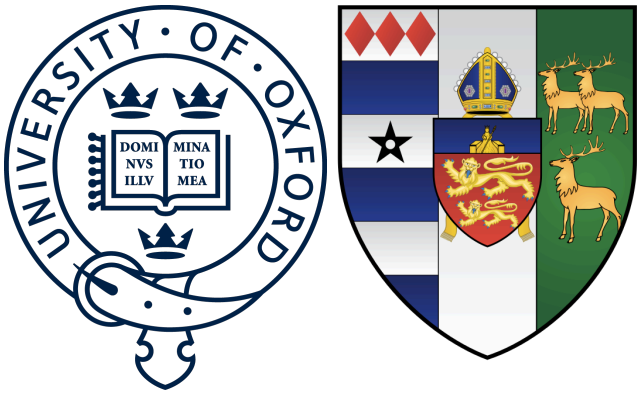
Investigation of Genetic Determinants of Drug Response in a *Plasmodium falciparum* Genetic Cross Using a High-Throughput Screening Method



Michael Allen Krause

Lincoln College

Trinity 2015

This thesis is submitted

in partial fulfillment of the requirements

for the degree Doctor of Philosophy

ABSTRACT

The emergence and rapid spread of artemisinin-resistant *Plasmodium falciparum* in southeast Asia highlights the importance of identifying genetic determinants of drug response and discovering novel potent antimalarials. In support of these efforts, the parental lines and progeny of the 803xGB4 genetic cross were characterized by pharmaceutical compound screening and whole-genome sequencing (WGS). Results of the compound screen identified 52 highly-active compounds with potential for development as antimalarials. A confirmatory screen was performed to more accurately define parasite responses to 384 of the 2,816 compounds in the primary screen. Linkage analysis revealed 25 genetic loci significantly associated with parasite response to 61 compounds. A locus on chromosome 5 containing *pfmdr1* was linked to response to six compounds, including the widely-used antimalarial mefloquine and the potential antimalarial triclosan. Progeny bearing the 803 allele at this locus were less sensitive to all six of these compounds. Results from RT-PCR experiments indicated increased transcription of *pfmdr1* in 803 in the absence of copy number variation (CNV), implicating a possible promoter polymorphism. Analysis of WGS data suggested the presence of several possible CNVs, including a deletion of *Rh2b* in 803 inherited by a proportion of progeny, and a non-Mendelian amplification of *Rh2a* and *Rh2b* in three progeny. Results from PCR and RT-PCR experiments confirmed the deletion event, and indicated increased *Rh2a* and *Rh2b* transcript levels in one of the progeny with the amplification. High-throughput compound screening and WGS enable the discovery of parasite biological responses that have the potential to further malaria elimination efforts.

ACKNOWLEDGMENTS

I am extremely grateful for the mentorship provided by Dr. Rick M. Fairhurst and Professor Dominic Kwiatkowski. Without their support, this project would not have been possible. I am also indebted to Dr. Thomas E. Wellems and Dr. Juliana M. Sá for the opportunity to work with the 803xGB4 genetic cross and for many rich discussions. I wish to thank the many coworkers at the Laboratory of Malaria and Vector Research at NIH and the Wellcome Trust Centre for Human Genetics in Oxford who gave me help and advice along the way. Finally, I thank Joe and Stella, who kept me sane.

STATEMENT OF WORK PERFORMED

Unless otherwise indicated, the work presented in this thesis is wholly my own. Notable exceptions are listed below.

The *Plasmodium falciparum* genetic cross described in this thesis was spearheaded by Dr. Juliana M. Sá in Dr. Thomas E. Wellems' laboratory, and involved a global collaboration of laboratories, institutes, and biotechnology companies. The *P. falciparum* line 803 was isolated from a patient as part of a clinical study of artemisinin resistance by Dr. Chanaki Amaratunga in Dr. Rick M. Fairhurst's laboratory. A complete list of contributors can be found in Appendix I.

I had assistance in performing the compound screen described in Chapters 3 and 4 from Sampada A. Shahane and Anna Liu, and in analyzing the results from Dr. Ruili Huang of Dr. Menghang Xia's laboratory.

Whole-genome sequence data was generated by the sequencing core at the Wellcome Trust Sanger Institute. I received substantial support from Alistair Miles in Professor Kwiatkowski's group in analyzing the whole-genome sequence data described in Chapter 5.

INDEX OF FIGURES

INDEX OF TABLES

List of abbreviations:

ABHom:	proportion of reference alleles
AQ:	amodiaquine
ARM:	artemisinin
ART:	artesunate
AVQ:	atovaquone
BaseQRankSum:	U-based, z-approximation from the Mann-Whitney Rank Sum Test for base qualities
chr:	chromosome
CQ:	chloroquine
CV:	coefficient of variation
DCP:	differential chemical phenotype
DHA:	dihydroartemisinin
dhfr:	dihydrofolate reductase
DMSO:	dimethyl sulfoxide
DNA:	deoxyribonucleic acid
DP:	total depth of reads that passed quality control
DV:	digestive vacuole
EDTA:	ethylenediaminetetraacetate
EIR:	entomological inoculation rate
FS:	Phred-scaled p-value using Fisher's Exact Test to detect strand bias
GWAS:	genome-wide association study
G6PD:	glucose-6-phosphate dehydrogenase
Hb:	hemoglobin
HbS:	sickle hemoglobin
HLF:	halofantrine
HEPES:	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
IC_{50}:	median inhibitory concentration
Indels:	insertions-deletions
kb:	kilobase
LOD:	logarithm of odds
LMF:	lumefantrine
LMVR:	Laboratory of Malaria and Vector Research
MDAQ:	monodesethylamodiaquine

MFQ:	mefloquine
MFU:	mean fluorescence unit
MQ:	root mean square of mapping quality of reads across all samples
MS:	microsatellite
PBS:	phosphate-buffered saline
PCR:	polymerase chain reaction
pfcrt:	*Plasmodium falciparum* chloroquine resistance transporter
PfEMP1:	*Plasmodium falciparum* erythrocyte membrane protein1
pfmdr1:	*Plasmodium falciparum* multidrug resistance 1
PPQ:	piperaquine
PYR:	pyrimethamine
QD:	variant confidence / unfiltered depth of non-reference samples
qHTS:	quantitative high-throughput screening
QN:	quinine
QTL:	quantitative trait loci
RBC:	red blood cell
Rh:	reticulocyte homologue
RNA:	ribonucleic acid
S/B:	signal-to-background ratio
SD:	standard deviation
SDS:	sodium dodecylsulfate
SNP:	single nucleotide polymorphism
SUL:	sulfadoxine
TSE:	Tris-saline-EDTA solution
VQSLOD:	variant quality score logarithm of odds
VQSR:	variant quality score recalibration
v/v:	volume per volume
WGS:	whole-genome sequence
w/v:	weight per volume

1 Introduction

1.1 P. falciparum *life cycle and pathogenesis*

Five *Plasmodium* species are known to cause illness in humans: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. The parasites infect red blood cells (RBCs) of humans and other mammals[1]. Malaria caused by *P. falciparum* is a global illness, with around half a billion cases and half a million deaths each year[2]. The other four species cause several hundred million additional cases, but without significant mortality[3]. This thesis will focus on malaria caused by *P. falciparum*.

The parasite life cycle begins when an infected mosquito bites a human host (Figure 1-1 #1). Sporozoites released into the skin from the mosquito's salivary gland migrate to the bloodstream[4] and travel to the liver, where they invade hepatocytes[5]. For 5-16 days, the sporozoites undergo several rounds of replication within the hepatocyte, which then bursts releasing thousands of merozoites into the bloodstream (Figure 1-1B)[6]. The merozoites initiate the asexual stage of the parasite life cycle, characterized by a 48-hour cycle of RBC invasion, maturation, sequestration in the microvasculature, rupture, and reinvasion[1,7–9]. This cycle continues until the infection is controlled by the immune system or cured with antimalarial therapy[1].

Sequestration of *P. falciparum*-infected RBCs is a key event in the life cycle[10]. This process enables the parasite to evade the human immune system, and to cause malaria symptoms[11]. Upon RBC invasion, the parasite initiates a complex series of structural modifications of the RBC cytoplasm and cell membrane (Figure 1-2)[12,13]. One important modification is the transport of cell-adhesion molecules to the RBC surface (mediated by unique structures called Maurer's clefts[14]), where they form knob-like protrusions (Figure 1-2B). These molecules facilitate parasitized RBC

adhesion to uninfected RBCs and endothelial cells of the microvasculature[13,15,16]. Binding to uninfected RBCs (termed rosetting[17]) is thought to enable the parasite to evade immune detection, while binding to endothelial cells (termed cytoadhesion) is thought to prevent clearance of mature parasites by the spleen[18]. Binding to endothelial cells also triggers host inflammatory responses that cause the symptoms of malaria[19–22]. The effect can be devastating, especially if parasites adhere in the microvasculature of the brain where they can impede blood flow and cause cerebral malaria[23–25].

The proteins responsible for cytoadhesion undergo a staggering level of antigenic variation[26]. Chief among them are members of the *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) family, encoded by *var* genes, which interact with specific receptors on endothelial cells[27]. This interaction mediates parasite sequestration in the microvasculature. Each parasite genome encodes ~60 different *var* genes, which are expressed in a mutually independent manner[28]. Switching between different *var* genes enables the parasite to escape immune detection and modify its tissue specificity[29]. The parasite's capacity for immune evasion is difficult to understate; on a global scale, the level of *var* gene diversity is immense[30,31], complicating efforts to develop a PfEMP1 targeting vaccine. One of the gravest results of *var* switching is the expression of the *var2csa* variant in pregnant women. This variant allows the parasite to bind to the placenta, leading to a constellation of complications including abortion and premature birth[32–34].

As the parasite develops within the RBC, it digests the cell's hemoglobin (Hb)[35]. The digestive process generates toxic free heme, which is polymerized by the parasite into a nontoxic crystalline form called hemozoin and stored in the digestive vacuole (DV) (Figure 1-2B)[36]. The parasite uses amino acids freed from hemoglobin digestion to mature within the RBC, first becoming a trophozoite, and then a schizont.

The schizont segments into separate merozoites, which will invade new RBCs upon schizont rupture[37].

A subset of merozoites develops into non-replicating gametocytes, initiating the sexual stage of the life cycle (Figure 1-1 C)[38]. Mature male and female gametocytes are taken up during a mosquito blood meal. Once in the mosquito midgut, the male gametocyte exflagellates and combines with the female gametocyte to form a zygote that develops into an ookinete; it is here that sexual recombination occurs. The zygote and ookinete are the only diploid stages in the parasite life cycle[39]. The ookinete traverses the midgut epithelium, entering the basal lamina and forming an oocyst[40]. Over the course of 7-21 days, the oocyst matures in the mosquito midgut and undergoes several rounds of division. After maturation, the oocyst ruptures, releasing thousands of sporozoites into the mosquito hemolymph. The sporozoites migrate to the mosquito's salivary gland, where they can infect a new host at the mosquito's next blood meal[41].

Malaria parasites are spread by *Anopheles* species of mosquitoes, which are found worldwide (Figure 1-3)[42]. In some regions in sub-Saharan Africa, humans are exposed to greater than 100 infectious bites per year, placing >80% of the population at risk of contracting malaria (Figure 1-4). Children bear the brunt of this burden and are much more likely to die following infection[2]. Indeed, malaria caused by *P. falciparum* has been so deadly over the course of human history that it has exerted selective pressure on the human genome[43].

J. B. S. Haldane first posited that beta-thalassemia represented a balanced polymorphism, in that the deleterious trait (life-threatening anemia) was balanced by a beneficial trait (protection afforded against fatal malaria)[44]. No epidemiological study has satisfactorily addressed his hypothesis, in large part because malaria is no longer endemic in areas where beta-thalassemia trait is prevalent (e.g., the

Mediterranean basin)[45]. In 1954, A. S. Allison expanded the concept to sickle hemoglobin (HbS), noting that, "the incidence of sickle-cell trait was higher in areas where malaria was prevalent than elsewhere."[46] This observation can be made by comparing the geographic distribution of the entomological inoculation rate (EIR, a measure of the number of infected bites per person per year) for *P. falciparum* (Figure 1-5)[47] with that of HbS allele frequency (Figure 1-6)[48]. HbS is caused by a glutamate-to-valine substitution at position 6 of the beta globin gene. HbSS homozygotes almost invariably die at a young age in Africa from complications of sickle cell disease. HbAS heterozygotes, however, have a reduced risk of severe malaria[49].

Further evidence of varying levels of clinical protection against severe malaria can be found for multiple other RBC polymorphisms. These include HbC[50,51] and HbE[52,53], alpha-thalassemia[54–56], glucose-6-phosphate dehydrogenase (G6PD) deficiency[57–59], and Southeast Asian ovalocytosis[60,61]. Together, RBC polymorphisms encode mankind's most common inherited disorders, offering a striking example of how profound an impact malaria has had on the human genome[43].

The mechanism by which these polymorphisms protect against severe malaria and death remains unclear, and is likely to involve the interaction of many genes and also depend upon environmental factors. The mechanism by which HbS protects against fatal malaria has been the subject of research for several decades. A range of theories have been put forward, including enhanced sickling of infected HbAS RBCs leading to splenic clearance[62,63], impaired parasite growth in HbAS RBCs[64,65], and enhanced phagocytosis of infected HbAS RBCs by monocytes[66,67]. Some investigators have attempted to develop broad mechanisms that could explain the protection afforded by multiple hemoglobinopathies. One theory postulates that hemoglobinopathic RBCs are structurally deficient[68], which prevents the parasite

from trafficking PfEMP1 to the RBC surface[69–72]. This deficiency prevents the parasite from triggering a strong immune reaction, reducing the risk of severe malaria[73]. The protective effect conferred by hemoglobinopathies may only occur in young children who are in high-transmission areas[74], highlighting the complex interaction between genetics and environment in determining risk for severe malaria.

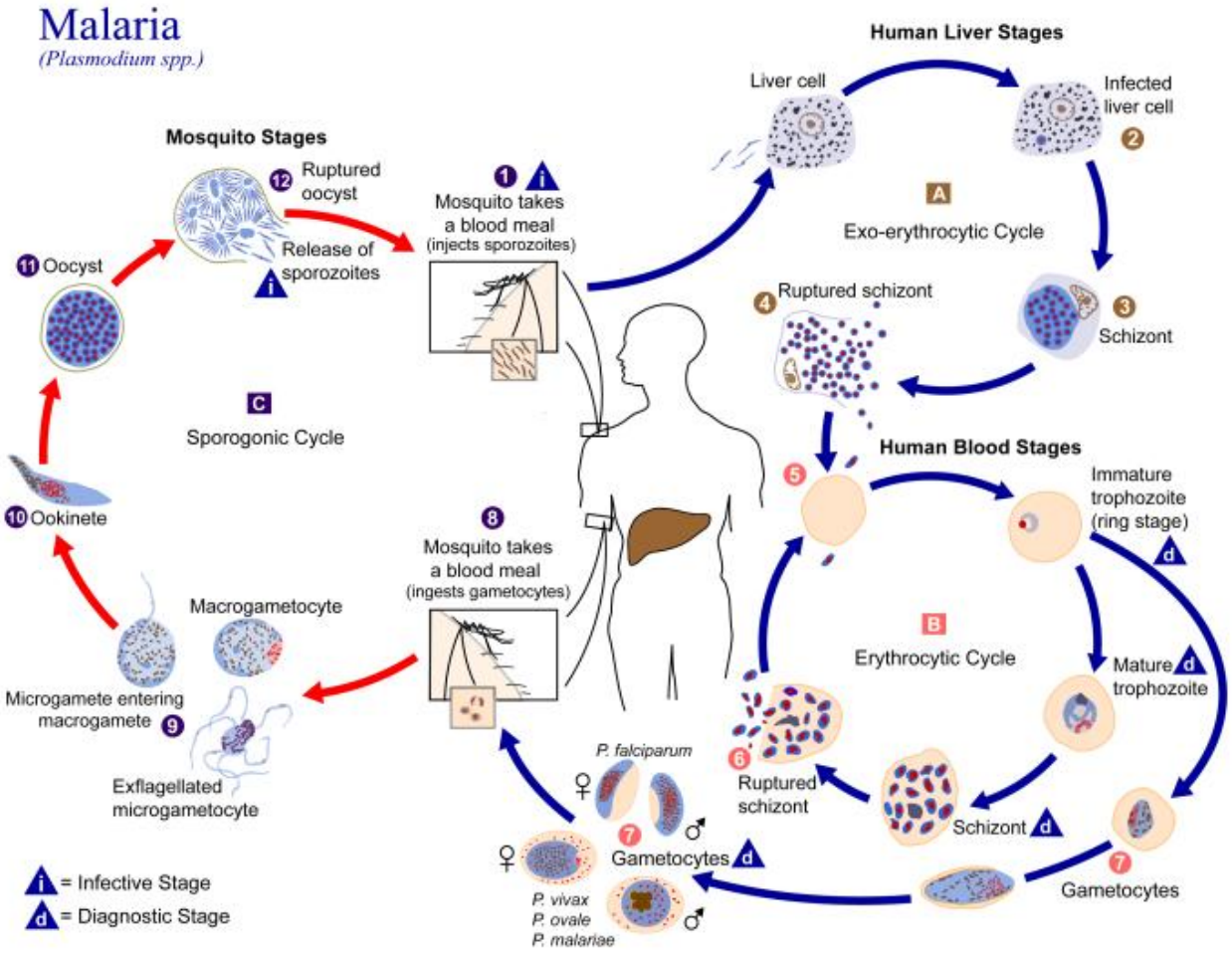


Figure 1-1 *P. falciparum* life cycle. (1) Infection begins when a mosquito bites a human host, injecting sporozoites. (**A**) The sporozoites invade a hepatocyte, replicate, and rupture. (**B**) Merozoites invade RBCs, initiating the erythrocytic cycle. Some merozoites form gametocytes (7), which are taken up by a mosquito during a blood meal (8). (**C**) Micro- and macrogametocytes merge to form an ookinete, triggering the sporogonic cycle in the mosquito. Figure from CDC DPDx – Malaria http://www.cdc.gov/dpdx/malaria/ (Visited 10 July 2015).

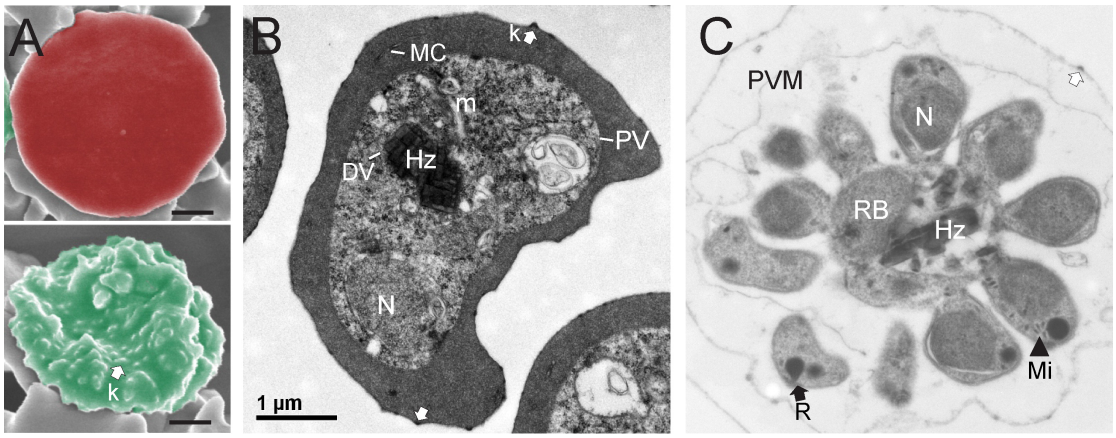


Figure 1-2 Morphology of *P. falciparum*-infected RBCs. (A) Scanning electron micrographs of uninfected (top) and infected (bottom) RBCs. (B and C) Transmission electron micrographs of RBCs showing a trophozoite (B) and a schizont (C). Visible features include: Maurer's cleft (MC), knob (k), mitochondrion (m), digestive vacuole (DV), hemozoin (Hz), parasitophorous vacuole (PV), parasitophorous vacuole membrane (PVM), nucleus (N), remnant body (RB), and rhoptry (R). Figure from Ref 13.

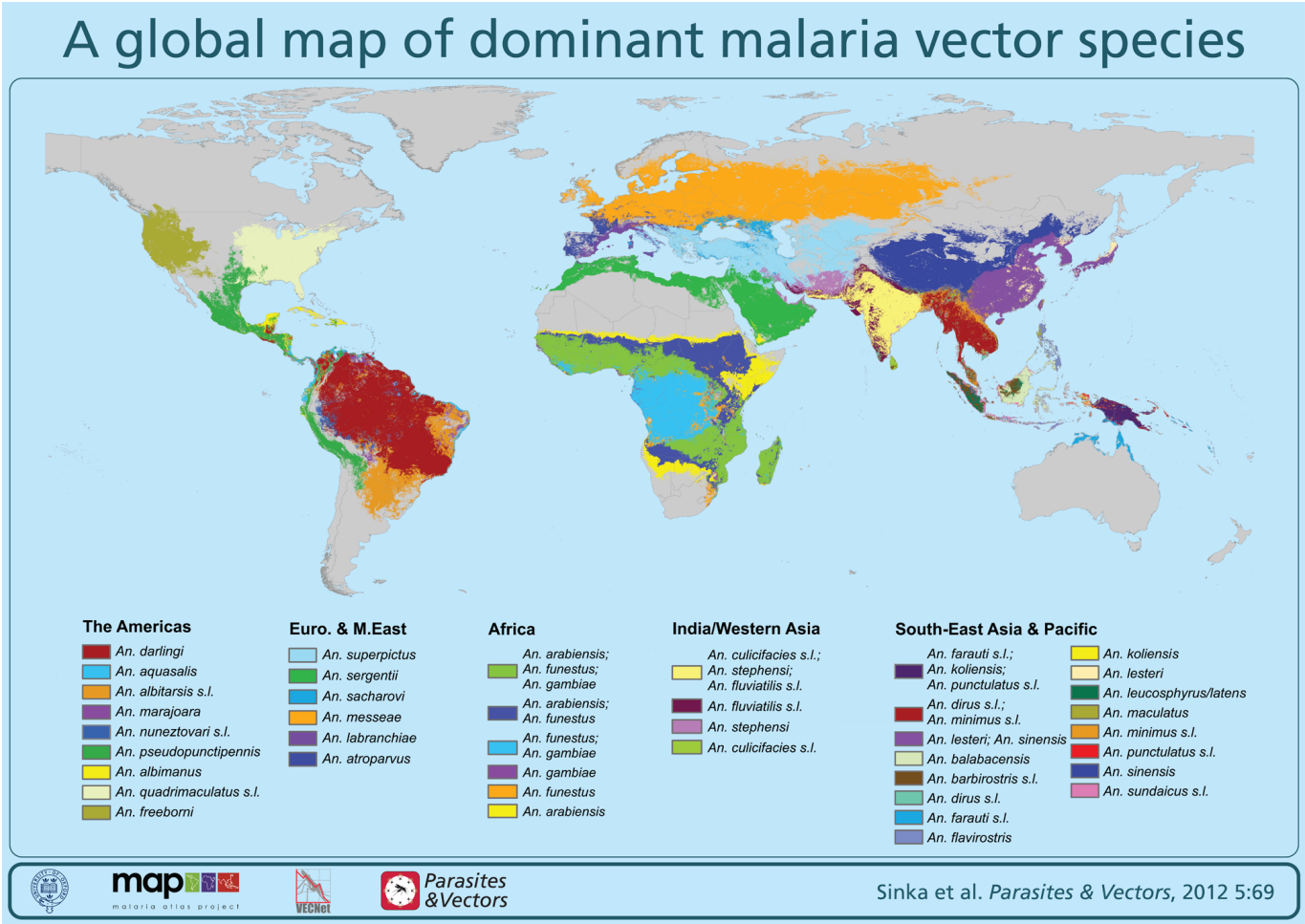


Figure 1-3 Global distribution of *Anopheles* malaria vector species. *Anopheles* species capable of transmitting malaria are found worldwide, even in countries like the United States and the United Kingdom that are no longer endemic for malaria. Figure from Ref 42.

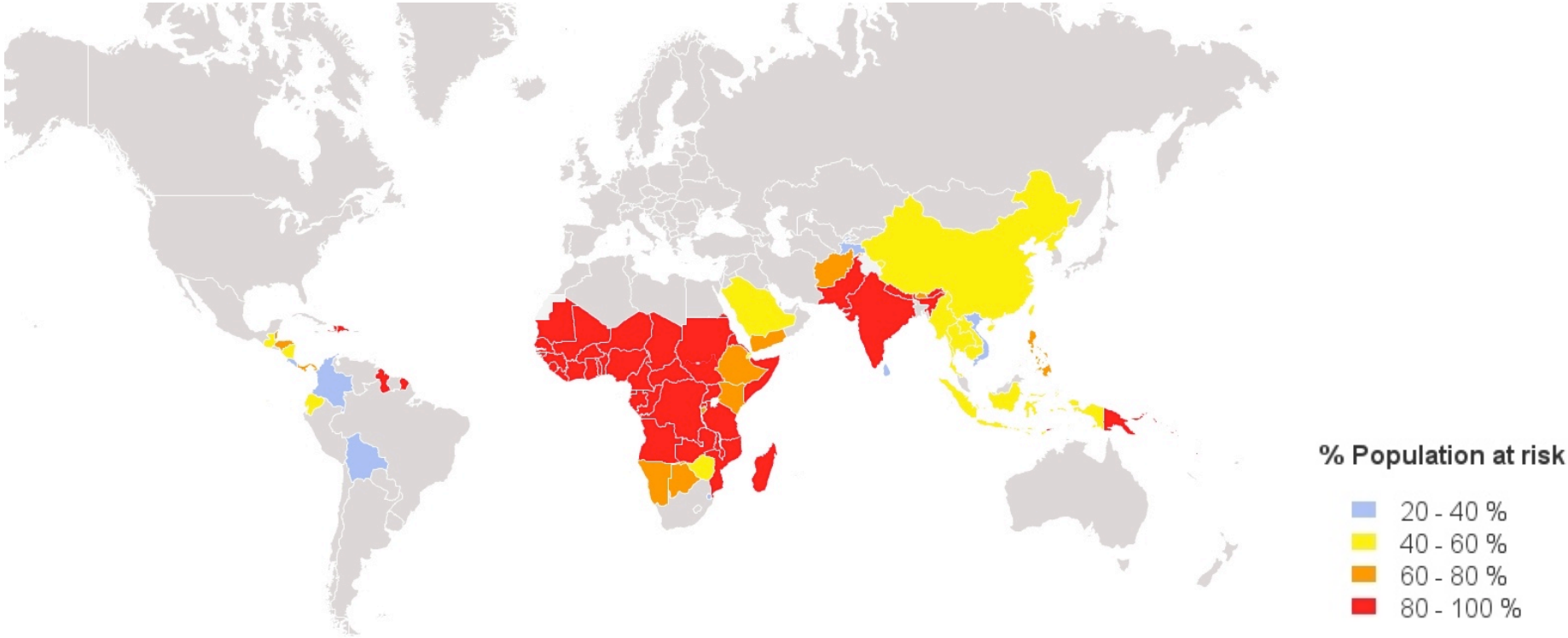


Figure 1-4 Percent of population at risk of malaria worldwide. Most of the world's population at risk of malaria resides in sub-Saharan Africa. Other regions of risk are South America and southeast Asia. Figure from Global Malaria Mapper, Medicines for Malaria Venture http://worldmalariareport.org/node/68 (Visited 10 July 2015).

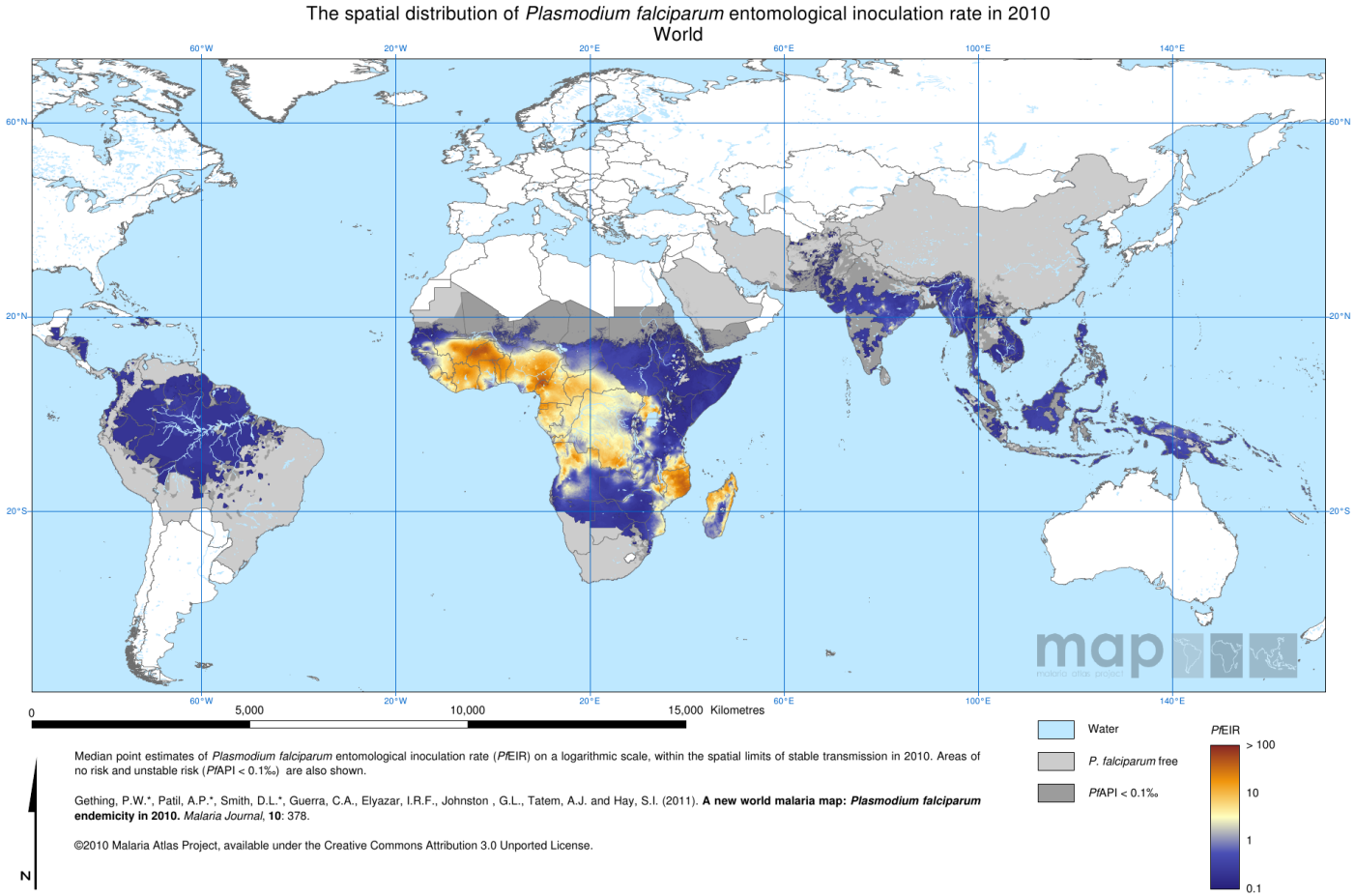


Figure 1-5 Global *P. falciparum* entomological inoculation rates (2010). Estimated number of *P. falciparum*-infected mosquito bites per person per year worldwide. The heaviest region of transmission is in sub-Saharan Africa. Figure from Ref 47.

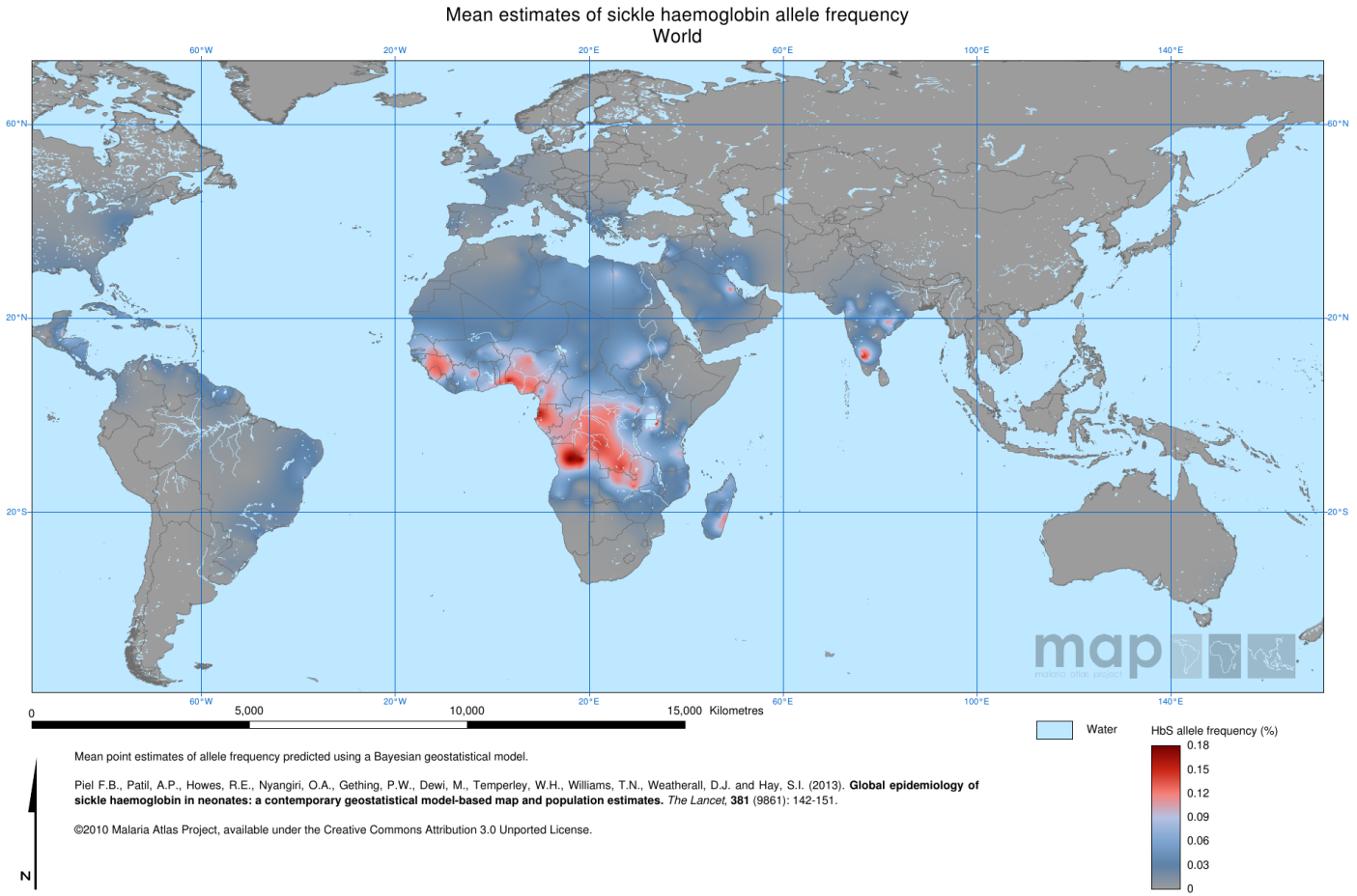


Figure 1-6 Global estimates of HbS allele frequency worldwide. The HbS allele is found predominantly in sub-Saharan Africa, with additional foci in Arabia and India. Figure from Ref 48.

1.2 Chemotherapy and resistance in P. falciparum

Malaria was a scourge to mankind for thousands of years before the causative agent was discovered, so it is no surprise that herbal remedies were devised to combat the illness[75]. One of the earliest recorded descriptions of the disease is a Chinese document from 2700 BCE. As early as 800-400 BCE, scholars were aware of the association between poor health, fevers, and enlarged spleens of those living in marshy areas[76]. At the time, some believed that miasmas arising from swampy land caused the disease, also called ague. Indeed, the term malaria is thought to derive from the Italian *mala aria*, meaning "bad air"[77]. Several centuries before Alphonse Laveran's 1880 discovery of the microbe that caused malaria[78], the Han of China and the Quechua of Peru were using herbal remedies to treat the symptoms of malaria, though almost certainly without intending to treat malaria itself[79]. These two herbal remedies, cinchona bark from Peru and infusions of *Artemisia annua* from China, provided two antimalarials still critical to treatment in the modern era: quinine (QN) and artemisinin (ARM). These compounds also set the basis for the chemical synthesis of a number of other antimalarials, including the 4-aminoquinolines chloroquine (CQ) and amodiaquine (AQ), the 8-aminoquinoline primaquine, and the ARM derivatives artesunate (ART) and OZ-439[80]. Both QN and ARM have a long and fascinating history.

1.2.1 Quinine and chloroquine

Among the Quechua in Peru, infusions of *Cinchona* bark (from which the active compound QN was eventually isolated) were used to treat shivering associated with damp, cool environments[81]. The arrival of Europeans in South America brought Jesuit priests, who became interested in the native remedy. They postulated that the

infusion might also treat the shivering associated with ague[82]. Jesuit missionaries brought the remedy back to Europe, where it became known as Jesuit's powder[79].

Some physicians in Europe successfully treated malaria with infusions of cinchona bark, and over time its popularity grew. In the early days of its use, therapeutic outcomes were inconsistent. Many merchants would sell any bark that tasted as bitter as cinchona, and different species of cinchona varied widely in their level of the active compound QN[83]. It was not until two seminal works were published, *Pyretelogia* by Richard Mortan in 1692[84] and *Therapeutice Speciales* by Francesco Torti in 1712[85], that use of QN-containing cinchona bark infusions became a widely accepted remedy for intermittent fever[86]. The latter work is noteworthy for Torti's depiction of a tree of fevers, with leafy branches indicating fevers that could be treated with bark, and nude branches indicating those that were refractory to treatment[87]. Demand for the cure became so great that some species of cinchona were almost harvested to extinction[88], which would have been a devastating loss for future generations.

Joseph Pelletier and Pierre-Joseph Caventou, two French chemists, extracted QN from Cinchona bark in 1820. A year later, French physicians were successfully using it to treat intermittent fever[89]. The heightened demand for drug led explorers to strike out into the wild to discover the species of *Cinchona* with the highest QN content. Charles Ledger was one such explorer; he sold seeds of *C. ledgeriana* to the Dutch in 1865[90]. With these seeds, the Dutch established massive *Cinchona* plantations in Java, eventually coming to monopolize the world's supply of QN, which had become the treatment of choice for intermittent fever[91].

At the dawn of World War II, it was clear to American generals that because the Japanese controlled the world's supply of QN, an alternative would have to be found if American soldiers were to fight in malarious regions[92]. Thus began the

curious story of CQ, discovered by the Germans, shared with the French, rediscovered by the Americans, and finally found to be easier to synthesize, more effective, and less dangerous than other antimalarials. Chemists at Bayer Pharmaceuticals in Germany first synthesized CQ, then named resochin, in 1934[93]. Preliminary studies in lower animals appeared to show that resochin was too toxic for human use, and so the compound was patented and shelved[92]. Through a series of business arrangements, the instructions for the synthesis of resochin were shared with both a French and an American pharmaceutical company. It was not until the early 1940s that scientists working for the US Government "rediscovered" CQ (then called SN 7618) as part of a massive 16,000-compound effort to find new antimalarials. Early success with SN 7618 in animal models led to successful human tests. It was at this point that it was discovered that the compound had already been synthesized, and that time and resources had been wasted. Nevertheless, the huge success of the human trials led to standardization of dose and treatment regimens in 1945[92,94]. For the next several decades, CQ became the drug of choice for the treatment of malaria.

Programs of CQ prophylaxis and vector control were so successful that scientists began to envision malaria eradication[95]. One such scientist was Mario Pinotti of Brazil, who implemented a clever form of mass drug administration by adding CQ to table salt. Pinotti's method was employed throughout South America, and in parts of Africa and Asia[96]. While initially effective in reducing the malaria burden in the countries where it was employed, by the 1960s evidence of CQ resistance had begun to emerge[97–99].

The spread of CQ resistance was devastating, with the number of malaria-related deaths doubling compared with prior to the emergence of resistance[100]. Genetic analysis has demonstrated that the mutation conferring CQ resistance has

independently arisen at least five times between 1957 and 1970[101,102]. It spread to Africa in 1978, and over the next two decades came to every endemic country on the continent[103] (Figure 1-7).

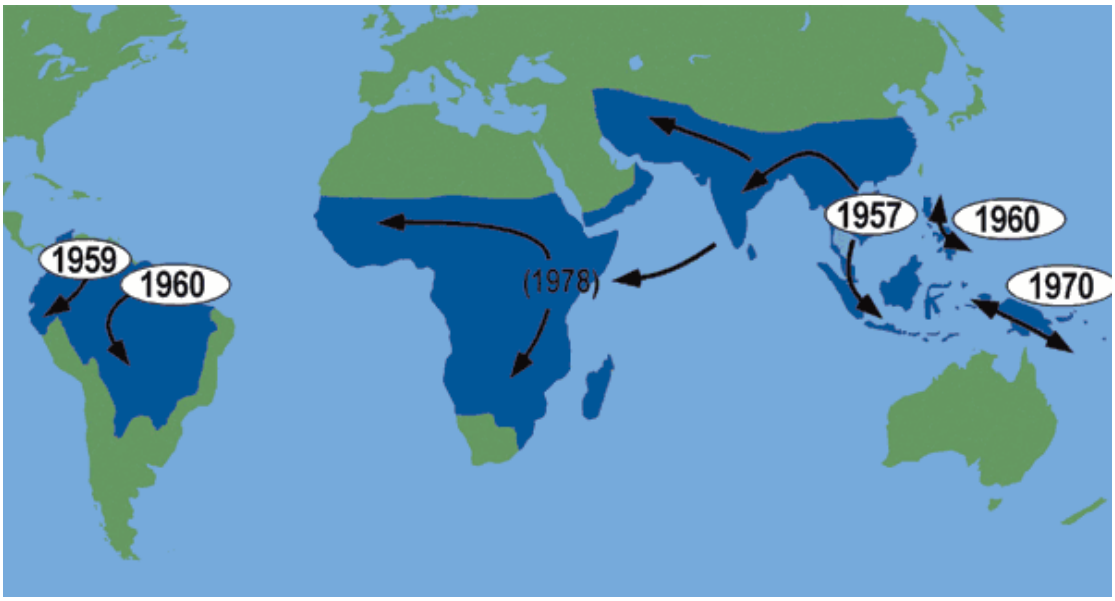


Figure 1-7 Origin and spread of CQ resistance. CQ resistance is postulated to have multiple different origins, with the earliest being in 1957 in southeast Asia. By 1978, it had spread to sub-Saharan Africa with devastating effects. Figure from Ref 104.

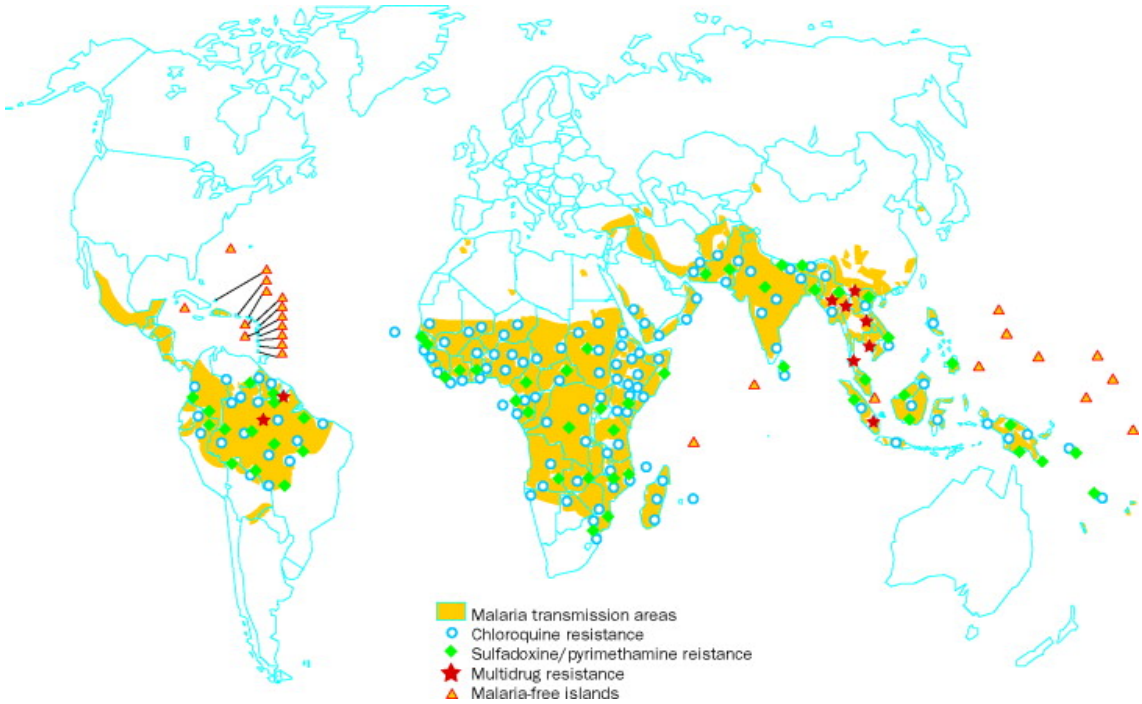


Figure 1-8 Global distribution of antimalarial resistance. Parasites with reduced sensitivity to CQ (blue circle) and sulfadoxine (SUL)/pyrimethamine (PYR) (green diamond). Figure from Ref 103.

1.2.2 Artemisinin and its derivatives

In China, infusions of the qinghao plant (*Artemisia annua* L.), from which ARM was derived, were used for thousands of years to treat a myriad of illnesses[105]. The earliest reference to using *A. annua* for treatment of malaria symptoms appeared in Ge Hong's (ca. 284 - 346 CE) *A Handbook of Prescriptions for Emergencies*[105,106]. It was over a millennium later before scientists in China provided empirical proof of its efficacy[107].

As with CQ, conflict was the catalyst for the discovery of ARM, and until recently, its history remained largely unknown in the West. The Vietnam War saw both American and Vietnamese troops stationed in an epicenter for the spread of CQ-resistant *P. falciparum*. The danger posed by this drug-resistant pathogen spurred research into novel antimalarials on both sides[106]. On the American side, the

result was mefloquine (MFQ), a drug capable of a single-dose cure of CQ-resistant *P. falciparum*[108]. The North Vietnamese, however, lacked the infrastructure to discover and develop new medicines, so they turned to their Chinese allies for assistance. This partnership resulted in the creation of Project 523, a consortium of hundreds of scientists with two main goals: 1) synthesis of new antimalarial drugs; and 2) research into and testing of traditional Chinese herbal remedies for antimalarial activity[106,107].

In pursuit of the second goal, researchers led by Tu Youyou carried out a large-scale study of ancient texts to identify traditional herbal remedies with possible antimalarial activity[109]. The group identified more than 2,000 Chinese herb preparations, and tested over 380 extracts in a mouse malaria model. Following years of effort, an extract of *A. annua* L. showed a promising level of inhibition in the mouse model. However, confirmatory screening failed to replicate this result. Puzzled, Tu went back to the source document, Ge Hong's *Handbook*. Rereading the primary source led Tu to postulate that the heat used in conventional extraction may have been degrading the active compound. Indeed, the low-temperature extract proved to be more effective. By 1972, Tu and her colleagues were able to extract and identify ARM as the active antimalarial component of *A. annua* L[107].

Hampered by the ongoing Cultural Revolution at the time, Tu and her colleagues had a difficult time testing ARM, and an even greater challenge in disseminating their work to the global community. With no formal protocol for initiating clinical trials in humans in place at the time, the investigators ingested ARM themselves to prove it was safe. Following these "safety" tests, the investigators received permission to begin field trials against *P. vivax* and *P. falciparum*. ARM proved to be more effective than CQ at reducing fever and eliminating parasites from the blood. The structure of ARM was determined in 1975, and in 1979 the

investigators received a National Invention Certificate from the Chinese government, which finally paved the way for sharing their results with the global malaria community. Collaboration with Western investigators began in the early 1980s[110], and it took nearly two decades for ARM combination therapy (ACT) to become the World Health Organization-recommended standard of care in Africa[111].

By the late 2000s, evidence of reduced efficacy of ARM began to appear in Cambodia[112–115]. In what is feared to be a harbinger of full-blown resistance, patients treated with ARM-based therapy were taking longer to clear parasites from their blood[116–118]. Hoping to prevent a repeat of the spread of frontline-drug resistance from southeast Asia to Africa, the global community was quick to act. The Tracking Resistance to Artemisinin Collaboration (TRAC) study was established at sites across Asia and Africa[119]. As a result of this study and of the pioneering work of Ariey et al., a polymorphism in the *kelch13* propeller domain was identified that correlated with the delayed-clearance phenotype observed in Cambodian patients[120]. It was hoped that this molecular marker would allow for surveillance and containment of artemisinin resistance within southeast Asia, as its spread to Africa would have dire consequences[121].

Further work exploring the nature of ARM resistance focused on the mechanism of resistance. It was established that multiple polymorphisms in the *kelch13* domain were capable of increasing parasite survival in response to ARM[120], indicating that the parasite is tolerant of perturbations in this protein,. A transcriptomics study of parasites isolated from patients associated ARM resistance with increased expression of genes involved in the unfolded protein response, as well as decelerated progression through the first stage of the intraerythrocytic cycle[122]. The authors postulated that ARM-resistant parasites are “primed” to deal with the protein damage caused by ARM[123]. Another group has postulated that

polymorphisms in *kelch13* prevent it from interacting with *P. falciparum* phosphatidylinositol-3-kinase (PfPI3K)[124]. They hypothesized that disruption of this interaction prevented degradation of PfPI3K, leading to higher levels of its lipid product phosphatidylinositol-3-phosphate (PI3P). They correlated levels of PI3P with parasite survival in response to ARM. Work is ongoing to better characterize the mechanism of parasite resistance to ARM, which will be critical to the development of new antimalarials.

The stories of CQ and ARM highlight two important facets in the fight to reduce malaria morbidity and mortality: 1) the need for a continuous robust pipeline for novel antimalarial drugs is needed; and 2) the importance of molecular tools with which to monitor and prevent the spread of drug-resistant parasites.

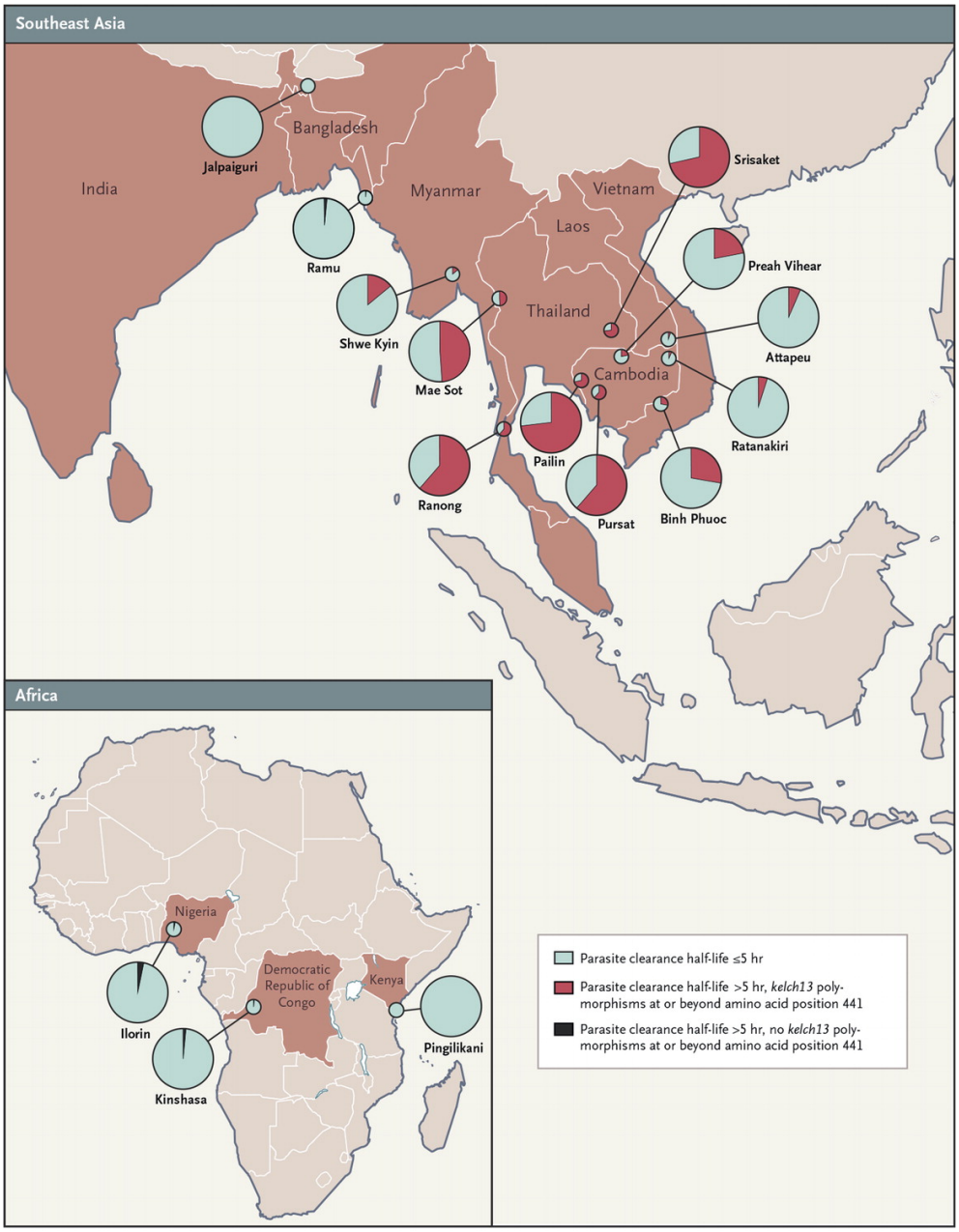


Figure 1-9 Distribution of TRAC sites in southeast Asia and Africa. The TRAC study showed that in some parts of southeast Asia, most parasites were slow-clearing *kelch13* mutants (red). Fortunately, at the conclusion of the study in 2013, no slow-clearing *kelch13* mutant parasites were identified in the three African sites. Figure from Ref 119.

1.2.3 Treatment of uncomplicated and severe malaria

The vast majority of malaria cases result in an uncomplicated course of disease. Symptoms rarely progress beyond fever and mild anemia, followed by splenomegaly. Treatment involves a course of ACT, which results in a rapid cure and complete recovery[125].

Severe malaria encompasses a broad range of pathologies, including coma, severe anemia, acidosis, hypoglycemia, pulmonary edema, and jaundice[126]. In Africa, severe disease occurs in roughly 0.5% of cases[127]. Of those cases, mortality in patients with the cerebral manifestation is 15-20%, even with treatment. In those patients with multi-organ dysfunction, treated mortality is >30%[128]. Children in Africa who survive cerebral malaria can suffer from persistent language deficits and an increased risk for epilepsy, and have a shorter lifespan[129].

QN used to be the sole treatment for severe disease, despite its narrow therapeutic window and life-threatening side effects (lethal hypotension, sterile abscess, sciatic nerve damage, hyperinsulinemic hypoglycemia[128,130]). The development of ART for treatment of severe malaria resulted in a 34% decrease in mortality in southeast Asia, and a 22% reduction in Africa. However, even with ART treatment, severe malaria is lethal in 14.7% and 8.5% of patients in Asia and Africa, respectively[131,132].

