Washington University in St. Louis

Washington University Open Scholarship

Summer 8-15-2018

Breath biomarkers and an expanded role for isoprenoids in Plasmodium falciparum

Chad Louis Schaber
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Microbiology Commons

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Microbiology and Microbial Pathogenesis

Dissertation Examination Committee:
Audrey R. Odom John, Chair
Jeffrey P. Henderson
Daniel S. Ory
Baranidharan Raman
L. David Sibley

Breath Biomarkers and an Expanded Role for Isoprenoids in *Plasmodium falciparum*
by
Chad L. Schaber

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2018
St. Louis, Missouri

Table of Contents

List of Figures

Chapter 4: Global proteomic analysis of prenylated proteins in Plasmodium falciparum using an alkyne-modified isoprenoid analogue

Appendix A: Malaria parasites produce volatile mosquito attractants

Appendix B: Malaria infection affects the volatile organic compound profile of sweat

Appendix C: Screen of farnesyl transferase inhibitors against *P. falciparum*

List of Tables

Appendix B: Malaria infection affects the volatile organic compound profile of sweat

Appendix C: Screen of farnesyl transferase inhibitors against *P. falciparum*

List of Abbreviations

AgOR, *Anopheles gambiae* odorant receptor

C15AlkOPP, alkyne modified farnesyl analogue

C20AlkOPP, alkyne modified geranylgeranyl pyrophosphate analogue

DMAPP, dimethylallyl pyrophosphate

EI, electron impact

EIC, extracted ion chromatogram

eNose, electronic nose

FPP, farnesyl pyrophosphate

FPPS, farnesyl pyrophosphate synthase

FSM, fosmidomycin

FT, farnesyl transferase

FTI, farnesyl transferase inhibitor

GC/MS, gas chromatography mass spectrometry

GGPP, geranylgeranyl pyrophosphate

GGT-1, geranylgeranyl transferase type I

GGT-2, geranylgeranyl transferase type II

GPP, geranyl pyrophosphate

HRP2, histidine-rich protein 2

IC_{50}, inhibitory concentration at 50% maximal activity

IPP, isopentenyl pyrophosphate

KNN, k-nearest neighbor

LC-MS, liquid chromatography mass spectrometry

LDA, linear discriminant analysis

LOD, limit of detection

MEP, methylerythritol phosphate

MW, molecular weight

PCA, principal component analysis

PCR, polymerase chain reaction

Pf, *Plasmodium falciparum*

pLDH, parasite lactate dehydrogenase

PLS, partial least squares

PoC, point of care

ppm/ppb/ppb, parts-per-million/-billion/-trillion

PrePS, Prenylation Prediction Suite

PTR, proton transfer reaction

RBC, red blood cell

RDT, rapid diagnostic test

RF, random forest

ROC, receiver operator characteristic

RT, retention time

SEM, standard error measurement

SIFT, selected ion flow tube

SPME, solid phase microextraction

SVM, support machine vector

TD, thermal desorption

TIC, total ion chromatogram

VOC, volatile organic compound

Acknowledgments

This work would not have been possible without the advice, support, and aid of my teachers, mentors, colleagues, family, and friends. Thank you all. A proper tribute would easily double the length of my dissertation.

A huge debt of gratitude is owed to my research mentor, Audrey. Exceptional, driven, insightful, and encouraging, Audrey did more than anyone else to guide me through my PhD. Whether bolstering my confidence in the face of setbacks or pushing me to take on challenging work and grow as a scientist, she has been an unflagging source of positivity and support. I cannot thank her enough.

My thanks to the members of my thesis committee. Their guidance helped me to avoid pitfalls and focus on the most meritorious avenues of research. My thanks to all of you.

A shower of thanks goes to all the members of the Odom John lab, past and present. I was lucky to have a group of people who filled my working hours with laughter and comradery day in and day out for six odd years. Each deserves their own Acknowledgements section. Special thanks go to Leah Imlay for being a continual source of aid and cheer, Chris Armstrong for being a great labmate and an even better friend, Dana Hodge for being a stalwart pillar stopping the lab from crashing down around our ears, and Ann Guggisberg for being a senior fellow-traveler who helped light my way.

Dedicated to my amazing parents. Thank you for your boundless wisdom, support, and love.

ABSTRACT OF THE DISSERTATION

Breath Biomarkers and an Expanded Role for Isoprenoids in *Plasmodium falciparum*

Chad L. Schaber

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Microbiology and Microbial Pathogenesis

Washington University in St. Louis, 2018

Professor Audrey R. Odom John, Chair

Malaria remains one of the deadliest infectious diseases worldwide, and efforts to combat it require novel insights into diagnostics, vector transmission, and drug inhibitor targets. Previous studies suggest that malaria infection causes hosts to preferentially attract the transmission vector, the *Anopheles* mosquito, but the mechanism and wider implication of these findings were not known. By analyzing the headspace above malaria parasite cultures, we identified several molecules that might engender mosquito attraction. We demonstrated that several of these molecules activate mosquito odorant receptors, including the known plant-emitted mosquito attractant α-pinene. During patient studies involving two independent pediatric clinical populations from Malawi, we observed that α-pinene and the related compound 3-carene are present at higher concentrations in the breath of infected patients versus uninfected patients. Furthermore, initial analysis of sweat samples identified additional malaria-induced volatile profile changes. These results provide a viable mechanism by which infection causes increased mosquito attraction, with potential applications for creating superior lures for mosquito vector elimination campaigns. With the same two populations, we evaluated two different methods for

breath collection. We found that inert sampling bags have superior performance to the inert Bio-VOC syringe. Fascinatingly, using data from the sampling bag study, we discovered that patient breath contains volatile biomarkers that can distinguish between infected and uninfected patients. Ascertaining that malaria breath biomarkers exist paves the way for future research on new non-invasive breath diagnostic tools.

The metabolic fate of isoprenoids, the chemical class to which α-pinene and 3-carene belong, also provides potential targets for antimalarial drugs. Previous research has shown prenylation, the attachment of specific isoprenoid groups to proteins to facilitate proper localization and function, is an essential pathway in the parasite. However, the protein prenylation substrates for the malaria parasite were not known. Using a novel metabolic labelling strategy with an alkyne modified prenyl analogue, we determined the entire complement of parasite prenylated proteins. These findings will aid ongoing efforts to design inhibitors against the parasite enzymes mediating prenylation. Initial evidence of a new series of parasite prenyl transferase, specifically farnesyl transferase, inhibitors is presented as well.

Chapter 1: Introduction

1.1 Overview

Malaria has killed countless numbers of people during the millennia in which it has been a part of the human experience. Even today, with advances in detection, treatment, and prevention having reduced the burden of the disease to a fraction of historic levels, malaria affects hundreds of millions and kills hundreds of thousands each year [1]. Maintaining gains and pushing ever closer to the goal of worldwide malaria elimination requires a thorough understanding of the causative agent with an eye to how new insights can aid anti-malarial campaigns.

This chapter aims to provide foundations and context for the remainder of the dissertation. It will establish the salient features of the malaria parasite, the state of malaria diagnostics, the method by which breath biomarkers are discovered and implemented, and the functioning of isoprenoids in the parasite. The chapter thereby prepares the reader to evaluate the merit and potential applications of my work as laid out in subsequent chapters.

1.2 Malaria parasite overview and life cycle

The causative agent of malaria is unicellular protozoan parasites of the genus *Plasmodium*. The two species that account for the vast majority of malaria infections in humans are *P. falciparum* and *P. vivax*. A minority of cases are caused by *P. malariae* and *P. ovale*. Occasionally, humans can become infected with non-human primate malaria species, most notably *P. knowlesi* [2]. The

focus of this dissertation will be on *P. falciparum*, which is thought to cause nearly all of the mortality associated with malaria and is the predominate species in Africa.

Throughout its life cycle, as reviewed in [3] and summarized here, the parasite occupies two hosts: humans and mosquitoes of the genus *Anopheles*. Upon biting a human, infected mosquitoes transfer the mature mosquito-stage parasites, called sporozoites, into the new human host. At this stage, the parasites are extra-cellular and motile, moving from the site of the bite to the bloodstream and then the liver. Upon reaching the liver, sporozoites invade hepatic cells and transition to the asexual replicating liver stage. In this life stage, each sporozoite produces up to 40,000 daughter parasites, called merozoites. Once the replicating parasites reach a critical point, the host liver cell ruptures, releasing the thousands of merozoites into the blood-stream and starting the next phase of the life cycle.

The blood stage of the parasite, which is the stage responsible for the symptoms and mortality associated with malaria, is characterized by a continuous asexual replication cycle. A merozoite enters a red blood cell, a.k.a. erythrocyte, and establishes itself inside a parasitophorous vacuole, a membrane-bound compartment that separates the parasite from the cell cytoplasm. The merozoite matures into a more metabolically active trophozoite, the parasite digesting hemoglobin from the host cell and glucose from the bloodstream to fuel its growth. The trophozoite then divides into 10-30 daughter merozoites, the red blood cell rupturing to release the daughter merozoites into the bloodstream to continue the cycle. The entire blood stage cycle from merozoite invasion to daughter merozoite release takes around 48 hours, and its continual repetition can result in more than 10% of a host's red blood cells being infected [4].

A subset of merozoites will diverge from this cycle, transitioning to sexual forms known as gametocytes. Maturation of gametocytes takes roughly two weeks. Once matured, the gametocytes are able to complete the life cycle when taken up in the blood meal of a mosquito. The gametocytes sexually recombine in the mosquito midgut and progress through a series of stages in the mosquito until they end as sporozoites in the salivary glands, thus closing the circle and ready to infect a new human host.

1.3 State of malaria diagnostics

Accurate and available diagnostic tools are key to treating and containing malaria while simultaneously preventing emergence of drug resistance. The WHO currently recommends that all patients suspected of having malaria be diagnosed prior to treatment [1]. Along with many other factors, unnecessary prescriptions raise the risk of resistance emerging by providing opportunities for the parasites to experience sub-therapeutic drug concentrations [5]. A long list of antimalarials have become partially or totally ineffective due to the rise and spread of drug resistance [6–8]. The current front-line treatment, artemisinin combination therapy (ACT), itself has started to report cases of resistance in Southeast Asia [9–12].

Of the several options for clinical malaria diagnosis, the gold standard for over 100 years has been Giemsa stained blood smear microscopy. While modified and optimized, this method has remained largely the same since Giemsa stain was first introduced in 1904 [13,14]. A sample of patient blood is drawn and spread across a microscope slide to produce a so-called "thick smear."

The slide is then fixed in methanol and dyed with Giemsa stain. When observed under a light microscope, infection can be determined by the presence or absence of purple-stained parasites inside pink-stained red blood cells. With expert operators, this type of diagnosis can be highly accurate down to low parasite levels of 20-100 parasites/μL blood [15]. However, it is not without drawbacks. The accuracy and limit of detection is highly dependent upon the skill of the person preparing and reading the slide as well as the quality of supplies; accuracy in the field can be as low as 30% [16–20]. The high initial costs in terms of training and equipment also limit microscopy as a viable option in remote and resource-limited settings.

These limitations have led to the development of rapid diagnostic tests (RDTs) that provide faster results with lower skill and cost requirements. RDTs act as lateral flow antigen detection tests. A small sample of patient blood is placed in a well along with a running buffer and dye-bonded antibodies to both a *Plasmodium* (test) and human (control) antigens. The solution is drawn along a strip of chromatography paper, and, if present, the test and control antigens are captured and concentrated along with their matching dye-containing antibodies in two separate locations. If enough antigen, and thus dye-bonded antibody, is present, the dye will accumulate to a level that makes it visible to the human eye as a colored band on the paper strip [21]. The parasite antigen in most currently employed RDTs is either histidine-rich protein 2 (HRP2) or parasite lactate dehydrogenase (pLDH). HRP2, which is specific to *P. falciparum*, is the most widely employed test antigen in Africa. In terms of accuracy, RDTs have proven equivalent or superior to microscopy in endemic regions [15,22–25] and are now the diagnostic of choice in multiple settings with 312 million RDTs delivered worldwide in 2017 [1]. Furthermore, multiple studies have shown that the introduction of RDTs decreases antimalarial over-prescription, albeit

from a high base rate [26–28]. However, while more robust to user errors and product failures

than microscopy, the accuracy of RDTs can also suffer dramatic reductions if not handled

properly, and several tests on the market are unreliable at parasite burdens below 200

parasites/µL blood [29–31]. Since HRP2 can remain at detectable levels in the bloodstream for

up to a month after infection clearance, RDTs with HRP2 as the target antigen also have an

intrinsic false positive rate [32].

Another mounting issue for HRP2-based RDTs is the spread of parasites that have stopped

producing the antigen. The first sub-population of parasites lacking HRP2 was documented in

2010 in Peru, a follow-up study demonstrating that the percentage of HRP2-negative parasites

had increased from 20% to 40% from 1998-2001 to 2003-2005 in the same region [33,34]. Since

then, significant HRP2-negative sub-populations have been reported in India and Africa; in some

cases, the HRP2-negative status has been linked to false negative HRP2-RDT results [35–39].

One mathematical modeling paper demonstrated that, given certain assumptions, HRP2-

negativity could spread throughout much of Africa [40]. However, the WHO and others have

urged caution in extrapolating from recent reports and suggested guidelines to better assess the

extent of HRP2-negativity moving forward [41,42]. Still, in the same report, the WHO called for

identifying new target antigens.

Beyond new antigens for lateral flow detection type RDTs, new technologies are needed to help

improve malaria diagnosis. Several new approaches have been proposed, with a few like loop-

mediated isothermal amplification and insulated isothermal PCR having shown success in initial

field trials (reviewed in [43]). The development of more accurate and reliable diagnostics should

help to mitigate the over-prescription of antimalarials due to lack of confidence in RDTs by health providers and/or patients [44]. Additionally, space exists for devices targeted at certain scenarios, such as case management in low-resource countries and malaria elimination surveillance efforts [45]. Non-invasive sampling has been identified specifically as a desirable feature for future diagnostics [43].

1.4 Introduction to breath VOCs

The idea that the odor of a patient, particularly breath, can aid in diagnosis has a long history. Diagnostic odors range from the sickly sweet smell associated with diabetes to the urine smell indicative of uremia or kidney failure [46]. The compounds behind these scents, or other undetectable gaseous molecules that might be a sign of disease, were not known. The first step toward a more holistic understanding of breath compounds came when Linus Pauling noted over two hundred compounds present in breath by mass spectrometry in 1971 [47]. Indeed, a typical human breath contains several hundred compounds with many hundreds of compounds having been noted across currently available work [48].

The nature of the compounds in the breath is understandably diverse, ranging from industrial by-products to biogenic molecules. All these compounds fall under the broad category of volatile organic compounds (VOCs), organic molecules that partition into the gas phase at approximately room temperature. VOCs range from one to eighteen carbons long, their volatility decreasing with size, amount of unsaturation, and number of functional groups [48]. While over a thousand VOCs have been noted in human breath, there are patterns and constants. Two compounds,

7

isoprene and acetone, are virtually universally present in human breath and are also the most abundant compounds in breath, with levels in the high parts-per-billion (ppb) to low parts-per-million range (ppm) [49–51]. All other VOCs, barring occasional high prevalence contaminants from diet or environmental scents or hazards, tend to be in the parts-per-trillion (ppt) to low ppb range with a large amount of observed variability [49]. This high variability in presence and abundance of VOCs is a fact of human breath research. For compounds originating in the environment or from ingested food, for example benzene (released from gasoline) and dimethyl sulfide (produced by garlic consumption), this variability can be explained by the large variety of environs and diets humans are exposed to [52,53]. Lifestyle choices can impact breath VOCs too, with smoking, for example, having especially dramatic effects on breath compounds [54]. For compounds thought to originate from human or microbiome metabolism, the causes for this variation are not well understood, but metabolism is exquisitely attuned to a wide range of stimuli, and the resulting VOC by-products reflect that fact. One example is the rapid change in isoprene levels brought on by physical exertion [55].

Determining whether a particular VOC is endogenous and what metabolic reaction produced it, though, is difficult. A common method to assess if a VOC is from an endogenous biological process versus the environment is to compare how prevalent it is in different sections of the breath [56]. The earliest part of the breath has the highest amount of contaminants from the environment and the oral cavity while the end, alveolar portion has the highest amount of compounds present in the blood or body cavity moving into the gas phase in the lungs. Comparisons can also be made between breath and room air samples. Regardless, if the VOC is higher in the end breath versus early breath (or breath versus room air), it is thought to be

8

endogenous. This method is not without its caveats (reviewed in [57]). Foremost among those caveats is that it does not discriminate between environmental VOC contaminants that become concentrated in the body versus VOCs produced in the body [58]. Even if a VOC is "endogenous," it could be a product of the microbiome rather than human metabolism [59]. Foreign pathogens could also contribute, a possibility explored in the next section.

Hundreds of breath VOCs are believed to have a human endogenous origin, but for most there is no evidence of their exact metabolic origins. Evidence for metabolic origins of VOCs tends to be correlative, such as ketogenic diets raising acetone levels, and thus acetone itself likely resulting from ketolytic processes [60]. Similarly, short chain fatty acids in breath were elevated after a glucose challenge and thus were linked to glucose metabolism [61]. Other lines of inquiry involve determining the VOCs generated by cell lines and then extrapolating to the metabolic processes known to be dominant in those lines (usually done in the context of cancer cell lines) [62–65]. More generally, the putative origins of VOCs are often based on either the molecule itself or a likely precursor being present in established human metabolic pathways [66].

1.5 Breath biomarkers of disease

Over the past decade, breath diagnostics have garnered increasing attention both for diseases lacking robust diagnostic options and as a non-invasive, easy-to-use alternative to established diagnostic tools (reviewed in [67,68]). To highlight a few notable examples, tuberculosis was diagnosed with over 90% accuracy in two recent field trials, one using a pilot point of care device and one using a potentially portable sensor system [69,70]. These point of care device

studies came after a string of studies demonstrating unique VOC profiles from both *Mycobacterium* cultures and infected patients [71–73]. Similarly, a set of sesquiterpenes was found to be indicative of *Aspergillus* in culture and then in breath samples from patients suffering from aspergillosis [74–76]. More broadly, multiple infectious agents have been linked to unique VOC profiles, whether from cultured specimens or infected patient breath (reviewed in [77,78]). Evidence for malaria is discussed in Chapter 3 and Appendix A. Outside of infection, chronic obstructive pulmonary disease (COPD) and lung cancer have been extensively investigated, and a large clinical trial for point-of-care lung cancer breath diagnosis is ongoing [79–82].

Discovery and validation of breath biomarkers plays a crucial role in establishing breath diagnostics for a given disease [67]. Discovery is the first step: a disease of interest has to be investigated to see whether it generates an observable change in affected patients' breath VOC profile prior to further work. These breath biomarker studies involve breath collection, VOC capture, analysis of VOCs, data processing, and determination of biomarkers by classification algorithms and comparative statistics.

While collecting breath may seem a straight-forward proposition, there are many variables to consider. Population size, the section of breath to be collected, and the mechanics of collecting the breath are all integral parts of a successful experimental design [83]. In exceptional cases where the disease of interest would be thought to influence VOCs in the upper airway or oral cavity, such as oral cancer or viral throat infections, alveolar breath is not the desired portion. The last section of the breath, a.k.a. the alveolar section, tends to be richest in VOCs reflecting

10

biological processes in the body and lowest in contaminants and thus is usually the most suitable breath portion for collection [56]. However, it can be difficult to consistently collect solely alveolar breath; sometimes this is achieved by having the patients exhale briefly before collection or by simply discarding the first few hundred milliliters of breath volume. Either approach can be hard to perform consistently across the entire test population though, so whole breath is sometimes collected to avoid such issues [57,83]. Recently, advanced devices that track CO_2 levels (which exist in a gradient across each breath) and breath volume in real time have made consistent collection more achievable [84–86]. These devices also act to transfer the desired portion of breath directly onto a VOC capturing material. In most other cases, an intermediate step is required, with breath first collected in an inert container and then transferred to VOC capturing absorbent material. Examples of the inert containers include glass bulbs, iron gas canisters, inert polymer sampling bags, and inert syringes (reviewed in [68,87]). Sampling bags and one type of inert syringe, the BioVOC, are the two most widely utilized methods; their comparative merits and drawbacks are the subject of Chapter 4.

After breath is collected, the VOCs present in the sample typically need to be captured for transport and later analysis. Samples can be maintained in sampling bags, glass bulbs, etc. for a period of time, but VOCs tend to diffuse out, react with the container surface, or otherwise degrade over a period of days [88,89]. Commonly, the breath sample will be stored for a few hours before VOC capture. For much breath research, capture involves driving the breath sample through a thermal desorption (TD) tube, which is a small glass or metal tube packed with tens to hundreds of milligrams of an absorbent resin that traps VOCs via absorption. Absorptive material can be thought of as a VOC sink: all VOCs with an affinity for the material will bind

11

until the material becomes saturated. When the time comes to analyze the bound VOCs, they are all released near-simultaneously by thermal desorption, which is achieved by rapidly heating the tube. The various properties, advantages, types of material, and permutations of TD tubes have been extensively reviewed by Woolfenden [90,91]. One important point is that VOCs bound in a TD tube stay quite stable for at least two weeks [92]. Finally, as an alternative to TD tubes, solid phase microextraction (SPME) fibers, which work via different physical processes, are also employed for capturing breath VOCs [87,93].

For analysis of captured VOCs, the most common platform is gas chromatography mass spectrometry (GC/MS). This technology has been reviewed extensively elsewhere [94,95]. In brief, compounds are separated in time by the GC portion as they pass through a long non-polar column. Thus separated, only one to a few compounds reach the MS detector at a time. Upon entering the MS, the compounds are fragmented in a characteristic and consistent manner by bombardment with high energy electrons, and the molecular weight of each fragment is then determined. The molecular fragment data (usually represented as mass to charge, or m/z) along with the retention time from GC separation is relatively unique to any given compound. The combined amount of all fragments also gives an abundance readout for a compound. Thus, even with a matrix of hundreds of compounds such as breath, the structural characteristics and quantity of all the compounds can be determined. Two common analytical alternatives to GC/MS are selected ion flow tube (SIFT) MS and proton transfer reaction (PTR) MS (reviewed in [96,97]). Unlike GC/MS, SIFT and PTR breath samples are directly injected without pre-concentration or capture. This method avoids the problems of VOC capture and degradation at the cost of not being able to concentrate VOCs or collect samples off-site. Additionally, SIFT

and PTR tend to have higher sensitivity, but provide less definitive structural data and fail to detect as broad a range of compounds. A final analytical option is sensor arrays and other so-called electronic noses (eNoses). While occasionally used in biomarker discovery, they provide minimal to no data on the identity and quantity of individual VOCs and are of more interest in later stages as a point-of-care device after biomarkers have already been established (see below).

After analyzing the samples, the raw data must be further processed to allow conclusions to be drawn. Due to the complex and often subtle interaction between a given disease and exhaled breath volatiles, the VOC changes are rarely able to be discerned "by eye." For GC/MS, raw data is deconvoluted and aligned, generating a list of compounds and their abundance in each sample (discussed in detail [98,99]). Except for *a priori* VOCs of interest, univariate statistical methods cannot grapple with the high dimensionality of breath data. Principle component analysis (PCA) is useful for visualizing the overall distribution of samples and flagging outliers, but disease state is rarely the main driver of VOC variation. Thus, machine learning techniques are the tool of choice for identifying potential biomarkers [99,100]. Most relevant to biomarker discovery are supervised techniques, algorithms that detect the compounds that best classify the samples based on a parameter, in this case, disease state. The compounds so identified can be further investigated with univariate techniques, or, ideally, the supervised classifier model can be validated with a second, independent data set.

After initial biomarker discovery, the next step in moving towards a true diagnostic is to validate the initial findings. Performing follow-on studies to ascertain whether biomarkers can be found across several study populations is critical due to the high variability inherent in breath data [67].

13

Assuming these studies are successful, the final step is creating a true point-of-care (PoC) devices that can be applied in a clinical setting. Most of the efforts to transition from laboratory techniques to PoC have been with electronic noses (eNoses). This moniker applies to a wide range of different sensor array systems, all of which output distinct patterns corresponding to the overall composition of a breath sample. Knowledge of the chemical classes and abundances of potential biomarkers help in selecting and tuning sensor systems for optimal performance. While increasingly cheap and reliable, eNoses are only now starting to become viable as clinical diagnostics, and there are still technical barriers to overcome [101]. If a few high abundance or chemically distinct biomarkers are established for a disease, a targeted sensor could be designed. However, it is rare for a disease to be diagnosable by a small number of breath VOCs. Alternative technologies based on miniaturized mass spectrometry devices are another potential avenue for developing PoC breath diagnostics [68].

1.6 Introduction to isoprenoids

Isoprenoids are perhaps the most diverse class of naturally occurring compounds with over 25,000 known, but all derive from the same precursor: isopentyl pyrophosphate (IPP), also called isopentyl diphosphate (IDP) [102]. There exist two dedicated pathways for generating IPP. In mammals, some bacteria, and plant cytosol, the mevalonate pathway is used. In Apicomplexa, the majority of eubacteria, and plant chloroplasts, the methylerythritol phosphate (MEP) pathway is used. Despite producing the same end product, the two pathways are metabolically distinct with no overlapping enzymes or intermediates. With Apicomplexa, including *Plasmodium*, the entire synthesis pathway is localized to the eponymous apicoplast, a non-photosynthetic plastid

organelle derived from a secondary endosymbiotic event with a red algae [103,104]. Notably, the MEP pathway enzymes are one of several plant-like systems retained from the endosymbiotic algal ancestor.

IPP itself is a five-carbon building block with a high-energy pyrophosphate group helping to drive various addition and rearrangement reactions. IPP and its isomer dimethylallyl pyrophosphate (DMAPP) are condensed to form a 10-carbon molecule (geranyl pyrophosphate, GPP) to which additional IPP molecules can be added to create 15- (farnesyl pyrophosphate, FPP) and then 20- (geranylgeranyl pyrophosphate, GGPP) carbon molecules [104]. These molecules can be further modified by cyclization, rearrangement, and/or addition of other biomolecules to form the numerous down-stream products. As might be expected for such a diverse class of compounds, isoprenoids play a multitude of roles, including as antibiotics, signaling molecules, and lipid membrane stabilizers [103]. Two classes of isoprenoids are of particular importance in the context of the following chapters: prenyl groups and small volatile terpenes.

Prenylation is the process of attaching a 15-carbon FPP or 20-carbon GGPP to the C-terminal end of a protein. Ubiquitous across eukaryotes, prenylation is an essential function; the prenyl groups facilitate membrane association and thus proper localization for substrate proteins. Important roles for prenylated proteins include extensive involvement in endomembrane trafficking as well as cell polarity and growth [105]. Several proteins become oncogenic due to improper prenylation, e.g. K-Ras, and prenylation inhibitors have garnered widespread interest as possible chemotherapeutics [106,107]. Multiple studies have noted inhibition of parasite

15

growth when treating *Plasmodium* with various prenylation inhibitors, leading to the conclusion that prenylation is essential for *Plasmodium* [108–110]. These studies have also engendered interest in "piggy-backing" prenylation inhibitors developed for cancer as potential new anti-malarials [111,112].

Three classes of proteins catalyze prenylation: farnesyl transferase (FT), geranylgeranyl transferase type I (GGT-1), and geranylgeranyl transferase type II (GGT-2). FT and GGT-1 recognize the "CaaX" box, a C-terminal domain consisting of a cysteine followed by two aliphatic amino acids and a variable terminal amino acid. The identity of the final three amino acids broadly determines whether the protein will be a substrate for a FT or GGT-1, though these rules are not definitive [113]. Both of the transferases dephosphorylate and conjugate a prenyl molecule (either a FPP or GGPP) to the cysteine in the CaaX box, and the final three amino acids are subsequently cleaved off by a separate enzyme [105]. Rather than a CaaX motif, the GGT-2 recognizes a terminal CC or CXC, with additional proximal amino acids contributing to substrate recognition. These type II transferases dephosphorylate and conjugate two GGPP molecules, one to each cysteine [105]. Despite annotations for all three transferases and evidence of prenylation occurring, only a handful of *Plasmodium* proteins were known to be prenylated until recently, as discussed in Chapter 4 [114,115].

In addition to prenyl groups, the subset of small, volatile, hydrocarbon isoprenoids known as terpenes is of particular interest to the current work. The two most common varieties, monoterpenes and sesquiterpenes, are formed, respectively, by geranyl pyrophosphate (GPP) or FPP being dephosphorylated to form a reactive carbocation before undergoing an intramolecular

attack and rearrangement. Typically, the result is a cyclic or even poly-cyclic hydrocarbon.

While all terpenes proceed by a carbocation intermediate, a wealth of possible end products are

possible, and hundreds of terpene synthase enzymes have been described, the exact product

dictated by the properties of the enzyme-binding pockets [116]. With the exception of a select

few non-plant eukaryotes, terpene synthesis is limited to prokaryotes, fungi, and plants [117].

One well-described role for these compounds is as defense molecules released to deter or kill

predators and parasites. Their other main role is as signaling molecules both to members of the

same species as well as to other, mutualist species [102]. It is in this capacity that terpenes are

recognized as the scents and aromas associated with plants, especially flowers and fruiting

bodies.

A key example of terpenes as signaling molecules is their use by plants to attract pollinating

insects, a phenomenon observed, or at least inferred, across multiple species [118]. Mosquitoes,

specifically the malaria vector genus *Anopheles*, are no exception, with several studies observing

mosquito attraction to terpene-containing blends mimicking plant aromas [119–121]. Mosquito

odorant receptor response to terpenes is discussed further in Appendix A. As female mosquitoes

both take blood meals and feed on plant nectar, they are also attracted to human-associated

scents. Both human scent mimicking and mosquito-attractive plant scents have been piloted in

mosquito mass trapping campaigns, a promising strategy for malaria control [122,123]. An

intriguing question has been whether infection status could serve to influence host odors in a way

that augments mosquito attraction. A number of studies have found evidence that this does occur,

but the mechanism was unknown [124–128]. How this intersects with terpenes is part of the data

explored in Chapters 2 and 3 and Appendixes A and B.

1.7 References

1. WHO Global Malaria Programme. World Malaria Report 2017. Geneva: WHO; 2017.

2. Ahmed MA, Cox-Singh J. *Plasmodium knowlesi* - an emerging pathogen. ISBT Sci Ser. 2015;10: 134–140.

3. Cowman AF, Healer J, Marapana D, Marsh K. Malaria: Biology and Disease. Cell. 2016;167: 610–624.

4. Gachot B, Ringwald P. Severe malaria. Rev du Prat. 1998;48: 273–278.

5. Stepniewska K, White NJ. Pharmacokinetic determinants of the window of selection for antimalarial drug resistance. Antimicrob Agents Chemother. 2008;52: 1589–1596.

6. Wells TNC, Van Huijsduijnen RH, Van Voorhis WC. Malaria medicines: A glass half full? Nat Rev Drug Discov. 2015;14: 424–442.

7. McClure NS, Day T. A theoretical examination of the relative importance of evolution management and drug development for managing resistance. Proc R Soc B Biol Sci. 2014;281: 1861.

8. Petersen I, Eastman R, Lanzer M. Drug-resistant malaria: Molecular mechanisms and implications for public health. FEBS Lett. Federation of European Biochemical Societies; 2011;585: 1551–1562.

9. Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, et al. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. N Engl J Med. 2014;371: 411–423.

10. Dondorp AM, Smithuis FM, Woodrow C, Seidlein L von. How to contain artemisinin- and multidrug-resistant falciparum malaria. Trends Parasitol. 2017;33: 353–363.

11. Ménard D, Khim N, Beghain J, Adegnika AA, Shafiul-Alam M, Amodu O, et al. A worldwide map of *Plasmodium falciparum* K13-propeller polymorphisms. N Engl J Med. 2016;374: 2453–2464.

12. Takala-Harrison S, Jacob CG, Arze C, Cummings MP, Silva JC, Dondorp AM, et al. Independent emergence of artemisinin resistance mutations among *Plasmodium falciparum* in Southeast Asia. J Infect Dis.

13. Fleischer B. Editorial: 100 years ago: Giemsa's solution for staining of plasmodia. Trop Med Int Heal 2004;9: 755–756.

14. Stain TG. History of Surgical Pathology The Giemsa Stain: Its history and applications. Int J Surg Pathol. 2009; 292–296.

15. Wongsrichanalai C, Barcus MJ, Muth S, Sutamihardja A, Wernsdorfer WH. A review of malaria diagnostic tools: microscopy and rapid diagnostic test (RDT). Am J Trop Med Hyg. 2007;77: 119–127.

16. Kahama-maro J, Acremont VD, Mtasiwa D, Genton B, Lengeler C. Kahama-Maro J, et al. Low quality of routine microscopy for malaria at different levels of the health system in Dar es Salaam. Malar J. 2011;10: 332.

17. Zurovac D, Midia B, Ochola SA, English M, Snow RW. Microscopy and outpatient malaria case management among older children and adults in Kenya. Trop Med Int Heal. 2006;11: 432–440.

18. The World Health Organization Regional Office for the Western Pacific. Malaria light microscopy. Creating a culture of quality. 2005.

19. K. Allen L, M. Hatfield J, J. Manyama M. Reducing microscopy based malaria misdiagnosis in a low- resource area of Tanzania. Tanzan J Health Res. 2013;15: 1–9.

20. Coleman RE, Maneechai N, Rachaphaew N, Kumpitak C, Miller RS, Soyseng V, et al. Comparison of field and expert laboratory microscopy for active surveillance for asymptomatic *Plasmodium falciparum* and *Plasmodium vivax* in Western Thailand. Am J Trop Med Hyg. 2002;67: 141–144.

21. Bell D, Wongsrichanalai C, Barnwell JW. Ensuring quality and access for malaria diagnosis: how can it be achieved? Nat Rev Microbiol. 2006;4: S7–S20.

22. Francony C, Sebastiao YV, Pires E, Gamboa D, Nery SV. Performance of microscopy and RDTs in the context of a malaria prevalence survey in Angola: a comparison using PCR as the gold standard. Malar J. 2014;12: 284.

23. Batwala V, Magnussen P, Nuwaha F. Are rapid diagnostic tests more accurate in diagnosis of *Plasmodium falciparum* malaria compared to microscopy at rural health centres? Malar J. 2010;9: 349.

24. Batwala V, Magnussen P, Nuwaha F. Comparative feasibility of implementing rapid diagnostic test and microscopy for parasitological diagnosis of malaria in Uganda. Malar J. 2011;10: 1–9.

25. Kyabayinze DJ, Zongo I, Cunningham J, Gatton M, Angutoko P, Ategeka J, et al. HRP2 and pLDH-based rapid diagnostic tests, expert microscopy, and PCR for detection of malaria infection during pregnancy and at delivery in areas of varied transmission: a prospective cohort study in Burkina Faso and Uganda. PLoS One. 2016;11: 1–15.

26. Yukich JO, Bennett A, Albertini A, Incardona S, Moonga H, Chisha Z, et al. Reductions in artemisinin-based combination therapy consumption after the nationwide scale up of routine malaria rapid diagnostic testing in Zambia. Am J Trop Med Hyg. 2012;87: 437–446.

27. Thiam S, Thior M, Faye B, Ndiop M, Diouf ML, Diouf MB, et al. Major reduction in anti-malarial drug consumption in Senegal after nation-wide introduction of malaria rapid diagnostic tests. PLoS One. 2011;6.

28. Zurovac D, Githinji S, Memusi D, Kigen S, Machini B, Muturi A, et al. Major improvements in the quality of malaria case-management under the "test and treat" policy in Kenya. PLoS One. 2014;9.

29. Maltha J, Gillet P, Jacobs J. Malaria rapid diagnostic tests in endemic settings. Clin Microbiol Infect. European Society of Clinical Infectious Diseases; 2013;19: 399–407.

30. Ruizendaal E, Dierickx S, Peeters Grietens K, Schallig HDFH, Pagnoni F, Mens PF. Success or failure of critical steps in community case management of malaria with rapid diagnostic tests: a systematic review. Malar J. 2014;13: 229.

31. World Health Organization (WHO). Malaria rapid diagnostic test performance. WHO. 2012;4: 140.

32. Wilson ML. Malaria rapid diagnostic tests. Clin Infect Dis. 2012;54: 1637–1641.

33. Gamboa D, Ho MF, Bendezu J, Torres K, Chiodini PL, Barnwell JW, et al. A large proportion of *P. falciparum* isolates in the Amazon region of Peru lack pfhrp2 and pfhrp3: Implications for malaria rapid diagnostic tests. PLoS One. 2010;5.

34. Akinyi S, Hayden T, Gamboa D, Torres K, Bendezu J, Abdallah JF, et al. Multiple genetic origins of histidine-rich protein 2 gene deletion in *Plasmodium falciparum* parasites from Peru. Sci Rep. 2013;3: 1–8.

35. Berhane A, Russom M, Bahta I, Hagos F, Ghirmai M, Uqubay S. Rapid diagnostic tests failing to detect *Plasmodium falciparum* infections in Eritrea: An investigation of reported false negative RDT results. Malar J. 2017;16: 1–6.

36. Kumar N, Pande V, Bhatt RM, Shah NK, Mishra N, Srivastava B, et al. Genetic deletion of HRP2 and HRP3 in Indian *Plasmodium falciparum* population and false negative malaria rapid diagnostic test. Acta Trop; 2013;125: 119–121.

37. Menegon M, L'Episcopia M, Nurahmed AM, Talha AA, Nour BYM, Severini C. Identification of *Plasmodium falciparum* isolates lacking histidine-rich protein 2 and 3 in Eritrea. Infect Genet Evol. 2017;55: 131–134.

38. Beshir KB, Sepúlveda N, Bharmal J, Robinson A, Mwanguzi J, Busula AO, et al. *Plasmodium falciparum* parasites with histidine-rich protein 2 (pfhrp2) and pfhrp3 gene deletions in two endemic regions of Kenya. Sci Rep. 2017;7: 1–10.

39. Koita OA, Doumbo OK, Ouattara A, Tall LK, Konaré A, Diakité M, et al. False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. Am J Trop Med Hyg. 2012;86: 194–198.

40.   Watson OJ, Slater HC, Verity R, Parr JB, Mwandagalirwa MK, Tshefu A, et al. Modelling the drivers of the spread of *Plasmodium falciparum* hrp2 gene deletions in sub-Saharan Africa. Elife. 2017;6: e25008.

