC_{50}$ (constant) and by pulsing for 6 hours every 24 hours at the same concentration (pulsed). **A.** Growth curves for both approaches, constant pressure and by pulsing at the IC_{50} for 3D7. **B.** Change in proportion over 18 days of competition for constant and pulsed DHA, and significance between the means of the different groups for the two conditions. **C.** Mean comparison of both approaches (constant and pulsed). The y-axis represents the log-2 of the change in proportion over the different timepoints.

4.3.2.4 Change in barcode counts after an artemisinin pulse similar to an RSA(0-3h) reflects higher tolerance in a *Pfkelch13* C580Y background

The success of an ACT depends ultimately on how effective its partner drug is, and given the short half-life of artemisinin, the partner should have a longer half-life to assure killing of remaining parasites after artemisinin levels have declined. Results from previous studies that were described in section 1.4.4, such as Noedl *et al.* 2008, showed that only a small fraction of the treatment failures with artemisinin as a monotherapy were due to resistance, combining a higher IC_{50} and a delayed parasite clearance in the circulating blood (Noedl *et al.* 2008). This standard inhibitory assay (IC_{50}), however, does not show a shift high enough to define a resistance phenotype, and is not always well correlated with slow-clearance from patients, partly because of the short half-life of the drug and the frequency of treatment failure (Dondorp *et al.* 2010). Other studies have shown that these standard *in vitro* assays might not reveal less sensitive strains because *P. falciparum* parasites get arrested during the first 24 hours of development (Tucker *et al.* 2012), however, these delayed rings have not been fully characterised. Artemisinin resistance therefore would be determined solely as the reduction in the clearance rate of parasites in peripheral blood (> 5 hours), until new *in vitro* assays were developed. An RSA$_{(0-3h)}$ relies on the treatment of early rings which were shown to be the stage most resilient to treatment, thus recapitulating the brief exposure to the concentration of artemisinin in circulating blood (700nM), as described in section 2.4.2. This assay was optimised by Witkowski *et al.* 2013, aimed at targeting 0-3-hour post-invasion rings, showing for the first time a clear shift in susceptibility of early rings in parasites isolated from Cambodia (Witkowski *et al.* 2013).

As shown by the DHA inhibitory concentration assay (Fig. 4.9C), no significant shift could be detected for DHA between strains despite a small increase in tolerance observed for the Cambodian strain, consistent with the observations made by Noedl *et al.* 2008 (Noedl *et al.* 2008). However, this minor difference was not sufficient to be detectable by BarSeq, as shown in Fig. 4.13A-B.

To develop a better method for using BarSeq to detect a phenotype for the Cambodian strain (CAM), containing the C580Y mutation in *Pfkelch13,* a different BarSeq experiment was developed which would recapitulate the conditions established for the RSA(0-3h). Fig. 4.14A shows survival rates for strains used in previous work by Ghorbal *et al.* 2014 and Straimer *et al.* 2015. A Cambodian isolate (Cam3.II), harbouring the C580Y mutation was published as having a 24% survival, whereas V1/S and NF54, both carrying the reference allele for *Pfkelch13*, had a 0.3% and 0.7% survival rate, respectively.

In order to optimize conditions to perform a BarSeq experiment by using a pulsing strategy similar to an RSA, we first attempted to perform a standard RSA(0-3h) to determine if lowering the concentration of the active metabolite dihydroartemisinin would increase the survival of the sensitive strains and avoid amplification bias from DNA coming from dead parasites. A concentration of 250 nM was used for pulsing the pool of barcoded parasites (t = 0 hours) for 6 hours, followed by drug washout, and recovery for 72 hours (t = 72 hours). Using these conditions, Fig. 4.14B shows a survival rate of 7.3% for 3D7 (a clone of NF54, table 2.1), and 33% for CAM, as measured by microscopy in triplicate after counting 10,000 cells (as in section 2.4.2.1). A t-test was performed showing no significant difference between the two strains, however the number of data points is low to perform any statistics, coming from technical triplicate values of parasitaemia at 72 hours (p-value 0.08).
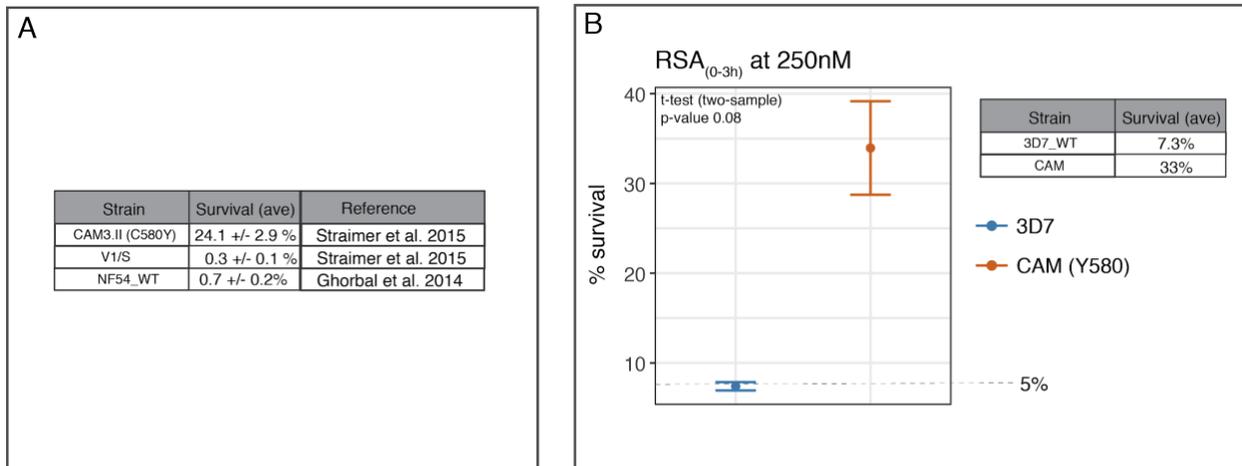
Figure 4.14 Theoretical and experimental RSA$_{(0-3h)}$ values. **A.** Reference survival rates for strains equivalent to the strains used for BarSeq. V1/S and CAM3.II have survival rates previously reported in Straimer *et al.* 2015. Survival of NF54 (parent line of 3D7) as reported by Ghorbal *et al.* 2014 or Straimer *et al.* 2015. The values for NF54 V1/S and CAM3.II are 0.7, 0.3 and 24.1%, respectively. **B.** RSA$_{(0-3h)}$ performed by using 250 nM DHA. Measured by standard light microscopy of 10,000 red blood cells for 3D7 and for CAM wild type.

The next step was to mix the panel of barcoded parasites and perform a BarSeq experiment using a three-hour exposure of 250nM of DHA, with gDNA collected at 0, 3, and 72 hours. To achieve the tightly synchronised early-ring culture required, the different barcoded strains were synchronised independently by performing multiple rounds of sorbitol treatment as in section 2.1.1.2, and mixed together at late trophozoite stages prior to a schizont purification using Percoll. Upon reinvasion, the pool of early rings was pulsed with 250 nM as described above. Fig. 4.15A-B shows the different strains that were included in the competition: two uniquely barcoded clones of 3D7 (blue) as internal replicates, CAM (orange), and one clone for V1/S (green). The growth curves for the four clones, measured by the barcode proportion over time (0-72 hours on the x-axis), show that at 72 hours the sensitive strains 3D7 and V1/S are outcompeted by CAM (Fig. 4.15A), demonstrating that tailoring the conditions for the BarSeq experiment to the drug profile can yield the expected phenotypes. This data also provides further evidence that there is indeed a loss in sensitivity to DHA by harbouring the K13-Y580 mutation in this Cambodian strain background, and also confirming the need to perform *in vitro* assays by pulsing for a short amount of time and with a high concentration,

similar to how the drug acts *in vivo.* Additionally, our results show that it is not necessary to go to the highest reported concentrations in order to observe a significant difference between tolerant and sensitive strains (p-value=0.042).

Fig. 4.15B shows two additional measurements that are typically used for RSA$_{(0-3h)}$ which are the growth rate and survival. The growth rate is taken by measuring the ratio of the untreated pool (non-exposed or NE) and the initial counts (INI), showing that the rates for both CAM and 3D7 are above one, indicating that over 72 hours they grew, whereas V1/S decreased in terms of the proportion of counts in the total reads and in absolute raw counts. Given the high synchronicity of the cultures, the results can be explained by these parasites making less progeny or by some internal competition for nutrients that might lead to this strain being outcompeted very rapidly, thus having a lower parasitaemia at the end of the competition than at the beginning. This had been observed by previous competition assays in this chapter (Fig. 4.7). For the drug-treated condition, the survival rate is the ratio of the proportion of counts of the total reads in the treated over the untreated, showing a fold change of >3 for the tolerant CAM strain, whereas <1 for the other strains, indicating they are dying.
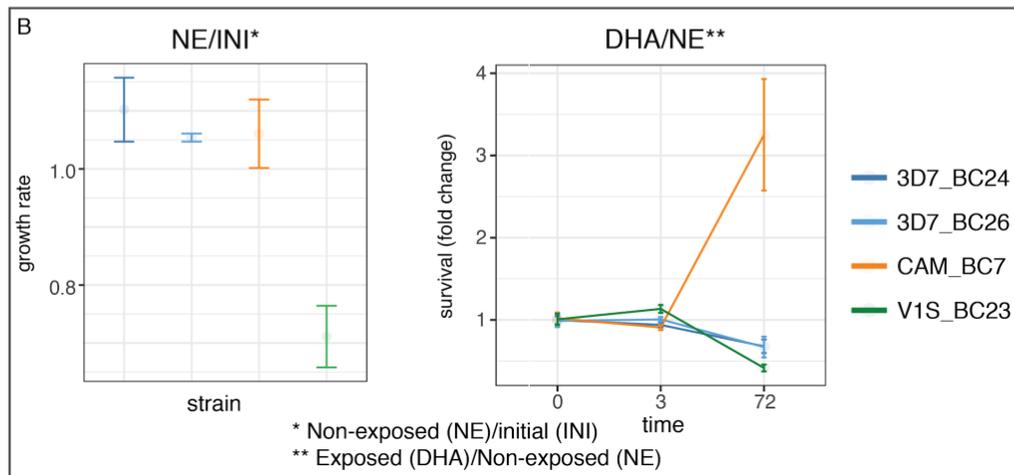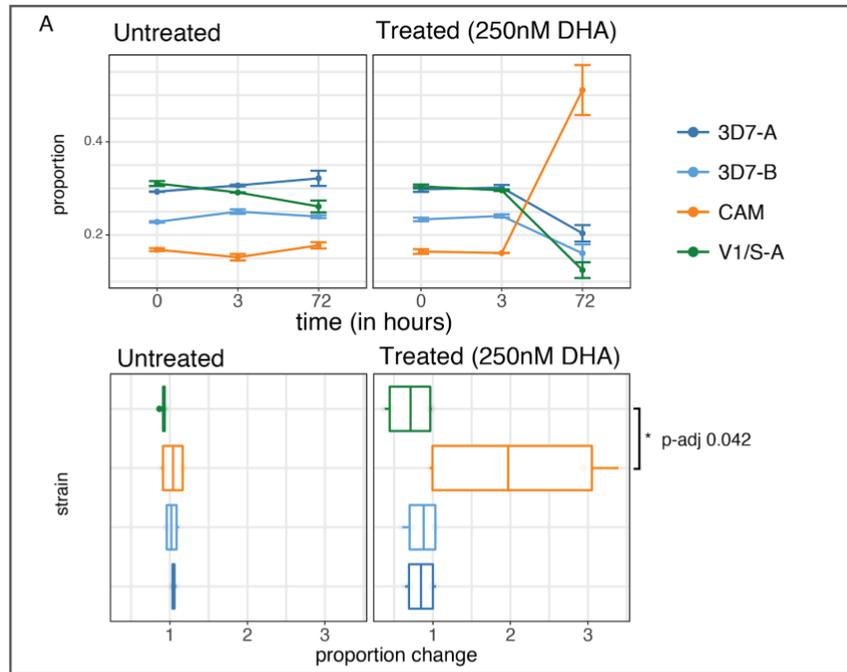
Figure 4.15 RSA-like assay using BarSeq. A. Competition assay with four barcoded clones (two clones for 3D7, one for V1/S and one for CAM). An RSA(0-3h)-like experiment was performed and proportion measured by BarSeq from reads in the total mix by high throughput sequencing (y-axis, top panel), over 72 hours at three independent time points (x-axis). The lower panel represents the change in the proportion for the four clones and significance between CAM and V1/S. **B.** Measurement of growth rates as the ratio between the untreated (NE), and the initial counts (INI) both measured by the proportion of read counts from the total, and right panel showing the fold change in survival as the ratio between the treated (DHA) and the untreated (NE). Technical triplicates were performed for this assay.

4.4 Discussion and future work

Understanding the mechanisms of action of antimalarial drugs is essential in order to validate present and future drug combinations. By increasing the knowledge on the biology of drug resistance, or the molecular target of specific compounds, the feasibility of combinations can provide information regarding which compounds might be synergistic or antagonistic.

Today, many phenotypic assays can be done in the lab as a result of the successful adaptation of *P. falciparum* into continuous culture (Trager and Jensen 1976), and advances in genome sequencing and genetic techniques have made studies aimed at identifying associations between phenotype and genotype possible. Regarding drug resistance, there are various examples that use these technologies to establish an association between population genetics and drug resistant parasites, for example as reported by Miotto *et al.* 2013 or Ariey *et al.* 2014 linking the *Pfkelch13* molecular marker of artemisinin resistance and the various polymorphisms present in the field with the clinical outcome of treatment failure (Miotto *et al.* 2013; Ariey *et al.* 2014). In addition to this, genome editing of malaria parasites has also shed light into some of the mechanisms of resistance to antimalarial compounds. This has been widely used to study molecular markers of artemisinin (*Pfkelch13*) and chloroquine (*Pfcrt*) resistance, amongst others, and the impact of mutations in these genes on resistance has been measured by different phenotypic assays (Ghorbal *et al.* 2014; Straimer, Lee *et al.* 2012).

An understanding of the fitness effects of drug resistance mutations is also essential to understand the epidemiology of resistance, but this has been much more sparsely studied. The different assays described in section 4.2.2 have been useful at determining the contribution of some particular alleles to fitness in different genetic backgrounds, but all only allow for phenotyping of two lines at the same time (Baragaña *et al.* 2015; Hayward, Saliba, and Kirk 2005; Nair *et al.* 2018).

To overcome this limitation and to pilot the way for large-scale phenotyping of parasites to measure both fitness and drug response, I developed a method by which we could generate libraries of parasites of diverse genetic backgrounds carrying a unique barcode inserted in a

constant region (*Pfrh3*). I had shown previously that large pools of plasmids can be maintained in a single transfection as stably replicating episomes. I then combined this approach with CRISPR/Cas9 editing to generate libraries of barcoded parasites that can be cloned by limiting dilution. Even though the efficiency of transfection varied, as shown in Fig. 4.1, parasites would reliably come back with barcodes inserted, indicating that the gRNA is efficient in initiating the homology directed repair pathway. Given the results from the observations from the previous chapter (section 3.3.4), which showed that DNA uptake events are far from statistically independent, I was able to show that when co-transfecting with Cas9 a pool of up to 96 uniquely barcoded plasmids, I obtained 5-9 successful integration events per transfection in the bulk culture, with individual parasites harbouring a single integrated barcode (Fig. 4.3).

Having successfully barcoded strains covering sufficient genetic diversity and drug response phenotypes, I also show the first attempts to apply BarSeq in *P. falciparum* by using these strains in a pool. These results highlight the potential power of this technology to measure inherent fitness of different strains and their drug response to known antimalarial compounds.

The BarSeq results show relatively subtle phenotypes, such as the fact that V1/S is hypersensitive to mefloquine, as had previously been shown by Duraisingh *et al.* 2000. Given that chloroquine and pyrimethamine resistance have spread worldwide, the finding that an epidemiologically relevant gene can be involved in modulating resistance to both chloroquine and mefloquine (*Pfmdr1*), as the phenotype observed from V1/S is fundamental. Previously this had been shown by standard inhibitory concentration assays only (Duraisingh *et al.* 2000). The results presented here highlight that the efficacy of mefloquine at concentrations that are sub-lethal to 3D7 can have dramatic effects on V1/S. Additionally, BarSeq shows that a recently adapted artemisinin-resistant strain originating from Cambodia (CAM) shows an intermediate phenotype to mefloquine (Fig. 4.10A), a result that brings some concern given the fact that mefloquine is being used as a combination drug with artemisinin derivatives to combat multidrug resistant parasites in Southeast Asia, where the CAM strain originated. I also used BarSeq to investigate the interaction of the strain panel with other compounds from the

quinoline class of antimalarials such as piperaquine (Fig. 4.10B), which is also used widely as a partner drug with dihydroartemisinin. The results showed no difference between strains, suggesting that it should be repeated including piperaquine resistant lines such as those which have emerged in Western Cambodia as part of the recent outbreak of multidrug resistant parasites (Amato *et al.* 2018). To that end, future work will aim to barcode additional parasites with diverse clinical phenotypes that will supplement and be studied in perspective with the pool generated in this chapter, by applying an alternative assay to measure piperaquine response as in Duru *et al.* 2015 (Duru *et al.* 2015).

The results shown also highlighted the use that this method has in unravelling the phenotype of increased tolerance to the frontline antimalarial artemisinin, by performing an RSA-like drug exposure for the pooled parasites in which early rings are pulsed with high concentrations of its main metabolite, dihydroartemisinin (Fig. 4.15). The results with BarSeq correlate well with standard RSA assays that are quantified by using light microscopy of Giemsa-stained slides (Fig. 4.14), and demonstrate that different exposure regimens can be employed with the BarSeq approach depending on the phenotype being examined.

In addition, my BarSeq results show striking evidence that some strains have an enhanced fitness when grown *in vitro* in the absence of any selective pressure. Measuring growth rates across strains has been done in the past using a variety of approaches. In a recent study by Murray *et al.* 2017, the multiplication rates of different strains were measured by growing them independently, and at different time points the number of genomes in subsequent generations was measured with qPCR. From the analysis, the authors concluded that Dd2 had higher exponential multiplication rates than other strains such as 3D7, D10 or HB3, over 48 hours (Murray *et al.* 2017). With the method I have developed growth rates are measured in competition rather than independently, and it was particularly striking to see that when I compared 3D7 to both V1/S and CAM in the absence of the drug (Fig. 4.7A), there was a clear fitness advantage of 3D7 over the other two strains.

A clear remaining question is what is the underlying biological basis of differences in fitness. Some ways by which I will explore fitness differences in the future will be to understand

whether there is a decrease in the number of merozoites (progeny) in the segmenting schizonts across strains, or if there is a difference in the length of the life cycle. Finally, I will assess whether there is an invasion phenotype that might explain the rapid decrease in the V1/S population when mixed with the other strains. Interestingly, across all the competition assays performed, V1/S exhibited lower fitness than a more recently adapted, multidrug resistant strain from Cambodia (CAM), consistent with the possibility that relatively robust fitness might play some part in the spread of this CAM parasite lineage among artemisinin-tolerant parasites in Southeast Asia (Amato *et al.* 2018). Because of the striking differences in inherent fitness across strains, which are non-attributable to any specific genetic differences, I am currently considering which strains to mix by carefully measuring their relative growth rates before pooling, in order to have equivalent growth rates over the course of the competition to prevent rapid loss of specific lines. This had been addressed in the *P. berghei* knock-out study by Bushell *et al.* 2017, from similar observations in which slow-growing but non-essential mutant phenotypes were rapidly outcompeted by other fast-growing mutants of redundant genes, precluding careful measurement of differences among slow growth phenotypes (Bushell *et al.* 2017).

This current work, and what I foresee will be an important use of this technology in the future is to be able to perform high throughput phenotyping of pools of potentially hundreds of barcoded parasites with isogenic backgrounds, carrying different alleles that might have an impact on resistance or fitness. In the absence of any drug pressure, I predict that some clones will have an enhanced fitness and will take over the population after a certain time. By contrast, in the presence of a selective pressure (e.g. drugs or any environmental stressor), the ratio of clones will change either in predictable ways based on what is known regarding markers of resistance, or in unexpected ways if there are no anticipated interactions between specific alleles and their outcome when exposed to a known or unknown antimalarial compound (Figure 4.16). The results so far show that a suitably diverse panel of strains will be sufficient to elucidate resistance and hypersensitivity phenotypes to different compounds, as seen with chloroquine and mefloquine (Fig. 4.7 and 4.10A, respectively). Current and future efforts include using this strategy to insert unique barcodes into different strains that have been

selected for resistance to many potential antimalarial compounds from the Malaria Box (Van Voorhis *et al.* 2016). The Malaria Drug Accelerator (MalDA) consortium, as an example, has been putting a large effort into drug target discovery by selecting for resistance to these candidate antimalarial compounds followed by whole genome sequencing, generating a large collection of resistant lines in the process. Working with the consortium, we are now in the process of implementing an approach to accurately measure fitness of all these mutants in a high-throughput manner by using this barcoding approach, as it fits within the scope of the vision of the consortium of using evolution and chemogenomics to accelerate drug discovery (Cowell *et al.* 2018).
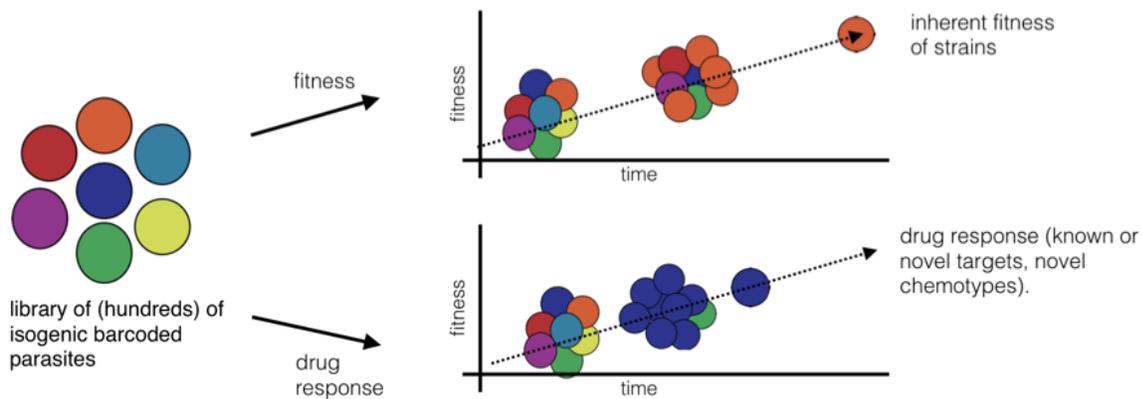
Figure 4.16 Schematic representation of the barcoding approach and its future applications. The scheme shows the strategy for performing multiplexed parallel phenotyping of libraries of barcoded *P. falciparum* clones. The top arrow refers to the use of this approach to measure the inherent fitness of strains in the absence of any drug pressure or environmental stress. The bottom arrow indicates that the technology can also be used to evaluate the contribution of specific alleles to resistance, and the potential use of it to identify cross-resistance, synergistic compounds, or to study the effect of novel chemotypes.

Chapter 5: Contribution of *Pfkelch13* alleles to parasite fitness and artemisinin resistance in different genetic backgrounds.