These mortality rates emphasize the urgent need for a better understanding of drug response in *P. falciparum*. Even with ART intervention, nearly 10% of severe malaria cases in Africa are fatal. Children that survive an episode of cerebral malaria are at increased risk for sequelae that reduce quality and length of life. The spread of ARM resistance in Africa would undo much of the progress made since the introduction of ACT, and leave QN as the sole therapy for severe disease.

1.3 *Genetic crosses in* P. falciparum

Genetic crosses in *P. falciparum* provide a powerful platform for investigation of the genetic determinants of a variety of parasite traits[133]. *P. falciparum* is haploid in the blood stage, and recombination only occurs during the sexual stage in the mosquito[7]. Genetic crosses in the laboratory take advantage of this trait by controlling which two parasites recombine to produce offspring. A genetic cross involves four major steps: 1) choosing two parasite lines that differ in a phenotype of interest and culturing gametocytes from those two lines; 2) infecting an *Anopheles* vector with a mixture of gametocytes from the two lines; 3) using the mosquito to infect a chimpanzee; and 4) recovering recombinant blood-stage progeny from the chimpanzee.

Completing a genetic cross requires recapitulating the parasite life cycle in the laboratory. As described above, the *P. falciparum* life cycle is incredibly complex, requiring multiple hosts and many cell types. Though very difficult, it can be completed in the laboratory. *In vitro* gametocyte culture is required, as is access to a secured insectary for infection of *Anopheles* mosquitoes. The most significant requirement, both in terms of cost and expertise, is a chimpanzee (*Pan troglodytes*), which is necessary for the liver stage of the parasite. Because of these difficulties, prior to the 803xGB4 genetic cross described in this thesis, only three other crosses have ever been completed: 3D7xHB3 in 1987[134], HB3xDd2 in 1990[135], and 7G8xGB4 in 2008[136].

The first genetic cross by Walliker et al.[134] was a proof of principle. The selected parental lines 3D7 and HB3 possessed differences in enzymes, pyrimethamine sensitivity, and infected RBC antigens. Following successful infection of the chimpanzee and recovery of blood stage progeny, the authors established that normal Mendelian inheritance operated and that the progeny inherited traits from one

parental line or the other[134]. In the second genetic cross, Wellems et al.[135] sought to characterize the genetic determinant of CQ resistance by crossing a CQ sensitive line with a CQ resistant line. They were able to show that CQ resistance was not linked to *pfmdr1*[135]. Further work identified polymorphisms in *pfcrt* that modulated parasite response to CQ[137,138]. The third genetic cross identified *Rh5*, a major invasion ligand of the parasite. All three of these crosses have proved to be invaluable resources to the research community.

The 803xGB4 genetic cross was the fourth and possibly final cross completed using the conventional method, as the National Institutes of Health has imposed a moratorium on most research involving chimpanzees. Currently, there is no alternative for completing the parasite life cycle, so it is conceivable that there will be no further genetic crosses. The 803xGB4 cross was designed to investigate artemisinin resistance, but given the impending ban on chimpanzee research, it was desired to cross two lines with distinct drug response profiles. Identifying parental lines with differential drug phenotypes increased the likelihood of identifying genetic loci associated with drug response in the progeny.

My contribution to the selection of the parental lines, as well as a detailed methodology for generating a genetic cross and recovering recombinant progeny, can be found in Appendix I.

1.4 Genetic analysis in P. falciparum

A number of tools have been developed to facilitate the study of the *P. falciparum* genome, including microsatellite markers, high-density single nucleotide polymorphism (SNP) arrays, and high-throughput Illumina-based whole-genome sequencing (WGS). These modalities have enabled phenotype-genotype linkage analysis, studies of *P. falciparum* population structure, and genome-wide association studies (GWAS). This thesis describes the use of microsatellite (MS) markers, a high-density SNP array, and WGS to characterize a new genetic cross and make observations about genome structure and genetic loci linked to drug response.

Microsatellites are genetic polymorphisms consisting of short-sequence repeats that can be identified by polymerase chain reaction (PCR) using a fluorescently-labeled primer and capillary electrophoresis[139]. In a *P. falciparum* genetic cross (see Appendix I), a set of microsatellite markers can be selected that are polymorphic between the parental lines. The progeny of the genetic cross can then be characterized by the pattern of MS markers they inherit[140,141]. With a sufficient number of markers, linkage analysis can be performed to identify genomic regions associated with a phenotype of interest[133]. Polymorphic MS can also be used to monitor for cross-contamination in cell culture. A set of 12 markers (described in Section 2.4.1) was used to ensure progeny cultures were clonal (and not cross-contaminated) prior to use in experiments. One advantage of MS markers is that they are fairly ubiquitous throughout the genome[140]. If enough polymorphic MS were identified for a given genetic cross, a potential genetic locus could be narrowed down to reduce the number of candidate genes. Disadvantages include the massive effort required to reach high resolution on each chromosome (each MS requires a separate PCR reaction for each progeny line), and the requirement for a specialized capillary electrophoresis machine.

High-density *P. falciparum* SNP microarrays (SNP arrays) were developed as a result of advances in WGS capabilities, including the publication of the *P. falciparum* genome in 2002[142]. Following the publication of the complete genome, over 60,000 unique SNPs were identified[143–145], growing to over a million less than 15 years later[146]. The first GWAS was performed using 189 drug-phenotyped parasites from Africa and a low-density array (3,257 SNPs)[147]. This study identified several regions under positive selection, likely from drug pressure, and identified several novel drug resistance candidates. Following on these efforts, van Tyne et al. developed a high-density array containing 17,582 validated SNPs[148]. Using this array, the authors were able to make observations on the global structure of the *P. falciparum* genome, and identify a novel locus on chromosome 10 associated with halofantrine (HLF) resistance. In this thesis, van Tyne et al.'s high-density SNP array was used to characterize the parents and progeny of a new genetic cross. A set of polymorphic SNPs differing between the parental lines was identified (described in Appendix I), and used to perform quantitative trait loci (QTL) analysis using data from a high-throughput compound screen (described in Chapters 3 and 4).

Advances in WGS technology and computational methods have transformed the analysis of the *P. falciparum* genome. High-throughput techniques are able to rapidly identify SNPs from clinical blood samples, allowing for analysis of *P. falciparum* genome variation on a global scale[149]. These tools have been used to make observations on the genetic architecture of ARM-resistant *P. falciparum*[150], perform GWAS to detect loci associated with parasite response to 22 drugs[151], and determine the effects of endemicity on selection in the *P. falciparum* genome[152]. The parental lines and progeny of the genetic cross described in this thesis were whole-genome sequenced, and a set of variants identified (described in Chapter 5). WGS

data were used to make observations about copy number variation (CNV) and recombination in this genetic cross, adding to the body of literature on this topic.

1.5 *Drug screening and* P. falciparum

1.5.1 Conventional drug screening methods

In-vitro screening of antimalarial compounds has a long history. One of the earliest semi-automated methods measured uptake of radiolabeled hypoxanthine, an essential nucleic acid in *P. falciparum*, against a dose titration of CQ, QN, MFQ, AQ, primaquine, PYR, and eight novel compounds[153]. This technique enabled the plotting of dose-titration curves, and estimation of the median inhibitory concentration (IC_{50}) for each compound. This quantitative measure was extremely valuable, as it enabled comparisons between different drugs and different parasite isolates to be made. Investigators worldwide began to take advantage of this new technique. It was used to investigate synergism between drugs[154,155], track resistance to existing drugs in the field, and investigate novel compounds. Since publication, this assay has become a standard in the field, and still remains in use[156,157]. Downsides include that it is expensive, time-consuming, and requires the use of radiation and special detection equipment.

Smilkstein et al. developed a modified technique that utilized the DNA dye SYBR Green I. In this assay, parasite growth is assessed by incubating parasites with drug for 72 hours, adding dye to each well, and then measuring the fluorescence level[158]. It is simpler and cheaper than the radiolabeled method, and has been widely adopted[159–164].

One area in which these methods have been applied is QTL analysis[133]. *P. falciparum* only undergoes sexual recombination in the mosquito; in all other stages of its life cycle, it is haploid (see Section 1.1 for a full description of the parasite life cycle). Investigators have been able to take advantage of this to study heritable traits in genetic crosses[165]. A detailed method for performing a genetic cross is presented

in Appendix I. Briefly, two parental lines that differ in a phenotype of interest are selected. Gametocytes from each of the parental lines are cultured and mixed, and then membrane fed to an *Anopheles* mosquito. Sporozoites are allowed to develop, and a susceptible non-human primate host is then infected. Following development of blood stage parasitemia, pools of recombinant progeny are collected and cloned by limiting dilution.

Using the screening methods described above and the parental lines and progeny from a genetic cross, QTL analysis is possible[166]. One of the most powerful applications of this method is in identifying genes linked to drug response. Detailed genetic marker information for the parental lines and progeny is required, along with IC_{50} data. QTL analysis is a statistical technique that identifies a genetic marker linked with variation in the phenotype of interest. Once the chromosomal region containing the marker has been identified, candidate genes can be identified and modified to confirm a role in modulating the phenotype. This technique has been used to identify genes associated with parasite response to numerous compounds, including PYR[167], CQ[137], QN[168], and AQ[169].

1.5.2 High-throughput drug screening

The concept of high-throughput screening developed as a result of the rapid expansion of chemical libraries in the late 1990s[170]. What originally took place in 96-well microtiter plates eventually moved to 384- and then 1536-well plates. Early on, screening on a large scale (~1,000,000 compounds) was sufficiently complex that compounds were only tested at one concentration, leading to low sensitivity and specificity[171]. As technology improved, including accurate small-volume aliquoting and plate-handling robotics, more sophisticated screening methodologies were developed. These rapid advances in screening required changes in the way assays were designed and data analyzed. One important innovation was screening large

compound libraries over a dose titration (Figure 1-10)[172]. With quantitative high-throughput screening (qHTS), each compound in the library is tested at several concentrations, which allows for dose-response curves to be plotted and IC_{50} values to be calculated. This method allows for active compounds to be identified and further classified by the level of their activity.

High-throughput screening methods have been adapted for use in malaria research, and have led to several important discoveries. Plouffe et al. screened over 1.5 million small molecules and identified ~6,000 with potent antimalarial activity (<1.25 μM), and were able to use in-silico methods to identify the mechanism of action of some of those compounds[173]. Weisman et al. took a more targeted approach by screening a smaller library of ~2,500 drug or drug-like compounds. With this screen, they were able to identify 36 novel inhibitors, 19 of which were current therapeutics[174]. Chong et al. screened a different library of ~2,500 pharmaceutical compounds and identified astemizole, an antihistamine, as a potential lead for antimalarial development[175]. Baniecki et al. screened 79,000 small molecules, and identified 181 that were highly active against drug-resistant *P. falciparum*[176]. In perhaps one of the best examples of the promise of this technique, Rottmann et al. identified a new class of antimalarial compounds in a screen of ~12,000 natural products and synthetic compounds. Spiroindolones were shown to be effective in the single nanomolar range against both *P. falciparum* and *P. vivax*, and were capable of a single-dose cure in a mouse model[177], and were recently used to treat *P. vivax* and *P. falciparum* infections in humans[178].

High-throughput screening has also been used to take advantage of the unique features of genetic crosses. Using the progeny from two genetic crosses along with 61 culture-adapted field isolates, Yuan et al. showed that >95% of the observed variation in drug response could be explained by polymorphisms in three

genes: *P. falciparum* chloroquine resistance transporter (*pfcrt*), *P. falciparum* multidrug resistance protein 1 (*pfmdr1*), and *P. falciparum* dihydrofolate reductase (*dhfr*)[179,180]. Use of the genetic crosses allowed Yuan et al. to more precisely map the genetic regions involved in drug response, which would not have been possible using culture-adapted field isolates alone.

Progeny from the 803xGB4 cross were screened against a library containing 2,816 pharmaceutical compounds[181]. This library was selected because it provided a broad range of drugs that are safe for human use, and are thus biologically relevant as potential antimalarials. Use of this library also allowed for comparison with work previously done by Yuan et al. using the other three genetic crosses. Another library option for use in screening was the Open Access Malaria Box, which consists of 400 compounds selected to represent a broad array of chemotypes[182]. Use of this library is possible in the future, as it could yield important insight about parasite response to potential antimalarial drugs.

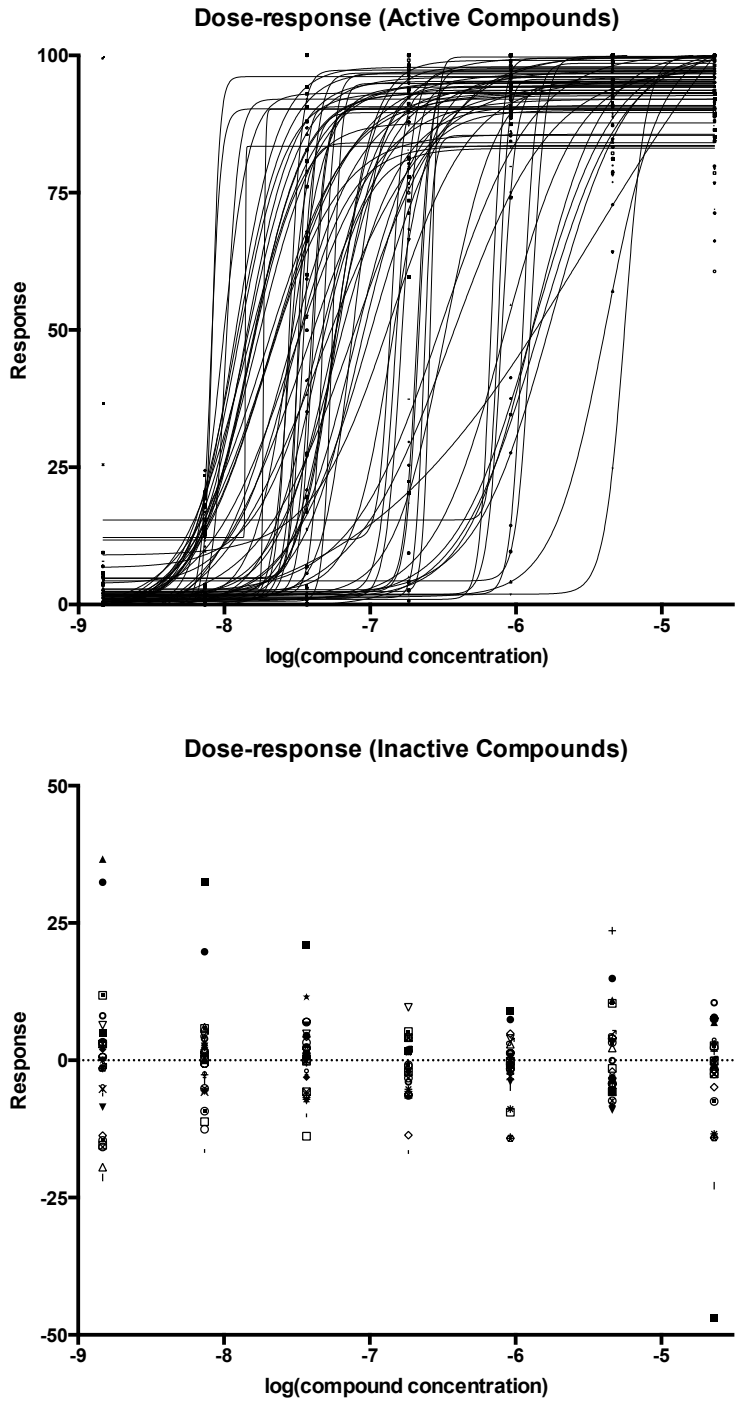


Figure 1-10 Schematic of high-throughput dose titration using data from 803. In the qHTS method, each 1536-well plate holds one concentration of each compound being screened in each well. The serial dilution takes place on a series of plates (represented by lightening of the plate color). Growth-inhibition curves for each compound are determined by grouping corresponding data from each plate, allowing for identification of active and inactive compounds.

1.6 Thesis structure and specific aims

This thesis seeks to identify genetic loci significantly associated with parasite drug response using the 803xGB4 genetic cross. Towards that goal, five Chapters and an Appendix will be presented. Chapter 2 describes the methodologies utilized in the experiments described. Chapters 3 and 4 describe the results and analysis of a high-throughput compound screen of the parental and progeny lines. Chapter 5 presents the most high-resolution analysis of a genetic cross to date. Appendix I details my contribution towards completing the genetic cross between 803 and GB4, including the selection of parental lines, recovery of recombinant clonal progeny, and analysis of those progeny my SNP microarray.

The following specific aims will be addressed:

- Recovery of recombinant progeny from the 803xGB4 genetic cross.
- Identification of highly active and differentially active compounds in a high-throughput screen.
- Identification of genetic loci significantly associated with parasite response to compounds in the pharmaceutical screen using QTL analysis.
- Analysis of WGS data for the parental lines and progeny of the 803xGB4 genetic cross, including identification of potential copy number variants, and characterization of variation and recombination.

2 Methods

2.1 In-vitro cultivation of P. falciparum

2.1.1 Laboratory lines and field isolates

GB4 was cloned from the Ghana III isolate and selected to be highly infectious to *Aotus* monkeys[183]. Cambodian parasites CP0803 (referred to herein as 803) and CP0657-2 were isolated from patients in western Cambodia in 2009[117]. Field isolates KN1068-4 and KN0897-6 were isolated directly from patients in Mali[184]. The laboratory line HB3 was cloned from the Honduran isolate Hondo-1[185]. Line Dd2 was cloned from the MFQ-selected line W2-MEF derived from Indochina III[186].

2.1.2 Conditions for in-vitro cultivation

Parasites were cultivated at 2.5% hematocrit in RPMI 1640 (KD Medical) supplemented with 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 50 µg/ml hypoxanthine, 0.21% sodium bicarbonate, 0.02 mg/ml gentamicin and 1% Albumax II (Invitrogen). Cultures were maintained in sterile, sealed-cap flasks (Corning) in an atmosphere of 5% O_2, 5% CO_2, and 90% N_2 at 37°C.

Whole blood provided by Interstate Blood Bank (Memphis, TN) was depleted of leukocytes by Sepacell filtration (Asahi Kasei Medical), and washed three times with incomplete RPMI prior to use in culture. Blood was stored at 4°C for up to 2 weeks before use.

2.1.3 Sorbitol synchronization

Ring-stage sorbitol synchronization[187] was performed when a culture contained at least 2% rings. Cultures were centrifuged at 800xg for 5 minutes and the supernatant was discarded. The pellet was resuspended in 10 ml of 5% w/v D-sorbitol in 1X phosphate buffered saline (PBS), and incubated at 37°C for 10

minutes. The culture was then centrifuged at 800xg for 5 minutes and the supernatant discarded. The pellet was washed in 10 ml of complete media and returned to culture.

2.1.4 Cryopreservation of parasites

For cryopreservation[188], a parasite culture was centrifuged at 800xg for 5 minutes and the supernatant discarded. Three to five times the pellet volume of sterile Gylcerolyte was added dropwise with gentle mixing. Aliquots of 500 µl were prepared in cryotubes (Nunc), and stored at -80°C for 24 hours to 1 week before transfer to liquid nitrogen for long-term storage.

2.1.5 Thawing of cryopreserved parasites

To revive cryopreserved isolates[188], the frozen pellet was allowed to thaw at room temperature and transferred to a 50 ml conical tube (ThermoFisher). Approximately 0.1X pellet volume of 12% sterile sodium chloride solution was added to the tube drop wise while agitating gently. The pellet was then incubated at room temperature for 5 minutes. Following incubation, 10X pellet volume of sterile 1.6% sodium chloride solution was added dropwise with constant agitation over 3 minutes.

After addition of the 1.6% sodium chloride solution, the tube was centrifuged at 800xg for 3 minutes with a low brake. The supernatant was discarded, and 10X pellet volume of 0.9% sodium chloride and 0.2% dextrose was added dropwise with agitation over 3 minutes. The parasite suspension was then centrifuged at 800xg for 3 minutes with a low brake. The supernatant was discarded. The pellet was resuspended in 5 ml of complete culture media with 200 µl of 50% hematocrit blood.

2.1.6 Cloning of progeny

Cloning by limiting dilution was performed by Prof. John Adams' laboratory at the University of South Florida. Pools of recombinant progeny were diluted and

added to 96-well plates such that the probability of any well containing a single parasite (a clone) was 25%. The wells were monitored by PCR and thin smear until positive and then cryopreserved. Over 400 potential recombinant clones were recovered. Genomic DNA and 200 µl of cryopreserved material of each clone were shipped to the Laboratory of Malaria and Vector Research (LMVR). A total of 70 lines were thawed and expanded, comprising 31 unique recombinants. The first batch of 46 lines was thawed in January 2013. The lines were grown and expanded until sufficient stocks could be cryopreserved for each line (see section 2.1.4). Genomic DNA was then extracted (see section 2.3) for Affymetrix SNP microarray analysis and WGS. A second batch of 24 lines was thawed and expanded in May 2014.

2.2 Assessment of P. falciparum *drug response*

2.2.1 Preparation of compounds for 96-well drug screening

Stock solutions at 5 mM of 12 antimalarials [MFQ, PPQ, LMF, HLF, PYR, AVQ, ART, dihydroartemisinin (DHA), CQ, QN, SUL, and monodesethylamodiaquine (MDAQ)] were prepared in the solvent indicated (Table 2-1), aliquoted, and stored at –80°C. MDAQ was generously provided by S. Ward (Liverpool School of Tropical Medicine). All other chemicals and solvents were purchased from Sigma-Aldrich.

Each compound was tested at 10 two-fold serial dilutions prepared in culture-grade water (Lonza) (Table 2-2). Fifty µl of each drug concentration was added to a 96-well flat-bottomed plate (Costar 3595, Corning) and dried overnight in a laminar flow hood. These assay plates were then stored for up to 1 month in the dark at 4°C.

Drug	MW (g/mol)	Solvent
Mefloquine (MFQ)	414.77	70% EtOH
Piperaquine (PPQ)	999.55	0.5% LA
Lumefantrine (LMF)	528.94	100% MeOH + 0.5% LA
Halofantrine (HLF)	536.88	70% EtOH
Pyrimethamine (PYR)	248.71	70% EtOH
Atovaquone (AVQ)	366.84	100% MeOH
Artesunate (ART)	384.42	70% EtOH
Dihydroartemisinin (DHA)	284.35	70% EtOH
Chloroquine (CQ)	515.86	70% EtOH
Quinine (QN)	324.42	70% EtOH
Sulfadoxine (SUL)	310.33	70% EtOH
Monodesethylamodiaquine (MDAQ)	400.73	70% EtOH

Table 2-1 Compounds and solvents. List of compounds used in the drug assay, their abbreviations, molecular weight (MW), and solvent used. EtOH = ethanol, LA = lactic acid, MeOH = methanol.

Drug	[Low] nM	[High] nM
CQ	2.4	2500
MDAQ	0.7	700
QN	4.8	5000
MFQ	1.2	1250
PPQ	2.4	2500
LMF	0.3	310
HLF	0.3	310
SUL	3.8	7000
PYR	0.5	500
AVQ	0.3	310
ART	0.1	100
DHA	0.1	100

Table 2-2 Compound concentrations used in 96-well screen. Table of compounds and the minimum and maximum concentrations in nM used in the 96-well drug assay. Abbreviations: CQ = chloroquine, MDAQ = monodesethylamodiaquine, QN = quinine, MFQ = mefloquine, PPQ = piperaquine, LMF = lumefantrine, HLF = halofantrine, SUL = sulfadoxine, PYR = pyrimethamine, AVQ = atovaquone, ART = artesunate, DHA = dihydroartemisinin.

2.2.2 96-well drug response assay

A modified version of the previously reported *in-vitro* drug response assay was performed[158]. The method utilizes an intercalating fluorescent dye (SYBR Green I) to measure the amount of DNA in the well, which is an estimate of parasite growth.

Parasite lines were sorbitol-synchronized at the ring stage two cycles prior to the drug assay (see section 2.1.3). Parasite cultures were adjusted to 0.5% parasitemia and 2% hematocrit. Assay plates were brought to room temperature, and 200 µl of the parasite suspension was added to each well. The plates were then incubated for 72 hours at 37°C in a humidified hypoxic chamber (Billups-Rothenberg) containing a mixture of 5% CO_2, 5% O_2, and 90% N_2. Following incubation, the plates were frozen overnight. After thawing, 100 µl of lysis buffer (20 mM Tris-HCl, 10 mM EDTA, 0.16% saponin w/v, 1.6% Triton-X v/v), and 10X SYBR Green I (supplied as 10,000X final concentration by Invitrogen) were added to each well. The plates were then incubated for 2 hours protected from light. Fluorescence intensity was measured using a FLUOstar Optima (BMG Labtech) at 485 nm excitation, 535 nm emission.

Data analysis was performed in Prism 6 (GraphPad Software). Raw fluorescence reads were normalized to the control wells containing no drug. Curves were fitted to the data by nonlinear regression to generate IC_{50} values. Each experiment was performed in three biological replicates. A mean ± standard deviation of three replicate IC_{50}s was reported for each parasite line for each compound.

2.2.3 Pharmaceutical compound library

The National Institutes of Health Chemical Genomics Center (NCGC) pharmaceutical compound library consists of a comprehensive library of drugs approved for human or veterinary use in the USA, Canada, or Japan. The collection was generated to allow for the repurposing of existing therapies[181].

2.2.4 1536-well drug response assay and analysis

The SYBR Green I parasite viability assay was adapted from methods described previously[179,180,173]. Three µl of growth medium were dispensed into 1,536-well black, clear-bottomed plates (Greiner BioOne) by Multidrop Combi

(ThermoFisher). Next, 23 nl of compound or control in DMSO was added to each well by pin tool (Kalypsys). Five µl of parasite suspension (0.3% parasitemia, 2.5% hematocrit) was added to each well. The plates were incubated in hypoxic chambers (Billups-Rothenberg) for 72 hours in a gas mixture of 5% CO_2, 5% O_2, and 90% N_2. Following incubation, 2 µl of lysis buffer (see Section 2.2.2) were added to each well. The plates were then agitated for 30 seconds and incubated protected from light for 1-2 hours. Fluorescence intensity in each well was measured at 485 nm excitation and 535 nm emission wavelengths by the Envision plate reader (PerkinElmer). Each compound in the library was tested at seven five-fold dilutions (29 µM-0.5 nM), with 2.3 µM ARM and 0.23% DMSO serving as positive and negative controls, respectively. Data were analyzed as described below. A set of 384 active compounds was tested in triplicate in a confirmatory screen at 11 three-fold dilutions (29 µM-0.5 nM). ARM and DMSO served as positive and negative controls, respectively.

Plate read data were analyzed as previously described[172]. Briefly, reads were normalized relative to the positive control (ARM) and negative control (DMSO) and corrected by an algorithm using DMSO-only plates run at the beginning and end of each assay. The DMSO plates were used to generate factors to correct for systematic variation. Data were then curve-fitted and assigned a curve class (Table 2-3). Curve class –1 was assigned to compounds with response curves with two asymptotes, and curve class –2 was assigned to compounds with response curves with only one asymptote. Compounds with > 80% efficacy were further categorized as either –1.1 or –2.1, while those with < 80% efficacy were assigned –1.2 or –2.2. Classes –1.1, –1.2, and –2.1 were considered active. Compounds with curve class –3 had single-point activity, and were considered inconclusive. Class 4 compounds were considered inactive, because either they did not have a curve fit, or the curve fit was below three SDs of the mean basal activity.

Table 2-3 Curve class designations.

Curve Class	Asymptotes	Activity
-1.1	2	>80%
-2.1	1	
-1.2	2	<80%
-2.2	1	
-3	-	Single point
4	Inactive	

2.2.5 Optimization of the 1536-well assay

In order to maximize the sensitivity, reproducibility, and robustness of the 1536-well assay, a round of optimization was performed. Two variables were tested: level of O_2 and concentration of ARM (the positive control). Three metrics were used to identify the ideal conditions. The first was the signal-to-background ratio (S/B) (Figure 3-1A). This ratio was calculated by dividing the mean fluorescence unit (MFU) in wells containing only DMSO (negative control) by the MFU in wells containing either 1.15 or 2.3 µM ARM. The second was the coefficient of variation (CV), which was calculated by dividing the SD by the mean of the wells containing either 1.15 or 2.3 µM ARM (Figure 3-1B). A CV of less than 10 was considered optimal for this assay. The third value was the Z-factor (Figure 3-1C)[189]. The Z factor is a dimensionless measure of assay quality. It was determined by the following equation:

$$\boldsymbol{Z = 1 - \frac{(3\sigma_{C+} + 3\sigma_{C-})}{|\mu_{C+} - \mu_{C-}|}}$$

where σ_{C+} is the SD of the positive control (either 1.15 or 2.3 µM ARM), σ_{C-} is the SD of the negative control (DMSO), μ_{C+} is the mean of the positive control, and μ_{C-} is the mean of the negative control.

2.3 Molecular biology techniques

2.3.1 Genomic DNA extraction

Parasites were cultured to ~5% parasitemia, and then centrifuged at 2000xg to collect the RBC pellet. Following aspiration of the supernatant, the pellet was resuspended in 30 ml of TSE (0.1M NaCl, 0.02M Tris-HCl, 0.50M EDTA) and 0.5 ml of 10% saponin. The lysed pellets were washed twice with TSE, then frozen at -20°C. DNA extractions occurred in batches of eight as follows.

The pellet was thawed at room temperature, and then resuspended in 1 ml TSE, with the addition of 50 µl of 0.5M EDTA, 50 µl of 10% SDS, 20 µl of RNAse (500 µg/ml), and 50 µl of proteinase K (10 mg/ml). The solution was then incubated at 37°C for 2 hours. After incubation, 1.5 ml of a solution containing phenol:chloroform (1:1) and 0.25M sodium perchlorate was added. The solution was rocked for 20 minutes.

Following rocking, the solution was centrifuged at 2000xg for 5 minutes. The aqueous layer was then transferred to a new tube and re-extracted with chloroform twice. Following re-extraction, the aqueous phase was transferred to 2 or 3 1.5 ml Eppendorf tubes, and 1 ml of ice-cold 100% ethanol was added to each tube. The DNA was then allowed to precipitate in ethanol overnight at -80°C.

After precipitation, the tubes were centrifuged at maximum speed for 10 minutes. The pellet was washed twice with 80% ethanol, and then allowed to dry at room temperature for 10 minutes.

After the pellet was dry, the DNA was suspended in 200 µl of Tris-EDTA. DNA concentration was measured by Nanodrop spectrophotometer (ThermoFisher), and purity was assessed by the ratio of absorbance at 260 nm and 280 nm.

2.3.2 RNA extraction

RNA was extracted from cultured *P. falciparum* using TRIzol (Life Technologies) according to the manufacturer's protocol. Briefly, 1 ml of TRIzol reagent was added to 200 µl of saponin-lysed culture pellet. The TRIzol mixture was gently agitated until the pellet dissolved, at which point the sample was stored at –80°C until multiple samples could be processed in bulk. The sample was then thawed at room temperature, and 200 µl of chloroform (Sigma) were added. The sample was agitated for 15 seconds, and then centrifuged at 14,000 rpm for 15 minutes at 4°C. Following centrifugation, the aqueous layer was removed to a separate tube. Five hundred µl of 100% isopropanol (Sigma) were added to each aqueous layer, and the sample was incubated for 10 minutes at room temperature. The sample was then centrifuged at 14,000 rpm for 10 minutes at 4°C. The supernatant was discarded, and the pellet was washed in 1 ml of 75% ethanol (Sigma). The sample was centrifuged at 10,000 rpm for 5 minutes at 4°C, and the supernatant discarded. The pellet was dried at room temperature for 10 minutes, resuspended in DEPC treated water (Sigma), and quantified by NanoDrop (Thermo Scientific).

For each microgram of RNA, 0.1X volume of 10X buffer and 1 µl of DNAse I (Life Technologies) were added, up to a volume of 10 µl. The reaction was incubated for 15 minutes at room temperature. DNAse treatment was repeated twice. The reaction was terminated by adding 0.1X 25 mM EDTA and incubating at 65°C for 10 minutes. RNA was then re-quantified by NanoDrop.

2.3.3 *Rh2a*, *Rh2b*, and *Rh6* copy number and transcription level

Primers specific to these *Rh2a*, *Rh2b*, and *Rh6* genes were developed. PCR reaction conditions were as follows. A 25-µl reaction consisted of 12.5 µl of Bioline MyTaq 2x Master mix, 0.2 µM forward and reverse primers, and PCR grade water to

bring the reaction volume to 25 µl. For Rh6, thermal cycling conditions were 94ºC for 5 minutes, 30 cycles of 94ºC for 20 seconds, 48ºC for 20 seconds, and 72ºC for 45 seconds, followed by 72ºC for 7 minutes. For Rh2a and Rh2b, thermal cycling conditions were 4ºC for 5 minutes, 30 cycles of 94ºC for 20 seconds, 46ºC for 20 seconds, and 72ºC for 1 minute 30 seconds, followed by 72ºC for 7 minutes.

The amplification in progeny 36D5, 36E5, and 39C3 was examined by qPCR. A 25-µl reaction consisted of 12.5 µl of 2X Rotor-Gene SYBR-Green I PCR mix, 1 µM forward and reverse primers (Table 2-4), and 50 ng of gDNA. Thermal cycling conditions were 95°C for 5 minutes, followed by 35 cycles of 95°C for 10 seconds and 60°C for 30 seconds (acquiring on the green channel). Results were analyzed using the $\Delta\Delta C_T$ method described in Section 2.4.4. Expression of *Rh2a* and *Rh2b* was assessed using the above protocol with cDNA generated using Reverse Transcriptase (Life Technologies) per the manufacturer's instruction.

Primer name	Sequence (5'-3')	Expected size (bp)
Rh6_F1	ATCCATCACGACAACATCCA	564
Rh6_R1	TGGGGCATGCTGTTATTCTT	
Rh2_SF1	CAGATAATATTAACAAGGCT	
Rh2a_R1	AGGTTTAATATCGACGAGTC	1173
Rh2b_R1	CAAACAACTCCTCCAGC	1779

Table 2-4 Rh PCR Primers. PCR primer sequences and expected product sizes for Rh6, Rh2a, and Rh2b.

2.3.4 Pfmdr1 *copy number and expression level*

Pfmdr1 copy number and expression level were assessed as previously described[190]. Briefly, copy number was assessed with 50 ng of gDNA in a 25-µl reaction consisting of 1X Rotor-Gene SYBR Green PCR mix, and 1µM of forward and reverse primer (Table 2-5). Thermal cycling was performed using the Rotor-Gene Q 5-plex HRM (Qiagen) with a 5-minute incubation at 95°C followed by 35 cycles of 95°C for 10 seconds and 60°C for 30 seconds (acquiring on the green channel). To assess *pfmdr1* expression levels, 50 ng of RNA were added to a 25-µl reaction mixture containing 1X Rotor-Gene probe RT-PCR mix (Qiagen), 0.8 µM forward and reverse primer, 0.2 µM probe (Table 2-5), and 0.25 µl of reverse transcriptase. Thermal cycling was performed using the Rotor-Gene Q 5-plex HRM (Qiagen), with a 10-minute step at 50°C, followed by 95°C for 5 minutes, and 35 cycles of 95°C for 10 seconds, and 60°C for 30 seconds (acquiring on the green channel).

The $\Delta\Delta C_T$ method[191] was used to quantify copy number and expression level of *pfmdr1*. $\Delta\Delta C_T$ is the difference between the ΔC_T of the reference sample (GB4) and the ΔC_T other sample (803 or Dd2). This method involved determining ΔC_T between *pfmdr1* and *pfldh* for each sample, and then calculating the relative copy number or expression level according to the formula:

$$\text{Relative target} = 2^{-\Delta\Delta Ct}$$

	Sequence (5'-3')
pfmdr1 f	TGCCCACAGAATTGCATCTATAA
pfmdr1 r	GACTGTACAAAGGTTCCATTTCGA
pfmdr1 probe	FAM-ACGATCAGACAAAATT-MGB
pfldh f	ACGATTTGGCTGGAGCAGAT
pfldh r	TCTCTATTCCATTCTTTGTCACTCTTTC
pfldh probe	FAM-AGTAATAGTAACAGCTGGATTTACCAAGGCCCCA-TAMRA

FAM: 6-carboxyfluorescein; TAMRA: 6-carboxytetramethylrhodamine

Table 2-5 Primer and probe sequences for *pfmdr1* and *pfldh* RT-PCR.

2.4 P. falciparum *genetic analysis*

2.4.1 Microsatellite genotyping

A set of 12 polymorphic MS markers was routinely used to monitor parasite cultures for contamination (because asexual stages are haploid, each clone will present a single band size for a given MS marker). These 12 markers were selected because they could robustly and reliably differentiate between the 803 and GB4 alleles. The protocol for MS genotyping was as follows.

A 25-µl reaction consisted of 1X Amplitaq Gold Master Mix (Life Technologies), 0.2 µM each of a 6-FAM-labeled forward primer and a reverse primer pair (synthesized by Eurofin Genomics; Table 2-6), and 1 µl of template gDNA (20-100 ng/µl).

The PCR cycle was as follows: 94°C for 2 minutes, followed by 42 cycles of 94°C for 20 seconds, 45°C for 10 seconds, 42°C for 10 seconds, and 60°C for 30 seconds, and a final extension at 60°C for 5 minutes. Following the PCR reaction, each sample was diluted 1:10 or 1:100 in water, and 1 µl of the product dilution was added to 10 µl of formamide and 0.2 µl of ROX Size Ladder (Life Technologies). The formamide mixture was then run on the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). MS band sizes were compared to the sizes listed in Table 2-6.

Marker	GB4	803	Δ Size	Forward Primer (5' to 3')	Reverse primer (5' to 3')
TAA81	120	114	6	TTTCACACAACACAGGATT	TGGACAAATGGGAAAGGATA
TAA87	99	114	15	ACATGTTCATATTACTCAC	AATGGCAACACCATTCAAC
TA60	89	82	7	AAAAAGGAGGATAAATACAT	TAGTAACGATGTTGACAA
TA1	170	167	3	TTTTATCTTCATCCCCAC	CCGTCATAAGTGCAGAGC
Pfg377	97	90	7	TTATCCCTACGATTAACA	GATCTCAACGGAAATTAT
1451458	213	200	13	GAGCATTATAAAATTGGCTA	AAAAAGAAAAGAGATGAACA
C3M54	215	264	49	AATATAATCATAAAGTCGTAC	CTAAGAGAAAAAATGGCTAT
C4M30	160	184	24	ATTGATGCTTTGTCTAATTAG	ATGACAAAACATGGTATGTA
753242	214	216	2	TGAAGTAGGGGATTAAAAA	AACCATATCAAAAATTTCAA
2549455	180	176	4	AAATATATGTTCTACTTTCAATCA	GAGGAAATAATATAATACACCA
401780	103	80	23	GAAACATAAAGGGATGTGTA	ACTTAGAAGAAATTCAATGC
299812	80	97	17	TAATATTCCTTACCTAACACAT	CGTTCACAATTTTATTTAATC

Table 2-6 MS primers. Name, parental band size, and forward and reverse primer sequence of 12 microsatellite markers used to uniquely identify each progeny line. Microsatellite band sizes are listed in base pairs. Forward primers carry a 6-FAM label on their 3' end.

Marker	Chr	Location (kb)	GB4	803	38G5	36D5	40E7	44E8	43A6	24G11	11C2	36E5	76H10	85D3	46G9	39C5	39C3	36H9	43H3	87E7	48C1
TAA81	5	1214	G	8	8	8	8	G	G	G	G	8	G	G	8	8	G	8	G	8	G
TAA87	6	374	G	8	8	8	G	8	8	8	G	8	G	8	8	8	8	G	8	G	8
TA60	13	2486	G	8	8	8	8	G	G	G	8	8	8	G	8	8	G	G	8	G	8
TA1	6	900	G	8	G	G	G	G	G	8	8	8	G	8	8	8	8	G	8	G	G
Pfg377	12	2046	G	8	G	8	8	G	G	G	G	G	8	8	G	8	8	G	G	8	8
1451458	14	1452	G	8	8	G	8	8	G	8	8	8	8	G	8	G	G	G	8	G	8
C3M54	3	909	G	8	8	G	8	8	G	G	8	G	G	G	8	G	G	G	G	8	G
C4M30	4	1087	G	8	G	8	8	G	G	8	G	8	G	G	8	8	8	8	8	G	8
753242	6	706	G	8	G	8	8	G	8	8	G	G	G	G	8	8	8	G	8	G	G
2549455	14	2550	G	8	G	8	8	8	G	8	8	G	8	8	G	8	G	8	G	8	8
401780	10	402	G	8	G	8	8	8	8	G	G	8	8	G	G	G	G	G	8	G	8
299812	8	1047	G	8	G	8	G	G	8	G	8	8	8	8	8	8	G	G	G	8	G

Table 2-7 Inheritance patterns of 17 progeny. Twelve microsatellite markers used to uniquely identify each of 17 progeny. Yellow fill indicates that the progeny inherited the GB4 allele; green fill indicates that the progeny inherited the 803 allele. Each progeny has a unique inheritance pattern, which allows for its identification. The presence of multiple bands for any marker indicates that the culture has been cross-contaminated and that any results must be discarded.

2.4.2 Analysis of Affymetrix array data

Genomic DNA was extracted (see section 2.3) from 70 progeny clones and the two parental lines for high-density Affymetrix-based SNP array hybridization. The selection of these clones was based on microsatellite genotyping results indicating that these clones were non-identical, meaning that they differed by at least one microsatellite marker. Array hybridization and genotype calling was performed by Dr. Steve Porcella's laboratory (Rocky Mountain Laboratory Genomics Unit, NIH). The array contained 74,656 markers, of which 17,582 were validated by whole genome sequence data[148]. Genotype calls were produced using the BRLMM-P algorithm[192].

Filtering of the 17,582 high-quality SNPs was performed as follows: SNPs that were non-polymorphic between 803 and GB4 were eliminated, as were SNPs which produced no signal in ten or more progeny and SNP calls that indicated non-parental genotypes (Mendelian error). The raw data contained many apparent double-crossovers occurring within small DNA segments (<10 kb), especially within the hypervariable regions. Often, multiple progeny possessed the same double-crossovers within a short window of markers. Assuming a recombination rate of 1% per 10 kb[193–195], the probability of having two crossovers in two consecutive marker intervals of roughly 10 kb is ~0.01. It follows that for a marker window of approximately 50 kb, the probability of a double crossover is 0.05. For another progeny to have a double crossover at exactly the same two markers is 0.01 x 0.01 = 0.0001. The probability that two progeny would have the same marker pattern within a 50-kb region is therefore approximately 5×10^{-6} (0.05 x 0.01 x 0.01). These apparent double-crossovers were assumed to be errors, and the intervening SNPs were eliminated.

2.4.3 WGS alignment and quality check

WGS data were generated by Illumina HiSeq using a PCR-free library preparation in the Sequencing Core of the Wellcome Trust Sanger Institute and then processed as follows (Figure 2-1). Raw FASTQ files were mapped to Pf3D7_v3 and formatted to create BAM files[196]. The Genome Analysis Toolkit UnifiedGenotyper was used to call variants, generating a raw VCF[197]. Raw VCF data were quality checked to ensure there were no issues in sample or data processing. A first-pass filter was applied to limit the analysis to the core genome (Table 2-8). A further filter was applied to limit the dataset to variants that were polymorphic between the parental lines. Variant quality score recalibration was performed to generate an analysis ready VCF (see section 2.4.5).

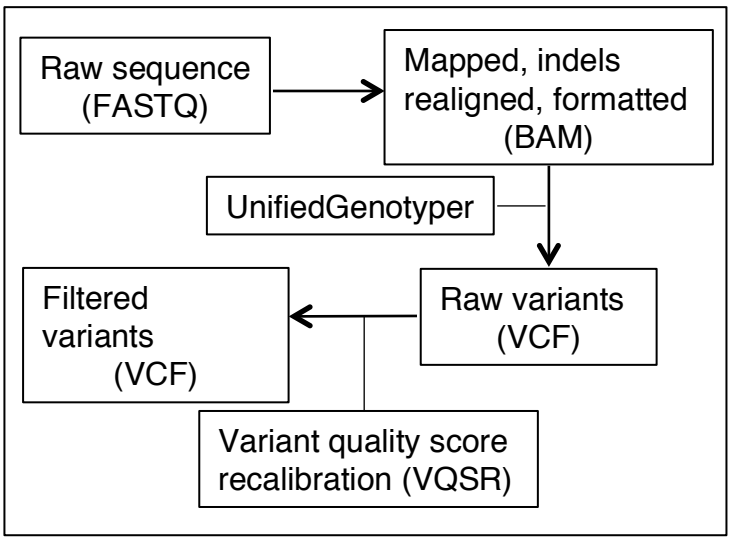


Figure 2-1 Processing of raw sequence data. Outline of the process required to process raw sequence in FASTQ format to an analysis ready callset consisting of high-quality SNPs and indels. Raw sequence data in FASTQ format are converted to BAM format. UnifiedGenotyper is run to generate a set of raw variants against the 3D7_v3 genome. VQSR is then performed to generate a filtered set of variants that can be used for further analysis.

Chr 1	1 - 27336	Subtelomeric Repeat
	27337 - 92900	Subtelomeric Hypervariable
	92901 - 457931	Core
	457932 - 460311	Centromere
	460312 - 575900	Core
	575901 - 616691	Subtelomeric Hypervariable
	616692 - 640851	Subtelomeric Repeat
Chr 2	1 - 23100	Subtelomeric Repeat
	23101 - 105800	Subtelomeric Hypervariable
	105801 - 447300	Core
	447301 - 450450	Centromere
	450451 - 862500	Core
	862501 - 925850	Subtelomeric Hypervariable
	925851 - 947102	Subtelomeric Repeat
Chr 3	1 - 34268	Subtelomeric Repeat
	34269 - 70630	Subtelomeric Hypervariable
	70631 - 597816	Core
	597817 - 600275	Centromere
	600276 - 1003060	Core
	1003061 - 1040961	Subtelomeric Hypervariable
	1040962 - 1067971	Subtelomeric Repeat
Chr 4	1 - 26511	Subtelomeric Repeat
	26512 - 91420	Subtelomeric Hypervariable
	91421 - 545800	Core
	545801 - 614900	Internal Hypervariable
	614901 - 642003	Core
	642004 - 644529	Centromere
	644530 - 935030	Core
	935031 - 983080	Internal Hypervariable
	983081 - 1143990	Core
	1143991 - 1182297	Subtelomeric Hypervariable
	1182298 - 1200490	Subtelomeric Repeat
Chr 5	1 - 18750	Subtelomeric Repeat
	18751 - 37900	Subtelomeric Hypervariable
	37901 - 455740	Core
	455741 - 457252	Centromere
	457253 - 1321390	Core
	1321391 - 1342974	Subtelomeric Hypervariable
	1342975 - 1343577	Subtelomeric Repeat
Chr 6	1 - 610	Subtelomeric Repeat
	611 - 72350	Subtelomeric Hypervariable
	72351 - 478652	Core
	478653 - 480971	Centromere
	480972 - 723117	Core
	723118 - 742800	Internal Hypervariable
	742801 - 1294830	Core
	1294831 - 1384651	Subtelomeric Hypervariable
	1384652 - 1418242	Subtelomeric Repeat
Chr 7	1 - 18000	Subtelomeric Repeat
	18001 - 77100	Subtelomeric Hypervariable
	77101 - 508360	Core
	508361 - 605650	Internal Hypervariable
	605651 - 809245	Core
	809246 - 811716	Centromere
	811717 - 1381600	Core
	1381601 - 1428410	Subtelomeric Hypervariable
	1428411 - 1445207	Subtelomeric Repeat
Chr 8	1 - 19100	Subtelomeric Repeat
	19101 - 73560	Subtelomeric Hypervariable
	73561 - 299079	Core
	299080 - 301403	Centromere
	301404 - 427430	Core
	427431 - 467340	Internal Hypervariable
	467341 - 1365730	Core
	1365731 - 1445690	Subtelomeric Hypervariable
	1445691 - 1472805	Subtelomeric Repeat
Chr 9	1 - 17955	Subtelomeric Repeat
	17956 - 79100	Subtelomeric Hypervariable
	79101 - 1242137	Core
	1242138 - 1244483	Centromere
	1244484 - 1473560	Core
	1473561 - 1505792	Subtelomeric Hypervariable
	1505793 - 1541735	Subtelomeric Repeat
Chr 10	1 - 26240	Subtelomeric Repeat
	26241 - 68970	Subtelomeric Hypervariable
	68971 - 1571815	Core
	1571816 - 1652190	Subtelomeric Hypervariable
	1652191 - 1687656	Subtelomeric Repeat
Chr 11	1 - 21990	Subtelomeric Repeat
	21991 - 110000	Subtelomeric Hypervariable
	110001 - 831968	Core
	831969 - 834245	Centromere
	834246 - 2003320	Core
	2003321 - 2037033	Subtelomeric Hypervariable
	2037034 - 2038340	Subtelomeric Repeat
Chr 12	1 - 14780	Subtelomeric Repeat
	14781 - 60300	Subtelomeric Hypervariable
	60301 - 766654	Core
	766655 - 780450	Internal Hypervariable
	780451 - 1282773	Core
	1282774 - 1285067	Centromere
	1285068 - 1688600	Core
	1688601 - 1745530	Internal Hypervariable
	1745531 - 2163700	Core
	2163701 - 2251150	Subtelomeric Hypervariable
	2251151 - 2271494	Subtelomeric Repeat
Chr 13	1 - 19160	Subtelomeric Repeat
	19161 - 74413	Subtelomeric Hypervariable
	74414 - 1168127	Core
	1168128 - 1170425	Centromere
	1170426 - 2791900	Core
	2791901 - 2894620	Subtelomeric Hypervariable
	2894621 - 2925236	Subtelomeric Repeat
Chr 14	1 - 1344	Subtelomeric Repeat
	1345 - 35774	Subtelomeric Hypervariable
	35775 - 1071523	Core
	1071524 - 1075089	Centromere
	1075090 - 3255710	Core
	3255711 - 3291511	Subtelomeric Hypervariable
	3291512 - 3291936	Subtelomeric Repeat

(Basepair positions according to Pf_3D7_v3)

Table 2-8 Core and hypervariable regions. Position in bp for subtelomeric repeats, core, and hypervariable regions for each of the 14 chromosomes.

2.4.4 Assessing Copy number variation

Copy number variation was inferred by fitting a hidden Markov model as previously described (Miles et al., in preparation). Briefly, average depth of coverage along each chromosome was calculated in 300-bp windows filtered for GC content greater than 20%. Coverage was then normalized to the mean coverage in the core region of chromosome 14 (Table 2-8) with GC > 20%. A Gaussian hidden Markov Model was fit to the normalized coverage data to infer CNVs. Results from the model were then plotted in a grid to allow for visualization of possible CNVs on each chromosome. A tutorial (kindly provided by Alistair Miles) for assessing copy number variation using a Hidden Markov Model can be found at:

http://nbviewer.ipython.org/github/alimanfoo/hmmcnv/blob/master/tutorial.ipynb

(Accessed 8 July, 2015).

2.4.5 Variant Quality Score Recalibration

The variant quality score recalibration (VQSR)[197] tool was part of the Genome Analysis Toolkit Version 3.3 provided by the Broad Institute (https://www.broadinstitute.org/gatk/ Accessed 10 July 2015). VQSR generated a metric for estimating the reliability of each variant in the callset. VQSR functions by creating a continuous, covarying estimate of the relationship between variant call annotations and the likelihood that a given variant is a true variant versus an artifact. Recalibration was performed separately for SNPs and indels.

The variant annotations used to recalibrate the SNPs were QD (Variant confidence / unfiltered depth of non-reference samples), MQ (root mean square of mapping quality of reads across all samples), DP (total depth of reads that passed quality control), and ABHom (proportion of reference alleles). Indels were recalibrated using the annotations QD, MQ, DP, BaseQRankSum (U-based, z-approximation from the Mann-Whitney Rank Sum Test for base qualities), and FS (phred-scaled p-value

using Fisher's Exact Test to detect strand bias). VQSR utilized a positive training set of variants (SNPs or indels) that had been previously identified as being high quality from the previous three crosses, and a negative training set of low quality variants. (Miles et al., in preparation)

The score generated by VQSR is called VQSLOD, the log odds ratio of the probability that a variant is true versus false under the trained Gaussian mixture model. VQSLOD thresholds for SNPs and indels were then set to filter out low-confidence variants from the callset.

2.4.6 Generation of linkage maps

The physical position of each polymorphic variant in the SNP Chip array data was determined from the reference genome Pf3D7_v3 from PlasmoDB (http://www.plasmodb.org). A Haldane mapping function was used to calculate the mapping unit (centiMorgan) for the 14 chromosomes according to the following equation:

$$r = \frac{1}{2}\left(1 - e^{-2\lambda d}\right)$$

where $0 \le r \le 0.5$ is the recombination fraction, λ is the genetic distance (Morgan), and d is the physical distance (in kilobases)[198].

2.4.7 Quantitative trait loci analysis

QTL analysis was performed in R using the R-package r/qtl[166,199] as previously described (Figure 2-2)[179,180,169]. The log(IC_{50}) for each compound was used as the phenotype. Cleaned Affymetrix SNP array data were used for the genotypes. QTL analysis generated logarithm of the odds (LOD) scores, which are the $\log_{10}$ of the likelihood ratio comparing probability that a QTL exists at a given marker against the probability that no QTL exists anywhere in the genome. P-values were assigned to LOD scores by performing a permutation test with 10,000 permutations.

For the primary screen, QTL analysis was performed for each compound in the library. A low false-discovery-rate (FDR)-corrected LOD threshold of 1.3 (FDR $p < 0.05$) was used to maximize the likelihood of including compounds with evidence of linkage in the confirmatory screen.

For the confirmatory screen, each parasite line was tested against the 384 compounds in triplicate (three replicate plates performed on the same assay day). QTL analysis was performed using each of 384 phenotypes and the microsatellite array genotype data. Compounds showing LOD scores above 3.5 were identified for each of the 14 chromosomes and p-values were generated by permutation test.