41.   World Health Organization (WHO). False-negative RDT results and implications of new reports of *P. falciparum* histidine-rich protein 2/3 gene deletions. Glob Malar Program. 2017; 1–11.

42.   Cheng Q, Gatton ML, Barnwell J, Chiodini P, McCarthy J, Bell D, et al. *Plasmodium falciparum* parasites lacking histidine-rich protein 2 and 3: a review and recommendations for accurate reporting. Malar J. 2014;13: 283.

43.   The malERA Consultative Group on Diagnoses. A research agenda for malaria eradication: Diagnoses and diagnostics. PLoS Med. 2011;8.

44.   World Health Organization (WHO). Universal access to malaria diagnostic testing: an operational manual. 2011.

45.   Bell D, Fleurent AE, Hegg MC, Boomgard JD, McConnico CC. Development of new malaria diagnostics: matching performance and need. Malar J. 2016;15: 406.

46.   Shirasu M, Touhara K. The scent of disease: Volatile organic compounds of the human body related to disease and disorder. J Biochem. 2011;150: 257–266.

47.   Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. Proc Natl Acad Sci. 1971;68: 2374–2376.

48.   de Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, et al. A review of the volatiles from the healthy human body. J Breath Res. 2014;8: 14001.

49.   Mochalski P, King J, Klieber M, Unterkofler K, Hinterhuber H, Baumann M, et al. Blood and breath levels of selected volatile organic compounds in healthy volunteers. Analyst. 2013;138: 2134–45.

50.   Schwarz K, Pizzini A, Arendacká B, Zerlauth K, Filipiak W, Schmid A, et al. Breath acetone - Aspects of normal physiology related to age and gender as determined in a PTR-MS study. J Breath Res. 2009;3.

51.   Turner C, Španěl P, Smith D. A longitudinal study of breath isoprene in healthy volunteers using selected ion flow tube mass spectrometry (SIFT-MS). Physiol Meas. 2006;27: 13–22.

52.   Wallace LA. Major sources of benzene exposure. Environ Health Perspect. 1989;82: 165.

53.   Taucher J, Hansel A, Jordan A, Lindinger W. Analysis of compounds in human breath after ingestion of garlic using proton-transfer-reaction mass spectrometry. J Agric Food Chem. 1996;44: 3778–3782.

54.   Buszewski B, Ulanowska A, Ligor T, Denderz N, Amann A. Analysis of exhaled breath from smokers, passive smokers and non-smokers by solidphase microextraction gas chromatography/mass spectrometry. Biomed Chromatogr. 2009;23: 551–556.

55.   King J, Kupferthaler A, Unterkofler K, Koc H, Teschl S, Teschl G, et al. Isoprene and acetone concentration profiles during exercise on an ergometer. J Breath Res. 2009;3.

56.   Van Den Velde S, Quirynen M, Van Hee P, Van Steenberghe D. Differences between alveolar air and mouth air. Anal Chem. 2007;79: 3425–3429.

57.   Pleil JD, Stiegel MA, Risby TH. Clinical breath analysis: Discriminating between human endogenous compounds and exogenous (environmental) chemical confounders. J Breath Res. 2013;7.

58.   Beauchamp J. Inhaled today, not gone tomorrow: Pharmacokinetics and environmental exposure of volatiles in exhaled breath. J Breath Res. 2011;5.

59.   Leja M, Amal H, Lasina I, Skapars R, Sivins A, Ancans G, et al. Analysis of the effects of microbiome-related confounding factors on the reproducibility of the volatolomic test. J Breath Res. 2016;10: 37101.

60.   Musa-Veloso K, Likhodii SS, Cunnane SC. Breath acetone is a reliable indicator of ketosis in adults consuming ketogenic meals. Am J Clin Nutr. 2002;76: 65–70.

61.   Gruber B, Keller S, Groeger T, Matuschek G, Szymczak W, Zimmermann R. Breath gas monitoring during a glucose challenge by a combined PTR-QMS/GCxGC-TOFMS approach for the verification of potential volatile biomarkers. J Breath Res. 2016;10: 36003.

62.   Calenic B, Filipiak W. Volatile organic compounds expression in different cell types: an in vitro approach. Int J Clin Tox. 2013; 43–51.

63.   Aksenov AA, Gojova A, Zhao W, Morgan JT, Sankaran S, Sandrock CE, et al. Characterization of volatile organic compounds in human leukocyte antigen heterologous expression systems: a cell's "chemical odor fingerprint." ChemBioChem. 2012;13: 1053–1059.

64.   Filipiak W, Mochalski P, Filipiak A, Ager C, Cumeras R, E. Davis C, et al. A compendium of volatile organic compounds (VOCs) released by human cell lines. Curr Med Chem. 2016;23: 2112–2131.

65.   Zimmermann D, Hartmann M, Moyer MP, Nolte J, Baumbach JI. Determination of volatile products of human colon cell line metabolism by GC/MS analysis. Metabolomics. 2007;3: 13–17.

66.   Hakim M, Broza YY, Barash O, Peled N, Phillips M, Amann A, et al. Volatile organic compounds of lung cancer and possible biochemical pathways. Chem Rev. 2012;112: 5949–5966.

67.   Boots AW, Bos LD, van der Schee MP, van Schooten FJ, Sterk PJ. Exhaled molecular fingerprinting in diagnosis and monitoring: validating volatile promises. Trends Mol Med. 2015;21: 633–644.

68.   Rattray NJW, Hamrang Z, Trivedi DK, Goodacre R, Fowler SJ. Taking your breath away: Metabolomics breathes life in to personalized medicine. Trends Biotechnol. 2014;32: 538–548.

69.   Coronel Teixeira R, Rodríguez M, Jiménez de Romero N, Bruins M, Gómez R, Yntema JB, et al. The potential of a portable, point-of-care electronic nose to diagnose tuberculosis. J Infect. 2017;75: 441–447.

70.   Zetola NM, Modongo C, Matsiri O, Tamuhla T, Mbongwe B, Matlhagela K, et al. Diagnosis of pulmonary tuberculosis and assessment of treatment response through analyses of volatile compound patterns in exhaled breath samples. J Infect. 2016;74: 367–376.

71.   Mourão MPB, Kuijper S, Dang NA, Walters E, Janssen HG, Kolk AHJ. Direct detection of *Mycobacterium tuberculosis* in sputum: A validation study using solid phase extraction-gas chromatography-mass spectrometry. J Chromatogr B Anal Technol Biomed Life Sci. 2016;1012–1013: 50–54.

72.   Syhre M, Chambers ST. The scent of *Mycobacterium tuberculosis*. Tuberculosis. 2008;88: 317–323.

73.   Phillips M, Cataneo RN, Condos R, Ring Erickson GA, Greenberg J, La Bombardi V, et al. Volatile biomarkers of pulmonary tuberculosis in the breath. Tuberculosis. 2007;87: 44–52.

74.   Heddergott C, Latgé JP, Calvo AM. The volatome of *Aspergillus fumigatus*. Eukaryot Cell. 2014;13: 1014–1025.

75.   Koo S, Thomas HR, Daniels SD, Lynch RC, Fortier SM, Shea MM, et al. A breath fungal secondary metabolite signature to diagnose invasive aspergillosis. Clin Infect Dis. 2014;59: 1733–1740.

76.   de Heer K, Kok M, Fens N, Weersink EJM, Zwinderman AH, van der Schee MP, et al. Detection of airway colonization by *Aspergillus fumigatus* using electronic nose technology in patients with cystic fibrosis. J Clin Microbiol. 2015;54: 569–575.

77.   Ahmed WM, Lawal O, Nijsen TM, Goodacre R, Fowler SJ. Exhaled volatile organic compounds of infection: a systematic review. ACS Infect Dis. 2017.

78.   Bos LDJ, Sterk PJ, Schultz MJ. Volatile metabolites of pathogens: a systematic review. PLoS Pathogens. 2013.

79.   Christiansen A, Davidsen JR, Titlestad I, Vestbo J, Baumbach J. A systematic review of breath analysis and detection of volatile organic compounds in COPD. J Breath Res.

2016;10.

80.   Queralto N, Berliner AN, Goldsmith B, Martino R, Rhodes P, Lim SH. Detecting cancer by breath volatile organic compound analysis: A review of array-based sensors. J Breath Res. 2014;8.

81.   Haick H, Broza YY, Mochalski P, Ruzsanyi V, Amann A. Assessment, origin, and implementation of breath volatile cancer markers. Chem Soc Rev. 2014;43: 1423–1449.

82.   van der Schee MPC, Boschmans J, Smith R, Parris R, Boyle B, Apthorp D, et al. Early detection of lung cancer through analysis of VOC biomarkers in exhaled breath: The LuCID study. Eur Respir J. 2017;50.

83.   Herbig J, Beauchamp J. Towards standardization in the analysis of breath gas volatiles. J Breath Res. 2014;8: 37101.

84.   Allsworth M, Apthorp D, Van Der Schee M, Smith R, Boschmans J, Kitchen S. Method for collecting a selective portion of a subject's breath. USA; 15/494,973, 2017.

85.   Doran S, Romano A, Hanna GB. Optimization of sampling parameters for standardized exhaled breath sampling. J Breath Res. 2017;12: 16007.

86.   Phillips M. Breath collection apparatus. US; 6,726,637, 2004.

87.   Lawal O, Ahmed WM, Nijsen TME, Goodacre R, Fowler SJ. Exhaled breath analysis: a review of "breath-taking" methods for off-line analysis. Metabolomics. 2017;13: 1–16.

88.   Kim Y-H, Kim K-H, Jo S-H, Jeon E-C, Sohn JR, Parker DB. Comparison of storage stability of odorous VOCs in polyester aluminum and polyvinyl fluoride Tedlar® bags. Anal Chim Acta. 2012;712: 162–7.

89.   Mochalski P, King J, Unterkofler K, Amann A. Stability of selected volatile breath constituents in Tedlar, Kynar and Flexfilm sampling bags. Analyst. 2013;138: 1405–18.

90.   Woolfenden E. Sorbent-based sampling methods for volatile and semi-volatile organic compounds in air Part 1: Sorbent-based air monitoring options. J Chromatogr A. 2010;1217: 2674–84.

91.   Woolfenden E. Sorbent-based sampling methods for volatile and semi-volatile organic compounds in air. Part 2. Sorbent selection and other aspects of optimizing air monitoring methods. J Chromatogr A. 2010;1217: 2685–94.

92.   van der Schee MP, Fens N, Brinkman P, Bos LD, Angelo MD, Nijsen TM, et al. Effect of transportation and storage using sorbent tubes of exhaled breath samples on diagnostic accuracy of electronic nose analysis. J Breath Res. 2013;7: 16002.

93.   Alonso M, Sanchez JM. Analytical challenges in breath analysis and its application to exposure monitoring. TrAC - Trends Anal Chem. 2013;44: 78–89.

94.   Sparkman OD, Penton Z, Kitson FG. Gas chromatography and mass spectrometry: a practical guide. 2011.

95.   Lopes AS, Cruz ECS, Sussulini A, Klassen A. Metabolomic strategies involving mass spectrometry combined with liquid and gas chromatography. In: Sussulini A, editor. Metabolomics: From Fundamentals to Clinical Applications. 2017. pp. 77–98.

96.   Chow KK, Short M, Zeng H. A comparison of spectroscopic techniques for human breath analysis. Biomed Spectrosc Imaging. 2012;1: 339–353.

97.   Smith D, Španěl P, Herbig J, Beauchamp J. Mass spectrometry for real-time quantitative breath analysis. J Breath Res. 2014;8.

98.   O'Callaghan S, De Souza DP, Isaac A, Wang Q, Hodkinson L, Olshansky M, et al. PyMS: A Python toolkit for processing of gas chromatography-mass spectrometry (GC-MS) data. Application and comparative study of selected tools. BMC Bioinformatics. 2012;13.

99.   Sugimoto M, Kawakami M, Robert M, Soga T. Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. Curr Bioinforma. 2012; 96–108.

100. Smolinska A, Hauschild AC, Fijten RRR, Dallinga JW, Baumbach J, Van Schooten FJ. Current breathomics - A review on data pre-processing techniques and machine learning in metabolomics breath analysis. J Breath Res. 2014;8.

101. Bikov A, Lázár Z, Horvath I. Established methodological issues in electronic nose research: How far are we from using these instruments in clinical settings of breath analysis? J Breath Res. 2015;9: 34001.

102. Gershenzon J, Dudareva N. The function of terpene natural products in the natural world. Nat Chem Biol. 2007;3: 408–14.

103. Guggisberg AM, Amthor RE, Odom AR. Isoprenoid biosynthesis in *Plasmodium falciparum*. Eukaryot Cell. 2014.

104. Imlay L, Odom AR. Isoprenoid metabolism in apicomplexan parasites. Curr Clin Microbiol Reports. 2014;1: 37–50.

105. Palsuledesai CC, Distefano MD. Protein prenylation: enzymes, therapeutics, and biotechnology applications. ACS Chem Biol. 2015;10: 51–62.

106. Berndt N, Hamilton AD, Sebti SM. Targeting protein prenylation for cancer therapy. Nat Rev Cancer. 2011;11: 775–791.

107. Ochocki JD, Distefano MD. Prenyltransferase inhibitors: treating human ailments from cancer to parasitic infections. Med Chem Commun. 2013;4: 476–492.

108. Glenn MP, Chang S-Y, Hornéy C, Rivas K, Yokoyama K, Pusateri EE, et al. Structurally

simple, potent, *Plasmodium* selective farnesyltransferase inhibitors that arrest the growth of malaria parasites. J Med Chem. 2006;49: 5710–5727.

109. Nallan L, Bauer KD, Bendale P, Rivas K, Yokoyama K, Hornéy CP, et al. Protein farnesyltransferase inhibitors exhibit potent antimalarial activity. J Med Chem. 2005;48: 3704–3713.

110. Wiesner J, Kettler K, Sakowski J, Ortmann R, Katzin AM, Kimura EA, et al. Farnesyltransferase inhibitors inhibit the growth of malaria parasites in vitro and in vivo. Angew Chemie Int Ed. 2004;43: 251–254.

111. Gelb MH, Van Voorhis WC, Buckner FS, Yokoyama K, Eastman R, Carpenter EP, et al. Protein farnesyl and N-myristoyl transferases: piggy-back medicinal chemistry targets for the development of antitrypanosomatid and antimalarial therapeutics. Mol Biochem Parasitol. 2003;126: 155–163.

112. Eastman RT, Buckner FS, Yokoyama K, Gelb MH, Van Voorhis WC. Fighting parasitic disease by blocking protein farnesylation. J Lipid Res. 2006;47: 233–240.

113. Maurer-Stroh S, Eisenhaber F. Refinement and prediction of protein prenylation motifs. Genome Biol. 2005;6: 1–15.

114. Chakrabarti D, Azam T, DelVecchio C, Qiu L, Park Y, Allen CM. Protein prenyl transferase activities of Plasmodium falciparum. Mol Biochem Parasitol. 1998;94: 175–184.

115. Chakrabarti D, Silva T Da, Barger J, Paquette S, Patel H, Patterson S, et al. Protein farnesyltransferase and protein prenylation in *Plasmodium falciparum*. J Biol Chem. 2002;277: 42066–42073

116. Chen F, Tholl D, Bohlmann J, Pichersky E. The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J. 2011;66: 212–229.

117. Chen X, Köllner TG, Jia Q, Norris A, Santhanam B, Rabe P, et al. Terpene synthase genes in eukaryotes beyond plants and fungi: Occurrence in social amoebae. Proc Natl Acad Sci. 2016;113.

118. Abbas F, Ke Y, Yu R, Yue Y, Amanullah S, Jahangir MM, et al. Volatile terpenoids: multiple functions, biosynthesis, modulation and manipulation by genetic engineering. Planta. 2017;246: 803–816.

119. Yu B-T, Ding Y-M, Mo J-C. Behavioural response of female *Culex pipiens pallens* to common host plant volatiles and synthetic blends. Parasites & Vectors. 2015;8: 598.

120. Nyasembe VO, Teal PEA, Mukabana WR, Tumlinson JH, Torto B. Behavioural response of the malaria vector *Anopheles gambiae* to host plant volatiles and synthetic blends. Parasit Vectors. 2012;5: 234.

121. Nikbakhtzadeh MR, Terbot JW, Otienoburu PE, Foster WA. Olfactory basis of floral preference of the malaria vector *Anopheles gambiae* (Diptera: Culicidae) among common African plants. J Vector Ecol. 2014;39: 372–383.

122. Homan T, Hiscox A, Mweresa CK, Masiga D, Mukabana WR, Oria P, et al. The effect of mass mosquito trapping on malaria transmission and disease burden (SolarMal): a stepped-wedge cluster-randomised trial. Lancet. 2016;388: 1193–1201.

123. Nyasembe VO, Tchouassi DP, Kirwa HK, Foster WA, Teal PEA, Borgemeister C, et al. Development and assessment of plant-based synthetic odor baits for surveillance and control of malaria vectors. PLoS One. 2014;9: 1–10.

124. Batista EPA, Costa EFM, Silva AA. *Anopheles darlingi* (Diptera: Culicidae) displays increased attractiveness to infected individuals with *Plasmodium vivax* gametocytes. Parasit Vectors. 2014;7: 251.

125. Busula AO, Verhulst NO, Bousema T, Takken W, de Boer JG. Mechanisms of *Plasmodium*-enhanced attraction of mosquito vectors. Trends Parasitol. 2017.

126. Cornet S, Nicot A, Rivero A, Gandon S. Malaria infection increases bird attractiveness to uninfected mosquitoes. Ecol Lett. 2013;16: 323–329.

127. De Moraes CM, Stanczyk NM, Betz HS, Pulido H, Sim DG, Read AF, et al. Malaria-induced changes in host odors enhance mosquito attraction. Proc Natl Acad Sci USA. 2014;111: 11079–84.

128. Lacroix R, Mukabana WR, Gouagna LC, Koella JC. Malaria infection increases attractiveness of humans to mosquitoes. PLoS Biol. 2005;3: e298.

Chapter 2: Breathprinting reveals malaria-associated biomarkers and mosquito attractants

Preface

The following work was performed by myself, Nalin Katta, Lucy B. Bollinger, Mwawi Mwale, Rachel Mlotha-Mitole, Indi Trehan, Barani Raman, and Audrey R. Odom John. I established the breath collection method, analyzed the breath samples by GC/MS, did preliminary analysis/quality control on the data, identified the elevated levels of monoterpenes, helped refine the data input for the correlation analysis, and interpreted the results of the classification model. LBB, IT, MM, and RM collected the breath samples and patient metadata. NK performed the correlation analysis and build the subsequent classification model. AOJ and BR helped myself and NK in designing the experiments. NK, LBB, AOJ, BR, IT, and I wrote the main manuscript text. NK prepared the Figures relating to the correlation analysis and I prepared all other Figures and Tables.

This chapter in its entirety has been published (Schaber CL, Katta N, Bollinger LB, Mwale M, Mlotha-Mitole R, Trehan I, Raman B, Odom John AR. Breathprinting Reveals Malaria-Associated Biomarkers and Mosquito Attractants. *Journal of Infectious Disease*. February 2018). Article available at: [http://dx.doi.org/10.1093/infdis/jiy072]. Supplemental Information can also be found at that web address. Reproduction is allowed by authors per the license agreement terms set out by *JID* and Oxford University Press.

For contributions to this work, the authors wish to thank Michelle Eckerle, Dana Hodge, Peter Kazembe, Robert Krysiak, Hans-Joerg Lang, Wentai Lou, Mark Manary, Jonathan Ngoma, Karl

2.1 Abstract

Current evidence suggests that malaria infection could alter patient breath metabolites, a phenomenon that could be exploited to create a breath-based diagnostic test. However, no study has explored this in a clinical setting. To investigate whether natural human malaria infection leads to a characteristic breath profile, we performed a field study in Malawi. Breath volatiles from children with and without uncomplicated falciparum malaria were analyzed by thermal desorption-gas chromatography/mass spectrometry. Using an unbiased, correlation-based analysis, we find that children with malaria have a distinct shift in overall breath composition. Highly accurate classification of infection status was achieved with a suite of six compounds. In addition, we find infection correlates with significantly higher breath levels of two mosquito-attractant terpenes, α-pinene and 3-carene. These findings attest to the viability of breath analysis for malaria diagnosis, identifies candidate biomarkers, and identifies plausible chemical mediators for increased mosquito attraction to malaria-infected patients.

2.2 Introduction

Malaria remains a critical global health concern that affects hundreds of millions of people each year [1]. The most deadly form, caused by the parasite *Plasmodium falciparum*, remains a particular burden throughout sub-Saharan Africa. Diagnostic testing for malaria is crucial for acute fever management in the clinic and also for public health campaigns aimed at monitoring and control [2]. Current clinical practice depends on the "gold standard" of microscopic examination of patient blood samples, with increasing use of rapid diagnostic tests (RDTs) based

31

on lateral flow format detection of parasite antigens [3,4]. Both methods can achieve high accuracy rates, but often face prohibitive cost and skill requirements in many endemic settings [3]. While RDTs demand fewer human and capital resources, a number of factors can lead to dramatically lower accuracy than microscopy [5]. Further, the most widespread RDTs, based on detection of the *P. falciparum* protein HRP2, have an intrinsic false positive rate, as the parasite-derived antigen remains in the bloodstream up to a month after infection clearance [6]. Worryingly, false negative results are now rising due to the spread of parasite populations lacking the HRP2 antigen in India, Peru, and Africa [7–12]. In some geographical regions, more than 20% of surveyed parasite infections already lack HRP2 [12]. In 2016, the WHO put out a call for "new test antigens" in response to growing concerns about current RDTs [13].

By investigating the existence and extent of breath biomarkers for malaria, new avenues for diagnostics become possible. Any given exhaled human breath contains hundreds of different molecules, known as volatile organic compounds (VOCs) due to their ready partition into the gas phase, and thousands of breath VOCs have been described [14]. Breath-based diagnosis operates on the presumption that pathological conditions create characteristic and reproducible changes in breath VOCs, as has been reported for an increasing number of malignancies and infectious diseases [15–18]. Determining if a given disease generates a unique, detectable breath VOC signature, i.e. a "breathprint," represents the first step in development of a breath-based diagnostic, which has the possibility to be non-invasive and easy to perform [19].

Preliminary studies indicate malaria could generate just such a breathprint. For example, a number of alterations in breath compounds were observed during experimental, sub-microscopic

32

malaria in volunteers [20]. However, no study has yet investigated whether these or other patterns are observed in clinical malaria episodes, where the parasite burden is at least one thousand times higher, the infection has been present longer, and the sexual stage of the parasite has had time to develop. Additional evidence that malaria is a prime candidate for breath-based diagnosis comes from studies of mosquito behavior. The *Plasmodium* parasite requires *Anopheles* spp. vector mosquitoes to sustain transmission [3]. Studies in human, mouse, and avian malaria have repeatedly demonstrated increased mosquito attraction to odors from infected vertebrate hosts [21–25]. Thus, *Plasmodium* infection may alter host VOCs, which might then be detected in the breath.

To evaluate for *P. falciparum*-specific changes in breath volatiles during natural human malaria infection, we performed unbiased breathprinting. We collected and analyzed breath volatiles from febrile Malawian children with and without uncomplicated *P. falciparum* malaria infection. In this work, we provide the first evidence that natural malaria infection correlates with global changes in breath volatiles that allow for accurate classification of infection status. Furthermore, we establish that volatile mosquito attractants are present at elevated levels in the breath of children with malaria.

2.3 Methods

2.3.1 Breath collection

Prior to enrollment, approval for this study was obtained from both the Malawi College of Medicine Research Ethics Committee (# P.05/14/1572) and the Institutional Review Board of

Washington University School of Medicine (#201504128). Patients were recruited from two ambulatory pediatric centers in Lilongwe, Malawi. Samples were collected over a two-week period during February 2016 from children ages 3-15 presenting for care. Children who had both a positive malaria rapid diagnostic test (RDT) and blood smear were classified as having malaria (n = 17), while those with both a negative RDT and blood smear were enrolled as uninfected controls (n = 18). After informed consent was obtained from caretakers, vital signs and anthropometry were taken and a brief demographic and health history form was completed. Inclusion and exclusion criteria are detailed further in the Supplemental Information. Parasitemia was quantified at a later date using fixed and stained thin smears. For each sample, one thousand red blood cells were counted and inspected for malaria parasites by an experienced parasitologist blinded to the patient's clinical status.

Breath collection was performed as previously reported with alterations detailed here [20]. In brief, ≥ 1 L of exhaled breath was collected in a 3 L SamplePro Flexfilm sample bag (SKC Inc.). Using a set flow pump (ACTI-VOC, Markes International), exactly 1 L of breath was pumped through an inert stainless-steel sorbent tube packed with Tenax 60/80, Carbograph 1 60/80, and Carboxen 1003 40/60 (Camsco). These are absorbent resins that capture VOCs present in the breath for transportation and later analysis. Sorbent tubes were stored at -20°C prior to shipment on freezer packs for off-site mass spectrometric analysis.

2.3.2 Gas chromatography-mass spectrometry (GC/MS) analysis of samples

Samples were analyzed by gas chromatography-mass spectrometry (GC/MS) one month after initial collection. All samples were run with a TurboMatrix 650 ATD (Perkin Elmer) connected

to a Leco Pegasus 4D GCxGC-TOFMS system. A gaseous standard mixture was added to each tube immediately prior to analysis. Raw data files along with patient infection status are available at the Metabolomics Workbench repository, where it has been assigned the Project ID PR000612 [26].

For analysis of the overall VOC profile, files were deconvoluted using MassHunter Qualitative Analysis (Agilent). Deconvoluted compound lists were imported into Mass Profiler Professional (Agilent) for alignment. Peaks were normalized to the 1,2-dichlorobenzene-D4 internal standard (m/z 150 @ 11.7 min). Compounds unique to individual samples were filtered out from further analysis, as were siloxane contaminants.

 The compounds α-pinene, 3-carene, isoprene, acetone, and the 1,2-dichlorobenzene-D4 internal standard were specifically identified and quantified in the GC/MS data files using OpenChrom [27]. The abundances of these compounds in each sample were calculated by integrating the respective base ion peaks. Peak areas were normalized to the base ion peak area of 1,2-dichlorobenzene-D4. For each specific compound, peaks with a normalized area of 0.0002 or less were considered at or below the limit of detection.

2.3.3 Classifier

Using the aligned, standardized compound list generated by Mass Profiler Professional, VOCs that were present in at least 20 participants at a raw signal of >20,000 counts were used in classifier analysis. Class labels were assigned to each subject based on their infection status and VOCs were sorted based on their correlation with infection status. The abundances of the six

most correlated VOCs were summed to create a cumulative abundance metric. Positively

correlated VOC abundances (that is, abundances of compounds that were higher in malaria-

positive patients) were added while negatively correlated VOC abundances (that were lower in

malaria-positive patients) were subtracted. A nearest mean classification algorithm (binary

classification) with leave-one-breath-sample-out cross validation scheme was followed to assign

predicted infection status. The predicted label [malaria (+) or malaria (-)], was compared with the

actual status in order to quantify the performance as shown in Figure 1. The classification

performance, as a function of number of VOCs included, was used to determine the optimal

number of VOCs needed for identification (Supplementary Figure 1D).

2.4 Results

2.4.1 Patient population characteristics and breath sample quality control

We performed a descriptive prospective case-control study of ambulatory pediatric patients in

Lilongwe, Malawi. Cases were defined as having malaria on the basis of both rapid diagnostic

testing and microscopic analysis of thick blood smears. Demographic and clinical characteristics

in the malaria-positive versus malaria-negative patient populations are shown in Table 1 (n =

35). Across all these characteristics, infected and uninfected cohorts were broadly similar,

specifically in regards to potential confounding factors like fever. Diet, which can have an

impact on breath volatiles, was fairly homogenous and was not markedly different between the

two groups (Supplementary Figure 2). For infected patients, the average parasitemia was 2.2%

(range: < 0.001% to 6%).

Breath volatiles were captured onto sorbent material and subsequently released by thermal desorption for analysis by gas chromatography/mass spectrometry (TD-GC/MS). To quality-control for successful breath collection, the levels of the two most common and abundant breath VOCs, isoprene and acetone, were compared to room air controls [28]. For each patient, we find that the abundance of isoprene and/or acetone was at least twice the level observed in room air controls, confirming successful breath collection (Supplementary Figure 3).

2.4.2 Correlation-based classifier identifies suite of six biomarkers with high diagnostic accuracy

Using an unbiased correlation based approach, we identified candidate biomarkers that best differentiated malaria positive breath samples from malaria negative breath samples (Figure 1A). Following pre-processing (Figure 1B), GC/MS data was used to correlate the abundance profile of each VOC with malaria infection status. This strategy identified VOCs that were both positively and negatively correlated with infection status, indicating that *P. falciparum* infection leads to a distinct breathprint marked by both increases and decreases in specific breath compounds (Figure 1C). While no individual compound served as an adequate classifier in isolation, the cumulative abundance across the six VOCs with the highest absolute correlation values proved to be a robust strategy to classify infection status (Figure 2A). All six malaria-associated VOCs—methyl undecane, dimethyl decane, trimethyl hexane, nonanal, isoprene, and tridecane—have been previously reported as present in human breath [14]. Isoprene is known to have an endogenous origin, while the other five VOCs are believed to be derived through oxidative stress-induced lipid peroxidation [28,29]. The three branched alkanes (methyl undecane, dimethyl decane, and trimethyl hexane) were annotated through manually curated

reference to a spectral library. The other three VOCs (nonanal, isoprene, and tridecane) were definitively identified by comparison to pure commercial standards. Characteristic data for all six compounds are provided in Supplementary Table 1.

Together, the six candidate biomarkers yield a cumulative abundance metric, which provides a more Gaussian distribution than individual component features (Figure 2C). Critically, with an appropriate cumulative abundance threshold, we classified malaria infection status with 83% accuracy (Figure 2B-D, Supplementary Figure 1). Potential confounding clinical characteristics (including sex, age, and malnutrition) were not found to be associated with significant differences in cumulative abundance of these six biomarkers (Supplementary Table 2). Thus, we have identified six specific breath compounds that represent candidate biomarkers, whose targeted detection may be used for noninvasive diagnosis of malaria.

2.4.3 Infected patients have elevated breath levels of mosquito-attractant terpenes

We expect that other combination of breath compounds may also have diagnostic utility. As illustrated in Supplementary Figure 1B-D, the changes in breath volatiles that we observed as a result of malaria infection are not limited to the top six compounds. Including additional compounds (up to the top 30 highest correlated compounds) does not result in lower accuracy, and, in validation studies, alternative highly correlative compounds may have improved reproducibility. The full list of compounds used for unbiased discovery, and their correlation values, can be found in the Supplementary Information.

From this extended list, two compounds in particular, the monoterpenes α-pinene and 3-carene, drew especial attention. Previous in vitro studies have identified that cultured *P. falciparum*-infected red blood cells produce a number of plant-like terpenes, including the monoterpene α-pinene [30,31]. Plant-produced terpenes in general influence *Anopheles* spp. mosquito attraction and feeding behavior [31]; these mosquitoes feed on plant-derived nectar in addition to the blood meals taken by females.

Using base ion peak areas, we find that the mean abundances of α-pinene and 3-carene were both significantly higher ($p = 0.04$ with 20% higher mean and $p = 0.01$ with a 28% higher mean, respectively) in the breath of children with malaria compared to uninfected children (Figure 3). To confirm that the changes in α-pinene and 3-carene did not reflect a general trend towards increased capture of monoterpenes during malaria infection, we evaluated levels of the structurally similar terpene (+)-limonene. We find that (+)-limonene was not increased in abundance in the breath of children with malaria (Supplementary Figure 4). Additionally, potential confounding clinical characteristics (including sex, age, and malnutrition) were not found to be associated with significant differences in α-pinene and 3-carene abundances (Supplementary Table 2). Using a receiver operator characteristic curve analysis, we found that breath levels of either α-pinene or 3-carene categorized malaria infection status with a maximum accuracy of 69% and 77%, respectively (Supplementary Figure 5).

2.5 Discussion

Despite impressive gains over the last two decades, only half of children with fever in Africa receive diagnostic testing for malaria as per WHO recommendations [1]. Given the costs of traditional microscopic blood smears and the caveats of intrinsic false positives and rising false negatives with HRP2-based RDTs, the case is clear for innovative alternatives. In this work, we provide the first report of candidate diagnostic biomarkers and elevated mosquito attractants in the breath of *P. falciparum*-infected children from a typical malaria-endemic clinical setting.

This study demonstrates the promise of breath testing for malaria diagnosis. We find robust and global differences in breath VOC composition based on infection status (Supplementary Figure 1B), with as few as six breath volatiles used to provide a classification accuracy of 83% (Figure 2). The patterns of breath volatiles identified in this population of Malawian children with uncomplicated falciparum malaria will require extensive validation in heterogeneous locations and populations. However, these initial studies provide a solid framework upon which to build a possible future diagnostic test. Targeted testing for specific volatiles may be feasible. Alternatively, so-called "eNose" technology may have features more suitable to rapid, field-stable, point-of-care testing in malaria-endemic settings. eNoses implement sensor arrays and pattern recognition technology to describe the chemical composition of complex volatile mixtures, such as breath [32]. One such existing commercial device, Aeonose, was used in a recent clinical study of pulmonary tuberculosis, achieving a diagnostic sensitivity of 88% and specificity of 92% [33]. Although technological barriers persist for clinical implementation [32], less extensive adaptation may be required to implement eNoses in scenarios in which noninvasive testing would be highly valuable, such as border-crossings and population-based screening efforts in elimination settings, a not insignificant need [2].

In the breath of children with malaria, we find increased levels of two terpenes, a class of biomolecules often used by plants for insect communication. Elevated quantities of specific breath terpenes (Figure 3) represent a biologically plausible chemical mechanism for the finding that malaria infection increases *Anopheles* spp. host attraction [22–24]. In particular, we propose that the monoterpene α-pinene represents a strong candidate to be considered as a malaria-induced volatile mosquito signal. In culture, increased α-pinene levels have been observed reproducibly upon *P. falciparum* infection of host cells [30,31]. In addition, this terpene is a direct, potent, and specific activator of *Anopheles gambiae* odorant receptors (AgOR21 and AgOR50), confirming that the primary mosquito vector expresses the biochemical machinery to detect this molecule [30]. Finally, several lines of evidence suggest that α-pinene specifically modulates mosquito feeding behavior. Both α-pinene and the related 3-carene are among the volatiles produced by mosquito-preferred nectar-providing plant species [34,35]. In addition, a blend of volatiles containing α-pinene enhanced *Anopheles* mosquito blood feeding to the same degree as *Plasmodium* infection [31]. Because malaria-induced volatiles are chemically identical to those produced by mosquito-preferred plants, our findings indicate that the malaria parasite may hijack mosquito behavior to increase transmission. Future studies are required to evaluate whether similar strategies may be used by additional vector-borne microbial pathogens. However, mosquito attraction is highly complex and the contribution of these elevated monoterpenes to the overall increased preference for malaria-infected hosts will require dedicated mosquito behavior testing.

Our work also highlights the potential utility of α-pinene and other terpenes as components of superior odor-baited mosquito traps. Successful mass trapping campaigns depend on human scent-mimicking odor baits [36], with some initial promise seen from lures composed of plant attractants, including α-pinene [37,38]. New odor baits blending human- and plant-derived attractant compounds may prove powerful tools for boosting the efficacy of malaria control efforts.

There are several potential limitations to our findings. Our study patients were largely homogeneous with respect to ethnicity, diet, and geographical location. Additional independent validation of our candidate biomarkers in both pediatric and adult patients in a variety of settings is necessary. Our results are also distinct from the previous breath metabolite findings reported by Berna et al. from experimentally-induced, submicroscopic, *P. falciparum*-infected, naïve healthy adults [20], which identified increased levels of four small thioethers as the best classifier of infection status. Using a similar collection protocol, these specific thioethers were only observed in the breath of a single patient, who tested negative for malaria. Thus, the thioethers may prove to be markers of the earliest stages of infection, but subside by the time an individual presents for care. The longer time between sample collection and analysis, as well as different sorbent material, may also explain an absence of thioethers in this study. Similarly, the failure to identify our six biomarkers and elevated terpenes by Berna et al. may be the result of the marked difference in parasite burden or age between the two study populations. The experimentally infected adults achieved a maximum parasitemia of $< 2.5*10^{-6}$ %, using a conversion factor of 4 million red blood cells per microliter, nearly one hundred thousand times less than the average parasitemia in our study. In addition, our study participants are likely to

have had multiple previous episodes of malaria. Prior parasite exposure may be required for

host-generated volatiles produced during *P. falciparum* infection. Finally, children with

uncomplicated falciparum malaria virtually always carry gametocytes, the sexual stage of the

parasite required for mosquito transmission [39–41]. Because gametocytes take more than a

week to mature, they are not present during experimental infections, and therefore examination

of experimental malaria patients will miss gametocyte-specific volatile changes. The increased

mosquito attraction observed during malaria infection appears to require the presence of

circulating gametocytes [22–24]; this was most recently highlighted in the largest experimental

cohort to date [21]. Future studies will evaluate the correlation between our candidate biomarkers

and parasite burden, prior parasite exposure, and gametocyte carriage.