5.1 Summary

The aim of this chapter is to study the impact on fitness of mutations in *Pfkelch13* in a KEL1 lineage, originally from Western Cambodia (Pursat). By combining CRISPR/Cas9 genome editing and high throughput sequencing I was able to provide evidence that parasites harbouring this genetic background are not only more permissive to resistance-associated mutations, but also have an enhanced fitness when compared to other strains from within SEA or from other continents (listed in table 2.1).

5.2 Background

5.2.1 Drug resistance and fitness: an evolutionary arms race

The efficacy of antimalarial compounds has historically been limited by resistance: independent origins of resistant parasites have occurred in Southeast Asia and South America, and they have spread to highly endemic regions in Sub-Saharan Africa. Both chloroquine and pyrimethamine have followed this path (Schwartz *et al.* 1983; Roper *et al.* 2004). With the current threat of the emergence of resistance to the frontline antimalarial artemisinin, and to some of the partner drugs used in artemisinin combination therapies (Leang *et al.* 2015), identification and validation of alleles that might be modulating a decrease in sensitivity or might be compensating for fitness losses associated with resistance (Kublin *et al.* 2003) will be essential to improve surveillance, and thus avoid the consequences of the loss of a frontline therapy similar to the spread chloroquine resistance (Trape 2001).

Fitness, or the ability of an organism to survive and reproduce in a given environment, can be negatively impacted by drug resistance, which has been a widely studied topic in the context of pathogenic microorganisms such as bacteria, yeast and protozoa. Generally,

pathogenic organisms have evolved mechanisms which allow them to optimize biochemical processes in order to cope with selective pressures such as drugs, which might incur an associated fitness cost (Andersson and Hughes 2010). Mutations acquired can impair the normal function of a protein in order to allow it to take on a new role. An example of this in the human immunodeficiency virus (HIV) is the acquisition of several mutations in the reverse transcriptase, which confer resistance to antiviral therapies but at the same time reduce fitness (Collins *et al.* 2004). Other evidence has shown fitness costs associated with drug resistance in *P. falciparum*. An example is mutations in *cytochrome b* that decreased susceptibility to the antimalarial atovaquone but negatively impacted fitness (Peters *et al.* 2002), or copy number amplifications in the multidrug resistance gene *pfmdr1* leading to increased survival upon exposure to the antimalarial drug mefloquine but reduced fitness (Preechapornkul *et al.* 2009). Another line of evidence that can be extrapolated from the field is the reversal of resistance to chloroquine in Malawi after the drug was removed from use, supporting the idea that mutations in the chloroquine resistance gene *pfcrt* had a fitness cost in these African parasites (Kublin *et al.* 2003).

The overall premise is that in absence of any drug pressure, the acquired drug-resistance associated polymorphisms will negatively impact fitness, which in fast-evolving organisms can eventually lead to a reversal of resistance (Hastings and Donnelly 2005). An alternative scenario, particularly under strong selective pressure, is the acquisition of compensatory mutations that will alleviate the fitness costs associated with the drug pressure, as described by Rosenthal (Rosenthal 2013). These compensatory mutations can restore the function of the mutated protein to alleviate any negative impact. The longer the selection pressure, given the effective population size, the more likely it is that there is a stepwise process of acquisition of compensatory mutations, eventually leading to some clones having an enhanced fitness almost to the point that they follow equivalent growth rates to the wild type clones, as shown by Marcusson *et al.* 2009 with *E. coli* (Marcusson *et al.* 2009).

In the historical context of antimalarial drug resistance, various examples support this idea: 1. The discovery that acquisition of multiple within-gene mutations in *pfcrt* was required

to reach the full chloroquine resistance phenotype (Gabryszewski *et al.* 2016); 2. A stepwise acquisition of mutations in *dhfr* is required to both increase tolerance to sulfadoxine-pirymethamine and compensate for reduced fitness (Lozovsky *et al.* 2009); and 3. A potential intergenic effect of a genetic background compensating for fitness in artemisinin resistant parasites (Miotto *et al.* 2015).

Candidate molecular markers of increased tolerance, such as mutations in the propeller domain of *Pfkelch13*, are present at high frequencies in Southeast Asia, and to some extent in Africa and South America (Amato *et al.* 2016). Different studies have shown that resistance to artemisinin has independent origins and there are different mutations that contribute to a loss of sensitivity (Straimer *et al.* 2017; Straimer *et al.* 2015; Ghorbal *et al.* 2014), evidencing a soft selective sweep. However, in particular, a lineage carrying mutation C580Y that emerged in Western Cambodia seven years ago, named KEL1 by Amato *et al.*, has been supplanting other artemisinin-resistant populations with other non-synonymous mutations in *Pfkelch13* (Amato *et al.* 2018). Additional evidence presented in the work by Amato *et al.* 2018 indicates that resistance to the partner drug piperaquine had the same origin. Tracing the evolution of this new lineage (KEL1/PLA1) reveals it is spreading from Western to Eastern Cambodia, and is also taking over other local parasite populations with resistance-associated mutations.

5.2.2 Epistatic interactions and their role in drug resistance

Some of the scenarios in which compensatory mechanisms can alleviate fitness costs associated with drug resistance can be explained by intergenic or intragenic epistatic interactions (Weinreich and Chao 2005; Keightley and Kacser 1987). This has been widely studied in the context of antibiotic resistance, and it was shown that a form of epistasis can come from the acquisition of compensatory mutations that alleviate costly mutations that might impair cellular function, which is often easier than reverting a mutation that has a high fitness cost (Maisnier-Patin *et al.* 2002). This scenario can often also explain the evolution of multidrug resistance in bacteria (Trindade *et al.* 2009). In addition to this, epistatic interactions can have a different outcome depending on the genetic background. Epistasis involving multiple genes acting together (magnitude epistasis), having a different outcome on phenotype

depending on the genetic background, can be defined as positive or negative epistasis. There are two scenarios for positive epistasis: the additive effect of two deleterious mutations alleviates the fitness costs associated with the negative effects of each independent mutation, or alternatively if the mutations are beneficial it can lead to an improvement in function greater than would be expected by the function of the independent mutations; both cases describing a positive epistasis (He *et al.* 2010). Negative epistasis, on the other hand, refers to the same additive effect but with a negative outcome on phenotype, which can also be a consequence of addition of either beneficial or deleterious mutations (Khan *et al.* 2011). Overall, both types of epistasis have been widely studied with the aim of shedding light on how the effect of variants can be impacted by the genetic background, and how this determines the fitness landscape by shaping the adaptive evolution of organisms. How the epistatic interactions affect fitness would, in theory, depend on the population itself: the effective population size (N_e), the extent to which populations recombine, mutation rates and interspecific competition, are all factors determining the directions organisms can take to either increase or decrease fitness. In the context of *P. falciparum*, given the emergence and spread of drug-resistance parasites from a very confined geographical location in Southeast Asia with a small N_e, it is important to understand the fitness landscapes of these parasites, which we hypothesize will be shaped by positive epistatic interactions of drug-resistance associated genes in these successful lineages. There have been some approaches undertaken to address these questions, considering the limitations in performing experimental genetics at scale for validation of gene function. Generally, studies have focused on the impact of single mutations on the outcome of drug susceptibility. However, how individual resistance mutations can impact the overall fitness can be different than the expected outcome if each mutation and its impact on fitness would be considered independently. An example was shown with the different pathways leading to resistance to the antifolate pyrimethamine in *P. falciparum*, by using a yeast model to introduce the different combinations of mutations seen in the field and studying their effect on fitness (Lozovsky *et al.* 2009).

Overall, resistance emerging in particular regions of the world can be a result of the local transmission intensity, or competition with other parasites, as recently shown by Bushman

et al. (Bushman *et al.* 2018), however it can also be linked to genetic backgrounds better tolerating resistance-associated polymorphisms: the latter is likely to play a key role in driving the epidemiology of drug resistance. One explanation is that epistatic interactions in the different parasite populations might have a different outcome: different genetic backgrounds might have an impact on the resistance phenotypes observed, and multidrug resistant strains in Southeast Asia, such as the KEL1/PLA1 lineage described by Amato *et al.* 2018, might be more likely to become resistant to other drugs, because of positive epistasis resulting from compensatory mechanisms emerging as a result of the local epidemiology, interventions, and evolutionary bottlenecks.

5.3 Methods and Results

5.3.1 Targeting vectors for CRISPR/Cas9-based editing of *Pfkelch13*

To evaluate the effect of different *Pfkelch13* alleles on parasite fitness in different backgrounds, multiple strategies were designed for generating constructs for mutating parasites from different strains.

5.3.1.1 Design of donor templates for CRISPR/Cas9 editing of *Pfkelch13*

Donor templates for CRISPR/Cas9 can be used in multiple different strategies to modify the genome. As described in chapters 3 and 4, a donor template can be used in order to disrupt a gene, in that particular case *Pfrh3* by insertion of a barcode. However, more elaborate donors can also be used to insert selectable markers, for example inserting a fluorescent tag in order to perform localisation experiments (Bushell *et al.* 2017; Sidik *et al.* 2016), or site specific changes to introduce specific synonymous or nonsynonymous substitutions as seen in Ghorbal *et al.* 2014. These approaches rely on a donor sequence either being provided as a separate plasmid from that expressing Cas9, or together as a single plasmid. The size of the donor region can vary, and it has been shown that a region of 0.5 to 1kb of homology is sufficient for initiation of the repair pathway present in this species (Knuepfer *et al.* 2017), however concrete evidence on the optimum length that leads to the highest efficiency does not exist for *P. falciparum*. In *P.*

berghei, as in several model eukaryotes, the general rule is that longer homology leads to higher efficiency in recombination (Pfander *et al.* 2011), although this data was generated in the absence of CRISPR cleavage.

Because the interest of this chapter is to introduce point mutations into *Pfkelch13*, several considerations were taken in order to design the donor sequences. First, in order to prevent further cleavage by the Cas9 nuclease, donor regions were designed to contain polymorphisms that protect the endogenous sequence, also referred to as "silent" or "shield" mutations. In chapters 3 and 4 this problem was addressed by deletion of the cleavage sites by the 100 bp sequence containing the barcode, thus protecting the edited parasites from further cleavage. For performing single nucleotide edits, however, since the intention of this chapter is to modify a gene predicted to be essential (Birnbaum *et al.* 2017; Schwach *et al.* 2015), the function of the protein needs to be maintained. Therefore, synonymous mutations were introduced at the cleavage site, closer to the PAM to better protect the endogenous gene.

Because the efficiency of repairing a double-strand break can vary between transfections, a simple PCR-based genotyping strategy is desirable, however designing selective primers to differentiate the wild type and edited locus when there are only a few differences might be challenging. To overcome this difficulty, a region within the donor template can be "recodonised", which introduces silent mutations in between the two homology regions. This strategy would also be helpful as by disrupting the homology between the cut site and desired mutation, the repair machinery needs to extend beyond the mutation of interest in order to find the region of homology. Additionally, it increases flexibility in terms of gRNA design, as multiple gRNAs can be covered within this recodonised region, thus increasing the overall editing efficiency.

In order to generate parasite lines harbouring different epidemiologically relevant polymorphisms in *Pfkelch13*, the two donor-design strategies were used for homology-directed repair, by using a two-plasmid approach as in chapters 3-4, or a single-plasmid approach. For the first donor strategy, following a similar approach from section 3.3.1.1, two homology arms were amplified from genomic DNA from the laboratory strain 3D7, using p265/p318 (homology

arm 1, 503 bp) and p267/p317 (homology arm 2, 450bp). Primers p267 and p265 had sequence overlap with each other and p318/p317 to previously digested vector pCC1 using restriction enzymes EcoRI and AvrII. The overlap between p267 and p265, as shown in Fig. 5.1A would contain the mutations to be introduced, which included the mutation C580 to Y, or to reverse the Y in a Cambodian background back to the reference allele (C). In addition to this polymorphism, two silent mutations were designed in the donor to prevent further cleavage by Cas9, at the PAM site (Fig. 5.3A). The two homology fragments and the digested pCC1 were gel-purified and assembled by Gibson cloning (section 2.3.2.7).

For the recodonised approach, the donor region was synthesised as a gBlock by Integrated DNA Technologies (IDT). The sequence of the recodonised region was informed by a Perl script, developed by Adam Reid at the Sanger Institute, that would consider the codon usage genome-wide and provide the most common alternative codon for a given sequence of any length (usually 100-150bp), for that amino acid in *P. falciparum*. Fig. 5.1B shows the general design of this approach. Two homology arms flank a 120bp recodonised region (in gray), with the entire donor flanked by 20bp overlap on each end (red) to the free plasmid ends for Gibson assembly. The plasmid used here had the pDC2 backbone shown in Fig. 5.1D, digested with EcoRI and AatII. Fig. 5.2 shows a chromatogram of the region around position 580. The top panel shows the sequence obtained by using the first approach, highlighting the two silent mutations at the PAM site, and codon TGT for cysteine. The bottom panel shows a section of the silent mutations across the recodonised region.

Fig. 5.1 C and D show the two versions of plasmids used for transfection. The first approach was to clone the donor region into pCC1 following the same protocol as in section 3.3.1.1 (C), whereas the second approach was performed on the pDC2-CAM-CoCas9-U6-hDHFR plasmid, which contains a codon-optimised Cas9 driven by the calmodulin promoter from *P. falciparum*, and a U6 cassette for the expression of a single gRNA as before. In the same plasmid, the donor sequence can be cloned by digesting the plasmid with EcoRI and AatII, and performing a direct Gibson reaction from either the two-arm homology version or by ordering

the synthetic genes containing 700 bp of homology and approximately 20bp at each side for homology into this vector, using universal primers p439/p440.
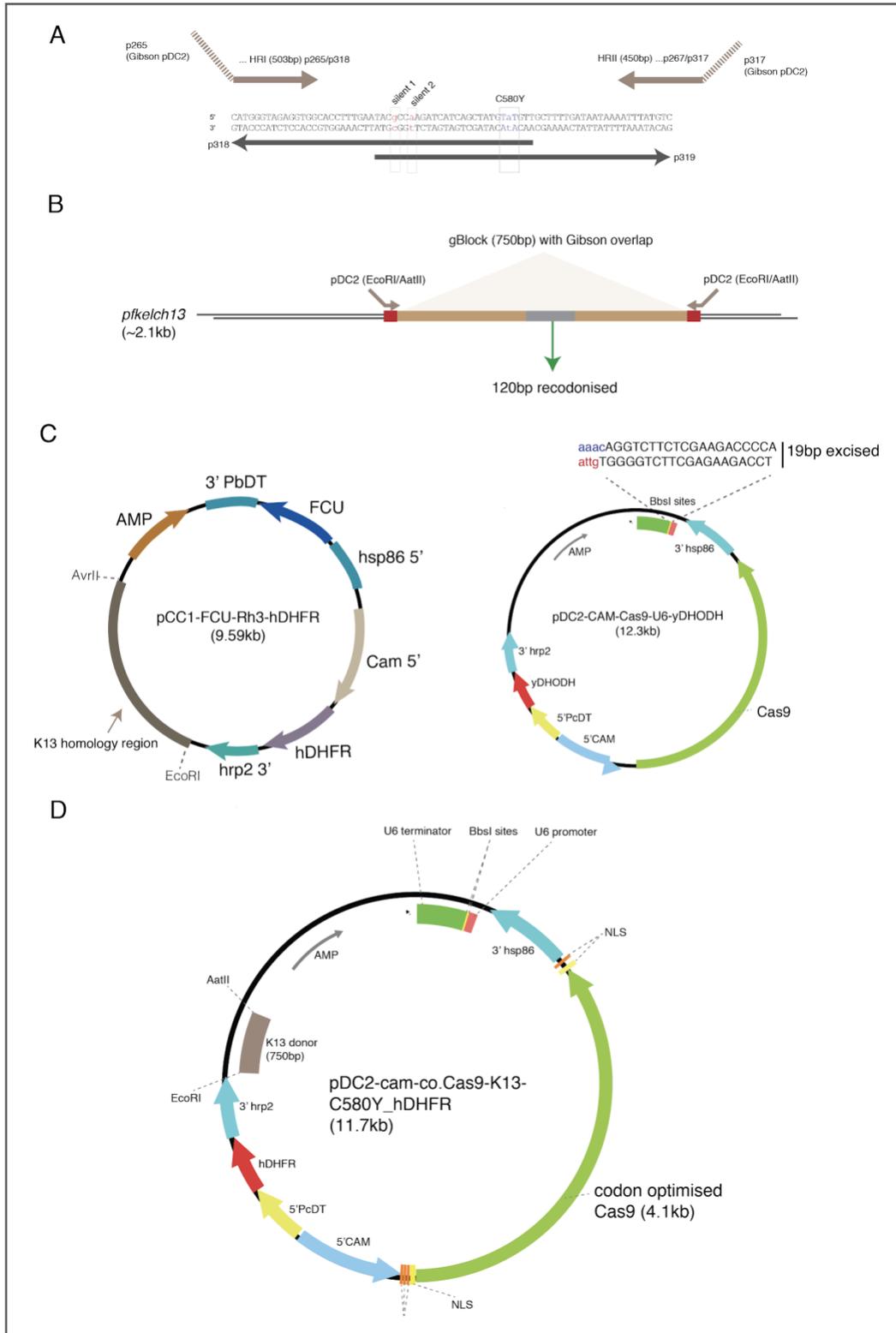
Figure 5. 1 Cloning of donor sequence using two strategies for editing of *Pfkelch13*
A. Figure illustrating the first donor design. A 750 bp donor sequence with homology to *Pfkelch13* was amplified with primers containing sequence overlap for Gibson assembly into the pDC2 backbone using primers (p265/p318 for HR1 and p267/317 for HR2). The HR1 -HR2 overlap encompassed mutations marked as silent 1 silent 2 and C580Y. **B.** Strategy for generating a donor sequence containing 100 bp of recodonised sequence as described, by ordering a synthetic sequence as a gBlock from IDT **C.** First editing strategy by utilising the pCC1 donor plasmid and the pDC2 Cas9-containing plasmid. The cloning strategy was as described earlier for pfrh3 in section 2.3.2.6. pCC1 contains the region of homology and both negative and positive selection cassettes. pDC2 contains Cas9 for expression under the control of the calmodulin promoter (Cam 5r), and expression of the gRNA from a U6 promoter. **D**. Single plasmid for expression of a codon optimised Cas9, a hDHFR selectable marker, the gRNA from a U6 promoter, and the K13 donor sequence for homologous recombination.

Fig. 5.2 shows a chromatogram of the region around position 580. The top panel shows the sequence obtained by using the first approach, highlighting the two silent mutations at the PAM site, and codon TGT for cysteine. The bottom panel shows a section of the silent mutations across the recodonised region.
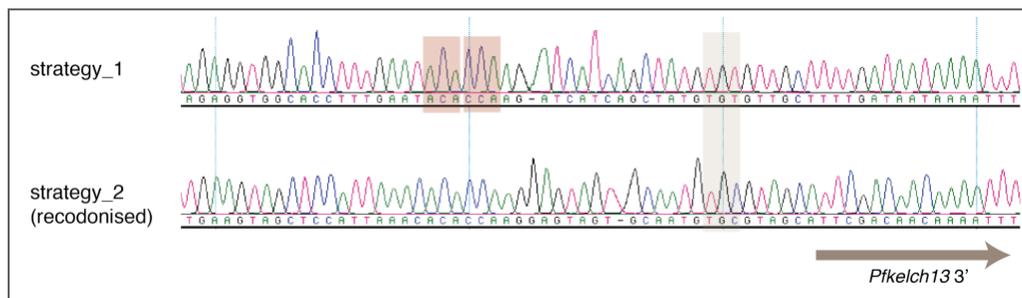
Figure 5. 2 Genotyping of plasmids with different donor sequences.
Sanger sequencing showing the two donor-design strategies, with two alternate C580 codons highlighted with the gray box. The top panel shows the standard strategy used in Ghorbal *et al.* 2014, by designing two silent mutation (red squares) at the gRNA binding site (Ghorbal *et al.* 2014). The bottom panel shows the strategy by which the 100 bp of the donor sequence are recodonised to 1). facilitate genotyping by PCR, 2). increase the opportunity to use multiple gRNAs targeting the same locus, and 3). drive recombination from the cut site past the mutation of interest.
.

5.3.1.2 Design of gRNAs targeting K13 and cloning into pDC2-Cas9 vector

Design of gRNA was performed in the same way as in section 3.3.1.2. Three independent gRNA sequences predicted by Benchling were selected based on distance to the

mutations of interest. The three red arrows on Fig. 5.3A indicate the gRNA sequences, next to the PAM sites, on the negative strand indicated by the orientation of the arrow. The strategy for gRNA cloning was the same as described in section 2.3.2.6, by standard T4 ligation with annealed oligonucleotides into a pDC2 backbone digested with BbsI (Fig. 5.3B). Three gRNAs were designed and cloned to target Pfkelch13, by using annealed primers p237/p238 (gRNA1), p614/p615 (gRNA2), and p616/p617 (gRNA3). They were all confirmed by Sanger sequencing by using p36 and are shown Fig. 5.3C with the BbsI site highlighted in blue.

Figure 5. 3 Design and cloning of gRNAs for targeting *Pfkelch13*.
 A. Wild type locus showing the two positions that were used as silent mutations for the donor design (silent), and position 580. Red arrows are the gRNAs used, with their respective PAM (B), identified by using Benchling. B. Primer ID and gRNA sequence, with restriction site BbsI (red and blue), and the cloning site flanked by the U6 promoter and the U6 terminator sequence. C. Sanger sequencing of three gRNAs using primer p36, with BbsI site highlighted in blue.

5.3.2 Mutations in *Pfkelch13* and their impact on resistance and fitness

5.3.2.1 CRISPR/Cas9-based editing of *Pfkelch13*

Once the correct insertion of the donor sequences into either the two or one plasmid approach described in Fig. 5.3 was performed, parasites were transfected using the different methods described (section 2.2). Overall ring transfections worked well for standard laboratory strain; however for recently adapted isolates (e.g. CAM), a preloading approach in which parasites were added to electroporated RBCs resulted in more efficient recovery. Nonetheless, as with other approaches, for the majority of transfections no parasites were recovered even after waiting for more than 50 days. More than 20 independent transfections were performed for each of the different strains used, and 10% of those had parasites become visible by microscopy 30 days or more post-transfection, all of which were subsequently confirmed to be successfully edited. Overall, the single plasmid approach (Fig. 5.1D) resulted in higher transfection efficiency than the two-plasmid system, however, given the low number of positive transfections we cannot assign causation to this targeting approach. The length of selection was the same as with the *Pfrh3* editing transfections described in Chapter 4: constant drug pressure for six days and then culture maintenance after transfection followed the protocol described in section 2.2.

