Debugging parasite genomes: Using metabolic modeling to accelerate antiparasitic drug development

Maureen A. Carey

Charlottesville, Virginia

Bachelors of Science, Lafayette College 2014

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Microbiology, Immunology, and Cancer Biology
University of Virginia

September, 2018

Abstract:

Eukaryotic parasites, like the casual agent of malaria, kill over one million people around the world annually. Developing novel antiparasitic drugs is a pressing need because there are few available therapeutics and the parasites have developed drug resistance. However, novel drug targets are challenging to identify due to poor genome annotation and experimental challenges associated with growing these parasites. Here, we focus on computational and experimental approaches that generate high-confidence hypotheses to accelerate labor-intensive experimental work and leverage existing experimental data to generate new drug targets. We generate genome-scale metabolic models for over 100 species to develop a parasite knowledgebase and apply these models to contextualize experimental data and to generate candidate drug targets.

UNIVERSITY of VIRGINIA | SCHOOL of MEDICINE

SCHOOL of MEDICINE REGISTRAR
Office of Student Affairs
PO Box 800739
434-924-5200

Final Examination Form

This form is to be completed by a student's committee and program to indicate that the student has passed the final defense of their thesis/dissertation, written or oral exams. This form must be scanned and emailed to the School of Medicine Registrar by the appropriate deadline: **December 1** for December, **May 1** for May, and **August 1** for August graduation.

Student Name: .Carey, Maureen, A. SIS ID (7 digits): _____
Last, First, Middle Initial (Print)

Degree Program: Microbiology Final Exam Passed Date: 9/19/18

PhD Final Examination: This committee, chaired by the primary advisor, will consist of a minimum of five tenured or tenure-track members of the faculty of the Biomedical Sciences Graduate Program. One member of the committee will serve as a representative of the School of Medicine to affirm that the student has been assessed fairly and with due rigor. Once these minimum requirements have been met, additional committee members from within the University or other institutions may be added.

Committee	Name	Department	Signature
Mentor	JENNIFER BULER Jason Papin	BIOLOGY BME	
Committee Chair	Alison Criss	MIC	
Committee Member	Young Hahn	MZC	
Committee Member (Outside Rep.)	Nasser Lacking	Pharm	
Outside Representative			
Committee Member (Optional)	Herré Agaisse	MIC	
Committee Member (Optional)			
Committee Member (Optional)			

Approved Title of Doctoral Dissertation:

Debugging parasite genomes: Using metabolic modeling to accelerate antiparasitic drug development.

Lucy Pemberton *Director of Graduate Studies (Print)*	*Signature*	9/20/18 *Date*
Kodi Ravichandran *Department Chair (Print)* (if required by the Program)	*Signature*	9/20/18 *Date*

Please submit this completed form to your BIMS Administrator

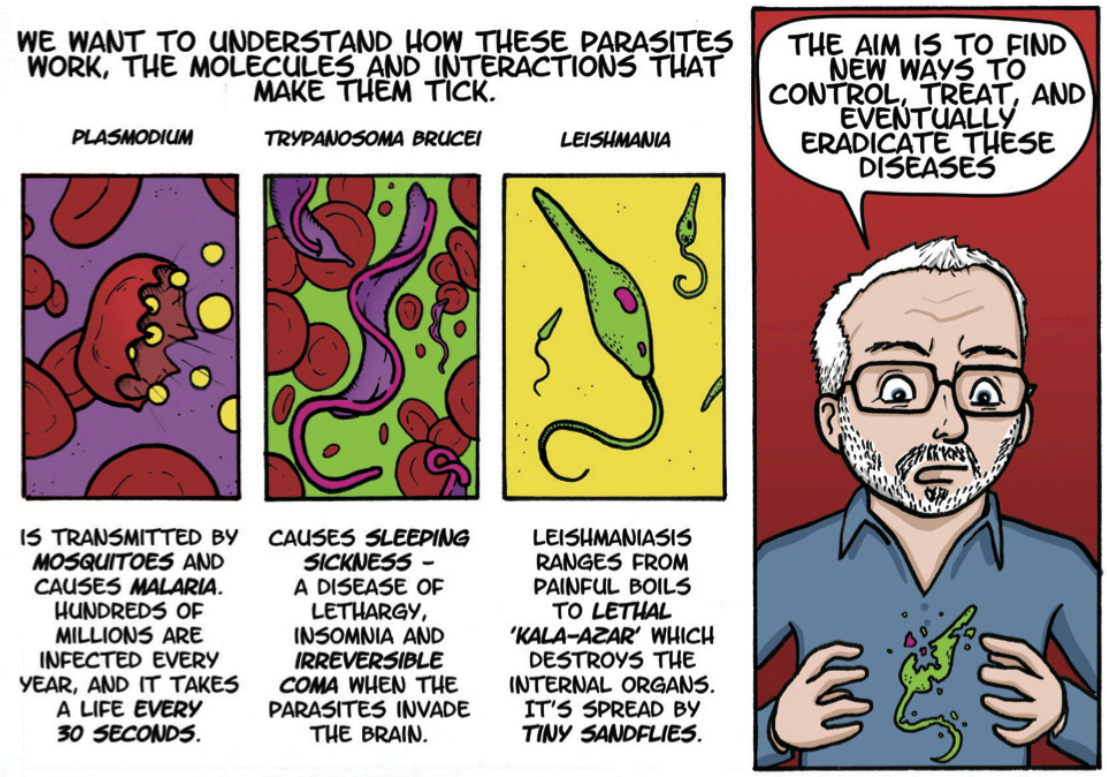


Figure 0.1: Image from blog.wellcome.ac.uk/2010/06/15/of-parasitology-and-comics/.

Preamble: Eukaryotic single-celled parasites cause diseases, such as malaria, African sleeping sickness, diarrheal disease, and leishmaniasis, with diverse clinical presentations and large global impacts. These infections result in over one million preventable deaths annually and contribute to a significant reduction in disability-adjusted life years. This global health burden makes parasitic diseases a top priority of many economic development and health advocacy groups. However, effective prevention and treatment strategies are lacking. Like most infectious disease problems, social, economic, and biological challenges converge, amplifying the disease burden and slowing the development of sustainable solutions. The work described here focuses on only one facet of this complex problem: the biology of the parasite, specifically its metabolism, during infection. However, this work broadly aims to increase our understanding of these parasites directly by studying disease-relevant phenotypes and indirectly by developing computational tools to study these organisms. Moreover, the tool development presented in this dissertation aims to increase accessibility and usability of computational biology tools and foster solutions sensitive to and compatible with the social and economic environment in which these diseases are most serious.

Acknowledgements:

Document was generated using R (R Core Team 2017), Rmarkdown (Allaire et al. 2018), and R packages: kableExtra, knitr, readxl, dplyr, ggpubr, tidyverse, and reshape2 (Zhu 2018; Xie 2014; Wickham and Bryan 2017; Wickham et al. 2015; Kassambara 2017; Wickham 2017, 2012), in addition to analytic and visualization software mentioned throughout the document. The research presented in this document was also dependent on the EuPathDB community for the database services and support. I use one of the EuPathDB databases nearly everyday and would be far less productive without the EuPathDB community's vision and hard work. The EuPathDB services inspired the knowledgebase project (Ch. 3), as I am sure they will inspire many more.

Thank you to my coauthors and the Guler, Papin, and Petri lab groups for their contributions to my work and evolution as a scientist. The Guler and Papin labs were my research homes and their thoughtful feedback and discussion was invaluable to the publications presented here and the work I will continue. The weekly Petri group meeting and friendships found there stretched my research in new and interesting directions. I am so looking forward to joining this group full time!

Thank you to my thesis committee, Drs. Hervé Agaisse, Alison Criss, Young Hahn, and Norbert Leitinger, for your time and feedback. I am so grateful that my committee meetings were always fun, interesting, and productive - what science should be! An extra thank you to Dr. Alison Criss for her support. On a personal note:

- To all of my (science and nonscience) advocates and mentors over the years - Thank you all for believing in me and for all of the lessons I will carry with me. In particular, thank you to Mr. O for teaching me how to teach myself and for being my archetype of a mentor. Drs. Kurt and Ho - thank you for helping me discover science!
- To my family and friends - Thank you for your support and for keeping me grounded in what matters most. Thank you Ron, Cami, Vicki, Scott, Brad, Donnie, Granddad, Aunt Debbie, Uncle Ed, Jenn, Nicole, Aunt Karen, Lela, Carson, Christina, Caitlin, Rusty, Mahmoud, Ali, Mom and Dad - I love you all! An extra thank you to the original Drs. Carey - Thank you for teaching me to be pleasantly persistent and showing me what leadership looks like. To my "twin" - I still hope I am as cool/kind/fun/impressive as you are when I grow up. Thank you for being my role model and leading me to where I am now.
- To my best friend and collaborator, Gregory - Without you, the sun would shine a bit less bright and my ideas would be half-baked. Thank you for everything.

None of this work would be possible without my team, and I thank you all!

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1 Introduction

In this introduction, I will contextualize the biological challenges associated with translational parasitology research by focusing on the diseases and then on the causal parasites. See **Table 1.1** for select parasitic diseases, their causal parasite, and statistics on global burden.

Table 1.1: Disease overview. Select diseases dicussed throughout this dissertation with causal parasite and disease burden noted. See Chapter 2 for more on malaria, toxoplasmosis, and cryptosporidiosis.

Disease	*Causal parasite*	Deaths annually	Endemic regions	Vector/ Transmission	Interesting disease/ parasite features
Malaria	*Plasmodium*	500,000	Tropics (3 billion people)	mosquitos	Infection induces a cyclical fever.
Leishmaniasis	*Leishmania*	20,000	Nearly worldwide (1 billion)	sandfly	Parasite has a mutable genome with frequent polyploidy.
Cryptosporidiosis (diarrheal disease)	*Cryptosporidium*	50,000+	Worldwide (disproportionally in low-income countries)	fecal-oral	Infection without diarrhea is associated with impaired growth.
Toxoplasmosis	*Toxoplasma gondii*	few	Worldwide	fecal-oral	Infection induces behavioral changes in smaller mammals.
Giardia (diarrheal disease)	*Giardia*	few	Worldwide	fecal-oral	Parasites have two nuclei.
Amebiasis (diarrheal disease)	*Entamoeba spp.*	100,000+	Worldwide (disproportionally in low-income countries)	fecal-oral	Malnutrition increases host susceptibility.

Disease control can be broken down into two major classes of efforts: prevention and treatment. To prevent serious infection, steps are taken to reduce exposure to the disease-causing agent and/or to prevent exposure from causing symptomatic infection. Chief among these are the development of vaccines to bolster immune recognition of pathogens, reducing both transmission and ameliorating disease severity, and reducing human contact with infected agents (such as insect vectors, contaminated water, and contagious individuals) (Gubler 1998). Vaccination has the potential to be extremely effective (*e.g.* malERA Consultative Group on Vaccines (2011)), as an effective vaccine paired with a global vaccination program eradicated smallpox (Belongia and Naleway 2003) and this approach has eliminated polio from many parts of the world (Kew et

al. 2005). Transmission blocking approaches include the use of bed nets to prevent exposure to the bite of malaria-infected mosquitoes, insecticides to kill infected vectors, and water chlorination (Gonçalves and Hunziker 2016). Insecticides specifically have been invaluable to the reduction of mosquito (malaria), also to a certain extent, tsetse fly (trypanosomosis) and sand fly (leishmaniasis) populations ("World Malaria Report 2017" 2017; Berg et al. 2012).

Few drugs exist for many of these diseases, drug resistance is common, and many drugs have stage specificity, leaving patients with few treatment options. Thus, the development of novel, effective therapeutics is a pressing need. Beyond the economic challenges associated with antimicrobial development, antiparasitic drug development is technically challenging for two primary reasons, namely, these parasites are eukaryotes and are challenging to grow *in vitro*. To elaborate, unlike prokaryotic pathogens, these parasites share many targetable features with their eukaryotic host and/or vector. To overcome the similarity between host and pathogen, strategies similar to the development of cancer therapeutics are necessary to minimize the negative effects on host. Enzyme kinetics can be leveraged such that the drug targets the pathogen's weak points while remaining below the lethal dose for host (Haanstra et al. 2017). Alternatively, selection of pharmacological treatment can synergize with the host immune response (*e.g.* Bogdan et al. (1991) and Kumaratilake et al. (1997)). For example, parasites must often survive high redox stress caused by host immune cells; a secondary redox stressor (*i.e.* drug) can synergize with this host response. Unique features like atypical organelles (*e.g.* Dahl et al. (2006)) can also be targeted if identified (see **Chapter 2**).

Drug development itself is hampered by experimental challenges shared by many eukaryotic pathogens. Chief among these challenges, many parasites have complicated life cycles in one or more hosts. For example, the malaria parasites infect multiple tissue types in host (primate, rodent, bird, or reptile) and vector (mosquito); all of these stages are required for complete development. These diverse environmental conditions are hypothesized to maintain redundancy in each parasite's genome, as genes and functions may only be utilized during some life cycles; thus, drug development must focus on function, not merely presence, of essential genes.

Some parasite species have unique experimental challenges due to their genomic traits and extreme host specificity. For example, *P. falciparum*, the most lethal human malaria parasite, was considered refractory to genetic modification until recently (Ghorbal et al. 2014; Lee and Fidock 2014) due to extremely low transfection efficiency. *E. histolytica*, a diarrheal pathogen, has also been refractory to efficient genetic manipulation; CRISPR technology was developed even later in this organism (*unpublished*). The genomes of *Leishmania*, which causes ulcers, develop significant

aneuploidy when under selective pressure due to genomic flexibility (Downing et al. 2011; Sterkers et al. 2012). *P. vivax*, another human malaria parasite, and *Cryptosporidium*, a major cause of diarrheal disease, have no long-term *in vitro* culture systems (Baydoun et al. 2017). Although many of these challenges can be circumvented with new technology, the use of clinical samples, and reductionist approaches, there are minimal historic data for reference.

Because these parasites are challenging to study, there is a lack of historic knowledge as well as profiling data (genome-wide essentiality screens, growth profiling in diverse media conditions, etc.). Accordingly, we do not have the knowledge to *rationally* design drugs. Untargeted and unbiased screens of chemical compounds for antiparasitic effects have proven useful (Boyom et al. 2014; Van Voorhis et al. 2016) but this approach makes predicting and understanding drug resistance and resistance mechanisms challenging.

As a result, data collected in one organism are frequently extrapolated to infer knowledge about another parasite, across and within genera. Trypanosomes and *Toxoplasma* are frequently used as model organisms for other parasites due to their experimental manipulatability. Mouse models of malaria and cryptosporidiosis are caused by different species than their human analogs. However, the modest characterization of functional differences among parasite species limits the utility of this extrapolation approach. The following chapters leverage comparative genomics and modeling approaches to address these challenges. This dissertation aims to provide tools for the study of these organisms and evaluation criteria to assess these tools, and an investigation of antimalarial drug resistance using these tools.

Chapter 2: Background

2 Background

In this chapter, I briefly discuss three genera of human pathogenic parasites. Specifically, I discuss the parasites that cause malaria, cryptosporidiosis, and toxoplasmosis, including the diseases as well as transmission, prevention, and treatment. Additionally, I discuss challenges associated with disease control and the history of the study of metabolism for each organism. This discussion is intended to motivate the value of studying these organisms for public health and some of the experimental challenges that are addressed by the computational approaches in this dissertation. Most importantly, I would like the reader to appreciate the history of antiparasitic repurposing and the lack of rigorous experimental comparison between parasite species.

2.1 *Plasmodium*

2.1.1 Malaria

Malaria remains a major public health burden around the globe with an estimated 216 million cases and 445,000 deaths in 2016 (**Table 1.1**; ("World Malaria Report 2017" 2017)). Human disease is caused by five species in the *Plasmodium* genus, *falciparum*, *vivax*, *malariae*, *knowlesi*, and *ovale*;[1] however, species of *Plasmodium* infect primates, small mammals, birds, and reptiles as well (Aikawa, Huff, and Sprinz 1969). Like many parasitic infections, malaria is a vector-borne disease; it is transmitted by female mosquitoes in the genus *Anopheles*. Although more than 70 species with a world-wide spread can carry malaria (Sinka et al. 2012), malaria infection only occurs frequently in Sub-Saharan Africa, South America, and Southeast Asia. In fact, 80% of the global disease burden is carried in 14 countries in Sub-Saharan Africa and India ("World Malaria Report 2017" 2017). The majority of malaria deaths occur in children under the age of 5 years ("World Malaria Report 2017" 2017), as adults in endemic regions develop protective immunity (Doolan, Dobaño, and Baird 2009). Despite this observation, there is not yet a clinically available vaccine for malaria.

2.1.2 Transmission and Pathogenesis

Humans acquire malaria following the bite of an infected mosquito; parasites in the mosquito's salivary gland are injected into the skin along with its itch-inducing saliva (Churcher et al. 2017). It takes as few as 1 parasite to cause disease, and a patient

[1] All citations in the following section reference research conducted in *Plasmodium falciparum*, unless noted otherwise.

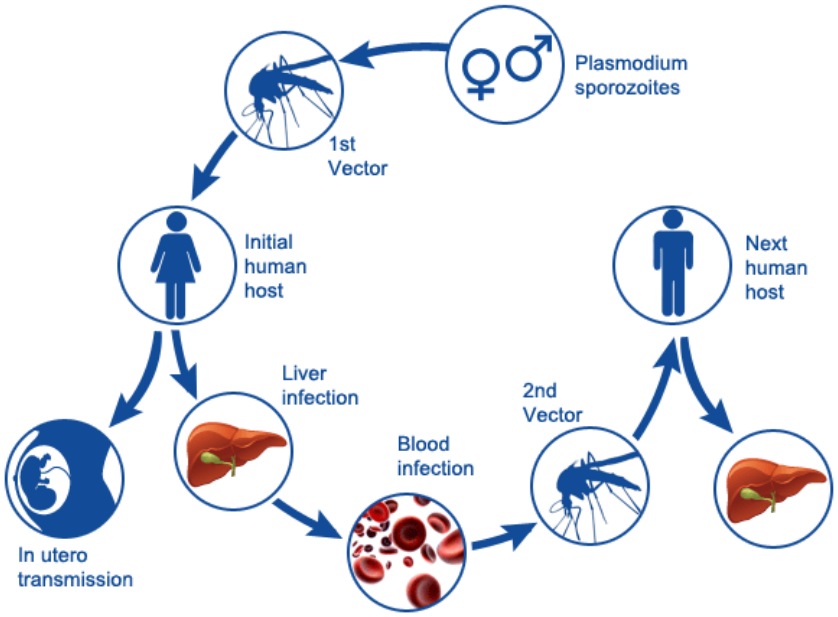


Figure 2.1: Malaria parasites require host and vector. Malaria is caused by *Plasmodium* parasites and is spread by mosquitoes and human-to-human via vertical transmission. The parasites infect two human tissues, first hepatocytes and then erythrocytes. Symptoms (including cyclical fevers and anemia) occur during blood infection. The parasite can be targeted at any stage of its lifecycle, but antimalarials most frequently target blood stage infection. The complete life cycle requires vector and each of these host tissues. Image from http://www.jenner.ac.uk/about-malaria.

without immunity to the disease has a 60% chance of infection if bitten even once by an infected mosquito with fewer than 100 parasites in its salivary gland (Churcher et al. 2017). Following injection into a human host, the parasite navigates to the liver and replicates asexually rapidly in hepatocytes before replicating in red blood cells (**Figure 2.1**). Both the liver and blood-stage parasite replicate asexually; gametocytes form from a small fraction of blood-stage parasites. The remainder of the parasite's sexual development occurs in the mosquito. The malaria parasites are obligate intracellular organisms when in the vertebrate host (**Figure 2.2**), requiring a host cell both to evade the immune system (Wright and Rayner 2014) and for nutritional support (Allary et al. 2007).

It is during asexual blood stage infection that symptoms are observed and parasites amplify from the thousands to the millions of cells. Mild *Plasmodium* infection confers flu-like symptoms with a cyclical fever as parasites synchronously rupture their host cells to invade neighboring cells (Kwiatkowski and Greenwood 1989; Garcia, Markus, and Madeira 2001).[2] As the parasite replicates in a red blood cell, it remodels the host

[2]Multiple *Plasmodium* species.

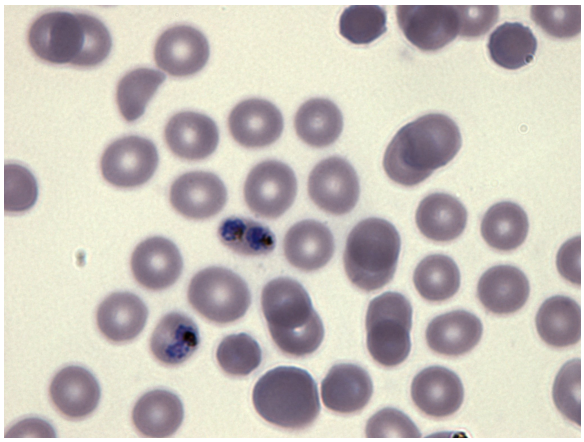


Figure 2.2: *Plasmodium falciparum.* Five species cause human malaria, with *Plasmodium falciparum* as the most lethal causative agent. Laboratory-adapted *P. falciparum* (BEI Resources, NIAID, NIH; *Plasmodium falciparum*, strain IPC 5202/IPC 5202/MRA-1240) shown. Prior to imaging, cells were fixed with methanol and stained with Giemsa stain for 15 min. Images were obtained on a Nikon Eclipse Ci microscope (100x) using an Imaging Source microscope camera. Representative image shown.

cell by building vesicular networks throughout the host's cytoplasm (Cyrklaff et al. 2011; Grüring et al. 2011). This makes the host red blood cell less flexible (Nash et al. 1989) and facilitates the export of parasite proteins to the host cell surface (Trelka et al. 2000), making the host cells cytoadhere to epithelial tissue (Su et al. 1995; Kyes et al. 1999). These two changes make the infected host cell become sequestered in small blood vessels (MacPherson et al. 1985; Ho and White 1999).

Severe infection is associated with systemic complications such as anemia, lactic acidosis, hypoglycemica, and coma (Idro et al. 2010), in part due to this sequestration of blood cells. Thus, these proteins, encoded by the var (Su et al. 1995), rifin (Kyes et al. 1999), or STEVOR gene families (Niang, Yan Yam, and Preiser 2009), can be viewed as virulence factors as they mediate cytoadherance (Magowan et al. 1988) and facilitate immune evasion (Craig and Scherf 2001; Bull et al. 1998) via frequent antigenic variation (Scherf et al. 1998).

2.1.3 Prevention and treatment

To prevent infection, transmission blocking tools are available. Both bed nets and the use of insecticides are effective by preventing human exposure to infected mosquitoes ("World Malaria Report 2017" 2017). Antimalarial drugs (**Table 2.1**) can be extremely effective if the causal parasite has not developed resistance to the treatment. In addition to alleviating symptoms and improving patient outcomes, antimalarial treatment also

effectively reduces circulating parasite burden, and thus reduces the likelihood of vertical or mosquito-mediated transmission. Moreover, select antimalarial drugs can be used as prophylaxis for travelers in malaria endemic regions (*e.g.* Overbosch et al. (2001)).

However, no vaccine currently exists for malaria, although several have been explored and one is in advanced clinical trials (RTS,S Clinical Trials Partnership 2015). The antigenic variation in *Plasmodium* surface protein expression mediates the parasite's immune evasion. The challenges associated with vaccine development are exacerbated by parasite-host species-specificities (Vaughan et al. 2012). Mouse models of malaria do not use the parasite species that infect humans; *Plasmodium berghei* is used to model severe (*falciparum*) malaria and *P. yoelli* is used to model less severe malaria, especially *vivax* malaria. Even humanized mouse models of disease do not support the complete *P. falciparum* life cycle (Minkah, Schafer, and Kappe 2018).

To treat infection, antimalarial drugs are available and active against all human infective species (**Table 2.1**). Combination therapies are used to slow the development of resistance (Eastman and Fidock 2009). However, antimalarial drug resistance is a pressing concern (discussed in the next section, **Parasite biology and previous research**).

2.1.4 Parasite biology and previous research

Humans have co-evolved with the malaria parasites for thousands of years, but the parasites themselves were first studied in 1880 (Cox 2010). In 1976, an *in vitro* culture system was developed for the direct study and manipulation of one species of these parasites (Trager and Jensen 1976). Since then, the *Plasmodium* parasites have moved from "neglected tropical" to "major human" pathogen, with billions or millions of dollars invested in disease control or research (respectively) each year ("World Malaria Report 2017" 2017). As a result, significant progress has been made in characterizing the parasites, despite some experimental challenges.

Plasmodium metabolic genes are better characterized than signaling pathways; for example, PlasmoDB identifies 43 genes in the genome of *P. falciparum* 3D7 associated with the term 'signaling' as opposed to 1112 genes associated with the term 'metabolism' (Aurrecoechea et al. 2009), and many antimalarials target metabolic functions (Fidock et al. (2000); Peterson, Walliker, and Wellems (1988); Phillips and Rathod (2010); Siregar et al. (2015)[3]). Moreover, metabolism has been described as the best-understood cellular process (Fuhrer et al. 2017), making interpreting metabolic

[3] *Plasmodium berghei*

Table 2.1: Clinically available antiparasitic drugs for apicomplexans.

Class	Drug	Organisms
amino alcohols	quinine	Plasmodium
4-aminoquinolones	chloroquine	Plasmodium
4-aminoquinolones	amodiaquone	Plasmodium
antifolates	pyrimethamine	Plasmodium and Toxoplasma
antifolates	proguanil	Plasmodium
sulonamides	sulphadoxine	Plasmodium
amino alcohols	mefloquine	Plasmodium
antimalarial, others	atovaquone	Plasmodium and Toxoplasma
8-aminoquinolines	primaquine	Plasmodium
endoperoxides	artemisinin (and others)	Plasmodium
amino alcohols	halofantrine	Plasmodium
antibiotics	doxycycline	Plasmodium
antibiotics	clindamycin	Plasmodium and Toxoplasma
antifolates	cycloguanil	Plasmodium
antibiotics	fosmidomycin	Plasmodium
antibiotics	tetracycline	Plasmodium
antibiotics	sulfadiazine	Plasmodium and Toxoplasma
antiviral/thiazolides	nitazoxanide	Cryptosporidium (sometimes)
antibiotics	azithromycin	Cryptosporidium (sometimes)

analyses more tractable. Additionally, NCBI's PubMed returns 21,218 hits when searching 'plasmodium metabolism'; although this number is inflated because several human metabolic disorders are associated with protection from malaria (*e.g.* glucose-6-phosphate dehydrogenase deficiency (Mbanefo et al. 2017)) or the direct result of infection (*e.g.* lactic acidosis (Sasi et al. 2007)), it still far exceeds the number of hits when searching 'plasmodium vaccine' (6027) or 'plasmodium signalling'/'plasmodium signaling' (1087). Thus, despite the poor annotation of the *Plasmodium* genomes (**Table 2.2**), metabolism is well studied. We focus on metabolism of the parasite for this reason and because many existing antimalarial drugs target metabolic functions.

Subcellular compartmentalization (*i.e.* organelles) facilitate organization of metabolic functions (**Figure 2.3**). *Plasmodium* are known to have a nucleus, apicoplast, mitochondria, food (or digestive) vacuole, Golgi apparatus, and endoplasmic reticulum (Singh et al. 2007; Fairhurst et al. 2003). The parasite also has a few unique

Table 2.2: Example genome statistics for three parasite genera. Each EuPathDB community (i.e. CryptoDB or PlasmoDB) guides the gene identifier nomenclature for that database. Because there has been no experimental system for the *Crypotosporidium* parasites, CryptoDB has not extensively utilized the 'putative' status as nearly genes remain untested and with putative function. ORFs = open reading frames.

species	strain	genome size (mbp)	ORFs	hypothetical genes	genes with putative function	genes with unknown function	percent hypothetical or unknown	percent putative
Plasmodium falciparum	3D7	23.33	5712	6	2053	1964	34.5%	35.9%
Toxoplasma gondii	ME49	65.67	8920	4285	729	1	48.0%	8.2%
Cryptosporidium parvum	Iowa II	9.10	4020	0	8	1480	36.8%	0.1%

organelles involved in invasion of the host cell, collectively called the apical complex (shared by all apicomplexans (Katris et al. 2014)). In the host red blood cell, the parasite creates a parasitophorus vacuole in which it resides and a network of vesicles, including the Maurers' cleft and transvesicular network (**Figure 2.3**), involved in protein secretion (Beck et al. 2014; Cooke et al. 2006) and nutrient acquisition (Lauer et al. 1997). The nucleus, apicoplast, and mitochondria each host their own genome (Gardner et al. (2002) and Aurrecoechea et al. (2009)[4]); although many genes that function in the apicoplast and mitochondria are encoded in the nucleus (Waller et al. 1998).

The apicoplast of *P. falciparum*, and presumably other *Plasmodium*, is only required for one function: to produce isopentenyl pyrophosphate (Yeh and DeRisi 2011). Isopentenyl pyrophosphate is used in the synthesis of ubiquinone and N-glycosylated and prenylated proteins (Yeh and DeRisi 2011). Antimalarials like fosmidomycin and antibiotics target this function in different ways. Fosmidomycin inhibits an upstream enzyme in isopentenyl pyrophosphate synthesis (Jomaa et al. 1999), and antibiotics inhibit the maintenance functions of the apicoplast (Dahl et al. 2006), rendering it nonfunctional. These maintenance functions (*i.e.* DNA replication and protein synthesis) are mechanistically more similar to bacteria than to eukaryotic cells (Kalanon and McFadden 2010; Lim and McFadden 2010), so antibiotics designed to kill prokaryotes are effective against this organelle. Under antibiotic pressure (like doxycycline) and in the presence of isopentenyl pyrophosphate supplementation, parasites can evolve to lose their apicoplast genome and nuclear-encoded apicoplast

[4]Multiple *Plasmodium* species

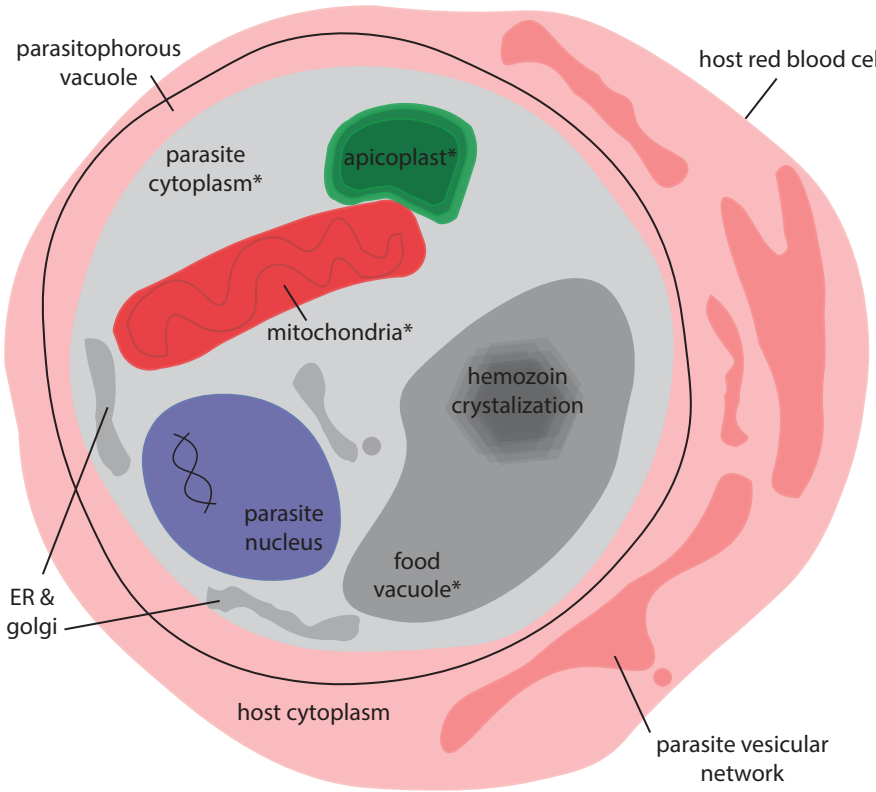


Figure 2.3: Compartmentalization in the malaria parasite. Relevant organelles in the malaria parasite are labeled. Organelles with star (*) are included in the model, as well as an extracellular compartment that describes all regions outside of the parasite, including the parasitophorous vacuole, host cytoplasm, and host extracellular environment.

genes localize elsewhere in the cell (Yeh and DeRisi 2011).

The *Plasmodium* apicoplast is home to steps in two other metabolic pathways: both fatty acid and heme synthesis. Both pathways also have steps in the mitochondria and cytosol. Enzymes in fatty acid synthesis have been explored as a drug targets (Surolia and Surolia 2001)[5] before discovering its nonessentiality in the blood stage (Vaughan et al. 2009).[6] Heme synthesis is also essential in some life stages of *Plasmodium* and explored as a drug target (Surolia and Padmanaban 1992); however it is not essential during the blood stage (Ke et al. 2014). It was previously assumed that host heme (from hemoglobin) was not accessible to the parasite for its own metabolic functions, but the parasites survival despite genetic knockdowns of heme synthesis enzymes indicates that the parasite can scavenge some form of host heme (Ke et al. 2014). Notably, these observations are host cell-dependent as reticulocytes but not erythrocytes are able to synthesize heme; experiments in Ke et al. (2014) were performed in erythrocytes where *de novo* heme synthesis does not occur but hemoglobin degradation may provide free heme.

The confusion on whether the apicoplast contains high-quality drug targets highlights the frequent off-target effects of enzyme inhibitors and the value of genetic modification to validate pathway essentiality. Both the inhibitors for fatty acid synthesis (triclosan) and heme synthesis (succinylacetone) must have off target effects; results of early inhibitor studies misled the field for years about the utility of targeting these pathways chemotherapeutically. Thus, whenever possible, we discuss genetic inhibition rather than pharmacologic inhibition in this dissertation. However, genetic modification is quite challenging. Genome-wide essentiality screens have been performed in *Plasmodium berghei* and *falciparum*, recently (Bushell et al. 2017; Zhang et al. 2018), dramatically transforming malaria research. However, these resources are not available for many parasites, especially the unculturable.

The mitochondria, however, is absolutely essential for *Plasmodium* growth; accordingly, several existing antimalarial drugs target enzymes in the mitochondria (*e.g.* atovaquone/proguanil, novel DSM compounds). However, unlike many other eukaryotic cells, mitochondrial function for energy production is not essential, at least during the asexual blood-stage of parasite development (Painter et al. 2007). The electron transport chain is essential in *Plasmodium* because dihydroorotate dehydrogenase is necessary for *de novo* nucleotide synthesis (specifically, pyrimidines) (Gutteridge and Trigg (1970)[7]; Reyes et al. (1982); Painter et al. (2007)). Moreover, these parasites are glycolytic during the blood stage. Enzymes within the tricarboxylic acid cycle

[5] Multiple *Plasmodium* species
[6] Tested in *Plasmodium yoelii* only.
[7] *Plasmodium knowlesi*

are nonessential (Ke et al. 2015) and, if dihydroorotate dehydrogenase is expressed in the cytoplasm, the electron transport chain is nonessential (Painter et al. 2007). As a result, *Plasmodium* are distinctly tumor-like, and not surprisingly proposals to repurpose antimalarials (specifically, artemisinin (Zhou, Li, and Xiao 2016; Efferth and Oesch 2004; Das 2015)) as anticancer chemotherapies and vice versa (*i.e.* trimetrexate or methotrexate (Nzila et al. 2010; Kiara et al. 2009)) have been explored.

Table 2.3: Non-inclusive experimentally characterized metabolic requirements or auxotrophies. *Cryptosporidium* data is genetically, not experimentally, derived.

Genus	Amino acids	Nucleotides	Lipids	Other
Plasmodium	isoleucine	purines	?	biotin, riboflavin, riboflavin
Toxoplasma	tryptophan, arginine	purines	?	riboflavin, thiamine
Cryptosporidium	all	all	?, lipoic acid	folates, thiamine diphosphate, pyridoxal phosphate

Plasmodium have a number of other unique metabolic features, such as an auxotrophy for isoleucine (Martin and Kirk (2007), see **Table 2.3** for others), a dependency on purine scavenging, using ubiquinone-8 rather than ubiquinone-10 (Skelton, Rietz, and Folkers 1970; Rietz et al. 1967)[8], and significant redundancy between metabolite scavenging and *de novo* synthesis (*e.g.* polyamines, and cholesterol). The blood-stage parasite also scavenges many amino acids from the host cell, both by direct import and via hemoglobin. The host red blood cell (erythrocytes for some species and reticulocytes for others) contains a large amount of hemoglobin; it is postulated that parasite digestion of hemoglobin in its food vacuole occurs both to make physical space for the parasite in the host cell (Krugliak, Zhang, and Ginsburg 2002) and to supply the parasite with heme and amino acids. Heme, however, is toxic at high concentrations and is detoxified through three known mechanisms: polymerization into hemozoin crystals, detoxification through interactions with hydrogen peroxide in the food vacuole, and a glutathione-mediated degradation process in the cytoplasm (Francis, Sullivan, and Goldberg 1997; Sigala and Goldberg 2014; Zhang, Krugliak, and Ginsburg 1999; Loria et al. 1999). Hemozoin crystallization is often thought to be the primary mechanism of heme detoxification as hemozoin can be seen in blood culture and patient blood smears; in fact, the antimalarial drug chloroquine binds to the growing ends of hemozoin crystals, preventing crystallization and detoxification (Sullivan et al. 1996; Pagola et al. 2000). However, between a third and 95% of

[8]Multiple *Plasmodium* species

heme released from hemoglobin is sequestered into hemozoin, suggesting that some is detoxified through alternative, less well characterized mechanisms (Loria et al. 1999; Nagababu, Chrest, and Rifkind 2003; Combrinck et al. 2002). Additionally, environmental factors might influence which mechanisms are used, as studies find very different values for the percent of heme crystallization. Moreover, *Plasmodium* uses only about 16% of the amino acids derived from hemoglobin (Krugliak, Zhang, and Ginsburg 2002); together these observations indicate much is still uncharacterized about hemoglobin digestion.

In addition to chloroquine, another antimalarial drug, artemisinin, interferes with hemoglobin digestion (Pandey et al. 1999). Unfortunately, the mechanism of action of artemisinin is not as clear as with chloroquine. Artemisinin in combination therapy is the current frontline antimalarial treatment, as it is highly effective against wild type parasites. In addition to interfering with hemoglobin digestion, artemisinin is known to have wide ranging effects including the generation of free radicals (Klonis et al. 2011; Juan Wang et al. 2010; Meshnick 2002; Efferth and Oesch 2004) that damage DNA, alkylate heme and proteins, and induce lipid peroxidation. Additionally, artemisinin depolarizes the parasite mitochondrial membrane potential (Antoine et al. 2014; Li et al. 2005) and potentially inhibits PfATP6 (an ortholog of the ER calcium pump SERCA (Eckstein-Ludwig et al. 2003; Arnou et al. 2011)). Artemisinin also seems to specifically target the parasite as it first invades the host cell as parasites are halted in that life stage upon treatment (Cheng, Kyle, and Gatton 2012; Klonis et al. 2013; Mok et al. 2011; Teuscher et al. 2010; Witkowski et al. 2010). Resistant parasites, however, will survive this cell cycle disruption, called dormancy (Peatey et al. 2015; Codd et al. 2011; Teuscher et al. 2010).

Artemisinin resistance is established in Southeast Asia (Ashley et al. 2014; Miotto et al. 2015; Straimer et al. 2015); in that part of the world, *P. falciparum* is the dominate malaria parasite species ("World Malaria Report 2017" 2017). The artemisinin resistance phenotype is correlated with mutations in the *P. falciparum Kelch13* gene (Ashley et al. 2014; Miotto et al. 2015; Ariey et al. 2014; Brown et al. 2015). *PfKelch13* mutations effect phosphoinositide 3-kinase (PI3K) signaling (Miotto et al. 2015; Straimer et al. 2015; Kamau et al. 2015; Isozumi et al. 2015) because *PfKelch13* facilitates the ubiquitination of PI3K and thus mutated *PfKelch13* stabilizes PI3K protein, leading to increased phosphoinositide 3-phosphate (Mbengue et al. 2015). Higher phosphoinositide 3-phosphate results in artemisinin resistance via unknown mechanisms (Mbengue et al. 2015). However, *PfKelch13* mutations alone do not confer resistance *in vitro* (Breglio et al. 2018; Siddiqui et al. 2017) and clinical resistance occurs without these mutations (Mukherjee et al. 2017).

2.1.5 Challenges

While the incidence rate of malaria has decreased by 18% globally between 2010 to 2016 ("World Malaria Report 2017" 2017), incidence of infection has increased since 2014 highlighting challenges associated with malaria control (**Figure 2.4** from ("World Malaria Report 2017" 2017)). Eradication of the disease is unlikely due to the lack of vaccine and the prevalence of multidrug resistance (to insecticides and to antimalarial treatment, shown in **Figure 2.5** from McClure and Day (2014)).

This dissertation will focus on antimalarial resistance. Because only one new antimalarial drug is in the clinical pipeline with a novel mechanism, the need for novel antimalarial drugs and combination therapies is pressing. However, it is also possible to leverage the weaknesses of antimalarial resistant parasites to develop therapeutics. Broadly, I aim to gain a new perspective on resistance by viewing it through a 'metabolic lens'. By characterizing the metabolic shifts that occur in association with resistance, we can begin to understand more about what it takes to support new functions, such as novel survival signaling, drug detoxification, or stage alterations in resistant parasites. Once we identify these compensatory changes, we can target them.

2.2 *Toxoplasma gondii*

2.2.1 Toxoplasmosis

Unlike *Plasmodium*, *Toxoplasma gondii* has high global prevalence with few fatalities. Between 4-84% of women of childbearing age are seropositive for *Toxoplasma*, with prevalence varying by country (Flegr et al. 2014). *Toxoplasma* is most frequently detected in immunocompromised or pregnant patients, as infection is asymptomatic in immunocompetent individuals. Acute infection is associated with flu-like symptoms (McAuley 2014) and the parasite is transmitted through contaminated food or water, vertically from mother to fetus, or via exposure to infected animals (Hill and Dubey 2002). Vertical transmission can be extremely dangerous (like *Plasmodium*) and, in the United States, congenital infection occurs between 1/3000 to 1/10,000 live births (McAuley 2014). In addition to the immunocompetence of the host, pathogenicity of *T. gondii* varies with parasite genetic background (Hill and Dubey 2018). Potential neurologic symptoms associated with infection have highlighted *Toxoplama* in pop science; preliminary research has implicated that host behavior shifts during chronic infection, making hosts less risk averse or at increased risk of mood disorders.

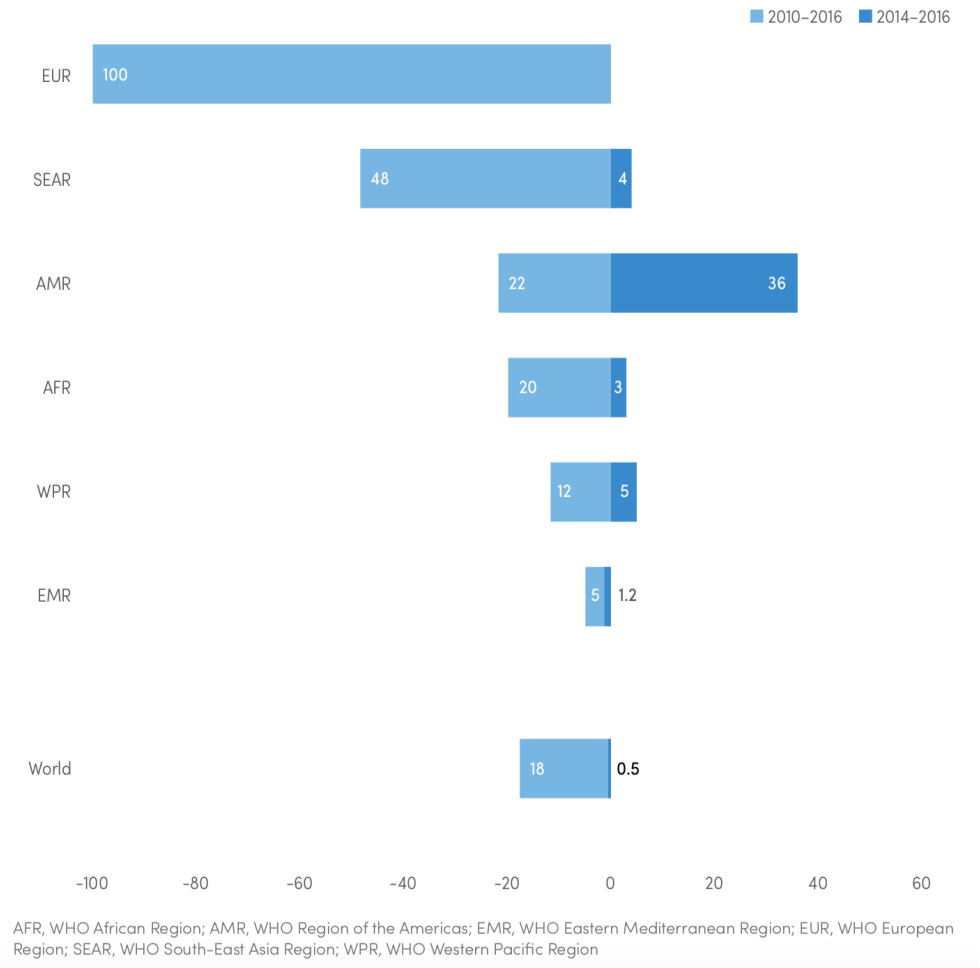


Figure 2.4: Trends in malaria incidence. Percentage change in malaria case incidence rate globally and by WHO region, 2010–2016 and 2014–2016. Image from the World Health Organization's World Malaria Report

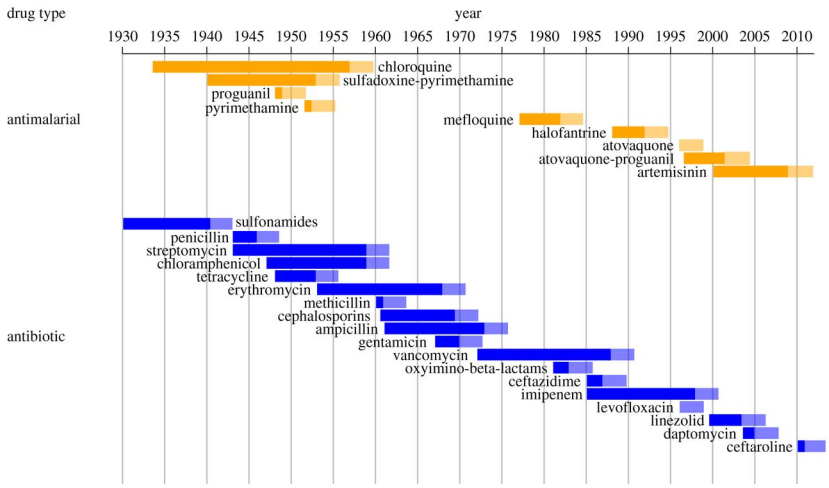


Figure 2.5: Drug resistance has emerged to every antimalarial on the market. Like the trends observed in antibiotic resistance (bottom), parasites have developed resistance to every antimalarial drug. Colored bars indicate introduction of a clinically approved antimalarial. Color changes to indicate when resistance was observed. Treatment efficacy sometimes persists after first observation of resistance (indicated by faded bar). Bar ends when the drug is no longer recommended for use. Figure from McClure and Day, 2014

2.2.2 Transmission and Pathogenesis

Toxoplama has an interesting life cycle. Although cats are the parasite's definite host, many mammals including humans can be infected (Hill and Dubey 2018). *T. gondii* can also promiscuously invade many different cell types. As a result, rapid replication throughout the body leads to a robust host immune response characterized by the production of IFN-gamma (Lieberman and Hunter 2002). In an immunocompentent host, the infection will be cleared quickly or the parasite can become dormant with recurrent reactivation of infection over time; this is called chronic infection. Dormant parasites can reside in a diverse set of tissues including skeletal and cardiac muscle, the brain, and eye, as well as (less frequently) the lungs, liver, and kidneys (Hill and Dubey 2018), and these cysts can reactivate, causing recurrent acute infection later in life. In an immunocompromized host, encephalitis is common (Randall and Hunter 2011). Lastly, in a pregnant host, the fetus can become infected causing birth defects or miscarriage; how the parasite invades the placenta is unknown (McAuley 2014).

2.2.3 Prevention and treatment

Despite the lack of a vaccine, prevention of toxoplasmosis is effective as the transmissible stage of the parasite is easily killed and avoidable. Specifically, the transmissible stage

is very sensitive to temperature change so cooking or freezing contaminated water or meat will prevent transmission (Hill and Dubey 2018). Additionally, interaction with infected animals can be avoided by at-risk individuals, such as the immunocompromised or pregnant (Hill and Dubey 2018). Acute infection is treated with antiparasitic drugs, such as atovaquone or other repurposed antimalarials (**Table 2.1**), but subclinical infections involving dormant parasites cannot be treated (Hill and Dubey 2018).

2.2.4 Parasite biology and previous research

Like the *Plasmodium* parasites, *Toxoplasma gondii* is an apicomplexan eukaryotic parasite, meaning the genera are phylogenetically related and both use an apical complex for invading a host cell. *Toxoplasma* also has the rare nonphotosynthetic plastid called an 'apicoplast.' Unlike *Plasmodium*, however, *T. gondii* can invade most nucleated cells in warm-blooded animals (Blader and Koshy 2014), including in immortalized human foreskin fibroblasts (Sidik et al. 2016). This growth promiscuity makes *T. gondii* comparatively easy to grow in laboratory culture and, as a result, *Toxoplasma* cell biology, and metabolism specifically, has been well studied. For example, NCBI's PubMed returns 7,587 hits when searching 'toxoplasma metabolism'; additionally, the genome has been better annotated than other parasites (**Table 2.2**).

There are many similarities in *Plasmodium* and *Toxoplasma* metabolism. For example, the apicoplast is essential for *T. gondii* survival (Fichera and Roos 1997) and contains the same metabolic pathways as the *Plasmodium* apicoplast. The apicoplast in *Toxoplasma* also houses enzymes involved in heme biosynthesis (Ralph et al. 2004) and fatty acid synthesis (McLeod et al. 2001). However, it has not been determined that isopentenyl pyrophosphate synthesis is the sole essential function in the *Toxoplasma* apicoplast like in *Plasmodium* (Yeh and DeRisi 2011).

In the mitochondria, *Toxoplasma gondii* has both the *de novo* and salvage pathways for select pyrimidines (Hortua Triana et al. 2016); however *de novo* synthesis is required for virulence (Fox and Bzik 2002). Thus, the antimalarial atovaquone, which targets an enzyme in the electron transport chain necessary for pyrimidine synthesis, is also effective against toxoplasmosis (Meneceur et al. 2008; McFadden et al. 2000). The parasite is auxotrophic for purines and must salvage them from the host (Chaudhary et al. 2004). *T. gondii* is also auxotrophic for tryptophan (Sibley, Messina, and Niesman 1994) and arginine (Fox, Gigley, and Bzik (2004); see **Table 2.3** for others). Similar to *Plasmodium*, *T. gondii* has maintained redundant biosynthesis and *de novo* synthesis pathways for many essential metabolites, like cholesterol (Coppens and Joiner 2001), lipoic acid (Crawford et al. 2006), and sphingolipids (Pratt et al. 2013). Unlike *Plasmodium* however, *Toxoplasma* actively uses the tricarboxylic acid

cycle during acute infection (MacRae et al. 2012) and it is essential (Fleige et al. 2008; Sidik et al. 2016).

2.2.5 Challenges

Principle challenges to reduce toxoplasmosis burden are (1) the parasite's broad host range facilitates a large reservoir of parasites and (2) chronic and recurrent disease associated with a dormant parasite phenotype. However, toxoplasmosis is a preventable and largely treatable disease, easily studied *in vitro* and in natural *in vivo* models of disease.

2.3 *Cryptosporidium*

2.3.1 Cryptosporidiosis

Cryptosporidium parasites inflict a major global health burden as a leading cause of enteric disease (Platts-Mills et al. 2015). Beyond diarrheal infection, both symptomatic and subclinical infections result in growth stunting and neurodevelopmental delay (Steiner et al. 2018). Cryptosporidiosis and malnutrition coexist in a vicious cycle: infection induces malnutrition (Korpe et al. 2016) and malnutrition is a risk factor for diarrheal disease (Liu et al. 2016).

2.3.2 Transmission and Pathogenesis

Cryptosporidium parasites are most frequently transmitted via the fecal-oral route and often through contaminated water because the transmissible stage is chlorine resistant (Korich et al. 1990). Like *Plasmodium* and *Toxoplasma*, these parasites have a complex life cycle with both asexual and sexual replication. However, this entire cycle occurs in the host's gut. Once the parasite has invaded host gut tissue, symptomatic infection presents with gastroenteritis, including severe watery diarrhea (Bouzid et al. 2013). Specific virulence factors are not known due to the lack of experimental manipulatability of the parasite, but different isolates are associated with variable symptoms such as intestinal villus blunting; how much of this variability is parasite or host-derived is unknown (Medema et al. 2009; Sayed et al. 2016). Disease in immunocompetent individuals is self-limiting (2-3 weeks), but children or immunocompromised individuals can become extremely dehydrated (Bouzid et al. 2013). Surprisingly, even subclinical infection is associated with negative patient outcomes in developing countries. Specifically, *Cryptosporidium* infection in children

is associated with growth stunting and neurodevelopmental delay and the mechanism of pathogenesis is not understood (Schnee et al. 2018; Steiner et al. 2018).

2.3.3 Prevention and treatment

Unfortunately, both prevention and treatment of cryptosporidiosis remain challenging. No vaccine exists for this disease (Sparks et al. 2015) and the transmissible stage of the parasite is resistant to chlorine treatment, making contaminated water a major mode of transmission (Korich et al. 1990). Importantly, the currently available drugs (nitazoxanide and azithromycin, **Table 2.1**) are ineffective in the most vulnerable populations (Sparks et al. 2015).

2.3.4 Parasite biology and previous research

Very limited data exists for these parasites due to the historic lack of *in vitro* culture system. Several experimental systems using organoids have been developed in the past year, but neither have been reproduced (Heo et al. 2018; Baydoun et al. 2017). Thus, what is known about *Cryptosporidium* is inferred from the genome and orthologous enzymes. Although the *Cryptosporidium* parasites are apicomplexans, they do not have an apicoplast like *Plasmodium* or *Toxoplasma* (Abrahamsen et al. 2004; Xu et al. 2004). Thus, the functions typically associated with this organelle, like fatty acid and heme synthesis are thought to occur in the cytoplasm for these parasites and are catalyzed by nonorthologous enzymes (Abrahamsen et al. 2004; Xu et al. 2004). *Cryptosporidium* do have a mitochondria, but it lacks a mitochondrial genome (Abrahamsen et al. 2004; Xu et al. 2004). The genomes lack enzymes from the tricarboxylic acid cycle and electron transport chain, and are accordingly thought to be glycolytic (Abrahamsen et al. 2004; Xu et al. 2004). Thus, energy metabolism is functionally similar to the blood-stage *Plasmodium* parasites.

Unlike *Plasmodium*, *Cryptosporidium* genomes do not encode any pyrimidine synthesis enzymes and, therefore, are assumed to scavenge host pyrimidines and purines for growth (Hyde 2008); thus, the adaptation of antimalarials like atovaquone to treat cryptosporidiosis is not possible as with toxoplasmosis (**Table 2.1**; Giacometti, Cirioni, and Scalise (1996)). *Cryptosporidium* genomes also do not contain any genes encoding genes involved in folate synthesis (Hyde 2008), unlike *Plasmodium* and *Toxoplasma* which both contain these genes and have been shown to synthesize and import folates (Salcedo-Sora et al. 2011; Metz 2007; Massimine et al. 2005; Aspinall et al. 2002). Accordingly, *Cryptosporidium* would not be sensitive to antimalarial drugs that target folate synthesis (**Table 2.1**).

Cryptosporidium parasites cannot synthesize amino acids *de novo* and encode many amino acid transporters (Abrahamsen et al. 2004; Xu et al. 2004). See **Table 2.3** for other known metabolic auxotrophies of *Cryptosporidium*. Interestingly, *Cryptosporidium* has laterally acquired several enzymes from neighboring gut bacteria, including tryptophan synthase B, which converts serine to tryptophan (Sateriale and Striepen 2016). This prokaryote-derived enzyme allows bacteria to tolerate a host immune response, specifically tryptophan starvation via IFN-gamma-mediated conversion of tryptophan to kyurenine, and might serve a similar role in *Cryptosporidium* survival (Sateriale and Striepen 2016). Additionally, these parasites have acquired inosine 5'-monophosphate dehydrogenase and thymidine kinase via lateral transfer; these enzymes mediate the conversion of nucleotides from scavenged precursors (Sateriale and Striepen 2016; Huang et al. 2004).

2.3.5 Challenges

Challenges associated with cryptosporidiosis control are numerous. First, disease surveillance is challenging. Diarrheal disease goes unreported as surveillance is often conducted through hospitals (*e.g.* Saluja et al. (2014)). Thus, many less severe cases are not reported. Additionally, it is challenging to pinpoint the causal pathogen in many diarrheal episodes (Platts-Mills et al. 2015). Many other diarrheal pathogens are spread by the same transmission mechanisms and, thus, coinfection is common in developing nations (Platts-Mills et al. 2015). Without understanding the scope of transmission or the burden of disease, it is challenging to unroll effective prevention and treatment campaigns. Second, no drugs are approved for vulnerable populations (Sparks et al. 2015).

Lastly, there are poor experimental systems for interrogation of the parasite and thus very limited data exists. There is no *in vitro* culture system for any *Cryptosporidium* species (Baydoun et al. 2017); efficient genetic and pharmacologic inhibitor screens for drug development cannot be performed without a culture system. Additionally, the historic models of disease, *C. parvum* in a protein deficient mouse (Liu et al. 2016) or in an immunocompromised mouse (*i.e.* Mead et al. (1991)) fail to replicate features of disease, such as a self-clearing infection. Thus, the potential drug targets against the parasite as well as much of the immune response to cryptosporidiosis remain uncharacterized. However, both a novel natural model of cryptosporidiosis is now available (*unpublished*, using *Cryptosporidium tyzzeri*) and an organoid-based ('mini gut') culture system (Heo et al. 2018) have been developed recently. These new technologies have the capacity to accelerate biological characterization of these important human pathogens and drug development.

2.4 Project motivation and shared challenges

Plasmodium, *Toxoplasma*, and *Cryptosporidium* parasites kill patients around the world despite there being multiple treatment options for each infection. Novel therapeutics are still necessary because (1) antiparasitic drugs have stage-specificities, meaning they cannot target all stages of a parasites development (*i.e.* they may not cure infection), (2) the parasites have evolved drug resistance, rendering old drugs useless, and (3) some treatment options cannot be used in children or pregnant women, a large proportion of the vulnerable patients. Novel therapeutics are challenging to identify because (1) parasite genomes are poorly annotated, especially regarding unique features of parasites, (2) many *Plasmodium* species and all *Cryptosporidium* species cannot be grown *in vitro* preventing the use genetic or inhibitor screens, and (3) immunological therapeutics cannot be extrapolated across parasite and host species, slowing vaccine and immunotherapy development. Thus, we focus on computational and experimental approaches that generate high-confidence hypotheses to accelerate labor-intensive experimental work and leverage existing experimental data to combat parasitic diseases.

Chapter 3: Building a parasite knowledgebase

The work presented in this chapter was augmented by the work of two coauthors, Gregory Medlock and Ana Untariou. Specifically, Gregory Medlock assisted with adapting analyses to a high-performance computing environment and Ana Untariou curated the *Plasmodium falciparum* 3D7 metabolic reconstruction regarding glutathione-dependent heme degradation.

3 Building a parasite knowledgebase

Modeling approaches, such as genome-scale metabolic modeling, highlight knowledge gaps, generate high-confidence data-driven hypotheses, and contextualize sparse data. Genome-scale metabolic models are built from genomic data and by inferring function to complete or connect metabolic pathways; these reconstructions are supplemented with functional genetic and biochemical studies. These models represent our best understanding of an organism's biochemistry and cell biology, and failed predictions illuminate knowledge gaps or unappreciated experimental complexities. Here, we focus on the construction of parasite genome-scale metabolic models from genetic data to develop a parasite knowledgebase, **Para**site **D**atabase for **G**enome-scale metabolic **M**odels (referenced as Paradigm) and the application of this knowledgebase to better utilize existing experimental data and for drug development.

3.1 Synopsis

Experimentally tractable model organisms are used to interrogate disease and parasite phenotypes, but characterization of functional differences between parasite species is limited to *post hoc* and single target studies. Each parasite genome encodes unique enzymes; however, it is unclear whether these differences arise from divergent metabolism or incomplete genome annotation. To address this challenge, we generated metabolic reconstructions from 162 parasite genomes; with these 162 metabolic reconstructions, representing 38 genera and 111 species, we compare metabolic capacity, gene essentiality, and pathway utilization. All parasite genomes encode unique metabolic functions, regardless of genome size, and parasites within the same genera tend to have similar network topology overall. Host cell type is associated with genetically-encoded transport ability from the extracellular environment to the parasite cytoplasm as well. Lastly, we highlight differences in kinase reactions among reconstructions and the potential effect on antiparasitic inhibitor screens.

3.2 Background

Data collected in one eukaryotic pathogen are frequently extrapolated to infer about another parasite, across and within genera. For example, model *in vitro* systems or *in vivo* systems, like mouse models of disease, are frequently used due to enhanced experimental manipulability. However, characterization of functional differences among parasite species is limited to *post hoc* and single target studies, limiting the utility of this extrapolation approach. To address this challenge, we generated metabolic

reconstructions for 162 genomes to serve as a knowledgebase for each parasite and to leverage comparative functional genomic approaches. These 162 genomes represent 111 species in the *Plasmodium*, *Toxoplasma*, *Cryptosporidium*, *Entamoeba*, *Trypanosoma*, *Leishmania*, and *Giardia* genera, and include all publically available genomes for human eukaryotic pathogens. We identify similarities and differences in gene essentiality and pathway utilization; this knowledge facilitates comparison of experimental findings. Using this approach, we can improve genome annotation, identify species-specific functions, interpret experimental results, and optimize selection of an experimental system for fastidious species.

3.3 Methods

All analytic code is available on my GitHub page, see Appendix. R (R Core Team 2017) and R packages tidyverse, ggdendro, seqinr, Biostrings, msa, and reshape2 were used for data processing and analysis (Wickham 2017, 2012; Vries and Ripley 2013; Charif and Lobry 2007; Pages et al., n.d.; Bodenhofer et al. 2015). For visualization of annotation similarity, we also used the package UpSetR (Gehlenborg 2017) and ggdendro (Vries and Ripley 2013). Python 3.6.4, pandas, CobraPy 0.13.0 (Ebrahim et al. 2013), and select code from CarveMe (Machado et al. 2018) and Memote (Lieven et al. 2018) were used for genome-scale metabolic modeling.

Genomic Analyses: Genomes and amino acid sequences were obtained from EuPathDB release 38 (Aurrecoechea et al. 2017). EuPathDB curates and compiles genome annotation for all genomes hosted by the database. Genome annotations are defined here as the identification of open reading frames through bioinformatic and experimental approaches followed by the mapping of these open reading frames to functional annotations, if possible. We used the open reading frames identified on EuPathDB and supplemented EuPathDB functional annotations with *de novo* Diamond annotations, described in the next paragraph. EuPathDB's OrthoMCL was also used for mapping orthology between *Plasmodium* species. In brief, orthology was mapped within each EuPathDB database by the 'map by orthology' tool from the genome of each organism with a curated reconstruction to all other genomes within that database. The search protocol was 'new search > genes > taxonomy > organism [pick] > transform by orthology'.

We mapped each organism's amino acid sequences using Diamond annotation (Buchfink, Xie, and Huson 2015) against proteins referenced in the BiGG databases (King et al. 2016) or against protein sequences obtained from OrthoMCL, part of EuPathDB that contains orthologous groups of parasite genes (Li, Stoeckert, and Roos 2003). Diamond is a similar approach to BLAST, with sensitive and fast performance

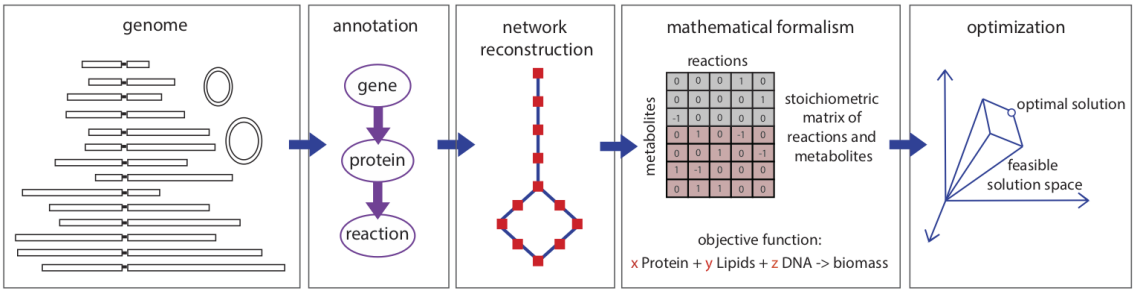


Figure 3.1: Metabolic modeling. Genome-scale metabolic models are build from an organism's genome. Open reading frames are annotated and mapped to enzymes. Enzymes are mapped to metabolic reactions, which are compiled into a metabolic reconstruction. This is represented mathematically such that linear optimization and mathematical constraints can be applied. This framework allows us to explore the metabolic capacity of an organism and to generate predictions about metabolic functions. Figure adapted from Medlock, Gregory Leonard, Maureen A. Carey, Dennis G. McDuffie, Michael B. Mundy, Natasa Giallourou, Jonathan R. Swann, Glynis Kolling, and Jason A. Papin. Metabolic mechanisms of interaction within a defined gut microbiota. *bioRxiv* (2018): 250860.

on protein annotations (Buchfink, Xie, and Huson 2015).

Model Generation: We generated draft reconstructions (**Figure 3.1**) by first annotating each organism's amino acid sequences, obtained from EuPathDB (Aurrecoechea et al. 2017), using Diamond annotation (Buchfink, Xie, and Huson 2015) against proteins referenced in the BiGG databases (King et al. 2016). We next mapped all functional annotations to reactions contained in the BiGG database (King et al. 2016), inspired by the approach conducted with the reconstruction pipeline CarveMe (Machado et al. 2018). Detailed methods are included in the analytic code hosted on my GitHub page, see Appendix.

Unlike the CarveMe approach (Machado et al. 2018), we included all high-scoring reactions. Rather than maximizing the number of high-scoring hits included in the network to build a functional network, we added *all* reactions identified via orthologous genes to draft models. This conservative approach generates broadly inclusive, but incomplete reconstructions (*i.e.* that are not able to produce biomass until gapfilled, see **Chapter 7.2.1**). However, this approach added redundant reactions from multiple different compartments (*e.g.* peroxisome, mitochondria, and cytosol) so all reaction versions other than the cytosolic version were pruned unless contained in a relevant compartment. Relevant compartments are defined in **Table 3.1** (see also **Figure 2.3**); for genera not included in **Table 3.1**, only the cytosol and extracellular space were used. For example, if a *Plasmodium* reconstruction contained a reaction in the cytosol, mitochondria, and chloroplast, only the cytoplasmic and mitochondrial versions would be kept.

Table 3.1: Subcellular compartments by genus.

Species/Database	Compartments
BiGG database	cytosol, extracellular, mitochondria, nucleus, lysosome, chloroplast, golgi, vacuole, endoplasmic reticulum, peroxisome/glyoxysome, flagellum, periplasm, thylakoid, thylakoid membrane, cytosolic membrane, carboxyzome, intermembrane space of mitochondria, eyespot, unidentified
Plasmodium	cytosol, extracellular, mitochondria, apicoplast, food vacuole
Leishmania	cytosol, extracellular, mitochondria, kinetoplast, glycosome
Cryptosporidium	cytosol, extracellular, pseudomitochondria
Toxoplasma	cytosol, extracellular, mitochondria, apicoplast
Giardia	cytosol, extracellular
Entamoeba	cytosol, extracellular
All else	cytosol, extracellular

Note:
Note: the BiGG database contains prokaryotes and eukaryotes.

Following this step, a large percentage of each reconstruction's reactions remained in unsupported compartments as there was no analogous cytosolic reaction. Thus, reactions only found in an unsupported compartment were moved to the extracellular space or cytosol; specifically, periplasmic metabolites were moved to the extracellular space and all internal subcompartment metabolites were moved to the cytosol. However, this step removed all reactions that summarized a transport reaction from the extracellular space to periplasm or from the cytosol to subcompartment. Note, the extracellular compartment of the model corresponds to the parasitophorous vacuole space contained within the host cell for intracellular parasites (*i.e. Plasmodium*, *Toxoplasma*, *Cryptosporidium*) and the host serum for extracellular parasites (*i.e. Trypanosoma in vitro*).

Manual Curation: We performed brief manual curation from literature sources, building on our curation conducted in (Carey, Papin, and Guler 2017) and presented in **Chapter 4**. **Table 3.2** contains all modifications resulting from our literature review; see code for implementation. Networks were manually curated with 8 types of modifications to improve our *Plasmodium falciparum* reconstruction iPfal17 (Carey, Papin, and Guler 2017); this manual curation (combined with automated curation conducted in **Chapter 6.2.2**) generated a new model, named iPfal18. In brief, we incorporate experimental evidence collected from the literature, leveraging *in vitro* experiments. For example, if no gene supports import of metabolite X, but media supplementation of metabolite X rescues inhibition of *de novo* synthesis of metabolite X, we added a transporter. For more information, see **Chapter 4**. Because manual

curation can (and should) be conducted on all models presented in this chapter, we present the majority of the curation process in detail in the next chapter (**Chapter 4**) to demonstrate the canonical modeling workflow (also visualized in **Figure 3.2A** and **B**).

Table 3.2: Manual modifications to our draft recontructions.

Model	Modification
iPfal17	switch _D_ to ___D_ to be BiGG compatible
iPfal17	update metabolite identifiers to be BiGG compatible
iPfal17	update gene IDs to newest PlasmoDB nomenclature
iPfal17	updated reaction IDs to be BiGG compatible
iPfal17	remove reaction (hcys_ex) because it is a duplicate with 'EX_hcys____L_e'
iPfal17	added necessary exchange reactions
iPfal17	added glutathione-dependent heme degradation
iPfal17	added H2O2 production by hemoglobin degradation

Automatic curation: We developed a novel automated curation approach using orthologous transformation, similar to the approach taken by Abdel-Haleem et al. (2018). This approach leverage the curation conducted in one organism for closely related organisms. We applied this approach to all *Plasmodium* reconstructions (generated using Diamond annotation to the BiGG database, see **Figure 3.2A**) using iPfal18 (**Figure 3.2B**, see **Chapter 4**). We first mapped orthology of *P. falicparum* to each other *Plasmodium* species to build an orthology thesaurus (**Figure 3.2C**). We then added genes and associated reactions from iPfal18 if there was an orthologous gene in the target species' reconstruction (**Figure 3.2D**) resulting in a mean 42 genes added (SD = 10.38) and a mean 113 reactions added (SD = 4.14) to the draft reconstruction (**Table 3.3**). Notably, this approach facilitates the compartmentalization of these reconstructions, a function many automated pipelines fail to include. This is particularly important for parasite-specific compartments like the apicoplast, which is not included in any database.

Gapfilling: Gapfilling is an analytic process used to bridge or complete genetically-supported metabolic pathways to permit the network to fulfill metabolic functions, and was used to generate functional *Plasmodium* models. The approach is described in more detail in **Chapter 7.2.1** and implications are described in **Biomass Formulation**. We scored gapfilled reactions to summarize the confidence of reaction addition. In short, each gapfilled reaction has a score associated with it for each type of gapfilling performed. We gapfilled for three or four objective functions (see next section) for ten iterations each. Gapfilling confidence is based on how

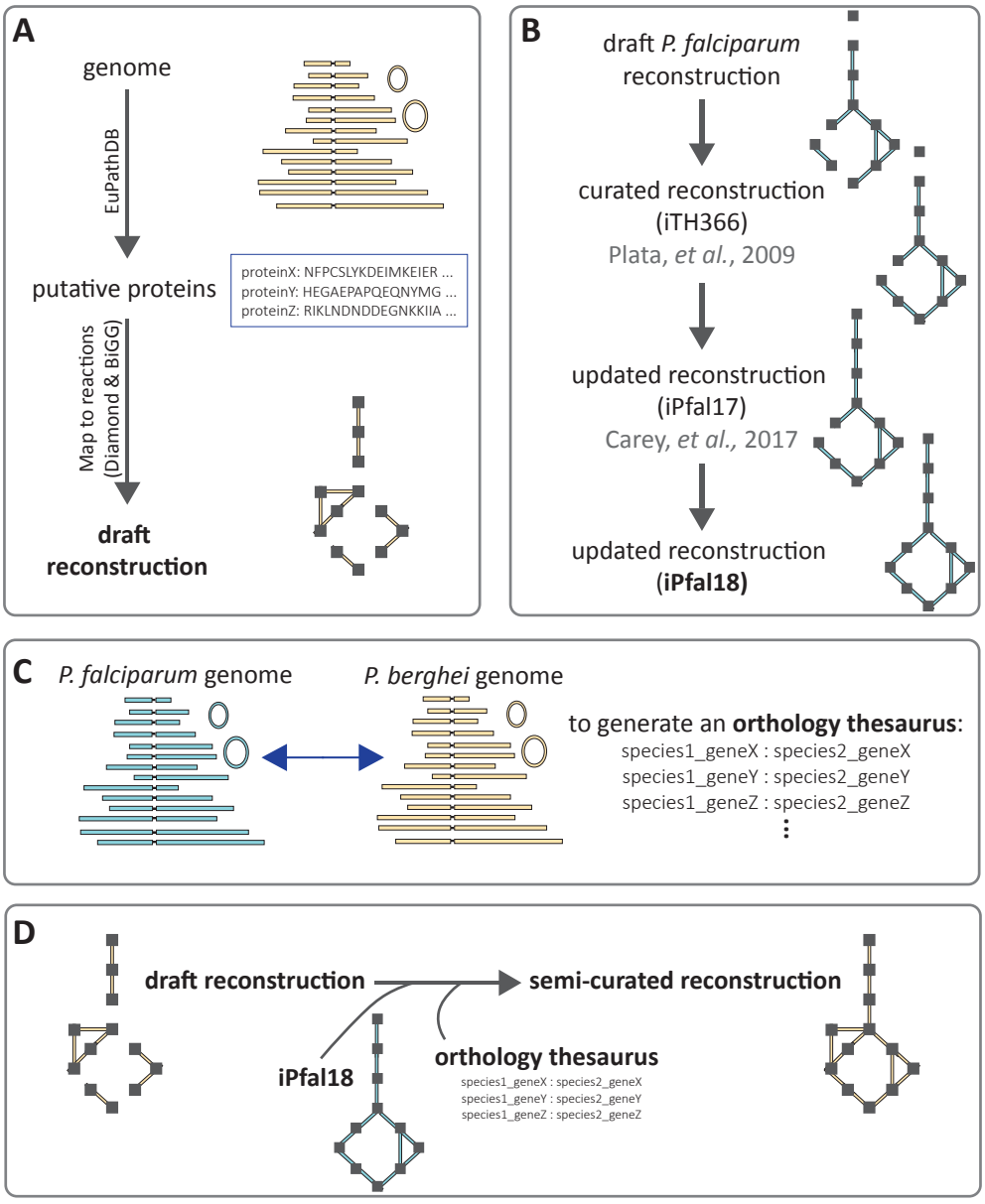


Figure 3.2: Parasite Database for Genome-Scale Metabolic Models (Paradigm). A: *De novo* reconstructions are built from putative protein sequences from EuPathDB for each organism. Amino acid sequences are mapped to functional domains using Diamond. Next draft reconstructions are built by mapping functional domains to known metabolic reactions. B: Reconstruction curation is an iterative process. For *P. falciparum.*, we have curated two versions of the reconstruction, based on a network published by Plata, et al. C: Mapped orthology from the EuPathDB databases are used to generated an orthology thesaurus, or a list of gene identifiers that have orthologous functions. D: Draft reconstructions are curated into a semi-curated reconstruction by adding all orthologous genes and associated reactions that are found in our well-curated reconstruction. This approach leverages the curation conducted in one organism to closely related organisms.

