UNIVERSITY OF CALIFORNIA, SAN DIEGO

Know thy Enemy: Exploring Pathogenic Evolution, Resistance, and Virulence in *Plasmodium falciparum* and *Leptospira interrogans*

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Victoria Catherine Corey

Committee in charge:

Professor Elizabeth Winzeler, Chair
Professor Trey Ideker
Professor Nathan Lewis
Professor James McKerrow
Professor Victor Nizet

2016

The Dissertation of Victoria Catherine Corey is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2016

TABLE OF CONTENTS

LIST OF SUPPLEMENTAL FILES

Chapter 2

Supplemental-Table-2-1_Corey.xlsx – *Summary of killing rate and selection success for compound panel.*

Supplemental-Table-2-2_Corey.xlsx – *Whole genome sequencing summary for MMV008149*

Supplemental-Table-2-3_Corey.xlsx – *Multi-stage activity of compound panel*

Supplemental-Table-2-4_Corey.xlsx – *Individual HaplotypeCaller Sequencing Data*

Supplemental-Table-2-5_Corey.xlsx – *Testing CNV caller with Dd2 parent samples*

Supplemental-Table-2-6_Corey.xlsx – *Individual Copy Number Variation (CNV) sequencing data*

Supplemental-Table-2-7_Corey.xlsx – *Enriched GO terms identified in PlasmoDB for SNVs, INDELs, and SNV/INDELs combined gene lists*

Supplemental-Table-2-8_Corey.xlsx – *Individual gene OPI term results for all 20 selection sets*

Supplemental-Table-2-9_Corey.xlsx – *Summary of OPI term overlapping within selection sets*

Supplemental-Table-2-10_Corey.xlsx – *Raw EC_{50} values for entire cross-resistance assay*

Chapter 3

Supplemental-Table-3-1_Corey.xlsx – *Whole genome sequencing haplotypecaller sequencing results for Gen16, Gen28, and Gen50*

Supplemental-Table-3-2_Corey.xlsx – *Summary lists of PfCARL allele mutations identified as conferring benzimidazolyl piperidine and imidazolopiperazine resistance*

Chapter 4

Supplemental-Table-4-1_Corey.xlsx – *Genome coordinates of SNVs with a change in allelic frequency*

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Elizabeth Winzeler, for her guidance and advice throughout my career at UCSD. The opportunity to discover my passion in bioinformatics and apply my newfound knowledge in the malarial field with a series of collaborators was pivotal to my development as a scientist and computational biologist. Without that, I would not be the scientist I am today. I would also like to thanks the past and present members of the Winzeler lab, particularly: Micah Manary and Taylor Bright, for the introduction into unix and whole genome sequencing analysis; Gregory Lamonte, playing the role of mentor for me and always being there as a sounding board from writing to experimental design; Stephan Meister, for the opportunity to gain exposure to compound clustering and further my knowledge in programming languages; and Matt Abraham, Justine Swann, Melanie Wree, and Erika Flannery, for their support and advice throughout my time in the lab.

Next, I would like to acknowledge all of the members of my dissertation committee, including Nathan Lewis, Trey Ideker, James McKerrow, and Victor Nizet. Each committee member has gone above and beyond, providing generous feedback and insight into not only my doctoral work but into future career opportunities. I would especially like to thank Nathan Lewis and Trey Ideker for their reflection into where I was

as a scientist and what I needed to execute in order to accomplish my professional goals.

This work has only been possible due to an extreme collaborative effort, and I would like to thank the labs of David Fidock, Dyann Wirth, Dan Goldberg, and Francisco Javier Gamo-Benito, with which this project became reality. I would additionally like to thank Marcus Lee, Amanda Lukens, Pamela Magistrado, and Eva Istvan, who not only helped me design the *in vitro* selection experimentation, but who were all willing to answer late night emails with varying paper drafts and provide sanity to this project.

I am grateful to the multiple sources of funding that have supported my graduate work throughout the last few years. The research presented in this dissertation was supported in part by a grant from the Bill and Melinda Gates Foundation (OPP1040406) and through an institutional training grant from the National Institute of General Medical Sciences (T32 GM008666).

In addition, I would like to thank my fellow classmates from the department of Biomedical Sciences, specifically Aaron Springer, Steve Rees, and Aaron Friedman. Each of them not only encouraged me to take breaks and "sharpen my axe," but also were fantastic scientists colleagues, always ready with a listening ear to help out a classmate in a

different field design an experiment or solve a problem. I would not be here without their constant support, and I hope we continue to have each other's scientific backs in the years to come.

Finally, I am extremely grateful for the support I received from my family. My parents, Victor and Diane Doroski, were always there to push me to be the best I could be, and were always that note of inspiration throughout my graduate career. My father subconsciously encouraged me to pursue genetics, a field I am now proud to be a part of. I would also like to thank my parent-in-laws, William and Hanalee Washburn and Craig and Linda Corey, for their relentless bragging of their daughter-in-law's work, and their constant questions which pushed me to truly be an expert in the malarial field. Above all, I would like to thank my husband, Colin Corey, for so many things. From his initial enthusiasm of moving to San Diego so I could work towards my doctorate, to his hugs and notes of encouragement when experiments and writing seemed to take a turn for the worst, I would not be where I am today without him.

Chapter 2, in full, has been submitted for publication for the material as it may appear in Nature Communications, 2016. Victoria C. Corey, Amanda K. Lukens, Eva S. Istvan, Marcus C.S. Lee, Virginia Franco, Pamela Magistrado, Olivia Coburn-Flynn, Tomoyo Sakata-Kato, Olivia Fuchs, Nina Gnadig, Greg Goldgof, Maria Linares, Maria G. Gomez-

Lorenzo, Maria Jose Lafuente-Monasterio, Sara Prats, Stephan Meister, Olga Tanaseichuk, Melanie Wree, Yingyao Zhou, Paul Willis, Francisco-Javier Gamo, Daniel E. Goldberg, David A. Fidock, Dyann F. Wirth, Elizabeth A. Winzeler, "A Broad Analysis of Resistance Development in the Malaria Parasite". The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been submitted for publication for the material as it may appear in American Chemical Society (ACS) Infectious Diseases, 2016. Victoria C. Corey and Pamela A. Magistrado, Amanda K. Lukens, Greg LaMonte, Erika Sasaki, Stephan Meister, Melanie Wree, Dyann F. Wirth, Elizabeth Winzeler, "Plasmodium falciparum Cyclic Amine Resistance Locus, PfCARL: A Resistance Mechanism for Two Distinct Compound Classes". The dissertation author was the co-primary investigator and author of this paper.

Chapter 4, in full, has been accepted for publication for the material as it may appear in the American Society of Tropical Medicine and Hygiene, 2016. Victoria C. Corey and Jason S. Lehmann, Jessica N. Ricaldi, Joseph M. Vinetz, Elizabeth A. Winzeler, Michael A. Matthias, "Whole Genome Shotgun Sequencing Shows Selection on *Leptospira* Regulatory Proteins During *in vitro* Culture Attenuation". The dissertation author was the co-primary investigator and author of this paper.

VITA

2010 Bachelor of Science, California Polytechnic State University, San Luis Obispo

2010-2012 Pharmacology Contractor, Achaogen, South San Francisco

2012-2016 Research Assistant, University of California, San Diego

2016 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

V. Corey and J. Lehmann, et al. *Whole Genome Shotgun Sequencing Shows Selection on Leptospira Regulatory Proteins During In vitro-Culture Attenuation. AJTMH 2015, publication in process*

V. Corey and P. Magistrado, et al. *Mutations in the Plasmodium falciparum cyclic amine resistance locus, pfcarl, confer resistance to two distinct compound classes - benzimidazolyl piperidines and imidazolopiperazines ACS Infect. Dis. 2016*, Article ASAP.

J. Swann, **V. Corey**, C. Scherer, E. Comer, et al. *A high-throughput luciferase-based assay identifies compounds with liver stage antimalarial activity ACS Infectious Diseases* 2016, 2(4):281-293.

E.L. Flannery, T. Wang, A. Akbari, **V.C. Corey**, F. Gunawan, et al. *Next-Generation sequencing of Plasmodium vivax patient samples shows*

evidence of direct evolution in drug-resistance genes. ACS Infect. Dis 2015, 1(8):367-379.

M. Manary, S. Singhakul, E. Flannery, S. Bopp, **V. Corey**, A. Bright, C. McNamera, J. Walker, and E. Winzeler. *Identification of pathogen genomic variants through an integrated pipeline.* BMC Bioinformatics 2014, 15:63.

C.L. Ng, G. Sicilano, M.C.S. Lee, M.J. Almedia, S.E. Bopp, L. Bertuccini, S. Wittlin, R. Kasdin, **V.C. Corey**, et al. *Cas9-modified pfmdr1 confers P. falciparum asexual blood stage and gametocyte resistance to piperazines and potentiates lumefantrine* Mol. Microbio. 2016, Article ASAP.

V. Corey, et al. *Comprehensive analysis of resistance development in the malaria parasite* Nat Comm, (2016): Manuscript in revisions

N. Kato, E. Comer, T. Sakata-Kato, M. Maetani, J. Bastien, **V. Corey**, D. Clarke, et al. *Diversity synthesis yields multistage antimalarial inhibitors including of a novel target that results in low single-dose cures in mice* Nature, (2016): Manuscript in revisions

M. Yi-Xiu Lim, G. LaMonte, M. Lee, E.D. Chow, B.H. Tan, B.F. Tjahjadi, A. Chua, **V. Corey**, M. Nachon, et al. *The P. falciparum UDP-galactose transporter and Acetyl-coA transporters are novel multidrug resistance genes against new classes of antimalarial compounds:* Manuscript submitted, Cell Host and Microbe

G. LaMonte, M. Yi-Xiu Lim, M. Wree, M. Nachon, **V. Corey,** D. Plouffe, A. Du, N. Figueroa, P. Bifani, E. Winzeler. *Characterization of the P. falciparum cyclic amine resistance locus, a multi-drug resistance gene against several classes of novel antimalarials:* Manuscript submitted, Cell Host and Microbe

A.H. Lee, **V.C. Corey**, E.A. Winzeler, and D.A. Fidock. *The mutability of multi-drug resistant southeast Asian Plasmodium falciparum malaria parasites:* Manuscript submitted ACS Infectious Diseases

FIELDS OF STUDY

Major Field: Biomedical Sciences (Genetics)

Professor Elizabeth Winzeler

ABSTRACT OF THE DISSERTATION

Know thy Enemy: Exploring Pathogenic Evolution, Resistance, and Virulence in *Plasmodium falciparum* and *Leptospira interrogans*

by

Victoria Catherine Corey

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2016

Professor Elizabeth Winzeler, Chair

We are in a constant battle against pathogens. Millions of individuals die every year due to inadequate diagnosis and emerging resistance to current therapeutics. Improvements in diagnosis and treatment strategies are needed, but these can be hindered by a lack of knowledge regarding a particular pathogen's molecular and resistance

mechanisms. *Plasmodium* and *Leptospira* are two such examples of widespread pathogenic organisms whose genomes are still poorly understood and additional knowledge is desperately needed to improve therapeutic strategies. This dissertation presents two evolution-based strategies that when coupled with whole genome sequencing can be used to identify virulence and resistance associated genes. These include a "forward" approach, which studies the development of resistance, and a "reverse" approach, which examines the loss of virulence.

In Chapter 2, the forward approach is used to examine novel targetable pathways in *Plasmodium falciparum*. Selectively evolving resistance to 50 novel antimalarial compounds, we successfully identify potential targets to 21 compounds and eight novel gene targets. Additionally, this chapter examines resistance development patterns against the compound set, and identifies fast-killing compounds may result in a slower onset of clinical resistance.

Chapter 3 focuses on PfCARL, one potential target identified in Chapter 2, which has been previously described as the target for KAF156, a drug currently in clinical trials. Our data demonstrate that *pfcarl* mutations confer resistance to two distinct compound classes – benzimidazolyl piperidines and imidazolopiperazines. However, these two classes appear to have different timing of action in the asexual blood

stage and different potencies against the liver and sexual blood stages, suggesting *pfcarl* is a multidrug resistance gene rather than a common target.

Finally, using the reverse approach, Chapter 4 identifies virulence-related genes in *Leptospira* by observing cumulative genomic changes occurring after serial *in vitro* passaging of a highly virulent *Leptospira interrogans* strain into a nearly avirulent isogenic derivative. Comparison between these two polyclonal strains identifies 15 non-synonymous single nucleotide variant (nsSNV) alleles that increased in frequency and 19 that decreased. These frequency changes likely contribute to the loss of virulence, and suggest new virulence-associated genes whose role in *Leptospira* pathogenesis should be further studied.

Chapter 1

Introduction: Plasmodium and Leptospira: the Approaches to Identifying Resistance and Virulence

1.1 The Origin of the Pathogen

The term "pathogen," coined in the 1880s, is derived from the Greek words *pathos* and *genēs,* translating poetically into "giving birth to suffering" [1]. Pathogens can refer to any organism that produces disease within a host, and covers a wide variety of organisms, including viruses, bacteria, fungi and parasites [2]. Though sometimes portrayed as a vindictive malcontent invader whose sole purpose is to wreck havoc on the host's body, pathogens are simply following an organism's basic biological need to procreate and live. Their survival, however, is at the expense of another, and when two organisms are vying for the same biological need, it can get hostile.

In the clinic, it seems we are in a constant war against pathogens. Even with the advantage of vaccines and chemotherapies, pathogens are highly adaptable and elusive, resulting in resistance to the various therapies thrown at them. In order to aid future therapeutic efforts, an understanding of how a particular pathogen attacks the host, adapts and develops resistance is vital. With the accessibility and affordability of whole genome sequencing we are in an ideal position to explore pathogenic genomes to answer these questions.

The work described here examines two pathogens, *Plasmodium* and *Leptospira*, whose genomes are still poorly annotated and resistance

mechanism/virulence factors poorly understood [3, 4]. We employ two main approaches: (1) studying the development of resistance (the "forward" approach); and, (2) examining the loss of virulence (the "reverse" approach). With the malaria parasite we utilize the forward approach, examining potential drug gene targets and resistance mechanisms for *Plasmodium falciparum* by using an *in vitro* selection method to develop resistant mutant derivatives and analyzing the genomic changes in the mutant derivative. Focusing next on the *Leptospira* bacteria, we apply the reverse approach to identify virulence factors by passaging a highly virulent *Leptospira interrogans* strain *in vitro* and examining genomic changes distinguishing the parental highly virulent strain from the attenuated derivative.

The first chapter of this dissertation provides an overview of *Plasmodium* and *Leptospira*, focusing on the life cycles and research need for both pathogens. The second chapter presents an examination of malarial chemical resistance: a systematical evalulation of whether or not pre-existing resistance mechanisms confer resistance to a set of 50 antimalarial compounds identified in phenotypic screens. Additionally, this chapter explores the genomic changes within the generated mutants, identifying potential compound gene targets that could be exploited in the design of future antimalarial therapies as well as elucidating possible

general resistance mechanisms in malaria parasites. Within the third chapter we focus on one particular gene identified in Chapter two, a cyclic amine ressitance locus (PfCARL) and investigate whether the gene is a direct target of the compound or a general resistance mechanism. Chapter three additionally examines the implications of PfCARL's function on a preclinical candidate, KAF156, which has resulted in PfCARL resistant parasites during *in vitro* selections. The fourth chapter examines the loss of virulence in an aggressive *Leptospira interrogans* strain, inferring novel virulence candidate genes in the organism. Finally, the last chapter provides concluding perspectives on the information gained for both pathogens, and the implications this work has on the field.

1.2 Plasmodium

1.2.1 Effect on the world

Malaria is a devastating disease that has ravished the world for centuries. Spread by mosquitos, malaria can be found throughout our history. Malarial antigens were found in Egyptian remains in 3200BC, and hundreds of thousand American Civil War soldiers fell ill to malarial infections in the mid 19th century [5, 6]. The initial upsurge of malaria 10,000 years ago has been linked to the increased breeding of the mosquito, as a rise in temperature in Africa and a rise in humidity 7,000-

12,00 years ago increased breeding grounds for the insect vector, allowing for the disease to propagate more freely across human populations [7]. Quinine (isolated from chichona bark), has been the primary treatment for malaria for hundreds of years, with the first malaria quinine treatment dating back to 1631 in Rome [5, 8]. Other forms of malaria treatment were not discovered until 1891, when methylene blue was developed [9, 10]. As the first synthetic drug for malaria, methylene blue launched further efforts in malaria drug design, leading to the development of chloroquine in the mid 1900s [11]. At the same time primaquine was synthesized to treat relapses in *P. vivax* patients [12, 13]. With this arsenal of antimalarials as well as insecticides to fight mosquitos, a war was waged from 1955-1960s to eliminate malaria. While this resulted in malarial elimination in some areas, such as the United States and Europe, it was not enough to eradicate the disease, and resistance began to emerge against the major antimalarials in the 1970s (chloroquine, proguanil, pyrimethamine, and sulfadoxine) [9].

Today, malaria is responsible for the plurality of disease-related deaths in children aged 1-4, and an estimated 214 million cases and ~438,000 deaths in 2015 alone [14, 15]. Economic development, seasonal climate (temperature, humidity, etc.), and transmission-reduction strategies such as spraying with dichloro-diphenyl-trichloroethane (DDT)

have all helped eliminate malaria in the United States, Europe, and other countries further from the equator. Impoverished regions closer to the equator such as parts of South America, sub-Saharan Africa, and South-East Asia, however, continue to struggle against malarial infections [15, 16].

1.2.2 Life cycle and biology

Malaria is caused by protozoan parasites of the genus *Plasmodium*, and is transmitted from person to person by *Anopheline* mosquitos. Four major species infect humans – *P. falciparum, P. vivax, P. ovale,* and *P. malariae,* – with *P. falciparum* having the highest mortality and *P. vivax* having the highest morbidity [17, 18]. *Plasmodium* has a complex lifecycle, shuffling between states depending on its environment (Fig. 1.1).

Plasmodium is initially transferred when an infected female mosquito bites a vertebrate host and sporozoites are introduced into the host's dermis and eventually the blood stream. From here, the sporozoites travel to the liver, where they invade hepatocytes and begin to develop and multiply over a span of 8-16 days depending on the *Plasmodium* species [19]. Some *Plasmodium* species, *P. vivax* being the most common, also develop hypnozoites – a dormant form of the parasite responsible for malarial relapse. Hypnozoites can remain in the liver for months to years before developing into merozoites [18].

Once fully developed, the liver merozoites are released back into the blood stream where they invade erythrocytes and initiate the blood stage of the infection. A majority of the merozoites undergo a ~48hr asexual cycle within red blood cells (RBCs), developing from a ring to a mature trophozoite and eventually dividing into ~20 merozoites during the schizont stage. The schizont ultimately ruptures the RBC, allowing the parasites to spread to other uninfected erythrocytes [17, 20]. This asexual stage is the primary cause of most malarial symptoms in patients, which include flu-like symptoms and anemia (due to the rupturing of blood cells) [21]. Complications of malaria can also include swelling of the brain (cerebral malaria), hypothesized to be due to sequestration of red blood cells in the cerebral microvascular endothelium, affecting blood flow [22].

While the asexual stage continues to cycle, a small fraction of the merozoites develop into gametocytes, the sexual stage of *Plasmodium* and the only transmittable form of the parasite. In *P. falciparum*, gametocytogenesis tends to start 7-15 days after the initiation of the asexual cycle, and takes approximately 10 days to result in mature gametocytes [23]. Gametocyte development is distringuished by morphological changes and is classifcally divied into five stages. Immature gametocytes (stage I-IV) are sequestered in bone marrow and are released back into the blood streem once they mature into a stage V

male or female gametocyte [24]. Once ingested by a mosquito, each gametocyte develops into 1 female or 8 male gametes in the mosquito midgut, based on their sex. A male and female gamete fuse together to form a zygote, which then turns into an ookinete. The ookinete burrows across the midgut wall and subsequently forms an oocyst, which enlarges over time and eventually bursts open to release sporozoites. These sporozoites travel to the salivary glands of the mosquito, and the vicious cycle repeats.

1.2.3 Research need

While malarial treatment began with an element used in a standard gin and tonic cocktail, today, we now supply malaria patients with a more complicated cocktail: a combination-based therapy. Given that the recently licensed Mosquirix vaccine provides only moderate protection [25], chemotherapy still constitutes the best clinical approach for the prevention and treatment of malaria. These consist of 4-aminoquinolines including chloroquine, piperaquine and related compounds; antifolates such as pyrimethamine and cycloguanil; alkanolamines such as halofantrine and lumefantrine; endoperoxides, such as artesunate, artemisinin, artemether; and newer synthetic compounds [26].

Antimalarial treatments are typically administered as combination therapies, and artemisinin-based combination therapies (ACTs, such as artemether-lumefantrine) currently represent the therapy class that is most effective and that is the standard of care recommended by the World Health Organization (WHO). Most drugs administered target the malaria parasite during its replicative cycle within human erythrocytes, the lifecycle stage that is associated with clinical manifestations of malaria [26].

Over the past fifteen years, malaria control measures, including insecticide treated bed nets, insecticide sprays, and ACTs, have resulted in a reduced global burden of malaria [27]. Globally, malaria cases have declined by 18% since the turn of the century, falling from an estimated 262 million in 2000 to 214 million in 2015, and malaria deaths have decreased by 48%, falling from ~839,000 to ~438,000 in the same time period [14]. Of the 100 malaria endemic countries, 35 are pursuing elimination [28]. While these results are encouraging, increased parasite resistance to the most effective drug, artemisinin, may undo our progress [29-33]. Clinical trials with artemisinin mono-therapies have shown that artemisinin is taking considerably longer to clear malaria infections in Southeast Asia—typically twice as long as observed a decade ago [34]. While efforts are being made to restrict the spread of artemisinin-resistant

parasites in Southeast Asia, resistance to other standard antimalarials, including chloroquine, pyrimethamine, and sulfadoxine emerged in the same region of Southeast Asia and subsequently spread to Africa [35]. Indeed, sporadic reports of artemisinin resistance in Africa and South America may foreshadow a widespread loss of artemisinin efficacy [36, 37].

In anticipation of eventual widespread ACT failure there has been a focused and coordinated effort to place new antimalarial candidates into the drug development pipeline (see http://www.mmv.org/research-development/rd-portfolio). The hope is to identify new compounds, ideally pan-lifecycle compounds with novel mechanisms of action, which can be used to define new classes of antimalarials. Two main strategies to identify new compounds are either a phenotypic screen, where millions of compounds are assayed against the parasite in a cell-based assay, or a target-based screen, where thousands of compounds are tested for the ability to directly inhibit an isolated protein target. A target-based approach allows one to screen fewer total compounds and identifies the MOA early on. However, this method requires one already has identified a critically essential targetable protein, and does not test a compound's efficacy against the entire organism. A phenotypic assay approach, on the other hand, requires screening significantly more compounds, and

does not identify how the compound is acting on the organism. However, unlike the target-based screens, a phenotypic assay is run against the entire organism, and compounds successfully identified in the screen already could be expected to act on an essential target. For *Plasmodium*, the genome is still largely unannotated and compounds are required to penetrate not only the parasite but typically also the invaded cell (erythrocyte or hepatocyte), a phenotypic approach is more desirable. Additionally, phenotypic screens provide the opportunity to identify compounds that may act on varying targets, minimizing the tunnel vision that occurs in target-based approaches where one is only focusing on a single protein. Therefore, efforts to identify novel antimalarials in recent years have primarily used a high throughput cell-based assay approach, leading to the screening of close to five million compounds against asexual stage *P. falciparum* over the last decade [38-40].

While some compounds have already been identified as potential antimalarials using a cell-based screening approach, many of these compounds have unknown mechanisms of action (MOA) in the *Plasmodium* parasite. While clinical candidates in the past have advanced despite having an unknown MOA, having this knowledge may be critical in developing alternative compounds if there are

complications in clinical trials. Additionally, beyond identifying new drug targets, there is also a need to gain a broader understanding of malarial resistance development and identify general mechanisms of resistance. Out of the ~5500 genes identified in the Plasmodium falciparum genome, only 40% have been assigned a Gene Ontology term, resulting in 60% of the genes being classified as hypothetical [3]. Because of this, we are unaware of how many potential pathways can be exploited in antimalarial therapies, and which mutations already present in the clinic may result in pre-existing resistance, drastically shortening the life of new therapeutics. Therefore, we are in need of an additional approach examining resistance development that not only identifies a compound's mechanism of action but also defines the parasite on a larger genomic scale.

1.3 Leptospirosis

1.3.1 Effect on the world

Leptospirosis, caused by pathogenic bacteria of the genus *Leptospira*, is an zoonotic infection of global distribution [41]. The Leptospirosis Burden Epidemiology Reference Group estimates there are over 500,000 hospitalized cases per year [42], which is likely an underestimate of the true burden of disease due primarily to inadequate

diagnostics, a lack of clinical awareness, and poor surveillance [43]. Transmission of *Leptospira* to humans occurs via exposure to contaminated water and wet soil or infected tissues and urine from chronically colonized reservoir hosts. Humans living in poverty with poor sanitation are at greatest risk of infection, particularly during seasonal flooding, monsoons, and tropical cyclones [41, 43].

1.3.2 Life cycle and biology

Leptospira are bacteria belonging to the phylum of spirochetes, known for their question-mark shape from their long thin morphology with a hooked end [44, 45]. Leptospires are all slow-growing obligate aerobes and for their sole carbon source utilize long-chain fatty acids which are metabolized via beta-oxidation [45, 46]. They also display a double membrane consisting of a lipopolysaccharide (LPS) outer membrane, similar in composition to that of a gram-negative bacteria, though having lower endotoxic activity [47].

The *Leptospira* genus includes at least 22 species classified into three large subgroups based on 16S rDNA phylogeny, *in vitro* growth characteristics, and virulence [48-51]. There are 15 recognized pathogenic species and seven saprophytic (non-infectious) species. Group I pathogens comprise > 250 serotypes and cause disease varying in severity, ranging from subclinical infections to severe disease—often

associated with renal failure and pulmonary hemorrhage—and death [52]. By contrast, group II species grow better in culture and cause predominantly mild, self-resolving illness without fatal complications.

Similar to *Plasmodium*, there is a transmission cycle for leptospirosis between a reservoir and non-reservoir host, however unlike *Plasmodium*, which passes infection to humans through a single vector organism (mosquitos), *Leptospira* can colonize multiple reservoir animals, including cattle, buffaloes, horses, sheep, goats, pigs, dogs, and rodents [45, 53]. The pathogen colonizes the renal tubes of these reservoir animals, and is routinely shed via urine into the environment [45]. *Leptospira* can transmit between reservoir hosts via sexual transmission, contact with infected urine, or by mother to susceptible offspring, and once colonized most reservoir hosts are chronically infected [46]. Transmission between reservoir and non-reservoir hosts occurs primarily through contact with infected urine or tissues, although indirect infection can occur through contaminated soil or water [41]. Most transmission to the host occurs via abrasions or cuts in the skin, or through exposed mucosal membranes, such as eyes or the nose. Transmission, however, can also occur through the skin in cases of prolonged exposure [46]. The pathogenic spirochete penetrates the host quickly, being detectable in the blood stream minutes after infection and detectable in the target organs by the third day post

infection [53, 54]. While the reservoir host typically sheds *Leptospira* in the urine for a lifetime, the non-reservoir host (ex: humans) has been found to only have renal colonization and leptospiruria for a few months, though there has been some documentation in which renal infections have lasted a year or longer [55].

1.3.3 Research need

Of the pathogenic species, *Leptospira interrogans*, the leading cause of leptospirosis-associated morbidity and mortality in humans, is the most extensively studied species [41, 42]. Nonetheless, very little is understood regarding the virulence mechanisms and even the basic biology of the causative agents of leptospirosis. Only four species for *Leptospira* have been previously assembled, which includes three pathogenic species (*L. interrogans, L. licerasiae,* and *L. borpetersenii*) and one saprophytic species (*L. biflexa*), and functions of most genes (59%) predicted in these genomes are unknown [48, 56-58]. There is additionally large variability between the genomes. Comparing three genomes of the four genomes (*L. interrogans, L. borpetersenii*, and *L. biflexa*), only 2052 proteins were shared, accounting for only 60% of the predicted genes found in *L. interrogans* alone [59]. Even between the two pathogenic species, 627 genes were found to be unique to *L. interrogans* [45].

Typical treatment of leptospirosis infections is with antibiotics, including doxycycline, azithromycin, penicillin, ceftriaxone, and cefotaxime [60]. Additionally, whole *Leptospira*-based vaccines have been used to protect livestock, domestic animals, and even human subjects from infection. Current vaccines, however, have multiple related adverse effects (including poor appetite, vomiting, diarrhea, and lethargy) and only confer short-term immunity to the specific *Leptospira* species used during vaccination [45]. Because of this, there is a strong need for more specific sub-unit vaccines, however a better understanding of the *Leptospira* virulence factors and associated genes is required for the development of these vaccines to occur.

1.4 Proposed Research Strategies to Identify Genes Involved in Resistance

1.4.1 Selecting for resistance in malaria: the forward approach

Elucidating gene function using forward genetics in malaria has been an uphill battle as there have been no practical ways until recently to perform genetic crosses or complementation in *Plasmodium [61]*. Notable successes, which took years to perform, include the identification of the gene involved in chloroquine resistance (chloroquine resistance transporter, *pfcrt*), which was achieved by crossing a drug resistant and

sensitive *P. falciparum* line [62]. Additionally, because malarial parasites need to grow within red blood cells, transfection methods have been difficult and extremely inefficient. Over the last decade, however, there has been significant success in using a whole-cell based chemical genomics approach to identify gene function in the context of identifying antimalarial drug targets [26, 63, 64]. This forward genetics approach using *in vitro* evolution has been well validated in identifying antimalarial gene targets (examples in Table 1.1), and provides a much more efficient technique to elucidate gene function when compared to previous cross studies. Briefly, *Plasmodium falciparum* is cultured in the presence of small molecules with antimalarial activity to encourage the emergence of resistant mutants. Following successful selection of resistant parasites, whole genome sequencing is used to analyze clones that have acquired resistance at a single nucleotide resolution. Performing this selection method on multiple independent cultures provides the identification of the genes involved in drug resistance based on mutations acquired in particular proteins, and in best cases occurs across all selected cultures. In numerous cases, *Plasmodium* has acquired resistance via non-synonymous mutations in a drug-binding pocket, thereby identifying the compounds MOA [65-68]. It should be noted, however, this type of selection method does not guarantee the mutated gene is in fact the

direct target of the compound used. Genes involved in general resistance have the potential to mutate as well, as illustrated by previous studies identifying single point mutations in *pfcrt* conferring resistance to mefloquine and halofantrine [69], or amplifications in a drug efflux pump (*pfmdr*), conferring resistance to chloroquine [70]. Additional experiments can be used to establish if a mutation detected by whole genome sequencing confers resistance. Without expressin the protein it can be challenging to distinguish between targets and resistance genes.

This method addresses questions regarding a compound target and is helpful even on a single compound level. The application of the forward approach on multiple compounds, however, opens up the possibility of analyzing resistance development on a wider scale in addition to identifying potential targets and/or resistance mechanisms. Typically more than one mutation is identified in a single resistant culture, and while initial focus is on gene similarity across a group of resistant strains, there is a potential to apply a broader analysis, identifying patterns in gene pathways or gene ontology information. Therefore this method will also allow for a more extensive understanding of the gene targets/pathways as well as resistance onset.

1.4.2 Loss of virulence in leptospira*: the reverse approach*

While *L. interrogans* is the most extensively studied *Leptospira* species, the molecular mechanisms underlying its pathogenesis remain largely unknown primarily because targeted gene knockouts in pathogenic *Leptospira* is inefficient and technically challenging [45]. Despite this barrier to progress in the field, transposon mutagenesis, first reported by Bourhy and others [71] and Murray and others [72], has been successful. Though technically difficult, targeted gene knockouts have also been described and used to validate a handful of *Leptospira* virulence-related genes (e.g., fliY, colA, mce) [73-75].

Given the difficulty of targeted gene knockouts, systems-based approaches, including transcriptome and comparative genome analysis, have been used to identify potential virulence candidates. Microarrays have been applied to investigate the transcriptional response of pathogenic *Leptospira* to various "host-like" conditions including temperature [76, 77], serum [78], physiological osmolarity [79], iron depletion [80], and host immune cells [81]. Recent RNA-seq experiments have further improved our understanding of global transcriptional responses during *Leptospira* growth *in vivo* [82, 83]. In addition, comparative genome analysis has been applied to identify 452

conserved pathogen-specific genes that likely play a role in *Leptospira* pathogenesis [48, 56, 57, 59, 84, 85].

With the accessibility of high throughput sequencing technology it is now possible to compare *Leptospira* virulent and avirulent strains and gain a larger understanding of virulence factors involved in the organism. By passaging a highly virulent *L. interrogans* strain until virulence is lost and comparing variant frequencies between the parent strain and its attenuated derivative, we are able to identify individual genes or combinations of genes which may be contributing to the overall virulence phenotype of pathogenic *Leptospira*.

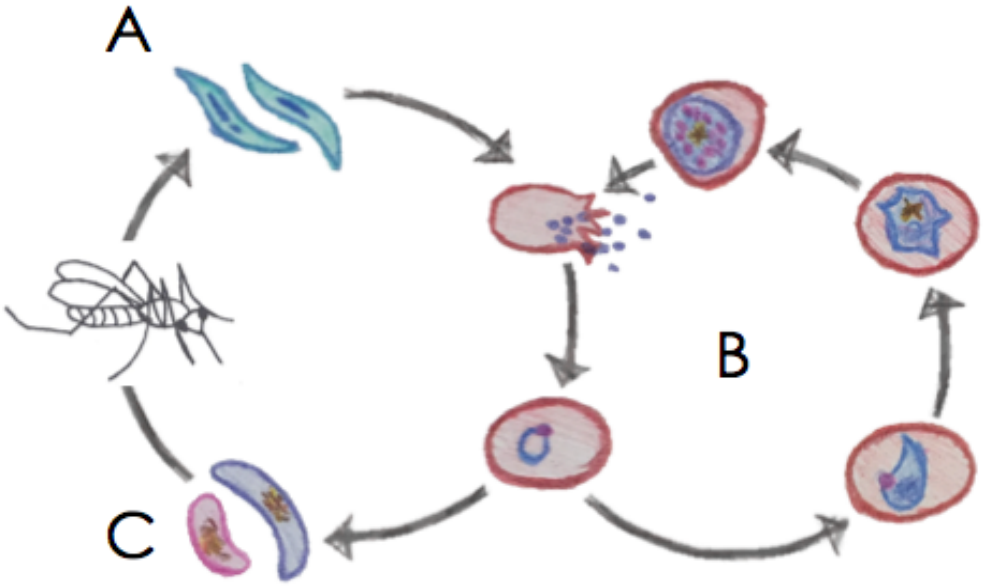


Figure 1.1 *Plasmodium falciparum* life cycle. The *Plasmodium falciparum* life cycle shuffles between a human host and a mosquito vector. *Plasmodium* sporozoites (A) are transferred from the mosquito to the human host, where they migrate to the liver to incubate and multiply in hepatocytes. Following this incubation, the merozoites migrate to the bloodstream where they asexually multiply (B). Finally, a small portion of the asexual blood stage parasites can develop into sexual stage gametocytes (C), which once fully matured can be taken up by the mosquito.

Table 1.1 Examples of target-discovery studies utilizing drug selection methods to elucidate potential drug targets

Drug	Presumed target	Genetic Mutations
Artemisinin	Kelch propeller domain	SNVs in 7 genes, including K13-propeller [32]
Atovaquone	electron transport	SNVs in cytochrome bc 1 [86]
Cladosporin	Lysyl-tRNA synthetase	CNVs on Chrom 13, including cytoplasmic lysyl-tRNA synthetase [87]
Decoquinate	electron transport	SNVs in mitochondrial cytochrome bc1 [88]
Fosmidomycin	inhibits DXR	CNV of region surrounding pfdxr [89]
GNF156/179	unknown	SNVs in cyclic amine resistance locus [90]
KAI407 (Imidazopyrazine)	Phosphatidylinositol 4-kinase	SNVs in *pfpi4k* and CNVs including *pfpi4k [91]*
KAE609 (spiroindolone)	P-type cation transporter (ATPase4)	SNVs in pfatp4 [92]
Piperaquine	digestive vacuole	SNV in pfcrt, CNV of multiple genes around pfmdr1 [93]
Thiaisoleucine	isoleucine analog	SNV in cytoplasmic isoleucyl tRNA synthase (IRS) [94]

Chapter 2

A Broad Analysis of Resistance Development in the Malaria Parasite

2.1 Introduction

With the need for new antimalarials, leads from phenotypic screens are being progressed into molecules that are suitable for testing in clinical trials. An open question, however, is whether small molecules from phenotypic screens will lead to the identification of new druggable targets and pathways in the parasite that do not rapidly lose effectiveness in the field because of acquired and pre-existing parasite resistance. Here we use a set of 50 antimalarial compounds identified in phenotypic screens [39, 40, 95, 96] to systematically evaluate whether resistant parasites can be selected and whether or not pre-existing resistance mechanisms confer resistance using a panel of strains containing mutations in a variety of genes, including *Plasmodium falciparum* cyclic amine resistance locus (*pfcarl)* [97], cytochrome *bc1* [98-100], and *Plasmodium falciparum* ATPase 4 (*pfatp4*) [92, 101, 102]. Our data show that pre-existing resistance is less likely to be a problem but that the *de novo* acquisition of resistance occurs rapidly for many compounds. Analyzing resistance development using whole genome sequencing methods, we identify the potential target for 19 compounds and provide a list of genes that can be exploited in therapeutic design. Finally, we highlight a set of antimalarial compounds that have thus far defied attempts to create drug-resistant parasites in a variety of different

laboratories and identify features that are shared by all, including a rapid rate of killing and lack of pre-existing resistance.

2.2 Results

2.2.1 Initial selection of compounds

To systematically investigate both the *de novo* acquisition and pre-existing landscape of drug resistance, we assembled a set of 50 diverse compounds selected from *P. falciparum* asexual phenotypic screens[39, 40, 95, 96]. Compounds were initially chosen based on potency, demonstrated by activity against asexual blood stages ranging from 23nM to 1.67μM EC_{50} with most compounds having an EC_{50} of <1μM in the *P. falciparum* strain 3D7 as measured by a hypoxanthine incorporation assay. To minimize non-novel pharmacophores, compounds were compared against the scaffolds of clinical antimalarials, eliminating candidates with similar structures.

Compounds in our set ranged in molecular weight from 261-574g•mol^{-1}, with 42 compounds having drug-like properties (compliant with Lipinski's rule of five) and the remaining 8 identified as probe-like compounds. We sought to maximize chemical diversity in our set by eliminating similar compounds as indicated by the Tanimoto coefficient (TC), since compounds with Tanimoto coefficients >0.85, a quantitative

measurement of chemical scaffold similarity, are thought to have similar biological activity to one another[103, 104]. The resulting compound set displayed an average TC of 0.186, ranging from 0.093-0.923 (Fig. 2.1). Although a few compounds were similar to one another (particularly two carbazoles: MMV009063 and MMV665882) a majority of the set was diverse possessing a variety of functional groups and heterocyclic substructures.