The genotyping strategy for the first donor described was to amplify the relevant region in the endogenous *Pfkelch13*, using primers flanking the donor sequence to avoid amplifying DNA from any remaining episomes. Fig. 5.4 shows chromatograms of the bulk population prior to cloning of the different edited lines that were generated. The reference strain 3D7, used as a wild type strain for the previous experiments using BarSeq, was selected for introducing the C580Y mutation in *Pfkelch13*. To better understand how mutations in this gene can impact fitness differentially across genetic backgrounds, and in particular in a KEL1 lineage, this latter strain was transfected with a version of the editing plasmid to revert the mutation Y580 back to the reference allele C580. In addition, another epidemiologically relevant SNP was introduced. The

nearby mutation A578S is not linked with any resistance phenotypes, but was reported in various studies as being at high frequency in Africa (Amato *et al.* 2016; Maiga-Ascofare and May 2016). The chromatogram indicates a high efficiency of editing using gRNA1 (Fig. 5.3C), which had been published before (Ghorbal *et al.* 2014). The two additional gRNAs designed (Fig. 5.3A-B) did not yield parasites after transfection. This may mean that they are inefficient in targeting Cas9 to the cut site, although it should be noted that the combination of low efficiency for targeting an essential gene with stochasticity in the transfections, means that the absence of editing with these gRNAs is not necessarily attributable to either plasmid design or the gRNAs themselves. Once editing at the bulk population level was confirmed, dilution cloning plates were made to isolate clonal lines, which were expanded for confirmation, phenotype validation and next generation sequencing.
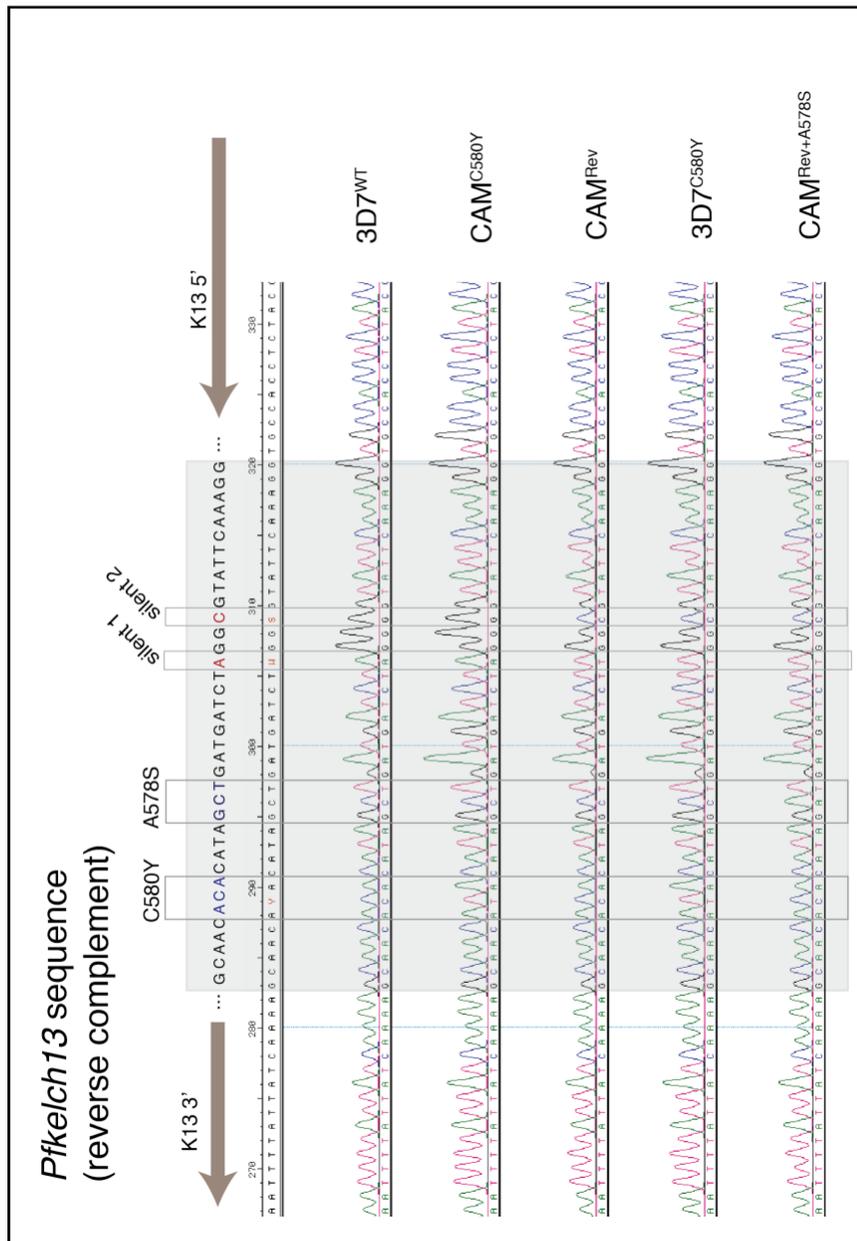
Figure 5. 4 Genotyping of edited locus from bulk culture.
Reverse complement sequence of *Pfkelch13* using sequencing primer p327 and showing the endogenous gene with some of the mutations of interested introduced using the first donor sequence strategy. These sequences are from the bulk population which was then cloned as in section 2.1.3. This reaction was performed by using KAPA Hot-Start Master Mix (section 2.3.2.2), with primer combinations p325/p328. PCR products were confirmed and purified as in section 2.3.2.5.

5.3.2.2 NGS analysis of CRISPR/Cas9-edited lines

Two clones per edited line (CAM^Rev and 3D7^C580Y were selected for whole genome sequencing. Genomic DNA was extracted as in section 2.3.1, and 1µg was provided to the Bespoke Team at the Sanger Institute for processing, library preparation and sequencing on an Illumina MiSeq platform. Once the fastq files were generated, I adapted scripts that allowed for mapping of the reads to the latest reference of 3D7 from PlasmoDB, using a Burrow-Wheeler Algorithm, as described in 1.2.1. Once BAM files were generated, they were indexed for both visualisation of reads and to perform SNP calling of the CRISPR-edited lines, using the pipeline from Snippy, which uses variant calling software Freebayes (https://github.com/tseemann/snippy). To visualise reads, the BAM files were directly loaded onto the genome viewer software IGV 2.4.8, from the Broad Institute.

Fig. 5.5A-B show an alignment of two independent clones for each strain that were sent for whole genome sequencing (C4, A5 for 3D7 and A5, B5 for CAM), together with the parent strains that were used for genetic modification. From the reads shown, the region shows good coverage and indicates a clonal population for all the transfected parasites, and highlighted are both the mutation of interest (amino acid position 580), and the two silent mutations designed as part of the donor sequence. The two sequences on the top are the two wild type strains used (parent strains), which do not show the silent mutations, and harbour the reference allele for *Pfkelch13* C580 for 3D7, or the Y580 allele in CAM that confers higher tolerance to artemisinin as measured by RSA$_{(0-3h)}$ (Straimer *et al.* 2015).

159
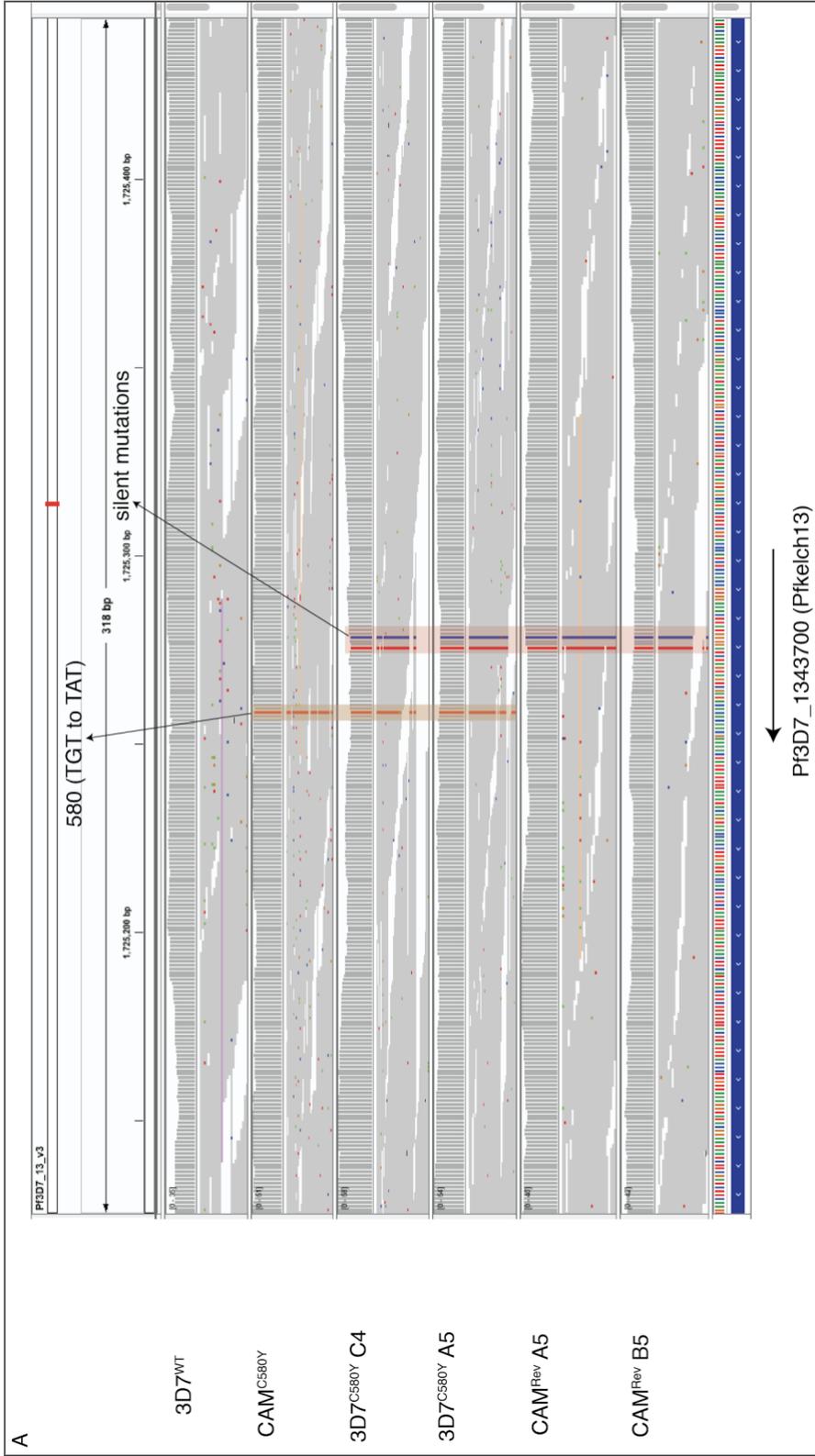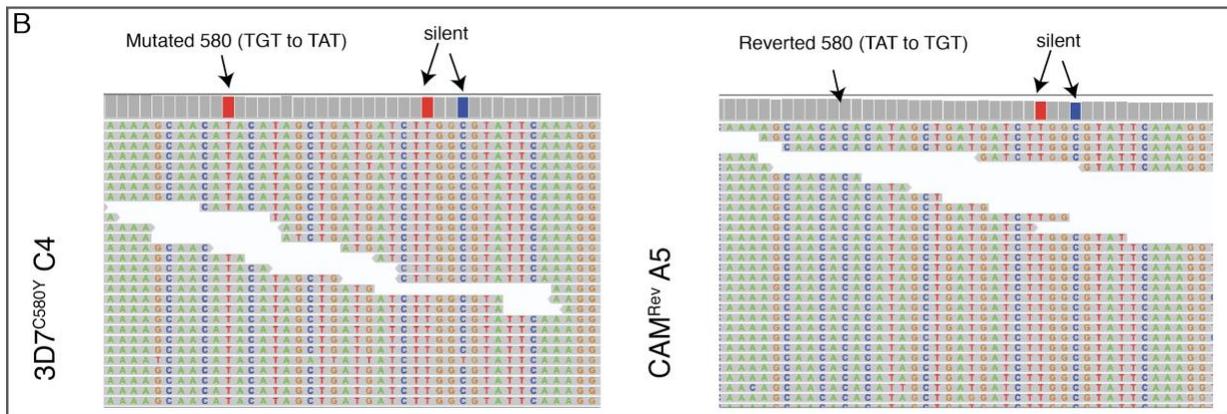
Figure 5. 5 Alignment of sequencing reads to the reference 3D7 using IGV
Reverse complement of whole genome sequencing reads of two independent clones for each CRISPR-edited line, using genome visualization software IGV from the Broad Institute (https://software.broadinstitute.org/software/igv/). **A.** Clones A5 and C4 for 3D7_C580Y MUT show successful editing of TGT to TAT (Cys to Tyr), and clones A5 and B5 show the TAT codon from CAM reverted back to the reference allele (codon TGT for Cys). Because all four clones are edited, the silent mutations are seen, marked with the red box. **B.** Zoom-in of edited region showing the mutated position 580, and the two silent mutations for a representative clone of each genetic background (clone C4 for 3D7 and clone A5 for CAM).

The results of the variant calling pipeline provide further evidence of the parent strain used for editing by aligning to the reference sequence of 3D7 (table 5.1). It also provided reassurance that the clones selected were actually clonal, by the absence of any wild-type reads aligning to the reference at this particular locus. The top row indicates the observed sequence in both lines for the introduced silent mutations (TGGC), and the second row the allele at codon 580. The last two columns show two representative clones and whether they contain the reference allele or an alternative, showing a C at position 580 for CAM[Rev], and a T in 3D7[C580Y], leading to the alternative amino acid at position 580.

Table 5. 1 Reference sequence for the position 580 and for the AA position 573.
Here the silent mutations were designed in the donor sequence. Variant calling software Snippy (http s://github.com/tseemann/snippy) was used to find the alternative sequence for both editing events.

LOCUS_TAG	POSITION	REFERENCE	ALTERNATIVE	AA POSITION	3D7^{C580Y} C4	CAM^{Rev} A5
exon_Pf3D7_1343700-E1	1725276	AGGG	TGGC	573	TGGC	TGGC
exon_Pf3D7_1343700-E1	1725259	C	T	580	T	C

From these results using whole genome sequencing analysis we were able to confirm editing using the CRISPR/Cas9 based strategy and the homogenous sequence reads allowed us to confirm the use of clonal populations for the subsequent experiments.

5.3.2.3 Amplicon sequencing strategy

In the previous chapters, I showed the first evidence that Barcode Sequencing could be adapted to *P. falciparum*, and that it was a useful technology to scale up drug resistance and parasite fitness phenotyping. I therefore aimed to perform a two-transfection approach to insert DNA barcodes into a non-essential gene (*Pfrh3*) within edited lines, but because of the limited number of selectable markers available for this species (Mamoun *et al.* 1999), and given the genetic bottlenecks that accompany each *P. falciparum* transfection (Hasenkamp, Russell, and Horrocks 2012), I developed an alternative approach for accomplishing a goal of measuring in vitro fitness of a limited number of competing parasites in a pool. Fig. 5.6 shows the strategy developed for this new approach, which entails amplicon sequencing of the *Pfkelch13* locus. The same principle applies as with the previous strategy but provides less overall flexibility in terms of the number of loci that can be monitored, as the amplification and counting is performed from the locus of interest rather than an independent and more complex barcode. With this approach, the mutations introduced into *Pfkelch13* were in effect used as a barcode to track the growth and appearance of different alleles over time. Fig. 5.6A shows the region

which is used as amplicon and the different combinations of mutations introduced into the genome of the two strains, 3D7 and CAM. The primers shown amplifying from the 5' and 3' region (in green and red, respectively), were designed as oligonucleotides p1492/p1493, and contain binding sequences for Illumina adapters shown as PCR2 in Fig. 5.6B. The box in Fig. 5.6B indicates the sequences that are used as barcodes for this approach, by employing the counting script described in section 2.5.1.2.

After confirmation of edited clones by standard genotyping methods and by next generation sequencing, they were mixed together at equal proportions and competition assays were performed as in section 4.3.2.
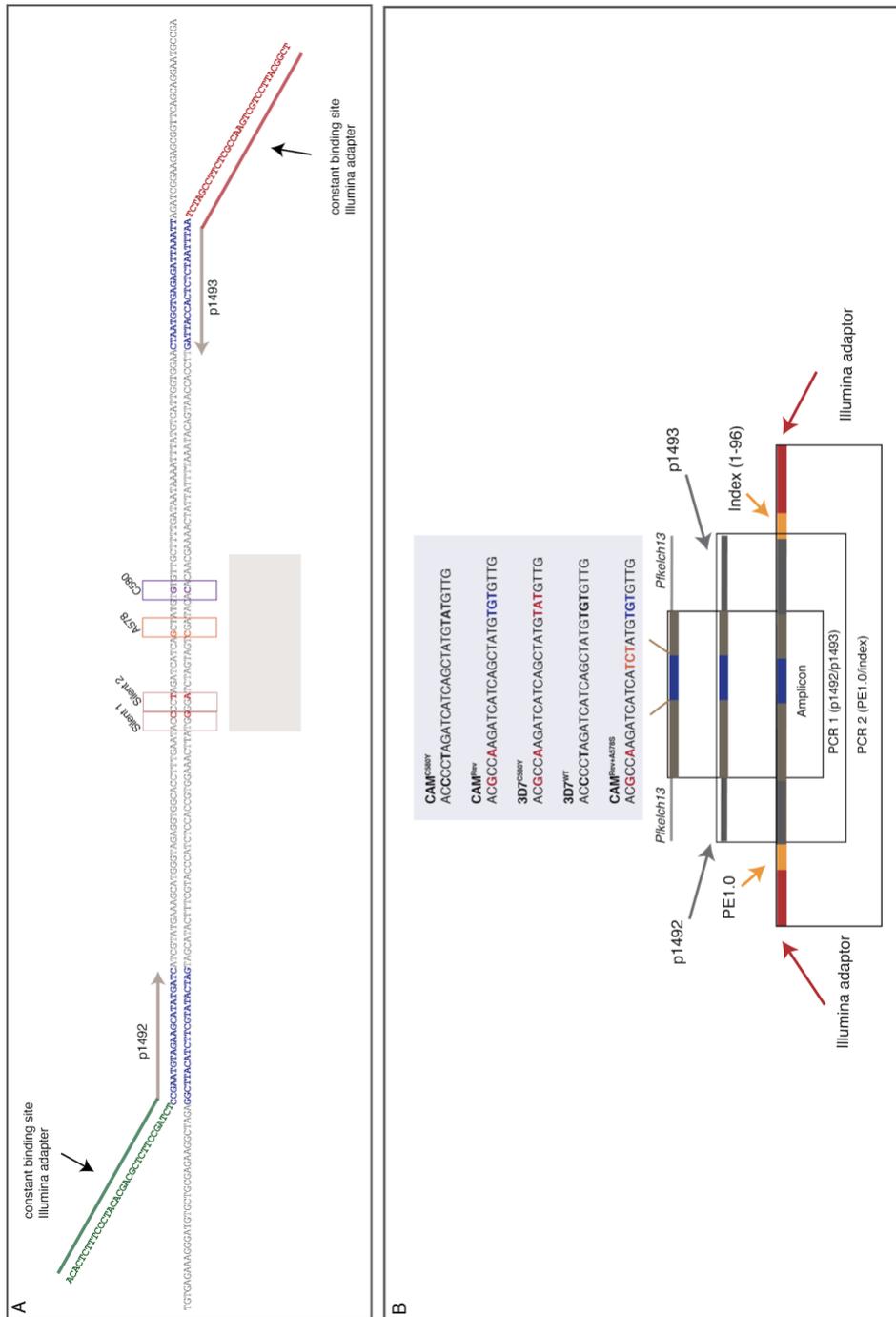
Figure 5. 6 Amplicon sequencing strategy for Illumina.
Overview of the method for generation of libraries compatible with Illumina platform MiSeq. **A.** Template for the nested PCR approach uses primers p1492 and p1493 containing constant binding sites for Illumina adapters (sequence in green). Gray box highlights the mutations introduced in different combinations for this donor-design strategy**. B.** Nested PCR reaction performed with the two sets of primers. Purple box highlighting the sequences used as ɛrcodes" for amplicon sequencing, from raw fastq reads (section 2.5.1.2).

5.3.2.4 A complex pool of CAM and 3D7 does not reveal any impact in growth by harbouring *Pfkelch13* mutations

From the previous chapter, I had shown that there are differences in inherent fitness amon g the strains used: V1/S for example, in the absence of drug pressure, is rapidly outcompeted by "fitter" strains such as 3D7, and even the more recently adapted strain CAM (Fig. 4.5 and 4.7A). In this section, the aim was to determine the contribution of *Pfkelch13* mutations to modulating fitness across the different strains, using the five different lines shown in Fig. 5.4. In the initial experiments, I mixed CRISPR edited lines from different strain backgrounds in order to assess whether this approach was sensitive enough to determine subtle differences coming from editing single mutations, or if the inherent fitness advantage of 3D7 would rapidly outcompete any CAM clone. Fig. 5.7 shows a competition assay lasting 35 days for the mixture of edited 3D7 and CAM lines. Similar to results from chapter 4, this assay shows an overall higher fitness for 3D7, compared to CAM, regardless of *Pfkelch13* haplotype (Fig. 5.7A). After day 7, the counts for CAM start to drop rapidly, an indication of it being outcompeted by 3D7, and at the later time points (from day 15), all CAM haplotypes are less than 10% of the total counts in the competition. Fig. 5.7A shows the proportion of each haplotype over time (left panel), and as a measure of fitness the mean proportion change over the entire time course is plotted on a log2 base scale (right panel). Below is a table showing the significance values obtained by comparing the mean across groups, showing significant differences between both 3D7 clones, and the three clones from C A M (p-value < 0.05). However, no significant differences were observed between isogenic clones for either of the strains. To understand the rate at which the different isogenic clones grew, the ratio of all the mutated alleles to its wild type was measured   until day 24 (Fig. 5.7B). The later measurements beyond day 20 were discarded from this analysis because the counts were below the threshold described earlier. The ratios for both mutated alleles of CAM relative to wild type declined, as measured by using a linear regression, and the mean slope of that regression show both have negative values, with CAM[Rev+A578S] having a lower average slope (adjusted p-value 6x10-4), indicating it is decreasing more rapidly than the other clones. Contrary to the CAM lines, the

slope for the mutated 3D7 parasite remained close to 0, showing that it is not changing when compared to the wild type allele, suggesting no appreciable loss of fitness by having the mutated version of *Pfkelch13* on the 3D7 background (Fig. 5.7C).
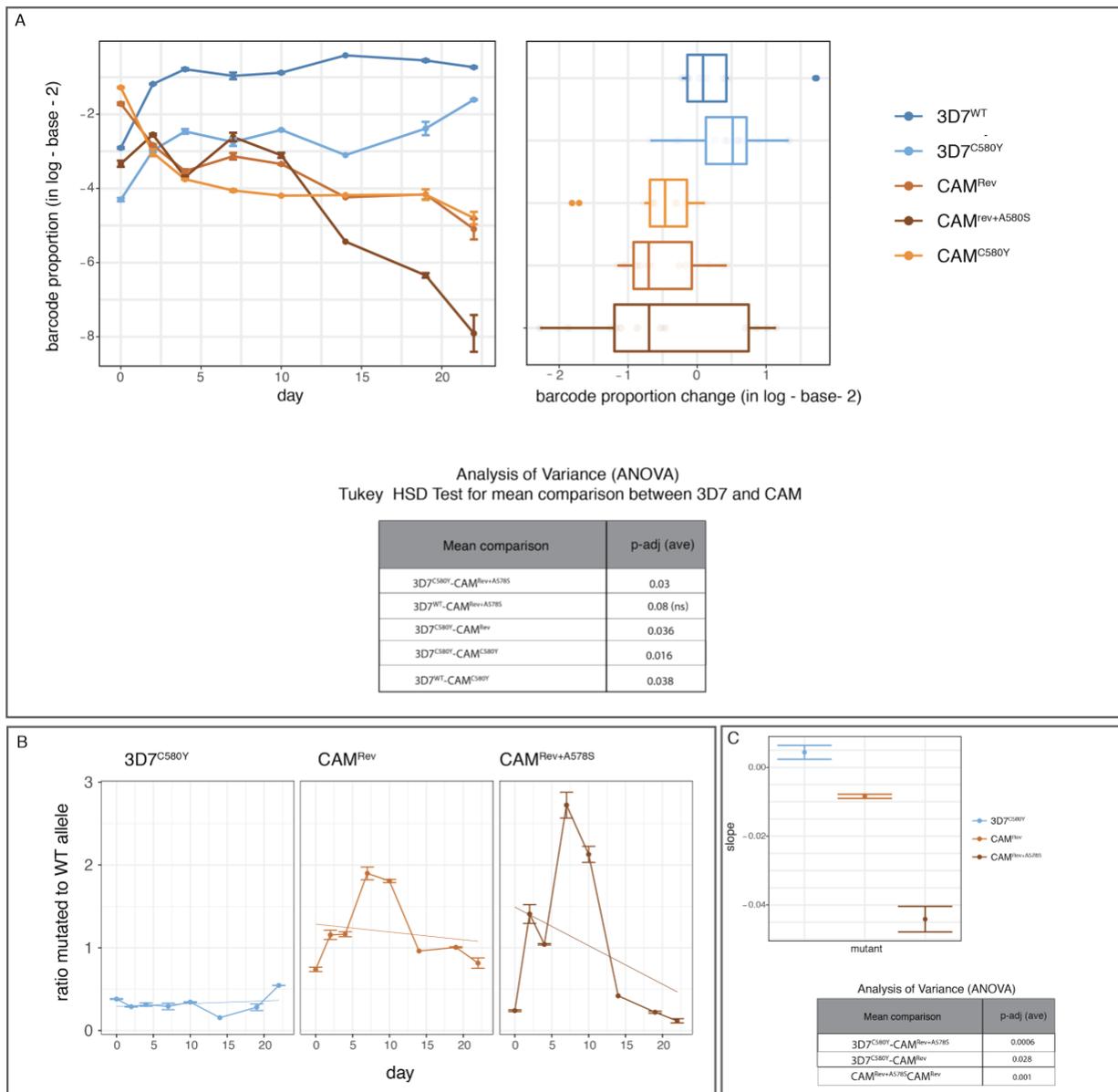
Figure 5. 7 Amplicon sequencing to measure the relative proportion of CRISPR-edited lines mixed in a pool.
A. Competition assay of five independent clones (3 for CAM and 2 for 3D7), carrying different *Pfkelch13* alleles, their growth over time (left), and change in relative proportion of total read counts in a log2 scale (right). Below the results of an ANOVA between groups and a TukeyHSD test to perform pairwise comparison of all means. **B.** Second approach to measure difference in growth across strains by performing a linear regression on the ratio of mutated to wild-type alleles, and **C.** the mean slope of the linear regression with p-values obtained as in (A).