2.6 Figures
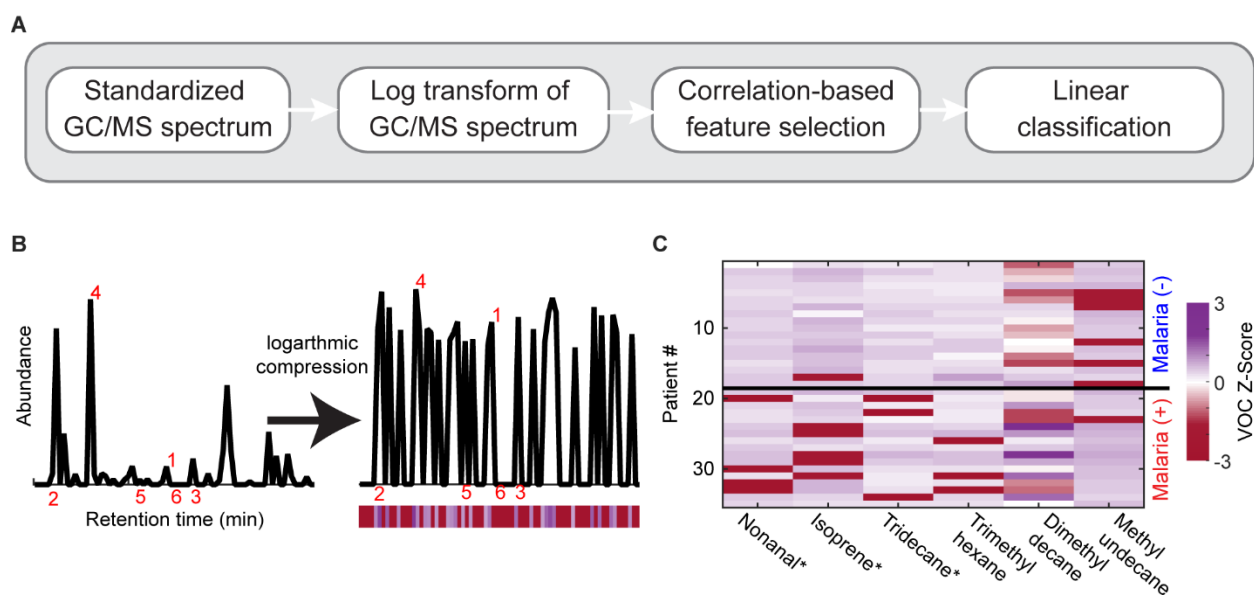
Figure 1. Presence of malaria infection corresponds with an altered breath VOC profile. *A*, Schematic of the analytical approach followed to classify breath samples. *B*, Representative GC/MS total ion chromatograph. After removal of contaminants and normalization, a $\log_2$ operator was applied to compress the abundance values obtained from each subject. For visualization, the abundance of each volatile organic compound (VOC) is represented as a color bar shown below the GC/MS chromatograph. Red numbers indicate the six VOCs with the largest absolute correlations. *C*, Z-score abundances for the six VOCs with the highest absolute correlation with malaria infection status are shown. Asterisk (*), compound identity confirmed by comparison to pure standard. Compound structures and properties provided in Supplementary Table 1.
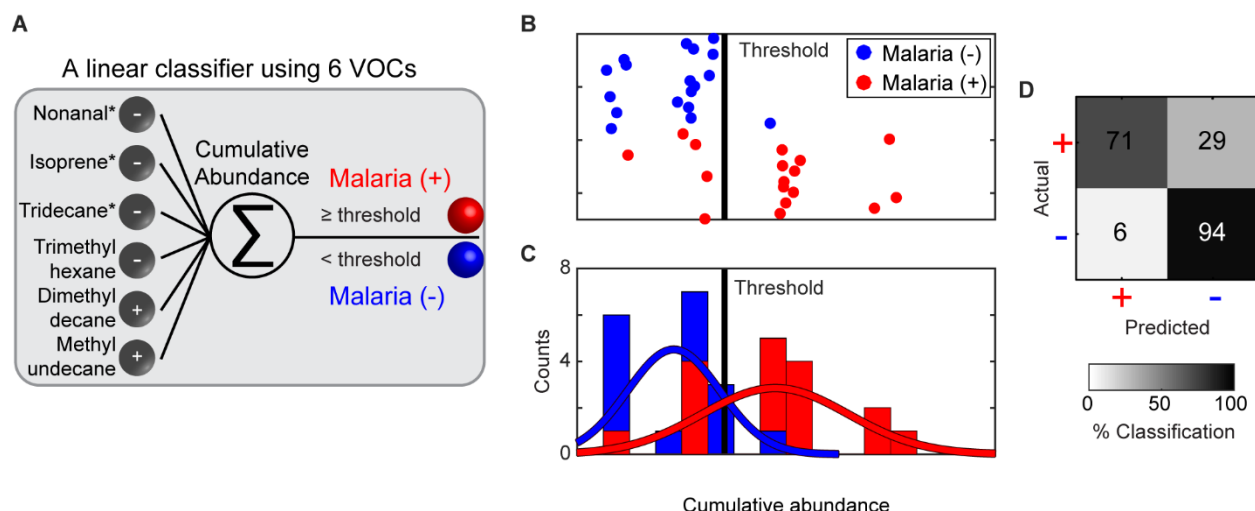
Figure 2. Accurate classification of falciparum malaria infection status achieved with six breath

VOCs. *A*, Schematic of the classification approach. The internal standard normalized abundance

values of the six VOCs are linearly combined to create a cumulative abundance metric.

Negatively correlated VOCs are subtracted rather than added. *B*, Distribution of cumulative

abundance of biomarkers from children with (red) or without (blue) falciparum malaria. *C*,

Cumulative abundance of the six VOCs across all subjects shows clear separation between the

two populations. *D*, Confusion matrix of actual and predicted malaria infection status. Displayed

are the percentages of patients in each class. 83% of classifications were correct, with a

specificity of 94% and sensitivity of 71%.
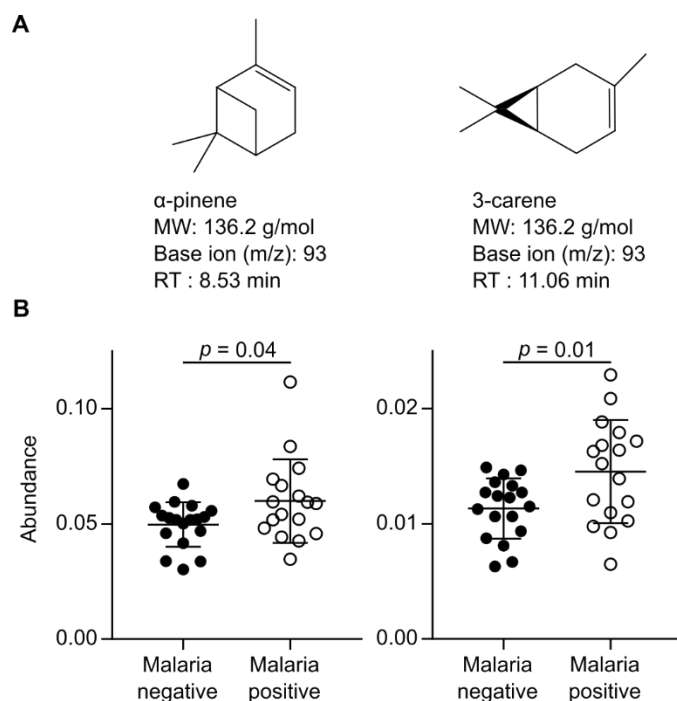
Figure 3. Malaria infection correlates with elevated levels of volatile mosquito-attractant terpenes. *A*, Structure and features of volatile terpenes α-pinene and 3-carene. MW = molecular weight, RT = retention time. *B*, Breath levels of α-pinene, left, and 3-carene, right, in children without ($n = 18$) and with ($n = 17$) falciparum malaria. Abundance quantified by peak area of base ion normalized to internal standard. Mean and standard deviation are shown. Student's t-test used to assess for significant difference between means.
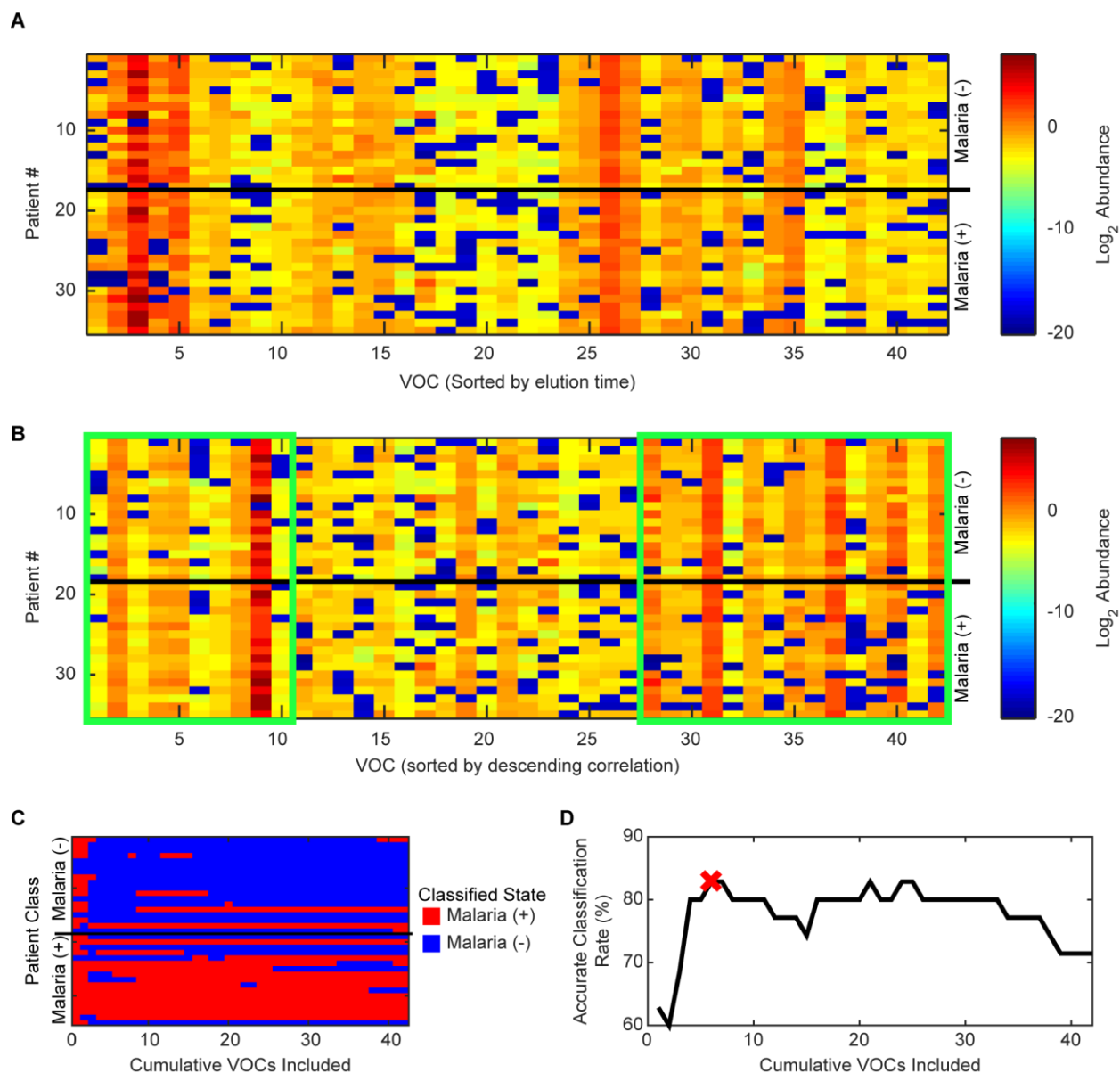
Supplementary Figure 1. Cumulative VOC abundance accurately diagnoses malarial infection.

A, Combined heat map of all patient VOC profiles shows no pattern between subject groups.

Rows are profiles of individual patients, columns are individual VOCs sorted in order of elution

time. *B*, Same as panel *A*, but with VOCs sorted in order of descending correlation with class

labels. Green boxes highlight regions of adundance difference between the two subject groups.

C, Classification map of all patients. Top half, malaria negative (-) subjects; bottom half, malaria

positive (+) subjects. Classifications do not change substantially with inclusion of additional

VOCs (up to 42). *D*, Diagnostic accuracy peaks at 83% using as few as six VOCs. Leave-one-out

cross validation yielded an accurate classification rate of 77%. Abundance values were

normalized to an internal standard.
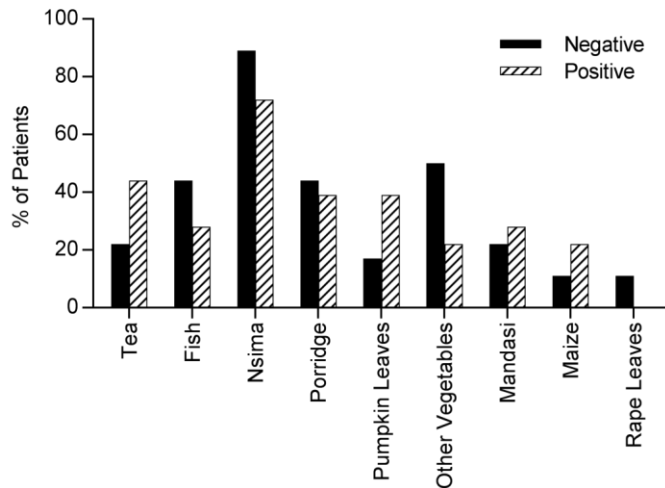
Supplementary Figure 2. Dietary recall does not show major differences based on infection status. Malaria infection status (Negative or Positive) indicated as in legend. Patients were asked to list all food and drink they had consumed in the previous 24 hours. Nsima is a maize-based thick porridge. "Other Vegetables" is a catch all for a number of green, leafy vegetables. Mandasi is a dough-based pastry.
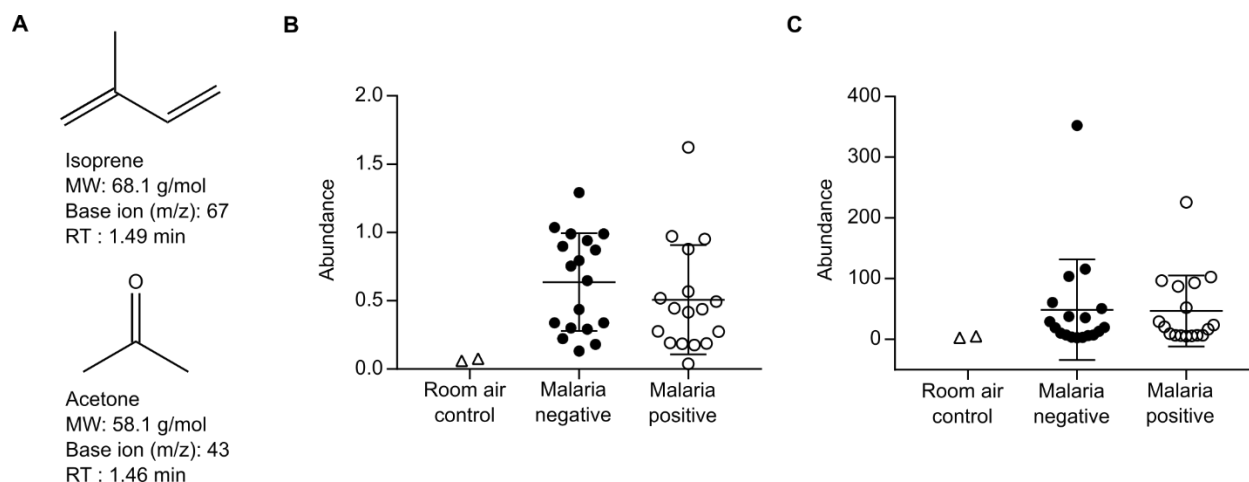
Supplementary Figure 3. Breath sample quality control confirms presence of common and abundant breath volatile organic compounds. *A*, Structure and features of isoprene and acetone. MW = molecular weight, RT = retention time. *B*, Abundance of isoprene in breath samples and room air. *C*, Abundance of acetone in breath samples and room air. In order to verify that breath samples were successfully collected, the abundance of isoprene and acetone were compared versus room air controls. Each breath sample had at least a two-fold higher abundance versus room air for isoprene, acetone, or both. Abundance quantified by the base ion peak area normalized to an internal standard. Room air control $n = 2$, malaria negative $n = 17$, malaria positive $n = 18$. Zero values were adjusted to the limit of quantification (0.0002). Mean and standard deviation are shown.
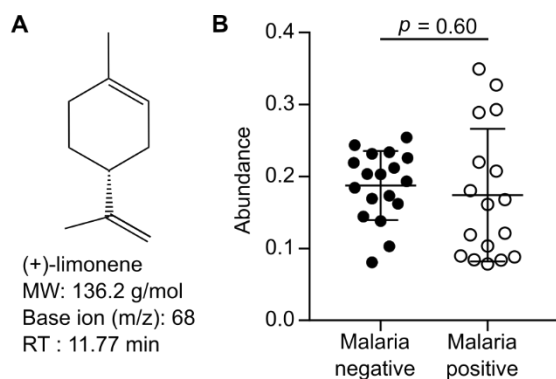
Supplementary Figure 4. Infected patients do not have higher breath abundance for all

monoterpenes. *A*, Structure and features of (+)-limonene, a structurally distinct monoterpene

from α-pinene and 3-carene. MW = molecular weight, RT = retention time. *B*, (+)-limonene

abundance in malaria negative and positive patients. Abundance quantified by peak area of base

ion normalized to internal standard. Mean and standard deviation are shown. Student's t-test

used to calculate significance. Malaria negative $n =17$, malaria positive $n = 18$. Zero values were

adjusted to the limit of quantification (0.0002). Identity of (+)-limonene confirmed by

comparison to true standard.
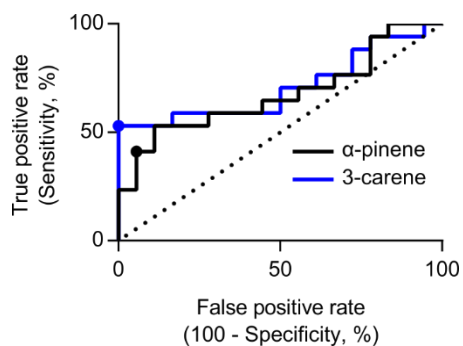
Supplementary Figure 5. Receiver operator characteristic (ROC) curves for terpenes of interest as malaria diagnostics. Curves were computed using internal standard normalized abundance values. Dotted line indicates expected results if predictive power is no better than random chance. Position with maximum likelihood ratio for each curve indicated by point. Using the threshold value at this point gave an overall accuracy of 69% for α-pinene and 77% for 3-carene.

2.7 Tables

Table 1. Patient demographic and clinical characteristics

	Malaria Positive (*n* = 17)	Malaria Negative (*n* = 18)	*p* value[1]
Demographics			
Age, median years (IQR)	8 (6-10)	7 (5-8.5)	0.33
Female, *n* (%)	8 (47)	10/17 (59)	0.73
Reported Symptoms, *n* (%)			
Fever	16 (94)	15 (83)	0.60
Diarrhea	0 (0)	2 (11)	0.49
Vomiting	5 (29)	4 (22)	0.71
Headache	16 (94)	14 (78)	0.34
Abdominal Pain	13 (76)	17 (94)	0.18
Muscle/Joint Pain	12 (71)	4 (22)	0.007
Other, *n* (%)			
Chronic Malnutrition[2]	5/16 (31)	3 (17)	0.43
Acute Malnutrition[2]	0/16 (0)	1 (6)	1
Uses Bednet	9 (53)	10 (56)	1
Malaria within past 3 months	3 (18)	5/17 (29)	0.69

Data represented as number (%) except for age. If one or more patients were excluded due to gaps in the record, number given is fraction of total. Abbreviation: IQR, interquartile range.

[1] Fisher's exact test or Mann-Whitney U-test used as appropriate to calculate *p* values.

[2] Chronic and acute malnutrition defined respectively as height-for-age Z-score or BMI-for-age Z-score two or more standard deviations below median.

Table S1. Properties of candidate diagnostic compounds identified by correlation analysis.
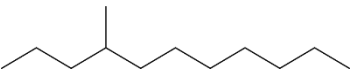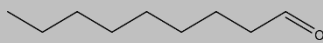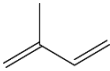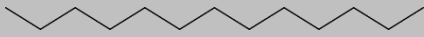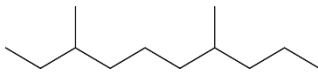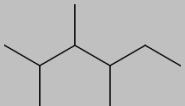
Compound Name	Structure	Base Ion (m/z)	Retention Time (min)	Retention Index[1]
4-methyl undecane		43	14.52	1113
Nonanal[2]		57	14.25	1104
Isoprene[2]		67	1.50	520 (lit.)
Tridecane[2]		57	20.26	1300
3,7-dimethyl decane		57	11.83	1031
2,3,4-trimethyl hexane		43	6.27	860

Table S2. Mean abundances and significance values for patients sorted by potential confounding factors.

	Six Compound Cumulative			α-Pinene			3-Carene		
	Mean Abundance (Negative)	Mean Abundance (Positive)	p value[1]	Mean Abundance (Negative)	Mean Abundance (Positive)	p value[1]	Mean Abundance (Negative)	Mean Abundance (Positive)	p value[1]
Malaria Infection	-41.5	-22.1	0.00002	0.0497	0.0599	0.04	0.0113	0.0145	0.01
Demographics									
Female	-33.8	-30.0	0.46	0.0546	0.0559	0.81	0.0128	0.0128	0.68
Reported Symptoms									
Fever	-40.9	-30.9	0.22	0.0516	0.0551	0.67	0.0129	0.0129	1
Diarrhea	-31.3	-44.7	0.23	0.0550	0.0494	0.62	0.0129	0.0130	0.98
Vomiting	-34.3	-25.6	0.14	0.0542	0.0560	0.78	0.0126	0.0136	0.52
Headache	-37.5	-31.2	0.40	0.0504	0.0554	0.50	0.0122	0.0130	0.67
Abdominal Pain	-29.8	-32.5	0.72	0.0585	0.0541	0.55	0.0136	0.0128	0.69
Muscle/ Joint Pain	-34.7	-29.0	0.28	0.0556	0.0536	0.71	0.0128	0.0130	0.89
Other									
Chronic Malnutrition[2]	-33.7	-28.7	0.42	0.0534	0.0551	0.78	0.0124	0.0135	0.47
Uses Bednet	-31.8	-32.3	0.92	0.0570	0.0528	0.42	0.0131	0.0127	0.76
Malaria within past 3 months	-30.6	-34.5	0.52	0.0535	0.0589	0.40	0.0127	0.0138	0.49
Diet									
Tea	-33.6	-29.2	0.42	0.0532	0.0575	0.43	0.0121	0.0144	0.09
Pumpkin Leaves	-34.9	-25.0	0.08	0.0530	0.0589	0.30	0.0125	0.0140	0.32

Values shown for the cumulative abundance calculated as per Figure 2A and abundances for α-pinene and 3-carene calculated as per Figure 3. The first column in each group is the mean for patients negative for the characteristic indicated by the row title. The second column in each group is the mean for patients positive for the characteristic indicated by the row title. Mean and significance were not calculated for acute malnutrition as only one patient presented with it.

[1] Student's t-test used to calculate p values.

[2] Chronic malnutrition defined as height-for-age Z-score two or more standard deviations below median.

2.8 References

1.   World Health Organisation. World Malaria Report 2016. Geneva: WHO; 2017.

2.   Bell D, Fleurent AE, Hegg MC, Boomgard JD, McConnico CC. Development of new malaria diagnostics: matching performance and need. Malar J. 2016;15: 406.

3.   Bell D, Wongsrichanalai C, Barnwell JW. Ensuring quality and access for malaria diagnosis: how can it be achieved? Nat Rev Microbiol. 2006;4: S7–S20.

4.   Wongsrichanalai C, Barcus MJ, Muth S, Sutamihardja A, Wernsdorfer WH. A review of malaria diagnostic tools: microscopy and rapid diagnostic test (RDT). Am J Trop Med Hyg. 2007;77: 119–127.

5.   Falade CO, Ajayi IO, Nsungwa-Sabiiti J, Siribié M, Diarra A, Sermé L, et al. Malaria rapid diagnostic tests and malaria microscopy for guiding malaria treatment of uncomplicated fevers in Nigeria and prereferral cases in 3 African countries. Clin Infect Dis. 2016;63: S290–S297.

6.   Wilson ML. Malaria rapid diagnostic tests. Clin Infect Dis. 2012;54: 1637–1641.

7.   Kumar N, Pande V, Bhatt RM, Shah NK, Mishra N, Srivastava B, et al. Genetic deletion of HRP2 and HRP3 in Indian *Plasmodium falciparum* population and false negative malaria rapid diagnostic test. Acta Trop. 2013;125: 119–121.

8.   Gamboa D, Ho MF, Bendezu J, Torres K, Chiodini PL, Barnwell JW, et al. A large proportion of *P. falciparum* isolates in the Amazon region of Peru lack pfhrp2 and pfhrp3: Implications for malaria rapid diagnostic tests. PLoS One. 2010;5.

9.   Berhane A, Russom M, Bahta I, Hagos F, Ghirmai M, Uqubay S. Rapid diagnostic tests failing to detect *Plasmodium falciparum* infections in Eritrea: an investigation of reported false negative RDT results. Malar J. 2017;16: 105.

10.   Koita OA, Doumbo OK, Ouattara A, Tall LK, Konaré A, Diakité M, et al. False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. Am J Trop Med Hyg. 2012;86: 194–198.

11.   Cheng Q, Gatton ML, Barnwell J, Chiodini P, McCarthy J, Bell D, et al. *Plasmodium falciparum* parasites lacking histidine-rich protein 2 and 3: a review and recommendations for accurate reporting. Malar J. 2014;13: 283.

12.   Watson OJ, Slater HC, Verity R, Parr JB, Mwandagalirwa MK, Tshefu A, et al. Modelling the drivers of the spread of *Plasmodium falciparum hrp2* gene deletions in sub-Saharan Africa. Elife. 2017;6: e25008.

13.   World Health Organization (WHO). False-negative RDT results and implications of new

reports of *P . falciparum* histidine-rich protein 2/3 gene deletions. Glob Malar Program Inf Note. 2016.

14. de Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, et al. A review of the volatiles from the healthy human body. J Breath Res. 2014;8: 14001.

15. Nakhleh MK, Amal H, Jeries R, Broza YY, Aboud M, Gharra A, et al. Diagnosis and classification of 17 diseases from 1404 subjects via pattern analysis of exhaled molecules. ACS Nano. 2017;11: 112–125.

16. Koo S, Thomas HR, Daniels SD, Lynch RC, Fortier SM, Shea MM, et al. A breath fungal secondary metabolite signature to diagnose invasive aspergillosis. Clin Infect Dis. 2014; 1–8.

17. Peng G, Tisch U, Adams O, Hakim M, Shehada N, Broza YY, et al. Diagnosing lung cancer in exhaled breath using gold nanoparticles. Nat Nanotechnol. 2009;4: 669–673.

18. Kahn N, Lavie O, Paz M, Segev Y, Haick H. Dynamic nanoparticle-based flexible sensors: diagnosis of ovarian carcinoma from exhaled breath. Nano Lett. 2015;15: 7023–7028.

19. Boots AW, Bos LD, van der Schee MP, van Schooten FJ, Sterk PJ. Exhaled molecular fingerprinting in diagnosis and monitoring: validating volatile promises. Trends Mol Med. 2015;21: 633–644.

20. Berna AZ, McCarthy JS, Wang RX, Saliba KJ, Bravo FG, Cassells J, et al. Analysis of breath specimens for biomarkers of *Plasmodium falciparum* infection. J Infect Dis. 2015;212: 1120–1128.

21. Busula AO, Bousema T, Mweresa CK, Masiga D, Logan JG, Sauerwein RW, et al. Gametocytaemia increases attractiveness of *Plasmodium falciparum*-infected Kenyan children to *Anopheles gambiae* mosquitoes. J Infect Dis. 2017.

22. Lacroix R, Mukabana WR, Gouagna LC, Koella JC. Malaria infection increases attractiveness of humans to mosquitoes. PLoS Biol. 2005;3: e298.

23. Batista EPA, Costa EFM, Silva AA. *Anopheles darlingi* (Diptera: Culicidae) displays increased attractiveness to infected individuals with *Plasmodium vivax* gametocytes. Parasit Vectors. 2014;7: 251.

24. De Moraes CM, Stanczyk NM, Betz HS, Pulido H, Sim DG, Read AF, et al. Malaria-induced changes in host odors enhance mosquito attraction. Proc Natl Acad Sci USA. 2014;111: 11079–84.

25. Cornet S, Nicot A, Rivero A, Gandon S. Malaria infection increases bird attractiveness to uninfected mosquitoes. Ecol Lett. 2013;16: 323–329.

26. Sud M, Fahy E, Cotter D, Azam K, Vadivelu I, Burant C, et al. Metabolomics

Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Res. 2016;44: D463–D470.

27. Wenig P, Odermatt J. OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data. BMC Bioinformatics. 2010;11: 405.

28. Mochalski P, King J, Klieber M, Unterkofler K, Hinterhuber H, Baumann M, et al. Blood and breath levels of selected volatile organic compounds in healthy volunteers. Analyst. 2013;138: 2134–45.

29. Hakim M, Broza YY, Barash O, Peled N, Phillips M, Amann A, et al. Volatile organic compounds of lung cancer and possible biochemical pathways. Chem Rev. 2012;112: 5949–5966.

30. Kelly M, Su CY, Schaber C, Crowley JR, Hsu FF, Carlson JR, et al. Malaria parasites produce volatile mosquito attractants. MBio. 2015;6: e00235-15.

31. Emami SN, Lindberg BG, Hua S, Hill SR, Mozuraitis R, Lehmann P, et al. A key malaria metabolite modulates vector blood seeking, feeding, and susceptibility to infection. Science. 2017;355: 1076–1080.

32. Fitzgerald JE, Bui ETH, Simon NM, Fenniri H. Artificial nose technology: status and prospects in diagnostics. Trends Biotechnol. 2017;35: 33–42.

33. Coronel Teixeira R, Rodríguez M, Jiménez de Romero N, Bruins M, Gómez R, Yntema JB, et al. The potential of a portable, point-of-care electronic nose to diagnose tuberculosis. J Infect. 2017; 1–7.

34. Nyasembe VO, Teal PEA, Mukabana WR, Tumlinson JH, Torto B. Behavioural response of the malaria vector *Anopheles gambiae* to host plant volatiles and synthetic blends. Parasit Vectors. 2012;5: 234.

35. Nikbakhtzadeh MR, Terbot JW, Otienoburu PE, Foster WA. Olfactory basis of floral preference of the malaria vector *Anopheles gambiae* (Diptera: Culicidae) among common African plants. J Vector Ecol. 2014;39: 372–383.

36. Homan T, Hiscox A, Mweresa CK, Masiga D, Mukabana WR, Oria P, et al. The effect of mass mosquito trapping on malaria transmission and disease burden (SolarMal): a stepped-wedge cluster-randomised trial. Lancet. 2016;6736: 207–211.

37. Yu B, Ding Y, Mo J. Behavioural response of female *Culex pipiens pallens* to common host plant volatiles and synthetic blends. Parasit Vectors; 2015;8: 598.

38. Nyasembe VO, Tchouassi DP, Kirwa HK, Foster WA, Teal PEA, Borgemeister C, et al. Development and assessment of plant-based synthetic odor baits for surveillance and control of malaria vectors. PLoS One. 2014;9: e89818.

39.   Eziefula AC, Bousema T, Yeung S, Kamya M, Owaraganise A, Gabagaya G, et al. Single dose primaquine for clearance of *Plasmodium falciparum* gametocytes in children with uncomplicated malaria in Uganda: A randomised, controlled, double-blind, dose-ranging trial. Lancet Infect Dis. 2014;14: 130–139.

40.   Von Seidlein L, Drakeley C, Greenwood B, Walraven G, Targett G. Risk factors for gametocyte carriage in Gambian children. Am J Trop Med Hyg. 2001;65: 523–527.

41.   Price R, Nosten F, Simpson JA, Luxemburger C, Phaipun L, Kuile F Ter, et al. Risk factors for gametocyte carriage in uncomplicated falciparum malaria. Am J Trop Med Hyg. 1999;60: 1019–1023.

Chapter 3: Breath biomarker discovery for malaria, a comparison of sampling methods

Preface

The following work was performed by myself, Lucy B. Bollinger, Mwawi Mwale, Rachel Mlotha-Mitole, Indi Trehan, and Audrey R. Odom John. LBB, IT, MM, and RM collected the breath samples and patient metadata. AOJ and I designed the experiments. I performed the GC/MS analysis and the data analysis. LBB, AOJ, IT, and I wrote the main manuscript text. I prepared all Figures and Tables.

This chapter in its entirety will be submitted to the *Journal of Breath Research*. Supplemental Information and the Supplemental Movies will be available online after publication.

I declare a potential conflict of interest as a co-inventor on U.S. Provisional Application Filed 62/550,283.

3.1 Abstract

Breath-based sampling technologies have enormous potential as simple and non-invasive point-of-care diagnostic devices. Breath biomarker discovery relies on robust and reproducible analyses of the volatile compounds present in patient breath. Although a number of different breath collection modalities have been employed for biomarker discovery, few studies have compared the relative strengths and weaknesses of these methods. In this report, we compare two of the most common breath collection systems, a Bio-VOC and sampling bags, in a field setting for malaria biomarker discovery. Malaria is a prime candidate for breath diagnosis, and our study location in Lilongwe, Malawi, represents real-world testing conditions. Two pediatric cohorts were recruited two months apart, the first study using a Bio-VOC and the second using sampling bags. The efficacy of breath collection was assessed by quantifying levels of two high prevalence breath compounds, acetone and isoprene, as well as determining the overall number of breath compounds collected and their cumulative abundance. By each metric, the sampling bags were superior to Bio-VOC. Use of sampling bags yielded 9- and 12-fold higher levels of acetone and isoprene, respectively, and an average of 10-fold more total volatiles detected in each individual. Sampling bags were likely superior due to both the greater volume of breath collected and their less diffusion-prone design. Additionally, malaria-infected patients had elevated breath levels of two mosquito-attractant terpenes (α-pinene and 3-carene), supporting previous observations.

3.2 Introduction

Test-and-treat strategies for control of infectious diseases depend upon simple and accurate diagnostics. For diseases such as malaria, which are endemic to resource-limited settings, large-scale population screening efforts would also greatly benefit from non-invasive diagnostic strategies that do not require blood sampling or highly skilled laboratory personnel. Breath-based diagnostics have the potential to meet this critical need. A typical exhaled human breath contains hundreds of volatile organic compounds (VOCs) [1], and an expanding number of pathologic states have been linked to unique exhaled breath VOC profiles [2,3]. Discovery and validation of candidate breath VOC biomarkers are the indispensable first steps towards development of breath-based diagnostics [2]. A variety of methodologic approaches to breath collection and analysis have been used to date, and a comprehensive analysis of their comparative utility in a field setting is lacking.

As a result of variability in the strategies used for breath collection, it is often challenging to compare results between studies. Breath collection methods are often determined by ease-of-use and cost considerations. The Bio-VOC^{TM} (Markes International, UK), originally developed to monitor environmental exposure to workplace hazards, is among the simplest and least expensive options for breath collection [4–6]. This breath collection apparatus boasts an easy, straightforward collection scheme and a reusable design. In addition, the Bio-VOC has the capability to enrich for collection of alveolar breath. Alveolar breath most closely reflects circulating bloodstream levels of compounds versus ambient air contaminants and is therefore of particular interest for disease biomarker discovery [7]. These features have led to the Bio-VOC being the second most common breath collection system [8]. The most common collection system is inert sampling bags made of materials such as polyvinyl fluoride and biaxially-oriented

polyethylene terephthalate [9]. Sampling bags can collect a larger volume of breath; however, the bags are single-use, more costly, less facile, and do not accurately harvest any particular portion of the exhaled breath. Another alternative for breath collection is the use of dedicated electronic breath sampling devices that track CO_2 levels and breath volume. Such devices allow for robust and highly accurate collection of particular breath portions. However, only a handful of these devices, including the BCA (Mensanna, New Jersey) and ReCIVA® (Owlstone, UK), are commercially available, there is little head-to-head comparative data, upfront costs are considerable, and they may not be suitable for field settings with high ambient temperatures, elevated humidity, and unreliable access to power [10,11].

Malaria, caused by infection with protozoan parasites of the genus *Plasmodium*, represents a particularly promising opportunity for development of breath-based diagnostics. Over the last decade, rapid diagnostic tests (RDTs) for malaria diagnosis have been increasingly deployed and have contributed to a substantial decrease in malaria incidence worldwide. Rapid diagnostic tests primarily rely upon detection of an exported parasite-derived protein, *Plasmodium falciparum* histidine-rich protein 2 (HRP2). Unfortunately, the emergence and spread of *P. falciparum* strains that lack HRP2 expression has put these diagnostics at risk [12,13] and fueled interest in alternative diagnostic methods, including breath-based diagnostics. Previous studies have demonstrated the biological feasibility of malaria-specific breath biomarkers, as cultured *P. falciparum*-infected red blood cells release characteristic volatiles that may be exhaled in the breath [14]. A pilot study of the breath composition of adults with experimental *P. falciparum* infection identified candidate breath biomarkers (small thioethers) of these early-stage, submicroscopic infections [15].

To evaluate the possibility of breath VOC biomarkers for pediatric malaria infection, we recently performed two independent clinical studies to characterize the breath volatile profiles of Malawian children with and without uncomplicated falciparum malaria. Because the two studies used distinct methods for breath collection, together they also provide an opportunity to directly compare the performance characteristics of each collection system, as downstream analyses in each study were otherwise identical. Because of ease-of-use and cost considerations, the Bio-VOC was used for breath collection in the first study. For the second, sampling bags were used in an effort to increase the volume of breath volatiles collected, and, using this data, we recently reported a suite of six breath VOCs that accurately diagnose uncomplicated pediatric malaria [16]. In the current work, we directly compare the results from these two sample collections, in order to interrogate the impact of breath collection method on VOC detection.

3.3 Methods

3.3.1 Breath collection

Prior to enrollment, approval for these studies was obtained from both the University of Malawi College of Medicine Research and Ethics Committee (# P.05/14/1572) and the Institutional Review Board of Washington University School of Medicine (#201504128). Two independent pediatric cohorts were recruited, both from ambulatory pediatric centers in Lilongwe, Malawi. Children aged 3-15 years were eligible for inclusion in the study if, in the course of routine care, the treating clinician determined a need for malaria testing. Exclusion criteria included severe malaria or any other condition requiring urgent medical intervention, having received

antimalarial therapy within the past week, known diabetes, kidney, or liver disease, or inability to cooperate with breath sampling.

The first recruitment period was in November 2015 and the second in February 2016, both during the "high" season for malaria. Children who had both a positive *P. falciparum* malaria RDT and blood smear were classified as having malaria, while those with both a negative rapid diagnostic test and blood smear were enrolled as controls. After informed consent was obtained from caretakers, vital signs and anthropometry were taken, and a brief demographic and health history form was completed.