2.2.2 Compound evaluation against a panel of resistant clones

Multidrug resistance alleles, including mutations in *P. falciparum* chloroquine resistance transporter (*pfcrt)* and *P. falciparum* multi-drug resistance transporter 1 (*pfmdr1*) are very common in field isolates. Therefore, we sought to assess the degree to which pre-existing resistance alleles would contribute to a loss of potency for this diverse collection of small molecule compounds. Fifteen clones derived from three main parent lines were chosen to maximize diversity in mutated pathways and genetic backgrounds, thus representing the variety of resistance seen in the field: 3D7[105], W2[106] or Dd2[107]. The 3D7 line originates from the clone of a Netherlands clinical isolate strain[105, 108] and is generally considered to be drug-sensitive, though it does convey resistance to sulfadoxine. Conversely, W2[106, 109] and Dd2[107] are multi-drug resistant lines originating from the Indochina III/CDC isolate, which contain

point mutations in *pfcrt* as well as amplifications in *pfmdr1* and GTP cyclohydrolase[109].

A total of nine known drug-resistance genes were represented by our strain set, containing validated critical single nucleotide variants (SNVs) or copy number variants (CNVs) as well as a handful of additional background mutations. One strain, TM90C2A, was a clinical isolate from Thailand[110], while the remaining 15 lines were created through *in vitro* evolution. Clones contained one or more resistance-conferring alleles in the folate pathway[110], *pfcrt*[111], *P. falciparum* dihydroorotate dehydrogenase (*pfdhodh*)[112], *pfcarl*[97], prolyl t-RNA synthetase[113], or heat shock protein 90 (*hsp90*). We also investigated multiple cytochrome bc_1[98-100] and *pfatp4*[92, 101, 102] alleles, including three alleles in the cytochrome bc_1 Q_o site, which confer resistance to atovaquone[98] or a tetracyclic benzothiazepine[99] and one allele in the cytochrome bc_1 Q_i site conferring resistance to a benzylsulfonamide[100]. The five PfATP4 mutant lines possess alleles that map to the transmembrane channel of the sodium-dependent ATP4 transporter, and were acquired by exposing parasites to sub-lethal concentrations of a spiroindolone[92], an aminopyrazole[101], or other scaffolds from the TCAMS library[102]. These two resistance genes were represented by multiple strains since critical SNV mutations may be

located within an active site or result in a change in membrane potential, resulting in each mutation only affecting a subset of PfATP4 or cytochrome *bc1* inhibitors.

To evaluate potential overlapping activity against known antimalarial drug targets, we performed dose-response studies with each compound against our assembled set of drug-resistant lines and their corresponding parents, in an asexual blood stage proliferation assay (Table 2.1). Comparisons were made to the drug-sensitive strain 3D7 when a parent was unavailable (TM90C2A, W2, and Dd2). As laboratories conducted EC_{50} assays with different strains and protocols, all comparisons were made between assays from the same laboratory and under the same assay conditions. Compounds displaying >5-fold EC_{50} shifts relative to the parent or sensitive strain were flagged as having potentially non-novel targets. Average fold shifts per strain ranged from 0.82-2.34, with a median of 1.23 indicating there was not common resistance to any one resistant clone. No resistance patterns were observed when classifying the compounds based on chemical structure, but this was not surprising given the diversity of compounds chosen. On the other hand, given that these confirmed compounds were derived from larger unconfirmed hit lists to which some filtration criteria may have

been applied (e.g. elimination of obvious dihydrofolate reductase inhibitors[38]), some pre-existing bias may exist in the set.

Alleles in the strain set generally did not confer resistance to the vast majority of the compounds (Fig. 2.2a), with only two compounds (MMV019066 and MMV008149) losing efficacy in some resistant lines relative to their parent clones (EC_{50} fold-shifts >5-fold). ATQ-R4, bearing a Q_o cytochrome *bc1* allele, was resistant to a propanamide, MMV019066, (p-value = 0.0016) with an EC_{50} fold-shift of 16x. In addition, CYTb-G131S, a cytochrome *bc1* Q_o allele, demonstrated complete resistance to MMV008149, a carboxamide, with a > 38-fold change in EC_{50}. Statistically significant resistance was seen against MMV019066 and MMV008149 in other cytochrome *bc1* Q_o mutants (ATQ-R5 and ATQ-R4, respectively), but the EC_{50} fold-shifts seen were under the threshold value set for the study. Neither compound shared any structural features with other cytochrome *bc1* inhibitors, including atovaquone and decoquinate (Fig. 2.2b).

As had been previously observed[101], mutations in PfATP4 resulted in parasites that were more sensitive to unrelated compounds: ATP4-Mut2 exhibited sensitivity to a sulfonamide (MMV009108 – p-value=0.0001; 23× fold-shift) and a carboxamide (MMV028038 – p-value=0.0007; 5× fold-shift). ATP4-Mut3 additionally conferred sensitivity to MMV009108 (p-value=0.0047; 5.2× fold-shift). Neither compound was structurally similar to

spiroindolones, pyrazoles, or dihydroisoquinolones, all previously identified PfATP4 inhibitors (NITD609[92], KAE678[92], GNF4492[101], and (+)-SJ733[102]), with a TC range of 0.127-0.34 (Fig. 2.2C).

Finally, two compounds (MMV665939 and MMV028895) lost efficacy in the parent strains W2 or Dd2 when compared to the 3D7 sensitive strain within the same lab. MMV665939 demonstrated fold shifts in W2 compared to 3D7 (EC_{50} fold-shift – 9.5×), as well as in the other two W2-based clones (PfATP4-Mut1: 16.6×, PfATP4-Mut2: 18.9×) compared to 3D7. Fold shifts in Dd2 compared to 3D7 were also seen in MMV665939, though these shifts were less severe (1.5-3.1×; average 2.4×). MMV028895 showed reduced efficacy in Dd2 (EC_{50} fold-shift 7.1×), which was also seen throughout the other Dd2 clones (2.0-7.5×; average – 5.5×), but unlike MMV665939 maintained potency against W2 strains. These efficacy changes in Dd2 and W2 are most likely due to *pfmdr*, given that W2 and Dd2 contain additional copies of the multi-drug resistance gene (2 and 3-4 copies, respectively) when compared to 3D7.

2.2.3 Selection of resistant parasites

While we found an overall lack of pre-existing resistance, analyzing the onset of resistance was critical, as compounds resulting in rapid resistance are not ideal for clinical development. We therefore sought to create drug-resistant parasites using a variety of different selection

methods. We implemented either a high-dose method or a ramp-up/pulse method for our *in vitro* selections. Two of the 50 compounds – MMV028895 and MMV665824 – were removed from the selection study during experimentation. MMV665824 was removed due to a significant loss of potency, while MMV028895 exhibited cross-resistance to resistant parasites generated by MMV007564. Out of the remaining 48 compounds, resistant parasites were obtained for 23 (Supplemental Table 2.1). As each selection was performed in triplicate with three independent cultures, we succeeded in generating a total of 66 resistant cultures. These 66 clones will be submitted to the Malaria Research and Reference Reagent Resource Center (MR4) for public availability. Parasites resistant to MMV026596 were not obtained despite 100 days of selection, but the cultures acquired hypersensitivity to mefloquine with a 10-fold reduction in EC_{50} compared to its parental line, 3D7. For the remaining 24 compounds, acquisition of resistant parasites was unsuccessful despite numerous attempts over an extensive period of time.

To determine if the cross-resistance assay was predictive, we sought to determine the target of MMV008149, the compound resulting in the largest EC_{50} fold-shift (>38x) within the cross-resistance assay set by fully sequencing six parasite clones that had acquired resistance to MMV008149. The EC_{50} values for the resistant lines ranged from 1.67μM to

10.06 μM, a 3-21x EC_{50}-fold change when compared to the Dd2 parent (EC_{50} = 485nM).

Genomic DNA (gDNA) was isolated from the Dd2 parent and six clones (two clones per resistant selection flask), and samples were prepared for whole genome sequencing (WGS). Samples were sequenced to >60x coverage using paired end reads and aligned to the 3D7 reference genome and variants were called with HaplotypeCaller (GATK). Comparing the nucleotide variation found in the resistant samples to the Dd2 parent clone, which had been isolated immediately before selections, we identified genomic changes that had presumably occurred during selection. Following this comparison, 19 mutations were identified: eleven SNVs and eight insertion/deletions (INDELs) (Table 2.2, Supplemental Table 2.2). INDEL mutations were comprised of intergenic (3), codon insertion/deletion (3), intronic (1), and frame-shift (1) mutations. The SNV set showed a mixture of intergenic (5), synonymous (3) and non-synonymous (3) mutations. Comparing all six clones, one gene was mutated across all samples: cytochrome *bc1*. Additionally, variant positions correlated well to EC_{50} fold changes, with the lower fold changes (3.4-5.9x) corresponding to the G131S amino acid change, and the higher EC_{50} fold-shifts (18-21x) corresponding to the Y126C and V284L amino acid changes. Two of the mutations, G131S and Y126C, were contained in the

Q_o site, while the third mutation (V284L) was not in either the Q_o or Q_i binding region[111]. G131S was also the major mutation found in CYTb-G131S, the cross-resistant strain predicting cytochrome *bc1* as a potential target for MMV008149. These results support that the cross-resistance assay was able to successfully identify compounds with overlapping targets, and made us more confident that our compound set consisted of novel targets.

2.2.4 Multi-stage activity profiling

As drugs that eliminate multiple stages of the parasite lifecycle will be critical components in global efforts to eliminate malaria[95], we were interested in knowing whether the compounds in our study had additional activities against either the liver and/or transmission stages of the parasite lifecycle. To determine this, we first evaluated the compounds in additional phenotypic assays. The first assay was a *P. berghei* hepatocyte invasion and development assay, which predicts causal prophylactic activity. Hepatocyte toxicity was additionally tested to identify false-positive activity in the liver stage due to host cell toxicity. Assays were executed in duplicate using a 12-point EC_{50} curve at starting concentrations of 5μM or 50μM (Supplemental Table 2.3). Positive and negative controls testing for hepatocyte toxicity and liver stage activity included atovaquone (PfLuc EC_{50}: 0.419 ± 0.044nM; HepG2 EC_{50}: 7720 ±

1370nM) and puromycin (PfLuc EC_{50}: 32.7 ± 29.7nM; HepG2 EC_{50}: 254 ± 36.6nM). Of the 47 compounds examined, fifteen were found to be active against liver-stage parasites (EC_{50} values < 1μM [10-880nM]), nine of which demonstrated a minimal 5-fold difference between parasite activity and hepatocyte toxicity (Fig. 2.3). Interestingly, six compounds were more potent in the liver stage than in the asexual stage. The most significant potency change was seen with MMV019066, a propanamide, which had a reduction in EC_{50} from 1.67μM to 0.31μM and previously demonstrated cross-resistance with cytochrome bc_1 alleles. This was expected given that cytochrome bc_1 inhibitors are highly active against liver stages. MMV024038, a quinoline sulfonamide, also demonstrated a reduction in EC_{50} from 228nM to 10nM. This compound, however, exhibited some hepatocyte toxicity (146nM); thus the indirect effects of the host-cell environment on parasite growth could not be ruled out.

The second assay was a late-stage (stage V) gametocyte survival assay, which predicts transmission-blocking activity. Assays were executed in duplicate using a 12-point EC_{50} curve at starting concentrations of 1.25μM or 12.5μM (Supplemental Table 2.3). Positive and negative control compounds included puromycin (0.61 ± 0.11μM) and atovaquone (>12.50μM), respectively. For the late-stage gametocyte assay twelve compounds resulted in EC_{50} values < 1μM (260-990nM), 50%

of which overlap with the fifteen compounds found to have activity in the liver-stage assay. Of the six compounds with complete multistage activity, 83% (5/6) demonstrated hepatocyte toxicity >10× the liver stage EC_{50}. Unlike the liver stage, none of the compounds had increased potency against sexual stage compared to the asexual parasites, which was expected as previous literature has found that most currently used antimalarials yield higher EC_{50} values against late stage gametocytes when compared to the asexual stage EC_{50} [114]. Five compounds, however, demonstrated late-stage gametocyte EC_{50} values within a 2-fold range of the asexual EC_{50} value.

2.2.5 Rate of killing assays

To identify whether selection success could be predicted by compound speed of action, and to gain a greater insight into the potential mechanisms of action, assays were performed to test the killing rate (Supplemental Table 2.1) [115]. To quantify the killing rate, viability time-course profiles for each compound were compared to antimalarials known to have fast (chloroquine), moderate (pyrimethamine), or slow (atovaquone) rate of action. These rates were then compared to other compound characteristics, including cross-resistance profile, selection success, potency and structural characteristics to identify any possible trends. Overall, the compound set resulted in a fairly even distribution of

fast, moderate, and slow speeds (Fig. 2.4A), implying there was no bias in the phenotypic screens for a particular killing rate. One compound's rate of killing, MMV666080, was unidentifiable due to compound availability. No correlation was seen between compound structure and speed of action. Additionally, there was no correlation found between killing rate and compound potency against 3D7, nor between potency and selection success (Fig. 2.5).

Comparing the selection success rate to compound rate of killing, we saw a significant direct positive correlation ($p < 0.05$) between slow speeds of action and successful selections (Fig. 2.4B-C). Of the twelve selections with compounds demonstrating a slow speed of action, 83% (10/12) were successful. This was in contrast to compounds with fast killing rates of which only 26% (5/19) were successful. Additionally, within each set of successful selections we noticed a trend within the length of time to generate resistant parasites. For the fast-acting compounds, 3/5 successful selections took >125 days, whereas 5/10 resistant parasites were successfully selected for within 50 days for slow-acting compounds. Therefore, slower compounds typically had greater success and required shorter periods of time to develop resistance when compared to fast-acting compounds. This trend remained relatively consistent within individual labs. These results confirm previous findings of faster acting

compounds having a lower propensity for developing *de novo* resistance[116]. Fast-acting compounds are already desirable in the clinic due to a quick clearance of parasites and alleviation of symptoms, as well as their propensity towards slower drug resistance development. The inability to develop resistance *in vitro*, however, even when using slow ramp up selection methods may imply that these benefits come from the target itself. These genes may have minimal mutational flexibility, inhibitors may target several genes, or inhibitors may have human host targets instead of parasite targets, making them ideal for exploitation in antimalarial development.

2.2.6 Whole genome sequencing analysis of the 23 successful selections

Single Nucleotide Polymorphisms

Having identified some physiochemical properties predictive of resistance development, we next sought to define how the parasites gained resistance, using WGS to analyze variants across each set of resistant strains. Implementing the sequencing method conducted for MMV008149 mutants, gDNA was isolated from the compound sensitive parent (3D7 or Dd2, depending on the selection) for each of the 23 resistant strain sets immediately prior to selections. As the selections resulted in polyclonal resistant cultures, clones wereisolated in a 96-well plate with a limiting dilution method, where parasites are diluted down to

one parasite per well and grown up over 2-3 weeks in the absence of compound pressure. Compound pressure was absent to additionally test the stability of the mutations. Following clonal isolation, clones were assayed to determine individual EC_{50} values, and gDNA was obtained from 1-3 clones per independent culture. Clonal samples were necessary for sequencing, as individual parasites in the selections may have gained different mutations. Additionally, *Plasmodium* is a haploid organism, and therefore analyzing a monoclonal sample would allow for a more accurate sequencing analysis, as mixed read variants could be filtered out as sequencing errors.

Having isolated the gDNA from both the parents and the resistant clones, samples were prepared for WGS, and sequenced using the same paired-end read protocol conducted previously (averaging 80x coverage). The sequences were aligned to a 3D7 reference genome (PlasmoDB_v13), resulting in 116 total resistant strains sequenced with an average coverage of 75x. For each compound strain set, variants were called with HaplotypeCaller (GATK), focusing on mutations acquired during selection by using the sensitive parent strain as an allele reference (Supplemental Table 2.4).

Overall, an average of three SNVs were seen per resistant sequenced sample, with 107/116 sequenced samples containing at least

one SNV (Figure 2.6). Of the samples containing SNVs, a majority of samples had at least one instance of a non-synonymous mutation (95/107), with intergenic (47/107) and synonymous (25/107) mutations seen at lower frequencies. Introns (9/107), start lost (4/107) and stop gained (2/107) mutations were even less frequent. There were a few cases where the total SNV count was significantly higher (8 total clones with SNV counts ≥10). Examining the mutations for these 8 resistant clones, however, showed that the increase in mutations was always due to an enrichment of mutations in subtelomeric regions, especially in variant surface antigen genic regions (*var* and *rifin*) known to recombine at higher frequencies [117-119].

A similar average of three INDELs per resistant sequenced sample was observed (Figure 2.6), though a lower proportion of samples contained at least one INDEL (93/116). Unlike the SNV mutations, however, which favored mutating genic regions, intergenic regions were the most frequent within the INDEL set (77/93), even though Plasmodium's genome is 52.6% coding. Following intergenic INDELs, codon deletions were the next frequent (31/93), followed by intronic mutations (28/93), frame-shifts (18/93), codon insertions (14/93), codon change plus codon insertions (11/93) and codon change plus codon deletion (7/93). Average lengths of INDELs also varied between INDEL types. The smallest

INDEL lengths were observed in codon change plus codon insertions (6.25bp), while the longest lengths were observed in the counterpart mutation (codon change plus codon deletions) at an average length of 27bp. Codon deletion and codon insertion INDEL lengths were observed at an average of 17.6bp and 10.9bp, respectively. Non-exomic regions had similar INDEL lengths, with averages of 10.2bp, 8.1bp, and 6.9bp observed for frame-shift, intergenic, and intronic mutations, respectively. Interestingly, a majority of frame-shift (16/18) and intergenic (54/77) INDELs were caused by deletions, while intronic INDELs had an even distribution (12/28 deletions).

Copy Number Variation

In addition to SNVs, we identified possible amplifications and deletions by analyzing average coverage across each gene in the *Plasmodium* genome. Given that *Plasmodium falciparum* is 90-95% AT-rich in intergenic regions [120], only genic coverage was analyzed, as highly AT-rich regions have a larger variability in coverage due to PCR-amplification bias. Alignment confidence is also reduced in intergenic regions due to AT-repeat segments, and therefore analyzing only genic regions would provide the most robust CNV analysis. Average coverage was calculated with GATK's DiagnoseTargets tool, supplying gene lengths and locations as the analyzed intervals. *Var, rifin*, and *stevor* family genes

were removed from the interval list, as these genes are known to have a high variability in copy numbers and misalignment issues [117-119]. Interval coverage was normalized to the average coverage across each chromosome on a per sample basis to minimize variability based on overall sample coverage. To normalize across runs, four 3D7 samples with varying coverage (85-141x coverage) and sequenced on independent runs were used as calibrators by normalizing the coverage across each interval to the average coverage across the calibrator samples.

To test our CNV caller, we examined four Dd2 samples with varying coverage (36-82x) and sequenced on different runs. For intervals to be considered, each 3D7 calibrator had to contain a coverage >0 and the entire gene needed to be coverage by at least one read. Defining amplifications as genes resulting in an average coverage of >2x across all four Dd2 samples, only two major amplification regions were identified (Supplemental Table 2.5). Chromosome 5 had an amplification of 2-3x across 14 genes (PF3D7_0521900-PF3D7_0523200), which agreed with previous reports [89, 121]. Another amplification of 3x on chromosome 12 was observed immediately before GTP cyclohydrolase I (*pfgch1*, PF3D7_1224000) across two genes (PF3D7_1223800-PF3D7_1223900), which was also in accordance to previous studies [89, 121]. Common deletions were also observed across sub-telomeric regions. Overall, the results we

found agreed with previous studies examining Dd2 CNVs, providing confidence in our CNV caller.

With our CNV caller tested, amplifications and deletions were examined for each selection set. While deletions were fairly straightforward and relatively easy to visually verify with IGV, amplifications were more complex, and therefore additional filters were placed to make our CNV caller more robust and minimize false positives. For our 3D7 samples, the drug sensitive strain or "parent" was required to have a CNV within the range of 0.8-1.2 following normalization to the 3D7 normalization set. Copy number levels outside this range suggested the region was hyper-variable, and were therefore deprioritized. For Dd2 samples, which tended to have larger variability, only a lower cut off requiring parent coverage of >0.6 was applied. To be considered an amplification, 3D7 samples needed a copy number of >2 compared to the sensitive strain. Dd2 samples were required to have a copy number of >2.5 or needed a copy number of >2 prior to comparing to the Dd2 parent if the parent sample was within 0.8-1.2 compared to the 3D7 normalization set. Finally, due to the low gene length in P. falciparum (average 2.3kb) and high gene density (1 gene per 4,338bp) we required CNVs to overlap a minimum of two sequential genes in order to have a

"high" confidence grade. Genes failing this requirement are still included in the CNV list, but were graded as "low" (Supplemental Table 2.6).

Overall, there were more amplifications than deletions. Only three samples contained deletions, and none of the samples were within the same selection set. Conversely, fourteen of the selection sets contained at least one amplification, with a total of 17 major amplifications detected across the selection sets. Interestingly, eight out of these seventeen amplifications were found in either chromosome 3 (4 amplifications) or 5 (4 amplifications), while the other nine amplifications were dispersed between chromosomes 2 (1 amplification), 10 (3 amplifications), 12 (3 amplifications), 13 (1 amplification), and 14 (1 amplification). All four of the chromosome 3 amplifications were localized to the same region, generically spanning across 19 genes (Pf3D7_0319000-Pf3D7_0320800), though spans varied between sets. The amplification on chromosome 3 was observed in 1/4 of the Dd2 selection sets, and in 3/19 of the 3D7 sets. Three of the four chromosome 5 amplifications were observed in 3D7 selection sets and were within the same region as the amplification typically seen in Dd2 samples overlapping *pfmdr1*, a common mechanism of resistance. The fourth chromosome 5 amplification was not found in this region but further downstream, and was observed in a Dd2 selection set.

Unlike the amplifications found in chromosome 3 and 5, none of the other amplifications overlapped between selection sets.

2.2.7 Identification of common mutated genes from whole genome sequencing data

Having obtained the sequencing data for each strain set, our first goal was to identify common genes mutated across each compound set, as these genes may play a critical role in resistance (Table 2.3). To do this, SNV, INDEL, and CNV mutations were compared across each resistant strain set, identifying genes mutated, amplified, or deleted at least once in each resistant sample. Out of the 23 compound sets sequenced, fifteen displayed common gene(s) identified primarily by SNV or INDEL mutations (Table 2.3). First, ten selections contained a single gene mutated across all selection samples. While some CNVs were found in a few of the ten selection sets, there were no common CNVs across samples, nor CNVs containing genes mutated via SNVs.

The other five selections displayed two genes mutated either across all samples or between samples, and were also primarily identified with SNVs or INDELs. Resistant clones from the MMV019719 selection resulted in a common SNV mutation in two separate genes: PF3D7_0404600 (conserved protein, unknown function) and PF3D7_1238800 (Acyl-CoA synthetase 11). Resistant cultures for MMV665924, MMV028038,

MMV011895, and MMV024114, on the other hand, did not have any genes commonly mutated across all samples, but instead displayed SNV-generated resistance in one of two genes. MMV665924 displayed mutations in two different acyl-CoA synthetases (ACS10 and ACS11), suggesting a common target of ACSs. MMV028038-resistant strains contained variants in an RND transporter (PF3D7_0107500) and in a putative nucleolar pre-ribosomal assembly protein (PF3D7_1028300), however the mutation pattern followed a distinct EC_{50} resistance pattern. RND transporter mutants displayed EC_{50} fold-shifts of 15-30x, whereas the nucleolar pre-ribosomal assembly protein mutants had significantly lower EC_{50} fold shifts of 3-5x, suggesting two major mechanisms of resistance. One other selection set (MMV011895) displayed SNV mutations in either *pfcrt* (PF3D7_0709000) or a putative amino acid transporter (PF3D7_0629500), however unlike MMV665924 and MMV028038, there was no discernable pattern in EC_{50} or other parasite attributes to justify multiple targets. It may be possible that both genes when mutated result in the same level of resistance, implying one gene if not both are in fact involved in general mechanisms of resistance. Finally, samples in the MMV024114 set contained either SNVs in *pfcrt* or frame-shifts in an aminopeptidase (PF3D7_1454400). The variant pattern followed discrepancies in the EC_{50} curve, suggesting both mutations resulted in

separate instances of resistance. Unlike the other three compounds, MMV024114 did contain some CNVs: primarily a 2-3x amplification over nine genes in chromosome 3 (PF3D7_0319300-PF3D7_0320100). This CNV, however, was only observed in clones from one of the three independent selections, and did not include any mutated genes from other MMV024114 selections.

For the remaining eight compound sets, CNVs played a more major role. Four of the selection sets (MMV009063, MMV665789, MMV007224, and MMV665882) displayed a common CNV across all samples. MMV009063 and MMV665789 both contained a 3x amplification spanning 8 genes on chromosome 5 (PF3D7_0522700 - PF3D7_0523400). Some of the CNVs found in MMV009063 only spanned 6 genes, starting at PF3D7_0522900. This amplification, commonly seen in Dd2, contains a well-known multi-drug resistance gene pfmdr1 (PF3D7_0523000). Interestingly, MMV665789 also contained a single SNV in pfmdr2 (PF3D7_1447900) shared across all samples. MMV665882 also contained an 1.5-2x amplification on chromosome 5, but this amplification started two genes downstream of pfmdr1, and was a bisected CNV, only amplifying two sets of two genes (PF3D7_0523200 - PF3D7_0523300; PF3D7_0523800 - PF3D7_0523900). MMV007224 contained a amplification spanning across an aminophospholipid-transporting P-type ATPase

(PF3D7_1219600), ranging from 2-6x. For two of the three samples, the CNV continued across two additional genes: a raf kinase inhibitor (PF3D7_1219700) and a conserved protein with unknown function (PF3D7_1219800), however the ATPase2 was thought to be the primarily cause of common resistance, since it was the only commonly amplified gene across the selection set. MMV023367 was also affected by CNVs, but unlike the first four compound sets, MMV023367 had one of two amplifications between samples. Four of the five selection samples contained 2x amplifications on chromosomes 10 (PF3D7_1007600-PF3D7_1007800) and 12 (PF3D7_1208400-PF3D7_1208500), spanning three and two genes, respectively. The fifth sample, on the other hand, contained a 3x amplification on chromosome 3, spanning 8 genes (PF3D7_0319200-PF3D7_0319900).

The final three selections (MMV665939, MMV673482, and MMV668399) had a common mutated gene, but these genes were affected by both SNVs and CNVs within each selection set. MMV665939 contained observed SNV mutations in 3/6 resistant strain samples in an ABC transporter (PF3D7_0319700; Y2079C and R2180P) and a 2x amplification was observed in the same gene for the remaining 3 resistant strains, implying a single target. All four MMV673482-resistant samples contained SNVs in phosphatidylinositol 4-kinase (PI4K – PF3D7_0509800),

but interestingly one sample contained an additional 3x amplification spanning PI4K and three other genes on chromosome 5 (PF3D7_0509800-PF3D7_0510100). Finally, only one resistant sample in the MMV668399 set had a deletion in a putative forkhead-associated (FHA) domain protein (PF3D7_0909700), however the remaining five samples displayed SNV mutations in the same gene (M1V and L95R), suggesting a common mutated gene.

2.2.8 Gene annotation enrichment across selections

Examining common mutations within each selection set aided in determining the primary variants resulting in resistance and identifying potential compound targets. Most resistant cultures, however, displayed more than one variant, some of which may be involved in general resistance mechanisms. Additionally, due to *Plasmodium* having a higher proportion of uncharacterized genes than many other sequenced organisms, some of the genes identified as possible targets have yet to be characterized but may be involved in essential pathways. We therefore sought to identify enriched pathways and processes involved in conferring resistance by broadening our examination across selection sets.

Combining the SNVs and INDELs across all 23 selections and removing any duplicate variants within the same selection set, we observed a total of 266 SNVs and 188 INDELs. Focusing first on SNVs, there

was a fairly even distribution of variants across the genome (Fig 2.7). There were a few instances in which non-synonymous mutations were clustered. However, this was due to different variants mapping to the same target, mostly originating from the same selection set. Surprisingly, there was a fairly even number of non-synonymous and intergenic mutations (102 and 111, respectively), despite 89% of samples displaying at least one non-synonymous mutation while only 44% displayed at least one intergenic mutations. This was largely due to multiple instances where the same non-synonymous mutation was identified across an entire selection set. In contrast, intergenic mutations were typically found in one sample. It was also observed that resistant strains displaying intergenic mutations were more likely to have more than one (57%) than samples displaying non-synonymous mutations (40%), but the different in frequency was not significant. The genome distribution continued to be fairly even when examining INDELs (Fig. 2.8). Despite a reduction in total INDELs compared to SNVs, INDELs appeared more widely distributed across the genome. Unlike SNVs, which tended to cluster around a given region within a selection set, there was a larger variability in the location of INDELs even in individual sets. Given that the *Plasmodium* genome is 81% AT-rich, it may be that some observed INDELs were a result of replication error or PCR amplification error as opposed to having an impact on

resistance. While this does not rule out INDEL mutations as having a potential effect on resistance, this resulted in deprioritizing INDELs in subsequent analyses, especially intergenic mutations where AT-richness reaches 90-95% and AT-repeats commonly occur [122].

Observing only a minor overlap of similar genes mutated across the selection sets, we next sought to identify enriched pathways and processes. Genes were assigned to every mutation, generating a list of genes corresponding to either SNV or INDEL mutations. Mutations located in genic regions (non-synonymous, synonymous, intronic, stop gained, start lost, codon changes and/or insertions/deletions, and frame-shifts) were assigned to the gene where the mutation was present, resulting in 74 and 77 unique genes for SNV and INDEL mutations, respectively. Intergenic mutations were assigned to the first downstream gene if the gene was within 1kb, as regulatory elements are known to have this bias in the *Plasmodium* genome [123]. Any intergenic mutations out of this range were removed from the analysis, resulting in 81 SNV mutations and 49 INDEL mutations not being assigned to any gene. A majority of the removed SNV mutations (67/81) were located in telomeric regions and found either after the last gene or more than 10kb upstream of the first gene in the chromosome, whereas this was only the case for 18/49 INDEL intergenic mutations. This resulted in 185 and 138 genes associated to

SNV and INDEL mutations, respectively. Pruning the list to only include unique genes lowered the gene count to 97 and 129 for SNV and INDEL mutations, respectively, further suggesting the INDEL mutations were more random.

To identify pathway enrichment, each gene list was loaded into PlasmoDB, an online genomic informational database specific to the *Plasmodium* genus [124]. Although many genes in the *Plasmodium* genome are uncharacterized, most contain Gene Ontology (GO) annotations, providing a more detailed annotation to genes identified as putative or having an unknown function. We therefore performed a Gene Ontology Enrichment analysis on each gene set (SNV only, INDEL only, or SNV and INDEL genes), analyzing enrichment in biological processes, molecular function, and cellular components and using a standard p-value cutoff of 0.05 (Supplemental Table 2.7). SNV and INDEL gene lists were originally separated, as INDELs are more likely to affect regulatory processes, given their primary presence in intergenic regions. However, we additionally ran the combined set to see which terms were still significantly enriched when the datasets were combined. The number of enriched GO terms was similar between the SNV group and INDEL group for molecular function (19 and 18, respectively) and cellular components (11 and 13, respectively). The two groups, however, did

differ in number of enriched terms in biological processes, with SNV-related genes displaying 60 enriched terms and INDEL-related genes displaying only 37. Combining the two gene sets and examining enrichment resulted in the enrichment of 49 biological process-, 7 molecular function-, and 15 cellular component-related GO terms.

To summarize and visualize the enrichment data, we submitted the GO term lists with their corresponding Benjamini-Hochberg corrected p-values to REVIGO, an online tool that removes redundant GO terms and visualizes GO similarity [125]. A number of pathways and process were identified as being enriched in the dataset (Fig. 2.9). GO terms involved in biological processes including pathogenesis and response to drug/stimulus were observed in the SNV set, which were expected as we selected resistant parasites with drug-like compounds. Pathogenesis was also observed in the INDEL set, but response to drug/stimulus was not seen, possibly due to INDEL mutations imposing more indirect effects, such as gene regulation. This theory was further supported by the enrichment of biological regulation-related terms within both the INDEL and combined set. Multiple terms related to cell adhesion (biological adhesion, cell adhesion, regulation of cell adhesion, etc.) were also enriched in all three sets.

For molecular function related GO terms, enrichment was observed for receptor activity in all three sets, as well as transmembrane receptor activity (including drug transmembrane transporter activity) in the SNV and combined set. This was not surprising, as efflux pumps are an effective way to quickly remove compound and promote resistance in *Plasmodium*. Cell adhesion molecular binding was additionally enriched in all three sets. Analyzing the gene sets, frequent mutations were observed in the erythrocyte membrane protein (PfEMP) *var* genes, which mediate cell adhesion to endothelial cells and are known to be highly polymorphic [126]. Finally, cellular component enrichment was observed in membrane related terms for all three sets. While this additionally may be related to efflux of compound, it was noted that a number of potential targets identified, including the RND transporter, cytochrome bc1, *PfCRT*, and ABC transporters, are all transmembrane-based proteins, and may have contributed to the enrichment.

2.2.9 Gene expression enrichment analysis within selections

In addition to analyzing the gene ontology enrichment across selections, which primarily focused on general and broad resistance mechanisms, it was important to analyze enrichment between selections to identify possible pathways specifically involved in a given compound's mechanism and interaction with the parasite. It was noted, however, that

many of the uncharacterized genes lacked sufficient gene ontology terms. Therefore, we sought to analyze pathway enrichment by combining gene expression data with gene ontology data. With a "guilt by association" strategy, one can identify uncharacterized genes by linking them with well-characterized genes demonstrating similar gene expression patterns. Previous work in our lab utilized an Ontology-based Pattern Identification (OPI) algorithm to cluster *Plasmodium* genes based on gene expression from 54 various stages in *P. falciparum* and *P. yoelii* [127]. Ruling out genes with limited gene expression data, this resulted in 4,713 genes clustered into 156 statistically enriched gene clusters. By applying these gene clusters to our SNV dataset, we sought to identify any pathway enrichment within selections sets.

In order to identify pathway enrichment, we focused on genes affected causal SNV mutations (non-synonymous, stop-gained, start-lost, intergenic, and intronic). Genes only affected by synonymous mutations were removed, as the mutation should not have resulted in any change to the gene or gene expression itself. Additionally, multi-variable genes (*var, rifin, stevor*) and genes known to be involved in general resistance (*pfmdr, pfcrt*) were removed, as these genes were known to not be involved in the compound's direct mechanism. This resulted in a unique gene list of 71 genes, and encompassed 20 of the 23 selection sets.

Applying the OPI clustering results, 37/71 genes were present in at least one of the 156 clusters (Supplemental Table 2.8). Many genes were present in more than one cluster, with an average of 6 OPI terms per gene (range 1-37). Sixteen of the twenty selections contained at least one gene successfully clustered, and nine of the sixteen contained more than one.

Having identified genes with OPI cluster terms, we next identified if any OPI terms overlapped between genes mutated in the same compound selection set (Supplemental Table 2.9). Of the nine selections clusters with >1 gene clustered, five compound sets displayed an overlap in OPI terms (MMV028038, MMV668399, MMV665939, MMV009108, and MMV024114). Four of the five sets contained minimal overlap, with 1-2 OPI terms overlapping. MMV028038 had one overlap in GO:PM15591202_Trp (Targeting malaria virulence and remodeling proteins to the host erythrocyte) between two non-synonymous mutations in a conserved protein, unknown function (PF3D7_1354900) and the RND transporter (PF3D7_0107500). MMV668399 contained a single overlap as well in GO:0016585 (chromatin remodeling complex) between two non-synonymous mutations between an inorganic anion exchanger (PF3D7_1471200) and a FHA domain protein, putative (PF3D7_0909700). MMV665939 had 2 term overlaps in GO:GNF0206 (cytoadherence) and

GO:0044419 (interaction between organisms) between an intergenic mutation upstream PAM18 (PF3D7_0724400) and a non-synonymous mutation in an ABC transporter (PF3D7_0319700). MMV024114 had one overlap in GO:0006259 (DNA metabolism) between two intergenic mutations upstream of either a putative exonuclease I (PF3D7_0725000) or a conserved protein with unknown function (PF3D7_1412100). The final selection set, MMV009108, had 29 overlaps between eight total genes. While multiple terms came up, one DNA metabolism and regulation of translation occurred multiple times. Additionally, some terms associated with sexual stage development were identified, which was of particular interest as MMV009108 was equally potent in sexual and asexual parasites.

These results were highly encouraging, as these GO terms may provide insight into how the compound is directly or indirectly acting on the parasite. It was additionally encouraging to see some overlap in phenotype and GO term, particularly sexual stage activity and MMV009108's sexual stage specific terms, as this further implies the enrichment observed is more compound specific and less focused on general mechanisms of resistance. Future experimentation is needed to verify compound involvement within these pathways, but the identification supplied by this analysis provides a significant first step in determining these compound's mechanisms of action.

2.3 Discussion and Conclusions

This study represents the first systematic analysis of cross-resistance in malaria parasites. We have assessed 50 antimalarial compounds with diverse chemical structures, rates of parasite killing, and stage specificity. With few exceptions, the compounds studied did not demonstrate significant cross-resistance to previously identified targets, indicating a large potential to identify additional druggable pathways in the parasite and further complement our arsenal of antimalarial therapeutics. A lack of cross-resistance to known drug targets does not mean that resistance will not quickly develop, however, and the resistance "life expectancy" and resulting fitness costs need to be considered for any potential clinical candidate.

Culturing parasites in the presence of compound and using whole genome sequencing, we identified the potential cause for resistance in 21 compound resistant strain sets. While some of the targets have been previously identified in other selection experiments, including PfCRT [128], *cytochrome bc1* [129], PI4K [91], and PfCARL [97]), six new genes were identified as potential compound targets, including phenyl-alanine t-RNA ligase, cathepsin dipeptidyl peptidase, and the RND transporter. While more work needs to be conducted to verify if these genes are involved directly with the compounds or instead provide general mechanisms of

resistance, our list presents a sizeable arsenal of genes that may be exploited in future therapeutic design.

Additional analysis using gene expression clustering information also identified potentially enriched pathways for five of the selection sets. Future experiments focused on these terms may elucidate some of these compound's mechanisms. Measuring the incorporation of labeled amino acids when translating genes, for example, may test regulation of translation in MMV009108. DNA repair and mutation frequencies may be measured in the absence and presence of compound to test DNA metabolism for MMV009108 and MMV024114. Additional experiments are necessarily to sufficiently determine some of the compound mechanisms, but the information provided here highlights important pathways, which should be tested first for compound mechanism against Plasmodium parasites.

Finally, in selecting for resistant mutants, we found that fast-acting compounds are harder to develop resistance against and generally have a longer onset of resistance when compared to slower acting compounds. Given that compound killing rates are thought to be determined primarily by their mode-of-action, this resistance feature may be largely due the target or targets themselves. In addition to killing rate, our compound set provides a list of various physicochemical and

structural features that may additionally be predictive of selection success, and the addition of our set to future screens may provide an eventual predictive model, focusing on compounds less likely to be prone to resistance development (Fig. 2.10, Fig. 2.11). Fast-acting compounds have already been a focus in therapeutic development as they rapidly stop disease progression and avoid severe complications, but the additional benefit of reduced evolution of resistance makes them even more attractive candidates for future antimalarial designs.

2.4 Materials and Methods

2.4.1 Compound origin and computational clustering

Compounds were all publically available and obtained from a variety of sources, including the MMV malaria box[95], the GlaxoSmithKline Tres Cantos Antimalarial Set (TCAMS)[39], the University of Dundee, and the Broad Institute's Diversity Oriented Synthesis libraries. Compound similarity was calculated using the Flexible MCS (FMCS) finder package in R. Clustering was conducted using hclust with a ward.D2 method setting. Pairwise distances were calculated between Tanimoto coefficient values from fmcsR.