5.3.2.5 Change in frequency of amplicons in pool reveal drug sensitivity profiles to known antimalarials

To understand whether this pool could reveal phenotypes of the different strains generated, and whether having the mutated allele would impact sensitivity to antimalarials, a panel of four known compounds were used, and short competition assays were performed in order to measure drug sensitivity. Fig. 5.8 shows the results of the competition assay lasting eight days: the y-axis representing the change in allele proportion measured in a log2 base. As expected, and consistent with the results from chapter 4, sensitivity of the panel of barcoded strains to chloroquine revealed resistance of CAM and inversely, sensitivity of 3D7 to this compound, as also shown in Fig. 4.7. Given the relevance of studying the impact of different *Pfkelch13* alleles in modulating drug response, three additional antimalarials were selected based on their clinical relevance in the context of multidrug-resistant parasites. Fig. 5.9 shows the allele proportions over time for atovaquone, lumefantrine, and halofantrine, tested at a concentration of 3XIC50, based on the values reported by Delves *et al.* 2012 for NF54 (parental line of clone 3D7). From the growth curves in the presence of these compounds, and from measuring the change of allele proportion across the eight days, atovaquone and halofantrine show no influence on the growth of specific lines, as their mean in log2 is 0. Strikingly, the wild type CAM Y580 line showed an increased survival when exposed to halofantrine in comparison to the CAM-Y580C mutant, indicating that the *Pfkelch13* allele may also modulate response to halofantrine. This compound is a candidate antimalarial with a similar structure to lumefantrine, however concerns have been raised regarding potential toxicity (Bouchaud *et al.* 2009). Based on these results, future work will entail repeating these assays as well as conventional dose-response experiments, with the aim of confirming the phenotype and to study potential interactions of this *Pfkelch13* allele with halofantrine, and whether this compound can synergise the effect of an artemisinin derivative in wild type *Pfkelch13* parasites. Given the importance of the C580Y mutation in modulating the response to the frontline antimalarial artemisinin, it will be important to continue studying the phenotype of novel

compounds in the context of drug-resistant parasites, and to evaluate their efficiency when used in combination.
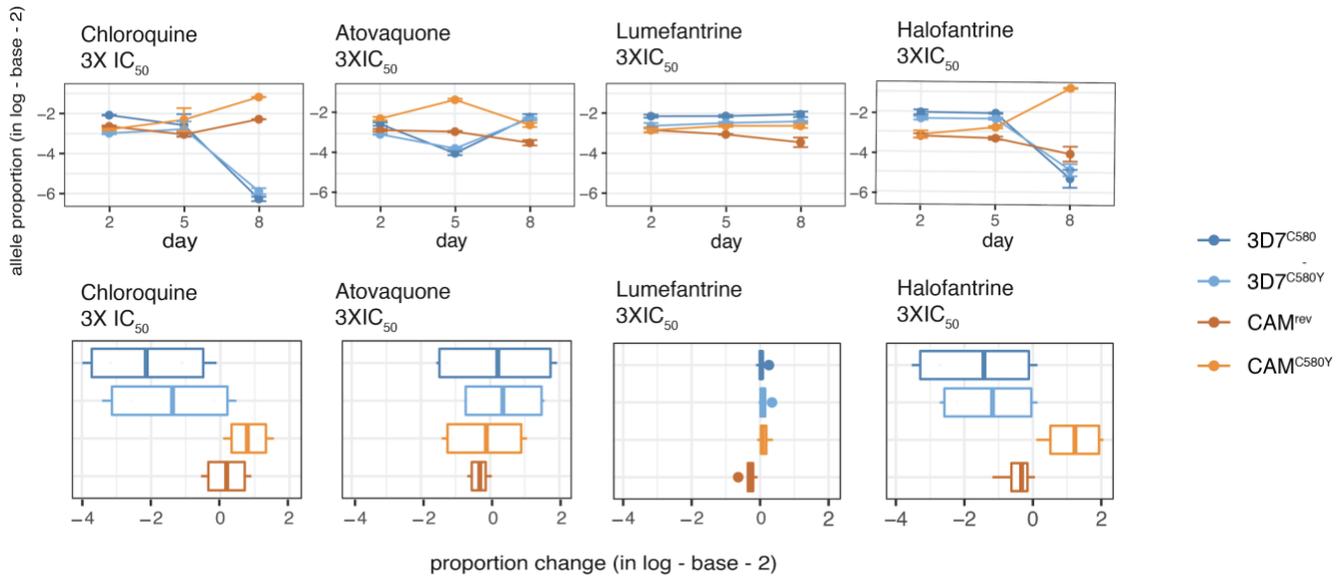
Figure 5. 8 Amplicon sequencing of CRISPR-edited lines in the presence of different antimalarials.
Growth curves of competition assays of four clones (top panel), with the y axis showing time, and the x-axis the proportion of each of the four alleles as a log - base - 2. Panel below shows the change in allele proportion over time, and is represented as the log - base - 2 on the x-axis. IC50 values were 0.6nM, 2.8nM, and 0.9 nM for atovaquone, lumefantrine and halofantrine, respectively (Delves *et al.* 2012), and 3 that value was used for the competition assay. Technical triplicates were performed for all conditions.

5.3.2.6 Competition between isogenic parasite clones carrying different Pfkelch13 alleles reveal a strain-dependent fitness cost

As discussed above in Fig. 5.7, when the two edited strains CAM and 3D7 are mixed together, the effect of the fittest strain (3D7) overshadows any minor difference that might be related to the contribution of specific *Pfkelch13* haplotypes to fitness. To overcome this and to shed light into their effect, only isogenic lines were pooled following the same strategy used in section 5.3.2.4. Two competition assays were performed with durations of 54 (Fig. 5.9A) and 35 (Fig. 5.9B) days, with genomic DNA samples collected every 3-4 days. Fig. 5.9A shows the growth curves of the 3D7 clones carrying the two different *Pfkelch13* alleles: WT (dark blue) and

C580Y (light blue). For 3D7, the competition experiment shows that the WT, or reference allele, grows better than the C580Y mutated allele. In a study undertaken by Straimer *et al.* 2017, a pyrosequencing approach was used to measure the contribution to parasite fitness of different *Pfkelch13* mutations that had been previously edited into different genetic backgrounds. By performing head-to-head competition assays with the wild type and mutant parasites of isogenic backgrounds, Straimer *et al.* also observed that the C580Y allele had a fitness cost in some genetic backgrounds, however for the two artemisinin-resistant Cambodian parasites there was no effect on fitness by having this mutation, when compared with the artemisinin-sensitive reference allele generated using nuclease-based genome editing. This would support the importance of certain genetic backgrounds in shaping the epidemiology of artemisinin resistance. In addition to this finding, the group of Tim Anderson recently showed that C580Y had a higher fitness cost than a less widespread mutation such as R561H, when introduced into a wild type Southeast Asian isolate, and thus supports the evidence provided by Straimer *et al.* 2017 (Nair *et al.* 2018; Straimer *et al.* 2017). These findings collectively are consistent with our hypothesis that mutation of *Pfkelch13* in a 3D7 background would result in impaired fitness, given that 3D7 is originally an African strain that has been adapted to culture for over 40 years, and has a wild type phenotype to all antimalarial compounds.

The observation of fitness differences was only apparent after more than 20 days of continuous co-culturing. Before day 20 for both experiments there was a considerable amount of noise that could originate from different sources. One potential source of initial variability may be phasing of the lifecycle stage of the two lines, with alternating replication and rupture cycles contributing to fluctuations, and future work could repeat the competition assay using more tightly synchronised parasite clones. Additionally, for these two experiments, the initial parasitaemia or hematocrit was not perfectly matched between clones, with the wild type clone being almost three times higher than the mutant clone. For the second experiment (Fig.5.9B), a similar fluctuation was observed as with the previous assay, however the striking peaks may be a result of a less frequent sampling of genomic DNA (only 4 time points until day 20). For these reasons, calculation of the average change in proportion was taken from day 20 until the last day of the competition. By doing this analysis, the second panel of both figures

shows an increase in the average change in proportion for the wild type allele. From both experimental set-ups, it is clear that the effect of the mutated allele in a 3D7 background is modest, and the way the assay conditions are established (e.g. length of time of competition) will strongly impact the ability to draw statistical significance. When performing a t-test for the first competition assay, no significant difference between the two clones on the selected time range (p-value = 0.07) was observed, explained by the fact that the proportion of read counts for the mutant line on the last days of competition remain at less than 10% of the total counts of the competition, effectively implying that at that point the mutant cl one is completely outcompeted. The difference for the two clones on the second competition assay, however, was significant (t-test, p-value=0.004), however given the time range selected, only three time points were considered, three replicates for each.

For the CAM strain belonging to the KEL1 lineage described by Amato *et al.* 2018, I hypothesised that reverting the artemisinin-resistant C580Y mutation back to the sensitive reference allele may not provide an enhanced fitness as it would in other genetic backgrounds, potentially related to differential epistatic interactions unique to these successful lineages (Amato *et al.* 2018). Competition assays between isogenic clones for CAM were performed by taking the same approach as with 3D7. The first experiment lasted 54 days, and the change in allele proportion is shown Fig. 5.10A. Unexpectedly, competing a CAM strain, harbouring the C580Y allele, to the reverted (CAM^Rev) resulted in a loss of fitness, indicating the drug-resistance allele was preferred. Unfortunately, for the first experiment, a 3D7 contamination became evident above the sequencing noise threshold after day 32 (dashed line), and similar to the observation from section 5.3.2.5, 3D7 rapidly outgrew both clones of CAM resulting in termination of the experiment. The second experiment shown in Fig. 5.10B shows a consistent and striking difference in the growth of the two clones, as the Y580C revertant is outcompeted by the wild type CAM. A t-test on the average change in proportion of the two clones in this experiment showed the difference to be statistically significant (p-value = 1.26x10-9).

The results from performing competition assays on variants generated in isogenic backgrounds show: 1. The artemisinin-resistance variant C580Y in a KEL1 lineage provides

enhanced fitness relative to the sensitive allele, whereas in a standard laboratory strain 3D7 the opposite effect on fitness is observed indicating that there might be an epistatic interaction that differs across strains, impacting differentially on fitness. 2. The importance of having isogenic control clones competing together when performing these assays is highlighted, as minor differences will only be evident in this context. When mixing highly diversified strains, such as 3D7 and CAM the resolution of the contribution of particular alleles to fitness is easily lost. In order to investigate the rate at which each clone decreased in proportion over time, a linear regression was used following the same approach as in section 5.3.2.5. Figures 5.9C and 5.10C show the ratio of the mutant to the wild type allele for 3D7 and CAM, respectively. For both cases, a linear regression of the ratio resulted in a negative slope, indicating a reduced growth for the mutated allele, for both experiments.
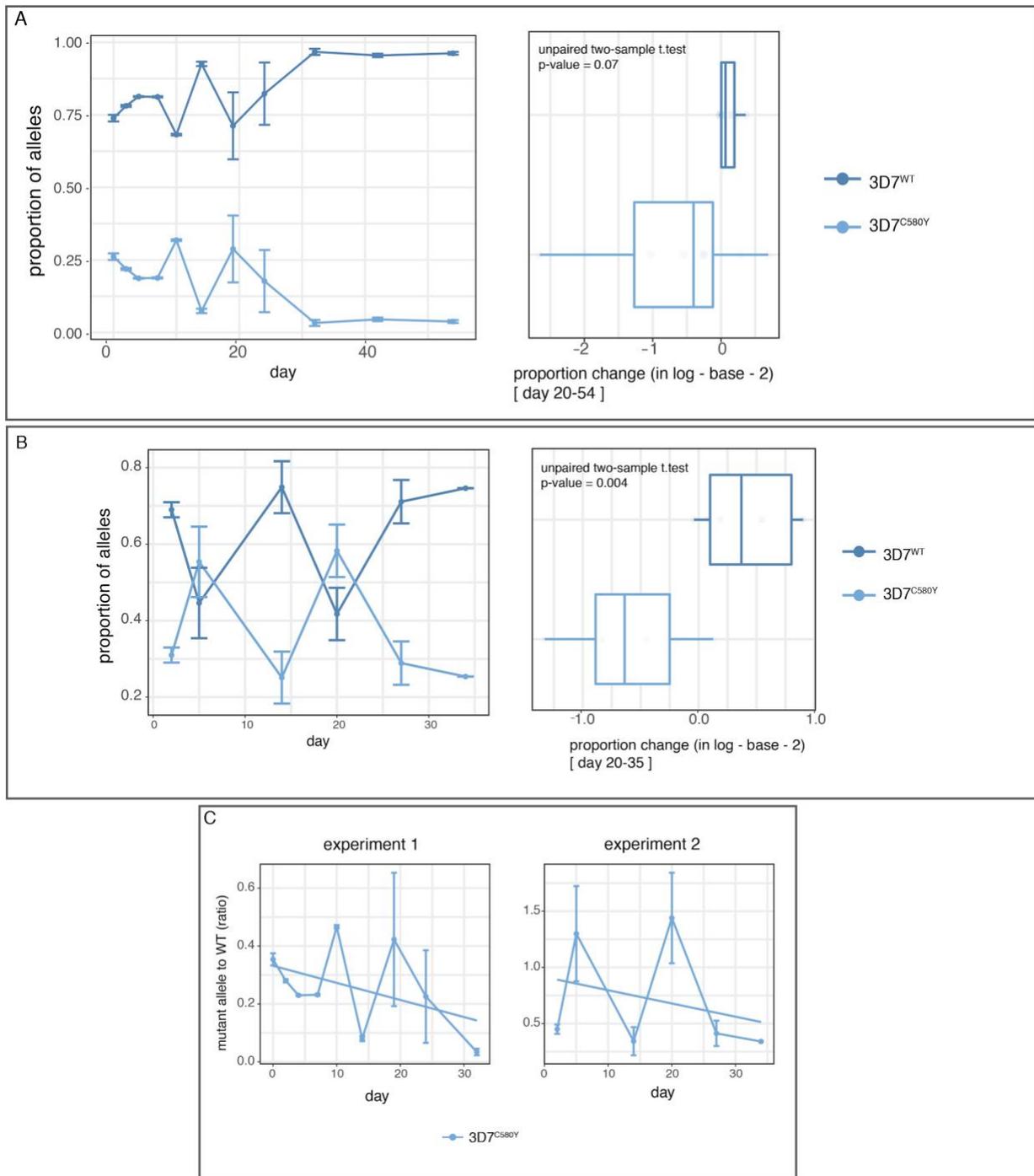
Figure 5. 9 Amplicon sequencing strategy used to measure the relative proportion of isogenic lines of 3D7, with Pfkelch13 alleles generated using CRISPR/Cas9-based editing.

A-B. Two independent experiments showing competition assays of two clones lasting 54 days (**A**), or 35 days (**B**) (time shown on x-axis), with technical triplicates. The relative allele proportion is shown on the y-axis, and for both A and B, on the right the change in relative proportion from day 20 until the end of the assay are shown in a log - base - 2 scale. An unpaired two-sample t-test was performed for both experiments, and significance levels are shown. **C**. A second approach to measure the difference in growth across strains by performing a linear regression on the ratio of mutated to wild-type alleles for the two experiments.

172
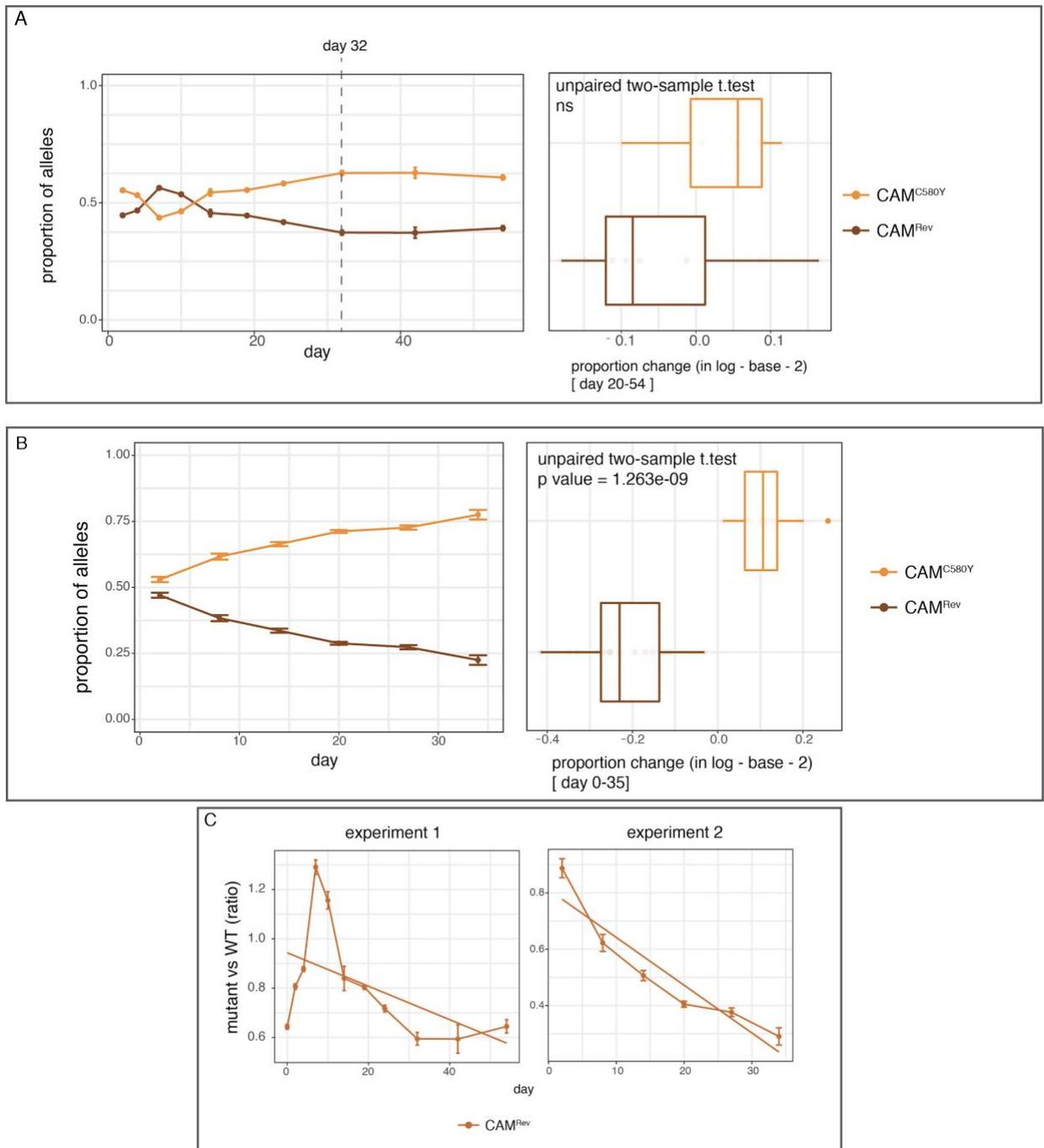
Figure 5. 10 Amplicon sequencing strategy used to measure the relative proportion of isogenic lines of CAM, with *Pfkelch13* alleles generated using CRISPR/Cas9-based editing.
A-B. Two independent experiments showing competition assays of two clones lasting 54 days (**A**), or 35 days (**B**) (time shown on x-axis), with technical triplicates. The dashed line on experiment 1 is explained in detail in section 5.3.2.6, and refers to a 3D7 contamination taking over the culture from this day. The relative allele proportion is shown on the y-axis, and for both A and B, on the right the change in relative proportion from day 20 until the end of the assay are shown in a log2 scale. An unpaired two-sample t-test was performed for both experiments, and significance levels are shown. **C**. A second approach to measure the difference in growth across strains by performing a linear regression on the ratio of mutated to wild-type alleles for the two experiments.