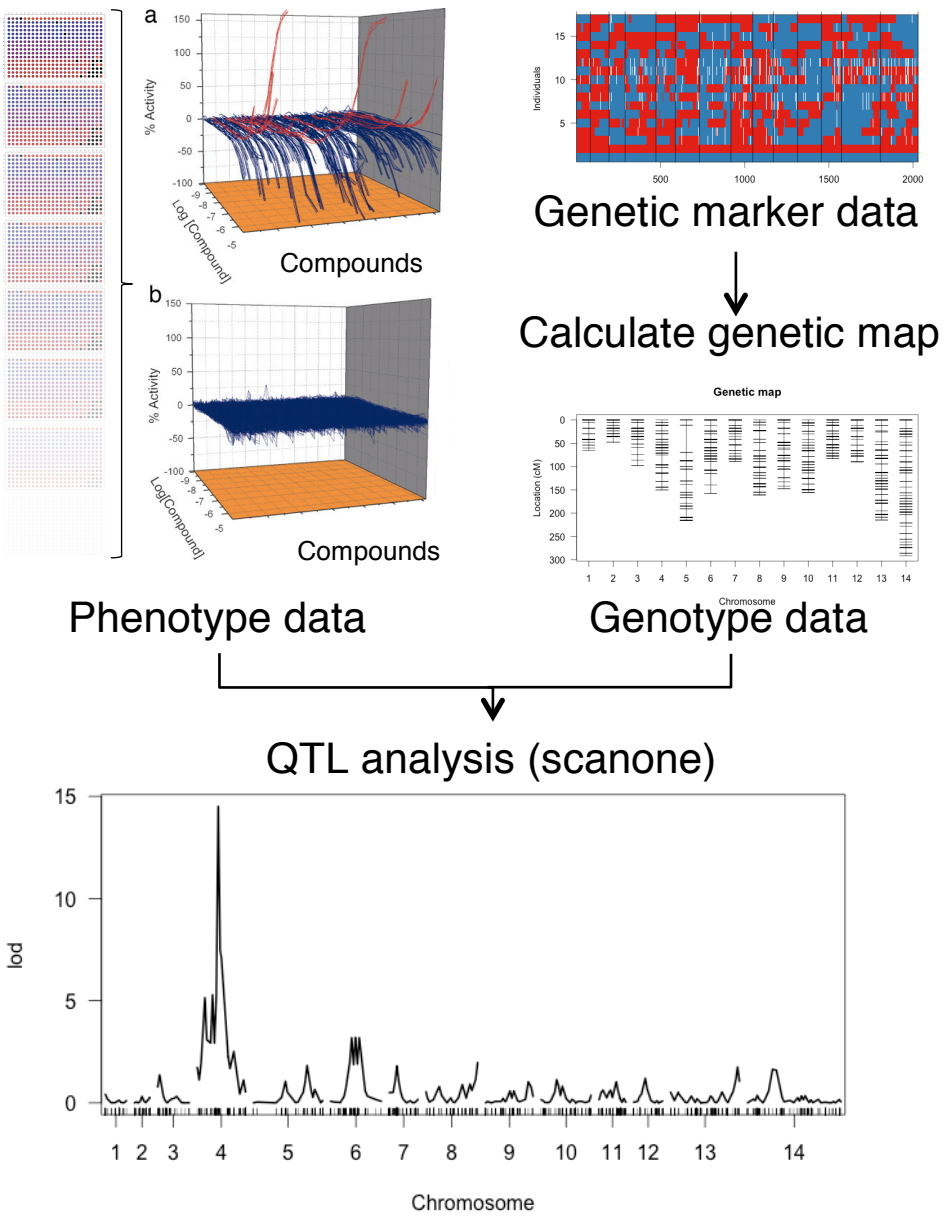


Figure 2-2 Schematic depicting QTL analysis. QTL analysis required detailed genetic marker data. These data were used to calculate a genetic map for each chromosome. The genetic map and phenotype data were used to perform a single QTL scan to identify regions associated with a phenotype of interest. The plot generated in this example identified a peak on chromosome 4 containing *dhfr* associated with antifolate response.

3 High-throughput compound screen

3.1 Summary

qHTS requires high levels of reproducibility in order to generate high-quality data for downstream analyses. Assay optimization is a challenging step that seeks to mitigate error introduced by human or machine error that can affect reproducibility. It begins by considering a number of parameters that could introduce variation, and then adjusting assay conditions to reduce that variation. Following optimization, parental lines and a set of progeny from the 803xGB4 genetic cross were screened against a 2,816-compound pharmaceutical library. From the full library, 384 compounds were selected for further analysis. 803xGB4 results were compared with previous genetic crosses screened using a similar compound library. Finally, a preliminary QTL analysis identified a locus on chromosome 4 containing *dhfr* associated with parasite response to antifolate compounds. Other potential loci associated with response to a diverse array of compounds were also identified.

3.2 Introduction

The last three decades have seen a paradigm change in terms of the size and scale of assays involving compounds with potential biological activity. What was once a fairly laborious process involving few compounds and animal or isolated tissue testing has evolved into a high-throughput technique that allows for screening of several thousand compounds simultaneously[172]. It began with advances in combinatorial chemistry and the commercial availability of enormous small compound libraries[170]. These techniques allowed for in excess of 1 million compounds to be tested, but technological and methodological shortcomings only allowed one concentration of each compound to be tested, with a high rate of false

positives and false negatives[171]. In the early 2000s, a new method was developed that allowed for screening of several thousand compounds at a number of different concentrations, making it possible to generate dose-response curves[172]. qHTS has led to discoveries in a diverse array of fields, including cancer biology, toxicology, and parasitology[179,180,173,177,200,201], and it provided a unique platform for characterizing the progeny of the 803xGB4 genetic cross.

High-throughput screening is complex both logistically and analytically, and optimization of assay conditions is required before high-throughput screening can occur. Because of the small (10 µl) assay volume and the large number of plates required for the screen, variation due to human or machine error was introduced. Optimization allowed for mitigation of these factors and increased confidence in the results. A series of challenges were encountered in optimizing the assay, including fungal contamination and poor parasite growth. Ultimately, two parameters, the incubation atmosphere and the selection of positive control, were found to be instrumental in determining assay quality.

To further explore genetic determinants of drug response, a pharmaceutical compound library[181] was used to screen the parental lines and progeny of the 803xGB4 cross using the optimized qHTS method. This unique library consisted of 2,816 pharmaceutical compounds approved for human or veterinary use by regulatory authorities in the US, Japan, and/or Canada. Results of this screen generated the most comprehensive drug response profiles to date for the parental lines and progeny of the 803xGB4 cross. A set of 384 compounds was selected for use in a confirmatory screen, which provided the framework for linkage analysis described in Chapter 4. Previous work by Yuan et al. examined the response of the parental lines of the three previous crosses to a similar library, which enabled comparison of the differential responses of each of the crosses.

3.3 Objectives

i. To optimize the qHTS assay for use with the parents and progeny of the 803xGB4 genetic cross.

ii. To screen parental lines and progeny of the 803xGB4 genetic cross against a pharmaceutical compound library and compare results with previous screens.

iii. To analyze the primary screen data and identify compounds for use in a confirmatory screen.

3.4 Results

3.4.1 Optimization of 1536-well assay

The high-throughput drug response assay was optimized and performed as described in Sections 2.2.4 and 2.2.5. To determine the optimal conditions for the assay, S/B ratio, CV, and Z-factor were calculated for hypoxic and ambient O_2 conditions using either 1.15 or 2.3 µM ARM as a positive control and 0.23% DMSO as negative control.

A S/B value of 3 was required in order to ensure that the assay could be replicated on a high-throughput scale (Figure 3-1A, dashed line). The S/B for both 803 and GB4 were below 2 for the ambient O_2 condition. In the hypoxic condition, the S/B was over 6. The assay could only move forward using hypoxic conditions.

A CV less than 10 was required to ensure a reproducible assay (Figure 3-1B, dashed line). CVs were over 10 for both lines at 1.15 and 2.3 µM ARM in the ambient O_2 condition. In the hypoxic condition, the CV for GB4 was below 10 for both the 1.15 and 2.3 µM ARM conditions. However, the CV for 803 was only less than 10 for the 2.3 µM ARM condition, indicating that hypoxic conditions with 2.3 µM ARM as the positive control was the optimal condition for the assay.

The value of Z-factor reflects the level of separation between the positive-control signal variation and the negative-control signal variation (Table 3-1). In this case, the signal was the fluorescence intensity. In both the 1.15 and 2.3 µM ARM conditions, the Z-factor was negative for the ambient O_2 condition, indicating overlap of the positive and negative control signal variation bands. A high-throughput screen with a negative Z-factor would produce unreliable data, making screening essentially impossible. In the 5% O_2 condition, both 1.15 and 2.3 µM ARM yielded a Z-factor > 0.5, meaning that the separation band between the positive and negative control signal variations was large, and that the assay was reliable (Figure 3-1C).

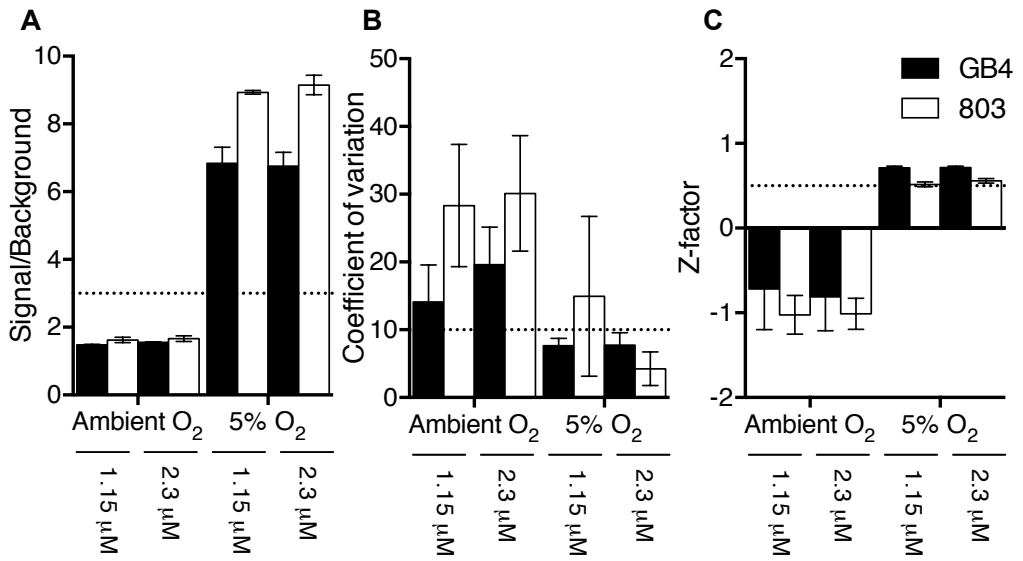


Figure 3-1 1536-well assay standardization metrics. Signal-to-background ratio (A), coefficient of variation (B), and Z-factor (C) of 803 and GB4 in ambient O_2 and 5% O_2 using 1.15 or 2.3 µM ARM as positive control.

Z-factor value	Structure of assay	Screening
1	SD = 0, or dynamic range → ∞	An ideal assay.
0.5 ≤ Z < 1	Separation band is large.	An excellent assay.
0 ≤ Z < 0.5	Separation band is small.	A double assay.
0	No separation band. The positive and negative control signal variation bands touch.	A yes/no assay.
Z < 0	The positive and negative control signal variation bands overlap.	Screening essentially impossible.

Table 3-1 Z-factor values and their implications for a given screen. (Table adapted from [189])

3.4.2 Comparison with previous compound screens

The primary screen was performed with 23 parasite lines and the compound library described in Section 2.2.3. The 23 lines consisted of the two parental lines and 17 unique progeny lines, with one additional member from each of four family groups. Families are groups of two or more parasites that are isogenic by SNP array

and WGS analysis (see Section 5.4.4 and Appendix I for more detail). The eight progeny that were members of the four family groups are as follows: GB4 = 34F5; 24G11 = 39A4; 85D3 = 61D3; 61E8 = 87E7. Results from the screen were analyzed, and compounds active in all lines tested were identified as described in section 2.2.3.

Data from the isogenic lines were used as a control for the reproducibility of the assay, since technical replicates were not performed at this stage. The log(IC_{50}) of compounds active in all 23 lines were plotted for each set of twins to observe for correlation (Figure 3-2). Linear regression was performed, with a slope of m = 1 indicating perfect correlation. The slopes for each set of twins were as follows: GB4:34F5 m = 0.76; 24G11:39A4 m = 0.83; 85D3:61D3 m = 0.81; and 61E8:87E7 m = 0.88. Correlation between the lines was similar to what was observed by Yuan et al. in a previous screen[180] using the parental lines of previous genetic crosses.

Yuan et al. identified 316 active compounds in the HB3x3D7 cross, 292 in the HB3xDd2 cross, and 745 in the 7G8xGB4 cross[180]. Using a similar library, 245 active compounds were identified in the 803xGB4 cross (Figure 3-3). The log(IC_{50} ratio) of the differential-response compounds from each of the four crosses was plotted (Figure 3-4). Nineteen compounds showed a DCP between HB3 and 3D7. The IC_{50} ratios for these 19 compounds was determined for the parental lines of the other three genetic crosses. Many of the compounds that showed a DCP between HB3 and 3D7 failed to produce a DCP in the other three crosses. A similar pattern was observed for the 24 compounds with DCPs from the Dd2xHB3 cross, the 30 compounds with DCPs from the 7G8xGB4 cross, and the 19 compounds with DCPs from the 803xGB4 cross. The only two compound classes that consistently showed DCPs in multiple crosses were the 4-aminoquinolines and the antifolates.

3.4.3 Identification of potent compounds

Amongst the parents and progeny of the 803xGB4 cross, 52 compounds with sub-micromolar mean IC_{50}s were identified (Figure 3-5, Appendix II Table S2-0-6). Of these compounds, nine were antineoplastic drugs, seven were antimalarials, seven were antiprotozoals, six were topical antiseptics, four were antiarrhythmics, three were microtubule stabilizers, three were antibacterials, two were antifolates, and two were antifungals. Also included were a tissue amoebacide, an antihelminthic, a histone deacetylase inhibitor, an immunosuppressant, a calcium channel blocker, and a tricyclic antidepressant. Three compounds could not be classified.

Among the highly active compounds, the lowest IC_{50} values (single nM scale) were observed for the antineoplastic actinomycin D and the antimalarial atovaquone. Actinomycin D is an inhibitor of transcription[202] used in the treatment of a number of human cancers. Side effects of the drug include bone marrow suppression, fatigue, and diarrhea[203]. Because of its mechanism of action and toxicity, it is unlikely to be a promising lead candidate for antimalarial devlopment[204]. Atovaquone is a well-characterized antimalarial drug that acts by collapsing the parasite's mitochondrial membrane potential[205]. Resistance to atovaquone results from mutations in cytochrome b[206].

Three compounds with <100 nM IC_{50} values were previously identified antipathogenics, indicating a potential for further development as antimalarials. Two of these compounds, azoxystrobin and buparvaquone, target the parasite's mitochondrion in a manner similar to atovaquone. Azoxystrobin, a fungicide, has been previously shown to act as a potent antimalarial[207]. Buparvaquone is a hydroxynaphthoquinone antiprotozoal used to treat bovine theileriosis[208], and also has previously reported antimalarial activity[208]. Acriflavinium is a topical antiseptic that is hazardous if ingested. However, structurally similar compounds have been

used therapeutically[209], indicating that this compound could be a lead for further development.

3.4.4 Preliminary linkage analysis

Following analysis of the primary screen (including QTL analysis described in Section 2.4.7), 384 compounds were selected for a confirmatory screen (Figure 3-6**;** Appendix II). Of 384 compounds, 356 met at least one of three criteria: 1) differential activity between 803 and GB4; 2) active in > 9 progeny lines with IC_{50} < 5 µM; 3) LOD score > 2.3. To maximize the chance of identifying linkage, the remaining 28 compounds had the next highest LOD scores. Of the 384 compounds selected for follow-up screening, 268 could be divided into general compound classes (Table S2-0-2), 78 had specific mechanisms of action (Table S2-0-4), and 41 had no classification or defined mechanism (Table S2-0-5). The largest compound classes represented in the follow-up screen were antihistamines (28 compounds), topical antiseptics (21 compounds), adrenergic receptor blockers (19 compounds), and antipsychotics (19 compounds).

Preliminary QTL analysis identified 28 compounds associated with compound response. A LOD threshold of 4 was established by performing a permutation test with 10,000 permutations at $\alpha = 0.025$. Results at this stage were considered suggestive, as replicates were not performed, and a limited number of progeny were screened. Of the 28 compounds with LOD scores > 4, there were five antifolates, two antispasmodics, two imidazole antifungals, two tricyclic antidepressants, two adrenergic receptor binders, two antibiotics and two antihistamines. Also included were an antiprotozoal, a psychedelic, an antipsychotic, and an opioid receptor binder, a serotonin receptor binder, and a dopamine receptor binder.

The strongest LOD scores were observed for the antifolate compounds pyrimethamine, trimethoprim, methodichlorophen, and triamterene, which are known

to inhibit *dhfr*, a key enzyme in the folate synthesis pathway[167,210,211]. Sequencing data indicated that 803 possessed three mutations in *dhfr* known to cause pyrimethamine treatment failure[212], while GB4 was wildtype. Four of the compounds elicited differential responses between 803 and GB4, and the progeny could clearly be classified as sensitive or resistant. Methotrexate, another antifolate, showed no differential response between 803 and GB4; however, previous reports have shown that methotrexate response is unaffected by natural *dhfr* polymorphisms[213]. When QTL analysis was performed using these phenotypes, a large peak was identified on chr 4, which contained *dhfr* (Figure 3-7), validating the initial results of the primary screen.

The other loci significantly associated with compound response produced more equivocal results. Some compounds produced a differential response only seen in 803 and three or fewer progeny, some compounds produced inconsistent results between twin lines, and some LOD scores were near the threshold of 4. These compounds were included in the confirmatory screen to validate the findings.

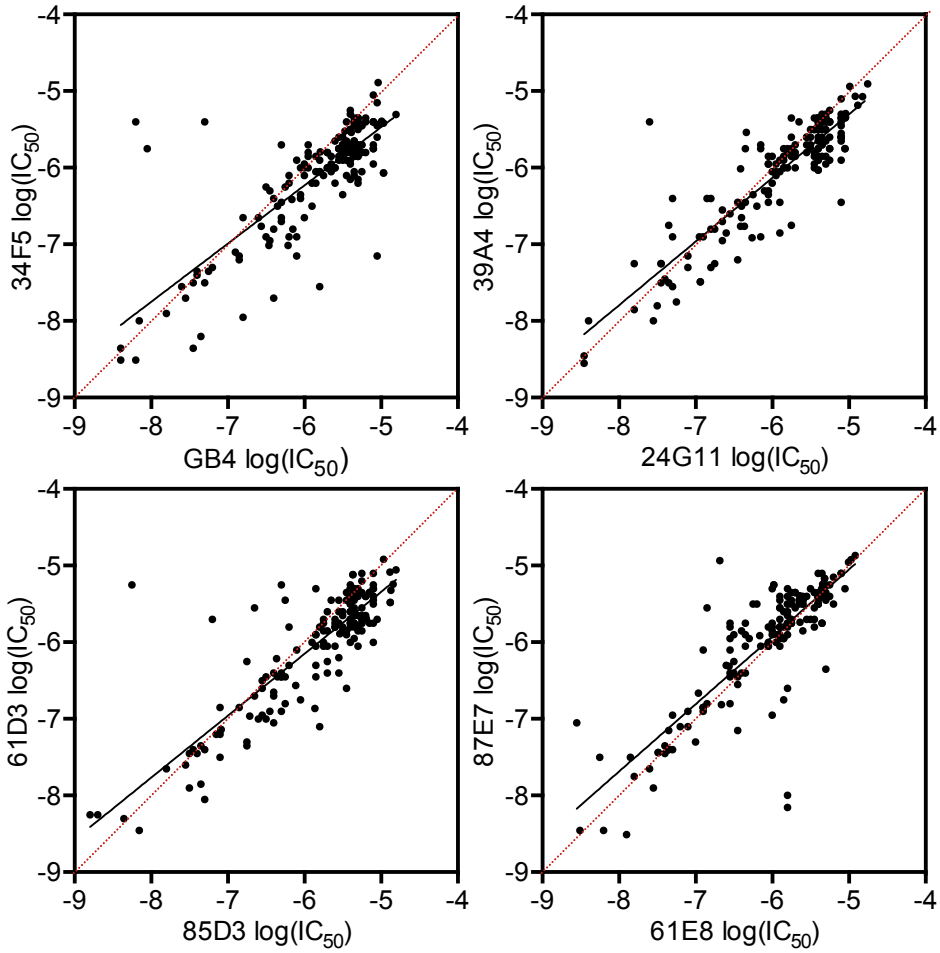


Figure 3-2 Active compound correlation of twin lines. Correlation of log(IC_{50}) between twin lines included in the primary screen. Regression lines (black) were plotted for each of the twin sets. Perfect correlation (y = x; red dashed line) was added for reference.

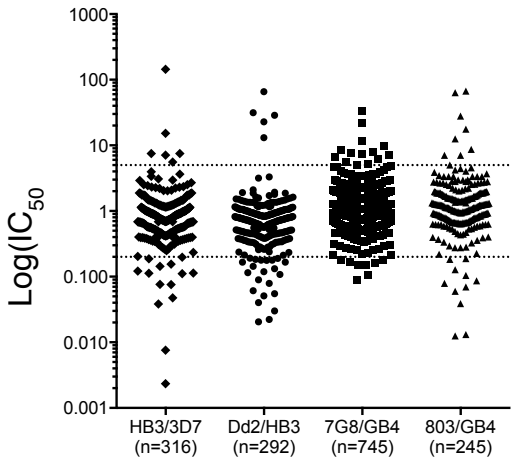


Figure 3-3 Active compounds from all crosses. Log(IC_{50}) ratio of parental lines from each of the four genetic crosses for active compounds. DCP thresholds (dotted line) at 5 and 0.2. Data adapted from [179,180].

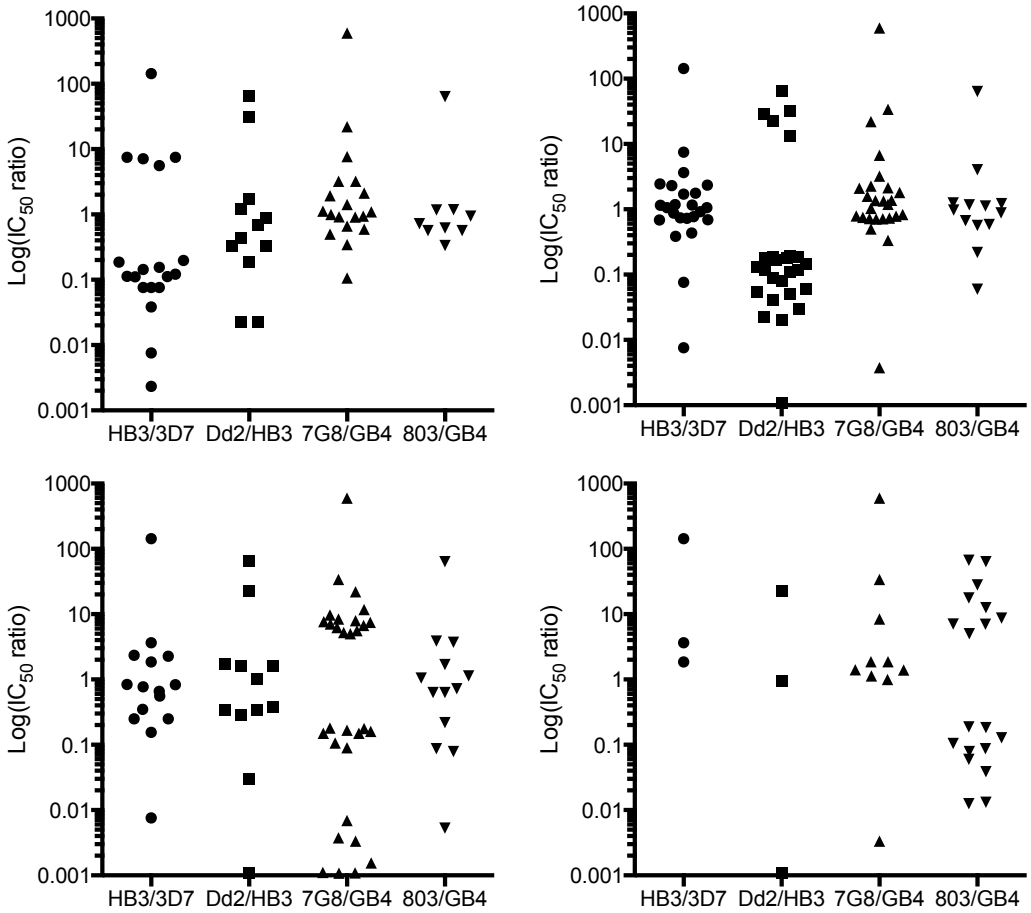


Figure 3-4 DCPs presented as ratios of IC_{50} values from each of the four crosses.

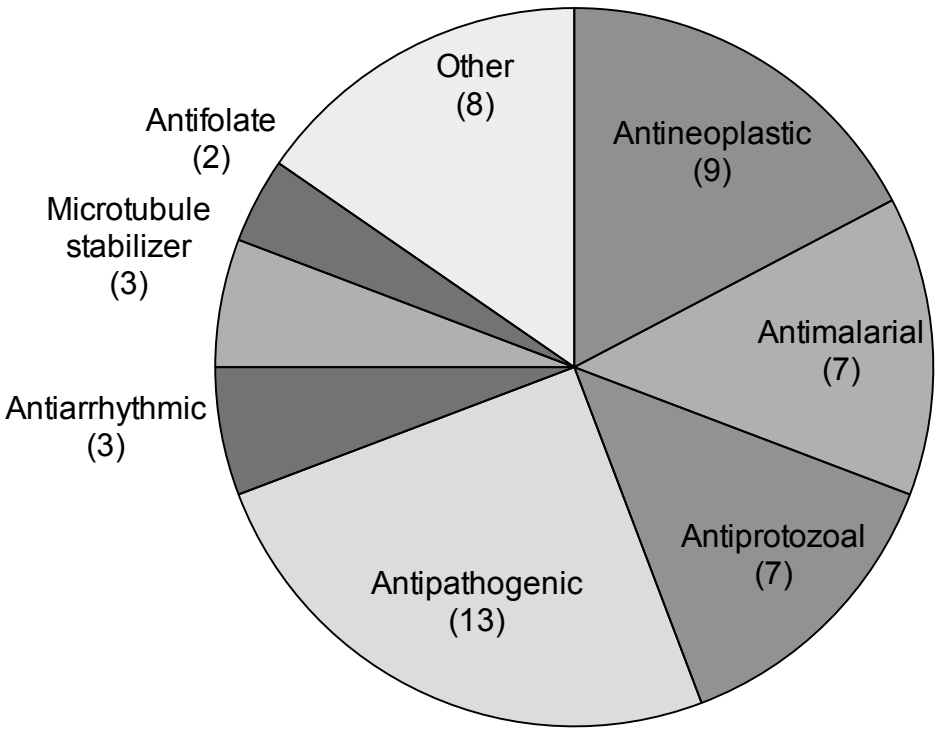


Figure 3-5 Highly active compounds identified in the primary screen. Classification of compounds for which the mean IC_{50} of all lines screened was < 1 µM.

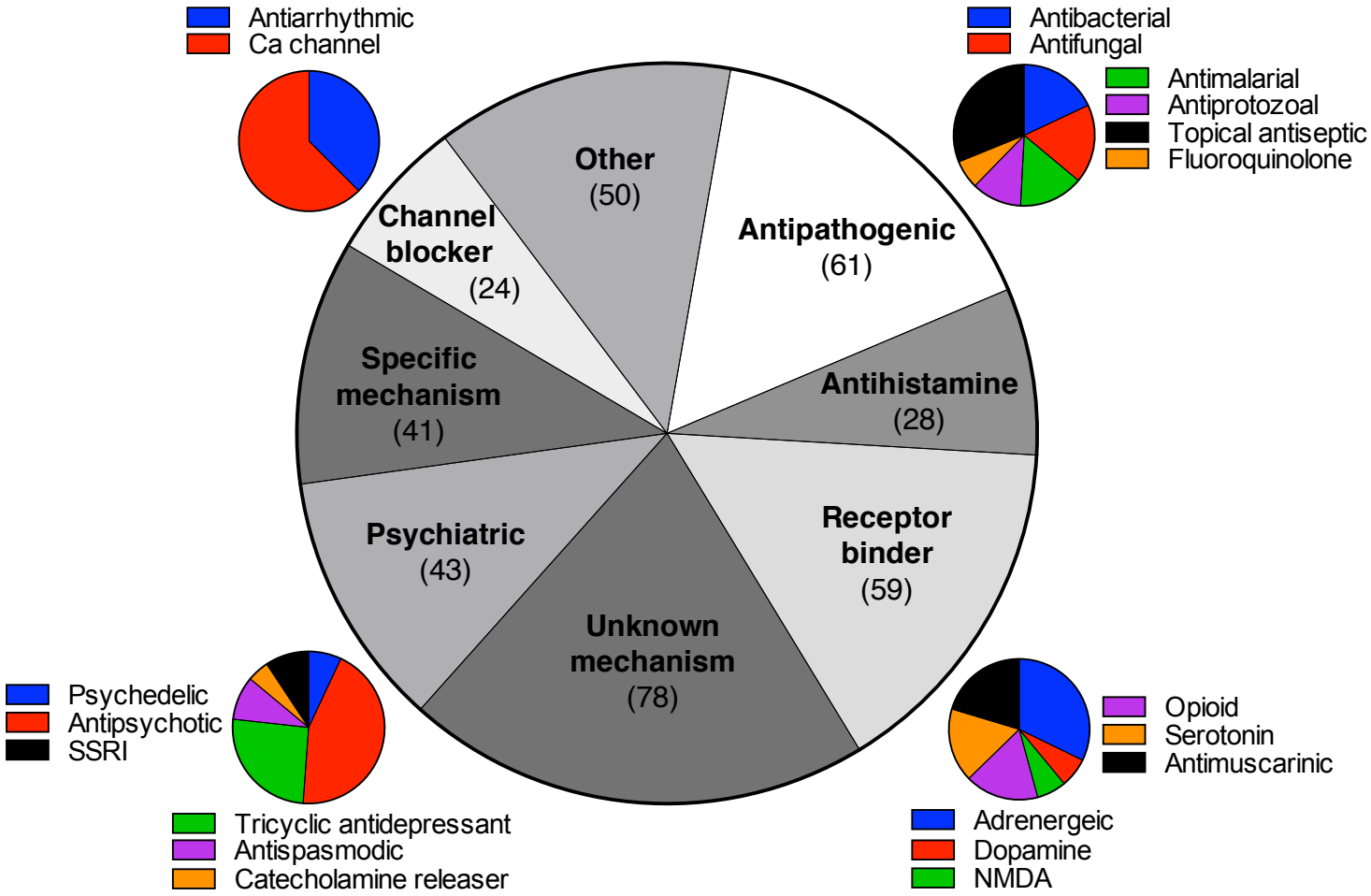


Figure 3-6 Compound classes in confirmatory screen. The 384 compounds selected for the confirmatory screen were divided by common mechanism of action.

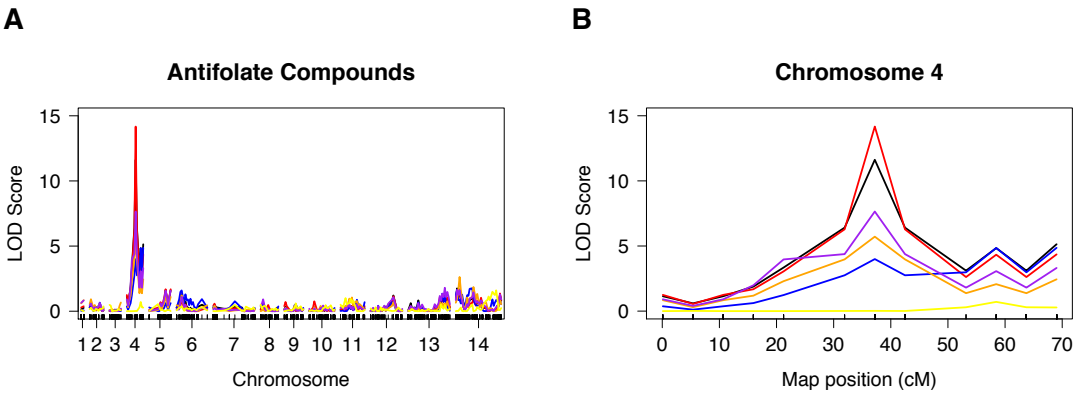


Figure 3-7 Chr 4 loci associated with antifolate response in the primary screen. QTL plot of (A) all 14 chromosomes and (B) chr 4. Trimethoprim (red), pyrimethamine (black), methodichlorophen (blue), triamterene (purple), methotrexate (yellow). LOD scores for pyrimethamine, trimethoprim, methodichlorophen, and triamterene were significant ($p < 0.05$).

3.5 Discussion

3.5.1 qHTS optimization

The nature of qHTS requires small reagent volumes and the testing of several thousand samples across dozens of 1536-well plates simultaneously. Variation is unavoidably introduced into each plate by instrumental and human error. However, to extract meaningful data from analysis of these plates, a certain degree of reproducibility is required. Optimizing the assay protocol can counteract this variation to allow for reliable interpretation of the results.

To optimize a qHTS assay to evaluate drug response in the parents and progeny of the 803xGB4 genetic cross, three variables were considered: S/B, CV, and Z-factor (see Section 2.2.5). Two variables were evaluated using these three metrics. Atmospheric oxygen content is known to impact *P. falciparum* growth[214]. Previous HTS screens with *P. falciparum* have been performed in ambient O_2 with 5% CO_2[179,180] or in hypoxic conditions with 5% O_2, 5% CO_2, and 90% N_2[173]. It was determined that hypoxic conditions offered the most robust and reliable results, generating a higher S/B ratio, a lower CV, and a Z-factor closer to 1. The other variable tested was the concentration of ARM (1.15 or 2.3 µM) to use as the positive control for parasite death. The highest Z-factor was achieved under the 2.3 µM ARM condition. Optimized qHTS conditions enabled reproducible screening parental lines and progeny of the genetic cross.

3.5.2 Comparison with previous screens

qHTS studies have been performed on the parental lines from the previous three crosses using a similar library of pharmaceutically active compounds. There were two key differences between Yuan et al.'s work and the work presented in this thesis. The first is that they screened only six parasite lines, so were able to perform

replicates of the primary screen. Twenty-three lines from the 803xGB4 cross were screened, including four pairs of isogenic lines. Screening a greater number of lines against the full compound library allowed for a more nuanced look at drug response in a variety of recombinant progeny, thereby maximizing the likelihood of detecting QTLs.

For comparison with Yuan et al.'s work, two analyses were performed. The first was to examine the concordance of drug response between the four pairs of isogenic lines, which were considered technical replicates. Only compounds classified as active were used in this analysis. Yuan et al. found concordance between replicates to range from 0.71 to 0.92, with a mean of 0.83. Concordance between the isogenic lines in the 803xGB4 screen was similar, ranging from 0.76 to 0.88 with a mean of 0.82.

Next, the active compounds from each of the four genetic crosses were identified. Overall, 803xGB4 was more similar to the 3D7xHB3 and Dd2xHB3 genetic crosses. 7G8xGB4 appeared to be the outlier, with >700 active compounds identified in the screen. This result was somewhat surprising, as it is over twice the number observed in the other crosses. For each of the crosses, the list of active compounds with a five-fold or greater difference in response between the parental lines was generated. For example, there were 19 compounds with a five-fold difference in response between 3D7 and HB3. The ratio of response to these 19 compounds (when available) was plotted for each of the other three crosses. This process was repeated using the list of differential-response compounds from each cross. One striking result is that compounds that generate a differential response in one cross are unlikely to do so in another. The only compound classes that did were the 4-aminoquinolines and the antifolates. This highlights the major finding of Yuan et al. that polymorphisms in *pfcrt*, *pfmdr1*, and *dhfr* control most differential drug

responses. Different polymorphisms between the parental lines in these three genes play a very nuanced role in drug response. This finding also has implications in drug design, as it makes it difficult to predict based on chemical structure whether a given compound is likely to be active or not.

3.5.3 Selection of compounds for confirmatory screening

A set of 384 compounds was selected for confirmatory screening, including potent compounds, and compounds that produced preliminary evidence of linkage in QTL analysis. Of the 52 compounds with sub-micromolar IC_{50}, 31 were previously identified as having antipathogenic activity. Twelve of the remaining compounds were identified as treatments for various human cancers, indicating a level of toxicity that would be inappropriate in an antimalarial. Of the remaining compounds, three were antiarrhythmic agents, including QN (also an antimalarial) and its stereoisomer quinidine, and clofilium tosylate. These compounds can be dangerous; both quinidine and clofilium tosylate can induce life-threatening arrhythmias[215,216]. As such, these two compounds may be unsuitable for development as antimalarial compounds. A final compound of interest was doxepin, a tricyclic antidepressant. Tricyclic antidepressants have long been of interest in malaria research[217,218]. They have been shown to revert the CQ resistance phenotype both *in vitro* and *in vivo*[217,219,220]. Doxepin has been shown to have single micromolar activity in previous screens[180]. As such, it may be a suitable compound for medicinal chemistry efforts to improve its efficacy.

In order to maximize the likelihood of detecting QTLs, one of the major criteria for selecting the 384 compounds for the confirmatory screen was evidence of linkage in the primary screen. To confirm that the primary screen was able to detect linkage, QTL analysis was performed using five antifolates. Because the polymorphisms in *dhfr* were known to modulate parasite response to antifolates[221], a significant

association with a locus on chromosome 4 containing *dhfr* was expected and observed for four of the five compounds. Parasite response to the fifth compound, methotrexate, has been previously shown to be unaffected by *dhfr* polymorphisms[213].

Preliminary analysis also identified several other QTLs for a variety of compounds. One compound, butoconazole, is an antifungal structurally related to clotrimazole, which has been studied for antimalarial activity[222,223]. Clotrimazole may target *P. falciparum* hemoperoxidase, resulting in oxidative stress and death of the parasite[224]. Preliminary data from the primary screen indicated that clotrimazole response differed between 803 and GB4, and that the 803 phenotype was inherited by three of the progeny. Resistance to azole antifungals has not been previously reported in *P. falciparum*, so this result was noteworthy with the caveat that the resistant phenotype was only observed in three lines, and that replicates were not performed at this stage.

4 Genetic loci associated with drug response

4.1 Summary

qHTS with the 803xGB4 genetic cross provided an exciting opportunity to further investigate the dynamics of the genetic basis parasite response. A set of 384 compounds was used for a confirmatory screen designed to refine measures of parasite response. Data from the confirmatory screen were used to perform QTL analysis to identify loci associated with drug response. Twenty-five loci were significantly associated with parasite response to 61 compounds, including several antimalarial compounds. Screening of further progeny refined the association of six compounds, including MFQ and triclosan, to a locus on chromosome 5 containing *pfmdr1*. qPCR analysis indicated that 803 transcribed *pfmdr1* at twice the level of GB4 in the absence of a copy number variation.

4.2 Introduction

Many studies have utilized qHTS to investigate aspects of *P. falciparum* biology, with results ranging from the discovery of a compound capable of single-dose cure in a mouse malaria model[177] to identification of compounds capable of disrupting parasite cytoadherence interactions[225]. The *P. falciparum* genetic crosses are particularly amenable to qHTS. Using this technique with the parental lines and progeny of three previous genetic crosses, Yuan et al. determined that mutations in *pfcrt*, *pfmdr1*, and *dhfr* determined parasite response to 96% of compounds with differential phenotypes between the parental lines[179,180]. These results indicated significant homogeneity in parasite response to a variety of compounds.

To further explore these findings, and to identify novel loci associated with drug response, the parents and progeny of the 803xGB4 genetic cross were screened

against a pharmaceutical compound library described in Chapter 3. Of the 2,816 compounds in the primary screen, 384 were selected for a confirmatory screen. This screen was designed to provide a more accurate measure of parasite response, as each compound was tested in triplicate, and at a larger range of concentrations (11 three-fold dilutions versus 7 five-fold dilutions).

Results of this screen were used to perform QTL analysis to identify regions significantly associated with drug response. Given Yuan et al.'s findings, it was expected that loci containing *dhfr* and *pfmdr1* would account for most of the QTLs observed. Because the parental lines possessed the same PfCRT haplotype, no linkage was expected to a locus containing *pfcrt*. Linkage analysis revealed a surprising number of loci associated with parasite response to a number of compounds, including *dhfr* and *pfmdr1*. However, these two genes only accounted for 26% of the observed associations. The locus containing *pfmdr1* was investigated further to determine if CNV or transcriptional changes might account for the linkage observed.

4.3 Objectives

i. To generate high-quality drug response data for use in QTL analysis.

ii. To identify genetic loci associated with drug response.

iii. Identify candidate genes within a locus that might modulate response to compounds of interest.

4.4 Results

4.4.1 Identification of loci associated with drug response

A confirmatory screen was performed in triplicate using 384 compounds and 18 additional progeny lines. Microarray data were cleaned and analyzed as described in Section 2.4.2. Of these lines, four were members of family groups: 34B1 and 40B12 were isogenic, and comprised a new family group; 34D8 was identical to the previously identified singleton 36H9, creating a new family group; 38A6 belonged to the previously identified 11H5 family, and 37F12 was a GB4 self cross. In total, 21 unique progeny were screened against 384 compounds.

QTL analysis was performed for each of the 384 compounds in the confirmatory screen. Compounds with LOD score > 3.5 or $p < 0.05$ were identified. In total, 25 genetic loci were significantly associated with parasite response to 61 compounds. These associations are briefly detailed below. More detailed analysis will be presented for loci on chromosomes 4 and 5.

Five compounds were linked to regions on chromosomes 2 (Figure 4-1**,** Table 4-1). Roxarsone (veterinary food additive) and monatepil (unclassified) were linked to a 37-kb region centered at 330 kb. 1-Phenylcyclohexylamine HCl (catecholamine releaser) and 2,5-Dimethoxy-4-ethylamphetamine (unclassified) were linked to a 250-kb region centered at 530 kb. Thonzonium Br (cationic surfactant) was linked to a 215-kb region centered at 630 kb.

Five compounds were linked to regions on chromosome 3 (Figure 4-2**,** Table 4-2). Atovaquone (antimalarial) was linked to a 100-kb region centered at 910 kb. Halofantrine (antimalarial) and aprofene (antimuscarinic) were linked to a 30-kb region centered at 950 kb. Premethadone (unknown) was linked to a 200-kb region centered at 725 kb. Amiodarone (antiarrhythmic) was linked to a 650-kb region centered at 580 kb.

Sixteen compounds were linked to a 600-kb region centered at 800 kb on chromosome 6 (Figure 4-5**,** Table 4-5). They included the antihistamines protmethazine, cyclizine, pheniramine, methapyrilene, and doxylamine succinate; the adrenergic receptor blockers alprenolol, depxropranolol, and efaroxan; the selective serotonin reuptake inhibitors sertraline, AMI-193, and alaproclate; metazocine (opioid receptor binder), amitryptiline (tricyclic antidepressant), diphenidol (antiemetic), pridinol (antispasmodic), and ipenoxazone (unclassified).

Five compounds were linked to regions on chromosome 7 (Figure 4-6**,** Table 4-6). Oxprenolol (adrenergic receptor blocker) and mexiletine (antiarrhythmic) were linked to a 250-kb region centered at 1300 kb. 1-Bis-4-fluorophenyl methyl piperazine (calcium channel blocker) and rimcazole (unknown) were linked to a 500-kb region centerted at 520 kb. Zolantadine (antihistamine) was linked to a 30-kb region centered at 1200 kb.

Five compounds were linked to regions on chromosome 8 (Figure 4-7**,** Table 4-7). Mibefradil (calcium channel blocker) and mosapramine (tricyclic antidepressant) were linked to a 100-kb region centered at 1300 kb. 1-Bis-4-fluorophenyl methyl piperazine (calcium channel blocker) was linked to a 30-kb region centered at 1345 kb. Primaquine (antimalarial) was linked to a 200-kb region centered at 700 kb. Dibenzepine (tricyclic antidepressant) was linked to a 70-kb region centered at 920 kb.

Twelve compounds were linked to a 61-kb region centered at 1418 kb on chromosome 9 (Figure 4-8**,** Table 4-8). These included the microtubule stabilizers docetaxel and vinorelbine, the calcium channel blockers nicardipine and azelnidipine, ciprofloxacin (antibiotic), vorinostat (histone deacetylase inhibitor), isoconazole (imidazole antifungal), bosutinib (antineoplastic), clenbuterol (adrenergic receptor binder), diponium (antihistamine), methotrexate (antifolate), and bifemelane

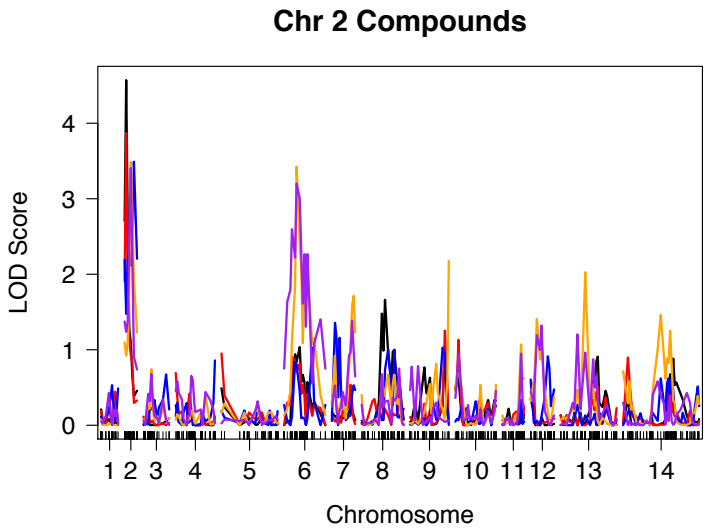
(nootropic). This region is notable because all of the progeny inherited the GB4 allele in this region (see Section 5.4.5 and Appendix I). The drug-resistant phenotype was only observed in 803. Confirmation of this phenotype in 803 may provide insight into the function of hypothetical genes within this region (see Appendix I for list of genes).

One compound, BIBX-1382 (epidermal growth factor receptor inhibitor), was linked to a 250-kb region centered at 414 kb on chromosome 10 (Figure 4-9**,** Table 4-9).

Five compounds were linked to regions on chromosome 13 (Figure 4-10**,** Table 4-10). Oxprenolol (adrenergic receptor binder) and mexiletine (antiarrhythmic) were linked to a 500-kb region centered at 1230 kb. U-62066 (opioid receptor binder) was linked to a 210-kb region centered at 550 kb. Tolteridine (antimuscarinic) and pronethalol (adrenergic receptor binder) were linked to a 900-kb region centered at 750 kb.

Six antifolates and an adrenergic receptor blocker were linked to regions on chromosome 4 (Figure 4-3**,** Table 4-3). Consistent with results of the primary screen, the antifolates mapped to a 60-kb region centered at 790 kb containing *dhfr* (Pf3D7_04_v3: 748,088 - 749,914). Progeny that inherited $dhfr_{803}$ showed a marked difference in IC_{50} compared with progeny that inherited $dhfr_{GB4}$ (Figure 4-11). The former were virtually insensitive to trimethoprim, with IC_{50} values recorded at 100 µM, which indicates that the true IC_{50} was above the highest concentrations tested in the assay (29 µM). Response to methodichlorophen also showed a distinct separation, with a ~100-fold difference in IC_{50} between the two alleles. The other four compounds exhibited a varying range of difference in IC_{50}, which indicated that there might be additional genes responsible for modulating the phenotype.

A



B

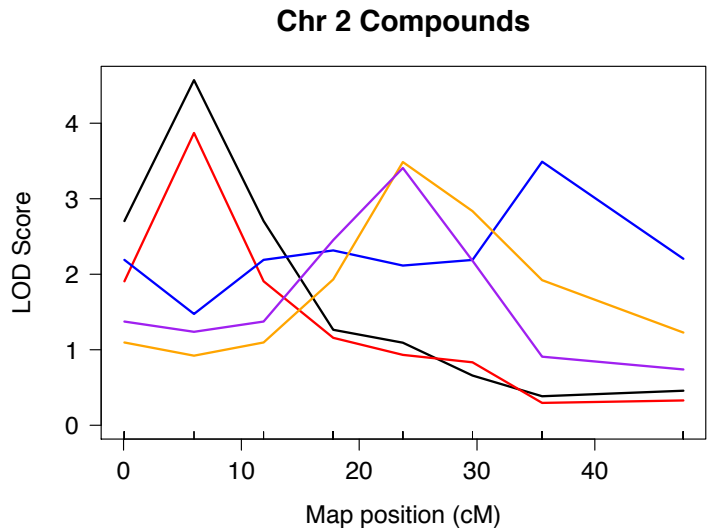


Figure 4-1 Chr 2 loci associated with drug response. QTL plots of (**A**) all 14 chromosomes and (**B**) chr 2. The x-axis represents the marker distribution (cM) along each of the 14 chromosomes. The y-axis represents the LOD score for each compound. Monatepil (black), roxarsone (red), thonzonium Br (blue), 2-5-Dimethoxy-4-ethylamphetamine HCl (orange), 1-Phenylcyclohexylamine HCl (purple).

	Chr 1	Chr 2 ↓	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	Chr 12	Chr 13	Chr 14
Monatepil	0.21	4.57*	0.09	0.26	0.49	1.04	0.25	1.66	1.03	0.64	0.36	0.91	0.88
Roxarsone	0.44	3.87*	0.20	0.69	0.95	0.91	0.54	0.45	1.25	1.13	0.55	0.51	0.90
Thonzonium Br	0.53	3.49*	0.67	0.86	0.27	1.03	1.36	1.00	1.03	0.32	0.91	0.27	0.62
2-5-Dimethoxy-4-ethylamphetamine HCl	0.08	3.49*	0.74	0.39	0.29	3.42*	1.71	0.67	2.18	0.72	1.41	2.03	1.46
1-Phenylcyclohexylamine HCl	0.43	3.41	0.68	0.65	0.32	3.20	1.39	0.75	0.78	0.94	1.32	1.20	0.58

Table 4-1 List of compounds linked to chr 2. Compounds with LOD > 3.0 or *p < 0.05 linked to regions on chr 2.

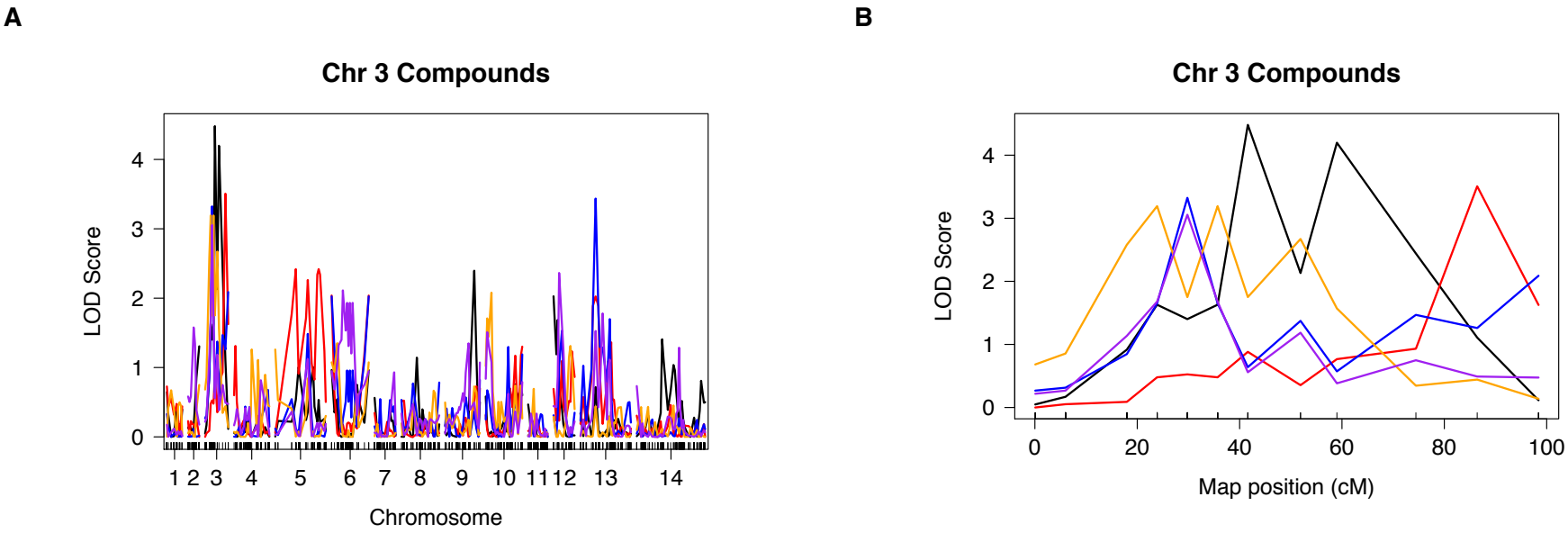


Figure 4-2 Chr 3 loci associated with drug response. QTL plots of (A) all 14 chromosomes and (B) chr 3. Atovaquone (black), halofantrine (blue), premethadone (red), amiodarone (orange), aprofene (purple).

	Chr 1	Chr 2	Chr 3 ↓	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	Chr 12	Chr 13	Chr 14
Atovaquone	0.65	1.31	4.48*	0.30	1.01	0.96	0.56	1.14	2.40	0.76	2.03	0.72	1.41
Halofantrine	0.73	0.26	3.51	1.31	2.42	2.04	0.35	0.52	0.73	1.30	1.24	2.02	0.46
Premethadone	0.33	0.11	3.32*	0.68	1.48	2.03	0.54	0.78	0.82	1.30	1.53	3.44	0.27
Amiodarone HCl	0.67	0.75	3.19*	1.26	1.26	1.35	0.19	0.67	0.73	2.08	1.31	0.44	0.45
Aprofene	0.18	1.58	3.06*	0.82	1.12	2.11	0.93	0.48	1.35	1.51	2.36	1.78	1.29

Table 4-2 List of compounds linked to chr 3. List of compounds with LOD > 3.0 or * $p < 0.05$ linked to regions on chr 3.

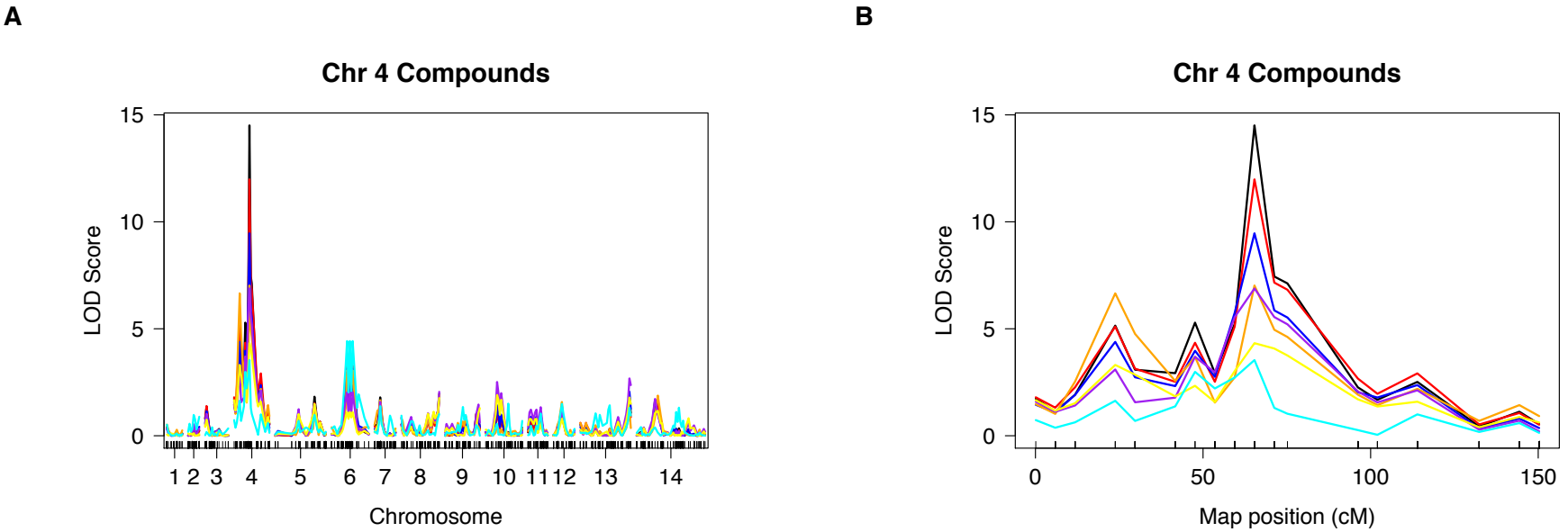


Figure 4-3 Chr 4 loci associated with drug response. QTL plots of (A) all 14 chromosomes and (B) chr 4. Methodichlorophen (black,), trimethoprim (red), trimetrexate (blue), pyrimethamine (orange), triamterene (purple), ormetoprim (yellow), dexpropanolol (cyan).