2.4.2 Strain culture origins and propagation

Plasmodium falciparum parent strains 3D7 and Dd2 used for selection were obtained from the labs of Dan Goldberg and David Fidock, respectively. Functional assay and cross-resistant strains were obtained from the labs of Elizabeth Winzeler, Dyann Wirth, David Fidock, and GlaxoSmithKline, as listed in Table 2.1.

Asexual parasites were grown with 5% hematocrit at 37°C in RPMI 1640 medium supplemented with 0.014mg•mL^{-1} hypoxanthine, 38.4mM HEPES, 0.2% $NaHCO_3$, 0.2% glucose, 3.4mM NaOH, 0.3g•L^{-1} glutamine, and 0.21% AlbuMAX II. Depending on lab preferences, parasites were grown either in the presence or absence of 0.05mg•mL^{-1} gentamicin and 4.2% human O^+ serum. Cultures were maintained in a gas mixture of 5% O_2, 5% CO_2, and 90% N_2. When not undergoing selections, cultures were maintained with media changes every other day, keeping parasitemia values at 0.3-4%.

Frozen stocks were prepared using one of two methods: (1) by freezing 100% RBCs at approximately 5% parasitemia with equal volumes of a freezing solution composed of 28% glycerol, 3% sorbitol, and 0.65% NaCl; (2) freezing 0.2mL RBCs at approximately 5% parasitemia with 0.3mL serum and 0.5mL of a glycerol solution. Stocks were thawed out by slowly adding 1/5 volume of 12% NaCl, followed by 5 volumes of 1.6% NaCl.

Parasites were then spun down at 800×g for 5min at room temperature and washed with supplemented media prior to standard culturing methods.

2.4.3 Cross-resistance and functional assays

EC_{50} assays were conducted using either a 48 or 72-hour assay based on lab preference and specialty. The 48-hour assay was carried out following standard methods using the ^{3}H-hypoxanthine incorporation assay[130]. Briefly, this assay relies on the parasite incorporation of labeled hypoxanthine that is proportional to *P. falciparum* growth. A culture of parasitized red blood cells (0.5% parasitemia with a percentage of ring stage higher than 70% of total parasitemia, 2.0% hematocrit) in RPMI-1640, 5% AlbuMAX and 5 µM hypoxanthine was exposed to drug serial dilutions. Plates were incubated for 24 hours at 37°C, 5% CO_2, 5% O_2, 90% N_2. After 24 hours of incubation, ^{3}H-hypoxanthine was added and plates were incubated for an additional 24-hour period. After that, parasites were harvested on a glass fiber filter using a TOMTEC Cell Harvester 96. Filters were dried and melt-on scintillator sheets were used to determine the incorporation of ^{3}H-hypoxanthine. Radioactivity was measured using a microbeta counter. Data were normalized using the incorporation of the positive (parasitized red blood cells without drug) and negative (same culture with artesunate at 2 µM) controls. All assays were conducted in

triplicate using three independently grown cultures, and EC_{50} values were determined using the XC50 module from Excel program (version 2.3.1), where parameter *a* is the min value (0%), parameter *b* is the max value (100%), parameter *c* is the curve point midway between the min and max value (50%), and parameter *d* is the curve slope. The XC50 equation is:

$$y = a + [(b - a) / (1 + (10^{x}/10^{c})^{d})]$$

Alternatively, a SYBR Green-based proliferation approach was used for the 72-hour assay, as previously described[38]. Briefly, a culture of parasitized red blood cells (0.3% parasitemia and 4% hematocrit) in screening medium (identical to supplemented medium above except no serum was added) was exposed to serial drug dilutions. Plates were incubated at 37°C and gassed with 93% N2, 4% CO2, and 3% O2 for 72 hours. Following incubation, 10x SYBR Green I (Invitrogen; supplied in 10,000x concentration) in lysis buffer (20nM Tris-HCL, 5mM EDTA, 0.16% Saponin wt per vol, 1.6% Triton X vol per vol) was added to the wells and the plates were incubated overnight at room temperature prior to plate reading. EC_{50} assays were repeated 3 times, taking the average EC_{50} value from each assay set. EC_{50} values for each strain were compared to those of the corresponding parent strain to determine fold-shift changes and identify resistance. A one-way ANOVA followed by a Dunnet's post-test was conducted in GraphPad Prism to identify significant changes

between the parent and resistant strain. For the purposes of this study and to minimize false positive, compounds displaying >5-fold EC_{50} shifts and determined to be significantly different from its corresponding parent by a one-way ANOVA analysis were identified as having potentially non-novel targets. All raw EC_{50} values are listed in Supplemental Table 2.7.

2.4.4 Cross-resistance computational analysis

Computational analysis of cross-resistance was conducted in R. Briefly, EC_{50} fold shift ratios were calculated using log transformed EC_{50} values from the parental and resistant strains. As EC_{50} assays were run across multiple labs, any strain fold-shift calculations were conducted between strains run within the same lab to minimize error due to assay variability. The natural log of the ratios was loaded into R, and a heat-map analysis was executed using heatmap.2 from the gplots package. Compounds were clustered by structural similarity (fmcsR) and strains were clustered by column mean values.

2.4.5 Evolution of compound-resistant lines

Based on the compound speed of action and lab specialty, selections were conducted using a high-pressure intermittent selection method, a step-wise selection method, or a constant selection method. For high-pressure selections, approximately 1-2×10^9 parasites were treated

at a concentration of 3-10× EC_{50} until parasites could not be seen by microscopy (2-10 days). Following treatment, compound pressure was removed and cultures were allowed to recover. Once healthy parasites were seen and parasitemia reached ~2%, compound pressure was reinstated. For step-wise selections, approximately 1×10^8 parasites were treated at a starting concentration resulting in a reduced growth rate of 50%. Cultures were examined daily by microscopy, increasing compound concentration in increments of 5-10% as needed to maintain a 50% reduced growth rate. Selections were carried out until cultures achieved a reproducible EC_{50} fold shift of >3×. Finally, constant selections were conducted in a similar manner to the high-pressure intermittent method with the exception that compound pressure was never removed. Following successful selection, cultures were cloned out using a limiting dilution method as previously described[131]. Selections were terminated after 200 days if resistance could not be obtained.

2.4.6 Library preparation and analysis of sequenced samples

Genomic DNA (gDNA) was obtained from parasites by washing infected RBCs with 0.05% saponin and isolating the gDNA using a DNeasy Blood and Tissue Kit (Qiagen), following the standard protocols. To prep the sequencing libraries, gDNA was tagmented and amplified with the Nextera XT kit (Cat. No FC-131-1024, Illumina) using the standard dual

index protocol, and sequenced on the Illumina HiSeq 2500 with a RapidRun mode, sequencing 100 base pairs deep on either end of the fragments. Following sequencing, reads were aligned to the *P. falciparum* 3D7 reference genome (PlasmoDB v. 13.0), following the Platypus pipeline as previously described (for SNVs and CNVs), with the exception that single nucleotide variants (SNVs) and insertion/deletions (INDELs) were called with GATK's HaplotypeCaller [132]. To identify valid variants, mutations were filtered using general recommendations from GATK (Table 2.4). Following the initial filtration, mutations where read coverage were < 5 and/or where mixed read ratios were > 0.2 (reference/total reads) across all samples were removed.

To calculate CNVs, our analysis primarily focused on genic regions. Using gene start and stop positions as intervals, average coverage across each gene was calculated with GATK's diagnoseTargets. Gene coverage was normalized and calibrated in MatLab. First, the average coverage across each chromosome was calculated. Next, each coverage interval was divided by the average coverage of the chromosome. Using four 3D7 samples, an average coverage for each interval was calculated, and each sample was normalized to these values by dividing each sample interval by the "calibrator" sample interval

value. Finally, to identify copy number changes between the sensitive parent strain and the resistant samples, fold shifts were calculated.

2.4.7 Accession codes

All relevant sequencing samples have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive database using NCBI accession number SRP069308.

2.4.8 Computational enrichment analysis

Gene lists were loaded onto PlasmoDB (release 27) to search for gene enrichment [124]. Gene enrichment analyses were generated for each individual ontology group (biological process, molecular function, and cellular component), utilizing both the InterPro domain database and the Annotation Center (which downloads GO annotations from sequencing centers, including GeneDB and JCVI) as GO association sources. Additionally, a p-value cutoff of 0.05 was required to in order to select enriched annotations. Finally, GO enrichment was visualized using REVIGO, an online tool which summarizes enrichment by removing redundant terms, using the default options (a median allowed similarity and a SimRel semantic similarity measure) and a *Plasmodium falciparum* GO term database [125].

2.4.9 Rate of killing and multi-stage activity assays

Rate of killing was determined for each compound following a previously described methodology, which uses the invasion of fresh erythrocytes as surrogate of parasite viability[115]. Briefly, parasites were treated with compounds for 48 hours. Compound was washed out and fresh-labeled erythrocytes added to the treated parasites. Double stained erythrocytes (RBC label plus parasite DNA label) were quantified and percent of survival determined.

Liver-stage activity was determined by pretreating hepatic human cells (HepG2) for 2 hours with compound in 1536 well plates infected with freshly dissected *P. berghei* sporozoites. After 48 hours of incubation, the viability of *P. berghei* exoerythrocytic forms (EEF) was measured by luminescence reaction light output using BrightGlo (Promega). Varying levels of compound concentration were used (5μM or 50μM) due to the stock concentration supplied.

To test sexual-stage activity, compounds were tested against late stage gametocytes using a MitoTracker fluorescent-based assay[114]. Specifically, synchronized stage V gametocytes were incubated with compound for 72 hours. MitoTracker® Red CMXRos (life technologies) was added to each well (final concentration: 500nM) together with saponin to lyse the red blood cells. Each plate was then imaged with an

Operetta High Content Imaging System (Perkin Elmer) for fluorescence (590-640nm). Varying levels of compound concentration were used (1.25μM or 12.5μM) due to the stock concentration supplied.

The gametocytocidal activity was measured using high content image analysis software (Harmony, Perkin Elmer). The readout was based on number of metabolically active gametocytes per well.

2.4.10 Cheminformatics predictors

We evaluated the association between acquisition of drug resistance and pharmacological, physicochemical and structural characteristics of the 48 compounds for which selection was attempted. Specifically, we considered the following properties: (1) pharmacological properties measured in this study: killing rate, toxicity and EC_{50} at three different stages of parasite development; (2) 51 physicochemical descriptors obtained with QikProp software[133] and 129 descriptors obtained with VolSurf+[134]; (3) 2694 extended-connectivity fingerprints (ECFP) that encode circular substructures with a bond diameter of 10, generated by ChemAxon software (ChemAxon, Kft.)[135]; and (4) 194 hierarchical scaffolds associated with the compounds generated by HierS software[136].

For each feature, we measured statistical significance of its relationship to the selection success, taking into account the nature of the

features. In particular, structural features (i.e. ECFP fingerprints and scaffolds) are binary and indicate the presence or absence of each structural element in a compound. For these reasons, hypergeometric statistical tests were applied to structural features and t-tests were applied to pharmacological and physiochemical features to measure statistical significance.

2.5 Acknowledgements

Chapter 2, in part, has been submitted for publication for the material as it may appear in Nature Communications, 2016. Victoria C. Corey, Amanda K. Lukens, Eva S. Istvan, Marcus C.S. Lee, Virginia Franco, Pamela Magistrado, Olivia Coburn-Flynn, Tomoyo Sakata-Kato, Olivia Fuchs, Nina Gnadig, Greg Goldgof, Maria Linares, Maria G. Gomez-Lorenzo, Maria Jose Lafuente-Monasterio, Sara Prats, Stephan Meister, Olga Tanaseichuk, Melanie Wree, Yingyao Zhou, Paul Willis, Francisco-Javier Gamo, Daniel E. Goldberg, David A. Fidock, Dyann F. Wirth, Elizabeth A. Winzeler, "A Broad Analysis of Resistance Development in the Malaria Parasite". The dissertation author was the primary investigator and author of this paper.

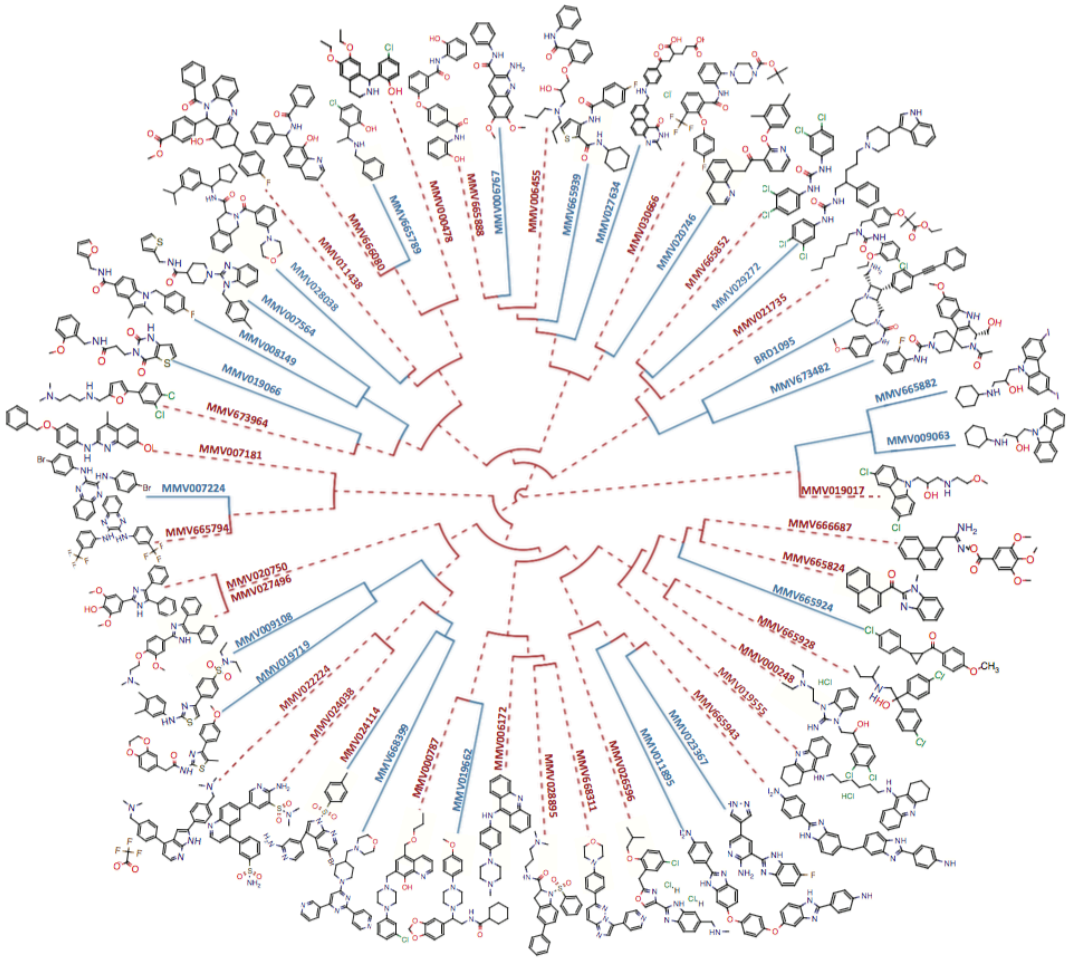


Figure 2.1 Fifty chemically diverse compound set Hierarchical clustering of the 50 compound set. Compounds were clustered by a maximum substructure similarity Tanimoto coefficient. In vitro selections that were successful in yielding resistant parasites are highlighted in blue, whereas compounds where resistance development was unsuccessful are highlighted in dashed red.

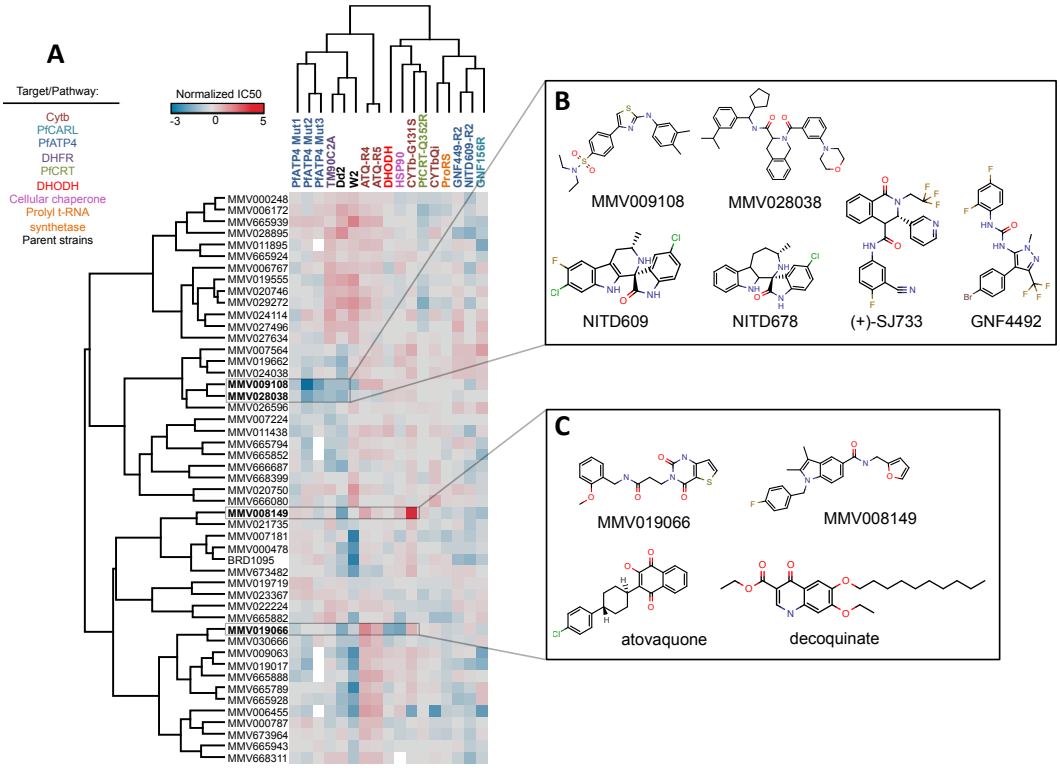


Figure 2.2 A luciferase-based high-throughput screening assay to identify malaria exoerythrocytic-stage inhibitors (A) A total of fifteen resistant strains were tested with each MMV compound to identify potential pre-existing cross-resistance. Calculating the fold shifts between each clone and either a corresponding parent or a drug sensitive 3D7 strain generated the heatmap above. To normalize conferred resistance and sensitivity, the natural log of each fold shift is displayed. Fold shifts instead of raw data were used as multiple assays were run with varying times and detection indicators. Incomplete cross-resistance assays are depicted in white. All assays were run in triplicate. (B) Chemical structures of the two MMV compounds (MMV009108 and MMV028038) with increased efficacy against one or more pfatp4 mutated clones. Both compounds displayed low structural similarity to a number of other known pfatp4 inhibitors. (C) Chemical structures of the two MMV compounds (MMV019066 and MMV008149) with decreased efficacy against one or more cytochrome bc1 mutated clones. Atovaquone and decoquinate, two other cytochrome bc1 inhibitors, were structurally significantly different.

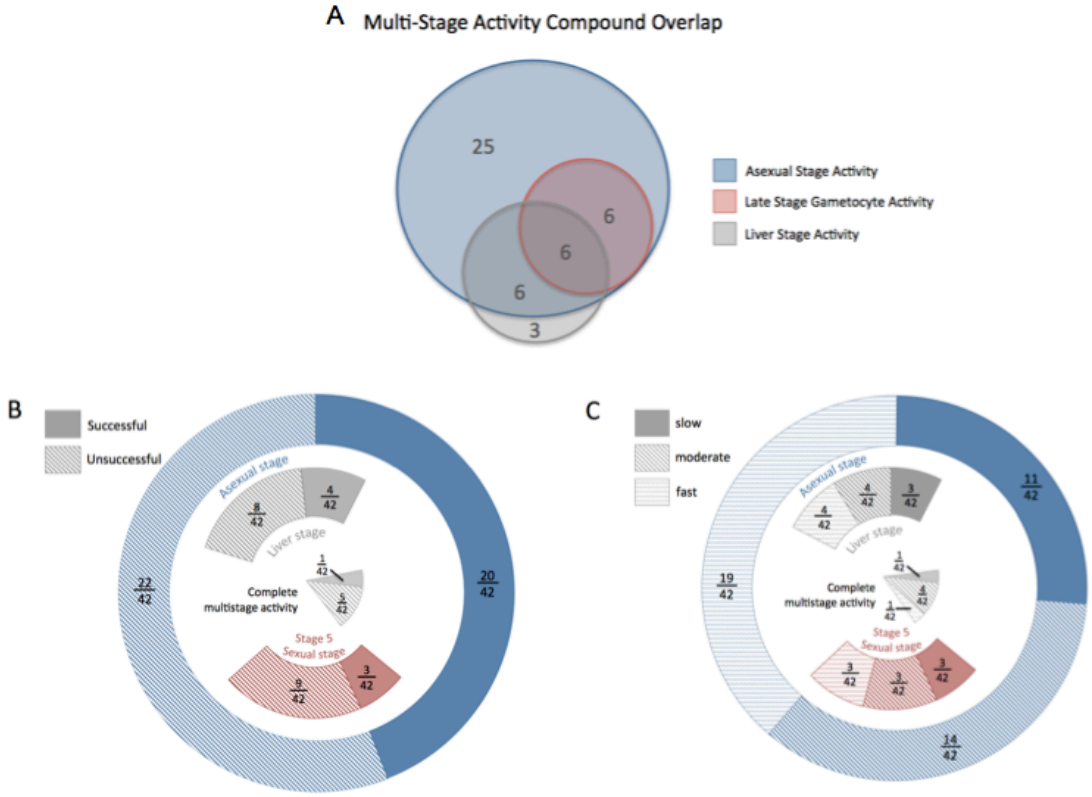


Figure 2.3 Multi-stage activity of compound set (A) Venn diagram summary of multi-stage activity. (B) Breakdown of activity based on selection success. Outer circle represents compounds with asexual activity, first inner circle represents dual stage activity and final pie represents complete multi-stage activity. A potency cut-off of 1μM was required to be considered active against any particular stage. (C) Breakdown of activity based on killing rate. Same pattern as seen in B applies.

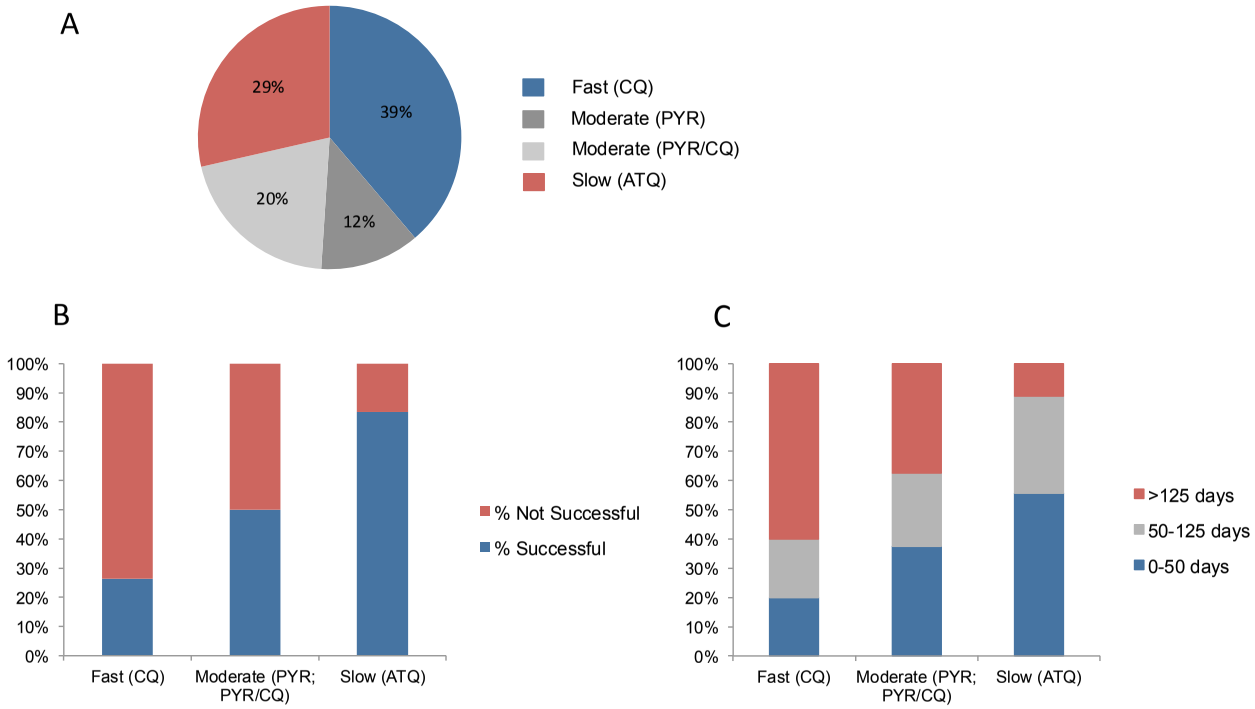


Figure 2.4 Killing rate trends in compound set (A) Proportion of fast (CQ), moderate (PYR or PYR/CQ), and slow (ATQ) compound killing rates in our compound set (49 compounds in analysis). (B) Proportion of failed and successful compound selections sorted by killing rate (47 compounds in analysis). (C) Successful selections further proportioned out based on amount of time required to result in resistant parasites. Trend in selection success and killing rate was found to be statistically significant ($p < 0.05$) by a chi squared test.

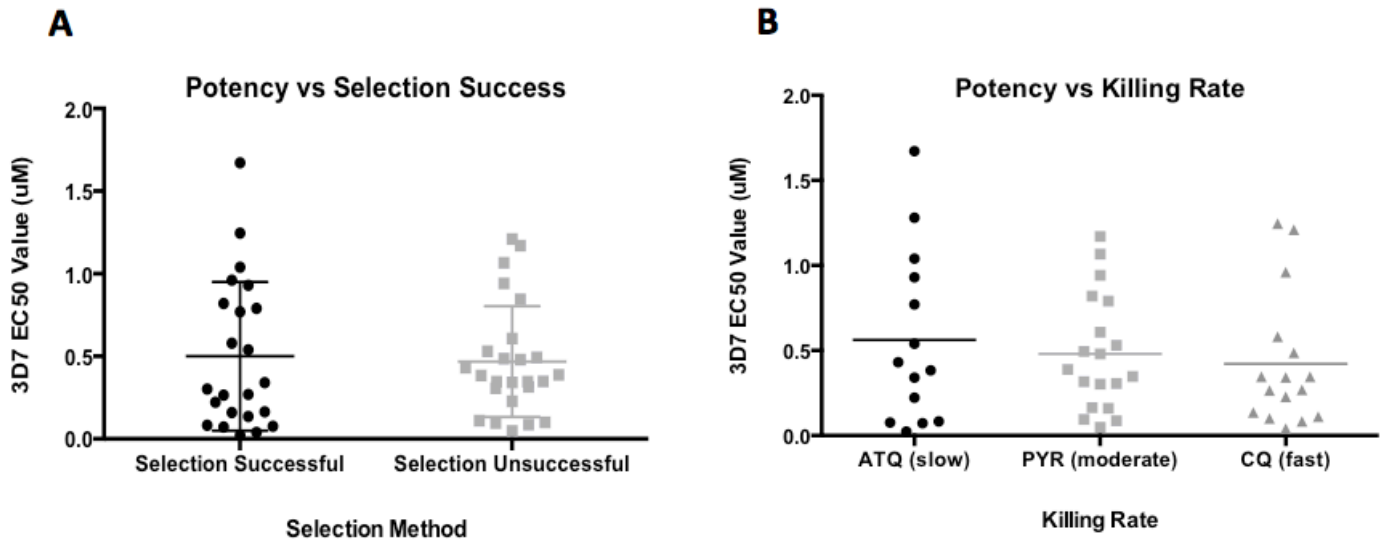


Figure 2.5 Compound potency vs selection success and killing rate Compound potency against a 3D7 parent strain was compared between groups divided by selection success (A) and killing rate (B) to determine if potency could be a predictor for either group.

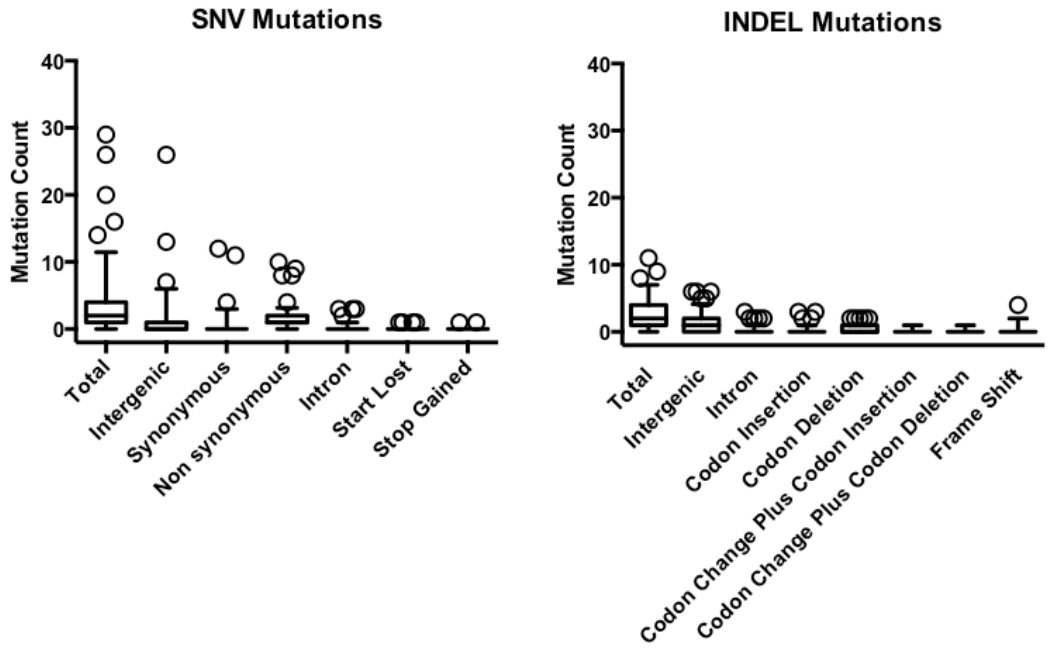


Figure 2.6 SNV and INDEL mutation frequencies across 116 sequenced resistant strains Frequencies were plotted in R for each mutation type. Box and whiskers were displayed for data points within 5-95 percentile.

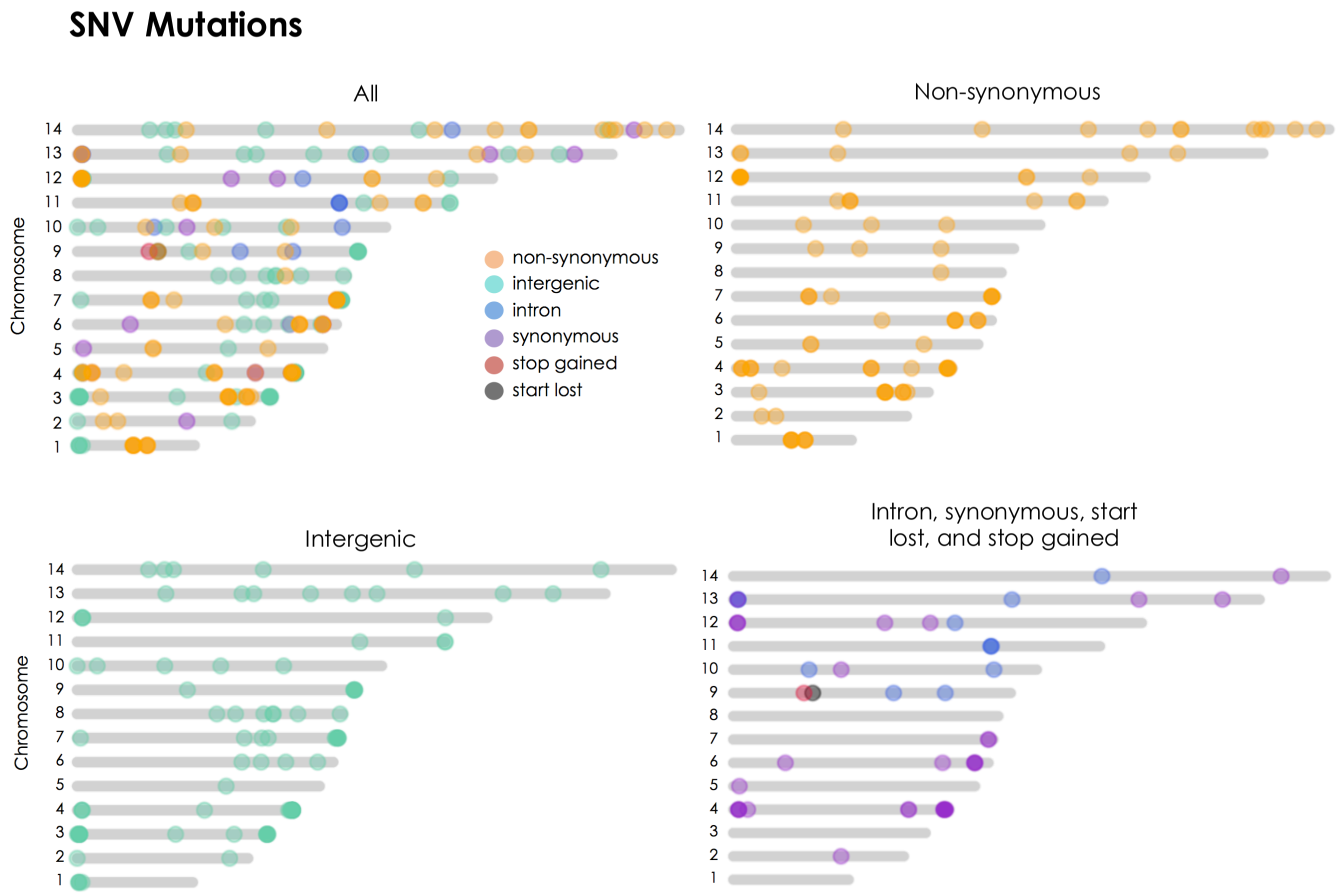


Figure 2.7 Distribution of single nucleotide variants across the *Plasmodium* genome Single nucleotide variants (SNVs) identified in the entire selection set (166 total strains) were graphically depicted on the 14 main *Plasmodium falciparum* chromosomes with an R plotting tool.

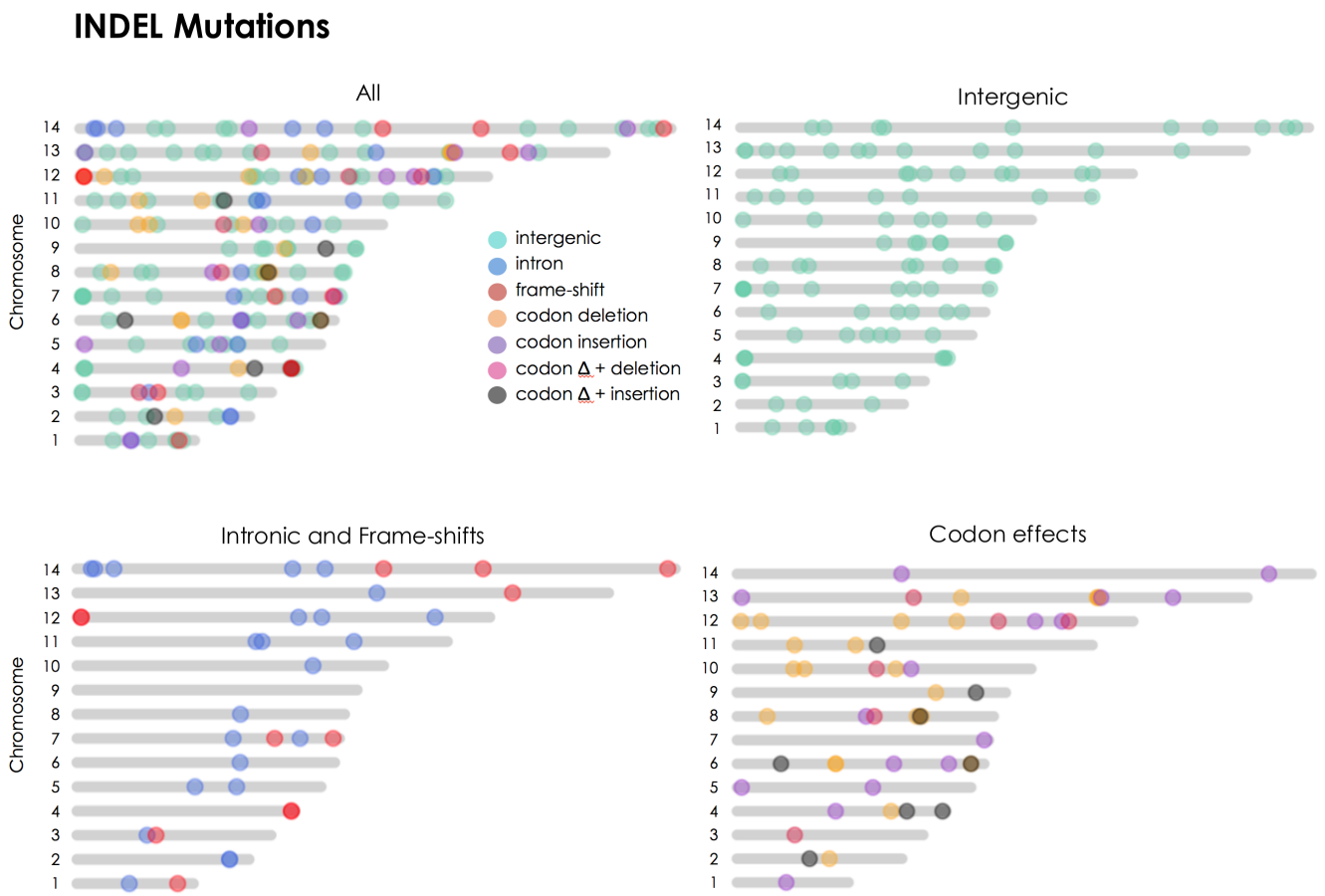


Figure 2.8 Distribution of insertion/deletions across the *Plasmodium* genome Insertion/deletions (INDELs) identified in the entire selection set (166 total strains) were graphically depicted on the 14 main *Plasmodium falciparum* chromosomes with an R plotting tool.