For the first study, samples were collected using a Bio-VOC breath sampler, a device which collects ~100 mL of exhaled, alveolar breath [17]. The Bio-VOC can be described as a chemically inert syringe open at both ends, pictured in Figure 1a. Study participants were instructed to take a normal breath in and then to exhale fully into the Bio-VOC. Since the volume of the syringe is smaller than the volume of a normal breath, the early portion of the breath is forced out by the later portion, thus semi-selectively retaining the alveolar portion of the breath. After each exhalation, a plunger is attached to one end of the Bio-VOC and a sorbent tube to the other. The retained breath is expelled through the sorbent tube by depression of the plunger. Three breath samples from each participant were collected and stored on a single sorbent tube. This process is diagrammed in Figure 1c and shown in Supplemental Movies 1 and 2.

The breath collection protocol for the second study has been previously described [16]. In brief, ≥ 1 L of exhaled breath was collected in a 3 L SamplePro Flexfilm sample bag (SKC Inc.,

Pennsylvania). Using a set flow pump (ACTI-VOC, Markes International), exactly 1 L of breath was pumped through a sorbent tube. This method is pictured and diagrammed in Figures 1b and 1d, and demonstrated in Supplemental Movie 3. Samples of room air were collected using both methods to assess possible environmental contaminants (n = 5 for Bio-VOC and n = 2 for sampling bags).

For both studies, breath VOCs were captured onto identical sorbent tubes: inert stainless steel packed with Tenax 60/80, Carbograph 1 60/80, and Carboxen 1003 40/60 (Camsco, Texas). Prior to sampling, sorbent tubes were conditioned by flushing with 120 mL/min He at 290°C for one hour, or with 100 mL/min He at 320°C for two hours. All breath samples were stored at -20°C prior to analysis.

Following specimen collection, study participants returned to usual care as per the recommendations of the treating clinician. Antimalarial medications were provided for participants with positive malaria rapid diagnostic test results. Demographic and anthropometric data was entered into a database and height-for-age and BMI-for-age Z scores were calculated using Anthro Plus software (World Health Organization, Switzerland).

3.3.2 Gas chromatography-mass spectrometry (GC/MS) analysis of samples

Samples were analyzed by GC/MS one month after initial collection. All samples were run with a TurboMatrix 650 ATD (Perkin Elmer, Massachusetts) connected to a Pegasus 4D GCxGC-TOFMS system (LECO, Michigan). A gaseous standard mixture was added to each tube

immediately prior to analysis. Detailed machine settings and protocol information were the same as previously detailed [16].

GC/MS data files were analyzed with OpenChrom [18]. The abundances of α-pinene, 3-carene, isoprene, acetone, and the 1,2-dichlorobenzene-D4 internal standard in each sample were calculated by integrating the respective base ion peaks. Peak areas were normalized to the base ion peak area of 1,2-dichlorobenzene-D4. Peaks with a normalized area of 0.0002 or less were considered at or below the limit of detection. One patient was removed from the malaria negative cohort for both terpenes using Grubbs' test for outliers ($\alpha = 0.0001$).

For analysis of the overall VOC profile, files were deconvoluted using MassHunter Qualitative Analysis (Agilent, California). Deconvoluted compound lists were imported into Mass Profiler Professional (Agilent) for alignment. Peaks were normalized to the 1,2-dichlorobenzene-D4 internal standard (m/z 150 @ 11.7 min). Compounds were given annotations using the "IDBrowser Identification" feature using the NIST v11 reference library. M/z expansion was set to -0.3 amu / +0.7 amu. Retention time (RT) matching was not employed. Internal standards, contaminants, and consistently marginal compounds were removed.

3.4 Results

3.4.1 Patient population characteristics and breath collection descriptions

To evaluate the performance characteristics of distinct breath collection methods, we compared the findings from two separate breath biomarker discovery studies. Both studies were performed

at the same site in Lilongwe, Malawi, three months apart, and both were conducted during the wet season and reached their enrollment targets within two weeks. The two study populations were demographically similar. The populations did not differ significantly with respect to potential confounding clinical criteria (Table 1), with the sole exception of the percentage of children reporting abdominal pain (64% in the Bio-VOC study; 86% in sampling bag study; $p = 0.04$), a finding of unclear significance.

The two studies were methodologically distinct in the approach to pediatric breath collection. In brief, for the first study three breath samples were sequentially collected for each patient using a Bio-VOC device, an inert syringe that semi-selectively captures a portion of alveolar (end expiratory) breath (Figure 1a, Supplemental Movie 1). In the second study, multiple full breaths per patient were collected in an inert sampling bag, until more than one liter of breath was obtained (Figure 1b, Supplemental Movie 3). The two studies were otherwise nearly identical in downstream sample processing, as breath samples were transferred to a sorbent tube within minutes of collection, either by manually expelling the sample (Bio-VOC) or using a set-flow pump to draw the sample from the bag and through the tube (sampling bags) (Figure 1c, Supplemental Movie 2). Thermal desorption-GC/MS, using identical equipment and settings, was used to release, identify, and quantify captured breath VOCs from both studies.

3.4.2 Comparison of efficacy of collection protocols

In order to compare the efficacy of breath collection, we first compared the detected levels of two well-characterized breath VOCs, acetone and isoprene. Several factors make these two compounds (Figure 2a) particularly useful in monitoring the efficiency of breath capture. Most

70

breath VOCs are present at concentrations in the parts-per-trillion (ppt) to low parts-per-billion (ppb) range. In contrast, acetone and isoprene are highly abundant in human breath, typically present in concentrations from tens to thousands of ppb [19]. Both compounds are therefore reliably detectable by GC/MS through a large dynamic concentration range. Furthermore, both acetone and isoprene are present at higher levels in the body and breath than in ambient air, such that breath levels are unlikely to represent changes in environmental conditions [7,19].

We find that both acetone and isoprene were reliably detected in breath samples collected by either method (Bio-VOC or sampling bags), with average abundance levels roughly ten-fold greater than levels found in room air controls (Figure S1). However, when breath was collected by Bio-VOC, levels of acetone and isoprene were both significantly reduced compared to breath samples collected via sampling bags (Figure 2b; $p < 0.0001$). Use of sampling bags resulted in 9-fold higher levels of acetone and 12-fold higher levels of isoprene compared to Bio-VOC collection. The markedly lower levels of acetone and isoprene captured by Bio-VOC suggested that other breath compounds, expected to be present at yet lower concentrations, would likely be at or below the limit-of-detection using this method.

To test this assumption, we thus qualitatively and quantitatively compared performance of each collection method across the range of breath VOCs. The raw data for GC/MS is visually represented as a total ion chromatogram (TIC), a plot of retention time versus detection signal strength. Comparing TICs from representative breath samples collected by Bio-VOC versus sampling bags (Figure 2c) illustrates the superior detection of breath VOCs using bag collection. When sampling bags are used, peak heights are higher and a greater number of peaks are visibly

71

present. To quantify these differences in overall VOC detection between the two collection systems, untargeted metabolomics was employed to determine the abundance of every compound present in the breath of each subject (see Supplemental Information and [16]). In order to minimize the impact of patient-to-patient variability, we restricted analyses to those compounds identified in the breath of at least 50% of all subjects for either study. By that metric, a stark difference was noted between the samples collected, depending on the breath collection method used (Figure 2d). Nearly 10 times as many compounds were detected with the sampling bag method compared to the Bio-VOC. This discrepancy in compounds detected was not offset by higher values for those peaks that were identified. Based on summing the peak areas for individual subjects, we find that the overall signal for the Bio-VOC was reduced as well, such that the mean total peak area was more than 7-fold lower than that of sampling bags. The restricted number of VOCs detected using Bio-VOC collection at least in part reflects the reduced volume of breath sampled. When a smaller breath volume is captured, fewer compounds will be collected at levels above the limit-of-detection by GC/MS (Figure 1c).

3.4.3 Attractant terpenes moderately elevated in infected patient breath

As expected, the inferior performance of the Bio-VOC hampers our ability to meaningfully compare breath profiling results between the two studies. Through analysis of breath compounds collected via sampling bags, we have previously reported a suite of six compounds that classified malaria infection status with 83% accuracy. In addition, we identified two potential mosquito attractant compounds, the terpenes α-pinene and 3-carene, which were present at higher levels in the breath of malaria-infected patients [16]. Importantly, we are unable to validate or reject these findings using the independent breath data obtained via Bio-VOC. When the Bio-VOC was used

for breath collection, five of the six classification biomarkers that were previously identified as consistent with malaria infection [16] were not detected in more than half of patients. The remaining biomarker, isoprene, was readily detected using Bio-VOC collection, but did not differ in abundance based on infection status (see Supplemental Information). Both potential mosquito-attractant terpenes, α-pinene and 3-carene, were successfully detected using Bio-VOC collection. However, as was found for the canonical breath compounds acetone and isoprene, detection of breath terpenes was also much reduced with Bio-VOC collection compared to sampling bags, with a 5-fold reduction in 3-carene and an 11-fold reduction in α-pinene levels across all patients between collection methods (p < 0.0001, Mann-Whitney U-test). Despite this limitation, we found that both terpenes demonstrate a pronounced trend towards higher abundance in the breath of malaria-positive patients. On average, individuals with malaria had breath levels of α-pinene and 3-carene that were 1.6-fold and 2.1-fold higher, respectively, than the breath levels of these compounds in uninfected individuals, although this finding did not reach statistical significance (Figure 3; p = 0.08 and p = 0.19).

Potential confounding biological variables were also noted in the Bio-VOC study. Malaria-infected subjects were more likely to be female (58% vs. 24% of malaria-negative; p = 0.04) and older (mean age 9 vs. 6; p = 0.02) (Table S1). However, we did not find that age correlated with terpene abundance (Figure S2), and neither α-pinene nor 3-carene levels differed significantly in female versus male subjects (p = 0.15 and 0.45, respectively; Table S2).

3.5 Discussion

Human breath profiling generates an information-rich metabolic "breathprint" that reflects individual-specific information about many facets of health and disease, from gastrointestinal disorders to neoplasms [3,20,21]. In particular, analysis of breath volatiles shows increasing promise as a non-invasive diagnostic strategy for a variety of infectious diseases, such as tuberculosis, invasive fungal infections, and malaria [15,16,22–24]. Here we compare results from two exploratory breath biomarker discovery studies that employed distinct collection methods, Bio-VOC and sampling bags. To the best of our knowledge, this is the first reported comparison of these two breath collection techniques in a field setting. While a single prior study did report the inferior performance of Bio-VOC for breath composition of an individual child (analyzed once using sampling bags and once using Bio-VOC) [25], comprehensive comparisons of breath collection methods with consistent downstream analyses have been lacking.

Understanding how breath metabolites change during a variety of pathological states is critical to informing the specificity of particular biomarkers. Several prior efforts have noted the number of VOCs observed across the sample population using a Bio-VOC method similar to the one we employed [7,26,27]. However, direct study-to-study comparisons have so far proven extremely challenging, due to the large number of additional variables involved in experimental designs. These differences include relatively modest experimental changes between studies, such as differences in absorbent resin and machine parameters. However, major confounders between studies include differences in the amount of breath collected, data pre-processing methodologies, and the type of detector utilized. For example, use of TD-GC/MS versus an alternative analytical technique, such as proton transfer reaction (PTR) mass spectrometry, brings in a range of trade-offs such as VOC detection limits, likelihood of sample degradation, and metabolite resolving

74

power [9,28]. Given the lack of standardizations, the difficulty in comparing studies has been well noted across the breath biomarker field [9,22,29].

Our current work provides strong evidence that Bio-VOC-based breath collection has inferior performance compared to breath collection using sampling bags. Breath collection by Bio-VOC yielded reduced levels of typical breath volatiles (acetone and isoprene) and fewer total VOCs present above the limit-of-detection (Figure 2). In part, these findings can be attributed to the reduced volume of breath collected (Figure 1c), as the Bio-VOC is estimated to have sampled between 264 and 387 mL total (based on 88-129 mL/breath), whereas the bags consistently sampled 1000 mL of breath [6,17]. Under field conditions, as in our study, actual collection volumes may be yet lower. Future studies are needed to determine if collecting additional breath (i.e. higher volumes) with the Bio-VOC can narrow the performance gap between it and sampling bags.

Interestingly, the reduced volume of breath collected by the Bio-VOC (approximately 30% of the volume of sample bags) does not fully account for the approximately 10-fold reduction in breath acetone and isoprene levels compared to sampling bags. While breath VOC levels exhibit substantial biological variability, neither acetone nor isoprene have greater variability (based on 95% CI, interquartile range, and coefficient of variation) in breath samples collected using Bio-VOC compared to sampling bags. In addition, we find that the inter-subject variability in both our studies are well within the typical range for other breath analyses [19]. One possible explanation for the poorer-than-expected Bio-VOC performance is that sampling bags may be more air-tight in design. Once the requisite liter of breath is collected, there is negligible

potential for leaking or diffusion with sampling bags. In contrast, when sampling breath with the Bio-VOC, the investigator must retrieve the device from the patient and transfer breath contents over a sorbent tube, potentially permitting some VOC diffusion out of the open ends of the device prior to capture. An additional consideration is whether the Bio-VOC detects fewer VOCs because of exclusion of ambient contaminants, which would be a desirable feature. Collecting breath via sampling bags is expected to result in a more mixed breath sample (oral plus alveolar breath) with an increased number of contaminants compared to the end-expiratory (alveolar) breath that is harvested by Bio-VOC. Arguing against this possibility, the canonical breath volatiles acetone and isoprene predominate within the body and not the environment [7,19]. Additionally, many of the other annotated VOCs unique to the sampling bag study are also believed to have endogenous (not environmental) origins, including hexanal and the methylated alkanes [30].

In our analysis of breath volatiles collected by sampling bags, we previously nominated six breath compounds as candidate biomarkers of malaria infection. Due to the inferior performance of the Bio-VOC, we were unable to evaluate those findings in this second independent cohort, as the majority of the previously proposed biomarkers were below the limit-of-detection. Our previous analysis also reported higher levels of two mosquito-attractant terpenes, α-pinene and 3-carene, in the breath of malaria-infected children [16]. In the cohort whose breath was collected using Bio-VOC, we also find higher breath levels of α-pinene and 3-carene in malaria-infected individuals compared to uninfected controls, although this did not reach statistical significance. This failure to reach significance could be the result of greater variability in the Bio-VOC α-pinene and 3-carene levels versus sampling bags levels. The coefficients of variation were 104%

and 130% for α-pinene and 3-carene for the Bio-VOC study, versus 28% and 30% for the terpenes for the sampling bags study. The greater variability is likely caused by measuring near the limit-of-detection with the Bio-VOC, versus well above it for the sampling bags. Regardless, the Bio-VOC cohort provides modest additional support for the hypothesis that malaria infection increases breath levels of mosquito-attractive terpene compounds. Together, both studies support the strong ongoing need for additional studies to validate malaria-induced breath changes, in hopes of pursuing development of much-needed new malaria diagnostics.

The current work has several potential limitations. Although our two patient populations were largely similar in terms of overall clinical characteristics, they may differ in some unaccounted-for manner that affects the number or diversity of VOCs detected. We also note that the Bio-VOC was designed for use by adults, not the pediatric population (ages 3 to 15 years) under investigation in this work. However, the average forced vital capacity of children in this age range readily exceeds the volume of the Bio-VOC, suggesting that the poor performance of the Bio-VOC is unlikely to be a specific concern with respect to pediatric breath collection [31]. Finally, the studies compared in this analysis both employed a common method for pre-concentrating breath VOCs prior to analysis via GC/MS, namely capture onto thermal desorption (TD) tubes. An alternative pre-concentration method, which has been used with Bio-VOCs, is solid phase microextraction (SPME) fibers [8]. Unlike TD tubes, SPME fibers are less dependent on the volume of breath collected [8,32]. Thus, additional investigation is necessary to confirm whether the conclusions reached here with regards to analysis via TD tubes may be extended to studies utilizing SPME fibers.

Altogether, our results strongly support the use of sampling bags over the Bio-VOC for breath biomarker discovery studies. Almost all human disease-specific VOCs that have been described to date are present in the breath in the parts-per-trillion (ppt) to low parts-per-billion (ppb) concentration range, and we find that the Bio-VOC demonstrates insufficient capacity to detect a diverse number of compounds within that range. The low cost and ease-of-use of the Bio-VOC cannot compensate for the improved performance of sampling bags. Specialized breath collection devices may prove more superior still. However, the cost barrier for these devices or even sampling bags calls for the development of a robust, low cost device for use in remote locations or for diseases with limited funding.

3.6 Figures
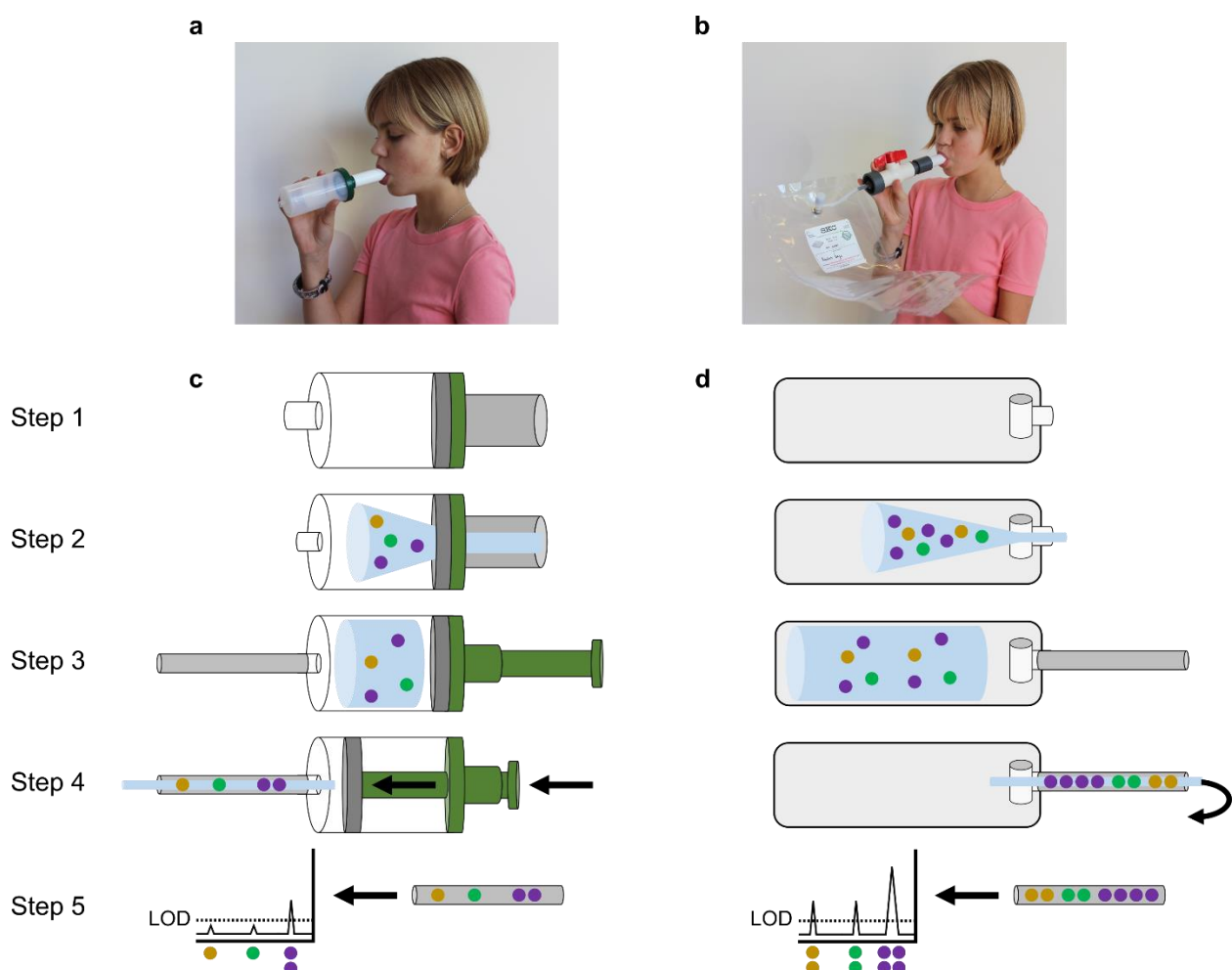
Figure 1. Comparison of Bio-VOC and sampling bag for breath collection. Pediatric breath collection via Bio-VOC (a) and sampling bag (b). For each method, field-site video of breath collection is presented in Movie S1 and S2, respectively. c) and d) Schematic of breath collection methods for Bio-VOC and sampling bag, respectively. Step 1, the empty collection device. Step 2, patient exhales into the collection device. Unique breath volatile organic compounds visualized as different color dots. Step 3, Sorbent tube is affixed to the collection device. Step 4, Breath volume driven through sorbent tube by depressing plunger or using an air pump. Breath

metabolites are captured on the sorbent tube. Step 5, Breath metabolites are released from the

sorbent tube by thermal desorption and measured by GC/MS. LOD: limit of detection. Overall,

the sampling bag (right) collects a larger volume of breath (Step 3) leading to a greater quantity

of breath metabolites captured. This in turn is reflected as higher signals by GC/MS, including

multiple breath metabolites that were undetectable via the Bio-VOC now being observed (Step

5).
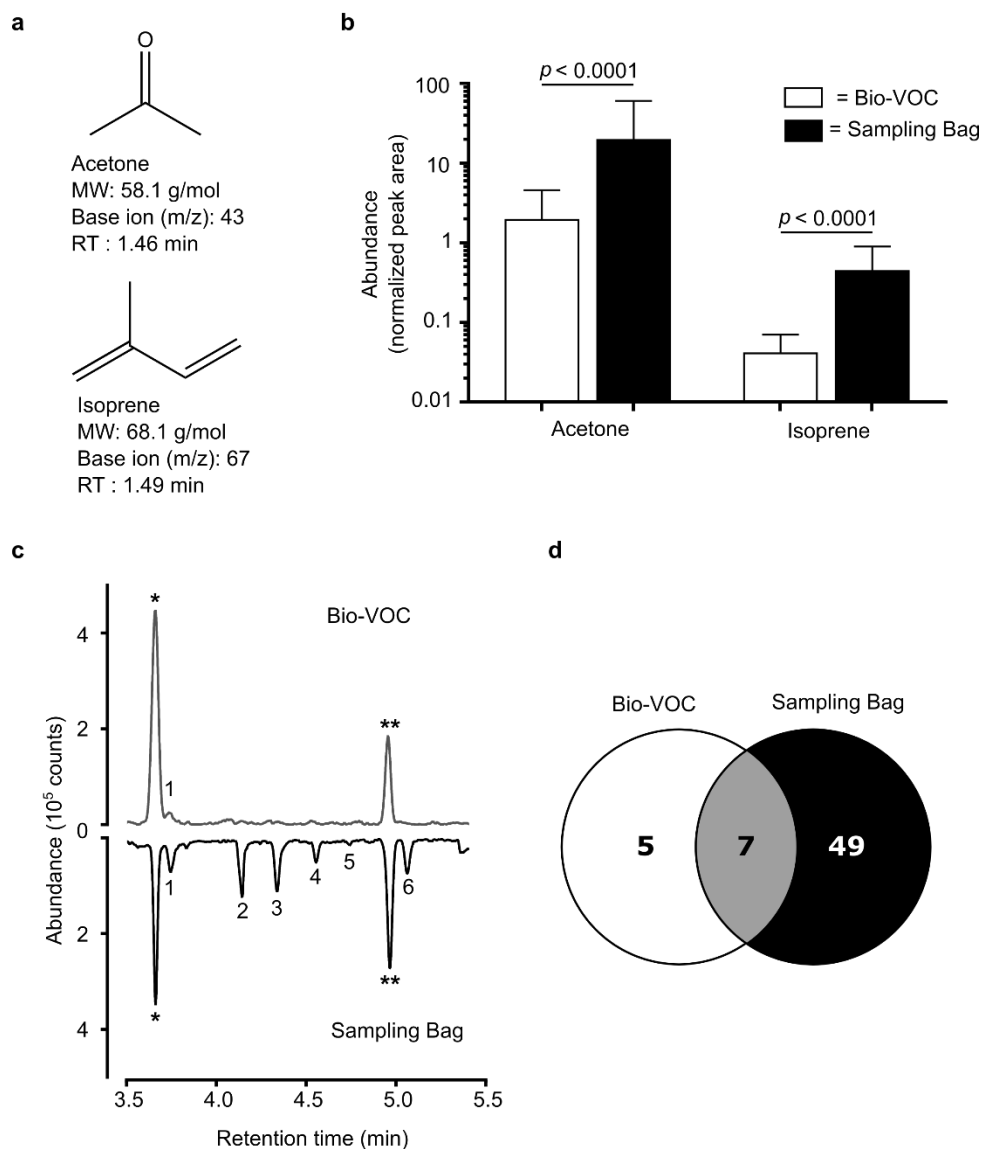
Figure 2. Comparison of breath volatile organic compounds (VOCs) using two different collection methods (Bio-VOC and sampling bags). a) Structure and chemical information for the two most common and abundant compounds in breath. MW = molecular weight, RT = retention time. b) Breath abundance of acetone and isoprene across the two studies. Median and interquartile range are shown. P-values, Mann Whitney U-tests. c) Portion of representative total ion chromatogram from both studies. * = internal standard, ** = column contaminant, number =

compound. Compound annotations are as follows: 1 = 1,3,5-cycloheptatriene, 2 = acetic acid, 3 = 4-ethylbenzamide, 4 = hexanal, 5 = propyl-propanedioic acid, and 6 = 2,3,4-trimethyl-hexane. d) Comparison of total VOCs detected above the limit-of-detection in more than half of patients in one or both studies.
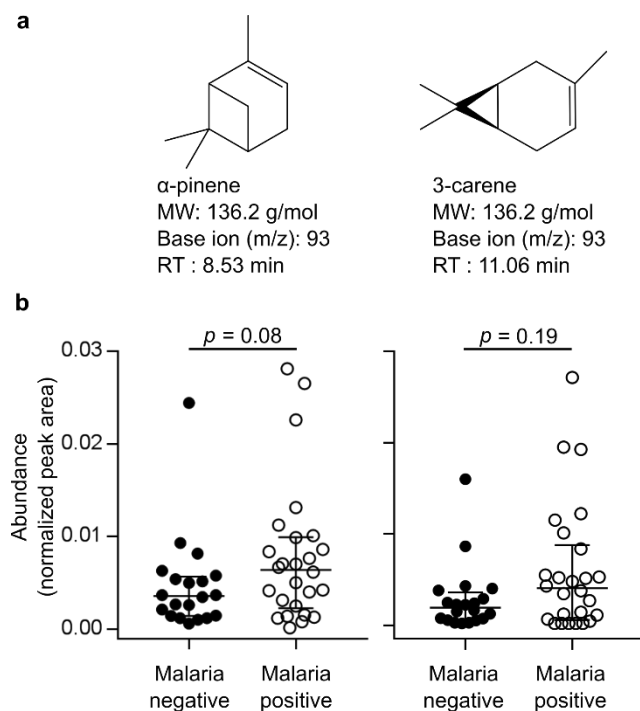
Figure 3. Breath abundance of candidate mosquito-attractant terpenes in Bio-VOC study. a) Structure and chemical information of the terpenes. MW = molecular weight, RT = retention time. b) Breath abundance of terpenes α-pinene, left, and 3-carene, right, in children without (n = 20) and with (n = 26) falciparum malaria. Median and interquartile range are shown. For subjects in which compounds were not detected, abundance values were adjusted to the limit-of-quantification (0.0002). P-values, Mann Whitney U-tests.
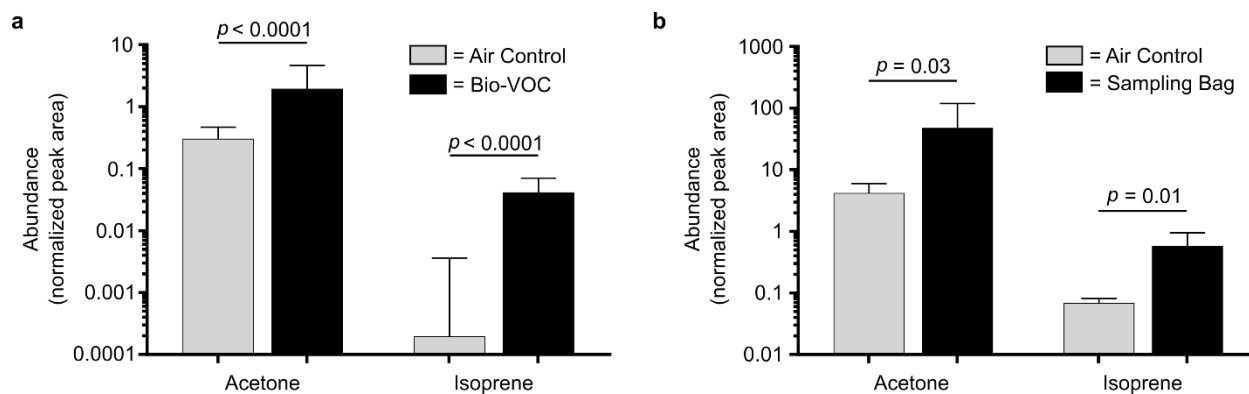
Supplementary Figure 1. Acetone and isoprene are significantly more abundant in breath samples versus room air. Higher levels of acetone and isoprene were seen in breath versus room air for both a) the Bio-VOC study and b) the sampling bag study. Median and interquartile range are shown. P-values, Mann Whitney U-tests.
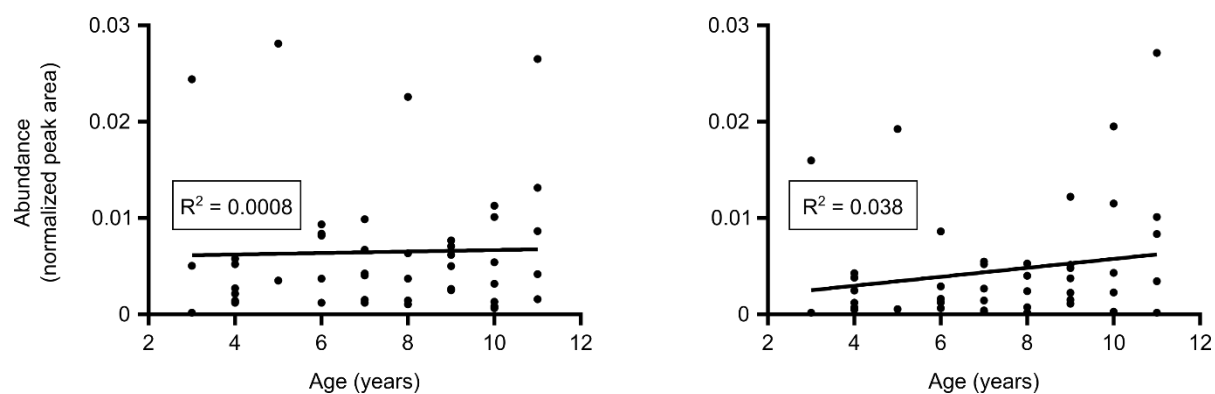
Supplementary Figure 2. Terpene levels do not correlate with age in Bio-VOC study.

Scatterplot for α-pinene (left) and 3-carene (right). Linear regression line with goodness-of-fit

(R^2) shown.

3.7 Tables

Table 1. Comparison of patient demographic and clinical characteristics between the two study populations.

	Bio-VOC Study (n = 47)	Sample Bag Study (n = 35)	p value[1]
Malaria Positive, n (%)	26 (55)	17 (49)	0.66
Demographics			
Age, median years (IQR)	8 (6-10)	8 (5-10)	0.86
Female, n (%)	20 (43)	18/34 (53)	0.38
Reported Symptoms, n (%)			
Fever	43 (91)	31 (89)	0.72
Diarrhea	6 (13)	2 (6)	0.46
Vomiting	19 (40)	9 (26)	0.24
Headache	35 (74)	30 (86)	0.28
Abdominal Pain	30 (64)	30 (86)	0.04
Muscle/Joint Pain	23 (49)	16 (46)	0.83
Other, n (%)			
Chronic Malnutrition[2]	9/46 (20)	8/34 (24)	0.78
Acute Malnutrition[2]	3/46 (7)	1 (3)	0.63
Uses Bednet	29 (62)	19 (54)	0.65
Malaria within past 3 months	9/45 (20)	8/34 (24)	0.79

Data represented as number (%) except for age. If one or more patients were excluded due to gaps in the record, number given is fraction of total. Abbreviation: IQR, interquartile range. Data for Sampling Bag Study reported previously [16].

[1] Fisher's exact test or Mann-Whitney U-test used as appropriate to calculate p values.

[2] Chronic and acute malnutrition defined respectively as height-for-age Z-score or BMI-for-age Z-score two or more standard deviations below median.

Table S1. Patient demographic and clinical characteristics between malaria positive and negative patients for Bio-VOC Study.

	Malaria Positive (*n* = 26)	Malaria Negative (*n* = 21)	*p* value[1]
Demographics			
Age, median years (IQR)	9 (7-10)	6 (4-8.5)	0.02
Female, *n* (%)	15 (58)	5 (24)	0.04
Reported Symptoms, *n* (%)			
Fever	25 (96)	18 (86)	0.31
Diarrhea	3 (12)	3 (14)	1
Vomiting	13 (50)	6 (29)	0.23
Headache	21 (81)	14 (67)	0.33
Abdominal Pain	16 (62)	14 (67)	0.77
Muscle/Joint Pain	16 (62)	7 (33)	0.08
Other, *n* (%)			
Chronic Malnutrition[2]	7 (27)	2/20 (10)	0.26
Acute Malnutrition[2]	1 (4)	2/20 (10)	0.57
Uses Bednet	11 (42)	18 (86)	0.003
Malaria within past 3 months	6/24 (25)	3 (14)	0.47

Data represented as number (%) except for age. If one or more patients were excluded due to gaps in the record, number given is fraction of total. Abbreviation: IQR, interquartile range.

[1] Fisher's exact test or Mann-Whitney U-test used as appropriate to calculate *p* values.

[2] Chronic and acute malnutrition defined respectively as height-for-age Z-score or BMI-for-age Z-score two or more standard deviations below median.

Table S2. Median abundances and significance values for patients sorted by potential confounding clinical variables for Bio-VOC study.

	α-Pinene			3-Carene		
	Median Abundance (Positive)	Median Abundance (Negative)	*p* value[1]	Median Abundance (Positive)	Median Abundance (Negative)	*p* value[1]
Malaria Infection	0.0064	0.0036	0.08	0.0040	0.0019	0.19
Demographics						
Female	0.0074	0.0037	0.15	0.0036	0.0023	0.45
Reported Symptoms						
Fever	0.0050	0.0019	0.07	0.0026	0.0009	0.26
Diarrhea	0.0032	0.0050	0.49	0.0027	0.0023	0.99
Vomiting	0.0063	0.0037	0.10	0.0043	0.0022	0.06
Headache	0.0050	0.0039	0.65	0.0023	0.0032	0.95
Abdominal Pain	0.0041	0.0062	0.27	0.0026	0.0023	0.47
Muscle/Joint Pain	0.0035	0.0054	0.20	0.0024	0.0029	0.45
Other						
Chronic Malnutrition[2]	0.0067	0.0039	0.49	0.0038	0.0020	0.30
Acute Malnutrition[2]	0.0027	0.0047	0.50	0.0025	0.0025	0.51
Uses Bednet	0.0044	0.0047	0.36	0.0019	0.0036	0.29
Malaria within past 3 months	0.0052	0.0037	0.61	0.0027	0.0024	0.97

Median Abundance (Negative) columns, the median normalized peak area values for patients negative for the characteristic indicated by the row title. Median Abundance (Positive) columns, median normalized peak area values for patients positive for the characteristic indicated by the row title.

[1] Mann-Whitney U-test used to calculate *p* values.

[2] Chronic or acute malnutrition defined as height-for-age Z-score or BMI-for-age Z-score (respectively) two or more standard deviations below median.

3.8 References

1. de Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, et al. A review of the volatiles from the healthy human body. J Breath Res. 2014;8: 14001.

2. Boots AW, Bos LD, van der Schee MP, van Schooten FJ, Sterk PJ. Exhaled molecular fingerprinting in diagnosis and monitoring: validating volatile promises. Trends Mol Med. 2015;21: 633–644.

3. Nakhleh MK, Amal H, Jeries R, Broza YY, Aboud M, Gharra A, et al. Diagnosis and classification of 17 diseases from 1404 subjects via pattern analysis of exhaled molecules. ACS Nano. 2017;11: 112–125.

4. Dyne D, Cocker J, Wilson HK. A novel device for capturing breath samples for solvent analysis. Sci Total Environ. 1997;199: 83–89.

5. Wilson HK, Monster AC. New technologies in the use of exhaled breath analysis for biological monitoring. Occup Environ Med. 1999;56: 753–7.

6. Application Note 013 The Bio-VOC – A low-cost , simple device for biological monitoring of VOCs in breath. 2015;44: 1–4.

7. Van Den Velde S, Quirynen M, Van Hee P, Van Steenberghe D. Differences between alveolar air and mouth air. Anal Chem. 2007;79: 3425–3429.

8. Lawal O, Ahmed WM, Nijsen TME, Goodacre R, Fowler SJ. Exhaled breath analysis: a review of "breath-taking" methods for off-line analysis. Metabolomics. 2017;13: 1–16.

9. Rattray NJW, Hamrang Z, Trivedi DK, Goodacre R, Fowler SJ. Taking your breath away: Metabolomics breathes life in to personalized medicine. Trends Biotechnol. 2014;32: 538–548.

10. Allsworth M, Apthorp D, Van Der Schee M, Smith R, Boschmans J, Kitchen S, inventors. Owlstone Medical Limited, assignee. Method for collecting a selective portion of a subject's breath. USA; 15/494,973, 2017.

11. Phillips M, inventor. Breath collection apparatus. USA; 6,726,637, 2004.

12. Odaga J, Sinclair D, Lokong J, Donegan S, Hopkins H, Garner P. Rapid diagnostic tests versus clinical diagnosis for managing people with fever in malaria endemic settings. Cochrane Database Syst Rev. 2014.

13. Bell D, Fleurent AE, Hegg MC, Boomgard JD, McConnico CC. Development of new malaria diagnostics: matching performance and need. Malar J. 2016;15: 406.

14. Kelly M, Su CY, Schaber C, Crowley JR, Hsu FF, Carlson JR, et al. Malaria parasites

produce volatile mosquito attractants. MBio. 2015;6: e00235-15.

15.   Berna AZ, McCarthy JS, Wang RX, Saliba KJ, Bravo FG, Cassells J, et al. Analysis of breath specimens for biomarkers of *Plasmodium falciparum* infection. J Infect Dis. 2015;212: 1120–1128.