	Chr 1	Chr 2	Chr 3	Chr 4 ↓	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	Chr 12	Chr 13	Chr 14
Methodichlorophen	0.42	0.31	1.37	14.51*	1.83	3.19	1.80	1.96	1.04	1.13	1.20	1.75	1.64
Trimethoprim	0.37	0.44	1.39	11.99*	1.52	3.28	1.52	1.69	1.00	1.10	1.27	1.52	1.44
Trimetrexate	0.31	0.49	1.15	9.46*	1.39	3.12	1.40	1.77	0.83	1.43	1.37	2.06	1.35
Pyrimethamine	0.52	0.07	0.81	7.02*	1.06	3.06	1.66	1.91	0.91	0.46	1.58	1.16	1.88
Triamterene	0.30	0.53	1.03	6.89*	1.48	2.03	1.58	2.05	1.46	2.51	1.10	2.68	1.72
Ormetoprim	0.11	0.12	0.75	4.33*	1.49	1.11	0.88	1.76	1.29	1.90	0.79	1.79	1.16
Dexpropranolol	0.52	0.97	0.39	3.54*	0.97	4.43	1.38	1.01	1.34	1.50	1.49	1.42	1.31

Table 4-3 List of compounds linked to chr 4. Compounds with LOD > 3.0 or *p < 0.05 linked to regions on chr 4.

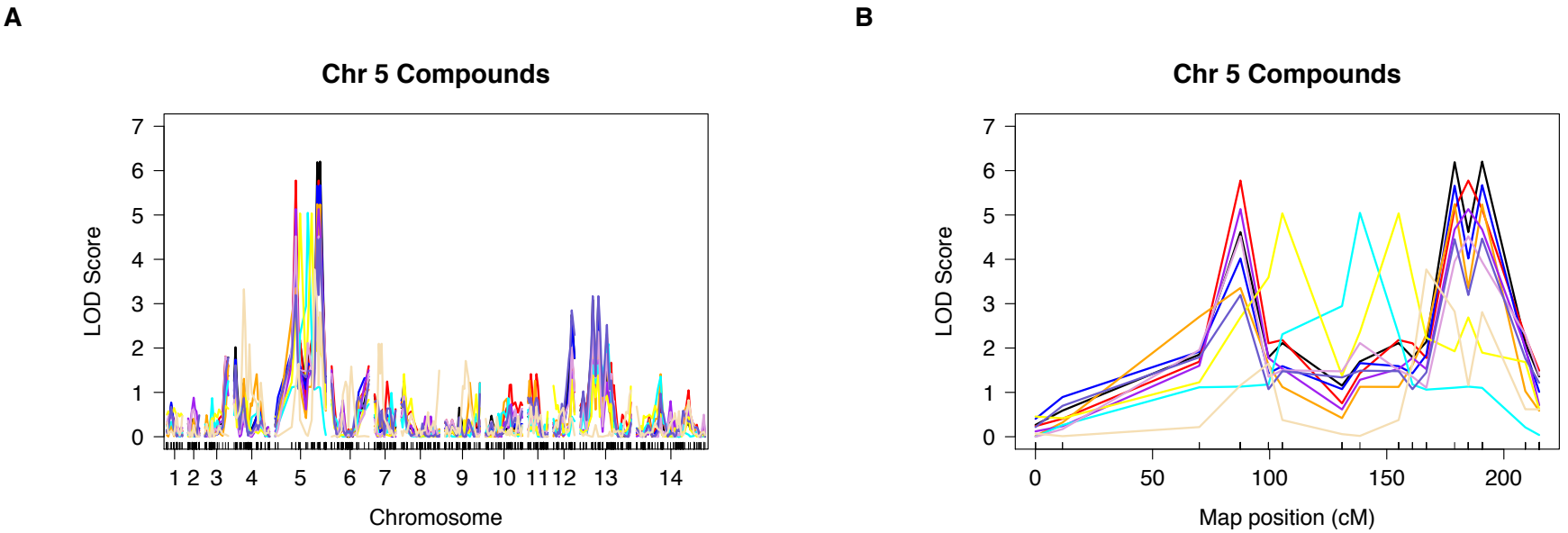


Figure 4-4 Chr 5 loci associated with drug response. QTL plots of (**A**) all 14 chromosomes and (**B**) chr 5. Clofoctol (black), suloctidil (red), triclosan (blue), 2-benzyl-4chlorophenol (orange), MFQ (purple), artemotil (cyan), artemisinin (yellow), difeterol (pink), cinnamedrine (grey), clobetasol (tan).

	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5 ↓	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	Chr 12	Chr 13	Chr 14
Clofoctol	0.72	0.39	1.79	2.02	6.20*	1.29	0.93	0.94	0.66	0.47	2.75	3.09	1.04
Suloctodil	0.61	0.20	1.81	1.53	5.78*	1.59	1.24	1.20	1.21	1.17	1.88	2.81	1.05
Triclosan	0.77	0.19	1.37	1.74	5.67*	1.35	0.50	1.26	0.27	0.53	2.61	2.76	1.17
2-Benzyl-4-chlorophenol	0.29	0.41	1.62	1.41	5.24*	1.06	0.49	1.14	1.22	0.86	1.03	2.31	1.41
Mefloquine HCl	0.51	0.88	1.48	1.37	5.13*	1.50	0.37	1.06	0.29	0.73	1.47	1.73	0.80
Artemotil	0.34	0.20	1.26	0.85	5.05*	0.73	0.99	0.62	1.21	0.87	0.45	2.09	1.36
Artemisinin	0.65	0.14	0.69	1.01	5.04*	0.83	0.80	1.41	0.96	0.35	1.29	1.39	0.70
Difeterol	0.55	0.61	1.80	1.53	4.52*	1.42	0.39	0.91	0.43	0.63	2.01	2.37	0.68
Cinnamedrine	0.67	0.48	1.78	1.61	4.47*	1.19	0.83	0.80	0.96	0.70	2.85	3.17	0.97
Clobetasol propionate	0.52	0.26	0.37	3.33	3.77*	1.58	2.10	1.49	1.71	0.47	0.66	1.05	0.83

Table 4-4 List of compounds linked to chr 5. Compounds with LOD > 3.0 or *p < 0.05 linked to regions on chr 5.

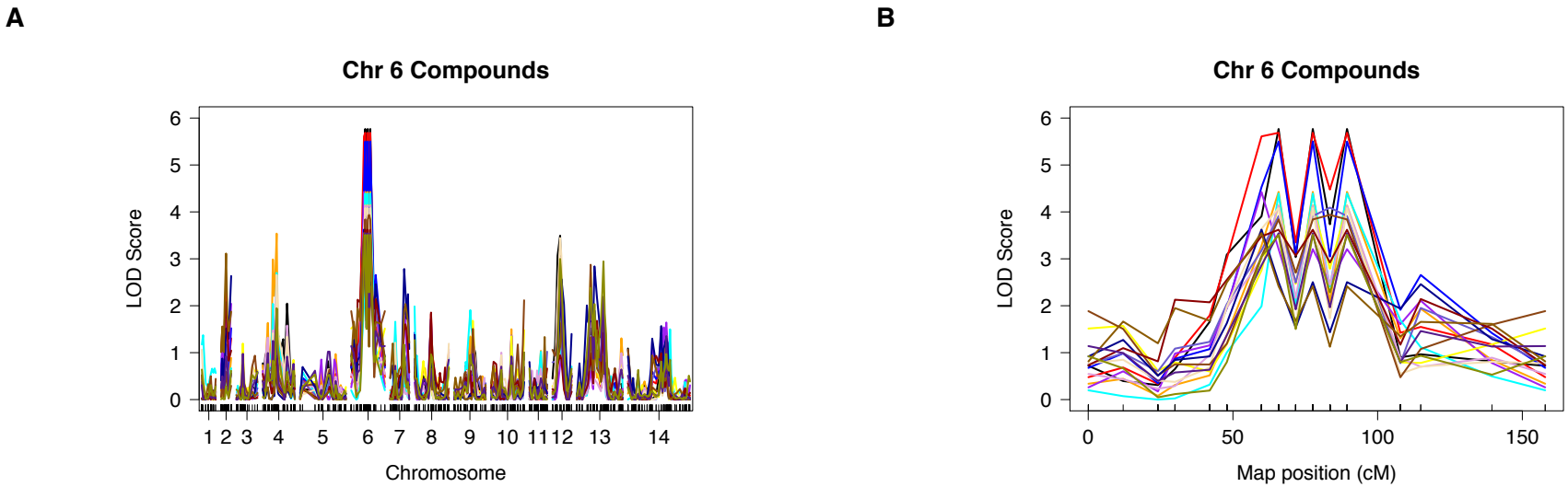


Figure 4-5 Chr 6 loci associated with drug response. QTL plots of (**A**) all 14 chromosomes and (**B**) chr 5. Promethazine HCl (black), metazocine fumarate (blue), alprenolol HCl (red), despropranolol (orange), doxylamine succinate (purple), amitriptyline HCl (cyan), cyclizine (yellow), ipenoxazone HCl (pink), diphenidol HCl (grey), sertraline HCl (tan), AMI-193 (brown), pheniramine (dark blue), pridinol (dark red), methapyrilene HCl (dark orange), alaproclate (dark purple), efaroxan HCl (dark yellow).

	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6 ↓	Chr 7	Chr 8	Chr 9	Chr 10	Chr 12	Chr 13	Chr 14
Promethazine HCl	0.45	2.03	0.88	2.10	0.82	5.77*	0.47	0.82	1.20	0.96	3.49	1.52	1.30
Metazocine fumarate	0.34	2.38	0.33	1.52	0.55	5.69*	0.89	1.82	1.30	0.79	1.95	1.00	1.28
Alprenolol HCl	0.46	1.52	0.24	1.52	0.54	5.50*	1.97	1.71	1.08	1.36	1.64	2.17	1.21
Dexpropranolol	0.52	0.97	0.39	3.54	0.97	4.43*	1.38	1.01	1.34	1.50	1.49	1.42	1.31
Doxylamine succinate	0.18	2.04	0.14	1.21	0.85	4.42*	1.80	1.10	0.76	1.36	0.99	1.31	1.64
Amitriptyline HCl	1.37	0.67	0.38	2.70	0.59	4.39*	1.66	1.99	1.91	1.17	1.96	1.14	1.50
Cyclizine	0.40	0.84	1.19	1.97	0.60	4.14*	0.90	0.86	1.68	1.48	2.22	2.71	0.64
Ipenoxazone HCl	0.57	1.44	0.58	2.55	0.50	4.13*	0.92	1.07	1.27	1.00	3.03	1.28	0.78
Diphenidol HCl	0.38	1.18	0.42	0.92	0.43	4.09*	1.00	1.52	1.09	1.00	1.53	1.35	1.03
Sertraline HCl	1.06	0.92	0.87	2.66	0.65	4.08*	1.21	1.34	0.94	0.95	3.44	2.24	0.82
AMI 193	0.52	0.59	0.94	0.93	0.64	3.94*	0.76	1.41	1.51	2.12	2.92	2.87	1.07
Pheniramine maleate	0.20	2.63	0.59	0.61	0.70	3.62*	2.78	1.38	0.21	0.92	2.57	2.84	1.57
Pridinol	0.12	2.07	0.28	0.87	0.27	3.62*	1.60	1.85	0.62	0.81	0.94	1.68	0.99
Methapyrilene HCl	0.33	3.11	0.42	0.71	0.25	3.55	2.04	0.59	0.48	0.72	1.64	1.48	0.94
Alaproclate	0.31	0.93	0.98	1.94	1.02	3.54*	0.86	0.87	0.97	0.91	2.50	2.29	1.35
Efaroxan HCl	0.64	0.61	0.58	1.95	0.53	3.51*	1.17	1.27	0.95	1.01	3.00	2.95	0.97

Table 4-5 List of compounds linked to chr 6. Compounds with LOD > 3.0 or *p < 0.05 linked to regions on chr 6.

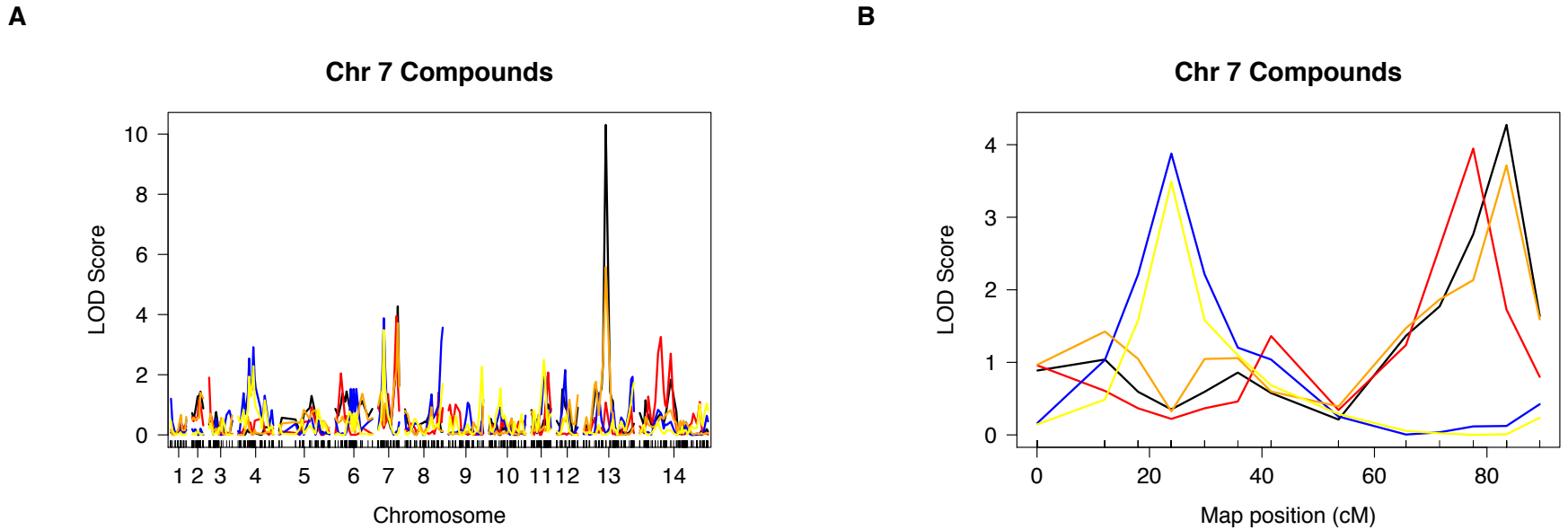


Figure 4-6 Chr 7 loci associated with drug response. QTL plots of (**A**) all 14 chromosomes and (**B**) chr 7. Oxprenolol (black), zolantidine (red), 1-bis-4-fluorophenyl methyl piperazine (blue), mexiletine (orange), rimcazole (yellow).

	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7 ↓	Chr 8	Chr 9	Chr 10	Chr 12	Chr 13	Chr 14
Oxprenolol HCl	0.18	1.44	0.76	0.26	1.30	1.45	4.27*	0.85	1.18	0.54	1.52	10.31	1.84
Zolantidine dimaleate	0.32	1.20	1.90	0.71	0.92	2.05	3.95*	1.41	1.01	0.58	0.11	1.08	3.26
1-Bis-4-fluorophenyl methyl piperazine	1.20	0.21	0.81	2.92	0.57	1.53	3.88*	3.57	1.07	0.92	2.16	1.93	0.83
Mexiletine HCl	0.67	1.40	0.71	0.61	0.83	1.37	3.71*	0.92	0.95	0.54	1.32	5.58	0.82
Rimcazole dihydrochloride	0.47	0.16	0.54	2.29	0.86	0.78	3.49	1.70	2.27	1.55	0.65	1.75	1.04

Table 4-6 List of compounds linked to chr 7. Compounds with LOD > 3.0 or *p < 0.05 linked to regions on chr 7.

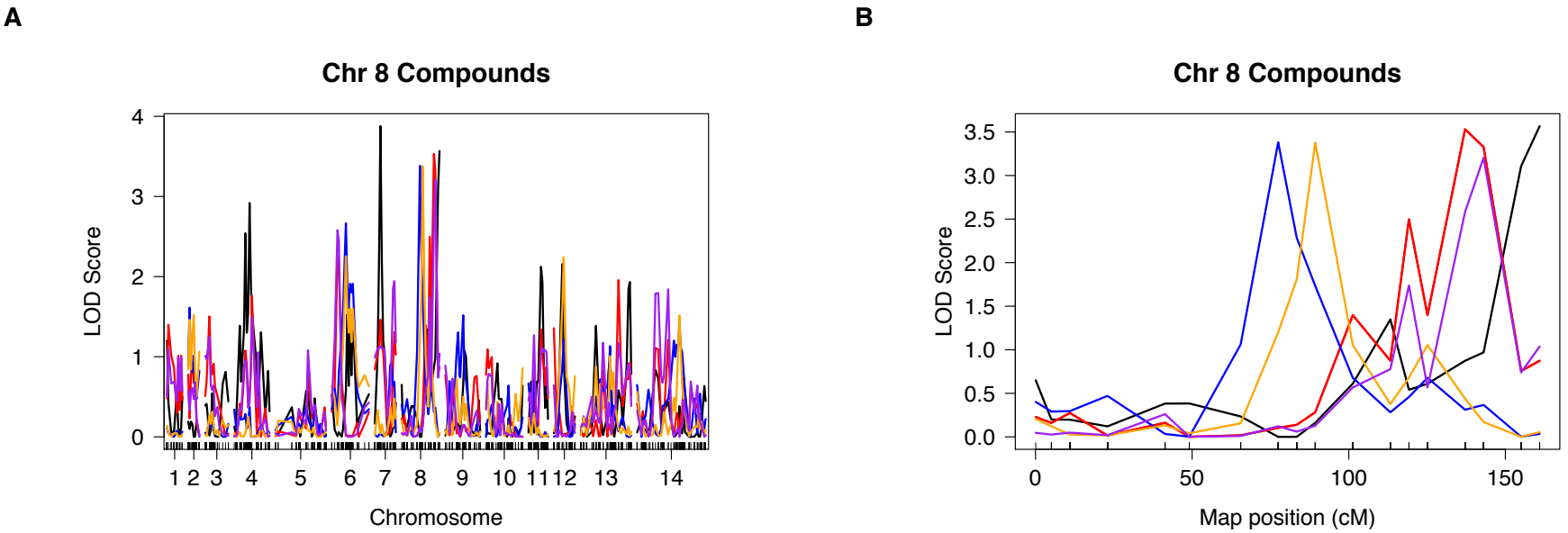


Figure 4-7 Chr 8 loci associated with drug response. QTL plots of (**A**) all 14 chromosomes and (**B**) chr 8. 1-Bis-4-fluorophenyl methyl piperazine (black), mibefradil HCl (red), primaquine phosphate (blue), dibezepine HCl (orange), mosapramine (purple).

	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8 ↓	Chr 9	Chr 10	Chr 12	Chr 13	Chr 14
1-Bis-4-fluorophenyl methyl piperazine	1.20	0.21	0.81	2.92	0.57	1.53	3.88*	3.57	1.07	0.92	2.16	1.93	0.83
Mibefradil HCl	1.40	0.93	1.50	1.76	0.82	2.25	1.46	3.53*	0.77	1.09	1.36	1.96	1.21
Primaquine phosphate	0.11	1.61	0.24	0.38	0.38	2.66	0.92	3.38*	1.52	0.64	1.22	1.13	1.41
Dibenzepine HCl	0.14	1.52	0.31	0.25	0.20	2.25	0.48	3.38*	0.72	0.86	2.24	1.01	1.51
Mosapramine	1.09	0.68	1.25	1.48	1.08	2.58	1.94	3.20	0.89	0.79	0.64	1.17	1.84

Table 4-7 List of compounds linked to chr 8. Compounds with LOD > 3.0 or *p < 0.05 linked to regions on chr 8.

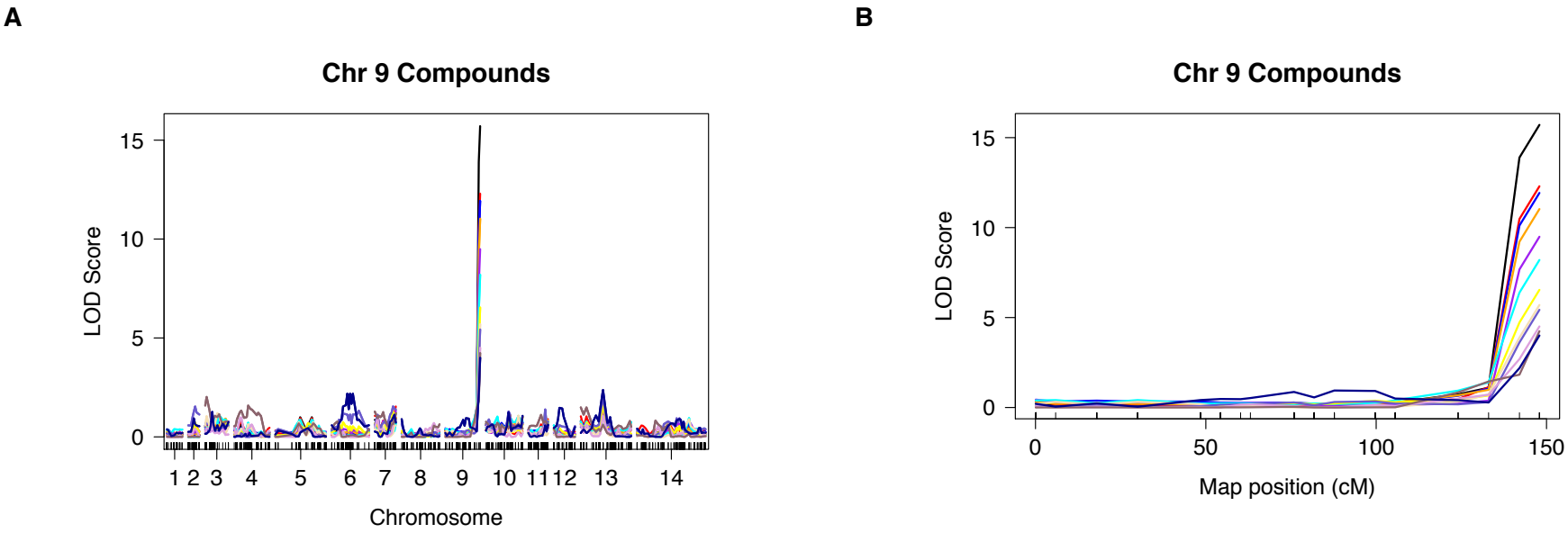
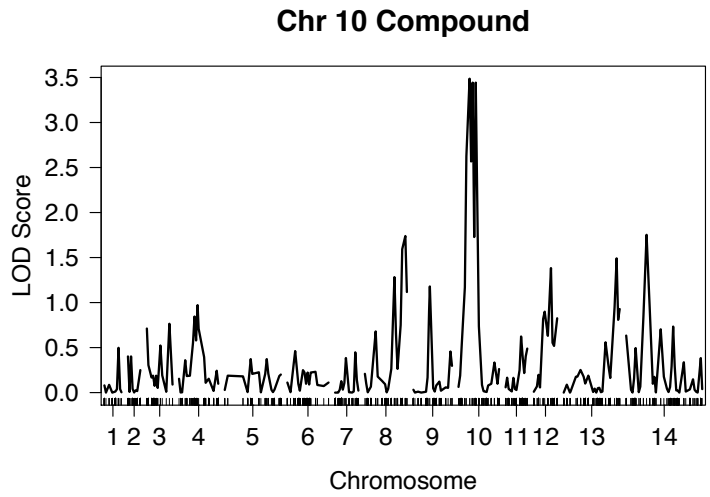


Figure 4-8 Chr 9 loci associated with drug response. QTL plots of (**A**) all 14 chromosomes and (**B**) chr 9. Docetaxel (black), ciprofloxacin (red), vinorelbine tartrate (blue), vorinostat (orange), isoconazole nitrate (purple), bosutinib (cyan), nicardipine (yellow), azelnidipine (tan), clenbuterol (grey), diponium bromide (pink), methotrexate (dark pink), bifemelane HCl (dark blue).

	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9 ↓	Chr 10	Chr 12	Chr 13	Chr 14
Docetaxel	0.18	0.20	0.86	0.61	0.99	0.38	1.22	0.23	15.71*	0.84	0.57	1.40	0.53
Ciprofloxacin	0.27	0.51	0.71	0.81	0.95	0.29	1.54	0.33	12.29	0.58	0.37	1.51	0.70
Vinorelbine tartrate	0.19	0.14	0.83	0.51	0.62	0.33	0.91	0.21	11.93*	0.82	0.36	1.25	0.65
Vorinostat	0.27	0.20	0.92	0.40	0.55	0.32	0.98	0.14	11.03	1.10	0.42	1.00	0.43
Isoconazole nitrate	0.38	0.35	0.51	0.36	0.58	0.41	1.31	0.26	9.49*	1.03	0.34	1.21	0.78
Bosutinib	0.39	0.00	1.09	0.72	0.87	0.29	0.69	0.39	8.20*	1.05	0.58	0.78	0.95
Nicardipine	0.24	0.72	0.93	0.29	0.69	0.74	1.25	0.30	6.54*	0.97	0.44	1.59	0.70
Azelnidipine	0.09	0.04	1.20	0.66	0.75	0.16	0.70	0.34	5.70	1.07	0.16	0.94	0.90
Clenbuterol HCl	0.20	1.54	0.67	0.23	0.42	1.32	1.54	0.65	5.43*	1.29	0.59	2.00	0.94
Diponium bromide	0.07	0.41	0.66	0.99	0.48	0.28	1.13	0.70	4.51*	0.57	0.27	1.06	0.65
Methotrexate	0.01	0.01	2.02	1.58	0.68	0.78	1.28	0.63	4.24*	0.76	0.46	1.49	0.82
Bifemelane HCl	0.38	0.94	1.26	0.23	0.44	2.18	1.12	0.89	4.00*	1.24	1.47	2.37	0.91

Table 4-8 List of compounds linked to chr 9. Compounds with LOD > 3.0 or *p < 0.05 linked to regions on chr 9.

A



B

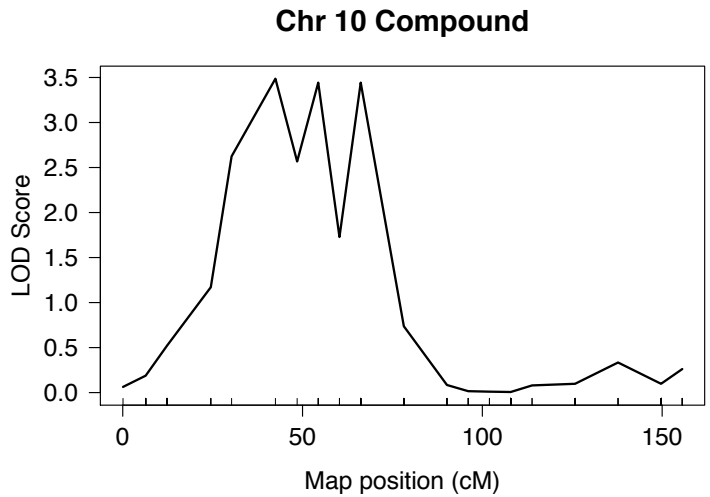


Figure 4-9 Chr 10 loci associated with drug response. QTL plot of (A) all 14 chromosomes and (B) chr 10. BIBX 1382 dihydrochloride (black).

	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10 ↓	Chr 12	Chr 13	Chr 14
BIBX 1382 dihydrochloride	0.50	0.40	0.76	0.97	0.37	0.46	0.45	1.74	1.18	3.49*	1.38	1.49	1.75

Table 4-9 List of compounds linked to chr 10. Compounds with LOD > 3.0 or *p < 0.05 linked to a region on chr 10.

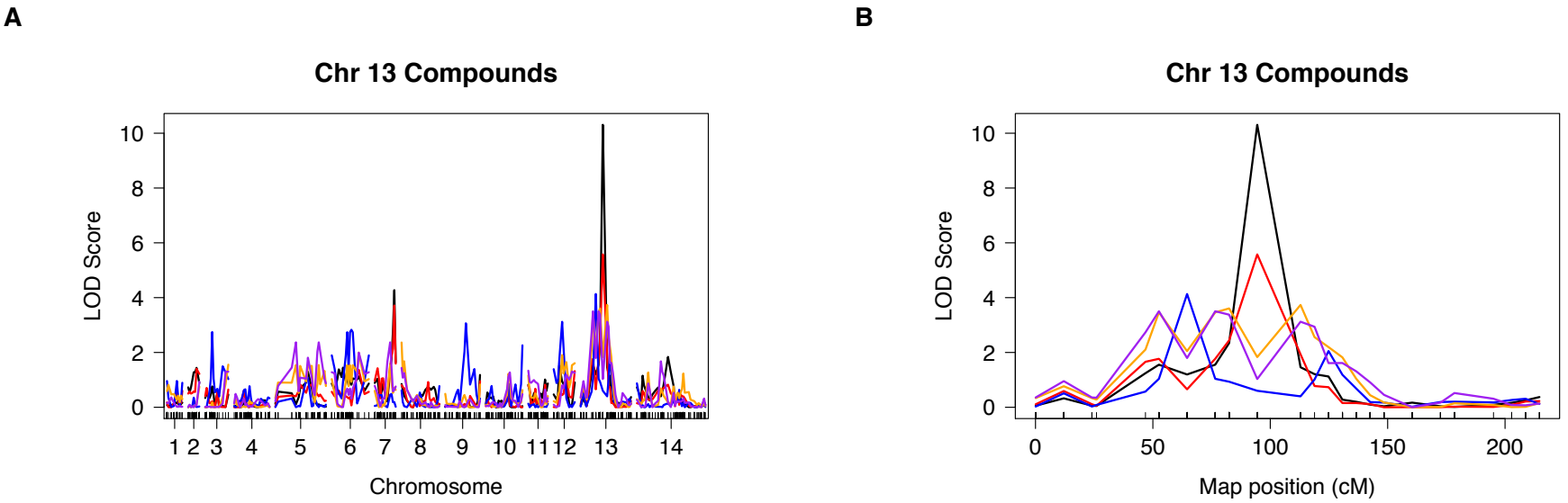


Figure 4-10 Chr 10 loci associated with drug response. QTL plot of (**A**) all 14 chromosomes and (**B**) chr 13. Oxprenolol HCl (black), mexiletine HCl (red), U 62066 (blue), tolterodine tartrate (orange), pronethalol (purple).

	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	Chr 12	Chr 13 ↓	Chr 14
Oxprenolol HCl	0.18	1.44	0.76	0.26	1.30	1.45	4.27	0.85	1.18	0.54	1.52	10.31*	1.84
Mexiletine HCl	0.67	1.40	0.71	0.61	0.83	1.37	3.71	0.92	0.95	0.54	1.32	5.58*	0.82
U 62066	0.96	0.35	2.75	0.77	1.09	2.83	0.49	0.77	3.06	2.26	3.12	4.13*	0.87
Tolterodine tartrate	0.82	0.75	1.55	0.30	1.54	2.00	1.54	2.36	0.83	1.22	1.90	3.73	1.27
Pronethalol HCl	0.36	0.93	1.30	0.49	2.37	1.99	2.37	1.44	0.29	1.27	1.30	3.50*	1.68

Table 4-10 List of compounds linked to chr 13. Compounds with LOD > 3.0 or *p < 0.05 linked to regions on chr 13.

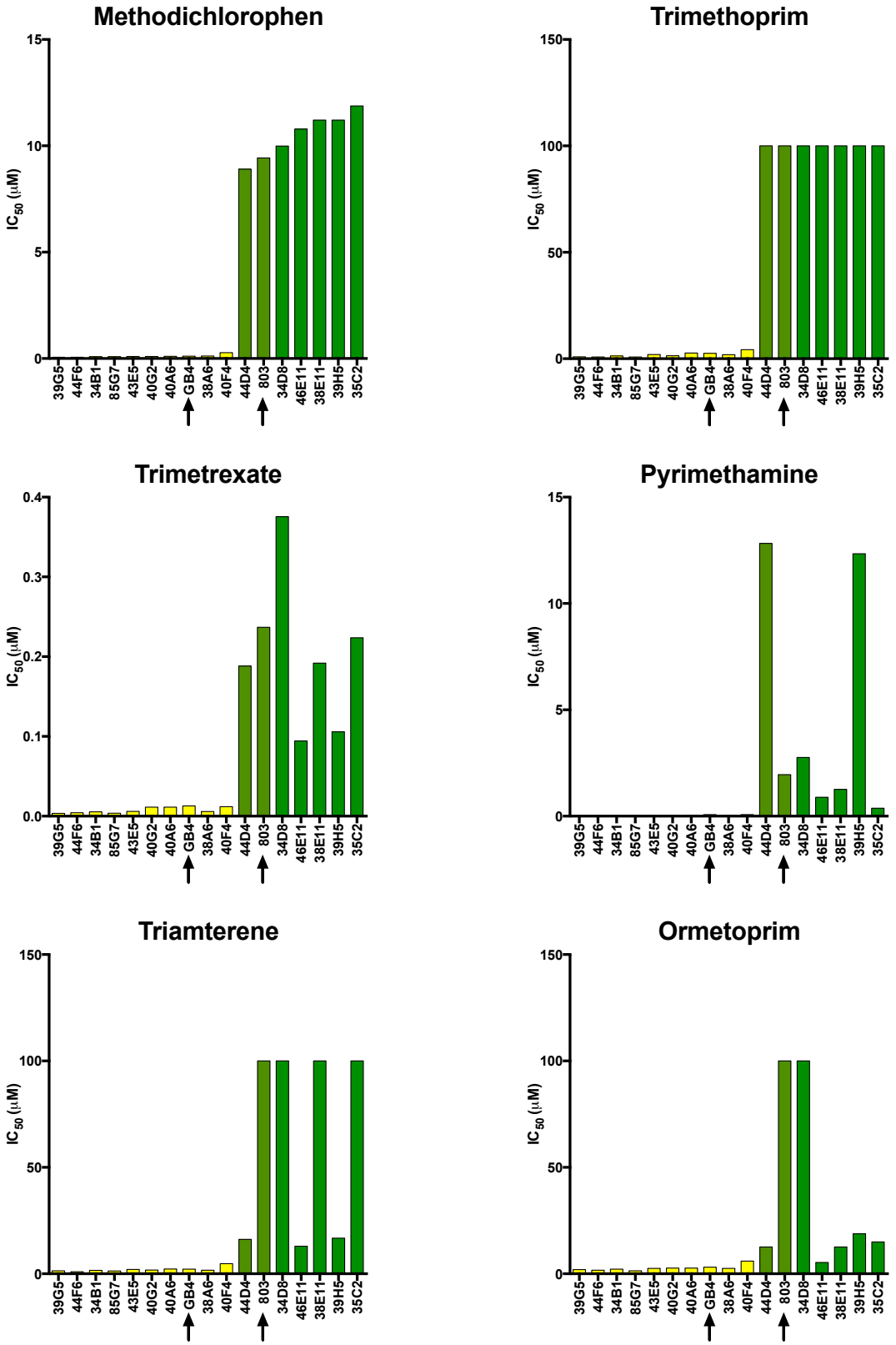


Figure 4-11 Response to antifolate compounds. IC_{50} values (µM) for six antifolate compounds. Arrows indicate the parental lines. Green indicates inheritance of the 803 DHFR and yellow indicates inheritance of the GB4 DHFR.

4.4.2 Increased expression level of **pfmdr1** *may modulate drug response*

Ten compounds were linked to regions on chromosome 5 (Figure 4-4**,** Table 4-4). These included the antibacterials clofoctol, triclosan, and 2-benzyl-4-chlorophenol; the antimalarials MFQ, ARM, and artemotil; and suloctidil (vasodilator), cinnamedrine (ephedrine sympathomimetic), clobetasol (corticosteroid), and difeterol (unclassified). Clofoctol, triclosan, 2-benzyl-4-chlorophenol, MFQ, suloctidil, cinnamedrine, clobetasol, and difeterol were linked to a 1-Mb region with two peaks centered at 55 kb and 1213 kb. ARM was linked to a 400-kb region centered at 550 kb. Artemotil was linked to a 5-kb region centered at 477 kb.

To confirm the linkage of these compounds to this region, seven additional progeny lines were screened. Following re-analysis of the dataset with the additional progeny, the QTLs for ARM and artemotil disappeared. The locus linked to clofoctol, triclosan, 2-benzyl-4-chlorphenol, suloctidil, difeterol, and MFQ was refined to a 146-kb region centered at 953 kb containing *pfmdr1* (Pf3D7_05_v3 from 957,890 to 962,149). The distribution of the IC_{50} values for these six compounds indicated that multiple genes may be involved in determining parasite response (Figure 4-12). Screening of more progeny may increase the power of the analysis sufficiently to detect additional loci.

To investigate whether *pfmdr1* copy number variation or expression level was involved in modulating parasite response to the six compounds above, qPCR was performed. Relative to GB4, 803 showed no evidence of copy number variation, whereas Dd2 showed the expected amplification of the locus[226] (Figure 4-13A). Interestingly, when *pfmdr1* expression was examined in ring-stage synchronized parasites, 803 showed a two-fold expression level difference over GB4 (Figure 4-13B), in the absence of copy number variation.

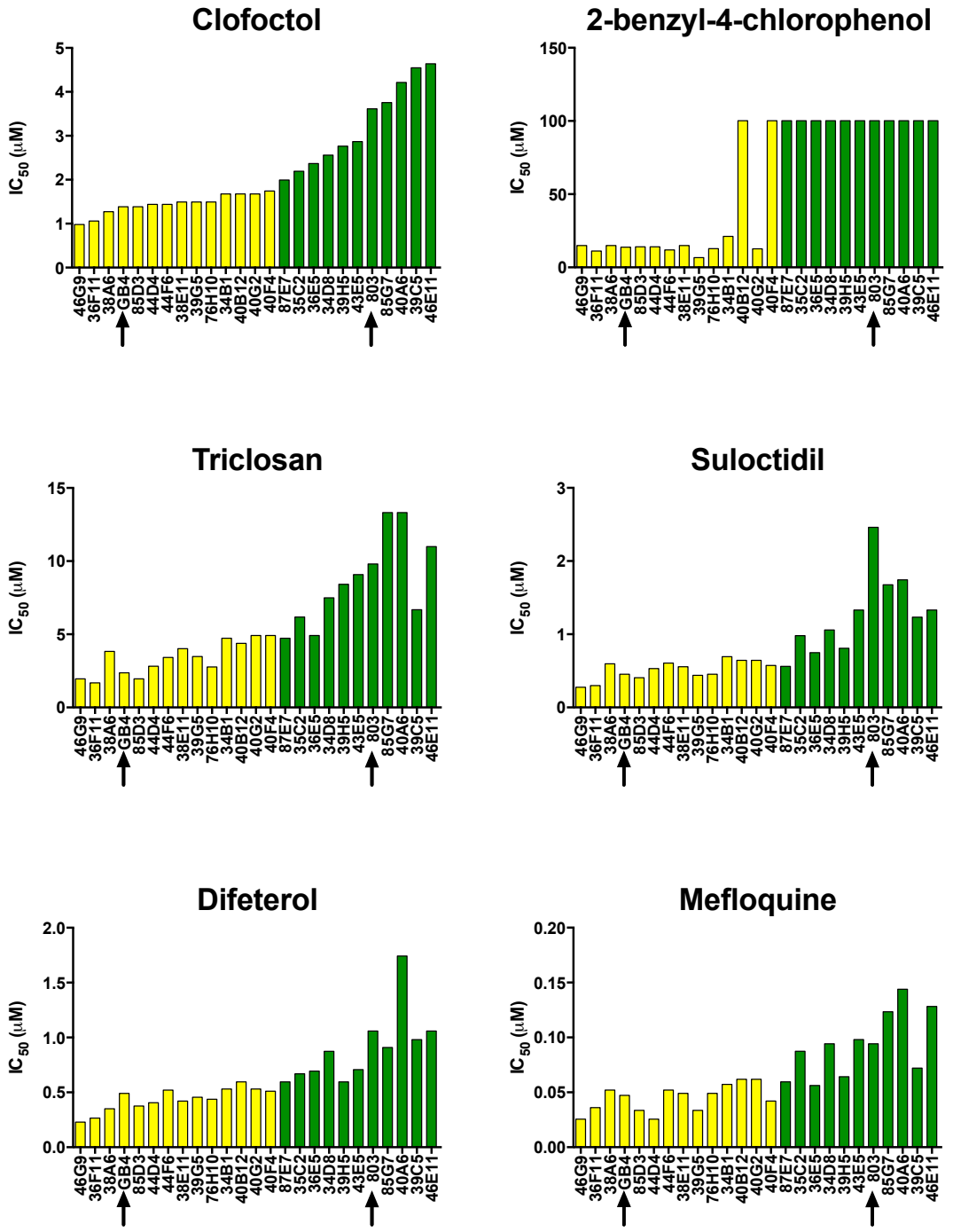


Figure 4-12 Response to 6 compounds linked to chromosome 5. IC_{50} (μM) values for six compounds linked to a locus on chromosome 5. Green indicates inheritance of the 803 allele and yellow indicates inheritance of the GB4 allele.

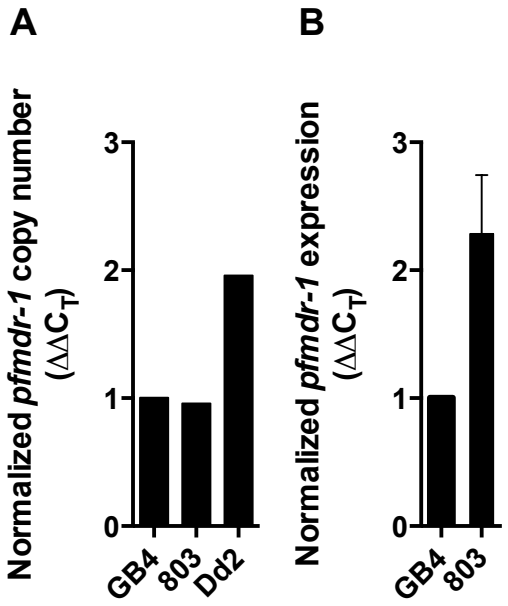


Figure 4-13 Copy number and expression level of *pfmdr1*. Both copy number and expression level were performed using *P. falciparum* lactate dehydrogenase as the control for the $\Delta\Delta C_T$ calculations. (**A**) Relative copy number was examined using gDNA for GB4, 803, and Dd2. Each reaction was performed in triplicate with n=1. (**B**) Expression level was determined by RT-PCR performed using RNA extracted from ring-stage synchronized cultures of 803 and GB4. Each reaction was performed in triplicate with n=5.

4.5 Discussion

4.5.1 Linkage to pfcrt*,* dhfr*, and* pfmdr1

Mutations in *pfcrt*, *dhfr*, and *pfmdr1* have been shown to account for 96% of differential drug responses observed between parental lines in previous *P. falciparum* genetic crosses. 803 and GB4 possess a different combination of mutations in these three genes than the parental lines of previous crosses (3D7xHB3, HB3xDd2, and 7G8xGB4). These unique differences allowed for further evaluation of Yuan et al.'s observation regarding the role these genes play in determining drug response. In total, loci containing *pfcrt*, *dhfr*, and *pfmdr1* were linked to parasite response for 16 of the 61 (26%) compounds identified by QTL analysis, much lower than the 96% described by Yuan et al.

In Yuan et al.'s study, a locus containing *pfcrt* was linked to response to 209 compounds. These response differences were attributed to polymorphisms in *pfcrt*. In

two previous crosses (HB3xDd2 and 7G8xGB4), the parental lines differed in PfCRT haplotype. HB3 is CQ-sensitive, Dd2 is CQ-resistant, and 7G8 is amodiaquine-resistant. 803 and GB4 share the same CQ-resistant $PfCRT_{72-76}$ haplotype (CVIET). This eliminated one of the drivers of differential drug response seen in Yuan et al.'s study. In this study, only two compounds, 1-bis-4-fluorophenyl methyl piperazine and rimcazole, were linked to a region on chromosome 7 containing *pfcrt*. The association was weak, with a p-value of 0.03 for 1-bis-4-fluorophenyl methyl piperazine, and 0.05 for rimcazole. Yuan et al. also linked rimcazole to the *pfcrt* locus. These findings may indicate that there are further mutations in *pfcrt* that have not been characterized, but that their role is minor in modulating drug response.

Yuan et al. linked response to 28 compounds to a locus containing *dhfr*. *Dhfr* polymorphisms were also present between the parental lines of the previous crosses: HB3 is mutant at codon 108, Dd2 is mutant at codons 51, 59, and 108. 803 also has three mutations in DHFR at codons 51, 59, and 108, rendering it virtually insensitive to some antifolate compounds. In the present study, six antifolate compounds were linked to a region on chromosome 4 containing *dhfr*, with p-values ranging from <0.001 for methodichlorophen to 0.02 for ormetoprim. When observing the variability in the IC_{50} values for the antifolate compounds, it appeared that genes other than *dhfr* may modulate response to some compounds. To examine this hypothesis, a second QTL scan was performed using additional progeny and controlling for the locus containing *dhfr*. The second scan failed to identify additional loci affecting the response to the six antifolates, which could be due to the relatively low number of progeny used in the analysis. In a yeast genetic cross, Bloom et al. found that 200 progeny were necessary to explain 50% of additive genetic variance affecting phenotype[227]. In the current genetic cross, 31 unique recombinants have been identified, with the potential for 2-3 times that many. Testing 30-60 additional progeny

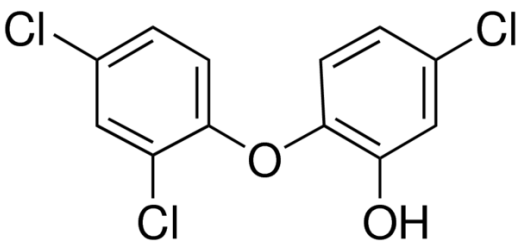
would greatly increase the power of the analysis, and provide valuable insight into the interaction of genes affecting drug response.

The locus containing *pfmdr1*, the third gene identified by Yuan et al., was linked to response to 50 compounds in their study. *Pfmdr1* was also polymorphic between the parental lines of previous crosses: HB3 is mutant at codons 184 and 1042; Dd2 is mutant at codon 86 and has a gene amplification; GB4 is mutant at codons 86 and 184; and 7G8 is mutant at codons 184, 1034, 1042, and 1246. Mutations in these five codons have been associated with varying levels of resistance to many antimalarial drugs[228–232]. Amplification and overexpression of *pfmdr1* has also been implicated in antimalarial resistance, especially to MFQ[226,233,234].

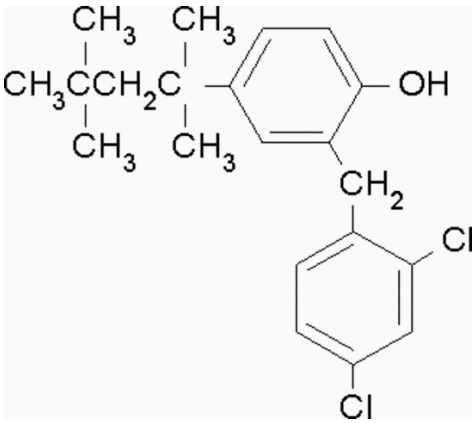
The current study identified a locus on chromosome 5 containing *pfmdr1* that was linked to parasite response to six compounds, including MFQ. While there was no evidence for amplification of *pfmdr1* in 803, there was evidence for increased *pfmdr1* transcription relative to GB4. Overexpression in the absence of amplification of the mammalian P-glycoprotein homologue of *pfmdr1* has been previously shown to occur in some human cancers[235]. In the bacteria *Pseudomonas aeruginosa*, triclosan resistance has been shown to be mediated by overexpression of a multidrug efflux pump mediated by mutations in its regulatory gene[236]. In *P. falciparum*, *pfmdr1* expression has been shown to increase following drug pressure, in what is termed an induction model[237]. However, little is known about the regulation of *pfmdr1* expression[238]. The finding of enhanced *pfmdr1* transcription merits further investigation, including transcription analysis of the progeny, confirmation of the effect on drug response, and further dissection of the regulation of *pfmdr1* expression.

4.5.2 Triclosan response linked to a locus containing pfmdr1

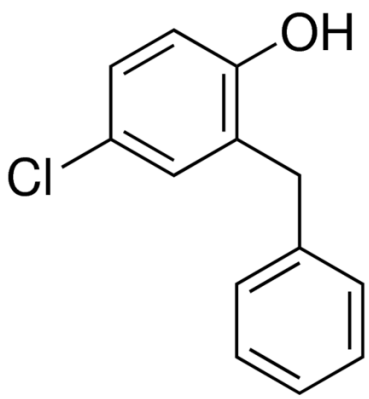
Parasite responses to triclosan and structurally related compounds 2-benzyl-4-chlorophenol and clofoctol (Figure 4-14) were linked to a locus containing *pfmdr1*. Triclosan is currently under investigation as a lead compound for antimalarial development[239,240]. Triclosan is an antibiotic that has been shown to offer protection against blood-stage *P. falciparum* by inhibiting enoyl-acyl carrier protein reductase[241–243]. It was previously linked to a locus containing *pfmdr1* by Yuan et al. through a GWAS with *P. falciparum* field isolates, but the authors were unable to detect linkage using QTL analysis and a genetic cross[179]. If triclosan resistance in the malaria parasite is mediated by increased *pfmdr1* expression, consideration will have to be given to its suitability as a lead compound for antimalarial development. For example, if triclosan were to be added to ACTs in southeast Asia, concerns about parasite resistance due to *pfmdr1* mutations would need to be addressed. If it were ineffective in this parasite population, it would not be a suitable candidate for ACT.



Triclosan



Clofoctol



2-benzyl-4-chlorophenol

Figure 4-14 Structures of triclosan, 2-benzyl-4-chlorophenol, and clofoctol. Chemical structures of three structurally related compounds linked to a locus containing *pfmdr1*. Triclosan and clofoctol share a 2,4-dichlorobenzene moiety. Triclosan and 2-benzyl-4-chlorophenol share a chlorophenol moiety.

4.5.3 QN and CQ response are not linked to known loci

An additional finding from this compound screen is that in spite of differential responses, QN and CQ failed to demonstrate linkage to any loci. Some studies have linked QN resistance to polymorphisms in *P. falciparum* Na^+/H^+ exporter (PfNHE)[168]. Follow up studies have variably supported and refuted this finding, indicating that the genetic mechanism behind QN resistance is complex[244–246]. In the current study, no linkage was found for QN response, even though the parental lines showed a 3.5-fold difference in QN IC_{50}. 803 and GB4 also displayed a two-fold difference in CQ response. Despite this difference, no linkage was found for CQ response. These findings serve to illustrate the fact that the interplay between genetic mutations and drug response is complex, and that there may be interacting loci at play that the current study is unable to detect.

4.5.4 ARM response not linked to known loci

In this screen, ARM and its derivatives were not linked to the previously identified *kelch13* gene on chromosome 13[120]. Polymorphisms within this gene have been shown to modulate parasite response to ARM derivative both *in vitro* and *in vivo*. The reason for this is because traditional *in vitro* assays fail to distinguish between ARM-resistant and ARM-sensitive parasites[117]. A specialized ring-survival assay (RSA) is required[247]. In this assay, 2-hour post-invasion rings are exposed to a 6 hour pulse of 700 nM ARM or derivative and then allowed to grow for a further 66 hours. Parasite growth is then assessed by comparing parasitemia in the treated well with the untreated well. In this assay, ARM resistant parasites exhibit >10% growth, while sensitive parasites exhibit <10% growth. QTL analysis was performed on RSA results from this genetic cross, and a large peak was observed on chromosome 13

that contains the *kelch13* gene (Sa et al, unpublished; see Appendix I for a list of contributors).

5 Analysis of WGS Data

5.1 Summary

WGS data analysis provided a wealth of information about the structure and variation of the genomes of the parents and progeny of this genetic cross. With the most in-depth analysis of a genetic cross to date, it was possible to infer CNV events genome-wide, which allowed for visualization of large-scale deletion or amplification events. A filtered set of variants also enabled further investigation of the scope and magnitude of the variation in this cross. WGS data were compared with Affymetrix Array microarray data, allowing for analysis of recombination parameters derived from each dataset.

5.2 Introduction

The *P. falciparum* genome is difficult to study for several reasons. It has an extreme nucleotide bias, with 80.6% A+T in the whole genome, and 90% in non-coding regions[142]. The genome is also highly repetitive, with short tandem repeats and low complexity sequence found in both coding and non-coding regions[248–250]. Furthermore, highly polymorphic RBC surface antigens undergo both meiotic and mitotic recombination[194,251], leading to a high level of diversity that is difficult to estimate[30,31]. These challenges notwithstanding, whole-genome deep sequencing of *P. falciparum* has provided a diverse array of insights into its natural diversity and population structure[149,252], mechanisms behind generation of antigenic diversity[251,253], and the genetic basis of artemisinin resistance[120,150,254].

The incredible plasticity of the *P. falciparum* genome is one of the greatest weapons in its arsenal. It has enabled the parasite to evolve resistance against most major classes of antimalarial drugs[101,114,255], facilitated the parasite's evasion of the

immune system[256], and thwarted efforts towards developing a vaccine[257,258]. Gaining insight into the mechanisms and parameters by which this plasticity operates will help to inform our strategies moving forward, whether it be drug development or vaccine design.

The 803xGB4 genetic cross provided a unique opportunity for in-depth study of the *P. falciparum* genome. High-coverage sequence data for both parental lines allowed for generation of a high-confidence callset of SNPs and Indels both in coding and non-coding regions. Coverage data was used to infer CNVs, which were detected throughout the genome. Such high resolution also allowed for fine mapping of recombination events, which were compared to recombination metrics inferred from SNP array data.

Generating a filtered callset of high-confidence variants was a challenging process. High quality gDNA was required for the parental lines and all of the progeny. Mapping raw sequence data to the reference genome (Pf3D7_v3) was difficult because of the hypervariable and repeat regions described above. Once mapped to the reference, variant calls were made which had to be filtered to generate a high-confidence set of SNPs and indels. It will then present results of an analysis of this callset, including evidence for CNV, SNP and indel prevalence, and recombination parameters.

5.3 Objectives

i. To generate a callset of variants from WGS data.

ii. To examine the WGS data for evidence of structural variation.

iii. To analyze the recombination parameters in this genetic cross using both WGS and SNP array data and to compare these results with previous genetic crosses.

5.4 Results

5.4.1 Sequence Alignment and Quality Check

Raw sequence data were processed as described in Section 2.4.1. Several factors were examined in the initial quality check of the WGS data. For each sample, total read length, percent of reads mapped and unmapped, and percent of paired and unpaired reads were examined to determine if any samples yielded aberrant results. Regions with near-complete coverage and unambiguous alignment to the 3D7 reference genome were labeled as the core genome. The core genome is punctuated by hypervariable regions containing paralogous and highly divergent sequence, and subtelomeric repeats containing low-complexity sequence.

Variant density across the genome can be used to visualize the difference between the core and non-core genome. As seen in previous crosses (Miles et al., in preparation), variant density was low in the core genome, and spiked in the hypervariable regions (Figure 5-1). The hypervariable regions contain highly polymorphic genes that encode proteins exposed to the human immune system. These genes undergo both meiotic and mitotic recombination, yielding highly divergent sequences. As a consequence, mapping these reads to a reference sequence (Pf3D7_v3) generates a huge spike in variant density in these regions.