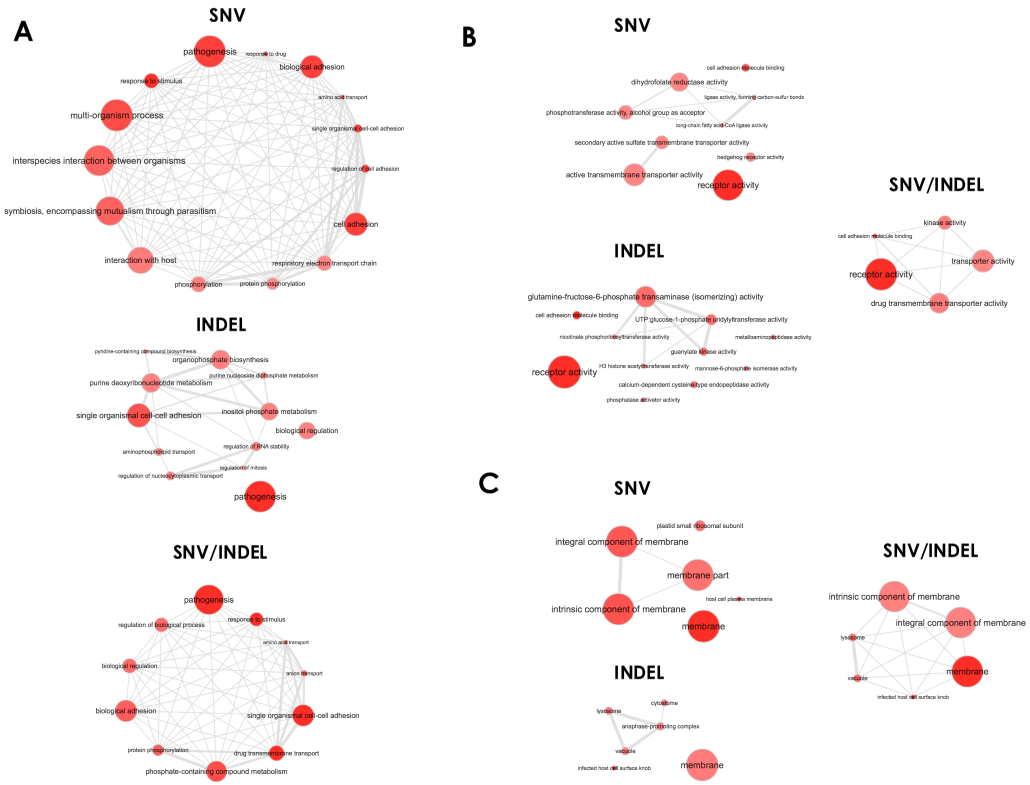


Figure 2.9 Gene ontology enrichment analysis of genes affected by SNV, INDEL, and SNV/INDEL mutations Analysis was performed in REVIGO for GO terms enriched in SNV mutations, INDEL mutations, and a combination of both SNV and INDEL mutations. Data are separated by the networks of non-redundant (A) biological processes, (B) molecular function, and (C) cellular components. P-values are visualized by the intensity of the red shading for each term, and bubble size correlates to the frequency of the GO term in the *Plasmodium falciparum* GO annotation database. Highly similar GO terms are linked via edges, where the degree of line width corresponds to the degree of similarity.

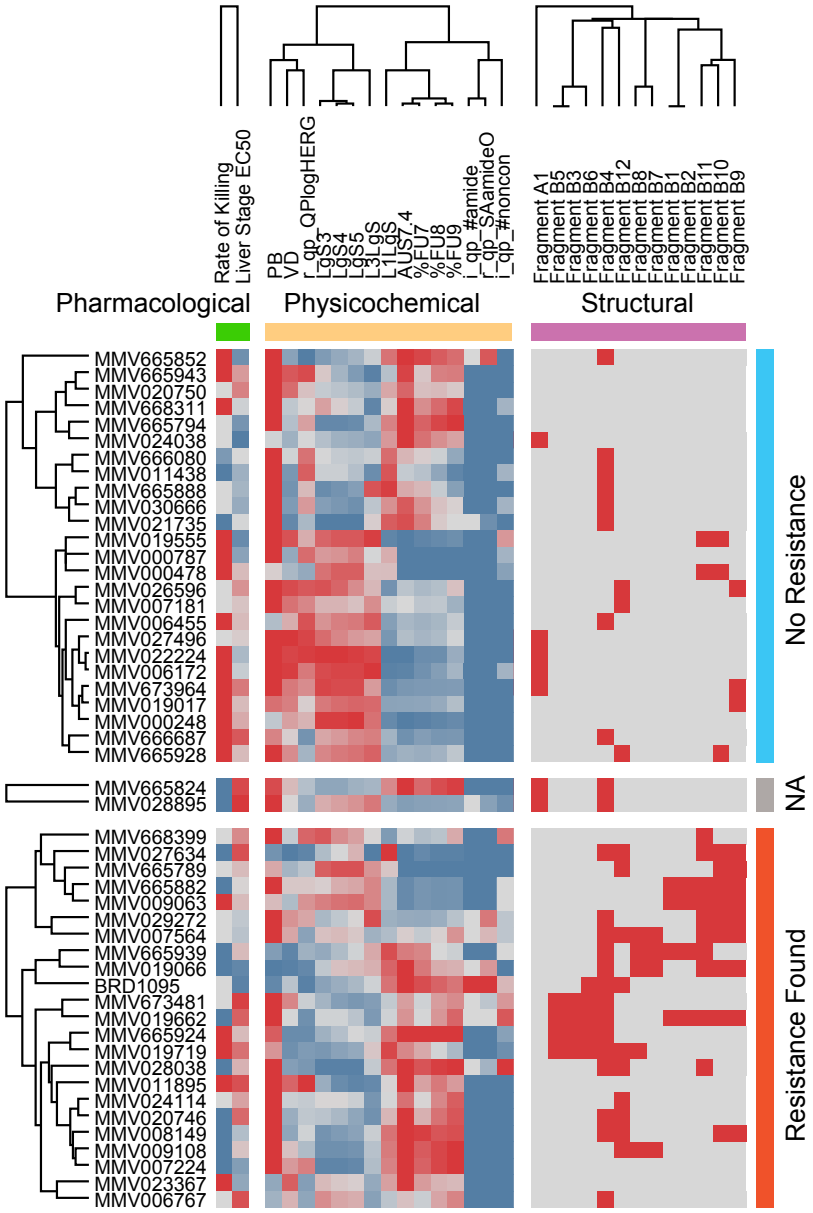


Figure 2.10 Pharmacological and *in silico* features enriched across selection groups The heat map consists of 48 compounds and a total of 30 statistically significant features ($p<0.05$) (Supplementary Table 1A). The 13 physicochemical features and two pharmacological features were first converted into rank values and then scaled into (-1,1)-interval (red = 1, blue = -1) for the heat map coloring. The presence and absence of structural fragments within corresponding compounds was colored in red and gray, respectively. Each of the three feature groups were independently hierarchically clustered (column trees), where Tanimoto similarity was used for structural features and Pearson correlation was used for pharmacological and physiochemical features. Compounds within each of the two groups (successful and unsuccessful selection) were independently hierarchically clustered using Pearson correlation across their feature profiles.

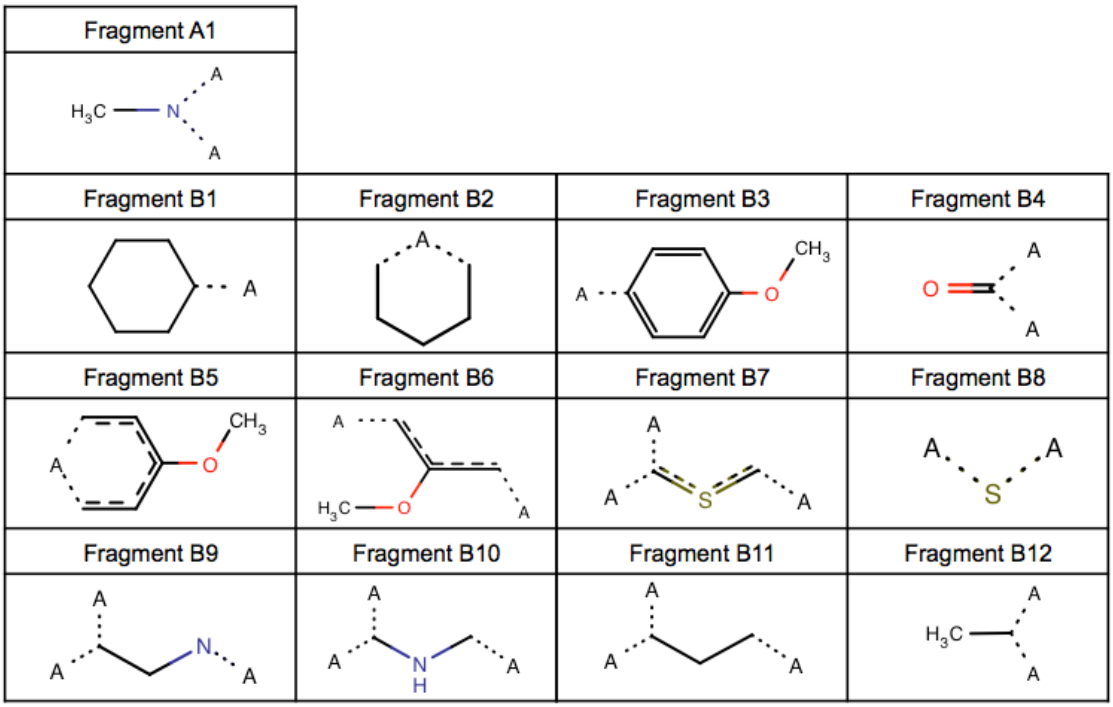


Figure 2.11 Twelve significantly enriched structural fragments Out of 2888 structural fragments, 13 were found to be significantly enriched within one of the two selection groups as per hypergeometric statistical tests ($p<0.05$). Twelve structural fragments (B1-B12) were enriched in compounds with successful selection, whereas 1 fragment (A1) was enriched in compounds where resistance was not obtained.

Table 2.1 Summary of mutated strains testing cross-resistance Fifteen mutated strains and their respective parents were used to test for potential cross-resistance present in our compound panel. Mutations (single nucleotide variants – SNVs) responsible for resistance are listed for each strain, along with their corresponding parents and origin. For TM90C21, there was no official parent strain, so the clone was compared to 3D7 sensitive strain for cross-resistance.

Strain	Target/Pathway	Parent Strain	SNVs	Reference
ATQ-R4	cytb	3D7	M133I, L144S	[98]
ATQ-R5	cytb	3D7	F267V	[98]
CYTb-G131S	cytb	Dd2	G131S	[99]
CYTbQi	cytb	Dd2	G33A	[100]
GNF156R	PfCARL	Dd2	L830V, S1076I, M81I	[97]
NITD609-R2	PfATP4	Dd2	T418N, P990R	[92]
GNF449-R2	PfATP4	Dd2	I203L, P990R	[101]
PfATP4-Mut2	PfATP4	W2	P412L	[102]
PfATP4-Mut1	PfATP4	W2	V178I	not published
PfATP4-Mut3	PfATP4	3D7	F917L	[102]
TM90C2A	DHFR (folate pathway)	3D7*	Unkn (MR4 origin)	[110]
PfCRT-Q352R	PfCRT	Dd2	Q352R	[111]
DHODH	DHODH	3D7	E182D	[112]
hsp90	cellular chaperone	Dd2	D88Y	not published
ProRS	prolyl t-RNA synthetase	Dd2	L482H	[113]

Table 2.2 SNV and INDEL mutations in MMV008149 SNV and INDEL mutations were called and filtered using HaplotypeCaller for six clones isolated from three MMV008149 resistant flasks.

	MMV008149-F1-ClB2	***MMV008149-F1-ClB7***	***MMV008149-F2-ClE2***	***MMV008149-F2-ClE3***	***MMV008149-F3-ClB1***	***MMV008149-F3-ClC10***
genome coverage (x)	84.13	81.56	74.81	90.93	82.73	72.93
% covered by 15 or more reads	95.6	95.7	95.6	95.7	95.8	95
SNVs identified						
total mutations	3	3	5	1	4	4
intergenic	0	2	4	0	1	3
intronic	0	0	0	0	0	0
synonymous	2	0	0	0	2	0
Non-synonymous	1	1	1	1	1	1
genes mutated in all samples (mutation)	*mal_mito_3 (cytochrome bc1)*					
INDELs identified						
total mutations	6	0	1	2	1	1
intergenic	1	0	1	2	0	1
intronic	1	0	0	0	0	0
frame shift	1	0	0	0	1	0
codon insertion/deletion	3	0	0	0	0	0
genes mutated in all samples (mutation)	*none*					

Table 2.3 Potential genes resulting in compound resistance

Compound	Potential Target (Gene ID) - common mutation, if applicable
	One Gene Mutated in All Samples
MMV006767	PfCRT (PF3D7_0709000) - S65R
MMV007564	PfCARL (PF3D7_0321900)
MMV008149	cytochrome bc1 (mal_mito_3)
MMV009108	RND Transporter (PF3D7_0107500) - A1108T
MMV019066	farnesyltransferase beta subunit (PF3D7_1147500)
MMV019662	RND transporter (PF3D7_0107500)
MMV020746	ABC Transporter (PF3D7_0319700)
MMV027634	bifunctional dihydrofolate reductase (PF3D7_0417200)
MMV029272	cathepsin dipeptidyl peptidase DPAP1 (PF3D7_1116700)
BRD1095	phenylalanine-tRNA ligase (PF3D7_0109800)
	Two Genes Mutated in All Samples
MMV019719	conserved protein, unkn func (PF3D7_0404600) - K2184I; Acyl-CoA synthetase ACS11 (PF3D7_1238800) - F387V
	Two Genes Mutated Between All Samples
MMV011895	amino acid transporter (PF3D7_0629500), CRT (PF3D7_0709000)
MMV665924	ACS10 (PF3D7_0525100), ACS11 (PF3D7_1238800)
MMV024114	PfCRT (PF3D7_0709000), aminopeptidase (PF3D7_1454400)
MMV028038	RND transporter; preribosomal assembly protein

Table 2.3 Potential genes resulting in compound resistance, continued

Compound	Potential Target (Gene ID) - common mutation, if applicable
	One Gene Amplified or Deleted Across All Samples
MMV007224	ATPase2 (Pf3D7_1219600) – 2-6x amplification in 3/3
MMV009063	PfMDR1 (PF3D7_0523000) - 3x amplification CNV
MMV665789	PfMDR1 (PF3D7_0523000) - 3x amplification CNV
MMV665882	putative transporter (PF3D7_0523800), conserved protein, unkn func. (PF3D7_0523900) - 2x amplification CNV
	Two Genes Amplified Between All Samples
MMV023367	Chrom3 (803592-833833) – 2-3x in 1/5; chrom10 (299280-311397) and chrom12 (390932-395986) – 2x in 4/5
	One Gene Either Amplified, Deleted or Mutated in All Samples
MMV665939	ABC Transporter (PF3D7_0319700) - SNVs in 3/6, 2X amplification in 3/6
MMV673482	PI4K (PF3D7_0509800) – SNVs in 4/4; 2.5x amplification in same gene in 1/4
MMV668399	FHA domain protein (PF3D7_0909700) - SNVs in 5/6, partial deletion in 1/6

Table 2.4 Applied HaplotypeCaller filters

SNV Filters		*INDEL Filters*	
Filter Name	*Filter Value*	*Filter Name*	*Filter Value*
ReadPosRankSum	> 8.0	*ReadPosRankSum*	< -20
	< -8.0	*QUAL*	< 500
QUAL	< 500	*QD*	< 2
QD	< 2	*DP*	< 7
MQRankSum	< -12.5		
DP	< 7		

Chapter 3

Plasmodium falciparum Cyclic Amine Resistance Locus, PfCARL: a Resistance Mechanism for Two Distinct Compound Classes

3.1 Introduction

One promising new class of multistage malaria drugs is the imidazolopiperazine class, including KAF156, a molecule currently in Phase-IIb clinical trials for both treatment and prevention of human malaria [137]. KAF156 is well tolerated and orally bioavailable [138], and with its potent multi-stage activity it has the potential to be the first new antimalarial drug with not only therapeutic but also prophylactic and transmission-blocking activities. KAF156 could thus provide significant aid to malaria eradication. The exact mechanism by which KAF156 exerts its antimalarial activity is unknown but parasite resistance to KAF156 is associated with the accumulation of mutations in the *Plasmodium falciparum* cyclic amine resistance transporter (*pfcarl*) [90, 97]. *Pfcarl* encodes a conserved protein of unknown function with seven conserved transmembrane domains. Analysis of yeast strains in which the *pfcarl* homolog was deleted [139] suggests that it plays a role in protein folding within the endoplasmic reticulum [140]. In addition, the protein has a conserved domain found in other eukaryotic organisms and mutations of the domain result in homeotic transformations during vertebrate development [141].

Here we present studies showing that resistance to an alternative compound class with a different phenotypic profile is also conferred

through mutations in *pfcarl*. Specifically, we show that treating parasites with sub-lethal concentrations of MMV007564, a benzimidazolyl piperidine identified from phenotypic screens (previously reported EC_{50}=0.5-0.9μM) against asexual stages of *P. falciparum* [95, 142], selects for parasites that have acquired mutations in *pfcarl*. Investigations into the timing of action in the asexual blood stage and potency in different life cycle stages for MMV007564 and KAF156 reveal that PfCARL is not a common target but a common resistance mechanism for these two chemically distinct compound classes.

3.2 Results

3.2.1 In vitro selection generated asexual blood stage P. falciparum lines resistant to MMV007564

The benzimidazolyl piperidine MMV007564 has previously been identified in asexual blood stage screens for activity against the drug-resistant W2 (EC_{50} = 0.3μM) and drug-sensitive 3D7 (EC_{50} = 0.5-0.9μM) strains of *P. falciparum* while having low cytotoxicity [95, 142]. However, liver and gametocyte-stage screens show that this compound has limited activity against non-asexual stages [143-148]. Due to its drug-like properties, asexual blood stage potency in a drug-resistant line, fast/moderate parasite killing rate [115], and novel chemical scaffold, we

sought to gain insight into benzimidazolyl piperidine mechanisms of action and/or resistance development. We generated MMV007564-resistant (MMV007564^R) asexual *P. falciparum* lines using an *in vitro* selection method (Fig. 3.1) that has been previously used to associate compounds with their targets [63]. A clone of the *P. falciparum* reference genome parasite line 3D7 was isolated and then three independent cultures derived from this clone were grown in the presence of compound at concentrations ranging from 3-10xEC_{50} (EC_{50} = 513 ± 56 nM). Parasite sensitivity to MMV007564 was evaluated throughout the 2-month selection period. All three MMV007564^R cultures demonstrated a 4-5 fold shift in EC_{50} compared to the parental 3D7 at generation thirteen after intermittent/continuous compound pressure at 3xEC_{50}, which increased to a 10-20 EC_{50} fold shift at generation 23 after continuous compound pressure was increased to 10xEC_{50}. A single clone for each independent culture was obtained by limiting dilution in the absence of the compound: MMV007564^R-F1-A3, MMV007564^R-F2-E5 and MMV007564^R-F3-E2. Each clone had at least a 10-fold shift in MMV007564 EC_{50} compared to the parent (Fig 3.2). In contrast, there was no shift in EC_{50} for other antimalarials, including quinine (0.86–1.1-fold shift), atovaquone (1.0-fold shift for all three strains), and mefloquine (0.93–1.0-fold shift), indicating that resistance was specifically generated to MMV007564 (Fig 3.2).

Additionally, the MMV007564^{R} clones had stable resistance to MMV007564 for 37 generations in culture without the compound (Fig. 3.1).

3.2.2 Whole genome sequencing identifies pfcarl as major mutated gene

In order to determine the genetic determinant of MMV007564 resistance, resistant parasites and the parental clone were analyzed using whole genome sequencing (WGS). First, clonal samples were isolated at generation 30 in the absence of compound pressure and analyzed at generation 50 (Gen50) to identify common genes mutated across all three independent cultures (Fig. 3.1). After sequencing the clones to >60x coverage using paired end reads, the sequences were aligned to the 3D7 reference genome and variants were called. To identify newly emerged genomic changes, the set of variants identified in each resistant clone was compared to the set of variants identified in the parent clone, which was isolated immediately prior to selections. After this comparison, only eight total mutations distinguished the resistant clones from their isogenic parent and these were evenly split between single nucleotide variants (SNVs) and insertion/deletions (INDELs) (Table 3.1). INDEL mutations were made up of three intergenic and one intronic mutation, whereas SNVs were comprised of one intergenic and three non-synonymous mutations. All three non-synonymous SNVs mapped to a single gene, Pf3D7_0321900, previously named the *Plasmodium falciparum* cyclic amine resistance

locus (*pfcarl*). Each clone encoded a different variant amino acid: L833I (7564-F2-E5), L1073Q (7564-F3-E2), and L1136P (7564-F1-A3) (Table 3.1).

To further investigate the temporal acquisition of resistance as well as to determine if the clones we selected at Gen50 were representative of the primary resistance mechanism of MMV007564, we sequenced two additional samples from un-cloned cultures taken before the sharp increase in EC_{50} (generation 16 - Gen16) and immediately before clonal isolation (generation 28 - Gen28). These bulk population samples were sequenced instead of individual clones, as we sought to not only identify when mutations arose, but additionally wanted to determine if multiple independent *pfcarl* mutations could be observed in one selected population.

Variants were identified using the same filters applied to the clonal set, removing positions that did not correspond to at least one sample with a mixed read ratio < 0.2 (reference allele reads/total reads). This resulted in fifteen variants in Gen16 (11 SNVs and 4 INDELs) and one SNV in Gen28 (Table 3.1, Supplemental Table 3.1). Gen16 was only comprised of intronic and intergenic mutations, however Gen28 contained one non-synonymous *pfcarl* mutation in Flask 3 (L1073Q), the same mutation found at Gen50 in the Flask 3 clone (7564-F3-E2). These results indicate that a

majority of the *pfcarl* mutations did not outcompete the entire population, but existed as sub-populations in each flask.

To further examine *pfcarl* mutations present in only a subset of the polyclonal population, we re-analyzed mutation calls in both generations while removing the read ratio filter applied previously (Table 3.2). This resulted in a total of 243 and 84 variants called in Gen16 and Gen28, respectively. A majority of variants were intergenic or intronic in Gen16, with only 31/243 mutations occurring within an exome region. Conversely, Gen28 had an even distribution of variants, with 49/84 coding region mutations. Three total mixed read SNV calls in *pfcarl* were identified in Gen16 and Gen28 cultures, with two different SNVs found at each time point. Calculating the binomial distribution, the frequency of mixed-read *pfcarl* variants were significant in both generations (Gen16: $p = 0.000021$; Gen28: $p = 0.000053$), suggesting the gene was selected for even in the polygenic populations. Flask 2 contained the SNV L833I at both Gen16 and Gen28 at mixed read ratios of 0.37 (46/124 and 24/64, respectively), which was the primary mutation found in Flask 2 at Gen50. All three flasks contained the SNV F1109L at Gen16 at mixed read ratios of 0.68 (15/22), 0.73 (90/122), and 0.25 (12/48) for Flask1, Flask2, and Flask3, respectively. This mutation was not seen, however, in any of the subsequent generations. Similarly, the SNV Q821H was seen at Gen28 in Flask1 and

Flask2 at heterozygous read ratios of 0.31 (28/88) and 0.62 (44/71), respectively, but this mutation was also not seen in any of the other generations (Fig. 3.1).

Our results demonstrate that parasites with the L833I-encoding variant emerged early on in the selection experiment, but that parasites bearing this mutation did not out-compete other parasites in Flask 2. The variant was only identified as a homogenous call once the culture was cloned. The L1073Q-coding variant likely rose to fixation between generation 16 and 28 in Flask 3, since this mutation was undetected at Gen16, yet had a mixed read ratio of 0.00 in Gen28 (0/32) and Gen50 (0/48). In analyzing the EC_{50} fold shifts across these generations it is most likely this mutation occurred between generations 16 and 22, as Flask 3 had an observable spike in EC_{50} during this time. L1136P was only found at Gen50 in Flask 1, implying the mutation arose between Gen28 and Gen50. The final two mutations, Q821H and F1109L, were not found at Gen50 but may represent mutations that resulted in initial resistance. F1109L was found in all three flasks but only in Gen16, implying the mutation was out-competed between Gen16 and Gen28. Q821H was found in Gen28, but was not observed in the Gen50 clones that were sequenced. These results demonstrate the importance of analyzing mixed cultures as well as multiple clones from each population, as it

provides a greater understanding of how the mutations developed over time and some clonal mutations may not have represented the entire population.

3.2.3 Differential cross-resistance of MMV007564 and KAF156-resistant asexual blood stage P. falciparum lines to imidazolopiperazines and benzimidazolyl piperidines

Previous *in vitro* selection studies in *P. falciparum* using the imidazolopiperazines GNF179 [90], GNF707, GNF452 and KAF156 [97] resulted in resistant lines with 13 different mutations in PfCARL (Fig. 3.3 and Supplemental Table 3.2) – occurring either singly or in combination. Except for the Q821H mutation found in the polyclonal MMV007564^{R} Flask 1 and 2 samples at Gen28, none of the mutations found in the MMV007564^{R} cultures overlapped with the PfCARL variants identified in the imidazolopiperazine-resistant lines [90, 97]. The vast majority of the PfCARL resistance mutations for both the imidazolopiperazines and MMV007564 were present at or near the predicted transmembrane domains of the protein and all PfCARL mutations found in the MMV007564^{R} clones (L833I, L1073Q and L1136P) were within ten bases of suspected resistance-conferring PfCARL alleles (L830V, E834D, M1069I, S1076N/R/I and I1139K) that emerged after imidazolopiperazine exposure (Fig. 3.3, Supplemental Table 3.2). These data suggest a shared structure-

function relationship. Similar to the imidazolopiperazine resistance mutations [90, 97], PfCARL mutations found in MMV007564^{R} lines were in a conserved region of this protein family (DUF747) found across all eukaryotic phyla, suggesting that these resistance loci play an important biological role in PfCARL function.

To test the theory of shared genetic resistance between the two compound classes, we characterized the extent of cross-resistance between MMV007564^{R} and a triple mutant KAF156-resistant line (KAF156^{R}: M81I, L830V, S1076I) using MMV007564, KAF156, and GNF179 (Table 3.3). KAF156 and its close analog GNF179 differ only in the halogen atom attached at the para position in the benzene ring (Fig. 3.4A) [90, 97]. Unlike the MMV007564^{R} lines, the KAF156^{R} line exhibited cross-resistance to all three compounds. Two of the MMV007564^{R} PfCARL mutant lines conferred resistance to either KAF156 (L833I) or GNF179 (L1136P), while the MMV007564^{R} PfCARL L1073Q mutant was not cross-resistant to either of the imidazolopiperazines tested. This suggests that various PfCARL mutations in the MMV007564^{R} lines are compound class specific and have different efficacies in conferring resistance to the imidazolopiperazines tested.

To reveal structure activity relationships (SAR), commercially available MMV007564 analogs were tested against the parental 3D7 cell

line (Fig. 3.4B-C, Table 3.3, Table 3.4). The analogs either had the methyl benzyl (substructure A) or the thiophen (substructure B) of MMV007564 replaced by different chemical groups (Figure 3.4B-C). A methyl benzyl in substructure A with the methyl at either the meta or para positions (MMV007564, **4** and **5**) appeared to be important for activity in 3D7 parasites (Table 3.3, Table 3.4). In general, having an aromatic ring in substructure B was important for activity (MMV007564, **6**, **7**, **8** and **9**) but the heteroatom nitrogen (**7**) or oxygen (**9**) in the aromatic ring was not preferred (Table 3.3, Table 3.4). Next, we tested cross-resistance of PfCARL mutant lines against these analogs. Interestingly, all MMV007564 analogs considered active against parental 3D7 (EC_{50} values < 3.5μM) had significantly reduced activity against both the KAD156^R line and all three MMV007564^R mutants, though at varying levels (Table 3.3). The KAF156^R line was ~370-fold resistant to imidazolopiperazines while the KAF156^R and MMV007564^R lines were 2- to 17-fold resistant to benzimidazolyl piperidines (Table 3.3). We speculate that the high level of imidazolopiperazine resistance of the KAF156^R line is due to the presence of triple mutations in PfCARL. As shown in previous studies, imidazolopiperazine resistant lines containing multiple PfCARL mutations were most likely to have ≥40-fold imidazolopiperazine resistance in contrast to lines with single PfCARL mutations [90, 97]. In the absence of

MMV007564^{R} lines with multiple PfCARL mutations, it is not known whether the same is true for benzimidazolyl piperidine resistance. Structural studies will help elucidate how imidazolopiperazine and benzimidazolyl piperidine interact with PfCARL and determine how the various mutations identified in this locus contribute to resistance. Nevertheless, our current data provide novel avenues to pursue SAR.

3.2.4 MMV007564 and imidazolopiperazines have varying potencies against the different P. falciparum life cycle stages

Since PfCARL mutations mediate resistance of asexual blood stage *P. falciparum* to both benzimidazolyl piperidines and imidazolopiperazines, we hypothesized that these two compound classes share a common target. *P. falciparum* has a tightly regulated intraerythrocytic developmental cycle (IDC) in which proteins reach peak expression at varying stages during the cycle. Parasites are therefore most susceptible to a small molecule inhibitor during the stage of the IDC in which the target of the molecule is expressed and generally unaffected during those stages when the target is not expressed. We therefore sought to determine whether asexual blood stage parasites were most susceptible to both benzimidazolyl piperidines and imidazolopiperazines at the same life-cycle stage, suggestive of a shared target (Fig. 3.5, Fig. 3.6). To test this, synchronized parasites were incubated with either

MMV007564 or GNF179 at varying times throughout one complete asexual IDC (0-46hr, 0-12hr, 12-24hr, 24-36hr, or 36-46hr) and parasite ring burden was calculated at the 46th hour.

Neither culture treated with compound from ring stage for 46 hours advanced past the trophozoite stage. In addition, there were no obvious morphological differences between the compound treated parasites. However, there was a striking difference in the effect of MMV007564 and GNF179 on asexual blood stage parasite development at 0-12hr, 12-24hr and 24-36hr treatments. At 0-12hr treatments, MMV007564-treated parasites doubled the ring parasitemia in the next generation while GNF179-treated parasites remained below the starting ring parasitemia. Interestingly, the 12-24hr and 24-36hr treatments had the opposite effect. At these time points, GNF179-treated parasites at least doubled in the next generation while MMV007564-treated parasites remained below the starting ring parasitemia. Therefore, 0-12hr asexual blood stage was most susceptible to GNF179 while 12-36hr asexual blood stage was most susceptible to MMV007564 indicating that these compounds likely target different proteins essential for asexual blood stage development.

To further define if the compound classes act on two distinct proteins, MMV007564 and KAF156 were assayed against *P. berghei* liver stages and *P. falciparum* stage V late sexual blood stages [149] to test for

multi-stage activity (Table 3.5). KAF156 remained equally potent in the liver and sexual stages with EC_{50}s not varying significantly from that of the asexual blood stage. In contrast, MMV007564 EC_{50}s for *P. berghei* liver stages and *P. falciparum* stage V gametocytes are 2 and 5-fold higher than the EC_{50} for the asexual blood stage, respectively. This suggests that the target for KAF156 is essential for all parasite life cycle stages while that of MMV007564 is essential only for asexual blood stage development.

3.3 Discussion and Conclusions

Results of this study suggest that the SNVs observed in *pfcarl* in our *in vitro* generated MMV007564^R lines are the genetic determinants for MMV007564 resistance. We have ruled out *pfcarl* SNVs as compensatory mutations due to the following observations: (1) mutations in different regions of *pfcarl* arose independently in 3 separate selection flasks; (2) MMV007564^R lines were cross-resistant with imidazolopiperazines, a distinct class of compounds previously shown to have *pfcarl* mutations as the genetic determinant of resistance; (3) an independently generated imidazolopiperazine-resistant line was cross-resistant with MMV007564 and other benzimidazolyl piperidines; and (4) this and previous studies with various *pfcarl* mutant lines have found a lack of cross resistance to standard antimalarials, including mefloquine, artemisinin [97], quinine and atovaquone, implying the cross resistance seen in this study is specific to

the benzimidazolyl piperidine and imidazolopiperazine compound classes.

Experiments to identify the cell cycle window in which each compound exerted its maximal effect revealed that GNF179 and MMV007564 were more effective at different asexual stages, implying that they have different cellular targets. Therefore, PfCARL is most likely a common resistance mechanism. Early asexual blood ring stages were most susceptible to GNF179 while late ring to trophozoite stages were most susceptible to MMV007564, suggesting that the compounds target different proteins crucial at different asexual blood stage development. Consistent with the possibility that the two compound classes have different targets, MMV007564 lost its potency in the liver and sexual blood stages while KAF156 did not, indicating that in contrast to the MMV007564 target, the KAF156 target is essential for the development of all parasite life cycle stages. It may be that PfCARL is a resistance mechanism that prevents the benzimidazolyl piperidines and imidazolopiperazines from acting on their yet unidentified targets.

It is not uncommon for unrelated antimalarials to have common genetic determinants of resistance, reflecting either a multidrug resistance mechanism or a common drug target. The two most well studied multidrug resistance mechanisms involve PfCRT and PfMDR1 [150], both

localized in the digestive vacuole membrane and members of the drug/metabolite and ATP-binding cassette transporter families, respectively. The K76T mutation in PfCRT is a validated marker for chloroquine resistance. PfCRT haplotypes also influence susceptibility to other antimalarials including amodiaquine, quinine and lumefantrine. Mutations and copy number variations in *pfmdr1* have been reported to influence susceptibility to lumefantrine, artemisinin, quinine, mefloquine, halofantrine and chloroquine. The targets of these antimalarials remain unclear but both PfCRT and PfMDR1 seem to be involved in the transport of these compounds away from the yet unidentified targets in the parasite [128, 150]. Recently, drug target identification efforts have identified other common genetic determinants of resistance for unrelated compound classes such as: (1) dihydroorotate dehydrogenase (PfDHODH) mutations conferring resistance to alkylthiophenes [112, 151, 152] and triazolopyrimidines [111, 112, 153, 154]; (2) P-type cation transporting ATPase (PfATP4) mutations conferring resistance to spiroindolones, pyrazoles, dihydroisoquinolones and a number of antimalarial agents in the Medicines for Malaria Venture (MMV) Malaria Box [155]; (3) cytochrome bc_1 mutations conferring resistance to atovaquone [65, 156], decoquinate [88], tetracyclic benzothiazepines [129], 4(*1H*) pyridones [157], a diversity-oriented synthesis probe [100], and

quinolone-3-diarylethers [158]; and (4) phosphatidylinositol-4-OH- kinase (PI4K) mutations conferring resistance to imidazopyrazines [91] and 2-aminopyradines [159]. For each of the PfDHODH, PfATP4, cytochrome bc_1 and PfPI4K inhibitors, there is functional and/or biochemical validation that these unrelated chemotypes directly interact with the mutated target, indicating common targets instead of multidrug resistance mechanisms [65, 88, 91, 129, 151-159]. It is unclear why diverse chemotypes repeatedly converge on the same targets, but it may indicate that such common targets play critical biological roles and that there are a limited number of accessible drug targets in *Plasmodium*.

The PfCARL mutations we have identified as conferring benzimidazolyl piperidine and imidazolopiperazine resistance have not been observed in 2,517 sequenced global field isolates (Figure 2 and Supplemental Table 3) [160, 161]. Unlike the resistance mutations that cluster at or near the transmembrane domains, natural diversity is mostly found elsewhere in the protein (Figure 2 and Supplemental Table 3). It would be interesting to know whether the field isolates with PfCARL mutations at or near the transmembrane domains are more resistant to benzimidazolyl piperidines and imidazolopiperazines compared with those with mutations in other regions of the protein.

MMV007564^R and KAF156^R parasites emerged relatively quickly. At 10^9 3D7 inoculum, MMV007564^R parasites emerged only 15 days after a short pulse at 3xEC_{50} followed by 3xEC_{50} continuous drug pressure. Similarly, KAF156^R parasites emerged after 17 days from a 10^9 Dd2 inoculum after 3xEC_{50} of continuous drug pressure [97]. This time to resistance evolution *in vitro* is comparable to that of atovaquone, known to generate resistant parasites readily (16-30 days with 10^8 inoculum) [158], and relatively fast compared with several other compounds (30-100 days with 10^9 inoculum) [93, 101, 112, 129]. It should be noted however, that evolution of resistance depends on parasite strain, concentration of selection agent, and number of mutations required for resistance, making direct comparisons between studies difficult [162, 163]. Furthermore, this and other studies have shown that a single amino acid change in PfCARL is enough to give rise to imidazolopiperazine and benzimidazolyl piperidine resistance *in vitro*. From these observations, we speculate that imidazolopiperazine and benzimidazolyl piperidine resistance might be easily acquired in the field and/or that there might be preexisting resistance to these inhibitors. This will have an implication in the field efficacy of KAF156, an imidazolopiperazine currently in clinical trials. Thus, the identification of a suitable partner drug for KAF156 will be crucial to protect its efficacy. This has been effective in prolonging the useful

lifespan of atovaquone, which despite resistance readily evolving *in vitro*, *in vivo* and in the field [65, 66, 110, 164], it has been used as an effective antimalarial with its partner drug proguanil [165].

PfCARL is a conserved protein of unknown function. Previous studies have found that PfCARL homologs may play a role in protein folding within the endoplasmic reticulum in yeast [139, 140] and homeotic transformations during vertebrate development [141]. Understanding the function of PfCARL will be important in order to develop strategies that circumvent this common drug resistance mechanism.

3.4 Methods and Materials

3.4.1 Source of compounds and parasite lines, and culturing

MMV007564 (synonym: GNF-Pf-4877 [149]) was obtained from GlaxoSmithKline and is freely available as part of Medicines for Malaria Venture's (MMV) Malaria Box. The following MMV007564 analogs and control antimalarials were commercially obtained: compounds **1** (C776-1565), **2** (C776-4170), **3** (D264-0067), **4** (C776-3796; synonym: MMV019741 [95], **5** (C776-3971), **6** (C776-3599; synonym: GNF-Pf-2535 [149]), **7** (C776-3523), **8** (C776-3534), **9** (C776-3561), **10** (C776-3585), **11** (C776-3587), **12** (C776-3670), **13** (C776-3575) and **14** (C776-3628) were from ChemDiv, USA; while quinine, atovaquone, and mefloquine were from Sigma-Aldrich,

USA. The imidazolopiperazines – KAF156 and GNF179 were synthesized as described previously [145, 166].

The following asexual blood stage *P. falciparum* strains were originally obtained through MR4 as part of the BEI resources Repository, NIAID, NIH: 3D7, MRA-102, deposited by DJ Carucci and Dd2, MRA-156, deposited by TE Wellems. The fast growing clone of 3D7 – IG06 [167] was used in this study. The Dd2 strain resistant to KAF156 and KAF179 was generated in a previous study (clone B3), and sequenced in our lab as the strain mutations were previously identified using microarray [90]. Parasites were cultured by standard methods [168] in RPMI 1640 medium supplemented with 28mM $NaHCO_3$, 25mM HEPES, 25ug/ml gentamicin and 0.5% AlbuMAX II (Life Technologies 11021-045).

Transgenic hepatoma cells (HepG2- A16-CD81-EGFP) were obtained from the laboratory of Dominique Mazier (INSERM, France) [169]. *P. berghei* luciferase sporozoites (strain MRA868) were obtained by dissection of infected *Anopheles stephensi* mosquito salivary glands supplied by the New York University Insectary. *P. falciparum* stage V gametocytes were obtained from SANARIA by Dr. Stephen L. Hoffman.

3.4.2 In vitro selection of MMV007564-resistant asexual blood stage P. falciparum lines

To generate MMV007564^{R} asexual blood stage *P. falciparum* lines, an initial intermittent followed by continuous drug pressure was applied to 3D7 (Fig 3.1). Three culture flasks, equivalent to 3 independent selections (Flasks 1, 2 and 3), each with 10^9 3D7 parasites were pressured with 3xEC_{50} (1.5μM) of MMV007564 for 3 days until no live parasites were detected by microscopy. After which the compound was washed off and parasites allowed to recover, with media changes every 2-3 days. When the parasitemia reached ~2% (10^9 inoculum), the cultures were once again drug pressured at 3xEC_{50}. Pressure was subsequently increased to 10xEC_{50} (5μM), as the parasites did not die off under continuous drug pressure at 3xEC_{50}. Due to limitations in MMV007564 availability, the inoculum size was reduced to 5 x 10^7 at this time. Parasites continued to grow at a typical multiplication rate at 10xEC_{50}. At different time points during the selection process, *in vitro* sensitivities to MMV007564 and control antimalarials by a SYBR Green I-based cell proliferation assay were performed to monitor MMV007564 resistance and aliquots of infected red blood cells (RBCs) were collected for genomic DNA (gDNA) extraction and whole genome sequencing to monitor the acquisition of mutations (Fig. 3.1). After MMV007564 resistance was confirmed for the parasites growing in

$10xEC_{50}$, each of the three independent selections were cloned by limiting dilution in the absence of the compound, with an inoculum size of 0.2 to 0.05 infected RBCs per well. Parasite clones were detected by microscopy after ~3 weeks of growth. Resistant parasites were cryopreserved in a solution composed of 28% glycerol, 3% sorbitol and 0.65% sodium chloride.