Table 3.3: Number of modifications added to each *Plasmodium* reconstruction based on orthology transformation from iPfal18.

species	Starting no. of genes	Reactions added	Metabolites added	Genes added
PpraefalciparumG0	414	113	83	39
PyoeliiyoeliiYNL	395	113	81	50
PinuiSanAntonio1	400	112	81	35
PbillcollinsiG01	396	113	83	39
PyoeliiyoeliiYX	404	113	81	44
PreichenowiG01	411	113	83	39
PcynomolgiB	382	112	81	36
Pgallinaceum8A	412	113	81	44
PvivaxP01	406	113	81	44
PgaboniG01	408	113	83	39
PyoeliiyoeliiYM	403	113	81	44
PreichenowiCDC	418	114	83	47
PadleriG01	415	113	83	39
PmalariaeUG01	402	113	81	45
PfalciparumIT	423	135	83	92
PovalecurtisiGH01	407	112	81	36
PgaboniSY75	412	113	83	39
PvinckeipetteriCR	413	112	81	36
Pchabaudichabaudi	414	113	81	44
PcynomolgiM	407	113	81	44
PknowlesiMalayanP	403	112	81	36
Pfalciparum3D7	422	110	80	33
PrelictumSGS1-like	410	113	81	44
PcoatneyiHackeri	402	112	81	35
PknowlesiH	408	113	81	44
Pvinckeivinckeivincl	404	112	81	35
PfragileNilgiri	405	112	81	35
PvivaxSal1	414	113	81	45
PbergheiANKA	401	113	81	44
PblacklockiG01	406	113	83	39

frequently a reaction is added in any of the gapfilling solutions and is noted as follows. For example, a reaction with the score 'OF3_1:1' appeared in 100% of solutions, but only one solution was generated, whereas a reaction with score 'OF3_3:2' as necessary in two out of three solutions. These scores are formatted as ObjectiveFunction_Y:X, with 'ObjectiveFunction' indicating which objective functions were used for gapfilling (note: this ranges from three to four for this study), 'X' indicating the number of times a reaction is added, and 'Y' indicating the number of iterations used to solve each gapfilling problem. Gapfilling was conducted following all steps involving compartmentalization, manual curation, or automated curation.

Objective functions: We use two classes of objective functions here to robustly

evaluate *Plasmodium* model performance. First, we maximize ATP production. Second, we use biomasses reactions, including a species-specific curated biomass reaction and a generic biomass reaction. *Plasmodium falicparum* has high-quality biomass reaction (Carey, Papin, and Guler 2017) and this was used for all *Plasmodium* models. Our generic biomass contains metabolites from several curated reconstructions (**Table 3.4**) and thus contains metabolites from the *Plasmodium falicparum*, *Leishmania major*, and *Cryptosporidium hominis* species-specific biomasses with the stoichometry contained in the iPfal18 biomass reaction; *Toxoplasma gondii* is excluded from this analysis due to lack of direct mapping of id nomenclature (*e.g.* metabolite ID C00498 in **Table 3.4**). This generic biomass was used to capture the (most conservatively defined) required biosynthetic capacity.

Table 3.4: Variable biomass components for published reconstructions.

C. hominis (2010)	Leishmania (2008)	T. gondii (2015)	iPfal17
adp	1-2-Diacylglycerol	(9Z)-Octadecenoic acid	10-Formyltetrahydrofolate
ala-L	ADP	1-Phosphatidyl-D-myo-inositol	2-Octaprenyl-6-hydroxyphenol
amp	AMP	Acetyl-CoA	2-Oxoglutarate
arg-L	ATP	Arachidonate	5-10-Methylenetetrahydrofolate
asn-L	Cardiolipin	ATP	5-6-7-8-Tetrahydrofolate
asp-L	CMP	ATP	ADP
atp	dAMP	C00498	Ammonium
cdlp_CT	dCMP	C00550	ATP
cdp	dGMP	C02679	biomass
cmp	dTMP	C04574	Coenzyme-A
ctp	Ergosterol	C06424	CTP
cys-L	Glycine	Cholesterol	dATP
d12dg_CT	GMP	CoA	dCTP
datp	H2O	CTP	dGTP
dctp	H	dATP	Diphosphate
dgtp	L-Alanine	dCTP	dTTP
dttp	L-Arginine	dGTP	Fe2
gdp	L-Asparagine	dTTP	Fe3
gln-L	L-Aspartate	FattyAcid	Flavin-adenine-dinucleotide-oxidized
glu-L	L-Cysteine	FMN	Glycine
gly	L-Glutamate	GDP-mannose	GTP
gmp	L-Glutamine	Geranylgeranyl diphosphate	H
gtp	L-Histidine	Glutathione	H2O
h	L-Isoleucine	Glycine	hemozoin[e]
his-L	L-Leucine	GTP	L-Alanine
ile-L	L-Lysine	H2O	L-Arginine
leu-L	L-Methionine	Heme	L-Asparagine
lys-L	L-Phenylalanine	Hexadecanoic acid	L-Aspartate

Table 3.4: Variable biomass components for published reconstructions. *(continued)*

C. hominis (2010)	Leishmania (2008)	T. gondii (2015)	iPfal17
lysylpgly_CT	L-Proline	L-Alanine	L-Cysteine
m12dg_CT	L-Serine	L-Arginine	L-Glutamate
met-L	L-Threonine	L-Asparagine	L-Glutamine
nad	L-Tryptophan	L-Aspartate	L-Histidine
nadp	L-Tyrosine	L-Cysteine	L-Isoleucine
nadph	L-Valine	L-Glutamate	L-Leucine
pgly_CT	Mannan	L-Glutamine	L-Lysine
phe-L	monoacylglycerol	L-Histidine	L-Malate
pi	Phosphate-HO4P	L-Isoleucine	L-Methionine
ppi	phosphatidyl-1D-myo-inositol	L-Leucine	L-Phenylalanine
pro-L	Phosphatidylcholine	L-Lysine	L-Proline
psetha_CT	phosphatidylethanolamin	L-Methionine	L-Serine
ser-L	Putrescine	L-Phenylalanine	L-Threonine
t12dg_CT	Spermidine	L-Proline	L-Tryptophan
thr-L	Triglyceride	L-Selenocysteine	L-Tyrosine
trp-L	UMP	L-Serine	L-Valine
tyr-L	Zymosterol	L-Threonine	lac-L[e]
utp		L-Tryptophan	lipid
val-L		L-Tyrosine	Nicotinamide-adenine-dinucleotide
		L-Valine	Nicotinamide-adenine-dinucleotide-phosphate
		Linoleate	Oxidized-glutathione
		NAD+	Phosphate
		NADP+	protein
		Octadecanoic acid	Protoheme
		Oxaloacetate	Putrescine
		Phosphatidylcholine	Pyridoxal-5-phosphate
		Phosphatidylethanolamin	Reduced-glutathione
		Phosphatidylserine	Riboflavin
		Protein N6-(lipoyl)lysine	S-Adenosyl-L-methionine
		Protein N6-(lipoyl)lysine	Spermidine
		Pyridoxal phosphate	Sulfate
		S-Adenosyl-L-methionine	Thiamine-diphosphate
		Tetrahydrobiopterin	UTP
		Tetrahydrofolate	
		Thiamin diphosphate	
		UDP-N-acetyl-alpha-D-glucosamine	
		UTP	

Model Comparison: Networks were compared by euclidean distance of reaction presence or of transporter capability. Transporter capability was identified by the presence of a reaction in the reconstruction (prior to gapfilling) that transported

a metabolite from the extracellular compartment to the intracellular compartment. Thus, only genetically-supported transporters were analyzed.

3.4 Results

Comparative genomics in the field of eukaryotic pathogens and apicomplexan parasites has primarily been limited to the study of parasite surface proteins that interact with the host. Accordingly, we first explore an unbiased comparative genomics analysis using 162 publically available genome sequences from the EuPathDB databases (**Table 3.5**); each EuPathDB is a rough phylogenetic grouping containing only organisms from one genus or several closely related genera.

Table 3.5: Genomes from EuPathDB.

species	database
AcastellaniiNeff	AmoebaDB
EdisparSAW760	AmoebaDB
EhistolyticaHM1IMSS	AmoebaDB
EhistolyticaHM1IMSS.A	AmoebaDB
EhistolyticaHM1IMSS.B	AmoebaDB
EhistolyticaHM3IMSS	AmoebaDB
EhistolyticaKU27	AmoebaDB
EinvadensIP1	AmoebaDB
EmoshkovskiiLaredo	AmoebaDB
EnuttalliP19	AmoebaDB
NfowleriATCC30863	AmoebaDB
Candersoni30847	CryptoDB
Chominis30976	CryptoDB
ChominisTU502	CryptoDB
ChominisTU502_2012	CryptoDB
ChominisUdeA01	CryptoDB
CmeleagridisUKMEL1	CryptoDB
CmurisRN66	CryptoDB
CparvumIowaII	CryptoDB
CtyzzeriUGA55	CryptoDB
Cubiquitum39726	CryptoDB
CveliaCCMP2878	CryptoDB
GniphandrodesUnknown	CryptoDB
VbrassicaformisCCMP3155	CryptoDB
GintestinalisAssemblageADH	GiardiaDB
GintestinalisAssemblageAWB	GiardiaDB
GintestinalisAssemblageBGS	GiardiaDB
GintestinalisAssemblageBGS_B	GiardiaDB
GintestinalisAssemblageEP15	GiardiaDB
SsalmonicidaATCC50377	GiardiaDB
AalgeraePRA109	MicrosporidiaDB
AalgeraePRA339	MicrosporidiaDB

Table 3.5: Genomes from EuPathDB. *(continued)*

species	database
EaedisUSNM41457	MicrosporidiaDB
EbieneusiH348	MicrosporidiaDB
EcuniculiEC1	MicrosporidiaDB
EcuniculiEC2	MicrosporidiaDB
EcuniculiEC3	MicrosporidiaDB
EcuniculiGBM1	MicrosporidiaDB
EhellemATCC50504	MicrosporidiaDB
EhellemSwiss	MicrosporidiaDB
EintestinalisATCC50506	MicrosporidiaDB
EromaleaeSJ2008	MicrosporidiaDB
MdaphniaeUGP3	MicrosporidiaDB
NausubeliERTm2	MicrosporidiaDB
NausubeliERTm6	MicrosporidiaDB
NbombycisCQ1	MicrosporidiaDB
NceranaeBRL01	MicrosporidiaDB
NdisplodereJUm2807	MicrosporidiaDB
NparisiiERTm1	MicrosporidiaDB
NparisiiERTm3	MicrosporidiaDB
OcolligataOC4	MicrosporidiaDB
PneurophiliaMK1	MicrosporidiaDB
Slophii42_110	MicrosporidiaDB
ThominisUnknown	MicrosporidiaDB
VcorneaeATCC50505	MicrosporidiaDB
Vculicisfloridensis	MicrosporidiaDB
BbigeminaBOND	PiroplasmaDB
BbovisT2Bo	PiroplasmaDB
BmicrotiRI	PiroplasmaDB
BovataMiyake	PiroplasmaDB
CfelisWinnie	PiroplasmaDB
TannulataAnkara	PiroplasmaDB
TequiWA	PiroplasmaDB
TorientalisShintoku	PiroplasmaDB
TparvaMuguga	PiroplasmaDB
PadleriG01	PlasmoDB
PbergheiANKA	PlasmoDB
PbillcollinsiG01	PlasmoDB
PblacklockiG01	PlasmoDB
Pchabaudichabaudi	PlasmoDB
PcoatneyiHackeri	PlasmoDB
PcynomolgiB	PlasmoDB
PcynomolgiM	PlasmoDB
Pfalciparum3D7	PlasmoDB
PfalciparumIT	PlasmoDB
PfragileNilgiri	PlasmoDB
PgaboniG01	PlasmoDB
PgaboniSY75	PlasmoDB
Pgallinaceum8A	PlasmoDB
PinuiSanAntonio1	PlasmoDB

Table 3.5: Genomes from EuPathDB. *(continued)*

species	database
PknowlesiH	PlasmoDB
PknowlesiMalayanPk1A	PlasmoDB
PmalariaeUG01	PlasmoDB
PovalecurtisiGH01	PlasmoDB
PpraefalciparumG01	PlasmoDB
PreichenowiCDC	PlasmoDB
PreichenowiG01	PlasmoDB
PrelictumSGS1.like	PlasmoDB
PvinckeipetteriCR	PlasmoDB
Pvinckeivinckeivinckei	PlasmoDB
PvivaxP01	PlasmoDB
PvivaxSal1	PlasmoDB
Pyoeliiyoelii17X	PlasmoDB
Pyoeliiyoelii17XNL	PlasmoDB
PyoeliiyoeliiYM	PlasmoDB
CcayetanensisCHN_HEN01	ToxoDB
CsuisWienI	ToxoDB
EacervulinaHoughton	ToxoDB
EbrunettiHoughton	ToxoDB
EfalciformisBayerHaberkorn1970	ToxoDB
EmaximaWeybridge	ToxoDB
EmitisHoughton	ToxoDB
EnecatrixHoughton	ToxoDB
EpraecoxHoughton	ToxoDB
EtenellaHoughton	ToxoDB
HhammondiHH34	ToxoDB
NcaninumLIV	ToxoDB
SneuronaSN3	ToxoDB
SneuronaSOSN1	ToxoDB
TgondiiARI	ToxoDB
TgondiiFOU	ToxoDB
TgondiiGAB2.2007.GAL.DOM2	ToxoDB
TgondiiGT1	ToxoDB
TgondiiMAS	ToxoDB
TgondiiME49	ToxoDB
Tgondiip89	ToxoDB
TgondiiRH	ToxoDB
TgondiiRUB	ToxoDB
TgondiiTgCatPRC2	ToxoDB
TgondiiVAND	ToxoDB
TgondiiVEG	ToxoDB
TvaginalisG3	TrichDB
BayalaiB08.376	TriTrypDB
CfasciculataCfCl	TriTrypDB
EmonterogeiiLV88	TriTrypDB
LaethiopicaL147	TriTrypDB
LarabicaLEM1108	TriTrypDB
LbraziliensisMHOMBR75M2903	TriTrypDB

Table 3.5: Genomes from EuPathDB. *(continued)*

species	database
LbraziliensisMHOMBR75M2904	TriTrypDB
LdonovaniBPK282A1	TriTrypDB
LenriettiiLEM3045	TriTrypDB
LgerbilliLEM452	TriTrypDB
LinfantumJPCM5	TriTrypDB
LmajorFriedlin	TriTrypDB
LmajorLV39c5	TriTrypDB
LmajorSD75	TriTrypDB
LmexicanaMHOMGT2001U1103	TriTrypDB
LpanamensisMHOMCOL81L13	TriTrypDB
LpanamensisMHOMPA94PSC1	TriTrypDB
LpyrrhocorisH10	TriTrypDB
LseymouriATCC30220	TriTrypDB
LspMARLEM2494	TriTrypDB
LtarentolaeParrotTarII	TriTrypDB
LtropicaL590	TriTrypDB
LturanicaLEM423	TriTrypDB
PconfusumCUL13	TriTrypDB
TbruceigambienseDAL972	TriTrypDB
TbruceiLister427	TriTrypDB
TbruceiTREU927	TriTrypDB
TcongolenseIL3000	TriTrypDB
TcruziCLBrener	TriTrypDB
TcruziCLBrenerEsmeraldo.like	TriTrypDB
TcruziCLBrenerNon.Esmeraldo.like	TriTrypDB
TcruzicruziDm28c	TriTrypDB
TcruziDm28c	TriTrypDB
TcruzimarinkelleiB7	TriTrypDB
TcruziSylvioX10.1	TriTrypDB
TcruziSylvioX10.1.2012	TriTrypDB
TevansiSTIB805	TriTrypDB
TgrayiANR4	TriTrypDB
TrangeliSC58	TriTrypDB
TtheileriEdinburgh	TriTrypDB
TvivaxY486	TriTrypDB

Sequence-based analyses investigating genetic similarity can be biased by AT content, genome alignment and assembly, and structural genomic variants such as genome size, presence of plastid genomes, or the number of chromosomes. These organisms contain variable genome sizes (*e.g.* **Table 2.2**), organellar genomes (*e.g.* **Table 3.1**), and chromosome number. Thus, we analyzed amino acid sequences by examining a conserved open reading frame across nearly all genomes, lactate dehydrogenase (**Figure 3.3**). Lactate dehydrogenase sequence clusters by genera, but it is challenging to interpret similarities and differences across genera (**Figure 3.3**).

Additionally, it can be challenging to infer function from sequence alone. To

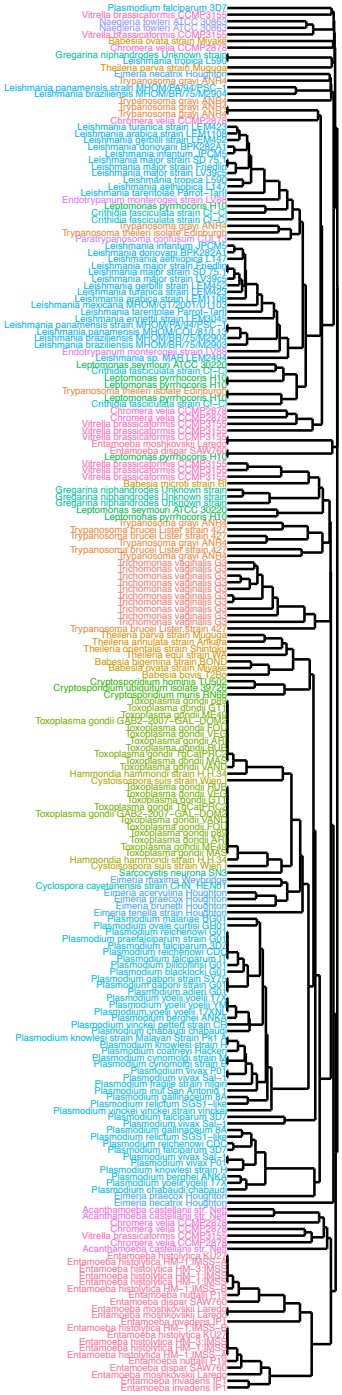


Figure 3.3: Sequence similarity. We compare genome (colored by genus) by clustering the amino acid sequence of a conserved open reading frame, the lactate dehydrogenase gene. Amino acid sequences were obtained by searching 'lactate dehydrogenase' on EuPathDB (omitting Fungi and Oomycetes), including both putative genes and validated enzymes.

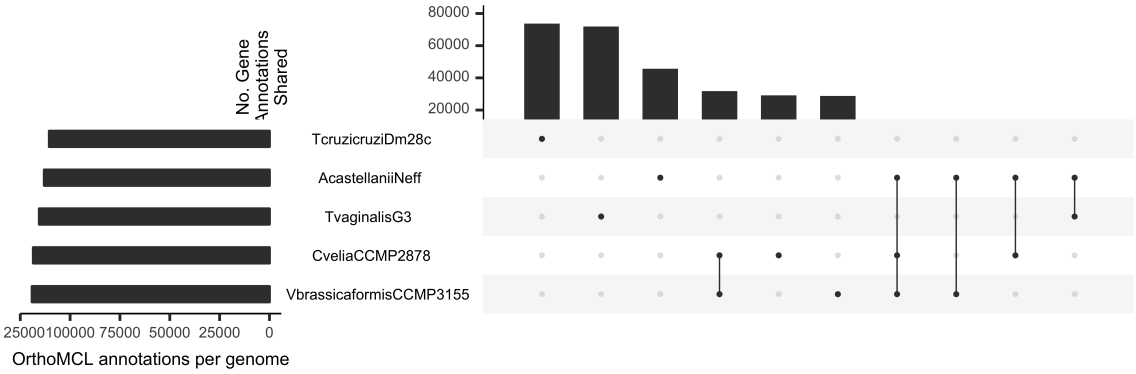


Figure 3.4: Genome reannotation by species with the largest gene sets shown. Genomes were reannotated to orthoMCL genes using Diamond. Plot can be read as a Venn Diagram. Bar plots show the intersection of all gene annotations in each species indicated with a dot. For example, the first column contains only gene annotations in *T. cruzi* cruzi Dm28c, whereas the fourth column contains annotations shared by only *C. velia* CCMP2878 and *V. brassicaformis* CCMP3155.

explore functional genomic content, we reannotated all genomes using Diamond against OrthoMCL genes and compared annotated genes in each genome. Each genome has unique gene annotations but many are shared (examples in **Figure 3.4**). For example, *Trichomonas vaginalis* G3 is the only complete genome on the *Trichomonas* database (TrichDB); this genome has the second most unique annotations by genome (**Figure 3.4**) and TrichDB has the third most unique annotations by genome despite containing only one genome (**Figure 3.5**). Unsurprisingly, some of the larger genomes, including *Chromera velia* CCMP2878 (CryptoDB, 193.4 megabases), *Acanthamoeba castellani* Neff (AmoebaDB, 42 megabases), and *T. vaginalis* G3 (176.3 megabases) have the most unique annotations (**Figure 3.4**). Annotation similarities can generate novel hypotheses about functional similarities. For example, the largest overlapping annotation group (CryptoDB and AmoebaDB) contains two types gut pathogens, the causes of cryptosporidiosis and amoebiasis, and these shared annotations may be consistent with mechanisms of gut pathogenesis (**Figure 3.5**). However, there are unique gene annotations associated with EuPathDB database, and it is unclear whether these differences arise from divergent metabolic functionality or incomplete genome annotation of these enzymes.

To address this challenge, we generated a metabolic reconstruction for each species (**Figure 3.6**). Genome-scale metabolic models are built from genomic data and by inferring function to complete or connect metabolic pathways; these reconstructions can be supplemented with functional genetic and biochemical studies (see **Methods**). Our *de novo* reconstructions contain only genetically supported information, and reconstruction size correlates with genome size (**Figure 3.7A**). Unsurprisingly, the

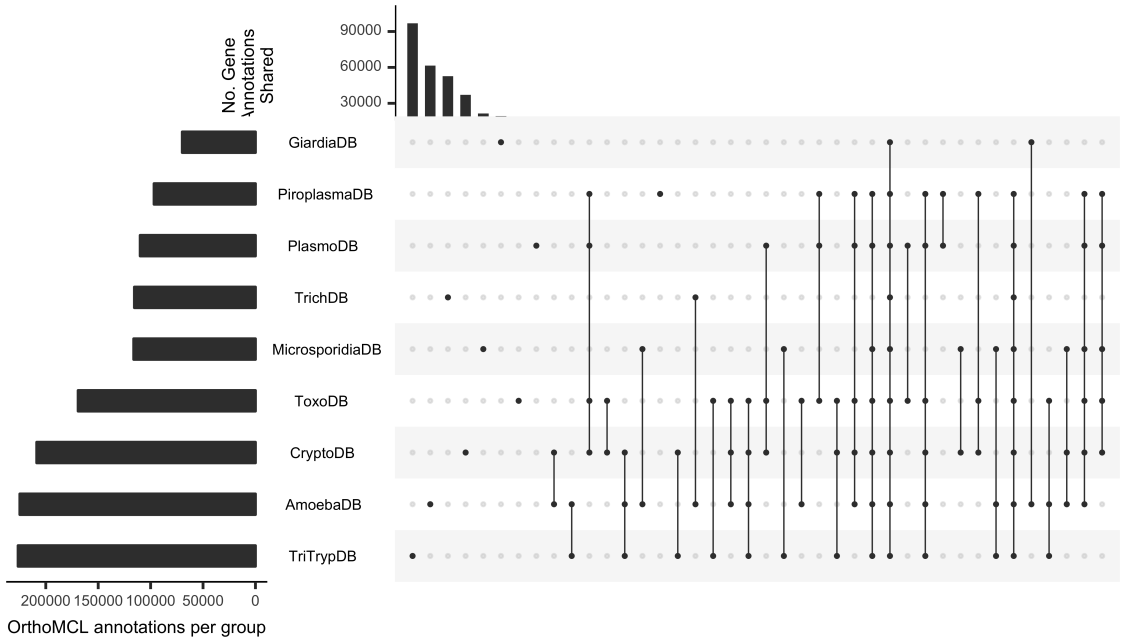


Figure 3.5: Genome reannotation by database. See previous figure for interpretation guidance.

large genome of *Chromera velia* CCMP2878 (CryptoDB, 31,799 ORFs, 2,943 reactions) has the most unique reactions (92, **Figure 3.6**). However, even small reconstructions contain unique reactions (**Figure 3.7B**). In fact, all reconstructions contain at least one unique reaction (**Figure 3.7B**), and small reconstructions do not have fewer unique reactions (**Figure 3.7C**). A core set of reactions are contained in all 162 reconstructions (right side of **Figure 3.7D**), and a large set of reactions are shared by only a few models (left side of **Figure 3.7D**). Reactions shared by all models include functions such as glycolysis.

We compare network structure and the predictions generated by each model, as we compared genomic content and annotations (**Figures 3.3**, **3.4**, and **3.5**). Network structures were minimally overlapping with 40 reactions shared by all reconstructions and 999 reactions in at least 50% of models. By comparing metabolic reactions in each reconstruction, we compare metabolic capacity of each species; two pairs, first *P. falciparum* 3D7 and IT and second *P yoelii* yoelii YM and 17X, were most similar and *C. velia* CCMP2878 and *T. cruzi* CL-Brener were most different, in contrast with the genetic similarity (**Figure 3.3**). As expected, models generated from genomes in the same genus contain similar sets of reactions (**Figure 3.8**).

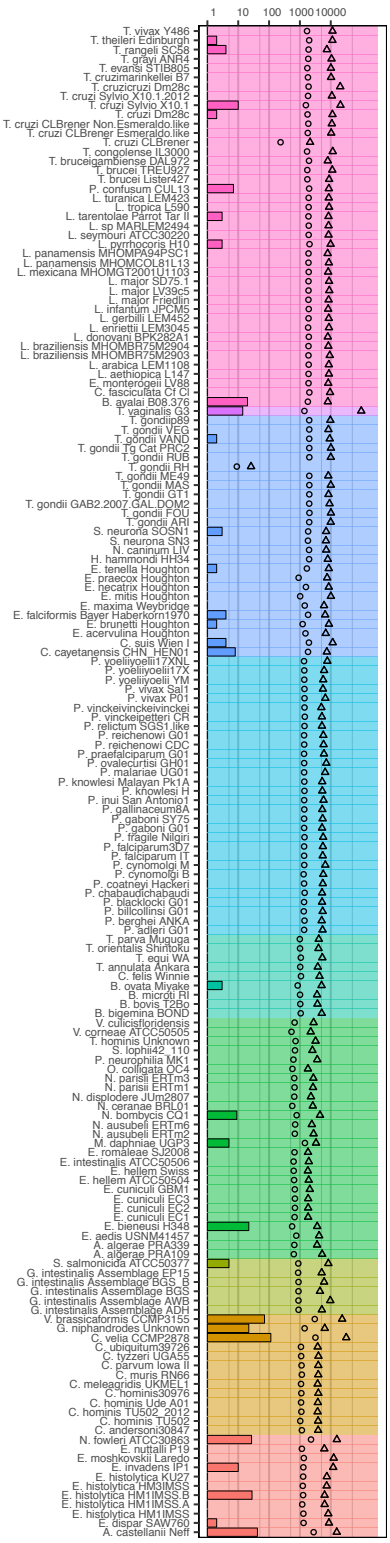


Figure 3.6: Draft reconstructions for 162 genomes. Triangles represent genome size (no. of open reading frames). Circle indicate model size (no. of reactions in the reconstruction). Bars represent the number of unique reactions per reconstruction. Genomes are color coded by EuPathDB database.

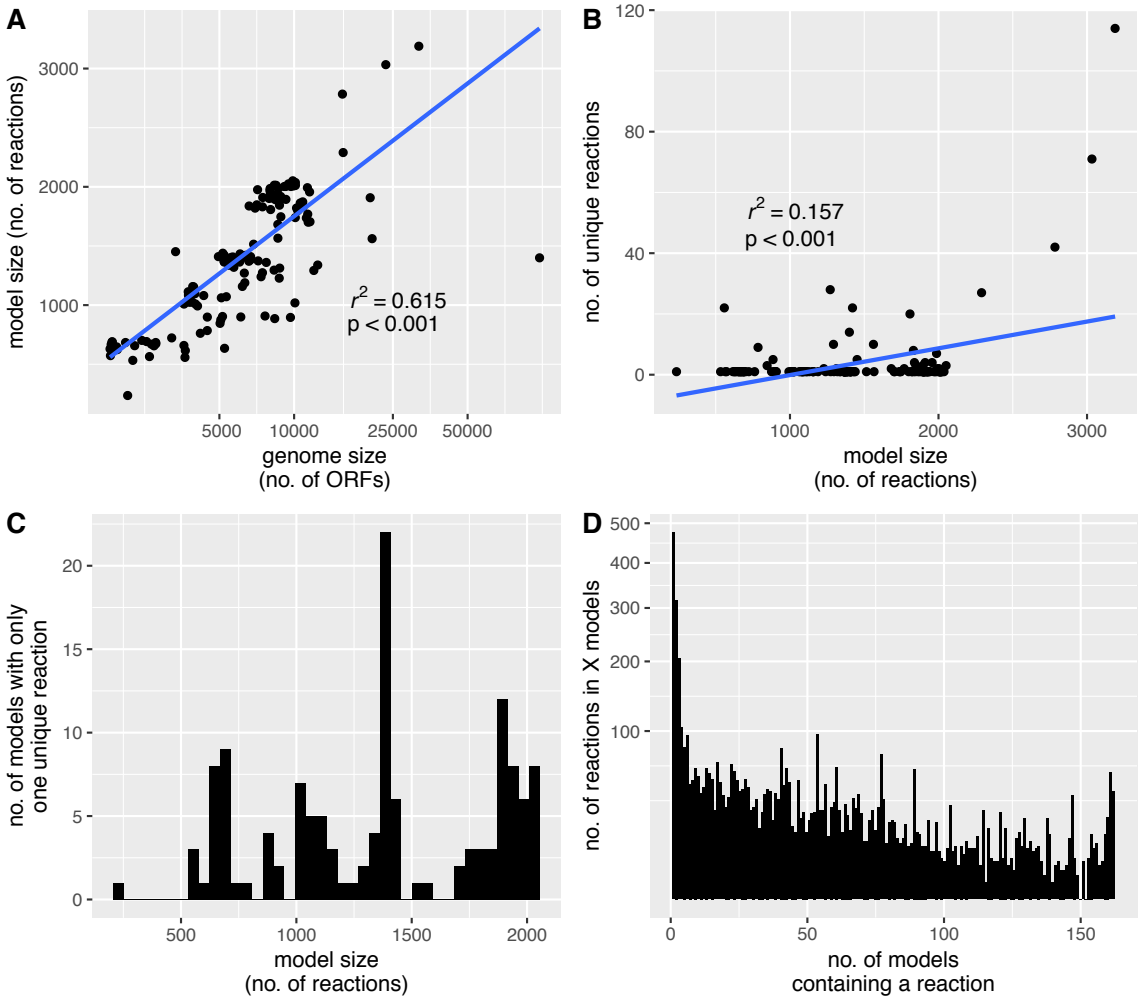


Figure 3.7: Draft reconstructions reveal unique functions in all organisms. A: Larger genomes generate larger reconstructions; size of genome measured by number of open reading frames, or ORFs. B: Larger reconstructions do not necessarily have more unique metabolic reactions, or reactions not found in any of the other 161 models. All reconstructions contain at least one unique reaction. C: Reconstructions of all sizes contain few unique features (defined as only one unique reaction), not just small reconstructions. D: Over 400 reactions are found in one reconstruction only (left side), whereas 40 reactions are found in all reconstructions (right side).

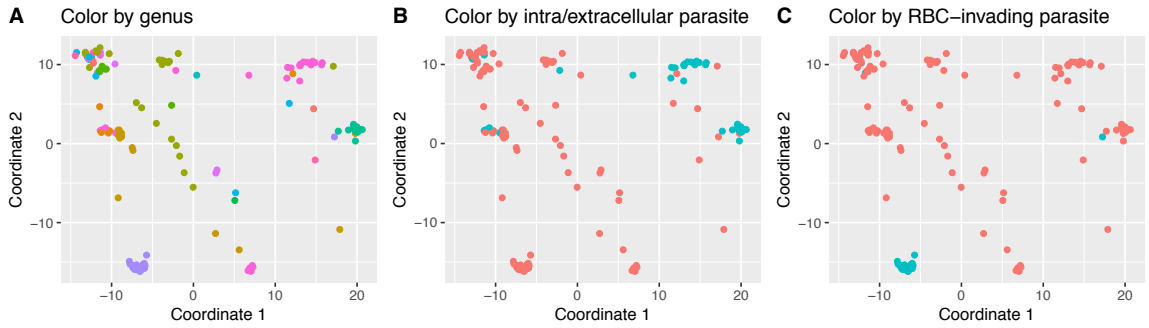


Figure 3.8: Network similarity. Networks were clustered based on reaction presence. Color indicates genus.

Table 3.6: Number and ratio of kinase-like reactions in each model.

species	number	percentage
BmicrotiRI	203	19.80
TgondiiGAB2-2007-GAL-DOM2	289	14.31
TgondiiTgCatPRC2	308	15.19
EbieneusiH348	181	31.53
PpraefalciparumG01	223	16.16
EromaleaeSJ2008	164	24.01
GintestinalisAssemblageADH	232	25.22
Pyoeliiyoelii17XNL	232	16.85
LmajorFriedlin	309	16.24
LpyrrhocorisH10	315	15.24
LmajorSD75	311	16.24
EmonterogeiiLV88	312	16.24
SsalmonicidaATCC50377	251	27.83
CtyzzeriUGA55	230	20.66
VcorneaeATCC50505	187	34.06
LdonovaniBPK282A1	310	15.83
GintestinalisAssemblageBGS	232	25.36
NausubeliERTm2	188	26.86
AalgeraePRA109	170	26.11
EpraecoxHoughton	206	22.29
PinuiSanAntonio1	248	17.83
CparvumIowaII	230	20.07
MdaphniaeUGP3	253	17.23
PbillcollinsiG01	225	16.63
CmeleagridisUKMEL1	232	21.07
NparisiiERTm3	162	24.11
TgondiiRUB	300	14.74
CfasciculataCfCl	296	14.50
LarabicaLEM1108	320	16.49
EhistolyticaHM1IMSS	263	20.05
TvaginalisG3	303	21.40
OcolligataOC4	159	26.99

Table 3.6: Number and ratio of kinase-like reactions in each model. *(continued)*

species	number	percentage
LtarentolaeParrotTarII	283	14.82
Pyoeliiyoelii17X	227	16.19
GniphandrodesUnknown	264	18.37
TorientalisShintoku	205	19.98
PreichenowiG01	219	15.87
TcruziCLBrenerEsmeraldo-like	256	14.10
TbruceiTREU927	288	16.43
TrangeliSC58	295	15.32
ThominisUnknown	218	29.50
LtropicaL590	304	15.71
EnuttalliP19	227	19.34
EcuniculiEC2	181	25.64
EtenellaHoughton	259	15.25
EinvadensIP1	259	19.79
EmaximaWeybridge	221	15.25
ChominisTU502_2012	224	19.84
TparvaMuguga	185	18.33
PcynomolgiB	240	17.96
LspMARLEM2494	321	16.49
LaethiopicaL147	306	15.93
Pgallinaceum8A	226	16.11
TgrayiANR4	278	14.80
AalgeraePRA339	186	27.56
EmoshkovskiiLaredo	266	19.63
CfelisWinnie	208	18.96
PvivaxP01	242	16.97
TevansiSTIB805	286	16.29
PgaboniG01	227	16.38
TgondiiGT1	302	14.88
LenriettiiLEM3045	315	16.93
NbombycisCQ1	184	22.97
Chominis30976	227	20.12
PconfusumCUL13	303	15.13
PyoeliiyoeliiYM	229	16.35
CmurisRN66	240	20.51
TgondiiVAND	295	14.61
PreichenowiCDC	220	15.79
TgondiiVEG	303	14.96
PadleriG01	228	16.46
TannulataAnkara	192	18.46
EhistolyticaHM1IMSS-B	245	19.04
EmitisHoughton	186	17.97
BbovisT2Bo	188	18.01
HhammondiHH34	290	14.49
EcuniculiEC3	180	25.53
TcruzimarinkelleiB7	284	15.46
TcongolenseIL3000	272	15.80
LmexicanaMHOMGT2001U1103	307	15.79

Table 3.6: Number and ratio of kinase-like reactions in each model. *(continued)*

species	number	percentage
EhistolyticaHM3IMSS	246	19.59
LbraziliensisMHOMBR75M2904	314	16.43
AcastellaniiNeff	367	13.11
EhistolyticaHM1IMSS-A	234	19.42
TcruziSylvioX10-1-2012	270	14.60
BayalaiB08-376	336	18.42
NfowleriATCC30863	332	14.40
PmalariaeUG01	229	16.51
EacervulinaHoughton	273	17.83
EhistolyticaKU27	255	19.77
PfalciparumIT	232	16.75
BovataMiyake	186	21.58
CsuisWienI	298	15.10
Tgondiip89	302	14.96
LseymouriATCC30220	295	15.22
SneuronaSN3	274	14.92
GintestinalisAssemblageAWB	212	23.25
TvivaxY486	268	15.61
PovalecurtisiGH01	243	17.48
PgaboniSY75	233	16.84
EnecatrixHoughton	253	15.99
LturanicaLEM423	314	16.24
PvinckeipetteriCR	238	16.36
EhellemATCC50504	173	26.09
TcruziDm28c	285	15.97
SneuronaSOSN1	263	14.11
NdisplodereJUm2807	170	25.26
CcayetanensisCHN_HEN01	295	15.97
Pchabaudichabaudi	241	16.74
EfalciformisBayerHaberkorn1970	265	14.29
GintestinalisAssemblageBGS_B	232	25.33
BbigeminaBOND	194	18.00
TbruceiLister427	293	16.62
EdisparSAW760	280	21.07
TgondiiFOU	305	14.86
LpanamensisMHOMPA94PSC1	318	16.58
PcynomolgiM	253	18.08
EcuniculiEC1	180	26.12
TgondiiME49	302	14.88
PneurophiliaMK1	191	30.22
EbrunettiHoughton	234	18.81
NceranaeBRL01	170	29.26
NausubeliERTm6	201	27.99
EcuniculiGBM1	181	25.86
LmajorLV39c5	314	16.40
LpanamensisMHOMCOL81L13	333	17.42
Candersoni30847	239	20.38
TcruzicruziDm28c	291	15.12

Table 3.6: Number and ratio of kinase-like reactions in each model. *(continued)*

species	number	percentage
PknowlesiMalayanPk1A	242	17.24
EhellemSwiss	178	27.55
Pfalciparum3D7	233	16.84
LgerbilliLEM452	305	15.89
TgondiiMAS	308	15.13
GintestinalisAssemblageEP15	227	25.45
PrelictumSGS1-like	234	16.44
ChominisTU502	224	21.50
TcruziSylvioX10-1	278	17.62
VbrassicaformisCCMP3155	366	12.00
Slophii42_110	185	26.13
PcoatneyiHackeri	242	17.52
ChominisUdeA01	226	20.43
PknowlesiH	244	17.22
TgondiiRH	5	20.00
TcruziCLBrenerNon-Esmeraldo-like	278	14.72
Pvinckeivinckeivinckei	233	16.34
EaedisUSNM41457	176	22.62
PfragileNilgiri	249	17.50
NparisiiERTm1	175	25.51
PvivaxSal1	242	17.04
TtheileriEdinburgh	320	15.93
TgondiiARI	303	14.97
Vculicisfloridensis	209	30.33
EintestinalisATCC50506	172	26.88
LinfantumJPCM5	304	15.59
PbergheiANKA	236	16.56
PblacklockiG01	233	16.88
TequiWA	198	18.22
CveliaCCMP2878	401	12.51
LbraziliensisMHOMBR75M2903	316	16.54
Cubiquitum39726	244	22.24
TbruceigambienseDAL972	312	15.72
NcaninumLIV	302	15.16
TcruziCLBrener	59	23.41

To further compare network structure, we explore the kinome and metabolite scavenging from the host. First, to approximate the kinome, we identified reactions involving phosphate transfers; networks contained between 5 and 401 reactions that transferred a phosphate from ATP to another metabolite. Between 12.004% and 34.06% of an organism's metabolic reactions included phosphate (**Table 3.6**). *Toxoplasma gondii* is often used as a model apicomplexan parasite because it is experimentally tractable (see **Chapter 2**) and some apicomplexan genera, such as *Cryptosporidium* and *Plasmodium*, had an enrichment of kinase reactions when compared to *Toxoplasma*.

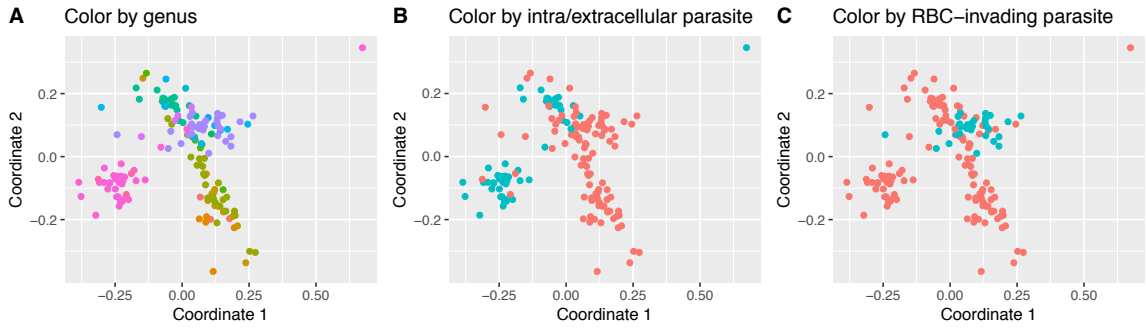


Figure 3.9: Genetically-supported transporters. Dimensionality reduction of capacity for imported metabolites. Dimensionality reduction is a technique to summarize variation in samples. Samples, or models, are represented by a point. Points that are close together have similar transporters. A: Reconstructions cluster by some genera. B: Extracellular parasites (blue) tend to have a negative coordinate 1 score. C: Red blood cell-infecting parasites (blue) tend to have a slight positive score for coordinates 1 and 2.

To explore each parasite's metabolic dependence on their host cells, we identified metabolites that could be imported via a genetically-supported transporter. We conducted a pairwise similarity between the set of metabolites that could be imported in each reconstructions. Following classical multidimensional scaling (or principal coordinates analysis, **Figure 3.9A**), we compared transporter topology between genera and parasite grouped by environmental niche. Reconstructions from organisms in the same genera had similar transport ability (**Figure 3.9A**); additionally, reconstructions separate roughly by some host cell types, like the organism's ability to divide extracellularly (**Figure 3.9B**) or in a host red blood cell (**Figure 3.9C**).

Next, we performed automated curation. All *Plasmodium* reconstructions were semi-curated using our automated curation pipeline and the curated reconstruction, iPfal18 of *Plasmodium falciparum* metabolism (**Figure 3.2**) and gapfilled to generate functional networks (*i.e.* networks that could product ATP and 'grow' as measured by the ability to produce biomass). Many modification were made to each *Plasmodium* reconstruction following semi-curation (**Figure 3.2C** and **Table 3.3**), greatly improving the genome-wide coverage of the reconstructions.

3.5 Discussion

Here, we presented 162 novel draft or semi-curated metabolic reconstructions for major human pathogens and closely related species and a pipeline for generating high-quality reconstructions from genomes, including automated curation by leveraging orthology (**Figure 3.2**). These reconstructions represent the first genome-scale metabolic recon-

structions for many of these organisms (previous reconstructions shown in **Table 3.7**), making Paradigm the broadest biochemical database for eukaryotic parasites to date. Our draft reconstruction approach contains key features to generate comprehensive networks for eukaryotic cells, making it unique among existing automated network reconstruction pipelines. Our semi-curation approach leverages the curation conducted in manually curated reconstructions for closely related organisms and genetic orthology, generating reconstructions that are more comprehensive than draft reconstructions. Both draft and semi-curated reconstructions can be used for comparative analyses, further curated by the modeling community, and applied to interrogate clinically and biologically relevant phenotypes.

Table 3.7: All previous metabolic reconstructions for eukaryotic parasites to our knowledge.

species	strain	previous reconstructions
Plasmodium falciparum	3D7	Plata, et al, 2010, Carey, et al, 2017, Chiappino-Pepe, et al, 2017 and Abdel-Haleem, et al, 2018
Plasmodium berghei	unspecified	Abdel-Haleem, et al, 2018
Plasmodium vivax	unspecified	Abdel-Haleem, et al, 2018
Plasmodium knowlesi	unspecified	Abdel-Haleem, et al, 2018
Plasmodium cynomolgi	unspecified	Abdel-Haleem, et al, 2018
Toxoplasma gondii	unspecified	Song, et al, 2013 and Tymoshenko, et al, 2015
Cryptosporidium hominis	unspecified	Vanee, et al, 2010
Trypanosoma cruzi	CL Brenner	Roberts, et al, 2009 (not genome scale)
Leishmania major	unspecified	Chavali, et al, 2008

Note:
There are several iterations of Plata, et al., 2010 not shown as they included only minor modifications.

Our approach has several key features tailored to eukaryotic pathogens. For example, discussion of biomass formulation is sorely lacking in many novel reconstruction papers and the assumptions used in formulated a biomass reaction for prokaryotes may not apply to eukaryotes. These assumptions are important as the objective function (like a biomass reaction) influences gapfilling and essentiality analyses. For example, in the first genome-scale metabolic model of any *Cryptosporidium* species, *C. hominis* (Vanee et al. 2010), 30 of 117 reactions involved in lipid synthesis were unsupported by genetic evidence; to generate biomass precursors, these 30 reactions were necessary and thus added despite lack of genetic evidence. The selection of biomass precursor metabolites like lipid species impact these results; for example, the

30 gapfilled reactions might not be added if ATP production only was used as an objective function.

Thus, to address these biomass-induced biases, we used multiple objective functions (ATP or biomass synthesis) and performed each gapfilling query 10 times to add confidence to our gapfilled reactions. For our semi-curated *Plasmodium* reconstructions, we also generated biomass reactions at multiple different scales: a universal and a genus-level biomass. We gapfilled each model to each of these objective reactions and added confidence scores to gapfilled reactions, corresponding to the number of gapfill solutions in which the reaction was added. These confidence scores inform our interpretation of model predictions (*i.e.* predictions involving low-confidence gapfilled reactions are low-confidence predictions) and highlight reactions for future manual curation. While including all gapfilled reactions (as opposed to just one possible solution) is not standard within the field, previous work has highlighted the uncertainty in network structure that gapfilling introduces (Biggs and Papin 2017). Thus, we believe that this uncertainty should be presented for future users and our confidence scores are a novel way to summarize this uncertainty.

Similarly, compartmentalization can induce biases in a model's predictions, as demonstrated at the end of **Chapter 5.1.5**. Compartmentalization is particularly relevant for generating reconstructions for eukaryotic organisms and a weak step of automated reconstruction approaches. To our knowledge, no automated approach directly addresses compartmentalization and, thus, compartmentalization is added manually. Both our *de novo* reconstruction and orthology-driven approaches addresses this. Compartmentalization was incorporated into our *de novo* reconstruction pipeline and implemented for several genera (**Table 3.1**). Furthermore, we used a curated model to inform the compartmentalization of each *Plasmodium* model; genes associated with compartmentalized reactions were mapped via orthology, assuming orthologous genes has comparable localization across species.

However, our *de novo* approach regarding compartmentalization yields one principle weakness; because genetically supported reactions were added to all feasible compartments, this adds plausible, but hypothetical, network functionality. For example, if a gene-encoded enzyme maps to mitochondrial and cytoplasmic reactions in an organism that contains a mitochondria, both versions will be included, adding network redundancy that may not be biologically accurate. Alternatively, if an enzyme maps to a chloroplast reaction that is not included in the BiGG database in any other subcellular compartment, we moved the reaction to the cytosol. It is plausible that chloroplast reactions like this example are not catalyzed by the parasite. However, it is likely that parasite do have functionality not well summarized in this database, which contains no parasite reconstructions, but 6 mammalian, 5 other

eukaryotic, 52 *E. coli*, and 12 other bacterial reconstructions (King et al. 2016). These modifications are encoded in our analytic pipeline for future reference (see code, https://github.com/maureencarey/). Consequently, our reconstructions will require manual network curation especially regarding pruning of excess functionality. This is also a weakness of our orthology-driven curation approach, which adds function without function removal, and of many modeling construction and validation (*i.e.* metabolic tasks) approaches as it is difficult to validate lack of function.

Directly answering our motivating biological question, we compared metabolic networks to identify divergent or conserved metabolic pathways to better leverage model systems for drug development. Network structures were quite unique with only 0.19952% of all reactions in more than 50% of the reconstructions; network topology did however clustered by genus (**Figure 3.8**) and transport ability is associated with host environment (**Figure 3.9**).

Despite structural similarities, minor topological differences in networks confer key metabolic strengths or weaknesses. For example, we identified a significant variation in the number of phosphate transferring reactions in *Toxoplasma gondii* and *Cryptosporidium* or *Plasmodium* (**Table 3.6**), making *T. gondii* a poor model system for kinase and phosphatase-based inhibitor screens for these species as off-target effects may vary significantly between organisms. For both *Plasmodium falciparum* and *Cryptosporidium*, inhibitors for phosphatidylinositol kinases are promising antiparasitics (Hassett and Roepe 2018; Manjunatha et al. 2017). However, kinase inhibitors are well-known for their promiscuity (Klaeger et al. 2017), and an enrichment of kinase-like reactions in *Cryptosporidium* may result in enhanced efficacy of the inhibitor. Fortunately, there was an insignificant difference in the ratio of kinase reactions in *Plasmodium falciparum* and *Plasmodium vivax* strains (**Table 3.6**). Because there is no *in vitro* culture system for *P. vivax*, inhibitor screens can only be conducted in *P. falciparum*; we predict the off-target effects would be comparable in both virulent *Plasmodium* species.

3.6 Conclusions

Here, we identified several novel findings, not readily apparent by genomic analysis alone. First, all parasite genomes encode unique metabolic functions, regardless of genome size, and parasites within the same genera tend to have similar network topology overall. Host cell type is associated with genetically-encoded transport ability from the extracellular environment to the parasite cytoplasm. Lastly, networks vary in the number and ratio of phosphate-using reaction they contain and the effect of this must be explored in inhibitor screens.

Most importantly, Paradigm provides a framework for organizing and interpreting our biochemical knowledge about eukaryotic parasites. This framework implements and builds on field-accepted standards for genome-scale metabolic modeling and the latest genome annotations in the parasitology field. Paradigm can be used broadly by the community and re-implemented iteratively to incorporate new genome sequences, novel datasets, and genome annotation updates. We call these networks 'semi-curated' to differentiate between the commonly used and referenced, uncurated 'draft' and well-curated network states. However, each reconstruction will require additional manual curation to maximize the utility and predictive accuracy, demonstrated in **Chapter 4**.

These reconstructions can be used to generate targeted experimental hypotheses for exploring differences between species and improving genome annotation by exploring differences between *in vitro* observations and *in silico* predictions. By applying this approach, we aim to develop a framework for identifying the best *in vitro* system or non-primate infection model of disease for drug development, and hypothesize that the best test system may vary by metabolic pathway for any one human pathogen.

Chapter 4: Curating a high quality reconstruction of parasite metabolism

The following text, figures, and tables have been adapted from Carey, Maureen A., Jason A. Papin, and Jennifer L. Guler. "Novel *Plasmodium falciparum* metabolic network reconstruction identifies shifts associated with clinical antimalarial resistance." *BMC Genomics* 18.1 (2017): 543.

4 Curating a high quality reconstruction of parasite metabolism

Draft and semi-curated models, as presented in the previous chapter, are manually curated to leverage biochemical data to expand our understanding of the organism's metabolism. Manual curation is an iterative process requiring updates as the field develops, perspectives evolve, and data are collected. Additionally, a model can be refined to ask particular biological questions or to represent a particular developmental stage, differentiated state of a cell, or condition-specific phenotype. While manual curation is time and labor intensive, it considered to be the best practice in the field for generating a high-quality and predictive network.

4.1 Synopsis

Here, we curated the metabolic reactions, gene-protein-reaction relationships, and subcellular compartmentalization contained in a genome-scale metabolic network reconstruction of the asexual blood-stage *P. falciparum* parasite to expand our understanding of the parasite's metabolism and, ultimately, to better understand the parasite's phenotype after developing resistance to clinically relevant antimalarial drugs. To do so, we manually curated an existing model and developed a framework for evaluating this iteration (and future iterations) of the model. We identified 11 metabolic tasks to evaluate iPfal17 performance. Predictions generated with iPfal17 are consistent with experimental literature, while generating novel hypotheses about parasite biology.

4.2 Methods

Manual curation: Manual curation of an existing *P. falciparum* metabolic network reconstruction (Plata et al. 2010) was conducted by a literature review and reference to generic and *Plasmodium*-specific databases (KEGG, Expasy, and PlasmoDB, MPMP) (Aurrecoechea et al. 2009; Kanehisa et al. 2016; Gasteiger et al. 2003; Ginsburg 2006). Data obtained from these sources were used to evaluate the inclusion of reactions as well as their stoichiometry, reversibility, localization, and gene annotations. Genetically and biochemically supported reactions were kept and new reactions were added. Reactions were removed if (1) explicitly determined to be false or (2) were nonfunctional and not supported biochemically or genetically. Spontaneous reactions (reactions that

occur without enzymes) are noted to differentiate from orphan reactions (reactions with unknown enzyme catalysts).

Flux balance analysis is an approach to explore metabolic phenotypes *in silico* (Orth, Thiele, and Palsson 2010). Flux balance analysis simulates steady-state flux values for each of the network's reactions that maximize subsequent flux through an objective function given a set of constraints. We chose biomass production as the objective reaction, consistent with previous studies interrogating gene essentiality (Plata et al. 2010; Thiele and Palsson 2010; Tymoshenko et al. 2015; Oberhardt et al. 2010), and permitted flux through all transport reactions. Constraints on the system include conservation of mass, reversibility of reactions, and reaction localization. Flux *variability* analysis uses a related approach to find the range of fluxes permissible given system constraints (Gudmundsson and Thiele 2010).

Objective reaction: In order to assess gene essentiality, we used a biomass reaction as the modeling objective function. The biomass reaction includes all metabolites known to be essential for growth. Thus, flux through this reaction was maximized for all *in silico* experimental procedures. We used the biomass reaction from a previous study (Plata et al. 2010) with modifications (outlined and motivated in the results section, **Metabolomics curation of biomass reaction**). The biomass reaction outlined in Plata et al. (2010) was created using field standard approaches; first it was based on the biomass of the closest phylogenetically related organism for which there is a metabolic reconstruction (at the time, *Saccharomyces cerevisiae*, Duarte, Herrgård, and Palsson (2004)), then it was adjusted to reflect experimental *Plasmodium* data such as lipid composition (Hsiao et al. 1991), genome and transcriptome nucleotide distribution (Llinás et al. 2006), and estimated proteome amino acid distribution.

We integrated new experimental data into the biomass formulation, specifically metabolomics studies (Cobbold et al. 2016; Gulati et al. 2015; Olszewski et al. 2009; Sana et al. 2013); we assumed any metabolite found in all metabolomics experiments were essential for cellular growth, consistent with field standards (Thiele and Palsson 2010; Lachance et al. 2018). However, if a metabolite was detected in a metabolomics experiment, but there are no known catabolism or import pathways for the compound, it was not added to the biomass reaction. Metabolite ratios were predicted from metabolomics data using measured abundance ratios.

Metabolic tasks: We simulated *in vitro* experiments and *in vivo* data to evaluate the model; these are our metabolic 'tasks' that the reconstruction should pass. We simulate *in vitro* growth requirements by modifying media components or access to particular metabolites. Metabolite import or production was eliminated from the reconstruction, and subsequent biomass production was observed. Effects of enzyme inhibition, gene knockouts, and metabolite production were also used to evaluate the

model. Lethal modifications were defined as changes that resulted in no production of biomass; growth-reducing modifications were defined as producing less than 90% of unconstrained flux value (Chavali et al. 2008; Oberhardt et al. 2010).

Table 4.1: Metabolites in the *in silico* extracellular environment.

1D-myo-Inositol 1-phosphate	glucocerebroside	oxidized glutathione
1D-myo-Inositol 1,3,4-trisphosphate	Glycero-3-phospho-1-inositol	Phosphate
1D-myo-Inositol 1,3,4,5-tetrakisphosphate	Glycero-3-phosphocholine	phosphatidyl 1D-myoinositol
1D-myo-Inositol 1,3,4,5,6-pentakisphosphate	Glycerophosphoserine	Phosphatidylcholine
1D-myo-Inositol 1,4-bisphosphate	hemozoin	phosphatidylethanolamine
1D-myo-Inositol 1,4,5-trisphosphate	L-homocystein	phosphotidylinositol
1D-myo-Inositol 1,4,5,6-tetrakisphosphate	L-lactate	prolyl glycine
1D-myo-Inositol 3,4-bisphosphate	L-Leucine	Propionate C30
1D-myo-Inositol 4-phosphate	L-Lysine	protein
2-Acyl sn-glycero-3-phosphoethanolamine C120	L-Malate	proteins with Asn-X-Ser/Thr residues
2-Acyl sn-glycero-3-phosphoethanolamine C140	L-Methionine	Putrescine
2-Acyl sn-glycero-3-phosphoethanolamine C141	L-Phenylalanine	Pyruvate
2-Acyl sn-glycero-3-phosphoethanolamine C160	L-Proline	R-Pantothenate
2-Acyl sn-glycero-3-phosphoethanolamine C161	L-Serine	Riboflavin
2-Acyl sn-glycero-3-phosphoethanolamine C180	L-Threonine	selenide
2-Acyl sn-glycero-3-phosphoethanolamine C181	L-Tryptophan	Spermidine
2-Acyl sn-glycero-3-phosphoglycerol C120	L-Tyrosine	sphingomyelin
2-Acyl sn-glycero-3-phosphoglycerol C140	L-Valine	Succinate
2-Acyl sn-glycero-3-phosphoglycerol C141	Maltose	Sulfate
2-Acyl sn-glycero-3-phosphoglycerol C160	Methylglyoxal	Superoxide anion
2-Acyl sn-glycero-3-phosphoglycerol C161	Nicotinamide	tetradecanoate C140
2-Acyl sn-glycero-3-phosphoglycerol C180	Nicotinate	tetradecenoate C141
2-Acyl sn-glycero-3-phosphoglycerol C181	Nitrate	Thiamine
2-dodecanoyl sn-glycerol-3-phosphate	Nitric oxide	Uracil
ADP-ribose	Nitrite	Urea
Ammonium	O2	Uridine
D-Mannose	octadecanoate C180	Xanthine
D-Sorbitol	octadecenoate C181	Xanthosine
dIMP	octanoate C80	Xanthosine 5-phosphate
dolichol	Ornithine	

Essentiality: We predicted essentiality by performing single deletion studies with both genes and reactions and double gene deletion studies in our curated model and each expression-constrained sensitive and resistant models. All simulations were performed in an *in silico* red blood cell environment (**Table 4.1**). Gene deletions were simulated by removing the gene of interest from the model. This change results in the inhibition of flux through all reactions that require that gene to function. If the model could not produce biomass with these constraints, the gene was deemed

Table 4.2: Asexual blood-stage *Plasmodium falciparum* parasite model, iPfal17, summary statistics. iTH366 from Plata, et al., 2010, iPfa from Chiappino-Pepe, et al., 2017.

	iTH366	iPfa	iPfal17
Reactions	1001	1066	1192
Enzymatic reactions	658	670	721
Reactions with gene annotations	657	586	672
Reactions with annotated citations	0	0	231
Metabolites	915	1258	991
Genes	366	325	482
Biomass components	51	73	82
Metabolites in extracellular environment	108	236	152

essential. Growth reducing phenotypes were also observed and noted. For reaction deletion studies, we removed reactions sequentially. Subsequent growth effects were used to determine reaction essentially.

The COBRA Toolbox 2015, Tiger Toolbox (version 1.3.1), and MATLAB R2013b were used for model generation and flux simulations in this section.

4.3 Results

Manual metabolic network curation: To maximize the predictive ability of the metabolic network model, we curated an existing, well-validated reconstruction of asexual blood-stage *P. falciparum* (iTH366, (Plata et al. 2010)) to improve its scope, and species- and stage-specificity. Our curated reconstruction, iPfal17, includes all metabolic reactions encoded by characterized genes in the parasite's genome, summarizing metabolic behavior during the asexual blood-stage parasite. It is larger in scope from the previously published version due to the addition of 268 reactions (**Table 4.2**, please see Appendix for Additional file 3, Table S1 & **4.3**), with 9.6% more enzymatic reactions and 2.3% more reactions with gene annotations. We also added 124 genes to the network (**Table 4.2** & see Appendix for Additional file 3, Table S1). It is larger in scope and gene coverage than a recent *de novo* reconstruction (**Table 4.2**). iPfal17 has gene annotations for 80.0% of enzymatic reactions, and 20.5% of transport and exchange reactions (**Figure 4.1**). iPfal17 includes 25.4% of the 1178 EC annotations in the *P. falciparum* genome, adding 14 EC numbers (Aurrecoechea et al. 2009) (see Appendix for Additional file 3, Table S1). We evaluated enzyme complex or isozyme status and replaced 7 gene-protein-reaction relationships (see Appendix for Additional file 3, Table S1).

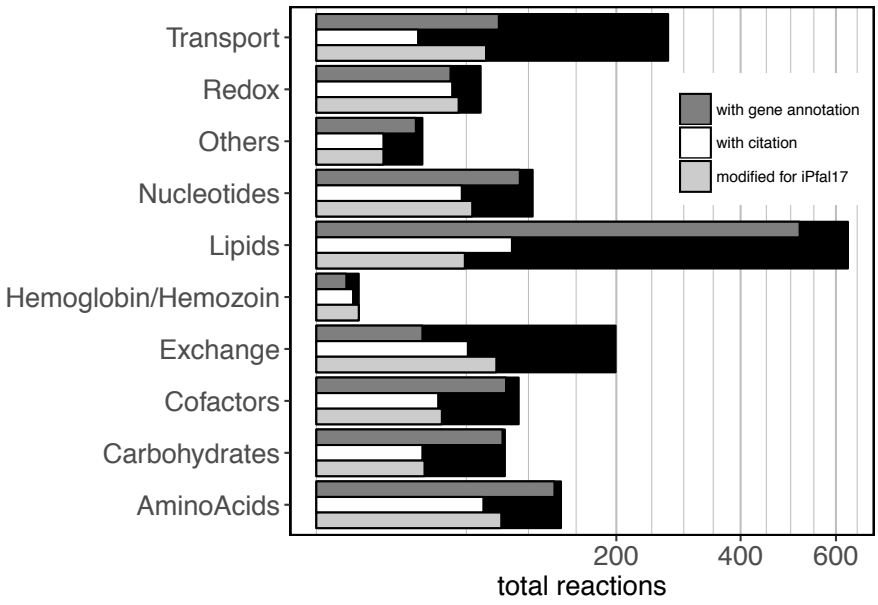


Figure 4.1: iPfal17 model curation is broad and comprehensive. Number of reactions in the *P. falciparum* reconstruction grouped by metabolic subsystems. Subsets of those reactions with gene annotations, literature citations, and modifications in the curation effort for this reconstruction are noted.

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17.

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
CYTDt	CYTDt	cytd[e] <=> cytd[c]	MAL8P1_32 PF13_0252 PF14_0662 PFA0160c	Transport base	NA	no pyrimidine salvage	NA
DCYTt	DCYTt	dcyt[e] <=> dcyt[c]	MAL8P1_32 PF13_0252 PF14_0662 PFA0160c	Transport base	NA	no pyrimidine salvage	NA
DTTPt	DTTPt	dttp[e] <=> dttp[c]	MAL8P1_32 PF13_0252 PF14_0662 PFA0160c	Transport base	NA	no pyrimidine salvage	NA
DURIt	DURIt	duri[e] <=> duri[c]	MAL8P1_32 PF13_0252 PF14_0662 PFA0160c	Transport base	NA	no pyrimidine salvage	NA
THYMDt1	THYMDt1	thymd[e] -> thymd[c]	MAL8P1_32 PF13_0252 PF14_0662 PFA0160c	Transport base	NA	no pyrimidine salvage	NA
URAt	URAt	ura[e] <=> ura[c]	MAL8P1_32 PF13_0252 PF14_0662 PFA0160c	Transport base	NA	no pyrimidine salvage	NA
URIt	URIt	uri[e] <=> uri[c]	MAL8P1_32 PF13_0252 PF14_0662 PFA0160c	Transport base	NA	no pyrimidine salvage	NA
NTD1pp	5'-nucleotidase	dump[c] + h2o[c] => duri[c] + pi[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
NTD10pp	5'-nucleotidase	h2o[c] + xmp[c] => pi[c] + xtsn[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA
NTD11pp	5'-nucleotidase	h2o[c] + imp[c] => ins[c] + pi[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA
NTD12pp	5'-nucleotidase	dimp[c] + h2o[c] => din[c] + pi[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA
NTD2pp	5'-nucleotidase	h2o[c] + ump[c] => pi[c] + uri[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA
NTD3pp	5'-nucleotidase	dcmp[c] + h2o[c] => dcyt[c] + pi[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA
NTD4pp	5'-nucleotidase	cmp[c] + h2o[c] => cytd[c] + pi[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA
NTD5pp	5'-nucleotidase	dtmp[c] + h2o[c] => pi[c] + thymd[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA
NTD6pp	5'-nucleotidase	damp[c] + h2o[c] => dad_2[c] + pi[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA
NTD7pp	5'-nucleotidase	amp[c] + h2o[c] => adn[c] + pi[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA
NTD8pp	5'-nucleotidase	dgmp[c] + h2o[c] => dgsn[c] + pi[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
NTD9pp	5'-nucleotidase	gmp[c] + h2o[c] => gsn[c] + pi[c]	PFL0305c	Nucleotides PurineMetabolis	3.1.3.5	duplicate	NA
PAPA120pp	phosphatidate phosphatase	h2o[c] + pa120[c] => 12dgr120[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	duplicate	NA
PAPA140pp	phosphatidate phosphatase	h2o[c] + pa140[c] => 12dgr140[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	duplicate	NA
PAPA141pp	phosphatidate phosphatase	h2o[c] + pa141[c] => 12dgr141[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	duplicate	NA
PAPA160pp	phosphatidate phosphatase	h2o[c] + pa160[c] => 12dgr160[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	duplicate	NA
PAPA161pp	phosphatidate phosphatase	h2o[c] + pa161[c] => 12dgr161[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	duplicate	NA
PAPA180pp	phosphatidate phosphatase	h2o[c] + pa180[c] => 12dgr180[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	duplicate	NA
PAPA181pp	phosphatidate phosphatase	h2o[c] + pa181[c] => 12dgr181[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	duplicate	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
PGPP120pp	phosphatidylglyc	h2o[c] + pgp120[c] => pg120[c] + pi[c]	NA	Lipids Phosphatidylethanola phatidylserineMetabolism	3.1.3.27	duplicate	NA
PGPP140pp	phosphatidylglyc	h2o[c] + pgp140[c] => pg140[c] + pi[c]	NA	Lipids Phosphatidylethanola phatidylserineMetabolism	3.1.3.27	duplicate	NA
PGPP141pp	phosphatidylglyc	h2o[c] + pgp141[c] => pg141[c] + pi[c]	NA	Lipids Phosphatidylethanola phatidylserineMetabolism	3.1.3.27	duplicate	NA
PGPP160pp	phosphatidylglyc	h2o[c] + pgp160[c] => pg160[c] + pi[c]	NA	Lipids Phosphatidylethanola phatidylserineMetabolism	3.1.3.27	duplicate	NA
PGPP161pp	phosphatidylglyc	h2o[c] + pgp161[c] => pg161[c] + pi[c]	NA	Lipids Phosphatidylethanola phatidylserineMetabolism	3.1.3.27	duplicate	NA
PGPP180pp	phosphatidylglyc	h2o[c] + pgp180[c] => pg180[c] + pi[c]	NA	Lipids Phosphatidylethanola phatidylserineMetabolism	3.1.3.27	duplicate	NA
PGPP181pp	phosphatidylglyc	h2o[c] + pgp181[c] => pg181[c] + pi[c]	NA	Lipids Phosphatidylethanola phatidylserineMetabolism	3.1.3.27	duplicate	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
GPDDA1pp	glycerophosphod phosphodi-esterase	g3pc[c] + h2o[c] => chol[c] + glyc3p[c] + h[c]	PF14_0060	Lipids PhosphatidylcholineMetabolis ; Lipids Phosphatidyle-tanolaminePhosphatidylser-ineMetabolism	3.1.4.46	duplicate	NA
GPDDA2pp	glycerophosphod phosphodi-esterase	g3pe[c] + h2o[c] => etha[c] + glyc3p[c] + h[c]	PF14_0060	Lipids PhosphatidylcholineMetabolis ; Lipids Phosphatidyle-tanolaminePhosphatidylser-ineMetabolism	3.1.4.46	duplicate	NA
GPDDA3pp	glycerophosphod phosphodi-esterase	g3ps[c] + h2o[c] => glyc3p[c] + h[c] + ser_L[c]	PF14_0060	Lipids PhosphatidylcholineMetabolis ; Lipids Phosphatidyle-tanolaminePhosphatidylser-ineMetabolism	3.1.4.46	duplicate	NA
GPDDA4pp	glycerophosphod phosphodi-esterase	g3pg[c] + h2o[c] => glyc3p[c] + glyc[c] + h[c]	PF14_0060	Lipids PhosphatidylcholineMetabolis ; Lipids Phosphatidyle-tanolaminePhosphatidylser-ineMetabolism	3.1.4.46	duplicate	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
GPDDA5pp	glycerophosphod phosphodi-esterase	g3pi[c] + h2o[c] => glyc3p[c] + h[c] + inost[c]	PF14_0060	Lipids Phosphatidyl-cholineMetabolis ; Lipids Phos-phatidyle-tanolaminePhos-phatidylser-ineMetabolism	3.1.4.46	duplicate	NA
CHTNASE	chitinase	chtn[c] + 2 h2o[c] -> 3 acgam[c]	PF07_0009 PFL2510w	Carbohydrates Aminosug-arsMetabolism	3.2.1.14	truncated gene; no biological evidence for reaction	NA
PSP_Lp	PSP Lp	h2o[c] + pser_L[c] -> pi[c] + ser_L[c]	PFI0880c	AminoAcids GlySer-Metabolism	3.1.3.2	Remove: required for invasion only	doi: 10.1128/EC.00244-10; 43 10.1016/j.jsb.2012.02.009
DM_HMFURN	DM HMFURN	hmfurn[c] ->	NA	Exchange	NA	hmfurn[c] never producted	NA
5AOPt2	5AOPt2	5aop[e] + h[e] -> 5aop[c] + h[c]	NA	Transport cofactpr	NA	No evidence for host 5aop; produced and consumed by parasite	NA
METS	methionine synthase	5mthf[c] + hcys_L[c] -> h[c] + met_L[c] + thf[c]	NA	AminoAcids Met-PolyamineMetab	2.1.1.13	no genetic evidence, and not needed for anything	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
AABTN	4-acetamidobutyra deacetylase	4aabutn[c] + h2o[c] -> 4abut[c] + ac[c]	NA	AminoAcids LysMetabolism	3.5.1.63	why is this here	NA
FRD_mt	succinate dehydrogenase	fadh2[m] + fum[m] -> fad[m] + succ[m]	PF10_0334 PFL0630w	Others Mitochondri-alTCACycle ; Carbohydrates Pyru-vateMetabolism ; Others Mitochondrial-ElectronFlow	1.3.99.1	remove (duplicate of SUCD2_u6m with SUCD3_u6m)	NA
NADH2_u6cm_	NADH dehydrogenase	h[m] + nadh[m] + q8[m] -> nad[m] + q8h2[m]	PFI0735c	Others Mitochondrial-ElectronFlow	1.6.99.3	Duplicate of NADH5	NA
PAPA120	phosphatidate phosphatase	h2o[c] + pa120[c] -> 12dgr120[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	PFC0995c is EC 2.3.1.20 (dag + acylCoA -> coA + tag)	NA
PAPA140	phosphatidate phosphatase	h2o[c] + pa140[c] -> 12dgr140[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	PFC0995c is EC 2.3.1.20 (dag + acylCoA -> coA + tag)	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
PAPA141	phosphatidate phosphatase	h2o[c] + pa141[c] -> 12dgr141[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	PFC0995c is EC 2.3.1.20 (dag + acylCoA -> coA + tag)	NA
PAPA160	phosphatidate phosphatase	h2o[c] + pa160[c] -> 12dgr160[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	PFC0995c is EC 2.3.1.20 (dag + acylCoA -> coA + tag)	NA
PAPA161	phosphatidate phosphatase	h2o[c] + pa161[c] -> 12dgr161[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	PFC0995c is EC 2.3.1.20 (dag + acylCoA -> coA + tag)	NA
PAPA180	phosphatidate phosphatase	h2o[c] + pa180[c] -> 12dgr180[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	PFC0995c is EC 2.3.1.20 (dag + acylCoA -> coA + tag)	NA
PAPA181	phosphatidate phosphatase	h2o[c] + pa181[c] -> 12dgr181[c] + pi[c]	PFC0995c	Lipids Utilization-Phospholipids	3.1.3.4	PFC0995c is EC 2.3.1.20 (dag + acylCoA -> coA + tag)	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
NTD5	5'-nucleotidase	dtmp[c] + h2o[c] -> pi[c] + thymd[c]	NA	Nucleotides PurineMetabolis	3.1.3.5	(remove duplicate) IMP specific gene; Remove reaction (no genetic evidence and thymd is not used)	Plasmodb
NTD4	5'-nucleotidase	cmp[c] + h2o[c] -> cytd[c] + pi[c]	NA	Nucleotides PurineMetabolis	3.1.3.5	(remove duplicate) IMP specific gene; Remove reaction (no genetic evidence and cytd is not used)	Plasmodb
NTD3	5'-nucleotidase	dcmp[c] + h2o[c] -> dcyt[c] + pi[c]	NA	Nucleotides PurineMetabolis	3.1.3.5	(remove duplicate) IMP specific gene; Remove reaction (no genetic evidence and dcyt is not used)	Plasmodb
EX_thymd(e)	EX thymd e	thymd[e] <=>	NA	Exchange	NA	Metabolite not used	NA
G6Pt6_2pp	G6Pt6 2pp	g6p[e] + 2 pi[c] -> g6p[c] + 2 pi[e]	NA	Transport	NA	No genetic evidence	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
F6Pt6_2pp	F6Pt6 2pp	f6p[e] + 2 pi[c] -> f6p[c] + 2 pi[e]	NA	Transport	NA	No genetic evidence; f6p can be made from fru (fructose), which import is supported by literature	NA
NCAMUP	NCAMUP	ncam[e] -> ncam[c]	NA	Transport cofactor	NA	Duplicate of EX_nicotinamic	NA
L_LACt2r	L LACt2r	h[e] + lac_L[e] <=> h[c] + lac_L[c]	NA	Transport	NA	No genetic evidence; reaction bypasses need for glycolysis	NA
2_1_1_103	phosphoethanola N-methyltransferas	3 ametam[c] + ethamp[c] <=> 3 ahcys[c] + cholp[c]	NA	AminoAcids Met-PolyamineMetab ; Lipids Phosphatidyl-cholineMetabolis	2.1.1.103	Not genetically supported; alternate product production pathways added; interferes with hypoxanthine being the only purine required for growth	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
GTPHs	GTPHs	gtp[c] + h2o[c] + h[c] -> nh4[c] + xtp[c]	NA	Nucleotides PurineMetabolis	NA	no genetic evidence, and not needed for anything	NA
GTHOr_mt	glutathione-disulfide reductase	gthox[m] + h[m] + nadph[m] <=> 2 gthrd[m] + nadp[m]	PF14_0192	Redox RedoxMetabolism ; Redox MitochondrialAntioxidantSystem	1.8.1.7	Mito version replaced with apico and cyto	DOI: 10.1371/journal.ppat.1001242
GTHP_mt	glutathione peroxidase	2 gthrd[m] + h2o2[m] <=> gthox[m] + 2 h2o[m]	PFL0595c	Redox RedoxMetabolism ; Redox MitochondrialAntioxidantSystem	1.11.1.9	Mito version replaced with apico and cyto	doi: 10.1371/journal.ppat.1001242; doi: 10.1111/j.1365-2958.2006.05303.x
DM_4HBA	NA	NA	NA	NA	NA	NA	NA
DM_HMFURN	DM HMFURN	hmfurn[c] ->	NA	Exchange	NA	Switch from Demand to Exchange	NA
DM_4HBA	DM 4HBA	4hba[c] ->	NA	Exchange	NA	Switch from Demand to Exchange	NA
GTHRDtmt	GTHRDtmt	gthrd[m] <=> gthrd[c]	NA	Transport MT	NA	NA	NA
EX_lac_L(e)	EX lac L e	lac_L[e] <=>	NA	Exchange	NA	Should not be reversible	NA
AKGtmt	AKGtmt	akg[m] + mal_L[c] -> akg[c] + mal_L[m]	PF08_0031	Transport MT	NA	Replaced with greater specificiity	NA

Table 4.3: Reactions deleted from Plata *et al.*, model (iTH366) in generating iPfal17. *(continued)*

Reaction name	Reaction description	Formula	Genes	Subsystem	EC Number	Notes	References
SUCD3_u6m_n	NA	NA	NA	NA	NA	Lumped reactions together - FAD cofactor, reaction halves are joined	NA
TALA	transaldolase	g3p[c] + s7p[c] <=> e4p[c] + f6p[c]	NA	Carbohydrates PentosePhosphateCycle	2.2.1.2	No genetic evidence; this reaction allows glyolysis to run without glucose	DOI: 10.1186/1475-2875-4-17
OMMBLHX3_n	OMMBLHX3 mt	2ommbl[m] + 2 atp[m] + 3 h2o[m] + nad[m] -> 2omhmbl[m] + 2 adp[m] + 3 h[m] + nadh[m] + 2 pi[m]	NA	Cofactors UbiquinoneMeta	NA	No genetic evidence; OMMBLHX uses same substrate for production of 2omhmbl	NA
Ht	Ht	h[c] <=> h[e]	NA	Transport	NA	NA	NA
ASPTA_mt	aspartate transaminase	akg[m] + asp_L[m] <=> glu_L[m] + oaa[m]	PFB0200c	AminoAcids AsnAsp-Metabolism ; Carbohydrates Pyru-vateMetabolism ; AminoAcids GluMetabolism	2.6.1.1	removed Mitochondrial version	Wrenger et al, J Mol Biol 2010
O2St	O2St	o2s[c] <=> o2s[e]	NA	Transport	NA	NA	NA

Following curation, the species and stage-specificity of the model was also improved. Gene annotations were evaluated against PlasmoDB resources (Aurrecoechea et al. 2009), resulting in 124 additional gene annotations. Importantly, we removed cellular import of pyrimidines from the host erythrocyte, as *P. falciparum* relies on *de novo* synthesis (**Table 4.3**) (Phillips and Rathod 2010; Painter et al. 2007). Blood-stage specificity was improved by removing genes only used in other life stages (specifically, the gene encoding chitinase (Langer and Vinetz 2001)). Additionally, 77 functionally unnecessary reactions were removed due to a lack of genetic and biochemical support (**Table 4.3**). Reactions necessary for growth were added manually. Reactions were individually curated, changing metabolite utilization and stoichiometry. See supplemental table for all manual modifications or additions, not presented here as it is too large to display, please see Appendix for Additional file 3, Table S1.