5.3.3 Why "Y"? – The role of genetic background in tolerating resistance-associated mutations.

5.3.3.1 CRISPR/Cas9-based editing and high throughput sequencing reveal the impact of genetic background in tolerating mutations in *pfkelch13*

From the results of the competition assays, it became clear that the Y580 allele in *Pfkelch13* differentially impacts fitness for two different backgrounds studied: 3D7 and CAM. There is a background-specific effect that allows for growth to be enhanced while sustaining the artemisinin-resistant allele, which could be a result of positive epistasis, as described earlier (He *et al.* 2010). In order to validate the phenotypes observed, and to bring some molecular evidence to support the striking epidemiology of the KEL1 lineage in Southeast Asia (Amato *et al.* 2018), a novel strategy was developed that combines CRISPR/Cas9-based editing and high throughput sequencing in order to generate mutant lines for *Pfkelch13* that have the potential to encode any codon at position 580, in effect mimicking spontaneous mutation at this important residue. But using different genetic backgrounds and an amplicon sequencing strategy to measure the complexity of mutants obtained post-transfection, the aim was to test whether differences in genetic background would result in different outcomes. The strategy is shown in Fig. 5.11. This donor template for this approach was synthesised so that it would contain a homology sequence of 750bp, containing 100bp recodonised as described in section 5.3.1.1, and at position 580 contain a random codon to repair to. To accomplish this, a gBlock (IDT) was synthesized (section 2.3.2.1), providing as template NNN in the described position, meaning that within the pool of donors, all possible codons should be represented. The donor sequence was assembled in triplicate reactions following the same protocol as shown in Fig. 5.1B using Gibson assembly, and then transformed into competent cells as in section 2.3.2.7 and recovered in a large culture volume to proceed directly to midiprep in order to maintain library complexity. To verify the complexity of the plasmid pool, a fraction of the transformation was plated and 12 minipreps were performed from colonies generated, for Sanger sequencing as shown in Fig. 5.12A. By sequencing 12 clones, an estimate of the

complexity of the pool could be assessed, showing 10 different codons, including stop codon TAG.

To fully assess that the plasmid mixture was a highly complex pool representing all possible amino acids, as was suggested from Sanger sequencing of a small number of individual clones, an amplicon sequencing strategy was developed as shown in Fig. 5.11A-B. The proportion of each of the 64 codons with the donor pool is shown as a log10 scale on the y-axis, and the codon on the x-axis. Leucine had three codons that were not well represented in the pool (CTN for Leu), however alternate codons TTA and TTG were well represented. From this result we were able to confirm a good representation of all possible amino acids (shown in Fig. 5.12C).
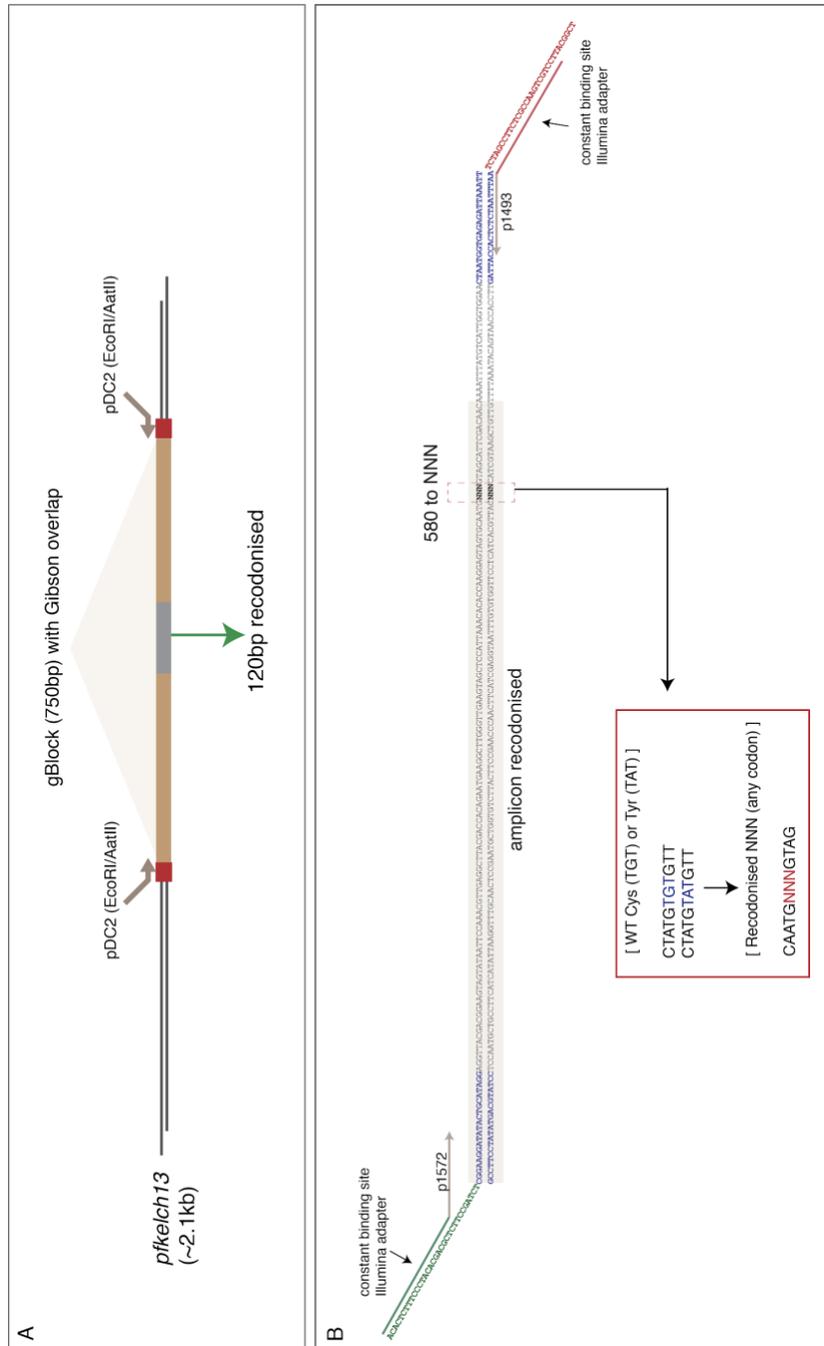
Figure 5. 11 Overview of method for generation of libraries compatible with Illumina platform MiSeq.
A. Design of donor sequence with 100bp of recodonised region, and flanking regions totaling 750bp. The fragment was synthesized as a gBlock to assemble using Gibson into a pDC2 backbone previously digested with EcoRI and AatII. **B.** Template for nested PCR approach uses primers p1572 and p1493 containing constant binding sites for Illumina adapters (sequence in green). Gray box highlights the recodonised region and position 580, which can encode any codon. Red box shows sequences used as "barcodes" for amplicon sequencing, from raw fastq reads (section 2.5.1.2), where NNN stands for any codon provided as a list.

Figure 5. 12 Genotyping of plasmid pool of pDC2 backbone with recodonised donor sequence for homology-directed repair of *Pfkelch13*, with NNN representing any codon to replace position 580.
A. Sanger sequencing of bacterial clones after transformation with plasmid codon pool", showing 12 independent clones and 10 unique codons, including stop codon TAG. **B.** Codon complexity of transfection-ready plasmid pool measured by Next Generation Sequencing, the y-axis showing the proportion of each codon on a log10 scale. Leucine codons CTN were underrepresented, however codons TTA and TTG are present for this amino acid. **C.** Pie chart showing the proportion of amino acids by taking the average of the different codons shown in (**B**).

Parasites were transfected with the pooled donors using the protocol described (2.2.1.2 and 2.2.1.3). Once parasites became visible by microscopy they were expanded and genotyped to confirm editing (as in 2.3.1), prior to amplicon sequence of the 580 position. Overall, more than 60 transfections were performed into different genetic backgrounds listed in table 2.1. The time it took the different strains to come up post-transfection varied across strains: transfections in a CAM background took over 35 days, whereas non-Cambodian backgrounds took between 20-30. Additionally, the integration efficiency was assessed by using primers that would amplify outside the region of homology, and a primer that would only amplify either the mutated, or wild type locus. Only transfections which showed a high editing efficiency as measured by PCR and Sanger sequencing were submitted for Next Generation Sequencing. Fig. 5.13 shows the proportion of each amino acid in the total read counts for each transfection for which a library was prepared using amplicon sequencing. It was shown previously in section 5.3.2.6 that for the CAM strain from the KEL1 parasite lineage, there was a fitness advantage by harbouring the Y580 allele in *Pfkelch13*. Notably, through this new approach that provides all possible codons for incorporation, when the codon complexity of the transfected parasites was analysed, the only strain that had incorporated tyrosine at the 580 position was CAM. This result provides strong support for the hypothesis that there is positive epistasis in this genetic background which may be facilitating the spread of resistant parasites due to enhanced fitness coupled with drug resistance. It should be noted that any native tyrosine would not be counted because only mutated parasites could be amplified using the recodonised region that was designed for the donor sequence, as shown in Fig. 5.11B.

Neither of the codons encoding for tyrosine were obtained in any of the other genetic backgrounds used for this study. In addition to cysteine, however, there were multiple codons

that were obtained by using this approach. From epidemiological studies, mutations in *Pfkelch13* which have been linked to clinical artemisinin resistance in Southeast Asia have mostly been mapped to the propeller domain of *Pfkelch13*, which has been associated in other organisms as being involved in protein homodimerization for nuclear translocation (Zipper and Mulcahy 2002). Because of the essential function of this domain, understanding how mutations conferring artemisinin resistance affect the structure of the protein is essential. In 2016, Singh *et al.* published a study in which these mutations were mapped to the six-blade structure of this protein. Their analysis was based on the crystal structure of Kelch13 of *P. falciparum* deposited in the Protein Database (PDB 4yy8 or 4zgc). Fig. 5.14 shows the structural mapping of these resistance mutations in the propeller domain performed by the authors, with residue 580 located at the central channel. The structural model shows residues falling into three clusters: main clusters 1 and 2 containing most residues (red and yellow, respectively), and residues C580 and Y493 falling in between the two clusters, labelled in green. Analysis of clinical parasite samples containing mutations in Pfkelch13 revealed that mutations in the second cluster were less common than cluster 1, and these residues are not surface exposed, therefore mutations would likely affect the structure of the protein. C580, the most prevalent residue mutated in Southeast Asia, was expected to form a disulfide bond with C532 (PDB 4zgc), and a hydrogen-bond with G533 (PDB 4yy8) (Singh *et al.* 2016).

My data indicates there are four major amino acids that can replace Cys at position 580, meaning that they can be structurally accepted, and the impact they might have on the different genetic backgrounds will be discussed in section 5.4. For the non-Cambodian, non-artemisinin tolerant genetic backgrounds, changes to amino acid serine (TCG), isoleucine (ATC) methionine (ATG), valine (GTT) and back to cysteine (TGT) were obtained. CAM, on the other hand, is the only background tolerating tyrosine (TAT), plus additional codons for cysteine (TGT), serine (TGC and AGC), glycine (GGA) and lysine (AAG), and their implications in terms of structure will also be discussed. Overall, these results provide evidence that epistasis might be impacting the phenotypic outcome differentially on these backgrounds.
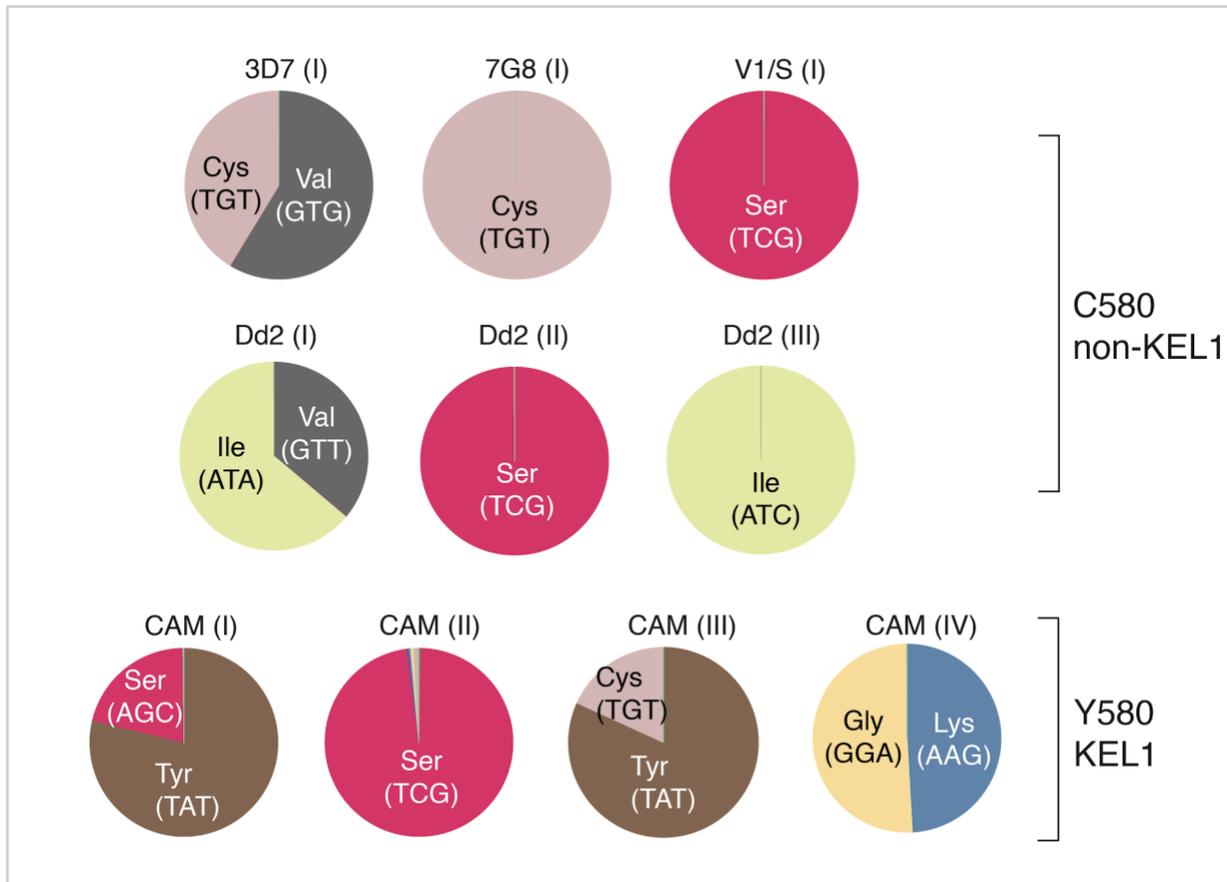
Figure 5. 13 Pie charts showing a summary of the different transfections analysed by amplicon sequencing.
The original *Pfkelch13* haplotype is noted on the right, and in brackets the transfection number, with multiple transfections performed for Dd2 and CAM. The proportion of alleles only represent mutated parasites and not wildtype residues, as the amplicons were generated using the nested PCR method that only amplifies the edited locus.
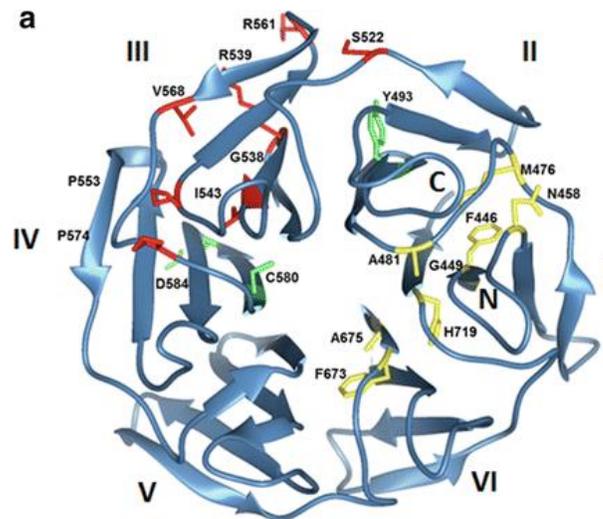
Figure 5. 14 Modeling of the propeller domain of K13 from Singh *et al.* 2016.
Figure shows 19 mutations identified in genome sequencing studies, some with clinical relevance measured as delayed parasite clearance or by performing in vitro or in vivo $RSA_{(0-3h)}$. Red and yellow indicate the two clusters identified. Green corresponds to the two residues that were not in any of the identified clusters: C580 and D584. The numbers indicate the different blades conforming to the propeller domain (Reproduced from Singh *et al.* 2016) (Singh *et al.* 2016).

To understand whether any of these alternate codons were found in other *Plasmodium* species, we downloaded all the genome assemblies available for this genus from NCBI RefSeq, using a script developed by NCBI (https://github.com/kblin/ncbi-genome-download). The sequence of the gene of interest for reference strain 3D7 (kelch, with RefSeq ID CAD52567.1) was blasted against all the sequences that were downloaded for the other species in order to generate a multiple sequence alignment, using the software MAFFT, from the European Bioinformatics Institute (https://www.ebi.ac.uk/Tools/msa/mafft/). This alignment was used to build a distance matrix for inferring phylogenies using FastTree (M. N. Price, Dehal, and Arkin 2009), and for visualization the ETE3 toolkit was used (http://etetoolkit.org/). Fig. 5.15 shows a phylogenetic tree and a region of 50 amino acids, with C580 highlighted in yellow. The cluster of *P. falciparum* is composed of all the sequences that were available from isolates deposited at NCBI, which did not contain any artemisinin resistant strain with the C580Y mutation. From our codon editing results, the most frequently obtained alternative codon was serine, with two codons observed: TCG and AGC. By building the tree of Kelch sequences shown in Fig. 5.15, we

observed that serine was the only amino acid that existed as an alternative to the highly conserved cysteine observed across all the *Plasmodium* species used for the alignment, and was found in *Plasmodium gallinaceum* and *Plasmodium relictum*, both avian malaria parasites. The presence of serine at position 580 in Kelch proteins from other *Plasmodium* species suggests structural tolerance for this residue.

Figure 5. 15 Phylogenetic tree using a distance matrix generated using FastTree from an alignment of K13 sequences across all the *Plasmodium* species.
Sequences used were available from NCBI. Only a segment containing 50 amino acids is shown, with residue 580 highlighted in yellow, with only one alternative amino acid present in two of the *Plasmodium* species used for this analysis.

5.4 Discussion and future work

The interest in this chapter was to characterise the impact of K13 mutations on a Cambodian strain belonging to the KEL1 lineage from Southeast Asia(Amato *et al.* 2018). I started with the hypothesis that spread of this lineage carrying the C580Y mutation in K13 could be related to fitness, with certain genetic backgrounds better tolerating these resistance-associated polymorphisms, and that this may be one of the factors underlying the current absence of K13-mediated artemisinin resistance in Africa (Henriques *et al.* 2015). Previous studies had shown that C580Y impacted fitness differently depending on the genetic background (Straimer *et al.* 2017), either in a negative or neutral manner. In addition, the C580Y mutation incurred a higher "cost" compared to the impact of a different, less widespread K13 variant seen in the Thai-Myanmar border(Nair *et al.* 2018). Overall, both studies ultimately highlighted two aims that should be considered carefully to improve drug resistance surveillance: 1. The importance of understanding the fitness landscapes of *P. falciparum* parasites at a local and global scale, and 2. That genetic background is important and can differentially modulate the phenotypic outcome when a selection pressure is applied. The starting point for the experiments performed in this chapter was, therefore, to understand why even though resistance-associated variants can emerge spontaneously (Amato *et al.* 2018), only certain genetic backgrounds might facilitate spread.

In this chapter I used a CRISPR/Cas9-based editing approach in two different strategies for site-specific genome editing of K13, a gene encoding for a protein shown to be involved in mechanisms of resistance to frontline antimalarial artemisinin in Southeast Asia. The first donor design strategy for making the targeting constructs contained the single positions to be mutated in the gene (e.g. C580Y, Y580C and A578S), and silent mutations at the PAM site to protect the locus from further cleavage by Cas9 (Ghorbal *et al.* 2014). The second strategy

introduced a recodonised ("silent") sequence of 100bp that would facilitate genotyping and would potentially improve efficiency of capturing the desired mutation. Both strategies worked for mutating K13, and were used to tackle questions regarding the impact of particular genetic backgrounds to parasite fitness.

The extent to which C580Y has a background-specific effect on the different strains used, and in particular on CAM, was studied following the first donor strategy described. Competition assays and amplicon sequencing for all five clones described (from both 3D7 and CAM), confirmed a need for using isogenic clones, rather than combining different strains, for long term co-culture. The latter conclusion came from the finding that the growth rate of 3D7 is higher than CAM, and that after less than ten cycles, it outcompetes the less fit strain. For this reason, minor differences between isogenic clones carrying these epidemiologically relevant alleles might be obscured (Fig. 5.7). The outcome of this first experiment led me to a similar conclusion as with the competition assays performed with the different strains barcoded in *Pfrh3*: difference in growth across different strains can be a consequence of varying multiplication rates, or an effect of an altered invasion phenotype. However, I also show evidence that combining the two strains, regardless of the different growth rates, can be used as a tool to perform short-term pooled assays (up to eight days) in the presence of antimalarial compounds to study the effect of K13 mutations on modulating response to other antimalarials, in a similar way as what was shown in section 4.3.2.2. Interestingly, these results showed that K13 Y580 in CAM results in higher tolerance to halofantrine, a result which will be validated given the important epidemiological consequences for the field (Fig. 5.8).

By competing isogenic clones head-to-head, I showed that C580Y displayed a fitness cost in the 3D7 background, however this effect only became significant after 20 days of continuous co-culturing. Future competition assays will explore the impact of changing the ratio of each clone to understand how it impacts the overall fitness of the less fit clone. For example, from the first competition assays, the initial parasitaemia of both lines was not the same, and the wild type strain started with an initial proportion of the total reads of 75% (Fig. 5.9). Several reasons might explain this. First, if there was a difference in stage between the two lines, that

might explain the initial fluctuations in total genomes until cultures became more asynchronous. However, this was carefully monitored by Giemsa smear, therefore this is unlikely to be the reason for the skew in the ratio WT/MUT. Second, if the haematocrit was not the same from the start of the experiment, it is possible that a 3-fold difference in the initial proportions might be partly due to an excess in red blood cells in the wild type clones. I think that this is one plausible explanation and, in the future, I will perform total cell counts using a haemocytometer before initiating competition assays. Overall however, the allele proportion remains unchanged over the first 20 days, indicating that perhaps a skew in the initial proportion would not impact the outcome of the competition or the study of the effect this particular allele (K13) has on fitness. A similar observation was published in two independent studies while this work was being performed, showing that Y580 has a fitness cost that differs based on genetic background (Judith Straimer *et al.* 2017), and that when introduced into a wild type parasite, it has a higher "cost" than other K13 mutations seen in the field (Nair *et al.* 2018). My results are consistent with both of these studies and with many other studies showing how resistance-associated mutations can impact fitness negatively in other organisms (Andersson and Hughes 2010).

To understand the extent to which C580Y is acting as a driver in the epidemiology of KEL1, I competed the C580 or Y580 alleles in the CAM background, which originally possessed a K13 Y580 haplotype. Strikingly, the results of the competition assay on these isogenic clones showed a distinct fitness cost when the allele was mutated to C580 (Fig. 5.10). The first experiment, however, did not reveal a strong phenotype because of a contamination with 3D7 that only became detectable after 20 days of competition, resulting in a drop in total counts for the CAM lines and potentially suppressing the growth difference, which in this experiment was not-significant (p-value ns). As described earlier, 3D7 has a higher growth rate than most strains used throughout this work, as seen in chapter 4 with the mixture that contained 3D7, V1/S and CAM. The second competition experiment, however, showed a striking difference between the two isogenic clones carrying the K13 alleles: after 35 days of competition, CAM[rev] was completely outcompeted, and a t-test between the two clones revealed a significant difference in the change in their relative proportion over time (p-value = 1.26x10-9). This result provides

the first direct experimental evidence that the establishment of the K13 Y580 allele on the KEL1 lineage has occurred on a supportive genetic background that is facilitating the outbreak and spread of this allele in Southeast Asia. I consider that these results indicate that a positive epistasis might be occurring in this genetic background in which multiple potentially deleterious or neutral mutations act in combination to enhancing fitness. Future work to understand the nature of the growth difference will include other phenotypic measurements such as merozoite numbers (Bunditvorapoom *et al.* 2018), invasion of host RBCs, and fitness assays on different clones to evaluate their phenotypes in competition with the wild type clone. We also plan to expand the analysis of the whole genome sequences and try to identify whether potential off-targets of Cas9 might have introduced other variants that might be impacting fitness. Although whole genome sequencing was performed, variants were called using the reference genome 3D7, meaning that it was not possible to detect variants that were different across CAM clones.