3.4.3 Genomic DNA extraction, preparation and analysis of sequenced samples

Parasites were isolated from erythrocytes by washing infected RBCs with 0.05% saponin. gDNA was subsequently isolated by following standard DNeasy Blood and Tissue Kit protocols (Qiagen). DNA libraries were prepped with the Nextera XT kit (Cat. No FC-131-1024, Illumina) using the standard dual index protocol, and sequenced on the Illumina HiSeq 2500 using a RapidRun mode and sequencing 100 base pairs deep on either end of the fragments. Paired-end reads were aligned to the *P. falciparum* 3D7 reference genome (PlasmoDB v. 13.0), following the Platypus pipeline as previously described [132]. Mutations, however, including single nucleotide variants (SNVs) and insertion/deletions (INDELS) were called using GATK's HaplotypeCaller, filtering mutations based on general recommendations from GATK (Table 3.2). Samples were additionally filtered by removing positions where read coverage was < 5

in the parent and any position where all samples had a heterozygous ratio > 0.2 (reference/total reads).

3.4.4 Binomial distribution calculation of polyclonal mutations

The binomial distribution was calculated for the mixed read variant calls in both Gen16 and Gen28. Calculations were conducted in Excel with the BIOM.DIST(x,trials,p,FALSE) function, where x was the number of successes (2 for Gen16 and Gen28), trials was the number of total exome region mutations (Gen16 = 31, Gen28 = 49), and p was the probability of randomly generating a mutation in *pfcarl*, which was the fraction of bases in *pfcarl* versus the entire genome (4968 / 23292622).

3.4.5 Dose response assay phenotyping for P. falciparum asexual blood stage

In vitro drug sensitivities of asexual blood stage MMV007564^{R} lines, the KAF156^{R} line and their respective 3D7 and Dd2 parents were determined using a SYBR Green I-based cell proliferation assay [170]. Twelve-point curve dilution series of the test compound were additionally carried out in duplicate or triplicate on the same day and replicated on at least three different days. The EC_{50} values were calculated using a nonlinear regression curve fit in Prism 5 (GraphPad Software Inc.) Parasite lines were considered resistant if their EC_{50} values significantly increased at least 2-fold from that of the parent.

3.4.6 Timing of action of MMV007564 and GNF179 in P. falciparum asexual blood stage

To compare the timing of action for MMV007564 and GNF179 in *P. falciparum* asexual blood stages, tightly synchronized 0-5hr old ring stage 3D7 parasites were prepared using the method of Witkowski et al (2013) with some modifications [171]. Schizonts were purified by a 35%/65% discontinuous Percoll (Sigma-Aldrich, USA) gradient, washed in RPMI-1640, and cultured for five hours to allow reinvasion into fresh RBCs. To remove remaining schizonts and hence purify 0-5hr old rings, the culture was treated with 5% D-sorbitol (Sigma-Aldrich, USA). The purified 0-5hr old rings were adjusted to 1% parasitemia, 2% hematocrit in 1 ml culture volumes and then incubated for 46hrs until the next ring stage. MMV007564 and GNF179 at 10xEC_{50} (5μM and 40nM, respectively) were added to the cultures at 0-12hr, 12-24hr, 24-36hr, 36-46hr, 0-24hr, 24-46hr and 0-46hr. A no compound control was included. Smears for microscopy were prepared at 0hr and every 10-12hrs for 46hrs. The experiment was replicated on three different days.

3.4.7 Dose response assay phenotyping for multi-stage activity

Liver-stage activity was determined by pretreating hepatic human cells (HepG2) for two hours with a 12-point curve dilution series of test compound in 1536 well plates infected with freshly dissected *P. berghei*

sporozoites. After 48 hours of incubation, the viability of *P. berghei* exoerythrocytic forms (EEF) was measured by luminescence reaction light output using BrightGlo (Promega). Dose response curves were repeated on three different days in order to calculate an average EC_{50} value.

Sexual-stage activity was determined by testing each test compound against late stage *P. falciparum* gametocytes using a MitoTracker fluorescent-based assay. Briefly, synchronized stage V gametocytes were incubated with compound for 72 hours, after which MitoTracker® Red CMXRos (life technologies) was added to each well together with saponin to lyse the RBCs. Following fixation, plates were imaged using an Operetta High Content Imaging System (Perkin Elmer) at fluorescence (590-640nm). High content image analysis software supplied by the Operetta was used to measure the gametocytocidal activity (Harmony, Perkin Elmer), with the readout based on the metabolically active gametocyte count per well.

3.4.8 Statistical analysis

Ordinary one-way ANOVA followed by Dunnett's multiple comparisons test was used to determine whether there is a significant difference in mean EC_{50} values between multiple resistant lines and their parent while two-tailed t-test was used for a single resistant line and its parent: $^{****}p < 0.0001$; $^{***}p < 0.001$; $^{**}p < 0.01$.

3.5 Acknowledgements

Chapter 3, in full, has been submitted for publication for the material as it may appear in American Chemical Society (ACS) Infectious Diseases, 2016. Victoria C. Corey and Pamela A. Magistrado, Amanda K. Lukens, Greg LaMonte, Erika Sasaki, Stephan Meister, Melanie Wree, Dyann F. Wirth, Elizabeth Winzeler, "Plasmodium falciparum Cyclic Amine Resistance Locus, PfCARL: A Resistance Mechanism for Two Distinct Compound Classes". The dissertation author was the co-primary investigator and author of this paper.

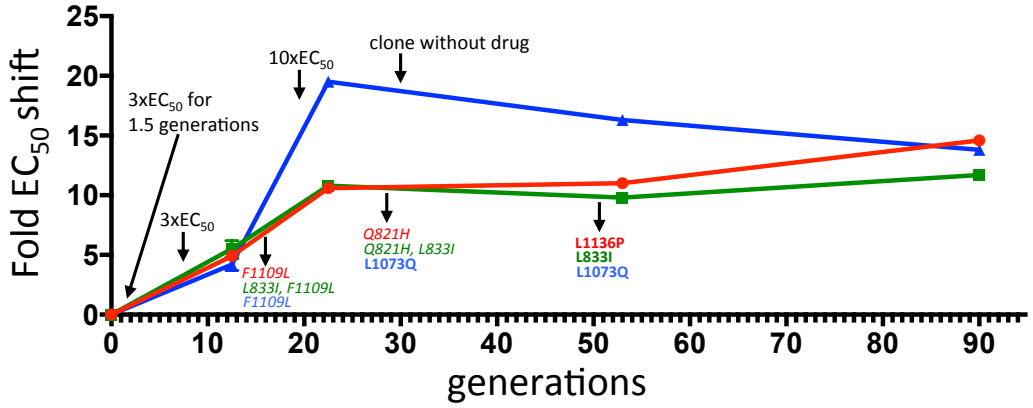


Figure 3.1 *In vitro* resistance selection timeline Selection with MMV007564 starting at 3xEC50 pulse for 1.5 generations followed by continuous drug pressure at 3xEC50 from generation 7 and 10xEC50 from generation 19, resulting in resistant lines as indicated by increasing fold EC50 shift from the 3D7 parent. Fold EC50 shifts are shown for 3 independent selections F1 (red), F2 (green) and F3 (blue). Time points when cloning without drug for single-cell isolation were performed and pfcarl mutations identified by whole genome sequencing analysis (heterozygous mutations in italics and homozygous mutations in bold) are indicated with arrows.

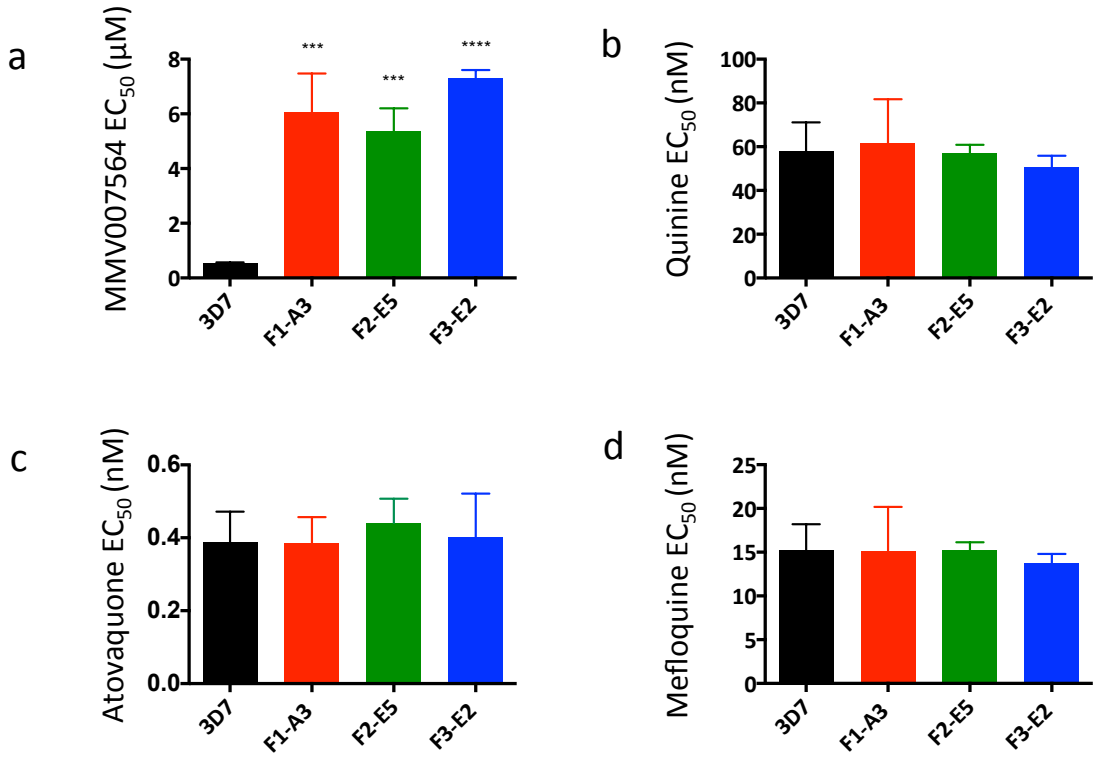


Figure 3.2 EC_{50} values of 3 clones from 3 independent selections Selections demonstrated at least a statistically significant 10-fold increase for MMV007564 (a) compared to the 3D7 parent but not for other antimalarials, quinine (b), atovaquone (c) and mefloquine (d). Bars are means of 3 independent experiments showing standard deviations. Statistical analyses were performed using ordinary one-way ANOVA followed by Dunnett's multiple comparisons test.

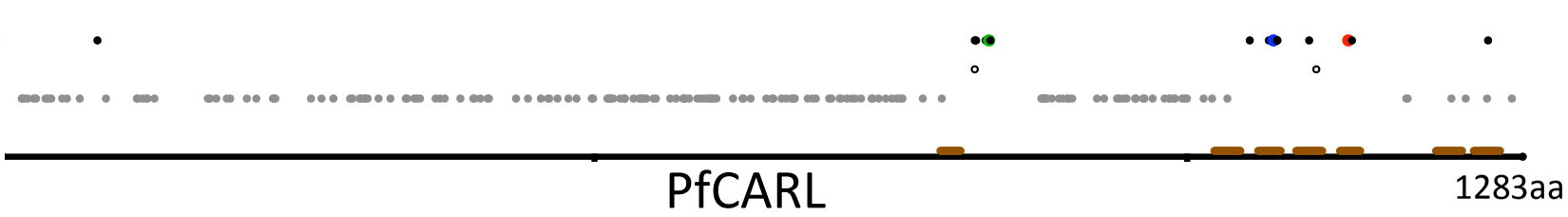


Figure 3.3 Single nucleotide variants (SNVs) in PfCARL Novel SNVs in MMV007564R clones F1-A3 (red solid circle), F2-E5 (green solid circle) and F3-E2 (blue solid circle), known SNVs implicated in imidazolopiperazine resistance (black solid circles) [90, 97], SNVs in polyclonal cultures of MMV007564R (black open circles) and SNVs in 2,517 sequenced global field isolates (gray solid circles) [161]. Transmembrane domains are indicated by brown broken lines. Amino acid mutations are indicated in Supplemental Table 3.

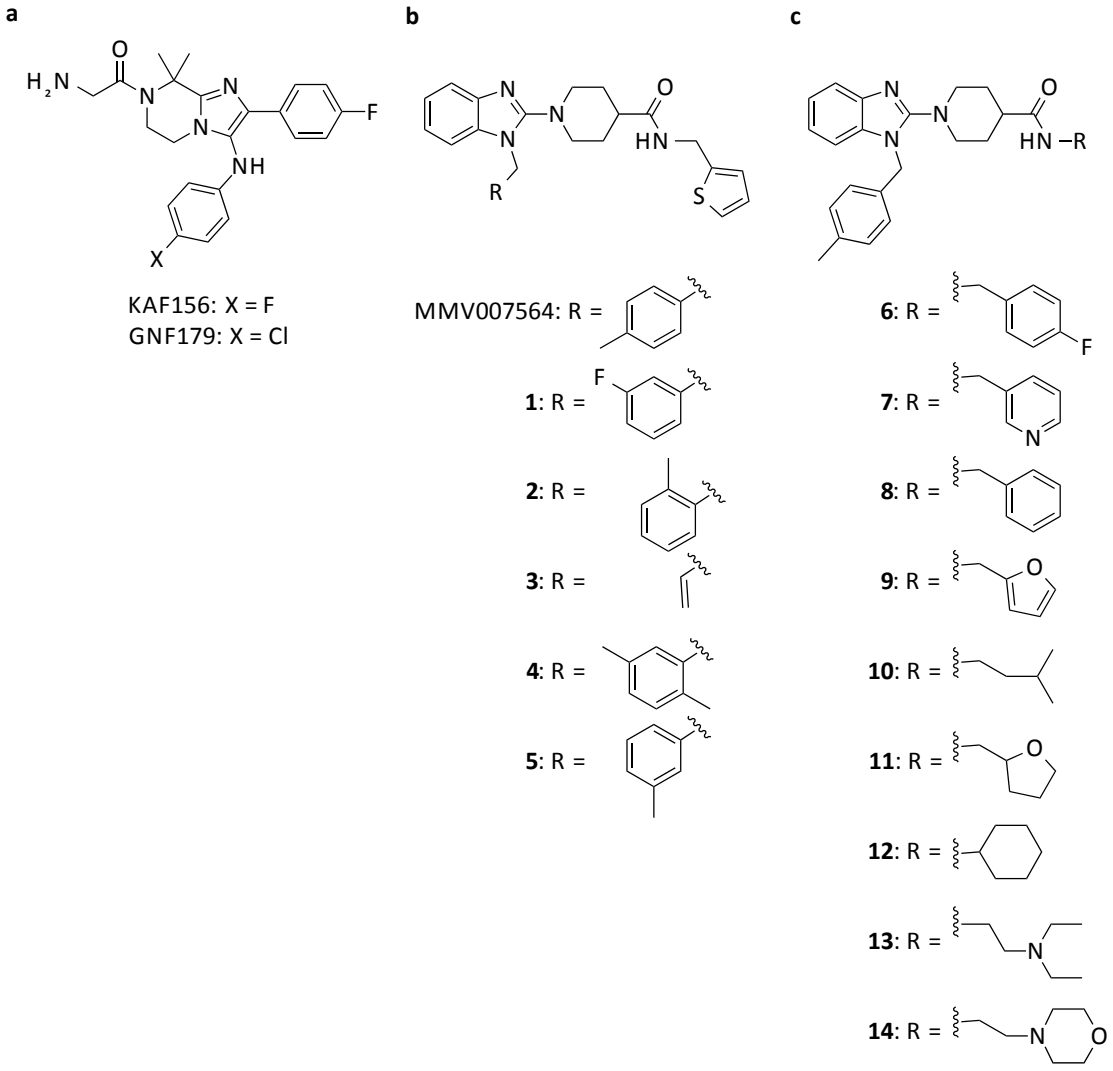


Figure 3.4 Chemical structures of imidazolopiperazines and benzimidazolyl piperidines (a) Imidazolopiperazines. (b) MMV007564 and other benzimidazolyl piperidines where the methyl benzyl group of MMV007564 is replaced with different chemical groups. (c) Benzimidazolyl piperidines where the thiophen of MMV007564 is replaced with different chemical groups.

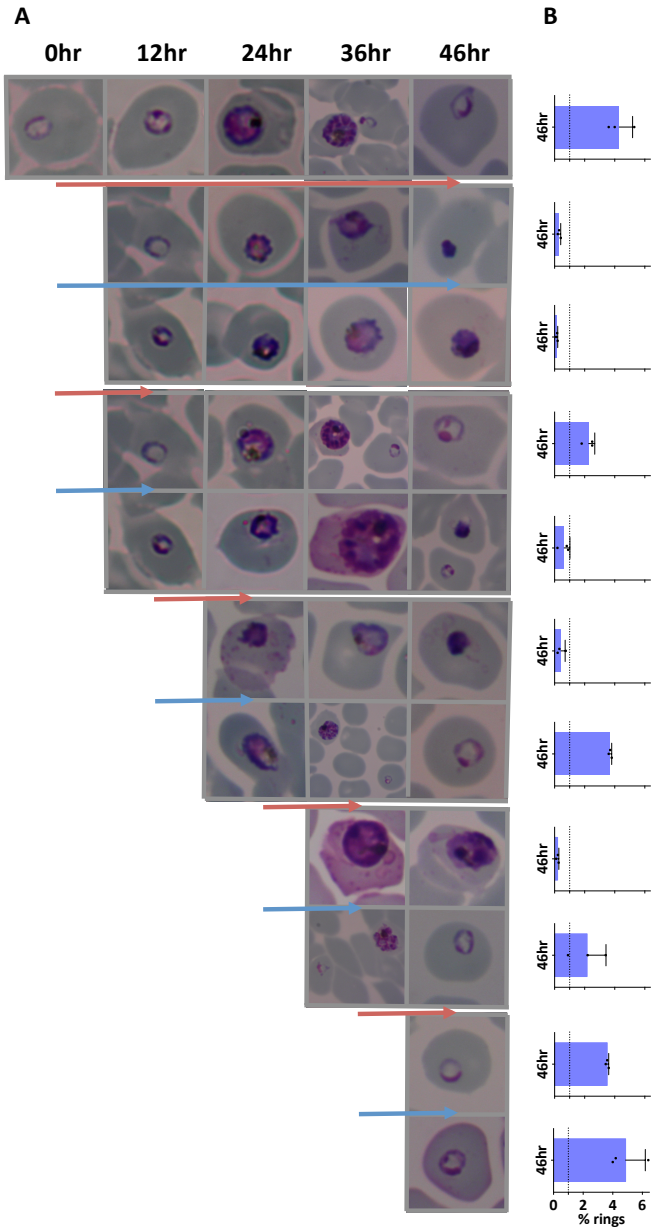


Figure 3.5 Effects MMV007564 and GNF179 on varying *P. falciparum* asexual life cycle stages (a) Susceptibility of the different asexual blood stages to 10xEC50 of MMV007564 (red arrow) and GNF179 (blue arrow) at different time points starting from tightly synchronized, 0-5hr old rings. Top panel is no compound. Arrows indicate the period when each compound was present in the culture (0-46hr, 0-12hr, 12-24hr, 24-36hr, 36-46hr). (b) Bars indicate average ring-stage parasitemia at the 46hr time point and error bars indicate standard deviation for 3 independent experiments. Broken lines at 1% rings indicate starting parasitemia at 0hr.

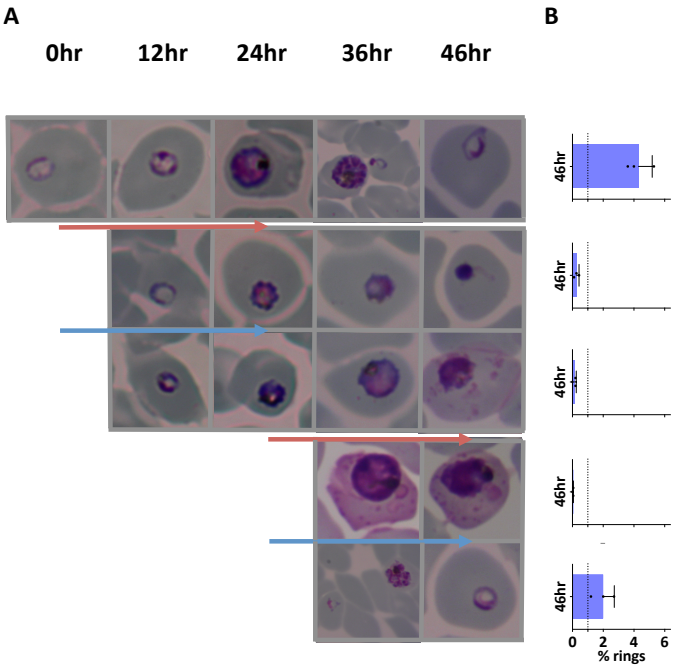


Figure 3.6 Effects MMV007564 and GNF179 on *P. falciparum* asexual life cycle stages at 0-24hr and 24-46hr pressures (a) Susceptibility of the different asexual blood stages to 10xEC50 of MMV007564 (red arrow) and GNF179 (blue arrow) at different time points starting from tightly synchronized, 0-5hr old rings. Top panel is no compound. Arrows indicate the period when each compound was present in the culture (0-24hr and 24-46hr). (b) Bars indicate average ring-stage parasitemia at the 46hr time point and error bars indicate standard deviation for 3 independent experiments. Broken lines at 1% rings indicate starting parasitemia at 0hr.

Table 3.1 SNV and INDEL mutation summary across generations

	Generation 16			***Generation 28***			***Generation 50***		
	7564-F1	***7564-F2***	***7564-F3***	***7564-F1***	***7564-F2***	***7564-F3***	***7564-F1***	***7564-F2***	***7564-F3***
genome coverage (x)	65.15	102.48	70.1	81.34	63.85	79.76	111.06	106.81	86.12
% covered by 15 or more reads	94.6	96.7	94.9	95	86.6	89.6	97.5	97.5	96.6
SNVs identified									
total mutations	6	1	9	0	0	1	2	1	1
intergenic	5	0	5	0	0	0	1	0	0
intronic	1	1	4	0	0	0	0	0	0
synonymous	0	0	0	0	0	0	0	0	0
Non-synonymous	0	0	0	0	0	1	1	1	1
genes mutated in all samples (mutation)		*none*			*none*		*Pf3D7_0321900: PfCARL*		
INDELs identified									
total mutations	1	2	2	0	0	0	1	1	2
intergenic	0	0	1	0	0	0	0	1	2
intronic	1	2	1	0	0	0	1	0	0
synonymous	0	0	0	0	0	0	0	0	0
Non-synonymous	0	0	0	0	0	0	0	0	0
genes mutated in all samples (mutation)		*none*			*none*			none	

Table 3.2 Applied filters for GATK's HaplotypeCaller

SNV Filters		*INDEL Filters*	
Filter Name	*Filter Value*	*Filter Name*	*Filter Value*
ReadPosRankSum	> 8.0	*ReadPosRankSum*	< -20
	< -8.0	*QUAL*	< 500
QUAL	< 500	*QD*	< 2
QD	< 2	*DP*	< 7
MQRankSum	< -12.5		
DP	< 7		

Table 3.3 Cross-resistance between MMV007564- and imidazolopiperazine-resistant lines with imidazolopiperazines and active benzimidazolyl piperidines with EC_{50} values < 3.5μM against the 3D7 parent Statistical analysis was performed using two-tailed t-test when comparing mean EC50s in the KAF156R line against the parental Dd2 and ordinary one-way ANOVA followed by Dunnett's multiple comparisons test when comparing mean EC50s in three MMV007564R lines against parental 3D7. Light to dark blue shading indicates significant 2-370 fold EC50 shift compared to parental line.

Compound	*EC_{50} (nM) ± SD*					
	Parent Dd2	*KAF156^R PfCARL: M81I, L830V, S1076I*	*Parent 3D7*	*MMV007564^R PfCARL: L1136P*	*MMV007564^R PfCARL: L833I*	*MMV007564^R PfCARL: L1073Q*
Imidazolopiperazines						
KAF156	9 ± 2	3331 ± 119****	10 ± 3	6 ± 0.3	48 ± 5****	9 ± 1
GNF179	3 ± 1	1024 ± 120***	4 ± 0.2	25 ± 4****	2 ± 0.09	5 ± 1
MMV007564 and other benzimidazolyl piperidines with the methyl benzyl of MMV007564 replaced with various chemical groups						
MMV007564	342 ± 145	1622 ± 134***	513 ± 56	6059 ± 1415***	5355 ± 848***	7319 ± 287****
4	180 ± 100	1688 ± 141***	822 ± 66	5896 ± 1881**	5926 ± 1948**	5449 ± 1002*
5	233 ± 73	1212 ± 170***	893 ± 118	8201 ± 1060*	9811 ± 2542**	11861 ± 5906**
Benzimidazolyl piperidines with the thiophen of MMV007564 replaced with various chemical groups						
6	174 ± 46	704 ± 77***	415 ± 30	3509 ± 476***	2046 ± 41*	6999 ± 1342****
8	679 ± 158	1378 ± 85**	2222 ± 593	8321 ± 1313***	8746 ± 1172***	8534 ± 1977***
9	956 ± 278	2176 ± 299**	3428 ± 960	11121 ± 5015*	11122 ± 3662*	9512 ± 3672*

Table 3.4 Inactive benzimidazolyl piperidines with EC_{50} values > 3.5μM against the 3D7 parent

Compound	*EC50 (nM) ± SD*			
	Parent 3D7	*MMV007564^{R} PfCARL: L1136P*	*MMV007564^{R} ^{D}PfCARL: L833I*	*MMV007564^{R} PfCARL :L1073Q*
Benzimidazolyl piperidines with the methyl benzyl of MMV007564 replaced with various chemical groups				
1	11612 ± 1686	12375 ± 2279	12011 ± 3727	9367 ± 2483
2	4466 ± 638	4298 ± 1426	3877 ± 1103	3016 ± 920
3	54883 ± 8276	61270 ± 11831	50773 ± 7835	49963 ± 4311
Benzimidazolyl piperidines with the thiophen of MMV007564 replaced with various chemical groups				
7	11488 ± 1875	17323 ± 7007	17240 ± 4820	15420 ± 1051
10	9437 ± 2344	9338 ± 2900	9408 ± 1677	7976 ± 1536
11	20090 ± 5409	21447 ± 6509	22180 ± 4132	17120 ± 3304
12	8522 ± 1741	7680 ± 2224	8490 ± 3015	6510 ± 1826
13	5714 ± 997	5520 ± 750	6047 ± 272	5648 ± 147
14	27917 ± 3530	27567 ± 4755	25710 ± 4692	23620 ± 2379

Table 3.5 Potency of MMV007564 and KAF156 at different *Plasmodium* life cycle stages Statistical analyses were performed using ordinary one-way ANOVA followed by Dunnett's multiple comparisons test to compare mean EC50s in liver and gametocyte stages against asexual blood stage for each compound.

Parasite life cycle stage	*EC_{50} (nM) ± SD*	
	MMV007564	*KAF156*
3D7 asexual blood stage	513 ± 56	10 ± 3
P. berghei liver stage	1250 ± 248**	10 ± 9
P. falciparum stage V gametocyte	2550 ± 439****	4 ± 1

Chapter 4

Whole Genome Shotgun Sequencing Shows Selection on *Leptospira* Regulatory Proteins During *in vitro* Culture Attenuation

4.1 Introduction

Leptospirosis is the most common zoonotic disease worldwide with an estimated 500,000 severe cases reported annually, and case fatality rates of 12–25%, due primarily to acute kidney and lung injuries. Despite its prevalence, the molecular mechanisms underlying leptospirosis pathogenesis remain poorly understood. In a previous independent study, we used reference-guided assemblies to identify inactivating non-synonymous single nucleotide variant (nsSNVs) in 11 putative virulence-associated genes that had emerged after passaging a P1 isolate for 18 subcultures including a family of virulence-modifying proteins upregulated during *in vivo* in an acute hamster infection model [172]. However, in this experiment, we considered only dominant alleles in P1 and P18 isolates. Here, in an independent attenuation experiment, we serially *in vitro* passaged the P1 isolate (LD_{50} < 100 *Leptospira*) into an avirulent derivative (P8A, LD_{50} > 108). We define the cumulative genome changes accompanying this observed loss of virulence by comparing the genomes of the parental strain and its isogenic, attenuated derivative through the use of next-generation sequencing and a custom SNV calling pipeline [132].

4.2 Results

4.2.1 Culture passage-based attenuation of L. interrogans *serovar Lai strain 56601*

The P1 isolate was derived from *L. interrogans* serovar Lai strain 56601 that had been serially passaged 3X in vivo to ensure a virulent phenotype [172]. The LD_{50} was determined to be < 102 *Leptospira* [172]. The P1 isolate was serially passaged *in vitro* in liquid Leptospira culture medium for ~400 generations (16 weeks) to become P8A. The LD_{50} of the P8A isolate was determined to be > 108 *Leptospira*, administered IP, indicating a complete loss of virulence. After *in vitro* passage, genomic DNA was isolated from both the P1 and P8A strains and frozen before sequencing.

4.2.2 Identification of SNV alleles differing in frequency between the attenuated and parental strains

Cumulative changes occurring during adaptation to *in vitro* growth, and associated with loss of virulence, were studied at the whole genome level. Genomic DNA from the polyclonal parental strain, P1, and from the attenuated isogenic derivative, P8A, was sequenced on an Illumina platform, using paired-end 100-bp reads to a mean coverage of greater than 250X. For strain P1, 15,492,436 reads were generated covering 99.4%

of the *L. interrogans* serovar Lai reference genome (4.689 Mb), and 15,651,273 reads were generated from P8A covering 99.9% of the reference genome (Table 4.1). In addition, > 99% of the reads from both the P1 and P8A samples aligned to the *L. interrogans* Lai strain 56601 genome, indicating high sample purity.

Variants were called and compared using a modified automated PLATYPUS genome analysis pipeline [132]. PLATYPUS aligned reads from each sequencing run (P1 and P8A) to the reference Lai genome [56] and identified SNVs using a default list of filters for each set of sequencing files. Given that the bacterial populations were not clonal, an allele frequency was calculated at each polymorphic site using the number of aligned reads metric for the P1 and P8A isolate (Supplemental Table 4.1). From this analysis, 99 SNVs were identified as having undergone a significant change in allele frequency between P1 and P8A, as determined by a two-proportional z-test before Bonferroni correction. Alternate nucleotides in these positions would result in 43 SNVs encoding synonymous amino acid substitutions, 34 encoding non-synonymous amino acid substitutions, and 25 intergenic SNVs (Fig. 4.1; see Supplemental Table 4.1 for complete listing of variants). In the P8A genome, all of these minor alleles had changed allele frequencies by at least 5% compared with the P1 genome and vice versa.

4.2.3 Analysis of nsSNVs with allelic frequencies that increased during attenuation

Since amino acid coding changes can alter overall functionality of the gene in which they reside, and may contribute to the observed loss of virulence in the P8A strain, we further examined the nsSNVs that were identified during our genomic comparisons. There were 15 genes that contained nsSNVs that increased in frequency during the course of the attenuation (Fig. 4.2A). To determine if the genes containing these nsSNVs were biased toward any particular biological function, they were organized by COG category [173], and the observed proportions were compared with their genome-wide expected frequencies. This approach identified a strong enrichment for genes involved in signal transduction mechanisms (Fig. 4.2B). Of the 3,683 genes in the genome, 233 are annotated as involved in signal transduction and comprised five of the 15 in our set (P = 0.01). We also noted that three genes, rbsK, mgtA, and mcm2 (encoding a putative ribokinase, a magnesium transporter, and methylmalonyl-CoA mutase, respectively), contained multiple SNVs. This is a higher number than that would be expected due to chance alone, and because these genes all have functions related to core metabolic pathways and cofactor biosynthesis, their allele frequency increase may be a result of bacterial adaptation to long-term *in vitro* culture conditions.

To infer additional possible biological significance of these genes, we performed a meta-analysis of previously published data showing transcriptional responses of *L. interrogans* under several surrogate *in vivo* conditions including temperature, physiological osmolarity, iron depletion, exposure to host innate immune cells, and peritoneal culture of pathogenic *Leptospira* in dialysis membrane chambers [76, 77, 79-82]. Of the 15 genes identified by this study as harboring nsSNVs of increasing allele frequency, six (LA_2704, LA_2930, LA_2950, LA_3455, LA_3725, and LA_3834) were previously reported to be upregulated in at least one set of *in vitro* surrogate experimental conditions.

To gain further insight into how these genes might contribute to the pathogenicity of *Leptospira* and their overall prevalence in the genus, the subcellular locations of the proteins they encode were predicted using PSORTb v. 3.0 (http://psort.org/psortb/index.html) [174], and the prevalence of gene homologs across all 20 species of the *Leptospira* pan-genome was also determined. Genes in other *Leptospira* species were considered homologous to our study-identified genes if they were reciprocal best BLAST hits using filters of 70% query length, e-value < 1e−3, and 30% identity match (Fig 4.3). This analysis revealed that five of the six genes (the subcellular location of LA_2950 could not be determined by the algorithm) were predicted to reside inside the bacterial cell,

indicating that these proteins are likely not the ultimate effectors of *Leptospira* pathogenesis, like toxins or other secreted factors, but may contribute to upstream signaling processes or metabolic capability. The pan-genus conservation analysis showed that three genes (LA_2930, LA_3725, and LA_3834) are found only in infectious *Leptospira* species and may have particularly relevant pathogenesis-related functions.

*4.2.4 Pan-*Leptospira *genomic analysis of amino acid residue conservation at nsSNV positions in homologs of attenuation-identified genes*

We conducted a three-part analysis of six genes of interest (Fig. 4.3) to determine if the nsSNVs in these genes caused amino acid changes in evolutionarily conserved residues. First, protein domain architecture was evaluated using SMART37 and PSIPRED [175, 176]. Next, we generated MSAs using homologous sequences from the 20 species pan-*Leptospira* genome for each of these genes. These MSAs were used to generate amino acid conservation scores for each residue in a respective gene based on JSD (scores above 0.8 are considered highly conserved, those less than 0.4 are considered disordered) [177]. Finally, we compared the proportion of sequencing reads from the P8A strain coding for the nsSNV amino acid to the proportion of the same mutant residue in homologs

from the entire pan-*Leptospira* genome using a Fisher's exact test with the following results.

LA_2704

Diguanylate cyclases participate in the formation of the ubiquitous second messenger, cyclic diguanylate monophosphate (cyclic-di-GMP), involved in bacterial virulence, biofilm formation, and persistence [178, 179]. The non-synonymous C311F substitution in this GGDEF, diguanylate cyclase is C-terminal to the catalytic core of this protein by one amino acid residue (Fig. 4.4A) [180]. The wild-type cysteine residue is conserved in every single homolog evaluated in this study, which is reflected by the high conservation score obtained from JSD analysis. The proportion of phenylalanine substitutions observed in the P8A strain represents a highly significant divergence from the genus-wide residue conservation at this position ($P < 0.001$).

LA_2930

The Y94F substitution in this Per-ARNT-Sim-(PAS)-GGDEF predicted signaling protein falls within the PAS sensor domain (Fig. 4.4B). PAS domains detect a large range of chemical and physical signals and then regulate the activity of their covalently linked effector domains, often by promoting the formation of dimers (a process required for proper GGDEF domain function). We could not deduce any insight into the particular ligands to which the PAS domain of this protein may bind, as the range of

potential signals is diverse (ranging from oxygen tension to small metabolites and to light itself) and on average, the pairwise identity shared between PAS domains is less than 20% [181]. Nonetheless, conservation analysis revealed that this position is highly conserved in *Leptospira* with significant divergence ($P < 0.03$) away from conservation status in the P8A attenuated strain.

LA_2950

Post synaptic density protein (PDZ) serine proteases are a unique family of proteins that form higher order oligmeric structures and have been demonstrated to degrade misfolded proteins in the periplasm of bacteria [182]. The P81R nsSNV in this PDZ serine protease was found to occur in an in-silico predicted coil to sheet transition, indicating that the wild-type proline residue may serve a structural role (Fig. 4.4C). JSD conservation analysis revealed the site to be moderately conserved within the Leptospira genus. The P8A arginine substitution at this residue was significant ($P < 0.05$), and was not observed in any of the protein homologs evaluated. Interestingly, domain architecture analysis revealed an N-terminal signal peptide indicating that this protein potentially has extracellular function.

LA_3455

This protein is a transmembrane nonselective transport channel found in the inner membrane of gram-negative bacteria that facilitates

the diffusion of glycerol M[183]. The S56P substitution in this protein was a significant divergence from genus-wide expected residues ($P < 0.01$) (Fig. 4.4D). The conserved residue position lies at the end of one of the eight α-helical regions of the aquaglyceroporin. The tight spatial arrangement of these helices to one another is essential for the proper function of the protein's glycerol-conducting channel [184], and the proline substitution in the P8A population of *Leptospira* could conceivably introduce a structural change that would alter its transport efficiency.

LA_3725

Domain analysis of the large LA_3725 protein revealed a single N-terminal transmembrane domain and a pre-toxin Hedgehog/Intein (HINT) domain (Pfam PF07591) nearer the C-terminal end of the coding region. The HINT superfamily belongs to a system of proteases that in bacteria are usually found N-terminal to a toxin module in polymorphic toxin systems [185, 186], and are believed to release the toxin domain via autoproteolysis. The L1624V nsSNV position lies at the in silico predicted transition of an α helix to a coiled secondary structure in a region of low sequence conservation (Fig. 4.4E). MSA analysis revealed that the P8A proportion of nsSNV reads was not statistically significant compared with genus-wide expected ratios.

LA_3834

The nsSNV position identified in the attenuated P8 strain codes for an L8F substitution in the N-terminal lipobox [187] of this protein (Fig. 4.4F). This amino acid substitution occurs at a moderately conserved residue according to JSD analysis, that is, seven residues upstream of the cysteine residue that is lipidated during export through the bacterial inner membrane. Although there appeared to be some flexibility in the amino acid conservation at the SNV position, genus-wide analysis revealed that no homologs contained the mutant phenylalanine at this position, indicating a significant divergence from expected proportions ($P < 0.03$).

4.2.5 Intergenic SNV analysis and novel ncRNA prediction

Analysis of SNV allele frequency differences between the P1 and P8A *L. interrogans* Lai strains revealed 25 intergenic SNVs, 22 on chromosome I, and three on chromosome II (Fig. 4.1, Supplemental Table 4.1). In previous whole genome surveys, several ncRNA loci were detected in the *L. interrogans* Lai genome [48], including three cobalamin riboswitches that are expressed both in vivo and in vitro [82]. Because these elements play vital roles in the regulation of gene expression, mutations within predicted ncRNAs could have functional implications potentially affecting virulence.