The iPfal17 reconstruction contains five compartments: extracellular space and four intracellular compartments (cytoplasmic, mitochondrial, apicoplast, and food vacuole; **Figure 2.3**). Few studies since the Plata et al. (2010) reconstruction (iTH366) investigated protein localization and therefore, few changes were made to compartmental assignments; the food vacuole compartment, containing two reactions, was added in this version of the reconstruction. As in iTH366, reactions with unknown localization were placed within the cytoplasm (Thiele and Palsson 2010). Again, similar to iTH366, a mitochondrial inner matrix was not added, as there is no evidence that the blood-stage parasite requires a proton gradient for energy production (Painter et al. 2007; Sturm et al. 2015; Ginsburg 2002). Nonpolar metabolites generated in one compartment and utilized in another were transported as needed for network functionality by assuming passive diffusion (Thiele and Palsson 2010).

We also included annotations that will accelerate future curation efforts. First, we did not remove blocked reactions (those that do not carry flux due to their lack of connectivity to other components of the network) because further research may add connectivity to these network components. iPfal17 contains 303 blocked reactions and 78 dead-end metabolites (specifically, 32 metabolites are not consumed and 46 are not produced). For example, 4-pyridoxate (a byproduct of vitamin B6 biosynthesis) is included; production is supported by bioinformatic analyses of the parasite genome, but the metabolite function or excretion pathway is not known. Second, citations are included within iPfal17 to identify the date of discovery and degree of literature support for each reaction (see Appendix for Additional file 3, Table S1 and **Table 4.3**). Literature support was only added to modified reactions, resulting in 231 citations (**Table 4.2** & see Appendix for Additional file 3, Table S1).

Metabolomics curation of biomass reaction: For the newly curated iPfal17 model, we modified the *Plasmodium* biomass reaction to better represent *in*

vitro data (**Table 4.4**). We added tRNA-ligated amino acids to the amino acid requirements to force protein production, rather than only demanding free amino acids. Additionally, lipid classes were added based on recently published metabolomics findings; phosphatidylinositol, phosphatidylglycerol, sphingomyelin, diacylglycerides, and triglycerides were added due to their observed increase in abundance between uninfected and infected erythrocytes (Gulati et al. 2015). Phosphatidylcholine ethers, acyl phosphatidylgycerol, lysophosphatidylinositol, bis(monoacyl-glyceryl)phosphate, and monosialodihexosylganglioside were excluded from the biomass reaction, as there is no known *Plasmodium* catabolism or import pathways for these lipids (Gulati et al. 2015). Analysis of metabolomics data enabled further curation of the biomass reaction with the addition of malate, alpha-ketoglutarate, and glutathione (both reduced and oxidized) (Cobbold et al. 2016; Olszewski et al. 2009; Sana et al. 2013). Importantly, we included the requirement for cellular export of lactate and hemozoin. Lactate is measured in extracellular *in vitro* metabolomics and *in vivo* via blood acidosis; it is the terminal product of glycolysis, the sole energy production pathway used by the blood-stage parasite (Biddau and Müller 2016; Ke et al. 2015; MacRae et al. 2013; Yeh et al. 2004). By requiring lactate export, we force the model to utilize glycolytic energy metabolism. Similarly, hemoglobin degradation is essential for the blood-stage parasite to produce free amino acids. Parasites can also import and synthesize some amino acids, but the breakdown of hemoglobin (and subsequent production of its byproduct, hemozoin) is necessary for growth (Dalal and Klemba 2015; Liu et al. 2006; Krugliak, Zhang, and Ginsburg 2002). Thus, by requiring hemozoin export, we force the *in silico* parasite to degrade hemoglobin as the primary pathway for amino acid production.

Table 4.4: Metabolic components of the biomass function. * = metabolites that have been added to the iPfal17 and are not present in the biomass function of iTH266.

Metabolites
Complex metabolites
protein,composed of tRNA-ligated amino acids (20)*
lipid, composed of sphingomyelin*, cholesterol, phosphatidyl choline, phosphatidyl ethanoloamine, triacylglycerides*, diacylglycerides*, phosphatidyl inositol*, phosphatidyl glycerol*, acyl phosphatidyl glycerol*
reduced & oxidized glutathione*
protoheme
Amino acids
alanine
asparagine
cysteine
glutamine
histidine
leucine
methionine

Table 4.4: Metabolic components of the biomass function. * = metabolites that have been added to the iPfal17 and are not present in the biomass function of iTH266. *(continued)*

Metabolites
phenylalanine
threonine
tyrosine
arginine
aspartate
glutamate
glycine
isoleucine
lysine
serine
proline
tryptophan
valine
Carbohydrates
malate*
alpha-ketoglutarate*
Nucleotides
ATP
CTP
GTP
UTP
dATP
dCTP
dGTP
dTTP
thiamine diphosphate
Excreted metabolites
lactate*
hemozoin*
Vitamins
pyridoxal
5-phosphate riboflavin
Others
spermidine
putrescine
s-adenosyl-methionine
2-octaprenyl 6-hydroyphenol
folates (mthf, f-thf, thf)
FAD
NAD
Fe2+ & Fe3+
SO4
coenzyme-A
water
NADP
NH4+

Table 4.5: Experimentally-derived metabolic tasks for evaluating iPfal17.

	Metabolic Task	In vitro	iPfal17	Hypothesis for in vitro/in silico discrepancies
1a	Growth in the presence of antimetabolite, riboflavin?	no	no	-
1b	Growth in the presence of antimetabolite, thiamine?	no	yes	Unknown antimetabolite mechanism; Off target effects of antimetabolite
1c	Growth in the presence of antimetabolite, nicotinamide?	no	yes	Unknown antimetabolite mechanism; Off target effects of antimetabolite
1d	Growth in the presence of antimetabolite, pyridoxine?	no	yes	Unknown antimetabolite mechanism; Off target effects of antimetabolite
2a	Grows without loops?	no	no	-
2b	ATP production if no exchange is allowed?	no	no	-
3a	Can produce purines?	yes	yes	-
3b	Growth with hypoxanthine as the only purine source?	yes	yes	-
3c	No growth if guanine, guanosine, inosine, adenine, or adenosine are only purine sources?	yes	60%	-
4	Growth with IPP supplementation and no apicoplast?	yes	no	Nuclear encoded proteins that function within the apicoplast may be expressed in the cytoplasm if the organelle is not present.
5a	Growth with glucose?	yes	yes	-
5b	Growth with alternative sugar source (no glucose, with ribose, mannose, fructose, galactose, or maltose)?	no	yes	Central carbon metabolism contains many reversible reactions. Carbon sources that support growth are debated.
6a	Can produce all amino acids except isoleucine?	yes	yes	-
6b	Is growth reduced without methionine, proline, tyrosine, cystine, glutamate, or glutamine supplementation?	yes	no	Model is not designed for growth reduction experiments.
6c	Growth without isoleucine supplementation?	no	no	-
7	Growth without calcium pantothenate?	no	no	-
8	Growth without p-aminobenzoic acid?	no	no	-
9	Cannot produce any metabolites if no exchange is allowed?	no	no	-

Table 4.5: Experimentally-derived metabolic tasks for evaluating iPfal17. *(continued)*

	Metabolic Task	In vitro	iPfal17	Hypothesis for in vitro/in silico discrepancies
10	Accuracy of experimental essentiality predictions	-	79.5%a	See table \ref{tab:tab_4} and associated supplemental
11	Accuracy of P. berghei essentiality predictions	-	61.4%a	See table \ref{tab:tab_s6}

iPfal17 validation and functional requirements: To validate the model against experimental results, essential metabolic tasks of blood-stage growth were identified and evaluated (**Table 4.5**). These tasks simulate experimental manipulations of the parasite or culturing environment, or clinical observations. For example, the parasite is able to grow *in vitro* with glucose as the sole carbon source and hypoxanthine as the purine source, and *in vivo* the parasite induces blood acidosis via lactate (Asahi et al. 1996; Geary et al. 1985; Miller et al. 2002). From these experimental or clinical observations, we develop a task to test if the *in silico* parasite behaves the same way. Additional tasks include the parasite's failure to grow in the presence of anti-metabolites for riboflavin, nicotinamide, thiamine, and pyridoxine (Geary, Divo, and Jensen 1985); an anti-metabolite is a compound that inhibits the use of the metabolite. We defined this set of *in silico* tasks to provide a framework for curation and validation efforts of future network reconstructions. Although iPfal17 fails to pass all metabolic task simulations, we believe this is the most comprehensive and accurate model to date due to the curation efforts and results from tests of the metabolic tasks. Failures generally exist in pathways that currently contain many reversible reactions (*i.e.* tasks 5a–b for glycolysis) or if the experimental evidence is not mechanistic (*i.e.* tasks 1a–d) or fully characterized (*i.e.* task 4; **Table 4.5**).

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*.

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
SERTRS	Serine-tRNA ligase	atp[c] + ser_L[c] + trnaser[c] => amp[c] + ppi[c] + sertrna[c]	6.1.1.11
ILETRS	Isoleucine-tRNA ligase	atp[c] + ile_L[c] + trnaile[c] => amp[c] + iletrna[c] + ppi[c]	6.1.1.5
LEUTRS	Leucine-tRNA ligase	atp[c] + leu_L[c] + trnaleu[c] => amp[c] + leutrna[c] + ppi[c]	6.1.1.4
VALTRS	Valine-tRNA ligase	atp[c] + trnaval[c] + val_L[c] => amp[c] + ppi[c] + valtrna[c]	6.1.1.9

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*. *(continued)*

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
FMETTRS	methionyl-tRNA formyltransferase	10fthf[c] + mettrna[c] => fmettrna[c] + h[c] + thf[c]	2.1.2.9
TYRTRS	Tyrosine-tRNA ligase	atp[c] + trnatyr[c] + tyr_L[c] => amp[c] + ppi[c] + tyrtrna[c]	6.1.1.1
ALATRS	Alanine-tRNA ligase	ala_L[c] + atp[c] + trnaala[c] => alatrna[c] + amp[c] + ppi[c]	6.1.1.7
CYSTRS	Cysteine-tRNA ligase	atp[c] + cys_L[c] + trnacys[c] => amp[c] + cystrna[c] + ppi[c]	6.1.1.16
HISTRS	Histidine-tRNA ligase	atp[c] + his_L[c] + trnahis[c] => amp[c] + histrna[c] + ppi[c]	6.1.1.21
THRTRS	Threonine-tRNA ligase	atp[c] + thr_L[c] + trnathr[c] => amp[c] + ppi[c] + thrtrna[c]	6.1.1.3
ENO	phosphopyruvate hydratase	2pg[c] <=> h2o[c] + pep[c]	4.2.1.11
GAPD	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)	g3p[c] + nad[c] + pi[c] <=> 13dpg[c] + h[c] + nadh[c]	1.2.1.12
LDH_L	L-lactate dehydrogenase	lac_L[c] + nad[c] <=> h[c] + nadh[c] + pyr[c]	1.1.1.27
TPI[ap]	Triose-phosphate isomerase	dhap[ap] <=> g3p[ap]	5.3.1.1
RPE	Ribulose-phosphate 3-epimerase	ru5p_D[c] <=> xu5p_D[c]	5.1.3.1
DPCOAK[ap]	dephospho-CoA kinase	atp[ap] + dpcoa[ap] => adp[ap] + coa[ap] + h[ap]	2.7.1.24
PNTK	pantothenate kinase	atp[c] + pnto_R[c] => 4ppan[c] + adp[c] + h[c]	2.7.1.33
PPCDC	Phosphopantothenoylcyst decarboxylase	4ppcys[c] + h[c] => co2[c] + pan4p[c]	4.1.1.36
PPNCL2	phosphopantothenate-cysteine ligase	4ppan[c] + ctp[c] + cys_L[c] => 4ppcys[c] + cmp[c] + h[c] + ppi[c]	6.3.2.5
PTPATi	pantetheine-phosphate adenylyltransferase	atp[c] + h[c] + pan4p[c] => dpcoa[c] + ppi[c]	2.7.7.3
DHFR	dihydrofolate reductase	dhf[c] + h[c] + nadph[c] <=> nadp[c] + thf[c]	1.5.1.3
DHFS	dihydrofolate synthase	atp[c] + dhpt[c] + glu_L[c] => adp[c] + dhf[c] + h[c] + pi[c]	6.3.2.12

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*. *(continued)*

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
DHPS2	dihydropteroate synthase	4abz[c] + 6hmhptpp[c] => dhpt[c] + ppi[c]	2.5.1.15
HPPK2	2-amino-4-hydroxy-6-hydroxymethyldihydropte diphosphokinase	6hmhpt[c] + atp[c] => 6hmhptpp[c] + amp[c] + h[c]	2.7.6.3
MTHFC	methenyltetrahydrofolate cyclohydrolase	h2o[c] + methf[c] <=> 10fthf[c] + h[c]	3.5.4.9
MTHFD	methylenetetrahydrofolat dehydrogenase (NADP+)	mlthf[c] + nadp[c] <=> methf[c] + nadph[c]	1.5.1.5
NADS2	NAD+ synthase (glutamine-hydrolysing)	atp[c] + dnad[c] + gln_L[c] + h2o[c] => amp[c] + glu_L[c] + h[c] + nad[c] + ppi[c]	6.3.5.1
NAMNPP	nicotinate phosphoribo-syltransferase	atp[c] + h2o[c] + nac[c] + prpp[c] => adp[c] + nicrnt[c] + pi[c] + ppi[c]	2.4.2.11
NNATr	nicotinate-nucleotide adenylyltransferase	atp[c] + h[c] + nicrnt[c] <=> dnad[c] + ppi[c]	2.7.7.18
RBFK	riboflavin kinase	atp[c] + ribflv[c] => adp[c] + fmn[c] + h[c]	2.7.1.26
CHORS	chorismate synthase	3psme[c] => chor[c] + pi[c]	4.2.3.5
DDPA	3-deoxy-7-phosphoheptulonate synthase	e4p[c] + h2o[c] + pep[c] => 2dda7p[c] + pi[c]	2.5.1.54
DHQS	3-dehydroquinate synthase	2dda7p[c] => 3dhq[c] + pi[c]	4.2.3.4
DHQTi	3-dehydroquinate dehydratase	3dhq[c] => 3dhsk[c] + h2o[c]	4.2.1.10
PSCVT	3-phosphoshikimate 1-carboxyvinyltransferase	pep[c] + skm5p[c] <=> 3psme[c] + pi[c]	2.5.1.19
SHK3Dr	shikimate dehydrogenase	3dhsk[c] + h[c] + nadph[c] <=> nadp[c] + skm[c]	1.1.1.25
CHRPL	chorismate lyase	chor[c] => 4hbz[c] + pyr[c]	4.1.3.40
R07456	R07456	g3p[c] + gln_L[c] + ru5p_D[c] <=> glu_L[c] + pydx5p[c]	NA
EX_cholesterol	EX cholesterol	chsterol[e] <=>	NA
EX_fe2(e)	EX fe2 e	fe2[e] <=>	NA
EX_ile_L(e)	EX ile L e	ile_L[e] <=>	NA
EX_phosphatidyl2	EX phosphatidyl2	pc[e] <=>	NA
EX_pnto_R(e)	EX pnto R e	pnto_R[e] <=>	NA
EX_ribflv1	EX ribflv1	ribflv[e] <=>	NA
EX_so4(e)	EX so4 e	so4[e] <=>	NA
FEROpp	ferroxidase	4 fe2[c] + 4 h[c] + o2[c] => 4 fe3[c] + 2 h2o[c]	1.16.3.1

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*. *(continued)*

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
lipid3	lipid3	dag[c] <=> 12dgr120[c] + 12dgr140[c] + 12dgr141[c] + 12dgr160[c] + 12dgr161[c] + 12dgr180[c] + 12dgr181[c]	NA
CDPMEK[ap]	4-(cytidine 5-diphospho)-2-C-methyl-D-erythritol kinase	4c2me[ap] + atp[ap] => 2p4c2me[ap] + adp[ap] + h[ap]	2.7.1.148
DMPPS[ap]	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	h2mb4p[ap] + h[ap] + nadph[ap] => dmpp[ap] + h2o[ap] + nadp[ap]	1.17.1.2
DXPRIi[ap]	1-deoxy-D-xylulose-5-phosphate reductoisomerase	dxyl5p[ap] + h[ap] + nadph[ap] => 2me4p[ap] + nadp[ap]	1.1.1.267
DXPS[ap]	1-deoxy-D-xylulose-5-phosphate synthase	g3p[ap] + h[ap] + pyr[ap] => co2[ap] + dxyl5p[ap]	2.2.1.7
IPDPS[ap]	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	h2mb4p[ap] + h[ap] + nadph[ap] => h2o[ap] + ipdp[ap] + nadp[ap]	1.17.1.2
MECDPDH2[ap]	MECDPDH2 ap	2mecdp[ap] + nadph[ap] => h2mb4p[ap] + h2o[ap] + nadp[ap]	1.17.7.1
MECDPS[ap]	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	2p4c2me[ap] => 2mecdp[ap] + cmp[ap]	4.6.1.12
MEPCT[ap]	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	2me4p[ap] + ctp[ap] + h[ap] => 4c2me[ap] + ppi[ap]	2.7.7.60
DMATT	dimethylallyltranstransfe	dmpp[c] + ipdp[c] => grdp[c] + ppi[c]	2.5.1.1
GRTT	geranyltranstransferase	grdp[c] + ipdp[c] => frdp[c] + ppi[c]	2.5.1.10
ADSL1r	adenylosuccinate lyase	dcamp[c] <=> amp[c] + fum[c]	4.3.2.2
GK1	guanylate kinase	atp[c] + gmp[c] <=> adp[c] + gdp[c]	2.7.4.8
HXPRT	hypoxanthine phospho-ribosyltransferase	hxan[c] + prpp[c] => imp[c] + ppi[c]	2.4.2.8
ADSS	Adenylosuccinate synthase	asp_L[c] + gtp[c] + imp[c] => dcamp[c] + gdp[c] + 2 h[c] + pi[c]	6.3.4.4

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*. *(continued)*

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
DHORTS	dihydroorotase	dhor_S[c] + h2o[c] <=> cbasp[c] + h[c]	3.5.2.3
DUTPDP	dUTP diphosphatase	dutp[c] + h2o[c] => dump[c] + h[c] + ppi[c]	3.6.1.23
OMPDC	Orotidine-5-phosphate decarboxylase	h[c] + orot5p[c] => co2[c] + ump[c]	4.1.1.23
ORPT	orotate phosphoribosyl-transferase	orot5p[c] + ppi[c] <=> orot[c] + prpp[c]	2.4.2.10
ASPCT	aspartate carbamoyltransferase	asp_L[c] + cbp[c] => cbasp[c] + h[c] + pi[c]	2.1.3.2
TMDS	thymidylate synthase	dump[c] + mlthf[c] => dhf[c] + dtmp[c]	2.1.1.45
TRDR	thioredoxin-disulfide reductase	h[c] + nadph[c] + trdox[c] => nadp[c] + trdrd[c]	1.8.1.9
FE2t	FE2t	fe2[e] => fe2[c]	NA
SO4ti	SO4ti	so4[e] => so4[c]	NA
DPCOAtap	DPCOAtap	dpcoa[ap] <=> dpcoa[c]	NA
NADPtap	NADPtap	nadp[ap] <=> nadp[c]	NA
PPItap	PPItap	ppi[ap] <=> ppi[c]	NA
CMPtap	CMPtap	cmp[ap] <=> cmp[c]	NA
CO2tap	CO2tap	co2[ap] <=> co2[c]	NA
CTPtap	CTPtap	ctp[ap] <=> ctp[c]	NA
DHAPtap	DHAPtap	dhap[c] + pi[ap] => dhap[ap] + pi[c]	NA
DMPPtap	DMPPtap	dmpp[ap] <=> dmpp[c]	NA
H2Otap	H2Otap	h2o[ap] <=> h2o[c]	NA
Htap	Htap	h[ap] <=> h[c]	NA
IPDPtap	IPDPtap	ipdp[ap] <=> ipdp[c]	NA
NADPHtap	NADPHtap	nadph[ap] <=> nadph[c]	NA
PEPPItap	PEPPItap	pep[c] + pi[ap] => pep[ap] + pi[c]	NA
PNTOt2	PNTOt2	h[e] + pnto_R[e] <=> h[c] + pnto_R[c]	NA
RIBFLVt2	RIBFLVt2	h[e] + ribflv[e] => h[c] + ribflv[c]	NA
4HBZtmt	4HBZtmt	4hbz[m] <=> 4hbz[c]	NA
DHORtmt	DHORtmt	dhor_S[m] <=> dhor_S[c]	NA
OCTDPtmt	OCTDPtmt	octdp[m] <=> octdp[c]	NA
OROTtmt	OROTtmt	orot[m] <=> orot[c]	NA
PHEMEtmt	PHEMEtmt	pheme[m] <=> pheme[c]	NA
PPItmt	PPItmt	ppi[m] <=> ppi[c]	NA
EX_hb	EX hb	=> hb[e]	NA
HBtr	HBtr	hb[e] => hb[c]	NA

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*. *(continued)*

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
CHSTEROLt	CHSTEROLt	atp[c] + chsterol[c] + h2o[c] <=> adp[c] + chsterol[e] + h[c] + pi[c]	NA
trna_ala	trna_ala	trnaala[c] <=>	NA
trna_asp	trna_asp	trnaasp[c] <=>	NA
trna_cys	trna_cys	trnacys[c] <=>	NA
trna_gly	trna_gly	trnagly[c] <=>	NA
trna_his	trna_his	trnahis[c] <=>	NA
trna_ile	trna_ile	trnaile[c] <=>	NA
trna_leu	trna_leu	trnaleu[c] <=>	NA
trna_lys	trna_lys	trnalys[c] <=>	NA
trna_pro	trna_pro	trnapro[c] <=>	NA
trna_ser	trna_ser	trnaser[c] <=>	NA
trna_thr	trna_thr	trnathr[c] <=>	NA
trna_tyr	trna_tyr	trnatyr[c] <=>	NA
trna_val	trna_val	trnaval[c] <=>	NA
trna_arg	trna_arg	trnaarg[c] <=>	NA
trna_met	trna_met	trnamet[c] <=>	NA
trna_asn	trna_asn	trnaasn[c] <=>	NA
trna_phe	trna_phe	trnaphe[c] <=>	NA
trna_trp	trna_trp	trnatrp[c] <=>	NA
OCTDPS	OCTDPS	frdp[c] + 5 ipdp[c] => octdp[c] + 5 ppi[c]	2.5.1.90
NADK	NADK	atp[c] + nad[c] => adp[c] + h[c] + nadp[c]	2.7.1.23
ARGTRS	ARGTRS	arg_L[c] + atp[c] + trnaarg[c] => amp[c] + ppi[c] + argtrna[c]	6.1.1.19
PGM	PGM	2pg[c] <=> 3pg[c]	5.4.2.1
HBZOPT_mt	HBZOPT_mt	4hbz[m] + octdp[m] => 3ophb[m] + 2 ppi[m]	2.5.1.39
HMGLB	HMGLB	hb[c] => 36 ala_L[c] + 6 arg_L[c] + 10 asn_L[c] + 15 asp_L[c] + 3 cys_L[c] + 4 gln_L[c] + 12 glu_L[c] + 20 gly[c] + 19 his_L[c] + 36 leu_L[c] + 22 lys_L[c] + 5 met_L[c] + 15 phe_L[c] + 14 pro_L[c] + 16 ser_L[c] + 16 thr_L[c] + 3 trp_L[c] + 6 tyr_L[c] + 31 val_L[c] + pheme[fv]	NA
HCO3E	HCO3E	co2[c] + h2o[c] => h[c] + hco3[c]	4.2.1.1

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*. *(continued)*

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
TRPTRS	TRPTRS	atp[c] + trnatrp[c] + trp_L[c] => amp[c] + ppi[c] + trptrna[c]	6.1.1.2
TKT2	TKT2	e4p[c] + xu5p_D[c] <=> f6p[c] + g3p[c]	2.2.1.1
DTMPK	DTMPK	atp[c] + dtmp[c] <=> adp[c] + dtdp[c]	2.7.4.9
UMPK	UMPK	atp[c] + ump[c] <=> adp[c] + udp[c]	2.7.4.14
NDPK3	NDPK3	atp[c] + cdp[c] => adp[c] + ctp[c]	2.7.4.6
NDPK4	NDPK4	atp[c] + dtdp[c] => adp[c] + dttp[c]	2.7.4.6
NDPK5	NDPK5	atp[c] + dgdp[c] => adp[c] + dgtp[c]	2.7.4.6
NDPK7	NDPK7	atp[c] + dcdp[c] => adp[c] + dctp[c]	2.7.4.6
NDPK8	NDPK8	atp[c] + dadp[c] => adp[c] + datp[c]	2.7.4.6
PPA	PPA	h2o[c] + ppi[c] => 2 h[c] + 2 pi[c]	3.6.1.1
OPHHX_mt	OPHHX_mt	2oph[m] + nadph[m] + 0.5 o2[m] => 2ohph[m] + h2o[m] + nadp[m]	1.14.13.8
CYTK1	CYTK1	atp[c] + cmp[c] <=> adp[c] + cdp[c]	2.7.4.14
SHKK	SHKK	atp[c] + skm[c] => adp[c] + h[c] + skm5p[c]	2.7.1.71
HMBZ	HMBZ	pheme[fv] => hemozoin[fv]	4.99.1.8
HMBZex	HMBZex	hemozoin[fv] => hemozoin[e]	NA
LAC	LAC	lac_L[c] => lac_L[e]	NA
METTRS	METTRS	atp[c] + met_L[c] + trnamet[c] => amp[c] + mettrna[c] + ppi[c]	6.1.1.10
ASNTRS	ASNTRS	asn_L[c] + atp[c] + trnaasn[c] => amp[c] + ppi[c] + asntrna[c]	6.1.1.22
FMNAT	FMNAT	atp[c] + fmn[c] + h[c] => fad[c] + ppi[c]	2.7.7.2
PHETRS	PHETRS	atp[c] + phe_L[c] + trnaphe[c] => amp[c] + ppi[c] + phetrna[c]	6.1.1.20
RNDR1	RNDR1	adp[c] + trdrd[c] => dadp[c] + h2o[c] + trdox[c]	1.17.4.1
RNDR2	RNDR2	gdp[c] + trdrd[c] => dgdp[c] + h2o[c] + trdox[c]	1.17.4.1

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*. *(continued)*

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
RNDR3	RNDR3	cdp[c] + trdrd[c] => dcdp[c] + h2o[c] + trdox[c]	1.17.4.1
PROTRS	PROTRS	atp[c] + pro_L[c] + trnapro[c] => amp[c] + ppi[c] + protrna[c]	6.1.1.15
ASPTRS	ASPTRS	asp_L[c] + atp[c] + trnaasp[c] => amp[c] + ppi[c] + asptrna[c]	6.1.1.12
GLYTRS	GLYTRS	atp[c] + gly[c] + trnagly[c] => amp[c] + ppi[c] + glytrna[c]	6.1.1.14
LYSTRS	LYSTRS	atp[c] + lys_L[c] + trnalys[c] => amp[c] + ppi[c] + lystrna[c]	6.1.1.6
OPHBDC_mt	OPHBDC_mt	3ophb[m] + h[m] => 2oph[m] + co2[m]	4.1.1.-
CYOOm_mt	CYOOm_mt	4 focytc[m] + 6 h[m] + o2[m] => 4 ficytc[m] + 2 h2o[m] + 6 h[c]	1.9.3.1
CYOR_u6m_mt	CYOR_u6m_mt	2 ficytc[m] + q8h2[m] => 2 focytc[m] + 2 h[m] + q8[m]	1.10.2.2
DHORD2_mt	DHORD2_mt	dhor_S[m] + q8[m] => orot[m] + q8h2[m]	1.3.5.2
PGK	PGK	13dpg[c] + adp[c] => 3pg[c] + atp[c]	2.7.2.3
ILEt2r	ILEt2r	ile_L[e] + leu_L[c] <=> ile_L[c] + leu_L[e]	NA
PRPPS	PRPPS	atp[c] + r5p[c] => amp[c] + h[c] + prpp[c]	2.7.6.1
PCt	PCt	pc[c] <=> pc[e]	NA
pc_prod	pc_prod	pc[c] => all_pc[c]	NA

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*. *(continued)*

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
Protein	Protein	0.1364 alatrna[c] + 0.0674 cystrna[c] + 0.1 fmettrna[c] + 0.3505 glutrna[c] + 0.1063 histrna[c] + 0.3566 iletrna[c] + 0.3405 leutrna[c] + 0.1019 mettrna[c] + 0.3055 sertrna[c] + 0.2045 thrtrna[c] + 0.1379 tyrtrna[c] + 0.203 valtrna[c] + 0.1349 argtrna[c] + 0.02095 trptrna[c] + 0.1353 glntrna[c] + 0.6245 asntrna[c] + 0.1823 phetrna[c] + 0.1085 protrna[c] + 0.2994 asptrna[c] + 0.1827 glytrna[c] + 0.5124 lystrna[c] => protein[c]	NA
Lipid_prod	Lipid_prod	0.519 chsterol[c] + 14 sphmyln[c] + 1.5 tag[c] + 35 all_pc[c] + 18 all_pe[c] + 4.25 all_pi[c] + 1.5 all_pg[c] + 0.2 all_apg[c] + 4 all_dgl[c] => lipid[c]	NA

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*. *(continued)*

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
biomass	biomass	0.000223 10fthf[c] + 0.000223 2ohph[m] + 0.01 akg[c] + 0.1364 ala_L[c] + 0.000223 amet[c] + 0.1349 arg_L[c] + 0.6245 asn_L[c] + 0.2994 asp_L[c] + 60.01 atp[c] + 0.000576 coa[c] + 0.0437 ctp[c] + 0.0674 cys_L[c] + 0.02721 datp[c] + 0.00655 dctp[c] + 0.00655 dgtp[c] + 0.02721 dttp[c] + 0.000223 fad[c] + 0.007106 fe2[c] + 0.007106 fe3[c] + 0.1353 gln_L[c] + 0.3505 glu_L[c] + 0.1827 gly[c] + 0.003 gthox[c] + 0.087 gthrd[c] + 0.06241 gtp[c] + 54.462 h2o[c] + 0.1063 his_L[c] + 0.3566 ile_L[c] + 0.3405 leu_L[c] + 0.5124 lys_L[c] + 0.01 mal_L[c] + 0.1019 met_L[c] + 0.000223 mlthf[c] + 0.001831 nad[c] + 0.000447 nadp[c] + 0.011843 nh4[c] + 0.1823 phe_L[c] + 0.000223 pheme[m] + 0.1085 pro_L[c] + 0.0479 ptrc[c] + 0.000223 pydx5p[c] + 0.000223 ribflv[c] + 0.3055 ser_L[c] + 0.003948 so4[c] + 0.00959 spmd[c] + 0.000223 thf[c] + 0.000223 thmpp[c] + 0.2045 thr_L[c] + 0.02095 trp_L[c] + 0.1379 tyr_L[c] + 0.2219 utp[c] + 0.203 val_L[c] + 0.1 hemozoin[e] + 0.01 lac_L[e] + 0.1 protein[c] + 0.1 lipid[c] => 59.81 adp[c] + 59.81 h[c] + 59.806 pi[c] + 0.7739 ppi[c] + biomass[c]	NA

Table 4.6: Predicted lethal reactions in wild-type blood-stage *Plasmodium falciparum*. *(continued)*

Essential Reaction	Reaction Name	Reaction formula	Reaction EC
biomass_s	biomass_s	biomass[c] =>	NA

Table 4.7: Predicted lethal genes in wild-type blood-stage *Plasmodium falciparum*.

MAL13P1_186	PF11_0410	PFA0225w	PFF1490w	PF14_0439
MAL13P1_221	PF13_0044	PFA0340w	PFI1170c	PFE1360c
MAL13P1_292	PF13_0140	PFA0555c	PFI1310w	MAL8P1_140
MAL13P1_67	PF13_0159	PFB0130w	PFI1365w	PF10_0150
MAL8P1_58	PF13_0179	PFB0210c	PFI1375w	PF14_0327
MAL8P1_81	PF13_0287	PFB0280w	PFI1420w	PFE0355c
PF07_0018	PF13_0354	PFB0295w	PFL0960w	PFE0370c
PF08_0068	PF14_0053	PFB0420w	PFL2465c	PF11_0381
PF08_0095	PF14_0060	PFC0831w	PF3D7_0714700	PF14_0574
PF10_0121	PF14_0097	PFD0830w	PF3D7_1370100	PF02_0059
PF10_0147	PF14_0198	PFE0150c	PF3D7_0706900	PF02_0060
PF10_0149	PF14_0248	PFE0410w	PF3D7_0702800	PF3D7_0206400
PF10_0154	PF14_0288	PFE0630c	PF3D7_0403000	PFB0279w
PF10_0155	PF14_0331	PFE1510c	PF3D7_0514400	PF14_0446
PF10_0221	PF14_0352	PFF0160c	PF3D7_1369900	mal_mito_2
PF10_0225	PF14_0373	PFF0370w	PFI0650c	mal_mito_1
PF10_0363	PF14_0415	PFF0450c	PF14_0015	PF10_0120
PF11_0059	PF14_0428	PFF0530w	PF14_0517	PF3D7_0211300
PF11_0169	PF14_0598	PFF1105c	PFI1570c	mal_mito_3
PF11_0270	PF14_0641	PFF1155w	PF11_0174	
PF11_0282	PF14_0697	PFF1410c	PFL2290w	
PF11_0295	PF14_0721	PFF1430c	PF13_0322	

Table 4.8: PlasmoGem (*Plasmodium berghei*) experimental results compared to iPfal17 predictions.

gene	PlasmoGem orthologous results	iPfal17 prediction
PFA0145c	essential	dispensible
PFA0340w	essential	essential
PFA0480w	essential	dispensible
PFB0130w	essential	essential
PFB0220w	essential	dispensible
PFB0295w	essential	essential
PFB0385w	essential	dispensible
PFB0435c	essential	dispensible
PFB0525w	essential	dispensible
PFC0470w	essential	dispensible
PFC0935c	essential	dispensible
PFC0995c	essential	dispensible
PFD0830w	essential	essential
PFE0150c	essential	essential

Table 4.8: PlasmoGem (*Plasmodium berghei*) experimental results compared to iPfal17 predictions. *(continued)*

gene	PlasmoGem orthologous results	iPfal17 prediction
PFE0370c	essential	essential
PFE0475w	essential	dispensible
PFE0485w	essential	dispensible
PFE0660c	essential	dispensible
PFE0765w	essential	dispensible
PFE1360c	essential	essential
PFE1510c	essential	essential
PFF0160c	essential	essential
PFF0370w	essential	essential
PFF0530w	essential	essential
PFF0895w	essential	dispensible
PFF1095w	essential	dispensible
PFF1115w	essential	dispensible
PFF1155w	essential	essential
PFF1190c	essential	dispensible
PFF1300w	essential	dispensible
PFF1350c	essential	dispensible
PFF1375c	essential	dispensible
PF07_0024	essential	dispensible
PF07_0073	essential	dispensible
MAL7P1.150	essential	dispensible
MAL8P1.150	essential	dispensible
MAL8P1.140	essential	essential
MAL8P1.13	essential	dispensible
PF3D7_0808200	essential	dispensible
PF08_0095	essential	essential
MAL8P1.17	essential	dispensible
PF08_0011	essential	dispensible
PFI0380c	essential	dispensible
PFI0680c	essential	dispensible
PFI0880c	essential	dispensible
PFI1020c	essential	dispensible
PFI1105w	essential	dispensible
PFI1140w	essential	dispensible
PFI1365w	essential	essential
PFI1370c	essential	dispensible
PFI1375w	essential	essential
PFI1375w	essential	essential
PF10_0051	essential	dispensible
PF10_0086	essential	dispensible
PF10_0120	essential	essential
PF10_0132	essential	dispensible
PF10_0150	essential	essential
PF10_0154	essential	essential
PF10_0155	essential	essential
PF10_0175	essential	dispensible
PF10_0221	essential	essential

Table 4.8: PlasmoGem (*Plasmodium berghei*) experimental results compared to iPfal17 predictions. *(continued)*

gene	PlasmoGem orthologous results	iPfal17 prediction
PF10_0322	essential	dispensible
PF10_0363	essential	essential
PF11_0051	essential	dispensible
PF11_0059	essential	essential
PF11_0173	essential	dispensible
PF11_0174	essential	essential
PF11_0208	essential	dispensible
PF11_0270	essential	essential
PF11_0295	essential	essential
PF11_0381	essential	essential
PF11_0407	essential	dispensible
PF11_0483	essential	dispensible
PF11_0483	essential	dispensible
PFL0110c	essential	dispensible
PFL0670c	essential	dispensible
PFL0675c	essential	dispensible
PFL0770w	essential	dispensible
PFL0900c	essential	dispensible
PFL1350w	essential	dispensible
PFL1515c	essential	dispensible
PFL1940w	essential	dispensible
PFL2000w	essential	dispensible
PFL2290w	essential	essential
PF13_0044	essential	essential
MAL13P1.56	essential	dispensible
MAL13P1.86	essential	dispensible
MAL13P1.118	essential	dispensible
MAL13P1.118	essential	dispensible
PF13_0133	essential	dispensible
PF13_0140	essential	essential
PF13_0141	essential	dispensible
PF13_0170	essential	dispensible
PF13_0182	essential	dispensible
PF13_0205	essential	dispensible
MAL13P1.210	essential	dispensible
PF13_0229	essential	dispensible
MAL13P1.221	essential	essential
PF13_0257	essential	dispensible
PF13_0322	essential	essential
MAL13P1.319	essential	dispensible
PF13_0344	essential	dispensible
PF13_0354	essential	essential
PF14_0053	essential	essential
PF14_0088	essential	dispensible
PF14_0097	essential	essential
PF14_0100	essential	dispensible
PF14_0166	essential	dispensible

Table 4.8: PlasmoGem (*Plasmodium berghei*) experimental results compared to iPfal17 predictions. *(continued)*

gene	PlasmoGem orthologous results	iPfal17 prediction
PF14_0198	essential	essential
PF14_0248	essential	essential
PF3D7_1430200	essential	dispensible
PF14_0288	essential	essential
PF14_0327	essential	essential
PF14_0331	essential	essential
PF14_0341	essential	dispensible
PF14_0352	essential	essential
PF14_0378	essential	dispensible
PF14_0381	essential	dispensible
PF14_0415	essential	essential
PF14_0425	essential	dispensible
PF14_0428	essential	essential
PF14_0446	essential	essential
PF14_0484	essential	dispensible
PF14_0517	essential	essential
PF14_0520	essential	dispensible
PF14_0589	essential	dispensible
PF14_0598	essential	essential
PF14_0641	essential	essential
PF14_0664	essential	dispensible
PF14_0721	essential	essential
PFA0160c	dispensible	dispensible
PFB0695c	dispensible	dispensible
PFB0890c	dispensible	dispensible
PFC0050c	dispensible	dispensible
PFC0430w	dispensible	dispensible
PFC0831w	dispensible	essential
PFC0910w	dispensible	dispensible
PFD0085c	dispensible	dispensible
PFD0465c	dispensible	dispensible
PFD0610w	dispensible	dispensible
PFE0405c	dispensible	dispensible
PFE0605c	dispensible	dispensible
PFE0775c	dispensible	dispensible
PFE0875c	dispensible	dispensible
PFE1050w	dispensible	dispensible
PFE1360c	dispensible	essential
PFF0360w	dispensible	dispensible
PFF0435w	dispensible	dispensible
PFF0945c	dispensible	dispensible
PFF1105c	dispensible	essential
PFF1130c	dispensible	dispensible
PFF1210w	dispensible	dispensible
PFF1210w	dispensible	dispensible
PFF1275c	dispensible	dispensible
PFF1335c	dispensible	dispensible

Table 4.8: PlasmoGem (*Plasmodium berghei*) experimental results compared to iPfal17 predictions. *(continued)*

gene	PlasmoGem orthologous results	iPfal17 prediction
PFF1360w	dispensible	dispensible
PFF1420w	dispensible	dispensible
PF07_0040	dispensible	dispensible
PF07_0073	dispensible	dispensible
PF07_0129	dispensible	dispensible
MAL8P1.13	dispensible	dispensible
MAL8P1.13	dispensible	dispensible
PF08_0077	dispensible	dispensible
PF08_0066	dispensible	dispensible
MAL8P1.81	dispensible	essential
PF08_0045	dispensible	dispensible
MAL8P1.13	dispensible	dispensible
PFI0735c	dispensible	dispensible
PFI0775w	dispensible	dispensible
PFI0815c	dispensible	dispensible
PFI0950w	dispensible	dispensible
PFI1110w	dispensible	dispensible
PFI1125c	dispensible	dispensible
PFI1170c	dispensible	essential
PFI1170c	dispensible	essential
PFI1195c	dispensible	dispensible
PFI1570c	dispensible	essential
PF10_0122	dispensible	dispensible
PF10_0137	dispensible	dispensible
PF10_0147	dispensible	essential
PF10_0150	dispensible	essential
PF10_0169	dispensible	dispensible
PF10_0175	dispensible	dispensible
PF10_0407	dispensible	dispensible
PF10_0409	dispensible	dispensible
PF10_0275	dispensible	dispensible
PF10_0289	dispensible	dispensible
PF3D7_1033800	dispensible	dispensible
PF10_0334	dispensible	dispensible
PF11_0036	dispensible	dispensible
PF11_0145	dispensible	dispensible
PF11_0157	dispensible	dispensible
PF11_0172	dispensible	dispensible
PF11_0174	dispensible	essential
PF11_0294	dispensible	dispensible
PF11_0453	dispensible	dispensible
PFL0035c	dispensible	dispensible
PFL0255c	dispensible	dispensible
PFL0285w	dispensible	dispensible
PFL0420w	dispensible	dispensible
PFL0480w	dispensible	dispensible
PFL0630w	dispensible	dispensible

Table 4.8: PlasmoGem (*Plasmodium berghei*) experimental results compared to iPfal17 predictions. *(continued)*

gene	PlasmoGem orthologous results	iPfal17 prediction
PFL0725w	dispensible	dispensible
PFL0770w	dispensible	dispensible
PFL0780w	dispensible	dispensible
PFL1155w	dispensible	dispensible
PFL1260w	dispensible	dispensible
PFL2210w	dispensible	dispensible
PFL2290w	dispensible	essential
PFL2570w	dispensible	dispensible
MAL13P1.40	dispensible	dispensible
PF13_0066	dispensible	dispensible
PF13_0089	dispensible	dispensible
PF13_0121	dispensible	dispensible
PF13_0128	dispensible	dispensible
PF13_0144	dispensible	dispensible
PF13_0234	dispensible	dispensible
MAL13P1.248	dispensible	dispensible
MAL13P1.284	dispensible	dispensible
PF13_0345	dispensible	dispensible
PF13_0353	dispensible	dispensible
MAL13P1.485	dispensible	dispensible
PF14_0015	dispensible	essential
PF14_0034	dispensible	dispensible
PF14_0164	dispensible	dispensible
PF14_0192	dispensible	dispensible
PF14_0192	dispensible	dispensible
PF14_0200	dispensible	dispensible
PF14_0282	dispensible	dispensible
PF14_0286	dispensible	dispensible
PF14_0354	dispensible	dispensible
PF14_0357	dispensible	dispensible
PF14_0368	dispensible	dispensible
PF14_0508	dispensible	dispensible
PF14_0541	dispensible	dispensible
PF14_0570	dispensible	dispensible
PF14_0573	dispensible	dispensible
PF3D7_1465700	dispensible	dispensible
PF14_0694	dispensible	dispensible
PF14_0751	dispensible	dispensible
PF14_0761	dispensible	dispensible

Table 4.9: Experimental knockouts compared to iPfal17 predictions. Tg = Toxoplasma gondii; Pf = Plasmodium falciparum; Pb = Plasmodium berghei; GR = growth reducing; L = lethal; NL = nonlethal; KO = knockout; Caret indicates contrary to published results; Star indicates known off target effects.

Enzyme	In vitro method	Species	Citation	Experimental Result	Plata	iPfal17
Pantothenate kinase	Inhibitors	Pf	Spry 2005	L	L	L
Dihydroptero sythetase	Inhibitors	Pf	Zhang and Meshnick 1991	L	L	L
Dihydrofolate reductase; thymidylate synthase	Inhibitor (1843U89)	Pf	Jiang 2000	L	L	L
Lactate deydroge-nase	Inhibitors (Gossypol, others)	Pf	Razakantoani 2000	L	L	L
1-Deoxy-D-xylulose-5-phosphate reductoiso-merase	Inhibitor (Fos-midomycin)	Pf	Cassera 2007	L	L	L
Ornithine decarboxylase Adenosylmeth decarboxy-lase	Inhibitors (3-Aminooxy-1-aminopropane DFMO)	Pf	Das Gupta 2005, Ramya 2006	NA	L^	Conditional KO: L
Spermidine synthase	Inhibitors (4MCHA, others)	Pf	Haider 2005	NA	L^	Conditional KO: L
FABI, enoyl-acyl carrier reductase	Inhibitor (Triclosan)*	Pf	Ramya 2007	L	-	NL
FABB/F 3-oxoacyl-acyl-carrier protein synthase I/II	Inhibitor (Cerulenin)	Pf	Ramya 2007	L	-	NL
Hypoxanthine phosphori-bosyl transferase	Antisense oligos	Pf	Dawson 1993	L	L	L

Table 4.9: Experimental knockouts compared to iPfal17 predictions. Tg = Toxoplasma gondii; Pf = Plasmodium falciparum; Pb = Plasmodium berghei; GR = growth reducing; L = lethal; NL = nonlethal; KO = knockout; Caret indicates contrary to published results; Star indicates known off target effects. *(continued)*

Enzyme	In vitro method	Species	Citation	Experimental Result	Plata	iPfal17
Adenylosuccin synthase	Inhibitor (hadacidin)	Pf	Webster 1984	L	L	L
Adenylosuccin lyase	Inhibitors (5-aminoimidazo 4-carboxamide ribonucleo-side)	Pf	Bulusu 2009	L	L	L
Dihydroorotat dehydroge-nase	RNAi; Inhibitors (several)	Pf	Deng 2009; McRobert & McConkey 2002	L	L	L
Adenosine deaminase	Inhibitor (methylth-ioco-formycin)	Pf	Ho 2009	L	L^*	Conditional KO: NL
Purine nucleoside phosphory-lase	Inhibitor (immucillin-H)	Pf	Kicska 2002	L	L^	Conditional KO: L
Ribonucleosid reductase	Oligodeoxynu phospho-rothioate	Pf	Chakrabarti 1993	L	L	L
Carbonic anhydrase	Inhibitors (several)	Pf	Krungkrai 2008	L	L	L
Carbamoyl-phosphate synthase	Ribozymes	Pf	Flores 1997	L	L	L
Deoxyuridine 5-triphosphate nucleotido-hydrolase	Inhibitors	Pf	Nguyen 2005	L	NL	L
Lactoylglutatl lyase	Inhibitor (S-p-bromobenzylg diethyl ester)	Pf	Thornalley 1994	L	NL	NL

Table 4.9: Experimental knockouts compared to iPfal17 predictions. Tg = Toxoplasma gondii; Pf = Plasmodium falciparum; Pb = Plasmodium berghei; GR = growth reducing; L = lethal; NL = nonlethal; KO = knockout; Caret indicates contrary to published results; Star indicates known off target effects. *(continued)*

Enzyme	In vitro method	Species	Citation	Experimental Result	Plata	iPfal17
Thioredoxin reductase (NADPH)	Gene KO	Pf	Krnajski 2002	L	L	L
3-Phosphoshikir 1-carboxyvinyl transferase	Inhibitor (glyphosate)	Pf	Roberts 1998	L	L	L
Chorismate synthase	RNAi	Pf	McRobert & McConkey 2002	L	L	L
Sphingomyelir	Inhibitor (Scypho-statin)	Pf	Hanada 2002	L	NL	NL (GR)
Plasmepsin II	Inhibitors	Pf	PubMed ID8816746	L	-	NL
cytosolic lysyl-tRNA synthetase	Inhibitor (cla-dosporin)	Pf	doi: 10.1016/j.cho	L	-	L
apicoplast-targeted isoleucyl-tRNA synthetase	Inhibitor (mupirocin)	Pf	doi: 10.1073/pnas.	L	-	L
cytosolic isoleucyl-tRNA synthetase	Inhibitor (Thi-aisoleucine)	Pf	doi: 10.1073/pnas.	L	-	L
gamma-Glutamylcyst synthase	Inhibitor (L-buthionine sulfox-imine); fail to Gene KO	Pf	doi: 10.1111/j.136 2958.2011.079	L	-	NL
glutathion s-transferase	Inhibitors (ellagic acid, others)	Pf	Sturm 2009	L	-	NL

Table 4.9: Experimental knockouts compared to iPfal17 predictions. Tg = Toxoplasma gondii; Pf = Plasmodium falciparum; Pb = Plasmodium berghei; GR = growth reducing; L = lethal; NL = nonlethal; KO = knockout; Caret indicates contrary to published results; Star indicates known off target effects. *(continued)*

Enzyme	In vitro method	Species	Citation	Experimental Result	Plata	iPfal17
glutathione reductase	Inhibitors; fail to KO	Pf	PMID: 23116403, Muller 2015	L	-	NL
Farnesyl diphosphate synthase	Inhibitors	Pf	Mukkamala 2008	L	L	L
Farnesyl diphosphate synthase	Inhibitors	Pf	Mukkamala 2008	L	L	L
Glycerol kinase	Gene KO	Pf	Schnick 2009	NL	NL	NL
Fructose-biphosphate aldolase	Antisense oligos	Pf	Wanidworanu 1999	NL	NL	NL
5-aminolevulinic acid synthase	Gene KO	Pf, Pb	doi: 10.1074/j and 10.1371/journal.ppat.1003	NL	-	NL
ferrochelatase	Gene KO	Pf	doi: 10.1074/j	NL	-	NL
aconitase	Gene KO; Inhibitor (Sodium fluoroacetate)	Pf	DOI: 10.1186/7007-11-67, DOI: 10.1016/j.celr	NL	-	NL
glutamate dehydrogenase a	Gene KO	Pf	DOI: 10.1186/1475-2875-10-193.	NL	-	NL
alpha-ketoglutarate dehydrogenase	Gene KO	Pf	doi: 10.1016/j.celr	NL	-	NL
succinyl-CoA synthetase	Gene KO	Pf	doi: 10.1016/j.celr	NL	-	NL
succinate dehydrogenase	Gene KO	Pf	doi: 10.1016/j.celr	NL	-	NL
citrate synthase	Gene KO	Pf	doi: 10.1016/j.celr	NL	-	NL

Table 4.9: Experimental knockouts compared to iPfal17 predictions. Tg = Toxoplasma gondii; Pf = Plasmodium falciparum; Pb = Plasmodium berghei; GR = growth reducing; L = lethal; NL = nonlethal; KO = knockout; Caret indicates contrary to published results; Star indicates known off target effects. *(continued)*

Enzyme	In vitro method	Species	Citation	Experimental Result	Plata	iPfal17
isocitrate dehydroge-nase	Gene KO	Pf	doi: 10.1016/j.celr	NL	-	NL
Protoporphyr oxidase	Inhibitor (Acifluor-fen)	Pf	Ramya 2007	L	L	NL
Acetyl-CoA carboxylase	Inhibitor; (Haloxy-fops, Fluazifops, Quizalo-fops)	Pf	Ramya 2007	L	-	NL
alpha-Amino-levulinic acid	Inhibitor	Pf	Ramya 2007	L	L	NL
dehydrogenas	(Succinyl acetone)	NA	NA	NA	NA	NA
Orotidine-monophospha	-	Pb	unpublished;	L	L	L
decarboxylase	NA	NA	Plata 2010	NA	NA	NA
glutathione synthetase	fail to Gene KO	Pb	Patzewitz, 2012 (doi: 10.1111/j.136 2958.2011.079	L	-	NL
Chitinase	Gene KO	Pb	Dessens 2001	NL	NL	NL
Arginase	Gene KO	Pb	Olszewski 2009	NL	NL	NL
Beta-Hydroxyacyl-ACP dehydratase	Gene KO	Pb	Vaughan 2009	NL	NL	NL
Beta-Hydroxyacyl-ACP dehydratase	Gene KO	Pb	Vaughan 2009	NL	NL	NL

Table 4.9: Experimental knockouts compared to iPfal17 predictions. Tg = Toxoplasma gondii; Pf = Plasmodium falciparum; Pb = Plasmodium berghei; GR = growth reducing; L = lethal; NL = nonlethal; KO = knockout; Caret indicates contrary to published results; Star indicates known off target effects. *(continued)*

Enzyme	In vitro method	Species	Citation	Experimental Result	Plata	iPfal17
Beta-Hydroxyacyl-ACP dehydratase	Gene KO	Pb	Vaughan 2009	NL	NL	NL
FABI, enoyl-acyl carrier reductase	Gene KO; siRNA	Pb	Vaughan 2009; Yu 2008	NL	NL	NL
FABB/F 3-oxoacyl-acyl-carrier protein synthase I/II	Gene KO	Pb	Vaughan 2009	NL	NL	NL
Thioredoxin peroxidase	Gene KO	Pb	Yano 2006 and 2008	NL	NL	NL
gamma-Glutamylcyst synthase	Gene KO	Pb	Vega-Rodriguez 2009	NL	NL	NL
Aquaglycerop	Gene KO	Pb	Promeneur 2007	NL	NL	NL
Phosphatidylc sterol acyltrans-ferase	Gene KO	Pb	Bhanot 2005	NL	NL	NL
glutathione reductase	Gene KO	Pb	doi: 10.1074/jbc.M	NL	-	NL

Table 4.10: Knockout predictions with experimental validations. Predictions for 18 enzymes of interest are included here. Tg = Toxoplasma gondii; Pf = Plasmodium falciparum; Pb = Plasmodium berghei; GR = growth reducing; L = lethal; NL = nonlethal; KO = knockout; Caret indicates contrary to published results; Star indicates known off target effects.

Enzyme	In vitro	In vitro method	Species	iTH366	iPfal17
Dihydrofolate reductase; thymidylate synthase	L	Inhibitor (1843 U89)	Pf	L	L

Table 4.10: Knockout predictions with experimental validations. Predictions for 18 enzymes of interest are included here. Tg = Toxoplasma gondii; Pf = Plasmodium falciparum; Pb = Plasmodium berghei; GR = growth reducing; L = lethal; NL = nonlethal; KO = knockout; Caret indicates contrary to published results; Star indicates known off target effects. *(continued)*

Enzyme	In vitro	In vitro method	Species	iTH366	iPfal17
FABI, enoyl-acyl carrier reductase	1) L, 2) NL	1) Inhibitor (Triclosan*), 2) Gene KO; siRNA	1) Pf, 2) Pb	-	NL&
FABB/F 3-oxoacyl-acyl-carrier protein synthase I/II	1) L, 2) NL	1) Inhibitor (Cerulenin), 2) Gene KO	1) Pf, 2) Pb	-	NL&
Dihydroorotate dehydrogenase	L	RNAi; Inhibitors (several)	Pf	L	L
Adenosine deaminase	L	Inhibitor (methylthioco-formycin)	Pf	L^$	cKO: NL%
Deoxyuridine 5-triphosphate nucleotido-hydrolase	L	Inhibitors (several)	Pf	NL	L
Lactoyl glutathione lyase	L	Inhibitor (S-p-bromobenzylglut diethyl ester)	Pf	NL	NL%
Sphingomyelinas	L	Inhibitor (Scyphostatin)	Pf	NL	NL (GR)%
Plasmepsin II	L	Inhibitors (several)	Pf	-	NL%
Cytosolic lysyl-tRNA synthetase	L	Inhibitor (cladosporin)	Pf	-	L
Gamma-Glutamylcystein synthase	1) L, 2) NL	1) Inhibitor (L-buthionine sulfoximine); fail to Gene KO, 2) Gene KO	1) Pf, 2) Pb	-	NL&
Glutathion s-transferase	L	Inhibitors (ellagic acid, others)	Pf	-	NL%

Table 4.10: Knockout predictions with experimental validations. Predictions for 18 enzymes of interest are included here. Tg = Toxoplasma gondii; Pf = Plasmodium falciparum; Pb = Plasmodium berghei; GR = growth reducing; L = lethal; NL = nonlethal; KO = knockout; Caret indicates contrary to published results; Star indicates known off target effects. *(continued)*

Enzyme	In vitro	In vitro method	Species	iTH366	iPfal17
Glutathione reductase	1) L, 2) NL	1) Inhibitors (several); fail to KO, 2) Gene KO	1) Pf, 2) Pb	-	NL&
5-Aminolevulinic acid synthase	1) NL, 2) L	1) Gene KO, 2) Inhibitor (Succinyl acetone)	1) Pf, Pb, 2) Pf	L	NL&
Aconitase	NL	Gene KO; Inhibitor (Sodium fluoroacetate)	Pf	-	NL
alpha-Ketoglutarate dehydrogenase	NL	Gene KO	Pf	-	NL
Succinyl-CoA synthetase	NL	Gene KO	Pf	-	NL
Protoporphyrino oxidase	L	Inhibitor (Acifluorfen)	Pf	L	NL%

We also evaluated predictions of the effects of gene knockouts and enzyme inhibitors using previously published experimental results (**Table 4.10** & **4.9**). Our updated model had improved accuracy of gene and reaction essentiality predictions, compared to previous models (**Table 4.10**). We predict that there are 159 essential reactions, and 107 lethal single gene knockouts (**Table 4.6** and **4.7**). Of experimentally validated knockouts, iPfal17 accurately predicts essentiality of 79.5% of genes and enzymes tested in *P. falciparum* and 61.4% for those tested in *P. berghei* (**Tables 4.5**, **4.10**, & **4.8**); predictions are also more accurate for gene knockouts and are less accurate in predicting enzyme inhibition (**Table 4.5** & **4.10**).

4.4 Discussion

Data-driven model curation improves predictive capability: Several *P. falciparum* reconstructions have been generated since the publication of iTH366, including those highlighting unique developmental stages within the blood-stage asexual cycle

by integrating stage-specific expression (Fang, Reifman, and Wallqvist 2014), *de novo* reconstructions to implement novel modeling approaches (Chiappino-Pepe et al. 2017), integrated host and pathogen networks (Wallqvist et al. 2016), and those exploring the other life stages of the parasite (Tymoshenko et al. 2013; Bazzani, Hoppe, and Holzhütter 2012). iPfal17 represents the most comprehensive and validated metabolic reconstruction of the asexual blood-stage malaria parasite, *P. falciparum*, to date. With iPfal17, we can simulate growth and predict gene and reaction essentiality and integrate datasets to probe targeted phenotypes, like resistance. It is larger in scope than previous models, includes more gene annotations, and documents literature citations associated with its components (**Table 4.2** & see Appendix for Additional file 3, Table S1, **Figure 4.1**). Moreover, invalid reactions have been removed, improving accuracy (**Table 4.3**). These curation efforts improve the model validity by better recapitulating experimental results, removing functions known to not occur in the asexual blood-stage parasite, and adding functions for which there is experimental evidence. Thus, gene and reaction knockout predictions generated with this model are more accurate. Moreover, iPfal17 has greater interpretability as reaction citations are included and accessible to users.

iPfal17 is similar in functional distribution and scope to other high quality models of apicomplexans, despite its reduced genome size. The *P. falciparum* genome is 23.3 MB and contains 5423 genes (excluding the antigenic var genes) (Aurrecoechea et al. 2009, 2017; Gardner et al. 2002); iPfal17 accounts for the function of 987 metabolites, 730 enzymatic reactions, 1195 total reactions, and 488 genes (**Table 4.2**). For reference, the network reconstruction for *Toxoplasma gondii*, with a genome of 80 MB with 8000 genes (Gajria et al. 2008; Xia et al. 2008), accounts for 1019 metabolites, 1089 enzymatic reactions, 3387 total reactions, and 527 genes (Tymoshenko et al. 2015); with a genome of 32.8 MB and 8272 genes (Ivens et al. 2005), the *Leishmania major* network reconstruction accounts for 1101 metabolites, 1047 enzymatic reactions, 1112 total reactions, and 560 genes (Chavali et al. 2008). These parasites all have notably poor genome annotation (40–60% of the genes are unknown) (Aurrecoechea et al. 2009, 2017) and, thus, have fewer associated genes than many other reconstructions (*e.g.* the *E. coli* and *S. cerevisiae* reconstructions account for 1366 and 910 genes, respectively (Orth et al. 2011; Heavner and Price 2015)).

Intracellular parasites, like *Plasmodium*, require more exchange and transport reactions as they obtain many nutrients from the host environment (Olszewski et al. 2009; McConville 2014; Wellems and Fairhurst 2012; Imlay and Odom 2014; Mazumdar and Striepen 2007). This reliance on the host for metabolic function permits the parasite to increase fitness by reducing its genome and hijacking host function. *P. falciparum* does just that: the parasite remodels the host erythrocyte, generating a vesicular network for protein translocation and increasing host cell permeability

for nutrient acquisition from the host serum (Lanzer et al. 2006; Bannister et al. 2000; Baumeister et al. 2006; Ginsburg et al. 1983; Staines et al. 2006). Thus, the apicomplexan network reconstructions include more transport reactions, many of which are not genetically mapped. Additionally, we chose to exclude an erythrocytic host compartment from the extracellular environment, despite the parasite's intra-host growth (Olszewski et al. 2009; Geary et al. 1985; Geary, Divo, and Jensen 1985; Divo et al. 1985). Other recent reconstructions (Wallqvist et al. 2016; Phaiphinit et al. 2016) have added this compartment, but the erythrocytic compartment is unlikely to improve model function due to the gross disruption of the host membrane as a barrier (Olszewski et al. 2009; Geary et al. 1985; Geary, Divo, and Jensen 1985; Divo et al. 1985).

We generated gene and reaction essentiality predictions with our curated network model, prior to integration of expression data, and found results largely consistent with previous models (Plata et al. 2010) (**Table 4.10**). We identified 159 essential reactions and 107 essential metabolic genes (**Table 4.6** & **4.7**); 24 of these have been empirically tested in cultured *P. falciparum* parasites (**Table 4**, and in *P. berghei*-**Table 4.8**). iPfal17 better predicts experimentally determined essential reactions than previous models, across a broad set of metabolic pathways (**Table 4** and *data not shown*). iPfal17 predictions fail when essential genes or reactions are involved closely with spontaneous reactions (*i.e.* lactoylglutathione lyase is downstream of a spontaneous reaction and upstream of non-metabolic redox products), are in pathways with uncharacterized mechanisms (*i.e.* plasmepsin II in hemoglobin degradation) or if experimental evidence is contradictory (*i.e.* heme biosynthesis pathway; **Table 4.10**).

Because pharmacological enzyme inhibition can be quite noisy and genetic modification has been challenging in *Plasmodium*, the development of CRISPR-Cas9 and other technologies will make it possible to integrate new experimental observations into the model with increasing accuracy (Ghorbal et al. 2014; Lee and Fidock 2014; Wagner et al. 2014; Lu et al. 2016). Until then, the model can be used to identify enzyme inhibitors with off-target effects. For example, within the heme biosynthesis pathway, pharmacological inhibition of aminolevulinic acid dehydrogenase and protoporphyrinogen oxidase kills blood-stage parasites (Ramya et al. 2007); however, disrupting the genes encoding the first (aminolevulinic acid dehydrogenase) and last (ferrochetalase) genes is not lethal in blood-stage parasites (Ke et al. 2014; Nagaraj et al. 2013). iPfal17 predictions are consistent with the gene knockout experiments in *P. falciparum* of Ke et al. (2015), suggesting that the enzyme inhibitors used by Ramya et al. (2007) have off target effects (**Table 4.10** and **Table 4.9**). iPfal17 also fails to predict the lethal nature of adenosine deaminase in purine-free conditions (Ho et al. 2009). Adenosine deaminase converts adenosine to hypoxanthine; as 38 reactions produce AMP, which then generate hypoxanthine products, we propose adenosine

deaminase may be essential for non-metabolic functions or the inhibitor of adenosine deaminase has off-target effects. Furthermore, these results generate hypotheses about the differential metabolic capabilities of *P. falciparum* and *P. berghei*, as experimental results in the rodent parasite conflict with some *P. falciparum* predictions (**Table 4.10** & **4.10**, **Table 4.8**).

4.5 Conclusions

Here, we curated a genome-scale metabolic reconstruction, iPfal17, to represent the metabolism of the asexual blood-stage *P. falciparum* malaria parasite and identified a set of metabolic tasks to evaluate both iPfal17 and future model iterations. Predictions generate with iPfal17 are consistent with experimental literature (codified in metabolic tasks), while generating novel hypotheses about parasite biology. iPfal17 was shared with our published manuscript for broader usage. In fact, several research groups have contacted us with questions about our approach and analysis implementation, demonstrating its utility.

Curation is an iterative process requiring updates as the field develops and as perspectives evolve; for example, this reconstruction is further curated in **Chapter 6.2.2** with metabolomics data and additional metabolic tasks were identified in **Chapter 3**. Future curation will include metabolomics-based curation, like in **Chapter 6.2.2**, as well as validation against a novel genome-wide essentiality screen conducted in *P. falciparum* (Zhang et al. 2018). The curation process, as well as model building and applying models to answer biological questions, expand our understanding of the parasite's metabolism. Moreover, the model can be viewed as a framework for storing thousands of hypotheses, such as hypotheses about gene-protein-reaction associations or enzyme function (*i.e.* reaction formulas), as well as rigorously evaluating and documenting these hypotheses.

Chapter 5: Applying metabolic models to better understand drug resistance

Some of the following text, figures, and tables has been adapted from Carey, Maureen A., Jason A. Papin, and Jennifer L. Guler. "Novel *Plasmodium falciparum* metabolic network reconstruction identifies shifts associated with clinical antimalarial resistance." *BMC Genomics* 18.1 (2017): 543. All analyses in the chloroquine section were performed by Ana Untariou.

5 Applying metabolic models to better understand drug resistance

Well-curated reconstructions, such as our *Plasmodium falciparum* model, can be used to contextualize sparse data (*i.e.* transcriptomics with few annotated transcripts) and make predictions about condition-specific phenotypes. The antimalarial resistant parasite is one such phenotype of great public health relevance. Malaria remains a major public health burden and resistance has emerged to every antimalarial on the market, including artemisinin, the frontline drug, and chloroquine, one of the oldest antimalarials. Our limited understanding of *Plasmodium* biology hinders the elucidation of resistance mechanisms. In this regard, systems biology approaches like genome-scale metabolic modeling can facilitate the integration of existing experimental knowledge and further understanding of these mechanisms.