16.   Schaber CL, Katta N, Bollinger LB, Mwale M, Mlotha-Mitole R, Trehan I, Raman B, Odom John AR. Breathprinting Reveals Malaria-Associated Biomarkers and Mosquito Attractants. J Infect Dis. 2018;217(10):1553-1560.

17.   Kwak J, Fan M, Harshman SW, Garrison CE, Dershem VL, Phillips JB, et al. Evaluation of Bio-VOC sampler for analysis of volatile organic compounds in exhaled breath. Metabolites. 2014;4: 879–888.

18.   Wenig P, Odermatt J. OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data. BMC Bioinformatics. 2010;11: 405.

19.   Mochalski P, King J, Klieber M, Unterkofler K, Hinterhuber H, Baumann M, et al. Blood and breath levels of selected volatile organic compounds in healthy volunteers. Analyst. 2013;138: 2134–45.

20.   Christiansen A, Davidsen JR, Titlestad I, Vestbo J, Baumbach J. A systematic review of breath analysis and detection of volatile organic compounds in COPD. J Breath Res. 2016;10: 34002.

21.   Queralto N, Berliner AN, Goldsmith B, Martino R, Rhodes P, Lim SH. Detecting cancer by breath volatile organic compound analysis: A review of array-based sensors. J Breath Res. 2014;8.

22.   Ahmed WM, Lawal O, Nijsen TM, Goodacre R, Fowler SJ. Exhaled volatile organic compounds of infection: a systematic review. ACS Infect Dis. 2017.

23.   Coronel Teixeira R, Rodríguez M, Jiménez de Romero N, Bruins M, Gómez R, Yntema JB, et al. The potential of a portable, point-of-care electronic nose to diagnose tuberculosis. J Infect. 2017; 1–7.

24.   Koo S, Thomas HR, Daniels SD, Lynch RC, Fortier SM, Shea MM, et al. A breath fungal secondary metabolite signature to diagnose invasive aspergillosis. Clin Infect Dis. 2014; 1–8.

25.   Caldeira M, Barros AS, Bilelo MJ, Parada A, Câmara JS, Rocha SM. Profiling allergic asthma volatile metabolic patterns using a headspace-solid phase microextraction/gas chromatography based methodology. J Chromatogr A. 2011;1218: 3771–3780.

26.   Dadamio J, Van den Velde S, Laleman W, Van Hee P, Coucke W, Nevens F, et al. Breath biomarkers of liver cirrhosis. J Chromatogr B. 2012;905: 17–22.

27.   Das MK, Bishwal SC, Das A, Dabral D, Varshney A, Badireddy VK, et al. Investigation

of gender-specific exhaled breath volatome in humans by GCxGC-TOF-MS. Anal Chem. 2014;86: 1229–1237.

28.   Bajtarevic A, Ager C, Pienz M, Klieber M, Schwarz K, Ligor M, et al. Noninvasive detection of lung cancer by analysis of exhaled breath. BMC Cancer. 2009;9: 348.

29.   Herbig J, Beauchamp J. Towards standardization in the analysis of breath gas volatiles. J Breath Res. 2014;8: 37101.

30.   Hakim M, Broza YY, Barash O, Peled N, Phillips M, Amann A, et al. Volatile organic compounds of lung cancer and possible biochemical pathways. Chem Rev. 2012;112: 5949–5966.

31.   Dockery D, Berkey C, Ware J, Speizer F, Ferris Jr B. Distribution of forced vital capacity and forced expiratory volume in one second in children 6 to 11 years of age. Am Rev Respir Dis. 1983;128: 405–412.

32.   Alonso M, Sanchez JM. Analytical challenges in breath analysis and its application to exposure monitoring. TrAC - Trends Anal Chem. 2013;44: 78–89.

Chapter 4: Global proteomic analysis of prenylated proteins in *Plasmodium falciparum* using an alkyne-modified isoprenoid analogue

Preface

The following work was performed by myself, Kiall F. Suazo, Charuta C. Palsuledesai, Audrey R. Odom John, and Mark D. Distefano. I performed the experiments to label the parasites with the prenyl analogue, purified the FPPS protein, and demonstrated the ability of FPPS to extend the analogue from a ten carbon to a fifteen carbon molecule. KS extracted labeled proteins from the parasite samples, visualized them by in-gel fluorescence, and identified them by pull-down and protein mass spectrometry. KS, CP, AOJ, MD, and I designed the experiments. KS and I prepared the Figures and Tables. KS, AOJ, MD, and I wrote the manuscript; my contributions were the Introduction and the Discussion sections on the biological context of the identified proteins, as well as the Results and Methods sections relating to the experiments I performed.

This chapter in its entirety has been published (Suazo KF, Schaber C, Palsuledesai CC, Odom John AR, Distefano MD. Global proteomic analysis of prenylated proteins in *Plasmodium falciparum* using an alkyne-modified isoprenoid analogue. *Scientific Reports*. December 2016). Article available at: [http://doi.org/10.1038/srep38615]. Supplemental Table 1 can also be found at that web address. Reproduction is allowed as per the Creative Commons Attribution 4.0 International License.

4.1 Abstract

Severe malaria due to *Plasmodium falciparum* infection remains a serious threat to health worldwide and new therapeutic targets are highly desirable. Small molecule inhibitors of prenyl transferases, enzymes that catalyze the post-translational isoprenyl modifications of proteins, exhibit potent antimalarial activity. The antimalarial actions of prenyltransferase inhibitors indicate that protein prenylation is required for malaria parasite development. In this study, we used a chemical biology strategy to experimentally characterize the entire complement of prenylated proteins in the human malaria parasite. In contrast to the expansive mammalian and fungal prenylomes, we find that *P. falciparum* possesses a restricted set of prenylated proteins. The prenylome of *P. falciparum* is dominated by Rab GTPases, in addition to a small number of prenylated proteins that also appear to function primarily in membrane trafficking. Overall, we found robust experimental evidence for a total of only thirteen prenylated proteins in *P. falciparum*, with suggestive evidence for an additional two probable prenyltransferase substrates. Our work contributes to an increasingly complete picture of essential, post-translational hydrophobic modifications in blood-stage *P. falciparum*.

4.2 Introduction

Over the past 15 years, improved efforts at controlling malaria, caused by infection with the protozoan parasite *Plasmodium falciparum*, have significantly decreased the overall number of cases and childhood deaths attributable to severe malaria [1]. However, there remain over 200

million infections and over half a million deaths due to malaria each year [2]. Access to highly effective antimalarial therapies remains a cornerstone of malaria control efforts. Unfortunately, widespread resistance to former first-line agents, such as chloroquine, and emerging resistance to newer treatments, such as the artemisinin-combination therapies, endangers control of malaria worldwide [3,4].

Target-based antimalarial drug development depends on identification of essential biological processes in *P. falciparum* that are amenable to small molecule inhibition. Development of therapeutics for the developing world is hampered by a relative lack of commercial pharmaceutical interest. Therefore, one strategy has been to identify potential antimalarial target proteins whose human homologs have themselves been explored as pharmaceutical targets. These kinds of repurposing approaches thus harness the power of previous large-scale small molecule screening and development pipelines, in hopes of reducing the effort and expense of developing novel antiparasitics for resource-limited settings.

Protein prenyltransferases have emerged as one such "piggybacking" target for antimalarial drug development [5,6]. Protein prenylation is the C-terminal modification of cellular proteins with either a farnesyl (15-carbon) or geranylgeranyl (20-carbon) isoprenyl group. Prenyl modification of proteins is catalyzed by several classes of cellular prenyltransferase enzymes, including farnesyl transferase (FT) and geranylgeranyltransferase type I and type II (GGT-1 and GGT-2) [7,8]. Prenylation is typically required for the membrane association and therefore the cellular activity of prenyltransferase substrates. For example, farnesylation of the small G-protein oncogene, K-Ras, is required for the transformation of many human cancers, including lung and

96

colon cancer [9]. For this reason, protein farnesyltransferase inhibitors have been extensively explored by the pharmaceutical industry as potential human chemotherapeutics [10,11]. Like most eukaryotic organisms, *P. falciparum* malaria parasites also possess protein prenyltransferase activity and have been found to incorporate both farnesyl and geranylgeranyl modifications into protein substrates [12,13]. Chemical inhibition of isoprenoid precursor biosynthesis in malaria parasites blocks protein prenylation and is lethal to cultured *P. falciparum*, suggesting that production of isoprenyl substrates for protein prenylation is an essential function of isoprenoid biosynthesis in the parasite [14]. In addition, inhibition of parasite prenyltransferase activity halts parasite replication [15–18], providing compelling evidence that protein prenylation is indispensable for malaria parasite growth.

Since protein prenyltransferase activity is required by *P. falciparum*, identification of prenyltransferase substrates will likely reveal additional antimalarial targets. Bioinformatic approaches have been previously used to predict a limited number of potential prenylated proteins in the malaria genome [19]. However, since *Plasmodium* spp. are evolutionarily divergent from organisms used to generate these models, and few prenylated proteins have been experimentally confirmed in malaria parasites, it is not clear how well bioinformatics algorithms perform in predicting prenyltransferase substrates for *Plasmodium* spp. In this work, we use a chemical labeling approach to metabolically tag, potentially, the full complement of prenylated proteins in asexual *P. falciparum* parasites. Our approach was to metabolically incorporate an alkyne-modified isoprenoid analogue into the pool of prenyltransferase protein substrates. This additional alkyne functional group permits selective binding of prenylated proteins to streptavidin beads, via click chemistry with biotin-azide. The resulting prenylated proteins were

97

identified by subsequent tryptic digestion and LC-MS analysis, coupled with bioinformatics analysis.

4.3 Methods

4.3.1 *P. falciparum* tissue culture

All culturing was done with *Plasmodium falciparum* genome reference strain 3D7. 3D7 was obtained from the Malaria Research and Reference Reagent Resource Center (strain MRA-102, contributed by D. J. Carucci, ATCC, Manassas, Virginia). Parasites were grown in RPMI-1640 media (Sigma-Aldrich, SKU R4130) supplemented with 27 mM sodium bicarbonate, 11 mM glucose, 5 mM HEPES, 1 mM sodium pyruvate, 0.37 mM hypoxanthine, 0.01 mM thymidine, 10 µg ml^{-1} gentamycin (Sigma-Aldrich) and 0.5% Albumax (Life Technologies) with a 2% suspension of human erythrocytes under an atmosphere of 5% CO_2, 5% O_2, balance N_2 and incubated at 37°C, as previously described [20,21]. For in-gel fluorescence, 40 mL of culture was used per replicate. Samples destined for mass spectroscopy analysis were derived from 200 mL of culture per replicate. For all experiments, cultures were adjusted to 4% of red blood cells infected (4% parasitemia) at experiment start. Cultures were treated with fosmidomycin (Life Technologies) to a final concentration of 600 nM (approximately half IC$_{50}$). Pyrophosphate probes or prenyl pyrophosphates (Echelon Biosciences) were added to a final concentration of 10 µM. After compounds were added, cultures were mixed thoroughly and incubated for 24 hours.

After 24 hours, cultures were saponin lysed as previously described [21], with modifications. In brief, cells were pelleted and washed with PBS before being lysed with 1% saponin in PBS.

Saponin lyses red blood cell (RBC) membranes but not parasite cell membranes, thus freeing the parasites from the RBCs. The lysed mixture was pelleted, and the loose RBC membrane layer and supernatant removed, leaving a parasite pellet. This pellet was washed with PBS, centrifuged again, the supernatant removed, and stored at -80°C.

RBC controls were performed with 5 mL per replicate of 2% hematocrit in supplemented RPMI media with no parasites. RBCs were pelleted without saponin since internal RBC proteins are released with lysis. This volume corresponds to the volume of saponin-freed parasites from 200mL total culture at 4% parasitemia.

4.3.2 In-gel fluorescence labeling

RBC controls and saponin-lysed pellets of *P. falciparum*, treated with or without FSM and FPP or C15AlkOPP, were suspended in 300 µL lysis buffer (10 mM PO_4^{3-}, 137 mM NaCl, 2.7 mM KCl, 2.4 µM PMSF, benzonase nuclease, protease inhibitor cocktail and 1% SDS) and sonicated 6 to 8 times for 2 seconds in 10-second intervals. Click reactions were performed on 100 µg samples of protein lysate (1 µg/µL) with 25 µM TAMRA-N_3, 1 mM TCEP, 0.1 mM TBTA, and 1 mM $CuSO_4$ at room temperature for one hour. Proteins were precipitated using a ProteoExtract precipitation kit (Calbiochem) to remove excess click chemistry reagents. Protein pellets were dissolved in 1X Laemlli loading buffer and heated at 95 °C for 5 minutes. Samples were fractionated using 12% SDS PAGE gels and imaged via in-gel fluorescence using a BioRad FX Molecular Imager with 542/568 nm excitation/emission wavelengths. Gels were stained with 1X Coomasie blue stain followed by destaining to visualize protein loading.

4.3.3 Cloning and expression of *Pf*FPPS

The gene *PfFPPS* (PF3D7_1128400) was PCR amplified from *P. falciparum* 3D7 cDNA (primers *Pf*FPPS Fwd 5'-CTCACCACCACCACCACCATGCUGAGAACGAGCAGAATAACCAAGATTC-3'; *Pf*FPPS Rev 5'-ATCCTATCTTACTCACTCAAGCGCCTGTAAACAAAATGTCC-3') and cloned into pBG1861 using ligation independent cloning as previously described.[22] The cloned *PfFPPS* sequence was verified by Sanger sequencing. The primers used add coding for a six histidine tag at the N-terminus of the gene to allow for nickel affinity purification.

Subsequently, pBG1861 was transformed into ArticExpress (DE3) RIL *E. coli* (Agilent Technologies). Cultures were grown in LB media with 100 µg/mL ampicillin at 37°C and 200 rpm until mid-log phase, at which point they were cooled to 8°C. Expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), 5 µM geraniol and 5 µM farnesol overnight at 8°C and 200 rpm. Following induction, cells were pelleted and then lysed by sonication in a solution of 25 mM Tris pH 7.5, 250 mM NaCl, 1 mM $MgCl_2$, 10% v/v glycerol, 20 mM imidazole, 1 mM dithiothreitol (DTT), 1 mg/mL lysozyme, 200 µM phenylmethylsulfonyl fluoride (PMSF) 0.3 U/mL benzonase nuclease (Novagen), and EDTA-free protease inhibitor (Roche). 6-histidine tagged protein was purified from soluble lysate over Ni-NTA resin (Goldbio). The resin with bound protein was washed with 25 mM Tris pH 7.5, 250 mM NaCl, 1 mM $MgCl_2$, 10% v/v glycerol, 20 mM imidazole, 1 mM dithiothreitol (DTT). Bound protein was then eluted with 25 mM Tris pH 7.5, 250 mM NaCl, 1 mM MgCl2, 10% v/v glycerol, 300 mM imidazole, 1 mM dithiothreitol (DTT).

Next, the elutant was further purified over a HiLoad 16/60 Superdex 200 gel filtration column (GE Healthcare) using an AKTAExplorer 100 FPLC (GE Healthcare). The FPLC buffer was 250 mM NaCl, 25 mM Tris pH 7.5, and 1 mM $MgCl_2$, 10% glycerol v/v. Fractions enriched with *Pf*FPPS, as seen by a strong band at ~ 44 kDa on a Coomassie-stained SDS-PAGE gel, were pooled and concentrated by centrifugation using Amicon Ultra-15 centrifugal filter units (EMD Millipore). Concentrated protein was supplemented with 1mM DTT, was flash-frozen in liquid N_2, and was then stored at -80°C prior to use. Protein concentration was measured by a BCA protein assay kit (Thermo Scientific).

4.3.4 Isoprenyl pyrophosphate synthase assay

Following purification, release of pyrophosphate during GGPP/C20AlkOPP synthesis from FPP/C15AlkOPP by *Pf*FPPS was monitored using the EnzChek phosphate assay kit (Life Technologies), as previously described [23]. Reactions were performed in a 50 µL volume, with final reagent concentrations as follows: 250 mM NaCl, 50 mM Tris pH 7.5, 1mM $MgCl_2$, 1 U/mL purine nucleoside phosphorylase (PNP), 0.2 mM 2-Amino-6-mercapto-7-methylpurine riboside (MESG), 0.1 U/mL yeast inorganic pyrophosphatase (New England Biolabs), and 2 µM purified *Pf*FPPS, and, where indicated, 100 µM IPP, FPP (Echelon Biosciences) and/or C15AlkOPP. All reagents save *Pf*FPPS were pre-warmed to 37°C. Reactions were initiated by the addition of *Pf*FPPS, after which absorbance at 360 nm was recorded over a 30 min period with a BMG POLARStar plate reader preheated to 37 °C. Absorbance monitoring was performed in clear 96-well flat-bottomed plates. Enzyme reactions were linear with respect to time and enzyme concentration. Absorbance units were converted to µM phosphate using a phosphate standard curve.

4.3.5 Pull-down of labeled proteins

Protein lysates (1.5 mg/mL) from parasites treated with FSM and FPP or C15AlkOPP were subjected to click reactions with 100 μM biotin-N_3, 50 mM TCEP, 10 mM TBTA, and 50 mM $CuSO_4$ for 90 minutes at room temperature. Excess reagents were removed by protein precipitation using 1 volume of chloroform, 4 volumes of CH_3OH, and 3 volumes of PBS. Proteins were precipitated in between two immiscible phases by centrifugation at 4,500 x g for 5 minutes. The aqueous layer was discarded and 4 volumes of CH_3OH was added, followed by centrifugation at 4,500 x g for 3 minutes to pellet the proteins. Proteins were dissolved in 1% SDS in PBS buffer (1.5 mg/mL) and incubated with 300 μL of NeutrAvidin® agarose resin (Thermo Scientific) for 90 minutes. Resin samples were washed to remove unbound proteins with 3-mL volumes of $3 \times 1\%$ SDS in PBS, $1 \times$ PBS, 3×8 M urea, and 3×50 mM NH_4HCO_3. Resin was suspended in 300 μL of 50 mM NH_4HCO_3 and combined with 5 μg trypsin (sequencing grade, Promega Corp.) for overnight digestion at 37 ℃. Supernatants were collected by washing the resin with 200 μL x 4 of 50 mM NH_4HCO_3 and samples were lyophilized.

4.3.6 Proteomic analysis

Sample preparation for MS/MS analysis. Lyophilized peptides were dissolved in 200 mM NH_4COO. Aliquots from resulting peptide solutions (20 μg) were obtained to prepare 0.25 μg/μL solutions. Each sample was loaded in SDB-XC extraction disk (3M, USA) packed in stage tips conditioned with 80% acetonitrile (ACN) and equilibrated with 200 mM NH_4HCO_2. Samples were washed with 200 mM NH_4HCO_2 and eluted into three fractions using 40 μL of 6%, 11%, and 17% ACN in H_2O. Each sample was dissolved in 100 μL of 5% ACN and 0.1% TFA in H_2O

and loaded onto packed extraction disks in stage tips that were conditioned (80% ACN and 0.1% TFA in H_2O) and equilibrated (5% ACN and 0.1% TFA in H_2O). Peptides were eluted with 80% ACN with 0.1% TFA in H_2O, lyophilized, and dissolved in 0.1% formic acid.

LC-MS/MS analysis of tryptic digested peptides. LC-MS/MS analyses were carried out using an RSLCnano System (Dionex, UK) and an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific). Samples were directly loaded and eluted at a flow rate of 300 nL/min onto a reverse-phase column (75 μm i.d., 450 mm) packed with ProntoSIL C18AQ 3 μm media (Bischoff, Germany) that was prepared in-house. The peptides were eluted with buffer A (0.1% formic acid in H_2O) and buffer B (0.1% formic acid in CH_3CN) in the following gradient segments of buffer B: 17 mins, 0-2%; 60 mins, 2-25%; 2 mins, 25-44%; 2 mins, 44-76%; 3 mins, 76%; and 2 mins, 76-2%. The eluted peptides from the column were sprayed into a nanospray ion source on an Orbitrap Fusion Tribrid mass spectrometer set to record single microscan FTMS scan events at a resolution of 30000 over the m/z range 300-1500 Da in positive ion mode, with charge states of 2-7 included. The top 15 data-dependent CID MS/MS were triggered from the FTMS scan and introduced into the Orbitrap- Fusion ion trap. The collision energy was set to 35% and activation Q to 0.25 with the scan range and ion trap scan rate both set to normal. The automatic gain control (AGC) target values were set to optimal conditions at 500,000 for MS1 and 5,000 for MS2 [24]. Dynamic exclusion was allowed once at a 90-second duration.

Proteomic data processing. The .raw files were searched using Sequest embedded in Proteome Discoverer (version 1.4.0.288, Thermo Scientific) against the *Plasmodium falciparum* 3D7 isolate (ID UP000001450) appended with *Homo sapiens* (ID UP000005640) from

UnipProtKB/SwissProt [25]. The precursor mass tolerance was set to 10 ppm and the fragment

mass tolerance was set to 0.6 Da. A variable modification was set as oxidized methionine. The

enzyme was set to Trypsin (Full) and up to 4 missed cleavages were allowed. A decoy search

was also performed.

The resulting .msf files were processed in Scaffold (version 4.4.1, Proteome Software Inc.,

Portland, OR) through searching with X! Tandem (version 2010.12.01.1, GPM Organization).

Glu → pyro-Glu of the N-terminus, ammonia-loss of the N-terminus, Gln → pyro-Glu of the N-

terminus and oxidation of methionine were specified in X! Tandem. Peptide identifications were

accepted if they could be established at greater than 95% by the Scaffold Local FDR algorithm.

Protein identifications were accepted if they could be established at greater than 99.0%

probability and contained at least 2 identified peptides. Protein probabilities were assigned by the

Protein Prophet algorithm [26]. Proteins that contained similar peptides and could not be

differentiated based on MS/MS alone were grouped to satisfy the principles of parsimony.

Proteins sharing significant peptide evidence were grouped into clusters. Fold changes in

enrichment between probe-treated and control samples were calculated using total weighted

spectra with 1 imputation.

4.4 Results

4.4.1 An alkyne-functionalized isoprenoid analogue is metabolically incorporated into malaria parasites

To identify the prenylated proteins in *P. falciparum* using an alkyne-containing isoprenoid analogue, we first tested for metabolic incorporation of the compound C15AlkOPP (Fig. 1A), which was previously employed to identify prenylated proteins in mammalian cells [27,28]. This probe structurally resembles the native isoprenoid substrates farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP) used for prenylation of proteins, and is a substrate for both mammalian FT and GGT-1 [29]. Red blood cells infected with *P. falciparum* were exposed to the probe in the presence or absence of fosmidomycin (FSM), an established inhibitor of isoprenoid biosynthesis in *P. falciparum* [21,30], followed by release of the intact parasites via mild detergent treatment. The free parasites were then lysed and the resulting lysates subjected to copper-catalyzed click reaction with TAMRA-N_3, which generates a stable cycloaddition product between the alkyne-tagged prenylated proteins and TAMRA-N_3 to allow visualization of labeled proteins (Fig. 1B). The samples were then fractionated via SDS-PAGE and subjected to in-gel fluorescence imaging (Fig. 1C, top panel).

Fluorescent protein bands were observed at approximately 25 and 50 kDa in samples obtained from parasites treated with C15AlkOPP (lane 3); a number of weaker bands, including species near 37 and 150 kDa, were also observed. A substantial enhancement of labeling ensued upon co-administration of FSM (lane 4), suggesting that depletion of the endogenous FPP pool results in increased incorporation of the analogue; similar results with C15AlkOPP have been observed in mammalian cells treated with lovastatin [27]. Replacing the probe with FPP showed only

limited labeling in these regions (lane 2), indicating that the alkyne analogue is a viable tool to tag cellular prenylated proteins.

It should be noted that *P. falciparum* parasites develop within human erythrocytes. Although human prenyltransferases have not been identified in the mature erythrocyte proteome [31], we evaluated for the possibility that human erythrocyte proteins could incorporate C15AlkOPP probe in the absence of parasite infection. Human red blood cells were thus also treated with C15AlkOPP and FSM, and did not demonstrate significant labeling (lane 1). The band observed above 25 kDa is almost certainly an artifact due to a highly abundant protein band in this region as shown in the Coomassie stained gel (Fig. 1C, bottom panel); given its size, that protein may represent a hemoglobin dimer [32]. Overall, these results indicate that the C15AlkOPP probe successfully labels prenylated proteins in the malaria parasite with minimal interference from human proteins.

4.4.2 The C15AlkOPP probe is elongated by *P. falciparum* FPPS/GGPPS

As noted above, C15AlkOPP is a substrate for the mammalian FT and GGT-1 enzymes. Previously reported metabolic labeling data suggests that C15AlkOPP can also be used to label substrates of GGT-2. However, those results do not preclude elongation of C15AlkOPP to C20AlkOPP prior to prenyltransferase-catalyzed incorporation. Hence, we next sought to investigate whether C15AlkOPP can be elongated to C20AlkOPP in malaria parasites. Unlike other organisms that express a series of prenyl synthases, *P. falciparum* produces a single multi-functional enzyme (*Pf*FPPS/GGPPS, hereafter *Pf*FPPS) that is responsible for synthesis of geranyl pyrophosphate, farnesyl pyrophosphate, and geranylgeranyl pyrophosphate [33,34]. To

evaluate whether the parasite enzyme may utilize the C15AlkOPP probe as a substrate, we assayed purified recombinant *Pf*FPPS protein. As expected, *Pf*FPPS uses IPP to elongate its natural substrate, FPP, thus generating GGPP. Similarly, we found that *Pf*FPPS also effectively uses C15AlkOPP as a substrate (Fig. 2). Thus, while it is not clear whether the elongated analogue is a substrate for the malarial GGTs, the production of that species maximizes the likelihood that our single probe strategy using C15AlkOPP will function to tag both farnesylated and geranylgeranylated proteins in the parasite.

4.4.3 Bioinformatic analysis of the *P. falciparum* proteome affords a list of putative prenylated proteins

To date, only a limited number of proteins with canonical C-terminal prenylation motifs have been demonstrated to be bona fide substrates in *P. falciparum* [14,35–37]. Prior to proteomic analysis, we performed a bioinformatic investigation to create a list of all possible prenyltransferase substrates present in the *P. falciparum* proteome. FASTA sequences of proteins from the *Plasmodium falciparum* genome reference isolate 3D7 proteome (UniProt ID UP000001450) with possible C-terminal –CaaX, –CXC, and –CC prenylation motifs were analyzed using Prenylation Prediction Suite (PrePS) [38]. Out of 90 protein sequences with Cys at the 4th position from the C-terminus (–CaaX), a total of 8 proteins were predicted to be prenylated (Table 1). This group contains 5 proteins whose molecular masses are close to 25, 37, 50, and 150 kDa, similar to what we observed experimentally via in-gel fluorescence labeling (Fig. 1C). Furthermore, five of these proteins contain basic residues upstream of the putatively prenylated Cys, a typical pattern observed for most prenylated proteins. In addition to these predicted proteins, we examined those that were not recognized by PrePS but met additional

criteria for prenylation: basic residues upstream of Cys at the -1 to -5 positions and a hydrophobic or aromatic residue at the +2 position. Seven proteins that satisfy these additional parameters were thus identified from the proteome database (Table 2). A previous study reported 8 proteins that satisfied these criteria based on the *P. falciparum* genome, which included a DEAD/DEAH box Helicase and Methionyl-tRNA formyltransferase (along with six predicted in our analysis) [36]. However, neither of these two proteins contains a C-terminal CaaX sequence in the reference *Plasmodium falciparum* isolate 3D7 proteome database (although one of them does in another database, see Figure S1). Interestingly, the remaining protein listed in Table 2 (but not identified in the aforementioned genomic analysis), PRL protein tyrosine phosphatase, does appear to be a real prenylated protein. Recent computational docking studies [39] suggested that it is a substrate and this has been confirmed through additional in vitro assays [36].

Proteins with C-terminal –CXC and –CC motifs were also extracted from the proteome and analyzed in PrePS. As expected, Rab proteins were predicted to be substrates of geranylgeranyl transferase type II (GGT-2) with high degrees of probability (Table 3). The molecular weights of these small GTPases are consistent with the bands observed near 25 kDa via in-gel fluorescence labeling (Fig. 1C). No proteins other than those belonging to the Rab family were identified as potential substrates, suggesting that there are no known proteins with these –CXC and –CC sequences that have the specific neighboring residues upstream of Cys necessary to render them substrates for GGT-2. In aggregate, our bioinformatics analysis suggests that the *P. falciparum* proteome contains 8 –CaaX, 3 –CXC, and 7 –CC proteins (18 total) that may be prenyltransferase substrates.

4.4.4 Proteomic analysis of the prenylated proteins in *P. falciparum*

After validating the use of C15AlkOPP as a prenylation probe in *P. falciparum*, we next sought to determine the molecular identity of prenylated proteins in the parasite. To reduce cellular production of FPP, which would compete with probe incorporation, cultured asexual *P. falciparum* were treated with sub-lethal (50% of the half-maximal inhibitory concentration) concentrations FSM. Fosmidomycin-treated parasites were grown in the presence of either C15AlkOPP probe or FPP as a negative control. Following labeling, parasites were freed from host erythrocytes, and parasite protein lysates were subjected to click reaction with biotin-N3; the affinity handle conferred by biotin permits the selective enrichment and isolation of labeled prenylated proteins upon pull-down through the strong biotin-avidin interaction (Fig. 1B). Proteins were washed under stringent conditions (1% SDS and 8 M urea) followed by on-bead trypsin digestion. Equal amounts of peptides from C15AlkOPP- and FPP-treated samples, were pre-fractionated and analyzed by nano-flow liquid chromatography and tandem mass spectrometry (MS/MS), followed by database searching against *Plasmodium falciparum* (UniProt ID UP000001450) and *Homo sapiens* proteomes (UniProt ID UP000005640). After data processing, 445 proteins were identified at 99.0% minimum probability with at least 2 identified peptides from 9073 spectra, and 98.0% minimum peptide confidence within a 1% false discovery rate, for both C15AlkOPP- and FPP-treated samples.

Our labeling strategy was designed to enrich for prenyltransferase substrates labeled with C15AlkOPP and/or C20AlkOPP using a biotin pull-down approach. Proteomic studies that use such enrichment methods are sometimes complicated due to nonspecific adsorption of proteins onto the avidin-coated beads used in these experiments. To address that issue, our approach

employed a quantitative comparison based on spectral counting between the samples treated with C15AlkOPP and those treated with FPP. We used the average total spectral counts for each identified protein to calculate the enrichment of proteins (fold-change) across three replicates of samples treated with C15AlkOPP probe versus FPP. The full list of spectral counts and identities are provided (Supplemental Table 1). Proteins with potential prenylation motifs (–CaaX, –CC, –CXC on their C-terminus) were extracted from that data and summarized in Table 4. A total of 15 prenylated proteins are listed, including 14 of the 18 predicted in our bioinformatics analysis. One additional protein (Q81LH7) bearing the –CaaX sequence CNFM, which was not predicted by PrePS, was also identified. Conversely, four of the –CaaX sequences predicted by PrePS were not observed (Q81583, Q81EC5, Q81E80 and Q81EK2 with –CaaX sequences CLVF, CTIM, CKQC and CNIM, respectively). All Rab proteins predicted to be prenylated by PrepPS were identified.

The list of all proteins identified in our analysis (Supplemental Table 1) is ordered based on the fold change between the C15AlkOPP- and FPP-treated samples. The top 12 proteins in that list correspond to the first 12 entries in Table 4 (9 to 33-fold change). For each of these top 12 hits, very few spectral counts were observed in the absence of probe, giving high confidence that these represent bona fide prenylated proteins. In contrast, numerous proteins were found below the 6-fold threshold. In those cases, substantial spectral counts were observed in the absence of probe; we attribute those proteins to nonspecific adsorption. However, the last 3 entries in Table 4, while manifesting low levels of spectral counts in the C15AlkOPP-treated samples, gave no spectral counts in the FPP-treated samples, suggesting that they may represent true, low abundance hits; one of those three, SNARE Ykt6.1, has been confirmed as a prenyltransferase

110

substrate [37]. It is worth noting that the aforementioned DEAD/DEAH box Helicase and Methionyl-tRNA formyltransferase suggested by previous investigators as possible prenylated proteins based on bioinformatics analysis [36] are not present in the list of proteins in Table S1. Finally, it should be noted that human Rab homologs were also identified and grouped into clusters with the parent proteins from *P. falciparum* (Supplementary Table S1 and Figure S2), since the data analysis was conducted using both the *H. sapiens* and *P. falciparum* databases. However, for each identified Rab, the change in spectral counts in the presence of probe was always higher for the candidate malarial proteins than their human homologs. Additionally, we found, in each case, that protein probabilities for parasite sequences were higher than those for the cognate human orthologs. Together, these characteristics indicate that the Rab proteins identified upon metabolic labeling of *P. falciparum* are indeed malarial in origin and that the human candidates are artifacts of sequence similarities.

4.5 Discussion

Due to the ongoing spread of drug resistance, there is a pressing need for new therapies to treat malaria. Evidence strongly suggests that protein prenylation is required for asexual development of the *P. falciparum* malaria parasite; several distinct chemotypes of prenyltransferase inhibitors exhibit potent antimalarial activity [15,16,18,40–42]. Given the essential nature of protein prenylation, it follows that the functions of prenyltransferase substrates themselves are necessary for parasite replication. Thus, identification of prenylated proteins in *P. falciparum* may reveal new, essential, and highly valuable targets for antimalarial drug development. Such targeting

could be accomplished either indirectly by interfering with their prenylation or directly via inhibition of their cognate functions.

Here we present the first experimentally determined catalog of prenylated proteins (the "prenylome") of blood stage *P. falciparum*. We have identified prenyltransferase substrates through the use of metabolic labeling with a novel, alkyne-containing isoprenoid analogue. We find that the *P. falciparum* farnesyl pyrophosphate synthase (FPPS) successfully elongates the probe (which is a derivative of FPP) to generate the cognate 20-carbon (GGPP derivative) probe. Therefore, we believe our in vivo metabolic labeling approach has likely captured the full complement of both farnesylated and geranylgeranylated proteins in *P. falciparum*, with the exception of prenylated proteins with very low levels of expression during blood-stage development that may not react in our derivatization strategy.

Eukaryotic systems possess three different protein prenyltransferases: farnesyltransferase (FT) and geranylgeranyltransferase type I (GGT-1) commonly recognize the same motif (the CaaX box) that includes the cysteine of their substrates they modify, and are thus referred to as CaaX prenyltransferases, whereas geranylgeranyltransferase type II (GGT-2, also called Rab geranylgeranyltransferase) recognizes an alternative motif [43]. Active prenyltransferases consist of two polypeptide subunits, α and β; FT and GGT-1 typically share an α subunit. As has previously been suggested [19], we find that experimentally confirmed prenyltransferase substrates of *P. falciparum* parasites possess canonical motifs that indicate the presence of both CaaX prenyltransferases and Rab geranylgeranyltransferases. *P. falciparum* lysate has previously been shown to possess both FT and GGT-1 activity [12]. The current annotation of the *P.*

falciparum genome indicates a full complement of genes encoding the candidate prenyltransferases, as follows: *Pf*FT [PF3D7_1242600 (α subunit) and PF3D7_1147500 (β subunit)], *Pf*GGT-1 [PF3D7_1242600 (α subunit; shared with *Pf*FT) and PF3D7_0602500 (β subunit)], and *Pf*GGT-2 [PF3D7_1442500 (α subunit) and PF3D7_1214300 (β subunit)].

Our work experimentally confirms that protein prenylation in *P. falciparum* reflects a more modest set of prenylated proteins than is observed in fungi or higher eukaryotes, including humans. According to PRENbase (http://mendel.imp.ac.at/PrePS/PRENbase/), a curated online database of protein prenylation across sequenced genomes, prenylated proteins in the human genome are spread into 43 clusters of paralogous proteins [44]. Biological functions of prenylated proteins are well conserved, even amongst unicellular eukaryotes, as 42 similarly defined clusters are present across fungal genomes. In stark contrast, we find robust experimental evidence for a total of only thirteen prenylated proteins in *P. falciparum*, with suggestive evidence for an additional two probable prenyltransferase substrates. While a restricted prenylome has been suggested bioinformatically for malaria parasites, our study provides important evidence that there are not unrecognized, non-canonical motifs used by the *P. falciparum* prenyltransferases.

During asexual replication, *P. falciparum* is an obligate parasite of human erythrocytes, and relies on vesicle-mediated trafficking of erythrocyte cytoplasm and hemoglobin, as well as export of essential proteins for remodeling of the erythrocyte membrane and cytoplasm. The importance of membrane trafficking to *P. falciparum* development is underscored by the restricted biological functions of the malaria prenylome. Interestingly, we find that the majority

113

of proteins in the *P. falciparum* prenylome belong to a single cluster of paralogous proteins, the Rab family of small GTPases, classic regulators of endomembrane trafficking. These proteins likely make up the broad band at 25 kDa found in our in-gel fluorescence images upon C15OPP labeling (Fig. 1C), as well as in studies using radiolabeling of parasites with [^{3}H]-geranylgeranyl pyrophosphate [45]. Unsurprisingly, of the eleven Rabs annotated in *P. falciparum*, we found all ten predicted to be geranylgeranylated by GGT-2 [46].

In addition, three more prenyltransferase substrates in *P. falciparum* (two SNARE proteins and a phosphatidylinositol 3-phosphate binding protein) are also likely to function in membrane trafficking. Notably absent from the *P. falciparum* prenylome are a number of GTPase superfamilies, including Ras and Rho, typical of other unicellular eukaryotes and metazoans. This restricted use of protein prenylation for a single biological function reflects the complement of small GTPases that has been suggested to have been present in the last common eukaryotic ancestor, prior to the dramatic expansions of paralogous GTPase gene families [47,48]. The limited collection of GTPases and prenylated proteins that we find in *P. falciparum* is therefore not unique to this parasite, but is shared with other Alveolates in this lineage, including several other important mammalian parasitic pathogens, such as *Cryptosporidium*, *Toxoplasma*, and *Eimeria*.