To evaluate whether any of our study identified intergenic SNVs resided in predicted ncRNA loci, we generated a list of predicted ncRNA in the *L. interrogans* Lai strain 56601 genome using RNAz [188] and the nocoRNAc pipeline [189]. Fifty candidate ncRNA loci were identified on chromosome I (cI replicon), and five on the cII replicon, none of which contained study-identified intergenic SNVs. Of the 55 candidate ncRNA loci identified, 41 were antisense to protein coding genes and 14 were found in intergenic regions (Table 4.2). Of these 14 ncRNA loci, none could be annotated using the Rfam database and could represent novel ncRNA genes.

4.3 Discussion and Conclusions

This study analyzed genomic changes in a polyclonal population of *L. interrogans* serovar Lai strain 56601 that occur during the culture-based attenuation of a highly virulent parent strain into a nearly avirulent isogenic derivative. This analysis was carried out using a modified PLATYPUS pipeline, originally designed to analyze eukaryotic genomes, which was readily and accurately adapted for the analysis of *Leptospira* genomes (prokaryotic). Novel, potentially virulence-related genes were identified in this study by analyzing nsSNV allele frequency changes accompanying in vitro, culture attenuation of *L. interrogans serovar Lai.* Because of the stochasticity of the underlying processes giving rise to

deleterious mutations in virulence-associated genes that are under neutral selection *in vitro*, future attenuation experiments would be most informative if whole genome sequencing data from several independent attenuated lineages are compared. The data summarized here will provide the foundation for future investigations to determine the role these genes play in the pathogenesis of leptospirosis.

Genome changes that occurred in the bacterial population during long-term *in vitro* culture passage attenuation of the virulent P1 *L. interrogans* Lai strain 56601 isolate into the avirulent P8A isolate likely occurred as the result of selection for rapid growth *in vitro* culture, likely to be in a tradeoff with virulence. After the isolation of the P1 strain from hamsters, the only selective pressure on the bacterial population became intrapopulation competition for growth *in vitro* in EMJH media. Therefore, the process of natural selection under these conditions would be expected to increase the population-level allelic frequencies of mutations beneficial to *in vitro* growth [190]. Such changes are often accompanied by allele-frequency increases of mutations in genes necessary for growth *in vivo* (i.e., relaxed selection on virulence genes would lead to the accumulation of deleterious mutations in these genes during growth *in vitro*). Accordingly, nsSNVs otherwise deleterious to virulence *in vivo*, which had previously been kept at low frequencies by *in vivo* selection, for

example, immune pressures of the host, would now be selectively neutral in vitro. These mutations would then be free to synchronously move with alleles under positive selection for growth in EMJH media in a type of genetic hitchhiking [191].

Interestingly, all single nucleotide variations identified in our genomic analysis of the attenuated P8A isolate originated from existing low-frequency subpopulation alleles in the virulent P1 isolate. We did not find any spontaneous mutations arising during the *in vitro* attenuation process. Only preexisting mutations expanded in frequency based on statistically significant thresholds, a phenomenon also noted previously in the apicomplexan parasite *Babesia bovis* [192]. The original process that generated these mutants appeared to have proceeded in a stochastic manner, SNVs appeared across the *L. interrogans Lai* genome with a nucleotide transition to transversion ratio, Ts/Tv, of approximately 0.5 (Table 4.2), suggesting that at a given position a substitution of one nucleotide was just as likely as any other. All nsSNVs identified existed as minor variants to wild-type alleles in the P8A population (Fig. 4.2A). Surprisingly, nsSNVs diverging from the reference sequence had an allele frequency of only 12% ± 4.97 (mean ± SD). It has been previously demonstrated in several other pathogens that microbial populations may harbor subpopulations that retain pathogenic capacity, despite being

attenuated at the population level [193-196]. Similarly, we were able to detect wild-type alleles in the majority of sequencing reads derived from the P8A isolate.

Two previous studies have examined genome differences between virulent and avirulent strains of *L. interrogans serovar Lai* to identify mutations that accompany the loss of the virulence of the parental strain. The first, which compared genome differences between *L. interrogans serovar Lai* strain IPAV (avirulent) and a non-isogenic isolate of *L. interrogans serovar Lai* [197], identified several hundred SNVs in gene-coding regions as well as dozens of insertions and deletions; interestingly many SNVs were found in genes related to signal transduction. The second study, recently reported from our group [172], identified a set of SNVs in 11 pathogen-specific genes of an attenuated isogenic derivative of *L. interrogans serovar Lai* strain 56601. There was no overlap in the genes identified in these earlier studies with those identified in this work, which mirrors results from another experimental evolution study in *Escherichia coli* that found few of the 115 strain replicates shared similar mutations [190]. This lack of overlap strongly underscores the stochastic nature of SNV expansion in vitro. Nonetheless, it should be noted that the sequencing coverage was approximately 2.5X higher in this study relative to our previous study, reads were substantially longer here (100 bases

versus 36) and paired ends were used. Also significant, this analysis detected mixed alleles, whereas our previous study focused only on dominant alleles. Because genes necessary for *in vivo* growth are under relaxed selection *in vitro*, the complement of putative virulence genes identified in attenuation experiments can differ substantially (i.e., no convergence), suggesting, that is, approach would be most informative if data derived from several independent attenuated lineages are analyzed.

Pathogenic *Leptospira* have evolved numerous signal transduction proteins to properly respond to environmental as well as in vivo host queues [85], in contrast to obligate parasites that have far fewer [198]. Because pathogenic *Leptospira* are transmitted by soil and surface water, they must transition between the external and host environment.

The identification of nsSNVs in two GGDEF di-guanylate cyclase (DGC) signal transduction genes (LA_2704 and LA_2930, both previously shown to be upregulated during exposure to *in vivo*-like conditions; Fig. 4.3) in our study was particularly intriguing. GGDEF domains catalyze the formation of the ubiquitous secondary messenger di-cyclic-GMP66 through a process of homo-dimerization of two DGC domains from separate proteins [180]. Intracellular concentrations of di-c-GMP have been demonstrated experimentally to regulate several pathogenesis-

related bacterial processes related to biofilm formation, motility, and virulence [199-201]. The *L. interrogans serovar Lai* genome contains genes for 14 distinct GGDEF domain containing proteins [85], and members of the genus are known to produce biofilms both *in vitro* and *in vivo* [202, 203]. The physiological effects of di-c-GMP levels have been reviewed previously [178]; while intracellular di-c-GMP levels promote biofilm formation, they might have differential effects on (i.e., promote or inhibit) other phenotypes. While there are currently no experimental data regarding DGCs and di-c-GMP in *Leptospira*, it should be noted that di-c-GMP levels appear to positively regulate motility and virulence in other spirochetes [204-206].

Data from a model biofilm system using *Pseudomonas aeruginosa* have demonstrated that increases in intracellular di-c-GMP levels, through the action of DGCs, lead to secretion of exopolysaccharide components required for biofilm formation [207]. These polysaccharides then act as signals for DGCs in neighboring bacteria to increase their di-c-GMP levels, encouraging further exopolysaccharide secretion in a positive feedback mechanism similar to paracrine signaling in eukaryotes. Whether impaired GGDEF signaling causes a similar nonautonomous trait in *L. interrogans* remains undetermined, it is interesting to consider whether a small percentage of mutant cells (i.e., those harboring LA_2704 or LA_2930

nsSNV mutations) would influence the *in vivo* survival of the Leptospira population as a whole through impaired biofilm production.

Bacterial lipoproteins have been suggested to be involved in pathogenesis including adhesion to host cells, immune modulation, and the translocation of virulence factors into host cells [208, 209], and there are several predicted in the genomes of spirochetes [187]. Thus, the identification of mutations in the putative lipoprotein LA_3834 in this study is intriguing. While the function of this protein has yet to be determined experimentally, several independent lines of evidence point to LA_3834 being a part of the *Leptospira* virulence gene repertoire. In addition to being transcriptionally upregulated during *in vivo* surrogate experiments [79, 81, 82], LA_3834 was recently demonstrated to be under the control of a transcriptional regulator (LB_139) that when knocked out, decreased expression of several genes (including LA_3834) and attenuated virulence in a hamster model of leptospirosis [83].

Two other study-identified genes (LA_2950 and LA_3455) with nsSNVs at conserved residues may also be important to *Leptospira* virulence and survival *in vivo*. LA_2950 encodes a protein predicted PDZ serine protease. Studies in *Salmonella typhimurium* have demonstrated that other PDZ serine proteases participate in the *in vivo* stress response to host microbicidal pressures [210-212]. Bacteria with mutations in these genes

were attenuated compared with wild-type parental strains, with decreased tissue burdens (up to a 105-fold decrease in one study [212]). LA_3455 encodes the Leptospira GlpF glycerol uptake facilitator protein. *L. interrogans* cannot use sugars as carbon sources, but instead, synthesizes sugars with de novo gluconeogenesis from glycerol [56]. Since the non-synonymous S56P SNV identified in the P8 strain of our study may introduce a strain in the secondary structure of a conserved helix essential for function, Leptospira cells harboring this mutation could conceivably experience an impaired acquisition of glycerol in vivo that could have downstream biosynthetic consequences.

To the best of our knowledge, our generation of a list of computationally predicted ncRNAs in *L. interrogans* is the first in the field. Although none of our study-identified intergenic SNVs mapped to these regions, small noncoding RNAs have recently been shown to regulate pathogenic mechanisms in bacteria [213-216]. Thus, it would be important to test whether similar mechanisms exist in pathogenic *Leptospira*.

This study has limitations, primarily in that further work needs to be done both qualitatively and quantitatively to describe the individual contribution of the genes identified here to *Leptospira* pathogenesis. It is likely that the genes identified in this study may be part of a virulence-related transcriptional profile, and the increase in nsSNV alleles seen may

collectively reduce the pathogenicity of *Leptospira* with these mutations. The ultimate mechanism of attenuation of *L. interrogans serovar Lai* in this study appears to be the additive effect of multiple mutant alleles, each subtracting from overall population fitness *in vivo*. The individual contributions of each of these genes to overall virulence is likely to remain hazy until more reliable methods of targeted mutagenesis are established for this important pathogen, until then attenuation-based studies are a reasonable alternative for identifying putative virulence-associated genes.

4.4 Materials and Methods

4.4.1 Ethics Statement

The experimental animal work was carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health in Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)-approved facilities, and was approved by the Institutional Animal Care and Use Committee of the University of California, San Diego under protocol S03128H.

4.4.2 Attenuation of L. interrogans *serovar Lai strain 56601*

Generation of the P1 isolate of L. interrogans *serovar Lai strain 56601*

Leptospira interrogans serovar Lai strain 56601 was kindly provided by David Haake (University of California Los Angeles, Los Angeles, CA), and was passaged through 3-week-old male Golden Syrian hamsters (*N* = 3, Charles Rivers Laboratories, Hollister, CA) to ensure a virulent phenotype. The initial three hamsters were each injected intraperitoneally (IP) with approximately 10^7 *Leptospira* in 1 mL of Ellinghausen-McCullough-Johnson-Harris *Leptospira* culture media (EMJH; BD Difco, Franklin Lakes, NJ). Four days post inoculation the animals were killed, the livers were harvested, macerated with a sterile scalpel blade, pooled in 5-mL sterile phosphate-buffered saline, then made into a slurry by repeatedly passing through a 22-gauge needle; 1 mL of this homogenate was then used to inject each of a second group (*N* = 3) of hamsters IP. The liver homogenization procedure was repeated 4 days later, and a third group (*N* = 3) of hamsters were injected, also IP. Four days after the IP injection of liver homogenate into the third group, the animals were killed, and livers harvested aseptically. Approximately 10 mg of minced liver tissue was then used to inoculate EMJH semisolid medium supplemented with 5-fl [53]. The semisolid culture was incubated at 25°C and monitored for *Leptospira* growth by dark field microscopy. Once growth occurred, 100

µL of this culture was used to inoculate 20 mL of sterile EMJH media, and the culture was incubated at 28°C on a rotary shaker, and was designated P1.

Genomic DNA isolation of P1 isolate

Approximately 10^7 *Leptospira* from 1 mL of EMJH P1 culture were spun down in a microcentrifuge (10,000 rpm, 5 minutes). Genomic DNA was then isolated from the cell pellet using the DNEasy Blood and Tissue kit (Qiagen, Valencia, CA) according to manufacturer's instructions. Eluted DNA was stored at -20°C for later sequencing.

LD_{50} determination of P1 isolate

Leptospira cells were counted using a Petroff-Hauser counting chamber (Hausser Scientific, Horsham, PA) under dark field microscopy. Challenge doses of 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, and 10^8 *Leptospira*/mL in sterile EMJH were then prepared based on observed counts. For each dilution group, 3-week-old male Golden Syrian hamsters (*N* = 3, Charles Rivers Laboratories) were each injected IP with 1 mL of the appropriate challenge dose. Animals were monitored for 21 days and euthanized when moribund. The LD_{50} was defined as the last dose in which two-thirds of the animals died after challenge.

In vitro EMJH culture-passage attenuation of the virulent P1 isolate into P8 isolate

The P1 isolate EMJH culture was subcultured by transferring 2 mL into 18 mL of sterile EMJH media (thus becoming P2A), and incubated at 28°C

on a rotary shaker for 14 days. This process was repeated iteratively for a total of seven subcultures, with the final subculture being designated P8A (~400 generations from the P1 parent culture). Genomic DNA extraction and LD_{50} determination were then performed exactly as described for the P1 isolate.

4.4.3 Genomic library preparation and assembly

Genomic DNA libraries were normalized to 0.2 ng/μL and prepared for sequencing using the Illumina Nextera XT Kit (Illumina, San Diego, CA) whole genome re-sequencing library according to manufacturer's instructions, using the Illumina protocol of tagmentation followed by ligation (v. 2013; Illumina, Inc., San Diego). DNA libraries were clustered and run on an Illumina HiSeq 2500 platform (Illumina) with PE100 on Rapid Run mode. Base calls were made using CASAVA v 1.8+ (Illumina).

Sequences were processed though the PLATYPUS pipeline (Winzeler Lab, UCSD, San Diego, CA) [132]. In brief, reads were aligned to the reference *L. interrogans* serovar Lai strain 56601 genome (NC_004342 and NC_004343) using Burrows-Wheeler Aligner [217], and unmapped reads were filtered using SAMtools [218]. SNVs were then initially called using Genome Analysis Toolkit [219, 220] and filtered using default filter values in PLATYPUS. Although the filters were initially designed for *Plasmodium falciparum*, they resulted in high sensitivity (93.4%) and specificity (91.2%)

for *L. interrogans* as well when screening for known SNVs between the 56601 and IPAV *L. interrogans* serovar Lai strains. After alignment, read depth per nucleotide identity at every position was called using SAMtools *mpileup*, which were then converted into proportional nucleotide identities per base. These proportions were then compared using a custom script testing for multi-comparison significant changes in allelic proportion across the entire genome. For two proportions x_1/n_1 and x_2/n_2 reads, our comparison statistic was the following:

$$z = \frac{\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right)}{\sqrt{\left(\frac{x_1 + x_2}{n_1 + n_2}\right)\left(1 - \frac{x_1 + x_2}{n_1 + n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

This statistic is an expansion of the simple two-proportional z-test for differences between two populations. This assumption is reasonable as each read serves as an independent random test of the nucleotide identity of the population, though significant error terms do exist. This number was then converted to a p-statistic using the total read depth and corrected using the Bonferroni method, as assumption about the independence of allelic frequency at multiple polymorphic sites may not hold. A list of sites that underwent statistically significant changes were then exported and annotated using a custom script.

4.4.4 Clusters of orthologous groups' functional category analysis of nsSNV-containing genes

Genes identified as containing nsSNVs with increasing allele frequencies in P8 were assigned to clusters of orthologous groups (COG) categories using the National Center for Biotechnology Information conserved domain webpage *(http://www.ncbi.nlm.nih.gov/Structure /cdd/cdd.shtml),* and compared with the genome-wide predicted COG frequencies for *L. interrogans* serovar Lai strain 56601 obtained from the spirochete genome browser webpage (http://sgb.fli-leibniz.de/cgi /index.pl). Statistical significance was assessed via χ^2 analysis using Fisher's exact test with a Bonferroni correction to account for multiple comparisons in Graphpad Prism (GraphPad Software, Inc., La Jolla, CA).

4.4.5 Pan genus comparative genome analysis of study-identified genes

The following genomes, consisting of a representative grouping all 20 *Leptospira* species, were used to analyze the presence of homologs of study-identified genes in other *Leptospira* species:

Leptospira alexanderi sv. Manhoa 3 str. L 60^T (Genbank: AHMT00000000), *Leptospira alstoni* sv. Pingchang str. 80-412 (Genbank: AOHD00000000), *Leptospira biflexa* sv. Patoc str. Patoc I Paris (Genbank: CP000786), *Leptospira borgpetersenii* sv. Javanica str. UI 09931 (Genbank: AHNP00000000), *Leptospira broomii* sv. Hurstbridge str. 5399^T (Genbank:

AHMO00000000), *Leptospira fainei* sv. Hurstbridge str. BUT 6^T (Genbank: AKWZ00000000), *Leptospira inadai* sv. Lyme str. 10^T (Genbank: AHMM00000000), *L. interrogans* sv. Copenhageni str. Fiocruz L1-130 (Genbank: AE016823), *L. interrogans* sv. Lai str. 56601 (Genbank: AE010300), *Leptospira kirschneri* sv. Cynopteri str. 3522 C^T (Genbank: AHMN00000000), *Leptospira kmetyi* sv. Malaysia str. Bejo-Iso9^T (Genbank: AHMP00000000), *Leptospira licerasiae* sv. Varillal str. VAR 010^T (Genbank: AHOO00000000), *Leptospira meyeri* sv. Hardjo str. Went 5 (Genbank: AKXE00000000), *Leptospira noguchii* sv. Panama str. CZ 214^T (Genbank: AKWY00000000), *Leptospira santarosai* sv. Shermani str. 1342K^T (AOHB00000000), *Leptospira terpstrae* sv. Hualin str. LT 11-33^T (Genbank: AOGW00000000), *Leptospira vanthielii* sv. Holland str. WaZ Holland (Genbank: AOGY00000000), *Leptospira weilii* sv. undetermined str. LNT 1234 (Genbank: AOHC00000000), *Leptospira wolbachii* sv. Codice str. CDC (Genbank: AOGZ00000000), *Leptospira wolffii* sv. undetermined str. Khorat-H2^T (Genbank: AKWX00000000), *Leptospira yanagawae* sv. Saopaulo str. Sao Paulo^T (Genbank: AOGX00000000).

Genes were considered homologs if they were bidirectional best hits [221, 222] using Basic Local Alignment Search Tool (BLAST) with cutoff values of 70% query coverage, e-values < $1e^{-3}$, and 30% identity.

4.4.6 Amino acid residue conservation analysis of study-identified nsSNV positions

Domain architecture analysis was performed on the protein sequences for LA_2704 (NP_712885.1), LA_2930 (713110.1), LA_2950 (NP_713130.1), LA_3455 (NP_713635.1), LA_3725 (NP_713905.1), and LA_3834 (NP_714014.1) using Simple Modular Architecture Research Tool (SMART) [223] and protein structure prediction server (PSIPRED) [175, 176], and represented graphically at *http://prosite.expasy.org/mydomains*.

Multiple sequence alignments (MSAs) of the homologs (defined by 70% query coverage, e-values < $1e^{-3}$, and 30% identity BLAST cutoffs) of each of these six genes were constructed by aligning sequences obtained from the Pathosystems Resource Integration Center database (*http://patricbrc.org*) using the CLUSTAL X alignment program freely available at *http://www.clustal.org/clustal2/#Download*. The accession numbers used in the alignments (Supplemental Table 3.1) are a representative collection of homolog sequences from each of the 20 species in the *Leptospira* genus in which homolog sequences could be identified. The LA_2704 alignment contained 40 homolog sequences, LA_2930 had 26 homologs, LA_2950 had 41 homologs, LA_3455 had 41 homologs, LA_3725 had 19 homologs, and LA_3834 had 45 homologs.

MSAs were then used to predict protein residue conservation based on Jensen-Shannon Divergence (JSD) [177]. Conservation scores were then graphed using Microsoft Excel (Redmond, WA).

The proportion of sequencing reads from the P8A strain coding for the nsSNV amino acid was compared with the proportion of the same mutant residues in homolog MSAs from the entire pan-*Leptospira* genome using a Fisher's exact test for each of the six genes. Results were considered statistically significant at $P < 0.05$.

4.4.7 Identification of potential ncRNAs in the L. interrogans *serovar Lai strain 56601 genome*

To identify novel ncRNA loci within the *L. interrogans* Lai strain 56601 genome, we first aligned the *L. interrogans* Lai 56601 (Genbank: AE010300), *L. kirschneri* Cynopteri 3522 C (Genbank: AHMN00000000), and *L. noguchii* Panama CZ214^T (Genbank: AKWY00000000) genomes using the progressive Cactus algorithm [224, 225]. The whole genome alignment was then used as input for RNAz (with default settings: –w 120 and –s 120) for prediction of structural RNAs [188] and then putative ncRNA loci identified and annotated using the nocoRNAc pipeline [189]. Predicted loci that could not be annotated using the Rfam database were considered potentially novel ncRNA genes.

4.5 Acknowledgements

Chapter 4, in full, has been accepted for publication for the material as it may appear in the American Society of Tropical Medicine and Hygiene, 2016. Victoria C. Corey and Jason S. Lehmann, Jessica N. Ricaldi, Joseph M. Vinetz, Elizabeth A. Winzeler, Michael A. Matthias, "Whole Genome Shotgun Sequencing Shows Selection on *Leptospira* Regulatory Proteins During *in vitro* Culture Attenuation". The dissertation author was the co-primary investigator and author of this paper.

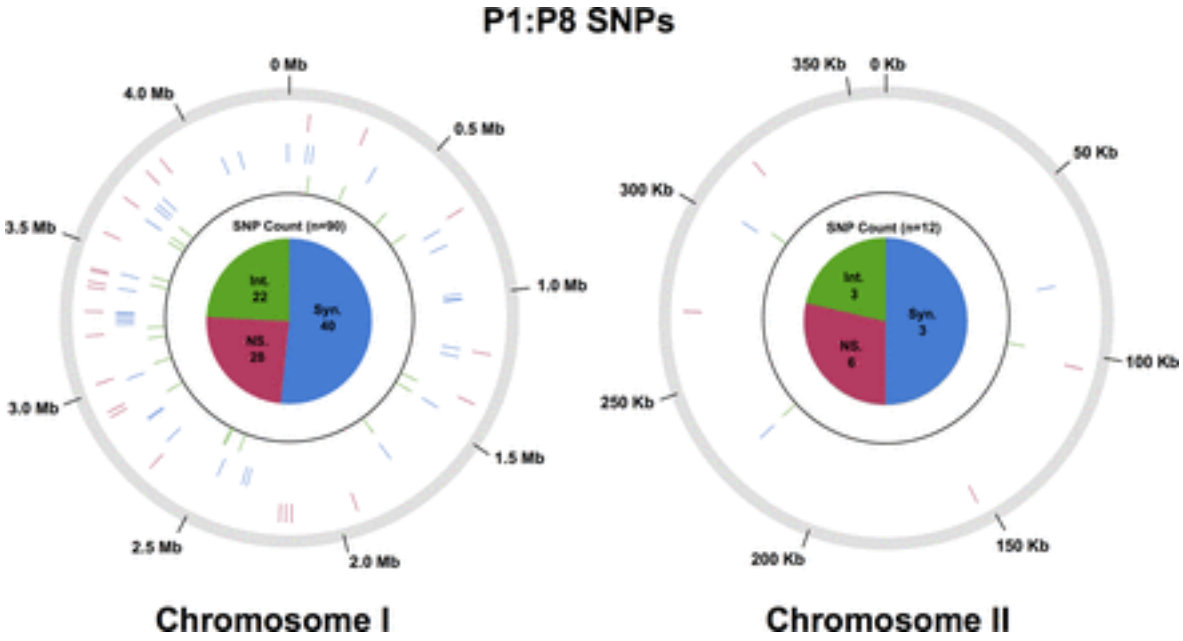


Figure 4.1 Genomic locations of single nucleotide variants (SNVs) changing allelic frequency from P1 to P8A Genomic location of study identified SNVs in the reference genome of Leptospira interrogans serovar Lai strain 56601 that significantly changed in allelic proportionality during culture attenuation from a virulent P1 strain to an attenuated P8A strain. Individual hash marks represent the genomic location of genes containing SNVs, and are color coded in concentric circles. Red = non-synonymous; blue = synonymous; green = intergenic. The total number of genes containing each type of SNV is represented by the pie chart in the center of each of the chromosome representations.

A

Protein	Function	Mutation	SNP Frequency Change	SNP Frequency Final	COG
LA_2704	two component response regulator	C311F	7.46%	18.75%	T
LA_2930	GGDEF family protein	Y94F	9.96%	13.66%	T
LA_3207	histidine kinase sensor protein	M856V	9.36%	20.93%	T
LA_3450	regulatory protein	L5F	5.92%	11.50%	T
LB_322	two-component hybrid sensor/regulator	S177P	5.60%	7.82%	T
LB_273	mcm2, methylmalonyl-CoA mutase	Y380F E394G	5.07% 6.80%	8.52% 23.46%	I
LA_2221	mgtA magnesium transport p-type atpase	M435L L440F L453F	6.66% 7.59% 8.40%	14.39% 18.49% 13.91%	P
LA_1392	rbsK ribokinase	V283G Y285F	5.40% 7.06%	7.49% 11.79%	G
LA_3455	glycerol uptake facilitator protein	S56P	6.31%	10.66%	G
LB_111	6-phosphofructokinase	T381R	7.60%	7.96%	G
LA_2950	hypothetical protein, PDZ serine protease	P81R	5.92%	8.06%	O
LA_3407	alanyl-tRNA synthetase	T53S	5.43%	7.42%	J
LA_3303	hypothetical protein, Ybbr like protein	T338K	5.78%	8.48%	S
LA_3725	cytoplasmic membrane protein	L1624V	6.83%	14.68%	S
LA_3834	hypothetical protein, beta propellar repeats	L8F	5.58%	8.33%	S

B

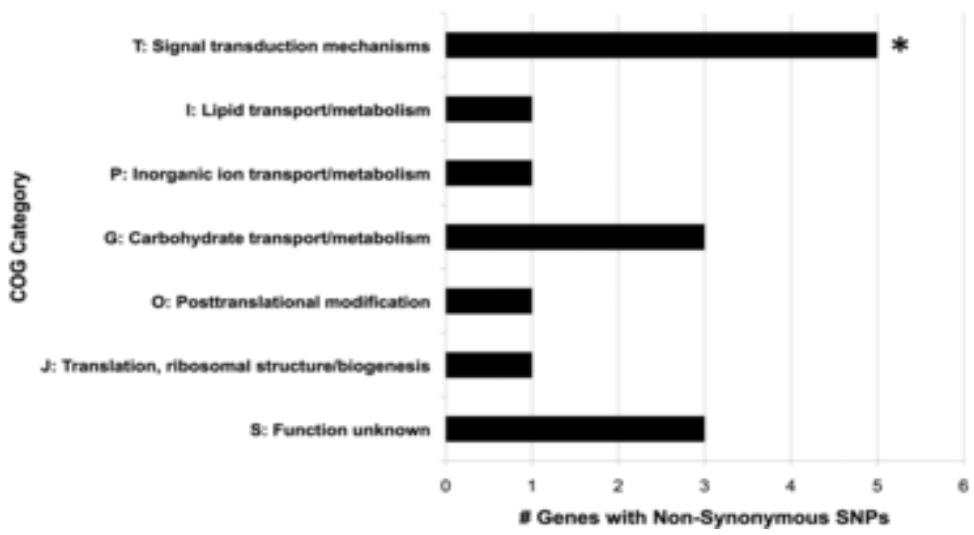
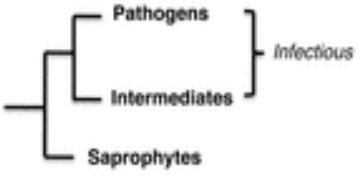


Figure 4.2 Clusters of orthologous group analysis of genes containing non-synonymous single nucleotide variants (nsSNVs) that increased allelic frequency from P1 to P8A The nsSNVs that increased allelic frequency during the attenuation of the virulent P1 strain of L. interrogans serovar Lai strain 56601 into the avirulent isogenic P8A strain are listed in (A). The nsSNV-containing genes in each group were then organized by clusters of orthologous groups (COG) category.45 Asterisks denote an enrichment of a particular COG category compared with genome-wide expected percentages of genes in each category (Fisher's exact test, P value given in figure) in (B). Total number of genes in Leptospira interrogans serovar Lai strain 56601 genome in represented COG categories: T = 233, I = 110, P = 134, G = 131, O = 118, J = 172, S = 853. Total number of predicted genes = 3,683.



Protein	Function	Upregulation	PSORTb	L. interrogans Copenhageni str. Fiocruz L1-130	L. kirschneri Cynopteri 3522 CT	L. noguchii Panama CZ 214T	L. alstoni serovar Pingchang str. 80-412	L. weilii serovar Ranarum str. ICFT	L. alexanderi Manhao 3 L 60T	L. borgpetersenii Javanica UI 09931	L. santarosai serovar Shermani str. 1342KT	L. kmetyi undetermined Bejo-Iso9T	L. fainei Hurstbridge BUT 6T	L. broomii undetermined 5399T	L. wolffii undetermined Khorat-H2T	L. licerasiae Varillal VAR 010	L. inadai Lyme 10T	L. weilbachii serovar Codice str. CDC	L. yanagawae serovar Saopaulo str. Sao Paulo	L. biflexa Patoc Patoc1	L. vanthielii serovar Holland str. Waz Holland	L. terpstrae serovar Hualin str. LT 11-33T	L. meyeri Hardjo Went 5
				P	P	P	P	P	P	P	P	P	I	I	I	I	I	S	S	S	S	S	S
LA_2704	two component response regulator	a	CM	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
LA_2930	GGDEF family protein	a	C	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
LA_2950	HtrA2, PDZ serine protease	a	UNK	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
LA_3455	glycerol uptake facilitator protein	a	CM	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
LA_3725	cytoplasmic membrane protein	a	CM	+	+	+	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-
LA_3834	putative lipoprotein	a,b,c	CM	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-

Figure 4.3 Homolog identification and characterization of potential virulence-associated genes in other *Leptospira* species Potential virulence-associated genes identified in this study include LA_2704 (NP_712885.1), LA_2930 (NP_713110.1), LA_2950 (NP_713130.1), LA_3455 (NP_713635.1), LA_3725 (NP_713905.1), and LA_3834 (NP_714014.1). Previous studies have shown these genes to be upregulated by Leptospira interrogans during exposure to host-like conditions (a = response to host innate immunity20; b = response to host physiological osmolarity,18 c = response to host cues during in vivo culture in intraperitoneal dialysis cassettes21). The PSORTb predicted subcellular locations of each of these proteins are listed. The presence of orthologous genes (defined as bidirectional best Basic Local Alignment Search Tool [BLAST] hits with minimum 70% query coverage, e-values < 1e−3, and 30% identity) was also determined for each of the 20 species in the Leptospira genus. P = pathogenic species; I = intermediate pathogens; S = saprophytic species). A schematic representation of these three clades of the genus was included for clarity.

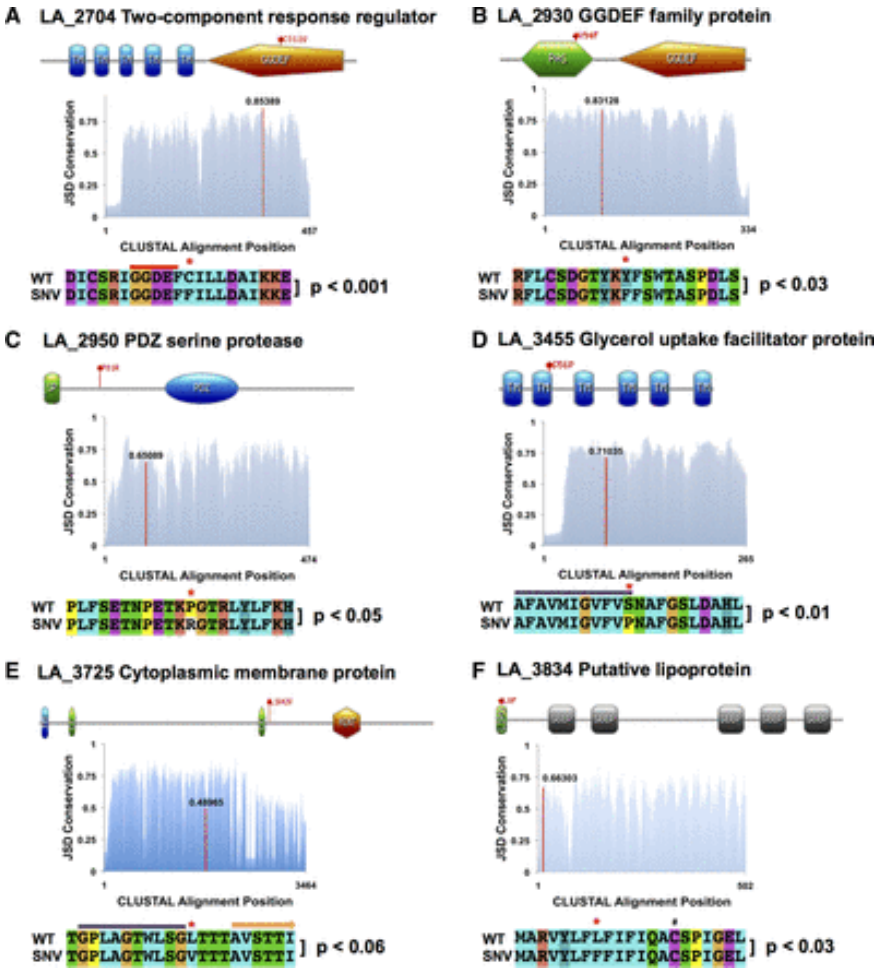


Figure 4.4 Amino acid conservation analysis at non-synonymous single nucleotide variants (nsSNV) positions identified gene homologs across the *Leptospira* genus Protein domain analysis was conducted for all genes and results are represented as diagrams at the top of panels A–E that include the nsSNV position for each gene (TM = transmembrane domain; PAS = Per-ARNT-Sim domain; GGDEF = diguanylate cyclases domain; SP = signal peptide; PDZ = PDZ serine protease domain; CC = coiled-coil domain; HINT = hedgehog intein domain; LIPO = Leptospira lipobox; SBBP = seven-bladed beta propeller domain). In addition, a Jensen-Shannon Divergence (JSD) estimate of amino acid residue conservation is represented graphically with the nsSNV residue highlighted as a red line with the conservation score indicated above it (scores above 0.8 indicate high conservation, those below 0.4 indicate disorder). Finally, at the bottom of each panel, an alignment schematic with the nsSNV position highlighted by a red asterisk is presented with the probability from a Fisher's exact test comparison of the number of P8A sequencing reads coding for the mutant amino acid to the genome-wide prevalence of that same residue in homologs of that particular gene. In (A), the red line above the alignment indicates the position of the catalytic residue of the protein. In (D), the purple line represents a conserved α helix. In (E), the purple line represents a predicted α helix, and the orange arrow represents a predicted beta-sheet secondary structure. In (F), the hash mark denotes the cysteine residue that is lipidated.

Table 4.1 Genome alignment statistics for *Leptospira interrogans* serovar Lai strains P1 and P8A

Sample	Median insert size (bp)	Total reads	Aligned reads (%)	Mean coverage	% Bases > 20X	% Bases > 40X	% Bases > 100X	% Bases > 130X	% Bases > 150X
P1	113	15,583,466	15,492,436 (99.4)	250.49	100	99.9	98.8	94.9	89.6
P8	129	15,668,032	15,651,273 (99.9)	263.53	100	99.9	98.4	95.5	91.8

Table 4.2 Predicted ncRNAs in *Leptospira interrogans* serovar Lai In silico predicted ncRNAs in the *L. interrogans* serovar Lai genome. The 14 predicted ncRNAs are listed by chromosome, start/stop positions, as well as the DNA strand the site resides on. Stress-induced duplex destabilization (SIDD) values near zero are highly destabilized states that promote helicase action and transcription. Terminator confidence is a percentage of certainty of transcription stop sites.

Chromosome	Start (bp)	Stop (bp)	Strand	SIDD value	Terminator confidence
CI	235,751	235,834	−	−0.5	73
CI	422,027	422,128	−	3.5	100
CI	538,202	538,259	+	2.5	72
CI	548,190	548,249	−	3.2	100
CI	1,181,513	1,181,570	+	2.0	74
CI	2,172,984	2,173,177	−	−0.2	76
CI	2,410,576	2,410,725	−	−0.3	78
CI	2,545,112	2,545,233	−	2.9	77
CI	2,823,520	2,823,576	−	−0.5	100
CI	2,876,638	2,876,810	−	−0.5	80
CI	2,935,474	2,935,522	−	−1.0	70
CI	3,382,029	3,382,080	+	2.2	89
CI	4,152,345	4,152,404	+	2.6	79
CII	73,532	73,627	−	2.5	86

Chapter 5

Conclusions and Future Perspectives

Treating any disease or pathogen requires a solid understanding of the organism, and the pathogenesis of the disease it produces. For virulence, understanding which genes are involved can aid in identifying antigens for vaccine development. For resistance development, identifying the factors involved in resistance and which genes have a larger constraint in mutational frequencies can help in the future development of therapeutics and provide a framework in therapy regimens. This dissertation focused on multiple methods used to gain this comprehensive understanding, analyzing both resistance development and virulence loss.

In Chapters 2 and 3, we first focus on *Plasmodium falciparum* and examine resistance development in the context of identifying novel antimalarial targets. In a systematic analysis consisting of 50 compounds with antimalarial activity, we assessed possible targetable genes pathways as well as general mechanisms of resistance. Focusing first on the effect of pre-existing resistance, our results demonstrate that parasites resistant to current antimalarials pose minimal pre-existing cross-resistance to this novel compound set. These results imply that there is still a large possibility other druggable pathways have yet to be identified and may be exploited in antimalarial therapeutic development. We next examined the process of resistance development, identifying which

compounds yielded resistant parasites after culturing and how quickly resistance developed. Out of the 50 compounds, we successfully generated resistant mutants against 23 novel antimalarials. Examining various physiochemical properties that may be predictive in resistance generation, we found compounds with a faster killing-rate were significantly less successful in generating resistant parasites. As fast-acting compounds are already desirable in the clinic given their ability to rapidly stop disease progression and avoid severe complications, these results only further support the need to have fast-acting compounds in the clinical development pipeline.

Analyzing the resistant strain sets generated in this study using whole genome sequencing, we identified the potential resistance-causing variants in 21 of the resistant strain sets. Some of the casual variants were located in genes that have already been established as compound targets, including *pfcrt*, *cytochrome bc1,* and *PI4K*. Encouragingly, in addition to previously known targets, we identified six new genes as potential compound targets: RND transporter (Pf3D7_0107500), cathepsin dipeptidyl peptidase (Pf3D7_1116700), phenylalanine-tRNA ligase (Pf3D7_0109800), FHA domain protein (Pf3D7_0909700), phd finger protein (Pf3D7_0310200), and acyl-CoA synthetase 10 and 11 (Pf3D7_0525100, PF3D7_1238800). While further work needs to be done to determine if

these genes are direct targets of the compounds or are general resistance mechanisms, the results are encouraging and suggest other unknown pathways can still be exploited in antimalarial therapy development.