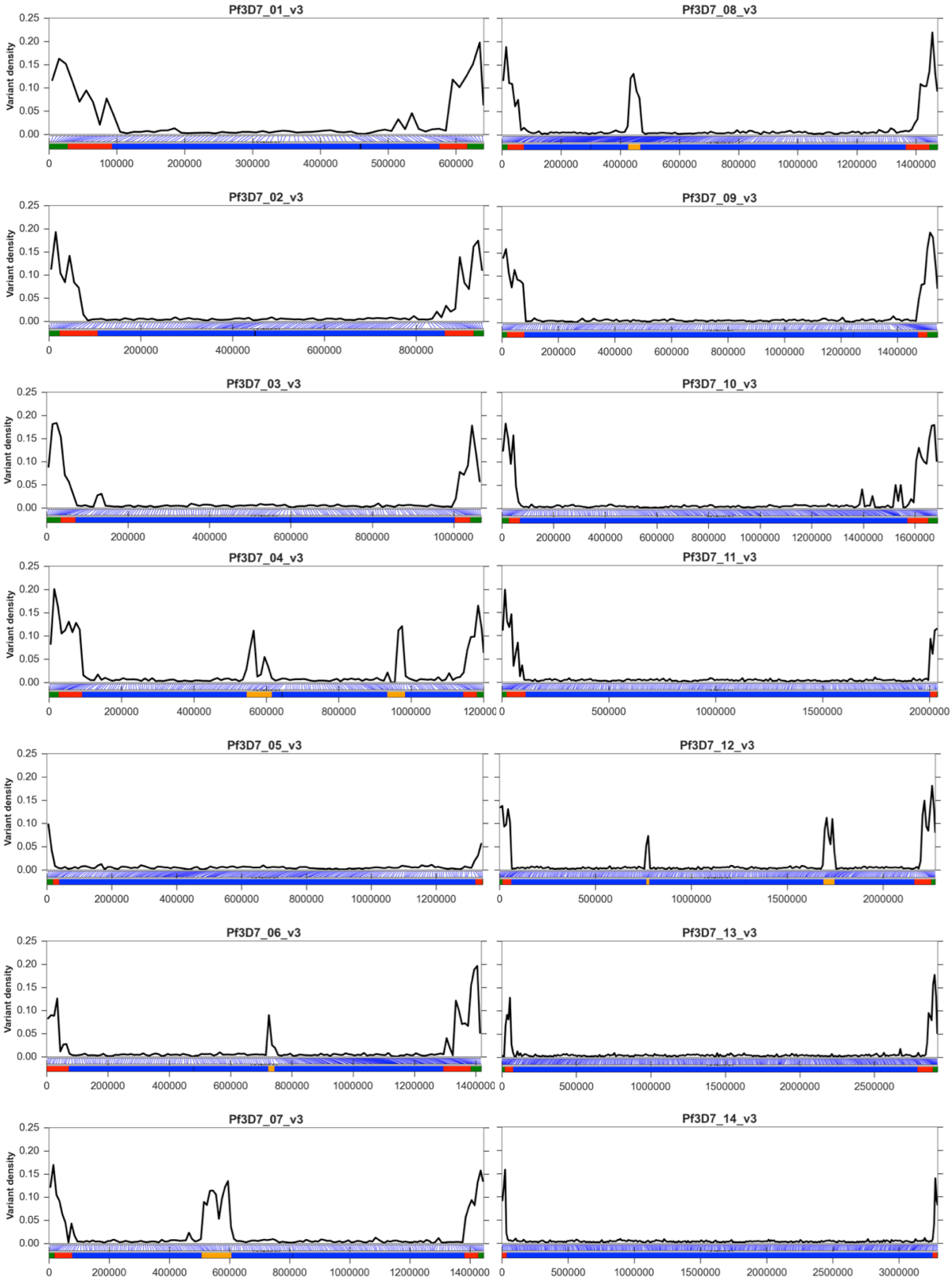


Figure 5-1 Variant density across each chromosome. Below each density plot, a variant locator marks the physical location of each variant along the chromosome. A painted model of the chromosome was added to indicate the locations of the subtelomeric repeats (green), subtelomeric hypervariable regions (red), internal hypervariable regions if present (orange), core genome (blue), and centromere (black).

5.4.2 Characterization of a potential copy number variant

The hidden Markov model (described in Section 2.4.4) identified several possible CNVs. As expected, the model was unreliable in subtelomeric and hypervariable regions, showing large numbers of deletions and amplifications due to the difficulty in mapping reads in these regions to the reference genome. Here, I will discuss five possible CNVs, three of which were non-Mendelian (not present in either parent), one which was Mendelian (present in one parent and inherited in some of the progeny), and one that was an interesting mixture of the two (a deletion that was present in one parent and a proportion of the progeny, and an amplification that was present only in three progeny). The last example will be explored in more depth, as it involved a possible genes implicated in erythrocyte invasion[259,260].

On chromosome 1, the model indicated that a ~30-kb non-Mendelian deletion had occurred (Figure 5-2A). It was observed in lines 38C7, 11C2, and 46G9, and spanned from ~94 to 115 kb. There are six genes in this region: *PF3D7_0102000* [*Plasmodium* exported protein (PEP), unknown function]; *PF3D7_0102100* (stevor); *PF3D7_0102200* [ring-infected surface antigen (RESA)]; *PF3D7_0102300* (PEP, unknown function); *PF3D7_0102400* (putative lysophospholipase); and *PF3D7_0102500* (erythrocyte biding antigen 181) (Figure 5-3A). Two other regions on chromosome 1 had apparent deletions. One, present in 803 and 40G11, contained only PfEMP1. Predicting CNVs in highly polymorphic genes and in regions with low complexity was problematic, because mapping reads in these regions was difficult. The other deletion indicated by the model was in 34F5, a twin of parental line GB4. This region spanned from 558 to 579 kb, and contained six genes: *PF3D7_0114500* (PEP, unknown function); *PF3D7_0114600* (stevor); *PF3D7_0114700* (rifin); *PF3D7_0114800* (PEP, unknown function); *PF3D7_0114900* (PEP, unknown function); and *PF3D7_0115000* (surfin).

On chromosome 13, the model showed a deletion from ~1422 to 1430 kb in 803 and several of the progeny, and an amplification from ~1430 to 1445 kb in three progeny (Figure 5-2B). This region contained three genes: *PF3D7_1335200* [reticulocyte binding homologue (Rh) 6, pseudogene]; *PF3D7_1335300* (*Rh2b*); and *PF3D7_1335400* (*Rh2a*) (Figure 5-3B). The model indicated a possible deletion of *Rh6* and *Rh2b* in 803 and the progeny that inherited its allele. In this region, there is zero coverage over *Rh6* and what appears to be an average of half coverage over part of *Rh2b* and all of *Rh2a*. In 39C3, 36D5, and 36E5, the model showed an amplification of *Rh2b* and *Rh2a*. There appears to be 2x coverage over these genes, which may indicate that the whole region has been amplified, resulting in two copies of *Rh2b* and two copies of *Rh2a*. The remaining progeny inherited the GB4 allele, with no indication of deletion or amplification in this region.

On chromosome 3, the model showed a Mendelian deletion present in GB4 and several of the progeny spanning from ~119 to 135 kb (Figure 5-2C). This region encoded two products: Pf3D7_0302300 (*P. falciparum* erythrocyte membrane protein 1); and PF3D7_0302200 [cytoadherence-linked antigen (clag) 3.2] (Figure 5-3C). The deletion of *Clag3.2* has been previously reported in GB4[261,262], so this finding served to validate the Hidden Markov model.

On chromosome 11, the model showed an interesting pattern of either deletion or amplification of a large region on the end of the chromosome (Figure 5-2D). The deletion/amplification event was non-Mendelian. Nine progeny showed a deletion in this region, six showed an amplification, and the rest were normal. The region spanned from 1930 to 2000 kb, and contained nine genes: *Pf3D7_1148700* (PEP, unknown function); *Pf3D7_1148800* (PEP, unknown function); *Pf3D7_1148900* (PEP, unknown function); *Pf3D7_1149000* (Duffy binding-like protein); *Pf3D7_1149100* (PEP, unknown function); *Pf3D7_1149200* (RESA);

Pf3D7_1149300 (serine/threonine protein kinase); *Pf3D7_1149400* (PEP, unknown function); and *Pf3D7_1149500* (RESA 2).

On chromosome 12, the model showed a non-Mendelian amplification from 970 to 980 kb in a proportion of the progeny (Figure 5-2E). This region contained five genes: *Pf3D7_3700* (putative integral membrane protein); *Pf3D7_1223800* (putative mitochondrial carrier protein); *Pf3D7_1223900* (putative 50S ribosomal protein L24); *Pf3D7_1224000* [GTP cyclohydrolase I (GCH-1)]; and *Pf3D7_1224100* (conserved *Plasmodium* protein, unknown function). GB4 has been shown to have a CNV in a region spanning GCH-1 (Miles et al., in preparation). The GB4 sequenced in this context was an independent isolation of genomic DNA from GB4 cultured in our laboratory.

The CNV on chromosome 13 was further investigated by PCR and quantitative PCR (qPCR) to confirm the deletion and amplification events, as described in Section 2.3.3. Five lines were chosen: 803, GB4, three lines with the amplification (36D5, 36E5, and 39C3), and one line that inherited the 803 allele (11C2). The deletions of *Rh6* and *Rh2b* were confirmed by PCR. Primers specific to *Rh6* yielded no product for 803 and 11C2, and product of the expected size for the remaining lines (Figure 5-4A). A shared forward primer and unique reverse primers for *Rh2a* and *Rh2b* confirmed that *Rh2b* was deleted in 803 and 11C2 (Figure 5-4B). Interestingly, the PCR product for *Rh2a* appeared to be smaller than expected in these two lines. PCR products for *Rh2a* and *Rh2b* were of the expected size for the remaining lines.

Rh2a and *Rh2b* copy number and expression level were assessed by qPCR as described in Section 2.3.3. As expected, single copies of *Rh2a* were present in 803 and GB4, and an amplification was observed in 39C3 (Figure 5-5A). *Rh2b* was deleted in 803 and amplified in 39C3 (Figure 5-5B). Relative to GB4, both 803 and 39C3 expressed higher levels of *Rh2a* (roughly 2.5-fold and 4.5-fold, respectively).

Expression of *Rh2b* was absent in 803, and increased roughly 12-fold in 39C3 relative to GB4.

Figure 5-2A: Chromosome 1

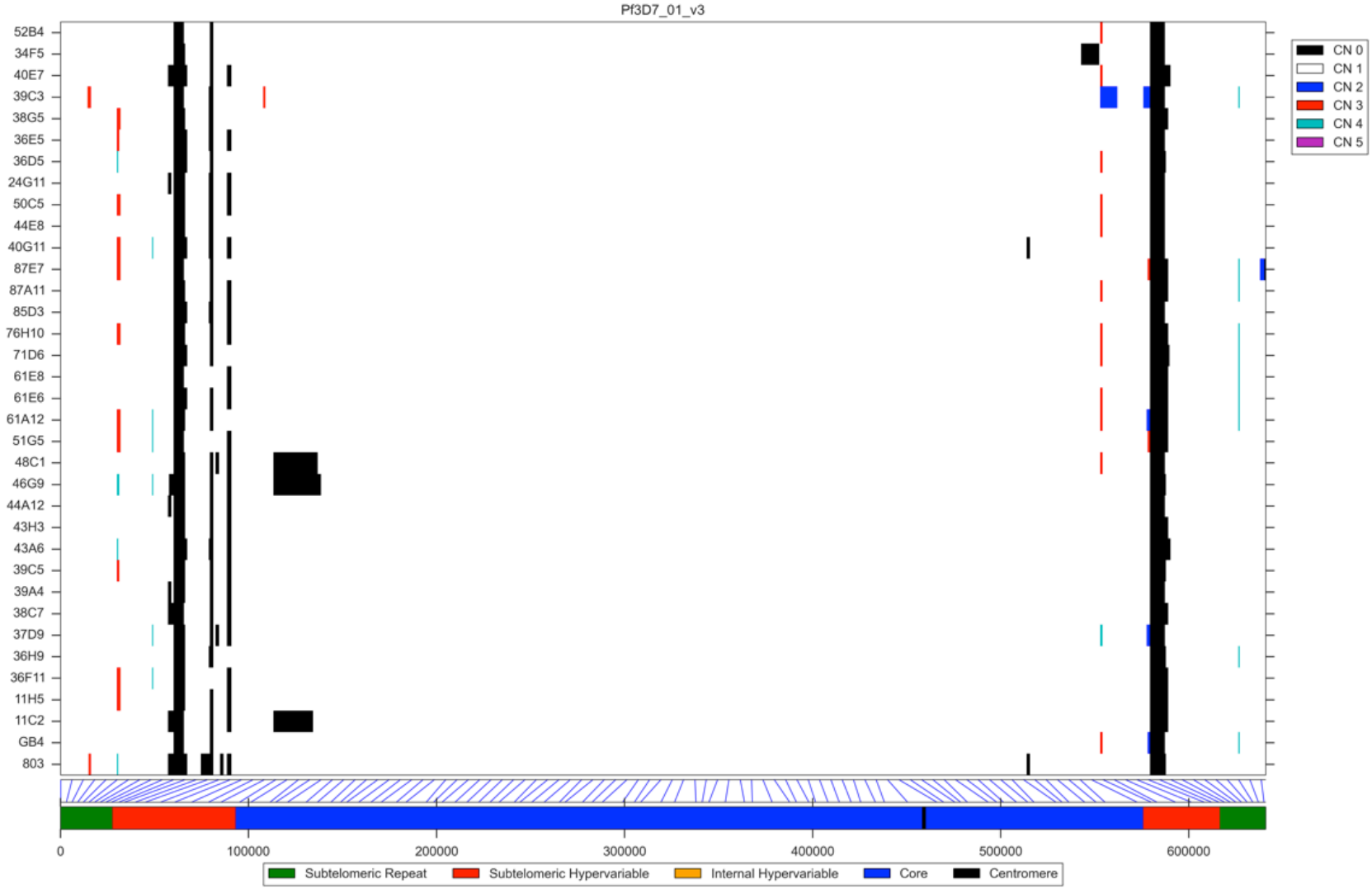


Figure 5-2**B: Chromosome 13**

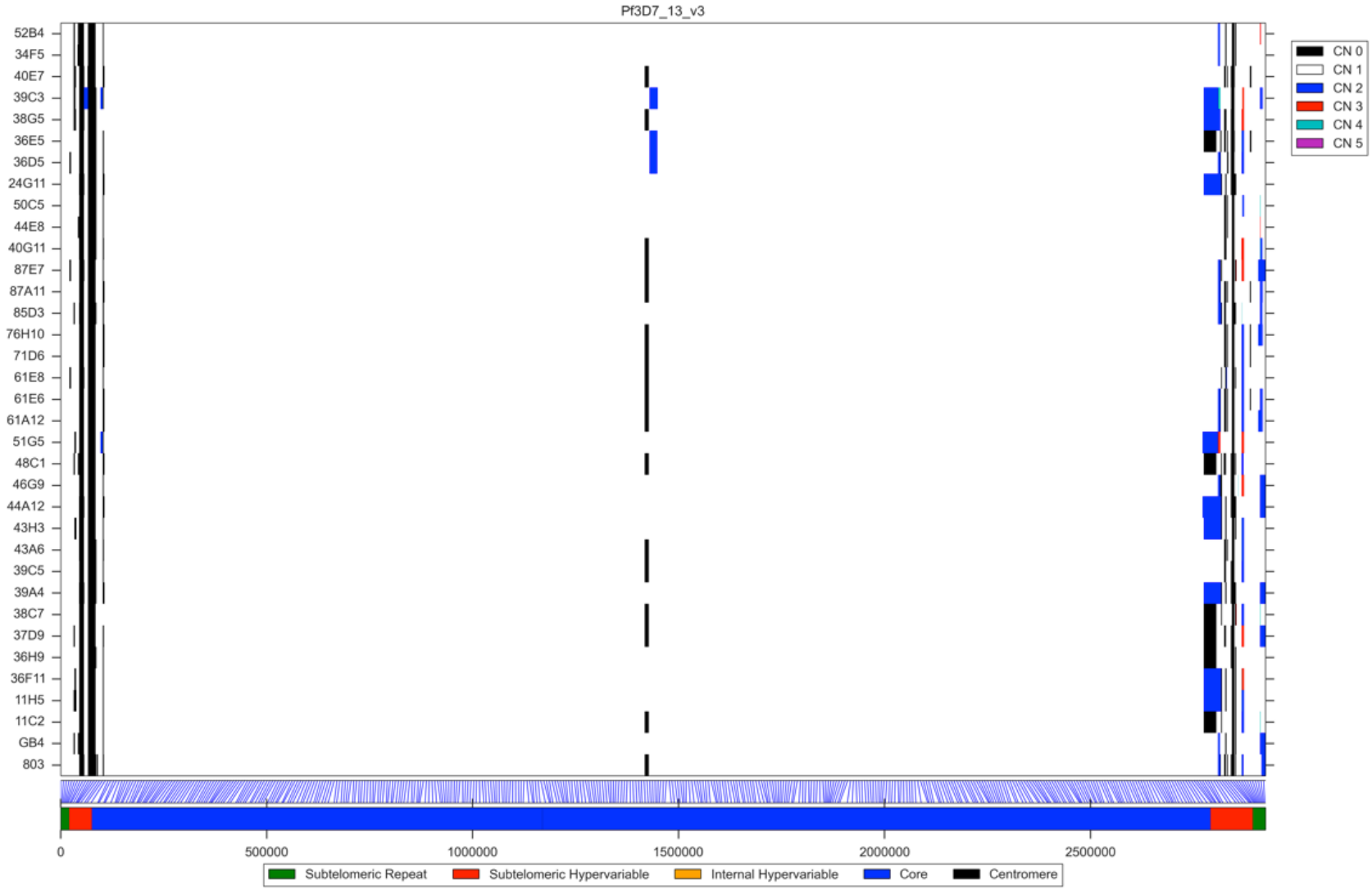


Figure 5-2C: Chromosome 3

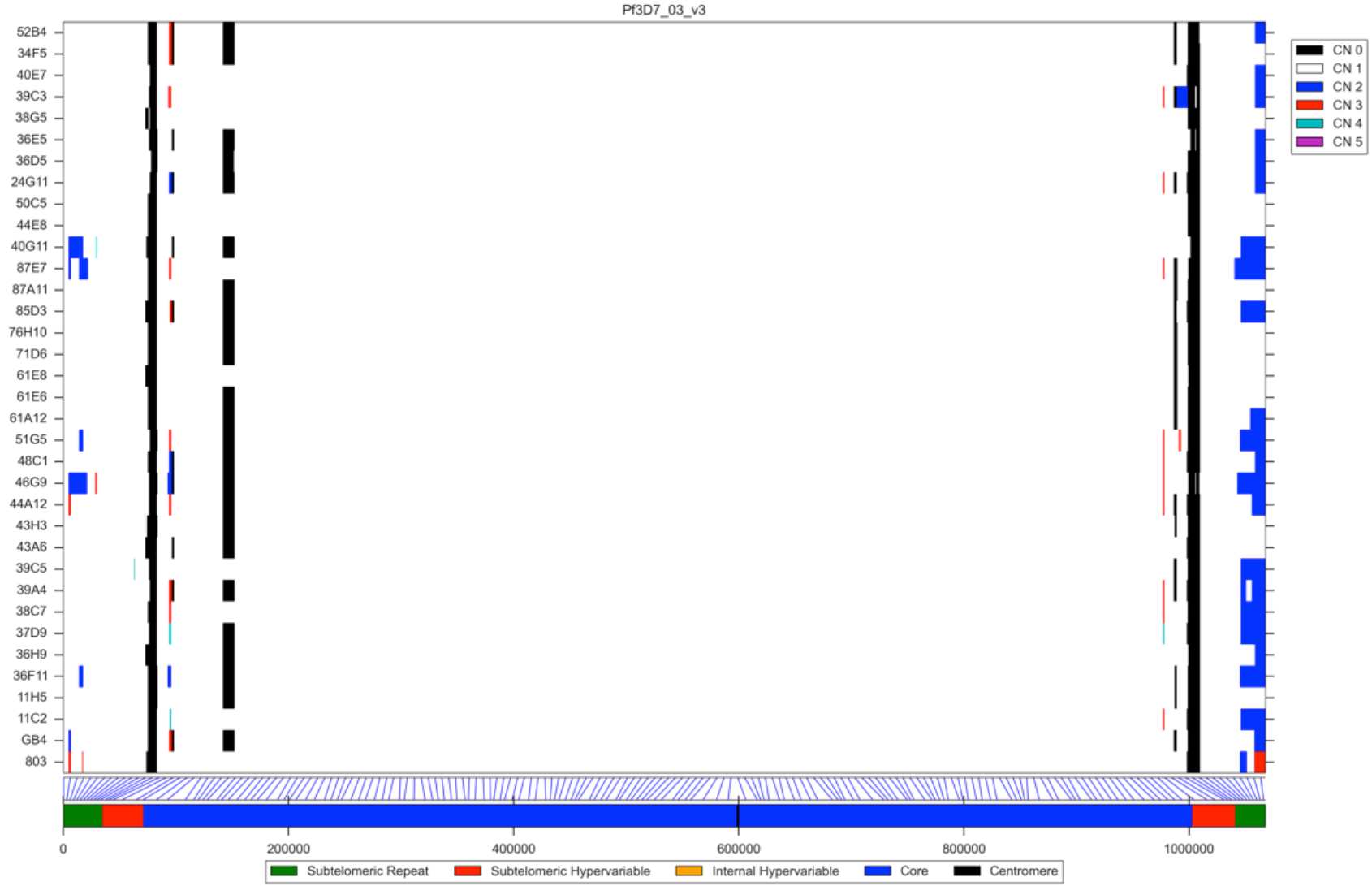


Figure 5-2D: Chromosome 11

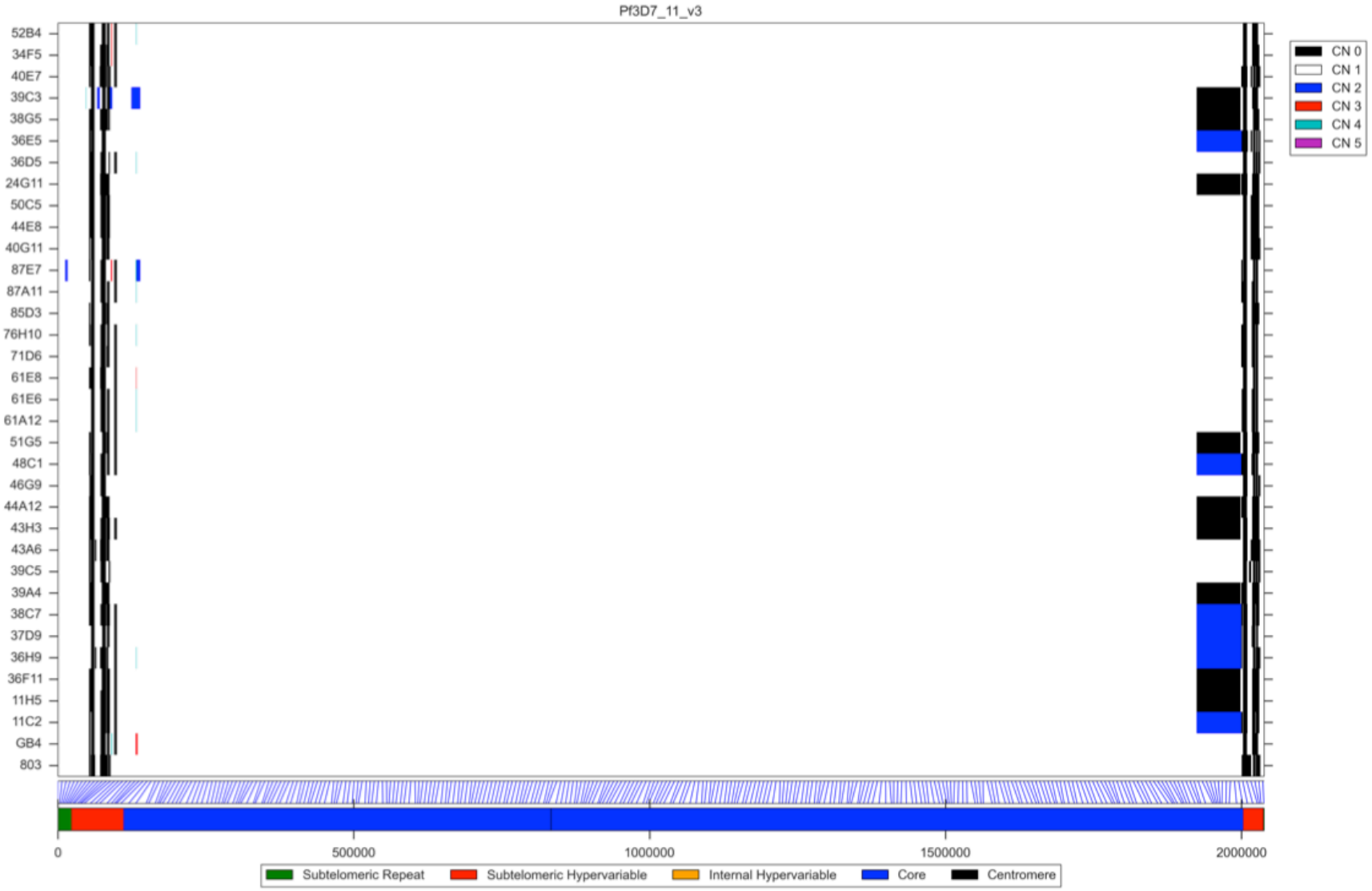


Figure 5-2**E: Chromosome 12**

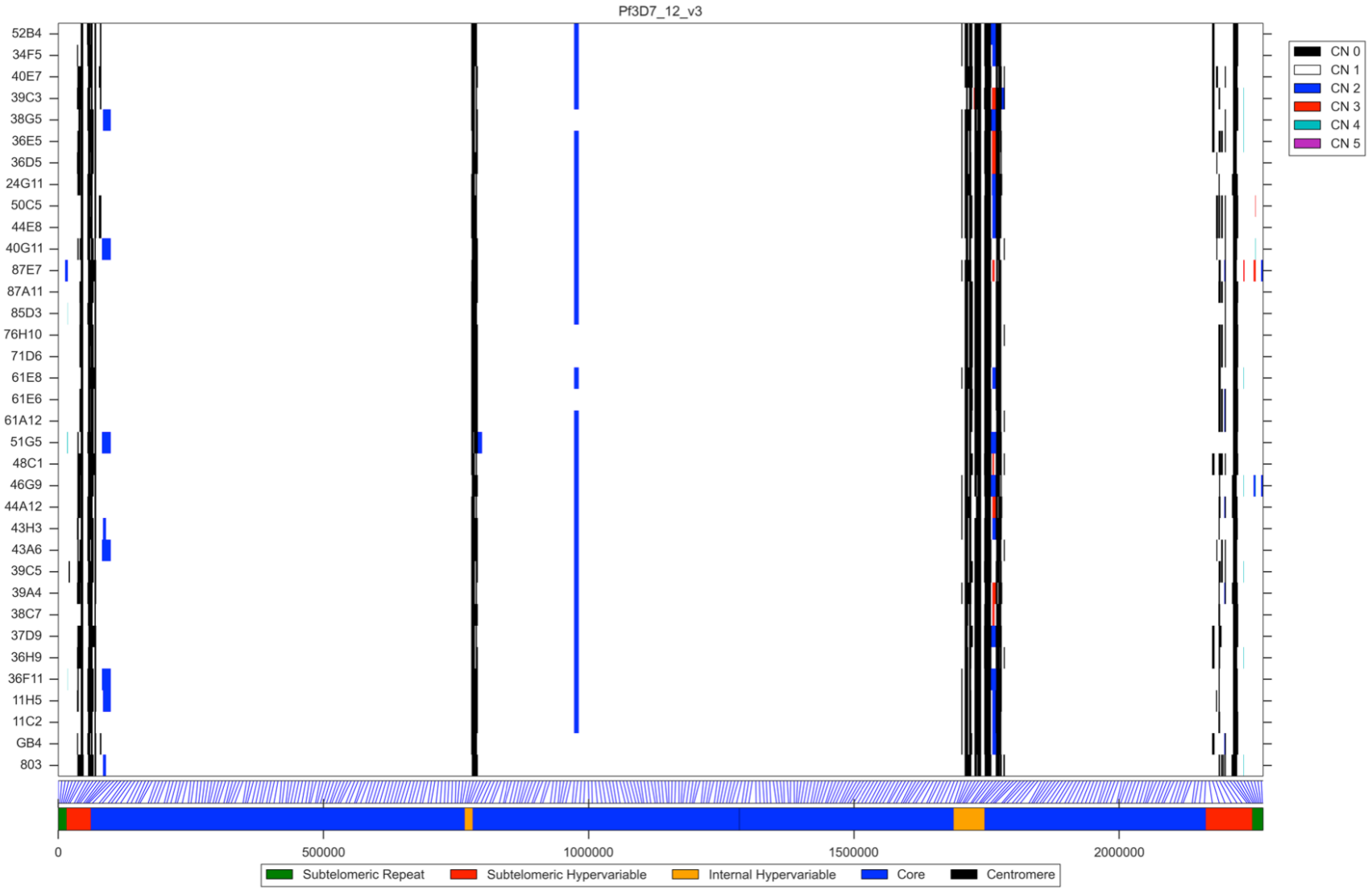


Figure 5-2 CNVs identified by hidden Markov model. (**A**) Chromosome 1; (**B**) Chromosome 13; (**C**) Chromosome 3; (**D**) Chromsome 11; (**E**) Chromsome 12. All progeny were included in order to identify any potential variation in sets of twins. A variant locator at the bottom of each panel marks the physical location of each variant in bp. Below each locator is a painted model of the chromosome indicating the location of the subtelomeric repeats, subtelomeric hypervariable regions, internal hypervariable regions, the core genome, and the centromere.

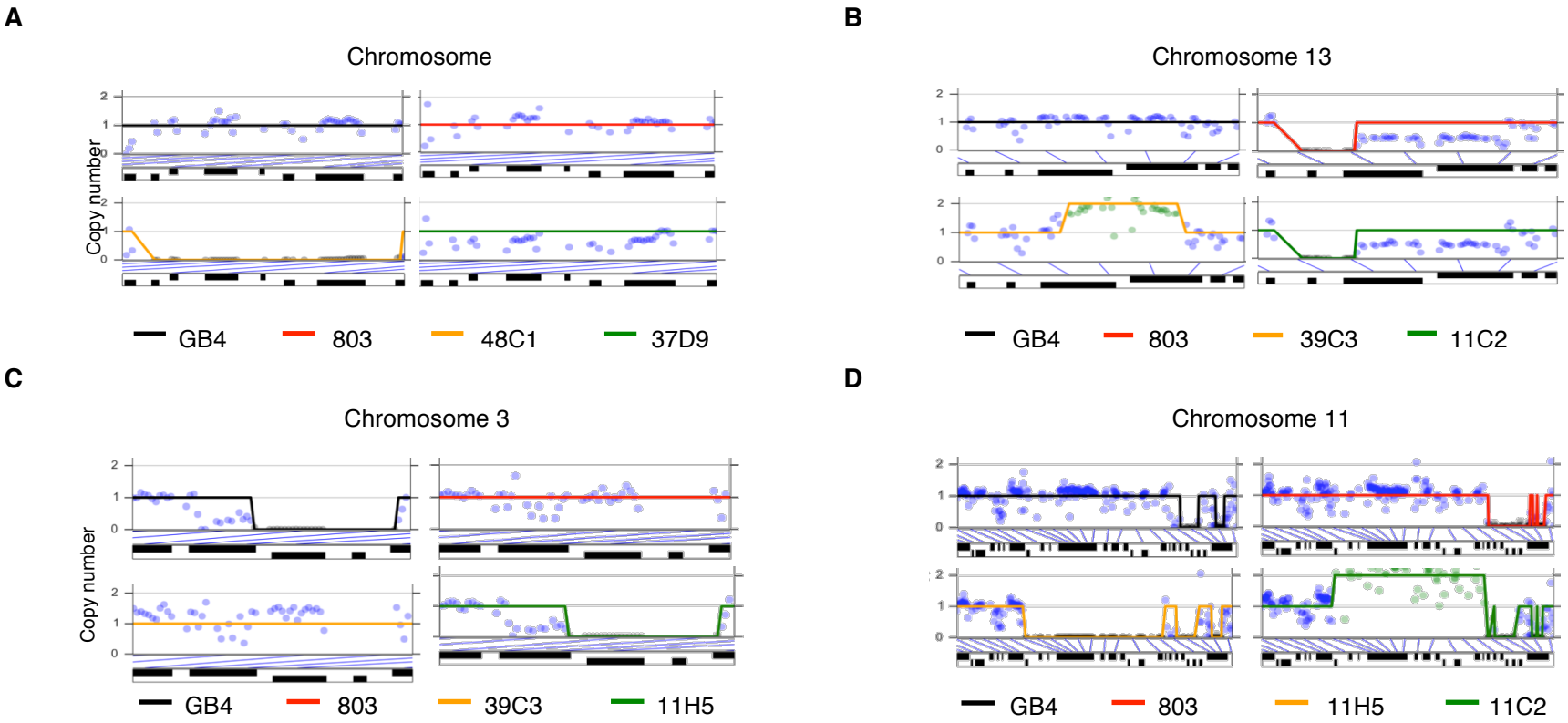


Figure 5-3 Coverage plots. Close-up coverage plots of the region of the CNVs. **(A)** A region on chromosome 1 with an apparent non-Mendelian deletion encompassing six genes; **(B)** A region on chromosome 13 involving three genes with a Mendelian deletion present in 803 and some progeny, and an amplification present in 39C3 and two other progeny; **(C)** A Mendelian deletion on chromosome 3 present in GB4 and some progeny that encompasses 2 genes; **(D)** A large region on chromosome 11 with apparent deletion in some progeny and amplification in other progeny. Neither deletion nor amplification is present in the parental lines.

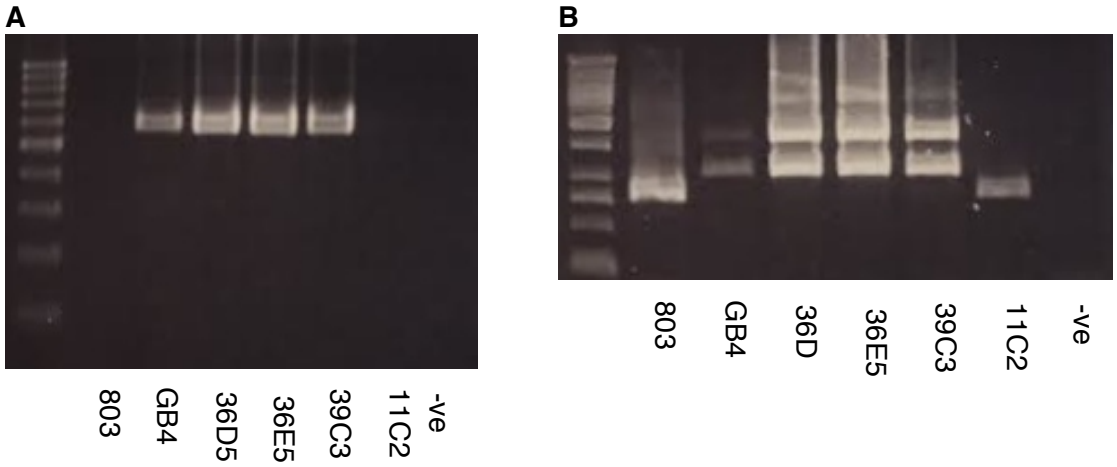


Figure 5-4 PCR validation of Rh deletions. Five parasite lines were chosen: 803, GB4, 36D6, 36E5, 39C3, and 11C2. (A) Agarose gel of PCR amplification products of *Rh6*. A 100-bp ladder was used with band sizes of 100 bp to 1 kb in 100-bp increments. The expected product size was 564 bp. (B) Agarose gel of PCR amplification products of *Rh2a* and *Rh2b*. A 1-kb ladder was used; clearly differentiated band sizes are 250 bp, 500 bp, 750 bp, 1 kb, 1.5 kb, and 2 kb. The expected band sizes are 1173 bp for *Rh2a* and 1779 bp for *Rh2b*.

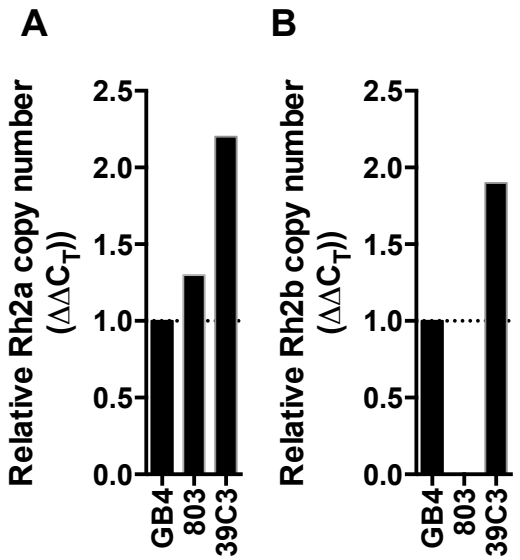


Figure 5-5 Relative *Rh2a* and *Rh2b* copy number. Relative copy number of (A) *Rh2a* and (B) *Rh2b*. Each reaction was performed in triplicate (n=1). $\Delta\Delta C_T$ was calculated relative to GB4.

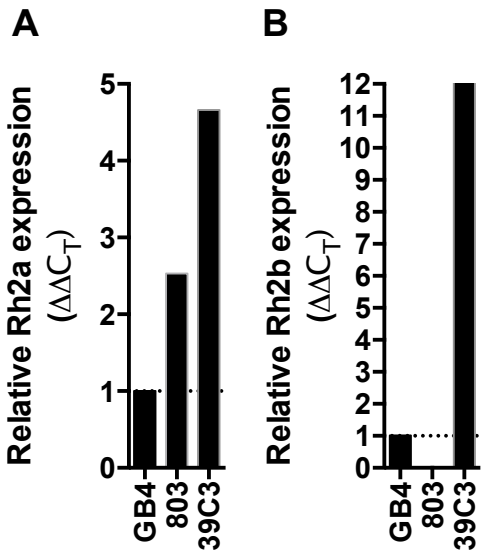


Figure 5-6 Relative expression of *Rh2a* and *Rh2b*. Relative expression of (A) *Rh2a* and (B) *Rh2b*. Each reaction was performed in triplicate (n=1). $\Delta\Delta C_T$ was calculated relative to GB4.

5.4.3 Variant filtering

VQSR was performed as described in Section 2.4.5. Following VQSR, the sensitivity (number of true variants identified) and 1 – specificity (percentage of Mendelian error calls) of the VQSLOD score were evaluated at several thresholds (4-6). The receiver-operator curves for several other variant metrics were plotted alongside VQSLOD in order to assess the recalibration. It was expected that the VQSLOD score would provide the highest number of variants identified with the lowest error rate. For SNPs, this proved to be the case (Figure 5-7A). A VQSLOD score of 5 identified approximately 13,000 SNPs with a Mendelian error rate of less than 0.1%. However, for indels, it was found that an additional filter was necessary to maximize the utility of the VQSLOD score. By filtering out calls with genotype quality (GQ) < 99, it was found that a VQSLOD threshold of 0 identified approximately 20,000 indels with an error rate of less than 0.1% (Figure 5-7D).

To illustrate the effect of this filtering step, an inheritance plot for chromosome 4 was generated to compare the raw variant dataset with the filtered dataset (Figure 5-8). From the raw VCF, over 28,000 variants were identified on chromosome 4

alone (Figure 5-8A). Many of these variants were missing in one or more progeny or were non-Mendelian inherited (i.e., present in the progeny but absent in either parent), which made the inheritance plot appear choppy with many apparent double crossovers. Applying the VQSLOD filter for SNPs and the GQ and VQSLOD filters for indels rendered the inheritance plot much more uniform (Figure 5-8), which is in line with what has been observed in the other three crosses. (Miles et al., in preparation).

A total of 33,649 variants passed the filters (Figure 5-9). A majority of these variants (20,765) were indels mainly found in intergenic regions. In contrast, most of the 12,884 SNPs were found in exons. Indels found within exons tended to occur in multiples of three, while those in introns occurred mostly in multiples of two (Figure 5-10). This made sense because indel lengths in multiples of three prevented frame-shifts from occurring.

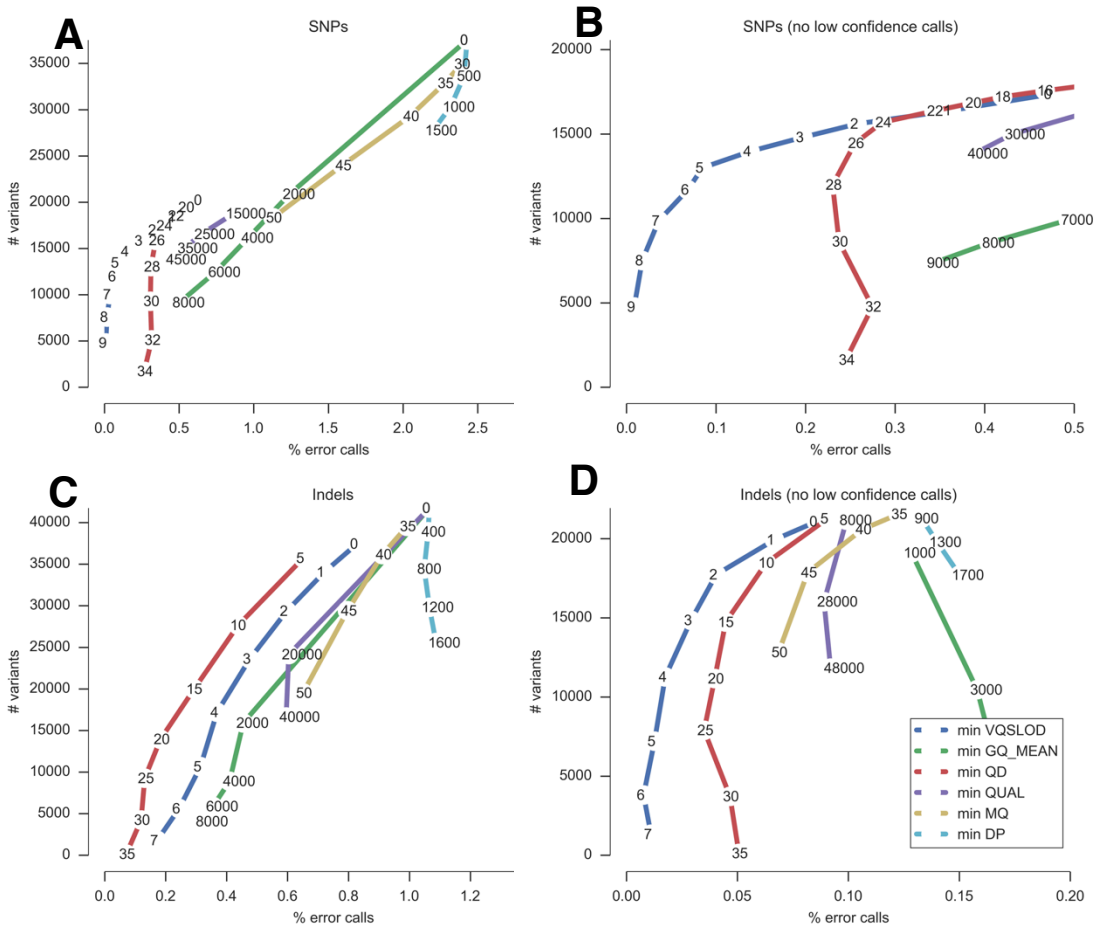


Figure 5-7 VQSR metrics for recalibration. Receiver Operating Characteristic (ROC) curves for variant metrics and VQSLOD. SNPs (A & B) and indels (C & D) were examined separately. Sensitivity (number of variants identified) and 1-Specificity (percentage of Mendelian error calls) were plotted at different thresholds of each metric. Eliminating low confidence calls (GQ < 99) greatly improved the specificity of the indel VQSLOD (D). Abbreviations: VQSLOD (Variant Quality Score Logarithm of Odds); GQ_Mean (phred-scaled confidence that the true genotype is the one provided in GT); QD (total unfiltered depth over all samples); QUAL (phred scaled probability that a REF/ALT polymorphism existed at this site given sequencing data); MQ (root mean square of the mapping quality over all samples); DP (Variant confidence from the QUAL field/unfiltered depth of non-reference samples).

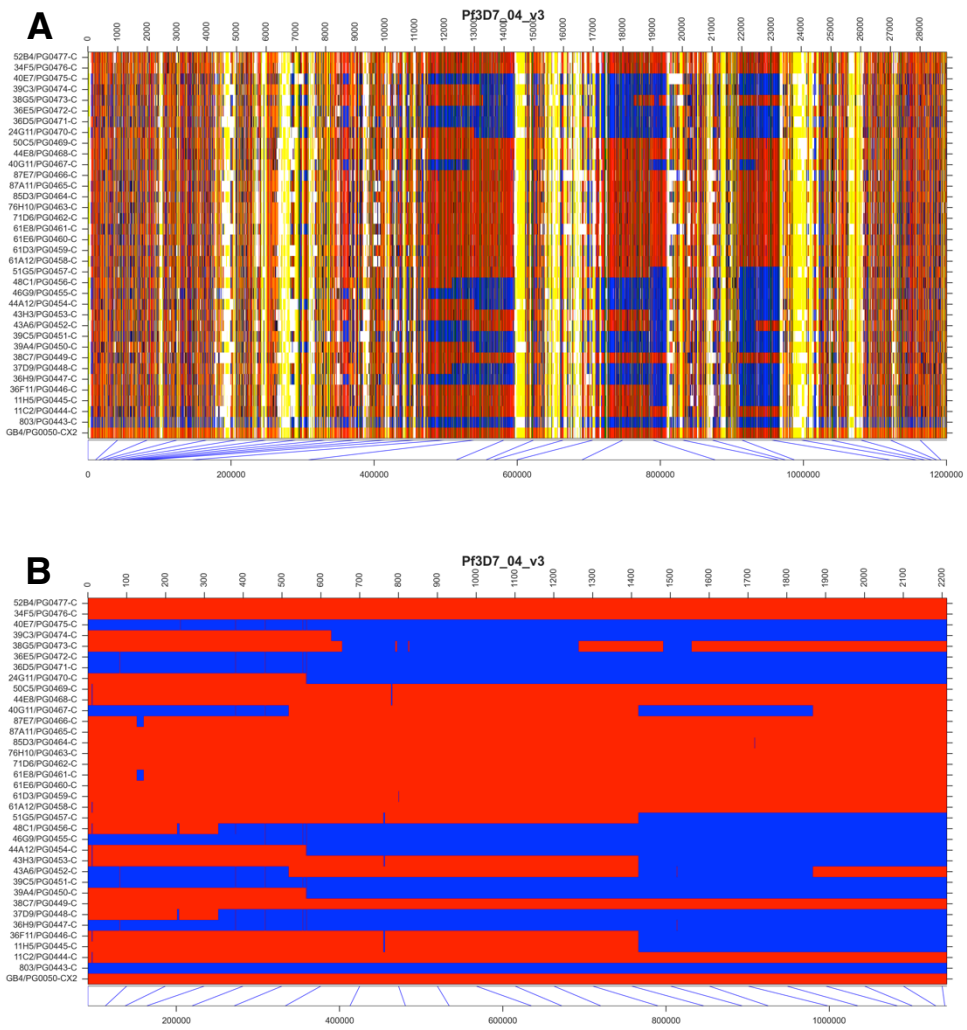


Figure 5-8 Colored inheritance plots. Comparison of inheritance of genotype calls (A) before and (B) after VQSR. Blue = inheritance from 803; red = inheritance from GB4; white = missing call; grey = filtered call; black = non-parental allele; yellow = one or both parents with a missing genotype call; orange = reference allele where parents are both reference; green = alternate allele where parents are both alternate.

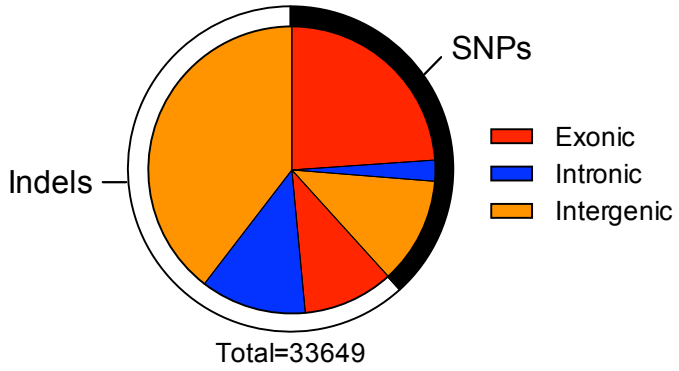


Figure 5-9 Breakdown of SNPs and indels. Summary of variants in the analysis-ready dataset. Of the 33,649 variants, 12,884 were SNPs (black) and 20,765 were indels (white). Of the SNPs, 8,042 were exonic, 796 were intronic, and 4,046 were intergenic. Of the indels, 3,437 were exonic, 4,011 were intronic, and 13,317 were intergenic.

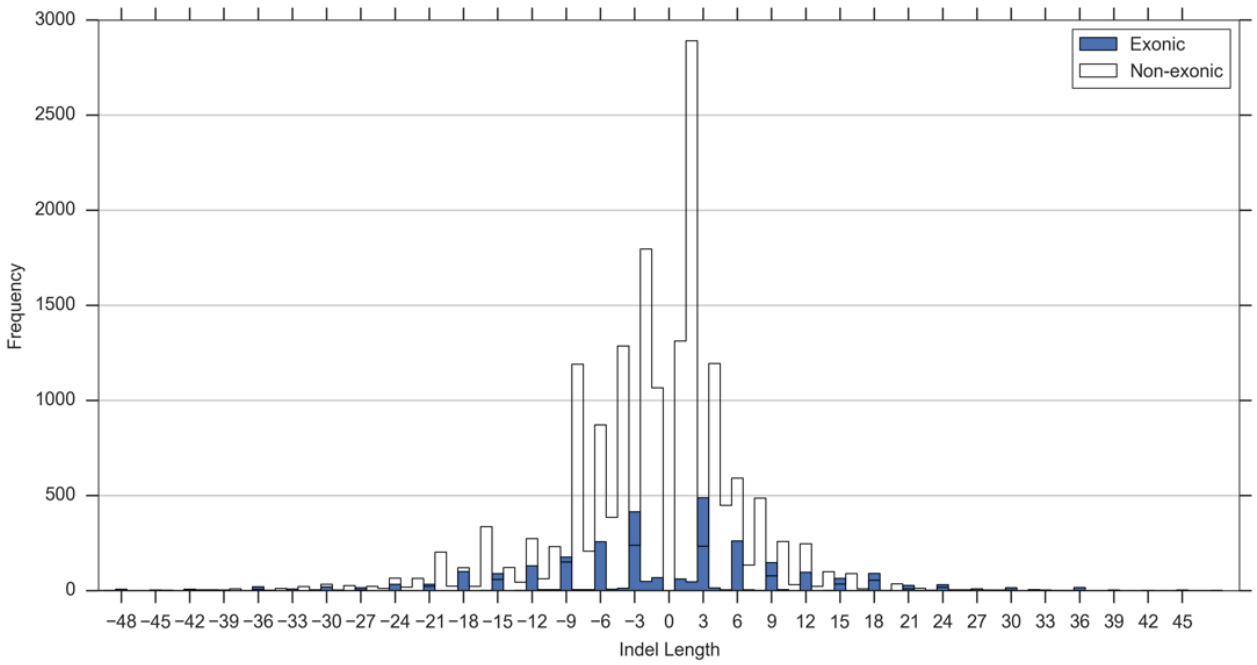


Figure 5-10 Indel length distribution. Histogram of indel lengths in exonic and non-exonic regions. Non-exonic regions refer to introns and intergenic regions. The majority of exonic indels occurred in multiples of 3, while non-exonic indels mostly occurred in multiples of 2.

5.4.4 Analysis of Isogenic lines

The filtered sets of variants were used to plot the pairwise distance between the progeny lines. From the results of the SNP array analysis, 17 unique progeny were expected (including 10 sets of family groups) in the 34 progeny sequenced. The results of the concordance analysis yielded the expected result. Next, the exact concordance between the isogenic lines was examined (Table 2-6). Overall, members within family groups were 99.95% identical, discordant by an average of 17 variants out of 33,649.

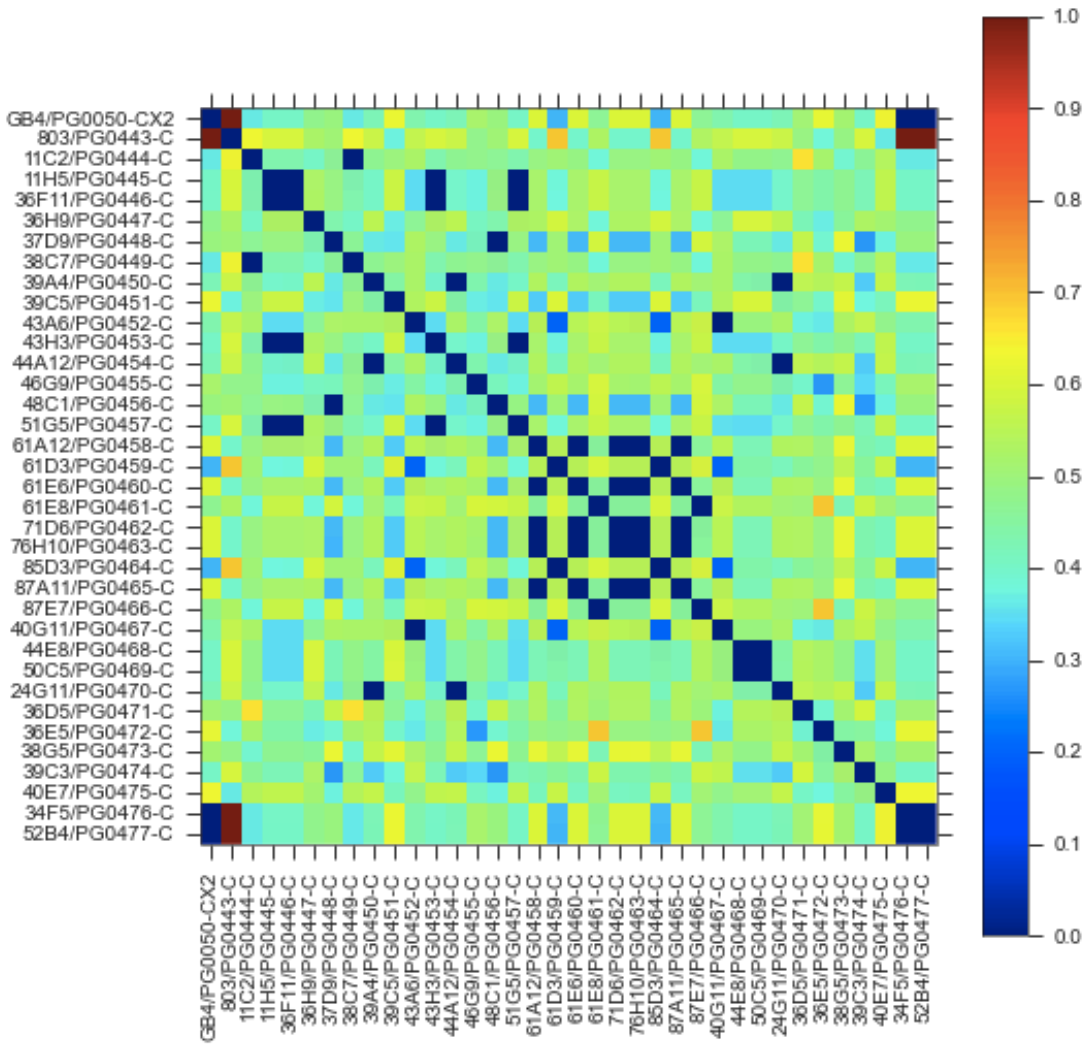


Figure 5-11 Concordance of twin lines. Pairwise distance plot of parents and progeny illustrating relatedness. The heatmap colors correlate with the degree of similarity between the parents and progeny lines, with intensity of blue indicating degree of relatedness and intensity of red indicating degree of difference. As expected, there were ten sets of family groups in the dataset: 1) GB4, 34F5, and 52B4; 2) 11C2 and 38C7; 3) 43H3, 11H5, 36F11, and 51G5; 4) 40G11 and 43A6; 5) 24G11, 39A4, and 44A12; 6) 71D6, 61A12, 87A11, 61E6, and 76H10; 7) 44E8 and 50C5; 8) 37D9 and 48C1; 9) 61E8 and 87E7; and 10) 85D3 and 61D3.