5.1 Artemisinin

5.1.1 Synopsis

Here, we utilize our genome-scale metabolic network reconstruction of the asexual blood-stage *P. falciparum* parasite, iPfal17, to expand our understanding of metabolic changes that support artemisinin resistance. Integration of clinical parasite transcriptomes into the iPfal17 reconstruction reveals patterns associated with antimalarial resistance. Flux balance analysis and simulation of gene knockouts and enzyme inhibition predict candidate drug targets unique to resistant parasites. These results predict that artemisinin sensitive and resistant parasites differentially utilize scavenging and biosynthetic pathways for multiple essential metabolites, including folate and polyamines. Our findings are consistent with experimental literature, while generating novel hypotheses about artemisinin resistance and parasite biology. We detect evidence that resistant parasites maintain greater metabolic flexibility, perhaps representing an incomplete transition to the metabolic state most appropriate for nutrient-rich blood. Using this systems biology approach, we can more productively analyze and interpret clinical expression data for the identification of candidate drug targets for the treatment of resistant parasites.

5.1.2 Background

Three billion people are at risk for malaria infection globally and treatment approaches are failing. Malaria is caused by *Plasmodium* parasites, and most deaths are associated

with human-infective *P. falciparum*. Without an efficacious vaccine, antimalarials are essential to combat the severity and spread of disease. Combination therapies are implemented to preserve antimalarial efficacy and slow resistance development (Schwartz and Lachish 2016; Olliaro and Taylor 2004; Eastman and Fidock 2009); despite this approach, this eukaryotic pathogen has developed resistance to every antimalarial on the market (Chakraborty 2016; Cowman et al. 2016; Plowe et al. 2007).

Typically, resistance is conferred by genomic changes that lead to drug export or impaired drug binding (for example (Sidhu, Verdier-Pinard, and Fidock 2002)); however, non-genetic mechanisms have also been implicated in *Plasmodium* resistance development (Guler et al. 2015; Herman et al. 2014; Gabryszewski et al. 2016) and other pathogenic organisms, such as *Pseudomonas aeruginosa* (Meylan et al. 2017) (reviewed in (El-Halfawy and Valvano 2012)). These laboratory-based studies provide insight into metabolic flexibility but the presence of relatively few examples limit our understanding of this method of adaptation, especially in malaria. Here, we aim to look beyond genetic mechanisms of resistance to identify resistance-associated metabolic adaptation. We hypothesize that metabolic changes must occur to support the resistance phenotype and resistance-conferring mutations. Ultimately, these changes, or 'shifts,' are required to increase the fitness of resistant parasites, or support the development of additional genetic changes that affect fitness. Metabolic or phenotypic 'background' could be as important as genetic background in the development of resistance.

In clinical malaria infections, artemisinin resistance is established in Southeast Asia (Ashley et al. 2014; Miotto et al. 2015; Straimer et al. 2015). This phenotype is correlated with mutations in the *P. falciparum Kelch13* gene (Ashley et al. 2014; Miotto et al. 2015; Ariey et al. 2014; Brown et al. 2015) and changes in both signaling pathways (Cheng, Kyle, and Gatton 2012; Codd et al. 2011; Straimer et al. 2015; Mbengue et al. 2015) and organellar function (Yang, Little, and Meshnick 1994; Dalal and Klemba 2015; Klonis et al. 2011; Juan Wang et al. 2010; Chen et al. 2014; Vega-Rodríguez et al. 2015; Cobbold et al. 2016; Peatey et al. 2015). Overall, due to the complexity of artemisinin's mechanism of killing (see citations above and (Meshnick 2002; Eckstein-Ludwig et al. 2003; Golenser et al. 2006; Efferth and Oesch 2004; Antoine et al. 2014; Sun et al. 2015; Li et al. 2005)), it has been challenging to separate the causes and effects of resistance. For this reason, there are few novel solutions to antimalarial resistance beyond altering the components of combination therapies to regain efficacy (*e.g.* artemisinin-atovaquone-proguanil (Schwartz and Lachish 2016)).

We aim to gain a new perspective on resistance by viewing it through a 'metabolic

lens'. By characterizing the metabolic shifts that occur during or after resistance acquisition, we can begin to understand more about what it takes to support new functions, such as novel signaling (*e.g.* PI3K signaling is affected by *PfKelch13* mutations (Miotto et al. 2015; Straimer et al. 2015, 2015; Kamau et al. 2015; Isozumi et al. 2015)), drug detoxification (*e.g.* regulating ROS stress associated with artemisinin treatment (Klonis et al. 2011; Juan Wang et al. 2010; Meshnick 2002; Efferth and Oesch 2004)), or stage alterations (*e.g.* dormancy of early ring stages (Cheng, Kyle, and Gatton 2012; Klonis et al. 2013; Mok et al. 2011; Teuscher et al. 2010; Witkowski et al. 2010)) in resistant parasites. Once we identify these compensatory changes, we can potentially target them. *Plasmodium* metabolic genes are better characterized than signaling pathways, as (for example) PlasmoDB identifies 43 3D7 genes associated with the term 'signalling' as opposed to 1112 3D7 genes associated with the term 'metabolism' (Aurrecoechea et al. 2009), and many antimalarials target metabolic functions (Fidock et al. 2000; Peterson, Walliker, and Wellems 1988; Siregar et al. 2015; Phillips and Rathod 2010). Moreover, metabolism has been described as the best-understood cellular process (Fuhrer et al. 2017), making interpreting metabolic analyses more tractable. Ultimately, if we can identify targetable conserved metabolic differences that arise with or in support of resistance, we can develop more robust antimalarial combination therapies aimed at preventing resistance.

Here, we use a systems biology approach to analyze the metabolic profile associated with resistant and sensitive parasites. Using constraint-based metabolic modeling, we integrated transcriptomic data from over 300 clinical isolates from Cambodia and Vietnam with varying levels of artemisinin sensitivity. This approach identified innate metabolic differences that arise with or in support of the resistant phenotype, despite large clinical variability, over multiple genetic backgrounds. Additionally, we were able to explore the functional consequences of expression changes by predicting essential enzymes within these distinct metabolic contexts; these enzymes are candidate drug targets for the prevention of drug resistance.

5.1.3 Methods

Normalized preprocessed data was obtained from GEO (GSE59097) (Mok et al. 2015). Probes on the microarray platform GPL18893 were annotated using NCBI's standalone BLAST correcting the gene labels for 647 probes. Only top hits were used; specifically, hits with greater than 95% identity, no gaps, and a score of over 100 were used (supplemental table too large to display, please see appendix for Additional file 3, Table S8). The R package limma was used to compare artemisinin sensitive and resistant samples collected from Cambodia and Vietnam (Ritchie et al. 2015). Samples with predominantly ring-stage parasites with no detectable gametocytes were used.

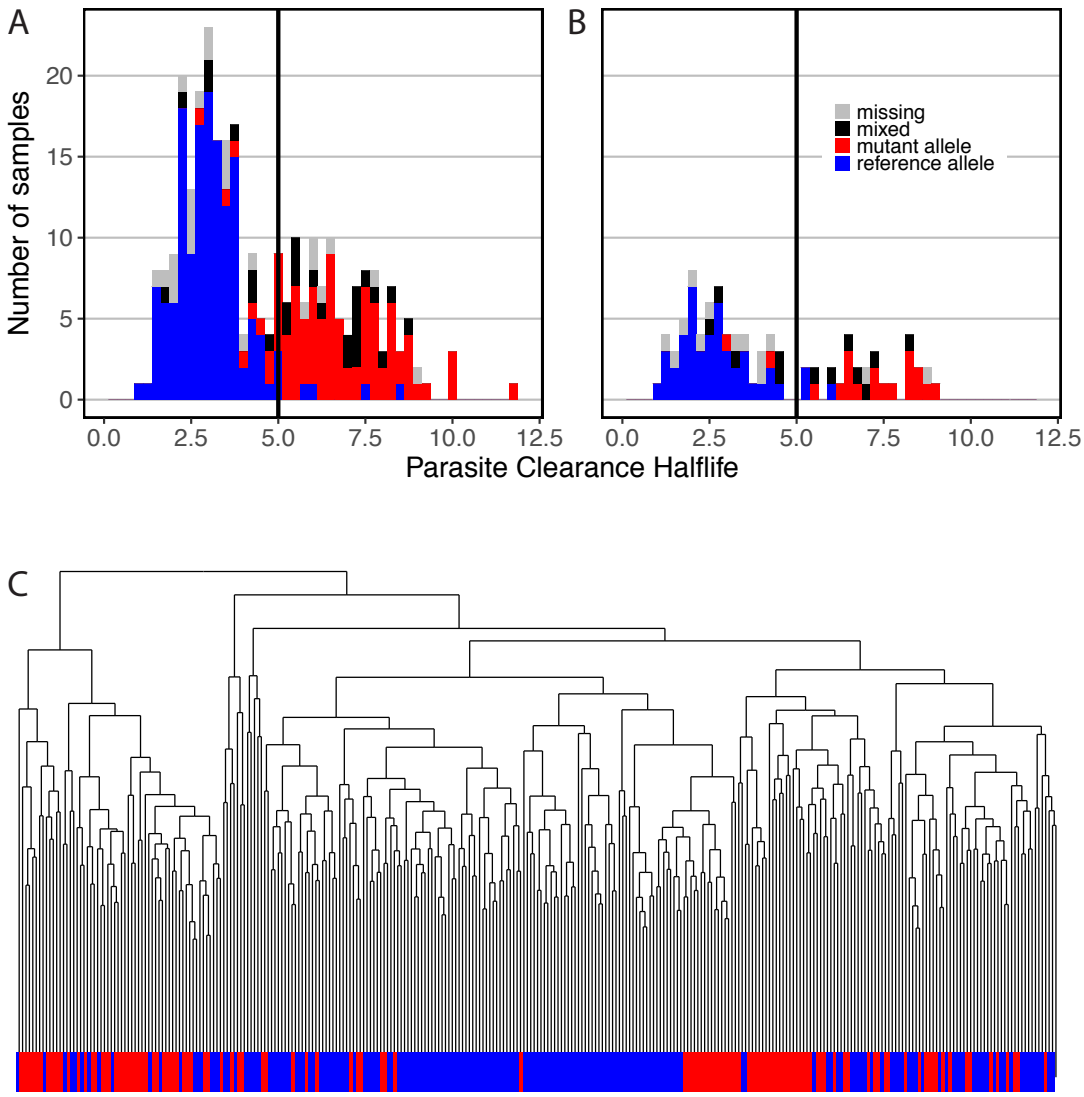


Figure 5.1: Ring-stage parasites are genotypically and phenotypically distinct, yet expression profiles fail to separate resistance phenotypes. Genotypic clustering: Genotypic (any mutation in *PfKelch13*) and phenotypic markers (parasite clearance half-life) were used to define artemisinin resistance in ring-stage parasites from GSE59097; using both markers, resistant and sensitive parasites from Cambodia (A) and Vietnam (B) separated into distinct populations. Genotype was identified in Mok, et al., 2015 with samples classified as containing the reference allele (blue), a mutant allele (red, any in the *PfKelch13* propeller domain), a mixed population (black, at least two reads from each the reference and mutant alleles), or missing (grey, no sequencing data or fewer than 5 reads). C: Phenotypic clustering: Resistant (red) and sensitive (blue) parasites from the two countries fail to cluster with consideration of genome-wide gene expression data (data not shown) or expression of metabolic genes alone.

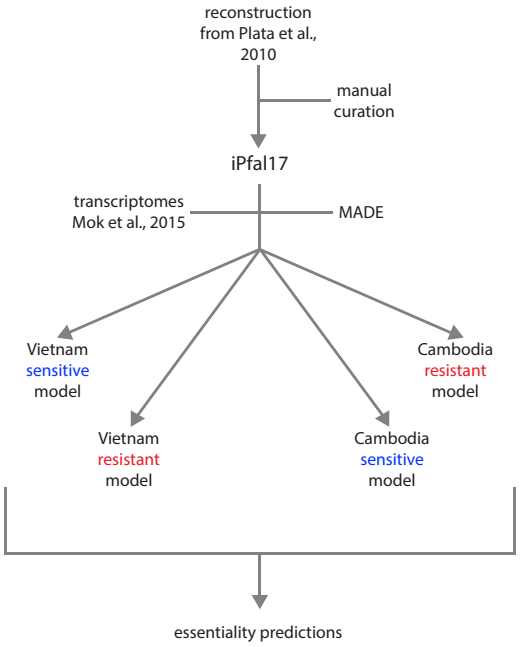


Figure 5.2: Computational pipeline. We curated an existing blood-stage *P. falciparum* reconstruction to generate our iPfal17 network reconstruction. We integrated transcriptomics data into this model using the MADE algorithm to generate four condition-specific models. We used these models to predict reaction essentiality; we highlight consensus results across resistant or sensitive models.

Resistant parasites were defined as both having at least one mutant *Kelch13* allele and a parasite clearance half-life of greater than 5 hours (**Figure 5.1**) (Mok et al. 2015; White et al. 2015). Sensitive parasites were defined by having at least no mutant *Kelch13* alleles and a parasite clearance half-life of less than 5 hours. Random Forest classifiers were built using the R package randomForest, using all ring-stage samples (Liaw, Wiener, and Others 2002). The metadata classifier used the variables listed in **Figure 5.4**, as outlined in the original study (Mok et al. 2015). Cambodian and Vietnamese ring-stage transcriptomes were compared separately to ensure patterns associated with resistance status were reproducible across phylogeny. These countries were chosen for large number of isolates and prevalence of resistance. Microarray probes were screened to remove non-metabolic genes and to keep only one probe per gene (consistent with standard practice). Multiple testing correction was conducted using a false discovery rate to adjust for the potential false positives associated with as standard p-value cutoff and multiple testing (Benjamini and Hochberg 1995; Benjamini and Yekutieli 2001).

Gene expression data with calculations of fold changes and associated adjusted p-value were incorporated into our curated model using the Metabolic Adjustment

for Differential Expression (MADE) algorithm. MADE utilizes statistical significance of gene expression changes along with network context to assign binary gene states ('on'/'off') to each metabolic gene. This constrains the network by limiting flux through reactions mapped to 'off' genes while maintaining growth, or a similar objective. An 80% growth threshold was used given that there is no reported evidence that resistant and sensitive parasites produce variable biomass as measured by the size of ring-stage parasites; while varying this threshold affects sensitive parasite biomass yield, it does not affect essentiality predictions (*data not shown*). Essential genes were predicted for the resultant condition-specific models (**Figure 5.2**) by conducting single gene and reaction deletions with established algorithms (Schellenberger et al. 2011). Consensus lethal gene and reaction deletions from the Cambodian and Vietnamese parasite models were used.

5.1.4 Results

Analysis of artemisinin sensitive and resistant transcriptomes: In order to investigate the presence of a distinct metabolic phenotype in artemisinin resistant parasites, we analysed a previously published expression dataset of clinical isolates from Southeast Asia (NCBI Gene Expression Omnibus accession: GSE59097). Patient blood samples were collected immediately prior to beginning artemisinin combination therapy, and their relative expression was evaluated via microarray (Mok et al. 2015). This dataset profiles (1) *in vivo* artemisinin naïve parasites, providing a view of the innate differences between sensitive and resistance parasites, and (2) a diverse population of parasites collected from multiple collection sites across two countries, allowing us to summarize variable resistant phenotypes that laboratory adapted parasites and *in vitro* assays cannot practically encompass.

We confined our analysis of this previously published expression data to ring-stage parasites from Cambodia and Vietnam, two countries that had clear resistant and sensitive parasite populations as defined by parasite clearance half-life, an *in vivo* phenotypic measure of resistance, and *PfKelch13* mutations, a commonly-used genetic marker of resistance (**Figure 5.1A** & **B**). There were 97 and 24 ring-stage resistant parasite expression profiles from Cambodia and Vietnam, respectively; resistant parasites are defined by both the presence of *PfKelch13* mutations and a parasite clearance half-life of more than 5 hours. There were 141 and 43 ring-stage sensitive parasite expression profiles from Cambodia and Vietnam, respectively, as defined by wild-type *PfKelch13* alleles and clearance half-life of less than 5 hours. Despite obvious genotypic and phenotypic separation (**Figure 5.1A** & **B**), artemisinin sensitive and resistant parasites do not separate well by hierarchical clustering of expression data (**Figure 5.1C**).

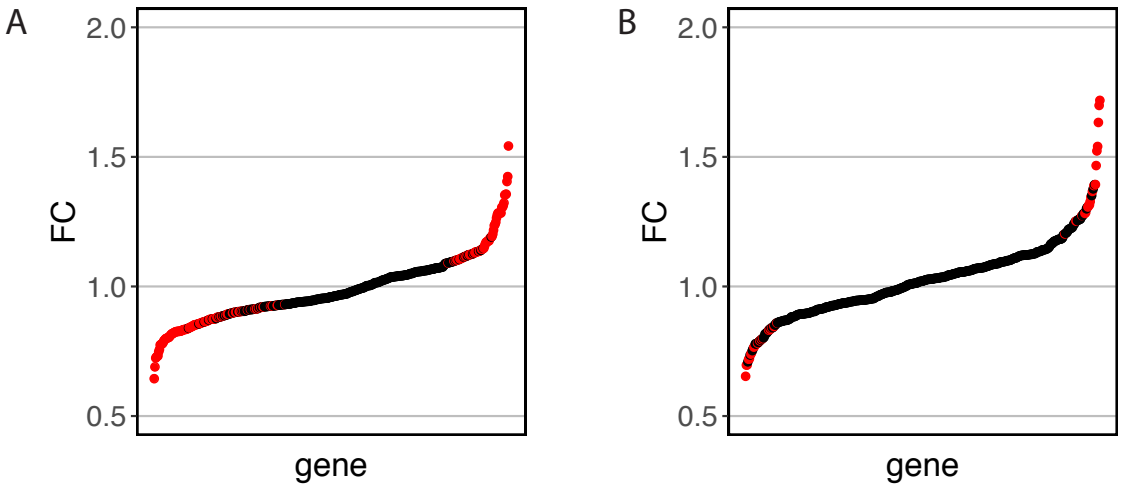


Figure 5.3: Distribution of genome-wide expression data demonstrates moderate differential expression between sensitive and resistant parasites. Fold change values from differential expression between sensitive and resistant parasites from Cambodia (A) and Vietnam (B) with significantly differentially expressed genes in red. Fold change is the ratio of mean expression in resistant parasites to mean expression in sensitive parasites, for each respective country.

Additionally, when comparing sensitive parasites to resistant parasites in either country, the fold change of transcript expression is moderate; no genes exhibited notable differential expression across both analyses (fold change >2 or <0.5 for both Cambodia and Vietnam sample sets, *data not shown*). Among metabolic genes specifically, expression differences are small (maximum fold change 0.6 and 1.6) and few are both significant and conserved between data sets (11 in common from 174 in Cambodia and 37 in Vietnam; **Figure 5.3A** & **B**). Large amounts of transcriptional variation (due to stage-dependent expression, genotypic variability, and host-pathogen interactions) across the population of clinical parasites may hide differences in the data sets. Moreover, we built a Random Forest classifier with expression data to predict resistance outcomes; the classifier predicted resistance poorly, with only 30.77% sensitivity (indicating only 30.77% of resistant samples were correctly identified) and 97.96% specificity (indicating 97.96% of sensitive samples were correctly identified; **Figure 5.4A**).

Although the expression data classifier performed poorly, a similar classifier built from metadata associated with each sample (patient and parasite characteristics) was highly predictive of resistance status with 85.71% sensitivity and 88.91% specificity (**Figure 5.4B**). In our analysis, two specific mutations and collection site were the most predictive of resistance status; removing any of these three variables decreased

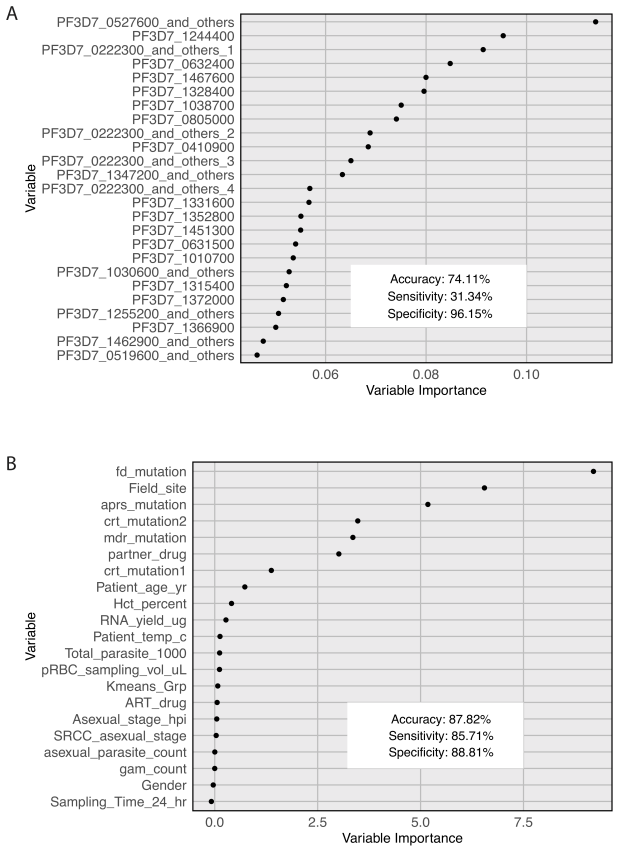


Figure 5.4: Artemisinin resistance is better predicted by metadata classifier than expression classifier. Using full expression profile (A) or metadata (B), including patient and parasite features (see Methods), we can classify samples as artemisinin sensitive or resistant by Random forest analysis. Of the top 25 most important variables (gene probes) in the expression classifier, 12 encoded exported proteins, four genes of complete unknown function, three encoded a putative kinase and putative phosphatases, one encoded a component of dynein, four were uncharacterized genes though to be involved in protein folding or trafficking, and one encoded a transcription factor. Abbreviation key (all from Mok, et al., 2015 or Miotto, et al., 2015) aprs_mutation: apicoplast ribosomal protein S10 (PF3D7_1460900.1) mutation; fd_mutation = ferredoxin (PF3D7_1318100) mutation. Field_site = location at which blood sample was collected; mdr_mutation = multidrug resistance protein 2 (PF3D7_1447900) mutation; partner_drug = Partner drug (Artemisinin based combination treatment) administered from day 3 onwards; crt_mutation2 = second CRT (PF3D7_0709000) mutation measured; crt_mutation1 = first CRT (PF3D7_0709000) mutation measured; Patient_age_yr = patient age in years; pRBC_sampling_vol_uL = Volume of packed RBC collected (uL); RNA_yield_ug = Amount of Total RNA isolated for each sample (ug); Patient_temp_c = patient temperature at time of admission in Celsius; ART_drug = Type and dosage of artemisinin drug given once a day on days 0, 1 and 2; asexual_parasite_count = Total asexual parasite densities per uL on admission; total_parasite_1000 = total number of parasites in whole sample of infected RBC collected (pRBC collection vol. x total parasite count per uL) divided by 1000; SRCC_asexual_stage = Spearman rank correlation coefficient of the gene expression for the isolate sample to the projected hpi; Kmeans_Grp = expression group; Asexual_stage_hpi = Projected hours post invasion (hpi) of the parasite asexual stage; Gender = Patient gender; gam_count = Total gametocyte parasite densities per uL on admission; Hct_percent = patient hematocrit (percent) on admission; Sampling_Time_24_hr = Time of sample collection in 24 hour format.

classifier accuracy by over 20%. If only parasite clearance half life is used to define resistance (omitting the complementary use of *Kelch13* mutations as done throughout chapter), *Kelch13* mutation is the most predictive feature in classifying resistance (*data not shown*). Thus, metadata better predicts resistance than expression data. In order to deconvolve this innate variability and identify functional cellular changes associated with varying levels of artemisinin sensitivity, we integrated metabolic expression data into a genome-scale metabolic model of blood-stage *P. falciparum*.

Integration of expression data into the metabolic model: With our curated metabolic network reconstruction, we integrated expression data from sensitive and resistant parasites collected in Cambodia and Vietnam into iPfal17 using the Metabolic Adjustment for Differential Expression algorithm (MADE (Jensen and Papin 2011)). MADE constrains gene utilization in the model to maximally account for statistically significant changes in expression while maintaining network functionality requirements (*e.g.* parasite viability). MADE integrates differential expression by minimizing the difference between significant expression changes (up/down) and model constraints (gene usage); this avoids arbitrary thresholding and ensures the condition-specific model is consistent with experimental data. Essential genes and genes supported by expression data (by having no change in expression or being upregulated in that condition) remain in the model that represents that condition. Conversely, if a gene is significantly down regulated in a condition and not functionally necessary for metabolism, the reactions catalyzed by the encoded enzyme will be removed from the model. Therefore, condition-specific models contain a reduced network with a subset of reactions annotated in the original curated reconstruction that are either necessary for network functionality (as defined by the objective function) and/or are supported by expression data (Jensen and Papin 2011). Thus, MADE generates functional condition-specific models representing the cell's metabolic capability given the condition-specific expression.

MADE integration of sensitive and resistant expression data from both countries generated four condition-specific models (**Figure 5.2**). By comparing these models, we identified differences in gene and pathway utilization between resistant and sensitive parasites that are consistent between the isolates from the two countries **Figure 5.5**. First, we conducted an enrichment analysis on genes that remain in (*i.e.* can be utilized by) each constrained model by comparing to the unconstrained curated model. As expected, all four models were enriched with genes involved in pathways with many essential reactions or little redundancy, such as transport reactions, tRNA synthesis, purine metabolism, and others (**Figure 5.6**). Sensitive (wild type) models corresponding to isolates from both Cambodia and Vietnam are uniquely enriched with the utilization of genes involved in the metabolism of nicotinate/nicotinamide (p-value = 0.0147), glutamate (p-value = 1.28e13), and selenocysteine (p-value =

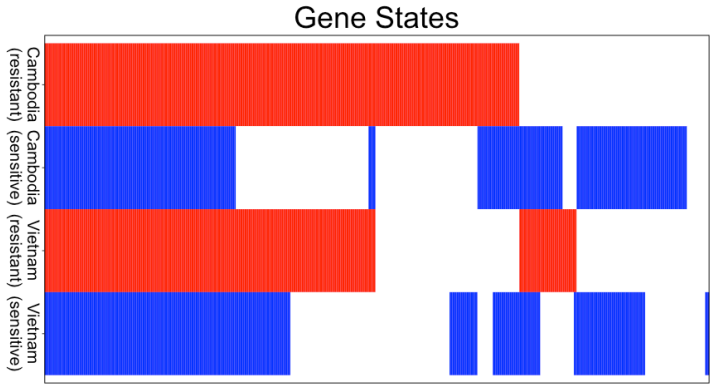


Figure 5.5: Functional differences in data-driven sensitive and resistant models. Gene states from four condition-specific models, the results of MADE integration, cluster by sensitivity not by location. Active genes in red/blue, with genes removed from expression-constrained models in white.

0.000585). Thus, sensitive models contain more reactions in these pathways than the unconstrained model, resulting from increased expression of these pathways in sensitive parasites (**Figure 5.6**). Resistant models from both countries are uniquely enriched with the utilization of genes involved in pyrimidine (p-value = 0.000000218), polyamine (p-value = 0.000439), redox reactions (p-value = 0.0000513), and central carbon metabolism (glycolysis [p-value = 0.000439] and the pentose phosphate pathway [p-value = 0.00606]). Thus, resistant models have a larger proportion of their total reactions associated with these pathways than the original unconstrained model, whereas sensitive models do not have this enrichment. This indicates that these pathways are upregulated in resistant parasites and may remain important for metabolism in the resistant state (**Figure 5.6**).

Table 5.1: Essential reactions unique to resistant parasites. All reactions in table are predicted to be lethal when removed from both Cambodia and Vietnam resistant models. Starred reactions are deleted from at least one resistant model due to expression constraints by MADE.

Reaction	Reaction Formula	EC Number	Reaction Function	Genes
CO2tmt	CO2[m] < => CO2[c]	-	CO2 transport	-
EX_folate4	p-aminobenzoate[e] <	-	p-aminobenzoate exchange	-
EX_fru(e)	fructose[e] < =>	-	fructose exchange	-

Table 5.1: Essential reactions unique to resistant parasites. All reactions in table are predicted to be lethal when removed from both Cambodia and Vietnam resistant models. Starred reactions are deleted from at least one resistant model due to expression constraints by MADE. *(continued)*

Reaction	Reaction Formula	EC Number	Reaction Function	Genes
EX_thm(e)	thiamine[e] < =>	-	thiamine exchange	-
FRUt1r	fructose[e] < => fructose[c]	-	fructose transport	PFB0210c
FUM_mt	fumarate [m] + H2O[m] < => malate[m]	4.2.1.2	fumarate hydratase in the TCA cycle	PFI1340w
FUMtmt	fumarate[m] < => fumarate [c]	-	fumarate transport into mitochondria	-
GHMT2r	serine[c] + thf[c] < = glyine[c] + H2O[c] +	2.1.2.1	serine hydroxymethyl-transferase in folate synthesis	PFL1720w
GLYCL_mt	glycine[m] + NAD[m CO2[m] + mlthf[m]	many	glycine cleavage system in folate synthesis and amino acid metabolism	PF13_0345a; PFL1550w; MAL13P1.390; PF14_0497; PF11_0339
GLYtmt	glycine[m] < => glycine[c]	-	glycine transport into mitochondria	-
HEX7	ATP[c] + fruc-tose[c] = > ADP[c] + fructo 6-phosphate[c] + h[c]	2.7.1.1	hexokinase of glycolysis	PFF1155w
MDHm	malate[m] + NAD[m h[m] + NADH[m] + aloacetate[m]	1.1.1.37	malate dehydrogenase in the TCA Cycle	PFF0895w
MLTHFtmt	mthf[m] < => mthf[c]	-	mthf transport into mitochondria	MAL8P1_13*; PF11_0172
NA	NA	NA	NA	NA
NADPHtmt	NADPH[c] < => NADPH[m]	-	NADPH transport into mitochondria	-
NADPtmt	NADPc] < => NADP[m]	-	NADP transport into mitochondria	-
NH4tmt	NH4[m] < => NH4[c]	-	NH4 transport into mitochondria	-
OAAtmt	oxaloacetate [m] < => oxaloacetate[c]	-	oxaloacetate into mitochondria	-
THFtmt	thf[m] < => thf[c]	-	thf into mitochondria	-
THMt3	h[c] + thi-amine[e] < => h[e] + thiamine[c]	-	thiamine import	-

Table 5.1: Essential reactions unique to resistant parasites. All reactions in table are predicted to be lethal when removed from both Cambodia and Vietnam resistant models. Starred reactions are deleted from at least one resistant model due to expression constraints by MADE. *(continued)*

Reaction	Reaction Formula	EC Number	Reaction Function	Genes
TMDPK	ATP[c] + thiamine[c] => AMP[c] + h[c] + thiamine diphosphate[c]	2.7.6.2	thiamine diphosphokinase in cofactor metabolism	PFI1195c
pABAt	p-aminobenzoate[e] < p-aminobenzoate[c]	-	p-aminobenzoate import	MAL8P1_13a; PF11_0172

Table 5.2: Essential reactions unique to sensitive parasites. All reactions in table are predicted to be lethal when removed from both Cambodia and Vietnam sensitive models. Starred reactions are deleted from at least one sensitive model due to expression constraints by MADE.

Reaction	Reaction Formula	EC Number	Reaction Function	Genes
2_7_8_3	CDP-choline[c] + ceramide[c] + h[c] => CMP[c] + sphingomyelin[c]	2.7.8.3	sphingomyelinase 2 in lipid metabolism	PFF1210w*;PFF1215w
AMETt2	adenosyl methionine[e] + h[e] => adenosyl methionine[c] + h[c]	-	adenosyl methionine import	PF11_0334; PFB0435c*; PFE0775c*; PFF1430c; PFL0420w; PFL1515c*
EX_o2(e)	O2[e] < =>	-	oxygen exchange	-
EX_ptrc(e)	putrescine[e] < =>	-	putrescine exchange	-
GAT_c	diacylglycerol[c] + acoenzyme-A[c] => coenzyme-A[c] + triacylglycerol[c]	2.3.1.20	diacylglycerol O-acyltransferase in lipid metabolism	PFC0995c
GPDDA4	glycerophosphoglyce[c] + H2O [c] => glycerol 3-phosphate[c] + glycerol[c] + h[c]	3.1.4.46	glycerophosphodiest phosphodiesterase in lipid metabolism and glycolysis	PF14_0060
O2t	O2[e] < => O2[c]	-	oxygen import	-
O2tmt	O2[m] < => O2[c]	-	oxygen transport into mitochondria	-
PItap	phosphate[ap] < => phosphate[c]	-	phosphate transport into apicoplast	-

Table 5.2: Essential reactions unique to sensitive parasites. All reactions in table are predicted to be lethal when removed from both Cambodia and Vietnam sensitive models. Starred reactions are deleted from at least one sensitive model due to expression constraints by MADE. *(continued)*

Reaction	Reaction Formula	EC Number	Reaction Function	Genes
PTRCt2	h[e] + putrescine[e] = > h[c] + putrescine[c]	-	putrescine import	-
PYK	ADP [c] + h[c] + phosphoenol pyruvate[c] = > ATP[c] + pyruvate[c]	2.7.1.40	pyruvate kinase in glycolysis	PFF1300w
amet_ex	adenosyl methionine[e] < =>	-	adenosyl methionine exchange	-

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models.

Reactions	Formula	EC	Subsystems	Genes
4HBZtmt	4hbz[m] <=> 4hbz[c]	NA	Transport MT	NA
ADSL1r	dcamp[c] <=> amp[c] + fum[c]	4.3.2.2	Nucleotides PurineMetabolism	PFB0295w
ADSS	asp_L[c] + gtp[c] + imp[c] => dcamp[c] + gdp[c] + 2 h[c] + pi[c]	6.3.4.4	Nucleotides PurineMetabolism ; AminoAcids AsnAspMetabolism	PF13_0287
ALATRS	ala_L[c] + atp[c] + trnaala[c] => alatrna[c] + amp[c] + ppi[c]	6.1.1.7	AminoAcids tRNA	PF13_0354
ARGTRS	arg_L[c] + atp[c] + trnaarg[c] => amp[c] + ppi[c] + argtrna[c]	6.1.1.19	AminoAcids ArgProMetabolism ; AminoAcids tRNA	PFL0900c PFI0680c
ASNTRS	asn_L[c] + atp[c] + trnaasn[c] => amp[c] + ppi[c] + asntrna[c]	6.1.1.22	AminoAcids AsnAspMetabolism ; AminoAcids tRNA	PFB0525w PFE0475w
ASPCT	asp_L[c] + cbp[c] => cbasp[c] + h[c] + pi[c]	2.1.3.2	Nucleotides PyrimidineMetabolism ; AminoAcids AsnAspMetabolism	MAL13P1_221
ASPTRS	asp_L[c] + atp[c] + trnaasp[c] => amp[c] + ppi[c] + asptrna[c]	6.1.1.12	AminoAcids AsnAspMetabolism ; AminoAcids tRNA	PFA0145c PFE0715w
CDPMEK[ap]	4c2me[ap] + atp[ap] => 2p4c2me[ap] + adp[ap] + h[ap]	2.7.1.148	Lipids IsoprenoidsMetabolism	PFE0150c
CHLPCTD	cholp[c] + ctp[c] + h[c] => cdpchol[c] + ppi[c]	2.7.7.15	Lipids PhosphatidylcholineMetabolism	MAL13P1_86

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models. *(continued)*

Reactions	Formula	EC	Subsystems	Genes
CHORS	3psme[c] => chor[c] + pi[c]	4.2.3.5	Cofactors ShikimateBiosynthesis	PFF1105c
CHRPL	chor[c] => 4hbz[c] + pyr[c]	4.1.3.40	Cofactors UbiquinoneMetabolism	NA
CHSTEROLt	atp[c] + chsterol[c] + h2o[c] <=> adp[c] + chsterol[e] + h[c] + pi[c]	NA	Transport e	NA
CMPtap	cmp[ap] <=> cmp[c]	NA	Transport AP	NA
CO2tap	co2[ap] <=> co2[c]	NA	Transport AP	NA
CTPtap	ctp[ap] <=> ctp[c]	NA	Transport AP	NA
CYOOm_mt	4 focytc[m] + 6 h[m] + o2[m] => 4 ficytc[m] + 2 h2o[m] + 6 h[c]	1.9.3.1	Others Mitochondrial-ElectronFlow	PF14_0288 PF14_0331 PF14_0721 PFI1365w PFI1375w mal_mito_2 mal_mito_1
CYOR_u6m_	2 ficytc[m] + q8h2[m] => 2 focytc[m] + 2 h[m] + q8[m]	1.10.2.2	Others Mitochondrial-ElectronFlow	PF14_0248 PF14_0373 PF10_0120 PF3D7_0211300 mal_mito_3
CYSTRS	atp[c] + cys_L[c] + trnacys[c] => amp[c] + cystrna[c] + ppi[c]	6.1.1.16	AminoAcids tRNA	PF10_0149
CYTK1	atp[c] + cmp[c] <=> adp[c] + cdp[c]	2.7.4.14	Nucleotides PyrimidineMetabolism	PFA0555c
DDPA	e4p[c] + h2o[c] + pep[c] => 2dda7p[c] + pi[c]	2.5.1.54	Cofactors ShikimateBiosynthesis	NA
DHAPtap	dhap[c] + pi[ap] => dhap[ap] + pi[c]	NA	Transport AP	PFE0410w
DHFR	dhf[c] + h[c] + nadph[c] <=> nadp[c] + thf[c]	1.5.1.3	Cofactors FolateBiosynthesis	PFD0830w
DHFS	atp[c] + dhpt[c] + glu_L[c] => adp[c] + dhf[c] + h[c] + pi[c]	6.3.2.12	Cofactors FolateBiosynthesis	PF13_0140
DHORD2_mt	dhor_S[m] + q8[m] => orot[m] + q8h2[m]	1.3.5.2	Nucleotides PyrimidineMetabolism ; Others Mitochondrial-ElectronFlow	PFF0160c
DHORTS	dhor_S[c] + h2o[c] <=> cbasp[c] + h[c]	3.5.2.3	Nucleotides PyrimidineMetabolism	PF14_0697
DHORtmt	dhor_S[m] <=> dhor_S[c]	NA	Transport MT	NA
DHPS2	4abz[c] + 6hmhptpp[c] => dhpt[c] + ppi[c]	2.5.1.15	Cofactors FolateBiosynthesis	PF08_0095
DHQS	2dda7p[c] => 3dhq[c] + pi[c]	4.2.3.4	Cofactors ShikimateBiosynthesis	NA
DHQTi	3dhq[c] => 3dhsk[c] + h2o[c]	4.2.1.10	Cofactors ShikimateBiosynthesis	NA
DMATT	dmpp[c] + ipdp[c] => grdp[c] + ppi[c]	2.5.1.1	Lipids TerpenoidMetabolism	PF11_0295
DMPPS[ap]	h2mb4p[ap] + h[ap] + nadph[ap] => dmpp[ap] + h2o[ap] + nadp[ap]	1.17.1.2	Lipids IsoprenoidsMetabolism	PFA0225w

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models. *(continued)*

Reactions	Formula	EC	Subsystems	Genes
DMPPtap	dmpp[ap] <=> dmpp[c]	NA	Transport AP	NA
DPCOAK[ap]	atp[ap] + dpcoa[ap] => adp[ap] + coa[ap] + h[ap]	2.7.1.24	Cofactors CoABiosynthesis	PF14_0415
DPCOAtap	dpcoa[ap] <=> dpcoa[c]	NA	Transport AP	NA
DTMPK	atp[c] + dtmp[c] <=> adp[c] + dtdp[c]	2.7.4.9	Nucleotides PyrimidineMetabolism	PFL2465c
DUTPDP	dutp[c] + h2o[c] => dump[c] + h[c] + ppi[c]	3.6.1.23	Nucleotides PyrimidineMetabolism	PF11_0282
DXPRIi[ap]	dxyl5p[ap] + h[ap] + nadph[ap] => 2me4p[ap] + nadp[ap]	1.1.1.267	Lipids IsoprenoidsMetabolism	PF14_0641
DXPS[ap]	g3p[ap] + h[ap] + pyr[ap] => co2[ap] + dxyl5p[ap]	2.2.1.7	Lipids IsoprenoidsMetabolism	MAL13P1_186
ENO	2pg[c] <=> h2o[c] + pep[c]	4.2.1.11	Carbohydrates Glycolysis	PF10_0155
EX_cholester	chsterol[e] <=>	NA	Exchange	NA
EX_fe2(e)	fe2[e] <=>	NA	Exchange	NA
EX_folate1	6hmhpt[e] <=>	NA	Exchange	NA
EX_folate2	6hmhpt[e] <=> 6hmhpt[c]	NA	Transport cofactor	NA
EX_hb	=> hb[e]	NA	Exchange	NA
EX_ile_L(e)	ile_L[e] <=>	NA	Exchange	NA
EX_nac(e)	nac[e] <=>	NA	Exchange	NA
EX_phosphat	pc[e] <=>	NA	Exchange	NA
EX_pnto_R(	pnto_R[e] <=>	NA	Exchange	NA
EX_ribflv1	ribflv[e] <=>	NA	Exchange	NA
EX_so4(e)	so4[e] <=>	NA	Exchange	NA
EX_spmd(e)	spmd[e] <=>	NA	Exchange	NA
FE2t	fe2[e] => fe2[c]	NA	Transport	PFF0450c
FEROpp	4 fe2[c] + 4 h[c] + o2[c] => 4 fe3[c] + 2 h2o[c]	1.16.3.1	Ion	NA
FMETTRS	10fthf[c] + mettrna[c] => fmettrna[c] + h[c] + thf[c]	2.1.2.9	Cofactors FolateBiosynthesis ; AminoAcids Met-PolyamineMetabolism	MAL13P1_67
FMNAT	atp[c] + fmn[c] + h[c] => fad[c] + ppi[c]	2.7.7.2	Cofactors RiboflavinMetabolism	PF10_0147
GAPD	g3p[c] + nad[c] + pi[c] <=> 13dpg[c] + h[c] + nadh[c]	1.2.1.12	Carbohydrates Glycolysis	PF14_0598
GK1	atp[c] + gmp[c] <=> adp[c] + gdp[c]	2.7.4.8	Nucleotides PurineMetabolism	PFI1420w
GLYTRS	atp[c] + gly[c] + trnagly[c] => amp[c] + ppi[c] + glytrna[c]	6.1.1.14	AminoAcids GlySerMetabolism ; AminoAcids tRNA	PF14_0198
GRTT	grdp[c] + ipdp[c] => frdp[c] + ppi[c]	2.5.1.10	Lipids TerpenoidMetabolism	PF11_0295

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models. *(continued)*

Reactions	Formula	EC	Subsystems	Genes
H2Otap	h2o[ap] <=> h2o[c]	NA	Transport AP	NA
HBZOPT_mt	4hbz[m] + octdp[m] => 3ophb[m] + 2 ppi[m]	2.5.1.39	Cofactors UbiquinoneMetabolism	PFF0370w
HBtr	hb[e] => hb[c]	NA	Transport	NA
HCO3E	co2[c] + h2o[c] => h[c] + hco3[c]	4.2.1.1	Nucleotides PyrimidineMetabolism ; Lipids FattyAcidSynthesis ; Carbohydrates PyruvateMetabolism	PF11_0410
HISTRS	atp[c] + his_L[c] + trnahis[c] => amp[c] + histrna[c] + ppi[c]	6.1.1.21	AminoAcids tRNA	PF14_0428
HMBZ	pheme[fv] => hemozoin[fv]	4.99.1.8	Hemozoin production	PF14_0446
HMBZex	hemozoin[fv] => hemozoin[e]	NA	Hemozoin production	NA
HMGLB	hb[c] => 36 ala_L[c] + 6 arg_L[c] + 10 asn_L[c] + 15 asp_L[c] + 3 cys_L[c] + 4 gln_L[c] + 12 glu_L[c] + 20 gly[c] + 19 his_L[c] + 36 leu_L[c] + 22 lys_L[c] + 5 met_L[c] + 15 phe_L[c] + 14 pro_L[c] + 16 ser_L[c] + 16 thr_L[c] + 3 trp_L[c] + 6 tyr_L[c] + 31 val_L[c] + pheme[fv]	NA	Hemoglobin digestion	MAL13P1_56 PFF1430c PF14_0076 PF14_0077 PF14_0078 PF14_0075 PF13_0133 PF3D7_0311700 PF3D7_1033800 PF3D7_1465700 PF3D7_1430200 PF3D7_0808200 PF11_0161 PF11_0162 PF11_0165 PF14_0015 PF14_0517 PFI1570c PF11_0174 PFL2290w PF13_0322 PF14_0439 PFE1360c MAL8P1_140 PF10_0150 PF14_0327 PFE0355c PFE0370c PF11_0381 PF14_0574
HPPK2	6hmhpt[c] + atp[c] => 6hmhptpp[c] + amp[c] + h[c]	2.7.6.3	Cofactors FolateBiosynthesis	PF08_0095
HXPRT	hxan[c] + prpp[c] => imp[c] + ppi[c]	2.4.2.8	Nucleotides PurineMetabolism	PF10_0121
Hfv	pheme[fv] => pheme[c]	NA	Hemoglobin digestion	NA
Htap	h[ap] <=> h[c]	NA	Transport AP	NA
ILETRS	atp[c] + ile_L[c] + trnaile[c] => amp[c] + iletrna[c] + ppi[c]	6.1.1.5	AminoAcids LeuIleValMetabolism	PF13_0179
ILEt2r	ile_L[e] + leu_L[c] <=> ile_L[c] + leu_L[e]	NA	Transport e	PF11_0334 PFB0435c PFE0775c PFF1430c PFL0420w PFL1515c

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models. *(continued)*

Reactions	Formula	EC	Subsystems	Genes
IPDPS[ap]	h2mb4p[ap] + h[ap] + nadph[ap] => h2o[ap] + ipdp[ap] + nadp[ap]	1.17.1.2	Lipids IsoprenoidsMetabolism	PFA0225w
IPDPtap	ipdp[ap] <=> ipdp[c]	NA	Transport AP	NA
LAC	lac_L[c] => lac_L[e]	NA	Export	PFB0465c PFI1295c
LDH_L	lac_L[c] + nad[c] <=> h[c] + nadh[c] + pyr[c]	1.1.1.27	Carbohydrates Glycolysis	PF13_0141 PF13_0144
LEUTRS	atp[c] + leu_L[c] + trnaleu[c] => amp[c] + leutrna[c] + ppi[c]	6.1.1.4	AminoAcids LeuIleValMetabolism	PF08_0011 PFF1095w
LYSTRS	atp[c] + lys_L[c] + trnalys[c] => amp[c] + ppi[c] + lystrna[c]	6.1.1.6	AminoAcids LysMetabolism ; AminoAcids tRNA	PF13_0262 PF14_0166
Lipid_prod	0.519 chsterol[c] + 14 sphmyln[c] + 1.5 tag[c] + 35 all_pc[c] + 18 all_pe[c] + 4.25 all_pi[c] + 1.5 all_pg[c] + 0.2 all_apg[c] + 4 all_dgl[c] => lipid[c]	NA	Lipids LipidProduction	NA
MECDPDH2[	2mecdp[ap] + nadph[ap] => h2mb4p[ap] + h2o[ap] + nadp[ap]	1.17.7.1	Lipids IsoprenoidsMetabolism	PF10_0221
MECDPS[ap]	2p4c2me[ap] => 2mecdp[ap] + cmp[ap]	4.6.1.12	Lipids IsoprenoidsMetabolism	PFB0420w
MEPCT[ap]	2me4p[ap] + ctp[ap] + h[ap] => 4c2me[ap] + ppi[ap]	2.7.7.60	Lipids IsoprenoidsMetabolism	PFA0340w
METTRS	atp[c] + met_L[c] + trnamet[c] => amp[c] + mettrna[c] + ppi[c]	6.1.1.10	AminoAcids Met-PolyamineMetabolism ; AminoAcids tRNA	PF10_0053 PF10_0340
MTHFC	h2o[c] + methf[c] <=> 10fthf[c] + h[c]	3.5.4.9	Cofactors FolateBiosynthesis	PFF1490w
MTHFD	mlthf[c] + nadp[c] <=> methf[c] + nadph[c]	1.5.1.5	Cofactors FolateBiosynthesis	PFF1490w
NACUP	nac[e] => nac[c]	NA	Transport cofactor	NA
NADK	atp[c] + nad[c] => adp[c] + h[c] + nadp[c]	2.7.1.23	Cofactors NicotinateNicotinamideMetabolism	PFI0650c
NADPHtap	nadph[ap] <=> nadph[c]	NA	Transport AP	NA
NADPtap	nadp[ap] <=> nadp[c]	NA	Transport AP	NA
NADS2	atp[c] + dnad[c] + gln_L[c] + h2o[c] => amp[c] + glu_L[c] + h[c] + nad[c] + ppi[c]	6.3.5.1	Cofactors NicotinateNicotinamideMetabolism	PFI1310w
NAMNPP	atp[c] + h2o[c] + nac[c] + prpp[c] => adp[c] + nicrnt[c] + pi[c] + ppi[c]	2.4.2.11	Cofactors NicotinateNicotinamideMetabolism	PFF1410c

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models. *(continued)*

Reactions	Formula	EC	Subsystems	Genes
NDPK3	atp[c] + cdp[c] => adp[c] + ctp[c]	2.7.4.6	Nucleotides PyrimidineMetabolism ; Nucleotides PurineMetabolism ; Lipids DolicholMetabolism	PF13_0349 PFF0275c
NDPK4	atp[c] + dtdp[c] => adp[c] + dttp[c]	2.7.4.6	Nucleotides PyrimidineMetabolism ; Nucleotides PurineMetabolism ; Lipids DolicholMetabolism	PF13_0349 PFF0275c
NDPK5	atp[c] + dgdp[c] => adp[c] + dgtp[c]	2.7.4.6	Nucleotides PyrimidineMetabolism ; Nucleotides PurineMetabolism ; Lipids DolicholMetabolism	PF13_0349 PFF0275c
NDPK7	atp[c] + dcdp[c] => adp[c] + dctp[c]	2.7.4.6	Nucleotides PyrimidineMetabolism ; Nucleotides PurineMetabolism ; Lipids DolicholMetabolism	PF13_0349 PFF0275c
NDPK8	atp[c] + dadp[c] => adp[c] + datp[c]	2.7.4.6	Nucleotides PyrimidineMetabolism ; Nucleotides PurineMetabolism ; Lipids DolicholMetabolism	PF13_0349 PFF0275c
NNATr	atp[c] + h[c] + nicrnt[c] <=> dnad[c] + ppi[c]	2.7.7.18	Cofactors NicotinateNicotinamideMetabolism	PF13_0159
OCTDPS	frdp[c] + 5 ipdp[c] => octdp[c] + 5 ppi[c]	2.5.1.90	Lipids TerpenoidMetabolism	PFB0130w
OCTDPtmt	octdp[m] <=> octdp[c]	NA	Transport MT	NA
OMPDC	h[c] + orot5p[c] => co2[c] + ump[c]	4.1.1.23	Nucleotides PyrimidineMetabolism	PF10_0225
OPHBDC_m	3ophb[m] + h[m] => 2oph[m] + co2[m]	4.1.1.-	Cofactors UbiquinoneMetabolism	NA
OPHHX_mt	2oph[m] + nadph[m] + 0.5 o2[m] => 2ohph[m] + h2o[m] + nadp[m]	1.14.13.8	Cofactors UbiquinoneMetabolism	PF08_0068
OROTtmt	orot[m] <=> orot[c]	NA	Transport MT	NA
ORPT	orot5p[c] + ppi[c] <=> orot[c] + prpp[c]	2.4.2.10	Nucleotides PyrimidineMetabolism	PFE0630c
PCt	pc[c] <=> pc[e]	NA	Transport lipid	NA
PEPPItap	pep[c] + pi[ap] => pep[ap] + pi[c]	NA	Transport AP	PFE1510c
PGK	13dpg[c] + adp[c] => 3pg[c] + atp[c]	2.7.2.3	Carbohydrates Glycolysis	PFI1105w MAL13P1.40

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models. *(continued)*

Reactions	Formula	EC	Subsystems	Genes
PGM	2pg[c] <=> 3pg[c]	5.4.2.1	Carbohydrates Glycolysis	PF11_0208 PFD0660w PFC0430w
PHEMEtmt	pheme[m] <=> pheme[c]	NA	Transport MT	NA
PHETRS	atp[c] + phe_L[c] + trnaphe[c] => amp[c] + ppi[c] + phetrna[c]	6.1.1.20	AminoAcids PheTyrMetabolism ; AminoAcids tRNA	PF11_0051 PFA0480w PFF0180w PFL1540c
PNTK	atp[c] + pnto_R[c] => 4ppan[c] + adp[c] + h[c]	2.7.1.33	Cofactors CoABiosynthesis	PF14_0200 PF14_0354
PNTOt2	h[e] + pnto_R[e] <=> h[c] + pnto_R[c]	NA	Transport cofactor	PF11_0059
PPA	h2o[c] + ppi[c] => 2 h[c] + 2 pi[c]	3.6.1.1	Nucleotides PurineMetabolism ; Lipids TerpenoidMetabolism	PF14_0541 PFC0710w
PPCDC	4ppcys[c] + h[c] => co2[c] + pan4p[c]	4.1.1.36	Cofactors CoABiosynthesis	MAL8P1_81
PPItap	ppi[ap] <=> ppi[c]	NA	Transport AP	NA
PPItmt	ppi[m] <=> ppi[c]	NA	Transport MT	NA
PPNCL2	4ppan[c] + ctp[c] + cys_L[c] => 4ppcys[c] + cmp[c] + h[c] + ppi[c]	6.3.2.5	Cofactors CoABiosynthesis	PF11_0036 PFD0610w
PROTRS	atp[c] + pro_L[c] + trnapro[c] => amp[c] + ppi[c] + protrna[c]	6.1.1.15	AminoAcids ArgProMetabolism ; AminoAcids tRNA	PFI1240c PFL0670c
PRPPS	atp[c] + r5p[c] => amp[c] + h[c] + prpp[c]	2.7.6.1	Carbohydrates PentosePhosphateCycle	PF13_0143 PF13_0157
PSCVT	pep[c] + skm5p[c] <=> 3psme[c] + pi[c]	2.5.1.19	Cofactors ShikimateBiosynthesis	PFB0280w
PTPATi	atp[c] + h[c] + pan4p[c] => dpcoa[c] + ppi[c]	2.7.7.3	Cofactors CoABiosynthesis	PF07_0018

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models. *(continued)*

Reactions	Formula	EC	Subsystems	Genes
Protein	0.1364 alatrna[c] + 0.0674 cystrna[c] + 0.1 fmettrna[c] + 0.3505 glutrna[c] + 0.1063 histrna[c] + 0.3566 iletrna[c] + 0.3405 leutrna[c] + 0.1019 mettrna[c] + 0.3055 sertrna[c] + 0.2045 thrtrna[c] + 0.1379 tyrtrna[c] + 0.203 valtrna[c] + 0.1349 argtrna[c] + 0.02095 trptrna[c] + 0.1353 glntrna[c] + 0.6245 asntrna[c] + 0.1823 phetrna[c] + 0.1085 protrna[c] + 0.2994 asptrna[c] + 0.1827 glytrna[c] + 0.5124 lystrna[c] => protein[c]	NA	AminoAcids ProteinProduction	NA
R07456	g3p[c] + gln_L[c] + ru5p_D[c] <=> glu_L[c] + pydx5p[c]	NA	Cofactors VitB6Metabolism	PF11_0169
RBFK	atp[c] + ribflv[c] => adp[c] + fmn[c] + h[c]	2.7.1.26	Cofactors RiboflavinMetabolism	MAL13P1_292
RIBFLVt2	h[e] + ribflv[e] => h[c] + ribflv[c]	NA	Transport cofactor	NA
RNDR1	adp[c] + trdrd[c] => dadp[c] + h2o[c] + trdox[c]	1.17.4.1	Nucleotides PyrimidineMetabolism ; Nucleotides PurineMetabolism ; Redox RedoxMetabolism	PF10_0154 PF14_0053 PF14_0352
RNDR2	gdp[c] + trdrd[c] => dgdp[c] + h2o[c] + trdox[c]	1.17.4.1	Nucleotides PyrimidineMetabolism ; Nucleotides PurineMetabolism ; Redox RedoxMetabolism	PF10_0154 PF14_0053 PF14_0352
RNDR3	cdp[c] + trdrd[c] => dcdp[c] + h2o[c] + trdox[c]	1.17.4.1	Nucleotides PyrimidineMetabolism ; Nucleotides PurineMetabolism ; Redox RedoxMetabolism	PF10_0154 PF14_0053 PF14_0352
RPE	ru5p_D[c] <=> xu5p_D[c]	5.1.3.1	Carbohydrates PentosePhosphateCycle	PFL0960w

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models. *(continued)*

Reactions	Formula	EC	Subsystems	Genes
SERTRS	atp[c] + ser_L[c] + trnaser[c] => amp[c] + ppi[c] + sertrna[c]	6.1.1.11	AminoAcids SelenocysteineMetabolism ; AminoAcids GlySerMetabolism	PF07_0073 PFL0770w
SHK3Dr	3dhsk[c] + h[c] + nadph[c] <=> nadp[c] + skm[c]	1.1.1.25	Cofactors ShikimateBiosynthesis	NA
SHKK	atp[c] + skm[c] => adp[c] + h[c] + skm5p[c]	2.7.1.71	Cofactors ShikimateBiosynthesis	PFB0280w PF02_0059 PF02_0060 PF3D7_0206400 PFB0279w
SMPD3l_host	h2o[c] + sphmyln_host[c] => cholp[c] + crm[c] + h[c]	3.1.4.12	Lipids SphingomyelinCeramideMetabolism	PFL1870c
SM_ex	sphmyln[e] <=>	NA	Exchange	NA
SM_host	sphmyln[e] => sphmyln_host[c]	NA	Transport e	NA
SO4ti	so4[e] => so4[c]	NA	Transport	NA
SPMDt2	h[e] + spmd[e] => h[c] + spmd[c]	NA	Transport	NA
THRTRS	atp[c] + thr_L[c] + trnathr[c] => amp[c] + ppi[c] + thrtrna[c]	6.1.1.3	AminoAcids tRNA	PF11_0270
TKT2	e4p[c] + xu5p_D[c] <=> f6p[c] + g3p[c]	2.2.1.1	Carbohydrates PentosePhosphateCycle	PFF0530w
TMDS	dump[c] + mlthf[c] => dhf[c] + dtmp[c]	2.1.1.45	Nucleotides PyrimidineMetabolism ; Cofactors FolateBiosynthesis	PFD0830w
TPI[ap]	dhap[ap] <=> g3p[ap]	5.3.1.1	Carbohydrates Glycolysis ; Lipids IsoprenoidsMetabolism	PFC0831w
TRDR	h[c] + nadph[c] + trdox[c] => nadp[c] + trdrd[c]	1.8.1.9	Redox RedoxMetabolism ; Redox MitochondrialAntioxidantSystem ; Nucleotides PyrimidineMetabolism ; Nucleotides PurineMetabolism	PFI1170c
TRPTRS	atp[c] + trnatrp[c] + trp_L[c] => amp[c] + ppi[c] + trptrna[c]	6.1.1.2	AminoAcids TrpMetabolism ; AminoAcids tRNA	PF13_0205 PFL2485c
TYRTRS	atp[c] + trnatyr[c] + tyr_L[c] => amp[c] + ppi[c] + tyrtrna[c]	6.1.1.1	AminoAcids PheTyrMetabolism	MAL8P1_125 PF11_0181
UMPK	atp[c] + ump[c] <=> adp[c] + udp[c]	2.7.4.14	Nucleotides PyrimidineMetabolism	PFA0555c
VALTRS	atp[c] + trnaval[c] + val_L[c] => amp[c] + ppi[c] + valtrna[c]	6.1.1.9	AminoAcids LeuIleValMetabolism	PF14_0589 PFC0470w

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models. *(continued)*

Reactions	Formula	EC	Subsystems	Genes
biomass_s	biomass[c] =>	NA	Exchange	NA
lipid3	dag[c] <=> 12dgr120[c] + 12dgr140[c] + 12dgr141[c] + 12dgr160[c] + 12dgr161[c] + 12dgr180[c] + 12dgr181[c]	NA	Lipids	NA
pc_prod	pc[c] => all_pc[c]	NA	Lipids LipidProduction	NA
trna_ala	trnaala[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0411500 PF3D7_0620800 PF3D7_0702700
trna_arg	trnaarg[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0529600 PF3D7_1370000 PF3D7_1369800 PF3D7_1341000
trna_asn	trnaasn[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0403000
trna_asp	trnaasp[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0714700
trna_cys	trnacys[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_1370100
trna_gly	trnagly[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_1103200 PF3D7_1370200
trna_his	trnahis[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0706900
trna_ile	trnaile[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0312700 PF3D7_0410200
trna_leu	trnaleu[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0510600 PF3D7_0527800 PF3D7_0620900 PF3D7_0714800 PF3D7_1103300
trna_lys	trnalys[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0707100 PF3D7_0707000
trna_met	trnamet[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_1438300 PF3D7_1339100
trna_phe	trnaphe[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0514400
trna_pro	trnapro[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_1216800 PF3D7_1418400 PF3D7_1339200
trna_ser	trnaser[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0410100 PF3D7_0621600 PF3D7_0714900 PF3D7_1337600
trna_thr	trnathr[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0706800 PF3D7_0730700 PF3D7_1355400
trna_trp	trnatrp[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_1369900

Table 5.3: Consensus predicted lethal reactions across 4 expression-constrained models. *(continued)*

Reactions	Formula	EC	Subsystems	Genes
trna_tyr	trnatyr[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0702800
trna_val	trnaval[c] <=>	NA	AminoAcids tRNA ; Exchange	PF3D7_0312600 PF3D7_0730600 PF3D7_1251900

Identification of conserved and uniquely essential pathways: Beyond general differences in pathway utilization, which encompasses both essentiality and pathway-level differences in expression, artemisinin sensitive and resistant parasites have unique essential genes and reactions. To identify these essential reactions and provide insight on targetable metabolic enzymes in the clinical isolates, we performed *in silico* single gene and reaction deletions with each of the four condition-specific models. Datasets from the parasites from each country were initially analyzed separately and then lists were compared to ensure resistance-associated trends are reproducible and observed in independent analyses. As expected, we identified many essential functions conserved in all models (**Table 5.3**), which is consistent with an active core metabolism required for basic parasite survival. Importantly, 21 reactions were essential in only resistant models, but not sensitive models (**Table 5.1**). Theoretically, drugs targeting these reactions would kill resistant parasites and have no effect on sensitive parasites; thus, there would be no selective pressure within the sensitive parasite population to develop resistance to these drugs. This list included serine hydroxymethyltransferase (PFL1720w in folate metabolism), the glycine cleavage system (PFL1550w and others in folate metabolism), thiamine diphosphokinase (PFI1195c in cofactor metabolism, specifically thiamine diphosphate), fumarate hydratase and malate dehydrogenase (PFI1340W and PFF0895w, respectively, in the mitochondrial electron transport chain and TCA cycle), and fructose hexokinase (PFF1155W in glycolysis; **Table 5.1**). We also identified 12 reactions that were essential only in artemisinin sensitive parasites (**Table 5.2**). Drugs targeting these reactions should not be combined with artemisinin, as they would not kill (and may select for) resistant parasites. Fortunately, no existing drug targets were found in this list of essential genes and reactions (**Table 5.2**). Among those identified were several transport reactions highlights the potential for condition-specific intra-organellar function (**Table 5.1** & **5.2**). Overall, our systems biology-based approach reveals unique metabolic phenotypes associated with artemisinin sensitivity; these differences were not detected in the original analysis of the expression dataset or by separately analyzing Cambodian or Vietnamese isolates ((Mok et al. 2015) and *data not shown*).

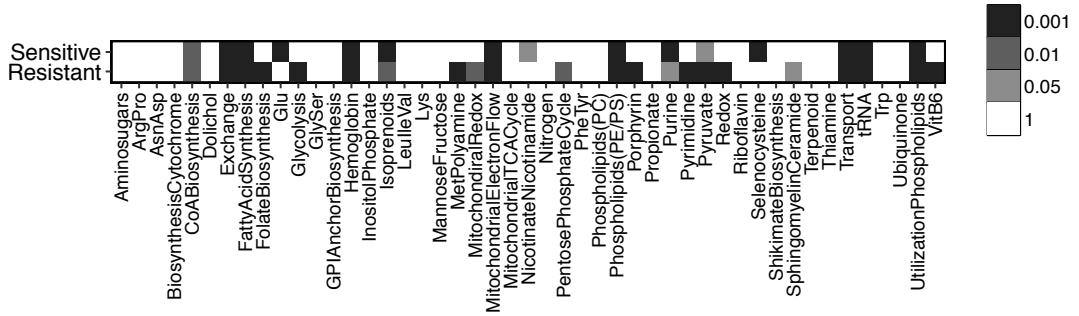


Figure 5.6: Artemisinin sensitive and resistant parasites utilize different metabolic genes and pathways. Enrichment analysis of gene utilization in sensitive and resistant parasite models demonstrates functional differences in expression data integration. Consensus gene utilization from resistant and sensitive models (both Cambodian and Vietnamese datasets) were used and compared to unconstrained model. Black = $p < 0.001$, grey = $p < 0.01$, light grey = $p < 0.05$, white = non significant. Abbreviation key: Aminosugars = amino sugar metabolism; ArgPro = arginine and proline metabolism; AsnAsp = asparagine and aspartate metabolism; BiosynthesisCytochrome = biosynthesis of cytochromes; CoABiosynthesis = coenzyme-A biosynthesis; Dolichol = dolichol metabolism; Exchange = exchange reactions; FattyAcidSynthesis = fatty acid synthesis; FolateBiosynthesis = folate biosynthesis; Glu = glutamate metabolism; Glycolysis = glycolysis; GlySer = glycine and serine metabolism; GPIAnchorBiosynthesis = GPI anchor biosynthesis; Hemoglobin = hemoglobin degradation (including hemozoin formation); InositolPhosphate = inositol phosphate metabolism; Isoprenoids = isoprenoid metabolism; LeuIleVal = leucine, isoleucine, and valine metabolism; Lys = lysine metabolism; MannoseFructose = mannose and fructose metabolism; MetPolyamine = methionine and polyamine metabolism; MitochondrialElectronFlow = mitochondrial electron transport chain; MitochondrialTCACycle = mitochondrial tricarboxylic acid cycle; NicotinateNicotinamide = nicotinate and nicotinamide metabolism; Nitrogen = nitrogen metabolism; PentosePhosphateCycle = pentose phosphate cycle; PheTyr = phenylalanine and tyrosine metabolism; Phosphatidylcholine = phosphatidylcholine metabolism; PhosphatidyletanolaminePhosphatidylserine = phosphatidyletanolamine and phosphatidylserine metabolism; Porphyrin = porphyrin metabolism; Propionate = propionate metabolism; Purine = purine metabolism; Pyrimidine = pyrimidine metabolism; Pyruvate = pyruvate metabolism; Redox = redox metabolism; RedoxMitochondrialAntioxidantSystem = mitochondrial redox metabolism; Riboflavin = riboflavin (vitamin B2) metabolism; Selenocysteine = selenocysteine metabolism; ShikimateBiosynthesis = shikimate biosynthesis; SphingomyelinCeramide = sphingomyelin and ceramide metabolism; Terpenoid = terpenoid metabolism; Thiamine = thiamine biosynthesis; Transport = transport reactions; tRNA = tRNA and protein synthesis; Trp = tryptophan metabolism; Ubiquinone = ubiquinone metabolism; UtilizationPhospholipids = utilization of phospholipids; VitB6 = pyridoxal (vitamin B6) metabolism.

5.1.5 Discussion

Systems biology approaches enable unbiased analyses of antimalarial resistance phenotypes. Here, we describe how a curated metabolic network reconstruction of the malaria parasite can serve as a platform for the analysis of gene expression and other 'omics data, and as a tool to generate testable hypotheses regarding essential genes and metabolic phenotypes. In particular, we used this network reconstruction to characterize key metabolic dependencies in resistant and sensitive parasites. We revealed emergent patterns in pathway activity, differential utilization of organelles, metabolic flexibility, and targetable weakness of resistant parasites.

Data integration reveals distinct metabolic patterns: The integration of expression data from clinical parasites into our network reconstruction highlights the differential utilization of metabolic genes and reveals metabolic shifts associated with variation in innate artemisinin sensitivity (**Figure 5.5** & **Figure 5.6**). Enriched metabolic pathways detected in sensitive and resistant models are consistent with previous experimental observations. For example, resistant models are uniquely enriched with genes involved in pyrimidine biosynthesis and mitochondrial redox reactions. This finding is consistent with the importance of mitochondrial function in surviving artemisinin stress (Chen et al. 2014; Peatey et al. 2015) and the physical interactions between artemisinin and proteins involved in glycolysis, nucleotide synthesis, and the mitochondria in mammalian cells and *P. falciparum* (Zhou, Li, and Xiao 2016; Ismail et al. 2016; Prieto et al. 2008). Additionally, the metabolic disruption of the redox reactions in the electron transport chain upon artemisinin treatment (via decreased production of orotate and fumarate, presumably via dihydroorotate dehydrogenase and succinate dehydrogenase enzymes (Yang, Little, and Meshnick 1994; Cobbold et al. 2016; Creek et al. 2016)) suggests that changes in these pathways may be important for survival in the presence of the drug. Thus, this metabolic network analysis approach allows us to filter out noise from diverse clinical isolates to identify alternative utilization of pathways associated with artemisinin resistance. However, due to the nature of this type of analysis, these enrichment results do not implicate specific reactions that are uniquely active in artemisinin sensitive or resistant parasites.

Condition-specific models have unique metabolic requirements: Upon integration of expression data and the identification of differentially utilized pathways above, we next used these models to predict targetable differences in sensitive and resistant parasites by identifying reactions that are essential within the context of the metabolic network (**Table 5.3**, **Table 5.1** & **5.2**). We identified (1) differences in intra-organellar function, (2) metabolic flexibility of scavenging and biosynthesis pathways, and (3) targetable weakness of resistant parasites. These metabolic shifts primarily

reside in mitochondrial metabolism, as well as folate and polyamine metabolism. Together, these results highlight the overall plasticity of *P. falciparum* metabolism and opportunities for further development of potential drug targets.

Interestingly, several transport reactions are found to be differentially essential in our constrained models (**Table 5.1** & **5.2**). Many transport reactions (79.5%) have no associated gene due to the incomplete characterization of the *P. falciparum* genome (**Figure 4.1**). They are included in the model due to biochemical evidence or functional necessity (*i.e.* a metabolite is produced in one compartment but it is a substrate for an enzyme in another). Transcriptomic data integration does not constrain their behavior explicitly: expression integration reduces the total number of reactions in a model, forcing transport of metabolites among organelles if within-compartment biosynthesis is non-functional. Function within organelles requires transport and loss of function reduces transport needs. Specifically, several mitochondrial and apicoplast transport reactions are uniquely essential in the sensitive and resistant parasite populations (**Figure 5.7**). In resistant models, this includes the mitochondrial transport of metabolites associated with the TCA cycle and electron transport chain (fumarate, oxaloacetate, and NADPH) and those involved in generation of folates (tetrahydrofolate, glycine, CO2, and NH4+) (**Figure 5.7A**). In sensitive models, apicoplast transport of ADP, ATP, and phosphate is essential (**Figure 5.7A**). Overall, these results indicate that sensitive and resistant parasites are differentially utilizing pathways within these organelles, and have unique requirements for transport of essential substrates. This observation is consistent with previous studies and our enrichment results highlighting the influence of mitochondrial metabolism on survival in the presence of artemisinin (Chen et al. 2014; Peatey et al. 2015). Moreover, oxygen transport into the cell and then into the mitochondria is only essential in sensitive parasites, further predicting differential use of the mitochondria in these parasites as oxygen serves as the terminal step in the electron transport chain. Resistant parasites are predicted to generate oxygen within the mitochondria via superoxide dismutase as opposed to transport (**Figure 5.7A**).

We also identify differential utilization of transport pathways from the extracellular environment into the parasite. *Plasmodium* metabolism contains redundancies; for many essential metabolites, the parasite's genome encodes one or more biosynthetic pathways, while there is also evidence for a parallel host-scavenging pathway (Ginsburg et al. 1983) (*e.g.* lipid (Gulati et al. 2015) and amino acid (Gulati et al. 2015; Liu et al. 2006) scavenging). Upon model integration, we find that artemisinin resistant and sensitive parasites utilize some of these metabolic pathways in alternative ways (**Figure 5.7A**). Bioinformatic analyses indicate *Plasmodium* can either scavenge or synthesize putrescine (Aurrecoechea et al. 2009) and adenosylmethionine (Aurrecoechea et al. 2009) (two essential polyamines and precursors to spermidine

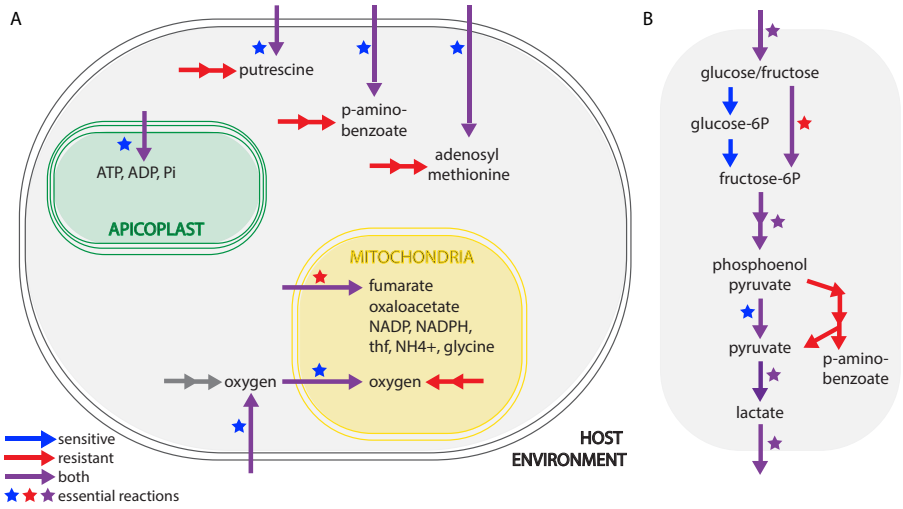


Figure 5.7: Artemisinin resistant and sensitive parasites have unique metabolite transport capabilities. A: Transport differences: Resistant parasites exhibit greater metabolic flexibility, allowing either import or biosynthesis of putrescine, p-aminobenzoate, adenosyl-methionine into the parasite's cytoplasm (grey). Sensitive parasites rely on import only Import or synthesis of ATP, ADP, and phosphate into the apicoplast (green organelle) is essential for sensitive parasites. Resistant parasites require transport of oxygen, fumarate, oxaloacetate, NADP, NADPH, tetrahydrofolate (thf), NH4, and glycine into the mitochondria, in yellow. B: p-aminobenzoate in glycolysis: Resistant parasites generate p-aminobenzoate via alternative components of the glycolysis pathway. Arrows colored for flux via flux variability analysis and stars for essentiality.