Chapter 3 in this dissertation went on to further characterize *pfcarl*, one of the genes identified in Chapter 2. As PfCARL has been previously identified as a potential target for KAF156, an imidazolopiperazine currently in Stage II clinical trials, it was intriguing that resistance selection with MMV007564, a benzimidazolyl piperidine, resulted in PfCARL-mutated cultures. Our results determine that PfCARL is in fact not a common target between the two compound classes, but most likely a common resistance mechanism. Examining the effects of MMV007564 and GNF179 (another imidazolopiperazine) on the parasitic asexual life cycle, our experimentation determined the two compound classes are most effective at different instances in the cell cycle. Early asexual blood ring stages were most susceptible to GNF179, whereas MMV007564 was more effective during the late ring to trophozoite stage, implying the two compound classes have different cellular targets. In addition to acting at different cell cycle windows of the asexual blood stage, KAF156 was found to be highly effective against both liver stage and sexual blood stage parasites, whereas the benzimidazolyl piperidines were only

effective against the asexual blood stage. Given that developing resistant parasites to MMV007564 or KAF156 can be done relatively quickly *in vitro* (15-17 days), it is possible resistance may develop quickly against KAF156 *in vivo*, and therefore these results identify a crucial need to supply KAF156 with a partner drug if and when it is used in the clinic in order to minimize resistance development.

Chapter 4 went to examine another pathogen, *Leptospira interrogans*, using a reverse approach where we examined genomic changes in a polyclonal population after a loss of virulence in order to identify possible virulence-associated genes. As virulent *Leptospira* strains are known to be extremely slow growing *in vitro*, allelic changes observed when culturing the parasite without selective pressure were most likely the result of rapid growth selection and in a tradeoff with virulence. Examining the changes in allelic frequency in non-synonymous single nucleotide variants (nsSNVs), our experimentation identified 15 genes containing nsSNVs that increased in frequency from the virulent to the attenuated strain. Analyzing this gene set for bias in a particular biological function, we found there was a strong enrichment (5/15) for genes involved in signal transduction mechanisms. Additionally, six genes identified in our set (LA_2704, LA_2930, LA_2950, LA_3455, LA_3725, and LA_3834) were previously reported to be up-regulated when *L. interrogans*

was examined under surrogate *in vivo* conditions, which tested the parasite's response to various stresses, including temperature, iron depletion, and exposure to host innate immune cells. Overall, our results suggest the ultimate mechanism of attenuation was the additive effect of multiple mutant alleles, each subtracting from overall population fitness *in vivo*. This work does have limitations, and further work need to be conducted to identify to individual gene contribution to virulence and *Leptospira* pathogenesis, though some has already been underway by conducting a comparative genomics analysis across Leptospira species [226]. The work described here, however, provides a solid first step in defining virulence-based genes in *Leptospira*.

Whole genome sequencing has only become affordable and accessible for high throughput experimentation in the last decade. With the revolution of this technology we now have vital tools available that can help answer questions we have struggled with before. Coupling whole genome sequencing to two evolutionary-based approaches, this dissertation presents examples where this technology was able to examine both virulence and resistance in both eukaryotic (*Plasmodium*) and prokaryotic (*Leptospira*) pathogens where additional genomic information is needed. Additionally, for *Plasmodium*, these methods provided information in a context that may be used for future therapy

design. For antimalarials, we are in a race against the inevitable development of drug resistance leading to reduced clinical effectiveness. Analyzing the organism in the context of drug resistance, we have been able to not only gain further knowledge of the *Plasmodium* genome, but also identify chemical properties associated with drug development. This information combined with therapeutic regimens may help select compounds with a longer lifespan in the clinic. As pathogens such as *Plasmodium* and *Leptospira* continue to emerge in the clinical setting, it will be vital to understand the processes underlying the virulence and adaptation of the parasite if we are to gain an upper hand in treatment and therapeutic development.

REFERENCES

1. Casadevall, A. & Pirofski, L.A., Microbiology: Ditch the term pathogen. *Nature*, 2014. **516**(7530): p. 165-166.

2. Alberts, B., Johnson, A. & Lewis, J., Molecular Biology of the Cell. 4th Edition. *New York: Garland Science*, 2002. **Introduction to Pathogens**.

3. Joubert, Y. & Joubert, F., A structural annotation resource for the selection of putative target proteins in the malaria parasite. *Malar J*, 2008. **7**: p. 90.

4. Zuerner, R.L., Host response to leptospira infection. *Curr Top Microbiol Immunol*, 2015. **387**: p. 223-250.

5. Institute of Medicine (US) Committee on the Economics of Antimalarial Drugs, *Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance*, ed. C.P. Kenneth J. Arrow, and Hellen Gelband. 2004: National Academies Press (US).

6. Miller, R.L., Ikram, S., Armelagos, G.J., Walker, R., Harer, W.B., Shiff, C.J., Baggett, D., Carrigan, M. & Maret, S.M., Diagnosis of Plasmodium falciparum infections in mummies using the rapid manual ParaSight-F test. *Trans R Soc Trop Med Hyg*, 1994. **88**(1): p. 31-32.

7. Carter, R. & Mendis, K.N., Evolutionary and Historical Aspects of the Burden of Malaria. *Clinical Microbiology Reviews*, 2002. **15**(4): p. 564-594.

8. Sfetcu, N., *Health & Drugs: Disease, Prescription & Medication*. 2014: Nicolae Sfetcu.

9. Wells, T.N., Hooft van Huijsduijnen, R. & Van Voorhis, W.C., Malaria medicines: a glass half full? *Nat Rev Drug Discov*, 2015. **14**(6): p. 424-442.

10. Guttmann, P. & P., E., On the effect of methylene blue on malaria. *Berliner Klinische Wochenschrift*, 1891. **28**: p. 953-956.

11. Krafts, K., Hempelmann, E. & Skorska-Stania, A., From methylene blue to chloroquine: a brief review of the development of an antimalarial therapy. *Parasitol Res*, 2012. **111**(1): p. 1-6.

12. Elderfield, R.C., Mertel, H.E., Mitch, R.T., Wempen, I.M. & Werble, E., Synthesis of Primaquine and Certain of its Analogs1. *Journal of the American Chemical Society*, 1955. **77**(18): p. 4816-4819.

13. Vale, N., Moreira, R. & Gomes, P., Primaquine revisited six decades after its discovery. *Eur J Med Chem*, 2009. **44**(3): p. 937-953.

14. World Health Organization, World Malaria Report. *WHO, Geneva*, 2015.

15. Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., Alvarado, M., Anderson, H.R., Anderson, L.M., Andrews, K.G., Atkinson, C., Baddour, L.M., Barker-Collo, S., Bartels, D.H., Bell, M.L., Benjamin, E.J., Bennett, D., Bhalla, K., Bikbov, B., Bin Abdulhak, A., Birbeck, G., Blyth, F., Bolliger, I., Boufous, S., Bucello, C., Burch, M., Burney, P., Carapetis, J., Chen, H., Chou, D., Chugh, S.S., Coffeng, L.E., Colan, S.D., Colquhoun, S., Colson, K.E., Condon, J., Connor, M.D., Cooper, L.T., Corriere, M., Cortinovis, M., de Vaccaro, K.C., Couser, W., Cowie, B.C., Criqui, M.H., Cross, M., Dabhadkar, K.C., Dahodwala, N., De Leo, D., Degenhardt, L., Delossantos, A., Denenberg, J., Des Jarlais, D.C., Dharmaratne, S.D., Dorsey, E.R., Driscoll, T., Duber, H., Ebel, B., Erwin, P.J., Espindola, P., Ezzati, M., Feigin, V., Flaxman, A.D., Forouzanfar, M.H., Fowkes, F.G., Franklin, R., Fransen, M., Freeman, M.K., Gabriel, S.E., Gakidou, E., Gaspari, F., Gillum, R.F., Gonzalez-Medina, D., Halasa, Y.A., Haring, D., Harrison, J.E., Havmoeller, R., Hay, R.J., Hoen, B., Hotez, P.J., Hoy, D., Jacobsen, K.H., James, S.L., Jasrasaria, R., Jayaraman, S., Johns, N., Karthikeyan, G., Kassebaum, N., Keren, A., Khoo, J.P., Knowlton, L.M., Kobusingye, O., Koranteng, A., Krishnamurthi, R., Lipnick, M., Lipshultz, S.E., Ohno, S.L., Mabweijano, J., MacIntyre, M.F., Mallinger, L., March, L., Marks, G.B., Marks, R., Matsumori, A., Matzopoulos, R., Mayosi, B.M., McAnulty, J.H., McDermott, M.M., McGrath, J., Mensah, G.A., Merriman, T.R., Michaud, C., Miller, M., Miller, T.R., Mock, C., Mocumbi, A.O., Mokdad, A.A., Moran, A., Mulholland, K., Nair, M.N., Naldi, L., Narayan, K.M., Nasseri, K., Norman, P., O'Donnell, M., Omer, S.B., Ortblad, K., Osborne, R., Ozgediz, D., Pahari, B., Pandian, J.D., Rivero, A.P., Padilla, R.P., Perez-Ruiz, F., Perico, N., Phillips, D., Pierce, K., Pope, C.A., 3rd, Porrini, E., Pourmalek, F., Raju, M., Ranganathan,

D., Rehm, J.T., Rein, D.B., Remuzzi, G., Rivara, F.P., Roberts, T., De Leon, F.R., Rosenfeld, L.C., Rushton, L., Sacco, R.L., Salomon, J.A., Sampson, U., Sanman, E., Schwebel, D.C., Segui-Gomez, M., Shepard, D.S., Singh, D., Singleton, J., Sliwa, K., Smith, E., Steer, A., Taylor, J.A., Thomas, B., Tleyjeh, I.M., Towbin, J.A., Truelsen, T., Undurraga, E.A., Venketasubramanian, N., Vijayakumar, L., Vos, T., Wagner, G.R., Wang, M., Wang, W., Watt, K., Weinstock, M.A., Weintraub, R., Wilkinson, J.D., Woolf, A.D., Wulf, S., Yeh, P.H., Yip, P., Zabetian, A., Zheng, Z.J., Lopez, A.D., Murray, C.J., AlMazroa, M.A. & Memish, Z.A., Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 2012. **380**(9859): p. 2095-2128.

16. Sachs, J. & Malaney, P., The economic and social burden of malaria. *Nature*, 2002. **415**(6872): p. 680-685.

17. Tilley, L., Dixon, M.W. & Kirk, K., The Plasmodium falciparum-infected red blood cell. *Int J Biochem Cell Biol*, 2011. **43**(6): p. 839-842.

18. Das, A., Sharma, M., Gupta, B. & Dash, A.P., Plasmodium falciparum and Plasmodium vivax: so similar, yet very different. *Parasitol Res*, 2009. **105**(4): p. 1169-1171.

19. Prudencio, M., Rodriguez, A. & Mota, M.M., The silent path to thousands of merozoites: the Plasmodium liver stage. *Nat Rev Microbiol*, 2006. **4**(11): p. 849-856.

20. Miller, L.H., Ackerman, H.C., Su, X.Z. & Wellems, T.E., Malaria biology and disease pathogenesis: insights for new treatments. *Nat Med*, 2013. **19**(2): p. 156-167.

21. Tangpukdee, N., Duangdee, C., Wilairatana, P. & Krudsood, S., Malaria Diagnosis: A Brief Review. *Korean J Parasitol*, 2009. **47**(2): p. 93-102.

22. Thonsranoi, K., Glaharn, S., Punsawad, C., Chaisri, U., Krudsood, S. & Viriyavejakul, P., Increased synapsin I expression in cerebral malaria. *Int J Clin Exp Pathol*, 2015. **8**(11): p. 13996-14004.

23. Eksi, S., Morahan, B.J., Haile, Y., Furuya, T., Jiang, H., Ali, O., Xu, H., Kiattibutr, K., Suri, A., Czesny, B., Adeyemo, A., Myers, T.G., Sattabongkot, J., Su, X.Z. & Williamson, K.C., Plasmodium falciparum gametocyte development 1 (Pfgdv1) and gametocytogenesis

early gene identification and commitment to sexual development. *PLoS Pathog*, 2012. **8**(10): p. e1002964.

24. Gardiner, D.L. & Trenholme, K.R., Plasmodium falciparum gametocytes: playing hide and seek. *Annals of Translational Medicine*, 2015. **3**(4): p. 45.

25. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *The Lancet*, 2015. **386**(9988): p. 31-45.

26. Flannery, E.L., Chatterjee, A.K. & Winzeler, E.A., Antimalarial drug discovery - approaches and progress towards new medicines. *Nat Rev Microbiol*, 2013. **11**(12): p. 849-862.

27. Galatas, B., Bassat, Q. & Mayor, A., Malaria Parasites in the Asymptomatic: Looking for the Hay in the Haystack. *Trends in Parasitology*.

28. Malaria Elimination Initiative, Country Briefing. *MEI, UCSF*, 2015.

29. Tun, K.M., Imwong, M., Lwin, K.M., Win, A.A., Hlaing, T.M., Hlaing, T., Lin, K., Kyaw, M.P., Plewes, K., Faiz, M.A., Dhorda, M., Cheah, P.Y., Pukrittayakamee, S., Ashley, E.A., Anderson, T.J.C., Nair, S., McDew-White, M., Flegg, J.A., Grist, E.P.M., Guerin, P., Maude, R.J., Smithuis, F., Dondorp, A.M., Day, N.P.J., Nosten, F., White, N.J. & Woodrow, C.J., Spread of artemisinin-resistant Plasmodium falciparum in Myanmar: a cross-sectional survey of the K13 molecular marker. *The Lancet Infectious Diseases*. **15**(4): p. 415-421.

30. Noedl, H., Se, Y., Schaecher, K., Smith, B.L., Socheat, D. & Fukuda, M.M., Evidence of artemisinin-resistant malaria in western Cambodia. *N Engl J Med*, 2008. **359**(24): p. 2619-2620.

31. Dondorp, A.M., Nosten, F., Yi, P., Das, D., Phyo, A.P., Tarning, J., Lwin, K.M., Ariey, F., Hanpithakpong, W., Lee, S.J., Ringwald, P., Silamut, K., Imwong, M., Chotivanich, K., Lim, P., Herdman, T., An, S.S., Yeung, S., Singhasivanon, P., Day, N.P.J., Lindegardh, N., Socheat, D. & White, N.J., Artemisinin Resistance in Plasmodium falciparum Malaria. *New England Journal of Medicine*, 2009. **361**(5): p. 455-467.

32. Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R.,

Duong, S., Sreng, S., Suon, S., Chuor, C.M., Bout, D.M., Menard, S., Rogers, W.O., Genton, B., Fandeur, T., Miotto, O., Ringwald, P., Le Bras, J., Berry, A., Barale, J.C., Fairhurst, R.M., Benoit-Vical, F., Mercereau-Puijalon, O. & Menard, D., A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature*, 2014. **505**(7481): p. 50-55.

33. Phyo, A.P., Nkhoma, S., Stepniewska, K., Ashley, E.A., Nair, S., McGready, R., Ier Moo, C., Al-Saai, S., Dondorp, A.M., Lwin, K.M., Singhasivanon, P., Day, N.P., White, N.J., Anderson, T.J. & Nosten, F., Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. *Lancet*, 2012. **379**(9830): p. 1960-1966.

34. Noedl, H., Se, Y., Sriwichai, S., Schaecher, K., Teja-Isavadharm, P., Smith, B., Rutvisuttinunt, W., Bethell, D., Surasri, S., Fukuda, M.M., Socheat, D. & Thap, L.C., Artemisinin Resistance in Cambodia: A Clinical Trial Designed to Address an Emerging Problem in Southeast Asia. *Clinical Infectious Diseases*, 2010. **51**(11): p. e82-e89.

35. Mita, T. & Tanabe, K., Evolution of Plasmodium falciparum drug resistance: implications for the development and containment of artemisinin resistance. *Jpn J Infect Dis*, 2012. **65**(6): p. 465-475.

36. Borrmann, S., Sasi, P., Mwai, L., Bashraheil, M., Abdallah, A., Muriithi, S., Fruhauf, H., Schaub, B., Pfeil, J., Peshu, J., Hanpithakpong, W., Rippert, A., Juma, E., Tsofa, B., Mosobo, M., Lowe, B., Osier, F., Fegan, G., Lindegardh, N., Nzila, A., Peshu, N., Mackinnon, M. & Marsh, K., Declining responsiveness of Plasmodium falciparum infections to artemisinin-based combination treatments on the Kenyan coast. *PLoS One*, 2011. **6**(11): p. e26005.

37. Vreden, S.G., Jitan, J.K., Bansie, R.D. & Adhin, M.R., Evidence of an increased incidence of day 3 parasitaemia in Suriname: an indicator of the emerging resistance of Plasmodium falciparum to artemether. *Mem Inst Oswaldo Cruz*, 2013. **108**(8): p. 968-973.

38. Plouffe, D., Brinker, A., McNamara, C., Henson, K., Kato, N., Kuhen, K., Nagle, A., Adrián, F., Matzen, J.T., Anderson, P., Nam, T.-g., Gray, N.S., Chatterjee, A., Janes, J., Yan, S.F., Trager, R., Caldwell, J.S., Schultz, P.G., Zhou, Y. & Winzeler, E.A., In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. *Proceedings of the National Academy of Sciences*, 2008. **105**(26): p. 9059-9064.

39. Gamo, F.-J., Sanz, L.M., Vidal, J., de Cozar, C., Alvarez, E., Lavandera, J.-L., Vanderwall, D.E., Green, D.V.S., Kumar, V., Hasan, S., Brown, J.R., Peishoff, C.E., Cardon, L.R. & Garcia-Bustos, J.F., Thousands of chemical starting points for antimalarial lead identification. *Nature*, 2010. **465**(7296): p. 305-310.

40. Guiguemde, W.A., Shelat, A.A., Bouck, D., Duffy, S., Crowther, G.J., Davis, P.H., Smithson, D.C., Connelly, M., Clark, J., Zhu, F., Jiménez-Díaz, M.B., Martinez, M.S., Wilson, E.B., Tripathi, A.K., Gut, J., Sharlow, E.R., Bathurst, I., Mazouni, F.E., Fowble, J.W., Forquer, I., McGinley, P.L., Castro, S., Angulo-Barturen, I., Ferrer, S., Rosenthal, P.J., DeRisi, J.L., Sullivan, D.J., Lazo, J.S., Roos, D.S., Riscoe, M.K., Phillips, M.A., Rathod, P.K., Van Voorhis, W.C., Avery, V.M. & Guy, R.K., Chemical genetics of Plasmodium falciparum. *Nature*, 2010. **465**(7296): p. 311-315.

41. Bharti, A.R., Nally, J.E., Ricaldi, J.N., Matthias, M.A., Diaz, M.M., Lovett, M.A., Levett, P.N., Gilman, R.H., Willig, M.R., Gotuzzo, E. & Vinetz, J.M., Leptospirosis: a zoonotic disease of global importance. *Lancet Infect Dis*, 2003. **3**.

42. Abela-Ridder, B., Sikkema, R. & Hartskeerl, R.A., Estimating the burden of human leptospirosis. *International Journal of Antimicrobial Agents*, 2010. **36, Supplement 1**: p. S5-S7.

43. Ko, A.I., Reis, M.G., Dourado, C.M.R., Johnson Jr, W.D. & Riley, L.W., Urban epidemic of severe leptospirosis in Brazil. *The Lancet*, 1999. **354**(9181): p. 820-825.

44. Paster, B.J., Dewhirst, F.E., Weisburg, W.G., Tordoff, L.A., Fraser, G.J., Hespell, R.B., Stanton, T.B., Zablen, L., Mandelco, L. & Woese, C.R., Phylogenetic analysis of the spirochetes. *J Bacteriol*, 1991. **173**(19): p. 6101-6109.

45. Ko, A.I., Goarant, C. & Picardeau, M., Leptospira: the dawn of the molecular genetics era for an emerging zoonotic pathogen. *Nat Rev Micro*, 2009. **7**(10): p. 736-747.

46. Lehmann, J.S., Matthias, M.A., Vinetz, J.M. & Fouts, D.E., Leptospiral Pathogenomics. *Pathogens*, 2014. **3**(2): p. 280-308.

47. Levett, P.N., Leptospirosis. *Clin Microbio Rev*, 2001. **14**.

48. Ricaldi, J.N., Fouts, D.E., Selengut, J.D., Harkins, D.M., Patra, K.P., Moreno, A., Lehmann, J.S., Purushe, J., Sanka, R., Torres, M., Webster, N.J., Vinetz, J.M. & Matthias, M.A., Whole Genome Analysis of Leptospira licerasiae Provides Insight into Leptospiral Evolution and Pathogenicity. *PLoS Negl Trop Dis*, 2012. **6**(10): p. e1853.

49. Matthias, M.A., Diaz, M.M., Campos, K.J., Calderon, M., Willig, M.R., Pacheco, V., Gotuzzo, E., Gilman, R.H. & Vinetz, J.M., Diversity of bat-associated Leptospira in the Peruvian Amazon inferred by bayesian phylogenetic analysis of 16S ribosomal DNA sequences. *Am J Trop Med Hyg*, 2005. **73**(5): p. 964-974.

50. Brenner, D.J., Kaufmann, A.F., Sulzer, K.R., Steigerwalt, A.G., Rogers, F.C. & Weyant, R.S., Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with a proposal for Leptospira alexanderi sp. nov. and four new Leptospira genomospecies. *International Journal of Systematic and Evolutionary Microbiology*, 1999. **49**(2): p. 839-858.

51. Bourhy, P., Collet, L., Brisse, S. & Picardeau, M., Leptospira mayottensis sp. nov., a pathogenic species of the genus Leptospira isolated from humans. *International Journal of Systematic and Evolutionary Microbiology*, 2014. **64**(12): p. 4061-4067.

52. Gouveia, E.L., Metcalfe, J., F. de Carvalho, A.L., Aires, T.S.F., Villasboas-Bisneto, J.C., Queirroz, A., Santos, A.C., Salgado, K., Reis, M.G. & Ko, A.I., Leptospirosis-associated Severe Pulmonary Hemorrhagic Syndrome, Salvador, Brazil. *Emerging Infectious Disease journal*, 2008. **14**(3): p. 505.

53. Faine, S., Adher, B., Bloin, C. & Perolat, P., *Leptospira and leptospirosis*. 1999, Australia: MedSci.

54. Lourdault, K., Aviat, F. & Picardeau, M., Use of quantitative real-time PCR for studying the dissemination of Leptospira interrogans in the guinea pig infection model of leptospirosis. *J Med Microbiol*, 2009. **58**(Pt 5): p. 648-655.

55. Ganoza, C.A., Matthias, M.A., Saito, M., Cespedes, M., Gotuzzo, E. & Vinetz, J.M., Asymptomatic Renal Colonization of Humans in the Peruvian Amazon by Leptospira. *PLoS Negl Trop Dis*, 2010. **4**(2).

56. Ren, S.X., Fu, G., Jiang, X.G., Zeng, R., Miao, Y.G., Xu, H., Zhang, Y.X., Xiong, H., Lu, G., Lu, L.F., Jiang, H.Q., Jia, J., Tu, Y.F., Jiang, J.X., Gu,

W.Y., Zhang, Y.Q., Cai, Z., Sheng, H.H., Yin, H.F., Zhang, Y., Zhu, G.F., Wan, M., Huang, H.L., Qian, Z., Wang, S.Y., Ma, W., Yao, Z.J., Shen, Y., Qiang, B.Q., Xia, Q.C., Guo, X.K., Danchin, A., Saint Girons, I., Somerville, R.L., Wen, Y.M., Shi, M.H., Chen, Z., Xu, J.G. & Zhao, G.P., Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing. *Nature*, 2003. **422**(6934): p. 888-893.

57. Nascimento, A.L., Ko, A.I., Martins, E.A., Monteiro-Vitorello, C.B., Ho, P.L., Haake, D.A., Verjovski-Almeida, S., Hartskeerl, R.A., Marques, M.V., Oliveira, M.C., Menck, C.F., Leite, L.C., Carrer, H., Coutinho, L.L., Degrave, W.M., Dellagostin, O.A., El-Dorry, H., Ferro, E.S., Ferro, M.I., Furlan, L.R., Gamberini, M., Giglioti, E.A., Goes-Neto, A., Goldman, G.H., Goldman, M.H., Harakava, R., Jeronimo, S.M., Junqueira-de-Azevedo, I.L., Kimura, E.T., Kuramae, E.E., Lemos, E.G., Lemos, M.V., Marino, C.L., Nunes, L.R., de Oliveira, R.C., Pereira, G.G., Reis, M.S., Schriefer, A., Siqueira, W.J., Sommer, P., Tsai, S.M., Simpson, A.J., Ferro, J.A., Camargo, L.E., Kitajima, J.P., Setubal, J.C. & Van Sluys, M.A., Comparative genomics of two Leptospira interrogans serovars reveals novel insights into physiology and pathogenesis. *J Bacteriol*, 2004. **186**(7): p. 2164-2172.

58. Bulach, D.M., Zuerner, R.L., Wilson, P., Seemann, T., McGrath, A., Cullen, P.A., Davis, J., Johnson, M., Kuczek, E., Alt, D.P., Peterson-Burch, B., Coppel, R.L., Rood, J.I., Davies, J.K. & Adler, B., Genome reduction in Leptospira borgpetersenii reflects limited transmission potential. *Proc Natl Acad Sci U S A*, 2006. **103**(39): p. 14560-14565.

59. Picardeau, M., Bulach, D.M., Bouchier, C., Zuerner, R.L., Zidane, N., Wilson, P.J., Creno, S., Kuczek, E.S., Bommezzadri, S., Davis, J.C., McGrath, A., Johnson, M.J., Boursaux-Eude, C., Seemann, T., Rouy, Z., Coppel, R.L., Rood, J.I., Lajus, A., Davies, J.K., Medigue, C. & Adler, B., Genome sequence of the saprophyte Leptospira biflexa provides insights into the evolution of Leptospira and the pathogenesis of leptospirosis. *PLoS One*, 2008. **3**(2): p. e1607.

60. Human Leptospirosis: guidance for diagnosis, surveillance, and control. *Geneva, World Health Organization/International Leptospirosis Society*, 2003.

61. Vaughan, A.M., Pinapati, R.S., Cheeseman, I.H., Camargo, N., Fishbaugher, M., Checkley, L.A., Nair, S., Hutyra, C.A., Nosten, F.H., Anderson, T.J., Ferdig, M.T. & Kappe, S.H., Plasmodium falciparum

genetic crosses in a humanized mouse model. *Nat Methods*, 2015. **12**(7): p. 631-633.

62. 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*, 1991. **88**(8): p. 3382-3386.

63. Flannery, E.L., Fidock, D.A. & Winzeler, E.A., Using genetic methods to define the targets of compounds with antimalarial activity. *J Med Chem*, 2013. **56**(20): p. 7761-7771.

64. Ross, L.S., Gamo, F.J., Lafuente-Monasterio, M., Singh, O.M.P., Rowland, P., Wiegand, R.C. & Wirth, D.F., In Vitro Resistance Selections for Plasmodium falciparum Dihydroorotate Dehydrogenase Inhibitors Give Mutants with Multiple Point Mutations in the Drug-binding Site and Altered Growth. *J Biol Chem*, 2014. **289**(26): p. 17980-17995.

65. Srivastava, I.K., Morrisey, J.M., Darrouzet, E., Daldal, F. & Vaidya, A.B., Resistance mutations reveal the atovaquone-binding domain of cytochrome b in malaria parasites. *Mol Microbiol*, 1999. **33**(4): p. 704-711.

66. Korsinczky, M., Chen, N., Kotecka, B., Saul, A., Rieckmann, K. & Cheng, Q., Mutations in Plasmodium falciparum cytochrome b that are associated with atovaquone resistance are located at a putative drug-binding site. *Antimicrob Agents Chemother*, 2000. **44**(8): p. 2100-2108.

67. Peterson, D.S., Milhous, W.K. & Wellems, T.E., Molecular basis of differential resistance to cycloguanil and pyrimethamine in Plasmodium falciparum malaria. *Proc Natl Acad Sci U S A*, 1990. **87**(8): p. 3018-3022.

68. Wang, P., Read, M., Sims, P.F. & Hyde, J.E., Sulfadoxine resistance in the human malaria parasite Plasmodium falciparum is determined by mutations in dihydropteroate synthetase and an additional factor associated with folate utilization. *Mol Microbiol*, 1997. **23**(5): p. 979-986.

69. Wilson, C.M., Serrano, A.E., Wasley, A., Bogenschutz, M.P., Shankar, A.H. & Wirth, D.F., Amplification of a gene related to mammalian mdr genes in drug-resistant Plasmodium falciparum. *Science*, 1989. **244**.

70. Djimde, A., Doumbo, O.K., Cortese, J.F., Kayentao, K., Doumbo, S., Diourte, Y., Coulibaly, D., Dicko, A., Su, X.Z., Nomura, T., Fidock, D.A., Wellems, T.E. & Plowe, C.V., A molecular marker for chloroquine-resistant falciparum malaria. *N Engl J Med*, 2001. **344**(4): p. 257-263.

71. Bourhy, P., Louvel, H., Saint Girons, I. & Picardeau, M., Random insertional mutagenesis of Leptospira interrogans, the agent of leptospirosis, using a mariner transposon. *J Bacteriol*, 2005. **187**(9): p. 3255-3258.

72. Murray, G.L., Morel, V., Cerqueira, G.M., Croda, J., Srikram, A., Henry, R., Ko, A.I., Dellagostin, O.A., Bulach, D.M., Sermswan, R.W., Adler, B. & Picardeau, M., Genome-wide transposon mutagenesis in pathogenic Leptospira species. *Infect Immun*, 2009. **77**(2): p. 810-816.

73. Liao, S., Sun, A., Ojcius, D.M., Wu, S., Zhao, J. & Yan, J., Inactivation of the fliY gene encoding a flagellar motor switch protein attenuates mobility and virulence of Leptospira interrogans strain Lai. *BMC Microbiology*, 2009. **9**(1): p. 1-10.

74. Kassegne, K., Hu, W., Ojcius, D.M., Sun, D., Ge, Y., Zhao, J., Yang, X.F., Li, L. & Yan, J., Identification of collagenase as a critical virulence factor for invasiveness and transmission of pathogenic Leptospira species. *J Infect Dis*, 2014. **209**(7): p. 1105-1115.

75. Zhang, L., Zhang, C., Ojcius, D.M., Sun, D., Zhao, J., Lin, X., Li, L., Li, L. & Yan, J., The mammalian cell entry (Mce) protein of pathogenic Leptospira species is responsible for RGD motif-dependent infection of cells and animals. *Molecular Microbiology*, 2012. **83**(5): p. 1006-1023.

76. Lo, M., Bulach, D.M., Powell, D.R., Haake, D.A., Matsunaga, J., Paustian, M.L., Zuerner, R.L. & Adler, B., Effects of temperature on gene expression patterns in Leptospira interrogans serovar Lai as assessed by whole-genome microarrays. *Infect Immun*, 2006. **74**(10): p. 5848-5859.

77. Qin, J.H., Sheng, Y.Y., Zhang, Z.M., Shi, Y.Z., He, P., Hu, B.Y., Yang, Y., Liu, S.G., Zhao, G.P. & Guo, X.K., Genome-wide transcriptional analysis of temperature shift in L. interrogans serovar lai strain 56601. *BMC Microbiol*, 2006. **6**: p. 51.

78. Patarakul, K., Lo, M. & Adler, B., Global transcriptomic response of Leptospira interrogans serovar Copenhageni upon exposure to serum. *BMC Microbiol*, 2010. **10**: p. 31.

79. Matsunaga, J., Lo, M., Bulach, D.M., Zuerner, R.L., Adler, B. & Haake, D.A., Response of Leptospira interrogans to physiologic osmolarity: relevance in signaling the environment-to-host transition. *Infect Immun*, 2007. **75**(6): p. 2864-2874.

80. Lo, M., Murray, G.L., Khoo, C.A., Haake, D.A., Zuerner, R.L. & Adler, B., Transcriptional response of Leptospira interrogans to iron limitation and characterization of a PerR homolog. *Infect Immun*, 2010. **78**(11): p. 4850-4859.

81. Xue, F., Dong, H., Wu, J., Wu, Z., Hu, W., Sun, A., Troxell, B., Yang, X.F. & Yan, J., Transcriptional responses of Leptospira interrogans to host innate immunity: significant changes in metabolism, oxygen tolerance, and outer membrane. *PLoS Negl Trop Dis*, 2010. **4**(10): p. e857.

82. Caimano, M.J., Sivasankaran, S.K., Allard, A., Hurley, D., Hokamp, K., Grassmann, A.A., Hinton, J.C. & Nally, J.E., A model system for studying the transcriptomic and physiological changes associated with mammalian host-adaptation by Leptospira interrogans serovar Copenhageni. *PLoS Pathog*, 2014. **10**(3): p. e1004004.

83. Eshghi, A., Becam, J., Lambert, A., Sismeiro, O., Dillies, M.A., Jagla, B., Wunder, E.A., Jr., Ko, A.I., Coppee, J.Y., Goarant, C. & Picardeau, M., A putative regulatory genetic locus modulates virulence in the pathogen Leptospira interrogans. *Infect Immun*, 2014. **82**(6): p. 2542-2552.

84. Adler, B., Lo, M., Seemann, T. & Murray, G.L., Pathogenesis of leptospirosis: the influence of genomics. *Vet Microbiol*, 2011. **153**(1-2): p. 73-81.

85. Nascimento, A.L., Verjovski-Almeida, S., Van Sluys, M.A., Monteiro-Vitorello, C.B., Camargo, L.E., Digiampietri, L.A., Harstkeerl, R.A., Ho, P.L., Marques, M.V., Oliveira, M.C., Setubal, J.C., Haake, D.A. & Martins, E.A., Genome features of Leptospira interrogans serovar Copenhageni. *Braz J Med Biol Res*, 2004. **37**(4): p. 459-477.

86. Bopp, S.E.R., Manary, M.J., Bright, A.T., Johnston, G.L., Dharia, N.V., Luna, F.L., McCormack, S., Plouffe, D., McNamara, C.W. & Walker,

J.R., Mitotic evolution of Plasmodium falciparum shows a stable core Genome but recombination in antigen families. *PLoS Genet*, 2013. **9**.

87. Hoepfner, D., McNamara, C.W., Lim, C.S., Studer, C., Riedl, R., Aust, T., McCormack, S.L., Plouffe, D.M., Meister, S. & Schuierer, S., Selective and specific inhibition of the plasmodium falciparum lysyl-tRNA synthetase by the fungal secondary metabolite cladosporin. *Cell Host Microbe*, 2012. **11**.

88. Nam, T.G., McNamara, C.W., Bopp, S., Dharia, N.V., Meister, S., Bonamy, G.M., Plouffe, D.M., Kato, N., McCormack, S., Bursulaya, B., Ke, H., Vaidya, A.B., Schultz, P.G. & Winzeler, E.A., A chemical genomic analysis of decoquinate, a Plasmodium falciparum cytochrome b inhibitor. *ACS Chem Biol*, 2011. **6**(11): p. 1214-1222.

89. Dharia, N.V., Sidhu, A.B., Cassera, M.B., Westenberger, S.J., Bopp, S.E., Eastman, R.T., Plouffe, D., Batalov, S., Park, D.J. & Volkman, S.K., Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in Plasmodium falciparum. *Genome Biol*, 2009. **10**.

90. Meister, S., Plouffe, D.M., Kuhen, K.L., Bonamy, G.M., Wu, T., Barnes, S.W., Bopp, S.E., Borboa, R., Bright, A.T. & Che, J., Imaging of Plasmodium liver stages to drive next-generation antimalarial drug discovery. *Science*, 2011. **334**.

91. McNamara, C.W., Lee, M.C., Lim, C.S., Lim, S.H., Roland, J., Nagle, A., Simon, O., Yeung, B.K., Chatterjee, A.K., McCormack, S.L., Manary, M.J., Zeeman, A.M., Dechering, K.J., Kumar, T.R., Henrich, P.P., Gagaring, K., Ibanez, M., Kato, N., Kuhen, K.L., Fischli, C., Rottmann, M., Plouffe, D.M., Bursulaya, B., Meister, S., Rameh, L., Trappe, J., Haasen, D., Timmerman, M., Sauerwein, R.W., Suwanarusk, R., Russell, B., Renia, L., Nosten, F., Tully, D.C., Kocken, C.H., Glynne, R.J., Bodenreider, C., Fidock, D.A., Diagana, T.T. & Winzeler, E.A., Targeting Plasmodium PI(4)K to eliminate malaria. *Nature*, 2013. **504**(7479): p. 248-253.

92. Rottmann, M., McNamara, C., Yeung, B.K., Lee, M.C., Zou, B., Russell, B., Seitz, P., Plouffe, D.M., Dharia, N.V., Tan, J., Cohen, S.B., Spencer, K.R., Gonzalez-Paez, G.E., Lakshminarayana, S.B., Goh, A., Suwanarusk, R., Jegla, T., Schmitt, E.K., Beck, H.P., Brun, R., Nosten, F., Renia, L., Dartois, V., Keller, T.H., Fidock, D.A., Winzeler, E.A. &

Diagana, T.T., Spiroindolones, a potent compound class for the treatment of malaria. *Science*, 2010. **329**(5996): p. 1175-1180.

93. Eastman, R.T., Dharia, N.V., Winzeler, E.A. & Fidock, D.A., Piperaquine resistance is associated with a copy number variation on chromosome 5 in drug-pressured Plasmodium falciparum parasites. *Antimicrob Agents Chemother*, 2011. **55**(8): p. 3908-3916.

94. Istvan, E.S., Dharia, N.V., Bopp, S.E., Gluzman, I., Winzeler, E.A. & Goldberg, D.E., Validation of isoleucine utilization targets in Plasmodium falciparum. *Proc Natl Acad Sci U S A*, 2011. **108**(4): p. 1627-1632.

95. Spangenberg, T., Burrows, J.N., Kowalczyk, P., McDonald, S., Wells, T.N. & Willis, P., The open access malaria box: a drug discovery catalyst for neglected diseases. *PLoS One*, 2013. **8**(6): p. e62906.

96. Crowther, G.J., Napuli, A.J., Gilligan, J.H., Gagaring, K., Borboa, R., Francek, C., Chen, Z., Dagostino, E.F., Stockmyer, J.B., Wang, Y., Rodenbough, P.P., Castaneda, L.J., Leibly, D.J., Bhandari, J., Gelb, M.H., Brinker, A., Engels, I., Taylor, J., Chatterjee, A.K., Fantauzzi, P., Glynne, R.J., Van Voorhis, W.C. & Kuhen, K.L., Identification of inhibitors for putative malaria drug targets amongst novel antimalarial compounds. *Molecular and biochemical parasitology*, 2011. **175**(1): p. 21-29.

97. Kuhen, K.L., Chatterjee, A.K., Rottmann, M., Gagaring, K., Borboa, R., Buenviaje, J., Chen, Z., Francek, C., Wu, T., Nagle, A., Barnes, S.W., Plouffe, D., Lee, M.C., Fidock, D.A., Graumans, W., van de Vegte-Bolmer, M., van Gemert, G.J., Wirjanata, G., Sebayang, B., Marfurt, J., Russell, B., Suwanarusk, R., Price, R.N., Nosten, F., Tungtaeng, A., Gettayacamin, M., Sattabongkot, J., Taylor, J., Walker, J.R., Tully, D., Patra, K.P., Flannery, E.L., Vinetz, J.M., Renia, L., Sauerwein, R.W., Winzeler, E.A., Glynne, R.J. & Diagana, T.T., KAF156 is an antimalarial clinical candidate with potential for use in prophylaxis, treatment, and prevention of disease transmission. *Antimicrob Agents Chemother*, 2014. **58**(9): p. 5060-5067.