	44A12	39A4
24G11	12	8
44A12		12

	34F5	52B4
GB4	11	18
34F5		11

	36F11	51G5	43H3
11H5	22	15	24
36F11		21	18
51G5			23

	37D9
48C1	27

	85D3
61D3	17

	76H10	71D6	61E6	61A12
87A11	16	7	8	12
76H10		15	14	18
71D6			7	15
61E6				12

	40G11
43A6	18

	87E7
61E8	26

Total: 33649 variants

	50C5
44E8	20

	11C2
38C7	29

Table 5-1 Discordance of twin lines. Number of discordant variants (SNPs and indels) between twin lines for the ten sets of twins analyzed.

5.4.5 Inheritance patterns and recombination parameters

Filtered WGS data were used to generate summary plots of crossover counts and inheritance of the 803 allele (Figure 5-12). The WGS data showed the same pattern of biased inheritance of the GB4 allele at the ends of chromosomes 7 and 9 as seen in the SNP array data (see Appendix I). Further investigation into the genes within these regions may be warranted. However, it is possible that the biased inheritance observed in these regions is simply an artifact of the low number of progeny included in this analysis. It is also noteworthy that the distribution of crossover counts is much more uniform than that of the SNP array data. This is likely

because the variant marker density is ~10-fold greater in the WGS data than in the SNP array data (~33,000 WGS versus ~3,000 SNP array). This greater resolution allows for more refined mapping of crossovers.

The filtered set of WGS variants along with the SNP array data were used to define the recombination parameters in this genetic cross. To start, WGS data were further filtered to eliminate any blocks of inheritance that were less than 10 kb. This was done to filter out both possible error calls, and conversion tracts possibly associated with non-crossover events (NCO). Analysis of the WGS data revealed a total of 239 crossovers (CO) (14.1 CO/progeny), less than the 268 CO (15.7 CO/progeny) identified in the analysis of the SNP array data. These values are lower than the reported 638 CO in the 32 progeny of the 7G8 x GB4 cross (20 CO / progeny). [194]

The Haldane genetic map generated from the SNP array genotyping data was 1738 cM, indicating an average map unit distance of 13.3 kb/cM. The number of CO per chromosome was plotted against the physical length of each chromosome (4-12). A regression line was fit to the WGS data (Figure 5-13A) and the SNP array data (Figure 5-13B) with origin at 0 CO at 0 Mb. The WGS data generally fit the linear trend, with a regression line slope of 10.7 CO/Mb. For the SNP data, the linear relationship (11.2 CO/Mb) held true for the majority of chromosomes, save chromosomes 1 and 7. Chromosome 1 only had one crossover event observed in two progeny. This observation is likely a product of the limited number of progeny being analyzed. Chromosome 7 had 29 crossovers, and a map length of 222 cM, making its calculated length longer than any other chromosome except chromosome 14. This observation may be explained by the fact that many of the observed crossovers on chromosome 7 occurred in a hypervariable region.

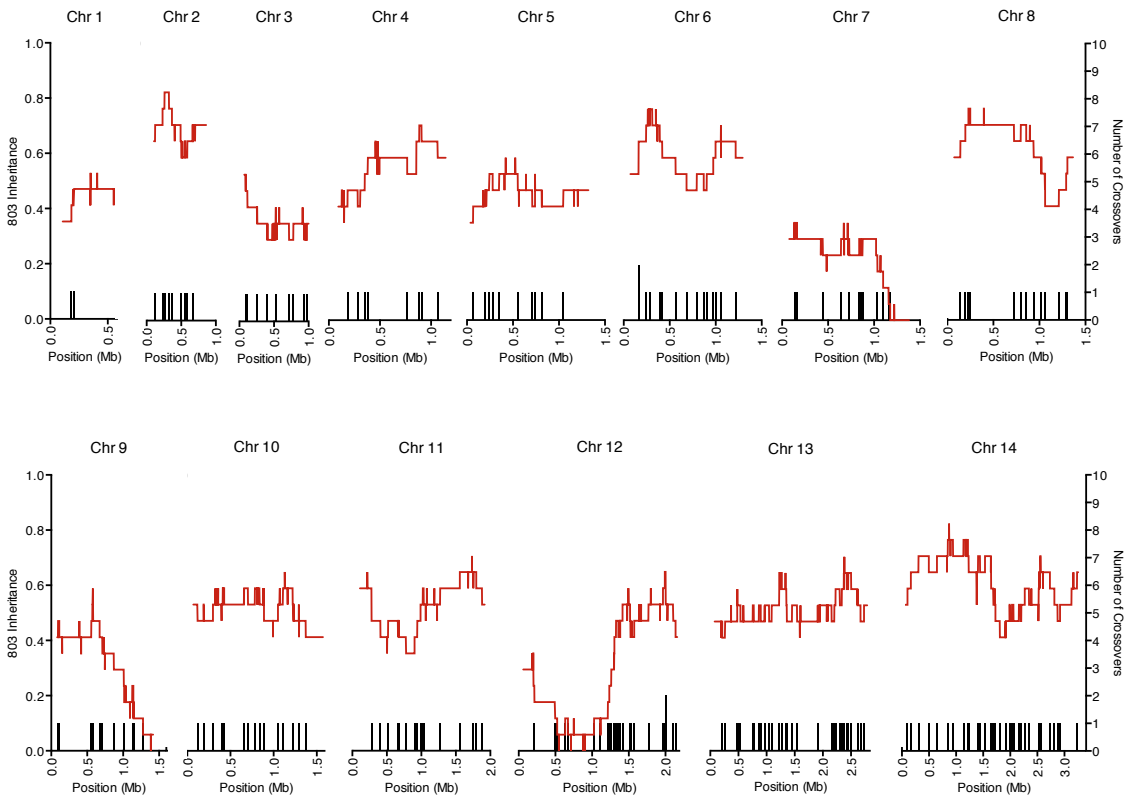


Figure 5-12 Crossover counts and inheritance bias. Crossover counts and inheritance of the 803 allele over the 14 chromosomes. The red lines indicate the proportion of progeny that inherited the 803 allele across the physical length of each chromosome. The black bars represent crossover counts.

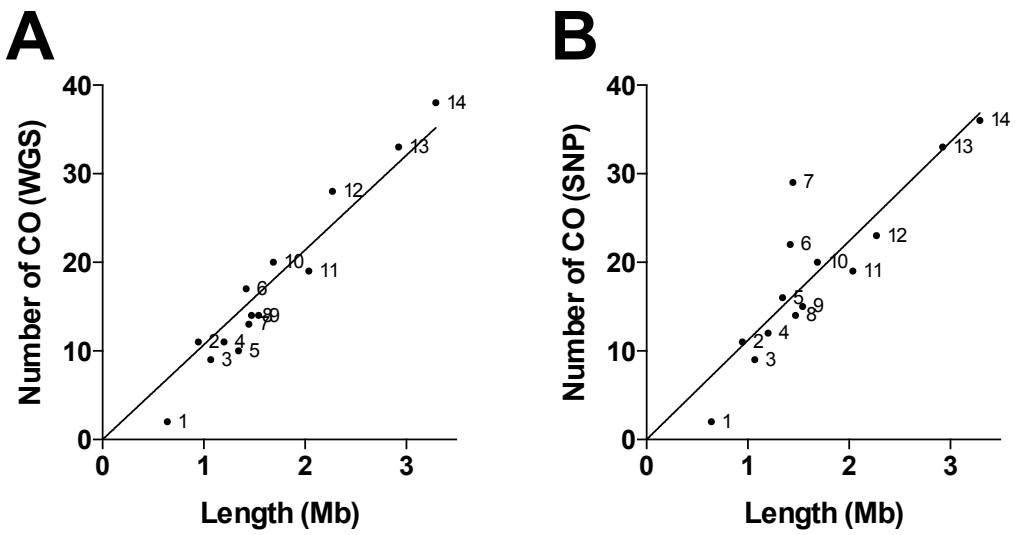


Figure 5-13 Recombination parameters. Number of crossovers observed in (A) WGS and (B) SNP array data plotted against the physical size of each chromosomes. Chromosome number is indicated next to each point. The regression line extrapolates to zero crossovers at zero physical length.

5.5 Discussion

5.5.1 Identification and validation of potential copy number variants

In this genetic cross, several possible CNVs were identified. Five examples, (Mendelian, non-Mendelian, or otherwise) were examined in more detail: a 30-kb deletion on chromosome 1, a 23-kb deletion/amplification on chromosome 13, a 16-kb deletion on chromosome 3, a 70-kb deletion/amplification on chromosome 11, and a 10-kb amplification on chromosome 12.

The deletion on chromosome 1 was non-Mendelian and involved a region containing six genes. One of the genes in this region, erythrocyte binding antigen (EBA)-181, has a role in erythrocyte invasion[263]. The three EBAs—EBA175, EBA140, and EBA181—are located in the micronemes, special organelles responsible for storage and release of erythrocyte invasion ligands[264]. Parasites variably express these invasion molecules such that any one is sufficient but not necessary for erythrocyte invasion. Switching between these invasion molecules may provide an important mechanism for immune evasion[265]. Therefore, it is possible that this deletion would have no deleterious effects on the parasite's ability to complete its lifecycle *in vitro*.

Interestingly, the deletion on chromosome 1 was only observed in three progeny, two of which were members of two family groups. However, the other members of the family did not possess the same deletion. This seemed to present two possibilities: 1) the model made an incorrect prediction; or 2) the model was correct, and three of the progeny spontaneously deleted this region. Looking at the raw coverage data indicated that the model was correct. The latter possibility raised two interesting questions about the biology of the parasite. The first question is: how should twin lines be defined? Family groups are composed of parasites that are

nearly isogenic, being identical for >99.95% of the 33,649 variants, and are assumed to be the product of a single meiosis. CNV was not considered when determining family group identities. At a minimum, CNVs add a caveat to the family group designation, without being sufficient to redefine the group structure. The second question is: how did this deletion arise? A study of CNV in *P. falciparum* laboratory lines showed that they were found commonly throughout the genome, and were most frequently observed in the subtelomeric regions[266]. However, it was unknown whether the observed CNVs reflected the absence of selection in *in-vitro* culture, or the natural proclivity of the parasite genome towards CNV generation[266]. The progeny lines in this thesis were in culture for less than a year at the time of gDNA extraction and WGS. This raised the possibility that some CNV generation could be a natural phenomenon that is not a result of lax selection in culture conditions.

The deletion/amplification event on chromosome 13 involved three Rh genes: *Rh6*, *Rh2b* and *Rh2a*. The Rh family consists of five genes and two pseudogenes that function in the invasion pathway[259,260]. Rh5, the most prominent member of the family, binds to the host cell ligand basigin, an interaction that is essential for RBC invasion[267]. In this cross, a deletion of *Rh6* and *Rh2b* in 803 and in progeny that inherited the 803 allele was observed, which was confirmed by PCR and qPCR. This pattern of deletion has been previously observed in Cambodian field isolates[150], making it likely that it did not arise *de novo* in 803. An amplification of the *Rh2b/Rh2a* locus was observed in three of the progeny. Confirmation of the amplification was undertaken with qPCR. Results indicated that the *Rh2a*/*Rh2b* locus was amplified in one of the progeny, and that there was increased expression of these two genes relative to GB4. These results are preliminary, as they represent a single experiment with only one of the three lines with evidence of amplification. Confirmation will be undertaken using a separate RNA extraction, and with RNA from the other two

progeny with the amplification. Identifying whether this amplification has any functional consequences will be achieved by performing *in-vitro* invasion assays, whereby parasite invasion efficiency is tested under a number of conditions[268].

The third example was a 16-kb deletion on chromosome 3 observed in GB4 and in progeny that inherited its allele. One of the genes in this region was *clag3.2*, which may have a role in nutrient transport[269]. This gene has been confirmed to be deleted in GB4 and in other parasites isolated from Ghana[261,262,270], further validating this CNV detection method.

The fourth example was a non-Mendelian deletion/amplification event involving the end of chromosome 11. The model indicated that nine of the progeny had a deletion in this region and six had an amplification, while the rest were normal. Eight of the nine genes in this region encoded exported proteins. One possible explanation is read-mapping error. *P. falciparum* genes that express infected-RBC surface proteins are known to have a high level of sequence diversity[143,145], and the recombination rate has been shown to be particularly high in the subtelomeric regions[145,193]. However, this hypothesis fails to explain the amplification. Another possible explanation is this was a natural phenomenon resulting from the proclivity of the telomeric regions towards CNV[266]. PCR and qPCR could be used to determine whether this pattern is real or artifact.

The fifth example was a curious incident involving a region containing *GCH-1*, which functions in the parasite's folate biosynthesis pathway[271]. *GCH-1* CNVs have been implicated in folate resistance[272]. Previous studies have shown a CNV at *GCH-1* in GB4[273]. Interestingly, the model in this thesis shows that GB4 is missing this amplification; however, it is present in many of the progeny lines. A possible explanation of this phenomenon is that GB4 lost the extra copy of the gene while in culture due to a fitness cost associated with the mutation. The GB4 that was used in

the cross may have still have had the amplification, which is why it is observed in the progeny. Another possible explanation is that this amplification arose *de novo* in the progeny. This, however, seems unlikely, as these lines have never been under folate drug pressure. A third possibility is that the model made the prediction in error, and GB4 does possess the amplification. The simplest test to confirm the CNV would be to observe the *GCH-1* expression levels of 803, GB4, and of progeny with and without the model-detected amplification by qPCR.

These five examples of CNV show that the *P. falciparum* genome has a layer of complexity that would benefit from further exploration. Advances in WGS along with an improving reference genome will make it possible to investigate CNV in natural parasite populations and to monitor the effects of drug interventions in a more thorough way.

5.5.2 Characterization of genome variation in the 803xGB4 genetic cross

In this genetic cross, 12,884 SNPs and 20,765 indels were identified for a total of 33,649 variants. These results were comparable to the number of variants identified in the previous three crosses (3D7xHB3: 15,388 SNPs and 26,699 indels; HB3xDd2: 14,885 SNPs and 21,576 indels; and 7G8xGB4: 14,392 SNPs and 20,079 indels) (Miles et al., in preparation). Consistent with the findings of the previous three crosses, indels outnumbered SNPs and were more likely to be found in intergenic regions, whereas SNPs were more likely found in exons (Miles et al., in preparation). The level to which indels outnumber SNPs in *P. falciparum* is aberrant when compared to other organisms for which data are available. For example, in humans, SNPs outnumber indels by closer to 10 to 1 (1000 Genomes Project, in review; http://www.1000genomes.org/home) .

Also consistent with previous findings, indels within genes tended to occur in multiples of three base pairs, reducing the likelihood of frameshifts. Those outside of genes tended to occur in multiples of two base pairs. Indels have the potential to disrupt protein function by shifting the reading frame, so it is reasonable that the majority would be found in intergenic regions, where they would be less likely to cause deleterious effects [274].

Multiple independent clones from each of the 10 family groups were sequenced. Members of each family group were > 99.9% identical to one another. This result suggests that members of each family group resulted from the expansion of a single meiotic product. Some family groups were quite large, with up to seven members, while ten lines were singletons.

There are several explanations as to why some family groups are larger than others. One possibility is that the larger groups have a proliferative advantage over the others. One method to investigate this question would be to count the number of merozoites that emerge from each ruptured schizont from each of the 17 clones[275]. If the more prolific family groups have a proliferative advantage, one would expect to see a greater number of merozoites per schizont. Disparate family group sizes could also be explained by other factors. Some family groups may have had a survival advantage in the chimpanzee or the mosquito. Exploring these hypotheses would involve more complex experiments that may no longer be possible, in part due to the moratorium on research involving chimpanzees. A third possibility is that members of the larger family groups were chosen for expansion by chance.

The level of variation detected in this genetic cross was in line with the results from the previous three crosses. A unique feature of this cross was the sequencing of members of each of ten family groups. Members of each family group were nearly identical for the 33,649 filtered variants; however, analysis of large scale deletions

and amplifications indicated an additional level of diversity that merits further exploration.

5.5.3 Recombination parameters in the 803xGB4 genetic cross

The recombination rate (RR) in *P. falciparum* is varied at both a population level and at a chromosomal level[145,193]. At a population level, a number of factors affect the RR, including intensity of transmission by mosquitoes and the diversity of the local parasite populations[276,277]. On a chromosomal level, RR is impacted by both the chromosomal region and by genetic marker locations. The subtelomeric regions of the chromosome are known to exhibit higher rates of recombination than other regions. Some chromosomes also have internal hypervariable regions that exhibit high RR[145]. Some genes are known to have high RR, including *var* genes, which encode the parasite cytoadherence ligand PfEMP1. Markers within these genes also impact RR calculations.

Using the Haldane mapping function, a recombination rate of 13.3 kb/cM was calculated from the SNP array data. This corresponds to an average 1% chance of observing a crossover in a 13.3-kb region in a single generation, similar to the rate observed in the other genetic crosses [9.6 kb/cM[194]; 17 kb/cM[141]; and 12.1-14.3 kb/cM (Miles et al., in preparation)]. This rate is approximately 50 times greater than the recombination rate observed in humans[278], further highlighting the extreme plasticity of the *P. falciparum* genome.

6 Conclusion

6.1 Summary

Frontline therapies against *P. falciparum* have repeatedly failed, illustrating the need both for a deeper understanding of the genetic mechanisms underlying resistance and for a robust pipeline of novel antimalarial drugs. To attempt to address these challenges, the fourth and possibly final genetic cross was conceived between an artemisinin-resistant parasite from Cambodia and an artemisinin-sensitive line from Africa. This genetic cross provided a phenomenal resource to study parasite drug response and to make detailed observations about genome structure with high-resolution WGS of the parental lines and progeny.

Chapters 3 describes the effort to optimize and perform a challenging qHTS method. This method was used to perform a high-throughput pharmaceutical compound screen to generate the most in-depth characterization of the parental lines and progeny of the 803xGB4 genetic cross to date. Analysis of the results of the compound screen revealed 52 compounds with sub-micromolar $IC_{50.}$ Results of this screen were also compared to data from the parental lines of the previous three crosses to make observations about differentially active compounds. A set of 384 compounds was selected for use in a confirmatory screen.

Chapter 4 describes efforts to screen additional progeny with the 384 selected compounds. This confirmatory screen was performed under more rigorous conditions to provide better estimates of drug response. Linkage analysis identified 25 loci significantly associated with parasite response to 61 compounds. Only 26% of these compounds were linked to *pfcrt*, *pfmdr1*, or *dhfr*, in contrast to the findings of Yuan et al[179,180]. These findings could be strengthened by screening additional progeny.

Chapter 5 describes the generation of WGS data for the parental lines and progeny of this genetic cross. Analysis of these data provided some new insights into variation in the parasite genome, including the identification of a CNV in a locus containing two genes, *Rh2a* and *Rh2b*, involved in RBC invasion. Further experiments will be performed to confirm the CNV in this region, and to determine if it has a functional consequence. CNVs were also identified that differed between otherwise isogenic lines; these CNVs may be a result of natural variation in the parasite genome, or may be an artifact of *in-vitro* culture. A set of 33,649 variants was identified, with indels being the greatest source of diversity. Inheritance patterns of the variants in the progeny lines were used to compare estimates of recombination rate with SNP array data.

There were several limitations to this study. One is that a limited number of progeny were available for analysis. qHTS optimization began at the same time as expansion of clonal progeny, and for much of the study, screening was performed in real time. As more progeny are made available for qHTS, the power of QTL analysis to detect loci associated with drug response will increase. With enough additional progeny, the analysis would be able to detect interacting loci[227], which has been impossible to date with *P. falciparum* crosses due to the limited number of progeny. Another limitation of this study is the qHTS method. Despite optimization, the risk for false positives still exists. Careful scrutiny must be applied to every QTL identified to ensure that the signal is real. This involves screening the parental lines and progeny in a more conventional format to confirm the pattern of differential response.

6.2 Future directions

Several of the compounds with sub-micromolar IC_{50} were previously identified antimalarials, or were antineoplastics, indicating a high level of toxicity and likely unsuitability for antimalarial development. Others, including the antiprotozoal

buparvaquone,[208] are lead compounds for antimalarial development. Pinacyanol chloride and toluidine blue, two compounds with undefined mechanism, would benefit from further characterization and medicinal chemistry efforts to increase efficacy.

Previous linkage studies with high-throughput compound screens identified three genes responsible for 96% of observed differential drug responses: *pfcrt*, *dhfr*, and *pfmdr1*. In this study, a locus containing *pfcrt* was only weakly associated with two compounds, one of which was identified in a previous study. This can be attributed to the fact that 803 and GB4 share a CQ-resistant PfCRT haplotype. Interestingly, though CQ was not linked to any loci in this screen, the parental lines displayed a 2-fold difference in CQ IC_{50}, indicating that there may be other genes involved in determining the phenotype. Six antifolates were linked to a locus containing *dhfr*. The association was strong for several compounds, with p-values <0.001.

Mutations in *pfmdr1* were previously identified as the third major driver of differential parasite response. In this study, 10 compounds were identified that were significantly associated to a locus on chromosome 5. Screening of additional progeny strengthened the association of six compounds to a 130-kb region containing *pfmdr1*. Since amplification and increased expression of *pfmdr1* have been linked to drug resistance in *P. falciparum*, copy number and expression levels were assessed in the two parental lines. Surprisingly, though 803 possessed a single copy of *pfmdr1* relative to GB4, it exhibited almost two-fold increased expression.

To confirm this finding, several experiments will be performed. To determine whether a basal increase in *pfmdr1* expression is heritable, three progeny with $pfmdr1_{803}$ and three progeny with $pfmdr1_{GB4}$ will be assessed for copy number and relative expression. These progeny will also be assessed for response to triclosan, clofoctol, and 2-benzyl-4-chlorophenol to confirm the findings of the high-throughput screen.

It is possible that polymorphisms in 803 within the *pfmdr1* promoter lead to higher levels of constitutive *pfmdr1* expression, giving rise to the difference in phenotype observed between 803 and GB4. A basal promoter of *pfmdr1* has been previously identified[237]. To examine this possibility, the sequence of the 5' UTR from 803 and GB4 would be compared to determine if there are any polymorphisms. If polymorphisms are identified, transient transfection would be performed to compare the 803 promoter and the GB4 promoter using a luciferase reporter. If no polymorphisms are identified, it is possible that a regulatory element exists outside of the 5' UTR of *pfmdr1*. It is possible that an inducer of *pfmdr1* expression is constitutively present in 803 and absent in GB4. Induction of *pfmdr1* expression in response to pressure by MFQ has been previously observed[237]. It is also possible that another gene within the locus identified by QTL analysis is responsible for the observed phenotype. Further examination of the genes within that region would then be undertaken to identify additional candidates.

If a promoter polymorphism were to be involved in increased *pfmdr1* expression, the role of increased expression in determining parasite response to the three compounds would be assessed. One method to test this is by inducing pseudo-heterozygosity by episomal expression of *pfmdr1*[138]. GB4 would be transiently transfected with a plasmid containing *pfmdr1* (including the promoter region) from either 803 or GB4. Response of the transiently transfected GB4 to triclosan, clofoctol, and 2-benzyl-4-chlorophenol would be assessed. To confirm the role of *pfmdr1*, a higher IC_{50} would be expected from the GB4 episomally expressing the 803 *pfmdr1* versus the GB4 *pfmdr1*.

Recent advances that have facilitated genome editing in *P. falciparum* include zinc-finger nucleases (ZFNs) and CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated nuclease) 9. Both techniques are

capable of permanently altering the sequence of a target gene, and would be useful in confirming the effect of any polymorphisms identified in the promoter of *pfmdr1* because they do not depend on episomal expression. ZFNs function by causing a double-strand break in a predefined locus and triggering homology-directed repair using a donor template[279]. Previous work has demonstrated that this technique can be used to achieve both complete allelic replacement and targeted site-specific editing in *P. falciparum*[280]. Both allelic exchange of the entire *pfmdr1* transcript (including the 5'-UTR) and targeted mutations (if identified) could be undertaken. One downside of ZFNs is that they are costly and time-consuming to produce.

The CRISPR-Cas9 system utilizes a guide RNA (gRNA) containing complementary nucleotides to a target sequence in the gene of interest. The gRNA recruits the Cas9 complex to the target in the genome, resulting in a double-strand break. A repair template containing the desired change is used in homology-directed repair, resulting in a modified gene[281]. The CRISPR-Cas9 system has been used in *P. falciparum* to silence genes[282], and to introduce point mutations in the gene responsible for ARM resistance[283]. This system has the advantage of not requiring engineered ZFNs, resulting in lower cost and effort.

Appendix I: Characterization of parental lines and recovery of recombinant progeny

Summary

This chapter outlines the selection of the parental lines for a genetic cross and the recovery of recombinant progeny. Phenotypically distinct parental lines were desired so as to increase the likelihood of identifying loci associated with drug response in the high-throughput screen. This chapter reviews the general procedure for performing a genetic cross of two *P. falciparum* lines, details the selection process for candidate parental isolates, presents data on the phenotypic differences between 803 and GB4, and describes the analysis and findings of a high-density SNP array used to identify unique recombinant progeny.

List of contributors

Juliana M. Sá[1], Sarah R. Kaslow[1], Michael A. Krause[1,2], Chanaki Amaratunga[1], Viviana Melendez-Muniz[1], Rebecca Salzman[1], Jiangbing Mu[1], Min Zhang[3], Kenneth Udenze[3], Suzanne Li[3], Alistair Miles[2], Menghang Xia[4], Ruili Huang[4], Sampada Shahane[4], Bingbing Deng[1], Richard T. Eastman[1], Ababacar Diouf[1], Kazutoyo Miura[1], Gregory S. Tullo[1], Justine S. Cummins-Oman[1], Soundarapandian Velmurugan[5], Adam Richman[5], Sumana Chakravarty[5], Yonas F. Abebe[5], Eric R. James[5], Peter F. Billingsley[5], Carole A. Long[1], Robert W. Gwadz[1], Patrick E. Duffy[6], B. Kim Lee Sim[5], Stephen L. Hoffman[5], Xinzhuan Su[1], Julian C. Rayner[7], Dominic P. Kwiatkowski[2,7], John H. Adams[3], Rick M. Fairhurst[1], Thomas E. Wellems[1]

[1]Laboratory of Malaria and Vector Research (LMVR), National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Rockville, MD
[2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
[3]University of South Florida, Tampa, FL
[4]National Center for Advancement of Translational Sciences, NIH, Rockville, MD
[5] Sanaria Inc., Rockville, MD
[6]Laboratory of Malaria Immunology and Vaccinology, NIAID, NIH, Rockville, MD
[7]Wellcome Trust Sanger Institute, Hinxton, UK

Introduction

The cross described in this thesis was designed to investigate the genetic mechanism behind ARM resistance. Delayed parasite clearance in response to treatment with ARM derivatives was first described along the Thailand-Cambodia border in the late 2000s, and was believed to herald the arrival of full-blown resistance[114,116,117]. Southeast Asia has long been a hotspot for emerging drug resistance in *P. falciparum*; indeed, CQ treatment failure first appeared in Thailand in 1957. By the 1970s, it had spread to sub-Saharan Africa and South America [284]. CQ-resistant *P. falciparum* led to the loss of many lives, and reversed years of progress towards controlling malaria[100].

When this genetic cross was conceived in 2012, there was no molecular marker for ARM resistance, nor was there an *in-vitro* assay to identify ARM-resistant parasites[114,117,247,285]. A genetic cross between an ARM-resistant and an ARM-sensitive parasite provided an ideal platform for identifying such a marker. It also provided an opportunity to explore genetic mechanisms of resistance to other compounds, and identify active highly active drug scaffolds.

Three previous genetic crosses have been completed: HB3 x 3D7[134], HB3 x Dd2[135], and 7G8 x GB4[136]. With phenotype and genotype information for the parents and progeny of a cross, one can perform QTL analysis to identify chromosomal regions significantly associated with the phenotype of interest[133,165]. Genes within those regions can be ranked according to the likelihood that they affect the phenotype. Genome editing strategies, such as CRISPR-Cas9 or zinc-finger nucleases, can then be employed to test whether polymorphisms in the selected gene are capable of modulating the phenotype of interest[280,282,283]. These crosses have been used to investigate the genetic basis of a wide range of traits, including invasion

tropisms, nutrient transport, and parasite resistance to a wide variety of compounds[180,169,269,286].

The 803xGB4 genetic cross was performed as follows (Figure S1-1)[134,136,165] (Sá et al. *in preparation*). Gametocytes from each parental line were cultured until mature (stage IV), and then mixed and fed to *Anopheles stephensi* mosquitoes. Parasites were allowed to mature in the mosquito, and then sporozoites were dissected out of the salivary gland and cryopreserved. After thawing, the sporozoites were injected intravenously into a chimpanzee. The chimpanzee was monitored by PCR and thin smear until parasite positive, after which a series of blood draws containing recombinant progeny were taken. Cryopreserved stocks were sent to University of South Florida and cloned as described in Section 2.1.6.

This chapter will focus on two aspects of the genetic cross. The first is the *in-vitro* assessment of the drug sensitivity of candidate parental lines. Phenotypically distinct parental lines maximized the likelihood of identifying loci associated with drug response in the progeny. The second is the recovery and expansion of recombinant clonal progeny. Over 200 progeny lines were recovered, and a total of 70 potential recombinants were expanded. Thirty-one unique recombinants have been identified to date and used in a qHTS to investigate genetic determinants of drug response, and for WGS analysis.

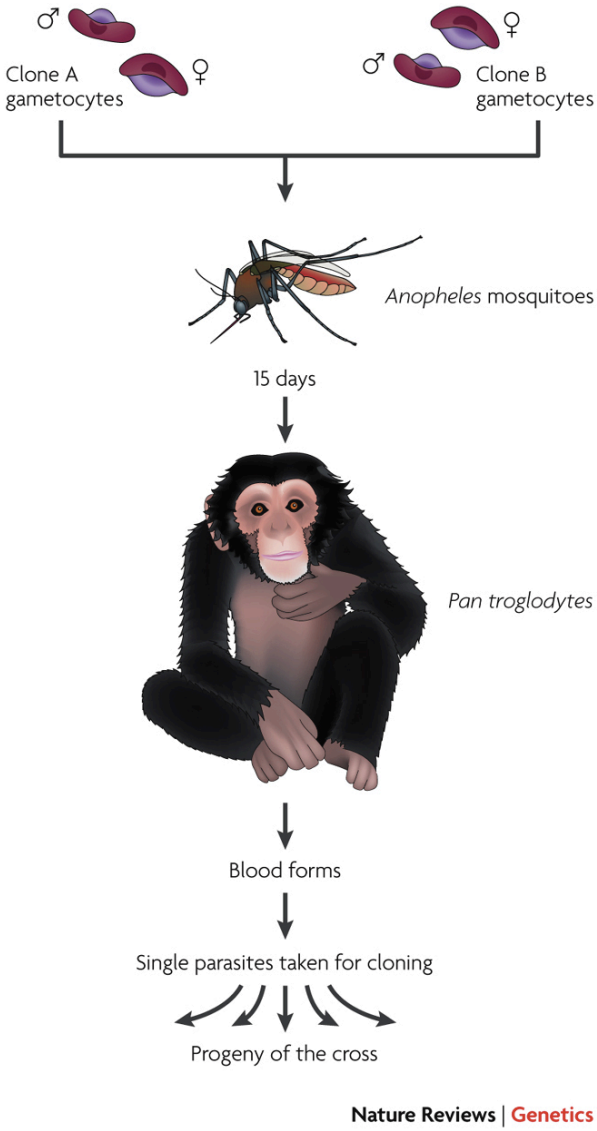


Figure S1-1 A genetic cross. Schematic outlining the steps towards completing a genetic cross. Gametocytes from each parental line are mixed and fed to an *Anopheles* mosquito. After 15 days, sporozoites are used to infect a chimpanzee. After blood stage parasitemia develops, pools of recombinant progeny are taken for cloning. Figure from Ref [165].

Objectives

i. To identify candidate parental isolates with differential responses to antimalarial compounds.

ii. To expand and identify unique recombinant progeny of the genetic cross for WGS and linkage analysis.

Results

Drug resistance profiles of candidate isolates

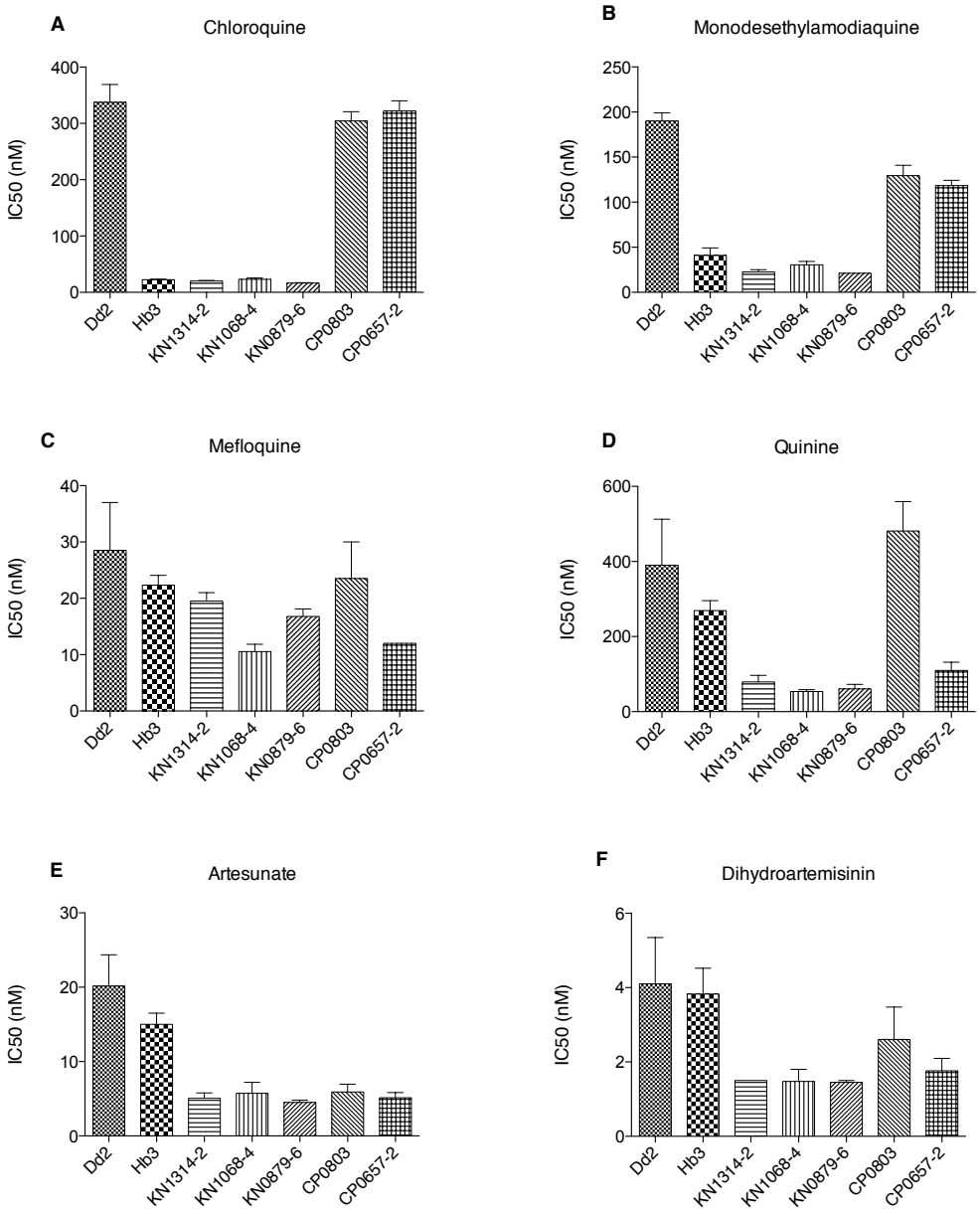
As part of the effort to characterize potential parental lines, several field isolates and laboratory lines were screened to identify candidates with promising differences in drug sensitivity. The laboratory lines used and culture conditions were described in sections 2.1.1 and 2.1.2. The preparation of compounds and the assay method were described in Sections 2.2 and 2.2.2. Twelve antimalarial compounds were tested in serial dilution (Table 2-2).

Mean IC_{50} values for each compound (three replicates) were plotted for two control laboratory lines (Dd2 and HB3), three Malian isolates, and two Cambodian isolates (Figure S1-2). Dd2 and HB3 have been used extensively in previous drug inhibition assays, and were therefore optimal controls for the assay[179,169,287,288]. The Malian and Cambodian isolates were obtained as described (Section 2.1.1).

There was a clear geographic distinction in antimalarial susceptibility in the isolates tested. Generally, the Cambodian isolates tended to be more resistant to the compounds tested than the Malian isolates. This was especially true for the quinoline drugs CQ (Figure S1-2A), MDAQ (Figure S1-2B), and QN (Figure S1-2D). One Cambodian line, CP0803, showed a ~2-fold higher IC_{50} for MFQ (Figure S1-2C) and DHA (Figure S1-2F). Interestingly, Cambodian isolate CP0657-2 was exquisitely sensitive to HLF (Figure S1-2G), and resistant to AVQ (Figure S1-2J). The isolates showed no meaningful differences for ART (Figure S1-2E), LMF (Figure S1-2H), and PPQ (Figure S1-2I). Unfortunately, the concentrations of SUL and PYR used in the assay failed to capture the IC_{50}s for all of the lines tested. As a result, the data for these two compounds were discarded.

Following failure of the Malian isolates to produce recombinant oocysts in the mosquito, GB4 was included as a candidate parental line. GB4 and CP0803

(hereafter called 803) were tested against eight antimalarials: CQ, MDAQ, QN, MFQ, PPQ, HLF, ART, and DHA (Figure S1-3A-H). GB4 followed the same pattern as the other African isolates, being generally more sensitive to the compounds tested than 803. The difference in response for CQ and MDAQ (Figure S1-3A, B) was especially interesting, given that GB4 and 803 possess the same CQ-resistant haplotype of *Plasmodium falciparum* chloroquine resistance transporter (PfCRT) (Table S1-6). The QN and MFQ (Figure S1-3C, D) responses were also promising. There was a large difference in sensitivity to HLF (Figure S1-3F) between 803 and GB4. PPQ (Figure S1-3E), ART (Figure S1-3G), and DHA (Figure S1-3H) showed no meaningful difference in IC_{50} between 803 and GB4.


A
Chloroquine
IC50 (nM)
400
300
200
100
0
Dd2
Hb3
KN1314-2
KN1068-4
KN0879-6
CP0803
CP0657-2
B
Monodesethylamodiaquine
IC50 (nM)
250
200
150
100
50
0
Dd2
Hb3
KN1314-2
KN1068-4
KN0879-6
CP0803
CP0657-2
C
Mefloquine
IC50 (nM)
40
30
20
10
0
Dd2
Hb3
KN1314-2
KN1068-4
KN0879-6
CP0803
CP0657-2
D
Quinine
IC50 (nM)
600
400
200
0
Dd2
Hb3
KN1314-2
KN1068-4
KN0879-6
CP0803
CP0657-2
E
Artesunate
IC50 (nM)
30
20
10
0
Dd2
Hb3
KN1314-2
KN1068-4
KN0879-6
CP0803
CP0657-2
F
Dihydroartemisinin
IC50 (nM)
6
4
2
0
Dd2
Hb3
KN1314-2
KN1068-4
KN0879-6
CP0803
CP0657-2

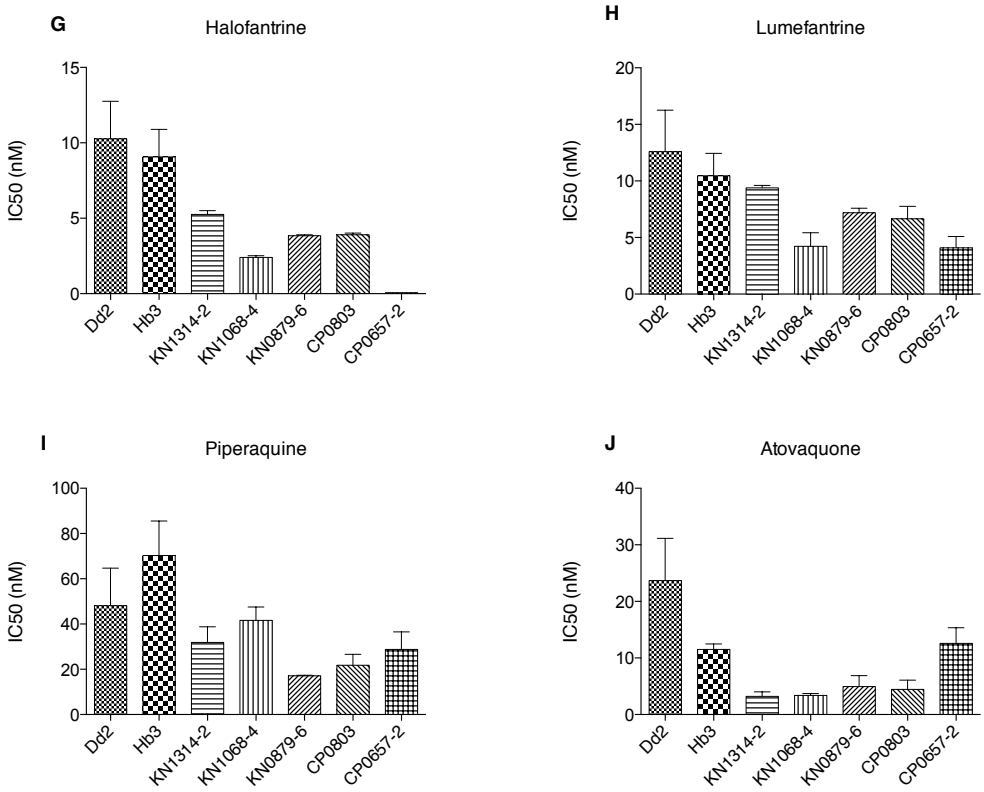


Figure S1-2 IC_{50} profiles of candidate parasite lines. Mean ± standard deviation IC_{50}s of two laboratory lines (Dd2 and HB3) and five field isolates (KN1314-2, KN0879-6, CP0803, and CP0657-2) of *P. falciparum*. The two parasites with the KN prefix were isolated from patients in Mali. In this chapter, the prefix CP was added to the numerical identifier of Cambodian isolates in order to distinguish them from Malian isolates, which were given the prefix KN. Each compound was tested in three biological replicates.

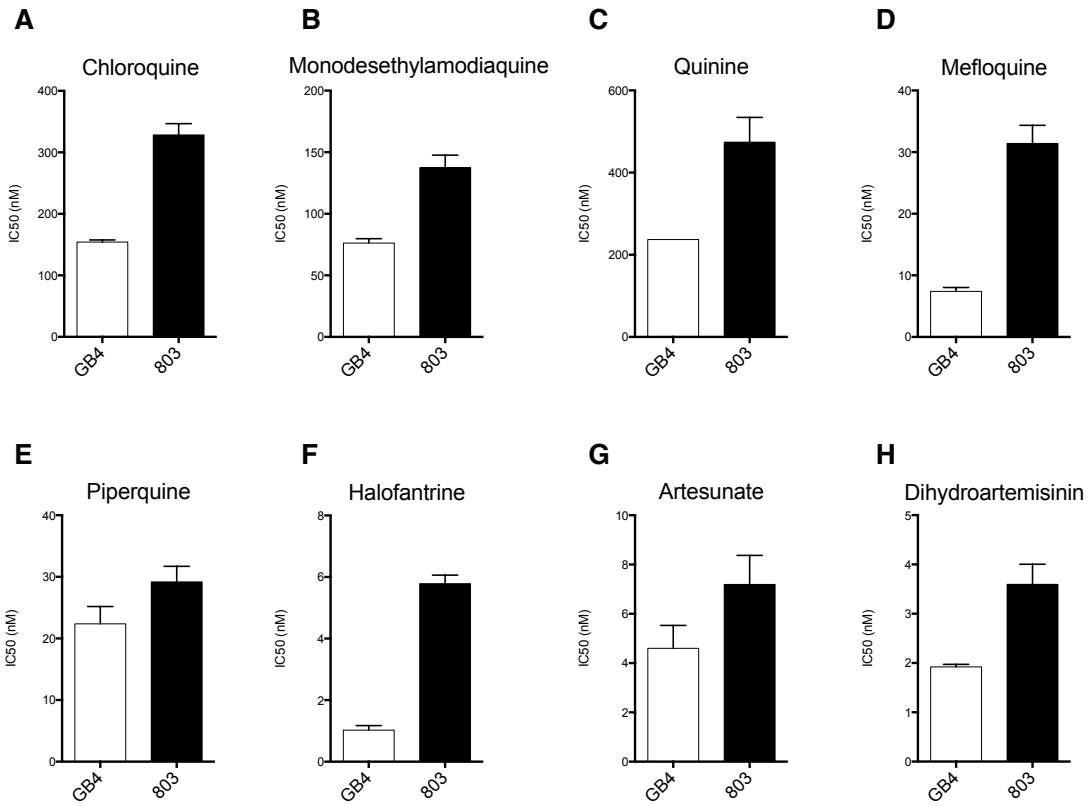


Figure S1-3 IC_{50} profiles of 803 and GB4. Mean ± standard deviation IC_{50}s of GB4 and 803 for eight antimalarial compounds (three biological replicates).

Inheritance and crossover patterns of recombinant progeny

Forty-six clonal progeny were originally selected to thaw and expand. Genomic DNA from these lines was hybridized to an Affymetrix SNP array. Analysis of the array data yielded 3,018 polymorphic markers, which segregated the 46 clones into 10 family groups and seven singletons. A family group was defined as two or more progeny that shared the same SNP inheritance pattern, meaning that there were 17 unique progeny that could be used for further experiments.

The 3,018 markers were spaced 7 kb ± 11 kb (mean ± standard deviation) apart along the 14 chromosomes. In the 17 progeny, 268 crossover events were identified (16 crossovers / progeny). Parental genotype inheritance was roughly even across most of the 14 chromosomes (Figure S1-4). Biased inheritance patterns were observed in five locations. In two locations, a 178-kb region on chromosome 7

containing 40 genes (Table S1-0-1) and a 119-kb region on chromosome 9 containing 20 genes (Table S1-0-2), only the GB4 allele was inherited in all 17 progeny. In three locations, a 14-kb region on chromosome 7 containing six genes (Table S1-0-3), a 125-kb region on chromosome 9 containing 35 genes (Table S1-0-4), and a 340-kb region on chromosome 12 (Table S1-0-5), 16 of the 17 progeny inherited the allele from GB4.

[Gene ID]	[Genomic Location(s)]	[Organism]	[Product Description]
PF3D7_0728100	07: 1,193,616 - 1,210,301 (-)	P. falciparum	conserved Plasmodium membrane protein,
PF3D7_0728200	07: 1,211,657 - 1,213,517 (+)	P. falciparum	actin-like protein, putative (ALP3)
PF3D7_0728300	07: 1,213,841 - 1,214,476 (-)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0728400.1	07: 1,215,174 - 1,215,763 (+)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0728400.2	07: 1,215,174 - 1,215,763 (+)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0728500	07: 1,216,088 - 1,217,201 (-)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0728600	07: 1,218,426 - 1,225,473 (+)	P. falciparum	zinc finger, C3HC4 type, putative
PF3D7_0728700	07: 1,226,029 - 1,228,714 (-)	P. falciparum	alpha/beta hydrolase, putative
PF3D7_0728800	07: 1,230,589 - 1,232,180 (+)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0728900	07: 1,232,272 - 1,234,672 (-)	P. falciparum	RNA binding protein, putative
PF3D7_0729000	07: 1,235,699 - 1,236,707 (+)	P. falciparum	signal recognition particle SRP9 (SRP9)
PF3D7_0729100	07: 1,238,446 - 1,246,188 (+)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0729200	07: 1,247,398 - 1,248,264 (+)	P. falciparum	1-cys peroxiredoxin (AOP)
PF3D7_0729300	07: 1,248,837 - 1,251,503 (-)	P. falciparum	60S ribosomal subunit export protein, putative
PF3D7_0729400	07: 1,252,847 - 1,254,097 (+)	P. falciparum	BRIX protein, putative
PF3D7_0729500	07: 1,255,371 - 1,257,653 (+)	P. falciparum	mRNA (N6-adenosine)-methyltransferase,
PF3D7_0729600	07: 1,258,787 - 1,259,730 (+)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0729700	07: 1,260,154 - 1,262,780 (-)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0729800	07: 1,264,445 - 1,264,738 (+)	P. falciparum	dynein light chain, putative
PF3D7_0729900	07: 1,265,910 - 1,281,347 (+)	P. falciparum	dynein heavy chain, putative
PF3D7_0730000	07: 1,281,785 - 1,282,322 (-)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0730100	07: 1,282,913 - 1,285,519 (+)	P. falciparum	tRNA pseudouridine synthase D, putative
PF3D7_0730200	07: 1,286,057 - 1,289,399 (-)	P. falciparum	adapter-related protein, putative
PF3D7_0730300	07: 1,297,459 - 1,301,454 (+)	P. falciparum	transcription factor with AP2 domain(s) (AP2-L)
PF3D7_0730400	07: 1,304,347 - 1,305,096 (+)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0730500	07: 1,305,894 - 1,314,215 (-)	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_0730600	07: 1,318,059 - 1,318,131 (+)	P. falciparum	tRNA Valine
PF3D7_0730700	07: 1,319,011 - 1,319,082 (-)	P. falciparum	tRNA Threonine
PF3D7_0730800.1	07: 1,321,306 - 1,322,367 (+)	P. falciparum	Plasmodium exported protein, unknown function
PF3D7_0730800.2	07: 1,321,306 - 1,322,367 (+)	P. falciparum	Plasmodium exported protein, unknown function
PF3D7_0730900	07: 1,325,375 - 1,331,813 (+)	P. falciparum	Plasmodium exported protein, unknown function
PF3D7_0731000	07: 1,333,175 - 1,334,938 (-)	P. falciparum	unspecified product
PF3D7_0731100	07: 1,338,959 - 1,341,790 (+)	P. falciparum	Plasmodium exported protein (PHISTc),
PF3D7_0731200	07: 1,342,842 - 1,343,996 (-)	P. falciparum	Plasmodium exported protein, unknown function
PF3D7_0731300	07: 1,350,032 - 1,351,149 (+)	P. falciparum	Plasmodium exported protein (PHISTb),
PF3D7_0731400	07: 1,353,758 - 1,355,482 (+)	P. falciparum	serine/threonine protein kinase, FIKK family,
PF3D7_0731500	07: 1,358,055 - 1,362,929 (+)	P. falciparum	erythrocyte binding antigen-175 (EBA175)
PF3D7_0731600	07: 1,368,161 - 1,370,596 (+)	P. falciparum	acyl-CoA synthetase (ACS5)
PF3D7_0731700	07: 1,372,063 - 1,372,760 (+)	P. falciparum	Plasmodium exported protein (hyp9), unknown
PF3D7_0731800	07: 1,378,618 - 1,380,769 (+)	P. falciparum	alpha/beta hydrolase, putative (GEXP08)

Table S1-0-1 Completely-biased inheritance on Chr 7. List of 40 genes within the 178-kb region on chromosome 7 with observed complete inheritance bias towards the GB4 allele. Data downloaded from PlasmoDB.org[289].

[Gene ID]	[Genomic Location(s)]	[Organism]	[Product Description]
PF3D7_093590	09: 1,420,483 - 1,422,807 (-)	P. falciparum	ring-exported protein 1 (REX1)
PF3D7_093600	09: 1,427,642 - 1,428,022 (-)	P. falciparum	ring-exported protein 2 (REX2)
PF3D7_093610	09: 1,430,607 - 1,431,146 (-)	P. falciparum	early transcribed membrane protein
PF3D7_093620	09: 1,433,332 - 1,434,108 (-)	P. falciparum	Plasmodium exported protein (hyp11), unknown
PF3D7_093630	09: 1,436,687 - 1,437,981 (-)	P. falciparum	ring-exported protein 3 (REX3)
PF3D7_093640	09: 1,440,783 - 1,441,660 (+)	P. falciparum	ring-exported protein 4 (REX4)
PF3D7_093650	09: 1,443,854 - 1,444,847 (-)	P. falciparum	Plasmodium exported protein, unknown function
PF3D7_093660	09: 1,446,929 - 1,447,947 (+)	P. falciparum	Plasmodium exported protein (PHISTb), unknown
PF3D7_093670	09: 1,454,422 - 1,455,720 (+)	P. falciparum	lysophospholipase, putative
PF3D7_093680	09: 1,458,392 - 1,459,862 (+)	P. falciparum	Plasmodium exported protein (PHISTc), unknown
PF3D7_093690	09: 1,462,400 - 1,463,659 (+)	P. falciparum	Plasmodium exported protein (PHISTb), unknown
PF3D7_093700	09: 1,465,904 - 1,467,055 (+)	P. falciparum	Plasmodium exported protein (PHISTb), unknown
PF3D7_093710	09: 1,468,546 - 1,469,417 (-)	P. falciparum	Plasmodium exported protein, unknown function
PF3D7_093720	09: 1,471,851 - 1,472,924 (+)	P. falciparum	lysophospholipase, putative
PF3D7_093730	09: 1,474,431 - 1,475,704 (+)	P. falciparum	rifin (RIF)
PF3D7_093740	09: 1,477,526 - 1,478,736 (+)	P. falciparum	rifin (RIF)
PF3D7_093750	09: 1,480,947 - 1,482,211 (-)	P. falciparum	rifin (RIF)
PF3D7_093760	09: 1,486,070 - 1,490,187 (+)	P. falciparum	erythrocyte membrane protein 1, PfEMP1 (VAR)
PF3D7_093770	09: 1,492,373 - 1,493,708 (+)	P. falciparum	rifin (RIF)
PF3D7_093780	09: 1,495,579 - 1,503,336 (-)	P. falciparum	erythrocyte membrane protein 1, PfEMP1 (VAR)

Table S1-0-2 Completely-biased inheritance on Chr 9. List of 20 genes within the 119-kb region on chromosome 9 with observed complete inheritance bias towards the GB4 allele. [289]

[Gene ID]	[Genomic Location(s)]	[Organism]	[Product Description]
PF3D7_0727200	07: 1,156,748 - 1,158,409 (-)	P. falciparum	cysteine desulfurase, putative (NFS)
PF3D7_0727300	07: 1,159,700 - 1,161,820 (-)	P. falciparum	DNA (cytosine-5)-methyltransferase, putative
PF3D7_0727400	07: 1,162,710 - 1,163,831 (+)	P. falciparum	proteasome subunit alpha type 5, putative
PF3D7_0727500	07: 1,164,455 - 1,166,578 (-)	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_0727600	07: 1,167,441 - 1,170,817 (+)	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_0727700	07: 1,171,286 - 1,174,156 (-)	P. falciparum	conserved Plasmodium protein, unknown function

Table S1-0-3 Partially-biased inheritance on Chr 7. List of six genes within a 14-kb region on chromosome 7 that displayed biased (16/17) inheritance of the GB4 allele. Data downloaded from PlasmoDB.org[289].