We identify only four confirmed prenylated proteins in *P. falciparum* that possess a canonical CaaX motif, which should serve for recognition and modification by either FT or GGT-1. Bioinformatic analyses are insufficient to indicate whether a given CaaX-containing protein is farnesylated or geranylgeranylated. However, experimental evidence suggests that at least one of

these proteins, PF14_0359, a Hsp40 analog, is specifically farnesylated. In *P. falciparum*, metabolic labeling with [3H]-farnesyl pyrophosphate, but not [3H]-geranylgeranyl pyrophosphate, identifies a dominant band at approximately 50 kDa [45]. As the remaining proteins in the malaria prenylome are between 23-38 kDa, this finding most likely represents farnesyl modification of *Pf*Hsp40 (48 kDa) [49].

P. falciparum expresses an expanded repertoire of molecular chaperones, comprising 2% of the overall genome, and including 49 Hsp40 superfamily members in total [50]. However, while other eukaryotes typically express up to five type I Hsp40s, *Pf*Hsp40 is the sole, cytosolic type I Hsp40 homolog in the malaria parasite. In other organisms, farnesylation of orthologous type I Hsp40s has been well described, and is required for the biological functions of these chaperones in mediating protein stability [51,52]. The cellular function of *Pf*Hsp40 in *P. falciparum* has yet to be explored, although immunofluorescence microscopy indicates that this protein is cytosolic [49]. However, data from proteomics and yeast two-hybrid studies indicate it may play a role in trafficking to the RBC membrane [53]. The extent to which farnesylation plays a role in the localization or functions of *Pf*Hsp40 remains to be explored.

Our work contributes to an increasingly complete picture of post-translational hydrophobic modifications in blood-stage *P. falciparum*. The identification of a limited set of CaaX proteins has important implications for understanding the evolution of this modification process, as well as the active work developing novel antimalarial therapies targeted to isoprenoid synthesis and prenyltransferases.

4.6 Figures
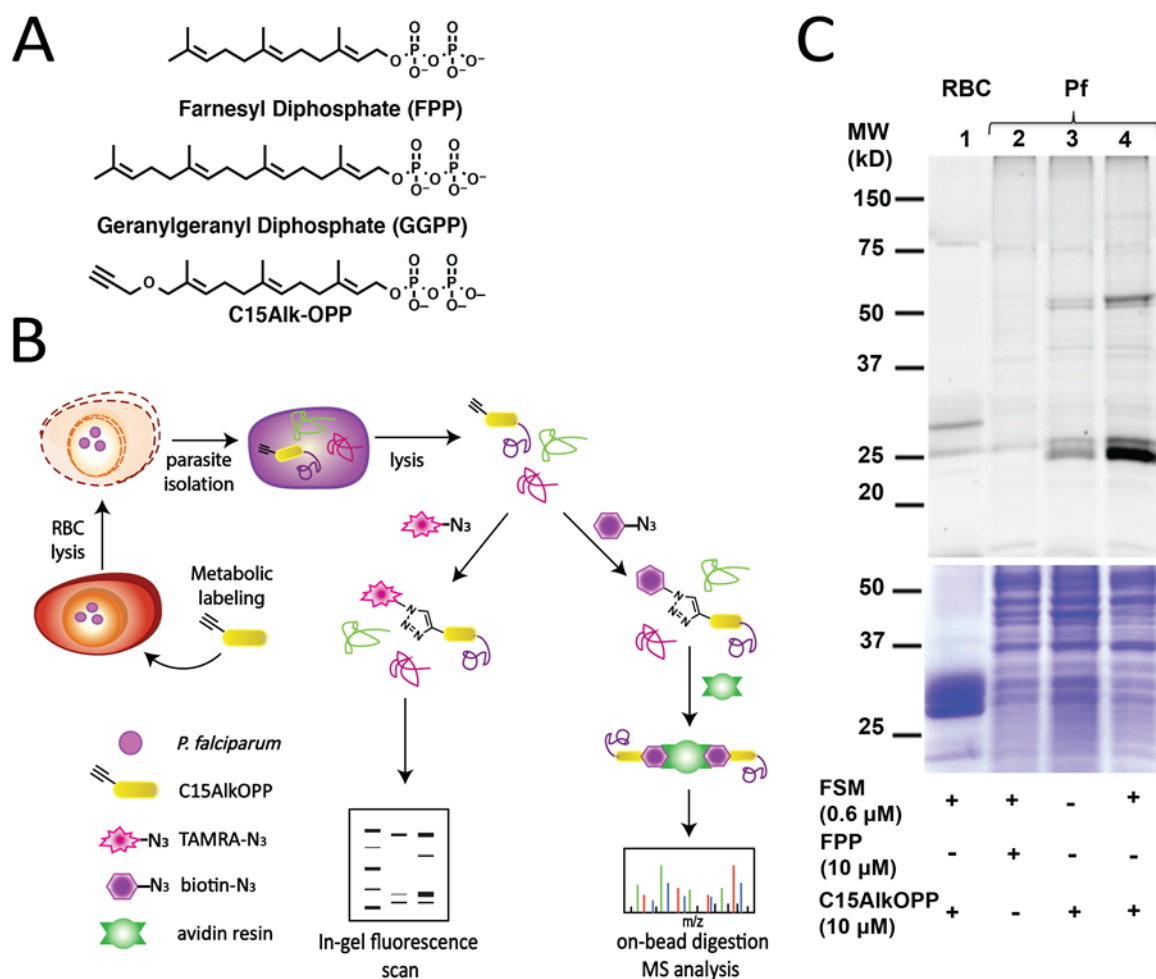
Figure 1. The C15AlkOPP probe allows tagging of prenylated proteins for in-gel fluorescence labeling and pulldown for proteomic analysis. (A) Structures of native isoprenoid substrates of prenylation, farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP), and the probe analog C15AlkOPP. (B) Scheme for metabolic labeling using the C15AlkOPP probe followed by selective labeling or enrichment using click chemistry. In-gel fluorescence and proteomic analysis of prenylated proteins were facilitated through click reactions with fluorescent reporter TAMRA-N3 and affinity handle biotin-N3, respectively. (C) Labeling of

prenylated proteins visualized through in-gel fluorescence (top panel) in P. falciparum lysates.

Lane 1: C15AlkOPP + FSM in red blood cells; lane 2: FPP + FSM in Pf; lane 3: C15AlkOPP in

Pf; lane 4: C15AlkOPP + FSM in Pf. Total protein loading by Coomassie blue stain is shown in

purple (bottom panel).
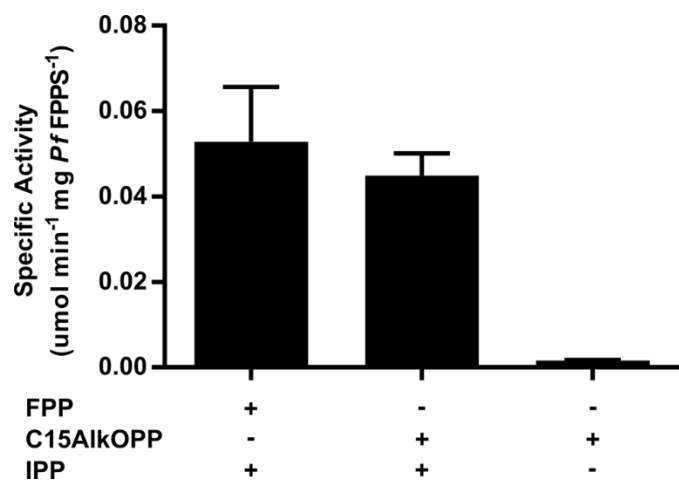
Figure 2. *P. falciparum* farnesyl pyrophosphate synthase (PfFPPS) accepts C15AlkOPP as a substrate. Specific activity (mean ± S.E.M.) of PfFPPS, measured using an absorbance-based phosphate release assay. Left to right, enzyme activity in the presence of FPP and IPP, C15AlkOPP probe and IPP, or C15AlkOPP alone. No activity was seen in absence of PfFPPS (data not shown). Three replicates were performed for each condition.

4.7 Tables

Table 1. Candidate prenyl transferase substrates from the *P. falciparum* 3D7 proteome, which contain canonical CaaX motifs and were identified by PrePS. PrePS was used to predict the likelihood of prenylation of the proteins identified: *low, **intermediate, ***high

Accession No.	Name	C-term	Mol. wt. (kDa)	PrePS
Q8I346	SNARE protein (putative)	KKQCCSIM	23	FT*** GGT-1***
Q8I583	Ulp1 protease (putative)	ISQGCLVF	123	FT* GGT-1*
Q8IKN0	Uncharacterized protein	QRRMCNIM	38	FT* GGT-1**
Q8IEC5	Uncharacterized protein	KKKKCTIM	95	FT*** GGT-1***
Q8IL88	Hsp40 subfamily A (putative)	GRVACAQQ	48	FT**
C0H5D3	SNARE protein (putative)	KNNQCCSLY	26	FT*
Q8IE80	Uncharacterized protein	NIAACKQC	34	FT*
Q8IEK2	Uncharacterized protein	NRLKCNIM	73	GGT-1**

Table 2. Candidate prenyl transferase substrates from the *P. falciparum* 3D7 proteome, which were not identified by PrePS but possess additional, promising CaaX features. PrePS was used to predict the likelihood of prenylation of the proteins identified: *low, **intermediate, ***high

Accession No.	Name	C-term	Mol. wt. (kDa)	Basic residue at -1 to -5	Hydrophobic residue at +2	Aromatic residue at +2
Q8IIN1	Protein tyrosine phosphatase	NCLRKCHFM	25	+	+	+
Q8ILH7	Uncharacterized protein	KKRNKCNFM	51	+	+	+
Q8I3K7	Membrane skeletal protein IMC1-related	QRNLYCSYA	34	+	+	+
Q8I0W8	Deoxyribodipyrimidine photolyase	KREKKCVAS	129	+	+	-
O77306	Ser/Thr protein kinase	NKKNSCAYT	157	+	+	-
Q9NLB7	Uncharacterized protein	NYNFLCLYI	10	-	+	+
O77380	CPSF	DLENMCSFL	339	-	+	+

Table 3. Candidate prenyl transferase substrates from the *P. falciparum* 3D7 proteome, which possess –CXC and –CC C-terminal motifs for possible geranylgeranylation. PrePS was used to predict the likelihood of prenylation of the proteins identified: *low, **intermediate, ***high

Accession No.	Name	C-term	Mol. wt. (kDa)	PrePS
-CXC				
Q8IL79	Copper transporter putative	ADPACCGC	27	No
Q8IM51	Secreted ookinete adhesive protein	ECSCSCSC	23	No
Q8IHR8	Rab6	NMLSKCLC	24	GGT-2**
Q7K6B0	Rab18	ESRSNCAC	23	GGT-2***
Q8I3W9	Rab1a	SPQSFCSC	24	GGT-2***
-CC				
Q8IHW0	Myosin heavy chain subunit, putative	ELNMFKCC	208	No
C6KST4	Uncharacterized protein	IKKKKMRCC	33	No
Q8II49	Conserved Plasmodium protein	TKKFFPCC	30	No
Q8IHW1	Conserved Plasmodium protein	KTKKCYCC	19	No
Q8IAL1	Uncharacterized protein	GKRFLGCC	9	No
Q7K6A8	Rab1b	KDTKKKCC	23	GGT-2***
O96193	Rab5a	TLSKKGCC	27	GGT-2***
Q8I274	Rab5c	EETKKKCC	24	GGT-2***
Q76NM4	Rab11a	TKKKNKCC	25	GGT-2***
Q8I5A9	Rab2	SRSGFSCC	24	GGT-2***
C0H516	Rab7	KMYKSRCC	24	GGT-2***
C0H5G2	Rab11b	NMNKVKCC	24	GGT-2**

Table 4. Prenylated proteins in *Plasmodium falciparum*, identified by C15AlkOPP labeling and proteomics, with –CaaX, -CXC, and –CC prenylation motifs. For each protein, fold-change indicates the ratio of the average total spectral counts obtained following C15AlkOPP- versus FPP-labeling, across three experimental replicates. An imputation of 1 was employed to calculate fold changes. Spectral counts and percent coverage for each replicate are shown. Two uncharacterized proteins (shown in red text) containing predicted prenylation motifs were identified with low spectral counts. [a]Protein accession numbers including theoretical molecular weights and [b]gene IDs were obtained from UniProt and PlasmoDB, respectively. [C]Unless otherwise referenced from previous reports, biological processes and cellular component of identified proteins were obtained from UniProt.[d] PrePS was used to predict the likelihood of prenylation of the proteins identified: *low, **intermediate, ***high.

Protein Name	Accession No.[a]/Gene ID[b]	Mol. wt. (kDa)[a]	Biological Process[c]	Cellular Component[c]	Spectral Count C15Alk-OPP	FPP	Fold Change	Percent coverage (%) C15Alk-OPP	FPP	C-term sequence[c]	Confirmed prenylation	PrePS[d]
Rab6	Q8IHR8 PF3D7_1144900	24	vesicular transport (putative)	trans Golgi, cytosolic [35]	33,32,33	0,0,0	33	70,70,59	0,0,0	DKNMLSKCLC	Yes [35]	GGT-2**
HSP40 subfamily A	Q8IL88 PF3D7_1437900	48	protein folding [49], protein translocation (putative) [54]	cytosolic, RBC membrane[49]	39,34,37	1,2,0	28	41,35,33	2,5,0	GRVACAQQ	No	FT**
Rab7	C0H616 PF3D7_0903200	24	retrograde transport [55]	endosomes [55]	25,27,25	0,0,0	26	59,66,66	0,0,0	KMYKSRCC	No	GGT-2***
Rab1b	Q7K6A8 PF3D7_0512600	23	ER to Golgi transport (putative) [46]	ER, Golgi (putative) [46]	19,12,17	0,0,0	16	37,35,31	0,0,0	PDTKKKCC	No	GGT-2***
Rab1a	Q8I3W9 PF3D7_0513800	24	ER to Golgi transport [46]	ER, Golgi [46]	11,10,12	0,0,0	11	17,21,22	0,0,0	SPQSFCSC	No	GGT-2***
Rab2	Q8I5A9 PF3D7_1231100	24	ER to Golgi transport (putative) [46]	ER, Golgi (putative) [46]	25,21,28	2,0,0	18	58,56,57	11,0,0	SRSGFSCC	No	GGT-2***
Rab5c	Q8I274 PF3D7_0106800	24	retrograde transport [14]	endosomes [14]	18,15,14	0,0,0	16	43,35,33	0,0,0	EETKKKCC	Yes [14]	GGT-2***
Rab11a	Q76NM4 PF3D7_1320600	25	formation of inner membrane complex [56], recycling endosome transport (putative)	vesicles [56]	12,17,12	0,0,0	14	41,42,41	0,0,0	TKKKNKCC	No	GGT-2***
Rab5a	O96193 PF3D7_0211200	27	retrograde transport [14]	endosomes [14]	9,9,11	0,0,0	9.7	25,29,29	0,0,0	TLSKKGCC	Yes [14]	GGT-2***
Rab18	Q7K6B0 PF3D7_0807300	23	mobilization of lipid droplets (putative) [57]	Parasitopho-rous vacuole membrane (putative) [57]	9,7,11	1,1,1	9	28,24,28	6,6,6	ESRSNCAC	No	GGT-2***
Rab11b	C0H5G2 PF3D7_1340700	24	transport (putative)	membrane (putative)	8,10,9	0,0,0	9	30,25,31	0,0,0	NMNKVKCC	No	GGT-2**
SNARE Ykt6.2	C0H5D3 PF3D7_1324700	26	vesicular fusion, transport (putative) [58]	membrane (putative)	5,6,7	0,0,0	6	14,18,24	0,0,0	KNNQCCSLY	No	FT*
SNARE Ykt6.1	Q8I346 PF3D7_0910600	23	vesicular fusion, transport (putative) [58]	Golgi, cytosolic, vacuolar [37]	3,2,2	0,0,0	2.3	17,13,11	0,0,0	LKQCCSIM	Yes [58]	FT*, GGT-1***
Uncharacterized	Q8IKN0 PF3D7_1460100	38	PI3P binding, trafficking to the food vacuole [59]	food vacuole lumen [59]	0,2,1	0,0,0	1.3	0,8,5	0,0,0	QRRMCNIM	No	FT*, GGT-1**
Uncharacterized	Q8ILH7 PF3D7_1428700	51			0,1,3	0,0,0	1.7	0,3,5	0,0,0	KRNKCNFM	No	-

4.8 References

1. Feachem RGA, Phillips AA, Hwang J, Cotter C, Wielgosz B, Greenwood BM, et al. Shrinking the malaria map: progress and prospects. Lancet. 2010;376: 1566–1578.

2. World Health Organization. World malaria report 2014. Geneva: WHO; 2014.

3. Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, et al. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. N Engl J Med. 2014;371: 411–423.

4. von Seidlein L, Dondorp A. Fighting fire with fire: mass antimalarial drug administrations in an era of antimalarial resistance. Expert Rev Anti Infect Ther. 2015;13: 715–730.

5. Eastman RT, Buckner FS, Yokoyama K, Gelb MH, Van Voorhis WC. Fighting parasitic disease by blocking protein farnesylation. J Lipid Res. 2006;47: 233–240.

6. Gelb MH, Van Voorhis WC, Buckner FS, Yokoyama K, Eastman R, Carpenter EP, et al. Protein farnesyl and N-myristoyl transferases: piggy-back medicinal chemistry targets for the development of antitrypanosomatid and antimalarial therapeutics. Mol Biochem Parasitol. 2003;126: 155–163.

7. Zhang FL, Casey PJ. Protein prenylation: molecular mechanisms and functional consequences. Annu Rev Biochem. 1996;65: 241–269.

8. Palsuledesai CC, Distefano MD. Protein prenylation: enzymes, therapeutics, and biotechnology applications. ACS Chem Biol. 2015;10: 51–62.

9. Bos JL. Ras oncogenes in human cancer: a review. Cancer Res . 1989;49: 4682–4689.

10. Berndt N, Hamilton AD, Sebti SM. Targeting protein prenylation for cancer therapy. Nat Rev Cancer. 2011;11: 775–791.

11. Ochocki JD, Distefano MD. Prenyltransferase inhibitors: treating human ailments from cancer to parasitic infections. Med Chem Commun. 2013;4: 476–492.

12. Chakrabarti D, Azam T, DelVecchio C, Qiu L, Park Y, Allen CM. Protein prenyl transferase activities of *Plasmodium falciparum*. Mol Biochem Parasitol. 1998;94: 175–184.

13. Chakrabarti D, Silva T Da, Barger J, Paquette S, Patel H, Patterson S, et al. Protein farnesyltransferase and protein prenylation in *Plasmodium falciparum*. J Biol Chem. 2002;277: 42066–42073.

14. Howe R, Kelly M, Jimah J, Hodge D, Odom AR. Isoprenoid biosynthesis inhibition disrupts Rab5 localization and food vacuolar integrity in *Plasmodium falciparum*.

Eukaryot Cell . 2013;12: 215–223.

15.   Wiesner J, Kettler K, Sakowski J, Ortmann R, Katzin AM, Kimura EA, et al. Farnesyltransferase inhibitors inhibit the growth of malaria parasites in vitro and in vivo. Angew Chemie Int Ed. 2004;43: 251–254.

16.   Nallan L, Bauer KD, Bendale P, Rivas K, Yokoyama K, Hornéy CP, et al. Protein farnesyltransferase inhibitors exhibit potent antimalarial activity. J Med Chem. 2005;48: 3704–3713.

17.   Glenn MP, Chang S-Y, Hucke O, Verlinde CLMJ, Rivas K, Hornéy C, et al. Structurally simple farnesyltransferase inhibitors arrest the growth of malaria parasites. Angew Chemie Int Ed. 2005;44: 4903–4906. doi:10.1002/anie.200500674

18.   Glenn MP, Chang S-Y, Hornéy C, Rivas K, Yokoyama K, Pusateri EE, et al. Structurally simple, potent, *Plasmodium* selective farnesyltransferase inhibitors that arrest the growth of malaria parasites. J Med Chem. 2006;49: 5710–5727.

19.   Buckner FS, Eastman RT, Yokoyama K, Gelb MH, Van Voorhis WC. Protein farnesyl transferase inhibitors for the treatment of malaria and African trypanosomiasis. Curr Opin Investig Drugs. 2005;6: 791.

20.   Trager W, Jensen JB. Human malaria parasites in continuous culture. Science. 1976;193: 673–675.

21.   Zhang B, Watts KM, Hodge D, Kemp LM, Hunstad DA, Hicks LM, et al. A second target of the antimalarial and antibacterial agent fosmidomycin revealed by cellular metabolic profiling. Biochemistry. 2011;50: 3570–3577.

22.   Mehlin C, Boni E, Buckner FS, Engel L, Feist T, Gelb MH, et al. Heterologous expression of proteins from *Plasmodium falciparum*: Results from 1000 genes. Mol Biochem Parasitol. 2006;148: 144–160.

23.   Imlay LS, Armstrong CM, Masters MC, Li T, Price KE, Edwards RL, et al. *Plasmodium* IspD (2-C-methyl-d-erythritol 4-Phosphate Cytidyltransferase), an essential and druggable antimalarial target. ACS Infect Dis. 2015;1: 157–167.

24.   Kalli A, Smith GT, Sweredoski MJ, Hess S. Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: Focus on LTQ-Orbitrap mass analyzers. J Proteome Res. 2013;12: 3071–3086.

25.   Consortium TU. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res . 2012;40: D71–D75.

26.   Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003;75: 4646–4658.

27.   DeGraw AJ, Palsuledesai C, Ochocki JD, Dozier JK, Lenevich S, Rashidian M, et al.

Evaluation of alkyne-modified isoprenoids as chemical reporters of protein prenylation. Chem Biol Drug Des. 2010;76: 460–471.

28. Palsuledesai CC, Ochocki JD, Markowski TW, Distefano MD. A combination of metabolic labeling and 2D-DIGE analysis in response to a farnesyltransferase inhibitor facilitates the discovery of new prenylated proteins. Mol BioSyst. 2014;10: 1094–1103.

29. Hosokawa A, Wollack JW, Zhang Z, Chen L, Barany G, Distefano MD. Evaluation of an alkyne-containing analogue of farnesyl diphosphate as a dual substrate for protein-prenyltransferases. Int J Pept Res Ther. 2007;13: 345–354.

30. Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, et al. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science. 1999;285: 1573–1576.

31. Kakhniashvili DG, Bulla LA, Goodman SR. The human erythrocyte proteome: analysis by ion trap mass spectrometry. Mol Cell Proteomics. 2004;3: 501–509.

32. Chiancone E, Gilbert LM, Gilbert GA, Kellett GL. Dissociation of Hemoglobin into Subunits: II. HUMAN OXYHEMOGLOBIN: GEL FILTRATION STUDIES . J Biol Chem. 1968;243: 1212–1219.

33. Artz JD, Wernimont AK, Dunford JE, Schapira M, Dong A, Zhao Y, et al. Molecular characterization of a novel geranylgeranyl pyrophosphate synthase from *Plasmodium* parasites. J Biol Chem. 2011;286: 3315–22.

34. Jordão FM, Gabriel HB, Alves JMP, Angeli CB, Bifano TD, Breda A, et al. Cloning and characterization of bifunctional enzyme farnesyl diphosphate/geranylgeranyl diphosphate synthase from *Plasmodium falciparum*. Malar J. 2013;12: 1–15.

35. Struck NS, Herrmann S, Schmuck-Barkmann I, De Souza Dias S, Haase S, Cabrera AL, et al. Spatial dissection of the cis- and trans-Golgi compartments in the malaria parasite *Plasmodium falciparum*. Mol Microbiol. 2008;67: 1320–1330.

36. Pendyala PR, Ayong L, Eatrides J, Schreiber M, Pham C, Chakrabarti R, et al. Characterization of a PRL protein tyrosine phosphatase from *Plasmodium falciparum*. Mol Biochem Parasitol. 2008;158: 1–10.

37. Ayong L, DaSilva T, Mauser J, Allen CM, Chakrabarti D. Evidence for prenylation-dependent targeting of a Ykt6 SNARE in *Plasmodium falciparum*. Mol Biochem Parasitol. 2011;175: 162–168.

38. Maurer-Stroh S, Eisenhaber F. Refinement and prediction of protein prenylation motifs. Genome Biol. 2005;6: 1–15.

39. Soni R, Sharma D, Patel S, Sharma B, Bhatt TK. Structure-based binding between protein farnesyl transferase and PRL-PTP of malaria parasite: an interaction study of prenylation process in *Plasmodium*. J Biomol Struct Dyn. 2016; 1–12.

40. Ohkanda J, Lockman JW, Yokoyama K, Gelb MH, Croft SL, Kendrick H, et al. Peptidomimetic inhibitors of protein farnesyltransferase show potent antimalarial activity. Bioorg Med Chem Lett. 2001;11: 761–764.

41. Olepu S, Suryadevara PK, Rivas K, Yokoyama K, Verlinde CLMJ, Chakrabarti D, et al. 2-Oxo-tetrahydro-1,8-naphthyridines as selective inhibitors of malarial protein farnesyltransferase and as anti-malarials. Bioorg Med Chem Lett. 2008;18: 494–497.

42. Jordão FM, Saito AY, Miguel DC, Peres VDJ, Kimura EA, Katzin AM. In vitro and in vivo antiplasmodial activities of risedronate and its interference with protein prenylation in *Plasmodium falciparum*. Antimicrob Agents Chemother. 2011;55: 2026–2031.

43. Casey PJ, Seabra MC. Protein Prenyltransferases. J Biol Chem. 1996;271: 5289–5292.

44. Maurer-Stroh S, Koranda M, Benetka W, Schneider G, Sirota FL, Eisenhaber F. Towards complete sets of farnesylated and geranylgeranylated proteins. Rost B, editor. PLoS Comput Biol. 2007;3: e66.

45. Moura IC, Wunderlich G, Uhrig ML, Couto AS, Peres VJ, Katzin AM, et al. Limonene arrests parasite development and inhibits isoprenylation of proteins in *Plasmodium falciparum*. Antimicrob Agents Chemother. 2001;45: 2553–2558.

46. Quevillon E, Spielmann T, Brahimi K, Chattopadhyay D, Yeramian E, Langsley G. The *Plasmodium falciparum* family of Rab GTPases. Gene. 2003;306: 13–25.

47. Dacks JB, Field MC. Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode. J Cell Sci. 2007;120: 2977–2985.

48. Rojas AM, Fuentes G, Rausell A, Valencia A. The Ras protein superfamily: Evolutionary tree and role of conserved amino acids. J Cell Biol. 2012;196: 189–201.

49. Botha M, Chiang AN, Needham PG, Stephens LL, Hoppe HC, Külzer S, et al. *Plasmodium falciparum* encodes a single cytosolic type I Hsp40 that functionally interacts with Hsp70 and is upregulated by heat shock. Cell Stress Chaperones. 2011;16: 389–401.

50. Pesce E-R, Cockburn IL, Goble JL, Stephens LL, Blatch GL. Malaria heat shock proteins: drug targets that chaperone other drug targets. Infect Disord. 2010;10: 147–157.

51. Flom GA, Lemieszek M, Fortunato EA, Johnson JL. Farnesylation of Ydj1 is required for in vivo interaction with Hsp90 client proteins. Mol Biol Cell . 2008;19: 5249–5258.

52. Summers DW, Douglas PM, Ren H-Y, Cyr DM. The Type I Hsp40 Ydj1 utilizes a farnesyl moiety and zinc finger-like region to suppress prion toxicity. J Biol Chem. 2009;284: 3628–3639.

53. Pesce E, Blatch GL. The Hsp40-Hsp70 chaperone machinery of *Plasmodium falciparum*. African J Biochem Res. 2009;3: 154–163.

54. Botha M, Pesce ER, Blatch GL. The Hsp40 proteins of *Plasmodium falciparum* and other apicomplexa: Regulating chaperone power in the parasite and the host. Int. J. Biochem. Cell Biol. 2007;39: 1781–1803.

55. Krai P, Dalal S, Klemba M. Evidence for a golgi-to-endosome protein sorting pathway in *Plasmodium falciparum*. PLoS One. 2014;9: e89771.

56. Agop-Nersesian C, Naissant B, Rached FB, Rauch M, Kretzschmar A, Thiberge S, et al. Rab11a-controlled assembly of the inner membrane complex is required for completion of apicomplexan cytokinesis. PLoS Pathog. 2009; 5: e1000270.

57. Langsley G, van Noort V, Carret C, Meissner M, de Villiers EP, Bishop R, et al. Comparative genomics of the Rab protein family in Apicomplexan parasites. Microbes Infect. 2008;10: 462–470.

58. Ayong L, Pagnotti G, Tobon AB, Chakrabarti D. Identification of *Plasmodium falciparum* family of SNAREs. Mol. Biochem. Parasitol. 2007;152: 113–122.

59. McIntosh MT, Vaid A, Hosgood HD, Vijay J, Bhattacharya A, Sahani MH, et al. Traffic to the malaria parasite food vacuole: a novel pathway involving a phosphatidylinositol 3-phosphate-binding protein. J. Biol. Chem. 2007;282: 11499–11508.

Chapter 5: Conclusion

5.1 Summary

The work presented here provides several valuable and actionable insights into *Plasmodium falciparum*. For the first time in a clinical population, malaria was shown to be associated with distinctive and potentially diagnostic changes in the breath volatile organic compound (VOC) profile of patients (Chapter 2). Preliminary analysis of sweat samples also found broad changes in the VOCs of infected patients (Appendix B). Of utility to future breath biomarker discovery studies, Chapter 3 gives the first head-to-head comparison of two prevalent breath collection methods, finding that sampling bags have distinct advantages over the Bio-VOC. Additionally, this dissertation elucidated a possible mechanism for the enhanced attraction of mosquitoes to infected hosts. Work in Appendix A discovered mosquito attractant terpenes in the air above malaria parasite cultures, namely the terpene α-pinene. It and the related terpene 3-carene were later found to be at higher levels in infected patient breath in two clinical populations (Chapters 2 and 3). The presence of higher terpene concentrations induced by malarial infection provides a plausible mechanism for increased mosquito attraction. This finding has implications for designing odorant lures for mosquito traps as well as for gaining a better understanding of the dynamics of malaria transmission. Finally, Chapter 4 used a novel labelling strategy to pinpoint all prenylated proteins in *P. falciparum*. This knowledge will prove valuable for understanding isoprenoid biology in the parasite as well the effects of antimalarials targeted at prenylation. A series of such inhibitors, specifically farnesyl transferase inhibitors (FTIs), is the subject of Appendix C.

5.2 Breath biomarkers of malaria

A few studies have noted VOC changes associated with animal model malaria and differences in mosquito attraction indicative of VOC changes [1–4]. One previous work found a set of small thioethers in breath associated with sub-clinical parasite levels during controlled human malaria infections [5]. However, our work was the first to evaluate breath changes in patients with active malaria. The discovery that a suite of six compounds could classify infection status with 83% accuracy provides viable evidence that breath could be a diagnostic medium for malaria. While our study provides groundwork for further studies, a single study is not definitive. A validation study in the same area with the same demographics will help confirm viability, but additional studies in populations of varying demographics and geographic locales are needed to determine whether our identified biomarkers are a widespread phenomenon. A consideration for these future studies is the possibility of better linking parasite density and life stage to VOC changes. Parasitemia, the metric of parasite burden used in our studies, is an imperfect measure. Conducting qPCR off-site, along with on-site diagnosis, will provide a highly accurate picture of parasite burden and the prevalence of the sexual stage gametocytes (discussed in section 5.4).

In future validation studies, the six compounds identified will provide a useful starting point but should not be the sole route of analysis. Beyond these six compounds, dozens of compounds correlated with infection status, and a portion of those VOCs may be more reliable across multiple studies. Indeed, while selective sensors for one or a few VOCs are possible, the current focus of point-of-care breath diagnosis is electronic nose (eNose) devices [6]. These eNoses detect the overall composition of the breath VOC profile and perform classification with the help

of machine learning algorithms built on detected abundances of numerous VOC sub-populations. Thus, continuing to use unbiased and aggregate analysis, whether with the correlation approach or other machine learning approaches like random forests, will maximize the chances of identifying VOC differences that will be useful for point-of-care diagnosis. In addition to being able to pick and tune the most appropriate eNose sensor system, the detailed and granular abundance and structural data of VOCs provided by GC/MS is still the most definitive approach to determining if breath profile changes exist. Any given eNose can fail in these early stages by virtue of performing poorly with the VOCs that happen to be important.

5.3 Breath collection methodology

As a growing but still niche field, breath biomarker discovery has seen an influx of new investigators. However, especially for studies on neglected tropical diseases, funds can be limited. Inexpensive, reliable tools for breath collection would help studies focusing on an expanding range of diseases. The Bio-VOC was seen as a low-cost, easy-to-handle alternative to the main method of breath collection, i.e. inert sampling bags. However, our findings indicate that the Bio-VOC's advantages are outweighed by its inferior performance compared to sampling bags. Surprisingly, no field comparison of the two methods had been performed previously, so such a comparative study is timely. Critically, the Bio-VOC fails to perform at a level that allows for reliable determination of breath biomarkers. Choosing it as the sole method for an initial study could erroneously exclude a disease as a candidate for breath diagnosis.

To expand on our findings, future studies should aim to test these differences in a rigorous lab setting. One helpful experiment would be to determine the effective limit of detection for both methods, a value that is known only approximately at present. Generalizing our findings, seen in pediatric patients, to all ages is also an important step. Another key test would be to evaluate the performance of the Bio-VOC when used to collect upwards of five breaths from each patient, as our study and other past studies have collected only three breaths. The increased capture volume may increase the limit-of-detection of the Bio-VOC to match that of the sampling bags, but variability associated with multiple breath samplings could arise. The potential for variability also points to an issue shared by sampling bags, the Bio-VOC, and other collection methods: a lack of standardization. While efforts have been made to standardize breath biomarker discovery in terms of method reporting [7], establishing a suggested minimum breath collection volume, based on breath VOC concentrations, would be a helpful step. Expanding standards to encompass agreed upon positive controls, such as isoprene levels or a basket of common breath VOCs, would help with inter-study comparability.

5.4 *P. falciparum* induced volatile changes

It was long assumed that mosquito transmission of *Plasmodium* parasites, and of numerous other pathogens, was a consequence of natural mosquito behavior without any input from the parasite. Only relatively recently did evidence emerge suggesting mosquitoes are disproportionately attracted to infected individuals, a fact now confirmed across several studies [8]. Given the designs of these studies, host odor seemed like the most likely mediator of increased attraction, but the exact mechanisms were not known, especially for human malaria. Our finding of known

133

plant-like attractants, especially α-pinene, emitted by parasite cultures and at higher levels in infected patient breath provides a plausible mechanism for increased mosquito attraction. However, the exact effects of α-pinene and 3-carene on mosquito behavior have not yet been elucidated. Testing these compounds against mosquitoes, probably as part of a blend that mimics human odors, should help settle the issue. The exact source of the terpenes is another outstanding question: we proposed that the parasite, which harbors multiple plant-like genes, may possess a plant-like terpene synthase. No such gene is annotated, but predicting terpene synthases from genome data is infamously difficult. Screening for the synthase via a toxic precursor build-up system in *E. coli* is one possibility, but *P. falciparum* genes do not regularly express and/or fold well in *E. coli*. If the parasite produces the terpenes, that would engender questions as to whether the terpenes made vary by geographic location to better match the preferences of local *Anopheles* species. Other mooted possibilities aside from parasite production include that the terpenes are arising from red blood cell metabolism [9] or are somehow being released from fatty deposits.

A linked question is the extent to which breath volatiles effect mosquito behavior in comparison to other body odors, most notably odors from sweat. A recent study found preferential mosquito attraction to sweat samples from infected patients, and sweat odors could be the dominant factor [10]. Disentangling this issue will require assessing mosquito attraction directly against collected breath samples or captured breath VOCs. Even if sweat VOCs have a greater role in attraction, breath VOCs will reflect those seen in sweat. An early indication of this is a putative monoterpene alcohol seen at higher levels in infected patient sweat samples in Appendix B (at an approximate retention time of 11 minutes).

A final area for future investigation is the life stage dependency of terpene prevalence and mosquito attraction. The malaria parasite can only complete its life cycle when taken up by a mosquito if the parasite has transitioned from the asexual blood stage cycle to the terminal sexual stage, i.e. gametocytes. Mature gametocytes take roughly two weeks after emergence from the liver to fully develop. Previous behavioral research indicates that the preferential attraction to infected hosts corresponds to the presence of mature gametocytes [11], though their presence is usually linked with peak overall parasite density. Knowledge of gametocyte levels via qPCR will help, but a more definitive approach would be to sample host volatiles before and after treatment that kills solely asexual parasites (a feature of some current frontline antimalarials). Regardless of life stage, if further confirmed, the alteration of host volatiles to increase vector attraction would be a new paradigm of parasite host manipulation that might apply beyond *P. falciparum*. Other eukaryotic parasites, like the causative agent of leishmaniasis transmitted by sandflies [12], are prime candidates to employ attraction tactics, but whether the phenomena extends to "simpler" pathogens such as arboviruses is an interesting avenue for exploration as well. Including increased preference for infected hosts also affects efforts to model malaria spread with regards to control strategies [13].

5.5 Roles of prenylation in *P. falciparum*

As with many eukaryotes, prenylation is an essential cellular process in *P. falciparum*. However, few protein substrates had been experimentally validated. By utilizing a comprehensive labeling technique, we present compelling evidence for the full suite of prenylated proteins in *P. falciparum*. Bolstering the labelling data, the substrates identified all possess features that predict

prenylation. A later, independent study using a similar labelling strategy found results broadly in agreement with our own, the few discrepancies explainable by the strictness of cut-off criteria [14]. Both studies found a limited number of prenylated proteins; based on predictive data the farnesyl transferase (FT) appears to have fewer than five substrates, in stark contrast to other organisms. This information has implications for drug development as farnesyl transferase activity has been shown to be essential for parasite growth; our findings imply that the essentiality hinges on a handful of downstream proteins. Since farnesyl transferase inhibitors (FTIs) are being explored as future antimalarials [15], knowledge of downstream processes will help to understand the mechanism of killing and how possible modes of resistance might arise. Hypothetically, inhibiting farnesylation of a single protein might kill the parasite, and targeting that specific prenylation event, or the substrate protein itself, may provide targets for new antimalarials.