By combining fitness and drug susceptibility assays throughout this dissertation, I have been able to characterize the impact of an important allele for artemisinin susceptibility in Southeast Asia, by parallel phenotyping and Next Generation Sequencing. In addition to this, in the current chapter I have been able to establish the importance of specific genetic backgrounds in modulating fitness, and hypothesized that a positive epistasis in a KEL1 lineage might be the reason for the spread of this allele in Southeast Asia. To try to strengthen the validity of this finding, I designed a novel experimental approach by which parasites from different genetic backgrounds were provided with a pool of plasmids containing any possible codon for replacement of 580 in K13. These results support the findings from the competition assays on isogenic clones, as the only strain for which the Y580 codon was obtained after transfection was CAM, implying a possible high fitness cost of this allele in non-KEL1 backgrounds. In addition, all strains tested showed a combination of codons comprising a small number of amino acids, for which the most common alternative was serine. Notably, serine is the only other amino acid apart from cysteine and tyrosine at position 580 observed across sequenced *Plasmodium* species (Fig. 5.17).

Amongst the next steps towards validating the relevance of the KEL1 genetic background for spread of artemisinin-resistant alleles, I have now generated clones for the different codons obtained and will perform RSA$_{(0-3h)}$ and fitness assays to test their impact on these phenotypes. In addition, I will transfect the pool of all possible codons into multiple parasite backgrounds belonging to the recent outbreak of parasites resistant to both DHA and piperaquine (a lineage described as KEL1/PLA1 by Amato *et al.*, 2018), as well as recent Cambodian non-KEL1/PLA1 lineages in order to validate the findings. Based on the modelling data from Singh *et al.* I will mutate residue C532 of K13, which based on the predicted structure of the propeller domain may form a disulfide bridge with C580. This disulfide bond would be disrupted by a C580 to Y580 mutation, and would lead to free SH group from the C532 residue, which could potentially act as a sensor of artemisinin resistance. If mutation of C532 in a artemisinin-sensitive background results in increased tolerance, or if in an artemisinin-resistant background C532 mutation reverts the resistant phenotype to sensitive, it would contribute to a mechanistic understanding of the impact of the C580Y polymorphism. Regardless of outcome, performing fitness assays at a large scale in the strains mentioned will be informative in dissecting the relevance of the genetic background for surveillance and control of the spread of these artemisinin resistant alleles in Southeast Asia.

In Southeast Asia, and particularly in Cambodia, the amount of time that some drug-resistance alleles remain within a population has been shown to be longer than in Africa. An example are the mutations in *Pfcrt* that still remain at high frequencies in Southeast Asia, particularly Thailand and Cambodia, where no chloroquine has been used for the treatment of *P. falciparum* malaria since the early 1990s (Lim *et al.* 2003; Mwai *et al.* 2009). On the other end, the example of chloroquine-resistant alleles in Malawi reverting back to the reference sequence, which led to sensitivity to this compound being regained, show local epidemiology and population structure can also impact maintenance of alleles with high fitness costs (Kublin *et al.* 2003). These are examples of the complex nature of the maintenance and spread of alleles associated with drug resistance. For the case of artemisinin, there is much less evidence of a population-specific effect on resistance and fitness, as K13-mediated artemisinin resistance has been successfully maintained in some parts in Southeast Asia (Woodrow and White 2016).

My results thus far and my ongoing experimental work will provide valuable evidence for the field on the importance of identifying genetic backgrounds tolerating resistance-associated mutations, and how population structure and genetic diversity might be facilitating spread of these alleles. It is also important to highlight that the work presented focuses on a genetic background which has dominated as part of the current outbreak of artemisinin resistance in Southeast Asia. This understanding is therefore relevant to malaria intervention strategies in highly endemic regions, as an active surveillance for other mechanisms of resistance or changes in the fitness landscape in African parasites will be important for containing the spread of resistance- or fitness-associated alleles. Even though genomics has made an enormous contribution to the understanding of drug resistant populations that have been under strong selective pressures from antimalarial drugs (Mackinnon and Marsh 2010), there is a need for experimental validation of these findings. Understanding pathways leading to an increased fitness in drug-resistant parasites or compensatory mutations, as we have seen on a KEL1 parasite lineage, will be essential, and genome-based editing approaches are reaching a point at which large-scale population genetics studies having become affordable and easier to validate these findings. Finally, the validity of these studies at this moment only applies to the in vitro experimental work presented. It will be important to translate these questions to other elements of parasite fitness in the natural environment. For example, understanding whether these variants impact transmission or if the fitness effects are restricted only to the blood stages. Therefore, expanding these studies to cover more field isolates and to be able to test them for fertility and transmission capabilities are all important steps that we hope this work will eventually lead to.

Chapter 6: Conclusions and future work

Over the past decade there has been a significant increase in the attention to and money spent on global malaria control, with local elimination and country-wide malaria eradication targeted in many areas. There is no question that these efforts have had an enormous effect on reducing mortality, due to more efficient vector control, improvement in diagnostics, and effective chemotherapies (Hemingway *et al.* 2016). However, each of these pillars supporting recent progress are under threat. Some examples are the spread of insecticide resistance (Anopheles gambiae 1000 Genomes Consortium *et al.* 2017) , the failure of rapid diagnostic tests coming from deletions in HRP2/3 (H. Gupta *et al.* 2017), and finally, the emergence and spread of antimalarial drug resistance (Amato *et al.* 2018). Parasite populations in Southeast Asia appear to be the epicentre of malaria drug resistance (Blasco *et al.* 2017) across multiple drug classes. Resistance to quinolines has been an ongoing problem, and there has been independent emergence of resistance in various places (Payne 1987). The example of chloroquine resistance is particularly notable, which reached such high levels that it had to be removed as treatment for *P. falciparum* malaria (Fidock *et al.* 2000). Other examples from the quinoline class where resistance has emerged to monotherapy are still being used in combination with the current frontline antimalarial artemisinin. Some of the combinations with quinolines which are currently in use include mefloquine-artesunate, lumefantrine-arthemeter, and piperaquine-dihydroartemisinin (Alker *et al.* 2007; Taylor and Juliano 2014), and resistance threatens all of these partner drugs. Resistance is now clearly an issue for artemisinin itself. Artemisinin is both a fast-acting drug, primarily on early ring-stages circulating in the blood, and a short half-life drug (Dondorp *et al.* 2010). This particular mode of action made it particularly important to anticipate resistance emerging from a monotherapy, which was studied by Noedl *et al.*, and later confirmed through various studies of populations of parasite isolates (Noedl *et al.* 2008; Miotto *et al.* 2015). Given this, partner drugs were deliberately chosen to act on more mature erythrocytic stages and to have a longer half-life, to act as cover for any parasites that remain after the short-lived artemisinin pulse has been eliminated from the body (Eastman and

Fidock 2009). Despite these precautions, treatment failure to ACTs have now been defined by different assays and are compound-specific: an *in vivo* prolonged parasite clearance for artemisinin treatment failure (Noedl *et al.* 2008), *in vitro* phenotype measured by RSA (Benoit Witkowski *et al.* 2013), and evidence of recrudescence upon treatment of parasites with a partner drug, piperaquine (Duru *et al.* 2015).

Identifying genetic markers associated with these phenotypes has been accelerated because access to whole genome sequencing has become more affordable and ultimately more ubiquitous, leading to thousands of parasite isolates being sequenced yearly through collective efforts. The expansion of *P. falciparum* genomics since deep sequencing on clinical isolates was first reported (Manske *et al.* 2012) is striking. These approaches have allowed for a careful surveillance of the genetics of *P. falciparum* isolates pre- and post-intervention policies. A study by Miotto *et al.* is an example in which genomics were used to dissect the epidemiology of artemisinin resistance. The authors used a Genome-Wide Association Study (GWAS) to help identify markers of a correlation of genotype-to-phenotype in populations with high levels of resistance, and identified K13 mutations that had reached high levels of fixation. Not only did these studies confirm the pioneering work of Ariey *et al.* identifying K13 as a molecular marker of artemisinin resistance, but also identified other markers linked to the resistance phenotype (section 1.4.5.3) (Ariey *et al.* 2014). Mutations in ferredoxin, PfCRT (chloroquine resistance transporter), ARPS10 (apicoplast ribosomal protein S10) and MDR2 (multidrug resistance protein 2) were all variants linked to a high-resistance phenotype in at least one of the sample groups selected. A reasonable explanation as to why these mutations are at high levels of fixation (F_{ST}) among these drug-resistant populations is that they might be compensating for a fitness loss, thus acting as a "genetic background" that can tolerate resistance-associated polymorphisms such as mutations in K13 (Miotto *et al.* 2015). Testing this hypothesis requires experimental studies to attribute function to mutations on resistant and sensitive parasite lineages from highly endemic countries, as well as understanding the process by which these mutations were acquired over time. Such experiments would allow us to tackle questions such as 1. did the genetic background have to be there for K13-C580Y to be successfully maintained, and 2. are those mutations epistatic?

We are currently at a time when experimental genetics can be performed at scale in many organisms (Wang *et al.* 2014). One of the obvious benefits of increasing the throughput of experimental approaches to *P. falciparum* will be to improve surveillance of this deadly disease, by functionally testing the impact of genomic variants that emerge in the field. Because of the urgent need to develop tools that can keep up with results from genomic studies, the first chapter of this dissertation was focused on understand transfection efficiency and dynamics by using a barcode sequencing approach in this species for the first time. The idea of generating dozens of mutants within a single transfection, as had been performed successfully in both *P. berghei* and *T. gondii*. (Sidik *et al.* 2016; Bushell *et al.* 2017), was the main objective of these experiments. Performing transfections with a pool of plasmids containing unique DNA barcodes revealed that while multiple different plasmids were successfully maintained episomally, by analysing the sequencing reads at the bulk and clonal levels it became clear that individual parasites were taking up multiple plasmids. These results are valuable for the field because of their implications for performing large-scale reverse genetic screens, for example, to target candidate genes involved in drug resistance mechanisms. Also, they could be valuable to understand the compensatory mechanisms taking place under certain selective pressures, or to validate GWAS hits and study the effects of particular alleles on a phenotype in a combinatorial manner, and measure their epistatic interactions and the phenotypic outcome.

In the same chapter, I was able to establish that the pseudogene *Pfrh3* could be used as a safe harbour for barcode insertion, which led to the generation of barcoded parasite strains, each with different sensitivities to antimalarials. By pooling these barcoded lines I was able to test different antimalarial drugs and use barcode sequencing (BarSeq) to measure growth phenotypes by analysing sequencing reads. While resistance phenotypes could be generated, the results also showed that strains varied in their relative growth rates when put in competition even in the absence of any drug, and drug resistant strains were rapidly outcompeted. The implications this might have on the epidemiology of drug resistance challenge the perceptions on what the best strategies for containment are, and suggest that the high fitness costs of K13-mediated artemisinin resistance might prevent these alleles from

spreading to geographical regions where transmission is high and competition from genetically diverse strains is also high.

To tackle this question, future work might entail using BarSeq technology systematically on parasite clones carrying different combinations of fitness-conferring mutations, or compensatory mutations such as those from the Miotto *et al.* study. Such an approach would allow us to understand the fitness landscapes of these highly-resistant parasites from Southeast Asia, and compare them with parasite isolates from epidemiologically distant regions. This would allow us to identify alleles, or more accurately combinations of alleles, that should be under careful monitoring. Overall, this chapter sets the grounds for performing parallel phenotyping at scale by using NGS in some of the lines selected to study the effects of particular genetic backgrounds on fitness in artemisinin-resistant parasites. These pilot studies led to the experimental design to address the main objective for the last results chapter.

Without a clear understanding of the effect of the "genetic background", the extent to which pathways leading to resistance will be the same in countries where parasite populations and local epidemiology vary greatly cannot be easily predicted. In the approach taken by Ariey *et al.*, a Tanzanian isolate (F32-ART5) was selected for resistance for over 120 generations with artemisinin. Strikingly, in this experimental design, other than K13 none of the genes from the Miotto *et al.* GWAS were acquired during selection (Ariey *et al.* 2014). How these long-term selections such as the one by Ariey *et al.* can be extrapolated to real-time evolution of parasites in the field remains unknown. One implication is that effectively, not all parasites are "made" in the same way, and that even though local emergence of resistance is possible in the different geographical regions (Alifrangis *et al.* 2014; Sá *et al.* 2009; Amato *et al.* 2016), certain strains are more efficient in spreading these variants, likely due to intrinsic differences in fitness. This study therefore highlights how the use of genomics at the population level can support other experimental approaches like the one used by Ariey *et al.* This also highlights the fact that more focus should be given to finding ways to validate the functional impact of natural variation under strong selective pressures, and scaling up experimental genetics is an obvious next step. Additional examples in which genomic studies were linked to drug resistance phenotypes came from two recent studies by Amato *et al.* 2017 and Witkowski *et al.* 2017(Amato *et al.* 2017;

Benoit Witkowski *et al.* 2017). In both, a surrogate marker of piperaquine resistance was identified: amplifications in two genes, plasmepsin 2 and 3 were present in genomic samples from recrudescent Cambodian isolates after treatment with DHA-PPQ. In the study by Amato *et al.*, two other genes showed an association with recrudescence: mutations in an exonuclease, and in an MCP (mitochondrial carrier protein), both of which are being studied in the Lee group at the Sanger Institute. The signal that was more strongly linked with the recrudescence phenotype, however, was the copy number variation in plasmepsins 2 and 3. Later studies showed that the recent outbreak of DHA/PPQ-resistant parasites arose on a K13-C580Y background (KEL1/PLA1)(Amato *et al.* 2018), emphasising that while local emergence of resistance in Southeast Asia occurs frequently, only some lineages are able to take over other local populations, suggesting special attention should be paid to particular combinations of alleles.

Several recent studies have begun to investigate these markers functionally. To date, amplification of plasmepsins 2 and 3 have not shown an appreciable effect on resistance, whereas emerging *Pfcrt* mutations in Cambodian parasites demonstrate a more robust drug sensitivity shift and recrudescence phenotype. Notably, Ross *et al.*, suggest that the amplification of the plasmepsin genes was instead a compensatory mechanism to alleviate the fitness cost in these parasites with novel *pfcrt* haplotypes (Ross *et al.* 2018). This example shows that genomic studies can help in surveillance efforts, however the extent to which they impact the parasite at the molecular level needs to be evaluated carefully with experimental approaches such as the ones developed through this dissertation, to inform any implications for malaria control.

Given the evidence that background has played a key role in the spread of the KEL1/PLA1 lineage, the aim of the last chapter was to tackle the questions of the impact of genetic background on the *in vitro* fitness of *P. falciparum* parasites (chapter 5), by using a strain belonging to the KEL1 background. Several different approaches could be taken to investigate this - some are envisioned as future work, while others were developed experimentally in this chapter. These include: 1) identify parasite isolates which contain the "genetic background" (KEL1 vs. non-KEL1), and study their fitness *in vitro* by performing assays

as described (section 5.3.2), 2) use CRISPR/Cas9-based editing to introduce combinations of these mutations into a standard laboratory strain like Dd2. This approach might be ideal because it uses a constant genetic background, but it has the limitation of given it critically relying on the ability to eliminate any remaining episomes in order to perform sequential transfections. The limited number of selectable markers could also make this appraoch challenging (Ganesan *et al.* 2011). 3) Transfecting multiple genetic backgrounds (section 5.3.3), and replacing codon 580 of K13 in all of them would shed light on which genetic backgrounds can tolerate this change that can lead to an increased survival to artemisinin, and would help prioritise alleles for functional analysis.

To start addressing these questions, CRISPR/Cas9-based gene editing was used to revert the C580Y mutation in K13 in a Cambodian KEL1 parasite strain referred to as CAM. In order to have a comparative study of the effect a particular allele has on fitness, the same mutation was introduced into the standard, laboratory strain 3D7. Unsurprisingly, and in concordance with recent evidence on the fitness costs of C580Y in non-KEL1 backgrounds (Straimer *et al.* 2017; Nair *et al.* 2018), this change had a moderate fitness cost in a 3D7 background (Fig 5.9). By contrast, by reverting the Y580 back to C580 (reference allele) in CAM, parasites became less fit (Fig. 5.10). This evidence supports the hypothesis of a positive epistasis in a KEL1 lineage, which resulted in an enhanced fitness that may have allowed these parasites to spread rapidly and take over other, less fit populations. Finally, I presented a radically different approach in which multiple genetic backgrounds were edited with a plasmid pool containing all possible codons to replace position 580 in K13, provided as a recodonised donor sequence. Supporting the finding from the competition assays using edited 3D7 and CAM lines, CAM was the only background that preferentially tolerated Y580, whereas all other backgrounds either emerged with the reference allele (Cys), or to potentially structurally tolerable changes (e.g. serine, Fig 5.14). Future work entailing phenotypic assays on all the viable allele changes will include assays to evaluate whether disrupting the predicted disulfide bridge between C580 and C532 has a background-specific phenotypic outcome by measuring both RSA$_{(0-3h)}$ and fitness via amplicon sequencing. In addition to this, *Pfkelch13*-editing transfections will be performed in other Cambodian backgrounds belonging to: 1) The recent KEL1/PLA1 lineage described by Amato *et*

al., to understand the impact of C580Y. Following the recent findings by Ross *et al.* suggesting amplifications in plasmepsins 2 and 3 are compensating for fitness losses from *pfcrt* mutations linked with piperaquine resistance, representative piperaquine-resistant strains will also be evaluated. 2) non-KEL1 backgrounds as a proof-of-concept to confirm the current preliminary findings in other genetic backgrounds (V1/S, Dd2, 3D7 and 7G8 in Fig. 5.14) that C580Y has a fitness cost except on the KEL1 background. The outcome of this work, therefore, will be valuable for the field as it will provide a broader picture of the fitness landscapes of K13-mediated artemisinin resistance in Southeast Asia.

Some of the aspects of this work that will be expanded in the future are the lack of replication for some of the experiments performed. Particularly, some of the competition assays do not have enough biological replicates, therefore it is not possible to determine, despite a clear general trend, what the relative fitness of the competing strains are whilst providing robust statistical power to the differences observed. In general, the analyses performed were using the data points available for each experiment, by comparing the means of the difference in growth rate over time. However, the limitations described do not allow for significance levels to be determined accurately. This particular issue is the major weakness of this work and will be at the epicenter for future experimental work in order to conclude it. In addition to this, there are multiple factors of the experimental design that require careful analysis, for example how to deal with biological noise, as observed in some of the competition assays presented. For this reason, novel approaches will be employed in an attempt to resolve the noise coming from the first days of competition. Performing a tighter synchronization, together with an accurate adjustment of parasitemia and hematocrit will be essential in order to resolve biological or technical noise at early days and to have a more accurate proportion of 1:1 in head-to-head competition assays. Finally, the limited number of parasite strains that were used for performing transfections do not allow for establishing what in the genetic backgrounds described is compensating for a loss of fitness, therefore replicating these experiments in strains belonging to the recently defined KEL1 lineage, and other representative examples of a non-artemisinin resistant parasite strain from Cambodia.

The idea of identifying hotspots of malaria transmission within geographical locations known to be at the epicentre of multidrug resistance is essential as, in theory, targeted mass drug administration (MDA) could be performed in order to eliminate the reservoir of parasites, thus interrupting transmission. MDA campaigns have been taking place in one form or another since 1932, starting in a rubber plantation in Liberia when plasmoquine was implemented (Rice 1932), to some of the largest MDA campaigns ever, undertaken in China, starting in the 1960s and spanning until today, with almost no cases of malaria now reported in this country (Tang 2000). This concept, however, has been challenged by the idea that MDA can lead to the spread of antimalarial drug resistance. Evidence of indirect MDA leading to this terrible outcome were followed by the use of pyrimethamine and chloroquine salts, which led to local replacement of sensitive alleles with resistant ones in both Southeast Asia and South America (von Seidlein and Greenwood 2003). A direct link between MDA and the spread of resistance is hard to prove however. It can also be argued that for the case of South America, a discontinuity in the treatment for political and socio-economic reasons also played a role in the emergence of chloroquine resistance. Similarly, in Southeast Asia, resistance emerged in some Cambodian provinces possibly also as a consequence of the political situation and the migration of non-immune workers to the local mines (von Seidlein and Greenwood 2003).

The fact that chloroquine resistance emerged independently in Southeast Asia and in South America, both showing a high linkage disequilibrium around *Pfcrt*, provided evidence that the emergence of resistance can occur and become established more efficiently in such regions before spreading to areas of high transmission such as Sub-Saharan Africa (Mita *et al.* 2009; Wootton *et al.* 2002). Because these resistance-associated variants have to take over the sensitive populations, genetic diversity will ultimately play a key role their ability to spread. In Southeast Asia, for example, diversity is lower than in Africa, primarily maintained by low transmission rates. This leads to recombination events being less frequent and accompanied by lower prevalence of polyclonal infections. Given this, if a resistant parasite arises in a high-transmission setting, where recombination is high, it is more likely that this parasite will be selected against, by being out-competed by "fitter" parasites, especially in a context in which acquired immunity is also higher. This is the main argument suggesting that while resistant

parasites can occur at low frequencies in Africa, they are less likely to establish. A modelling approach taken in a recent study by Bushman *et al.* supported the differences between low vs high transmission settings for the establishment and spread of drug resistance. Interestingly, the authors propose that consistent with all the historical evidence on the emergence and spread of antimalarial drug resistance (e.g. chloroquine or pyrimethamine), low-transmission settings usually promote maintenance of drug-resistant genotypes, however, they found based on the modeling performed that high-transmission settings would actually facilitate spread (Bushman *et al.* 2018). If this is the case, one of the scenarios of the recent outbreak of multidrug resistant parasites in Southeast Asia (KEL1/PLA1), with enhanced fitness, could be an example of a successfully maintained parasite lineage.

A key consideration from for the field is how parasite populations in Africa will manifest resistance to the frontline antimalarial artemisinin. Even though reasonable attention should be paid to the markers that have been identified in the various genomic studies in Southeast Asia, such as those discussed here, it will be important to also focus efforts on studying carefully how parasite populations are changing in the different low-transmission settings in Africa. Given the importance of low transmission for drug resistance, it is possible that there are currently some evolutionary bottlenecks in certain populations that we might be unaware of, and the emergence of a resistance-associated polymorphism, possibly non-K13 mediated, might expand in frequency given the decreased genetic diversity of these areas. The lessons learned from this thesis, about how genomic sequencing and experimental genetics can be brought together, must be applied to such location.