[Gene ID]	[Genomic Location(s)]	[Organism]	[Product Description]
PF3D7_093	Pf3D7_09_v3: 1,293,832 -	P. falciparum	DHHC-type zinc finger protein, putative
PF3D7_093	Pf3D7_09_v3: 1,297,285 -	P. falciparum	apicoplast ribosomal protein S6, putative (RPS6)
PF3D7_093	Pf3D7_09_v3: 1,298,865 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,303,390 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,309,467 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,311,440 -	P. falciparum	mRNA processing protein, putative
PF3D7_093	Pf3D7_09_v3: 1,314,552 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,317,827 -	P. falciparum	calcyclin binding protein, putative
PF3D7_093	Pf3D7_09_v3: 1,320,345 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,323,893 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,325,248 -	P. falciparum	gamma-tubulin complex component, putative
PF3D7_093	Pf3D7_09_v3: 1,331,349 -	P. falciparum	mitochondrial-processing peptidase subunit beta, putative
PF3D7_093	Pf3D7_09_v3: 1,334,115 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,335,740 -	P. falciparum	delta tubulin, putative
PF3D7_093	Pf3D7_09_v3: 1,338,175 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,340,065 -	P. falciparum	histidine--tRNA ligase, putative
PF3D7_093	Pf3D7_09_v3: 1,343,385 -	P. falciparum	DNA excision-repair helicase, putative
PF3D7_093	Pf3D7_09_v3: 1,346,858 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,348,022 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,349,790 -	P. falciparum	transcription factor with AP2 domain(s), putative (ApiAP2)
PF3D7_093	Pf3D7_09_v3: 1,350,836 -	P. falciparum	vacuolar ATP synthase subunit e, putative
PF3D7_093	Pf3D7_09_v3: 1,354,581 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,357,411 -	P. falciparum	UBX domain, putative
PF3D7_093	Pf3D7_09_v3: 1,361,935 -	P. falciparum	cAMP-dependent protein kinase catalytic subunit (PKAc)
PF3D7_093	Pf3D7_09_v3: 1,364,256 -	P. falciparum	conserved Plasmodium protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,366,569 -	P. falciparum	small nuclear ribonucleoprotein (snRNP), putative
PF3D7_093	Pf3D7_09_v3: 1,367,396 -	P. falciparum	conserved protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,368,594 -	P. falciparum	sec-1 family protein
PF3D7_093	Pf3D7_09_v3: 1,373,018 -	P. falciparum	phosphatidylinositol N-acetylglucosaminyltransferase
PF3D7_093	Pf3D7_09_v3: 1,377,953 -	P. falciparum	gametocyte development protein 1 (GDV1)
PF3D7_093	Pf3D7_09_v3: 1,400,800 -	P. falciparum	Plasmodium exported protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,405,192 -	P. falciparum	gametocytogenesis-implicated protein (GIG)
PF3D7_093	Pf3D7_09_v3: 1,409,199 -	P. falciparum	Plasmodium exported protein, unknown function
PF3D7_093	Pf3D7_09_v3: 1,413,840 -	P. falciparum	cytoadherence linked asexual protein 9 (CLAG9)

Table S1-0-4 Partially-biased inheritance on Chr 9. List of 35 genes on chromosome 9 that displayed biased (16/17) inheritance of the GB4 allele. Data downloaded from PlasmoDB.org[289].

[Gene ID]	[Genomic Location(s)]	[Organism]	[Product Description]
PF3D7_1216900	Pf3D7_12_v3: 670,447 -	P. falciparum	DNA-binding chaperone, putative
PF3D7_1217000	Pf3D7_12_v3: 674,349 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1217100.1	Pf3D7_12_v3: 675,456 -	P. falciparum	type IIB DNA topoisomerase, putative
PF3D7_1217100.2	Pf3D7_12_v3: 675,456 -	P. falciparum	type IIB DNA topoisomerase, putative
PF3D7_1217200	Pf3D7_12_v3: 678,602 -	P. falciparum	snoRNA-associated small subunit rRNA
PF3D7_1217300	Pf3D7_12_v3: 682,837 -	P. falciparum	GTP binding protein, putative
PF3D7_1217400	Pf3D7_12_v3: 686,307 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1217500	Pf3D7_12_v3: 693,067 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1217600	Pf3D7_12_v3: 695,559 -	P. falciparum	anaphase promoting complex subunit 10,
PF3D7_1217700	Pf3D7_12_v3: 696,895 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1217800	Pf3D7_12_v3: 698,771 -	P. falciparum	zinc finger protein, putative
PF3D7_1217900	Pf3D7_12_v3: 700,531 -	P. falciparum	conserved protein, unknown function
PF3D7_1218000	Pf3D7_12_v3: 703,895 -	P. falciparum	thrombospondin-related apical membrane
PF3D7_1218100	Pf3D7_12_v3: 706,227 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1218200	Pf3D7_12_v3: 708,936 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1218300	Pf3D7_12_v3: 717,955 -	P. falciparum	adaptor protein subunit, putative
PF3D7_1218400	Pf3D7_12_v3: 720,886 -	P. falciparum	triose or hexose phosphate/phosphate
PF3D7_1218500	Pf3D7_12_v3: 725,841 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1218600	Pf3D7_12_v3: 732,174 -	P. falciparum	arginine--tRNA ligase, putative
PF3D7_1218700	Pf3D7_12_v3: 736,335 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1218800	Pf3D7_12_v3: 738,357 -	P. falciparum	secreted ookinete protein, putative (PSOP17)
PF3D7_1218900	Pf3D7_12_v3: 740,994 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1219000	Pf3D7_12_v3: 746,002 -	P. falciparum	formin 2
PF3D7_1219100	Pf3D7_12_v3: 757,649 -	P. falciparum	clathrin heavy chain, putative
PF3D7_1219200	Pf3D7_12_v3: 764,448 -	P. falciparum	rifin, pseudogene (RIF)
PF3D7_1219300	Pf3D7_12_v3: 766,654 -	P. falciparum	erythrocyte membrane protein 1, PfEMP1 (VAR)
PF3D7_1219400	Pf3D7_12_v3: 776,510 -	P. falciparum	erythrocyte membrane protein 1 (PfEMP1),
PF3D7_1219500	Pf3D7_12_v3: 784,018 -	P. falciparum	erythrocyte membrane protein 1 (PfEMP1),
PF3D7_1219600	Pf3D7_12_v3: 785,571 -	P. falciparum	aminophospholipid-transporting P-ATPase
PF3D7_1219700	Pf3D7_12_v3: 792,877 -	P. falciparum	raf kinase inhibitor (RKIP)
PF3D7_1219800	Pf3D7_12_v3: 795,157 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1219900	Pf3D7_12_v3: 796,500 -	P. falciparum	D-ribulose-5-phosphate 3-epimerase, putative
PF3D7_1220000	Pf3D7_12_v3: 798,304 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1220100	Pf3D7_12_v3: 802,544 -	P. falciparum	pre-mRNA splicing factor, putative
PF3D7_1220200	Pf3D7_12_v3: 804,963 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1220300	Pf3D7_12_v3: 807,560 -	P. falciparum	Cell cycle associated protein, putative
PF3D7_1220400	Pf3D7_12_v3: 816,909 -	P. falciparum	debranching enzyme-associated ribonuclease,
PF3D7_1220500	Pf3D7_12_v3: 818,703 -	P. falciparum	ribosome biogenesis protein TSR3, putative
PF3D7_1220600	Pf3D7_12_v3: 820,586 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1220700	Pf3D7_12_v3: 822,958 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1220800	Pf3D7_12_v3: 825,697 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1220900	Pf3D7_12_v3: 831,252 -	P. falciparum	heterochromatin protein 1 (HP1)
PF3D7_1221000	Pf3D7_12_v3: 836,912 -	P. falciparum	histone-lysine N-methyltransferase, H3 lysine-4
PF3D7_1221100	Pf3D7_12_v3: 847,437 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1221200	Pf3D7_12_v3: 849,840 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1221300	Pf3D7_12_v3: 851,012 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1221400	Pf3D7_12_v3: 857,097 -	P. falciparum	inner membrane complex protein 1h, putative
PF3D7_1221500	Pf3D7_12_v3: 860,262 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1221600	Pf3D7_12_v3: 863,810 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1221700	Pf3D7_12_v3: 865,726 -	P. falciparum	FbpA domain protein, putative
PF3D7_1221800	Pf3D7_12_v3: 869,260 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1221900	Pf3D7_12_v3: 872,361 -	P. falciparum	conserved Plasmodium membrane protein,
PF3D7_1222000	Pf3D7_12_v3: 876,916 -	P. falciparum	conserved protein, unknown function
PF3D7_1222100	Pf3D7_12_v3: 879,630 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1222200	Pf3D7_12_v3: 881,970 -	P. falciparum	small nucleolar RNA snoR15
PF3D7_1222300	Pf3D7_12_v3: 883,567 -	P. falciparum	endoplasmin, putative (GRP94)
PF3D7_1222400	Pf3D7_12_v3: 890,581 -	P. falciparum	transcription factor with AP2 domain(s) (ApiAP2)
PF3D7_1222500	Pf3D7_12_v3: 899,198 -	P. falciparum	ATP-binding protein, putative
PF3D7_1222600	Pf3D7_12_v3: 907,203 -	P. falciparum	transcription factor with AP2 domain(s) (ApiAP2)
PF3D7_1222700	Pf3D7_12_v3: 918,722 -	P. falciparum	glideosome-associated protein 45 (GAP45)
PF3D7_1222800	Pf3D7_12_v3: 920,790 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1222900	Pf3D7_12_v3: 923,527 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1223000	Pf3D7_12_v3: 925,375 -	P. falciparum	conserved Plasmodium protein, unknown

PF3D7_1223100	Pf3D7_12_v3: 927,825 -	P. falciparum	cAMP-dependent protein kinase regulatory
PF3D7_1223200	Pf3D7_12_v3: 932,917 -	P. falciparum	nucleus export protein BRR6, putative
PF3D7_1223300	Pf3D7_12_v3: 934,971 -	P. falciparum	DNA gyrase subunit A (GyrA)
PF3D7_1223400	Pf3D7_12_v3: 941,125 -	P. falciparum	phospholipid-transporting ATPase, putative
PF3D7_1223500	Pf3D7_12_v3: 946,864 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1223600	Pf3D7_12_v3: 962,631 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1223700	Pf3D7_12_v3: 968,980 -	P. falciparum	integral membrane protein, putative
PF3D7_1223800	Pf3D7_12_v3: 971,268 -	P. falciparum	mitochondrial carrier protein, putative
PF3D7_1223900	Pf3D7_12_v3: 972,710 -	P. falciparum	50S ribosomal protein L24, putative
PF3D7_1224000	Pf3D7_12_v3: 974,372 -	P. falciparum	GTP cyclohydrolase I (GCH1)
PF3D7_1224100	Pf3D7_12_v3: 976,820 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1224200	Pf3D7_12_v3: 981,301 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1224300	Pf3D7_12_v3: 988,628 -	P. falciparum	polyadenylate-binding protein, putative (PABP)
PF3D7_1224400	Pf3D7_12_v3: 994,255 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1224500	Pf3D7_12_v3: 998,353 -	P. falciparum	histone chaperone ASF1, putative (ASF1)
PF3D7_1224600	Pf3D7_12_v3: 1,000,679 -	P. falciparum	cytochrome c heme lyase, putative
PF3D7_1224700	Pf3D7_12_v3: 1,003,201 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1224800	Pf3D7_12_v3: 1,009,289 -	P. falciparum	conserved Plasmodium protein, unknown
PF3D7_1224900	Pf3D7_12_v3: 1,010,832 -	P. falciparum	splicing factor 3b subunit, putative (SF3B14)
PF3D7_1225000	Pf3D7_12_v3: 1,012,310 -	P. falciparum	conserved Plasmodium protein, unknown

Table S1-0-5 Partially-biased inheritance Chr 12. List of 83 genes within the 340-kb region of chromosome 12 that displayed biased (16/17) inheritance of the GB4 allele. Data downloaded from PlasmoDB.org[289].

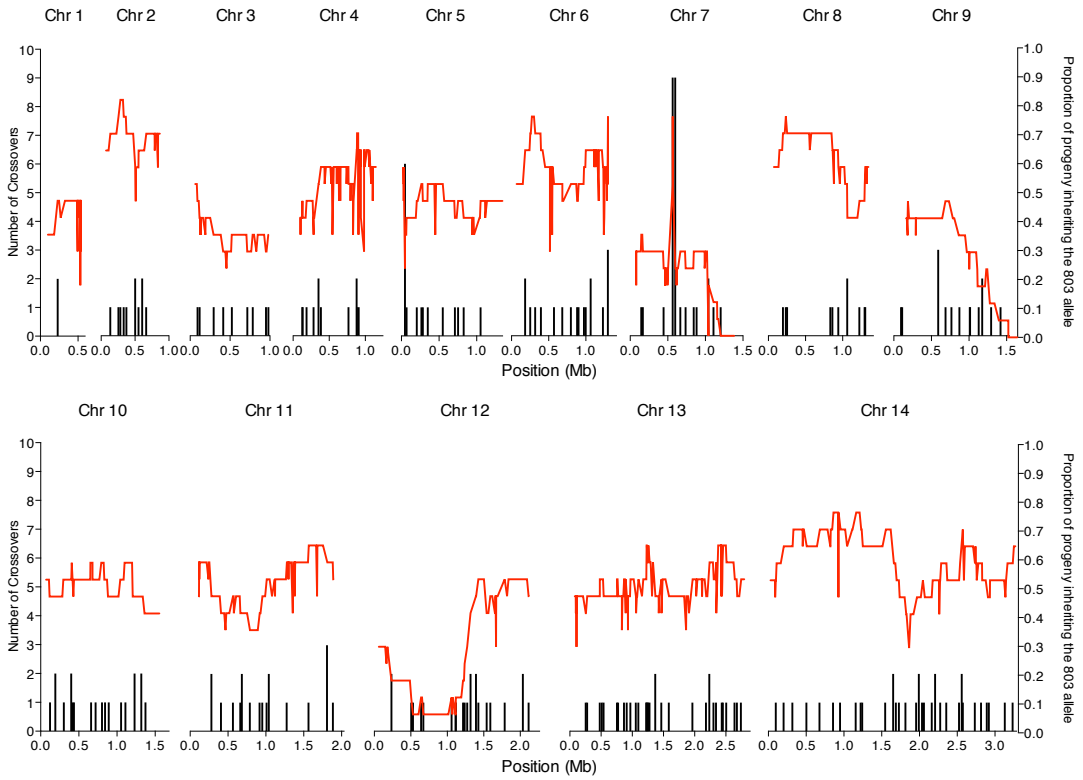


Figure S1-4 Crossover counts and inheritance patterns. The black bars denote the number of crossovers at each position across the 14 chromosomes. The red lines represent the proportion of progeny inheriting the 803 allele at each position.

Discussion

Characterization of phenotypically distinct parental isolates

A priority when designing the genetic cross was to identify phenotypically distinct parental lines. To that end, two Cambodian isolates with delayed clearance, three Malian isolates with rapid clearance, and GB4, a laboratory clone isolated from Ghana (see section 2.1.1), were tested against several antimalarials. The original goal was to cross a rapid-clearing isolate from Mali with a delayed-clearing isolate from Cambodia. GB4 was included as a candidate isolate because it had been used successfully in a previous cross, and was capable of infecting *Aotus* monkeys[136].

The 12 antimalarial compounds for screening candidate parental lines were chosen for one of two reasons: 1) the parasite's genetic determinant of resistance was well defined, thereby making the compound an ideal positive control; or 2) the genetic determinant was completely unknown or not fully understood.

CQ, MDAQ, SUL, PYR, and AVQ fell into the first category. Much is known about how these compounds act and how the parasite has evolved to combat them. The geographic spread of resistance mutations to each of the compounds has also been well documented, meaning predictions could be made about how the candidate parental isolates would be expected respond[169,221,290,291].

The 4-aminoquinolines CQ and MDAQ are thought to inhibit the polymerization of toxic byproducts of hemoglobin digestion, ultimately leading to the death of the parasite[292–295]. The K76T polymorphism in PfCRT has been linked to resistance to 4-aminioquinolines[137,138,296]. This mutation has nearly reached fixation in Southeast Asia[297,298], but not in our field site in Kenieroba, Mali where CQ-sensitive parasites are still prevalent, and where the Malian candidate lines were isolated (Krause MA, unpublished data, [299]).

SUL and PYR are antifolates, and act by inhibiting dihydropteroate synthetase (DHPS) and dihydrofolate reductase (DHFR), respectively, thereby interfering with purine metabolism[167,210]. Resistance mutations in these two genes are widespread[300], and have led to treatment failure and the withdrawal of SUL and PYR from many malaria-endemic countries. The wide difference in response between resistant and sensitive parasites made pyrimethamine an ideal control compound. The Cambodian isolates failed to respond to the concentrations of PYR used in this assay. A wider range of concentrations that captured the IC_{50} of 803 was included in the library of compounds described in Chapter 3.

AVQ is a napthoquinone that causes breakdown of the mitochondrial membrane potential[205]. Atovaquone resistance has been linked to mutations in *cytochrome b* and in *dhfr*[206]. Interestingly, mutations in *cytochrome b* can evolve *de novo* during treatment[301]. All parasite lines tested were sensitive to AVQ, so it was not a useful compound to include in further screening.

Three of the five compounds tested (CQ, MDAQ, PYR) proved to be useful as control compounds because their mechanisms of action were well documented[179,180,169], and because the resistance-mediating polymorphisms of the laboratory lines were well-characterized (Table S1-6). The K76T PfCRT mutation that is strongly linked with CQ resistance has been well defined[137,138]. The PfCRT haplotypes of the candidate parental lines tested were known, allowing CQ to be used as a control in the assay. As expected, parasites possessing the resistance mutation had much higher IC_{50} values for CQ than sensitive parasites. Similarly, MDAQ response has been linked to mutations in PfCRT and PfMDR1[169]. As with CQ, parasite response to MDAQ correlated with PfCRT haplotype. GB4 and 803, despite sharing the same PfCRT haplotype, showed a moderate difference in response to CQ and MDAQ, indicating that there might be other genes affecting the phenotype.

Because the concentration of PYR used in this assay was too narrow to capture the IC_{50} of 803, it provided limited utility in this assay. However, further testing as described in Chapter 3 revealed that GB4 was sensitive to PYR, while 803 was highly resistant.

QN, MFQ, LMF, HLF, PPQ, ART, and DHA fell into the second category, with incompletely understood mechanisms of action or resistance. QN, an extract of cinchona bark, has been in use as an antimalarial for almost 200 years[302,303]. QN resistance is not widespread, and the parasite's resistance mechanism is still not firmly defined, although polymorphisms in *pfcrt*, pfmdr1, and *pfnhe1* may be involved[304–307]. MFQ, LMF, and HLF are structurally related derivatives of QN[308]. The exact mechanism of action of these three compounds is unclear, but they likely act in a similar fashion. It has been noted that resistance to these compounds is inversely related to CQ sensitivity[234]. Parasite resistance to MFQ, LMF, and HLF has been hypothesized to be due to increased copy number of *pfmdr1*[234,309–312].

PPQ, a bisquinoline, has a poorly understood mechanism of action, and a genetic determinant for PPQ resistance has not been identified[313,314]. Parasite resistance to PPQ is incompletely understood, and there remains much debate in the literature about cross-resistance between PPQ and other antimalarials[313,315]. As such, it was unfortunate that the candidate parental isolates did not have differential sensitivity to PPQ.

ART and DHA, derivatives of the sesquiterpene lactone ARM, are poorly understood. These compounds are thought to act by causing oxidative damage to the parasite and killing it[316,317]. Mutations in the kelch propeller domain on chromosome 13 were recently correlated to delayed parasite clearance and *in-vitro* artemisinin resistance[120]. Traditional *in-vitro* assays are unable to distinguish between ARM-sensitive and ARM-resistant parasites[117], a finding corroborated here. There is less

than a two-fold difference in IC_{50} between 803 and GB4 for DHA and ART. This narrow window is not sufficient for classifying a parasite as resistant or sensitive. A novel assay developed by Witkowski and Amaratunga can distinguish between ARM-sensitive and ARM-resistant parasites. The ring-stage survival assay (RSA) measures parasite growth 66 hours after a 6-hour pulse with a high dose of DHA. Resistant parasites show ≥10% growth relative to DMSO-treated controls, whereas sensitive parasites show <1% growth[247]. The parents and progeny of this genetic cross showed differential sensitivity in the RSA, which was mapped to a region on chromosome 13 containing the kelch propeller gene. (Sá et al., in preparation)

All three of the Malian isolates tested had favorable chemical sensitivity profiles, meaning that they were several times more sensitive to the compounds tested than the Cambodian lines. In particular, the large QN IC_{50} difference between 803 and the Malian lines was very promising. However, further experiments with the Malian isolates indicated that they failed to produce gametocytes *in vitro*, failed to infect laboratory *Anopheline* mosquitos, or failed to recombine with Cambodian isolates in the mosquito midgut. (Sá et al., in preparation)

GB4 was tested against eight of the 10 antimalarial compounds described above. LMF and AVQ were excluded because results in the previous assay indicated that there was no pattern of differential sensitivity to these compounds. IC_{50} values for QN, MFQ, and HLF showed favorable differences between 803 and GB4. GB4 was also successfully used in a previous genetic cross, and was capable of infecting *Aotus* monkeys, allowing for the possibility of *in-vivo* studies in the progeny. Because of these reasons, GB4 was eventually selected as the other parental isolate.

	PfCRT										PfMDR1					DHFR					
Clone	**72**	**74**	**75**	**76**	**97**	**220**	**271**	**326**	**356**	**371**	**86**	**184**	**1034**	**1042**	**1246**	**51**	**59**	**108**	**164**	**298**	**306**
3D7	C	M	N	K		A	Q	N	I	R	N	Y	S	N	D	N	C	S	I	E	S
HB3	C	**I**	**E**	**T**		A	Q	N	I	R	N	**F**	S	**D**	D	N	C	**N**			
GB4	C	**I**	**E**	**T**	H	**S**	**E**	**S**	**T**	**I**	**Y**	**F**	S	N	D	N	C	S			
Dd2	C	**I**	**E**	**T**	H	**S**	**E**	**S**	**T**	**I**	**Y**	Y	S	N	D	**I**	**R**	**T**			
CP0803	C	**I**	**E**	**T**		**S**	**E**	**S**	**T**	**I**	N	Y		N	D	**I**	**R**	**N**	I	E	S
CP0657-2	C	**I**	**E**	**T**								**F**			D			**N**	**L**	E	S
KN1068-4	C	M	N	K																	
KN0879-6	C	M	N	K																	
KN1314-2	C	M	N	K																	

[180,167,169,254]; Krause et al. *unpublished*

Table S1-6 Haplotypes of selected antimalarial resistance genes. Amino acid identities at select codons of PfCRT, PfMDR1, and DHFR. 3D7 provides the reference allele. Underlined bold-face indicates change from reference.

Identification of unique recombinant progeny by SNP array analysis

Examination of the SNP inheritance patterns revealed 17 unique progeny out of the 46 clones tested. One reason for such a low number of unique recombinants may be due to unreliable microsatellite markers initially used to select clones for expansion. Each of the 46 clones was selected because it differed by at least one genetic marker from other clones. It was discovered that some of the markers produced erratic results, meaning that what were thought to be unique progeny were actually identical. This problem was remedied by testing a new set of markers that produced reliable and reproducible band-size differences between 803 and GB4, which allowed for generation of inheritance "barcodes" for each of the progeny (see section 2.4.1).

Further analysis of the SNP inheritance patterns yielded a number of examples of biased inheritance of the GB4 allele. Two regions, one on chromosome 7 and one on chromosome 9, showed complete inheritance of the GB4 allele for all 17 progeny (Figure S1-4). The region on chromosome 7 (Table S1-0-1) is located on the

chromosome end. Interestingly, this is the same region that showed biased inheritance in the 7G8xGB4 genetic cross, except that the 7G8 allele was exclusively inherited[194]. The authors implied that this biased inheritance could be explained by the presence of genes that gave progeny with the 7G8 allele a competitive advantage in the cross. There are 40 genes within this region that code for a variety of proteins[289], many of which have unknown functions. However, nine of the 40 genes encode exported proteins. This may indicate that one or more of these proteins is involved in completing the parasite's life cycle in the mosquito or the chimpanzee. It may also indicate that this region is resistant to recombination.

The region on chromosome 9 is also located on the chromosome end, and contains 20 genes (Table S1-0-2). Twelve of the 20 genes code for exported proteins or membrane proteins, two code for putative lysophospholipases, and the remainder are highly polymorphic *rifin* and *var* genes. The chromosome ends of *P. falciparum* chromosomes are known to have high recombination rates[193,194], so it is very interesting that there was no recombination in this region on chromosome 9. This might indicate that the conditions of the genetic cross exerted selective pressure on this region, or it may indicate that filtering of the array data eliminated subtelomeric SNPs.

Three regions exhibited inheritance of the GB4 allele in 16 of the 17 progeny. The regions on chromosome 7 and 9 are both adjacent to the regions with complete inheritance of the GB4 allele. The region on chromosome 12 is large, containing 83 genes. Many of these genes have unknown functions. Several encode proteins of note, including transporters, integral membrane proteins, and secreted proteins. Again, the inheritance pattern may point to the GB4 allele yielding a survival advantage to the progeny.

Appendix II

Table S2-0-1 Classified compounds selected for confirmatory screen.

Compound class	Compound	Log(IC50)	DCP	LOD
Adrenergic receptor binder	Oxprenolol hydrochloride	4.68	No	8.11
	Nifenalol	5.16	No	5.35
	Clenbuterol HCl	5.31	No	2.71
	Tulobuterol hydrochloride	4.97	Yes	2.69
	Efaroxan hydrochloride	5.10	No	2.61
	Nebivolol hydrochloride	5.30	No	2.58
	Mabuterol hydrochloride	5.13	Yes	2.52
	Ifenprodil tartrate	5.09	No	2.50
	Dexpropranolol	5.36	No	2.35
	Indoramin hydrochloride	4.67	No	2.29
	Xibenolol	4.93	Yes	2.22
	Bupranolol	4.90	No	2.20
	Atipamezole hydrochloride	5.57	No	2.05
	Detomidine hydrochloride	5.25	Yes	1.99
	Guanfacine hydrochloride	5.39	No	1.96
	Nylidrin	5.43	No	1.93
	Carvedilol	5.36	No	1.68
	Alprenolol hydrochloride	5.18	No	1.51
	Pronethalol hydrochloride	5.34	No	1.49
Antiarrhythmic	Mexiletine hydrochloride	4.47	No	3.42
	Quinine hydrochloride dihydrate	6.56	No	2.71
	Dronedarone hydrochloride	5.40	No	2.07
	Amiodarone hydrochloride	5.04	No	2.04
	Quinidine hydrochloride	6.99	No	1.91
	Aprindine hydrochloride	5.45	No	1.87
	Pirmenol	5.55	No	1.69
	Propafenone hydrochloride	6.28	No	1.64
	Clofilium tosylate	6.29	No	1.52
Antibacterial	Nitrofurazone	4.67	No	7.29
	Lincomycin hydrochloride	5.72	No	3.91
	Ciprofloxacin	5.21	No	3.43
	Nitrofurantoin	4.44	No	3.17
	Nifuroxazide	5.76	No	3.13
	Clofoctol	5.79	No	2.49
	Prulifloxacin	5.12	Yes	2.06
	Minocycline hydrochloride	5.06	Yes	1.90
	Ethacridine lactate hydrate	6.02	Yes	1.74
	Proflavin hemisulfate	5.23	No	1.65
	Acriflavinium hydrochloride	7.06	No	1.44
Antifolate	Trimethoprim	4.96	Yes	11.78
	Triamterene	5.15	Yes	11.38
	Pyrimethamine	5.57	Yes	9.06
	Diaveridine	4.90	No	8.08
	Methodichlorophen	4.98	Yes	5.70
	Ormetoprim	4.31	Yes	5.38
	Trimetrexate	7.26	Yes	3.46
	Methotrexate	7.21	No	2.39
Antifungal	Dipyrithione	5.95	No	1.88
	Phenylmercuric acetate	5.47	No	1.86
	Cycloheximide	6.88	No	1.53

Table S2-0-2 *continued*

Compound class	Compound	Log(IC50)	DCP	LOD
Antihistamine	Doxylamine succinate	5.13	Yes	5.29
	Cyclizine	5.13	No	4.23
	Orphenadrine hydrochloride	5.79	No	3.96
	Tesmilifene hydrochloride	4.85	No	3.09
	Thonzylamine hydrochloride	5.04	Yes	2.92
	Halopyramine	5.44	No	2.73
	Clemastine fumarate	5.39	No	2.59
	Meclizine hydrochloride	4.98	No	2.56
	Bromazine	5.06	No	2.45
	Desloratadine	5.17	No	2.43
	Promethazine hydrochloride	5.30	No	2.37
	(S)-(+)-Dimethindene maleate	5.41	No	2.37
	Azelastine hydrochloride	4.99	No	2.36
	Cyproheptadine hydrochloride	4.75	Yes	2.34
	Azatadine maleate	5.31	No	2.31
	Trimeprazine tartrate	4.99	No	2.25
	Triprolidine hydrochloride	5.22	No	2.17
	Phenyltoloxamine citrate salt	5.47	Yes	2.17
	Astemizole	5.61	No	2.13
	Diponium bromide	5.01	No	2.05
	Diphenhydramine hydrochloride	5.73	No	1.83
	Pheniramine maleate	5.50	No	1.82
	Flunarizine hydrochloride	5.12	Yes	1.76
	Ketotifen	5.56	Yes	1.74
	Mequitazine	5.38	No	1.67
	Carbinoxamine maleate	5.93	No	1.64
	Methapyrilene hydrochloride	5.43	No	1.57
	Tripelennamine citrate	5.54	No	1.55
Antimalarial	Mefloquine hydrochloride	7.30	No	2.69
	Primaquine phosphate	5.48	No	2.64
	Lumefantrine	7.23	Yes	2.27
	Atovaquone	8.39	No	2.21
	Proguanil	5.92	No	2.09
	Artemotil	7.37	No	2.03
	Chloroquine	6.54	No	2.02
	Artemisinin	7.68	No	1.94
	Halofantrine	6.64	No	1.70
Antimuscarinic	Procyclidine hydrochloride	5.09	No	6.25
	Timepidium Bromide	5.37	Yes	3.65
	Prifinium bromide	5.53	No	3.39
	Metixene hydrochloride	5.03	No	3.28
	Oxyphenonium bromide	5.29	No	2.86
	Oxapium iodide	5.54	No	2.70
	Methantheline bromide	4.28	No	2.39
	Trihexyphenidyl hydrochloride	4.90	No	2.32
	Aprofene	5.04	No	2.25
	Piperidolate hydrochloride	5.49	No	2.12
	Trospium chloride	5.82	Yes	1.78
	Tolterodine tartrate	5.61	No	1.77

Table S2-0-2 *continued*

Compound class	Compound	Log(IC50)	DCP	LOD
Antineoplastic	Nemorubicin	7.58	No	2.96
	Actinomycin D	8.41	No	2.56
	Flavopiridol	5.02	No	2.41
	Pirarubicin	5.71	No	2.37
	Epirubicin hydrochloride	6.31	Yes	2.31
	Bosutinib	6.23	No	2.30
	Mitoxantrone	5.71	No	2.29
	Zosuquidar trihydrochloride	5.91	No	2.16
	Bortezomib	6.67	No	2.06
	Vinblastine sulfate salt	6.89	Yes	2.01
	Carfilzomib	6.59	No	1.91
	Idarubicin hydrochloride	6.02	No	1.88
	Daunorubicin	5.98	No	1.67
Antiprotozoal	Clopidol	4.90	No	6.70
	Decoquinate	6.35	No	2.69
	Pentamidine isethionate	6.30	No	2.42
	Imidocarb dipropionate	6.44	No	2.12
	Maduramicin ammonium	6.32	No	2.07
	Quinacrine dihydrochloride	7.11	No	2.05
	Buparvaquone	7.30	No	1.72
Antipsychotic	Mesoridazine besylate	4.99	No	4.31
	Propionylpromazine HCl	5.31	No	3.01
	Thioridazine hydrochloride	5.27	No	2.96
	Prochlorperazine	4.68	No	2.73
	Chlorprothixene hydrochloride	4.67	No	2.67
	Perphenazine	5.27	No	2.56
	Amperozide hydrochloride	5.39	No	2.43
	Trifluperidol hydrochloride	5.45	No	2.32
	Thioproperazine dimesylate	4.91	No	2.19
	Ziprasidone hydrochloride	4.92	No	2.16
	Flupenthixol dihydrochloride	5.31	No	2.08
	Pimozide	5.27	No	2.01
	Fluphenazine hydrochloride	5.36	No	1.69
	Chlorpromazine	5.34	No	1.61
	Bromperidol	5.46	No	1.60
	Haloperidol	5.54	No	1.60
	Promazine hydrochloride	5.57	Yes	1.59
	Sertindole	5.32	No	1.49
	Penfluridol	5.36	No	1.43
Antispasmodic	Ethaverine hydrochloride	4.83	No	8.90
	Butinoline	4.97	No	5.92
	Camylofin	5.34	No	3.06
	Octylonium bromide	5.38	No	2.16

Table S2-0-2 *continued*

Compound class	Compound	Log(IC50)	DCP	LOD
Calcium channel blocker	Fendiline	4.71	Yes	3.03
	Lidoflazine	4.81	No	2.91
	Bepridil	5.34	No	2.74
	Azelnidipine	5.53	No	2.71
	Lomerizine	5.09	No	2.34
	BFMP	4.89	No	2.32
	Lercanidipine hydrochloride	5.25	No	2.29
	Methoxyverapamil HCl	5.04	No	2.19
	Isradipine	5.18	No	2.19
	(S)-(+)-Niguldipine hydrochloride	5.21	No	2.18
	Mibefradil HCl	5.76	Yes	2.08
	Nicardipine	5.73	No	2.06
	Vatanidipine	6.08	Yes	1.92
	Manidipine	5.35	No	1.88
	Verapamil	5.69	Yes	1.80
Catecholamine releaser	NEA	4.45	No	2.94
	DMEA	5.13	No	1.65
Corticosteroid	Amelometasone	4.56	Yes	2.53
	Clobetasol propionate	5.15	No	2.50
	Halcinonide	5.00	No	2.24
	Resocortol butyrate	5.06	No	2.23
	Clobetasone butyrate	5.98	Yes	2.15
	Hydrocortisone 17-butyrate	4.36	Yes	1.58
Cough suppressant	Butamirate citrate	4.76	Yes	2.31
	Caramiphen edisylate	5.48	No	2.23
	Cloperastine hydrochloride	5.62	No	2.14
	Clobutinol hydrochloride	5.44	No	1.84
	Carbetapentane citrate	5.43	No	1.83
Dopamine receptor binder	Domperidone	5.09	No	5.16
	Clebopride maleate	5.33	No	2.36
	Apomorphine	4.73	No	2.35
	Droperidol	5.63	Yes	1.88
Fluoroquinolone	Difloxacin hydrochloride	4.63	No	4.86
	Danofloxacin	5.49	No	3.52
	Gemifloxacin mesylate	5.54	No	1.82
	Ulifloxacin	5.36	No	1.30
Imidazole antifungal	Butoconazole nitrate	5.13	No	6.38
	Climbazole	4.93	No	5.93
	Econazole nitrate	5.70	No	3.45
	Posaconazole	4.94	No	2.83
	Isoconazole nitrate	4.75	No	2.16
	Tioconazole	5.05	No	2.16
	Sulconazole nitrate	5.84	No	2.08
	Clotrimazole	5.46	No	2.07
Immunosuppressive	Mycophenolic acid	5.23	No	2.51
	Azathioprine	4.80	No	2.29
	Cyclosporine A	6.36	No	2.09
Microtubule stabilizer	Vincristine sulfate	7.24	No	3.38
	Docetaxel	4.98	Yes	2.65
	Docetaxel	4.98	No	2.65
	Vinorelbine tartrate	6.32	No	2.10
	Vinblastine sulfate	6.10	No	1.92

Table S2-0-2 *continued*

Compound class	Compound	Log(IC50)	DCP	LOD
NMDA receptor antagonist	Memantine hydrochloride	5.06	No	3.29
	CNS-1102	5.51	No	2.30
	Remacemide hydrochloride	5.08	Yes	1.85
	Eliprodil	5.76	No	1.81
NSAID	Glafenine	4.58	No	3.27
	Bufexamac	5.99	Yes	2.87
	Proglumetacin	5.66	No	2.23
	Tinoridine	5.65	No	2.14
Opioid receptor binder	Meptazinol hydrochloride	5.23	No	4.35
	Levallorphan tartrate salt	4.79	No	2.97
	Buprenorphine	5.02	No	2.55
	Anileridine	3.96	No	2.49
	Tramadol hydrochloride	5.03	No	2.48
	Dextromoramide tartrate	4.91	Yes	2.37
	(+)-Metazocine fumarate (2:1)	4.61	No	2.20
	U-62066	4.77	No	2.19
	Sufentanil citrate	5.48	No	1.88
	(-)-Pentazocine succinate	5.38	Yes	1.85
Psychedelic	Ibogaine hydrochloride	4.83	No	4.62
	Lysergide	4.57	No	2.68
	N,N-Dimethyltryptamine fumarate	4.49	No	2.38
Serotonin receptor binder	AMI-193	5.18	No	4.39
	Ketanserin	4.90	No	3.06
	Cinanserin hydrochloride	4.64	No	2.55
	Dihydroergotamine	5.64	No	2.51
	SB-206553	5.37	No	2.39
	Eletriptan	3.79	No	2.37
	Altanserin hydrochloride	5.31	No	2.28
	MODPT	4.88	Yes	2.14
	Sarpogrelate	5.50	Yes	1.72
	Mosapride citrate	5.24	No	1.57
SSRI	Alaproclate	5.47	No	3.27
	Sertraline hydrochloride	5.46	No	2.89
	Fluvoxamine maleate	5.13	No	2.43
	Paroxetine	4.81	No	2.42

Table S2-0-2 *continued*

Compound class	Compound	Log(IC50)	DCP	LOD
Topical antiseptic	Ciclopirox	5.16	Yes	3.21
	Malachite Green chloride	6.24	Yes	3.15
	Cetrimonium bromide	5.48	No	2.63
	Propamidine	6.22	No	2.58
	Domiphen bromide	5.56	No	2.58
	Triclosan	5.54	Yes	2.57
	Triclosan	5.54	Yes	2.57
	Nitromersol	5.72	No	2.55
	Aminoquinuride dihydrochloride	5.57	No	2.53
	2-Benzyl-4-chlorophenol	4.08	No	2.51
	Dibucaine hydrochloride	4.84	No	2.43
	Thonzonium bromide	5.21	No	2.35
	Fipronil	5.26	No	2.33
	Fipronil	5.26	No	2.33
	Myralact	4.59	No	2.18
	Methylrosaniline chloride	6.39	No	2.08
	EHDMA	5.58	No	2.06
	Benzethonium chloride	5.76	No	2.00
	Methylbenzethonium chloride	5.71	No	1.87
	Alexidine dihydrochloride	6.52	No	1.50
	Chlorhexidine	6.37	No	1.43
Tricyclic antidepressant	Amitryptiline hydrochloride	5.17	No	5.64
	Cyclobenzaprine hydrochloride	5.06	No	4.94
	Dibenzepine hydrochloride	5.43	No	3.49
	Imipramine hydrochloride	5.37	No	2.92
	Trimipramine maleate	5.24	No	2.81
	Amoxapine	5.32	No	2.53
	Desipramine hydrochloride	5.49	No	2.43
	Opipramol dihydrochloride	5.13	Yes	1.81
	Doxepin hydrochloride	6.08	Yes	1.72
	Mosapramine	5.77	No	1.66
	Clocapramine	5.62	No	1.60
Tyrosine kinase inhibitor	Lapatinib	5.69	No	2.32
	Sunitinib malate	5.53	No	2.22
Vasodilator	Nafronyl oxalate	4.83	No	2.89
	Etafenone	5.49	No	2.18
	Suloctidil	5.64	No	1.80
Veterinary food additive	Roxarsone	5.21	No	3.18
	Nitrovin	5.47	No	2.46

Table S2-0-2 Classified compounds selected for confirmatory screen. List of compounds selected for the follow-up screen. The compounds have been grouped by class according to *Martindale: The Complete Drug Reference*[203]. Log(IC50) represents the average log(IC50) value for the 23 lines tested in the primary screen. DCP refers to differential chemical phenotype, defined as a five-fold difference in IC50 between 803 and GB4. Abbreviations: NMDA = N-methyl-D-aspartate; NSAID = non-steroidal anti-inflammatory drug; SSRI = selective serotonin reuptake inhibitor; MODPT = 5-methoxy-N,N-diisopropyltryptamine hydrochloride; EHDMA = Ethylhexadecyldimethylammonium bromide; NEA = (S)-(+)-N-Ethylamphetamine hydrochloride; DMEA = (+)-2,5-Dimethoxy-4-ethylamphetamine hydrochloride.

Table S2-0-3 Compounds without a general class.

Compound	Mechanism	Log(IC50)	DCP	LOD
Melengestrol acetate	Progestogen used in cattle feed.	4.93	No	5.49
Topotecan hydrochloride	Inhibitor of DNA topoisomerase I.	5.43	No	5.15
Clomifene citrate	Estrogenic and antiestrogenic.	5.24	No	4.80
Bromindione	Anticoagulant	5.67	Yes	3.88
Gimatecan	Topoisomerase I inhibitor.	5.42	No	3.47
Gemcitabine	DNA synthesis inhibitor.	5.62	No	3.37
Reserpine	Antihypertensive.	4.87	No	3.30
Melinamide	Lipid regulating drug.	4.97	No	3.26
Chlorphentermine hydrochloride	Anoretic	5.11	Yes	3.22
Rapacuronium bromide	Competitive neuromuscular blocker.	5.51	No	3.14
Mifepristone	Antiprogestogenic abortant.	4.83	Yes	3.07
Metochalcone	Choleretic.	5.50	No	3.00
Exifone	Cognitive supporter.	5.32	No	2.94
Gabexate mesilate	Proteolytic enzyme inhibitor.	4.86	No	2.89
Rimantadine	Antiviral agent.	5.52	No	2.85
Pimobendan	Phosphodiesterase type 3 inhibitor.	5.01	No	2.76
Donepezil HCl	ACE inhibitor.	4.94	No	2.76
Xylazine HCl	Veterinary sedative.	4.79	No	2.73
Bazedoxifene	Selective oestrogen receptor modulator.	5.81	No	2.73
Cantharidin	Vesicant.	5.19	No	2.66
Cinnamedrine	Sympathomimetic of ephedrine.	5.93	No	2.62
Orlistat	Lipase inhibitor.	5.13	No	2.57
1,2-Phenylenediamine	Used in hair color preparations.	4.68	No	2.53
Haloxon	Veterinary antihelminthic.	5.04	No	2.53
Cetylpyridinium bromide	Oral antiseptic.	5.52	No	2.47
Decamethonium bromide	Depolarizing neuromuscular blocker.	5.49	No	2.45
Bifemelane hydrochloride	Nootropic.	5.64	No	2.41
Zolantidine dimaleate	Central-acting antihistamine.	5.32	Yes	2.41
Cinacalcet hydrochloride	Calcimimetic agent.	5.06	No	2.38
Carmofur	Orally active derivative of fluorouracil.	5.20	No	2.36
Nefopam	Central-acting analgesic.	4.91	No	2.36
PP-242	Beta lactam antibacterial.	5.11	No	2.36
Trequinsin hydrochloride	Phosphodiesterase inhibitor.	4.88	No	2.35
Thonzonium bromide	Cationic surfactant.	5.21	No	2.35
Devazepide	Benzodiazepene.	5.05	No	2.35
Reboxetine mesylate	Noradrenaline reuptake inhibitor.	4.57	No	2.34
Proadifen hydrochloride	Cytochrome P450 inhibitor.	5.36	No	2.32
Homidium bromide	Trypanocide.	7.48	Yes	2.29
Levobupivacaine hydrochloride	Local anaesthetic.	5.05	Yes	2.29
Hexylcaine hydrochloride	Local anaesthetic; sodium channel blocker.	5.31	Yes	2.28
Epifibatide	Antiplatelet, anticoagulant.	5.17	No	2.23
Piperacetazine	Sedative; antipsychotic.	4.98	No	2.21
Mitomycin	Alkylating antineoplastic and antibiotic.	6.29	Yes	2.19
Propofol	Short-acting anaesthetic.	5.01	No	2.19
Loperamide hydrochloride	Antidiarrheal.	5.71	No	2.16
Oxethazaine	Amide anaesthetic.	5.43	No	2.15
Vorinostat	Histone deacetylase inhibitor.	6.31	No	2.08
Azoxystrobin	Agricultural fungicide.	7.43	No	2.06
Hycanthone	Schistosomicide.	5.85	No	2.05
Flupirtine maleate	Analgesic.	5.25	No	2.05
Perhexiline	Anti-anginal.	5.45	No	2.03
Bupropion hydrochloride	Weak SSRI	5.65	No	2.02
Dipyridamole	Adenosine reuptake inhibitor	5.34	No	2.01
Rotigotine hydrochloride	Dopamine agonist.	5.47	No	2.00
Benzethonium chloride	Cationic surfactant.	5.76	No	2.00
Tilorone	Antiviral; interferon inducer.	5.26	Yes	2.00

Table S2-0-4 *continued*

Compound	Mechanism	Log(IC50)	DCP	LOD
Almitrine dimethanesulfonate	Respiratory stimulant.	5.22	No	1.98
Spiramycin II	Macrolide antibacterial.	5.38	No	1.92
Diphenidol hydrochloride	Antiemetic.	5.94	No	1.90
Methylene blue	Antiseptic.	7.50	No	1.89
BIBX 1382 dihydrochloride	EGFR inhibitor (Manufacturer)	5.05	Yes	1.88
Berberine	Broad-spectrum antimicrobial	6.54	Yes	1.87
Emetine	Tissue amoebacide.	6.59	No	1.83
Nafamostat mesilate	Anticoagulant; prtoease inhibitor.	5.27	Yes	1.80
Diphenoxylate HCl	Antidiarrheal.	5.01	No	1.78
Sofalcone	Cytoprotective.	4.08	Yes	1.76
Hydroxy progesterone caproate	Potent progestogen.	5.69	No	1.74
Ambenonium dichloride	Cholinesterase inhibitor.	5.50	No	1.74
3,3 -Diethylthiacarbocyanine	Antihelminthic.	6.57	No	1.72
Benfluorex hydrochloride	Anoretic and hypolipidemic agent	4.27	Yes	1.68
Nizofenone fumarate	Nootropic and vasodilator.	5.31	No	1.62
Pridinol	Central-acting antispasmodic.	5.34	No	1.59
Cridanimod	Antiviral agent.	5.23	No	1.58
Hexetidine	Bactericidal and fungicidal antiseptic.	5.21	Yes	1.56
Clorprenaline HCl	Sympathomimetic.	5.31	No	1.52
Rifampicin	Antimycobacterial.	5.79	No	1.47
Tilmacoxib	COX-2 inhibitor.	4.06	Yes	x
Viloxazine hydrochloride	Noradrenalin reuptake inhibitor.	4.35	Yes	x

Table S2-0-4 Compounds without a general class. List of compounds selected for follow-up screening without a general class or with a more specific mechanism. Mechanisms were assigned according to *Martindale: The Complete Drug Reference*[203]. Log(IC50) represents the average log(IC50) value for the 23 lines tested in the primary screen. DCP refers to differential chemical phenotype, defined as a five-fold difference in IC50 between 803 and GB4. Abbreviations: ACE = acetylcholinesterase; SSRI = selective serotonin reuptake inhibitor; EGFR = epidermal growth factor receptor; COX = cyclooxygenase.

Compound	Log(IC50)	DCP	LOD
Nifuraldezone	4.31	No	4.62
N-(2-chloroethyl)pyrrolidine	3.86	No	4.48
Octoclothepin maleate salt	4.79	No	4.34
Ipidacrine hydrochloride hydrate	4.87	No	3.27
1,3-Diphenylguanidine	5.02	Yes	3.14
Indole-3-carbinol	4.84	No	2.98
Premethadone	4.99	No	2.98
Dorastine	4.97	No	2.64
1,5-Naphthalenediamine	4.83	No	2.64
(-)-alpha-Methadol hydrochloride	4.89	No	2.45
Pranolium chloride	5.32	No	2.38
PTEP	4.72	No	2.37
Etonitazene	4.86	No	2.37
Dichloralurea	4.67	No	2.36
Minopafant	5.21	No	2.31
Benzyldimethyldodecylammonium chloride	5.95	No	2.30
Difeterol	6.51	No	2.30
Thioglycolic acid anilide	4.59	No	2.28
Diphenylpyraline hydrochloride	5.36	No	2.27
HPEP	4.48	No	2.25
Anacolin	5.32	No	2.25
Rimcazole dihydrochloride	5.25	No	2.20
2-[2-(Aminomethyl)phenylthio]benzol	4.74	No	2.19
Cepharanthine	5.41	No	2.16
MitMAB	5.48	No	2.13
Roxindole hydrochloride	5.23	No	2.09
Ipenoxazone hydrochloride	5.58	No	2.07
Me(2-phenoxyethyl)-2-thenylNH4 iodide	5.50	No	2.07
Pinacyanol chloride	6.51	No	2.04
MG 624	5.91	No	2.01
Cloflucarban	5.62	No	1.95
Tocophersolan	5.02	No	1.92
EACPC	5.36	No	1.91
Monatepil	5.43	No	1.88
Benzyldimethyltetradecylammonium Cl	5.68	No	1.85
Fenaminosulf	5.47	No	1.80
1-Phenylcyclohexylamine hydrochloride	5.46	No	1.76
Rescimetol	5.98	No	1.71
Nitarsone	5.36	No	1.61
Toluidine blue	6.59	Yes	1.46
Ellagic acid	4.16	Yes	x

Table S2-0-5 Compounds without entry. List of compounds without entry in *Martindale: The Complete Drug Reference*[203]. Log(IC50) represents the average log(IC50) value for the 23 lines tested in the primary screen. DCP refers to differential chemical phenotype, defined as a five-fold difference in IC50 between 803 and GB4. Abbreviations: PTEP = N-phenyl-N-[1-[2-(2-thienyl)ethyl]-4-pip]propanamide hydrochloride; HPEP = N-[1-(2-hydroxy-2-phenyl-ethyl)-4-pip]-N-phe-propanNH3 hydrochloride; EACPC = Ethyl 4-(2-amino-4-chloroanilino)piperidine-1-carboxylate.

Mechanism	Compound	Log(IC50)	DCP	LOD
Antineoplastic	Actinomycin D	8.41	No	2.56
Antimalarial	Atovaquone	8.39	No	2.21
Antimalarial	Artemisinin	7.68	No	1.94
Antineoplastic	Nemorubicin	7.58	No	2.96
Antiseptic.	Methylene blue	7.50	No	1.89
Trypanocide	Homidium bromide	7.48	Yes	2.29
Agricultural fungicide.	Azoxystrobin	7.43	No	2.06
Antimalarial	Artemotil	7.37	No	2.03
Antiprotozoal	Buparvaquone	7.30	No	1.72
Antimalarial	Mefloquine hydrochloride	7.30	No	2.69
Antifolate	Trimetrexate	7.26	Yes	3.46
Microtubule stabilizer	Vincristine sulfate	7.24	No	3.38
Antimalarial	Lumefantrine	7.23	Yes	2.27
Antifolate	Methotrexate	7.21	No	2.39
Antiprotozoal	Quinacrine dihydrochloride	7.11	No	2.05
Antibacterial	Acriflavinium hydrochloride	7.06	No	1.44
Antiarrhythmic	Quinidine hydrochloride	6.99	No	1.91
Antineoplastic	Vinblastine sulfate salt	6.89	Yes	2.01
Antifungal	Cycloheximide	6.88	No	1.53
Antineoplastic	Bortezomib	6.67	No	2.06
Antimalarial	Halofantrine	6.64	No	1.70
Antineoplastic	Carfilzomib	6.59	No	1.91
-	Toluidine blue	6.59	Yes	1.46
Tissue amoebacide	Emetine	6.59	No	1.83
Antihelminthic.	3,3 -Diethylthiacarbocyanine	6.57	No	1.72
Antiarrhythmic	Quinine hydrochloride dihydrate	6.56	No	2.71
Antimalarial	Chloroquine	6.54	No	2.02
Antimicrobial	Berberine	6.54	Yes	1.87
Topical antiseptic	Alexidine dihydrochloride	6.52	No	1.50
-	Difeterol	6.51	No	2.30
-	Pinacyanol chloride	6.51	No	2.04
Antiprotozoal	Imidocarb dipropionate	6.44	No	2.12
Topical antiseptic	Methylrosaniline chloride	6.39	No	2.08
Topical antiseptic	Chlorhexidine	6.37	No	1.43
Immunosuppressive	Cyclosporine A	6.36	No	2.09
Antiprotozoal	Decoquinate	6.35	No	2.69
Antiprotozoal	Maduramicin ammonium	6.32	No	2.07
Microtubule stabilizer	Vinorelbine tartrate	6.32	No	2.10
Antineoplastic	Epirubicin hydrochloride	6.31	Yes	2.31
HDAC	Vorinostat	6.31	No	2.08
Antiprotozoal	Pentamidine isethionate	6.30	No	2.42
Antiarrhythmic	Clofilium tosylate	6.29	No	1.52
Antineoplastic	Mitomycin	6.29	Yes	2.19

Mechanism	Compound	Log(IC50)	DCP	LOD
Topical antiseptic	Malachite Green chloride	6.24	Yes	3.15
Antineoplastic	Bosutinib	6.23	No	2.30
Topical antiseptic	Propamidine	6.22	No	2.58
Microtubule stabilizer	Vinblastine sulfate	6.10	No	1.92
Ca channel blocker	Vatanidipine	6.08	Yes	1.92
Tricyclic antidepressant	Doxepin hydrochloride	6.08	Yes	1.72
Antibacterial	Ethacridine lactate hydrate	6.02	Yes	1.74
Antineoplastic	Idarubicin hydrochloride	6.02	No	1.88

Table S2-0-6 Compounds with sub-micromolar IC_{50}.

Bibliography

1. Greenwood, B. M., Bojang, K., Whitty, C. J. M. & Targett, G. A. T. Malaria. *Lancet* **365,** 1487–98 (2005).

2. WHO. *World Malaria Report (2014)*. (2014).

3. Murray, C. J. L. *et al.* Global malaria mortality between 1980 and 2010: a systematic analysis. *Lancet* **379,** 413–31 (2012).

4. Vanderberg, J. P. Imaging mosquito transmission of Plasmodium sporozoites into the mammalian host: Immunological implications. *Parasitol. Int.* **63,** 150–164 (2014).

5. Coppi, A. *et al.* Heparan sulfate proteoglycans provide a signal to Plasmodium sporozoites to stop migrating and productively invade host cells. *Cell Host Microbe* **2,** 316–27 (2007).

6. Meis, J. F. G. M. & Verhave, J. P. Exoerythrocytic Development of Malarial Parasites. *Adv. Parasitol.* **27,** 1–61 (1988).

7. Bannister, L. & Mitchell, G. The ins, outs and roundabouts of malaria. *Trends Parasitol.* **19,** 209–213 (2003).

8. Glushakova, S., Yin, D., Li, T. & Zimmerberg, J. Membrane transformation during malaria parasite release from human red blood cells. *Curr. Biol.* **15,** 1645–1650 (2005).

9. Lew, V. L. Malaria: Endless fascination with merozoite release. *Curr. Biol.* **15,** 760–761 (2005).

10. Pouvelle, B., Buffet, P. a, Lépolard, C., Scherf, a & Gysin, J. Cytoadhesion of Plasmodium falciparum ring-stage-infected erythrocytes. *Nat. Med.* **6,** 1264–1268 (2000).

11. Miller, L. H., Ackerman, H. C., Su, X. & Wellems, T. E. Malaria biology and disease pathogenesis: insights for new treatments. *Nat. Med.* **19,** 156–67 (2013).