(Plata et al. 2010; Pretzel et al. 2016)). Similar redundancy has been identified for the acquisition of p-aminobenzoate, a folate precursor generated by branch of glycolysis necessary for nucleotide synthesis ((Aurrecoechea et al. 2009; Salcedo-Sora et al. 2011); **Figure 5.7A** & **B**, **Table 5.1** & **5.2**). These metabolites are measurable via blood sample metabolomics (Salcedo-Sora et al. 2011; Wishart et al. 2007); therefore, host scavenging is a viable option for blood-stage parasites. We predict that sensitive parasites rely on the import of putrescine, adenosylmethionine, and p-aminobenzoate. Resistant parasite expression supports either host scavenging or direct biosynthesis due to parasite survival upon reaction knockout *in silico*. Thus, we expect that resistant parasites are more metabolically flexible for these metabolites; perhaps resistant parasites have failed to appropriately modulate their transition to the nutrient-rich blood-stage environment, and this unexpected flexibility is evolutionarily beneficial once confronted with artemisinin treatment.

Interestingly, recent metabolomics studies demonstrate that intra-parasitic putrescine levels are decreased upon artemisinin treatment (Creek et al. 2016). Furthermore, protein interaction studies indicate artemisinin covalently binds with spermidine synthase and adenosylmethionine synthetase (Wang et al. 2015). Activity in both the biosynthetic and scavenging pathway of putrescine and adenosylmethionine may allow resistant parasites to compensate for artemisinin's effect on polyamines. The essential role of polyamines is well established in *Plasmodium* (Singh et al. 1997; Assaraf et al. 1987). In other organisms, these compounds stabilize DNA and RNA (Stevens 1970) and signal a pause in the cell cycle (Mandal et al. 2013). In the presence of artemisinin, perhaps polyamines act to stabilize the genome from oxidative stress (Klonis et al. 2011; Meshnick 2002; Efferth and Oesch 2004) and trigger dormancy (Cheng, Kyle, and Gatton 2012; Codd et al. 2011). As resistant parasites are more likely to survive dormancy, flexibility in polyamine metabolism could provide more routes for artemisinin survival (Peatey et al. 2015; Teuscher et al. 2012).

Our systems biology approach also identifies metabolic weaknesses of resistant parasites; these weaknesses can be used to identify drug targets for combination therapies (**Figure 5.8**). For example, we identified the mitochondrial import of fumarate and subsequent conversion to oxaloacetate (via fumarate hydratase, PFI1340W, and malate dehydrogenase, PFF0895W) to be uniquely essential in resistant parasites (**Figure 5.8A**, **Table 5.1**). Expression data from sensitive parasites supports mitochondrial import of malate and utilization of malate:quinone oxidoreductase (PFF0815W) to generate oxaloacetate from malate, bypassing the need for fumarate and the associated enzymes, fumarate hydratase and malate dehydrogenase. We predict that inhibitors of fumarate transport or fumarate hydratase and malate dehydrogenase would specifically kill artemisinin resistant parasites, offering an example of enhanced metabolic flexibility of sensitive parasites and a potential artemisinin-combination therapy target. The

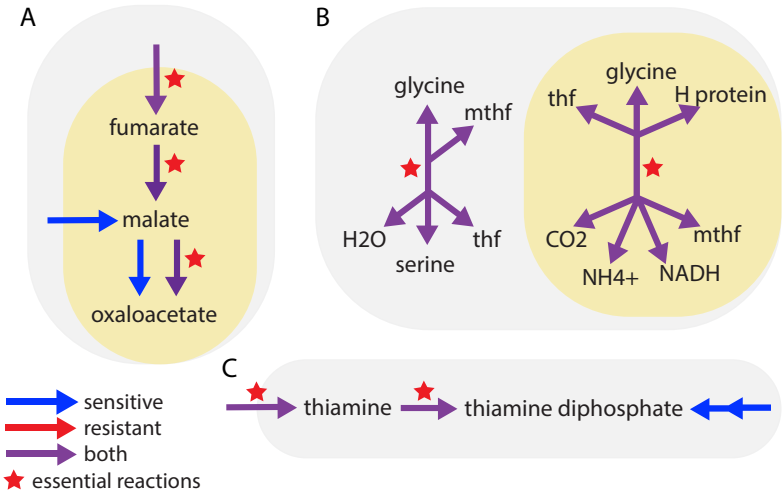


Figure 5.8: Artemisinin resistant parasites display unique metabolic weaknesses. A: Trycarboxylic acid cycle: Resistant parasites rely on generation of oxaloacetate from the conversion of fumarate to malate, using fumarate hydratase and malate dehydrogenase, in the mitochondria. Sensitive parasites can also import malate into the mitochondria and use an alternative enzyme (malate:quinone oxidoreductase) to convert malate to oxaloacetate. B: Folate metabolism: Inhibition of the serine hydroxylmethltransferase (SHMT) enzyme (left) and the glycine cleavage system (right) is lethal in resistant parasites. Sensitive parasites can use either of these enzyme complexes interchangeably to produce methyltetrahydrofolate and tetrahydrofolate. C: Cofactor synthesis: The import of thiamine and the conversion of thiamine to thiamine diphosphate via thiamine thiphosphokinase is essential in resistant parasites. Sensitive parasites can also synthesize thiamine diphosphate de novo. Arrows colored for flux via flux variability analysis and stars for essentiality. Gray background indicates cytosolic localization, yellow indicates mitochondrial localization.

TCA cycle is essential during the mosquito-stage of parasite development (MacRae et al. 2013; Srivastava et al. 2016), but not the blood-stage (Ke et al. 2015; MacRae et al. 2013); this once again highlights the possibility that resistant parasites exhibit incomplete transition to the metabolic state most appropriate for nutrient-rich blood.

Additionally, we identified serine hydroxymethyltransferase (SHMT) and thiamine diphosphokinase as potential drug targets of resistant parasites (**Table 5.1**, **Figure 5.8B & C**); see below for discussion of SHMT. Both the import of thiamine and thiamine diphosphokinase are essential only in resistant parasites (**Figure 5.8C**), and we predict inhibition of import or enzyme activity would specifically target resistant parasites. These reactions are relatively uncharacterized as the parasite can likely synthesize thiamine diphosphate (vitamin B1) *de novo* (Müller and Kappes 2007). Thus, this approach can generate novel hypotheses and be utilized for the identification of novel drug targets, and, importantly, targets to help prevent the development of resistance.

Data-driven model implementation highlights knowledge gaps: Although iPfal17 represents our best understanding of intra-erythrocytic *P. falciparum* biochemistry as the most comprehensive reconstruction to date, predictions occasionally contradict published experimental results. These results illuminate experimental complexities and incompletely characterized pathways. For example, our model predicted that cytosolic SHMT is only essential in resistant parasites (**Figure 5.8B** left). In sensitive parasites, the essential metabolites can be generated by SHMT or the mitochondrial glycine cleavage system, given the reversible nature of these enzymes (Maenpuen et al. 2009; Salcedo, Sims, and Hyde 2005). Therefore, in our sensitive models, neither SHMT nor the glycine cleavage system is essential when knocked out individually. This observation conflicts with the literature, as SHMT is essential in cultured parasites (Maenpuen et al. 2009; Pornthanakasem et al. 2012; França et al. 2005). Thus, iPfal17 is unable to predict this intricacy of parasite metabolism, revealing interesting regulatory effects, an uncharacterized location dependency for metabolite generation, or *in vivo*/*in vitro* differences in enzyme reversibility.

Similarly, model integration reveals that protein localization influences essentiality predictions. We predicted that the cyclical oxidization and reduction of glutathione, a key regulator of oxidative stress (Schulz et al. 2000; Mittler 2002; Becker et al. 2004; Färber et al. 1996), and supporting reactions were essential only in resistant parasites when the glutathione redox system was located within the mitochondria (*data not shown*). This is consistent with artemisinin's induction of reactive oxygen species, the parasite's obvious need to survive this stress (Klonis et al. 2011; Efferth and Oesch 2004; Antoine et al. 2014; Sun et al. 2015), data showing artemisinin sensitivity is

correlated with glutathione levels in rodent *Plasmodium* (Vega-Rodríguez et al. 2015), and artemisinin's inhibition of mammalian glutathione s-transferases (Zhou, Li, and Xiao 2016). However, upon moving these reactions to the cytosolic and apicoplast compartments (as supported by (Kehr et al. 2010)), these reactions were no longer essential. Thus, model analysis challenges the integration of previously incomparable datasets by demonstrating that this localization and role of glutathione yield different predictions. Future studies will be required to clarify these findings.

5.1.6 Conclusions

Here, we have presented an investigation of the metabolic differences between artemisinin sensitive and resistant parasites. Antimalarial resistance is a major public health problem and we demonstrate that constraint-based modeling can be used to reveal metabolic shifts that arise with or in support of the resistant phenotype and discrepancies between otherwise incomparable datasets. We find inherent differences in artemisinin resistant and sensitive parasite metabolism, even before artemisinin treatment. Artemisinin resistant parasites have major metabolic shifts in the mitochondria and in the synthesis of folates and polyamines, indicating incomplete transition to the metabolic state most appropriate for the blood-stage environment. These findings generate novel hypotheses about *Plasmodium* biochemistry and perspectives on antimalarial resistance.

5.1.7 Experimental interrogation

In silico predictions can be explored experimentally. We focused on the subset of predictions presented in **Figure 5.7** highlighting the different use of nutrient scavenging and *de novo* synthesis in artemisinin sensitive and resistant parasites. Specifically, we predicted the differential use of putrescine synthesis (further elaborated in **Figure 5.9**) by sensitive and resistant parasites. Resistant parasites are simulated to use both scavenging and synthesis pathways to generate a putrescine pool for conversion to spermidine and spermine; thus, we predicted that resistant parasites were not sensitive to inhibition of scavenging or of *de novo* synthesis, due to their redundancy. Sensitive parasites, however, have decreased expression of *de novo* enzymes and are accordingly simulated to rely only on scavenging. We predict these parasites would be sensitive to inhibition of putrescine transport. As previously mentioned, metabolomics studies reveal putrescine abundance is decreased upon artemisinin treatment (Creek et al. 2016) and artemisinin interacts with downstream polyamine enzymes (spermidine synthase and adenosylmethionine synthetase (Wang et al. 2015)). Thus, we performed

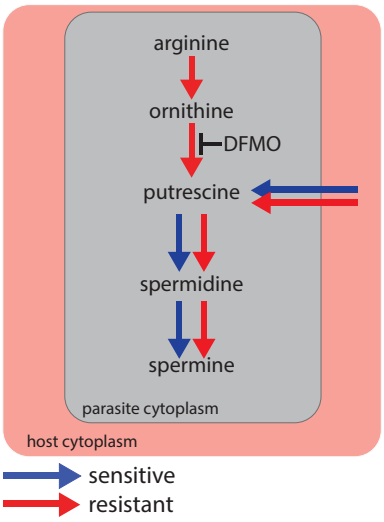


Figure 5.9: Putrescine scavenging and synthesis in *Plasmodium falciparum*. Reactions in blue are used by artemisinin sensitive clinical isolates and reactions in red are used by resistant isolates, prior to drug treatment.

three sets of experiments on artemisinin sensitive and resistant parasites to explore differences in putrescine metabolism.

To determine if putrescine supplementation enhances growth, parasites (*P. falciparum* from BEI Resources, NIAID, NIH, strains IPC 5202/MRA-1240 and IPC 4884/MRA-1238, contributed by Didier Menard) were growth under normal growth conditions (see **Parasite cultivation** in **Chapter 6.3**) with 200uM putrescine. Putrescine enhanced resistant parasite growth significantly in one experiment (n = 16-24, t-test p-value = 8.6x10ˆ-8); however, this failed to replicate (n = 48, p-value = 0.7305). Putrescine did not enhance sensitive parasite growth in either experiment.

To determine if sensitive and resistant parasites are differentially affected by an inhibitor of *de novo* putrescine synthesis, parasites were growth with difluoromethylornithine (DFMO; **Figure 5.9**). Sensitive and resistant parasites were plated into a 96-well plate at a parasite density of approximately 1% parasitemia and 0.5% hematocrit. Parasites were treated with 0.01-27mM DFMO for 72 hours, and grown in triplicate at each DFMO concentration. Following growth, total growth was measured using Sybr green-based flow cytometry (Bei et al. 2010). Parasite proliferation was expressed as a percentage of the media-only control. Resultant normalized growth values were fit to a nonlinear regression with four parameters and fit with a least square fit in R, per standard in the field, and an EC50 was calculated. Artemisinin

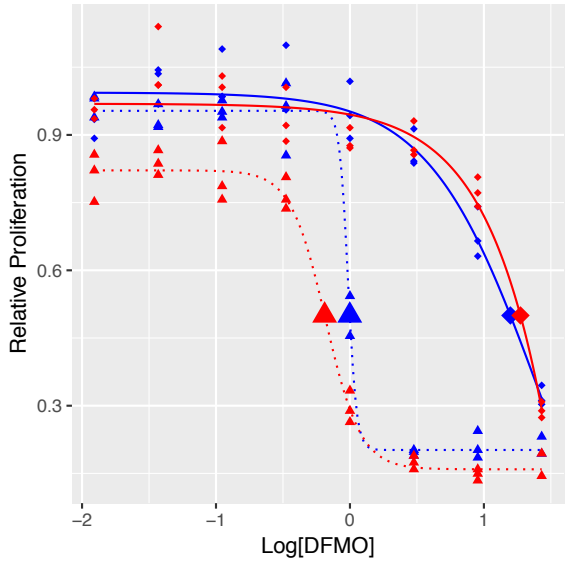


Figure 5.10: Artemisinin sensitive and resistant parasites are sensitive to DFMO treatment. Sensitive parasites in red and resistant parasites in blue. DFMO treatment in dotted line, DFMO with putrescine rescue in solid line. Calculated EC50 is shown with large point.

sensitive and resistant parasites were equally sensitive to DFMO treatment (dotted lines, **Figure 5.10**). Consistent with others in the field, we also rescued DFMO treatment by added exogenous putrescine (200uM), and putrescine could rescue the growth defect in both clonal lines (solid lines, **Figure 5.10**). However, independent experimental replicates (*data not shown*) yielded unpredictable growth of untreated parasite, making results challenging to interpret without adequate replicates.

Lastly, to test if the larger putrescine pool generated by redundant use of scavenging and *de novo* synthesis mediates artemisinin sensitivity, we tested parasite sensitivity to artemisinin in the presence of putrescine using a ring stage survival assay (Witkowski et al. 2013). In brief, a laboratory-adapted parasites (NF54; BEI Resources, NIAID, NIH, strain MRA-1000) were tightly synchronized to the ring stage (see **Chapter 6.3** for methods) and exposed to artemisinin for 6 hours, with or without putrescine treatment (200uM). After 6 hours, parasites were spun and washed in fresh media to remove artemisinin. Growth was measured using Sybr green-based flow cytometry (Bei et al. 2010) and microscopy. Sensitive parasites remained sensitive to artemisinin even with concurrent putrescine treatment (**Figure 5.11**).

In conclusion, clinical isolates with varying sensitivity to artemisinin have differential expression of the enzymes involved in putrescine synthesis prior to drug treatment (**Figure 5.9**) and, following drug treatment, putrescine abundance drops

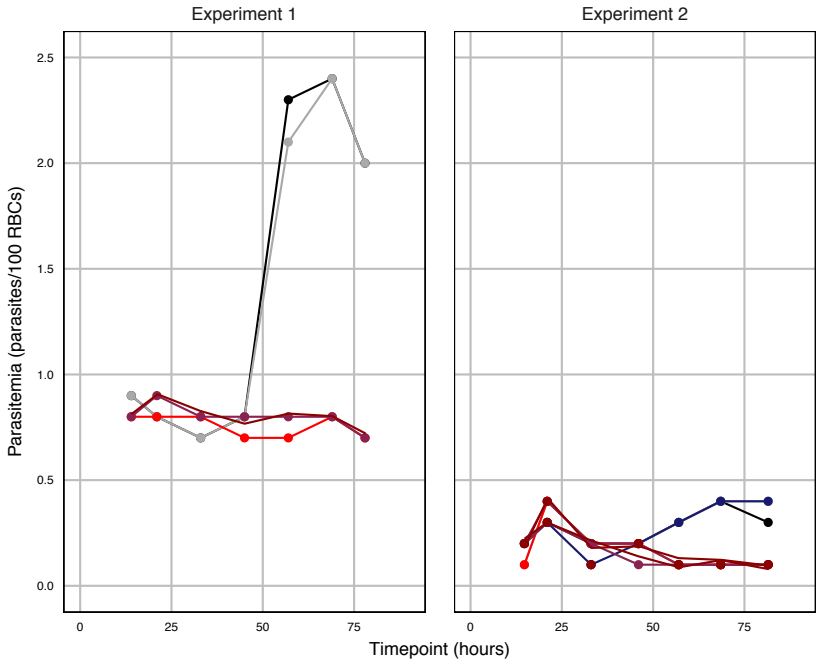


Figure 5.11: Putrescine does not affect sensitivity to artemisinin. Artemisinin was administered at time 0, and washed at 6 hours (beginning of curve shown). Black = untreated control, grey = putrescine-only control, red = artemisinin only control, maroon = artemisinin with putrescine treatment.

in *in vitro* parasite lines (Creek et al. 2016). Artemisinin interacts with the enzyme that consumes putrescine (Wang et al. 2015). We now show, in unreplicated studies, that artemisinin sensitive and resistant *P. falciparum* strains (IPC 5202/MRA-1240 and IPC 4884/MRA-1238) are both sensitive to inhibitors of *de novo* putrescine synthesis and this can be rescued by exogenous putrescine supplementation (**Figure 5.10**). Exogenous putrescine may enhance resistant parasite growth but a larger putrescine pool does not mediate artemisinin resistance (**Figure 5.11**). Future work remains to replicate these findings and continue to profile the role of putrescine in artemisinin resistance. However, as polyamine concentrations (putrescine, spermine, and spermidine, specifically) change as human erythrocytes age (Cooper, Shukla, and Rennert 1976) and likely vary from individual-to-individual (Wishart et al. 2007), it is possible that this variability results from distinct and dynamic host polyamine pools. Blood batch-matched results will aid in the interpretation of these experiments.

5.2 Chloroquine

5.2.1 Synopsis

Using a similar systems biology approach and analytic pipeline, we identified metabolic shifts in chloroquine resistant parasites when treated with chloroquine. All analyses in this section were performed by a talented biomedical engineering undergraduate student, Ana Untariou. We worked together for three years and I am grateful for this experience, as well as training in mentorship and 'PIship.'

5.2.2 Background

First introduced in 1934, chloroquine was a front-line antimalarial until the late 1950s when its heavy usage led to emergence of resistant *P. falciparum* strains near the Cambodia-Thailand border (Payne 1987). Chloroquine resistance has now been confirmed in over 40 countries, making resistance to this drug a global concern (Payne 1987). Despite efforts to introduce novel and effective antimalarial therapeutics, efficacy is diminished by the development of multi-drug resistance, including resistance for artemisinin-based combination therapies, mefloquine, quinine, atovaquone, and sulfadoxine-pyrimethamine (Ashley et al. 2014; Price et al. 2004; Zalis et al. 1998; Färnert et al. 2003; Rønn et al. 1996). Interestingly, when chloroquine usage is removed from a community, the parasite population becomes resensitized to the drug (Hayward, Saliba, and Kirk 2005; Ord et al. 2007). Accordingly, reinstating chloroquine is appealing due to the lack of novel antimalarials. The mechanism

of action and the resistance mechanism of chloroquine are well-studied; however, leveraging these mechanisms has not yet realized combination therapies. Thus, it is appealing to develop combination therapies that target chloroquine resistant parasites under drug pressure as these parasites are weakened; this may slow the development of resistance.

5.2.3 Methods

Normalized microarray expression values of *in vitro* K1 parasites untreated or treated with IC50 concentrations of chloroquine for 4hr and 24hrs were used to investigate the transcriptional effects of chloroquine treatment (GEO accession number: GSE31109, Kritsiriwuthinan et al. (2011)). Expression data was analyzed with limma (Ritchie et al. 2015) to calculate differential expression (specifically, fold change and associated p-value). Condition-specific models were produced by integrating gene expression from the 4 and 24 hour treatment conditions into our blood-stage *P. falciparum* genome-scale metabolic reconstruction (**Chapter 4**) using the Metabolic Adjustment for Differential Expression (MADE) algorithm (Jensen and Papin 2011). MADE constrains the network by limiting flux through reactions mapped to lowly expressed genes while maintaining growth. Essential reactions were predicted for the resultant condition-specific models by conducting single gene and reaction deletions with established algorithms. See **Chapter 4** for more methodological details regarding reaction essentiality and flux balance analysis. Exposure for 4 and 24 hours were defined as short and long-term treatment, respectively.

5.2.4 Results and Discussion

To understand the effect of chloroquine treatment on resistant parasites, we implemented our analytic pipeline described in the previous section to compare transcriptomics data (GSE31109) from chloroquine resistant parasites in the presence or absence of the drug. Strikingly, these resistant parasites downregulated gene expression following treatment (**Figure 5.12**). This indicates that despite being resistant to the drug, these parasites are affected by chloroquine treatment. Longer exposure time exacerbates this effect, as illustrated by the increase in differentially expressed genes with the 24 hour drug treatment (**Figure 5.12A** to **B**). This observation is consistent with growth defects observed in chloroquine-treated resistant parasites (Yayon et al. 1983).

We next integrated these expression profiles into our *P. falciparum* metabolic model to generate condition-specific models for chloroquine-resistant (untreated) par-

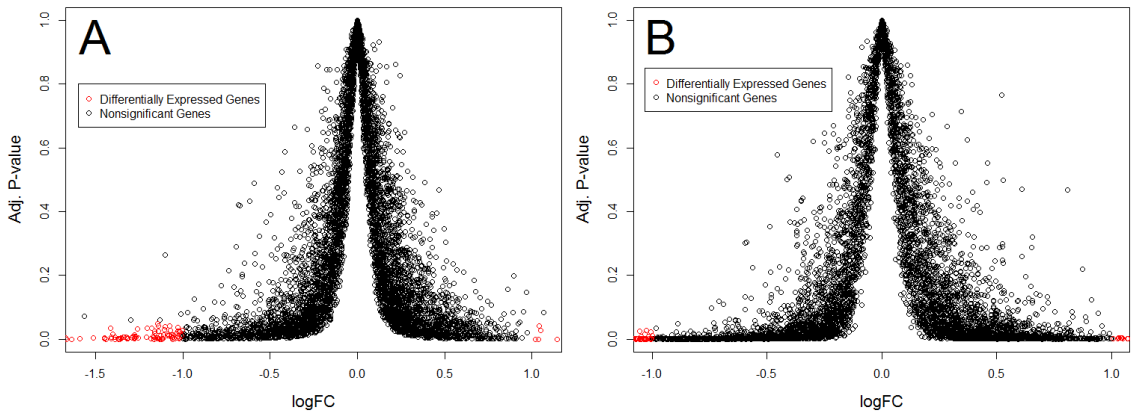


Figure 5.12: Resistant parasites respond to chloroquine by downregulating expression. Adjusted p-values represent the significance level of changes in expression. Fold change quantifies the variation in the gene expression relative to untreated resistant parasite expression during short term treatment (A) or long-term treatment (B).

asites, and chloroquine-resistant parasites treated with short or long-term chloroquine treatment. With the resultant condition-specific models, we identified essential reactions to find targetable pathways in resistant parasites (**Figure 5.13A**). One hundred and eighty metabolic reactions (out of 1192) are essential in all three models (**Figure 5.13A**, center). Consistent with previous studies, this work identified an increased importance of lipid synthesis (**Figure 5.13B**) in chloroquine-treated resistant parasites. *De novo* synthesis of thiamine diphosphate, the active form of vitamin B1, is uniquely essential during short-term chloroquine treatment (**Figure 5.13C**), when normally these cells can import or synthesize thiamine. During long-term chloroquine treatment, the conversion of chorismate into 4-aminobenzoate for folate metabolism and import of oxidized glutathione are essential (**Figure 5.13D**).

Flux levels of these reactions generated from flux balance analysis were investigated to understand thiamine diphosphate usage, in light of the thiamine production switch occurring in **Figure 5.13C**. Because thiamine diphosphate-dependent enzymes are in glycolysis and isoprenoid metabolism, two essential pathways, these reactions are constitutively essential in all condition-specific models. Thus, we use flux balance analysis to simulate steady-state flux values through the network's reactions to see if the flux in these essential pathways varies. We focus on isoprenoid metabolism because nonessential steps in this pathway become essential in the long-term chloroquine treatment model (**Figure 5.13D**). In response to drug treatment, flux through a thiamine diphosphate-dependent reaction and several other reactions in isoprenoids metabolism are consistently greater in response to chloroquine treatment (**Figure**

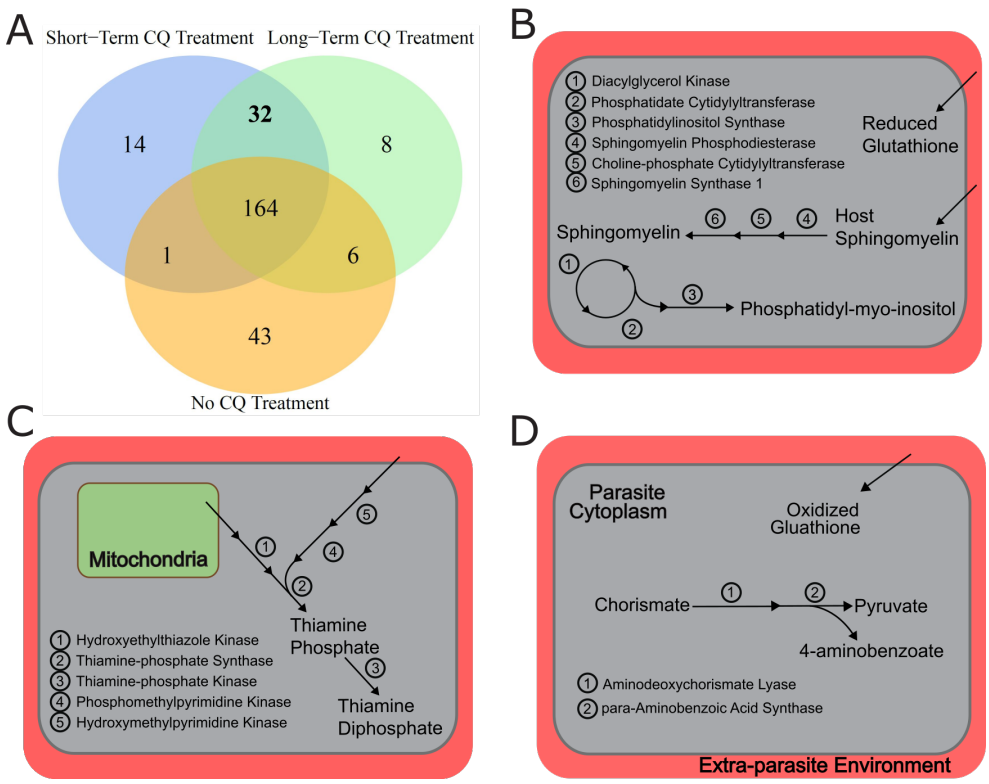


Figure 5.13: Chloroquine-treated resistant parasites have new metabolic weaknesses. (A) Comparison of essential reactions in the three condition models (short-term and long-term chloroquine treatment, and untreated condition). (B-D) Illustrations of common essentiality predictions (bolded in A) between the drug-treatment models. Panel B represents sphingomyelin and ceramide metabolism and inositol phosphate metabolism. Panel C display essential reactions observed only in short-term treatment, representing thiamine metabolism. Panel D displays essential reaction in only long-term treatment, especially folate metabolism. Red depicts the host red blood cell, grey is the parasite's cytoplasm, and green represents the parasite's mitochondria.

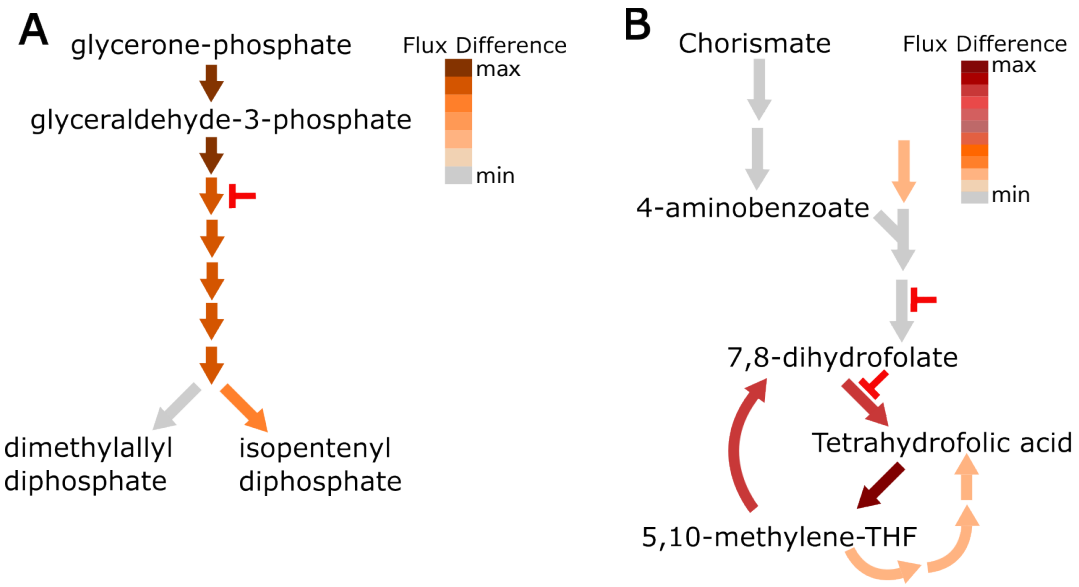


Figure 5.14: Flux shifts in isoprenoids metabolism and folate biosynthesis in response to chloroquine. Mean differences in flux values in (A) isoprenoids metabolism and (B) folate metabolism reactions in the chloroquine-treated versus untreated models. Clinically available drugs target these pathways; drug targets are labeled in the figure with red inhibitor arrows.

5.14A).

Thus, we predict a novel role for isoprenoids synthesis in chloroquine resistant parasites. Under chloroquine treatment, there is increased flux through reactions in the non-mevalonate pathway for isoprenoids metabolism (**Figure 5.14A**), the only synthesis pathway for isoprenoids in *P. falciparum* (Vial, Philippot, and Wallach 1984; Grellier et al. 1994). This pathway is thiamine diphosphate-dependent and we also observed a switch in thiamine scavenging to *de novo* synthesis early in chloroquine treatment (**Figure 5.13C**), highlighting the dynamic state of these pathways. Our computational analysis suggests that chloroquine-resistant parasites have increased susceptibility to non-mevalonate pathway inhibitors due to increased flux in this pathway. This pathway is targetable using drugs such as fosmidomycin (**Figure 5.14A**) and its derivative, FR-900098 (Jomaa et al. 1999). Fosmidomycin is moderately effective against chloroquine-resistant parasites (Tahar and Basco 2007); we propose that these parasites are even more susceptible to fosmidomycin while under chloroquine treatment. Thus, we have propose a novel combination therapies involving readily available antimalarials that may inhibit the growth of chloroquine resistant parasites.

We also predict differential flux in folate metabolism. Folate metabolism is needed for DNA synthesis and metabolism of certain amino acids (Metz 2007). During long-term chloroquine treatment, synthesis of 4-aminobenzoate from chorismate is used, rather than importing 4-aminobenzoate from the host cell (**Figure 5.13D**). Interestingly, downstream steps in folate metabolism, including dihydropteroate synthase and dihydrofolate reductase, carry more flux during chloroquine treatment (**Figure 5.14B**), implying they are necessary for survival or are overexpressed during treatment. This suggests that this pathway has increased importance under chloroquine treatment and could be targeted in combination therapies.

Recent clinical use of such a combination therapy supports this conclusion; chloroquine in combination with inhibitors of dihydrofolate reductase and dihydropteroate synthase (sulfadoxine-pyrimethamine) is effective against chloroquine-resistant parasites (Bustos et al. 1999; Hallett et al. 2006; Menard et al. 2005). Our results suggest that chloroquine-resistant parasites are more susceptible to these drugs than sensitive parasites. Thus, we generated the novel observation that resistant parasites carried greater metabolic flux in isoprenoids and folate metabolism when treated with chloroquine. Inhibitors for these areas of metabolism have been identified for *P. falciparum* and represent candidate drugs for chloroquine combination therapies. Experimental validation of these hypotheses are underway.

5.2.5 Conclusions

New cellular functions (*e.g.* tolerance of artemisinin or chloroquine) result in collateral changes to the genome and to cellular metabolism. We identified inherent differences in artemisinin resistant and sensitive parasite metabolism, even before artemisinin treatment, and unique ways chloroquine resistant parasites respond to drug treatment. These findings generate areas of future research to elucidate *Plasmodium* biochemistry, understand the evolution of resistant parasites, and tackle antimalarial resistance.

Chapter 6: Extending modeling with metabolomics

Some of the following text, figures, and tables has been adapted from Carey, Maureen A., Vincent Covelli, Audrey Brown, Gregory L. Medlock, Mareike Haaren, Jessica G. Cooper, Jason A. Papin, and Jennifer L. Guler. "Influential Parameters for the Analysis of Intracellular Parasite Metabolomics." *mSphere* 3.2 (2018): e00097-18. Many thanks to the other authors for their contributions to this work. Specifically, experimental work was conducted by Vincent Covelli, Audrey Brown, Mareike Haaren, and Jessica G. Cooper. Gregory L. Medlock assisted with the design of machine learning analyses. Ana Untariou implemented the analyses presented in Figures 6.1 and 6.2.

6 Extending modeling with metabolomics

Models serve as tools to contextualize data and to generate data-driven, condition-specific hypotheses. Throughout this dissertation, we have demonstrated the utility of GENREs for data contextualization and interpretation. Moreover, with these modeling approaches, we have identified divergent pathways in parasite metabolism that can be leveraged for informing drug development (**Chapter 3**) and novel targets of drug resistant parasites (**Chapter 5**). However, the predictions generated by these models are limited by the quality and quantity of data used for their construction. Thus, integrating high-quality, stage-specific data will improve their predictive capabilities. Accordingly, parasite models are weakly predictive when compared to GENREs of 'model' organisms (*i.e. E. coli, S. cerevisiae*) due to the lack of historic data and poor genome annotation.

Data integration into GENREs (like in **Chapter 5**) improves predictions but offers additional challenges. Transcriptomic data is most frequently used to constrain GENREs to generate condition-specific models. However, regulation occurs at the translational level as well as transcriptional level, so constraining a GENRE with proteomics data may generate more accurate models and, therefore, predictions than transcriptomic-derived models. We explore the effects of using transcriptomic or proteomic data to constrain metabolic reconstructions.

To investigate the difference between models derived from either transcriptomic or proteomic data, annotated enzyme expression from matched transcriptomics and proteomics data (from five biological conditions, with 3 replicates each from GSE65209 and PXD001659) was integrated into a GENRE (Bosi et al. 2016). We compared reaction essentiality predictions generated to understand the effects of data type on resulting predictions. As transcriptomic datasets have higher coverage than many proteomics datasets, the effects of data coverage were investigated by sampling subsets of data (10-100% of the dataset). These 'matched' predictions were, surprisingly, no more similar than predictions generated from 'mismatched' models (**Figure 6.1**).

Transcriptomic-derived models predicted more essential reactions than proteomics models, even when controlling for scope (*i.e.* number of data points, or enzymes, integrated; **Figure 6.2**). This indicates transcriptomic-derived models are more constrained than proteomics-derived models. This set of analyses identified inconsistencies between individual predictions generated by 'matched' proteomics- and transcriptomics-derived conditional models and the scope of predictions generated by these models. These results emphasize the importance of interpreting predictions as both condition- and datatype-specific results and highlight the need for a better understanding of the intricacies of datatype-specific results.

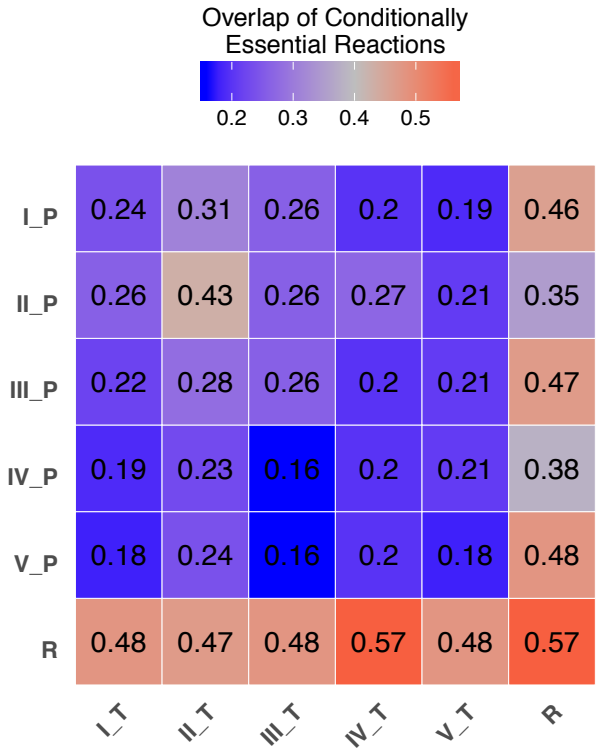


Figure 6.1: Poor predictions overlap between transcriptomics- and proteomics-derived models. Proportions of essential reaction overlap between condition-specific constrained models. Sample numbers (I_, II_, etc.) indicate the timepoint for data collection and sample letter (_P, _T) indicate the datatype, proteomics and transcriptomics, respectively. R indicates randomly generated datasets were used to constrain the model, ten times. The diagonal column represents proportion of the essential reaction overlap among transcriptomics- and proteomics-derived conditional models from matched datasets. The rightmost vertical column represents the median prediction overlap among proteomics-derived condition-specific models and matching random models (10 total random models). The bottom horizontal column represents the median prediction overlap among transcriptomics-derived condition-specific models and matching random models (10 total random models). Of note, the matched datasets (diagonal comparison) are no more similar than models constrained with unmatched datasets (off-diagonal comparison).

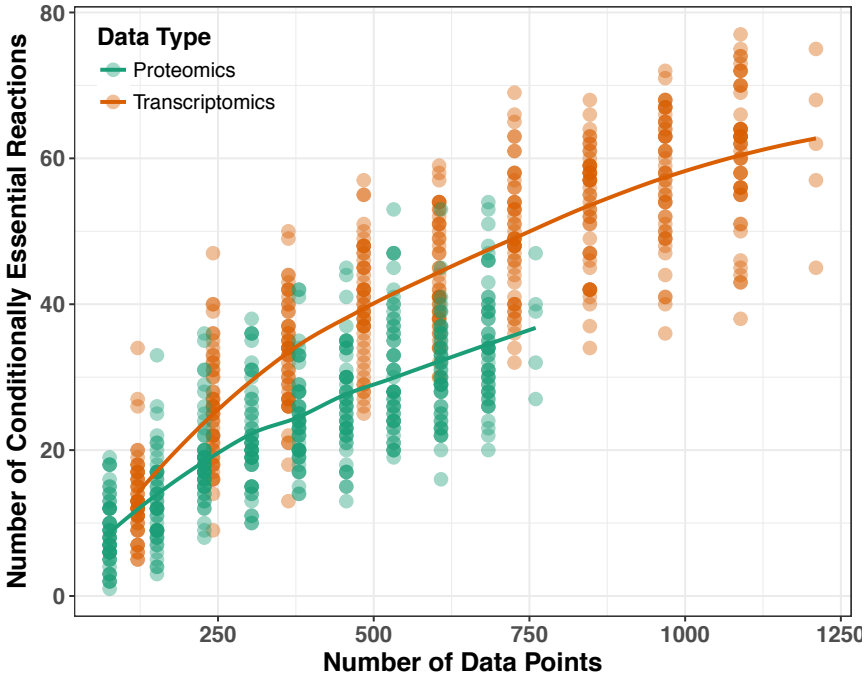


Figure 6.2: Transcriptomic data integration generates more constrained models. Each point represents a model generated from a subset of data (for one biological condition, using the mean of three biological replicates). To subset data, the data were bootstrapped 10 times at each subsetting threshold (10-100% of the dataset). The X-axis represents the number of enzymes that were constrained. Proteomics data are lower coverage than transcriptomic data; here, proteomics contained fewer than 800 enzymes, whereas the transcriptomics contained data for over 1200 enzymes. The Y-axis indicates how many essential reactions were predicted by each model.

Fortunately, metabolomics data offers a new profiling method that is closer to the phenotype than expression data and agnostic to genome annotation. Molecular characterization of pathogens such as the malaria parasite can lead to improved biological understanding and novel treatment strategies. Untargeted metabolomics is a promising approach to learn about pathogen biology. By measuring many small molecules in the parasite at once, we gain a better understanding of important pathways that contribute to the parasite's response to perturbations such as drug treatment. Although increasingly popular, approaches for intracellular parasite metabolomics and subsequent analysis are not well explored. The findings presented in this chapter emphasize the critical need for improvements in these areas to limit misinterpretation due to host metabolites and to standardize biological interpretation. Such improvements will aid both basic biological investigations and clinical efforts to understand important pathogens.

6.1 Synopsis

Utilizing metabolomics data to constrain a reconstruction is a common approach in prokaryotes (Medlock et al. 2018), but no analytic tools are available to perform this analysis and it has not been implemented for any eukaryotic parasites. Here, we constrain our *Plasmodium falciparum* GENRE (iPfal18) with publically available metabolomics data; this new analysis codifies experimental data to offer a reproducible approach for data integration. Moreover, we discuss the collection and analysis of metabolomics data and the unique caveats of metabolomics in intracellular parasites.

6.2 Metabolomics as model constraints

6.2.1 Background

Genome-scale metabolic models summarize the metabolic *capability* of an organism. Thus, they may overpredict metabolic functions (*i.e.* metabolic functions that are condition-specific) and under-predicting can be used to focus curation. If a model cannot perform a function, it must be added. Accordingly, metabolomics data is well suited to curate these models as metabolomics measures the biochemical compounds in a particular sample.

6.2.2 Methods

Metabolomics data analysis: Metabolomics data were collected by Lewis, Baska, and Llinas (*unpublished*, but publically available), and stored on PlasmoDB (Aurrecoechea et al. 2009). Metabolites were measured from *in vitro* blood-stage culture supernatant or cell lysate from three experimental groups at three pHs. Experimental groups included enriched infected red blood cells containing parasites, parasites isolated from host cells, and uninfected red blood cells. To elaborate, blood-stage *Plasmodium faliparum* parasites (strain not provided) were grown in human red blood cells and RPMI media. Blood culture of *Plasmodium faliparum* is maintained with fewer than 3% infected red blood cells, leaving 97% (or more) of the cells uninfected; thus, infected cells can be enriched to increase the ratio of parasitized cells to uninfected cells. For the enriched infected red blood cell experimental group here, infected cells were enriched using Percoll, leveraging the density different between uninfected and infected cells. To isolate parasites from host cells, infected red blood cells were specifically lysed with saponin. See **Chapter 6.3** for detailed metabolomics protocols for reference. Next, enriched infected red blood cells, isolated parasites, or uninfected red blood cells were incubated with media containing U-13 C glutamine at three different pHs (6.4, 7.4, and 8.4). Samples were pelleted to measure the extracellular metabolome from the resultant supernatant. All samples were analyzed using mass spectrometry. When a compound had multiple isotopomers, all isotopomers were summed to provide the compound abundance.

For our analysis, we obtained metabolite supernatant abundances from PlasmoDB (Aurrecoechea et al. 2009). We calculated a mean metabolite abundance for the uninfected samples, at all pHs and then calculated a shared mean abundance for the enriched infected red blood cells and isolated parasite samples at all pHs. From these mean abundances, we identified metabolites that had a greater than 2 fold change in either group. If the fold change was greater than 2 when comparing uninfected cells to the parasitized groups, these metabolites were consumed by parasites. If the fold change was greater than 2 when comparing parasitized groups to uninfected cells, these metabolites were produced by parasites. This approach is extremely conservative as it requires the metabolite trends to be conserved across pHs and to be conserved in all parasitized samples (enriched infected cells and purified parasites). These lists were then subsetted such that only metabolites contained in the BiGG reference database (King et al. 2016) were included. Our analytic code is provided on my Github page, github.com/maureencarey.

Automated metabolomics-driven curation: Automated curation of iPfal18 was performed by gapfilling for metabolites identified to be produced or consumed by the parasite. First, excretion or import of these metabolites were added to the

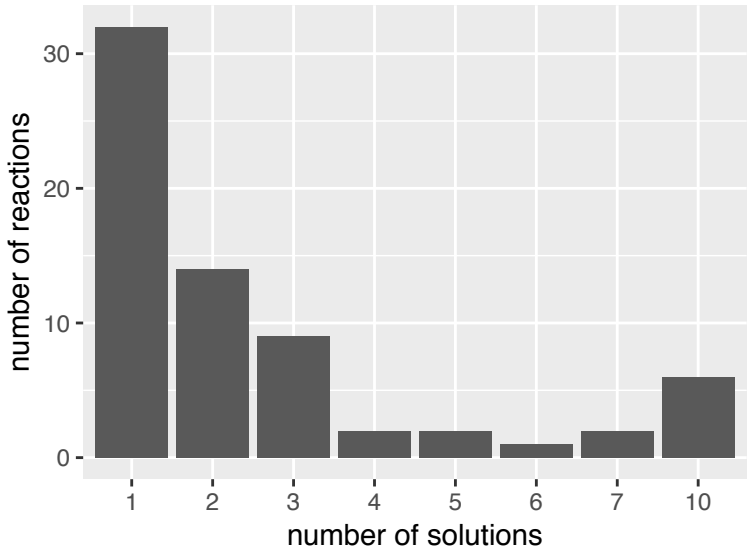


Figure 6.3: Metabolomics-based curation of iPfal17. Curation yields a subset of reproducible reaction additions.

reconstruction. Next, the model objective was changed to each of these excretion or import reactions sequentially and was gapfilled ten times per objective; this ensures synthesis or consumption of each of these measured metabolites. Reaction added due to gapfilling were given confidence scores. See **gapfilling** section of **Chapters 3** and **7.2.1** for more information on this analytic tool and confidence scores.

6.2.3 Results and Discussion

Three metabolites (dTTP, GTP, and 5-methylthioadenosine) less abundant in the supernatant in infected samples, indicating either the parasite consumes or the parasite induces host consumption of these compounds. The metabolic network contained these metabolites but none of their import reactions; thus, these three metabolic functions were added to iPfal18.

Forty-five metabolites were more abundant in the supernatant of infected samples; thus, these metabolites were produced by the parasite or by the infected host. Of these, only 29 were contained in the database used for model construction. Five of these were not yet in the reconstruction and were added. Six were contained in the original reconstruction but could not be exported. The remaining metabolites were found in the network, and could be exported.

For the six metabolites that could not yet be exported, we gapfilled for these

function. Solutions included between one and 7 new reactions, including a total of 68 total reactions and 35 metabolites. In sum, 127 reactions and 68 metabolites were added upon metabolomics-driven curation of the iPfal18 network.

Here, we define a gapfilling query as a particular task (*e.g.* exporting metabolite X) and a gapfilling solution as one iteration of a query. For each gapfilling query, we obtained 10 solutions, with 1-10 unique solutions; if multiple reactions were needed for one solution, there tended to be more unique solutions (left side of **Figure 6.3**). Six reactions were added with high confidence because they were required for every solution for a particular query. This confidence score will inform future curation efforts.

The minority (28%) of reactions added were transport reactions, and only 17% of high-confidence reactions were transport reactions. Thus, to replicate the *in vitro* results, the *in silico* parasite needed increased metabolic capacity, indicating the network did not support production or consumption of the metabolites. Additionally, the reconstruction lacked the reactions to move the synthesized metabolites to or from the media (*i.e.* extracellular environment), highlighting the underannotation of transporters and export pathways, as previously discussed in the literature (Kenthirapalan et al. 2016; Martin et al. 2005).

6.2.4 Conclusions

Here, we used metabolomics data to curate our model of *Plasmodium falciparum* metabolism, adding network functionality and improving model predictions. However, these results are biased by methodological challenges associated with metabolomics. The data discussed here were collected by measuring metabolite abundance in media following culture of *Plasmodium falciparum*-infected red blood cells (using multiple experimental conditions) or uninfected red blood cells. Thus, the measured metabolites might have been derived from the host cell, if metabolite abundance is dependent on parasite infection. Moreover, it is plausible that the metabolites measured in infected cultures were generated by host cells and only released during host cell lysis, as the parasite ruptures its host during the replication process. Improved metabolomics methods could disentangle the individual behavior of both host and parasite, but current approaches are unable to distinguish between these possible interpretations. The following section explores analytic choices affecting metabolomic interpretation. We hypothesized that analytic controls could be used to classify metabolome into host and parasite derived features.

6.3 Intracellular pathogen metabolomics methodology

6.3.1 Background

Malaria continues to be responsible for hundreds of thousands of deaths annually, most of which result from infection with the protozoan parasite *Plasmodium falciparum*. Characterization of the biology of this important pathogen can lead to improved treatment strategies. Omics approaches, such as genomics, transcriptomics, and proteomics, are widely used, but the limited annotation of the parasite's genome makes these data sets challenging to interpret. One way to alleviate this lack of functional knowledge is to use network-based modeling to contextualize noisy or sparse data and facilitate the interpretation of complex data (Carey, Papin, and Guler 2017). Additionally, the measurement of direct mediators of the phenotype, such as signaling and biosynthetic metabolites, can improve the ability to characterize phenotypes mediated by proteins that are not yet annotated in the genome. For this reason, metabolomics is becoming increasingly popular in studies of intraerythrocytic stages of *P. falciparum* (Allman et al. 2016; Babbitt et al. 2012; Creek et al. 2016; Olszewski et al. 2009; Park et al. 2015; Parvazi et al. 2016; Sana et al. 2013; Sengupta et al. 2016; Teng et al. 2014; Siddiqui et al. 2017). These studies have improved our understanding of malaria pathogenesis (Park et al. 2015), strain-specific phenotypes (Teng et al. 2014), and host-parasite interactions (Sana et al. 2013). Recent studies have successfully identified metabolic signatures that correlate well with biological function, such as time- and dose-dependent responses to antimalarial treatment (Allman et al. 2016; Creek et al. 2016) and resistance-conferring mutations (Siddiqui et al. 2017).

Previous studies on *P. falciparum* have been confined to the larger, late-intraerythrocyte-stage parasites. This is mainly due to the characteristics of the available purification approaches used; for example, magnetic purification specifically enriches late-stage parasites that contain paramagnetic hemozoin while excluding early ring stages and uninfected host cells (Paul et al. 1981). Accordingly, the study of the smaller, early-ring stage parasite is more challenging due to an inability to isolate adequate amounts of parasite material from host material (Siddiqui et al. 2017). However, specific functionality (*i.e.*, artemisinin resistance) can be observed only in the early parasite stages and metabolic details would greatly advance our understanding of such phenotypes.

There are distinct challenges that need to be considered in performing metabolomic studies in obligate intracellular pathogens such as *P. falciparum*; chief among these are acquiring adequate material and the potential for contamination from host cells. Due to inefficient purification methods, samples typically have few parasites and yet

abundant host erythrocyte material. Uninfected host cells are often 10 times more prevalent than *P. falciparum*-infected host cells in laboratory culture and clinical infections, and the host erythrocyte contains up to 10-fold more cellular material (Langreth et al. 1978; Canham and Burton 1968).

In this study, we sought to define critical parameters that can be used to overcome these challenges and facilitate the collection of high-quality metabolomics data. We chose to investigate an extreme case, namely, metabolically perturbing early-ring-stage *P. falciparum* parasites, to determine if the extensive extraparasite contamination present after employment of commonly used isolation methods can be removed analytically. We show that both the choice of analytic parameters (in particular, the normalization approach) and extraparasite contamination heavily influence the interpretation of metabolic changes. However, even appropriate normalization fails to remove environmental noise completely. Contamination from the media and host cells is as influential on the metabolome as sample treatment. Thus, we propose that the combination of improved purification and improved analytic parameters could generate more-accurate measures of the metabolome, increasing the utility of untargeted metabolomics to investigate intracellular parasite biology.

6.3.2 Methods

Table 6.1: Parasite sample reference table. Parasite samples were quantified by protein, DNA, parasite number, parasitemia, and stage distribution.

Sample	Clone	Blood batch	Antimalarial	Protein (mg)	DNA (mg/mL)	Parasitemia (%)	Parasite no.	Stage
BAT-A (+)	MRA-1240	1	700nM DHA	91.0	0.927	1.1	3270000	97% early rings
BAT-A (-)	MRA-1240	1	None	67.1	0.587	1.1	3270000	97% early rings
BAT-B (+)	MRA-1240	1	700nM DHA	121.2	0.476	1.0	2890000	98% early rings
BAT-B (-)	MRA-1240	1	None	118.1	1.216	1.0	2890000	98% early rings
BAT-C (+)	MRA-1240	2	700nM DHA	119.2	0.985	1.0	2930000	98% early rings
BAT-C (-)	MRA-1240	2	None	87.9	1.739	1.0	2930000	98% early rings

Table 6.1: Parasite sample reference table. Parasite samples were quantified by protein, DNA, parasite number, parasitemia, and stage distribution. *(continued)*

Sample	Clone	Blood batch	Antimalarial	Protein (mg)	DNA (mg/mL)	Parasitemia (%)	Parasite no.	Stage
BAT-D (+)	MRA-1240	2	700nM DHA	98.3	0.656	1.9	6950000	98% early rings
BAT-D (-)	MRA-1240	2	None	130.1	0.557	1.9	6950000	98% early rings
BAT-E (+)	MRA-1240	3	700nM DHA	125.9	1.326	2.2	6510000	93% early rings
BAT-E (-)	MRA-1240	3	None	128.5	2.083	2.2	6510000	93% early rings
PUR-A (+)	MRA-1238	1	700nM DHA	120.2	0.325	0.6	1310000	96% early rings
PUR-A (-)	MRA-1238	1	None	125.4	0.547	0.6	1310000	96% early rings
PUR-B (+)	MRA-1238	1	700nM DHA	123.3	0.259	1.0	2500000	98% early rings
PUR-B (-)	MRA-1238	1	None	121.0	0.673	1.0	2500000	98% early rings
PUR-C (+)	MRA-1238	2	700nM DHA	104.6	0.648	0.6	2260000	97% early rings
PUR-C (-)	MRA-1238	2	None	100.4	0.543	0.6	2260000	97% early rings
PUR-D (+)	MRA-1238	3	700nM DHA	120.7	0.599	1.0	3320000	96% early rings
PUR-D (-)	MRA-1238	3	None	110.3	1.563	1.0	3320000	96% early rings
PUR-E (+)	MRA-1238	3	700nM DHA	128.5	0.869	1.0	2630000	96% early rings
PUR-E (-)	MRA-1238	3	None	114.5	1.118	1.0	2630000	96% early rings

Parasite cultivation. Laboratory-adapted *P. falciparum* clonal lines were cultured in RPMI 1640 (Thermo Fisher Scientific, Waltham, MA) containing HEPES (Sigma-Aldrich, St. Louis, MO) supplemented with 0.5% AlbuMAX II lipid-rich bovine serum albumin (Sigma-Aldrich, St. Louis, MO) and 50 mg/liter hypoxanthine (Thermo

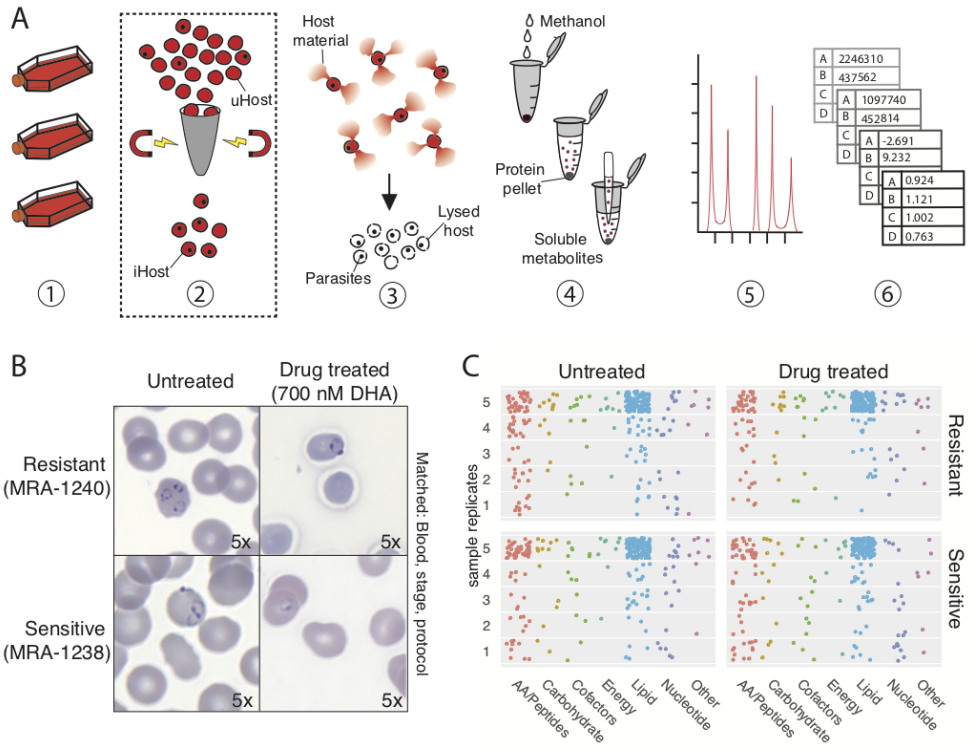


Figure 6.4: Metabolomics pipeline and metabolite identification. A: Metabolomics purification and analysis pipeline. Step 1: Laboratory-adapted *Plasmodium falciparum* clones are cultured in host erythrocytes. Parasite count is collected at this step (total erythrocyte number multiplied by percent parasitemia yields total parasite value; see Materials and Methods). Step 2: If enriching for late-stage parasites is desired, cultures are passed through a magnetic column to retain paramagnetic late-stage-infected erythrocytes. Note that this was not done for the present study. iHost, infected host; uHost, uninfected host. Step 3: Host erythrocytes are lysed using saponin; parasites remain intact. Samples are washed to remove hemoglobin and other intracellular host material and quenched on liquid nitrogen. Total protein is quantified here (prior to freezing). Step 4: Soluble metabolites are extracted from precipitated protein using methanol and centrifugation. Double-stranded DNA is quantified here. Step 5: Metabolites are separated via liquid chromatography and identified using mass spectroscopy. Metabolite spectra are compared to a library of authenticated standard metabolites for high-confidence identification. Step 6: Abundance data for each metabolite are normalized to an appropriate parameter (*i.e.*, DNA content or parasite number), log transformed, centered with respect to the median, and scaled with respect to variances, prior to employing statistical comparisons. B: Experimental comparison. All samples were grown in RPMI media supplemented with AlbuMAX and hypoxanthine and with one of three blood batches (matched across treatment conditions). At the early ring stage (3 h postinvasion), 10 samples were treated with dihydroartemisinin (DHA; 700 nM) for 6 h and 10 samples were matched with respect to protocol and condition (blood batch, medium batch, and stage) without drug treatment. Images shown were taken at the 6h time point (100x magnification); dormancy was observed at 24 h. C: Summary of identified metabolites. Metabolites (each represented by one point) from various metabolic subgroups were not uniformly detected in all five replicates for any sample group. How frequently a metabolite was measured across replicates is indicated by the metabolite point placed in data corresponding to 1 to 5 replicates (y axis). The majority of metabolites detected were lipid species, as indicated by the large number of blue dots. A full list of identified metabolites is provided in the supplemental material.

Fisher Scientific, Waltham, MA). Parasite cultures were maintained at 3% hematocrit and diluted with human red blood cells (blood batch noted in **Table 6.1**) to maintain parasitemia at between 1% and 3%, with changes of culture medium every other day (**Figure 6.4A; step 1**). Cultures were incubated at 37 degrees C with 5% oxygen, 5% carbon dioxide, and 90% nitrogen (Trager and Jensen 1976). Some samples were treated with artemisinin, an antimalarial with metabolic effects (dihydroartemisinin; see antimalarial treatment details in **Table 6.1**) (Allman et al. 2016; Creek et al. 2016). Cultures were tested for *Mycoplasma* monthly using a LookOut Mycoplasma PCR detection kit (Sigma-Aldrich); none tested positive.

Parasite isolation. Two distinct laboratory-adapted clinical isolates of *P. falciparum* (BEI Resources, NIAID, NIH; *Plasmodium falciparum* strains IPC 5202/MRA-1240 and IPC 4884/MRA-1238, contributed by Didier Menard) containing mixed stages with >50% rings were synchronized using 5% sorbitol (Sigma-Aldrich, St. Louis, MO) (Lambros and Vanderberg 1979). The resultant early-stage cultures were incubated at 37 degrees C in AlbuMAX media to allow the development of a predominantly schizont population. After the late-stage population was confirmed using microscopy, cultures were checked every 1 to 2h for the development of newly invaded ring-stage parasites. If the parasites were treated with dihydroartemisinin, the treatment was performed at this stage. Fourteen 25cm3 flasks containing early ring-stage parasites (less than 3h postinvasion, treated with dihydroartemisinin or left untreated) were subsequently lysed from the erythrocyte membrane using 0.15% saponin, as previously described (Moll et al. 2008) (**Figure 6.4A; step 3**). Prior to lysis, a sampling of parasite material was taken for determination of erythrocyte count (hemocytometer) and parasitemia (Sybr green-based flow cytometry (Bei et al. 2010)), which contributed to parasite number determination (total number of erythrocytes x percent parasitemia yields the total parasite count). Additional samples were obtained following erythrocyte lysis for protein quantification using Bradford reagent (Sigma-Aldrich, St. Louis, MO). A series of three wash steps were then performed using 1x phosphate-buffered saline (PBS) (Sigma-Aldrich, St. Louis, MO) and centrifugation at 2,000g to remove soluble erythrocyte metabolites. Purified material was kept on ice until it was flash frozen using liquid nitrogen (to quench metabolism), followed by storage at -80 degrees C until sent for analysis. This procedure was performed five times for both parasite clonal lines (strains IPC 5202/MRA-1240 and IPC 4884/MRA-1238) to provide 10 drug-treated replicates for metabolomic analysis. Additionally, matched parasites (same parasite lineage, medium type, stage, blood batches, and purification methods) were also grown without drug treatment (**Table 6.1**) to generate 10 additional control samples (see comparison in **Figure 6.4B**).

Metabolite preparation, analysis, and identification. Metabolites were identified using ultra-high-performance liquid chromatography coupled with tandem

mass spectroscopy (UPLC/MS-MS) by Metabolon, Inc. (Durham, NC). All sample preparations and metabolite identifications were performed according to standard protocols of Metabolon, Inc. (briefly summarized here). Double-stranded DNA was quantified in all samples using a Quant-it PicoGreen dsDNA assay kit (Thermo Fisher, Waltham, MA) according to the manufacturer's instructions. Proteins were precipitated using methanol for 2 min with vigorous shaking and then centrifuged for extraction (**Figure 6.4A; step 4**). Sample extracts were separated into aliquots, dried, and suspended in appropriate standard-containing solvents for analysis by four methods. These four methods facilitate the measurement of metabolites with different biochemical properties and include two reverse-phase UPLC/MS-MS methods, one with positive ion electrospray ionization (ESI) optimized for hydrophilic compounds and one optimized for hydrophobic compounds, and a third method with negative-ion-mode ESI. Additionally, a UPLC/MS-MS method with negative-ion-mode ESI following elution from a hydrophilic interaction chromatography column was used. Waters Acquity ultraperformance liquid chromatography and a Thermo Scientific Q Exactive high-resolution/accurate mass spectrometer were used for all metabolite detection procedures (**Figure 6.4A; step 5**).

To evaluate the quality of the mass spectrometry pipeline, several controls were used. Ultrapure water or the solvent alone or both were used as blank samples to control for nonspecific signals in the pipeline. Technical controls were employed to ensure that the instruments were working within specifications; a pooled sample of human plasma and a pooled aliquot of experimental samples were used to distinguish biological from technical variability. A set of recovery and internal standards were also used to quantify variability and instrument performance. Variability scores for all runs included in this analysis met the acceptance criteria specified by Metabolon, Inc.

Raw data were extracted using hardware and software developed by Metabolon, Inc. Metabolites were quantified using the area under the curve and were identified by comparison to a library of several thousands of preexisting entries of purified standards or recurrent unknown compounds. Each library standard was uniquely authenticated by retention time/indices, mass-to-charge ratios, and chromatographic data. Named metabolites corresponded to library standards or were predicted with confidence according to the standard protocols specified by Metabolon, Inc.

DNA quantification. Measurement of host-derived dsDNA levels was performed by incubating uninfected erythrocytes at 3% hematocrit for 48 h in PBS or RPMI 1640 alone or RPMI 1640 with 50 mg/liter hypoxanthine or RPMI with 50 mg/liter hypoxanthine and 0.5% AlbuMAX II lipid-rich BSA. Erythrocytes were subjected to saponin lysing and washed prior to dsDNA quantification using a Quant-it PicoGreen dsDNA assay kit as described above.

Microscopy. Laboratory-adapted *P. falciparum* clones (BEI Resources, NIAID, NIH; *Plasmodium falciparum*, patient line strain E/MRA-1000 or strain IPC 5202/MRA-1238, contributed by Didier Menard) at 1.5% parasitemia with >60% rings were lysed using 0.15% saponin, as previously described (Moll et al. 2008). Samples were washed twice using 1x PBS (Sigma-Aldrich, St. Louis, MO) and centrifugation at 2,000xg for 5 min. For bright-field images, parasites were fixed with methanol and stained with Giemsa stain for 15 min. Images were obtained on a Nikon Eclipse Ci microscope (100x) using an Imaging Source microscope camera and Nikon NIS Elements imaging software. Representative images are shown. For production of fluorescent images, samples were stained on slides with either DAPI (4,6-diamidino-2-phenylindole) (Sigma-Aldrich, St. Louis, MO) at 1:20,000 or CD235a-phycoerythrin (CD235a-PE) antibody (Thermo Fisher Scientific, Waltham, MA) at 1:100. Fluorescent images were acquired using an Evos FL cell imaging system (Thermo Fisher Scientific, Waltham, MA). Representative images are shown, and quantification of 1,214 parasites associated with erythrocyte membranes was performed for 11 preparations.

Data preprocessing and statistical analysis. Following the analytic protocol outlined in reference Xia and Wishart (2011), we first preprocessed metabolite abundances for each sample by imputing missing values corresponding to half of the lowest detectable metabolite abundance. Next, we normalized metabolite abundances by sample features **Figure 6.7**, followed by normalization using metabolite features with log transformation, centering, and scaling (**Figure 6.4A, step 6**) (Sugimoto et al. 2012).

Specifically, to limit intersample variability, metabolite abundances for each replicate were normalized to the sample value for double-stranded DNA, protein, or parasite number. To limit intermetabolite variability, metabolite abundances were log transformed, centered with respect to the median (Evans et al. 2009), and scaled by standard deviation (**Figure 6.4A; step 6**).

The resultant processed metabolite abundances were used for calculation of univariate and multivariate statistics, as well as for classification. All analyses were conducted using R with tidyverse (Wickham 2017), knitr (Xie 2014), reshape2 (Wickham 2007, 2012), pracma (Borchers 2015), grid and gridExtra (Auguie 2016), extrafont (Chang 2014), and RSvgDevice (Luciani, Decorde, and Lise 2009) for data wrangling and visualization and vegan (Wagner 2015) and base R (Team 2015) for analysis. Analyses of variance (ANOVAs) were used to compare group means for determinations of differential abundances, and P values were adjusted using the false-discovery rate (Benjamini and Hochberg) (Benjamini and Hochberg 1995) to correct for multiple testing. The significance cutoff was 0.05. PERMANOVAs were used to compare population separation data. Correlations were conducted using a two-sided Pear-

son's product moment correlation with false-discovery rate (Benjamini and Hochberg) in R. See the supplemental material for code and detailed analysis available at github.com/gulermalaria/metabolomics.

Random forest analysis. Random forest analysis is a machine learning technique and was used here to classify sample groups. Within a random forest classifier, individual trees are built from subsets of the data and internally validated with respect to the remaining data set. With this approach, variables (metabolites) are ranked by their effect on classifier accuracy, as measured by a change in performance following removal of the variable. Classifiers were built with each data normalization method to predict drug treatment or blood batch. These analyses were conducted in R using the RandomForest package (Liaw, Wiener, and Others 2002) and base R (Team 2015). See the supplemental material for code and detailed analysis available at github.com/gulermalaria/metabolomics.

6.3.3 Results

Metabolomics. We conducted metabolomics on early-ring-stage (0h to 3h) *Plasmodium falciparum* parasites lysed from host erythrocytes. Two parasite clones were grown in matched conditions, lysed and washed from the host cell, and analyzed via ultra-high-performance liquid chromatography coupled with mass spectrometry (UPLC/MS) (**Figure 6.4A**). Prior to isolation, each clone (representing either a drug-sensitive or a drug-resistant line) was either left untreated or treated with 700 nM dihydroartemisinin (for 6h), generating four sample groups with matched blood batches, media, and purification approaches (**Figure 6.4B**). Dihydroartemisinin, the active component of the antimalarial artemisinin, is a known metabolic disruptor (Allman et al. 2016; Creek et al. 2016; Cobbold et al. 2016). Both sensitive and resistant parasites are known to enter a unique metabolic state, called dormancy, following treatment. Dormancy is characterized by reduced metabolic activity (Chen et al. 2014; Cheng, Kyle, and Gatton 2012; Peatey et al. 2015); thus, treated ring-stage parasites should have a metabolome distinct from that seen with untreated parasites.

Mass spectrometry analysis of these samples detected 297 identifiable metabolites; 155 metabolites were detected in every sample. Samples contained between 182 and 267 metabolites. The detected metabolites represented 10 energy-associated metabolites, 159 lipid species, 108 peptides and amino acids, 40 nucleotides, 28 cofactors, 20 carbohydrates, and 10 in other categories (see raw data available at github.com/gulermalaria/metabolomics). Lipid species were the most consistently detected metabolites in every sample (as measured by the percentage of metabolite found in every sample), and amino acids were often unique to individual samples

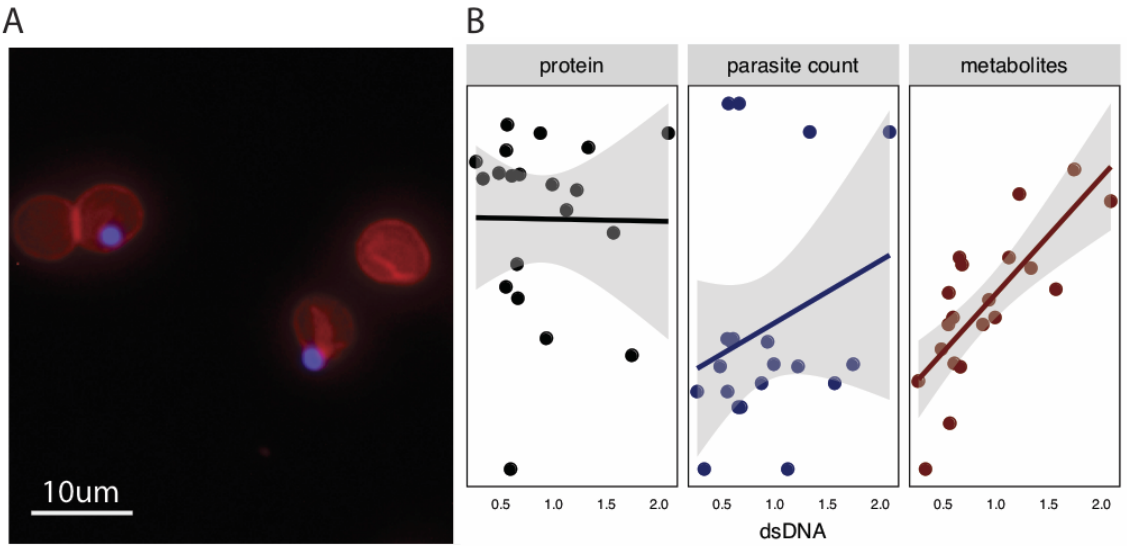


Figure 6.5: Host persistence is detected using multiple approaches. A: Visualization of parasites within erythrocyte ghosts. Fluorescent imaging (40x magnification) reveals parasites (blue, DAPI) retained within erythrocyte ghosts (red, phycoerythrin-conjugated CD235a antibody) following saponin treatment. Approximately 70% of the parasites remain associated with host membranes. B: Sample characteristics. Samples were evaluated for levels of double-stranded DNA (dsDNA; quantified in micrograms per milliliter on the x axis), protein amounts (black; quantified in micrograms on the y axis, ranging from 67.0641 to 130.0936ug, in the left panel), and parasite counts (blue; quantified on the y axis, ranging from 1,306,500 to 6,946,875 parasites, in the center panel) prior to analysis. The total number of metabolites detected per sample (red; quantified on the y axis, ranging from 182 to 267 metabolites, in the right panel) was significantly correlated with sample dsDNA quantification ($P = 0.000098$; $r2 = 0.76$). Protein amount and parasite count were not significantly correlated with dsDNA. The fit line uses a linear model, and the shaded region represents the standard error.