7. Appendix

7.1 List of primers used

p36: GTTGTGTGGAATTGTGAGCGG
p171: CATGGACCCACTCCAGACGAACCTACCGACTTCATG
p172: CTAGCATGAAGTCGGTAGGTTCGTCTGGAGTGGGTC
p173: CATGGCAAGAACCTCATCCTACGCTATCCCGACCAG
p174: CTAGCTGGTCGGGATAGCGTAGGATGAGGTTCTTGC
p191: GGGTAAACTAGAATATGCTATACCGG
p194: ACCATCACGGGATAAAGTAACTGG
p195: CACCATTGAAGAAGAATCTCCTAAG
p196: CACTTGTTCCTTTCTCATTTGAATC
p203: AGAACATATTTATTAAATCTACAACTTGTAAAAGCAGAAGAAG
p204: CGCTCATAGCCCATGGCCTCTATAACCTTCTTAGGAG
p205: TTATAGAGGCCATGGGCTATGAGCGCTAGCGGTATACATCATTAATTGGAAAGG
p206: CGCATCAGGCGCCAGCTGGTACATGTTTGAATACTAAC
p212: CAATTAATGATGTATACCGCCTTCAATTTCGATGGGTAC
p215: ATTGTGGTAATACAGAAATGGATG
p216: AAACCATTTATTGCTTCCGCATC
p217: ATTGGATGCGGAAGCAATAAATG
p218: AAACCATCCATTTCTGTATTACCA
p219: CTAAGAAGGTTATAGAGGCGTAATTCGTGCGCGTCAG
p239: TATCTCTATAGTCGCAGTAGGCG
p267: AATACACCAAGATCATCAGCTATGTGTGTTGCTTTTGATAATAAAATTTATG
p317: CGCATCAGGCGCCAGCCTAGGAACGGAGTGACCAAATCTGGG
p265: AGAACATATTTATTAAATCTTCCGTTAACTATACCCATACC
p318: TACACCAAGATCATCAGCTATGTATGTTGCTTTTGATAATAAAATTTATGTC
p439: GAGGTACCGAGCTCGAATTCACTCAAGCTTGGGGGGATCC
p440: AGTGCCACCTGACGTCTTAGCTAAGCATGCGGGCCC
p325: CGTATGATAGGGAATCTGGTGG
p328: GCTGCCATTCATTTGTATCTGG
p327: CTTCGCCATTTTCTCCTCCTG
p326: CCGAATGTAGAAGCATATGATC
p1492: ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGAATGTAGAAGCATATGATC
p1493: CTAATGGTGAGAGATTAAATTAGATCGGAAGAGCGGTTCAGCAGGAATGCCGA

91_Illumina (Arg445): TCGGCATTCCTGCTGAACCGCTCTTCCGATCTGTAATTCGTGCGCGTCAG
97_Illumina (Arg444): ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTTCAATTTCGATGGGTAC
PE1.0_Illumina: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T

8. References

Abu Bakar, N. *et al.*, 2010. Digestive-vacuole genesis and endocytic processes in the early intraerythrocytic stages of Plasmodium falciparum. *Journal of cell science*, 123(Pt 3), pp.441–450.

Achan, J. *et al.*, 2011. Quinine, an old anti-malarial drug in a modern world: role in the treatment of malaria. *Malaria journal*, 10, p.144.

Ahmadian, A., Ehn, M. & Hober, S., 2006. Pyrosequencing: history, biochemistry and future. *Clinica chimica acta; international journal of clinical chemistry*, 363(1-2), pp.83–94.

Aikawa, M. *et al.*, 1966. The feeding mechanism of avian malarial parasites. *The Journal of cell biology*, 28(2), pp.355–373.

Alifrangis, M. *et al.*, 2014. Independent origin of plasmodium falciparum antifolate super-resistance, Uganda, Tanzania, and Ethiopia. *Emerging infectious diseases*, 20(8), pp.1280–1286.

Alker, A.P. *et al.*, 2007. Pfmdr1 and in vivo resistance to artesunate-mefloquine in falciparum malaria on the cambodian–thai border. *The American journal of tropical medicine and hygiene*, 76(4), pp.641–647.

Aly, A.S.I., Vaughan, A.M. & Kappe, S.H.I., 2009. Malaria parasite development in the mosquito and infection of the mammalian host. *Annual review of microbiology*, 63, pp.195–221.

Amaratunga, C. *et al.*, 2016. Dihydroartemisinin-piperaquine resistance in Plasmodium falciparum malaria in Cambodia: a multisite prospective cohort study. *The Lancet infectious diseases*, 16(3):357-65

Amaratunga, C., Neal, A.T. & Fairhurst, R.M., 2014. Flow cytometry-based analysis of artemisinin-resistant Plasmodium falciparum in the ring-stage survival assay. *Antimicrobial agents and chemotherapy*, 58(8), pp.4938–4940.

Amato, R. *et al.*, 2017. Genetic markers associated with dihydroartemisinin-piperaquine failure in Plasmodium falciparum malaria in Cambodia: a genotype-phenotype association study. *The Lancet infectious diseases*, 17(2), pp.164–173.

Amato, R. *et al.*, 2016. Genomic epidemiology of artemisinin resistant malaria. *eLife*, 5, p.e08714.

Amato, R. *et al.*, 2018. Origins of the current outbreak of multidrug-resistant malaria in southeast Asia: a retrospective genetic study. *The Lancet infectious diseases*, 18(3), pp.337–345.

Andersson, D.I. & Hughes, D., 2010. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature reviews. Microbiology*, 8(4), pp.260–271.

Anopheles gambiae 1000 Genomes Consortium *et al.*, 2017. Genetic diversity of the African malaria vector Anopheles gambiae. *Nature*, 552(7683), pp.96–100.

Antoine, T. *et al.*, 2014. Rapid kill of malaria parasites by artemisinin and semi-synthetic endoperoxides involves ROS-dependent depolarization of the membrane potential. *The Journal of antimicrobial chemotherapy*, 69(4), pp.1005–1016.

Ariey, F. *et al.*, 2014. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature*, 505(7481):50-5.

Ashley, E.A. *et al.*, 2014. Spread of artemisinin resistance in Plasmodium falciparum malaria. *The New England journal of medicine*, 371(5), pp.411–423.

Baird, J.K., 2013. Evidence and implications of mortality associated with acute Plasmodium vivax malaria. *Clinical microbiology reviews*, 26(1), pp.36–57.

Bannister, L. & Mitchell, G., 2003. The ins, outs and roundabouts of malaria. *Trends in parasitology*, 19(5), pp.209–213.

Baragaña, B. *et al.*, 2015. A novel multiple-stage antimalarial agent that inhibits protein synthesis. *Nature*, 522(7556), pp.315–320.

Bentley, D.R. *et al.*, 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), pp.53–59.

Billker, O. *et al.*, 2004. Calcium and a calcium-dependent protein kinase regulate gamete formation and mosquito transmission in a malaria parasite. *Cell*, 117(4), pp.503–514.

Birnbaum, J. *et al.*, 2017. A genetic system to study Plasmodium falciparum protein function. *Nature methods*, 14, p.450.

Blackman, M.J. *et al.*, 1990. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion-inhibiting antibodies. *The Journal of experimental medicine*, 172(1), pp.379–382.

Blasco, B., Leroy, D. & Fidock, D.A., 2017. Antimalarial drug resistance: linking Plasmodium falciparum parasite biology to the clinic. *Nature medicine*, 23(8), pp.917–928.

Bouchaud, O. *et al.*, 2009. Fatal cardiotoxicity related to halofantrine: a review based on a worldwide safety data base. *Malaria journal*, 8, p.289.

Bousema, T. *et al.*, 2014. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nature reviews. Microbiology*, 12(12), pp.833–840.

Bozdech, Z. *et al.*, 2003. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS biology*, 1(1), p.E5.

Bunditvorapoom, D. *et al.*, 2018. Fitness Loss under Amino Acid Starvation in Artemisinin-Resistant Plasmodium falciparum Isolates from Cambodia. *Scientific reports*, 8(1), p.12622.

Bushell, E. *et al.*, 2017. Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes. *Cell*, 170(2), pp.260–272.e8.

Bushman, M. *et al.*, 2018. Within-host competition can delay evolution of drug resistance in malaria. *PLoS biology*, 16(8), p.e2005712.

Calhoun, S.F. *et al.*, 2017. Chromosome End Repair and Genome Stability in Plasmodium falciparum. *mBio*, 8(4). pii: e00547-17.

Carter, L.M. *et al.*, 2013. Stress and sex in malaria parasites: Why does commitment vary? *Evolution, medicine, and public health*, 2013(1), pp.135–147.

Chugh, M. *et al.*, 2013. Protein complex directs hemoglobin-to-hemozoin formation in Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America*, 110(14), pp.5392–5397.

Claessens, A. *et al.*, 2014. Generation of antigenic diversity in Plasmodium falciparum by structured rearrangement of Var genes during mitosis. *PLoS genetics*, 10(12), p.e1004812.

Collins, J.A. *et al.*, 2004. Competitive fitness of nevirapine-resistant human immunodeficiency virus type 1 mutants. *Journal of virology*, 78(2), pp.603–611.

Cowell, A.N. *et al.*, 2018. Mapping the malaria parasite druggable genome by using in vitro evolution and chemogenomics. *Science*, 359(6372), pp.191–199.

Crabb, B.S. *et al.*, 1997. Stable transgene expression in Plasmodium falciparum. *Molecular and biochemical parasitology*, 90(1):131-44.

Cong, L. & Zhang, F., 2015. Genome Engineering Using CRISPR-Cas9 System. In S. M. Pruett-Miller, ed. *Chromosomal Mutagenesis*. Methods in Molecular Biology. Springer New York, pp. 197–217.

Cruaud, P. *et al.*, 2017. High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Scientific reports*, 7, p.41948.

Dalrymple, U., Mappin, B. & Gething, P.W., 2015. Malaria mapping: understanding the global endemicity of falciparum and vivax malaria. *BMC medicine*, 13, p.140.

Davis, L. & Maizels, N., 2014. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proceedings of the National Academy of Sciences*, 111(10), pp.E924–E932.

Deitsch, K., Driskill, C. & Wellems, T., 2001. Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. *Nucleic acids research*, 29(3): 850–853.

Delves, M. *et al.*, 2012. The activities of current antimalarial drugs on the life cycle stages of Plasmodium: a comparative study with human and rodent parasites. *PLoS medicine*, 9(2), p.e1001169.

Denis, M.B. *et al.*, 2006. Surveillance of the efficacy of artesunate and mefloquine combination for the treatment of uncomplicated falciparum malaria in Cambodia. *Tropical medicine & international health: TM & IH*, 11(9), pp.1360–1366.

De Silva, E.K. *et al.*, 2008. Specific DNA-binding by apicomplexan AP2 transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 105(24), pp.8393–8398.

Dhingra, S.K. *et al.*, 2017. A Variant PfCRT Isoform Can Contribute to Plasmodium falciparum Resistance to the First-Line Partner Drug Piperaquine. *mBio*, 8(3): e00303-17.

Didier Ménard *et al.*, 2008. Dihydrofolate Reductase I164L Mutation in Plasmodium falciparum, Madagascar. *Emerging Infectious Disease journal*, 14(7), p.1166.

Dinant, C., Houtsmuller, A.B. & Vermeulen, W., 2008. Chromatin structure and DNA damage repair. *Epigenetics & chromatin*, 1(1), p.9.

Dluzewski, A.R. *et al.*, 2008. Formation of the food vacuole in Plasmodium falciparum: a potential role for the 19 kDa fragment of merozoite surface protein 1 (MSP1(19)). *PLoS one*, 3(8), p.e3085.

Doench, J.G. *et al.*, 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature biotechnology*, 32(12), pp.1262–1267.

Dogovski, C. *et al.*, 2015. Targeting the cell stress response of Plasmodium falciparum to overcome artemisinin resistance. *PLoS biology*, 13(4):e1002132.

Dondorp, A.M. *et al.*, 2010. Artemisinin resistance: current status and scenarios for containment. *Nature reviews. Microbiology*, 8(4), pp.272–280.

Duraisingh, M.T. *et al.*, 2000. Increased sensitivity to the antimalarials mefloquine and artemisinin is conferred by mutations in the pfmdr1 gene of Plasmodium falciparum. *Molecular microbiology*, 36(4), pp.955–961.

Duraisingh, M.T., Triglia, T. & Cowman, A.F., 2002. Negative selection of Plasmodium falciparum reveals targeted gene deletion by double crossover recombination. *International journal for parasitology*, 32(1), pp.81–89.

Duru, V. *et al.*, 2015. Plasmodium falciparum dihydroartemisinin-piperaquine failures in Cambodia are associated with mutant K13 parasites presenting high survival rates in novel piperaquine in vitro assays: retrospective and prospective investigations. *BMC medicine*, 13:305.

Eastman, R.T. *et al.*, 2011. Piperaquine resistance is associated with a copy number variation on chromosome 5 in drug-pressured Plasmodium falciparum parasites. *Antimicrobial agents and chemotherapy*, 55(8), pp.3908–3916.

Eastman, R.T. & Fidock, D.A., 2009. Artemisinin-based combination therapies: a vital tool in efforts to eliminate malaria. *Nature reviews. Microbiology*, 7(12), pp.864–874.

Eckstein-Ludwig, U. *et al.*, 2003. Artemisinins target the SERCA of Plasmodium falciparum. *Nature*, 424(6951), pp.957–961.

Liu, W. *et al.*, 2010. Origin of the human malaria parasite Plasmodium falciparum in gorillas. *Nature*, 467(7314), pp.420–425.

Erlich, Y. *et al.*, 2008. Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nature methods*, 5(8), pp.679–682.

Escalante, A.A. & Ayala, F.J., 1994. Phylogeny of the malarial genus Plasmodium, derived from rRNA gene sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 91(24), pp.11373–11377.

Eziefula, A.C., 2016. Artesunate-mefloquine: a malaria treatment for African children? *The Lancet infectious diseases*, 16(10), pp.1086–1087.

Faust, C. & Dobson, A.P., 2015. Primate malarias: Diversity, distribution and insights for zoonotic Plasmodium. *One health (Amsterdam, Netherlands)*, 1, pp.66–75.

Fidock, D.A. *et al.*, 2000. Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Molecular cell*, 6(4), pp.861–871.

Flannery, E.L., Chatterjee, A.K. & Winzeler, E.A., 2013. Antimalarial drug discovery [mdash] approaches and progress towards new medicines. *Nature reviews. Microbiology*, 11(12), pp.849–862.

Gabryszewski, S.J. *et al.*, 2016. Combinatorial Genetic Modeling of pfcrt-Mediated Drug Resistance Evolution in Plasmodium falciparum. *Molecular biology and evolution*, 33(6), pp.1554–1570.

Ganesan, S.M. *et al.*, 2011. Yeast dihydroorotate dehydrogenase as a new selectable marker for Plasmodium falciparum transfection. *Molecular and biochemical parasitology*, 177(1), pp.29–34.

Gardner, M.J. *et al.*, 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, 419(6906), pp.498–511.

Ghorbal, M. *et al.*, 2014. Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nature biotechnology*, 32(8), pp.819–821.

Giaever, G. & Nislow, C., 2014. The yeast deletion collection: a decade of functional genomics. *Genetics*, 197(2), pp.451–465.

Gnügge, R. & Symington, L.S., 2017. Keeping it real: MRX-Sae2 clipping of natural substrates. *Genes & development*, 31(23-24), pp.2311–2312.

Godiska, R. *et al.*, 2010. Linear plasmid vector for cloning of repetitive or unstable sequences in Escherichia coli. *Nucleic acids research*, 38(6), p.e88.

Goldenberg, A.M. & Wexler, L.F., 1988. Quinine overdose: review of toxicity and treatment. *Clinical cardiology*, 11(10), pp.716–718.

Gomes, A.R. *et al.*, 2015. A genome-scale vector resource enables high-throughput reverse genetic screening in a malaria parasite. *Cell host & microbe*, 17(3), pp.404–413.

Guinet, F. *et al.*, 1996. A developmental defect in Plasmodium falciparum male gametogenesis. *The Journal of cell biology*, 135(1), pp.269–278.

Guo, Z., 2016. Artemisinin anti-malarial drugs in China. *Acta pharmaceutica Sinica. B*, 6(2), pp.115–124.

Gupta, D.K. *et al.*, 2016. DNA damage regulation and its role in drug-related phenotypes in the malaria parasites. *Scientific reports*, 6, p.23603.

Gupta, H. *et al.*, 2017. Molecular surveillance of pfhrp2 and pfhrp3 deletions in Plasmodium falciparum isolates from Mozambique. *Malaria journal*, 16(1), p.416.

Hadley, T.J. *et al.*, 1987. Falciparum malaria parasites invade erythrocytes that lack glycophorin A and B (MkMk). Strain differences indicate receptor heterogeneity and two pathways for invasion. *The Journal of clinical investigation*, 80(4), pp.1190–1193.

Harper, K. & Armelagos, G., 2010. The Changing Disease-Scape in the Third Epidemiological Transition. *International journal of environmental research and public health*, 7(2), pp.675–697.

Hasenkamp, S., Russell, K.T. & Horrocks, P., 2012. Comparison of the absolute and relative efficiencies of electroporation-based transfection protocols for Plasmodium falciparum. *Malaria journal*, 11, p.210.

Hastings, I.M. & Donnelly, M.J., 2005. The impact of antimalarial drug resistance mutations on parasite fitness, and its implications for the evolution of resistance. *Drug resistance updates: reviews and commentaries in antimicrobial and anticancer chemotherapy*, 8(1-2), pp.43–50.

Hayward, R., Saliba, K.J. & Kirk, K., 2005. pfmdr1 mutations associated with chloroquine resistance incur a fitness cost in Plasmodium falciparum. *Molecular microbiology*, 55(4), pp.1285–1295.

Hegreness, M. *et al.*, 2008. Accelerated evolution of resistance in multidrug environments. *Proceedings of the National Academy of Sciences of the United States of America*, 105(37), pp.13977–13981.

Hemingway, J. *et al.*, 2016. Tools and Strategies for Malaria Control and Elimination: What Do We Need to Achieve a Grand Convergence in Malaria? *PLoS biology*, 14(3), p.e1002380.

Henriques, G. *et al.*, 2015. The Mu subunit of Plasmodium falciparum clathrin-associated adaptor protein 2 modulates in vitro parasite response to artemisinin and quinine. *Antimicrobial agents and chemotherapy*, 59(5), pp.2540–2547.

He, X. *et al.*, 2010. Prevalent positive epistasis in Escherichia coli and Saccharomyces cerevisiae metabolic networks. *Nature genetics*, 42(3), pp.272–276.

Hsu, P.D. *et al.*, 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9), pp.827–832.

Ibraheem, Z.O. *et al.*, 2014. Role of Different Pfcrt and Pfmdr-1 Mutations in Conferring Resistance to Antimalaria Drugs in Plasmodium falciparum. *Malaria research and treatment*, 2014, p.950424.

Kafsack, B.F.C. *et al.*, 2014. A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature*, 507(7491):248-52

Kahrstrom, C.T., 2015. Parasite physiology: The missing link in artemisinin resistance. *Nature reviews. Microbiology*, 13(6), pp.328–329.

Kebaier, C., Voza, T. & Vanderberg, J., 2009. Kinetics of mosquito-injected Plasmodium sporozoites in mice: fewer sporozoites are injected into sporozoite-immunized mice. *PLoS pathogens*, 5(4), p.e1000399.

Keightley, P.D. & Kacser, H., 1987. Dominance, pleiotropy and metabolic structure. *Genetics*, 117(2), pp.319–329.

Kessl, J.J., Meshnick, S.R. & Trumpower, B.L., 2007. Modeling the molecular basis of atovaquone resistance in parasites and pathogenic fungi. *Trends in parasitology*, 23(10):494-501

Khan, A.I. *et al.*, 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, 332(6034), pp.1193–1196.

Kirkman, L.A., Lawrence, E.A. & Deitsch, K.W., 2014. Malaria parasites utilize both homologous recombination and alternative end joining pathways to maintain genome integrity. *Nucleic acids research*, 42(1), pp.370–379.

Klonis, N. *et al.*, 2007. Evaluation of pH during cytostomal endocytosis and vacuolar catabolism of haemoglobin in Plasmodium falciparum. *Biochemical Journal*, 407(3), pp.343–354.

Knuepfer, E. *et al.*, 2017. Generating conditional gene knockouts in Plasmodium - a toolkit to produce stable DiCre recombinase-expressing parasite lines using CRISPR/Cas9. *Scientific reports*, 7(1), p.3881.

Koenderink, J.B. *et al.*, 2010. The ABCs of multidrug resistance in malaria. *Trends in parasitology*, 26. (9):440-6.

Krogh, B.O. & Symington, L.S., 2004. Recombination proteins in yeast. *Annual review of genetics*, 38, pp.233–271.

Kublin, J.G. *et al.*, 2003. Reemergence of chloroquine-sensitive Plasmodium falciparum malaria after cessation of chloroquine use in Malawi. *The Journal of infectious diseases*, 187(12), pp.1870–1875.

Lambros, C. & Vanderberg, J.P., 1979. Synchronization of Plasmodium falciparum erythrocytic stages in culture. *The Journal of parasitology*, 65(3), pp.418–420.

Leang, R. *et al.*, 2015. Evidence of Plasmodium falciparum Malaria Multidrug Resistance to Artemisinin and Piperaquine in Western Cambodia: Dihydroartemisinin-Piperaquine Open-Label Multicenter Clinical Assessment. , 59(8), pp.4719–4726.

Lee, A.H. & Fidock, D.A., 2016. Evidence of a Mild Mutator Phenotype in Cambodian Plasmodium falciparum Malaria Parasites. *PloS one*, 11(4), p.e0154166.

Lee, A.H., Symington, L.S. & Fidock, D.A., 2014. DNA repair mechanisms and their biological roles in the malaria parasite Plasmodium falciparum. *Microbiology and molecular biology reviews: MMBR*, 78(3), pp.469–486.

Lee, M.C.S. *et al.*, 2008. Plasmodium falciparum Sec24 marks transitional ER that exports a model cargo via a diacidic motif. *Molecular microbiology*, 68(6), pp.1535–1546.

Lee, M. & Fidock, D., 2014. CRISPR-mediated genome editing of Plasmodium falciparum malaria parasites. *Genome medicine*, 6(8), pp.1–4.

Le Roch, K.G., Chung, D.-W.D. & Ponts, N., 2012. Genomics and integrated systems biology in Plasmodium falciparum: a path to malaria control and eradication. *Parasite immunology*, 34(2-3), pp.50–60.

Levy, S.F. *et al.*, 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542), pp.181–186.

Lewis, I.A. *et al.*, 2014. Metabolic QTL analysis links chloroquine resistance in Plasmodium falciparum to impaired hemoglobin catabolism. *PLoS genetics*, 10(1), p.e1004085.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* , 25(14), pp.1754–1760.

Lim, P. *et al.*, 2013. Ex vivo susceptibility of Plasmodium falciparum to antimalarial drugs in western, northern, and eastern Cambodia, 2011-2012: association with molecular markers. *Antimicrobial agents and chemotherapy*, 57(11), pp.5277–5283.

Lim, P. *et al.*, 2003. pfcrt polymorphism and chloroquine resistance in Plasmodium falciparum strains isolated in Cambodia. *Antimicrobial agents and chemotherapy*, 47(1), pp.87–94.

Li, W. *et al.*, 2005. Yeast Model Uncovers Dual Roles of Mitochondria in the Action of Artemisinin. *PLoS genetics*, 1(3), p.e36.

Lozovsky, E.R. *et al.*, 2009. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proceedings of the National Academy of Sciences of the United States of America*, 106(29), pp.12025–12030.