12. Moxon, C. a., Grau, G. E. & Craig, A. G. Malaria: Modification of the red blood cell and consequences in the human host. *Br. J. Haematol.* **154,** 670–679 (2011).

13. Tilley, L., Dixon, M. W. A. & Kirk, K. The Plasmodium falciparum-infected red blood cell. *Int. J. Biochem. Cell Biol.* **43,** 839–42 (2011).

14. Mundwiler-Pachlatko, E. & Beck, H.-P. Maurer's clefts, the enigma of Plasmodium falciparum. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 19987–94 (2013).

15. Miller, L. H. Distribution of mature trophozoites and schizonts of Plasmodium falciparum in the organs of Aotus trivirgatus, the night monkey. *Am. J. Trop. Med. Hyg.* **18,** 860–865 (1969).

16. Luse, S. A. & Miller, L. H. Plasmodium falciparum malaria. Ultrastructure of parasitized erythrocytes in cardiac vessels. *Am. J. Trop. Med. Hyg.* **20,** 655–660 (1971).

17. David, P. H., Handunnetti, S. M., Leech, J. H., Gamage, P. & Mendis, K. N. Rosetting: A new cytoadherence property of malaria-infected erythrocytes. *Am. J. Trop. Med. Hyg.* **38,** 289–297 (1988).

18. Buffet, P. a *et al.* The pathogenesis of Plasmodium falciparum malaria in humans: Insights from splenic physiology. *Blood* **117,** 381–392 (2011).

19. Kwiatkowski, D. *et al.* TNF concentration in fatal cerebral, non-fatal cerebral, and uncomplicated Plasmodium falciparum malaria. *Lancet* **336,** 1201–1204 (1990).

20. Tripathi, A. K., Sha, W., Shulaev, V., Stins, M. F. & Sullivan, D. J. Plasmodium falciparum-infected erythrocytes induce NF-κB regulated inflammatory pathways in human cerebral endothelium. *Blood* **114,** 4243–4252 (2009).

21. Anstey, N. M. *et al.* Lung injury in vivax malaria: pathophysiological evidence for pulmonary vascular sequestration and posttreatment alveolar-capillary inflammation. *J. Infect. Dis.* **195,** 589–596 (2007).

22. Ioannidis, L. J., Nie, C. Q. & Hansen, D. S. The role of chemokines in severe malaria: more than meets the eye. *Parasitology* **141,** 602–13 (2014).

23. MacPherson, G. G., Warrell, M. J., White, N. J., Looareesuwan, S. & Warrell, D. a. Human cerebral malaria. A quantitative ultrastructural analysis of parasitized erythrocyte sequestration. *Am. J. Pathol.* **119,** 385–401 (1985).

24. Dorovini-Zis, K. *et al.* The neuropathology of fatal cerebral malaria in Malawian children. *Am. J. Pathol.* **178,** 2146–2158 (2011).

25. Ponsford, M. J. *et al.* Sequestration and microvascular congestion are associated with coma in human cerebral malaria. *J. Infect. Dis.* **205,** 663–671 (2012).

26. Kirkman, L. a. & Deitsch, K. W. Antigenic variation and the generation of diversity in malaria parasites. *Curr. Opin. Microbiol.* **15,** 456–462 (2012).

27. Su, X. Z. *et al.* The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. *Cell* **82,** 89–100 (1995).

28. Guizetti, J. & Scherf, A. Silence, activate, poise and switch! Mechanisms of antigenic variation in Plasmodium falciparum. *Cell. Microbiol.* **15,** 718–726 (2013).

29. Smith, J. D. *et al.* Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82,** 101–110 (1995).

30. Barry, A. E. *et al.* Population genomics of the immune evasion (var) genes of Plasmodium falciparum. *PLoS Pathog.* **3,** 1–9 (2007).

31. Rask, T. S., Hansen, D. a., Theander, T. G., Pedersen, A. G. & Lavstsen, T. Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes - divide and conquer. *PLoS Comput. Biol.* **6,** (2010).

32. Fried, M. & Duffy, P. E. Adherence of Plasmodium falciparum to chondroitin sulfate A in the human placenta. *Science* **272,** 1502–1504 (1996).

33. Salanti, A. *et al.* Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. *J. Exp. Med.* **200,** 1197–1203 (2004).

34. Whitty, C. J. M., Edmonds, S. & Mutabingwa, T. K. Malaria in pregnancy. *BJOG* **112,** 1189–1195 (2005).

35. Coronado, L. M., Nadovich, C. T. & Spadafora, C. Malarial hemozoin: From target to tool. *Biochim. Biophys. Acta - Gen. Subj.* **1840,** 2032–2041 (2014).

36. Francis, S. E., Sullivan, D. J. & Goldberg, D. E. Hemoglobin metabolism in the malaria parasite Plasmodium falciparum. *Annu. Rev. Microbiol.* **51,** 97–123 (1997).

37. Francia, M. E. & Striepen, B. Cell division in apicomplexan parasites. *Nat. Rev. Microbiol.* **12,** 125–36 (2014).

38. Bruce, M. C., Alano, P., Duthie, S. & Carter, R. Commitment of the malaria parasite Plasmodium falciparum to sexual and asexual development. *Parasitology* **100 Pt 2,** 191–200 (1990).

39. Sinden, R. E., Hartley, R. H. & Winger, L. The development of Plasmodium ookinetes in vitro: an ultrastructural study including a description of meiotic division. *Parasitology* **91 (Pt 2),** 227–244 (1985).

40. Angrisano, F., Tan, Y. H., Sturm, A., McFadden, G. I. & Baum, J. Malaria parasite colonisation of the mosquito midgut - Placing the Plasmodium ookinete centre stage. *Int. J. Parasitol.* **42,** 519–527 (2012).

41. Aly, A. S. I., Vaughan, A. M. & Kappe, S. H. I. Malaria parasite development in the mosquito and infection of the mammalian host. *Annu. Rev. Microbiol.* **63,** 195–221 (2009).

42. Sinka, M. E. *et al.* A global map of dominant malaria vectors. *Parasit. Vectors* **5,** 69 (2012).

43. Kwiatkowski, D. P. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet.* **77,** 171–92 (2005).

44. Bengtsson, B. O. & Tunlid, A. The 1948 International Congress of genetics in Sweden: People and politics. *Genetics* **185,** 709–715 (2010).

45. Nozais, J. P. Malaria in the Mediterranean world. The historical and present-day distribution. *Bull. Soc. Pathol. Exot. Filiales* **81,** 854–860 (1988).

46. Allison, A. C. Protection afforded by sickle-cell trait against subtertian malareal infection. *Br. Med. J.* **1,** 290–4 (1954).

47. Gething, P. W. *et al.* A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malar. J.* **10,** 378 (2011).

48. Piel, F. B. *et al.* Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *Lancet* **381,** 142–51 (2013).

49. Taylor, S. M., Parobek, C. M. & Fairhurst, R. M. Haemoglobinopathies and the clinical epidemiology of malaria: a systematic review and meta-analysis. *Lancet. Infect. Dis.* **12,** 457–68 (2012).

50. Agarwal, A. *et al.* Hemoglobin C associated with protection from severe malaria in the Dogon of Mali, a West African population with a low prevalence of hemoglobin S. *Blood* **96,** 2358 (2000).

51. Modiano, D. *et al.* Haemoglobin C protects against clinical Plasmodium falciparum malaria. *Nature* **414,** 305–308 (2001).

52. Chotivanich, K. *et al.* Hemoglobin E: a balanced polymorphism protective against high parasitemias and thus severe P falciparum malaria. *Blood* **100,** 1172–1176 (2002).

53. Naka, I. *et al.* Lack of association of the HbE variant with protection from cerebral malaria in Thailand. *Biochem. Genet.* **46,** 708–711 (2008).

54. Flint, J. *et al.* High frequencies of alpha-thalassaemia are the result of natural selection by malaria. *Nature* **321,** 744–750 (1986).

55. Williams, T. N. *et al.* High incidence of malaria in alpha-thalassaemic children. *Nature* **383,** 522–525 (1996).

56. Allen, S. J. *et al.* alpha+-Thalassemia protects children against disease caused by other infections as well as malaria. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 14736–14741 (1997).

57. Bienzle, U., Lucas, A., Ayeni, O. & Luzzatto, L. Glucose-6-Phosphate Dehydrogenase and Malaria. *Lancet* **299,** 107–110 (1972).

58. Ruwende, C. & Hill, a. Glucose-6-phosphate dehydrogenase deficiency and malaria. *J. Mol. Med. (Berl).* **76,** 581–588 (1998).

59. Guindo, A., Fairhurst, R. M., Doumbo, O. K., Wellems, T. E. & Diallo, D. a. X-linked G6PD deficiency protects hemizygous males but not heterozygous females against severe malaria. *PLoS Med.* **4,** 516–522 (2007).

60. Foo, L. C., Rekhraj, V., Chiang, G. L. & Mak, J. W. Ovalocytosis protects against severe malaria parasitemia in the Malayan aborigines. *Am. J. Trop. Med. Hyg.* **47,** 271–275 (1992).

61. Allen, S. J. *et al.* Prevention of cerebral malaria in children in Papua New Guinea by Southeast Asian ovalocytosis band 3. *Am. J. Trop. Med. Hyg.* **60,** 1056–1060 (1999).

62. Luzzatto, L., Nwachuku-Jarrett, E. S. & Reddy, S. Increased sickling of parasitised erythrocytes as mechanism of resistance against malaria in the sickle-cell trait. *Lancet* **295,** 319–321 (1970).

63. Roth, E. F. *et al.* Sickling rates of human AS red cells infected in vitro with Plasmodium falciparum malaria. *Science* **202,** 650–652 (1978).

64. Friedman, M. J. Erythrocytic mechanism of sickle cell resistance to malaria. *Proc. Natl. Acad. Sci. U. S. A.* **75,** 1994–1997 (1978).

65. Pasvol, G., Weatherall, D. J. & Wilson, R. J. Cellular mechanism for the protective effect of haemoglobin S against P. falciparum malaria. *Nature* **274,** 701–703 (1978).

66. Ayi, K., Turrini, F., Piga, A. & Arese, P. Enhanced phagocytosis of ring-parasitized mutant erythrocytes: A common mechanism that may explain protection against falciparum malaria in sickle trait and beta-thalassemia trait. *Blood* **104,** 3364–3371 (2004).

67. Arese, P., Turrini, F. & Schwarzer, E. Band 3/Complement-mediated Recognition and Removal of Normally Senescent and Pathological Human Erythrocytes. *Cell. Physiol. Biochem.* 133–146 (2005).

68. Destro-Bisol, G., Giardina, B., Sansonetti, B. & Spedini, G. Interaction between oxidized hemoglobin and the cell membrane: A common basis for several falciparum malaria-linked genetic traits. *Am. J. Phys. Anthropol.* **101,** 137–159 (1996).

69. Fairhurst, R. M. *et al.* Abnormal display of PfEMP-1 on erythrocytes carrying haemoglobin C may protect against malaria. *Nature* **435,** 1117–21 (2005).

70. Cholera, R. *et al.* Impaired cytoadherence of Plasmodium falciparum-infected erythrocytes containing sickle hemoglobin. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 991–6 (2008).

71. Amaratunga, C. *et al.* A role for fetal hemoglobin and maternal immune igg in infant resistance to plasmodium falciparum malaria. *PLoS One* **6,** 1–9 (2011).

72. Krause, M. a. *et al.* A-Thalassemia Impairs the Cytoadherence of Plasmodium Falciparum-Infected Erythrocytes. *PLoS One* **7,** 1–7 (2012).

73. Fairhurst, R. M., Bess, C. D. & Krause, M. A. Abnormal PfEMP1/knob display on Plasmodium falciparum-infected erythrocytes containing hemoglobin variants: fresh insights into malaria pathogenesis and protection. *Microbes Infect.* **14,** 851–62 (2012).

74. Williams, T. N. *et al.* An immune basis for malaria protection by the sickle cell trait. *PLoS Med.* **2,** 0441–0445 (2005).

75. Dobson, M. J. Bitter-sweet solutions for malaria: exploring natural remedies from the past. *Parassitologia* **40,** 69–81 (1998).

76. Neghina, R., Neghina, A. M., Marincu, I. & Iacobiciu, I. Malaria, a journey in time: in search of the lost myths and forgotten stories. *Am. J. Med. Sci.* **340,** 492–8 (2010).

77. Cox, F. E. History of the discovery of the malaria parasites and their vectors. *Parasit. Vectors* **3,** 5 (2010).

78. Bruce-Chwatt, L. J. Alphonse Laveran's discovery 100 years ago and today's global fight against malaria. *J. R. Soc. Med.* **74,** 531–536 (1981).

79. Meshnick, S. R. & Dobson, M. J. in *Antimalarial Chemotherapy: Mechanisms of Action, Resistance, and New Directions in Drug Discovery* (ed. Rosenthal, P. J.) 15–25 (Humana Press, 2001).

80. Sá, J. M., Chong, J. L. & Wellems, T. E. Malaria drug resistance: new observations and developments. *Essays Biochem.* **51,** 137–60 (2011).

81. Guerra, F. The introduction of cinchona in the treatment of malaria. *J. Trop. Med. Hyg.* **80,** 112–118; 135–140 (1977).

82. Greenwood, D. The quinine connection. *J. Antimicrob. Chemother.* **30,** 417–427 (1992).

83. Bruce-Chwatt, L. J. *et al. Chemotherapy of Malaria. Second edition. World Health Organization, Monograph Series. No. 27* (1986).

84. Trail, R. R. Richard Morton (1637-1698). *Med. Hist.* **14,** 166–174 (1970).

85. Roncalli Amici, R. The history of Italian parasitology. *Vet. Parasitol.* **98,** 3–30 (2001).

86. Smith, D. C. Quinine and fever: The development of the effective dosage. *J. Hist. Med. Allied Sci.* **31,** 343–367 (1976).

87. Hempelmann, E. & Krafts, K. Bad air, amulets and mosquitoes: 2,000 years of changing perspectives on malaria. *Malar. J.* **12,** 232 (2013).

88. Sullivan, M. S. & Swingland, I. R. Extinction risk: Predicting and redressing the threat. *Biodivers. Conserv.* **15,** 2009–2016 (2006).

89. Delepine, M. Joseph Pelletier and Joseph Caventou. *J. Chem. Educ.* **28,** 454–(1951).

90. Gramiccia, G. Ledger's cinchona seeds: a composite of field experience, chance, and intuition. *Parassitologia* **29,** 207–220 (1987).

91. Fosberg, F. R. Cinchona plantation in the new world. *Econ. Bot.* **1,** 330–333 (1947).

92. Coatney, G. R. Pitfalls in a discovery: the chronicle of chloroquine. *Am. J. Trop. Med. Hyg.* **12,** 121–128 (1963).

93. Jensen, M. & Mehlhorn, H. Seventy-five years of Resochin in the fight against malaria. *Parasitol. Res.* **105,** 609–627 (2009).

94. Smith, A. *Chloroquine, nonproprietary name for SN 7618*. *JAMA : the Journal of the American Medical Association* **130,** (1946).

95. Soper, F. L. Nation-wide malaria eradication projects in the Americas. V. General principles of the eradication programs in the Western Hemisphere. *J. Natl. Malar. Soc.* **10,** 183–194 (1951).

96. Giglioli, G., Rutten, F. J. & Ramjattan, S. Interruption of malaria transmission by chloroquinized salt in Guyana, with observations on a chloroquine-resistant strain of Plasmodium falciparum. *Bull. World Health Organ.* **36,** 283–301 (1967).

97. WHO. *Chemotherapy of Malaria: Report of a Technical Meeting*. (1961).

98. Eyles, D. E., Hoo, C. C., Warren, M. & Sandosham, A. A. Plasmodium falciparum resistant to chloroquine in Cambodia. *Am. J. Trop. Med. Hyg.* **12,** 840–843 (1963).

99. Payne, D. Spread of chloroquine resistance in Plasmodium falciparum. *Parasitol. Today* **3,** 241–6 (1987).

100. Trape, J. F. The public health impact of chloroquine resistance in Africa. *Am. J. Trop. Med. Hyg.* **64,** 12–7 (2001).

101. Wootton, J. C. *et al.* Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. *Nature* **418,** 320–3 (2002).

102. Chen, N. *et al.* pfcrt Allelic Types with Two Novel Amino Acid Mutations in Chloroquine-Resistant Plasmodium falciparum Isolates from the Philippines. *Antimicrob. Agents Chemother.* **47,** 3500–3505 (2003).

103. Wongsrichanalai, C., Pickard, A. L., Wernsdorfer, W. H. & Meshnick, S. R. Epidemiology of drug-resistant malaria. *Lancet. Infect. Dis.* **2,** 209–18 (2002).

104. Wellems, T. E. Transporter of a malaria catastrophe. *Nat. Med.* **10,** 1169–1171 (2004).

105. Hsu, E. The history of qing hao in the Chinese materia medica. *Trans. R. Soc. Trop. Med. Hyg.* **100,** 505–508 (2006).

106. Miller, L. H. & Su, X. Artemisinin: discovery from the Chinese herbal garden. *Cell* **146,** 855–8 (2011).

107. Tu, Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat. Med.* **17,** 1217–1220 (2011).

108. Trenholme, C. M. *et al. Mefloquine (WR 142,490) in the treatment of human malaria. Science (New York, N.Y.)* **190,** (1975).

109. Zhang, J. F. *et al. A detailed chronological record of project 523 and the discovery and development of qianghaosu (artemisinin)*. (Yangcheng Evening News Publishing Company, 2006).

110. Klayman, D. L. Qinghaosu (artemisinin): an antimalarial drug from China. *Science* **228,** 1049–1055 (1985).

111. World Health Organization. Guidelines for the Treatment of Malaria, Second Edition. (2010).

112. Alker, A. P. *et al.* PFMDR1 and in vivo resistance to artesunate-mefloquine in falciparum malaria on the Cambodian-Thai border. *Am. J. Trop. Med. Hyg.* **76,** 641–647 (2007).

113. Denis, M. B. *et al.* Surveillance of the efficacy of artesunate and mefloquine combination for the treatment of uncomplicated falciparum malaria in Cambodia. *Trop. Med. Int. Heal.* **11,** 1360–1366 (2006).

114. Dondorp, A. M. *et al.* Artemisinin resistance in Plasmodium falciparum malaria. *N. Engl. J. Med.* **361,** 455–67 (2009).

115. Noedl, H. *et al.* Evidence of artemisinin-resistant malaria in western Cambodia. *N. Engl. J. Med.* **359,** 2619–2620 (2008).

116. Phyo, A. P. *et al.* Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. *Lancet* **379,** 1960–6 (2012).

117. Amaratunga, C. *et al.* Artemisinin-resistant Plasmodium falciparum in Pursat province, western Cambodia: a parasite clearance rate study. *Lancet Infect. Dis.* **12,** 851–8 (2012).

118. Kyaw, M. P. *et al.* Reduced susceptibility of Plasmodium falciparum to artesunate in southern Myanmar. *PLoS One* **8,** e57689 (2013).

119. Ashley, E. A. *et al.* Spread of Artemisinin Resistance in Plasmodium falciparum Malaria. *N. Engl. J. Med.* **371,** 411–423 (2014).

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

121. Plowe, C. V. Malaria: Resistance nailed. *Nature* **505,** 30–1 (2014).

122. Hu, G. *et al.* Transcriptional profiling of growth perturbations of the human malaria parasite Plasmodium falciparum. *Nat. Biotechnol.* **28,** 91–8 (2010).

123. Meshnick, S. R., Taylor, T. E. & Kamchonwongpaisan, S. Artemisinin and the antimalarial endoperoxides: from herbal remedy to targeted chemotherapy. *Microbiol. Rev.* **60,** 301–315 (1996).

124. Alassane Mbengue *et al.* A molecular mechanism of artemisinin resistance in Plasmodium falciparum malaria. *Nature* (2015). doi:10.1038/nature14412

125. White, N. J. *et al.* Seminar Malaria. **383,** (2014).

126. Dondorp, A. M. *et al.* The relationship between age and the manifestations of and mortality associated with severe malaria. *Clin. Infect. Dis.* **47,** 151–157 (2008).

127. Snow, R. W. *et al.* Severe childhood malaria in two areas of markedly different falciparum transmission in East Africa. *Acta Trop.* **57,** 289–300 (1994).

128. World Health Organization. Severe falciparum malaria. *Trans. R. Soc. Trop. Med. Hyg.* **94,** 1–90 (2000).

129. Birbeck, G. L. *et al.* Blantyre Malaria Project Epilepsy Study (BMPES) of neurological outcomes in retinopathy-positive paediatric cerebral malaria survivors: A prospective cohort study. *Lancet Neurol.* **9,** 1173–1181 (2010).

130. Yen, L. M. *et al.* Role of quinine in the high mortality of intramuscular injection tetanus. *Lancet* **344,** 786–787 (1994).

131. Dondorp, A. *et al.* Artesunate versus quinine for treatment of severe falciparum malaria: A randomised trial. *Lancet* **366,** 717–725 (2005).

132. Dondorp, A. M. *et al.* Artesunate versus quinine in the treatment of severe falciparum malaria in African children (AQUAMAT): An open-label, randomised trial. *Lancet* **376,** 1647–1657 (2010).

133. Ranford-Cartwright, L. C. & Mwangi, J. M. Analysis of malaria parasite phenotypes using experimental genetic crosses of Plasmodium falciparum. *Int. J. Parasitol.* **42,** 529–34 (2012).

134. Walliker, D. *et al.* Genetic analysis of the human malaria parasite Plasmodium falciparum. *Science* **236,** 1661–6 (1987).

135. Wellems, T. E. *et al.* Chloroquine resistance not linked to mdr-like genes in a Plasmodium falciparum cross. *Nature* **345,** 253–5 (1990).

136. Hayton, K. *et al.* Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of Plasmodium falciparum invasion. *Cell Host Microbe* **4,** 40–51 (2008).

137. Wellems, T. E., Walker-Jonah, a & Panton, L. J. Genetic mapping of the chloroquine-resistance locus on Plasmodium falciparum chromosome 7. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 3382–6 (1991).

138. Fidock, D. A. *et al.* Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell* **6,** 861–71 (2000).

139. Mansfield, E. S. *et al.* Sensitivity, reproducibility, and accuracy in short tandem repeat genotyping using capillary array electrophoresis. *Genome Res.* **6,** 893–903 (1996).

140. Su, X. z & Wellems, T. E. Toward a high-resolution Plasmodium falciparum linkage map: polymorphic markers from hundreds of simple sequence repeats. *Genomics* **33,** 430–44 (1996).

141. Su, X. *et al.* A Genetic Map and Recombination Parameters of the Human Malaria Parasite Plasmodium falciparum. *Science (80-.).* **286,** 1351–1353 (1999).

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

143. Volkman, S. K. *et al.* A genome-wide map of diversity in Plasmodium falciparum. *Nat. Genet.* **39,** 113–9 (2007).

144. Jeffares, D. C. *et al.* Genome variation and evolution of the malaria parasite Plasmodium falciparum. *Nat. Genet.* **39,** 120–125 (2007).

145. Mu, J. *et al.* Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome. *Nat. Genet.* **39,** 126–130 (2007).

146. Kwiatkowski, D. P. Malaria genomics: tracking a diverse and evolving parasite population. *Int. Health* **7,** 82–84 (2015).

147. Mu, J. *et al.* Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat. Genet.* **42,** 268–271 (2010).

148. Van Tyne, D. *et al.* Identification and functional validation of the novel antimalarial resistance locus PF10_0355 in Plasmodium falciparum. *PLoS Genet.* **7,** e1001383 (2011).

149. Manske, M. *et al.* Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* (2012). doi:10.1038/nature11174

150. Miotto, O. *et al.* Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nat. Genet.* **47,** 226–234 (2015).

151. Wendler, J. P. *et al.* A genome wide association study of Plasmodium falciparum susceptibility to 22 antimalarial drugs in Kenya. *PLoS One* **9,** (2014).

152. Mobegi, V. A. *et al.* Genome-wide analysis of selection on the malaria parasite plasmodium falciparum in West African populations of differing infection endemicity. *Mol. Biol. Evol.* **31,** 1490–1499 (2014).

153. Desjardins, R. E., Canfield, C. J., Haynes, J. D. & Chulay, J. D. Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. *Antimicrob. Agents Chemother.* **16,** 710–718 (1979).

154. Chulay, J. D., Atkins, W. M. & Sixsmith, D. G. Synergistic antimalarial activity of pyrimethamine and sulfadoxine against Plasmodium falciparum in vitro. *Am. J. Trop. Med. Hyg.* **33,** 325–330 (1984).

155. Gershon, P. & Howells, R. Combination of the antibiotics erythromycin and tetracycline with 3 standard antimalarials against Plasmodium falciparum in vitro. *Ann. Trop. Med. Parasitol.* **78,** 1–11 (1984).

156. Chugh, M. *et al.* Identification and deconvolution of cross-resistance signals from antimalarial compounds using multidrug-resistant Plasmodium falciparum strains. *Antimicrob. Agents Chemother.* **59,** 1110–8 (2015).

157. Ke, H. *et al.* The heme biosynthesis pathway is essential for Plasmodium falciparum development in mosquito stage but not in blood stages. *J. Biol. Chem.* **289,** 34827–37 (2014).

158. Smilkstein, M., Sriwilaijaroen, N., Kelly, J. X., Wilairat, P. & Riscoe, M. Simple and Inexpensive Fluorescence-Based Technique for High-Throughput Antimalarial Drug Screening. *Antimicrob. Agents Chemother.* **48,** 1803–1806 (2004).

159. Vossen, M. G., Pferschy, S., Chiba, P. & Noedl, H. The SYBR green I malaria drug sensitivity assay: Performance in low parasitemia samples. *Am. J. Trop. Med. Hyg.* **82,** 398–401 (2010).

160. Biagini, G. a. *et al.* Generation of quinolone antimalarials targeting the Plasmodium falciparum mitochondrial respiratory chain for the treatment and prophylaxis of malaria. *Proc. Natl. Acad. Sci.* **109,** 8298–8303 (2012).

161. Griffin, C. E. *et al.* Mutation in the Plasmodium falciparum CRT protein determines the stereospecific activity of antimalarial Cinchona alkaloids. *Antimicrob. Agents Chemother.* **56,** 5356–5364 (2012).

162. Lim, P. *et al.* Ex vivo susceptibility of Plasmodium falciparum to antimalarial drugs in western, northern, and eastern Cambodia, 2011-2012: association with molecular markers. *Antimicrob. Agents Chemother.* **57,** 5277–83 (2013).

163. Tanaka, T. Q. *et al.* Potent Plasmodium falciparum Gametocytocidal Activity of Diaminonaphthoquinones, Lead Antimalarial Chemotypes Identified in an Antimalarial Compound Screen. *Antimicrob. Agents Chemother.* **59,** 1389–1397 (2015).

164. Erath, J. *et al.* Small-Molecule Xenomycins Inhibit All Stages of the Plasmodium Life Cycle. *Antimicrob. Agents Chemother.* **59,** 1427–1434 (2015).

165. Su, X., Hayton, K. & Wellems, T. E. Genetic linkage and association analyses for trait mapping in Plasmodium falciparum. *Nat. Rev. Genet.* **8,** 497–506 (2007).

166. Broman, K. W., Wu, H., Sen, Ś. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19,** 889–890 (2003).

167. Peterson, D. S., Walliker, D. & Wellems, T. E. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. *Proc. Natl. Acad. Sci. U. S. A.* **85,** 9114–8 (1988).

168. Ferdig, M. T. *et al.* Dissecting the loci of low-level quinine resistance in malaria parasites. *Mol. Microbiol.* **52,** 985–97 (2004).

169. Sá, J. M. *et al.* Geographic patterns of Plasmodium falciparum drug resistance distinguished by differential responses to amodiaquine and chloroquine. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 18883–9 (2009).

170. Schnecke, V. & Boström, J. Computational chemistry-driven decision making in lead generation. *Drug Discov. Today* **11,** 43–50 (2006).

171. Malo, N., Hanley, J. A., Cerquozzi, S., Pelletier, J. & Nadon, R. Statistical practice in high-throughput screening data analysis. *Nat. Biotechnol.* **24,** 167–75 (2006).

172. Inglese, J. *et al.* Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 11473–8 (2006).

173. Plouffe, D. *et al.* In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 9059–64 (2008).

174. Weisman, J. L. *et al.* Searching for new antimalarial therapeutics amongst known drugs. *Chem. Biol. Drug Des.* **67,** 409–16 (2006).

175. Chong, C. R., Chen, X., Shi, L., Liu, J. O. & Sullivan, D. J. A clinical drug library screen identifies astemizole as an antimalarial agent. *Nat. Chem. Biol.* **2,** 415–6 (2006).

176. Baniecki, M. L., Wirth, D. F. & Clardy, J. High-throughput Plasmodium falciparum growth assay for malaria drug discovery. *Antimicrob. Agents Chemother.* **51,** 716–23 (2007).

177. Rottmann, M. *et al.* Spiroindolones, a potent compound class for the treatment of malaria. *Science* **329,** 1175–80 (2010).

178. White, N. J. *et al.* Spiroindolone KAE609 for Falciparum and Vivax Malaria. *N. Engl. J. Med.* **371,** 403–410 (2014).

179. Yuan, J. *et al.* Chemical genomic profiling for antimalarial therapies, response signatures, and molecular targets. *Science* **333,** 724–9 (2011).

180. Yuan, J. *et al.* Genetic mapping of targets mediating differential chemical phenotypes in Plasmodium falciparum. *Nat. Chem. Biol.* **5,** 765–71 (2009).

181. Huang, R. *et al.* The NCGC pharmaceutical collection: a comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics. *Sci. Transl. Med.* **3,** 80ps16 (2011).

182. Spangenberg, T. *et al.* The Open Access Malaria Box: A Drug Discovery Catalyst for Neglected Diseases. *PLoS One* **8,** (2013).

183. Sullivan, J. S. *et al.* Adaptation of a strain of Plasmodium falciparum from Ghana to Aotus lemurinus griseimembra, A. nancymai, and A. vociferans monkeys. *Am. J. Trop. Med. Hyg.* **69,** 593–600 (2003).

184. Lopera-Mesa, T. M. *et al.* Effect of red blood cell variants on childhood malaria in Mali: a prospective cohort study. *Lancet Haematol.* **2,** 140–149 (2015).

185. Bhasin, V. K. & Trager, W. Gametocyte-forming and non-gametocyte-forming clones of Plasmodium falciparum. *Am. J. Trop. Med. Hyg.* **33,** 534–7 (1984).

186. Oduola, A. M., Milhous, W. K., Weatherly, N. F., Bowdre, J. H. & Desjardins, R. E. Plasmodium falciparum: induction of resistance to mefloquine in cloned strains by continuous drug exposure in vitro. *Exp. Parasitol.* **67,** 354–60 (1988).

187. Lambros, C. & Vanderberg, J. P. Synchronization of Plasmodium falciparum Erythrocytic Stages in Culture. *J. Parasitol.* **65,** 418 (1979).

188. Moll, K., Kaneko, A., Scherf, A. & Wahlgren, M. *Methods in Malaria Research*. (EviMalaR/MR4, 2013).

189. Zhang, J. H., Chung, T. & Oldenburg, K. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* **4,** 67–73 (1999).

190. Bohórquez, E. B., Juliano, J. J., Kim, H. S. & Meshnick, S. R. Mefloquine exposure induces cell cycle delay and reveals stage-specific expression of the pfmdr1 gene. *Antimicrob. Agents Chemother.* **57,** 833–839 (2013).

191. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25,** 402–408 (2001).

192. Affymetrix White Paper, A. BRLMM-P : a Genotype Calling Method for the SNP 5.0 Array. 1–16 (2007).

193. Mu, J. *et al.* Recombination hotspots and population structure in Plasmodium falciparum. *PLoS Biol.* **3,** e335 (2005).

194. Jiang, H. *et al.* High recombination rates and hotspots in a Plasmodium falciparum genetic cross. *Genome Biol.* **12,** R33 (2011).

195. Kosambi, D. The estimation of map distances from recombination values. *Ann. Eugen.* **12,** 172–175 (1943).

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

197. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–303 (2010).

198. Haldane, J. The combination of linkage values and the calculation of distances between linked factors. *J Genet* **8,** 299–309 (1919).

199. Gentleman, R. I. and R. R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* **5,** 299–314 (1996).

200. Shen, M. *et al.* Identification of Therapeutic Candidates for Chronic Lymphocytic Leukemia from a Library of Approved Drugs. *PLoS One* **8,** (2013).

201. Huang, R. *et al.* Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. *Sci. Rep.* **4,** 5664 (2014).

202. Sobell, H. M. Actinomycin and DNA transcription. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 5328–5331 (1985).

203. *Martindale: The Complete Drug Reference*. (Pharmaceutical Press, 2014).

204. Hefti, F. F. Requirements for a lead compound to become a clinical candidate. *BMC Neurosci.* **9 Suppl 3,** S7 (2008).

205. Srivastava, I. K., Rottenberg, H. & Vaidya, a. B. Atovaquone, a broad spectrum antiparasitic drug, collapses mitochondrial membrane potential in a malarial parasite. *J. Biol. Chem.* **272,** 3961–6 (1997).

206. Korsinczky, M. *et al.* Mutations in Plasmodium falciparum cytochrome b that are associated with atovaquone resistance are located at a putative drug-binding site. *Antimicrob. Agents Chemother.* **44,** 2100–8 (2000).

207. Witschel, M., Rottmann, M., Kaiser, M. & Brun, R. Agrochemicals against Malaria, Sleeping Sickness, Leishmaniasis and Chagas Disease. *PLoS Negl. Trop. Dis.* **6,** (2012).

208. Gokhale, N. H. *et al.* Transition metal complexes of buparvaquone as potent new antimalarial agents: 1. Synthesis, X-ray crystal-structures, electrochemistry and antimalarial activity against Plasmodium falciparum. *J. Inorg. Biochem.* **95,** 249–258 (2003).

209. Wainwright, M. Acridine-a neglected antibacterial chromophore. *J. Antimicrob. Chemother.* **47,** 1–13 (2001).

210. Plowe, C. V. *et al.* Mutations in Plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. *J. Infect. Dis.* **176,** 1590–6 (1997).

211. Mita, T. *et al.* Indigenous evolution of Plasmodium falciparum pyrimethamine resistance multiple times in Africa. *J. Antimicrob. Chemother.* **63,** 252–5 (2009).

212. Abdul-Ghani, R., Farag, H. F. & Allam, A. F. Sulfadoxine-pyrimethamine resistance in Plasmodium falciparum: A zoomed image at the molecular level within a geographic context. *Acta Tropica* **125,** 163–190 (2013).

213. Walter, R. D., Bergmann, B., Kansy, M., Wiese, M. & Seydel, J. K. Pyrimethamin-resistant Plasmodium falciparum lack cross-resistance to methotrexate and 2,4-diamino-5-(substituted benzyl) pyrimidines. *Parasitol. Res.* **77,** 346–350 (1991).

214. Schuster, F. L. Cultivation of plasmodium spp. *Clinical Microbiology Reviews* **15,** 355–364 (2002).

215. Huston, M. & Levinson, M. Are one or two dangerous? Quinine and quinidine exposure in toddlers. *J. Emerg. Med.* **31,** 395–401 (2006).

216. Eckardt, L. *et al.* Drug-related torsades de pointes in the isolated rabbit heart: comparison of clofilium, d,l-sotalol, and erythromycin. *J. Cardiovasc. Pharmacol.* **32,** 425–434 (1998).

217. Bitonti, A. J. *et al.* Reversal of chloroquine resistance in malaria parasite Plasmodium falciparum by desipramine. *Science* **242,** 1301–1303 (1988).

218. Dutta, P., Pinto, J. & Rivlin, R. Antimalarial properties of imipramine and amitriptyline. *J. Protozool.* **37,** 54–58 (1990).

219. Basco, L. K. & Le Bras, J. Reversal of chloroquine resistance with desipramine in isolates of Plasmodium falciparum from Central and West Africa. *Trans. R. Soc. Trop. Med. Hyg.* **84,** 479–481

220. Salama, A. & Facer, C. A. Desipramine reversal of chloroquine resistance in wild isolates of Plasmodium falciparum. *Lancet* **335,** 164–165 (1990).

221. Maïga, O. *et al.* A shared Asian origin of the triple-mutant dhfr allele in Plasmodium falciparum from sites across Africa. *J. Infect. Dis.* **196,** 165–72 (2007).

222. Saliba, K. J. & Kirk, K. Clotrimazole inhibits the growth of Plasmodium falciparum in vitro. *Trans. R. Soc. Trop. Med. Hyg.* **92,** 666–667 (1998).

223. Tiffert, T., Ginsburg, H., Krugliak, M., Elford, B. C. & Lew, V. L. Potent antimalarial activity of clotrimazole in in vitro cultures of Plasmodium falciparum. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 331–336 (2000).

224. Trivedi, V. *et al.* Clotrimazole inhibits hemoperoxidase of Plasmodium falciparum and induces oxidative stress: Proposed antimalarial mechanism of clotrimazole. *J. Biol. Chem.* **280,** 41129–41136 (2005).

225. Gullingsrud, J. *et al.* High-Throughput Screening Platform Identifies Small Molecules That Prevent Sequestration of Plasmodium falciparum-Infected Erythrocytes. *J. Infect. Dis.* **211,** 1134–1143 (2014).

226. Wilson, C. M. *et al.* Amplification of a gene related to mammalian mdr genes in drug-resistant Plasmodium falciparum. *Science* **244,** 1184–1186 (1989).

227. Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494,** 234–7 (2013).

228. Foote, S. J. *et al.* Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in Plasmodium falciparum. *Nature* **345,** 255–8 (1990).

229. Duraisingh, M. T. *et al.* The tyrosine-86 allele of the pfmdr1 gene of Plasmodium falciparum is associated with increased sensitivity to the anti-malarials mefloquine and artemisinin. *Mol. Biochem. Parasitol.* **108,** 13–23 (2000).

230. Ngo, T. *et al.* Analysis of pfcrt, pfmdr1, dhfr, and dhps mutations and drug sensitivities in Plasmodium falciparum isolates from patients in Vietnam before and after treatment with artemisinin. *Am. J. Trop. Med. Hyg.* **68,** 350–356 (2003).

231. Pickard, A. L. *et al.* Resistance to Antimalarials in Southeast Asia and Genetic Polymorphisms in pfmdr1 Resistance to Antimalarials in Southeast Asia and Genetic Polymorphisms in pfmdr1. *Antimicrob. Agents Chemother.* **47,** 2418–2423 (2003).

232. Reed, M. B., Saliba, K. J., Caruana, S. R., Kirk, K. & Cowman, a F. Pgh1 modulates sensitivity and resistance to multiple antimalarials in Plasmodium falciparum. *Nature* **403,** 906–909 (2000).

233. Cowman, A. F., Galatis, D. & Thompson, J. K. Selection for mefloquine resistance in Plasmodium falciparum is linked to amplification of the pfmdr1 gene and cross-resistance to halofantrine and quinine. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 1143–1147 (1994).

234. Peel, S. A., Bright, P., Yount, B., Handy, J. & Baric, R. S. A strong association between mefloquine and halofantrine resistance and amplification,

overexpression, and mutation in the P-glycoprotein gene homolog (pfmdr) of Plasmodium falciparum in vitro. *Am. J. Trop. Med. Hyg.* **51,** 648–58 (1994).

235. Esteller, M., Martinez-Palones, J. M., García, A., Xercavins, J. & Reventós, J. High rate of MDR-1 and heterogeneous pattern of MRP expression without gene amplification in endometrial cancer. *Int. J. cancer.* **63,** 798–803 (1995).

236. Chuanchuen, R. *et al.* Cross-resistance between triclosan and antibiotics in Pseudomonas aeruginosa is mediated by multidrug efflux pumps: Exposure of a susceptible mutant strain to triclosan selects nfxB mutants overexpressing MexCD-OprJ. *Antimicrob. Agents Chemother.* **45,** 428–432 (2001).

237. Myrick, A., Munasinghe, A., Patankar, S. & Wirth, D. F. Mapping of the Plasmodium falciparum multidrug resistance gene 5'-upstream region, and evidence of induction of transcript levels by antimalarial drugs in chloroquine sensitive parasites. *Mol. Microbiol.* **49,** 671–683 (2003).

238. Horrocks, P., Wong, E., Russell, K. & Emes, R. D. Control of gene expression in Plasmodium falciparum - Ten years on. *Mol. Biochem. Parasitol.* **164,** 9–25 (2009).

239. Fidock, D. a, Rosenthal, P. J., Croft, S. L., Brun, R. & Nwaka, S. Antimalarial drug discovery: efficacy models for compound screening. *Nat. Rev. Drug Discov.* **3,** 509–20 (2004).

240. Belluti, F. *et al.* Design, synthesis, and biological and crystallographic evaluation of novel inhibitors of Plasmodium falciparum enoyl-ACP-reductase (PfFabI). *J. Med. Chem.* **56,** 7516–26 (2013).

241. Perozzo, R. *et al.* Structural elucidation of the specificity of the antibacterial agent triclosan for malarial enoyl acyl carrier protein reductase. *J. Biol. Chem.* **277,** 13106–13114 (2002).

242. Surolia, N. & Surolia, A. Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of Plasmodium falciparum. *Nat. Med.* **7,** 167–173 (2001).

243. McLeod, R. *et al.* Triclosan inhibits the growth of Plasmodium falciparum and Toxoplasma gondii by inhibition of apicomplexan Fab I. *Int. J. Parasitol.* **31,** 109–113 (2001).

244. Ménard, D. *et al.* Global analysis of Plasmodium falciparum Na+/H+ exchanger (pfnhe-1) allele polymorphism and its usefulness as a marker of in vitro resistance to quinine. *International Journal for Parasitology: Drugs and Drug Resistance* **3,** 8–19 (2013).

245. Poyomtip, T. *et al.* Polymorphisms of the pfmdr1 but not the pfnhe-1 gene is associated with in vitro quinine sensitivity in Thai isolates of Plasmodium falciparum. *Malaria Journal* **11,** 7 (2012).

246. Pelleau, S. *et al.* Differential association of Plasmodium falciparum Na +/H + exchanger polymorphism and quinine responses in field- and culture-adapted

isolates of Plasmodium falciparum. *Antimicrob. Agents Chemother.* **55,** 5834–5841 (2011).

247. Witkowski, B. *et al.* Novel phenotypic assays for the detection of artemisinin-resistant Plasmodium falciparum malaria in Cambodia: in-vitro and ex-vivo drug-response studies. *Lancet. Infect. Dis.* **13,** 1043–9 (2013).

248. Kemp, D. J., Coppel, R. L. & Anders, R. F. Repetitive proteins and genes of malaria. *Annu. Rev. Microbiol.* **41,** 181–208 (1987).

249. DePristo, M. A., Zilversmit, M. M. & Hartl, D. L. On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins. *Gene* **378,** 19–30 (2006).

250. Zilversmit, M. M. *et al.* Low-complexity regions in plasmodium falciparum: Missing links in the evolution of an extreme genome. *Mol. Biol. Evol.* **27,** 2198–2209 (2010).

251. Claessens, A. *et al.* Generation of Antigenic Diversity in Plasmodium falciparum by Structured Rearrangement of Var Genes During Mitosis. *PLoS Genet.* **10,** e1004812 (2014).

252. Ocholla, H. *et al.* Whole-Genome Scans Provide Evidence of Adaptive Evolution in Malawian Plasmodium falciparum Isolates. *J. Infect. Dis.* jiu349–(2014). doi:10.1093/infdis/jiu349

253. Lemieux, J. E. *et al.* Genome-wide profiling of chromosome interactions in Plasmodium falciparum characterizes nuclear architecture and reconfigurations associated with antigenic variation. *Mol. Microbiol.* **90,** 519–537 (2013).

254. Miotto, O. *et al.* Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. *Nat. Genet.* **45,** 648–655 (2013).

255. Rathod, P. K., McErlean, T. & Lee, P. C. Variations in frequencies of drug resistance in Plasmodium falciparum. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 9389–93 (1997).

256. Templeton, T. J. The varieties of gene amplification, diversification and hypervariability in the human malaria parasite, Plasmodium falciparum. *Mol. Biochem. Parasitol.* **166,** 109–16 (2009).

257. Chan, J.-A., Fowkes, F. J. I. & Beeson, J. G. Surface antigens of Plasmodium falciparum-infected erythrocytes as immune targets and malaria vaccine candidates. *Cell. Mol. Life Sci.* **71,** 3633–57 (2014).

258. Barry, A. E. & Arnott, A. Strategies for designing and monitoring malaria vaccines targeting diverse antigens. *Front. Immunol.* **5,** 359 (2014).

259. Sahar, T. *et al.* Plasmodium falciparum reticulocyte binding-like homologue protein 2 (PfRH2) is a key adhesive molecule involved in erythrocyte invasion. *PLoS One* **6,** (2011).

260. Duraisingh, M. T. *et al.* Phenotypic variation of Plasmodium falciparum merozoite proteins directs receptor targeting for invasion of human erythrocytes. *EMBO J.* **22,** 1047–1057 (2003).

261. Sepúlveda, N. *et al.* A Poisson hierarchical modelling approach to detecting copy number variation in sequence coverage data. *BMC Genomics* **14,** 128 (2013).

262. Chung, W. Y. *et al.* The CLAG/RhopH1 locus on chromosome 3 of Plasmodium falciparum: Two genes or two alleles of the same gene? *Mol. Biochem. Parasitol.* **151,** 229–232 (2007).

263. Persson, K. E. M. *et al.* Erythrocyte-binding antigens of Plasmodium falciparum are targets of human inhibitory antibodies and function to evade naturally acquired immunity. *J. Immunol.* **191,** 785–94 (2013).

264. Cowman, A. F. & Crabb, B. S. Invasion of red blood cells by malaria parasites. *Cell* **124,** 755–766 (2006).

265. Persson, K. E. M. *et al.* Variation in use of erythrocyte invasion pathways by Plasmodium falciparum mediates evasion of human inhibitory antibodies. *J. Clin. Invest.* **118,** 342–351 (2008).

266. Cheeseman, I. H. *et al.* Gene copy number variation throughout the Plasmodium falciparum genome. *BMC Genomics* **10,** 353 (2009).

267. Crosnier, C. *et al.* Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum. *Nature* 4–8 (2011). doi:10.1038/nature10606

268. Theron, M., Hesketh, R. L., Subramanian, S. & Rayner, J. C. An adaptable two-color flow cytometric assay to quantitate the invasion of erythrocytes by Plasmodium falciparum parasites. *Cytom. Part A* **77,** 1067–1074 (2010).

269. Nguitragool, W. *et al.* Malaria parasite clag3 genes determine channel-mediated nutrient uptake by infected red blood cells. *Cell* **145,** 665–77 (2011).

270. Iriko, H. *et al.* Diversity and evolution of the rhoph1/clag multigene family of Plasmodium falciparum. *Mol. Biochem. Parasitol.* **158,** 11–21 (2008).

271. Salcedo-Sora, J. E. & Ward, S. A. The folate metabolic network of Falciparum malaria. *Mol. Biochem. Parasitol.* **188,** 51–62 (2013).

272. Heinberg, A. *et al.* Direct evidence for the adaptive role of copy number variation on antifolate susceptibility in Plasmodium falciparum. *Mol. Microbiol.* **88,** 702–712 (2013).

273. Samarakoon, U. *et al.* The landscape of inherited and de novo copy number variants in a Plasmodium falciparum genetic cross. *BMC Genomics* **12,** 457 (2011).

274. Taylor, M. S., Ponting, C. P. & Copley, R. R. Occurrence and consequences of coding sequence insertions and deletions in Mammalian genomes. *Genome Res.* **14,** 555–66 (2004).

275. Glushakova, S. *et al.* Hemoglobinopathic erythrocytes affect the intraerythrocytic multiplication of Plasmodium falciparum in vitro. *J. Infect. Dis.* **210,** 1100–1109 (2014).

276. Hill, W. G., Babiker, H. a, Ranford-Cartwright, L. C. & Walliker, D. Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malaria parasites. *Genet. Res.* **65,** 53–61 (1995).

277. Pollak, E. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* **117,** 353–360 (1987).

278. Jensen-Seaman, M. I. *et al.* Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res.* **14,** 528–538 (2004).

279. Carroll, D. Genome engineering with zinc-finger nucleases. *Genetics* **188,** 773–782 (2011).

280. Straimer, J. *et al.* Site-specific genome editing in Plasmodium falciparum using engineered zinc-finger nucleases. *Nat. Methods* **9,** 993–8 (2012).

281. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review of Genetics* **45,** 273–297 (2011).

282. Wagner, J. C., Platt, R. J., Goldfless, S. J., Zhang, F. & Niles, J. C. Efficient CRISPR-Cas9–mediated genome editing in Plasmodium falciparum. *Nat. Methods* **11,** 1–6 (2014).

283. Ghorbal, M. *et al.* Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nat. Biotechnol.* **32,** (2014).

284. Packard, R. M. *The Making of a Tropical Disease: A Short History of Malaria*. *Emerging Infectious Diseases* **14,** (Johns Hopkins University Press, 2007).

285. Noedl, H. *et al.* Artemisinin resistance in Cambodia: a clinical trial designed to address an emerging problem in Southeast Asia. *Clin. Infect. Dis.* **51,** e82–9 (2010).

286. Hayton, K. *et al.* Various PfRH5 polymorphisms can support Plasmodium falciparum invasion into the erythrocytes of owl monkeys and rats. *Mol. Biochem. Parasitol.* **187,** 103–10 (2013).

287. Echeverry, D. F. *et al.* Short report: polymorphisms in the pfcrt and pfmdr1 genes of Plasmodium falciparum and in vitro susceptibility to amodiaquine and desethylamodiaquine. *Am. J. Trop. Med. Hyg.* **77,** 1034–8 (2007).

288. Nkrumah, L. J. *et al.* Probing the multifactorial basis of Plasmodium falciparum quinine resistance: evidence for a strain-specific contribution of the sodium-proton exchanger PfNHE. *Mol. Biochem. Parasitol.* **165,** 122–31 (2009).

289. Aurrecoechea, C. *et al.* PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.* **37,** D539–43 (2009).

290. Mita, T., Tanabe, K. & Kita, K. Spread and evolution of Plasmodium falciparum drug resistance. *Parasitol. Int.* **58,** 201–9 (2009).

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

292. Combrinck, J. M. *et al.* Insights into the role of heme in the mechanism of action of antimalarials. *ACS Chem. Biol.* **8,** 133–7 (2013).

293. Chou, A. C., Chevli, R. & Fitch, C. D. Ferriprotoporphyrin IX fulfills the criteria for identification as the chloroquine receptor of malaria parasites. *Biochemistry* **19,** 1543–9 (1980).

294. Egan, T. J., Ross, D. C. & Adams, P. A. Quinoline anti-malarial drugs inhibit spontaneous formation of beta-haematin (malaria pigment). *FEBS Lett.* **352,** 54–7 (1994).

295. Rush, M. a *et al.* Colorimetric high-throughput screen for detection of heme crystallization inhibitors. *Antimicrob. Agents Chemother.* **53,** 2564–8 (2009).

296. Holmgren, G. *et al.* Amodiaquine resistant Plasmodium falciparum malaria in vivo is associated with selection of pfcrt 76T and pfmdr1 86Y. *Infect. Genet. Evol.* **6,** 309–14 (2006).

297. Takahashi, N. *et al.* Large-scale survey for novel genotypes of Plasmodium falciparum chloroquine-resistance gene pfcrt. *Malar. J.* **11,** 92 (2012).

298. Uhlemann, A. C. & Krishna, S. Antimalarial multi-drug resistance in Asia: mechanisms and assessment. *Curr. Top. Microbiol. Immunol.* **295,** 39–53 (2005).

299. Lopera-Mesa, T. M. *et al.* Plasmodium falciparum Clearance Rates in Response to Artesunate in Malian Children With Malaria: Effect of Acquired Immunity. *J. Infect. Dis.* **207,** 1655–63 (2013).

300. Kublin, J. G. *et al.* Molecular markers for failure of sulfadoxine-pyrimethamine and chlorproguanil-dapsone treatment of Plasmodium falciparum malaria. *J. Infect. Dis.* **185,** 380–8 (2002).

301. Kuhn, S., Gill, M. J. & Kain, K. C. Emergence of atovaquone-proguanil resistance during treatment of Plasmodium falciparum malaria acquired by a non-immune north American traveller to west Africa. *Am. J. Trop. Med. Hyg.* **72,** 407–9 (2005).

302. Meshnick, S. R. Why does Quinine still work after 350 years of use? *Parasitol. Today* **13,** 89–90 (1997).

303. Jarcho, S. & Torti, F. *Quinine's predecessor: Francesco Torti and the early history of cinchona*. (Johns Hopkins University Press, 1993). at <http://books.google.co.uk/books?id=0eFsAAAAMAAJ>

304. Pukrittayakamee, S., Supanaranond, W., Looareesuwan, S., Vanijanonta, S. & White, N. J. Quinine in severe falciparum malaria: evidence of declining efficacy in Thailand. *Trans. R. Soc. Trop. Med. Hyg.* **88,** 324–7 (1994).

305. Le Bras, J. & Durand, R. The mechanisms of resistance to antimalarial drugs in Plasmodium falciparum. *Fundam. Clin. Pharmacol.* **17,** 147–53 (2003).

306. Foley, M. & Tilley, L. Quinoline antimalarials: mechanisms of action and resistance. *Int. J. Parasitol.* **27,** 231–40 (1997).

307. Cheruiyot, J. *et al.* Polymorphisms in Pfmdr1, Pfcrt, and Pfnhe1 genes are associated with reduced in vitro activities of quinine in Plasmodium falciparum isolates from western Kenya. *Antimicrob. Agents Chemother.* **58,** 3737–43 (2014).

308. Schlitzer, M. Malaria chemotherapeutics part I: History of antimalarial drug development, currently used therapeutics, and drugs in clinical development. *ChemMedChem* **2,** 944–86 (2007).

309. Price, R. N. *et al.* Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet* **364,** 438–47 (2004).

310. Wilson, C. M. *et al.* Amplification of pfmdr 1 associated with mefloquine and halofantrine resistance in Plasmodium falciparum from Thailand. *Mol. Biochem. Parasitol.* **57,** 151–60 (1993).

311. Nateghpour, M., Ward, S. a & Howells, R. E. Development of halofantrine resistance and determination of cross-resistance patterns in Plasmodium falciparum. *Antimicrob. Agents Chemother.* **37,** 2337–43 (1993).

312. Duraisingh, M. T. & Cowman, A. F. Contribution of the pfmdr1 gene to antimalarial drug-resistance. *Acta Trop.* **94,** 181–90 (2005).

313. Davis, T. M. E., Hung, T., Sim, I., Karunajeewa, H. A. & Ilett, K. F. Piperaquine: a resurgent antimalarial drug. *Drugs* **65,** 75–87 (2005).

314. Gargano, N., Cenci, F. & Bassat, Q. Antimalarial efficacy of piperaquine-based antimalarial combination therapies: Facts and uncertainties. *Trop. Med. Int. Heal.* **16,** 1466–1473 (2011).

315. Basco, L. K. & Ringwald, P. In vitro activities of piperaquine and other 4-aminoquinolines against clinical isolates of Plasmodium falciparum in Cameroon. *Antimicrob. Agents Chemother.* **47,** 1391–4 (2003).

316. O'Neill, P. M., Barton, V. E. & Ward, S. a. The molecular mechanism of action of artemisinin--the debate continues. *Molecules* **15,** 1705–21 (2010).

317. Haynes, R. & Cheu, K. Considerations on the Mechanism of Action of Artemisinin Antimalarials: Part 1-The 'Carbon Radical' and 'Heme' Hypotheses. *Infect. Disord. Targets* **13,** 217–277 (2013).