(**Figure 6.4C**). Several metabolites were measured that are not known to be part of *P. falciparum* metabolism, including kynurenine (detected in 25% of samples), phenol red (phenolsulfonphthalein; detected in 95% of samples), and HEPES (detected in all samples; see raw data available at github.com/gulermalaria/metabolomics).

Table 6.2: Host entrapment of purified parasites. Parasites remain in host cells following purification. Laboratory adapt *(continued)*

Image preparation date	Free Parasites	Total Parasites Counted

Table 6.2: Host entrapment of purified parasites. Parasites remain in host cells following purification. Laboratory adapted *Plasmodium falciparum* clones (BEI Resources, NIAID, NIH: *Plasmodium falciparum*, Strain Patient line E/MRA-1000 or IPC 4884/MRA-1238, contributed by Didier Ménard) at >60centrifugation at 2000xg for 5 minutes. Samples were then stained on slides with either DAPI at 1:20,000 (Sigma Aldrich, St Louis, MO) and CD235a-PE antibody at 1:100 (Thermo Fisher Scientific, Waltham, MA) for fluorescence microscopy. Fluorescent images were acquired using the EVOS FL Cell Imaging System (Thermo Fisher Scientific, Waltham, MA). Parasites not associated with erythrocyte membranes were counted.

Image preparation date	Free Parasites	Total Parasites Counted
8.22.17	3	21
9.22.17	34	77
9.25.17	20	58
9.25.17	36	78
9.27.17	10	49
9.27.17	23	58
12.15.17	41	242
12.16.17	37	153
1.9.18	60	131
1.11.18	65	204
1.12.18	50	143
TOTAL	379	1214

Host contamination. Despite implementation of the current best practices, including erythrocyte lysis and washing steps to remove parasites from their intracellular milieu (**Figure 6.4A**; see, *e.g.*, references Allman et al. (2016) and Parvazi et al. (2016)), parasite separation from the host is poor. Microscopy confirmed that the parasites lysed from host cells remained embedded in erythrocyte membranes and that washes failed to isolate parasite material (**Figure 6.5A**) (Langreth 1977). Importantly, over 68% of parasites remained associated with the host membrane (**Table 6.2**). This result emphasized that erythrocyte "ghosts" (cell membranes with associated metabolites) remained abundant in the sample and could have heavily contributed to the metabolome. Thus, we sought analytic approaches to remove host contamination post hoc.

Normalization. We first explored the use of normalization with three distinct approaches. Metabolomics preprocessing methods can influence results (Ejigu et al. 2013; Kohl et al. 2012), but the role of normalization, particularly in intracellular pathogens, has not been extensively explored. Both host- and parasite-derived metrics (double-stranded DNA or dsDNA, protein, and parasite levels) were eval-

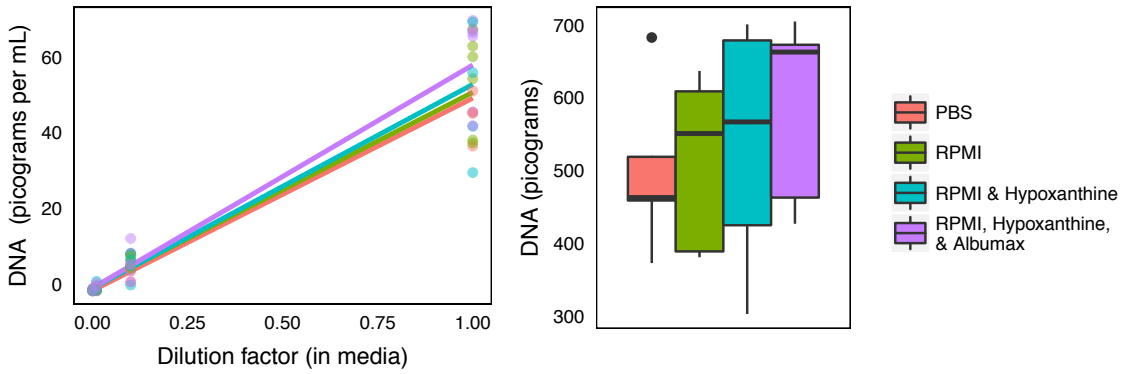


Figure 6.6: DNA contribution from host erythrocyte and media. Erythrocytes contribute DNA despite being anucleated. Measurement of host-derived dsDNA levels was performed by incubating uninfected erythrocytes at 3% hematocrit for 48 h in PBS or RPMI 1640 alone or RPMI 1640 with 50 mg/liter hypoxanthine or RPMI with 50 mg/liter hypoxanthine and 0.5% AlbuMAX II lipid-rich BSA. Erythrocytes were saponin lysed and washed twice with PBS prior to dsDNA quantification using a Quant-it PicoGreen dsDNA assay kit as described in Materials and Methods. Nondetectable values (below the limit of detection) were imputed as 0. Data in the left panel demonstrate that DNA abundance is concentration-dependent and does not represent mere instrument noise. Data in the right panel demonstrate that components of media (such as AlbuMAX II lipid-rich BSA) contribute to DNA quantification but that erythrocytes in PBS contribute the majority of the measured DNA.

uated in the experimental setup (**Figure 6.4A**). Sample replicates contained 1.3 to 6.9 million parasites (**Table 6.1**). As expected, no two normalization metrics were correlated across samples (**Figure 6.5B**; see code for the full analysis available at github.com/gulermalaria/metabolomics). Metabolite yield (as measured by the number of identified metabolites) was correlated only with DNA abundance ($p = 0.000098$, $r2=0.76$) (**Figure 6.5B**), indicating that DNA abundance is associated best with total biomass.

Initially, we anticipated that dsDNA should come primarily from the parasite fraction, as host erythrocytes are anucleated and growth medium does not contain any intact DNA; however, we found that host cells and AlbuMAX (a medium component) did contribute to sample dsDNA (**Figure 6.6**). Protein was likely also derived from all three culture components, namely, parasite, host erythrocyte, and media (via AlbuMAX supplementation). Although parasite counts represent a direct measure of the parasite fraction, this variable was collected several steps upstream of metabolome quantification (**Figure 6.4A**) and may have been suboptimal compared to metrics collected later in the pipeline.

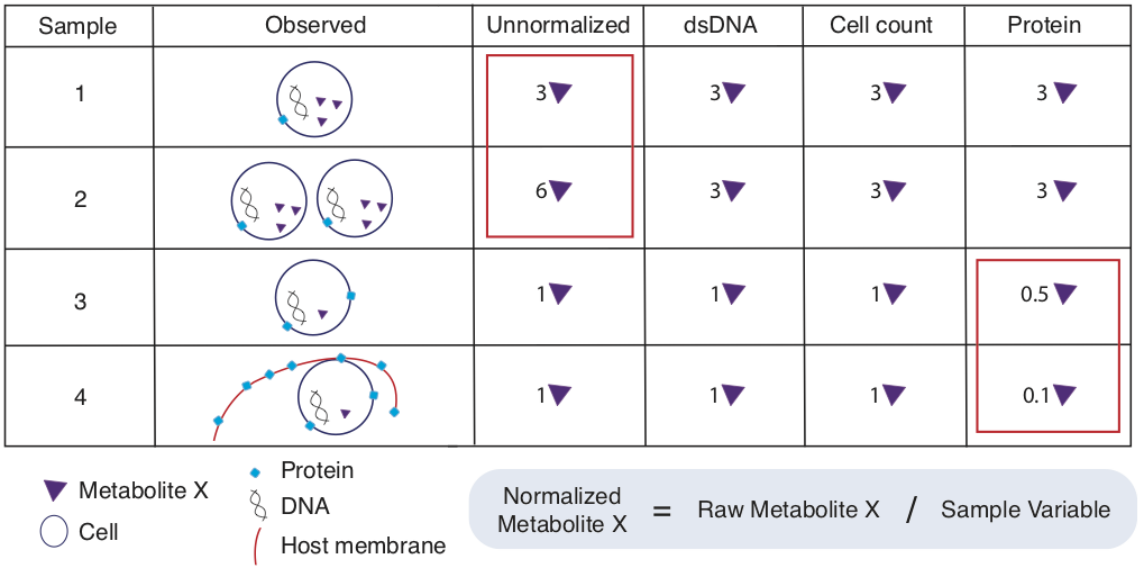


Figure 6.7: Normalization approaches impact the final metabolite abundance. Normalization controls for sample-to-sample variation were performed. Normalization requires sample metabolite abundance to be divided by the quantified normalization factor, the sample variable (the equation is in the blue box; normalization factors are shown above the box). The examples of results shown in the table indicate abundances of X metabolites given several different sample metrics for normalization. For example, identical samples with difference cell counts (sample 1 and sample 2) reveal the importance of normal- ization; without it, the data corresponding to the identical samples show a 2-fold difference in the values determined for metabolite X. The values determined for identical parasite samples 3 and 4 also show a nearly 2-fold difference in metabolite abundance after normalizing to protein levels, due to host bias for protein measures.

We normalized metabolomes with respect to these parasite-derived and host-derived metrics to determine if normalization reduces extraparasite noise to reveal parasite metabolomes. Normalization of metabolite levels can be calculated by a variety of methods (**Table 6.3; Figure 6.7**), all aiming to enhance interpretation of results by controlling for technical or nonbiological variation. To normalize, we divide the value representing the abundance of each metabolite in a sample by the corresponding sample variable to control for sample-to-sample variation (**Figure 6.7**). As illustrated in **Figure 6.7**, normalization can significantly affect interpretation of results and should be selected carefully based on experimental design and knowledge of samples.

Table 6.3: Parameters in metabolomics analysis of intracellular parasites, including *Plasmodium*. Most parameters are dependent on experimental design. *(continued)*

Parameter	Options	Factors to consider

Table 6.3: Parameters in metabolomics analysis of intracellular parasites, including *Plasmodium*. Most parameters are dependent on experimental design.

Parameter	Options	Factors to consider
Growth conditions	Ring stage	Limited biomass (12µm, Figs. 5.4A and B), haploid genome
		Few enrichment options
	Late stage	Larger in size (310µm), polyploid genome
		Can use magnetic enrichment (Fig. 5.4A)
	Mixed stages	Consider effects of stage variation on data
	Media batches	Relevant if using serumbased media formulations
	Blood batches	Must be recorded and matched within comparisons (Table S3)
		Useful to assess host contamination levels (Fig. 5.9 & 5.10)
Additional controls	Uninfected erythrocytes	Use to identify or control for host metabolites
		Used in addition to normalization
Enrichment methods	Saponin, other lytic reagents	Compatible with all stages (Fig. 5.4A)
		Parasites remain in erythrocyte ghosts (Fig. 5.5A)
		(Need improved methods that isolate parasite from host cell)
	Magnetic purification	Increases parasite to host ratio (Fig. 5.4A)
Metabolite Detection	NMR	Limited metabolite detection but higher confidence
	Mass Spectrometry	Industry standard for broad detection
	Radio labeling	Targeted approach with high confidence
	Single metabolite assays	Highconfidence, targeted approach with low throughput
Preanalysis normalization	Cell number normalization	Can be combined with any postanalysis normalization but requires sample manipulation
Postanalysis normalization	Parasite derived parameters	i.e. parasite number
		Selection requires knowledge of experimental design
	Parameters with mixed derivation (host, parasite)	i.e. protein, DNA; Can fail to remove undesired noise (Fig. 5.5 & 5.8)
	Internal standards	Dependent on metabolomics facilities
Centering	Mean	Standard centering
	Median	Less sensitive to outliers
	Other	See van den Berg, et. al, 2006 for summary of alternative approaches
	Within group SD	Requires no additional samples

Table 6.3: Parameters in metabolomics analysis of intracellular parasites, including *Plasmodium*. Most parameters are dependent on experimental design. *(continued)*

Parameter	Options	Factors to consider
Scaling	Zscoring	Requires control samples (i.e. untreated or uninfected erythrocytes)
Statistical analysis	Univariate	Requires multiple comparison corrections
	Multivariate	Reveals group differences based on multiple variables
	Machine learning (e.g. Random Forest)	Classification is more stringent than univariate tests, but can identify nonlinear effects

Because the effect of normalization has not been explored in intracellular parasites, we normalized to parasite number (parasite derived), dsDNA amount (parasite, medium, and host derived), and total protein amount (parasite, medium, and host derived) and then performed principal component analysis with all sample metabolomes (**Figure 6.8A-D**). The normalization methods all yield distinct principal component structures, and yet none clearly separate the four sample groups (as measured by clustering of the sample groups by permutational multivariate analysis of variance [PERMANOVA]; p values are provided in the figure under the *Normalization* heading). However, with DNA normalization, we are able to separate drug-treated parasites from untreated parasites or clonal groups (**Figure 6.8B**); with parasite number normalization, we can distinguish clonal groups (**Figure 6.8D**).

Consistent with the lack of distinct separation, univariate statistical analysis revealed no metabolites that were differentially abundant among the four groups (see code for the full analysis available at github.com/gulermalaria/metabolomics). When normalization is employed, metabolome differences between groups are highly dependent on the approach; the top differentially abundant metabolites are normalization method-dependent (*data not shown*; see code for the full analysis available at github.com/gulermalaria/metabolomics). These findings emphasize that biological interpretations can change significantly depending on the chosen analytic parameters and thus that the selected normalization metric is a critical parameter and must be shared for analytic reproducibility.

Data filtering. We next examined and removed extraparasite metabolites in our data set in order to explore the effect of sample contamination. Because there are no unique metabolites associated with the host, we explored medium-specific metabolites, specifically, phenol red and HEPES. Both phenol red (a pH indicator) and HEPES (a buffer) are components of the growth medium and should not be utilized by cells. These metabolites are routinely excluded from metabolomics analysis for this reason.

Interestingly, the abundances of 82 (of a total of 298) metabolites were correlated

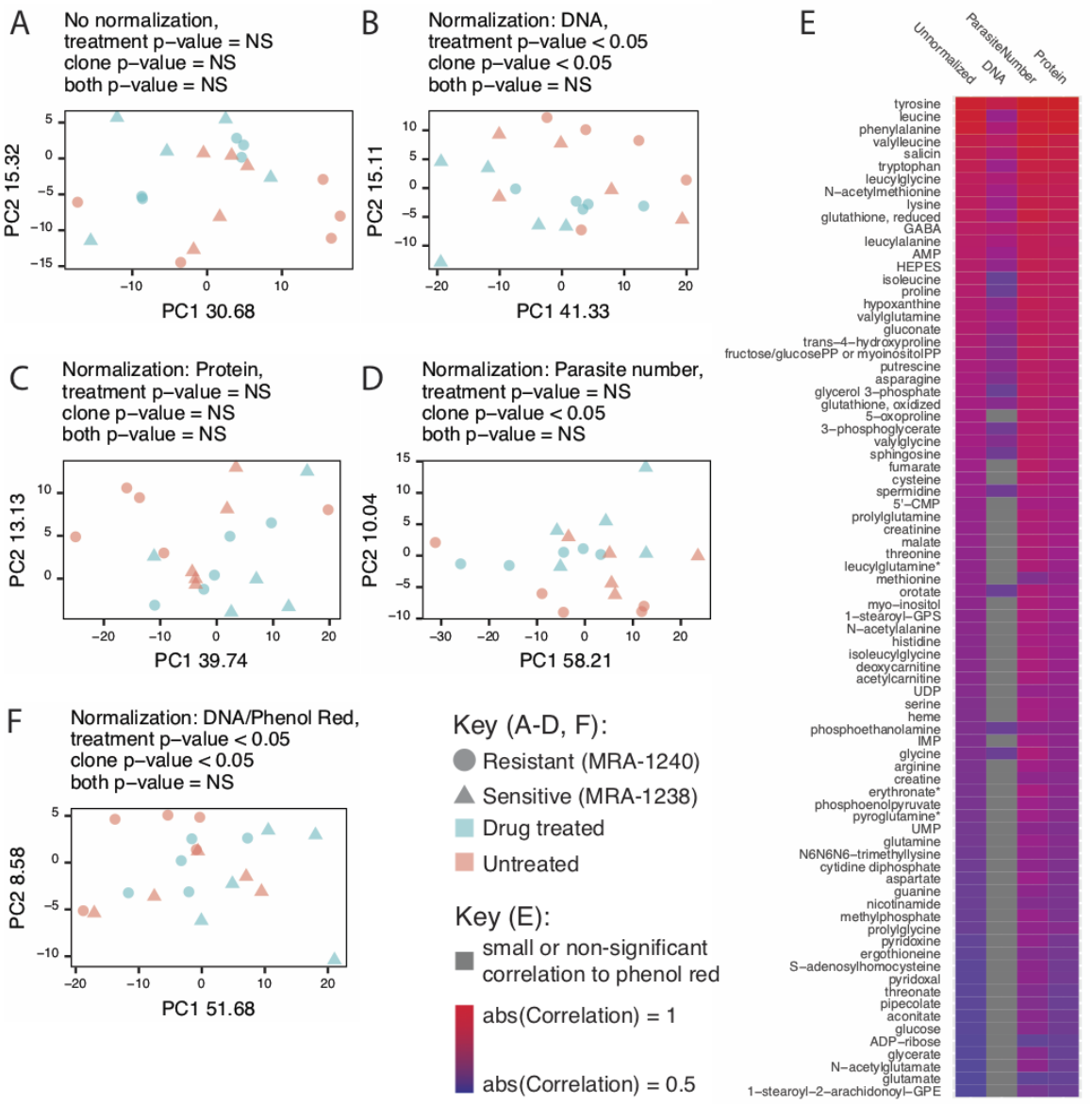


Figure 6.8: Metabolomes are dependent on the normalization approach and are influenced by extraparasite contamination. A-D: Normalization affects metabolome similarity. Principal-component (PC) analysis was performed prior to normalization (A) as well as after using three different normalization methods (B: DNA normalization, C: total protein normalization, and D: parasite count normalization) on all identified metabolites. PERMANOVA significance is listed for each grouping. E: Metabolites associated with components of media. The raw abundance of 82 metabolites was correlated with phenol red levels (unnormalized column), using a two-sided Pearson's product moment correlation with Benjamini and Hochberg false-discovery rate correction. These associations were not removed with parasite number and protein normalization. DNA normalization best removes associations with components of media (increases in numbers of insignificant correlations, in grey); only 39% of correlations remain. F: Removal of medium-associated metabolites. Principal-component analysis (PCA) of DNA-normalized samples with phenol red-correlated metabolites removed from the data set yielded no improvement in sample clustering.

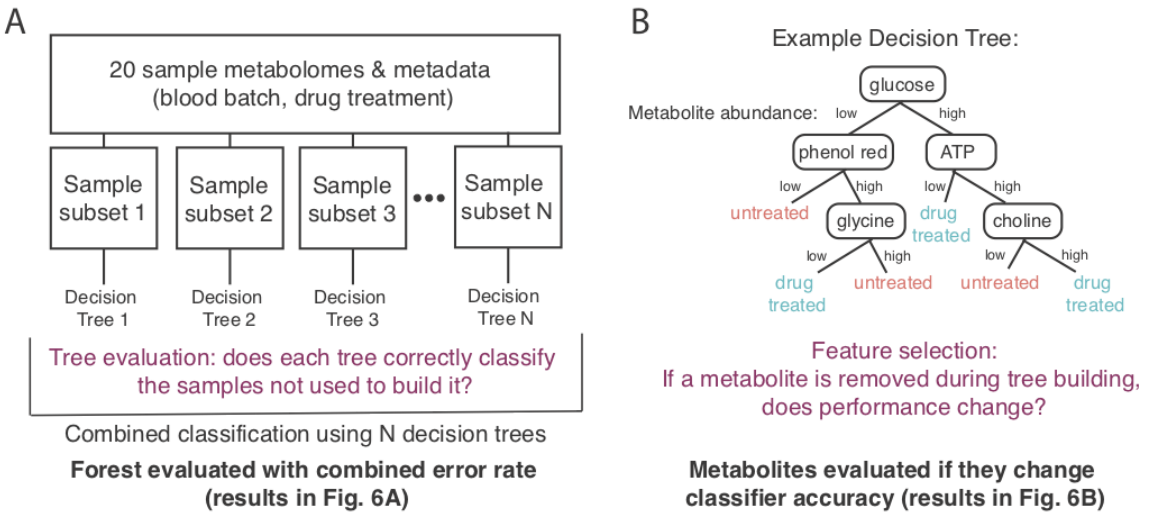


Figure 6.9: Random forest analysis. A: Building a random forest classifier. Samples are randomly classified into subsets (training and test data sets); from the training subsets, decision trees are built to separate samples into groups (see panel B). Trees are evaluated by testing classification performance on the remaining samples from the test data sets. See Materials and Methods for more details on the analyses. B: Evaluating metabolite importance. Metabolite importance is calculated by determining the effect of removal of the metabolite from the data set on classifier performance. See Methods for further details.

with phenol red (**Figure 6.8E**) and the abundances of 76 metabolites were correlated with HEPES (*data not shown*); the abundances of 59 metabolites were correlated with both compounds. Many (39%) of these metabolites remained correlated with the components of the media even after normalization (phenol red data are shown in **Figure 6.8E**).

Because phenol red and HEPES appeared to increase in abundance in drug-treated samples (nonsignificant trend; *data not shown*), we argue that this extraparasitic fraction may influence the interpretation of drug treatment data. If we remove these medium-associated metabolites from our analysis, surprisingly, sample separation into the four treatment groups does not provide an improvement in comparison to the results seen with DNA normalization alone (based on the remaining 216 metabolites; see **Figure 6.8F** compared to **Figure 6.8B**). Thus, both *post hoc* data filtering methods were insufficient to remove the effect of extraparasite contamination in our low-powered study.

Machine learning. We next used machine learning to attempt to separate the extraparasite-associated metabolome from the parasite metabolome. Here, we leveraged the multiple blood batches used in parasite culture (**Figure 6.4A**). Our four

sample groups were grown in three different blood batches (**Table 6.1**). Univariate statistical analysis revealed only one metabolite with differential abundance results among the blood batches (1-arachidonoyl-GPE; see code for the full analysis available at github.com/gulermalaria/metabolomics). To further explore the host contribution to the metabolome, we built random forest classifiers for analysis of blood batch and drug treatment data (**Figure 6.9A**). Random forest analysis is an internally validated machine learning approach, used here to classify samples into groups based on their metabolome (**Figure 6.9A**) and to identify individual variables that are important for prediction accuracy (**Figure 6.9B**).

We first built classifiers for analysis of blood batch data across all samples. Ninety-five metabolites (of 298) improved classifier accuracy in analysis of blood batch data (using the DNA normalized data set; see code for the full analysis available at github.com/gulermalaria/metabolomics). Many of these metabolites are correlated in abundance with the components of the media explored (**Figure 6.8E**), including CDP-ethanolamine, AMP, ADP-ribose, and aspartate, which are among the top 10 most influential metabolites in this classifier. The remaining metabolites (203 in total) had no effect on the performance of the classifier or worsened its predictive ability, indicating they are not associated with blood batch due to high variability or association with other features that differentiate samples. The classifier built from DNA-normalized metabolomes predicted blood batch data with a 30% error rate (**Figure 6.10A**). We also built a blood batch classifier from each of the other normalization approaches (**Figure 6.10A**).

To determine if blood batch is as influential on the metabolome data as a potent antimalarial drug treatment, we built similar classifiers for analysis of artemisinin treatment. Parasites were classified into two treatment conditions with a 30% class error rate using DNA-normalized metabolomes (**Figure 6.10A**). A total of 118 metabolites (of 298) improved the accuracy of this classification, including medium-correlated metabolites such as pipecolate, several dipeptides, and phenol red (see code for the full analysis available at github.com/gulermalaria/metabolomics).

The performance of our classifier (**Figure 6.10A**) was relatively poor due to the small sample size, and the results indicated that only a subset of the measured metabolomes was predictive of blood batch or drug treatment. Classifiers built from data under alternative normalization approaches were comparable in performance, but different metabolites contributed to their accuracy (**Figure 6.10B**). Removal of phenol red and associated metabolites from the data set (listed as phenol red correction data; **Figure 6.10A**) reduced blood batch classifier performance more than it did treatment classifier performance; this result is not surprising, because both the components of the media and the host cells are extraparasitic. Thus, by

A Combined classifier error rate

	Blood	Treatment
Unnormalized	30%	20%
DNA	30%	30%
Parasite Number	40%	25%
Protein	50%	20%
DNA & Phenol Red	45%	25%

B Metabolite effect on Classifier Accuracy

Metabolite	Blood	Treatment
CDP-ethanolamine	↑↑↑↑↑	↓↓↓↓-
valyl leucine	↓↓↓-	↑↑↑↑
succinate	↓↓↓↑-	↓↓↑↑-
hypoxanthine	↓↓--	↓↓↑-
phenol red	↑↑↑↑	↑↑↑↑

Figure 6.10: Blood batch and antimalarial treatment influence metabolomes. A: Classifier performance. Classifiers were built to predict blood batch or treatment conditions using the metabolomics data with or without 4 normalization approaches. The classifier error rate varies with the normalization approach. B: The normalization method determines the important metabolites. A sample consisting of five metabolites associated with improved or worsened classifier accuracy is shown. These metabolites are shown in accordance with their importance in classifier performance and their interesting behavior across classifiers. Upward-pointing arrows indicate that the metabolite improves classifier accuracy in one classifier, and downward-pointing arrows indicate they worsen accuracy in one classifier (arrows represent the normalization approaches from panel A); if the metabolite does not improve or worsen accuracy, a dash is shown. Contradictory results (both upward-pointing and downward-pointing arrows for one metabolite) indicate that the normalization method changes the importance of the metabolite. Note that valyl leucine, hypoxanthine, and phenol red were removed upon phenol red filtering and, therefore, are present in only 4 classifiers, as indicated by the four arrows and dashes.

removing medium contamination, we may also be removing host contamination and data associated with the blood batch. However, phenol red is associated with both blood batch classifier accuracy and treatment classifier accuracy (**Figure 6.10B**); this result supports the idea of the necessity of removing extraparasitic metabolites during sample preparation, as they can skew meaningful biological interpretation.

Interestingly, when the classifier was built using a different normalization approach, the set of metabolites that most extensively contributed to accuracy changed (representative examples are shown in **Figure 6.10B**; code for the full analysis is available at github.com/gulermalaria/metabolomics). Although some metabolites (such as CDP-ethanolamine or valyl leucine) were consistently associated with blood classifier accuracy or treatment classifier accuracy, respectively, some metabolites (such as succinate and hypoxanthine) gave contradictory results depending on the data normalization approach (**Figure 6.10B**). 1-Arachidonoyl-GPE, identified by univariate statistics, was not among the top most predictive metabolites in any classifier but did contribute to accuracy in some blood batch classifiers. Thus, sample metabolome can classify both blood batch and sample group, indicating that sample treatment and blood batch influence the metabolome and that this is normalization approach-dependent.

6.3.4 Discussion

The lifestyle of intracellular parasites presents challenges to implementing traditional metabolomics protocols, predominately due to host metabolite contamination and limitations in the amounts of parasite material. These challenges are exacerbated when studying early parasite stages (such as the *Plasmodium* ring stage studied here), when the parasite is smallest. In our study, we conducted a detailed assessment of the impact of extraparasite contamination and investigated analytic approaches to improve metabolome interpretation. We recommend improved discussion of normalization methods in the metabolomics field, especially for intracellular parasites, as normalization significantly effects the interpretation of a data set. Additionally, we propose several analytic approaches to explore the effect of host contamination.

Metabolome interpretation is normalization approach-dependent. Normalization limits nonbiological variation and is absolutely essential for biological interpretation (**Figure 6.7**). Normalization factors can be calculated using a variety of methods, and normalization is implemented either before or after metabolite quantification and identification (described as preanalysis or postanalysis) (**Table 6.3**) (Ejigu et al. 2013; Kohl et al. 2012). Often, preanalysis normalization is conducted by isolating the same number of cells for analysis (Hoffmann, Seidl, and Dugas 2002) but this

is not typically used in the study of *P. falciparum* as generating adequate biomass can be challenging (Allman et al. 2016; Creek et al. 2016; Siddiqui et al. 2017). Furthermore, the use of inaccurate quantification methods may negate the utility of this step by introducing more variability. Postanalysis normalization methods are also routinely used; these include the use of internal standards (Babbitt et al. 2012; Ejigu et al. 2013), corrections for protein amounts (often used for supernatant or cell-free metabolomics (Silva, Cordeiro-da-Silva, and Coombs 2011)), DNA content (an approach validated in mammalian cells (Silva et al. 2013) and applied to bacterial cells (Medlock et al. 2018)), or cell number (typically used for bacterial populations (Wu and Li 2016)).

To our knowledge, normalization was never described in detail in previous metabolomics studies of *P. falciparum*, perhaps due to the technical challenges that we explored here. We evaluated three postanalysis normalization approaches, namely, the protein, double-stranded DNA, and parasite number approaches (**Figure 6.7** and **Figure 6.8A-D**). Overall, we conclude that normalization significantly affects the interpretation of results (**Figure 6.8A and 6.10**). The normalization approach influences the metabolites with the greatest differential abundances (data not shown because they did not reach significance) and the metabolites predictive of sample group shift with data normalization (**Figure 6.10**).

In the present studies, only the parasite count data were entirely parasite derived. The extracellular environment (including components of media and host erythrocyte) likely contributes heavily to protein abundance. Accordingly, parasite count and protein abundance are not correlated. We also show that the host cell contributes to dsDNA levels, despite lacking a nucleus (**Figure 6.6**). This material may be contributed by the small proportion of dying white blood cells that remain after erythrocyte preparation. Despite this finding, our analysis shows that dsDNA normalization of early-ring-stage metabolomes best distinguished sample and treatment groups and removed medium contamination (**Figure 6.8**). Much variability still remained after this step; we did not identify any differentially abundant metabolites even though artemisinin has been reported to have metabolic effects on late-stage parasites (Allman et al. 2016; Creek et al. 2016; Cobbold et al. 2016) and dormancy induces metabolic shifts in ring-stage parasites (Chen et al. 2014; Cheng, Kyle, and Gatton 2012; Peatey et al. 2015). Although dsDNA normalization was the most effective approach in our data set, it is not appropriate for all experimental cases; for example, this type of analysis would introduce variability in comparisons of groups of different parasite stages due to known genome copy number differences (Janse et al. 1986; Le Roch et al. 2004).

Media and host contribute to the measured metabolome. We found that extra-

parasite material contributed by host erythrocytes and components of media can also heavily impact the metabolome. Many studies employ erythrocyte lysis prior to sample purification (Allman et al. 2016; Parvazi et al. 2016; Sana et al. 2013; Siddiqui et al. 2017). However, several results from our study show that this step does not eliminate the potential for host contamination.

First, lipid species were the major class of metabolites detected in our analysis (**Figure 6.4C**), perhaps due to the abundance of the erythrocyte membranes or "ghosts" present in the preparations (**Figure 6.5A**). Second, more than a quarter of the metabolome is correlated with the components of the media (phenol red, **Figure 6.8E**, and HEPES, *data not shown*). Unlike HEPES (Teng et al. 2014), phenol red has not been shown to be imported into the parasite; neither metabolite is produced or biochemically consumed by the parasite. Thus, it is likely that these medium-derived metabolites remained associated with cells following *in vitro* culture in medium. This medium also contains high levels of other metabolites such as glutathione, hypoxanthine, glutamine, and many amino acids, which are correlated with phenol red and/or HEPES abundances. Third, we measured metabolites not expected to be produced or consumed by *Plasmodium* (Carey, Papin, and Guler 2017). For example, kynurenine is present in erythrocytes, derived from the amino acid L-tryptophan (Y. Wang et al. 2010; Hartai et al. 2005), and is not known to be involved in *P. falciparum* metabolism (Ginsburg 2006). Lastly, the only differentially abundant metabolite in our entire analysis that reached significance was associated with the blood batch (1-arachidonoyl-GPE). This metabolite has not been studied in the context of erythrocyte or *Plasmodium* metabolism but can be explored as a potential marker of host contamination.

In fact, we were able to predict a set of metabolites that are most likely to be influenced or derived from the host erythrocyte by identifying the metabolites that are most predictive of blood batch (**Figure 6.9B** and **6.10B**; see figures in code available at github.com/gulermalaria/metabolomics for a comprehensive list). Going forward, it may be possible to use specific metabolite markers to assess levels of host contamination and parasite sample purity and to control for host contamination during analysis.

Future recommendations. Parasite metabolomics is a rapidly expanding field; thus, well-documented methodologies and rigorous evaluation criteria will enhance data reproducibility and the quality of metabolomics-derived observations. In this study, we compiled evidence of host erythrocyte and medium contamination in untargeted metabolomics studies of intracellular parasites and explored the analytic decisions that influence metabolome interpretation. We showed that analytic approaches can improve the accuracy and interpretability of intracellular parasite metabolomes but

that, ultimately, better methods are needed to extract biological differences from samples.

A common approach used in the study of *P. falciparum* involves the use of an uninfected erythrocyte control to adjust for the presence of host metabolites (Babbitt et al. 2012; Olszewski et al. 2009; Park et al. 2015; Sana et al. 2013; Sengupta et al. 2016; Teng et al. 2014), but even with the use of this control, interpretation of data remains challenging (see, *e.g.*, (Olszewski et al. 2010)). Uninfected erythrocyte controls are used for z score metabolite abundance calculations (infected relative to uninfected), for differential abundance calculations (infected divided by uninfected), or for calculations involving subtraction of "host" metabolite data from infected-population data. However, we hypothesize that, in some cases, the use of the uninfected erythrocyte control alone is not sufficient; as we show in **Figure 6.8F** and **6.10**, correcting the data set by removing extraparasite contamination data (medium-associated metabolites) fails to improve treatment classification. We suggest that the quantitative analytic methods applied here must also be used to evaluate the efficacy of the uninfected erythrocyte control.

Another common analytic step involves the removal of extraparasitic metabolites, such as phenol red, as they are considered to represent noise from culture media. However, these metabolites contain valuable information about experimental variation and could be used for quality control, as indicated by the frequent correlation between phenol red abundance and other metabolites (**Figure 6.8E**). For this reason, these metabolites should not be excluded from the data set and subsequent analysis.

We suggest a set of considerations and recommendations for enhancing the accuracy of parasite metabolomics (**Table 6.3** and below). First, samples must be better purified away from host material. Purification could involve enrichment methods to increase parasitemia prior to lysis (reducing the ratio of uninfected host cells to parasites) or the direct removal of host material postlysis. Currently, enrichment approaches exist only for late-stage malaria parasites. Second, markers of host contamination must be used to evaluate the level of medium and host contamination. The number of metabolites with abundances correlated with phenol red or HEPES can be used to assess the contribution of the media. The visual detection of ghost material (via microscopy) combined with assessment of host-specific metabolite markers is an effective option to assess sample purity. Additionally, analytic approaches (such as blood batch classification) can be used to identify remaining or experiment-specific markers of contamination. Finally, data must be normalized to appropriate measurements to maximize the metabolome signal associated with the treatment of interest; subsequent subtraction of metabolites associated with host or media (*e.g.*, uninfected erythrocyte control or known components of media) can further reduce

metabolite influence mediated by extraparasite conditions. Importantly, we propose that, similarly to studies in Leishmania (Rojo et al. 2015; Westrop et al. 2015; Akpunarlieva et al. 2017), normalization and discussion of the chosen normalization metrics should become standard during metabolomics analysis of intraerythrocytic parasites. With these considerations, metabolomics has the potential to become a powerful tool in the study of intracellular parasites.

6.4 Conclusions

Metabolomics data offers a new profiling method that is agnostic to genome annotation or many experimental challenges, an ideal approach for eukaryotic pathogens; however, best practices for the analyses of these data has not been widely discussed. Parasite metabolomics is a rapidly expanding field and, thus, well-documented methodology and rigorous evaluation criteria will enhance data reproducibility and the quality of metabolomics-derived observations. In this chapter, we compiled evidence of host cell and media contamination in untargeted metabolomics studies of intracellular parasites and explore analytic decisions that influence metabolome interpretation. We showed that analytic approaches can improve the accuracy and interpretability of intracellular parasite metabolomes but, ultimately, better methods are needed to extract biological differences from samples.

Chapter 7: Becoming a computational biologist

Some of the following text, figures, and tables has been adapted from Carey, Maureen A., and Jason A. Papin. "Ten simple rules for biologists learning to program." *PLoS Computational Biology* 14.1 (2018): e1005871, which has now been viewed over 41,000 times with 29 percent of views leading to download.

7 Becoming a computational biologist

This section contains 'side projects' essential to the completion of this work.

7.1 Learning to program

I began my PhD with no programming experience, and quickly learned that would need to change. I remember a spring afternoon when a labmate helped me write a short script to analyze and reorganize an expression dataset and I thought, 'this is magic. *literally* magic.' From that moment on, I decided that I could still become a wizard even if I had missed the boat on Hogwarts. I read tutorials and muddled through to learn basic R and MATLAB and to implement analyses, and took a course in programming ('Computing as a Research tool' with Ed Hall, highly recommended). Most importantly, I devoted many hours of work and frustration to the cause. I now think my data wrangling and analytics (or, the application of programming) are my strongest skills. One visit home, my parents asked me what I wished I had known before beginning this journey - inspiring this article for PLOS Computational Biology's Ten Simple Rules series (Carey and Papin 2018).

7.1.1 Ten Simple Rules for biologists learning to program

7.1.1.1 Introduction

As big data and multi-omics analyses are becoming mainstream, computational proficiency and literacy are essential skills in a biologist's tool kit. All "omics" studies require computational biology: the implementation of analyses requires programming skills, while experimental design and interpretation require a solid understanding of the analytical approach. While academic cores, commercial services, and collaborations can aid in the implementation of analyses, the computational literacy required to design and interpret omics studies cannot be replaced or supplemented. However, many biologists are only trained in experimental techniques. We write these 10 simple rules for traditionally trained biologists, particularly graduate students interested in acquiring a computational skill set.

7.1.1.2 Rule 1: Begin with the end in mind

When picking your first language, focus on your goal. Do you want to become a programmer? Do you want to design bioinformatic tools? Do you want to implement

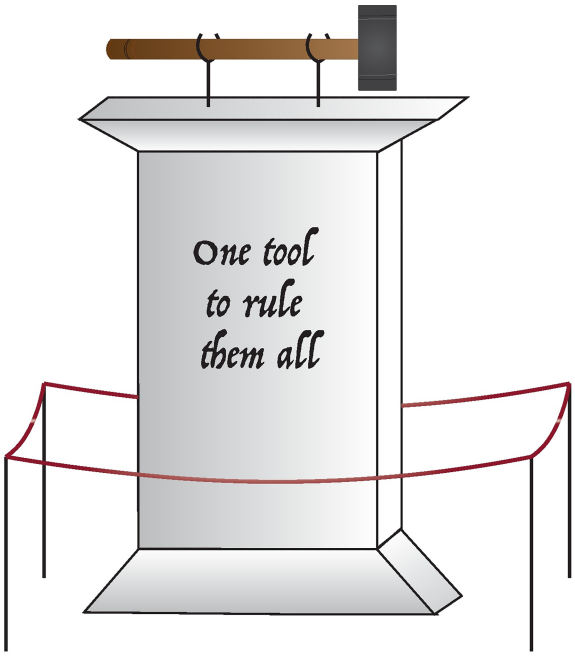


Figure 7.1: The one tool to rule them all (or: how programming languages do not work).

tools? Do you want to just get these data analyzed already? Pick an approach and language that fits your long- and short-term goals.

Languages vary in intent and usage. Each language and package was created to solve a particular problem, so there is no universal "best" language (**Figure 7.1**). Pick the right tool for the job by choosing a language that is well suited for the biological questions you want to ask. If many people in your field use a language, it likely works well for the types of problems you will encounter. If people in your field use a variety of languages, you have options. To evaluate ease of use, consider how much community support a language has and how many resources that community has created, such as prevalence of user development, package support (documentation and tutorials), and the language's "presence" on help pages. Practically, languages vary in cost for academic and commercial use. Free languages are more amenable to open source work (*i.e.*, sharing your analyses or packages). See **Table 7.2** for a brief discussion of several programming languages, their key features, and where to learn more.

7.1.1.3 Rule 2: Baby steps are steps

Once you've begun, focus on one task at a time and apply your critical thinking and problem solving skills. This requires breaking a problem down into steps. Analyzing

Language	Key features	Documentation	Sample tutorials	Community groups
Bash	• Most common Unix shell • Practical for execution of scripts written in all other languages • Versatile • Easy to delete files or make other drastic changes • Weaknesses include executing math and limited data structures • Default for macOS and most Linux distributions	• gnu.org/software/bash/manual/ • On macOS's terminal, type "man <command>" to get the manual for any command (and "q" to exit manual page)	• The Linux Documentation Project's Beginner's guide: tldp.org/LDP/Bash-Beginners-Guide/html/ • Ubuntu's documentation: help.ubuntu.com/community/Beginners/BashScripting • Azet's GitHub page: github.com/azet/community_bash_style_guide	• Google Plus: plus.google.com/communities/110832059019676429606 • GitHub community resources page: github.com/awesome-lists/awesome-bash
Python	• General purpose language • Considered easy to learn due to readability • Flexible syntax considered both a strength and weakness • Interpreted language	• docs.python.org	• Google's Python class: developers.google.com/edu/python/ • The Hitchhiker's Guide to Python: docs.python-guide.org/	• Python Users Group: wiki.python.org/moin/LocalUserGroups • Python Special Interest Groups: python.org/community/sigs/
R	• Community involvement • Application-focused development • Easy to learn by coupling basic programming and applications • Well-developed visualization • Variable package quality • "Tidy data" community • Interpreted language	• rdocumentation.org • r-project.org • cran.r-project.org	• R for cats: rforcats.net • Books by Hadley Wickham: hadley.nz • R Tutorial's introduction: r-tutor.com/r-introduction • Cyclismo's R Tutorial: cyclismo.org/tutorial/R/	• R-Ladies: rladies.org • R Users Group: many
SAS	• Statistical computing • High-quality development of statistical functions by commercial and academic developers • Domain-specific usage • Free for students only • Typically a compiled language	• support.sas.com	• Boston University's SAS Training for Statistics: bu.edu/stat/bu-student-chapter-of-the-asa/sas-training/	• SAS User Groups: sas.com/en_us/connect/user-groups.html
MATLAB	• Well-developed applications in engineering • Maintained professionally • Interpreted language • Discounted academic license	• mathworks.com/help/matlab	• Cyclismo's MATLAB Tutorial: cyclismo.org/tutorial/matlab/ • For purchase courses offered at: matlabacademy.mathworks.com	• MATLAB Central: mathworks.com/matlabcentral/
Perl	• General purpose language • Handles text well • Waning community involvement • Syntax modelled after human language • Interpreted language	• perl.org • cpan.org	• Beginning Perl: perl.org/books/beginning-perl/ • Perl maven's tutorial: perlmaven.com • Perl::Learn: learn.perl.org	• Perl Mongers: pm.org • Perl Monks: perlmonks.org
Fortran	• Numeric computation • Fast • Often used for high-performance computing • Limited development • Compiled language	• fortranwiki.org	• many at Fortran wiki: fortranwiki.org/fortran/show/Tutorials	• Fortran Friends: fortran.orpheusweb.co.uk
C/C++	• Low-level language • Powerful, used for source code of many other languages • Challenging to learn as it requires explicit syntax • Explicit syntax enforces good programming habits • Compiled language	• devdocs.io/c • cppreference.com	• C programming's tutorial: cprogramming.com/tutorial/ • Learn-C's web-based tutorial: learn-c.org	• Standard C++ Foundation: isocpp.org • C/C++ Users Group (CUG): hal9k.com/cug

https://doi.org/10.1371/journal.pcbi.1005871.t001

Figure 7.2: A noninclusive discussion of programming languages. A shell is a command line (*i.e.*, programming) interface to an operating system, like Unix operating systems. Low-level programming languages deal with a computer's hardware. The process of moving from the literal processor instructions towards human-readable applications is called 'abstraction.' Low-level languages require little abstraction. Interpreted languages are quicker to test (*e.g.*, to run a few lines of code), facilitating learning through trial and error. Interpreted languages tend to be more human readable. Compiled languages are powerful because they are often more efficient and can be used for low-level tasks. The distinction between interpreted and compiled languages is not always rigid. All languages presented below are free unless noted otherwise. The Wiki page on programming languages provides a great overview and comparison of languages.

omics data may sound challenging, but the individual steps do not: *e.g.*, read your data, decide how to interpret missing values, scale as needed, identify comparison conditions, divide to calculate fold change, calculate significance, correct for multiple testing. Break a large problem into modular tasks and implement one task at a time. Iteratively edit for efficiency, flow, and succinctness. Mistakes will happen. That's ok; what matters is that you find, correct, and learn from them.

7.1.1.4 Rule 3: Immersion is the best learning tool

Don't stitch together an analysis by switching between or among languages and/or point and click environments (Excel, Microsoft, https://www.microsoft.com/en-us/, etc.). While learning, if a job can be done in one language or environment, do it all there. For example, importing a spreadsheet of data (like you would view in Excel) is not necessarily straightforward; Excel automatically determines how to read text, but the method may differ from conventions in other programming languages. If the import process "misreads" your data (*e.g.*, blank cells are not read as blank or "NA," numbers are in quotes indicating that they are read as text, or column names are not maintained), it can be tempting to return to Excel to fix these with search-and-replace strategies. However, these problems can be fixed by correctly reading the data and by understanding the language's data structures. Just like a spoken language (Genesee 1994, 2014), immersion is the best learning tool (Campbell and Bolker 2002; Guzdial 2004). In addition to slowing the learning curve, transferring across programs induces error. See References (Zeeberg et al. 2004; Ziemann, Eren, and El-Osta 2016; Linke 2009) for additional Excel or word processing–induced errors.

Eventually, you may identify tasks that are not well suited to the language you use. At that point, it may be helpful to pick up another language in order to use the right tool for the job (see **Rule 1**). In fact, understanding one language will make it easier to learn a second. Until then, however, focus on immersion to learn.

7.1.1.5 Rule 4: Phone a friend

There are numerous online resources: tutorials, documentation, and sites intended for community Q and A (StackOverflow, StackExchange, Biostars, etc.), but nothing replaces a friend or colleague's help. Find a community of programmers, ranging from beginning to experienced users, to ask for help. You may want to look for both technical support (*i.e.*, a group centered around a language) and support regarding a particular scientific application (*e.g.*, a group centered around omics analyses). Many universities have scientific computing groups, housed in the library or information technology (IT) department; these groups can be your starting point. If your lab or

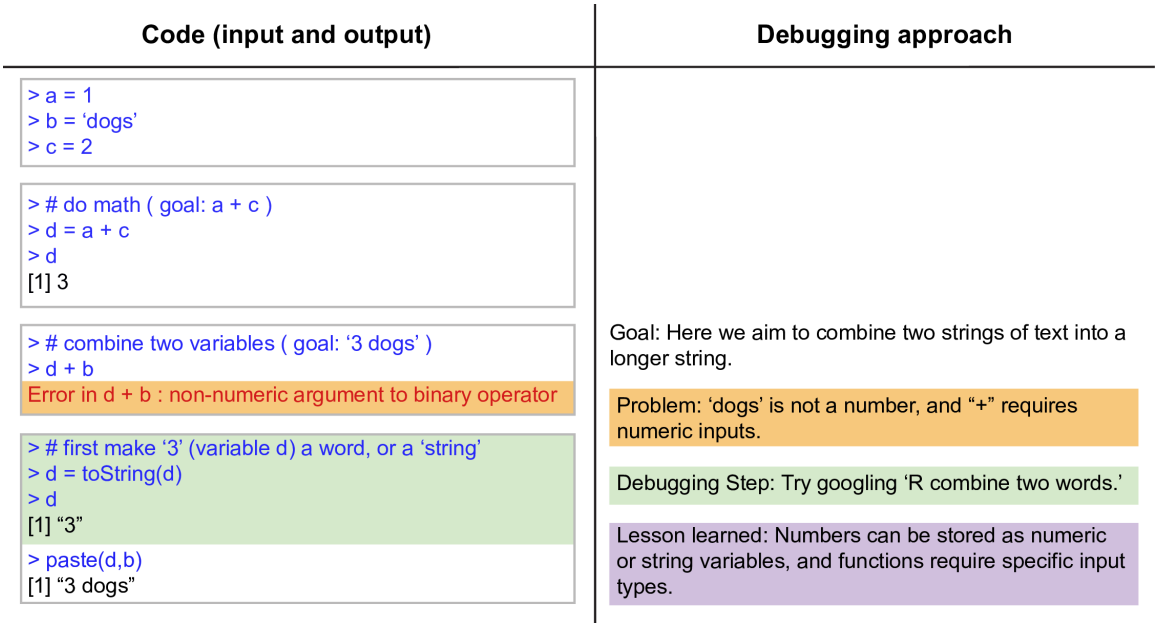


Figure 7.3: Anatomy of an error message, Part 1 (or: How to write more than one line of code). Here we show an example of the debugging process in R using the RStudio environment, with the goal of concatenating two words.

university does not have a community of programmers, seek them out virtually or locally. Coursera courses, for example, have comment boards for students to answer each other's questions and learn from their peers. Organizations like Software and Data Carpentry or language user groups have mailing lists to connect members. Many cities have events organized by language-specific user groups or interest groups focused on big data, machine learning, or data visualization. These can be found through meetup.com, Google groups, or through a user group's website; some are included in **Table 7.2**.

Once you find a community, ask for help. At the beginning stages, in-person help to deconstruct or interpret an online answer is invaluable. Additionally, ask a friend for code. You wouldn't write a paper without first reading a lot of papers or begin a new project without shadowing a few experimenters. First, read their code. Implement and interpret, trying to understand each line. Return to discuss your questions. Once you begin writing, ask for edits.

7.1.1.6 Rule 5: Learn how to ask questions

There's an answer to almost anything online, but you have to know what to ask to get help. In order to know what to ask, you have to understand the problem.

Start by interpreting an error message. Watch for generic errors and learn from them. Identify which component of your error message indicates what the issue is and which component indicates where the issue is (**Figure 7.3, Figure 7.4, Figure 7.5, Figure 7.6**). Understanding the problem is essential; this process is called "debugging." Without truly understanding the problem, any "solution" will ultimately propagate and escalate the mistake, making harder-to-interpret errors down the road. Once you understand the problem, look for answers. Looking for answers requires effective googling. Learn the vocabulary (and meta-vocabulary) of the language and its users. Once you understand the problem and have identified that there is no obvious (and publicly available) solution, ask for answers in programming communities (see **Rule 4** and **Table 7.2**). When asking, paraphrase the fundamental problem. Include error messages and enough information to reproduce the problem (include packages, versions, data or sample data, code, etc.). Present a brief summary of what was done, what was intended, how you interpret the problem, what troubleshooting steps were already taken, and whether you have searched other posts for the answer.

See the following website for suggestions: http://codereview.stackexchange.com/help/how-to-ask and (Torres 2017). End with a "thank you" and wait for the help to arrive.

7.1.1.7 Rule 6: Don't reinvent the wheel

Rule 6 can also be found in "Ten Simple Rules for the Open Development of Scientific Software" (Prlić and Procter 2012), "Ten Simple Rules for Developing Public Biological Databases" (Helmy, Crits-Christoph, and Bader 2016), "Ten Simple Rules for Cultivating Open Science and Collaborative R&D" (Masum et al. 2013), and "Ten Simple Rules To Combine Teaching and Research" (Vicens and Bourne 2009). Use all resources available to you, including online tutorials, examples in the language's documentation, published code, cool snippets of code your labmate shared, and, yes, your own work. Read widely to identify these resources. Copy-and-paste is your friend. Provide credit if appropriate (*i.e.*, comment "adapted from so-n-so's X script") or necessary (*e.g.*, read through details on software licenses). Document your scripts by commenting in notes to yourself so that you can use old code as a template for future work. These comments will help you remember what each line of code intends to do, accelerating your ability to find mistakes.

7.1.1.8 Rule 7: Develop good habits early on

Computational research is research, so use your best practices. This includes maintaining a computational lab notebook and documenting your code. A compu-

Code (input and output)

In[1]:
```
# load the 'pandas' package, which allows us to use the data
structure called 'DataFrame'
import pandas
# load dataset
df = pandas.read_csv('http://bioconnector.org/workshops/data/
gapminder.csv')
# print top 5 rows of our dataset
print df.head()
```

Out[1]:

	country	continent	year	lifeExp	pop	gdpPercap
0	Afghanistan	Asia	1952	28.801	8425333	779.445314
1	Afghanistan	Asia	1957	30.332	9240934	820.853030
2	Afghanistan	Asia	1962	31.997	10267083	853.100710
3	Afghanistan	Asia	1967	34.020	11537966	836.197138
4	Afghanistan	Asia	1972	36.088	13079460	739.981106

In[2]:
```
# let's extract only the data regarding countries in Africa
df['continent'] == 'Africa'
```

Out[2]:
```
0          False
1          False
2          False
       ...
1702       True
1703       True
```

In[3]:
```
# let's use the boolean statement to subset the dataframe
df.loc[(df['continent'] == 'Africa')]
df.head()
```

Out[3]:

	country	continent	year	lifeExp	pop	gdpPercap
24	Algeria	Africa	1952	43.077	9279525	2449.008185
25	Algeria	Africa	1957	45.685	10270856	3013.986023
26	Algeria	Africa	1962	48.303	11000948	2550.816880
27	Algeria	Africa	1967	51.407	12760499	3256.991771
28	Algeria	Africa	1972	54.518	14760787	4182.663766

Debugging approach

Goal: Here we aim to extract the data associated with countries in Africa. We want to subset the dataframe to extract these datapoints.

Hint 1: If your command can be executed, there will not be an error message. Check output to see if your results are what you aim to get.

Problem 1: We identified which datapoints we want, but did not extract the data.

Debugging step 1: Google 'python pandas selecting data,' and read documentation.

Lesson learned: View output and compare to your expected results to identify 'silent' errors.

Figure 7.4: Anatomy of an error message, Part 2 (or: Just because it works, doesn't mean it's right). Here we provide more examples of the debugging process. Examples shown in the next three figures are conducted in Python using a Jupyter notebook. Environments like RStudio (previous figure) and Jupyter notebooks are two examples of integrated development environments; these environments offer additional support, including built-in debugging tools. First, we show an error that does not induce an error message, but the user must debug nonetheless.

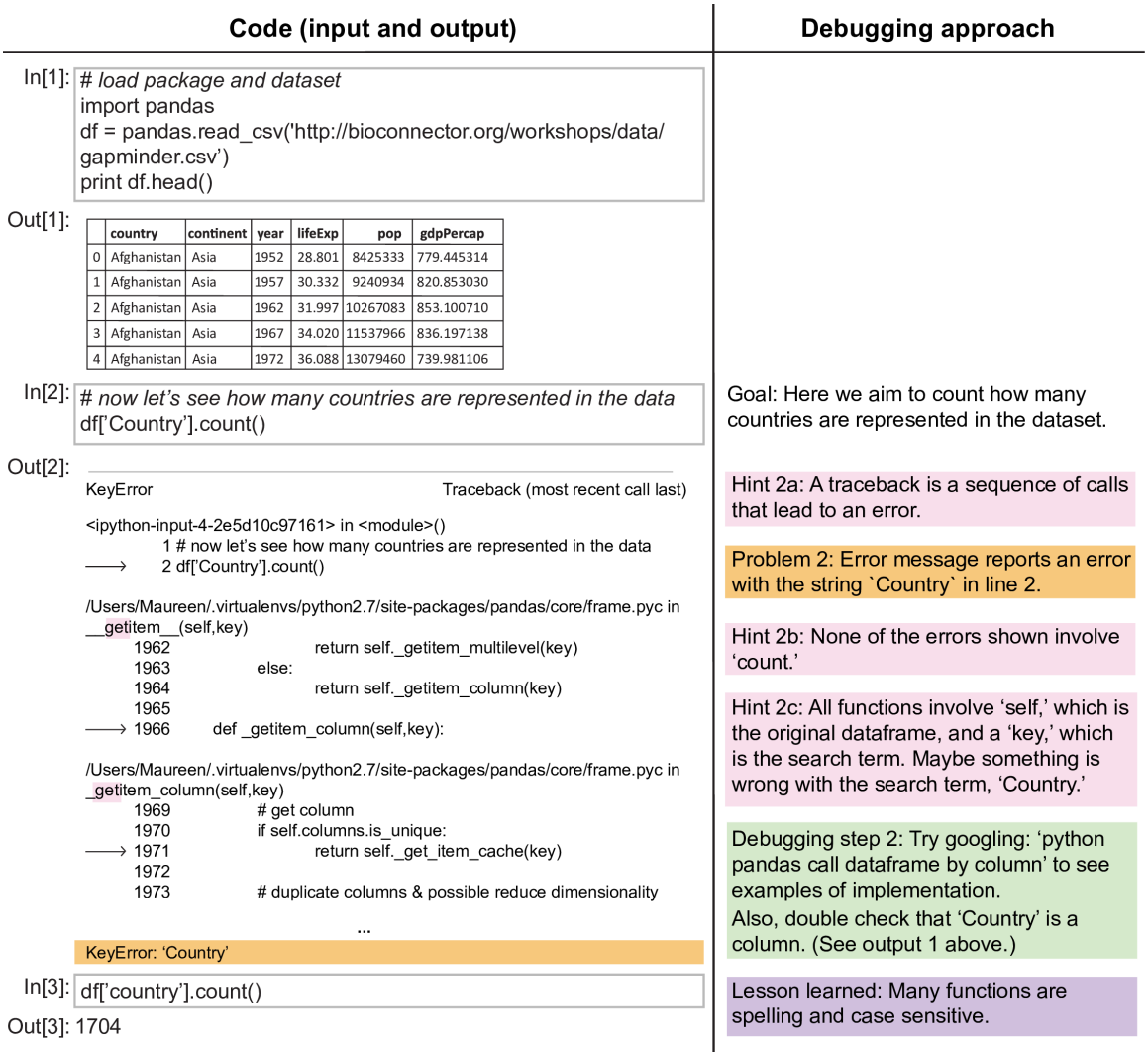


Figure 7.5: Anatomy of an error message, Part 3 (or: Trace your way back to the problem). Here we show an explicit error message.

Code (input and output)

In[1]:
```
# load package and dataset
import pandas
df = pandas.read_csv('http://bioconnector.org/workshops/data/
gapminder.csv')
```

In[2]:
```
# let's see if the mean life expectancy (lifeExp) changes by year
new_df = df.groupby('year',as_index = False)['lifeExp'].mean()
new_df.head()
```

Out[2]:

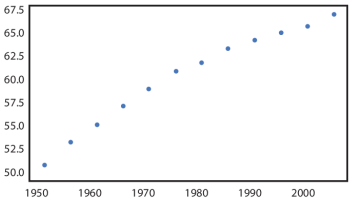
	year	lifeExp
0	1952	49.057620
1	1957	51.507401
2	1962	53.609249
3	1967	55.678290
4	1972	57.647386

In[3]:
```
# We found a solution, but what does it mean? Let's break it down.
# df.groupby: the function groupby is applied to the dataframe df
# groupby requires a grouping variable ('year')
# Next we take the mean of column 'lifeExp' of this grouped df
# but then what does the 'as_index = False' do?
# let's try changing 'False' to 'True' to see how the results change
new_df = df.groupby('year',as_index = True)['lifeExp'].mean()
new_df.head()
```

Out[3]:
```
year
1952  49.057620
1957  51.507401
1962  53.609249
1967  55.678290
1972  57.647386
Name: lifeExp, dtype: float64
```

In[4]:
```
# if we want both variables as columns, we use 'as_index = False'
plot_df = df.groupby('year',as_index = False)['lifeExp'].mean()
# to make a plot of these data, first load a plotting package. We
can specify the shorthand 'plt' to make accessing functions easier
import matplotlib.pyplot as plt
# make plot by setting plot type (scatter) and x and y variables
plt.scatter(x = plot_df['year'], y = plot_df['lifeExp'])
plt.show()
```

Out[4]:



Debugging approach

Goal: Here we aim to find the mean life expectancy each year. We found the code by googling 'python pandas find mean of groups,' but we do not understand the line of code that provides this result. Thus, we will debug the line to understand it.

Debugging step 3: Break down each component of code into knowns and unknowns. Use documentation and experimentation to understand your unknowns.

Hint 3b: The dataframe's index is an axis label. Try googling 'python pandas index' to learn more. The new output is fine, unless you want both variables (year and mean life expectancy) as columns. We want each variable in a column to make plotting easier.

Lesson learned: Debug to understand your solution, not just the problem.

Figure 7.6: Anatomy of an error message, Part 4 (or: Debugging a solution). Lastly, we show how to debug a solution to understand a line of code found on the internet.

tational lab notebook is by definition a lab notebook: your lab notebook includes protocols, so your computational lab notebook should include protocols, too. Computational protocols are scripts, and these should include the code itself and how to access everything needed to implement the code. Include input (raw data) and output (results), too. Figures and interpretation can be included if that's how you organize your lab notebook. Develop computational "place habits" (file-saving strategies). It is easier to organize one drawer than it is to organize a whole lab, so start as soon as you begin to learn to program. If you can find that experiment you did on June 12, 2011—its protocol and results—in under five minutes, you should be able to find that figure you generated for lab meeting three weeks ago, complete with code and data, in under five minutes as well. This requires good version control or documentation of your work. Like with protocols, each time you run a script, you should note any modifications that are made. Document all changes in experimental and computational protocols. These habits will make you more efficient by enhancing your work's reproducibility. For specific advice, see "Ten Simple Rules for a Computational Biologist's Laboratory Notebook" (Schnell 2015), "Ten Simple Rules for Reproducible Computational Research" (Sandve et al. 2013), and "Ten Simple Rules for Taking Advantage of Git and GitHub" (Perez-Riverol et al. 2016).

7.1.1.9 Rule 8: Practice makes perfect

Use toy datasets to practice a problem or analysis. Biological data get big, fast. It's hard to find the computational needle-in-a-haystack, so set yourself up to succeed by practicing in controlled environments with simpler examples. Generate small toy datasets that use the same structure as your data. Make the toy data simple enough to predict how the numbers, text, etc., should react in your analysis. Test to ensure they do react as expected. This will help you understand what is being done in each step and troubleshoot errors, preparing you to scale up to large, unpredictable datasets. Use these datasets to test your approach, your implementation, and your interpretation. Toy datasets are your negative control, allowing you to differentiate between negative results and simulation failure.

7.1.1.10 Rule 9: Teach yourself

How would you teach you if you were another person? You would teach with a little more patience and a bit more empathy than you are practicing now. You are not alone in your occasional frustration (**Figure 7.7**). Learning takes time, so plan accordingly. Introductory courses are helpful to learn the basics because the basics are easy to neglect in self-study. Articulate clear expectations for yourself and benchmarks for success. Apply some of the structure (deadlines, assignments, etc.) you would

Figure 7.7: "How to exit the vim editor?" (or: We all get stuck at some point). Now viewed >1.6 million times; see: http://stackoverflow.com/questions/11828270/how-to-exit-the-vim-editor.

provide a student to help motivate and evaluate your progress. If something isn't working, adjust; not everyone learns best by any one approach. Explore tutorials, online classes, workshops, books like Practical Computing for Biologists (Haddock and Dunn 2011), local programming meetups, etc., to find your preferred approach.

7.1.1.11 Rule 10: Just do it

Just start coding. You can't edit a blank page.

Learning to program can be intimidating. The power and freedom provided in conducting your own computational analyses bring many decisions points, and each decision brings more room for mistakes. Furthermore, evaluating your work is less black-and-white than for some experiments. However, coding has the benefit that failure is risk free. No resources are wasted—not money, time (a student's job is to learn!), or a scientific reputation. *In silico*, the playing field is leveled by hard work and conscientiousness. So, while programming can be intimidating, the most intimidating step is starting.

7.1.1.12 Conclusion

Markowetz recently wrote, "Computational biologists are just biologists using a different tool" (Markowetz 2017). If you are a traditionally trained biologist, we intend these 10 simple rules as instruction (and pep talk) to learn a new, powerful, and exciting tool. The learning curve can be steep; however, the effort will pay dividends. Computational experience will make you more marketable as a scientist (see "Top

N Reasons To Do A Ph.D. or Post-Doc in Bioinformatics/Computational Biology" (Bergman 2012)). Computational research has fewer overhead costs and reduces the barrier to entry in transitioning fields (Kwok 2013), opening career doors to interested researchers. Perhaps most importantly, programming skills will make you better able to implement and interpret your own analyses and understand and respect analytical biases, making you a better experimentalist as well. Therefore, the time you spend at your computer is valuable. Acquiring programming expertise will make you a better biologist.

7.1.1.13 Acknowledgments

Thank you to Ed Hall, Pat Schloss, Matthew Jenior, Angela Zeigler, Jhansi Leslie, and Gregory Medlock for their feedback.

7.2 Contributioning to open source software

All modeling work presented in this dissertation used Cobra software. CobraPy is an open source software toolkit for **Co**straint-**B**ased **R**econstruction and **A**nalysis in the (freely available) **Py**thon programming language. An analogous Cobra package, written for the MATLAB programming language is also available. Both packages, CobraPy available at github.com/opencobra/cobrapy, depend on user contributions for development as few professional software developers are involved in the project. This section presents several contributions I have made to the software.

7.2.1 Gapfilling

7.2.1.1 What is gapfilling?

Genome-scale metabolic reconstructions are built from genomic data - reactions are added if there is genetic evidence for the associated enzyme. However, because not every enzyme is annotated in the genome (due to incomplete genome annotation), the reconstruction may not be complete. To increase the scope of a reconstruction (*i.e.* to add reactions), we perform 'gapfilling' to ensure that the reconstruction can complete a particular task (**Figure 7.8**). This optimization problem adds reactions to allow the reconstruction to carry flux under given constraints.

7.2.1.2 Modifications and testing of the gapfilling code

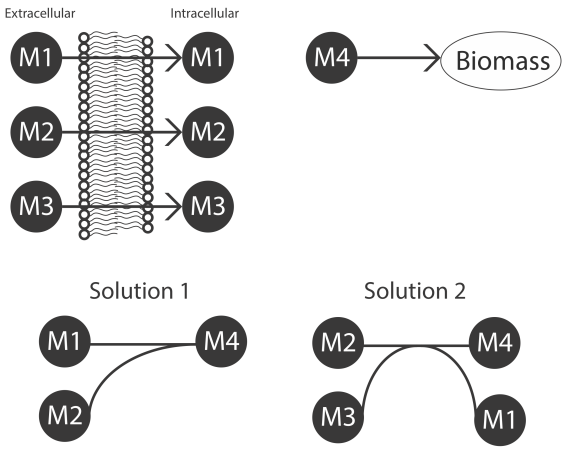


Figure 7.8: Gapfilling is a process to ensure functionality of a network by adding a reaction or reactions to fill a gap. Here the functionality needed is the generation of biomass from Metabolite 4 (M4). Three metabolites (M1, M2, and M3) can be imported into the cell, but the conversion from these metabolites to M4 is not annotated into the genome and, therefore, is not known. Reactions from a reaction database (Solution 1 and Solution 2) could complete the network to ensure that the cell can generate biomass from metabolites M1, M2, and M3. Alternative solutions are sometimes found, as demonstrated here. Figure from Gregory Medlock.

This function was an existing part of CobraPy, but had several flaws. Most importantly, if new metabolites were introduced during the gapfilling solution, no solution could be found. Thus, I also changed the code for how it searched and evaluated exchange reactions; these reactions will create or consume a metabolite from nothing. Biologically, these represent the process of gas exchange from the environment to a culture media or of changing media in an *in vitro* culture system. Lastly, I added new tests to evaluate the new code.

7.2.2 Renaming genes

7.2.2.1 Why would you want to rename genes?

EuPathDB, and PlasmoDB in particular, has updated gene identifiers several times. For example, the gene encoding lactate dehydrogenase in *Plasmodium falciparum* 3D7 was not included in the first generation of gene identifiers; in the second generation, the open reading frame was named 'PF13_0141' as it was the 141st gene on the 13th chromosome of *Plasmodium falciparum*. It is now named 'PF3D7_1324900', to highlight the species (Pf, meaning *P. falciparum*), the strain (3D7), and the chromosome. Expression datasets tend to use only one generation of gene identifiers, whereas the original model had a mixture of gene identifier generations. Single target

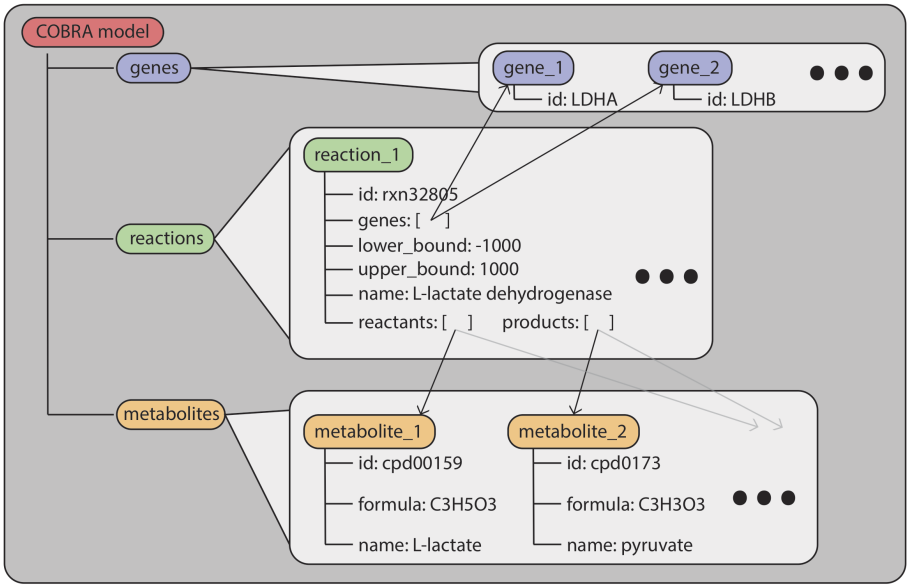


Figure 7.9: Model structures. The Cobra package, a COnstraint-Based Reconstruction and Analysis package, stores models as a series of objects, including gene, reaction, and metabolite objects. Each object is mapped to other related objects. For example, the reaction object describing the reaction catalyzed by lactate dehydrogenase maps to two genes and multiple metabolites, specifically the reactions's reactants and products. Figure from Gregory Medlock.

experimental datasets also tend to use a mixture of gene identifiers, typically using the most 'popular' identifier, or the identifier that was used during the gene's initial characterization. To keep iPfal18 up to date with the most current nomenclature, I updated all gene identifiers to the current generation. For some data integration, however, it is useful to move to a previous generation of identifiers.

7.2.2.2 What was wrong with the rename genes code?

Cobra models have several attributes, like enzyme-encoding genes, reactions, and metabolites, and these attributes are all interconnected (**Figure 7.9**). That is, a reaction feature points to the metabolites that are involved in the reaction as well as to the genes that govern the reaction. To change any of these attributes, the feature itself must be changed and the pointers between attributes must be updated. Thus, the process of updating gene identifiers requires a fair bit of code.

The existing rename_genes CobraPy function did not allow a subset of genes to be updated. If only some genes were being renamed, all 'old' names were removed. For example, if 'gene1' was being updated to 'gene_1' but 'gene_2' was not being updated, the resultant model would only contain 'gene_1.' Accordingly, I added code

to address this problem. One remaining problem exists however. If multiple genes are being replaced with the same new identifier, the new identifier will be duplicated in the gene-reaction rule. For example, if gene1_transcriptA and gene1_transcriptB are part of a gene-reaction rule 'gene1_transcriptA AND gene1_transcriptB' (indicating both gene1_transcriptA and gene1_transcriptB are necessary for the reaction to proceed) and both are being replaced with the new gene identifier 'gene1_transcripts', then the new gene-reaction rule will be 'gene1_transcripts AND gene1_transcripts'. However, I have identified no functional consequences to this problem.

7.2.3 Additional examples

Additional examples of new functions added to the CobraPy software can be found on my Github page, https://github.com/maureencarey/.

Chapter 8: Reflections and Future Directions

8 Reflections and Future Directions

The work presented here could springboard many scientific careers, and I hope to pursue many of the remaining questions, directly or indirectly, throughout my career. I will take this opportunity to focus on a few of my favorite future directions, both scientific and professional.

8.1 Parasite phenotyping with host in mind

The malaria parasite resides within the human red blood cell, stealing its nutrients and eventually rupturing the cell to invade neighbors. The parasite's demands for glucose are so great that it inhibits glycolysis in neighboring uninfected red blood cells (Mehta, Sonawat, and Sharma 2006), effectively starving the host (Sasi et al. 2007). This idea has intrigued me since I began reading about these parasites. *In vitro* blood stage culture of the malaria parasite is often treated as static, but the *in vivo* environment is certainly much more dynamic. This kind of nutritional competition and interaction with the host, as well as sheer stress and nutritional variation are major variables not address in *in vitro* culture. I am interested in leveraging clinical profiling data to address these *in vitro* limitations as my career progresses.

Also, experimental systems can be better characterized to address these *in vivo*/*in vitro*, especially the metabolome of the aging red blood cell. The red blood cell is often treated as a metabolically inactive entity (and directly referenced as such in Yeh and DeRisi (2011); Lamb (2012); Pellé et al. (2015)), but these cells contain over 2000 proteins by some reports (Bryk and Wiśniewski 2017) with physiologically relevant variation (Bordbar et al. 2015). *In vitro*, we grow *P. falciparum* in donor blood derived from one individual (termed a 'blood batch') for up to 30 days. Metabolite abundance varies amongst older and younger red blood cells *in vivo* (*e.g.* Cooper, Shukla, and Rennert (1976)) and the metabolome of red blood cells changes over 30 days of storage (Paglia et al., n.d.).[1] We found in **Chapter 6.3** that blood batch was more predictive of sample metabolome than artemisinin drug treatment and experimental variation hypothesized to be associated with blood age in **Chapter 5**. Thus, I hypothesize there is a large amount of experimental variation explained by the assumption of a "metabolically inactive" erythrocyte. Simple metabolomics experiments and growth phenotyping in multiple blood conditions will address this.