Lyko, B. *et al.*, 2012. A high-throughput method to detect Plasmodium falciparum clones in limiting dilution microplates. *Malaria journal*, 11, p.124.

Mackinnon, M.J. & Marsh, K., 2010. The selection landscape of malaria parasites. *Science*, 328(5980), pp.866–871.

Mackintosh, C.L., Beeson, J.G. & Marsh, K., 2004. Clinical features and pathogenesis of severe malaria. *Trends in parasitology*, 20(12), pp.597–603.

Maier, A.G. *et al.*, 2008. Exported proteins required for virulence and rigidity of Plasmodium falciparum-infected human erythrocytes. *Cell*, 134(1), pp.48–61.

Maier, A.G. *et al.*, 2006. Negative selection using yeast cytosine deaminase/uracil phosphoribosyl transferase in Plasmodium falciparum for targeted gene deletion by double crossover recombination. *Molecular and biochemical parasitology*, 150(1), pp.118–121.

Maiga-Ascofare, O. & May, J., 2016. Is the A578S Single-Nucleotide Polymorphism in K13-propeller a marker of emerging resistance to artemisinin among Plasmodium falciparum in Africa? *The Journal of infectious diseases*, 213(1):165-6.

Maisnier-Patin, S. *et al.*, 2002. Compensatory adaptation to the deleterious effect of antibiotic resistance in Salmonella typhimurium. *Molecular microbiology*, 46(2), pp.355–366.

Malhotra, I. *et al.*, 2009. Can prenatal malaria exposure produce an immune tolerant phenotype? A prospective birth cohort study in Kenya. *PLoS medicine*, 6(7), p.e1000116.

Mamoun, C.B. *et al.*, 1999. A set of independent selectable markers for transfection of the human malaria parasite Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America*, 96(15), pp.8716–8720.

Manske, M. *et al.*, 2012. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature*, 487(7407), pp.375–379.

Marcusson, L.L., Frimodt-Møller, N. & Hughes, D., 2009. Interplay in the selection of fluoroquinolone resistance and bacterial fitness. *PLoS pathogens*, 5(8), p.e1000541.

Martiney, J.A., Cerami, A. & Slater, A.F., 1995. Verapamil reversal of chloroquine resistance in the malaria parasite Plasmodium falciparum is specific for resistant parasites and independent of the weak base effect. *The Journal of biological chemistry*, 270(38), pp.22393–22398.

Mbengue, A. *et al.*, 2015. A molecular mechanism of artemisinin resistance in Plasmodium falciparum malaria. *Nature*, 520(7549):683-7.

McMahon, M. *et al.*, 2003. Keap1-dependent proteasomal degradation of transcription factor Nrf2 contributes to the negative regulation of antioxidant response element-driven gene expression. *The Journal of biological chemistry*, 278(24), pp.21592–21600.

McNamara, C.W. *et al.*, 2013. Targeting Plasmodium PI(4)K to eliminate malaria. *Nature*, 504(7479), pp.248–253.

Medicine, Institute of, 1991. *Malaria: Obstacles and Opportunities*, Washington, DC: The National Academies Press.

Meshnick, S.R., 2002. Artemisinin: mechanisms of action, resistance and toxicity. *International journal for parasitology*, 32(13), pp.1655–1660.

Miller, L.H. *et al.*, 2002. The pathogenic basis of malaria. *Nature*, 415(6872), pp.673–679.

Mimitou, E.P. & Symington, L.S., 2011. DNA end resection—Unraveling the tail. *DNA repair*, 10(3), pp.344–348.

Miotto, O. *et al.*, 2015. Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nature genetics*, 47(3):226-34.

Miotto, O. *et al.*, 2013. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. *Nature genetics*, 45(6), pp.648–655.

Mita, T., Tanabe, K. & Kita, K., 2009. Spread and evolution of Plasmodium falciparum drug resistance. *Parasitology international*, 58(3), pp.201–209.

Mok, S. *et al.*, 2015. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science*, 347(6220):431-5.

Moonen, B. *et al.*, 2010. Operational strategies to achieve and maintain malaria elimination. *The Lancet*, 376(9752), pp.1592–1603.

Moon, R.W. *et al.*, 2013. Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. *Proceedings of the National Academy of Sciences of the United States of America*, 110(2), pp.531–536.

Moura de Sousa, J. *et al.*, 2017. Multidrug-resistant bacteria compensate for the epistasis between resistances. *PLoS biology*, 15(4), p.e2001741.

Murray, L. *et al.*, 2017. Multiplication rate variation in the human malaria parasite Plasmodium falciparum. *Scientific reports*, 7(1), p.6436.

Mwai, L. *et al.*, 2009. Chloroquine resistance before and after its withdrawal in Kenya. *Malaria journal*, 8, p.106.

Nair, S. *et al.*, 2018. Do fitness costs explain the rapid spread of kelch13-C580Y substitutions conferring artemisinin resistance? *Antimicrobial agents and chemotherapy*, 62(9). pii: e00605-18.

Nguyen, T.T. *et al.*, 2014. Robust dose-response curve estimation applied to high content screening data analysis. *Source code for biology and medicine*, 9(1), p.27.

Noedl, H. *et al.*, 2008. Evidence of Artemisinin-Resistant Malaria in Western Cambodia. *The New England journal of medicine*, 359(24), pp.2619–2620.

Nosten, F. *et al.*, 2000. Effects of artesunate-mefloquine combination on incidence of Plasmodium falciparum malaria and mefloquine resistance in western Thailand: a prospective study. *The Lancet*, 356(9226), pp.297–302.

Nosten, F.H., 2010. Pyronaridine-artesunate for uncomplicated falciparum malaria. *The Lancet*, 375(9724), pp.1413–1414.

Noubiap, J.J.N., 2014. Shifting from quinine to artesunate as first-line treatment of severe malaria in children and adults: saving more lives. *Journal of infection and public health*, 7(5), pp.407–412.

Ooi, Y.-S. *et al.*, 2008. Recombineering linear DNA that replicate stably in E. coli. *Plasmid*, 59(1), pp.63–71.

Orr-Weaver, T.L., Szostak, J.W. & Rothstein, R.J., 1981. Yeast transformation: a model system for the study of recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 78(10), pp.6354–6358.

Ott, K.J., 1967. Malaria Parasites and Other Haemosporidia. *Science*, 157(3792), pp.1029–1029.

Paloque, L. *et al.*, 2016. Plasmodium falciparum: multifaceted resistance to artemisinins. *Malaria journal*, 15(1), pp.1–12.

Payne, D., 1987. Spread of chloroquine resistance in Plasmodium falciparum. *Parasitology today* , 3(8), pp.241–246.

Peters, J.M. *et al.*, 2002. Mutations in cytochrome b resulting in atovaquone resistance are associated with loss of fitness in Plasmodium falciparum. *Antimicrobial agents and chemotherapy*, 46(8), pp.2435–2441.

Pfander, C. *et al.*, 2011. A scalable pipeline for highly effective genetic modification of a malaria parasite. *Nature methods*, 8(12), pp.1078–1082.

Pigott, D.M. *et al.*, 2012. Funding for malaria control 2006–2010: A comprehensive global assessment. *Malaria journal*, 11(1), pp.1–11.

Pinder, J., Salsman, J. & Dellaire, G., 2015. Nuclear domain "knock-in" screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic acids research*, 43(19), pp.9379–9392.

Pollack, Y. *et al.*, 1982. The genome of Plasmodium falciparum. I: DNA base composition. *Nucleic acids research*, 10(2), pp.539–546.

Porteus, M.H. & Carroll, D., 2005. Gene targeting using zinc finger nucleases. *Nature biotechnology*, 23(8), pp.967–973.

Preechapornkul, P. *et al.*, 2009. Plasmodium falciparum pfmdr1 amplification, mefloquine resistance, and parasite fitness. *Antimicrobial agents and chemotherapy*, 53(4), pp.1509–1515.

Price, M.N., Dehal, P.S. & Arkin, A.P., 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution*, 26(7), pp.1641–1650.

Price, R.N. *et al.*, 2004. Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *The Lancet*, 364(9432):438-447.

Price, R.N. & Douglas, N.M., 2009. Artemisinin combination therapy for malaria: beyond good efficacy. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 49(11), pp.1638–1640.

Radfar, A. *et al.*, 2009. Synchronous culture of Plasmodium falciparum at high parasitemia levels. *Nature protocols*, 4(12), pp.1899–1915.

Ran, F.A. *et al.*, 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11), pp.2281–2308.

Rathod, P.K., McErlean, T. & Lee, P.C., 1997. Variations in frequencies of drug resistance in Plasmodium falciparum. *Proceedings of the National Academy of Sciences of the United States of America*, 94(17), pp.9389–9393.

Rayner, J.C. *et al.*, 2000. Two Plasmodium falciparum genes express merozoite proteins that are related to Plasmodium vivax and Plasmodium yoelii adhesive proteins involved in host cell selection and invasion. *Proceedings of the National Academy of Sciences*, 97(17), pp.9648–9653.

Rice, J.B., 1932. Experiments on Plasmoquine as a Prophylactic Among West African Negroes Exposed to Bites of A. Costalis Infected with Subtertian Malaria. *Annals of tropical medicine and parasitology*, 26(4), pp.555–576.

Ritz, C. *et al.*, 2015. Dose-Response Analysis Using R. *PloS one*, 10(12), p.e0146021.

Roper, C. *et al.*, 2004. Intercontinental spread of pyrimethamine-resistant malaria. *Science*, 305(5687), p.1124.

Rosario, V., 1981. Cloning of naturally occurring mixed infections of malaria parasites. *Science*, 212(4498), pp.1037–1038.

Rosenthal, P.J., 2013. The interplay between drug resistance and fitness in malaria parasites. *Molecular microbiology*, 89(6), pp.1025–1038.

Ross, L.S. *et al.*, 2018. Emerging Southeast Asian PfCRT mutations confer Plasmodium falciparum resistance to the first-line antimalarial piperaquine. *Nature communications*, 9(1), p.3314.

Rovira-Graells, N. *et al.*, 2016. New Assays to Characterise Growth-Related Phenotypes of Plasmodium falciparum Reveal Variation in Density-Dependent Growth Inhibition between Parasite Lines. *PloS one*, 11(10), p.e0165358.

Rug, M. & Maier, A.G., 2013. Transfection of Plasmodium falciparum. *Methods in molecular biology* , 923, pp.75–98.

Russell, B. *et al.*, 2013. Field-based flow cytometry for ex vivo characterization of Plasmodium vivax and P. falciparum antimalarial sensitivity. *Antimicrobial agents and chemotherapy*, 57(10), pp.5170–5174.

Rybak, M.J. & McGrath, B.J., 1996. Combination antimicrobial therapy for bacterial infections. Guidelines for the clinician. *Drugs*, 52(3), pp.390–405.

Sá, J.M. *et al.*, 2009. Geographic patterns of Plasmodium falciparum drug resistance distinguished by differential responses to amodiaquine and chloroquine. *Proceedings of the National Academy of Sciences of the United States of America*, 106(45), pp.18883–18889.

Saliba, K.S. & Jacobs-Lorena, M., 2013. Production of Plasmodium falciparum gametocytes in vitro. *Methods in molecular biology* , 923, pp.17–25.

Sanchez, C.P. *et al.*, 2010. Transporters as mediators of drug resistance in Plasmodium falciparum. *International journal for parasitology*, 40(10):1109-18.

Sander, J.D. & Joung, J.K., 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology*, 32(4), pp.347–355.

Sanger, F. *et al.*, 1973. Use of DNA Polymerase I Primed by a Synthetic Oligonucleotide to Determine a Nucleotide Sequence in Phage f1 DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 70(4), pp.1209–1213.

Schlitzer, M., 2007. Malaria chemotherapeutics part I: History of antimalarial drug development, currently used therapeutics, and drugs in clinical development. *ChemMedChem*, 2(7), pp.944–986.

Schwach, F. *et al.*, 2015. PlasmoGEM, a database supporting a community resource for large-scale experimental genetics in malaria parasites. *Nucleic acids research*, 43(D1), pp.D1176–D1182.

Schwartz, I.K. *et al.*, 1983. In-vivo and in-vitro assessment of chloroquine-resistant Plasmodium falciparum malaria in Zanzibar. *The Lancet*, 1(8332), pp.1003–1005.

von Seidlein, L. & Greenwood, B.M., 2003. Mass administrations of antimalarial drugs. *Trends in parasitology*, 19(10), pp.452–460.

Setthaudom, C. *et al.*, 2011. Role of Plasmodium falciparum chloroquine resistance transporter and multidrug resistance 1 genes on in vitro chloroquine resistance in isolates of Plasmodium falciparum from Thailand. *The American journal of tropical medicine and hygiene*, 85(4), pp.606–611.

Shandilya, A. *et al.*, 2013. A plausible mechanism for the antimalarial activity of artemisinin: A computational approach. *Scientific reports*, 3, p.2513.

Sidhu, A.B.S. *et al.*, 2006. Decreasing pfmdr1 copy number in plasmodium falciparum malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. *The Journal of infectious diseases*, 194(4), pp.528–535.

Sidik, S.M. *et al.*, 2016. A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. *Cell*, 166(6), pp.1423–1435.e12.

Singh, G.P., Goel, P. & Sharma, A., 2016. Structural mapping of Kelch13 mutations associated with artemisinin resistance in malaria. *Journal of structural and functional genomics*, 17(2-3), pp.51–56.

Sinha, A. *et al.*, 2014. A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. *Nature*, 507(7491), pp.253–257.

Skinner-Adams, T.S. *et al.*, 2003. Comparison of Plasmodium falciparum transfection methods. *Malaria journal*, 2(1), pp.1–4.

Smilkstein, M. *et al.*, 2004. Simple and inexpensive fluorescence-based technique for high-throughput antimalarial drug screening. *Antimicrobial agents and chemotherapy*, 48(5), pp.1803–1806.

Smith, A.M. *et al.*, 2009. Quantitative phenotyping via deep barcode sequencing. *Genome research*, 19(10), pp.1836–1842.

Solyakov, L. *et al.*, 2011. Global kinomic and phospho-proteomic analyses of the human malaria parasite Plasmodium falciparum. *Nature communications*, 2, p.565.

Straimer, J. *et al.*, 2015. Drug resistance. K13-propeller mutations confer artemisinin resistance in Plasmodium falciparum clinical isolates. *Science*, 347(6220):428-31.

Straimer, J. *et al.*, 2015. K13-propeller mutations confer artemisinin resistance in Plasmodium falciparum clinical isolates. *Science*, 347(6220), pp.428–431.

Straimer, J. *et al.*, 2017. Plasmodium falciparum K13 Mutations Differentially Impact Ozonide Susceptibility and Parasite Fitness In Vitro. *mBio*, 8(2). pii: e00172-17.

Straimer, J., Lee, M.C.S., Lee, A.H., Zeitler, B., Williams, A.E., Pearl, J.R., Zhang, L., Rebar, E.J., Gregory, P.D., Llinas, M., *et al.*, 2012. Site-specific genome editing in Plasmodium falciparum using engineered zinc-finger nucleases. *Nature methods*, 9(10), pp.993–998.

Straimer, J., Lee, M.C.S., Lee, A.H., Zeitler, B., Williams, A.E., Pearl, J.R., Zhang, L., Rebar, E.J., Gregory, P.D., Llinás, M., *et al.*, 2012. Site-specific genome editing in Plasmodium falciparum using engineered zinc-finger nucleases. *Nature methods*, 9(10), pp.993–998.

Struck, N.S. *et al.*, 2008. Spatial dissection of the cis- and trans-Golgi compartments in the malaria parasite Plasmodium falciparum. *Molecular microbiology*, 67(6), pp.1320–1330.

Summers, R.L., Nash, M.N. & Martin, R.E., 2012. Know your enemy: understanding the role of PfCRT in drug resistance could lead to new antimalarial tactics. *Cellular and molecular life sciences: CMLS*, 69(12):1967-95.

Sundararaman, S.A. *et al.*, 2016. Genomes of cryptic chimpanzee Plasmodium species reveal key evolutionary events leading to human malaria. *Nature communications*, 7, p.11078.

Sutherland, C.J. *et al.*, 2010. Two nonrecombining sympatric forms of the human malaria parasite Plasmodium ovale occur globally. *The Journal of infectious diseases*, 201(10), pp.1544–1550.

Su, X. *et al.*, 1997. Complex polymorphisms in an~ 330 kDa protein are linked to chloroquine-resistant P. falciparum in Southeast Asia and Africa. *Cell*, 91(5):593-603.

Taguchi, K., Motohashi, H. & Yamamoto, M., 2011. Molecular mechanisms of the Keap1–Nrf2 pathway in stress response and cancer evolution. *Genes to cells: devoted to molecular & cellular mechanisms*, 16(2), pp.123–140.

Tang, L., 2000. Progress in malaria control in China. *Chinese medical journal*, 113(1), pp.89–92.

Tavares, J. *et al.*, 2013. Role of host cell traversal by the malaria sporozoite during liver infection. *The Journal of experimental medicine*, 210(5), pp.905–915.

Taylor, H.M. *et al.*, 2001. Plasmodium falciparum homologue of the genes for Plasmodium vivax and Plasmodium yoelii adhesive proteins, which is transcribed but not translated. *Infection and immunity*, 69(6), pp.3635–3645.

Taylor, S.M. & Juliano, J.J., 2014. Artemisinin combination therapies and malaria parasite drug resistance: the game is afoot. *The Journal of infectious diseases*, 210(3), pp.335–337.

Tewari, R. *et al.*, 2010. The Systematic Functional Analysis of Plasmodium Protein Kinases Identifies Essential Regulators of Mosquito Transmission. *Cell host & microbe*, 8(4), pp.377–387.

Titulaer, H.A.C., Zuidema, J. & Lugt, C.B., 1991. Formulation and pharmacokinetics of artemisinin and its derivatives. *International journal of pharmaceutics*, 69(2), pp.83–92.

Trager, W. & Jensen, J.B., 1976. Human malaria parasites in continuous culture. *Science*, 193(4254), pp.673–675.

Trape, J.F., 2001. The public health impact of chloroquine resistance in Africa. *The American journal of tropical medicine and hygiene*, 64(1-2 Suppl), pp.12–17.

Trindade, S. *et al.*, 2009. Positive epistasis drives the acquisition of multidrug resistance. *PLoS genetics*, 5(7), p.e1000578.

Tucker, M.S. *et al.*, 2012. Phenotypic and genotypic analysis of in vitro-selected artemisinin-resistant progeny of Plasmodium falciparum. *Antimicrobial agents and chemotherapy*, 56(1), pp.302–314.

Tu, Y., 2011. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nature medicine*, 17(10), pp.1217–1220.

Udeinya, I.J. *et al.*, 1983. Plasmodium falciparum: effect of time in continuous culture on binding to human endothelial cells and amelanotic melanoma cells. *Experimental parasitology*, 56(2), pp.207–214.

Urnov, F.D. *et al.*, 2010. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*, 11(9), pp.636–646.

Valen, V. & L, 1973. A new evolutionary law. *Evol Theory*, 1, pp.1–30.

Van Voorhis, W.C. *et al.*, 2016. Open Source Drug Discovery with the Malaria Box Compound Collection for Neglected Diseases and Beyond. *PLoS pathogens*, 12(7), p.e1005763.

Vijaykadga, S. *et al.*, 2006. In vivo sensitivity monitoring of mefloquine monotherapy and artesunate--mefloquine combinations for the treatment of uncomplicated falciparum malaria in Thailand in 2003. *Tropical medicine & international health: TM & IH*, 11(2), pp.211–219.

Wagner, J.C. *et al.*, 2014. Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum. *Nature methods*, 11(9), pp.915–918.

Walliker, D. *et al.*, 1987. Genetic analysis of the human malaria parasite Plasmodium falciparum. *Science*, 236(4809), pp.1661–1666.

Wang, T. *et al.*, 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, 343(6166), pp.80–84.

Warhurst, D.C. & Williams, J.E., 1996. ACP Broadsheet no 148. *Journal of clinical pathology*, 49(7), pp.533–538.

Waters, A.P., Higgins, D.G. & McCutchan, T.F., 1991. Plasmodium falciparum appears to have arisen as a result of lateral transfer between avian and human hosts. *Proceedings of the National Academy of Sciences of the United States of America*, 88(8), pp.3140–3144.

Weinreich, D.M. & Chao, L., 2005. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. *Evolution; international journal of organic evolution*, 59(6), pp.1175–1182.

Wellems, T.E. *et al.*, 1990. Chloroquine resistance not linked to mdr-like genes in a Plasmodium falciparum cross. *Nature*, 345(6272), pp.253–255.

Wellems, T.E. & Plowe, C.V., 2001. Chloroquine-resistant malaria. *The Journal of infectious diseases*, 184(6), pp.770–776.

Wellems, T.E., Walker-Jonah, A. & Panton, L.J., 1991. Genetic mapping of the chloroquine-resistance locus on Plasmodium falciparum chromosome 7. *Proceedings of the National Academy of Sciences of the United States of America*, 88(8), pp.3382–3386.

Williamson, D.H. *et al.*, 2002. Topology and replication of a nuclear episomal plasmid in the rodent malaria Plasmodium berghei. *Nucleic acids research*, 30(3), pp.726–731.

Winzeler, E.A. *et al.*, 1999. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science*, 285(5429), pp.901–906.

Witkowski, B. *et al.*, 2017. A surrogate marker of piperaquine-resistant Plasmodium falciparum malaria: a phenotype–genotype association study. *The Lancet infectious diseases*, 17(2), pp.174–183.

Witkowski, B. *et al.*, 2010. Increased tolerance to artemisinin in Plasmodium falciparum is mediated by a quiescence mechanism. *Antimicrobial agents and chemotherapy*, 54(5):1872-7.

Witkowski, B. *et al.*, 2013. Novel phenotypic assays for the detection of artemisinin-resistant Plasmodium falciparum malaria in Cambodia: in-vitro and ex-vivo drug-response studies. *The Lancet infectious diseases*, 13(12), pp.1043–1049.

Witkowski, B. *et al.*, 2013. Reduced artemisinin susceptibility of Plasmodium falciparum ring stages in western Cambodia. *Antimicrobial agents and chemotherapy*, 57(2):914-23.

Woodrow, C.J. & White, N.J., 2016. The clinical impact of artemisinin resistance in Southeast Asia and the potential for future spread. *FEMS microbiology reviews*, 41(1):34-48.

Wootton, J.C. *et al.*, 2002. Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. *Nature*, 418(6895), pp.320–323.

World Health Organization, 2017. *World Malaria Report* World Health Organization, ed., Geneva, Switzerland.

World Health Organization, 2016. *World Malaria Report 2015*, World Health Organization.

Wu, Y. *et al.*, 1995. Transfection of Plasmodium falciparum within human red blood cells. *Proceedings of the National Academy of Sciences of the United States of America*, 92(4):973-7.

Wu, Y., Kirkman, L.A. & Wellems, T.E., 1996. Transformation of Plasmodium falciparum malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *Proceedings of the National Academy of Sciences of the United States of America*, 93(3):1130-4.

Zhang, M. *et al.*, 2018. Uncovering the essential genes of the human malaria parasite Plasmodium falciparum by saturation mutagenesis. *Science*, 360(6388). pii: eaap7847.

Zipper, L.M. & Mulcahy, R.T., 2002. The Keap1 BTB/POZ dimerization function is required to sequester Nrf2 in cytoplasm. *The Journal of biological chemistry*, 277.