In service of identifying antimalarial targets, but also expanding the understanding of prenylation as a biological process, several future studies are warranted. Evolutionarily, *P. falciparum* exists before the expansion of the Ras superfamily [16], which explains the few predicted geranylgeranyl transferase type II (GGT-2) substrates. For the protein substrates of geranylgeranyl transferase type I (GGT-1) and FT, it is unclear if their reduced number is in line with other species of the same lineage or if it is the result of *P. falciparum*'s parasitic lifestyle. Expanding prenylome studies to multiple species will help answer this question; currently only *P. falciparum* and various human cell types have been explored with comprehensive labelling. It is worth noting that *P. falciparum* may have additional prenylated proteins that are exclusive to mosquito or liver stage parasites, though predictive data suggest that numerous additional

136

substrates are unlikely to be discovered. Determining which proteins of the prenylome are

farnesylated specifically will be essential for understanding FTI activity. Strong evidence exists

that Hsp40 is farnesylated; performing labelling while administering FTIs should help identify

the others. The protein Hsp40 warrants special investigation as it is easily the most prevalent

farnesylated protein and could be a key essential enzyme. Attempting to knock out Hsp40 or

altering its prenylation site will help identify the role that prenylation plays for this enzyme,

which, unusually for a prenylated protein, appears to be mostly cytosolic.

5.6 Final thoughts

Advances in diagnostics, mosquito vector control, and antimalarial treatments are crucial in

decreasing and ultimately eliminating the malaria burden worldwide. The work presented in this

dissertation sets the ground for innovations in all three areas by establishing the plausibility of

breath diagnosis, identifying malaria induced attractants with implications for mosquito traps and

malaria transmission, and elucidating the substrates for a metabolic process targeted by a class of

potential antimalarial compounds. As this work built upon previous findings, so must future

studies validate and expand upon what is presented here to achieve the goal of translating

observations into actions to defeat malaria.

5.7 References

1.	Cornet S, Nicot A, Rivero A, Gandon S. Malaria infection increases bird attractiveness to uninfected mosquitoes. Ecol Lett. 2013;16: 323–329.

2.	Batista EPA, Costa EFM, Silva AA. *Anopheles darlingi* (Diptera: Culicidae) displays increased attractiveness to infected individuals with *Plasmodium vivax* gametocytes. Parasit Vectors. 2014;7: 251.

3.	Lacroix R, Mukabana WR, Gouagna LC, Koella JC. Malaria infection increases attractiveness of humans to mosquitoes. PLoS Biol. 2005;3: e298.

4.	De Moraes CM, Stanczyk NM, Betz HS, Pulido H, Sim DG, Read AF, et al. Malaria-induced changes in host odors enhance mosquito attraction. Proc Natl Acad Sci USA. 2014;111: 11079–84.

5.	Berna AZ, McCarthy JS, Wang RX, Saliba KJ, Bravo FG, Cassells J, et al. Analysis of breath specimens for biomarkers of *Plasmodium falciparum* infection. J Infect Dis. 2015;212: 1120–1128.

6.	Rattray NJW, Hamrang Z, Trivedi DK, Goodacre R, Fowler SJ. Taking your breath away: Metabolomics breathes life in to personalized medicine. Trends Biotechnol. 2014;32: 538–548.

7.	Herbig J, Beauchamp J. Towards standardization in the analysis of breath gas volatiles. J Breath Res. 2014;8: 37101.

8.	Busula AO, Verhulst NO, Bousema T, Takken W, de Boer JG. Mechanisms of *Plasmodium*-enhanced attraction of mosquito vectors. Trends Parasitol. 2017; 1–13.

9.	Emami SN, Lindberg BG, Hua S, Hill SR, Mozuraitis R, Lehmann P, et al. A key malaria metabolite modulates vector blood seeking, feeding, and susceptibility to infection. Science. 2017;355: 1076–1080.

10.	Robinson A, Busula AO, Voets MA, Beshir KB, Caulfield JC, Powers SJ, et al. *Plasmodium*-associated changes in human odor attract mosquitoes. Proc Natl Acad Sci USA. 2018.

11.	Busula AO, Bousema T, Mweresa CK, Masiga D, Logan JG, Sauerwein RW, et al. Gametocytaemia increases attractiveness of *Plasmodium falciparum*-infected Kenyan children to *Anopheles gambiae* mosquitoes. J Infect Dis. 2017.

12.	O'Shea B, Rebollar-Tellez E, Ward RD, Hamilton JGC, El Naiem D, Polwart A. Enhanced sandfly attraction to Leishmania-infected hosts. Trans R Soc Trop Med Hyg. 2002;96: 117–118.

13.	Buonomo B, Vargas-De-León C. Effects of mosquitoes host choice on optimal

intervention strategies for malaria control. Acta Appl Math. 2014;132: 127–138.

14. Gisselberg JE, Zhang L, Elias JE, Yeh E. The prenylated proteome of *Plasmodium falciparum* reveals pathogen-specific prenylation activity and drug mechanism-of-action. Mol Cell Proteomics. 2017;16: S54–S64.

15. Ochocki JD, Distefano MD. Prenyltransferase inhibitors: treating human ailments from cancer to parasitic infections. Med Chem Commun. 2013;4: 476–492.

16. Rojas AM, Fuentes G, Rausell A, Valencia A. The Ras protein superfamily: Evolutionary tree and role of conserved amino acids. J Cell Biol. 2012;196: 189–201.

Appendix A: Malaria parasites produce volatile mosquito attractants

Preface

The following work was performed by myself, Megan Kelly, Chih-Ying Su, Jan R. Crowley, Fong-Fu Hsu, John R. Carlson, and Audrey R. Odom John. My main contribution was using a new approach to re-analyze the SPME data in order to better establish the *P. falciparum* specific volatiles. I updated Figure 2 and created Figure S2 to reflect the new conclusions drawn from this analysis. I also aided in updating portions of the manuscript, and I contributed a small amount of new material. Otherwise, for the experiments in *P. falciparum* AOJ and MK conceived and designed the experiments; AOJ, MK, JRC, FH, and myself performed the experiments and analyzed the data. For the work done in *Anopheles gambiae* odorant receptors, J Carlson and CY-S designed the experiments; CY-S performed the experiments; and CY-S and J Carlson analyzed the data. MK, CY-S, J Carlson, and AOJ wrote the manuscript.

A.1 Abstract

The malaria parasite, *Plasmodium falciparum*, contains a non-photosynthetic plastid organelle that possesses plant-like metabolic pathways. Plants use the plastidial isoprenoid biosynthesis pathway to produce volatile odorants, known as terpenes. In this work, we describe the volatile chemical profile of cultured malaria parasites. Among the identified compounds are several plant-like terpenes and terpene derivatives, including known mosquito attractants. We establish the molecular identity of the odorant receptors of the malaria mosquito vector, *Anopheles gambiae,* which respond to these compounds. The malaria parasite produces volatile signals that are recognized by mosquitoes and may thereby mediate host attraction and facilitate transmission.

A.2 Introduction

Malaria remains an enormous burden to human health worldwide. There are over 250 million cases of malaria each year, and nearly one million deaths [1]. A single protozoan species, *Plasmodium falciparum*, is responsible for the most severe and deadly cases of this disease. Widespread and emerging drug resistance has contributed to a resurgence of malaria and to increased international attention to malaria control [2, 3]. Because *P. falciparum* is transmitted through the bite of mosquitoes of the genus *Anopheles*, mosquito vector management has remained a key component of most malaria reduction efforts [4].

Female mosquitoes choose their mammalian hosts based in part on complex chemical cues. Some of these signals, such as carbon dioxide, have been well characterized on a molecular level. For example, carbon dioxide is not only a potent mosquito stimulant, but also augments mosquito feeding behaviors and modulates attraction to other human body odors [5]. However, *Anopheles gambiae* strains that lack functional CO_2 receptors are still capable of locating human hosts [6], indicating that additional chemical signals also drive host preference. Several recent studies have demonstrated that malaria-infected hosts, including humans and rodents, are more attractive to *Anopheles* spp. [7-9]. Analysis of volatile organic compounds emitted by *Plasmodium*-infected mice revealed a global increase in volatiles, including host-derived compounds that enhance mosquito attraction [9].

While female mosquitoes depend on protein-rich blood meals for egg maturation, both male and female mosquitoes are also attracted to and feed from plants. Plant nectar is an important, carbohydrate-rich nutrient source that provides essential energy for flight, and, for some mosquito species, overwintering [10, 11]. This phytoattraction has been successfully harnessed by malaria control efforts through "attractive nectar baiting" strategies, in which mosquitoes are lured to sugar-water blends spiked with plant volatiles and insecticides [12, 13]. Suspected preferred host plants for *Anopheles gambiae* include *Asteracaeae* spp. and *Ricinus communis* [14]. Analysis of purified odorants from these plants has revealed enrichment of volatile compounds known as terpenes, including 10-carbon monoterpenes such as pinene and limonene. At low concentrations, these purified terpenes directly mediate attraction of *Anopheles* spp. [14].

Terpenes are low vapor-pressure hydrocarbons that belong to a class of compounds known as isoprenoids. Over 200,000 isoprenoids have been described, and this large group of biomolecules exhibits dramatic structural and functional diversity [15]. All isoprenoids are produced downstream of two common 5-carbon precursors, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). Animals and fungi generate isoprenoids through a biosynthetic route that proceeds through mevalonate. In contrast, eubacteria and plastid-containing eukaryotes use an alternate metabolic route, the non-mevalonate or methylerythritol phosphate (MEP) pathway. Plants utilize both the mevalonate and MEP pathways; however it is the chloroplast-localized MEP pathway that is used for biosynthesis of the terpene volatiles that constitute their characteristic flavors and fragrances [16]. For many species of insects, not just mosquitoes, chemodetection of plant-derived terpenes directly modulates herbivory and pollination behaviors (reviewed in [17]).

The malaria parasite, *Plasmodium falciparum*, contains an unusual plastid organelle, called the apicoplast, of similar endosymbiotic evolutionary origin to the plant chloroplast [18]. While the apicoplast retains several plant-like metabolic pathways, evidence suggests that the MEP pathway may be the only essential function of this organelle during intraerythrocytic development [19-21]. In this work, we examine the possibility that, like plants, *Plasmodium falciparum* parasites might utilize the MEP pathway to produce terpenes. We determined the volatile chemical composition of headspace gas from cultured *P. falciparum*, and thus identified parasite-produced terpene molecules that represent known mosquito phytoattractants. In addition, we have established the molecular identity of the *Anopheles gambiae* odorant receptors that respond to these plant-like terpenes. Together, our studies provide evidence that malaria

parasites produce specific volatile compounds, and anopheline mosquitoes that transmit malaria contain the cellular machinery necessary for detecting and responding to these compounds. Thus, plant-like terpenes produced by *P. falciparum* may represent semiochemicals for mediating anopheline mammalian host preference.

A.3 Methods

A.3.1 *P. falciparum* culture and strains

All *P. falciparum strains* were cultured *in vitro* in human erythrocytes [22], at 2% hematocrit. The culture conditions were as described previously [23], with the following modifications: a 5% O_2–5% CO_2–90% N_2 atmosphere in RPMI 1640 medium supplemented with 27 mM sodium bicarbonate, 11 mM glucose, 5 mM HEPES, 1 mM sodium pyruvate, 0.37 mM hypoxanthine, 0.01 mM thymidine, 0.25 mg/ml gentamicin (Goldbio), and 0.5% Albumax (Invitrogen). Wildtype strain 3D7 (MRA-102) was obtained from the Malaria Research and Reference Reagent Resource Center (MR4). 3D7-IG was kindly provided by Daniel Goldberg, MD PhD, Washington University School of Medicine.

A.3.2 Headspace sampling

Plasmodium falciparum strain 3D7-MR4 was cultured in a cell bioreactor bag (GE Lifesciences) for 48 hours with a volume of 200mL at 2% hematocrit and 2% parasitemia (infected erythrocytes / total erythrocytes). The culture was injected into the bag via syringe through a liquid injection port in a sterile environment. Uninfected samples contained erythrocytes and media, and blank controls represented sampling from empty bags without media or erythrocytes.

The two injection ports with attached airtight filters were then used to fill the bag with a 5% O_2–5% CO_2–90% N_2 atmosphere. The biobag was secured to a tilting plate and connected to 0.63" sterile plastic tubing (Cole Parmer) through two injection ports. The ends of the tubing were connected to luer pieces, which were secured to the biobag ports using parafilm. Of note, the biobag ports do not contain luer locks, but all other pieces of tubing in the system are connected with interlocking luer pieces. One piece of tubing was connected directly to a Bio-Rad Econo Pump, and the other was fed through an airtight hole in a 250 mL media bottle (Kimax). The bottle also contained openings for fiber insertion and outgoing plastic tubing. This tubing was connected to the other end of the peristaltic pump, completing the closed loop. A carboxen/polydimethylsiloxane SPME microfiber (Sigma Aldrich), inside a manual holder, was placed through an adaptor into the media bottle. Parafilm was used to secure the fiber and fiber-holder in place and provide an airtight seal. Each experiment was performed in a temperature-maintained 37°C room for optimal malarial growth. Sampling was initiated by opening the clamps on the two biobag injection ports, initiating peristalsis, and extending the fiber from inside the holder to its exposed position in the bottle. The fiber was exposed to the sampling conditions for 48 hours. After sampling, the fiber was re-sheathed and analyzed by GC/MS as detailed below.

A.3.3 GC/MS analysis of SPME fiber extracts

Samples were analyzed on an Agilent 7890A gas chromatograph interfaced to an Agilent 5975C mass spectrometer. The GC column used for the study was an Agilent HP-5MS (30 m, 0.25mm i.d., 0.25 μm film thickness. Samples were injected in a splitless mode with injector and transfer line temperatures set at 300°C. A linear temperature gradient was started with an initial

temperature of 60ºC, held for 2 minutes, increased to 300º C at 10ºC /minute, and held for 1

minute. The ion source temperature, electron energy, and emission current were set at 230ºC,

70eV and 300μA, respectively, to obtain EI mass spectra.

A.3.4 Manual analysis of GC/MS data

The raw data were analyzed by the Automated Mass Spectral Deconvolution and Identification

System (AMDIS), which provided an output of the GC trace with deconvoluted mass spectrum

extracted from each trace. Each mass spectrum represents a potential compound at a specific

point in the trace. Every sample (headspace gas and parasite extract) yielded an average of 700

mass spectra per analysis. The structures of the compounds in each GC peaks were identified by

database search (NIST mass spectrum library) using AMDIS software. Background peaks that

represented known biologically irrelevant contaminants, such as polysiloxane arising from

SPME fibers, were excluded from further analysis, as were compounds that did not possess

consistent, parasite-unique ion spectra at given retention times.

A.3.5 Saponin lysis of *P. falciparum* cultures

Parasites were freed from RBCs through lysis with saponin at a final concentration of 0.1%

(vol/vol), followed by centrifugation at 2500 rpm and resuspension in 4 mLs of PBS. Pellets

were washed in an additional 1 mL of PBS, followed by centrifugation at 14,000 for 1 minute.

Dry pellets were stored at -80ºC until analysis.

A.3.6 Organic extraction

Extraction of isolated parasite cells was performed as described in the original Folch procedure [24], with the following modifications. Saponin-lysed parasite pellets were suspended in 1 mL of 2:1 (v/v) chloroform/methanol. The suspensions were sonicated for 30 seconds and then iced for 30 seconds for 3 cycles. Samples were vortexed at 600 rpm for 1 hour after sonication. Samples were then centrifuged at 14000 rpm and the supernatant was recovered. 300 μL of 0.9% NaCl was added to induce phase separation of the sample. The organic phase was recovered for analysis. The organic phase was not evaporated after extraction to avoid loss of volatile compounds.

A.3.7 GC/MS analysis of extracted samples

GC/MS analyses were conducted on a Thermo ISQ 1300 GC/MS system with Xcalibur operation system (San Jose, CA, USA). Separation was achieved by a Thermo 30m TG SQC column (0.25mm i.d., 0.25 u film thickness) at a flow rate of 1ml/min with He as the carrier gas. The GC temperature was started at 50°C for 2 min, raised to 150°C at 10°C/min, and to the final temperature of 300°C at a rate of 20°C/min. The samples were injected in a splitless mode, and the EI mass spectra were acquired in the mass range of 40 to 450 Da at a rate of 0.2sec/scan. The injector, transfer line, and ion source temperatures were set at 240°C, 250°C and 210°C, respectively.

A.3.8 Single-unit electrophysiological recordings

All experiments were performed on adult female flies, 5 days after eclosion. Flies were reared at 25°C in an incubator with a 12-hr light-dark cycle. "Empty neuron" recordings were from flies of

genotype *w Δhalo/Δhalo; Or22a-GAL4/UAS-AgOrX*. The ab3A mutant flies and *Or22a-GAL4* and *UAS-AgOr* transgenic lines were described previously [25]. Fourteen AgOrs (AgOr11, 18, 20, 21, 26, 27, 30, 31, 46, 48, 50, 56, 57, 75), previously found to respond to terpenes [25], were selected to test their responsiveness to additional terpene compounds (α-(+)-pinene, 26870; β-(+)-pinene, 80607; α-(-)-pinene, 305715; β-(-)-pinene, 402753; R-(+)-limonene, 183164; S-(-)-limonene, 218367, Sigma Aldrich). Odorants were diluted in paraffin oil (10^{-2}, v/v) and odor stimuli (50 μl applied to a filter disc) were delivered from a Pasteur pipette via a 500-ms pulse of air (200 ml/min) into the main air stream (2000 ml/min) as described previously [25]. Extracellular single-unit recordings were performed essentially as described [25]. Briefly, electrical activity of the ORNs was recorded extracellularly by placing a sharp electrode filled with Ringer solution into a sensillum and the reference electrode filled with the same Ringer solution was placed in the eye. AC signals (300-2000 Hz) were recorded on an Iso-DAM amplifier (World Precision Instruments) and digitized at 5 kHz with Axoscope 10.2 (Molecular Devices). ORN spike responses were quantified offline and averaged from 6 different neurons. Baseline spike frequency (calculated from spike activity 1 sec prior to odor stimulus) was subtracted.

A.4 Results

A.4.1 Plant-like volatile compounds in *Plasmodium falciparum* headspace gas

We hypothesized that malaria parasites might produce volatile organic compounds, including terpenes. We therefore evaluated the chemical composition of the headspace gas above asynchronous *P. falciparum* parasites, cultured in human red blood cells (RBCs). Because

previous studies of volatile emissions from *P. berghei*-infected mice (average blood volume 2-4mL) [9] or low volume *P. falciparum* cultures [26] did not detect malaria-specific volatiles, we utilized large-volume cultures (200mL) to increase the likelihood of detecting small quantities of *Plasmodium*-produced compounds. In addition, because terpenes are present at low levels in human serum [27], we utilized media supplemented with a lyophilized serum substitute (Albumax, Invitrogen), which does not contain detectable terpenes. For headspace sampling, we employed solid-phase microextraction (SPME) fibers, which selectively bind and concentrate nonpolar organic compounds, as is typically performed to evaluate plant-derived volatiles (reviewed in [28]).

Fibers were exposed to a controlled atmosphere conditioned by *P. falciparum* for forty-eight hours, and then were desorbed and analyzed by electron impact (EI) gas chromatography-mass spectrometry (GC/MS). As is typical of complex volatile samples, component peaks overlap and are not well resolved by visual inspection (representative traces, Fig. S1). For this reason, resulting chromatograms were deconvoluted to isolate overlapping peaks and to extract and annotate component mass spectra. When distinguishing parasite-specific compounds, we aimed to identify compounds qualitatively present in parasite-infected samples compared to controls. Therefore we conservatively selected compounds present in a majority of independent biological replicates of parasite-infected RBC samples and excluded entities also present in either uninfected RBC samples or blank controls that contained neither RBCs nor media. Four compounds specific to parasite-infected samples were thus identified, including two terpenes (Fig. 1). These identified compounds have previously been identified as typical components of plant essential oils and/or fungal volatile profiles [29-32].

A.4.2 Terpenes are present in malaria-infected erythrocytes

We identified several entities that were annotated as terpenes and were present exclusively in the headspace gas of malaria parasites, compared to control uninfected erythrocytes or blank samples. Since terpenes with closely related chemical structures give rise to similar mass spectra, variability in compound annotation is typical and expected. The dominant malaria-specific terpenes were annotated as a 15-carbon sesquiterpene (4,5,9,10-dehydro isolongifolene) and its close derivative (8,9-dehydro-9-formyl cycloisolongifolene) (Fig. 2; Fig. S2). No commercial standards or known synthesis routes are described for either compound; however, the structural annotations are supported by consistent database match factors from 654-774.

In addition, each malaria-infected sample contained at least one 10-carbon monoterpene. Monoterpene annotations varied between samples, but included the structurally related compounds limonene and pinanediol (α-pinene derivative) (Fig. S3). To confirm the identity of these monoterpenes, we extracted nonpolar organic compounds from cultured *P. falciparum* and performed GC/MS analysis. *P. falciparum*-infected cultures, but not uninfected RBC or blank controls, contained a single peak suggestive of a monoterpene, with a retention time of 2.39 minutes, identical to that of an α-pinene (monoterpene) standard (Fig. 3). Comparison of the mass spectra of the observed parasite-specific peak with that of a purified standard established that the parasite-specific compound is α-pinene (Fig. 3E), a terpene compound previously shown to be produced by *Anopheles*-preferred plant species and attractive to *A. gambiae* [14].

A.4.3 Terpenes are produced by *de novo* isoprenoid biosynthesis in malaria parasites

To evaluate whether terpenes in malaria-infected samples were produced *de novo* by the parasite, we utilized fosmidomycin, a phosphonic acid antibiotic that inhibits the first dedicated enzyme of the MEP pathway, deoxyxylulose phosphate reductoisomerase [19]. Previous metabolic profiling of fosmidomycin-treated parasites has established that fosmidomycin reduces concentrations of isoprenoid precursors in *P. falciparum* [23]. Upon fosmidomycin treatment of cultured *P. falciparum*, pinene peak abundance decreased dramatically (Fig. 3D). Proteomic studies of mature RBCs indicate that these cells do not possess the enzymatic machinery to produce isoprenoid precursors required for terpene synthesis [33, 34]. In addition, RBCs do not appear to contain substantial stores of IPP, since malaria parasites that cannot produce IPP themselves are unable to survive [21, 23]. Together this evidence strongly supports that the monoterpenes emitted by *Plasmodium*-infected RBCs arise from the MEP pathway of the malaria parasite.

A.4.4 *Anopheles* odorant receptors respond to malaria-produced terpenes

P. falciparum is transmitted person-to-person through the bite of anopheline mosquitoes. To locate plant and mammalian nutrient sources, *A. gambiae* detect volatile compounds, which signal through ligand-gated voltage channels known as odorant receptors (AgORs) [25]. Electrophysiological and behavioral studies indicate that *A. gambiae* detect and are attracted to plant volatiles. While high concentrations of terpenes often repel mosquitoes, pinene and limonene directly attract *A. gambiae* at low concentrations and are the dominant volatile organic compounds found in the extracts of mosquito-preferred plant species [14].

To determine the biochemical mechanism by which *A. gambiae* detect plant- and malaria-produced terpenes, we assayed a panel of mosquito odorant receptors (AgORs) for pinene and/or limonene ligand-activated electrical activity. Using the *Drosophila* "empty neuron" *in vivo* expression system [35, 36], we found that AgOR75 was dramatically stimulated by (+)-limonene, while AgORs 21 and 50 were substantially stimulated by pinene (Fig. 4, Fig. S3). These odorant receptors are differentially expressed in *Anopheles* chemosensory tissues. Specifically, AgORs 21 and 50 are highly expressed in both male and female antennae [37]. These studies confirm that the primary African malaria vector mosquito can distinguish monoterpenes produced by *P. falciparum*. In addition, our studies also establish the molecular identity of the monoterpene-specific odorant receptors of *A. gambiae*.

A.5 Discussion

Our studies indicate that *Plasmodium falciparum* malaria parasites produce a repertoire of plant-like volatile compounds. These compounds may represent interspecies chemical signals, or semiophores, that modulate the attraction of vector mosquitoes to hosts. Among the parasite-specific compounds we identified, terpenes are bioavailable molecules that readily pass through membranes and partition into alveolar gas in the lung. Terpenes, likely from dietary sources, have previously been identified in exhaled breath samples of humans [38]. Upon malaria infection, parasite-produced terpenes are likely to be detected outside of infected individuals, since the total number of parasites in a typical human infection well exceeds the number sampled in culture in these studies [39, 40].

Previous studies have suggested that *P. falciparum* infection of *Anopheles* spp. mosquitoes may reduce fitness and alter feeding behaviors [41-43]. Over time, selective pressures might enrich for mosquitoes with a decreased tendency to feed from malaria-infected individuals. Therefore, any chemical signals that increase attraction of mosquitoes to infected individuals must be difficult to select against and resistant to evolutionary pressures. This hypothesis is consistent with the finding that malaria infection increases production of typical mammalian host odorants [9]. Our studies suggest an additional strategy by *P. falciparum* for overcoming selection against biting infected hosts, in which the malaria parasite compensates by imitating the volatile components of plants preferred by *Anopheles* spp. The parasite thus hijacks a highly selected signaling response that is necessary for mosquito nectar-feeding behavior and survival. Since *Plasmodium* infection increases nectar attraction in *Anopheles* [44], the parasite appears to facilitate transmission both by generating a mosquito chemoattractant and by sensitizing the mosquito to detect this signal. Interruption of parasite-mediated volatile signaling to mosquitoes will be a potent means of blocking this critical step in the malaria life cycle.

P. falciparum has well-characterized biosynthetic machinery to produce isoprenoid building blocks and prenyl diphosphates [45-47]. In other systems, such as plants, terpenes are produced by terpene synthases, which generate terpenes by catalyzing intramolecular cyclization of prenyl diphosphate substrates [48]. This promiscuous reaction typically produces a variety of chemically related terpene variants from a single enzyme, a cardinal feature of this enzyme class [49]. Consistent with this product diversity, the large protein family of terpene synthases (PFAM 01397) exhibits remarkable sequence diversity. Our studies strongly suggest that terpenes are

produced *de novo* in *P. falciparum*, since chemical inhibition of parasite-specific isoprenoid biosynthesis reduces terpene production. No unambiguous terpene synthase ortholog is present in *P. falciparum* by domain or phylogenetic analyses, but is likely to be represented among the nearly half of parasite genes that remain unannotated. The diversity of terpenes present in *P. falciparum*-conditioned gas suggests that there is at least one monoterpene and one sesquiterpene synthase.

Here we report a repertoire of volatile organic compounds that are specific to *P. falciparum*-infected cultures. These compounds are not likely to represent all possible malaria-specific volatiles, because our conservative data filtering necessarily excludes compounds that are parasite-specific but exhibit significant biological variability. The volatile fingerprint of *P. falciparum* represents not only a target for the development of inhibitors that will interrupt malaria transmission, but also an untapped strategy for malaria diagnostics. The parasite-specific compounds we have identified may represent volatile biomarkers of malaria infection. Ongoing studies will establish the presence and identity of these compounds in human *P. falciparum* infection.

A.6 Figures
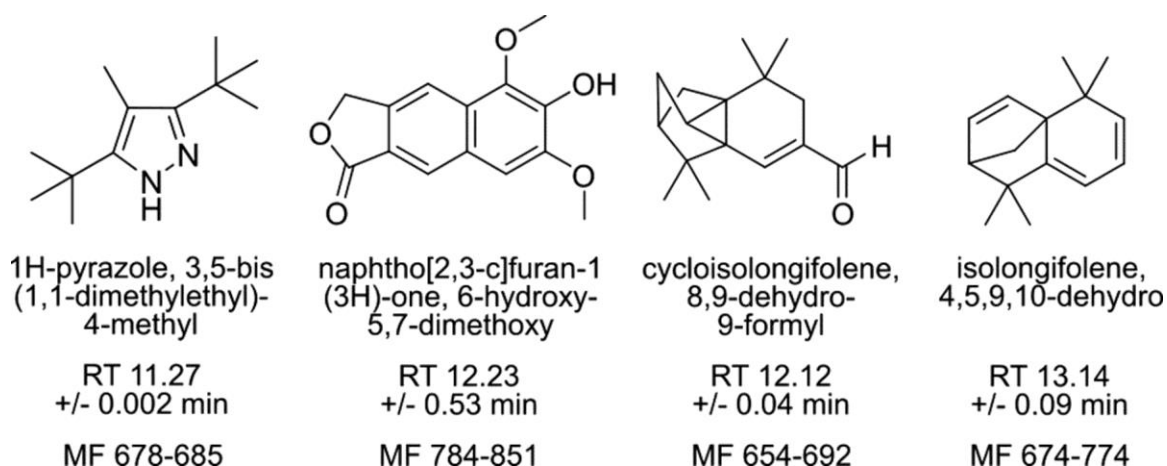
Figure 1. *Plasmodium*-specific VOCs. Compounds annotated in three or more *P. falciparum*-infected SPME sampling replicates (total n=5) and not in uninfected red blood cell samples (n=3) or blank controls (n=6). For each compound, the average retention time (RT) and the range of match factors are indicated. Match factors (MF, 0-999) describe how well sample spectrum agrees with the database spectrum; values >650 indicate close identity.
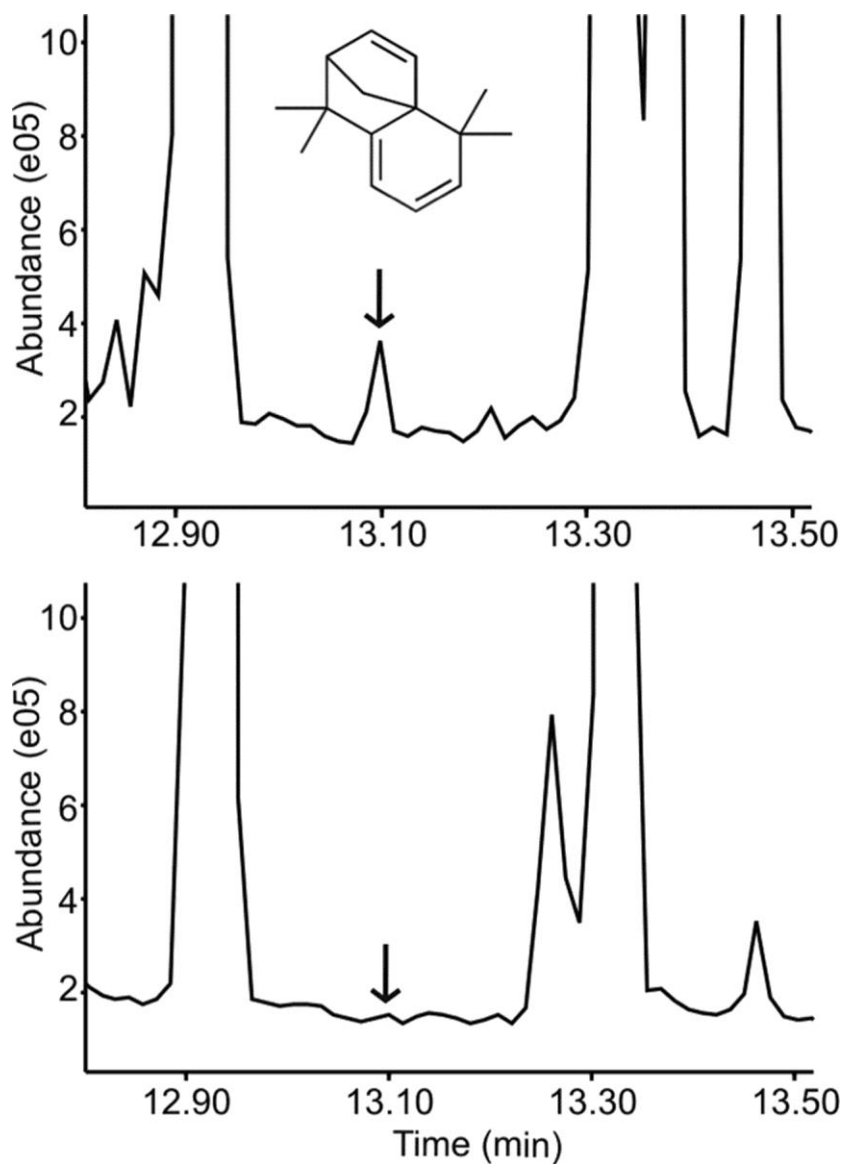
Figure 2. 4,5,9,10-dehydro isolongifolene is present in the headspace gas of *Plasmodium*-infected red blood cells (RBCs). Top: Total ion chromatogram (TIC) of SPME fibers conditioned with headspace gas from *P. falciparum*-infected human RBCs. Arrow, retention time of 13.101 min (typical of 4,5,9,10-dehydro isolongifolene). Bottom: TIC of SPME fibers conditioned with headspace gas from uninfected human RBCs.
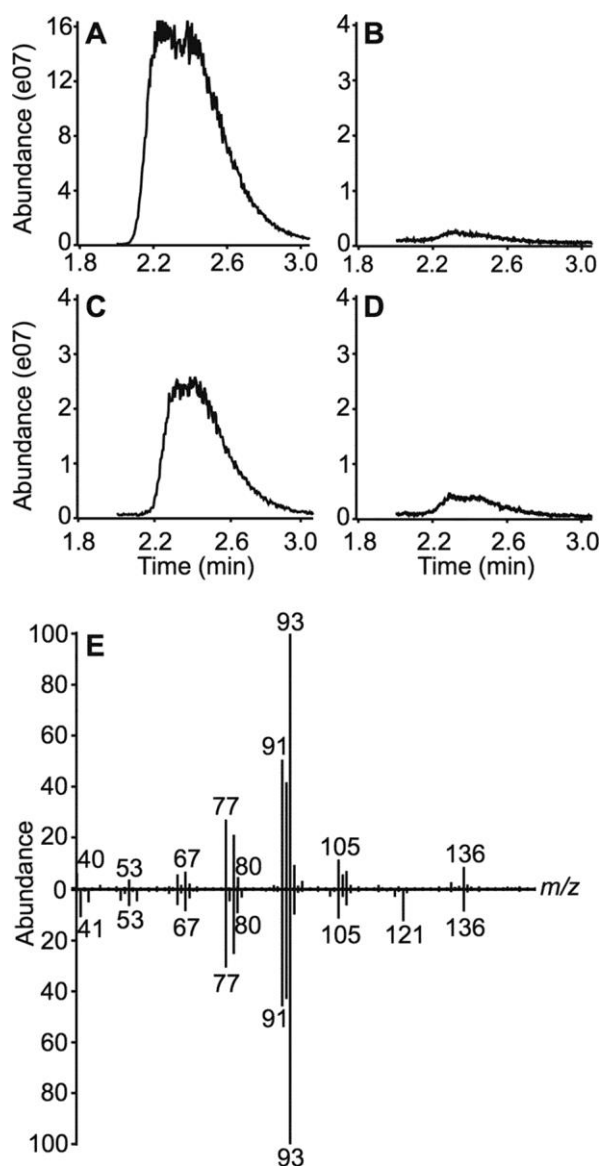
Figure 3. Pinene is produced in malaria parasites by de novo isoprenoid biosynthesis. A-D

Extracted ion chromatograms (EIC) of the total ion chromatograms (TIC) at $m/z = 93$, the base

peak in the ion spectra of pinene. A α-(+)-pinene standard (positive control). B uninfected RBCs

(negative control). C *Plasmodium*-infected RBCs. D *Plasmodium*-infected RBCs treated with 5

μM fosmidomycin (inhibitor of parasite isoprenoid biosynthesis). E Electron impact (EI) mass

spectra of the observed pinene peak from *P. falciparum* in panel C (top) compared to the purified