Additionally, while the health consequences of acute infection are often well characterized, I am also interested in the effects of chronic low level exposure to

[1] This was not tested using the same storage conditions used for *Plasmodium* blood culture.

these parasites. Many of these parasites have co-evolved with their human host (and, presumably, vector host as well). As disease fatality decreases with widespread treatment campaigns, the effect of 'asymptomatic' or low level infection will become even more important to understand. Thus, profiling these parasites in non-acute or sub-clinical infections will become increasingly valuable.

8.2 Moving from malaria to more neglected tropical diseases

I began my PhD focusing on *Plasmodium falciparum*, the most lethal, common, and well studied of the malaria parasites ("World Malaria Report 2017" 2017). Malaria, too, is one of the most lethal and well studied of the parasitic diseases; this put me in an excellent position to leverage computational tools to study a moderately understudied organism. By 'moderately understudied,' I mean that there are enough genotypic and phenotypic data on the parasite to build relatively high quality models, but these models could still be useful in contextualizing sparse datasets. There are enough data to learn the techniques (data analysis, model building, data interpretation) while illuminating the Pandora's box of the yet-undiscovered.

The other parasites presented in **Chapter 3** are even more understudied *Plasmodium*, making tool development for the interpretation and contextualization of sparse data even more important. I am grateful that I had the opportunity to learn in *Plasmodium* but I am excited to work in the field of the neglected tropical parasites. I aim to apply the models generated in **Chapter 3** to better understand these organisms, especially the *Cryptosporidium* parasites, much like I have used our *Plasmodium* models in **Chapter 5**.

8.3 Making modeling accessible

Biological modeling is currently a niche technique, the way that flow cytometry, microfluidics, and even microscopy once were. I aim to make models and modeling approaches more accessible. To do this, high quality models must be shared and the technical approaches communicated to both traditionally and nontraditionally-trained computational biologists. All models generated over the course of this work are or will be made publically available.

In addition, I am working on an approach to communicate modeling concepts to aspiring computational biologists. Inspired by both a 'toy' metabolic model generated by several colleagues (Rawls, Dougherty, Blais, *et al.*, in preparation) and the questions I often received when presenting my research to non-modelers, I became interested in

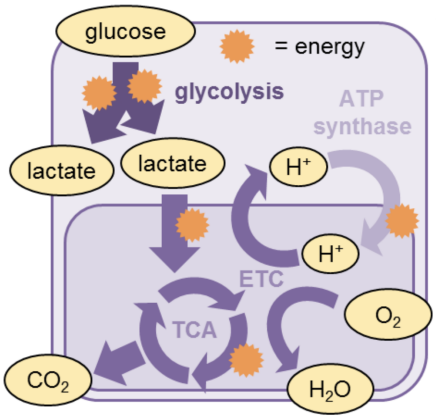


Figure 8.1: Textbook-like visualization of glycolysis and respiration summarized in our 'toy' metabolic network. Individual steps of glycolysis, the tricarboxylic acid (TCA) cycle, and the electron transport chain (ETC) are aggragated into summary reactions. Major metabolites (in yellow) are presented, as well as ATP generation (notated here as energy). Figure from Rawls, Dougherty, Blais, and Papin, in preparation.

using simple models to demonstrate modeling concepts without the use of programming. By decoupling the acts of learning the technical concepts and learning to program, more researchers can be exposed to genome-scale metabolic modeling enabling them to evaluate the utility of this modeling framework for their own biological questions.

I have to quote favorite saying in my family, "you have to pick the right tool for the job." Knives will cut almost anything, but sometimes scissors are a much more appropriate tool. However, learning *which* tool is right requires an understanding of the problem as well as usage of the tool and what assumptions that tool makes. To continue our scissors analogy, you have to know both what you are trying to cut and which end to hold the scissors. At minimum, basic programming is necessary to use most computational tools, and thus, earning to program is a major obstacle to obtaining this technical knowledge and picking the right tool.

For this reason, I am working in collaboration with Michal Stolarcyzk, a visiting masters student, to develop a web-based tool that visualizes a small metabolic network summarizing glycolysis (**Figure 8.1**) and performs common simulations on it (**Figure 8.2**). The application, documentation linked here, https://github.com/michalstolarczyk/shinyapp is presented from a biologist's point of view. For example, simulations are presented as three classes of experiments: changing the model's reaction bounds are described as (1) changing growth media, (2) under/overexpressing enzymes, and (3) knocking out genes. This unprecedented (but simple) shift in language retains all precision and accuracy of the technical details

but is much more approachable to biologists. My goal with this project is to help researchers decide if metabolic modeling is right for them and to motivate the learning process.

8.4 Using computational biology to 'level the playing field'

One of my first research experiences was in the field of public health; at an internship with the International AIDS Society, I was asked to compile a literature review to define 'community-based research' and explore the degree to which community-based research practices were used in the HIV/AIDS field (the project eventually evolved into the following publication: Brizay et al. (2015)). I quickly learned the benefit of developing local capacity for science and advocacy, as the key stakeholders of any public health problem will have some of the most creative and practical solutions. For this reason, paired with my enjoyment of programming and desire to broadly increase diversity in the sciences, I plan to work to increase capacity for computational biosciences, particularly to address infectious disease public health problems, locally and globally.

Computational tools allow researchers to generate high-confidence, data-driven hypotheses before stepping in the lab - and are thus well-suited for use in research groups with limited funding. In addition to a lower start-up cost, computational research provides researchers more flexibility over their schedules, making it well-suited for those with additional responsibilities (*i.e.* parents, students with additional jobs, etc.). Computational sciences are underrepresented both in the international research community and by minorities in the United States. In the United States, less than 20% of computer science or engineering undergraduate degrees are granted to women. The statistics for people of color are worse: less than 10% for both computer science or engineering undergraduate degrees. Representation drops for higher degrees ("Women, Minorities, and Persons with Disabilities in Science and Engineering," n.d.). Moreover, few biologists are trained in programming or basic computer science, and this excludes the application-driven students from entering the field of computational biology, leaving only students who entered the field from a technical background. Thus, I aim to increase diversity in science by providing engaged mentorship to support young scientists and research opportunities for aspiring scientists, and by investing in traditional diversity initiatives.

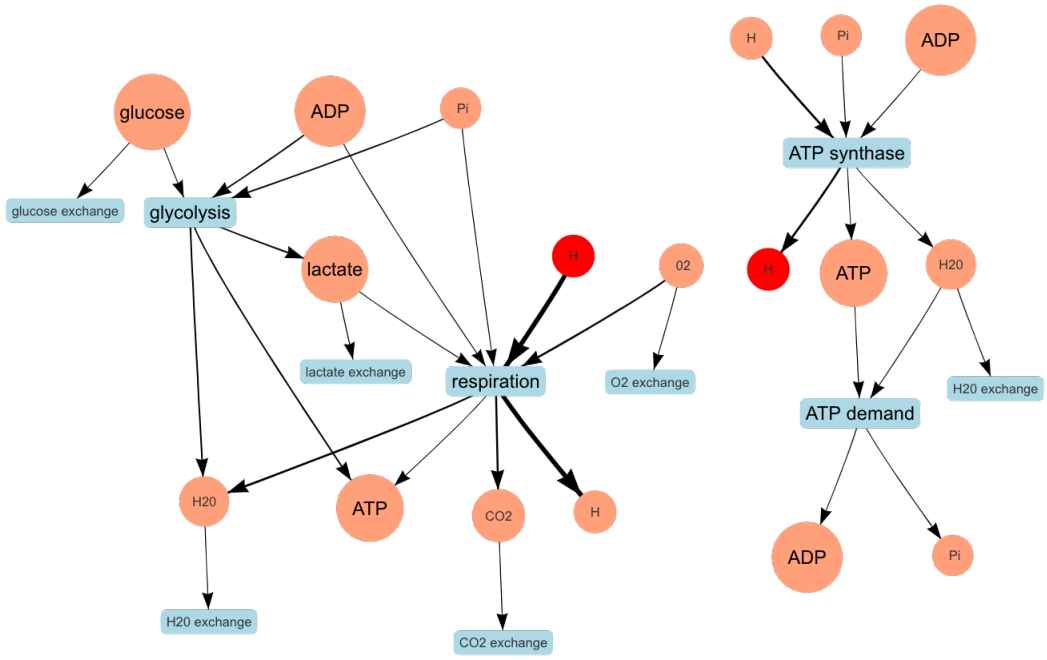


Figure 8.2: Model visualization depicts all reactions contained in the network. Here, metabolites are in orange or red, and reactions in blue. Flux of each reaction is indicated by the arrow weight.

Chapter 9: Conclusions

9 Conclusions

In this dissertation, I present (1) Paradigm, a framework based on genome-scale metabolic modeling for organizing and interpreting biochemical knowledge about eukaryotic parasites, (2) the curation of an individual parasite metabolic reconstruction for the most lethal malaria parasite, (3) the application of network-based modeling approaches to understand and target antimalarial resistance, and (4) network-based analytics for novel 'omics measurements and the current limitations of such analyses, as well as highlight a few of my career aspirations. Throughout the chapters of this dissertation, I demonstrate how network-based analyses are advantageous as they reveal complex or emergent trends not readily apparent from more "traditional" analytic methods.

Biological observations, whether they are species of parasitic eukaryotes, genes in a genome, expressed transcripts, or metabolites detected in a cell, can be viewed either as independent entities or nodes in a biological network. Statistical approaches consistently view these observations as independent entities. For example, expression data is often analyzed by calculating differential expression and associated significance, and rank ordering transcript fold change; the largest significant changes are typically the focus of downstream analyses and interpretation. In univariate statistical analyses, each variable is analyzed in isolation, and *post hoc* corrections are used to account for the fact that there are often multiple variables measured.

Network-based approaches, in contrast, treat these observations as nodes. Expression data might be analyzed using a network-based approach by creating a co-expression network or by using network connectivity and function to set variable-specific thresholds for interpreting expression changes (used here). These approaches typically emphasize *patterns* of change, instead of or in addition to *degree* of change. Multivariate statistical analyses and machine learning meet univariate statistical and network-based approaches somewhere in the middle, while adding new alternative assumptions, limitations, and benefits. Many forms of machine learning view these biological observations from a less biased perspective, asking both "are there changes?" and "are there patterns of changes?" rather than "what are the changes or patterns of changes?".

Throughout this dissertation, a combination of statistical (univariate and multivariate), network-based, and machine learning approaches are used. I hope it is clear that more knowledge can be gleaned by using all of these analytic approaches than any one in isolation. Most eukaryotic parasites are unculturable and/or not experimentally tractable, limiting the collection of data. Thus, to better leverage data collected in these organisms to understand clinically relevant parasite phenotypes,

we must integrate data (*i.e.* diverse data types to better predict patterns of change) or focus analyses on the interpretable subset of data (*e.g.* the annotated subset of genes in a genome). Multivariate statistical, network-based, and machine learning approaches are best suited for these goals. This dissertation emphasizes network-based approaches because they are underutilized in the field (**Table 3.7**) and they have immense potential to accelerate the drug development process (*e.g.* Agren et al. (2014); Raškevičius et al. (2018); Kim et al. (2011)).

We first built Paradigm, a **Para**site **D**atabase for **G**enome-scale metabolic **M**odels, to provide the field a framework for organizing and interpreting our biochemical knowledge about eukaryotic parasites (**Chapter 3**). Both our approach and resultant models can be used broadly by the computational and parasitology fields. These reconstructions can be used to generate targeted hypotheses for exploring differences between species and improving genome annotation. Paradigm will be made publicly available following the examples of the Eukaryotic Pathogens Database (Aurrecoechea et al. 2017) and the Biochemical, Genetic and Genomic database (King et al. 2016). By sharing these tools, other research groups can use these models to answer their own research questions as done with other large collections of metabolic models (King et al. 2016; Magnúsdóttir et al. 2017).

We applied our Paradigm framework to begin to identify the most representative *in vitro* system or non-primate infection model of disease for drug development. We hypothesize that the best test system may vary by the metabolic pathway being targeted. We highlight some such suggestions, especially regarding kinase inhibitor screens, and future work will include simulations of gene essentiality to quantify the accuracy of model predictions. Paradigm contains draft reconstructions, unlike the curated *P. falciparum* 3D7 reconstruction curated in **Chapter 4** and applied in **Chapter 5**, and thus these models will require some manual curation as well. Following curation, we will identify a conserved list of essential genes as well as 'model pathways' that are comparable between model organisms and disease-causing parasites.

We focus on these common drug targets because drug repurposing has been extremely useful in the fight against parasitic disease due to the biological challenges in targeting eukaryotic cells and the limited funding available for drug development. Antimalarials have been used to treat toxoplasmosis, and antibiotics have been used to treat several parasitic diseases (**Table 2.1**). Anticancer drugs also have potential to become antiparasitic therapies, and vice versa (Huijsduijnen et al. 2013; Nzila et al. 2010). Additionally, it has been challenging to incentivise development and even production of species-specific antiparasitics for the neglected tropical disease especially. For example, an antiparasitic drug for African sleeping sickness was developed but medical production stopped due to its cost; once eflornithine was discovered to also be

an effective way of removing facial hair, the commercial potential encouraged cosmetic companies to resume production (McNeil, n.d.). Thus, the economic and biological constraints on antiparasitic drug development make drugs that target multiple parasites appealing.

Using genome-scale metabolic models to identify drug targets also lends insight into the drug's mechanism of action. Knowing the mechanism of action of a drug can accelerate the understanding of resistance mechanisms and development of new drugs. For example, the antimalarial atovaquone inhibits an enzyme in the electron transport chain that is essential for nucleotide synthesis (Painter et al. 2007). By understanding the drug target (Fry and Pudney (1992) and others), genetic association studies can be focused on mutations within the drug's target to identify the causal changes (Syafruddin, Siregar, and Marzuki (1999) and others). Following the identification of the causal mutation, researchers were able to identify that this mutation was not transmissible as parasites with the mutation could not survive in the mosquito vector (Goodman et al. 2016), and thus has reduced capacity to spread throughout communities. Thus, rational drug design, especially using genome-scale metabolic models, has advantages to accelerate the development and implementation of effective therapies.

Next, we curated a genome-scale metabolic reconstruction, iPfal17, to represent the metabolism of the asexual blood-stage *P. falciparum* malaria parasite and identified a set of metabolic tasks to evaluate both iPfal17 and future model iterations (**Chapter 4**). This curation pipeline is an example of the manual steps that can be taken to improve the quality of models generated in Paradigm. Both the processes of model building and applying models to answer biological questions expand our understanding of the parasite's metabolism. Moreover, the model can be viewed as a framework for storing thousands of hypotheses, such as hypotheses about gene-protein-reaction associations or enzyme function (*i.e.* reaction formulas), as well as rigorously evaluating and documenting these hypotheses.

However, all models have limitations and thus model construction is an iterative process requiring regular curation for improvement. For example, iPfal17 was further curated using metabolomics data in **Chapter 6.2.2** adding functionality to our already manually curated network. Future curation will include metabolomics-based curation, like in **Chapter 6.2.2**, as well as validation against a novel genome-wide essentiality screen conducted in *P. falciparum* (Zhang et al. 2018). Zhang et al. (2018) offers an unprecedented opportunity for network curation in this organism. Curation will require sequential knockout of genes *in silico* to identify essential genes, and the comparison of this gene list to essential genes identified experimentally. These discrepancies will be used to guide and focus experiments that identify the unknown

biological phenomena that lead to the experimental result.

With our curated *Plasmodium falciparum* reconstruction, we generated new hypotheses about antimalarial drug resistance. We find that inherent differences exist in artemisinin resistant and sensitive parasite metabolism, even before artemisinin treatment. Artemisinin resistant parasites have major metabolic shifts in the mitochondria and in the synthesis of folates and polyamines, indicating incomplete transition to the metabolic state most appropriate for the blood-stage environment. We next focused on polyamine synthesis due to literature support for this pathway being implicated in artemisinin resistance. We experimentally interrogated putrescine scavenging and *de novo* synthesis in artemisinin sensitive and resistant malaria parasites. Inconclusive results highlight the role of the host cell, the human erythrocyte, as a metabolic buffer for the parasite. Human-to-human and temporal variation in the metabolome of blood cells and serum may influence the development of resistance if polyamine levels influence artemisinin resistance.

Similarly, we implicated collateral metabolic shifts in chloroquine resistance. We find resistant parasites mount a metabolic response to chloroquine treatment, particularly in thiamine, lipid, isoprenoid, and folate metabolism. Thiamine is metabolized into thiamine diphosphate, an essential cofactor for isoprenoid synthesis. As a result of this finding and the availability of experimental tools to manipulate thiamine diphosphate levels, we are actively interrogating thiamine metabolism experimentally. Specifically, we are using thiamine-free media, oxythiamine (an analog of thiamine diphosphate that inhibits thiamine diphosphate-dependent enzymes), and direct measurement of thiamine, thiamine monophosphate, and the active thiamine diphosphate using precolumn derivatization, reversed-phase liquid chromatography and fluorescence detection as previously described (Gerrits et al. 1997; Stuetz et al. 2012). By growing sensitive and resistant parasites in the presence or absence of exogenous thiamine and/or oxythiamine, we will determine if sensitive and resistant parasites predominately use *de novo* synthesis or scavenge thiamine.

The accuracy of these model predictions is limited by the quality of the model construction and by how well suited the model is to address the biological question. Thus, prediction accuracy can be improved by improving the quality of the model via model curation. We have clearly address model improvement via curation in **Chapter 4**. To expand upon on our manual curation, we used metabolomics data to curate our model of *Plasmodium falciparum* metabolism, adding network functionality and improving model predictions in **Chapter 6.2.2**.

Alternatively, prediction accuracy can be improved by better matching the model to the biological question. This is performed by the integration of high-quality condition-specific data into the model with well validated integration algorithms, as

implemented in **Chapter 5**. For example, to simulate essential enzymes *in vivo*, *in vivo* expression data would better constrain the model than an *in vitro* dataset. Again, intracellular metabolomics data is appealing as there are fewer layers of regulation between the metabolome and phenotype, than (for example) there are between the transcriptome and phenotype.

However, exploration of metabolomics data generated in the Guler lab highlighted that these metabolomics results are biased by methodological challenges associated with metabolomics (**Chapter 6.3**). Currently, it is not possible to distinguish metabolites derived from the host cell versus metabolites derived from the parasite, for any intracellular pathogen. Improved metabolomics methods could disentangle the individual behavior of both host and parasite, but current approaches are unable to distinguish between the two (Carey et al. (2018), see also, Olszewski et al. (2010)). We hypothesize that enrichment of parasite material away from host material and using uninfected host cells as a control will improve our ability to distinguish host from parasite metabolome. Although we show that analytic approaches can improve the accuracy and interpretability of intracellular parasite metabolomes, ultimately better experimental methods are needed to extract biological differences from samples. Thus, metabolomics data can be better used to profile the parasite phenotype and to improve model predictions when the host and parasite metabolomes can be separated.

In sum, throughout this dissertation, I present my contributions to the fields of parasitology and computational biology and demonstrate the utility of computational tools including network-based modeling, multivariate statistics, and machine learning in the study of the biology of eukaryotic parasites. All of these computational tools are reliant on the user's computational proficiency, and thus I also emphasize the importance of learning to program and making the introduction of computational techniques accessible to biologists.

10 Appendix

This appendix contains links to the supplementary information referenced throughout this document.

10.1 Supplementary Tables

Some tables were excluded from the for-print version of this document due to size limitations. They can be obtained via the following links.

Chapter 4, Additional file 3, Table S1: See https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-3905-1 for Additional file 3, Table S1

Chapter 4, Additional file 3, Table S8: See https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-3905-1 for Additional file 3, Table S8

10.2 Code and analyses

All code used to generate the analyses and resultant tables and figures is hosted on GitHub, either my personal page, the Guler Malaria Lab page, or the (Papin) Computational Systems Biology Lab page. If you are reading this document electronically, all links throughout document are hyperlinked.

My personal page (hyperlinked here): https://github.com/maureencarey/

Guler Malaria Lab (hyperlinked here): https://github.com/gulermalaria/

Computational Systems Biology Lab (hyperlinked here): https://github.com/csbl

References

Abdel-Haleem, Alyaa M, Hooman Hefzi, Katsuhiko Mineta, Xin Gao, Takashi Gojobori, Bernhard O Palsson, Nathan E Lewis, and Neema Jamshidi. 2018. "Functional Interrogation of Plasmodium Genus Metabolism Identifies Species- and Stage-Specific Differences in Nutrient Essentiality and Drug Targeting." *PLoS Comput. Biol.* 14 (1). journals.plos.org: e1005895. https://doi.org/10.1371/journal.pcbi.1005895.

Abrahamsen, Mitchell S, Thomas J Templeton, Shinichiro Enomoto, Juan E Abrahante, Guan Zhu, Cheryl A Lancto, Mingqi Deng, et al. 2004. "Complete Genome Sequence of the Apicomplexan, Cryptosporidium Parvum." *Science* 304 (5669). science.sciencemag.org: 441–45. https://doi.org/10.1126/science.1094786.

Agren, Rasmus, Adil Mardinoglu, Anna Asplund, Caroline Kampf, Mathias Uhlen, and Jens Nielsen. 2014. "Identification of Anticancer Drugs for Hepatocellular Carcinoma Through Personalized Genome-scale Metabolic Modeling." *Mol. Syst. Biol.* 10 (3). EMBO Press: 721. https://doi.org/10.1002/msb.145122.

Aikawa, M, C G Huff, and H Sprinz. 1969. "Comparative Fine Structure Study of the Gametocytes of Avian, Reptilian, and Mammalian Malarial Parasites." *J. Ultrastruct. Res.* 26 (3). Elsevier: 316–31.

Akpunarlieva, Snezhana, Stefan Weidt, Dhilia Lamasudin, Christina Naula, David Henderson, Michael Barrett, Karl Burgess, and Richard Burchmore. 2017. "Integration of Proteomics and Metabolomics to Elucidate Metabolic Adaptation in Leishmania." *J. Proteomics* 155 (February). Elsevier: 85–98. https://doi.org/10.1016/j.jprot.2016.12.009.

Allaire, J J, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, and Winston Chang. 2018. "Rmarkdown: Dynamic Documents for R."

Allary, Marina, Jeff Zhiqiang Lu, Liqun Zhu, and Sean T Prigge. 2007. "Scavenging of the Cofactor Lipoate Is Essential for the Survival of the Malaria Parasite Plasmodium Falciparum." *Mol. Microbiol.* 63 (5). Wiley Online Library: 1331–44. https://doi.org/10.1111/j.1365-2958.2007.05592.x.

Allman, Erik L, Heather J Painter, Jasmeet Samra, Manuela Carrasquilla, and Manuel Llinás. 2016. "Metabolomic Profiling of the Malaria Box Reveals Antimalarial Target Pathways." *Antimicrob. Agents Chemother.* 60 (11). Am Soc Microbiol: 6635–49. https://doi.org/10.1128/AAC.01224-16.

Antoine, Thomas, Nicholas Fisher, Richard Amewu, Paul M O'Neill, Stephen A Ward, and Giancarlo A Biagini. 2014. "Rapid Kill of Malaria Parasites by Artemisinin

and Semi-Synthetic Endoperoxides Involves ROS-dependent Depolarization of the Membrane Potential." *J. Antimicrob. Chemother.* 69 (4). academic.oup.com: 1005–16. https://doi.org/10.1093/jac/dkt486.

Ariey, Frédéric, Benoit Witkowski, Chanaki Amaratunga, Johann Beghain, Anne-Claire Langlois, Nimol Khim, Saorin Kim, et al. 2014. "A Molecular Marker of Artemisinin-Resistant Plasmodium Falciparum Malaria." *Nature* 505 (7481). nature.com: 50–55. https://doi.org/10.1038/nature12876.

Arnou, Bertrand, Cédric Montigny, Jens Preben Morth, Poul Nissen, Christine Jaxel, Jesper V Møller, and Marc le Maire. 2011. "The Plasmodium Falciparum ca(2+)-ATPase PfATP6: Insensitive to Artemisinin, but a Potential Drug Target." *Biochem. Soc. Trans.* 39 (3): 823–31. https://doi.org/10.1042/BST0390823.

Asahi, H, T Kanazawa, Y Kajihara, K Takahashi, and T Takahashi. 1996. "Hypoxanthine: A Low Molecular Weight Factor Essential for Growth of Erythrocytic Plasmodium Falciparum in a Serum-Free Medium." *Parasitology* 113 (Pt 1) (July). cambridge.org: 19–23.

Ashley, Elizabeth A, Mehul Dhorda, Rick M Fairhurst, Chanaki Amaratunga, Parath Lim, Seila Suon, Sokunthea Sreng, et al. 2014. "Spread of Artemisinin Resistance in Plasmodium Falciparum Malaria." *N. Engl. J. Med.* 371 (5). Mass Medical Soc: 411–23. https://doi.org/10.1056/NEJMoa1314981.

Aspinall, Tanya V, David H M Joynson, Edward Guy, John E Hyde, and Paul F G Sims. 2002. "The Molecular Basis of Sulfonamide Resistance in Toxoplasma Gondii and Implications for the Clinical Management of Toxoplasmosis." *J. Infect. Dis.* 185 (11): 1637–43. https://doi.org/10.1086/340577.

Assaraf, Y G, L Abu-Elheiga, D T Spira, H Desser, and U Bachrach. 1987. "Effect of Polyamine Depletion on Macromolecular Synthesis of the Malarial Parasite, Plasmodium Falciparum, Cultured in Human Erythrocytes." *Biochem. J* 242 (1). ncbi.nlm.nih.gov: 221–26.

Auguie, B. 2016. "GridExtra: Miscellaneous Functions for' Grid' Graphics. R Package Version 2.2. 1." *Google Scholar*.

Aurrecoechea, Cristina, Ana Barreto, Evelina Y Basenko, John Brestelli, Brian P Brunk, Shon Cade, Kathryn Crouch, et al. 2017. "EuPathDB: The Eukaryotic Pathogen Genomics Database Resource." *Nucleic Acids Res.* 45 (D1): D581–D591.

Aurrecoechea, Cristina, John Brestelli, Brian P Brunk, Jennifer Dommer, Steve Fischer, Bindu Gajria, Xin Gao, et al. 2009. "PlasmoDB: A Functional Genomic Database for Malaria Parasites." *Nucleic Acids Res.* 37 (Database issue): D539–43.

https://doi.org/10.1093/nar/gkn814.

Babbitt, Shalon E, Lindsey Altenhofen, Simon A Cobbold, Eva S Istvan, Clare Fennell, Christian Doerig, Manuel Llinás, and Daniel E Goldberg. 2012. "Plasmodium Falciparum Responds to Amino Acid Starvation by Entering into a Hibernatory State." *Proc. Natl. Acad. Sci. U. S. A.* 109 (47). National Acad Sciences: E3278–87. https://doi.org/10.1073/pnas.1209823109.

Bannister, L H, J M Hopkins, R E Fowler, S Krishna, and G H Mitchell. 2000. "A Brief Illustrated Guide to the Ultrastructure of Plasmodium Falciparum Asexual Blood Stages." *Parasitol. Today* 16 (10). Elsevier: 427–33.

Baumeister, Stefan, Markus Winterberg, Christophe Duranton, Stephan M Huber, Florian Lang, Kiaran Kirk, and Klaus Lingelbach. 2006. "Evidence for the Involvement of Plasmodium Falciparum Proteins in the Formation of New Permeability Pathways in the Erythrocyte Membrane." *Mol. Microbiol.* 60 (2). Wiley Online Library: 493–504. https://doi.org/10.1111/j.1365-2958.2006.05112.x.

Baydoun, Martha, Sadia Benamrouz Vanneste, Colette Creusy, Karine Guyot, Nausicaa Gantois, Magali Chabe, Baptiste Delaire, et al. 2017. "Three-Dimensional (3D) Culture of Adult Murine Colon as an in Vitro Model of Cryptosporidiosis: Proof of Concept." *Sci. Rep.* 7 (1): 17288.

Bazzani, Susanna, Andreas Hoppe, and Hermann-Georg Holzhütter. 2012. "Network-Based Assessment of the Selectivity of Metabolic Drug Targets in Plasmodium Falciparum with Respect to Human Liver Metabolism." *BMC Syst. Biol.* 6 (August). bmcsystbiol.biomedcentral.com: 118. https://doi.org/10.1186/1752-0509-6-118.

Beck, Josh R, Vasant Muralidharan, Anna Oksman, and Daniel E Goldberg. 2014. "PTEX Component HSP101 Mediates Export of Diverse Malaria Effectors into Host Erythrocytes." *Nature* 511 (7511). nature.com: 592–95. https://doi.org/10.1038/nature13574.

Becker, Katja, Leann Tilley, Jonathan L Vennerstrom, David Roberts, Stephen Rogerson, and Hagai Ginsburg. 2004. "Oxidative Stress in Malaria Parasite-Infected Erythrocytes: Host–Parasite Interactions." *Int. J. Parasitol.* 34 (2). Elsevier: 163–89. https://doi.org/10.1016/j.ijpara.2003.09.011.

Bei, Amy K, Tiffany M DeSimone, Aida S Badiane, Ambroise D Ahouidi, Tandakha Dieye, Daouda Ndiaye, Ousmane Sarr, Omar Ndir, Souleymane Mboup, and Manoj T Duraisingh. 2010. "A Flow Cytometry-Based Assay for Measuring Invasion of Red Blood Cells by Plasmodium Falciparum." *Am. J. Hematol.* 85 (4). Wiley Online Library: 234–37.

Belongia, Edward A, and Allison L Naleway. 2003. "Smallpox Vaccine: The Good, the Bad, and the Ugly." *Clin. Med. Res.* 1 (2): 87–92.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *J. R. Stat. Soc. Series B Stat. Methodol.* 57 (1). [Royal Statistical Society, Wiley]: 289–300.

Benjamini, Yoav, and Daniel Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing Under Dependency." *Ann. Stat.* 29 (4). Institute of Mathematical Statistics: 1165–88.

Bergman, Casey. 2012. "Top N Reasons to Do a Ph.D. Or Post-Doc in Bioinformatics/Computational Biology." https://caseybergman.wordpress.com/2012/07/31/top-n-reasons-to-do-a-ph-d-or-post-doc-in-bioinformaticscomputational-biology/.

Berg, Henk van den, Morteza Zaim, Rajpal Singh Yadav, Agnes Soares, Birkinesh Ameneshewa, Abraham Mnzava, Jeffrey Hii, Aditya Prasad Dash, and Mikhail Ejov. 2012. "Global Trends in the Use of Insecticides to Control Vector-Borne Diseases." *Environ. Health Perspect.* 120 (4): 577–82. https://doi.org/10.1289/ehp.1104340.

Biddau, Marco, and Sylke Müller. 2016. "Carbon Metabolism of Plasmodium Falciparum." *Compr Anal Parasite Biol: Metab to Drug Discov* 371. books.google.com.

Biggs, Matthew B, and Jason A Papin. 2017. "Managing Uncertainty in Metabolic Network Structure and Improving Predictions Using EnsembleFBA." *PLoS Comput. Biol.* 13 (3). Public Library of Science: e1005413. https://doi.org/10.1371/journal.pcbi.1005413.

Blader, Ira J, and Anita A Koshy. 2014. "Toxoplasma Gondii Development of Its Replicative Niche: In Its Host Cell and Beyond." *Eukaryot. Cell* 13 (8): 965–76. https://doi.org/10.1128/EC.00081-14.

Bodenhofer, Ulrich, Enrico Bonatesta, Christoph Horejš-Kainrath, and Sepp Hochreiter. 2015. "Msa: An R Package for Multiple Sequence Alignment." *Bioinformatics* 31 (24): 3997–9. https://doi.org/10.1093/bioinformatics/btv494.

Bogdan, Christian, Steffen Stenger, Martin Röllinghoff, and Werner Solbach. 1991. "Cytokine Interactions in Experimental Cutaneous Leishmaniasis. Interleukin 4 Synergizes with Interferon-γ to Activate Murine Macrophages for Killing of Leishmania Major Amastigotes." *Eur. J. Immunol.* 21 (2). Wiley Online Library: 327–33.

Borchers, H W. 2015. "Pracma: Practical Numerical Math Functions."

Bordbar, Aarash, Douglas McCloskey, Daniel C Zielinski, Nikolaus Sonnenschein, Neema Jamshidi, and Bernhard O Palsson. 2015. "Personalized Whole-Cell Kinetic

Models of Metabolism for Discovery in Genomics and Pharmacodynamics." *Cell Syst* 1 (4): 283–92. https://doi.org/10.1016/j.cels.2015.10.003.

Bosi, Emanuele, Jonathan M Monk, Ramy K Aziz, Marco Fondi, Victor Nizet, and Bernhard Ø Palsson. 2016. "Comparative Genome-Scale Modelling of Staphylococcus Aureus Strains Identifies Strain-Specific Metabolic Capabilities Linked to Pathogenicity." *Proceedings of the National Academy of Sciences* 113 (26): E3801–E3809. https://doi.org/10.1073/pnas.1523199113.

Bouzid, Maha, Paul R Hunter, Rachel M Chalmers, and Kevin M Tyler. 2013. "Cryptosporidium Pathogenicity and Virulence." *Clin. Microbiol. Rev.* 26 (1). Am Soc Microbiol: 115–34. https://doi.org/10.1128/CMR.00076-12.

Boyom, Fabrice F, Patrick V T Fokou, Lauve R Y Tchokouaha, Thomas Spangenberg, Alvine N Mfopa, Ruffin M T Kouipou, Cedric J Mbouna, Valerie F Donkeng Donfack, and Paul H A Zollo. 2014. "Repurposing the Open Access Malaria Box to Discover Potent Inhibitors of Toxoplasma Gondii and Entamoeba Histolytica." *Antimicrob. Agents Chemother.* 58 (10): 5848–54. https://doi.org/10.1128/AAC.02541-14.

Breglio, Kimberly F, Rifat S Rahman, Juliana M Sá, David J Roberts, and Thomas E Wellems. 2018. "Kelch Mutations in Plasmodium Falciparum Protein K13 Do Not Modulate Dormancy After Artemisinin Exposure and Sorbitol Selection in Vitro." *Antimicrob. Agents Chemother.* 62 (5). https://doi.org/10.1128/AAC.02256-17.

Brizay, Ulrike, Lina Golob, Jason Globerman, David Gogolishvili, Mara Bird, Britt Rios-Ellis, Sean B Rourke, and Shirin Heidari. 2015. "Community-Academic Partnerships in HIV-related Research: A Systematic Literature Review of Theory and Practice." *J. Int. AIDS Soc.* 18 (1). Wiley Online Library: 19354.

Brown, Tyler S, Christopher G Jacob, Joana C Silva, Shannon Takala-Harrison, Abdoulaye Djimdé, Arjen M Dondorp, Mark Fukuda, et al. 2015. "Plasmodium Falciparum Field Isolates from Areas of Repeated Emergence of Drug Resistant Malaria Show No Evidence of Hypermutator Phenotype." *Infect. Genet. Evol.* 30 (March). Elsevier: 318–22. https://doi.org/10.1016/j.meegid.2014.12.010.

Bryk, Agata H, and Jacek R Wiśniewski. 2017. "Quantitative Analysis of Human Red Blood Cell Proteome." *J. Proteome Res.* 16 (8): 2752–61. https://doi.org/10.1021/acs.jproteome.7b00025.

Buchfink, Benjamin, Chao Xie, and Daniel H Huson. 2015. "Fast and Sensitive Protein Alignment Using DIAMOND." *Nat. Methods* 12 (1). nature.com: 59–60. https://doi.org/10.1038/nmeth.3176.

Bull, P C, B S Lowe, M Kortok, C S Molyneux, C I Newbold, and K Marsh. 1998.

"Parasite Antigens on the Infected Red Cell Surface Are Targets for Naturally Acquired Immunity to Malaria." *Nat. Med.* 4 (3). nature.com: 358–60.

Bushell, Ellen, Ana Rita Gomes, Theo Sanderson, Burcu Anar, Gareth Girling, Colin Herd, Tom Metcalf, et al. 2017. "Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes." *Cell* 170 (2): 260–272.e8.

Bustos, D G, C J Canfield, E Canete-Miguel, and D B Hutchinson. 1999. "Atovaquone-Proguanil Compared with Chloroquine and Chloroquine-Sulfadoxine-Pyrimethamine for Treatment of Acute Plasmodium Falciparum Malaria in the Philippines." *J. Infect. Dis.* 179 (6): 1587–90. https://doi.org/10.1086/314770.

Campbell, W, and E Bolker. 2002. "Teaching Programming by Immersion, Reading and Writing." In *32nd Annual Frontiers in Education*, 1:T4G–23–T4G–28vol.1. ieeexplore.ieee.org. https://doi.org/10.1109/FIE.2002.1158015.

Canham, P B, and A C Burton. 1968. "Distribution of Size and Shape in Populations of Normal Human Red Cells." *Circ. Res.* 22 (3). Am Heart Assoc: 405–22.

Carey, Maureen A, Vincent Covelli, Audrey Brown, Gregory L Medlock, Mareike Haaren, Jessica G Cooper, Jason A Papin, and Jennifer L Guler. 2018. "Influential Parameters for the Analysis of Intracellular Parasite Metabolomics." *mSphere* 3 (2). https://doi.org/10.1128/mSphere.00097-18.

Carey, Maureen A, and Jason A Papin. 2018. "Ten Simple Rules for Biologists Learning to Program." *PLoS Comput. Biol.* 14 (1): e1005871. https://doi.org/10.1371/journal.pcbi.1005871.

Carey, Maureen A, Jason A Papin, and Jennifer L Guler. 2017. "Novel Plasmodium Falciparum Metabolic Network Reconstruction Identifies Shifts Associated with Clinical Antimalarial Resistance." *BMC Genomics* 18 (1): 543. https://doi.org/10.1186/s12864-017-3905-1.

Chakraborty, Arnish. 2016. "Emerging Drug Resistance in Plasmodium Falciparum: A Review of Well-Characterized Drug Targets for Novel Antimalarial Chemotherapy." *Asian Pacific Journal of Tropical Disease* 6 (7). Elsevier: 581–88. https://doi.org/10.1016/S2222-1808(16)61090-3.

Chang, W. 2014. "Extrafont: Tools for Using Fonts." R package version 0.17.

Charif, Delphine, and Jean R Lobry. 2007. "SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis." In *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, edited by Ugo Bastolla, Markus Porto, H Eduardo Roman, and Michele

Vendruscolo, 207–32. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-35306-5/_10.

Chaudhary, Kshitiz, John A Darling, Leah M Fohl, William J Sullivan Jr, Robert G K Donald, Elmer R Pfefferkorn, Buddy Ullman, and David S Roos. 2004. "Purine Salvage Pathways in the Apicomplexan Parasite Toxoplasma Gondii." *J. Biol. Chem.* 279 (30): 31221–7. https://doi.org/10.1074/jbc.M404232200.

Chavali, Arvind K, Jeffrey D Whittemore, James A Eddy, Kyle T Williams, and Jason A Papin. 2008. "Systems Analysis of Metabolism in the Pathogenic Trypanosomatid Leishmania Major." *Mol. Syst. Biol.* 4 (March). msb.embopress.org: 177. https://doi.org/10.1038/msb.2008.15.

Chen, Nanhua, Alexis N LaCrue, Franka Teuscher, Norman C Waters, Michelle L Gatton, Dennis E Kyle, and Qin Cheng. 2014. "Fatty Acid Synthesis and Pyruvate Metabolism Pathways Remain Active in Dihydroartemisinin-Induced Dormant Ring Stages of Plasmodium Falciparum." *Antimicrob. Agents Chemother.* 58 (8). Am Soc Microbiol: 4773–81. https://doi.org/10.1128/AAC.02647-14.

Cheng, Qin, Dennis E Kyle, and Michelle L Gatton. 2012. "Artemisinin Resistance in Plasmodium Falciparum: A Process Linked to Dormancy?" *Int. J. Parasitol. Drugs Drug Resist.* 2 (December). Elsevier: 249–55. https://doi.org/10.1016/j.ijpddr.2012.01.001.

Chiappino-Pepe, Anush, Stepan Tymoshenko, Meriç Ataman, Dominique Soldati-Favre, and Vassily Hatzimanikatis. 2017. "Bioenergetics-Based Modeling of Plasmodium Falciparum Metabolism Reveals Its Essential Genes, Nutritional Requirements, and Thermodynamic Bottlenecks." *PLoS Comput. Biol.* 13 (3). journals.plos.org: e1005397. https://doi.org/10.1371/journal.pcbi.1005397.

Churcher, Thomas S, Robert E Sinden, Nick J Edwards, Ian D Poulton, Thomas W Rampling, Patrick M Brock, Jamie T Griffin, et al. 2017. "Probability of Transmission of Malaria from Mosquito to Human Is Regulated by Mosquito Parasite Density in Naïve and Vaccinated Hosts." *PLoS Pathog.* 13 (1): e1006108. https://doi.org/10.1371/journal.ppat.1006108.

Cobbold, Simon A, Hwa H Chua, Brunda Nijagal, Darren J Creek, Stuart A Ralph, and Malcolm J McConville. 2016. "Metabolic Dysregulation Induced in Plasmodium Falciparum by Dihydroartemisinin and Other Front-Line Antimalarial Drugs." *J. Infect. Dis.* 213 (2). academic.oup.com: 276–86. https://doi.org/10.1093/infdis/jiv372.

Codd, Andrea, Franka Teuscher, Dennis E Kyle, Qin Cheng, and Michelle L Gatton. 2011. "Artemisinin-Induced Parasite Dormancy: A Plausible Mechanism for Treatment Failure." *Malar. J.* 10 (March). malariajournal.biomedcentral.com: 56. https:

//doi.org/10.1186/1475-2875-10-56.

Combrinck, Jill M, Egan Joanne, Giovanni R Hearne, Helder M Marques, Skhumbuzo Ntenteni, B Trevor Sewell, Peter J Smith, et al. 2002. "Fate of Haem Iron in the Malaria Parasite Plasmodium Falciparum." *Biochem. J* 365 (2). Portland Press Limited: 343–47.

Cooke, Brian M, Donna W Buckingham, Fiona K Glenister, Kate M Fernandez, Lawrence H Bannister, Matthias Marti, Narla Mohandas, and Ross L Coppel. 2006. "A Maurer's Cleft–Associated Protein Is Essential for Expression of the Major Malaria Virulence Antigen on the Surface of Infected Red Blood Cells." *J. Cell Biol.* 172 (6). Rockefeller University Press: 899–908. https://doi.org/10.1083/jcb.200509122.

Cooper, K D, J B Shukla, and O M Rennert. 1976. "Polyamine Distribution in Cellular Compartments of Blood and in Aging Erythrocytes." *Clin. Chim. Acta* 73 (1): 71–88.

Coppens, Isabelle, and Keith A Joiner. 2001. "Parasite–Host Cell Interactions in Toxoplasmosis: New Avenues for Intervention?" *Expert Rev. Mol. Med.* 3 (2). Cambridge University Press: 1–20. https://doi.org/10.1017/S1462399401002277.

Cowman, Alan F, Julie Healer, Danushka Marapana, and Kevin Marsh. 2016. "Malaria: Biology and Disease." *Cell* 167 (3). Elsevier: 610–24. https://doi.org/10.1016/j.cell.2016.07.055.

Cox, Francis Eg. 2010. "History of the Discovery of the Malaria Parasites and Their Vectors." *Parasit. Vectors* 3 (1): 5. https://doi.org/10.1186/1756-3305-3-5.

Craig, A, and A Scherf. 2001. "Molecules on the Surface of the Plasmodium Falciparum Infected Erythrocyte and Their Role in Malaria Pathogenesis and Immune Evasion." *Mol. Biochem. Parasitol.* 115 (2). Elsevier: 129–43.

Crawford, Michael J, Nadine Thomsen-Zieger, Manisha Ray, Joachim Schachtner, David S Roos, and Frank Seeber. 2006. "Toxoplasma Gondii Scavenges Host-derived Lipoic Acid Despite Its de Novo Synthesis in the Apicoplast." *EMBO J.* 25 (13). EMBO Press: 3214–22. https://doi.org/10.1038/sj.emboj.7601189.

Creek, Darren J, Hwa H Chua, Simon A Cobbold, Brunda Nijagal, James I MacRae, Benjamin K Dickerman, Paul R Gilson, Stuart A Ralph, and Malcolm J McConville. 2016. "Metabolomics-Based Screening of the Malaria Box Reveals Both Novel and Established Mechanisms of Action." *Antimicrob. Agents Chemother.* 60 (11). Am Soc Microbiol: 6650–63. https://doi.org/10.1128/AAC.01226-16.

Cyrklaff, Marek, Cecilia P Sanchez, Nicole Kilian, Cyrille Bisseye, Jacques Simpore, Friedrich Frischknecht, and Michael Lanzer. 2011. "Hemoglobins S and C Interfere

with Actin Remodeling in Plasmodium Falciparum–Infected Erythrocytes." *Science* 334 (6060). American Association for the Advancement of Science: 1283–6. https://doi.org/10.1126/science.1213775.

Dahl, Erica L, Jennifer L Shock, Bhaskar R Shenai, Jiri Gut, Joseph L DeRisi, and Philip J Rosenthal. 2006. "Tetracyclines Specifically Target the Apicoplast of the Malaria Parasite Plasmodium Falciparum." *Antimicrob. Agents Chemother.* 50 (9): 3124–31. https://doi.org/10.1128/AAC.00394-06.

Dalal, Seema, and Michael Klemba. 2015. "Amino Acid Efflux by Asexual Blood-Stage Plasmodium Falciparum and Its Utility in Interrogating the Kinetics of Hemoglobin Endocytosis and Catabolism in Vivo." *Mol. Biochem. Parasitol.* 201 (2). Elsevier: 116–22. https://doi.org/10.1016/j.molbiopara.2015.07.002.

Das, A K. 2015. "Anticancer Effect of AntiMalarial Artemisinin Compounds." *Ann. Med. Health Sci. Res.* 5 (2): 93–102. https://doi.org/10.4103/2141-9248.153609.

Divo, A A, T G Geary, N L Davis, and J B Jensen. 1985. "Nutritional Requirements of Plasmodium Falciparum in Culture. I. Exogenously Supplied Dialyzable Components Necessary for Continuous Growth." *J. Protozool.* 32 (1). Wiley Online Library: 59–64.

Doolan, Denise L, Carlota Dobaño, and J Kevin Baird. 2009. "Acquired Immunity to Malaria." *Clin. Microbiol. Rev.* 22 (1). Am Soc Microbiol: 13–36, Tableof Contents. https://doi.org/10.1128/CMR.00025-08.

Downing, Tim, Hideo Imamura, Saskia Decuypere, Taane G Clark, Graham H Coombs, James A Cotton, James D Hilley, et al. 2011. "Whole Genome Sequencing of Multiple Leishmania Donovani Clinical Isolates Provides Insights into Population Structure and Mechanisms of Drug Resistance." *Genome Res.* 21 (12). genome.cshlp.org: 2143–56. https://doi.org/10.1101/gr.123430.111.

Duarte, Natalie C, Markus J Herrgård, and Bernhard Ø Palsson. 2004. "Reconstruction and Validation of Saccharomyces Cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model." *Genome Res.* 14 (7): 1298–1309. https://doi.org/10.1101/gr.2250904.

Eastman, Richard T, and David A Fidock. 2009. "Artemisinin-Based Combination Therapies: A Vital Tool in Efforts to Eliminate Malaria." *Nat. Rev. Microbiol.* 7 (12). nature.com: 864–74. https://doi.org/10.1038/nrmicro2239.

Ebrahim, Ali, Joshua A Lerman, Bernhard O Palsson, and Daniel R Hyduke. 2013. "COBRApy: COnstraints-Based Reconstruction and Analysis for Python." *BMC Syst. Biol.* 7 (August). bmcsystbiol.biomedcentral.com: 74. https://doi.org/10.1186/1752-0509-7-74.

Eckstein-Ludwig, U, R J Webb, I D A Van Goethem, J M East, A G Lee, M Kimura, P M O'Neill, P G Bray, S A Ward, and S Krishna. 2003. "Artemisinins Target the SERCA of Plasmodium Falciparum." *Nature* 424 (6951). nature.com: 957–61. https://doi.org/10.1038/nature01813.

Efferth, Thomas, and Franz Oesch. 2004. "Oxidative Stress Response of Tumor Cells: Microarray-Based Comparison Between Artemisinins and Anthracyclines." *Biochem. Pharmacol.* 68 (1). Elsevier: 3–10. https://doi.org/10.1016/j.bcp.2004.03.003.

Ejigu, Bedilu Alamirie, Dirk Valkenborg, Geert Baggerman, Manu Vanaerschot, Erwin Witters, Jean-Claude Dujardin, Tomasz Burzykowski, and Maya Berg. 2013. "Evaluation of Normalization Methods to Pave the Way Towards Large-Scale LC-MS-based Metabolomics Profiling Experiments." *OMICS* 17 (9). online.liebertpub.com: 473–85. https://doi.org/10.1089/omi.2013.0010.

El-Halfawy, Omar M, and Miguel A Valvano. 2012. "Non-Genetic Mechanisms Communicating Antibiotic Resistance: Rethinking Strategies for Antimicrobial Drug Design." *Expert Opin. Drug Discov.* 7 (10). Taylor & Francis: 923–33. https://doi.org/10.1517/17460441.2012.712512.

Evans, Anne M, Corey D DeHaven, Tom Barrett, Matt Mitchell, and Eric Milgram. 2009. "Integrated, Nontargeted Ultrahigh Performance Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry Platform for the Identification and Relative Quantification of the Small-Molecule Complement of Biological Systems." *Anal. Chem.* 81 (16). ACS Publications: 6656–67.

Fairhurst, Rick M, Hisashi Fujioka, Karen Hayton, Kathleen F Collins, and Thomas E Wellems. 2003. "Aberrant Development of Plasmodium Falciparum in Hemoglobin CC Red Cells: Implications for the Malaria Protective Effect of the Homozygous State." *Blood* 101 (8): 3309–15. https://doi.org/10.1182/blood-2002-10-3105.

Fang, Xin, Jaques Reifman, and Anders Wallqvist. 2014. "Modeling Metabolism and Stage-Specific Growth of Plasmodium Falciparum HB3 During the Intraerythrocytic Developmental Cycle." *Mol. Biosyst.* 10 (10). pubs.rsc.org: 2526–37. https://doi.org/10.1039/c4mb00115j.

Färber, P M, K Becker, S Müller, R H Schirmer, and R M Franklin. 1996. "Molecular Cloning and Characterization of a Putative Glutathione Reductase Gene, the PfGR2 Gene, from Plasmodium Falciparum." *Eur. J. Biochem.* 239 (3). Wiley Online Library: 655–61.

Färnert, Anna, Johan Lindberg, Pedro Gil, Göte Swedberg, Yngve Berqvist, Mita M Thapar, Niklas Lindegårdh, Sándor Berezcky, and A Björkman. 2003. "Evidence of Plasmodium Falciparum Malaria Resistant to Atovaquone and Proguanil Hydrochlo-

ride: Case Reports." *BMJ* 326 (7390): 628–29. https://doi.org/10.1136/bmj.326.7390.628.

Fichera, M E, and D S Roos. 1997. "A Plastid Organelle as a Drug Target in Apicomplexan Parasites." *Nature* 390 (6658): 407–9. https://doi.org/10.1038/37132.

Fidock, D A, T Nomura, A K Talley, R A Cooper, S M Dzekunov, M T Ferdig, L M Ursos, et al. 2000. "Mutations in the P. Falciparum Digestive Vacuole Transmembrane Protein PfCRT and Evidence for Their Role in Chloroquine Resistance." *Mol. Cell* 6 (4). Elsevier: 861–71.

Flegr, Jaroslav, Joseph Prandota, Michaela Sovičková, and Zafar H Israili. 2014. "Toxoplasmosis–a Global Threat. Correlation of Latent Toxoplasmosis with Specific Disease Burden in a Set of 88 Countries." *PLoS One* 9 (3). Public Library of Science: e90203.

Fleige, Tobias, Nils Pfaff, Uwe Gross, and Wolfgang Bohne. 2008. "Localisation of Gluconeogenesis and Tricarboxylic Acid (TCA)-cycle Enzymes and First Functional Analysis of the TCA Cycle in Toxoplasma Gondii." *Int. J. Parasitol.* 38 (10): 1121–32. https://doi.org/10.1016/j.ijpara.2008.01.007.

Fox, Barbara A, and David J Bzik. 2002. "De Novo Pyrimidine Biosynthesis Is Required for Virulence of Toxoplasma Gondii." *Nature* 415 (6874): 926–29. https://doi.org/10.1038/415926a.

Fox, Barbara A, Jason P Gigley, and David J Bzik. 2004. "Toxoplasma Gondii Lacks the Enzymes Required for de Novo Arginine Biosynthesis and Arginine Starvation Triggers Cyst Formation." *Int. J. Parasitol.* 34 (3): 323–31. https://doi.org/10.1016/j.ijpara.2003.12.001.

Francis, S E, D J Sullivan, and D E Goldberg. 1997. "Hemoglobin Metabolism in the Malaria Parasite Plasmodium Falciparum." *Annu. Rev. Microbiol.* 51: 97–123. https://doi.org/10.1146/annurev.micro.51.1.97.

França, Tanos C C, Pedro G Pascutti, Teodorico C Ramalho, and Jose D Figueroa-Villar. 2005. "A Three-Dimensional Structure of Plasmodium Falciparum Serine Hydroxymethyltransferase in Complex with Glycine and 5-Formyl-Tetrahydrofolate. Homology Modeling and Molecular Dynamics." *Biophys. Chem.* 115 (1). Elsevier: 1–10. https://doi.org/10.1016/j.bpc.2004.12.002.

Fry, Mitchell, and Mary Pudney. 1992. "Site of Action of the Antimalarial Hydroxynaphthoquinone, 2-[Trans-4-(4'-Chlorophenyl) Cyclohexyl]-3- Hydroxy-1,4-Naphthoquinone (566C80)." *Biochem. Pharmacol.* 43 (7): 1545–53. https://doi.org/10.1016/0006-2952(92)90213-3.

Fuhrer, Tobias, Mattia Zampieri, Daniel C Sévin, Uwe Sauer, and Nicola Zamboni. 2017. "Genomewide Landscape of Gene–Metabolome Associations in Escherichia Coli." *Mol. Syst. Biol.* 13 (1). EMBO Press: 907. https://doi.org/10.15252/msb.20167150.

Gabryszewski, Stanislaw J, Satish K Dhingra, Jill M Combrinck, Ian A Lewis, Paul S Callaghan, Matthew R Hassett, Amila Siriwardana, et al. 2016. "Evolution of Fitness Cost-Neutral Mutant PfCRT Conferring P. Falciparum 4-Aminoquinoline Drug Resistance Is Accompanied by Altered Parasite Metabolism and Digestive Vacuole Physiology." *PLoS Pathog.* 12 (11): e1005976. https://doi.org/10.1371/journal.ppat.1005976.

Gajria, Bindu, Amit Bahl, John Brestelli, Jennifer Dommer, Steve Fischer, Xin Gao, Mark Heiges, et al. 2008. "ToxoDB: An Integrated Toxoplasma Gondii Database Resource." *Nucleic Acids Res.* 36 (Database issue). academic.oup.com: D553–6. https://doi.org/10.1093/nar/gkm981.

Garcia, C R, R P Markus, and L Madeira. 2001. "Tertian and Quartan Fevers: Temporal Regulation in Malarial Infection." *J. Biol. Rhythms* 16 (5). journals.sagepub.com: 436–43. https://doi.org/10.1177/074873001129002114.

Gardner, Malcolm J, Neil Hall, Eula Fung, Owen White, Matthew Berriman, Richard W Hyman, Jane M Carlton, et al. 2002. "Genome Sequence of the Human Malaria Parasite Plasmodium Falciparum." *Nature* 419 (6906). nature.com: 498–511. https://doi.org/10.1038/nature01097.

Gasteiger, Elisabeth, Alexandre Gattiker, Christine Hoogland, Ivan Ivanyi, Ron D Appel, and Amos Bairoch. 2003. "ExPASy: The Proteomics Server for in-Depth Protein Knowledge and Analysis." *Nucleic Acids Res.* 31 (13). academic.oup.com: 3784–8.

Geary, T G, A A Divo, L C Bonanni, and J B Jensen. 1985. "Nutritional Requirements of Plasmodium Falciparum in Culture. III. Further Observations on Essential Nutrients and Antimetabolites." *J. Protozool.* 32 (4). Wiley Online Library: 608–13.

Geary, T G, A A Divo, and J B Jensen. 1985. "Nutritional Requirements of Plasmodium Falciparum in Culture. II. Effects of Antimetabolites in a Semi-Defined Medium." *J. Protozool.* 32 (1). Wiley Online Library: 65–69.

Gehlenborg, Nils. 2017. "UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets."

Genesee, Fred. 1994. "Integrating Language and Content: Lessons from Immersion." National Center for Research on Cultural Diversity; Second Language Learning; Center for Applied Linguistics.

———. 2014. “Second Language Learning in School Settings: Lessons from Immersion.” In *Bilingualism, Multiculturalism, and Second Language Learning: The McGill Conference in Honour of Wallace E. Lambert*, edited by Allan G Reynolds, 20. Psychology Press.

Gerrits, J, H Eidhof, J W I Brunnekreeft, and J Hessels. 1997. “[9] Determination of Thiamin and Thiamin Phosphates in Whole Blood by Reversed-Phase Liquid Chromatography with Precolumn Derivatization.” In *Methods in Enzymology*, 279:74–82. Academic Press. https://doi.org/10.1016/S0076-6879(97)79011-6.

Ghorbal, Mehdi, Molly Gorman, Cameron Ross Macpherson, Rafael Miyazawa Martins, Artur Scherf, and Jose-Juan Lopez-Rubio. 2014. “Genome Editing in the Human Malaria Parasite Plasmodium Falciparum Using the CRISPR-Cas9 System.” *Nat. Biotechnol.* 32 (8). nature.com: 819–21. https://doi.org/10.1038/nbt.2925.

Giacometti, A, O Cirioni, and G Scalise. 1996. “In-Vitro Activity of Macrolides Alone and in Combination with Artemisin, Atovaquone, Dapsone, Minocycline or Pyrimethamine Against Cryptosporidium Parvum.” *J. Antimicrob. Chemother.* 38 (3): 399–408.

Ginsburg, Hagai. 2002. “Abundant Proton Pumping in Plasmodium Falciparum, but Why?” *Trends Parasitol.* 18 (11). Elsevier: 483–86.

———. 2006. “Progress in in Silico Functional Genomics: The Malaria Metabolic Pathways Database.” *Trends Parasitol.* 22 (6). Elsevier: 238–40. https://doi.org/10.1016/j.pt.2006.04.008.

Ginsburg, H, M Krugliak, O Eidelman, and Z I Cabantchik. 1983. “New Permeability Pathways Induced in Membranes of Plasmodium Falciparum Infected Erythrocytes.” *Mol. Biochem. Parasitol.* 8 (2). Elsevier: 177–90.

Golenser, Jacob, Judith H Waknine, Miriam Krugliak, Nicholas H Hunt, and Georges E Grau. 2006. “Current Perspectives on the Mechanism of Action of Artemisinins.” *Int. J. Parasitol.* 36 (14). Elsevier: 1427–41. https://doi.org/10.1016/j.ijpara.2006.07.011.

Gonçalves, Daniel, and Patrick Hunziker. 2016. “Transmission-Blocking Strategies: The Roadmap from Laboratory Bench to the Community.” *Malar. J.* 15 (February): 95. https://doi.org/10.1186/s12936-016-1163-3.

Goodman, Christopher D, Josephine E Siregar, Vanessa Mollard, Joel Vega-Rodríguez, Din Syafruddin, Hiroyuki Matsuoka, Motomichi Matsuzaki, et al. 2016. “Parasites Resistant to the Antimalarial Atovaquone Fail to Transmit by Mosquitoes.” *Science* 352 (6283): 349–53. https://doi.org/10.1126/science.aad9279.

Grellier, P, A Valentin, V Millerioux, J Schrevel, and D Rigomier. 1994. “3-Hydroxy-

3-Methylglutaryl Coenzyme a Reductase Inhibitors Lovastatin and Simvastatin Inhibit in Vitro Development of Plasmodium Falciparum and Babesia Divergens in Human Erythrocytes." *Antimicrob. Agents Chemother.* 38 (5): 1144–8. https://doi.org/10.1128/AAC.38.5.1144.

Grüring, Christof, Arlett Heiber, Florian Kruse, Johanna Ungefehr, Tim-Wolf Gilberger, and Tobias Spielmann. 2011. "Development and Host Cell Modifications of Plasmodium Falciparum Blood Stages in Four Dimensions." *Nat. Commun.* 2 (January). nature.com: 165. https://doi.org/10.1038/ncomms1169.

Gubler, D J. 1998. "Resurgent Vector-Borne Diseases as a Global Health Problem." *Emerg. Infect. Dis.* 4 (3): 442–50. https://doi.org/10.3201/eid0403.980326.

Gudmundsson, Steinn, and Ines Thiele. 2010. "Computationally Efficient Flux Variability Analysis." *BMC Bioinformatics* 11 (September). bmcbioinformatics: 489. https://doi.org/10.1186/1471-2105-11-489.

Gulati, Sonia, Eric H Ekland, Kelly V Ruggles, Robin B Chan, Bamini Jayabalasingham, Bowen Zhou, Pierre-Yves Mantel, et al. 2015. "Profiling the Essential Nature of Lipid Metabolism in Asexual Blood and Gametocyte Stages of Plasmodium Falciparum." *Cell Host Microbe* 18 (3). Elsevier: 371–81. https://doi.org/10.1016/j.chom.2015.08.003.

Guler, Jennifer L, John White 3rd, Margaret A Phillips, and Pradipsinh K Rathod. 2015. "Atovaquone Tolerance in Plasmodium Falciparum Parasites Selected for High-Level Resistance to a Dihydroorotate Dehydrogenase Inhibitor." *Antimicrob. Agents Chemother.* 59 (1). Am Soc Microbiol: 686–89. https://doi.org/10.1128/AAC.02347-14.

Gutteridge, W E, and P I Trigg. 1970. "Incorporation of Radioactive Precursors into DNA and RNA of Plasmodium Knowlesi in Vitro." *J. Protozool.* 17 (1): 89–96.

Guzdial, Mark. 2004. "Programming Environments for Novices." *Computer Science Education Research* 2004. books.google.com: 127–54.

Haanstra, Jurgen R, Albert Gerding, Amalia M Dolga, Freek J H Sorgdrager, Manon Buist-Homan, François du Toit, Klaas Nico Faber, et al. 2017. "Targeting Pathogen Metabolism Without Collateral Damage to the Host." *Sci. Rep.* 7 (January): 40406. https://doi.org/10.1038/srep40406.

Haddock, Steven Harold David, and Casey W Dunn. 2011. *Practical Computing for Biologists*. Sinauer Associates Sunderland, MA.

Hallett, Rachel L, Samuel Dunyo, Rosalynn Ord, Musa Jawara, Margaret Pinder, Anna Randall, Ali Alloueche, et al. 2006. "Chloroquine/Sulphadoxine-Pyrimethamine for

Gambian Children with Malaria: Transmission to Mosquitoes of Multidrug-Resistant Plasmodium Falciparum." *PLoS Clin. Trials* 1 (3): e15. https://doi.org/10.1371/journal.pctr.0010015.

Hartai, Zsuzsanna, Peter Klivenyi, Tamas Janaky, Botond Penke, Laszlo Dux, and Laszlo Vecsei. 2005. "Kynurenine Metabolism in Plasma and in Red Blood Cells in Parkinson's Disease." *J. Neurol. Sci.* 239 (1). jns-journal.com: 31–35. https://doi.org/10.1016/j.jns.2005.07.006.

Hassett, Matthew R, and Paul D Roepe. 2018. "PIK-ing New Malaria Chemotherapy." *Trends Parasitol.*, June. Elsevier. https://doi.org/10.1016/j.pt.2018.06.003.

Hayward, Rhys, Kevin J Saliba, and Kiaran Kirk. 2005. "Pfmdr1 Mutations Associated with Chloroquine Resistance Incur a Fitness Cost in Plasmodium Falciparum." *Mol. Microbiol.* 55 (4). Wiley Online Library: 1285–95. https://doi.org/10.1111/j.1365-2958.2004.04470.x.

Heavner, Benjamin D, and Nathan D Price. 2015. "Comparative Analysis of Yeast Metabolic Network Models Highlights Progress, Opportunities for Metabolic Reconstruction." *PLoS Comput. Biol.* 11 (11). journals.plos.org: e1004530. https://doi.org/10.1371/journal.pcbi.1004530.

Helmy, Mohamed, Alexander Crits-Christoph, and Gary D Bader. 2016. "Ten Simple Rules for Developing Public Biological Databases." *PLoS Comput. Biol.* 12 (11). journals.plos.org: e1005128. https://doi.org/10.1371/journal.pcbi.1005128.

Heo, Inha, Devanjali Dutta, Deborah A Schaefer, Nino Iakobachvili, Benedetta Artegiani, Norman Sachs, Kim E Boonekamp, et al. 2018. "Modelling Cryptosporidium Infection in Human Small Intestinal and Lung Organoids." *Nat Microbiol* 3 (7). nature.com: 814–23. https://doi.org/10.1038/s41564-018-0177-8.

Herman, Jonathan D, Daniel P Rice, Ulf Ribacke, Jacob Silterra, Amy A Deik, Eli L Moss, Kate M Broadbent, et al. 2014. "A Genomic and Evolutionary Approach Reveals Non-Genetic Drug Resistance in Malaria." *Genome Biol.* 15 (11). genomebiology.biomedcentral.com: 511. https://doi.org/10.1186/PREACCEPT-1067113631444973.

Hill, D, and J P Dubey. 2002. "Toxoplasma Gondii: Transmission, Diagnosis and Prevention." *Clin. Microbiol. Infect.* 8 (10): 634–40. https://doi.org/10.1046/j.1469-0691.2002.00485.x.

Hill, D E, and J P Dubey. 2018. "Toxoplasma Gondii." In *Foodborne Parasites*, edited by Ynés R Ortega and Charles R Sterling, 119–38. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-67664-7/_6.

Ho, Meng-Chiao, María B Cassera, Dennis C Madrid, Li-Min Ting, Peter C Tyler, Kami Kim, Steven C Almo, and Vern L Schramm. 2009. "Structural and Metabolic Specificity of Methylthiocoformycin for Malarial Adenosine Deaminases." *Biochemistry* 48 (40). ACS Publications: 9618–26. https://doi.org/10.1021/bi9012484.

Ho, M, and N J White. 1999. "Molecular Mechanisms of Cytoadherence in Malaria." *Am. J. Physiol.* 276 (6 Pt 1). physiology.org: C1231–42.

Hoffmann, Reinhard, Thomas Seidl, and Martin Dugas. 2002. "Profound Effect of Normalization on Detection of Differentially Expressed Genes in Oligonucleotide Microarray Data Analysis." *Genome Biol.* 3 (7). genomebiology.biomedcentral.com: RESEARCH0033.

Hortua Triana, Miryam Andrea, Daniela Cajiao Herrera, Barbara H Zimmermann, Barbara A Fox, and David J Bzik. 2016. "Pyrimidine Pathway-Dependent and -Independent Functions of the Toxoplasma Gondii Mitochondrial Dihydroorotate Dehydrogenase." *Infect. Immun.* 84 (10): 2974–81. https://doi.org/10.1128/IAI.00187-16.

Hsiao, L L, R J Howard, M Aikawa, and T F Taraschi. 1991. "Modification of Host Cell Membrane Lipid Composition by the Intra-Erythrocytic Human Malaria Parasite Plasmodium Falciparum." *Biochem. J* 274 (Pt 1) (February): 121–32.

Huang, Jinling, Nandita Mullapudi, Cheryl A Lancto, Marla Scott, Mitchell S Abrahamsen, and Jessica C Kissinger. 2004. "Phylogenomic Evidence Supports Past Endosymbiosis, Intracellular and Horizontal Gene Transfer in Cryptosporidium Parvum." *Genome Biol.* 5 (11): R88. https://doi.org/10.1186/gb-2004-5-11-r88.

Huijsduijnen, Rob Hooft van, R Kiplin Guy, Kelly Chibale, Richard K Haynes, Ingmar Peitz, Gerhard Kelter, Margaret A Phillips, Jonathan L Vennerstrom, Yongyuth Yuthavong, and Timothy N C Wells. 2013. "Anticancer Properties of Distinct Antimalarial Drug Classes." *PLoS One* 8 (12). Public Library of Science: e82962. https://doi.org/10.1371/journal.pone.0082962.

Hyde, John E. 2008. "Fine Targeting of Purine Salvage in Cryptosporidium Parasites." *Trends Parasitol.* 24 (8): 336–39. https://doi.org/10.1016/j.pt.2008.05.001.

Idro, Richard, Kevin Marsh, Chandy C John, and Charles R J Newton. 2010. "Cerebral Malaria: Mechanisms of Brain Injury and Strategies for Improved Neurocognitive Outcome." *Pediatr. Res.* 68 (4): 267–74. https://doi.org/10.1203/PDR.0b013e3181eee738.

Imlay, Leah, and Audrey R Odom. 2014. "Isoprenoid Metabolism in Apicomplexan Parasites." *Curr Clin Microbiol Rep* 1 (3-4). Springer: 37–50. https://doi.org/10.

1007/s40588-014-0006-7.

Ismail, Hanafy M, Victoria Barton, Matthew Phanchana, Sitthivut Charoensutthivarakul, Michael H L Wong, Janet Hemingway, Giancarlo A Biagini, Paul M O'Neill, and Stephen A Ward. 2016. "Artemisinin Activity-Based Probes Identify Multiple Molecular Targets Within the Asexual Stage of the Malaria Parasites Plasmodium Falciparum 3D7." *Proc. Natl. Acad. Sci. U. S. A.* 113 (8). National Acad Sciences: 2080–5. https://doi.org/10.1073/pnas.1600459113.

Isozumi, Rie, Haruki Uemura, Isao Kimata, Yoshio Ichinose, John Logedi, Ahmeddin H Omar, and Akira Kaneko. 2015. "Novel Mutations in K13 Propeller Gene of Artemisinin-Resistant Plasmodium Falciparum." *Emerg. Infect. Dis.* 21 (3). ncbi.nlm.nih.gov: 490–92. https://doi.org/10.3201/eid2103.140898.

Ivens, Alasdair C, Christopher S Peacock, Elizabeth A Worthey, Lee Murphy, Gautam Aggarwal, Matthew Berriman, Ellen Sisk, et al. 2005. "The Genome of the Kinetoplastid Parasite, Leishmania Major." *Science* 309 (5733). science.sciencemag.org: 436–42. https://doi.org/10.1126/science.1112680.

Janse, C J, P F van der Klooster, H J van der Kaay, M van der Ploeg, and J P Overdulve. 1986. "DNA Synthesis in Plasmodium Berghei During Asexual and Sexual Development." *Mol. Biochem. Parasitol.* 20 (2). Elsevier: 173–82.

Jensen, Paul A, and Jason A Papin. 2011. "Functional Integration of a Metabolic Network Model and Expression Data Without Arbitrary Thresholding." *Bioinformatics* 27 (4). academic.oup.com: 541–47. https://doi.org/10.1093/bioinformatics/btq702.

Jomaa, Hassan, Jochen Wiesner, Silke Sanderbrand, Boran Altincicek, Claus Weidemeyer, Martin Hintz, Ivana Türbachova, et al. 1999. "Inhibitors of the Nonmevalonate Pathway of Isoprenoid Biosynthesis as Antimalarial Drugs." *Science* 285 (5433): 1573–6. https://doi.org/10.1126/science.285.5433.1573.

Kalanon, Ming, and Geoffrey I McFadden. 2010. "Malaria, Plasmodium Falciparum and Its Apicoplast." *Biochem. Soc. Trans.* 38 (3): 775–82. https://doi.org/10.1042/BST0380775.

Kamau, Edwin, Susana Campino, Lucas Amenga-Etego, Eleanor Drury, Deus Ishengoma, Kimberly Johnson, Dieudonne Mumba, et al. 2015. "K13-Propeller Polymorphisms in Plasmodium Falciparum Parasites from Sub-Saharan Africa." *J. Infect. Dis.* 211 (8). academic.oup.com: 1352–5. https://doi.org/10.1093/infdis/jiu608.

Kanehisa, Minoru, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. 2016. "KEGG as a Reference Resource for Gene and Protein Annotation." *Nucleic Acids Res.* 44 (D1). academic.oup.com: D457–62. https://doi.org/10.1093/

nar/gkv1070.

Kassambara, A. 2017. "Ggpubr:'Ggplot2' Based Publication Ready Plots." *R Package Version 0. 1* 6.

Katris, Nicholas J, Giel G van Dooren, Paul J McMillan, Eric Hanssen, Leann Tilley, and Ross F Waller. 2014. "The Apical Complex Provides a Regulated Gateway for Secretion of Invasion Factors in Toxoplasma." *PLoS Pathog.* 10 (4). journals.plos.org: e1004074. https://doi.org/10.1371/journal.ppat.1004074.

Ke, Hangjun, Ian A Lewis, Joanne M Morrisey, Kyle J McLean, Suresh M Ganesan, Heather J Painter, Michael W Mather, Marcelo Jacobs-Lorena, Manuel Llinás, and Akhil B Vaidya. 2015. "Genetic Investigation of Tricarboxylic Acid Metabolism During the Plasmodium Falciparum Life Cycle." *Cell Rep.* 11 (1). Elsevier: 164–74. https://doi.org/10.1016/j.celrep.2015.03.011.

Ke, Hangjun, Paul A Sigala, Kazutoyo Miura, Joanne M Morrisey, Michael W Mather, Jan R Crowley, Jeffrey P Henderson, Daniel E Goldberg, Carole A Long, and Akhil B Vaidya. 2014. "The Heme Biosynthesis Pathway Is Essential for Plasmodium Falciparum Development in Mosquito Stage but Not in Blood Stages." *J. Biol. Chem.* 289 (50). ASBMB: 34827–37. https://doi.org/10.1074/jbc.M114.615831.

Kehr, Sebastian, Nicole Sturm, Stefan Rahlfs, Jude M Przyborski, and Katja Becker. 2010. "Compartmentation of Redox Metabolism in Malaria Parasites." *PLoS Pathog.* 6 (12). journals.plos.org: e1001242. https://doi.org/10.1371/journal.ppat.1001242.