98. Bopp, S.E.R., Manary, M.J., Bright, A.T., Johnston, G.L., Dharia, N.V., Luna, F.L., McCormack, S., Plouffe, D., McNamara, C.W., Walker, J.R., Fidock, D.A., Denchi, E.L. & Winzeler, E.A., Mitotic Evolution of <italic>Plasmodium falciparum</italic> Shows a Stable Core

Genome but Recombination in Antigen Families. *PLoS Genet*, 2013. **9**(2): p. e1003293.

99. Dong, C.K., Urgaonkar, S., Cortese, J.F., Gamo, F.J., Garcia-Bustos, J.F., Lafuente, M.J., Patel, V., Ross, L., Coleman, B.I., Derbyshire, E.R., Clish, C.B., Serrano, A.E., Cromwell, M., Barker, R.H., Jr., Dvorin, J.D., Duraisingh, M.T., Wirth, D.F., Clardy, J. & Mazitschek, R., Identification and validation of tetracyclic benzothiazepines as Plasmodium falciparum cytochrome bc1 inhibitors. *Chem Biol*, 2011. **18**(12): p. 1602-1610.

100. Lukens, A.K., Heidebrecht, R.W., Jr., Mulrooney, C., Beaudoin, J.A., Comer, E., Duvall, J.R., Fitzgerald, M.E., Masi, D., Galinsky, K., Scherer, C.A., Palmer, M., Munoz, B., Foley, M., Schreiber, S.L., Wiegand, R.C. & Wirth, D.F., Diversity-oriented synthesis probe targets Plasmodium falciparum cytochrome b ubiquinone reduction site and synergizes with oxidation site inhibitors. *J Infect Dis*, 2015. **211**(7): p. 1097-1103.

101. Flannery, E.L., McNamara, C.W., Kim, S.W., Kato, T.S., Li, F., Teng, C.H., Gagaring, K., Manary, M.J., Barboa, R., Meister, S., Kuhen, K., Vinetz, J.M., Chatterjee, A.K. & Winzeler, E.A., Mutations in the P-type cation-transporter ATPase 4, PfATP4, mediate resistance to both aminopyrazole and spiroindolone antimalarials. *ACS Chem Biol*, 2015. **10**(2): p. 413-420.

102. Jimenez-Diaz, M.B., Ebert, D., Salinas, Y., Pradhan, A., Lehane, A.M., Myrand-Lapierre, M.E., O'Loughlin, K.G., Shackleford, D.M., Justino de Almeida, M., Carrillo, A.K., Clark, J.A., Dennis, A.S., Diep, J., Deng, X., Duffy, S., Endsley, A.N., Fedewa, G., Guiguemde, W.A., Gomez, M.G., Holbrook, G., Horst, J., Kim, C.C., Liu, J., Lee, M.C., Matheny, A., Martinez, M.S., Miller, G., Rodriguez-Alejandre, A., Sanz, L., Sigal, M., Spillman, N.J., Stein, P.D., Wang, Z., Zhu, F., Waterson, D., Knapp, S., Shelat, A., Avery, V.M., Fidock, D.A., Gamo, F.J., Charman, S.A., Mirsalis, J.C., Ma, H., Ferrer, S., Kirk, K., Angulo-Barturen, I., Kyle, D.E., DeRisi, J.L., Floyd, D.M. & Guy, R.K., (+)-SJ733, a clinical candidate for malaria that acts through ATP4 to induce rapid host-mediated clearance of Plasmodium. *Proc Natl Acad Sci U S A*, 2014. **111**(50): p. E5455-5462.

103. Martin, E.J., Blaney, J.M., Siani, M.A., Spellmeyer, D.C., Wong, A.K. & Moos, W.H., Measuring Diversity: Experimental Design of Combinatorial Libraries for Drug Discovery. *Journal of Medicinal Chemistry*, 1995. **38**(9): p. 1431-1436.

104. Matter, H., Selecting Optimally Diverse Compounds from Structure Databases: A Validation Study of Two-Dimensional and Three-Dimensional Molecular Descriptors. *Journal of Medicinal Chemistry*, 1997. **40**(8): p. 1219-1229.

105. Walliker, D., Quakyi, I.A., Wellems, T.E., McCutchan, T.F., Szarfman, A., London, W.T., Corcoran, L.M., Burkot, T.R. & Carter, R., Genetic analysis of the human malaria parasite Plasmodium falciparum. *Science*, 1987. **236**(4809): p. 1661-1666.

106. Oduola, A.M., Weatherly, N.F., Bowdre, J.H. & Desjardins, R.E., Plasmodium falciparum: cloning by single-erythrocyte micromanipulation and heterogeneity in vitro. *Exp Parasitol*, 1988. **66**(1): p. 86-95.

107. Wellems, T.E., Panton, L.J., Gluzman, I.Y., do Rosario, V.E., Gwadz, R.W., Walker-Jonah, A. & Krogstad, D.J., Chloroquine resistance not linked to mdr-like genes in a Plasmodium falciparum cross. *Nature*, 1990. **345**(6272): p. 253-255.

108. Ponnudurai, T., Leeuwenberg, A.D. & Meuwissen, J.H., Chloroquine sensitivity of isolates of Plasmodium falciparum adapted to in vitro culture. *Trop Geogr Med*, 1981. **33**(1): p. 50-54.

109. Campbell, C.C., Collins, W.E., Nguyen-Dinh, P., Barber, A. & Broderson, J.R., Plasmodium falciparum gametocytes from culture in vitro develop to sporozoites that are infectious to primates. *Science*, 1982. **217**(4564): p. 1048-1050.

110. Looareesuwan, S., Viravan, C., Webster, H.K., Kyle, D.E., Hutchinson, D.B. & Canfield, C.J., Clinical studies of atovaquone, alone or in combination with other antimalarial drugs, for treatment of acute uncomplicated malaria in Thailand. *Am J Trop Med Hyg*, 1996. **54**(1): p. 62-66.

111. Lukens, A.K., Ross, L.S., Heidebrecht, R., Javier Gamo, F., Lafuente-Monasterio, M.J., Booker, M.L., Hartl, D.L., Wiegand, R.C. & Wirth, D.F., Harnessing evolutionary fitness in Plasmodium falciparum for drug discovery and suppressing resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 2014. **111**(2): p. 799-804.

112. Ross, L.S., Javier Gamo, F., Lafuente-Monasterio, M.J., eacute, Singh, O.M.P., Rowland, P., Wiegand, R.C. & Wirth, D.F., In Vitro Resistance

Selections for Plasmodium falciparum Dihydroorotate Dehydrogenase Inhibitors give Mutants with Multiple Point Mutations in the Drug-Binding Site and Altered Growth. *Journal of Biological Chemistry*, 2014.

113. Herman, J.D., Pepper, L.R., Cortese, J.F., Estiu, G., Galinsky, K., Zuzarte-Luis, V., Derbyshire, E.R., Ribacke, U., Lukens, A.K., Santos, S.A., Patel, V., Clish, C.B., Sullivan, W.J., Jr., Zhou, H., Bopp, S.E., Schimmel, P., Lindquist, S., Clardy, J., Mota, M.M., Keller, T.L., Whitman, M., Wiest, O., Wirth, D.F. & Mazitschek, R., The cytoplasmic prolyl-tRNA synthetase of the malaria parasite is a dual-stage target of febrifugine and its analogs. *Sci Transl Med*, 2015. **7**(288): p. 288ra277.

114. Plouffe, David M., Wree, M., Du, Alan Y., Meister, S., Li, F., Patra, K., Lubar, A., Okitsu, Shinji L., Flannery, Erika L., Kato, N., Tanaseichuk, O., Comer, E., Zhou, B., Kuhen, K., Zhou, Y., Leroy, D., Schreiber, Stuart L., Scherer, Christina A., Vinetz, J. & Winzeler, Elizabeth A., High-Throughput Assay and Discovery of Small Molecules that Interrupt Malaria Transmission. *Cell Host & Microbe*. **19**(1): p. 114-126.

115. Linares, M., Viera, S., Crespo, B., Franco, V., Gomez-Lorenzo, M., Jimenez-Diaz, M., Angulo-Barturen, I., Sanz, L. & Gamo, F.-J., Identifying rapidly parasiticidal anti-malarial drugs using a simple and reliable in vitro parasite viability fast assay. *Malaria Journal*, 2015. **14**(1): p. 441.

116. Sanz, L.M., Crespo, B., De-Cózar, C., Ding, X.C., Llergo, J.L., Burrows, J.N., García-Bustos, J.F. & Gamo, F.-J., <italic>P. falciparum In Vitro</italic> Killing Rates Allow to Discriminate between Different Antimalarial Mode-of-Action. *PLoS ONE*, 2012. **7**(2): p. e30949.

117. Bultrini, E., Brick, K., Mukherjee, S., Zhang, Y., Silvestrini, F., Alano, P. & Pizzi, E., Revisiting the Plasmodium falciparum RIFIN family: from comparative genomics to 3D-model prediction. *BMC Genomics*, 2009. **10**(1): p. 1-16.

118. Peters, J., Fowler, E., Gatton, M., Chen, N., Saul, A. & Cheng, Q., High diversity and rapid changeover of expressed var genes during the acute phase of Plasmodium falciparum infections in human volunteers. *Proc Natl Acad Sci U S A*, 2002. **99**(16): p. 10689-10694.

119. Claessens, A., Hamilton, W.L., Kekre, M., Otto, T.D., Faizullabhoy, A., Rayner, J.C. & Kwiatkowski, D., Generation of antigenic diversity in

Plasmodium falciparum by structured rearrangement of Var genes during mitosis. *PLoS Genet*, 2014. **10**(12): p. e1004812.

120. Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E. & Bowman, S., Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, 2002. **419**.

121. Jiang, H., Yi, M., Mu, J., Zhang, L., Ivens, A., Klimczak, L.J., Huyen, Y., Stephens, R.M. & Su, X.-z., Detection of genome-wide polymorphisms in the AT-rich Plasmodium falciparum genome using a high-density microarray. *BMC Genomics*, 2008. **9**(1): p. 1-15.

122. Kensche, P.R., Maria Hoeijmakers, W.A., Toenhake, C.G., Bras, M., Chappell, L., Berriman, M. & Bártfai, R., The nucleosome landscape of Plasmodium falciparum reveals chromatin architecture and dynamics of regulatory sequences. *Nucleic Acids Research*, 2015.

123. Gunasekera, A.M., Myrick, A., Militello, K.T., Sims, J.S., Dong, C.K., Gierahn, T., Le Roch, K., Winzeler, E. & Wirth, D.F., Regulatory motifs uncovered among gene expression clusters in Plasmodium falciparum. *Mol Biochem Parasitol*, 2007. **153**(1): p. 19-30.

124. Gutierrez, J.B., Harb, O.S., Zheng, J., Tisch, D.J., Charlebois, E.D., Stoeckert, C.J., Jr. & Sullivan, S.A., A Framework for Global Collaborative Data Management for Malaria Research. *Am J Trop Med Hyg*, 2015. **93**(3 Suppl): p. 124-132.

125. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T., REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, 2011. **6**(7): p. e21800.

126. Noviyanti, R., Brown, G.V., Wickham, M.E., Duffy, M.F., Cowman, A.F. & Reeder, J.C., Multiple var gene transcripts are expressed in Plasmodium falciparum infected erythrocytes selected for adhesion. *Mol Biochem Parasitol*, 2001. **114**(2): p. 227-237.

127. Zhou, Y., Ramachandran, V., Kumar, K.A., Westenberger, S., Refour, P., Zhou, B., Li, F., Young, J.A., Chen, K., Plouffe, D., Henson, K., Nussenzweig, V., Carlton, J., Vinetz, J.M., Duraisingh, M.T. & Winzeler, E.A., Evidence-based annotation of the malaria parasite's genome using comparative expression profiling. *PLOS One*, 2008. **3**.

128. Bellanca, S., Summers, R.L., Meyrath, M., Dave, A., Nash, M.N., Dittmer, M., Sanchez, C.P., Stein, W.D., Martin, R.E. & Lanzer, M., Multiple drugs compete for transport via the Plasmodium falciparum chloroquine resistance transporter at distinct but interdependent sites. *J Biol Chem*, 2014. **289**(52): p. 36336-36351.

129. Dong, C., Urgaonkar, S., Cortese, J.F., Gamo, F.J., Garcia-Bustos, J.F., Lafuente, M.J., Patel, V., Ross, L., Coleman, B.I., Derbyshire, E.R., Clish, C.B., Serrano, A.E., Cromwell, M., Barker, R.H., Dvorin, J.D., Duraisingh, M.T., Wirth, D.F., Clardy, J. & Mazitschek, R., Identification and validation of tetracyclic benzothiazepines as Plasmodium falciparum cytochrome bc(1) inhibitors. *Chemistry & biology*, 2011. **18**(12): p. 1602-1610.

130. 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*, 1979. **16**(6): p. 710-718.

131. Goodyer, I.D. & Taraschi, T.F., Plasmodium falciparum: a simple, rapid method for detecting parasite clones in microtiter plates. *Exp Parasitol*, 1997. **86**(2): p. 158-160.

132. Manary, M.J., Singhakul, S.S., Flannery, E.L., Bopp, S.E., Corey, V.C., Bright, A.T., McNamara, C.W., Walker, J.R. & Winzeler, E.A., Identification of pathogen genomic variants through an integrated pipeline. *BMC Bioinformatics*, 2014. **15**(1): p. 1-14.

133. QikProp, version 4.2. *Schrödinger, LLC, New York, NY*, 2014.

134. Cruciani, G., Crivori, P., Carrupt, P.A. & Testa, B., Molecular fields in quantitative structure–permeation relationships: the VolSurf approach. *Journal of Molecular Structure: THEOCHEM*, 2000. **503**(1–2): p. 17-30.

135. Rogers, D. & Hahn, M., Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 2010. **50**(5): p. 742-754.

136. Wilkens, S.J., Janes, J. & Su, A.I., HierS: hierarchical scaffold clustering using topological chemical graphs. *J Med Chem*, 2005. **48**(9): p. 3182-3193.

137. Novartis. *Efficacy, Safety, Tolerability and Pharmacokinetics of KAF156 in Adult Patients With Acute, Uncomplicated Plasmodium*

Falciparum or Vivax Malaria Mono-infection. Available from: http://clinicaltrials.gov/show/NCT01753323

138. Leong, F.J., Zhao, R., Zeng, S., Magnusson, B., Diagana, T.T. & Pertel, P., A first-in-human randomized, double-blind, placebo-controlled, single- and multiple-ascending oral dose study of novel Imidazolopiperazine KAF156 to assess its safety, tolerability, and pharmacokinetics in healthy adult volunteers. *Antimicrob Agents Chemother*, 2014. **58**(11): p. 6437-6443.

139. Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., Chu, A.M., Connelly, C., Davis, K., Dietrich, F., Dow, S.W., El Bakkoury, M., Foury, F., Friend, S.H., Gentalen, E., Giaever, G., Hegemann, J.H., Jones, T., Laub, M., Liao, H., Liebundguth, N., Lockhart, D.J., Lucau-Danila, A., Lussier, M., M'Rabet, N., Menard, P., Mittmann, M., Pai, C., Rebischung, C., Revuelta, J.L., Riles, L., Roberts, C.J., Ross-MacDonald, P., Scherens, B., Snyder, M., Sookhai-Mahadeo, S., Storms, R.K., Veronneau, S., Voet, M., Volckaert, G., Ward, T.R., Wysocki, R., Yen, G.S., Yu, K., Zimmermann, K., Philippsen, P., Johnston, M. & Davis, R.W., Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science*, 1999. **285**(5429): p. 901-906.

140. Jonikas, M.C., Collins, S.R., Denic, V., Oh, E., Quan, E.M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J.S. & Schuldiner, M., Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science*, 2009. **323**(5922): p. 1693-1697.

141. Howell, G.R., Shindo, M., Murray, S., Gridley, T., Wilson, L.A. & Schimenti, J.C., Mutation of a Ubiquitously Expressed Mouse Transmembrane Protein (Tapt1) Causes Specific Skeletal Homeotic Transformations. *Genetics*, 2007. **175**(2): p. 699-707.

142. Plouffe, D., Brinker, A., McNamara, C., Henson, K., Kato, N., Kuhen, K., Nagle, A., Adrian, F., Matzen, J.T., Anderson, P., Nam, T.G., Gray, N.S., Chatterjee, A., Janes, J., Yan, S.F., Trager, R., Caldwell, J.S., Schultz, P.G., Zhou, Y. & Winzeler, E.A., In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. *Proc Natl Acad Sci U S A*, 2008. **105**(26): p. 9059-9064.

143. Bowman, J.D., Merino, E.F., Brooks, C.F., Striepen, B., Carlier, P.R. & Cassera, M.B., Antiapicoplast and gametocytocidal screening to identify the mechanisms of action of compounds within the malaria box. *Antimicrob Agents Chemother*, 2014. **58**(2): p. 811-819.

144. Duffy, S. & Avery, V.M., Identification of inhibitors of Plasmodium falciparum gametocyte development. *Malar J*, 2013. **12**: p. 408.

145. Meister, S., Plouffe, D.M., Kuhen, K.L., Bonamy, G.M.C., Wu, T., Barnes, S.W., Bopp, S.E., Borboa, R., Bright, A.T., Che, J., Cohen, S., Dharia, N.V., Gagaring, K., Gettayacamin, M., Gordon, P., Groessl, T., Kato, N., Lee, M.C.S., McNamara, C.W., Fidock, D.A., Nagle, A., Nam, T.-g., Richmond, W., Roland, J., Rottmann, M., Zhou, B., Froissard, P., Glynne, R.J., Mazier, D., Sattabongkot, J., Schultz, P.G., Tuntland, T., Walker, J.R., Zhou, Y., Chatterjee, A., Diagana, T.T. & Winzeler, E.A., Imaging of Plasmodium Liver Stages to Drive Next-Generation Antimalarial Drug Discovery. *Science*, 2011. **334**(6061): p. 1372-1377.

146. Plouffe, D.M., Wree, M., Du, A.Y., Meister, S., Li, F., Patra, K., Lubar, A., Okitsu, S.L., Flannery, E.L., Kato, N., Tanaseichuk, O., Comer, E., Zhou, B., Kuhen, K., Zhou, Y., Leroy, D., Schreiber, S.L., Scherer, C.A., Vinetz, J. & Winzeler, E.A., High-Throughput Assay and Discovery of Small Molecules that Interrupt Malaria Transmission. *Cell Host Microbe*, 2016. **19**(1): p. 114-126.

147. Ruecker, A., Mathias, D.K., Straschil, U., Churcher, T.S., Dinglasan, R.R., Leroy, D., Sinden, R.E. & Delves, M.J., A male and female gametocyte functional viability assay to identify biologically relevant malaria transmission-blocking drugs. *Antimicrob Agents Chemother*, 2014. **58**(12): p. 7292-7302.

148. Sun, W., Tanaka, T.Q., Magle, C.T., Huang, W., Southall, N., Huang, R., Dehdashti, S.J., McKew, J.C., Williamson, K.C. & Zheng, W., Chemical signatures and new drug targets for gametocytocidal drug development. *Scientific reports*, 2014. **4**: p. 3743.

149. Gagaring K, B.R., Francek C, Chen Z, Buenviaje J, Plouffe D, Winzeler E, Brinker A, Diagana T, Taylor J, Glynne R, Chatterjee A, Kuhen K. Genomics Institute of the Novartis Research Foundation (GNF): 10675 John Jay Hopkins Drive, San Diego CA 92121, USA and Novartis Institute for Tropical Disease, 10 Biopolis Road, Chromos # 05–01, 138 670 Singapore.

150. Petersen, I., Eastman, R. & Lanzer, M., Drug-resistant malaria: Molecular mechanisms and implications for public health. *FEBS Letters*, 2011. **585**(11): p. 1551-1562.

151. Patel, V., Booker, M., Kramer, M., Ross, L., Celatka, C.A., Kennedy, L.M., Dvorin, J.D., Duraisingh, M.T., Sliz, P., Wirth, D.F. & Clardy, J., Identification and characterization of small molecule inhibitors of Plasmodium falciparum dihydroorotate dehydrogenase. *J Biol Chem*, 2008. **283**(50): p. 35078-35085.

152. Booker, M.L., Bastos, C.M., Kramer, M.L., Barker, R.H., Jr., Skerlj, R., Sidhu, A.B., Deng, X., Celatka, C., Cortese, J.F., Guerrero Bravo, J.E., Crespo Llado, K.N., Serrano, A.E., Angulo-Barturen, I., Jimenez-Diaz, M.B., Viera, S., Garuti, H., Wittlin, S., Papastogiannidis, P., Lin, J.W., Janse, C.J., Khan, S.M., Duraisingh, M., Coleman, B., Goldsmith, E.J., Phillips, M.A., Munoz, B., Wirth, D.F., Klinger, J.D., Wiegand, R. & Sybertz, E., Novel inhibitors of Plasmodium falciparum dihydroorotate dehydrogenase with anti-malarial activity in the mouse model. *J Biol Chem*, 2010. **285**(43): p. 33054-33064.

153. Gujjar, R., Marwaha, A., El Mazouni, F., White, J., White, K.L., Creason, S., Shackleford, D.M., Baldwin, J., Charman, W.N., Buckner, F.S., Charman, S., Rathod, P.K. & Phillips, M.A., Identification of a metabolically stable triazolopyrimidine-based dihydroorotate dehydrogenase inhibitor with antimalarial activity in mice. *J Med Chem*, 2009. **52**(7): p. 1864-1872.

154. Phillips, M.A., Gujjar, R., Malmquist, N.A., White, J., El Mazouni, F., Baldwin, J. & Rathod, P.K., Triazolopyrimidine-based dihydroorotate dehydrogenase inhibitors with potent and selective activity against the malaria parasite Plasmodium falciparum. *J Med Chem*, 2008. **51**(12): p. 3649-3653.

155. Spillman, N.J. & Kirk, K., The malaria parasite cation ATPase PfATP4 and its role in the mechanism of action of a new arsenal of antimalarial drugs. *International Journal for Parasitology: Drugs and Drug Resistance*, 2015. **5**(3): p. 149-162.

156. Fry, M. & Pudney, M., Site of action of the antimalarial hydroxynaphthoquinone, 2-[trans-4-(4'-chlorophenyl) cyclohexyl]-3-hydroxy-1,4-naphthoquinone (566C80). *Biochem Pharmacol*, 1992. **43**(7): p. 1545-1553.

157. Bueno, J.M., Herreros, E., Angulo-Barturen, I., Ferrer, S., Fiandor, J.M., Gamo, F.J., Gargallo-Viola, D. & Derimanov, G., Exploration of 4(1H)-pyridones as a novel family of potent antimalarial inhibitors of the plasmodial cytochrome bc1. *Future Med Chem*, 2012. **4**(18): p. 2311-2323.

158. Nilsen, A., LaCrue, A.N., White, K.L., Forquer, I.P., Cross, R.M., Marfurt, J., Mather, M.W., Delves, M.J., Shackleford, D.M., Saenz, F.E., Morrisey, J.M., Steuten, J., Mutka, T., Li, Y., Wirjanata, G., Ryan, E., Duffy, S., Kelly, J.X., Sebayang, B.F., Zeeman, A.-M., Noviyanti, R., Sinden, R.E., Kocken, C.H.M., Price, R.N., Avery, V.M., Angulo-Barturen, I., Jiménez-Díaz, M.B., Ferrer, S., Herreros, E., Sanz, L.M., Gamo, F.-J., Bathurst, I., Burrows, J.N., Siegl, P., Guy, R.K., Winter, R.W., Vaidya, A.B., Charman, S.A., Kyle, D.E., Manetsch, R. & Riscoe, M.K., Quinolone-3-Diarylethers: A New Class of Antimalarial Drug. *Science Translational Medicine*, 2013. **5**(177): p. 177ra137-177ra137.

159. Ghidelli-Disse, S., Lafuente-Monasterio, M., Waterson, D., Witty, M., Younis, Y., Paquet, T., Street, L., Chibale, K., Gamo-Benito, F., Bantscheff, M. & Drewes, G., Identification of Plasmodium PI4 kinase as target of MMV390048 by chemoproteomics. *Malaria Journal*, 2014. **13**(Suppl 1): p. P38.

160. Aurrecoechea, C., Brestelli, J., Brunk, B.P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G. & Harb, O.S., PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res*, 2009. **37**.

161. *The Pf3K Project: pilot data release 4*. http://www.malariagen.net/data/pf3k-4. 2015.

162. Rathod, P.K., McErlean, T. & Lee, P.C., Variations in frequencies of drug resistance in Plasmodium falciparum. *Proc Natl Acad Sci U S A*, 1997. **94**(17): p. 9389-9393.

163. Ding, X., Ubben, D. & Wells, T., A framework for assessing the risk of resistance for anti-malarials in development. *Malaria Journal*, 2012. **11**(1): p. 292.

164. Chiodini, P.L., Conlon, C.P., Hutchinson, D.B., Farquhar, J.A., Hall, A.P., Peto, T.E., Birley, H. & Warrell, D.A., Evaluation of atovaquone in the treatment of patients with uncomplicated Plasmodium falciparum malaria. *J Antimicrob Chemother*, 1995. **36**(6): p. 1073-1078.

165. McKeage, K. & Scott, L., Atovaquone/proguanil: a review of its use for the prophylaxis of Plasmodium falciparum malaria. *Drugs*, 2003. **63**(6): p. 597-623.

166. Nagle, A., Wu, T., Kuhen, K., Gagaring, K., Borboa, R., Francek, C., Chen, Z., Plouffe, D., Lin, X., Caldwell, C., Ek, J., Skolnik, S., Liu, F., Wang, J., Chang, J., Li, C., Liu, B., Hollenbeck, T., Tuntland, T., Isbell, J., Chuan, T., Alper, P.B., Fischli, C., Brun, R., Lakshminarayana, S.B., Rottmann, M., Diagana, T.T., Winzeler, E.A., Glynne, R., Tully, D.C. & Chatterjee, A.K., Imidazolopiperazines: Lead Optimization of the Second-Generation Antimalarial Agents. *Journal of Medicinal Chemistry*, 2012. **55**(9): p. 4244-4273.

167. Babbitt, S.E., Altenhofen, L., Cobbold, S.A., Istvan, E.S., Fennell, C., Doerig, C., Llinas, M. & Goldberg, D.E., Plasmodium falciparum responds to amino acid starvation by entering into a hibernatory state. *Proc Natl Acad Sci U S A*, 2012. **109**(47): p. E3278-3287.

168. Trager, W. & Jensen, J.B., *Science*, 1976. **193**(4254): p. 673.

169. Silvie, O., Rubinstein, E., Franetich, J.F., Prenant, M., Belnoue, E., Renia, L., Hannoun, L., Eling, W., Levy, S., Boucheix, C. & Mazier, D., Hepatocyte CD81 is required for Plasmodium falciparum and Plasmodium yoelii sporozoite infectivity. *Nat Med*, 2003. **9**(1): p. 93-96.

170. Johnson, J.D., Dennull, R.A., Gerena, L., Lopez-Sanchez, M., Roncal, N.E. & Waters, N.C., Assessment and Continued Validation of the Malaria SYBR Green I-Based Fluorescence Assay for Use in Malaria Drug Screening. *Antimicrob Agents Chemother*, 2007. **51**(6): p. 1926-1933.

171. Witkowski, B., Amaratunga, C., Khim, N., Sreng, S., Chim, P., Kim, S., Lim, P., Mao, S., Sopha, C., Sam, B., Anderson, J.M., Duong, S., Chuor, C.M., Taylor, W.R., Suon, S., Mercereau-Puijalon, O., Fairhurst, R.M. & Menard, D., 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*, 2013. **13**(12): p. 1043-1049.

172. Lehmann, J.S., Fouts, D.E., Haft, D.H., Cannella, A.P., Ricaldi, J.N., Brinkac, L., Harkins, D., Durkin, S., Sanka, R., Sutton, G., Moreno, A., Vinetz, J.M. & Matthias, M.A., Pathogenomic inference of virulence-

associated genes in Leptospira interrogans. *PLoS Negl Trop Dis*, 2013. **7**(10): p. e2468.

173. Tatusov, R.L., Koonin, E.V. & Lipman, D.J., A genomic perspective on protein families. *Science*, 1997. **278**(5338): p. 631-637.

174. Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., Foster, L.J. & Brinkman, F.S., PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, 2010. **26**(13): p. 1608-1615.

175. Jones, D.T., Protein secondary structure prediction based on position-specific scoring matrices1. *Journal of Molecular Biology*, 1999. **292**(2): p. 195-202.

176. Buchan, D.W., Minneci, F., Nugent, T.C., Bryson, K. & Jones, D.T., Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res*, 2013. **41**(Web Server issue): p. W349-357.

177. Capra, J.A. & Singh, M., Predicting functionally important residues from sequence conservation. *Bioinformatics*, 2007. **23**(15): p. 1875-1882.

178. Romling, U., Galperin, M.Y. & Gomelsky, M., Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev*, 2013. **77**(1): p. 1-52.

179. Ryan, R.P., Cyclic di-GMP signalling and the regulation of bacterial virulence. *Microbiology*, 2013. **159**(Pt 7): p. 1286-1297.

180. Chan, C., Paul, R., Samoray, D., Amiot, N.C., Giese, B., Jenal, U. & Schirmer, T., Structural basis of activity and allosteric control of diguanylate cyclase. *Proc Natl Acad Sci U S A*, 2004. **101**(49): p. 17084-17089.

181. Moglich, A., Ayers, R.A. & Moffat, K., Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure*, 2009. **17**(10): p. 1282-1294.

182. Singh, A. & Rosenthal, P.J., Selection of cysteine protease inhibitor-resistant malaria parasites is accompanied by amplification of falcipain genes and alteration in inhibitor transport. *J Biol Chem*, 2004. **279**.

183. Maurel, C., Reizer, J., Schroeder, J.I., Chrispeels, M.J. & Saier, M.H., Jr., Functional characterization of the Escherichia coli glycerol facilitator, GlpF, in Xenopus oocytes. *J Biol Chem*, 1994. **269**(16): p. 11869-11872.

184. Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J. & Stroud, R.M., Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*, 2000. **290**(5491): p. 481-486.

185. Zhang, D., de Souza, R.F., Anantharaman, V., Iyer, L.M. & Aravind, L., Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct*, 2012. **7**: p. 18.

186. Zhang, D., Iyer, L.M. & Aravind, L., A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic Acids Res*, 2011. **39**(11): p. 4532-4552.

187. Setubal, J.C., Reis, M., Matsunaga, J. & Haake, D.A., Lipoprotein computational prediction in spirochaetal genomes. *Microbiology*, 2006. **152**(Pt 1): p. 113-121.

188. Gruber, A.R., Findeiss, S., Washietl, S., Hofacker, I.L. & Stadler, P.F., RNAz 2.0: improved noncoding RNA detection. *Pac Symp Biocomput*, 2010: p. 69-79.

189. Herbig, A. & Nieselt, K., nocoRNAc: characterization of non-coding RNAs in prokaryotes. *BMC Bioinformatics*, 2011. **12**: p. 40.

190. Tenaillon, O., Rodriguez-Verdugo, A., Gaut, R.L., McDonald, P., Bennett, A.F., Long, A.D. & Gaut, B.S., The molecular diversity of adaptive convergence. *Science*, 2012. **335**(6067): p. 457-461.

191. Lang, G.I., Rice, D.P., Hickman, M.J., Sodergren, E., Weinstock, G.M., Botstein, D. & Desai, M.M., Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, 2013. **500**(7464): p. 571-574.

192. Mazuz, M.L., Molad, T., Fish, L., Leibovitz, B., Wolkomirsky, R., Fleiderovitz, L. & Shkap, V., Genetic diversity of Babesia bovis in virulent and attenuated strains. *Parasitology*, 2012. **139**(3): p. 317-323.

193. Bawden, F.C., Reversible changes in strains of tobacco mosaic virus from leguminous plants. *J Gen Microbiol*, 1958. **18**(3): p. 751-766.

194. Callow, L.L., Mellors, L.T. & McGregor, W., Reduction in virulence of Babesia bovis due to rapid passage in splenectomized cattle. *Int J Parasitol*, 1979. **9**(4): p. 333-338.

195. Wong, M.M., Karr, S.L., Jr. & Chow, C.K., Changes in the virulence of Naegleria fowleri maintained in vitro. *J Parasitol*, 1977. **63**(5): p. 872-878.

196. Ebert, D., Experimental evolution of parasites. *Science*, 1998. **282**(5393): p. 1432-1435.

197. Zhong, Y., Chang, X., Cao, X.-J., Zhang, Y., Zheng, H., Zhu, Y., Cai, C., Cui, Z., Zhang, Y., Li, Y.-Y., Jiang, X.-G., Zhao, G.-P., Wang, S., Li, Y., Zeng, R., Li, X. & Guo, X.-K., Comparative proteogenomic analysis of the Leptospira interrogans virulence-attenuated strain IPAV against the pathogenic strain 56601. *Cell Res*, 2011. **21**(8): p. 1210-1229.

198. Galperin, M.Y., Nikolskaya, A.N. & Koonin, E.V., Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol Lett*, 2001. **203**(1): p. 11-21.

199. Tischler, A.D. & Camilli, A., Cyclic diguanylate regulates Vibrio cholerae virulence gene expression. *Infect Immun*, 2005. **73**(9): p. 5873-5882.

200. Simm, R., Morr, M., Kader, A., Nimtz, M. & Romling, U., GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Mol Microbiol*, 2004. **53**(4): p. 1123-1134.

201. Tischler, A.D. & Camilli, A., Cyclic diguanylate (c-di-GMP) regulates Vibrio cholerae biofilm formation. *Mol Microbiol*, 2004. **53**(3): p. 857-869.

202. Ristow, P., Bourhy, P., Kerneis, S., Schmitt, C., Prevost, M.C., Lilenbaum, W. & Picardeau, M., Biofilm formation by saprophytic and pathogenic leptospires. *Microbiology*, 2008. **154**(Pt 5): p. 1309-1317.

203. Brihuega, B., Samartino, L., Auteri, C., Venzano, A. & Caimi, K., In vivo cell aggregations of a recent swine biofilm-forming isolate of

Leptospira interrogans strain from Argentina. *Rev Argent Microbiol*, 2012. **44**(3): p. 138-143.

204. He, M., Zhang, J.J., Ye, M., Lou, Y. & Yang, X.F., Cyclic Di-GMP receptor PlzA controls virulence gene expression through RpoS in Borrelia burgdorferi. *Infect Immun*, 2014. **82**(1): p. 445-452.

205. Bian, J., Liu, X., Cheng, Y.Q. & Li, C., Inactivation of cyclic Di-GMP binding protein TDE0214 affects the motility, biofilm formation, and virulence of Treponema denticola. *J Bacteriol*, 2013. **195**(17): p. 3897-3905.

206. Novak, E.A., Sultan, S.Z. & Motaleb, M.A., The cyclic-di-GMP signaling pathway in the Lyme disease spirochete, Borrelia burgdorferi. *Front Cell Infect Microbiol*, 2014. **4**: p. 56.

207. Irie, Y., Borlee, B.R., O'Connor, J.R., Hill, P.J., Harwood, C.S., Wozniak, D.J. & Parsek, M.R., Self-produced exopolysaccharide is a signal that stimulates biofilm formation in Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A*, 2012. **109**(50): p. 20632-20636.

208. Kovacs-Simon, A., Titball, R.W. & Michell, S.L., Lipoproteins of bacterial pathogens. *Infect Immun*, 2011. **79**(2): p. 548-561.

209. Ristow, P., Bourhy, P., da Cruz McBride, F.W., Figueira, C.P., Huerre, M., Ave, P., Girons, I.S., Ko, A.I. & Picardeau, M., The OmpA-like protein Loa22 is essential for leptospiral virulence. *PLoS Pathog*, 2007. **3**(7): p. e97.

210. Chatfield, S.N., Strahan, K., Pickard, D., Charles, I.G., Hormaeche, C.E. & Dougan, G., Evaluation of Salmonella typhimurium strains harbouring defined mutations in htrA and aroA in the murine salmonellosis model. *Microb Pathog*, 1992. **12**(2): p. 145-151.

211. Sinha, K., Mastroeni, P., Harrison, J., de Hormaeche, R.D. & Hormaeche, C.E., Salmonella typhimurium aroA, htrA, and aroD htrA mutants cause progressive infections in athymic (nu/nu) BALB/c mice. *Infect Immun*, 1997. **65**(4): p. 1566-1569.

212. Johnson, K.S., Charles, I.G., Dougan, G., Miller, I.A., Pickard, D., O'Goara, P., Costa, G., Ali, T. & Hormaeche, C.E., The role of a stress-response protein in bacterial virulence. *Res Microbiol*, 1990. **141**(7-8): p. 823-825.

213. Toledo-Arana, A., Repoila, F. & Cossart, P., Small noncoding RNAs controlling pathogenesis. *Curr Opin Microbiol*, 2007. **10**(2): p. 182-188.

214. Papenfort, K. & Vogel, J., Regulatory RNA in bacterial pathogens. *Cell Host Microbe*, 2010. **8**(1): p. 116-127.

215. Storz, G., Opdyke, J.A. & Zhang, A., Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol*, 2004. **7**(2): p. 140-144.

216. Gong, H., Vu, G.P., Bai, Y., Chan, E., Wu, R., Yang, E., Liu, F. & Lu, S., A Salmonella small non-coding RNA facilitates bacterial invasion and intracellular replication by modulating the expression of virulence factors. *PLoS Pathog*, 2011. **7**(9): p. e1002120.

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

218. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R., The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 2009. **25**(16): p. 2078-2079.

219. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D. & Daly, M.J., A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, 2011. **43**(5): p. 491-498.

220. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M.A., The Genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*, 2010. **20**.

221. Altenhoff, A.M. & Dessimoz, C., Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS computational biology*, 2009. **5**(1): p. e1000262.

222. Wolf, Y.I. & Koonin, E.V., A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes. *Genome Biol Evol*, 2012. **4**(12): p. 1286-1294.

223. Schultz, J., Milpetz, F., Bork, P. & Ponting, C.P., SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A*, 1998. **95**(11): p. 5857-5864.

224. Paten, B., Diekhans, M., Earl, D., John, J.S., Ma, J., Suh, B. & Haussler, D., Cactus graphs for genome comparisons. *J Comput Biol*, 2011. **18**(3): p. 469-481.

225. Paten, B., Earl, D., Nguyen, N., Diekhans, M., Zerbino, D. & Haussler, D., Cactus: Algorithms for genome multiple sequence alignment. *Genome Res*, 2011. **21**(9): p. 1512-1528.

226. Fouts, D.E., Matthias, M.A., Adhikarla, H., Adler, B., Amorim-Santos, L., Berg, D.E., Bulach, D., Buschiazzo, A., Chang, Y.-F., Galloway, R.L., Haake, D.A., Haft, D.H., Hartskeerl, R., Ko, A.I., Levett, P.N., Matsunaga, J., Mechaly, A.E., Monk, J.M., Nascimento, A.L.T., Nelson, K.E., Palsson, B., Peacock, S.J., Picardeau, M., Ricaldi, J.N., Thaipandungpanit, J., Wunder, E.A., Jr., Yang, X.F., Zhang, J.-J. & Vinetz, J.M., What Makes a Bacterial Species Pathogenic?:Comparative Genomic Analysis of the Genus Leptospira. *PLoS Negl Trop Dis*, 2016. **10**(2): p. e0004403.