Kenthirapalan, Sanketha, Andrew P Waters, Kai Matuschewski, and Taco W A Kooij. 2016. "Functional Profiles of Orphan Membrane Transporters in the Life Cycle of the Malaria Parasite." *Nat. Commun.* 7 (January). nature.com: 10519. https://doi.org/10.1038/ncomms10519.

Kew, Olen M, Roland W Sutter, Esther M de Gourville, Walter R Dowdle, and Mark A Pallansch. 2005. "Vaccine-Derived Polioviruses and the Endgame Strategy for Global Polio Eradication." *Annu. Rev. Microbiol.* 59. annualreviews.org: 587–635. https://doi.org/10.1146/annurev.micro.58.030603.123625.

Kiara, Steven M, John Okombo, Victor Masseno, Leah Mwai, Isabella Ochola, Steffen Borrmann, and Alexis Nzila. 2009. "In Vitro Activity of Antifolate and Polymorphism in Dihydrofolate Reductase of Plasmodium Falciparum Isolates from the Kenyan Coast: Emergence of Parasites with Ile-164-Leu Mutation." *Antimicrob. Agents Chemother.* 53 (9): 3793–8. https://doi.org/10.1128/AAC.00308-09.

Kim, Hyun Uk, Soo Young Kim, Haeyoung Jeong, Tae Yong Kim, Jae Jong Kim, Hyon E Choy, Kyu Yang Yi, Joon Haeng Rhee, and Sang Yup Lee. 2011. "Integra-

tive Genome-Scale Metabolic Analysis of Vibrio Vulnificus for Drug Targeting and Discovery." *Mol. Syst. Biol.* 7 (January): 460. https://doi.org/10.1038/msb.2010.115.

King, Zachary A, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A Lerman, Ali Ebrahim, Bernhard O Palsson, and Nathan E Lewis. 2016. "BiGG Models: A Platform for Integrating, Standardizing and Sharing Genome-Scale Models." *Nucleic Acids Res.* 44 (D1). academic.oup.com: D515–22. https://doi.org/10.1093/nar/gkv1049.

Klaeger, Susan, Stephanie Heinzlmeir, Mathias Wilhelm, Harald Polzer, Binje Vick, Paul-Albert Koenig, Maria Reinecke, et al. 2017. "The Target Landscape of Clinical Kinase Drugs." *Science* 358 (6367). science.sciencemag.org. https://doi.org/10.1126/science.aan4368.

Klonis, Nectarios, Maria P Crespo-Ortiz, Iveta Bottova, Nurhidanatasha Abu-Bakar, Shannon Kenny, Philip J Rosenthal, and Leann Tilley. 2011. "Artemisinin Activity Against Plasmodium Falciparum Requires Hemoglobin Uptake and Digestion." *Proc. Natl. Acad. Sci. U. S. A.* 108 (28). National Acad Sciences: 11405–10. https://doi.org/10.1073/pnas.1104063108.

Klonis, Nectarios, Stanley C Xie, James M McCaw, Maria P Crespo-Ortiz, Sophie G Zaloumis, Julie A Simpson, and Leann Tilley. 2013. "Altered Temporal Response of Malaria Parasites Determines Differential Sensitivity to Artemisinin." *Proc. Natl. Acad. Sci. U. S. A.* 110 (13). National Acad Sciences: 5157–62. https://doi.org/10.1073/pnas.1217452110.

Kohl, Stefanie M, Matthias S Klein, Jochen Hochrein, Peter J Oefner, Rainer Spang, and Wolfram Gronwald. 2012. "State-of-the Art Data Normalization Methods Improve NMR-based Metabolomic Analysis." *Metabolomics* 8 (Suppl 1). Springer: 146–60. https://doi.org/10.1007/s11306-011-0350-z.

Korich, D G, J R Mead, M S Madore, N A Sinclair, and C R Sterling. 1990. "Effects of Ozone, Chlorine Dioxide, Chlorine, and Monochloramine on Cryptosporidium Parvum Oocyst Viability." *Appl. Environ. Microbiol.* 56 (5): 1423–8.

Korpe, Poonum S, Rashidul Haque, Carol Gilchrist, Cristian Valencia, Feiyang Niu, Miao Lu, Jennie Z Ma, et al. 2016. "Natural History of Cryptosporidiosis in a Longitudinal Study of Slum-Dwelling Bangladeshi Children: Association with Severe Malnutrition." *PLoS Negl. Trop. Dis.* 10 (5): e0004564.

Kritsiriwuthinan, Kanyanan, Sastra Chaotheing, Philip J Shaw, Chayaphat Wongsombat, Porntip Chavalitshewinkoon-Petmitr, and Sumalee Kamchonwongpaisan. 2011. "Global Gene Expression Profiling of Plasmodium Falciparum in Response to the Anti-Malarial Drug Pyronaridine." *Malar. J.* 10 (August): 242. https:

//doi.org/10.1186/1475-2875-10-242.

Krugliak, Miriam, Jianmin Zhang, and Hagai Ginsburg. 2002. "Intraerythrocytic Plasmodium Falciparum Utilizes Only a Fraction of the Amino Acids Derived from the Digestion of Host Cell Cytosol for the Biosynthesis of Its Proteins." *Mol. Biochem. Parasitol.* 119 (2). Elsevier: 249–56.

Kumaratilake, L M, A Ferrante, B S Robinson, T Jaeger, and A Poulos. 1997. "Enhancement of Neutrophil-Mediated Killing of Plasmodium Falciparum Asexual Blood Forms by Fatty Acids: Importance of Fatty Acid Structure." *Infect. Immun.* 65 (10). Am Soc Microbiol: 4152–7.

Kwiatkowski, D, and B M Greenwood. 1989. "Why Is Malaria Fever Periodic? A Hypothesis." *Parasitol. Today* 5 (8). cell.com: 264–66.

Kwok, Roberta. 2013. "Computing: Out of the Hood." *Nature* 504 (7479). nature.com: 319–21.

Kyes, S A, J A Rowe, N Kriek, and C I Newbold. 1999. "Rifins: A Second Family of Clonally Variant Proteins Expressed on the Surface of Red Cells Infected with Plasmodium Falciparum." *Proc. Natl. Acad. Sci. U. S. A.* 96 (16). National Acad Sciences: 9333–8.

Lachance, Jean-Christophe, Jonathan M Monk, Colton J Lloyd, Yara Seif, Bernhard O Palsson, Sebastien Rodrigue, Adam M Feist, Zachary A King, and Pierre-Etienne Jacques. 2018. "BOFdat: Generating Biomass Objective Function Stoichiometric Coefficients from Experimental Data." *bioRxiv*. https://doi.org/10.1101/243881.

Lamb, Tracey. 2012. *Immunity to Parasitic Infection*. John Wiley & Sons.

Lambros, C, and J P Vanderberg. 1979. "Synchronization of Plasmodium Falciparum Erythrocytic Stages in Culture." *J. Parasitol.* 65 (3). JSTOR: 418–20.

Langer, R C, and J M Vinetz. 2001. "Plasmodium Ookinete-Secreted Chitinase and Parasite Penetration of the Mosquito Peritrophic Matrix." *Trends Parasitol.* 17 (6). Elsevier: 269–72.

Langreth, S G. 1977. "Electron Microscope Cytochemistry of Host-Parasite Membrane Interactions in Malaria." *Bull. World Health Organ.* 55 (2-3). ncbi.nlm.nih.gov: 171–78.

Langreth, S G, J B Jensen, R T Reese, and W Trager. 1978. "Fine Structure of Human Malaria in Vitro." *J. Protozool.* 25 (4). Wiley Online Library: 443–52.

Lanzer, Michael, Hannes Wickert, Georg Krohne, Laetitia Vincensini, and Catherine Braun Breton. 2006. "Maurer's Clefts: A Novel Multi-Functional Organelle in the

Cytoplasm of Plasmodium Falciparum-Infected Erythrocytes." *Int. J. Parasitol.* 36 (1). Elsevier: 23–36. https://doi.org/10.1016/j.ijpara.2005.10.001.

Lauer, S A, P K Rathod, N Ghori, and K Haldar. 1997. "A Membrane Network for Nutrient Import in Red Cells Infected with the Malaria Parasite." *Science* 276 (5315): 1122–5.

Lee, Marcus Cs, and David A Fidock. 2014. "CRISPR-mediated Genome Editing of Plasmodium Falciparum Malaria Parasites." *Genome Med.* 6 (8). genomemedicine.biomedcentral: 63. https://doi.org/10.1186/s13073-014-0063-9.

Le Roch, Karine G, Jeffrey R Johnson, Laurence Florens, Yingyao Zhou, Andrey Santrosyan, Munira Grainger, S Frank Yan, et al. 2004. "Global Analysis of Transcript and Protein Levels Across the Plasmodium Falciparum Life Cycle." *Genome Res.* 14 (11). genome.cshlp.org: 2308–18. https://doi.org/10.1101/gr.2523904.

Li, Li, Christian J Stoeckert Jr, and David S Roos. 2003. "OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes." *Genome Res.* 13 (9). genome.cshlp.org: 2178–89. https://doi.org/10.1101/gr.1224503.

Li, Wei, Weike Mo, Dan Shen, Libo Sun, Juan Wang, Shan Lu, Jane M Gitschier, and Bing Zhou. 2005. "Yeast Model Uncovers Dual Roles of Mitochondria in the Action of Artemisinin." *PLoS Genet.* 1 (3). Public Library of Science: e36. https://doi.org/10.1371/journal.pgen.0010036.

Liaw, Andy, Matthew Wiener, and Others. 2002. "Classification and Regression by randomForest." *R News* 2 (3). researchgate.net: 18–22.

Lieberman, Linda A, and Christopher A Hunter. 2002. "The Role of Cytokines and Their Signaling Pathways in the Regulation of Immunity to Toxoplasma Gondii." *Int. Rev. Immunol.* 21 (4-5): 373–403.

Lieven, Christian, Moritz Emanuel Beber, Brett G Olivier, Frank T Bergmann, Parizad Babaei, Jennifer A Bartell, Lars M Blank, et al. 2018. "Memote: A Community-Driven Effort Towards a Standardized Genome-Scale Metabolic Model Test Suite." *bioRxiv*. https://doi.org/10.1101/350991.

Lim, Liting, and Geoffrey Ian McFadden. 2010. "The Evolution, Metabolism and Functions of the Apicoplast." *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365 (1541): 749–63. https://doi.org/10.1098/rstb.2009.0273.

Linke, Dirk. 2009. "Commentary: Never Trust Your Word Processor." *Biochem. Mol. Biol. Educ.* 37 (6). Wiley Online Library: 377. https://doi.org/10.1002/bmb.20340.

Liu, J, D T Bolick, G L Kolling, Z Fu, and R L Guerrant. 2016. "Protein Malnutrition

Impairs Intestinal Epithelial Cell Turnover, a Potential Mechanism of Increased Cryptosporidiosis in a Murine Model." *Infect. Immun.* 84 (12): 3542–9.

Liu, Jun, Eva S Istvan, Ilya Y Gluzman, Julia Gross, and Daniel E Goldberg. 2006. "Plasmodium Falciparum Ensures Its Amino Acid Supply with Multiple Acquisition Pathways and Redundant Proteolytic Enzyme Systems." *Proc. Natl. Acad. Sci. U. S. A.* 103 (23). National Acad Sciences: 8840–5. https://doi.org/10.1073/pnas.0601876103.

Llinás, Manuel, Zbynek Bozdech, Edith D Wong, Alex T Adai, and Joseph L DeRisi. 2006. "Comparative Whole Genome Transcriptome Analysis of Three Plasmodium Falciparum Strains." *Nucleic Acids Res.* 34 (4): 1166–73. https://doi.org/10.1093/nar/gkj517.

Loria, P, S Miller, M Foley, and L Tilley. 1999. "Inhibition of the Peroxidative Degradation of Haem as the Basis of Action of Chloroquine and Other Quinoline Antimalarials." *Biochem. J* 339 (Pt 2): 363–70.

Lu, Junnan, Ying Tong, Jiaqiang Pan, Yijun Yang, Quan Liu, Xuefang Tan, Siting Zhao, Li Qin, and Xiaoping Chen. 2016. "A Redesigned CRISPR/Cas9 System for Marker-Free Genome Editing in Plasmodium Falciparum." *Parasit. Vectors* 9 (April). parasitesandvectors.biomedcentral: 198. https://doi.org/10.1186/s13071-016-1487-4.

Luciani, T J, M Decorde, and V Lise. 2009. "RSvgDevice: An R SVG Graphics Device."

Machado, D, S Andrejev, M Tramontano, and K R Patil. 2018. "Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities." *bioRxiv*. biorxiv.org.

MacPherson, G G, M J Warrell, N J White, S Looareesuwan, and D A Warrell. 1985. "Human Cerebral Malaria. A Quantitative Ultrastructural Analysis of Parasitized Erythrocyte Sequestration." *Am. J. Pathol.* 119 (3). ncbi.nlm.nih.gov: 385–401.

MacRae, James I, Matthew Wa Dixon, Megan K Dearnley, Hwa H Chua, Jennifer M Chambers, Shannon Kenny, Iveta Bottova, Leann Tilley, and Malcolm J McConville. 2013. "Mitochondrial Metabolism of Sexual and Asexual Blood Stages of the Malaria Parasite Plasmodium Falciparum." *BMC Biol.* 11 (June). bmcbiol.biomedcentral.com: 67. https://doi.org/10.1186/1741-7007-11-67.

MacRae, James I, Lilach Sheiner, Amsha Nahid, Christopher Tonkin, Boris Striepen, and Malcolm J McConville. 2012. "Mitochondrial Metabolism of Glucose and Glutamine Is Required for Intracellular Growth of Toxoplasma Gondii." *Cell Host Microbe* 12 (5): 682–92. https://doi.org/10.1016/j.chom.2012.09.013.

Maenpuen, Somchart, Kittipat Sopitthummakhun, Yongyuth Yuthavong, Pimchai Chaiyen, and Ubolsree Leartsakulpanich. 2009. "Characterization of Plasmodium Falciparum Serine Hydroxymethyltransferase: A Potential Antimalarial Target." *Mol. Biochem. Parasitol.* 168 (1). Elsevier: 63–73. https://doi.org/10.1016/j.molbiopara.2009.06.010.

Magnúsdóttir, Stefanía, Almut Heinken, Laura Kutt, Dmitry A Ravcheev, Eugen Bauer, Alberto Noronha, Kacy Greenhalgh, et al. 2017. "Generation of Genome-Scale Metabolic Reconstructions for 773 Members of the Human Gut Microbiota." *Nat. Biotechnol.* 35 (1): 81–89. https://doi.org/10.1038/nbt.3703.

Magowan, C, W Wollish, L Anderson, and J Leech. 1988. "Cytoadherence by Plasmodium Falciparum-Infected Erythrocytes Is Correlated with the Expression of a Family of Variable Proteins on Infected Erythrocytes." *J. Exp. Med.* 168 (4). jem.rupress.org: 1307–20.

malERA Consultative Group on Vaccines. 2011. "A Research Agenda for Malaria Eradication: Vaccines." *PLoS Med.* 8 (1): e1000398. https://doi.org/10.1371/journal.pmed.1000398.

Mandal, Swati, Ajeet Mandal, Hans E Johansson, Arturo V Orjalo, and Myung Hee Park. 2013. "Depletion of Cellular Polyamines, Spermidine and Spermine, Causes a Total Arrest in Translation and Growth in Mammalian Cells." *Proc. Natl. Acad. Sci. U. S. A.* 110 (6). National Acad Sciences: 2169–74. https://doi.org/10.1073/pnas.1219002110.

Manjunatha, Ujjini H, Sumiti Vinayak, Jennifer A Zambriski, Alexander T Chao, Tracy Sy, Christian G Noble, Ghislain M C Bonamy, et al. 2017. "A Cryptosporidium PI(4)K Inhibitor Is a Drug Candidate for Cryptosporidiosis." *Nature* 546 (May). Macmillan Publishers Limited, part of Springer Nature. All rights reserved.: 376. https://doi.org/10.1038/nature22337.

Markowetz, Florian. 2017. "All Biology Is Computational Biology." *PLoS Biol.* 15 (3). journals.plos.org: e2002050. https://doi.org/10.1371/journal.pbio.2002050.

Martin, Rowena E, Roselani I Henry, Janice L Abbey, John D Clements, and Kiaran Kirk. 2005. "The'permeome'of the Malaria Parasite: An Overview of the Membrane Transport Proteins of Plasmodium Falciparum." *Genome Biol.* 6 (3). BioMed Central: R26.

Martin, Rowena E, and Kiaran Kirk. 2007. "Transport of the Essential Nutrient Isoleucine in Human Erythrocytes Infected with the Malaria Parasite Plasmodium Falciparum." *Blood* 109 (5): 2217–24. https://doi.org/10.1182/blood-2005-11-026963.

Massimine, Kristen M, Lanxuan T Doan, Chloé A Atreya, Timothy T Stedman, Karen S Anderson, Keith A Joiner, and Isabelle Coppens. 2005. "Toxoplasma Gondii Is Capable of Exogenous Folate Transport. A Likely Expansion of the BT1 Family of Transmembrane Proteins." *Mol. Biochem. Parasitol.* 144 (1): 44–54. https://doi.org/10.1016/j.molbiopara.2005.07.006.

Masum, Hassan, Aarthi Rao, Benjamin M Good, Matthew H Todd, Aled M Edwards, Leslie Chan, Barry A Bunin, Andrew I Su, Zakir Thomas, and Philip E Bourne. 2013. "Ten Simple Rules for Cultivating Open Science and Collaborative R&D." *PLoS Comput. Biol.* 9 (9). journals.plos.org: e1003244. https://doi.org/10.1371/journal.pcbi.1003244.

Mazumdar, Jolly, and Boris Striepen. 2007. "Make It or Take It: Fatty Acid Metabolism of Apicomplexan Parasites." *Eukaryot. Cell* 6 (10). Am Soc Microbiol: 1727–35. https://doi.org/10.1128/EC.00255-07.

Mbanefo, Evaristus Chibunna, Ali Mahmoud Ahmed, Afaf Titouna, Ahmed Elmaraezy, Nguyen Thi Huyen Trang, Nguyen Phuoc Long, Nguyen Hoang Anh, et al. 2017. "Association of Glucose-6-Phosphate Dehydrogenase Deficiency and Malaria: A Systematic Review and Meta-Analysis." *Sci. Rep.* 7 (April): 45963. https://doi.org/10.1038/srep45963.

Mbengue, Alassane, Souvik Bhattacharjee, Trupti Pandharkar, Haining Liu, Guillermina Estiu, Robert V Stahelin, Shahir S Rizk, et al. 2015. "A Molecular Mechanism of Artemisinin Resistance in Plasmodium Falciparum Malaria." *Nature* 520 (7549). nature.com: 683–87. https://doi.org/10.1038/nature14412.

McAuley, James B. 2014. "Congenital Toxoplasmosis." *J Pediatric Infect Dis Soc* 3 Suppl 1 (September): S30–5. https://doi.org/10.1093/jpids/piu077.

McClure, Nathan S, and Troy Day. 2014. "A Theoretical Examination of the Relative Importance of Evolution Management and Drug Development for Managing Resistance." *Proc. Biol. Sci.* 281 (1797). https://doi.org/10.1098/rspb.2014.1861.

McConville, Malcolm. 2014. "Open Questions: Microbes, Metabolism and Host-Pathogen Interactions." *BMC Biol.* 12 (March). bmcbiol.biomedcentral.com: 18. https://doi.org/10.1186/1741-7007-12-18.

McFadden, Diane C, Stanislas Tomavo, Edward A Berry, and John C Boothroyd. 2000. "Characterization of Cytochrome B from Toxoplasma Gondii and Qo Domain Mutations as a Mechanism of Atovaquone-Resistance." *Mol. Biochem. Parasitol.* 108 (1): 1–12. https://doi.org/10.1016/S0166-6851(00)00184-5.

McLeod, R, S P Muench, J B Rafferty, D E Kyle, E J Mui, M J Kirisits, D G Mack, et

al. 2001. “Triclosan Inhibits the Growth of Plasmodium Falciparum and Toxoplasma Gondii by Inhibition of Apicomplexan Fab I.” *Int. J. Parasitol.* 31 (2). Elsevier: 109–13.

McNeil, Donald G, Jr. n.d. “Cosmetic Saves a Cure for Sleeping Sickness.” *The New York Times*.

Mead, J R, M J Arrowood, R W Sidwell, and M C Healey. 1991. “Chronic Cryptosporidium Parvum Infections in Congenitally Immunodeficient SCID and Nude Mice.” *J. Infect. Dis.* 163 (6). JSTOR: 1297–1304.

Medema, G, P Teunis, M Blokker, D Deere, A Davison, Ph Charles, and J F Loret. 2009. “Risk Assessment of Cryptosporidium in Drinking Water.” *World Health Organization*, 143.

Medlock, Gregory Leonard, Maureen A Carey, Dennis G McDuffie, Michael B Mundy, Natasa Giallourou, Jonathan R Swann, Glynis Kolling, and Jason A Papin. 2018. “Metabolic Mechanisms of Interaction Within a Defined Gut Microbiota.” *bioRxiv*, January.

Mehta, Monika, Haripalsingh M Sonawat, and Shobhona Sharma. 2006. “Glycolysis in Plasmodium Falciparum Results in Modulation of Host Enzyme Activities.” *J. Vector Borne Dis.* 43 (3): 95–103.

Menard, Didier, Nestor Madji, Alexandre Manirakiza, Djibrine Djalle, Max Roger Koula, and Antoine Talarmin. 2005. “Efficacy of Chloroquine, Amodiaquine, Sulfadoxine-Pyrimethamine, Chloroquine-Sulfadoxine-Pyrimethamine Combination, and Amodiaquine-Sulfadoxine-Pyrimethamine Combination in Central African Children with Noncomplicated Malaria.” *Am. J. Trop. Med. Hyg.* 72 (5): 581–85.

Meneceur, Pascale, Marie-Anne Bouldouyre, Dominique Aubert, Isabelle Villena, Jean Menotti, Virginie Sauvage, Jean-François Garin, and Francis Derouin. 2008. “In Vitro Susceptibility of Various Genotypic Strains of Toxoplasma Gondii to Pyrimethamine, Sulfadiazine, and Atovaquone.” *Antimicrob. Agents Chemother.* 52 (4): 1269–77. https://doi.org/10.1128/AAC.01203-07.

Meshnick, Steven R. 2002. “Artemisinin: Mechanisms of Action, Resistance and Toxicity.” *Int. J. Parasitol.* 32 (13). Elsevier: 1655–60.

Metz, Jack. 2007. “Folic Acid Metabolism and Malaria.” *Food Nutr. Bull.* 28 (4 Suppl): S540–549. https://doi.org/10.1177/15648265070284S407.

Meylan, Sylvain, Caroline B M Porter, Jason H Yang, Peter Belenky, Arnaud Gutierrez, Michael A Lobritz, Jihye Park, Sun H Kim, Samuel M Moskowitz, and James J Collins. 2017. “Carbon Sources Tune Antibiotic Susceptibility in Pseudomonas Aeruginosa

via Tricarboxylic Acid Cycle Control." *Cell Chem Biol* 24 (2). Elsevier: 195–206. https://doi.org/10.1016/j.chembiol.2016.12.015.

Miller, Louis H, Dror I Baruch, Kevin Marsh, and Ogobara K Doumbo. 2002. "The Pathogenic Basis of Malaria." *Nature* 415 (6872). nature.com: 673–79. https://doi.org/10.1038/415673a.

Minkah, Nana K, Carola Schafer, and Stefan H I Kappe. 2018. "Humanized Mouse Models for the Study of Human Malaria Parasite Biology, Pathogenesis, and Immunity." *Front. Immunol.* 9 (April): 807. https://doi.org/10.3389/fimmu.2018.00807.

Miotto, Olivo, Roberto Amato, Elizabeth A Ashley, Bronwyn MacInnis, Jacob Almagro-Garcia, Chanaki Amaratunga, Pharath Lim, et al. 2015. "Genetic Architecture of Artemisinin-Resistant Plasmodium Falciparum." *Nat. Genet.* 47 (3). nature.com: 226–34. https://doi.org/10.1038/ng.3189.

Mittler, Ron. 2002. "Oxidative Stress, Antioxidants and Stress Tolerance." *Trends Plant Sci.* 7 (9). Elsevier: 405–10.

Mok, Sachel, Elizabeth A Ashley, Pedro E Ferreira, Lei Zhu, Zhaoting Lin, Tomas Yeo, Kesinee Chotivanich, et al. 2015. "Drug Resistance. Population Transcriptomics of Human Malaria Parasites Reveals the Mechanism of Artemisinin Resistance." *Science* 347 (6220). science.sciencemag.org: 431–35. https://doi.org/10.1126/science.1260403.

Mok, Sachel, Mallika Imwong, Margaret J Mackinnon, Joan Sim, Ramya Ramadoss, Poravuth Yi, Mayfong Mayxay, et al. 2011. "Artemisinin Resistance in Plasmodium Falciparum Is Associated with an Altered Temporal Pattern of Transcription." *BMC Genomics* 12 (August). bmcgenomics.biomedcentral.com: 391. https://doi.org/10.1186/1471-2164-12-391.

Moll, K, I Ljungström, H Perlmann, A Scherf, and others. 2008. "Methods in Malaria Research." https://www.researchgate.net/profile/Abhinav_Sinha/publication/312495167_Tight_synchronisation_protocol_for_in_vitro_cultures_of_Plasmodium_falciparum/links/587f233808ae9275d4ebaab5.pdf; researchgate.net.

Mukherjee, Angana, Selina Bopp, Pamela Magistrado, Wesley Wong, Rachel Daniels, Allison Demas, Stephen Schaffner, et al. 2017. "Artemisinin Resistance Without Pfkelch13 Mutations in Plasmodium Falciparum Isolates from Cambodia." *Malar. J.* 16 (1): 195. https://doi.org/10.1186/s12936-017-1845-5.

Müller, Sylke, and Barbara Kappes. 2007. "Vitamin and Cofactor Biosynthesis Pathways in Plasmodium and Other Apicomplexan Parasites." *Trends Parasitol.* 23 (3). Elsevier: 112–21. https://doi.org/10.1016/j.pt.2007.01.009.

Nagababu, Enika, Francis J Chrest, and Joseph M Rifkind. 2003. "Hydrogen-Peroxide-

Induced Heme Degradation in Red Blood Cells: The Protective Roles of Catalase and Glutathione Peroxidase." *Biochimica et Biophysica Acta (BBA) - General Subjects* 1620 (1): 211–17. https://doi.org/10.1016/S0304-4165(02)00537-8.

Nagaraj, Viswanathan Arun, Balamurugan Sundaram, Nandan Mysore Varadarajan, Pradeep Annamalai Subramani, Devaiah Monnanda Kalappa, Susanta Kumar Ghosh, and Govindarajan Padmanaban. 2013. "Malaria Parasite-Synthesized Heme Is Essential in the Mosquito and Liver Stages and Complements Host Heme in the Blood Stages of Infection." *PLoS Pathog.* 9 (8). journals.plos.org: e1003522. https://doi.org/10.1371/journal.ppat.1003522.

Nash, G B, E O'Brien, E C Gordon-Smith, and J A Dormandy. 1989. "Abnormalities in the Mechanical Properties of Red Blood Cells Caused by Plasmodium Falciparum." *Blood* 74 (2). Am Soc Hematology: 855–61.

Niang, Makhtar, Xue Yan Yam, and Peter Rainer Preiser. 2009. "The Plasmodium Falciparum STEVOR Multigene Family Mediates Antigenic Variation of the Infected Erythrocyte." *PLoS Pathog.* 5 (2). journals.plos.org: e1000307. https://doi.org/10.1371/journal.ppat.1000307.

Nzila, Alexis, John Okombo, Ruy Perez Becker, Roma Chilengi, Trudie Lang, and Tim Niehues. 2010. "Anticancer Agents Against Malaria: Time to Revisit?" *Trends Parasitol.* 26 (3): 125–29. https://doi.org/10.1016/j.pt.2009.12.002.

Oberhardt, Matthew A, Joanna B Goldberg, Michael Hogardt, and Jason A Papin. 2010. "Metabolic Network Analysis of Pseudomonas Aeruginosa During Chronic Cystic Fibrosis Lung Infection." *J. Bacteriol.* 192 (20). Am Soc Microbiol: 5534–48. https://doi.org/10.1128/JB.00900-10.

Olliaro, P L, and W R Taylor. 2004. "Developing Artemisinin Based Drug Combinations for the Treatment of Drug Resistant Falciparum Malaria: A Review." *J. Postgrad. Med.* 50 (1). annalsafrmed.org: 40–44.

Olszewski, Kellen L, Michael W Mather, Joanne M Morrisey, Benjamin A Garcia, Akhil B Vaidya, Joshua D Rabinowitz, and Manuel Llinás. 2010. "Branched Tricarboxylic Acid Metabolism in Plasmodium Falciparum." *Nature* 466 (7307). nature.com: 774–78. https://doi.org/10.1038/nature09301.

Olszewski, Kellen L, Joanne M Morrisey, Daniel Wilinski, James M Burns, Akhil B Vaidya, Joshua D Rabinowitz, and Manuel Llinás. 2009. "Host-Parasite Interactions Revealed by Plasmodium Falciparum Metabolomics." *Cell Host Microbe* 5 (2). Elsevier: 191–99. https://doi.org/10.1016/j.chom.2009.01.004.

Ord, Rosalynn, Neal Alexander, Sam Dunyo, Rachel Hallett, Musa Jawara, Geoffrey

Targett, Christopher J Drakeley, and Colin J Sutherland. 2007. "Seasonal Carriage of Pfcrt and Pfmdr1 Alleles in Gambian Plasmodium Falciparum Imply Reduced Fitness of Chloroquine-Resistant Parasites." *J. Infect. Dis.* 196 (11). academic.oup.com: 1613–9. https://doi.org/10.1086/522154.

Orth, Jeffrey D, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard Ø Palsson. 2011. "A Comprehensive Genome Scale Reconstruction of Escherichia Coli Metabolism—2011." *Mol. Syst. Biol.* 7 (1). EMBO Press: 535. https://doi.org/10.1038/msb.2011.65.

Orth, Jeffrey D, Ines Thiele, and Bernhard Ø Palsson. 2010. "What Is Flux Balance Analysis?" *Nat. Biotechnol.* 28 (March). Nature Publishing Group: 245. https://doi.org/10.1038/nbt.1614.

Overbosch, D, H Schilthuis, U Bienzle, R H Behrens, K C Kain, P D Clarke, S Toovey, et al. 2001. "Atovaquone-Proguanil Versus Mefloquine for Malaria Prophylaxis in Nonimmune Travelers: Results from a Randomized, Double-Blind Study." *Clin. Infect. Dis.* 33 (7). academic.oup.com: 1015–21. https://doi.org/10.1086/322694.

Pages, H, P Aboyoun, R Gentleman, and S DebRoy. n.d. "Biostrings: String Objects Representing Biological Sequences, and Matching Algorithms. R Package. 2014."

Paglia, Giuseppe, Angelo D'Alessandro, Ottar Rolfsson, Olafur E Sigurjonsson, Aarash Bordbar, Sirus Palsson, Travis Nemkov, Kirk C Hansen, Sveinn Gudmundsson, and Bernhard O Palsson. n.d. "Biomarkers de Ning the Metabolic Age of Red Blood Cells During Cold Storage." https://doi.org/10.1182/blood-2016-06.

Pagola, S, P W Stephens, D S Bohle, A D Kosar, and S K Madsen. 2000. "The Structure of Malaria Pigment Beta-Haematin." *Nature* 404 (6775): 307–10. https://doi.org/10.1038/35005132.

Painter, Heather J, Joanne M Morrisey, Michael W Mather, and Akhil B Vaidya. 2007. "Specific Role of Mitochondrial Electron Transport in Blood-Stage Plasmodium Falciparum." *Nature* 446 (7131). nature.com: 88–91. https://doi.org/10.1038/nature05572.

Pandey, A V, B L Tekwani, R L Singh, and V S Chauhan. 1999. "Artemisinin, an Endoperoxide Antimalarial, Disrupts the Hemoglobin Catabolism and Heme Detoxification Systems in Malarial Parasite." *J. Biol. Chem.* 274 (27): 19383–8.

Park, Youngja H, Ya Ping Shi, Bill Liang, Carl Angelo D Medriano, Young Ho Jeon, Eucaris Torres, Karan Uppal, Laurence Slutsker, and Dean P Jones. 2015. "High-Resolution Metabolomics to Discover Potential Parasite-Specific Biomarkers in a Plasmodium Falciparum Erythrocytic Stage Culture System." *Malar. J.* 14 (March).

malariajournal.biomedcentral.com: 122. https://doi.org/10.1186/s12936-015-0651-1.

Parvazi, Shirin, Sedigheh Sadeghi, Mehri Azadi, Maryam Mohammadi, Mohammad Arjmand, Farideh Vahabi, Somye Sadeghzadeh, and Zahra Zamani. 2016. "The Effect of Aqueous Extract of Cinnamon on the Metabolome of Plasmodium Falciparum Using 1HNMR Spectroscopy." *J. Trop. Med.* 2016 (January). Hindawi. https://doi.org/10.1155/2016/3174841.

Paul, F, S Roath, D Melville, D C Warhurst, and J O Osisanya. 1981. "Separation of Malaria-Infected Erythrocytes from Whole Blood: Use of a Selective High-Gradient Magnetic Separation Technique." *Lancet* 2 (8237). Elsevier: 70–71.

Payne, D. 1987. "Spread of Chloroquine Resistance in Plasmodium Falciparum." *Parasitol. Today* 3 (8): 241–46.

Peatey, Christopher L, Marina Chavchich, Nanhua Chen, Karryn J Gresty, Karen-Ann Gray, Michelle L Gatton, Norman C Waters, and Qin Cheng. 2015. "Mitochondrial Membrane Potential in a Small Subset of Artemisinin-Induced Dormant Plasmodium Falciparum Parasites in Vitro." *J. Infect. Dis.* 212 (3). academic.oup.com: 426–34. https://doi.org/10.1093/infdis/jiv048.

Pellé, Karell G, Rays H Y Jiang, Pierre-Yves Mantel, Yu-Ping Xiao, Daisy Hjelmqvist, Gina M Gallego-Lopez, Audrey OT Lau, Byung-Ho Kang, David R Allred, and Matthias Marti. 2015. "Shared Elements of Host-Targeting Pathways Among Apicomplexan Parasites of Differing Lifestyles." *Cell. Microbiol.* 17 (11). Wiley Online Library: 1618–39.

Perez-Riverol, Yasset, Laurent Gatto, Rui Wang, Timo Sachsenberg, Julian Uszkoreit, Felipe da Veiga Leprevost, Christian Fufezan, et al. 2016. "Ten Simple Rules for Taking Advantage of Git and GitHub." *PLoS Comput. Biol.* 12 (7). journals.plos.org: e1004947. https://doi.org/10.1371/journal.pcbi.1004947.

Peterson, D S, D Walliker, and T E Wellems. 1988. "Evidence That a Point Mutation in Dihydrofolate Reductase-Thymidylate Synthase Confers Resistance to Pyrimethamine in Falciparum Malaria." *Proc. Natl. Acad. Sci. U. S. A.* 85 (23). National Acad Sciences: 9114–8.

Phaiphinit, Suthat, Sittiporn Pattaradilokrat, Chidchanok Lursinsap, and Kitiporn Plaimas. 2016. "In Silico Multiple-Targets Identification for Heme Detoxification in the Human Malaria Parasite Plasmodium Falciparum." *Infect. Genet. Evol.* 37 (January). Elsevier: 237–44. https://doi.org/10.1016/j.meegid.2015.11.025.

Phillips, Margaret A, and Pradipsinh K Rathod. 2010. "Plasmodium Dihydroorotate Dehydrogenase: A Promising Target for Novel Anti-Malarial Chemotherapy." *Infect.*

Disord. Drug Targets 10 (3). ingentaconnect.com: 226–39.

Plata, German, Tzulin Hsiao, Kellen L Olszewski, Manuel Llinas, and Dennis Vitkup. 2010. "Reconstruction and Flux Balance Analysis of the Plasmodium Falciparum Metabolic Network." *Mol. Syst. Biol.* 6 (1). EMBO Press: 408. https://doi.org/10.1038/msb.2010.60.

Platts-Mills, James A, Sudhir Babji, Ladaporn Bodhidatta, Jean Gratz, Rashidul Haque, Alexandre Havt, Benjamin Jj McCormick, et al. 2015. "Pathogen-Specific Burdens of Community Diarrhoea in Developing Countries: A Multisite Birth Cohort Study (MAL-ED)." *Lancet Glob Health* 3 (9): e564–75. https://doi.org/10.1016/S2214-109X(15)00151-5.

Plowe, Christopher V, Cally Roper, John W Barnwell, Christian T Happi, Hema H Joshi, Wilfred Mbacham, Steven R Meshnick, et al. 2007. "World Antimalarial Resistance Network (WARN) III: Molecular Markers for Drug Resistant Malaria." *Malar. J.* 6 (September). malariajournal.biomedcentral.com: 121. https://doi.org/10.1186/1475-2875-6-121.

Pornthanakasem, Wichai, Darin Kongkasuriyachai, Chairat Uthaipibull, Yongyuth Yuthavong, and Ubolsree Leartsakulpanich. 2012. "Plasmodium Serine Hydroxymethyltransferase: Indispensability and Display of Distinct Localization." *Malar. J.* 11 (November). malariajournal.biomedcentral.com: 387. https://doi.org/10.1186/1475-2875-11-387.

Pratt, Steven, Nilu K Wansadhipathi-Kannangara, Catherine R Bruce, John G Mina, Hosam Shams-Eldin, Josefina Casas, Kentaro Hanada, Ralph T Schwarz, Sabrina Sonda, and Paul W Denny. 2013. "Sphingolipid Synthesis and Scavenging in the Intracellular Apicomplexan Parasite, Toxoplasma Gondii." *Mol. Biochem. Parasitol.* 187 (1): 43–51. https://doi.org/10.1016/j.molbiopara.2012.11.007.

Pretzel, Jette, Marina Gehr, Maike Eisenkolb, Lihui Wang, Karin Fritz-Wolf, Stefan Rahlfs, Katja Becker, and Esther Jortzik. 2016. "Characterization and Redox Regulation of Plasmodium Falciparum Methionine Adenosyltransferase." *J. Biochem.* 160 (6). academic.oup.com: 355–67. https://doi.org/10.1093/jb/mvw045.

Price, Ric N, Anne-Catrin Uhlemann, Alan Brockman, Rose McGready, Elizabeth Ashley, Lucy Phaipun, Rina Patel, et al. 2004. "Mefloquine Resistance in Plasmodium Falciparum and Increased Pfmdr1 Gene Copy Number." *Lancet* 364 (9432): 438–47. https://doi.org/10.1016/S0140-6736(04)16767-6.

Prieto, Judith Helena, Sasa Koncarevic, Sung Kyu Park, John Yates 3rd, and Katja Becker. 2008. "Large-Scale Differential Proteome Analysis in Plasmodium Falciparum Under Drug Treatment." *PLoS One* 3 (12). journals.plos.org: e4098. https://doi.org/

10.1371/journal.pone.0004098.

Prlić, Andreas, and James B Procter. 2012. "Ten Simple Rules for the Open Development of Scientific Software." *PLoS Comput. Biol.* 8 (12). journals.plos.org: e1002802. https://doi.org/10.1371/journal.pcbi.1002802.

Ralph, Stuart A, Giel G van Dooren, Ross F Waller, Michael J Crawford, Martin J Fraunholz, Bernardo J Foth, Christopher J Tonkin, David S Roos, and Geoffrey I McFadden. 2004. "Tropical Infectious Diseases: Metabolic Maps and Functions of the Plasmodium Falciparum Apicoplast." *Nat. Rev. Microbiol.* 2 (3): 203–16. https://doi.org/10.1038/nrmicro843.

Ramya, T N C, Satyendra Mishra, Krishanpal Karmodiya, Namita Surolia, and Avadhesha Surolia. 2007. "Inhibitors of Nonhousekeeping Functions of the Apicoplast Defy Delayed Death in Plasmodium Falciparum." *Antimicrob. Agents Chemother.* 51 (1). Am Soc Microbiol: 307–16. https://doi.org/10.1128/AAC.00808-06.

Randall, L M, and C A Hunter. 2011. "Parasite Dissemination and the Pathogenesis of Toxoplasmosis." *Eur. J. Microbiol. Immunol.* 1 (1): 3–9. https://doi.org/10.1556/EuJMI.1.2011.1.3.

Raškevičius, Vytautas, Valeryia Mikalayeva, Ieva Antanavičiūtė, Ieva Ceslevičienė, Vytenis Arvydas Skeberdis, Visvaldas Kairys, and Sergio Bordel. 2018. "Genome Scale Metabolic Models as Tools for Drug Design and Personalized Medicine." *PLoS One* 13 (1): e0190636. https://doi.org/10.1371/journal.pone.0190636.

R Core Team. 2017. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing.

Reyes, P, P K Rathod, D J Sanchez, J E Mrema, K H Rieckmann, and H G Heidrich. 1982. "Enzymes of Purine and Pyrimidine Metabolism from the Human Malaria Parasite, Plasmodium Falciparum." *Mol. Biochem. Parasitol.* 5 (5): 275–90.

Rietz, P J, F S Skelton, K Folkers, and Others. 1967. "Occurrence of Ubiquinones-8 and-9 in Plasmodium Lophurae." *Internationale Zeitschrift Fur Vitaminforschung* 37. cabdirect.org: 405–11.

Ritchie, Matthew E, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-sequencing and Microarray Studies." *Nucleic Acids Res.* 43 (7). academic.oup.com: e47. https://doi.org/10.1093/nar/gkv007.

Rojo, David, Gisele A B Canuto, Emerson A Castilho-Martins, Marina F M Tavares, Coral Barbas, Ángeles López-Gonzálvez, and Luis Rivas. 2015. "A Multiplatform Metabolomic Approach to the Basis of Antimonial Action and Resistance in Leishmania

Infantum." *PLoS One* 10 (7). journals.plos.org: e0130675. https://doi.org/10.1371/journal.pone.0130675.

RTS,S Clinical Trials Partnership. 2015. "Efficacy and Safety of RTS,S/AS01 Malaria Vaccine with or Without a Booster Dose in Infants and Children in Africa: Final Results of a Phase 3, Individually Randomised, Controlled Trial." *Lancet* 386 (9988). Elsevier: 31–45. https://doi.org/10.1016/S0140-6736(15)60721-8.

Rønn, A M, H A Msangeni, J Mhina, W H Wernsdorfer, and I C Bygbjerg. 1996. "High Level of Resistance of Plasmodium Falciparum to Sulfadoxine-Pyrimethamine in Children in Tanzania." *Trans. R. Soc. Trop. Med. Hyg.* 90 (2): 179–81. https://doi.org/10.1016/S0035-9203(96)90129-7.

Salcedo, Enrique, Paul F G Sims, and John E Hyde. 2005. "A Glycine-Cleavage Complex as Part of the Folate One-Carbon Metabolism of Plasmodium Falciparum." *Trends Parasitol.* 21 (9). Elsevier: 406–11. https://doi.org/10.1016/j.pt.2005.07.001.

Salcedo-Sora, J Enrique, Edwin Ochong, Susan Beveridge, David Johnson, Alexis Nzila, Giancarlo A Biagini, Paul A Stocks, et al. 2011. "The Molecular Basis of Folate Salvage in Plasmodium Falciparum: Characterization of Two Folate Transporters." *J. Biol. Chem.* 286 (52). ASBMB: 44659–68. https://doi.org/10.1074/jbc.M111.286054.

Saluja, T, S D Sharma, M Gupta, R Kundu, S Kar, A Dutta, M Silveira, et al. 2014. "A Multicenter Prospective Hospital-Based Surveillance to Estimate the Burden of Rotavirus Gastroenteritis in Children Less Than Five Years of Age in India." *Vaccine* 32 Suppl 1 (August): A13–9. https://doi.org/10.1016/j.vaccine.2014.03.030.

Sana, Theodore R, D Benjamin Gordon, Steven M Fischer, Shane E Tichy, Norton Kitagawa, Cindy Lai, William L Gosnell, and Sandra P Chang. 2013. "Global Mass Spectrometry Based Metabolomics Profiling of Erythrocytes Infected with Plasmodium Falciparum." *PLoS One* 8 (4). journals.plos.org: e60840. https://doi.org/10.1371/journal.pone.0060840.

Sandve, Geir Kjetil, Anton Nekrutenko, James Taylor, and Eivind Hovig. 2013. "Ten Simple Rules for Reproducible Computational Research." *PLoS Comput. Biol.* 9 (10). journals.plos.org: e1003285. https://doi.org/10.1371/journal.pcbi.1003285.

Sasi, Philip, Shamus P Burns, Catherine Waruiru, Michael English, Claire L Hobson, Christopher G King, Moses Mosobo, et al. 2007. "Metabolic Acidosis and Other Determinants of Hemoglobin-Oxygen Dissociation in Severe Childhood Plasmodium Falciparum Malaria." *Am. J. Trop. Med. Hyg.* 77 (2): 256–60.

Sateriale, Adam, and Boris Striepen. 2016. "Beg, Borrow and Steal: Three Aspects of Horizontal Gene Transfer in the Protozoan Parasite, Cryptosporidium Parvum."

PLoS Pathog. 12 (3): e1005429. https://doi.org/10.1371/journal.ppat.1005429.

Sayed, Fatma G, Amany I Hamza, Lamia A Galal, Douaa M Sayed, and Mona Gaber. 2016. "Virulence of Geographically Different Cryptosporidium Parvum Isolates in Experimental Animal Model." *Ann. Parasitol.* 62 (3): 221–32. https://doi.org/10.17420/ap6203.56.

Schellenberger, Jan, Richard Que, Ronan M T Fleming, Ines Thiele, Jeffrey D Orth, Adam M Feist, Daniel C Zielinski, et al. 2011. "Quantitative Prediction of Cellular Metabolism with Constraint-Based Models: The COBRA Toolbox V2.0." *Nat. Protoc.* 6 (9). nature.com: 1290–1307. https://doi.org/10.1038/nprot.2011.308.

Scherf, A, R Hernandez-Rivas, P Buffet, E Bottius, C Benatar, B Pouvelle, J Gysin, and M Lanzer. 1998. "Antigenic Variation in Malaria: In Situ Switching, Relaxed and Mutually Exclusive Transcription of Var Genes During Intra-Erythrocytic Development in Plasmodium Falciparum." *EMBO J.* 17 (18). EMBO Press: 5418–26.

Schnee, Amanda E, Rashidul Haque, Mami Taniuchi, Jashim Uddin, Masud Alam, Jie Liu, Elizabeth T Rogawski, et al. 2018. "Identification of Etiology-Specific Diarrhea Associated with Linear Growth Faltering in Bangladeshi Infants." *Am. J. Epidemiol.*, May. https://doi.org/10.1093/aje/kwy106.

Schnell, Santiago. 2015. "Ten Simple Rules for a Computational Biologist's Laboratory Notebook." *PLoS Comput. Biol.* 11 (9). journals.plos.org: e1004385. https://doi.org/10.1371/journal.pcbi.1004385.

Schulz, J B, J Lindenau, J Seyfried, and J Dichgans. 2000. "Glutathione, Oxidative Stress and Neurodegeneration." *Eur. J. Biochem.* 267 (16). Wiley Online Library: 4904–11.

Schwartz, Eli, and Tamar Lachish. 2016. "Artemisinin-Based Combination Therapy (ACT) Versus Atovaquone-Proguanil: Do Not Choose Between but, Rather, Combine Them." *Evid. Based. Med.* 21 (2). ebm.bmj.com: 64. https://doi.org/10.1136/ebmed-2015-110183.

Sengupta, Arjun, Soumita Ghosh, Bidyut K Das, Abhinash Panda, Rina Tripathy, Sylviane Pied, B Ravindran, Sulabha Pathak, Shobhona Sharma, and Haripalsingh M Sonawat. 2016. "Host Metabolic Responses to Plasmodium Falciparum Infections Evaluated by 1 H NMR Metabolomics." *Mol. Biosyst.* 12 (11). Royal Society of Chemistry: 3324–32.

Sibley, L D, M Messina, and I R Niesman. 1994. "Stable DNA Transformation in the Obligate Intracellular Parasite Toxoplasma Gondii by Complementation of Tryptophan Auxotrophy." *Proc. Natl. Acad. Sci. U. S. A.* 91 (12): 5508–12.

Siddiqui, Ghizal, Anubhav Srivastava, Adrian S Russell, and Darren J Creek. 2017. "Multi-Omics Based Identification of Specific Biochemical Changes Associated with PfKelch13-Mutant Artemisinin-Resistant Plasmodium Falciparum." *J. Infect. Dis.* 215 (9). academic.oup.com: 1435–44. https://doi.org/10.1093/infdis/jix156.

Sidhu, Amar Bir Singh, Dominik Verdier-Pinard, and David A Fidock. 2002. "Chloroquine Resistance in Plasmodium Falciparum Malaria Parasites Conferred by Pfcrt Mutations." *Science* 298 (5591). science.sciencemag.org: 210–13. https://doi.org/10.1126/science.1074045.

Sidik, Saima M, Diego Huet, Suresh M Ganesan, My-Hang Huynh, Tim Wang, Armiyaw S Nasamu, Prathapan Thiru, et al. 2016. "A Genome-Wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes." *Cell* 166 (6): 1423–1435.e12.

Sigala, Paul A, and Daniel E Goldberg. 2014. "The Peculiarities and Paradoxes of Plasmodium Heme Metabolism." *Annu. Rev. Microbiol.* 68: 259–78. https://doi.org/10.1146/annurev-micro-091313-103537.

Silva, Ana Marta, Anabela Cordeiro-da-Silva, and Graham H Coombs. 2011. "Metabolic Variation During Development in Culture of Leishmania Donovani Promastigotes." *PLoS Negl. Trop. Dis.* 5 (12). journals.plos.org: e1451. https://doi.org/10.1371/journal.pntd.0001451.

Silva, Leslie P, Philip L Lorenzi, Preeti Purwaha, Valeda Yong, David H Hawke, and John N Weinstein. 2013. "Measurement of DNA Concentration as a Normalization Strategy for Metabolomic Data from Adherent Cell Lines." *Anal. Chem.* 85 (20). ACS Publications: 9536–42. https://doi.org/10.1021/ac401559v.

Singh, Sanjay, Sunil K Puri, Shio K Singh, Ragini Srivastava, Ram C Gupta, and Vikas C Pandey. 1997. "Characterization of Simian Malarial Parasite (Plasmodium Knowlesi)-Induced Putrescine Transport in Rhesus Monkey Erythrocytes a Novel Putrescine Conjugate Arrests in Vitro Growth of Simian Malarial Parasite (Plasmodium Knowlesi) and Cures Multidrug Resistant Murine Malaria (Plasmodium Yoelii) Infection in Vivo." *J. Biol. Chem.* 272 (21). ASBMB: 13506–11.

Singh, Subhash, Matthew Plassmeyer, Deepak Gaur, and Louis H Miller. 2007. "Mononeme: A New Secretory Organelle in Plasmodium Falciparum Merozoites Identified by Localization of Rhomboid-1 Protease." *Proc. Natl. Acad. Sci. U. S. A.* 104 (50): 20043–8. https://doi.org/10.1073/pnas.0709999104.

Sinka, Marianne E, Michael J Bangs, Sylvie Manguin, Yasmin Rubio-Palis, Theeraphap Chareonviriyaphap, Maureen Coetzee, Charles M Mbogo, et al. 2012. "A Global Map of Dominant Malaria Vectors." *Parasit. Vectors* 5 (April): 69. https://doi.org/10.1186/1756-3305-5-69.

Siregar, Josephine E, Genji Kurisu, Tamaki Kobayashi, Motomichi Matsuzaki, Kimitoshi Sakamoto, Fumika Mi-ichi, Yoh-Ichi Watanabe, et al. 2015. "Direct Evidence for the Atovaquone Action on the Plasmodium Cytochrome Bc1 Complex." *Parasitol. Int.* 64 (3). Elsevier: 295–300. https://doi.org/10.1016/j.parint.2014.09.011.

Skelton, F S, P J Rietz, and K Folkers. 1970. "Coenzyme Q. CXXII. Identification of Ubiquinone-8 Biosynthesized by Plasmodium Knowlesi, P. Cynomolgi, and P. Berghei." *J. Med. Chem.* 13 (4). ACS Publications: 602–6.

Sparks, Hayley, Gayatri Nair, Alejandro Castellanos-Gonzalez, and A Clinton White Jr. 2015. "Treatment of Cryptosporidium: What We Know, Gaps, and the Way Forward." *Curr Trop Med Rep* 2 (3): 181–87.

Srivastava, Anubhav, Nisha Philip, Katie R Hughes, Konstantina Georgiou, James I MacRae, Michael P Barrett, Darren J Creek, Malcolm J McConville, and Andrew P Waters. 2016. "Stage-Specific Changes in Plasmodium Metabolism Required for Differentiation and Adaptation to Different Host and Vector Environments." *PLoS Pathog.* 12 (12). journals.plos.org: e1006094. https://doi.org/10.1371/journal.ppat.1006094.

Staines, Henry M, Stephanie Ashmore, Hannah Felgate, Jessica Moore, Trevor Powell, and J Clive Ellory. 2006. "Solute Transport via the New Permeability Pathways in Plasmodium Falciparum–Infected Human Red Blood Cells Is Not Consistent with a Simple Single-Channel Model." *Blood* 108 (9). American Society of Hematology: 3187–94. https://doi.org/10.1182/blood-2006-02-001693.

Steiner, Kevin L, Shahnawaz Ahmed, Carol A Gilchrist, Cecelia Burkey, Heather Cook, Jennie Z Ma, Poonum S Korpe, et al. 2018. "Species of Cryptosporidia Causing Subclinical Infection Associated with Growth Faltering in Rural and Urban Bangladesh- a Birth Cohort Study." *Clin. Infect. Dis.*, April. academic.oup.com. https://doi.org/10.1093/cid/ciy310.

Sterkers, Yvon, Laurence Lachaud, Nathalie Bourgeois, Lucien Crobu, Patrick Bastien, and Michel Pagès. 2012. "Novel Insights into Genome Plasticity in Eukaryotes: Mosaic Aneuploidy in Leishmania." *Mol. Microbiol.* 86 (1). Wiley Online Library: 15–23. https://doi.org/10.1111/j.1365-2958.2012.08185.x.

Stevens, L. 1970. "The Biochemical Role of Naturally Occurring Polyamines in Nucleic Acid Synthesis." *Biol. Rev. Camb. Philos. Soc.* 45 (1). Wiley Online Library: 1–27.

Straimer, Judith, Nina F Gnädig, Benoit Witkowski, Chanaki Amaratunga, Valentine Duru, Arba Pramundita Ramadani, Mélanie Dacheux, et al. 2015. "Drug Resistance. K13-Propeller Mutations Confer Artemisinin Resistance in Plasmodium Falciparum Clinical Isolates." *Science* 347 (6220). science.sciencemag.org: 428–31. https://doi.

org/10.1126/science.1260867.

Stuetz, Wolfgang, Verena Ilona Carrara, Rose McGready, Sue Jean Lee, Hans Konrad Biesalski, and François Henry Nosten. 2012. "Thiamine Diphosphate in Whole Blood, Thiamine and Thiamine Monophosphate in Breast-Milk in a Refugee Population." *PLoS One* 7 (6): e36280. https://doi.org/10.1371/journal.pone.0036280.

Sturm, Angelika, Vanessa Mollard, Anton Cozijnsen, Christopher D Goodman, and Geoffrey I McFadden. 2015. "Mitochondrial ATP Synthase Is Dispensable in Blood-Stage Plasmodium Berghei Rodent Malaria but Essential in the Mosquito Phase." *Proc. Natl. Acad. Sci. U. S. A.* 112 (33). National Acad Sciences: 10216–23. https://doi.org/10.1073/pnas.1423959112.

Su, X Z, V M Heatwole, S P Wertheimer, F Guinet, J A Herrfeldt, D S Peterson, J A Ravetch, and T E Wellems. 1995. "The Large Diverse Gene Family Var Encodes Proteins Involved in Cytoadherence and Antigenic Variation of Plasmodium Falciparum-Infected Erythrocytes." *Cell* 82 (1). Elsevier: 89–100.

Sugimoto, Masahiro, Masato Kawakami, Martin Robert, Tomoyoshi Soga, and Masaru Tomita. 2012. "Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis." *Curr. Bioinform.* 7 (1). ingentaconnect.com: 96–108. https://doi.org/10.2174/157489312799304431.

Sullivan, D J, I Y Gluzman, D G Russell, and D E Goldberg. 1996. "On the Molecular Mechanism of Chloroquine's Antimalarial Action." *Proc. Natl. Acad. Sci. U. S. A.* 93 (21): 11865–70.

Sun, Chen, Jian Li, Yu Cao, Gongbo Long, and Bing Zhou. 2015. "Two Distinct and Competitive Pathways Confer the Cellcidal Actions of Artemisinins." *Microb. Cell Fact.* 2 (1). ncbi.nlm.nih.gov: 14–25. https://doi.org/10.15698/mic2015.01.181.

Surolia, N, and G Padmanaban. 1992. "De Novo Biosynthesis of Heme Offers a New Chemotherapeutic Target in the Human Malarial Parasite." *Biochem. Biophys. Res. Commun.* 187 (2). Elsevier: 744–50.

Surolia, N, and A Surolia. 2001. "Triclosan Offers Protection Against Blood Stages of Malaria by Inhibiting enoyl-ACP Reductase of Plasmodium Falciparum." *Nat. Med.* 7 (2). nature.com: 167–73. https://doi.org/10.1038/84612.

Syafruddin, D, J E Siregar, and S Marzuki. 1999. "Mutations in the Cytochrome B Gene of Plasmodium Berghei Conferring Resistance to Atovaquone." *Mol. Biochem. Parasitol.* 104 (2). Elsevier: 185–94.

Tahar, Rachida, and Leonardo K Basco. 2007. "Molecular Epidemiology of Malaria in Cameroon. XXV. In Vitro Activity of Fosmidomycin and Its Derivatives Against

Fresh Clinical Isolates of Plasmodium Falciparum and Sequence Analysis of 1-Deoxy-d-Xylulose 5-Phosphate Reductoisomerase." *Am. J. Trop. Med. Hyg.* 77 (2): 214–20.

Team, R C. 2015. "R Foundation for Statistical Computing; Vienna, Austria: 2014."*: A Language and Environment for Statistical Computing*.

Teng, Rongwei, Adele M Lehane, Markus Winterberg, Sarah H Shafik, Robert L Summers, Rowena E Martin, Donelly A van Schalkwyk, Pauline R Junankar, and Kiaran Kirk. 2014. "1H-NMR Metabolite Profiles of Different Strains of Plasmodium Falciparum." *Biosci. Rep.* 34 (6). bioscirep.org: e00150. https://doi.org/10.1042/BSR20140134.

Teuscher, Franka, Nanhua Chen, Dennis E Kyle, Michelle L Gatton, and Qin Cheng. 2012. "Phenotypic Changes in Artemisinin-Resistant Plasmodium Falciparum Lines in Vitro: Evidence for Decreased Sensitivity to Dormancy and Growth Inhibition." *Antimicrob. Agents Chemother.* 56 (1). Am Soc Microbiol: 428–31. https://doi.org/10.1128/AAC.05456-11.

Teuscher, Franka, Michelle L Gatton, Nanhua Chen, Jennifer Peters, Dennis E Kyle, and Qin Cheng. 2010. "Artemisinin-Induced Dormancy in Plasmodium Falciparum: Duration, Recovery Rates, and Implications in Treatment Failure." *J. Infect. Dis.* 202 (9). Oxford University Press: 1362–8. https://doi.org/10.1086/656476.

Thiele, Ines, and Bernhard Ø Palsson. 2010. "A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction." *Nat. Protoc.* 5 (1). nature.com: 93–121. https://doi.org/10.1038/nprot.2009.203.

Torres, Leonardo Collado. 2017. "How to Ask for Help for Bioconductor Packages." http://lcolladotor.github.io/2017/03/06/How-to-ask-for-help-for-Bioconductor-packages/#.WwLeZ1MvzUp.

Trager, W, and J B Jensen. 1976. "Human Malaria Parasites in Continuous Culture." *Science* 193 (4254). science.sciencemag.org: 673–75.

Trelka, Darin P, Timothy G Schneider, John C Reeder, and Theodore F Taraschi. 2000. "Evidence for Vesicle-Mediated Trafficking of Parasite Proteins to the Host Cell Cytosol and Erythrocyte Surface Membrane in Plasmodium Falciparum Infected Ertythrocytes." *Mol. Biochem. Parasitol.* 106 (1). Elsevier: 131–45. https://doi.org/10.1016/S0166-6851(99)00207-8.

Tymoshenko, Stepan, Rebecca D Oppenheim, Rasmus Agren, Jens Nielsen, Dominique Soldati-Favre, and Vassily Hatzimanikatis. 2015. "Metabolic Needs and Capabilities of Toxoplasma Gondii Through Combined Computational and Experimental Analysis." *PLoS Comput. Biol.* 11 (5): e1004261.

Tymoshenko, Stepan, Rebecca D Oppenheim, Dominique Soldati-Favre, and Vassily Hatzimanikatis. 2013. "Functional Genomics of Plasmodium Falciparum Using Metabolic Modelling and Analysis." *Brief. Funct. Genomics* 12 (4). academic.oup.com: 316–27. https://doi.org/10.1093/bfgp/elt017.

Vanee, Niti, Seth B Roberts, Stephen S Fong, Patricio Manque, and Gregory A Buck. 2010. "A Genome-Scale Metabolic Model of Cryptosporidium Hominis." *Chem. Biodivers.* 7 (5). Wiley Online Library: 1026–39.

Van Voorhis, Wesley C, John H Adams, Roberto Adelfio, Vida Ahyong, Myles H Akabas, Pietro Alano, Aintzane Alday, et al. 2016. "Open Source Drug Discovery with the Malaria Box Compound Collection for Neglected Diseases and Beyond." *PLoS Pathog.* 12 (7). Public Library of Science: e1005763. https://doi.org/10.1371/journal.ppat.1005763.

Vaughan, Ashley M, Stefan H I Kappe, Alexander Ploss, and Sebastian A Mikolajczak. 2012. "Development of Humanized Mouse Models to Study Human Malaria Parasite Infection." *Future Microbiol.* 7 (5): 657–65. https://doi.org/10.2217/fmb.12.27.

Vaughan, Ashley M, Matthew T O'Neill, Alice S Tarun, Nelly Camargo, Thuan M Phuong, Ahmed S I Aly, Alan F Cowman, and Stefan H I Kappe. 2009. "Type II Fatty Acid Synthesis Is Essential Only for Malaria Parasite Late Liver Stage Development." *Cell. Microbiol.* 11 (3). Wiley Online Library: 506–20. https://doi.org/10.1111/j.1462-5822.2008.01270.x.

Vega-Rodríguez, Joel, Rebecca Pastrana-Mena, Keila N Crespo-Lladó, José G Ortiz, Iván Ferrer-Rodríguez, and Adelfa E Serrano. 2015. "Implications of Glutathione Levels in the Plasmodium Berghei Response to Chloroquine and Artemisinin." *PLoS One* 10 (5). journals.plos.org: e0128212. https://doi.org/10.1371/journal.pone.0128212.

Vial, Henri J, Jean R Philippot, and Donald F H Wallach. 1984. "A Reevaluation of the Status of Cholesterol in Erythrocytes Infected by Plasmodium Knowlesi and P. Falciparum." *Mol. Biochem. Parasitol.* 13 (1): 53–65. https://doi.org/10.1016/0166-6851(84)90101-4.

Vicens, Quentin, and Philip E Bourne. 2009. "Ten Simple Rules to Combine Teaching and Research." *PLoS Comput. Biol.* 5 (4). journals.plos.org: e1000358. https://doi.org/10.1371/journal.pcbi.1000358.

Vries, A de, and B D Ripley. 2013. "Ggdendro: Tools for Extracting Dendrogram and Tree Diagram Plot Data for Use with Ggplot."

Wagner, Helene. 2015. "Vegan: Community Ecology Package." *R Package*.

Wagner, Jeffrey C, Randall J Platt, Stephen J Goldfless, Feng Zhang, and Jacquin

C Niles. 2014. "Efficient CRISPR-Cas9-mediated Genome Editing in Plasmodium Falciparum." *Nat. Methods* 11 (9). nature.com: 915–18. https://doi.org/10.1038/nmeth.3063.

Waller, R F, P J Keeling, R G Donald, B Striepen, E Handman, N Lang-Unnasch, A F Cowman, G S Besra, D S Roos, and G I McFadden. 1998. "Nuclear-Encoded Proteins Target to the Plastid in Toxoplasma Gondii and Plasmodium Falciparum." *Proc. Natl. Acad. Sci. U. S. A.* 95 (21). National Acad Sciences: 12352–7.

Wallqvist, Anders, Xin Fang, Shivendra G Tewari, Ping Ye, and Jaques Reifman. 2016. "Metabolic Host Responses to Malarial Infection During the Intraerythrocytic Developmental Cycle." *BMC Syst. Biol.* 10 (1). bmcsystbiol.biomedcentral.com: 58. https://doi.org/10.1186/s12918-016-0291-2.

Wang, Jigang, Chong-Jing Zhang, Wan Ni Chia, Cheryl C Y Loh, Zhengjun Li, Yew Mun Lee, Yingke He, et al. 2015. "Haem-Activated Promiscuous Targeting of Artemisinin in Plasmodium Falciparum." *Nat. Commun.* 6 (December). nature.com: 10111. https://doi.org/10.1038/ncomms10111.

Wang, Juan, Liying Huang, Jian Li, Qiangwang Fan, Yicheng Long, Ying Li, and Bing Zhou. 2010. "Artemisinin Directly Targets Malarial Mitochondria Through Its Specific Mitochondrial Activation." *PLoS One* 5 (3). journals.plos.org: e9582. https://doi.org/10.1371/journal.pone.0009582.

Wang, Yutang, Hanzhong Liu, Gavin McKenzie, Paul K Witting, Johannes-Peter Stasch, Michael Hahn, Dechaboon Changsirivathanathamrong, et al. 2010. "Kynurenine Is an Endothelium-Derived Relaxing Factor Produced During Inflammation." *Nat. Med.* 16 (3). nature.com: 279–85. https://doi.org/10.1038/nm.2092.

Wellems, Thomas E, and Rick M Fairhurst. 2012. "An Evolving Picture of the Interactions Between Malaria Parasites and Their Host Erythrocytes." *Cell Res.* 22 (3). nature.com: 453–56. https://doi.org/10.1038/cr.2012.14.

Westrop, Gareth D, Roderick A M Williams, Lijie Wang, Tong Zhang, David G Watson, Ana Marta Silva, and Graham H Coombs. 2015. "Metabolomic Analyses of Leishmania Reveal Multiple Species Differences and Large Differences in Amino Acid Metabolism." *PLoS One* 10 (9). journals.plos.org: e0136891. https://doi.org/10.1371/journal.pone.0136891.

White, Lisa J, Jennifer A Flegg, Aung Pyae Phyo, Ja Hser Wiladpai-ngern, Delia Bethell, Christopher Plowe, Tim Anderson, et al. 2015. "Defining the in Vivo Phenotype of Artemisinin-Resistant Falciparum Malaria: A Modelling Approach." *PLoS Med.* 12 (4). journals.plos.org: e1001823. https://doi.org/10.1371/journal.pmed.1001823.

Wickham, H. 2007. "Reshaping Data with the Reshape Package." had.co.nz.

———. 2012. "Reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package."

Wickham, Hadley. 2017. "Tidyverse: Easily Install and Load'tidyverse'packages."

Wickham, Hadley, and J Bryan. 2017. "Readxl: Read Excel Files. R Package Version 1.0. 0." URL https://CRAN. R-project. org/package= readxl.

Wickham, Hadley, Romain Francois, Lionel Henry, and K Müller. 2015. "Dplyr: A Grammar of Data Manipulation." *R Package Version 0. 4* 3.

Wishart, David S, Dan Tzur, Craig Knox, Roman Eisner, An Chi Guo, Nelson Young, Dean Cheng, et al. 2007. "HMDB: The Human Metabolome Database." *Nucleic Acids Res.* 35 (Database issue). academic.oup.com: D521–6. https://doi.org/10.1093/nar/gkl923.

Witkowski, Benoit, Chanaki Amaratunga, Nimol Khim, Sokunthea Sreng, Pheaktra Chim, Saorin Kim, Pharath Lim, et al. 2013. "Novel Phenotypic Assays for the Detection of Artemisinin-Resistant Plasmodium Falciparum Malaria in Cambodia: In-Vitro and Ex-Vivo Drug-Response Studies." *Lancet Infect. Dis.* 13 (12): 1043–9. https://doi.org/10.1016/S1473-3099(13)70252-4.

Witkowski, Benoit, Joel Lelièvre, María José López Barragán, Victor Laurent, Xin-Zhuan Su, Antoine Berry, and Françoise Benoit-Vical. 2010. "Increased Tolerance to Artemisinin in Plasmodium Falciparum Is Mediated by a Quiescence Mechanism." *Antimicrob. Agents Chemother.* 54 (5). Am Soc Microbiol: 1872–7. https://doi.org/10.1128/AAC.01636-09.

"Women, Minorities, and Persons with Disabilities in Science and Engineering." n.d. https://www.nsf.gov/statistics/2017/nsf17310/.

"World Malaria Report 2017." 2017. Geneva: World Health Organization.

Wright, Gavin J, and Julian C Rayner. 2014. "Plasmodium Falciparum Erythrocyte Invasion: Combining Function with Immune Evasion." *PLoS Pathog.* 10 (3): e1003943. https://doi.org/10.1371/journal.ppat.1003943.

Wu, Yiman, and Liang Li. 2016. "Sample Normalization Methods in Quantitative Metabolomics." *J. Chromatogr. A* 1430 (January). Elsevier: 80–95. https://doi.org/10.1016/j.chroma.2015.12.007.

Xia, Dong, Sanya J Sanderson, Andrew R Jones, Judith H Prieto, John R Yates, Elizabeth Bromley, Fiona M Tomley, et al. 2008. "The Proteome of Toxoplasma Gondii: Integration with the Genome Provides Novel Insights into Gene Expression

and Annotation." *Genome Biol.* 9 (7). genomebiology.biomedcentral.com: R116. https://doi.org/10.1186/gb-2008-9-7-r116.

Xia, Jianguo, and David S Wishart. 2011. "Web-Based Inference of Biological Patterns, Functions and Pathways from Metabolomic Data Using MetaboAnalyst." *Nat. Protoc.* 6 (6). nature.com: 743–60. https://doi.org/10.1038/nprot.2011.319.

Xie, Y. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." *Implement Reprod Res*. books.google.com.

Xu, Ping, Giovanni Widmer, Yingping Wang, Luiz S Ozaki, Joao M Alves, Myrna G Serrano, Daniela Puiu, et al. 2004. "The Genome of Cryptosporidium Hominis." *Nature* 431 (7012). nature.com: 1107–12. https://doi.org/10.1038/nature02977.

Yang, Y Z, B Little, and S R Meshnick. 1994. "Alkylation of Proteins by Artemisinin. Effects of Heme, pH, and Drug Structure." *Biochem. Pharmacol.* 48 (3). Elsevier: 569–73.

Yayon, A, J A Vande Waa, M Yayon, T G Geary, and J B Jensen. 1983. "Stage-Dependent Effects of Chloroquine on Plasmodium Falciparum in Vitro." *J. Protozool.* 30 (4): 642–47.

Yeh, Ellen, and Joseph L DeRisi. 2011. "Chemical Rescue of Malaria Parasites Lacking an Apicoplast Defines Organelle Function in Blood-Stage Plasmodium Falciparum." *PLoS Biol.* 9 (8). journals.plos.org: e1001138. https://doi.org/10.1371/journal.pbio.1001138.

Yeh, Iwei, Theodor Hanekamp, Sophia Tsoka, Peter D Karp, and Russ B Altman. 2004. "Computational Analysis of Plasmodium Falciparum Metabolism: Organizing Genomic Information to Facilitate Drug Discovery." *Genome Res.* 14 (5). genome.cshlp.org: 917–24. https://doi.org/10.1101/gr.2050304.

Zalis, M G, L Pang, M S Silveira, W K Milhous, and D F Wirth. 1998. "Characterization of Plasmodium Falciparum Isolated from the Amazon Region of Brazil: Evidence for Quinine Resistance." *Am. J. Trop. Med. Hyg.* 58 (5): 630–37. https://doi.org/10.4269/ajtmh.1998.58.630.

Zeeberg, Barry R, Joseph Riss, David W Kane, Kimberly J Bussey, Edward Uchio, W Marston Linehan, J Carl Barrett, and John N Weinstein. 2004. "Mistaken Identifiers: Gene Name Errors Can Be Introduced Inadvertently When Using Excel in Bioinformatics." *BMC Bioinformatics* 5 (June). bmcbioinformatics: 80. https://doi.org/10.1186/1471-2105-5-80.

Zhang, J, M Krugliak, and H Ginsburg. 1999. "The Fate of Ferriprotorphyrin IX in Malaria Infected Erythrocytes in Conjunction with the Mode of Action of Antimalarial

Drugs." *Mol. Biochem. Parasitol.* 99 (1): 129–41.

Zhang, Min, Chengqi Wang, Thomas D Otto, Jenna Oberstaller, Xiangyun Liao, Swamy R Adapa, Kenneth Udenze, et al. 2018. "Uncovering the Essential Genes of the Human Malaria Parasite Plasmodium Falciparum by Saturation Mutagenesis." *Science* 360 (6388). science.sciencemag.org. https://doi.org/10.1126/science.aap7847.

Zhou, Yiqing, Weichao Li, and Youli Xiao. 2016. "Profiling of Multiple Targets of Artemisinin Activated by Hemin in Cancer Cell Proteome." *ACS Chem. Biol.* 11 (4). ACS Publications: 882–88. https://doi.org/10.1021/acschembio.5b01043.

Zhu, Hao. 2018. "KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax."

Ziemann, Mark, Yotam Eren, and Assam El-Osta. 2016. "Gene Name Errors Are Widespread in the Scientific Literature." *Genome Biol.* 17 (1). genomebiology.biomedcentral.com: 177. https://doi.org/10.1186/s13059-016-1044-7.