α-(+)-pinene standard in panel A (bottom).
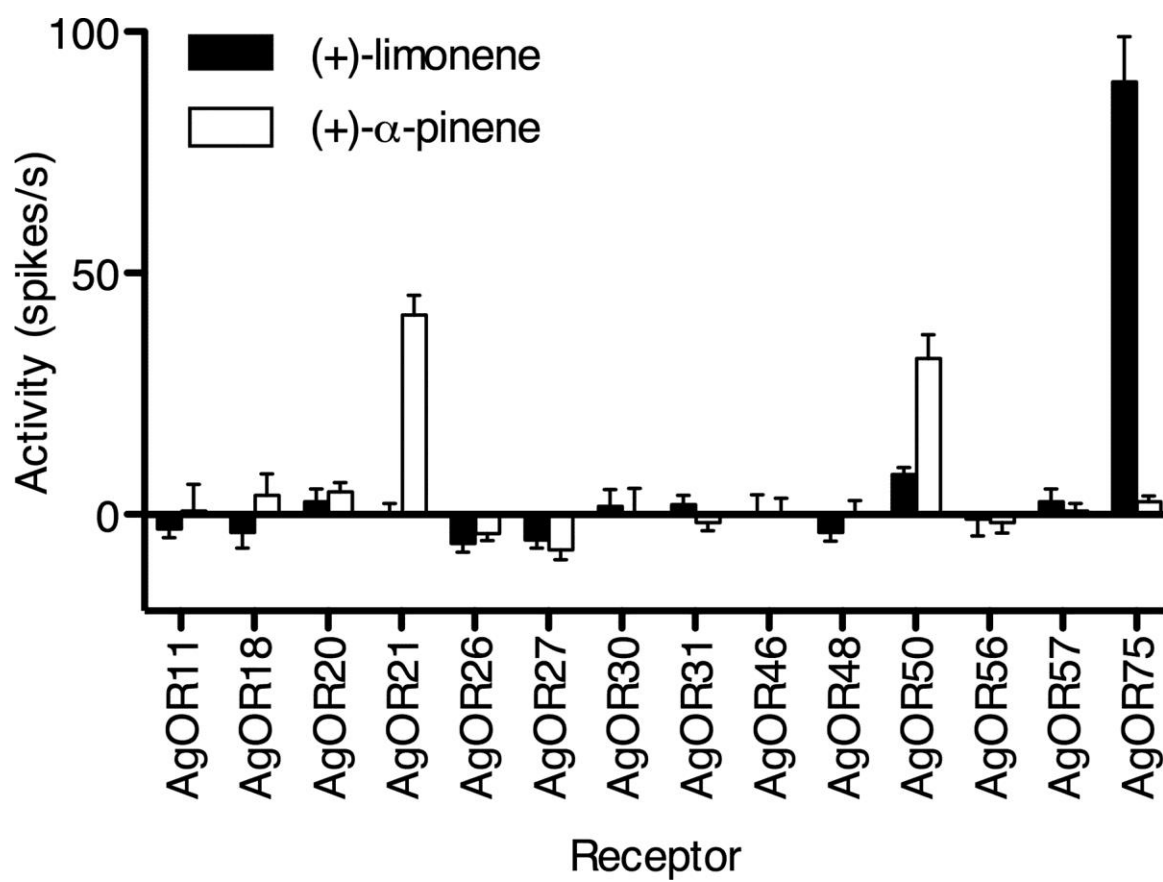
Figure 4. *Anopheles gambiae* odorant receptors respond to malaria terpenes. *Anopheles* odorant receptors were expressed in the *Drosophila* "empty neuron" in vivo expression system and exposed to (+)-limonene and α-(+)-pinene. (+)-limonene specifically activated receptor 75. A (+)-pinene activated receptors 21 and 50 (n=6, error bar: ±S.E.M).
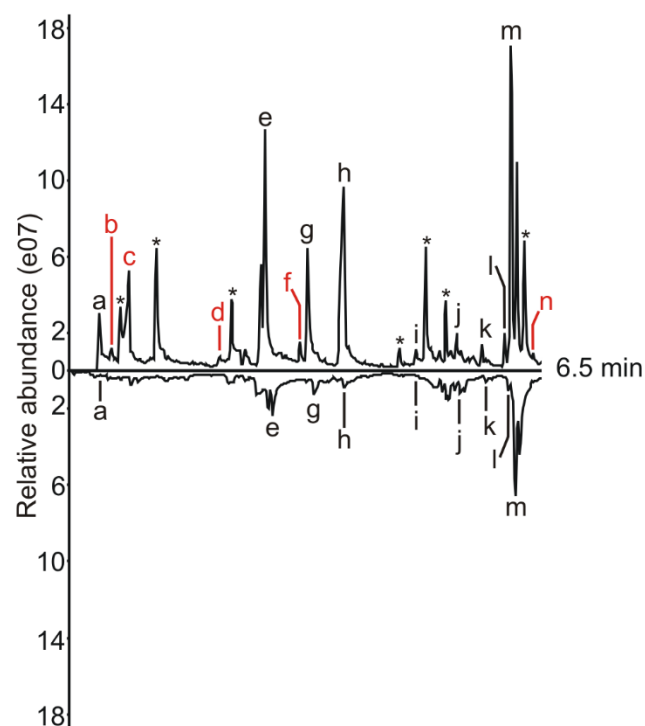
Supplemental Figure 1. Total ion chromatograph (TIC) of *Plasmodium* parasite and red blood cell VOCs. Top: Typical chromatograph of SPME fibers exposed to *Plasmodium falciparum*-infected red blood cells. Marked peaks represent dominant compounds in each trace, many of which represent known contaminant compounds (asterisks, silyated background compounds). Red, parasite-specific peaks from a representative parasite sample compared to uninfected control. Bottom: Typical chromatograph of SPME fibers exposed to uninfected red blood cells. A hydrogen azide, B n-butane, C n-hexane, D toluene, E 2,4-dimethyl heptane, F 2,3-dimethyl heptane, G 4-methyl octane, H methoxy-phenyl-oxime, I benzaldehyde, J octanal, K 2-ethyl 1-hexanol, L 2,3,6,7-tetramethyl octane, M dodecane, N 1,4-dimethyl-, trans-cyclooctane.
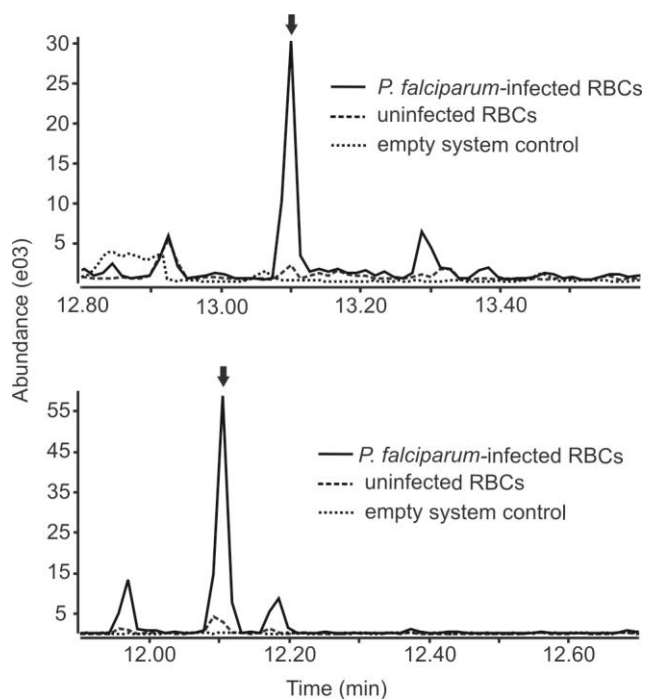
Supplemental Figure 2. Terpenes exclusive to headspace gas of infected red blood cells. Top: Overlay of typical extracted ion chromatograms for the base peak ion of 4,5,9,10-dehydro isolongifolene, m/z = 143. Arrow indicates retention time of 4,5,9,10-dehydro isolongifolene, 13.1 min. Bottom: Overlay of typical extracted ion chromatograms for the parent peak ion of 8,9 dehydro-9-formyl cycloisolongifolene, m/z = 230. Arrow indicates retention time of 8,9 dehydro-9-formyl cycloisolongifolene, 12.1 min.

isolongifolene, 4,5,9,10-dehydro
cycloisolongifolene, 8,9-dehydro-9-formyl
limonen-7-ol, pivalate
(+)-epi-bicyclosesquiphellandrene
(1S,2S,3R,5S)-(+)-pinanediol
10s,11s-himachala-3(12),4-diene
7-epi-α-eudesmol
cadalene
d-limonene
lupa-12,20(29)-dien-3-one
megastigmatrienone
α-cubebene
β-carotene
guaiol
1,3,8-p-menthatriene
camphene
ar-himachalen-2-ol
lilial

compound ID
match probability

0   0.40   0.80

Supplemental Figure 3. Heatmap of terpene annotation predictions. Each column represents compounds identified from SPME sampling replicates from five independent biological samples of *P. falciparum*-infected RBCs. Not represented are compounds also present in RBCs or identified solely from nonpolar organic extractions. The gray scale represents the confidence in the match ID given by NIST, based on the similarity between the ion spectra in the sample and the reference spectra from the NIST library. White represents 0% probability, indicating the spectrum was not found in the TIC, and black represents 80% probability (the highest match of all annotated compounds), indicating the terpene was found in the TIC with an ion spectrum that matched closely to the reference spectrum in the NIST library. Heatmap includes all entities annotated as terpenes; monoterpenes indicated in bold.
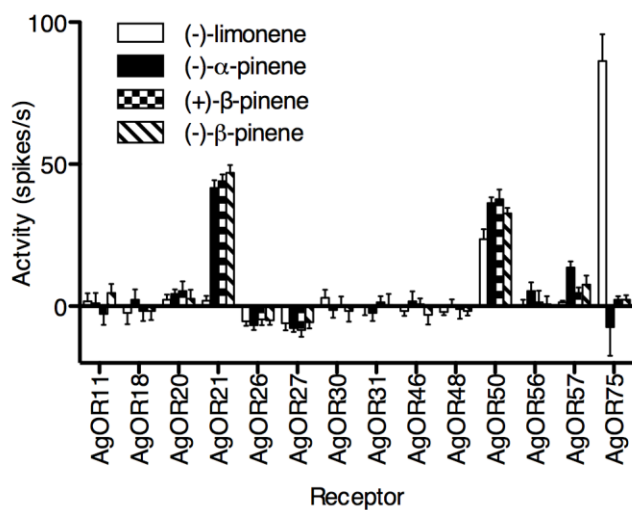
Supplemental Figure 4. *Anopheles* odorant receptor stimulation by parasite monoterpenes. Responses of *Anopheles* receptors in the presence of additional terpene isomers. (-)-limonene had no major differences in its activation profile from (+)-limonene, and α-(-)-pinene, β-(+)-pinene, β-(-)-pinene had no major differences in its activation profile from α-(+)-pinene.

A.7 References

1.   World Health Organization. WHO malaria report 2012. 2012.

2.   Dondorp AM, Fairhurst RM, Slutsker L, MacArthur JR, Guerin PJ, Wellems TE, et al. The threat of artemisinin-resistant malaria. N Engl J Med. 2011;365(12):1073-5.

3.   Fairhurst RM, Nayyar GM, Breman JG, Hallett R, Vennerstrom JL, Duong S, et al. Artemisinin-resistant malaria: research challenges, opportunities, and public health implications. Am J Trop Med Hyg. 2012;87(2):231-41.

4.   World Health Organization. Malaria entomology and vector control: World Health Organization; 2013.

5.   Dekker T, Geier M, Cardé RT. Carbon dioxide instantly sensitizes female yellow fever mosquitoes to human skin odours. J Exp Biol. 2005;208(15):2963-72.

6.   McMeniman CJ, Corfas RA, Matthews BJ, Ritchie SA, Vosshall LB. Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans. Cell. 2014;156(5):1060-71.

7.   Lacroix R, Mukabana WR, Gouagna LC, Koella JC. Malaria infection increases attractiveness of humans to mosquitoes. PLoS Biol. 2005;3(9):e298.

8.   Cornet S, Nicot A, Rivero A, Gandon S. Malaria infection increases bird attractiveness to uninfected mosquitoes. Ecol Lett. 2013;16(3):323-9.

9.   De Moraes CM, Stanczyk NM, Betz HS, Pulido H, Sim DG, Read AF, et al. Malaria-induced changes in host odors enhance mosquito attraction. Proc Natl Acad Sci USA. 2014;111(30):11079-84.

10.   Foster WA. Mosquito sugar feeding and reproductive energetics. Annu Rev Entomol. 1995;40:443-74.

11.   Gu W, Müller G, Schlein Y, Novak RJ, Beier JC. Natural plant sugar sources of *Anopheles* mosquitoes strongly impact malaria transmission potential. Plos One. 2011;6(1):e15996.

12.   Beier JC, Müller GC, Gu W, Arheart KL, Schlein Y. Attractive toxic sugar bait (ATSB) methods decimate populations of *Anopheles* malaria vectors in arid environments regardless of the local availability of favoured sugar-source blossoms. Malar J. 2012;11:31.

13.   Nyasembe VO, Tchouassi DP, Kirwa HK, Foster WA, Teal PEA, Borgemeister C, et al. Development and assessment of plant-based synthetic odor baits for surveillance and control of malaria vectors. Plos One. 2014;9(2).

14. Nyasembe VO, Teal PEA, Mukabana WR, Tumlinson JH, Torto B. Behavioural response of the malaria vector *Anopheles gambiae* to host plant volatiles and synthetic blends. Parasit Vector. 2012;5.

15. Gershenzon J, Dudareva N. The function of terpene natural products in the natural world. Nat Chem Biol. 2007;3(7):408-14.

16. Gutensohn M, Nagegowda DA, Dudareva N. Involvement of compartmentalization in monoterpene and sesquiterpene biosynthesis in plants. Isoprenoid Synthesis in Plants and Microorganisms; 2013. p. 155-69.

17. Hick AJ, Luszniak MC, Pickett JA. Volatile isoprenoids that control insect behaviour and development. Nat Prod Rep. 1999;16(1):39-54.

18. van Dooren GG, Striepen B. The algal past and parasite present of the apicoplast. Ann Rev Micro. 2013;67:271-89.

19. Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, et al. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science. 1999;285(5433):1573-6.

20. Odom AR, Van Voorhis WC. Functional genetic analysis of the *Plasmodium falciparum* deoxyxylulose 5-phosphate reductoisomerase gene. Mol Biochem Parasitol. 2010;170(2):108-11.

21. Yeh E, DeRisi JL. Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage *Plasmodium falciparum*. Plos Biology. 2011;9(8).

22. Trager W, Jensen JB. Human malaria parasites in continuous culture. Science. 1976;193(4254):673-5.

23. Zhang B, Watts KM, Hodge D, Kemp LM, Hunstad DA, Hicks LM, et al. A second target of the antimalarial and antibacterial agent fosmidomycin revealed by cellular metabolic profiling. Biochem. 2011;50(17):3570-7.

24. Folch J, Lees M, Sloane Stanley GH. A simple method for the isolation and purification of total lipides from animal tissues. J Biol Chem. 1957;226(1):497-509.

25. Carey AF, Wang G, Su CY, Zwiebel LJ, Carlson JR. Odorant reception in the malaria mosquito *Anopheles gambiae*. Nature. 2010;464(7285):66-71.

26. Wong R, Flematti GR, Davis T. Investigation of volatile organic biomarkers derived from *Plasmodium falciparum* in vitro. Malaria J. 2012;11:1-8.

27. Ashley DL, Bonin MA, Cardinali FL, McCraw JM, Wooten JV. Blood concentrations of volatile organic compounds in a nonoccupationally exposed US population and in groups with suspected exposure. Clin Chem. 1994;40(7):1401-4.

28.	Yang C, Wang J, Li D. Microextraction techniques for the determination of volatile and semivolatile organic compounds from plants: A review. Analytica Chimica Acta. 2013;799:8-22.

29.	Xie Z, Zhao Z, Wu G, Ye X, Shi L, Guo J. Component analysis of the root exudates at different growth stages in chili pepper. Acta Agriculturae Boreali-Occidentalis Sinica. 2012;8:031.

30.	Chen C-L, Hostettler F. Phenolic constituents of ELM wood: 2-Naphthoic acid derivatives from *Ulmus thomasii*. Tetrahedron. 1969;25(16):3223-9.

31.	Ze-kun L. GC-MS analysis of volatile oils from Bupleurum chinense DC. f. vanheurckii (Muell.-Arg.) Shan et Y. Li. J Med Plants Res. 2012;6(5):926-8.

32.	Dũng NX, Chính TD, Rãng DD, Leclercq PA. Chemical composition of the leaf oil of *Alpinia breviligulata* Gagnep. from Vietnam. J Essential Oil Res. 1994;6(2):181-2.

33.	Alvarez-Llamas G, de la Cuesta F, Barderas MG, Zubiri I, Posada-Ayala M, Vivanco F. Characterization of membrane and cytosolic proteins of erythrocytes. Vascular Proteomics. 2013: 71-80.

34.	Goodman SR, Daescu O, Kakhniashvili DG, Zivanic M. The proteomics and interactomics of human erythrocytes. Exp Biol Med. 2013;238(5):509-18.

35.	Hallem EA, Ho MG, Carlson JR. The molecular basis of odor coding in the *Drosophila* antenna. Cell. 2004;117(7):965-79.

36.	Dobritsa AA, van Naters WG, Warr CG, Steinbrecht RA, Carlson JR. Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. Neuron. 2003;37(5):827-41.

37.	Pitts RJ, Rinker DC, Jones PL, Rokas A, Zwiebel LJ. Transcriptome profiling of chemosensory appendages in the malaria vector *Anopheles gambiae* reveals tissue-and sex-specific signatures of odor coding. BMC Genomics. 2011;12(1):271.

38.	Phillips M, Herrera J, Krishnan S, Zain M, Greenberg J, Cataneo RN. Variation in volatile organic compounds in the breath of normal humans. J Chromatogr B Biomed Sci Appl. 1999;729(1-2):75-88.

39.	Lu JZ, Lee PJ, Waters NC, Prigge ST. Fatty acid synthesis as a target for antimalarial drug discovery. Comb Chem High Throughput Screen. 2005;8(1):15-26.

40.	Adu-Gyasi D, Adams M, Amoako S, Mahama E, Nsoh M, Amenga-Etego S, et al. Estimating malaria parasite density: assumed white blood cell count of 10,000/mu l of blood is appropriate measure in Central Ghana. Malaria J. 2012;11.

41.   Anderson R, Knols B, Koella J. *Plasmodium falciparum* sporozoites increase feeding-associated mortality of their mosquito hosts *Anopheles gambiae* sl. Parasitol. 2000;120(04):329-33.

42.   Koella JC. An evolutionary view of the interactions between anopheline mosquitoes and malaria parasites. Microbes and Infection. 1999;1(4):303-8.

43.   Rossignol P, Ribeiro J, Spielman A. Increased biting rate and reduced fertility in sporozoite-infected mosquitoes. Am J Trop Med Hyg. 1986;35(2):277-9.

44.   Nyasembe VO, Teal PEA, Sawa P, Tumlinson JH, Borgemeister C, Torto B. *Plasmodium falciparum* infection increases *Anopheles gambiae* attraction to nectar sources and sugar uptake. Curr Biol. 2014;24(2):217-21.

45.   Cassera MB, Gozzo FC, Fabio L, Merino EF, Del Portillo HA, Peres VJ, et al. The methylerythritol phosphate pathway is functionally active in all intraerythrocytic stages of *Plasmodium falciparum*. J Biol Chem. 2004;279(50):51749-59.

46.   Jordao FM, Gabriel HB, Alves JMP, Angeli CB, Bifano TD, Breda A, et al. Cloning and characterization of bifunctional enzyme farnesyl diphosphate/geranylgeranyl diphosphate synthase from *Plasmodium falciparum*. Malaria J. 2013;12.

47.   Artz JD, Wernimont AK, Dunford JE, Schapira M, Dong AP, Zhao Y, et al. Molecular characterization of a novel geranylgeranyl pyrophosphate synthase from *Plasmodium* parasites. J Biol Chem. 2011;286(5):3315-22.

48.   Bohlmann J, Meyer-Gauen G, Croteau R. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. P Natl Acad Sci USA. 1998;95(8):4126-33.

49.   Degenhardt J, Kollner TG, Gershenzon J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry. 2009;70(15-16):1621-37.

Appendix B: Malaria infection affects the volatile organic compound profile of sweat

Preface

The following work was performed by myself, Lucy B. Bollinger, Mwawi Mwale, Rachel Mlotha-Mitole, Indi Trehan, and Audrey R. Odom John. LBB, IT, MM, and RM collected the sweat samples and patient metadata. I performed the GC/MS and data analysis. AOJ and I designed the experiments.

For contributions to this work, the authors wish to thank Amalia Berna, Michelle Eckerle, Leah Imlay, Peter Kazembe, Robert Krysiak, Hans-Joerg Lang, Wentai Lou, Mark Manary, Jonathan Ngoma, Brigida Rusconi, Karl Seydel, Terrie Taylor, Malcolm Tobias, the NIH/NIGMS Biomedical Mass Spectrometry Resource at Washington University, and the Center for High Performance Computing at Washington University.

B.1 Methods

B.1.1 Sweat sample collection

Prior to enrollment, approval for the study was obtained from both the Malawi College of Medicine Research Ethics Committee (# P.05/14/1572) and the Institutional Review Board of Washington University School of Medicine (#201504128). A pediatric cohort was recruited from ambulatory pediatric centers in Lilongwe, Malawi, from November 4th to November 16th, 2015. Eligibility, exclusion criteria, enrollment procedures, and malaria diagnosis are the same as discussed in Chapters 2 and 3.

No Nonsense® brand tan nylon leggings (Kayser-Roth Corp., NC) had been cut into "sleeves" roughly 7 in (17.8 cm) long. Pairs of sleeves were placed in individual sealable Ziploc® Double Zipper bags (S.C. Johnson & Sons, WI) and shipped to Malawi. After obtaining consent, the patient had a sleeve put on each arm between the armpit and the elbow. Patients and caretakers were instructed to refrain from touching the sleeves, which were worn for one hour. After the hour had elapsed, the pair of sleeves were removed and placed in a new, labeled Ziploc® bag. The air was pressed out of the bag, which was then sealed and stored at -20°C. In addition to patient samples, room air control samples were collected by placing a pair of sleeves on a clean bench at the collection site for one hour. Sleeves were always handled with clean gloves.

B.1.2 Gas chromatography/mass spectrometry (GC/MS) analysis

Samples were shipped back to Saint Louis, MO, on dry ice and stored at -20°C until analysis, which was performed in two periods 5 weeks and 12 weeks after collection. For each patient sample, the pair of sleeves was placed in a 250 mL Pyrex bottle with a septum cap. All bottles, caps, and septa were baked at 165°C for 2 hours before use to remove possible contaminants. Septa were only used once and then discarded. The headspace above each pair of sleeves was sampled for 24 hours by an 85 µm CAR/PDMS Stableflex 24 Ga SPME fiber (MilliporeSigma, MA). The fibers were baked in the GC/MS inlet at 300°C for 45 min prior to sampling.

After sampling for 24 hours, the fiber was run with an Agilent 7890A GC coupled to a 7200 qTOF MS (Agilent, CA). The GC inlet temperature was set to 300°C, operating in Splitless mode with purge flow split to vent set at 50 mL/min at 0.5 min. The GC had a 30 m length x 0.25 mm ID x 0.25 µm film thickness DB-5 column (Agilent, CA). The GC oven program was hold at 60°C for 2 min, ramp 10°C/min to 300°C, then hold at 300°C for 4 min. The column had a constant flow rate of 1.2 mL/min. The MS was operated in EI mode, source temperature set to 280°C, source emission current set to 35 µA, and electron energy set to 70 eV. The acquisition rate was 5 Hz with an acquisition time of 200 ms/spectrum and mass range of 20 to 750 m/z.

B.1.3 Feature extraction and statistical analysis

Feature extraction was performed using the XCMS Online platform [1]. A Pairwise Job was submitted with the sample data divided on the basis of malaria infection status: malaria positive samples (n = 30) and malaria negative patients (n = 24) acting as the two datasets. The

parameters used are listed below. The Annotation and Identification settings are not relevant as

they are not designed for GC/qTOF-MS data.

- General
    - Ret. time format: minutes
    - Polarity: positive

- Feature Detection
    - Method: centWave
    - ppm: 40
    - min. peak width: 0.18
    - max. peak width: 4
    - mzdiff: 0.01
    - S/N threshold: 6
    - Integration method: 1
    - prefilter peaks: 3
    - prefilter intensity: 100
    - Noise Filter: 0

- Retention Time Correction
    - Method: obiwarp
    - profStep: 1

- Alignment
    - bw: 2
    - minfrac: 0.5
    - mzwid: 0.025
    - minsamp: 1
    - max: 100

- Statistics
    - Stat test: Welch t-test
    - p-value threshold: 0.01
    - FC threshold: 1.5
    - p-value threshold: 0.05
    - value: into
    - Normalization: None

- Misc.

      o   Bypass file sanity check: checked

The output feature list was additionally analyzed in MATLAB (MathWorks, MA) to generate principal component analysis (PCA) plots.

Further statistical analysis was performed using the caret package for R [2], which allows for building and testing predictive models with a variety of underlying algorithms. Random forest (RF), partial least squares (PLS), k-nearest neighbor (KNN), linear discriminant analysis (LDA), and support vector machine (SVM), models were built in quintuplicate using a new set of seeds with each iteration. The input data was restricted to features that had an m/z of less than 250, a retention time of less than 23 minutes, and a p-value of less than 0.2 (Welch's t-test comparing between malaria positive and negative samples). The data was centered and scaled, features with near-zero variation were removed, and missing values were filled using k-nearest neighbor imputation. Models were set to tune based on overall accuracy. Validation was performed using repeated cross validation with 10 folds and 10 repeats. The code was run on the cluster at the Center for High Performance Computing at Washington University. Sample code is presented below:

```
#Loading caret package and setting directory where data is located

library(caret)
setwd("/sample directory")

#Importing the sample data and formatting as a data frame. The .csv file should have a column
#with patient identifiers, a column with "Infection" status, and a column for each extracted
#feature

sample_data <- read.csv("sample_data.csv")
sample_data.df<-data.frame(sample_data)
```

```
#Creating a list to store the models

PLS_models_list <- list()


#Creating a loop to build the model 5 times with a new set of seeds each time

for(i in 1:5){

  set.seed(i*8)

  {my_seeds <- vector(mode = "list", length = 101)
   for(j in 1:100) my_seeds[[j]] <- sample.int(2000, 10)
   my_seeds[[101]] <-50}

#Setting the validation method to be used. Note it is done with a specified set of seeds
#(my_seeds) to allow comparisons between models

  my_control <- trainControl(method = "repeatedcv", number = 10, repeats = 10,
                  classProbs = TRUE,
                  savePredictions = TRUE,
                  seeds = my_seeds)

 #Setting the model to be built along with relevant parameters

  model_PLS<-train(Infection~.,
           metric = "Accuracy",
           tuneLength = 5,
           data = breath.df,
           method = "pls",
           trControl = my_control,
           preProcess = c("center","scale", "nzv","knnImpute"))


  PLS_models_list[[i]] <- model_PLS

}

#Outputting list of all 5 models built.

save(PLS_models_list, file ="PLS_models_list.RData")
```

B.2 Figures
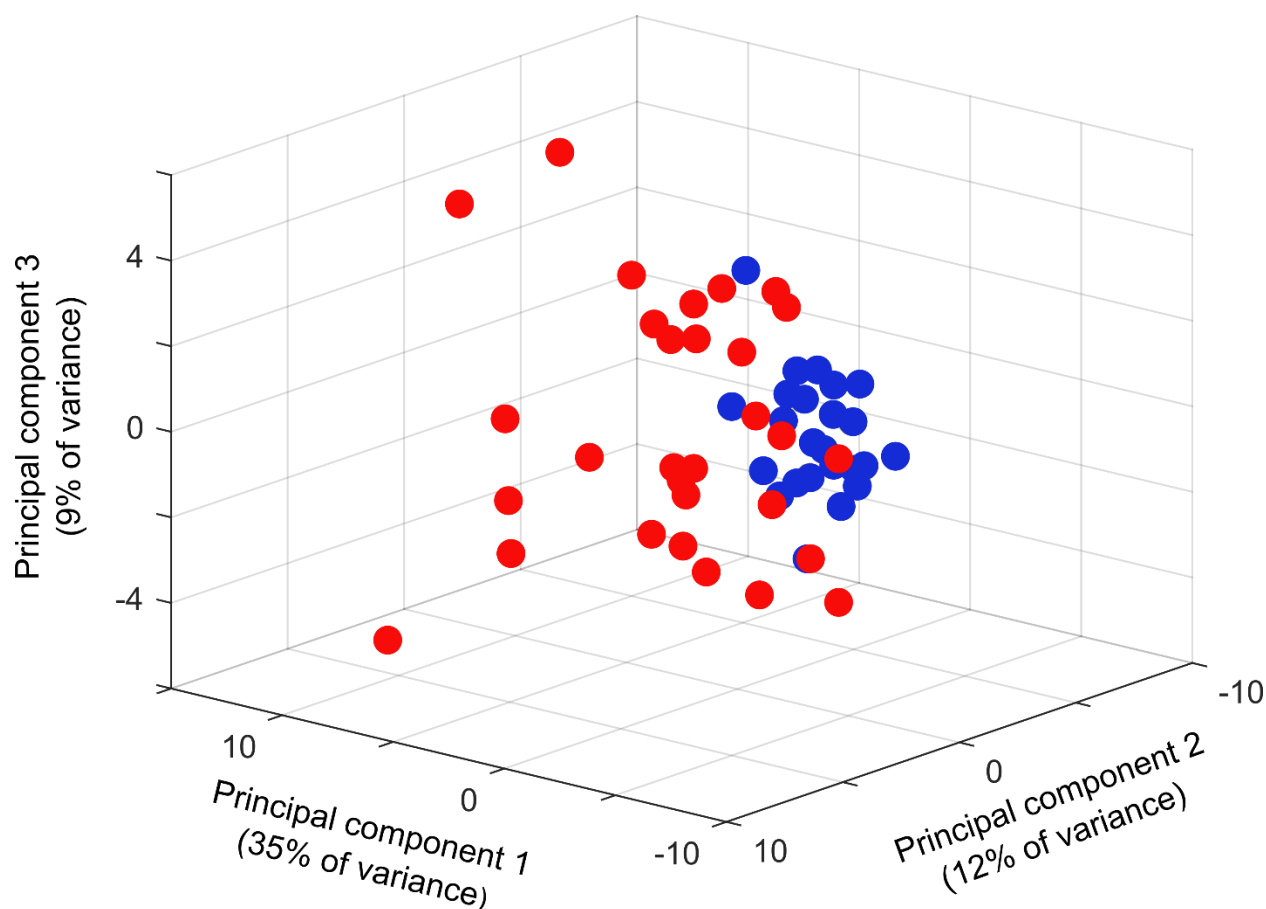
Figure 1 Principal component analysis (PCA) plot of sweat samples shows moderate separation by infection status. Malaria infected patient samples are represented by red dots, and malaria negative samples by blue dots. The spatial position of each dot represents its relative "score" for each principle component, a composite of correlated variables that accounts for a given percentage of the total variability in the sample population. Calculated using all features extracted by XCMS analysis.
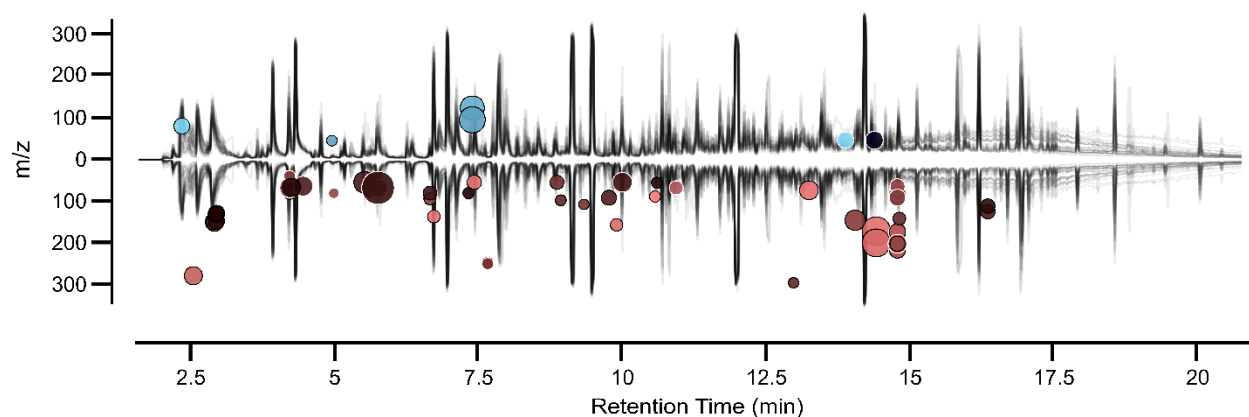
Figure 2 Metabolomic cloud plot visualizing potential features of interest differing based on infection status. Each dot represents a feature (m/z and retention time) extracted by XCMS with a mean abundance higher/lower by at least 1.5-fold and significantly different (Welsh's t-test, p < 0.01) based on infection status. Features higher in infected patients are shown below, in red. Features ≥ 1.5-fold higher in uninfected patients are shown above, in blue. The darker the dot, the greater the degree of significance. The larger the dot, the larger the fold change difference. All total ion chromatograms for the malaria negative patient samples (top) and positive patient samples (mirrored, bottom) are overlaid to demonstrate the overall signal at the retention times were features of interest are located.

B.3 Tables

Table 1 Patient demographic and clinical data for sweat analysis study.

	Malaria Positive (*n* = 30)	Malaria Negative (*n* = 24)	*p* value[1]
Demographics			
Age, median years (IQR)	8.5 (5-10)	6.5 (5-8.8)	0.11
Female, *n* (%)	17 (57)	8 (33)	0.11
Reported Symptoms, *n* (%)			
Fever	28 (93)	20 (83)	0.39
Diarrhea	3 (10)	2 (8)	1
Vomiting	16 (53)	8 (33)	0.18
Headache	23 (77)	13 (54)	0.09
Abdominal Pain	20 (67)	17 (71)	0.78
Muscle/Joint Pain	15 (50)	9 (38)	0.42
Other, *n* (%)			
Chronic Malnutrition[2]	8 (27)	2/23 (9)	0.16
Acute Malnutrition[2]	0 (0)	4/23 (17)	0.03
Uses Bednet	12 (40)	16 (67)	0.06
Malaria within past 3 months	7/28 (25)	6 (25)	1

Data represented as number (%) except for age. If one or more patients were excluded due to gaps in the record, number given is fraction of total. Abbreviation: IQR, interquartile range.

[1] Fisher's exact test or Mann-Whitney U-test used as appropriate to calculate *p* values.

[2] Chronic and acute malnutrition defined respectively as height-for-age Z-score or BMI-for-age Z-score two or more standard deviations below median.

Table 2 Performance of models trained on sweat sample data to classify patient infection status. The input data was restricted to features that met certain criteria (see Methods). Each iteration was executed with a different specified seed value. PLS = partial least squares, LDA = linear discriminant analysis, RF = random forest, KNN = k-nearest neighbor, and SVM = support vector machine.

	Classification Accuracy (%)				
	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
PLS	81	84	82	82	83
LDA	74	73	72	73	75
RF	68	69	69	70	69
KNN	68	69	69	67	68
SVM	77	79	78	78	80

B.4 References

1. Gowda H, Ivanisevic J, Johnson CH, Kurczy ME, Benton HP, Rinehart D, et al. Interactive XCMS online: Simplifying advanced metabolomic data processing and subsequent statistical analyses. Anal Chem. 2014;86: 6931–6939.

2. Kuhn M. caret Package. J Stat Softw. 2008;28: 1–26.

Appendix C: Screen of farnesyl transferase inhibitors against *P. falciparum*

Preface

The following work was performed by myself, Feng Xu (University of Minnesota), Mark

Distefano (University of Minnesota), and Audrey Odom John. Compounds were formulated and

synthesized by FX. I performed the screening against *P. falciparum* and subsequent data

analysis.

C.1 Methods

C.1.1 Drug formulation and synthesis

The farnesyl transferase inhibitor (FTI) drugs were formulated and synthesized by the Distefano lab (University of Minnesota). Compounds were based on two parent structures, A and B, that have known farnesyl transferase inhibitor activity in non-*P. falciparum* organisms. Structures were confirmed with electron spray ionization mass spectrometry, H^1 nuclear magnetic resonance, and analytical high pressure liquid chromatography. The drugs were resuspended in pure dimethyl sulfoxide (DMSO), then shipped on dry ice and stored at -20°C.

C.1.2 *P. falciparum* tissue culture

All culturing was done with *Plasmodium falciparum* genome reference strain 3D7. 3D7 was obtained from the Malaria Research and Reference Reagent Resource Center (strain MRA-102, contributed by D. J. Carucci, ATCC, Manassas, Virginia). Parasites were grown in RPMI-1640 media (Sigma-Aldrich, SKU R4130) supplemented with 27 mM sodium bicarbonate, 11 mM glucose, 5 mM HEPES, 1 mM sodium pyruvate, 0.37 mM hypoxanthine, 0.01 mM thymidine, 10 µg ml^{-1} gentamycin (Sigma-Aldrich) and 0.5% Albumax (Life Technologies) with a 2% suspension of human erythrocytes under an atmosphere of 5% CO_2, 5% O_2, balance N_2 and incubated at 37°C, as previously described [1,2].

C.1.3 Quantifying drug growth inhibition

Growth inhibition assays were performed in opaque bottom 96-well plates with 100 µL supplemented media (see above) at 2% dimethyl sulfoxide (DMSO) and 2% human erythrocytes. The highest tested FTI drug concentration depended on the stock provided, ranging from 40 to 480 µM. Starting with the highest concentration each FTI was serially diluted 1:2 nine times to give a range of drug concentrations at which to test inhibitory activity. A "no drug" control well was tested along with each drug concentration series. Each drug series was done in duplicate on the same plate. Asynchronous *P. falciparum* culture, diluted to a final parasitemia of 0.75%, was added to all wells except no parasite control wells, and was allowed to grow for 72 hours at 37°C under atmosphere. After 72 hours, DNA content was quantified as a proxy for parasite growth using Picogreen (Thermo Fisher Scientific, CA) as previously described [3]. The fluorescence of DNA bound Picogreen dye was measured using a FLUOstar Omega microplate reader (BMG Labtech, Germany) with the excitation wavelength set to 485 nm and the recorded emission wavelength set to 528 nm. The IC_{50} values were calculated using the non-linear regression feature of the software Prism 7 (Graphpad Software, CA). Wells where the red blood cells had visibly lysed were not used in calculations. Three replicates were performed for each drug.
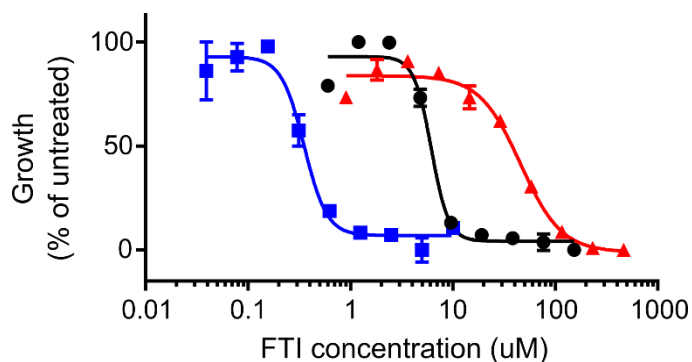
C.2 Figure

Figure 1 Representative malaria parasite growth curves for farnesyl transferase inhibitors (FTIs). The three curves are characteristic of FTIs that have high (FXU2-50, blue), moderate (FXU2-66, black), and low (FXU2-3, red) activity against the malaria parasite. Each point is the mean of two technical replicates testing at a particular concentration. Each curve, fitted by a non-linear regression model, calculates the concentration at which each FTI has 50% of its maximum inhibitory activity (IC_{50}), which corresponds to 0.35 µM for FXU2-50, 6 µM for FXU2-66, and 49 µM for FXU2-3.

C.3 Table

Table 1 Antimalarial activity for screened farnesyl transferase inhibitors (FTIs). IC_{50} = inhibitory concentration at 50% maximal activity. SEM = standard error measurement. Each IC_{50} value is the mean of three replicates (two for those marked with a "*").

FTI Identifier	IC_{50} (μM)	SEM (μM)	Parent Structure
FXU2-50	0.35	0.005	B
FXU2-49	0.45	0.040	B
FXU1-120	1.40	0.200	A
FXU2-54	1.70	0.200	B
FXU1-99	2.20	0.400	A
FXU2-4	2.50	0.200	A
FXU2-11	2.60	0.500	B
FXU1-140	2.70	0.100	A
FXU2-44	3.00	0.100	B
FXU2-90	3.06	0.170	A
FXU2-132	3.12	0.240	A
FXU1-107	3.20	0.400	A
FXU2-122	5.07	0.200	A
FXU2-134	5.21	0.160	A
FXU2-7	5.70	1.000	A
FXU2-66	6.06	0.300	B
FXU2-92	7.10	0.320	A
FXU2-91	7.13	0.230	A
FXU2-45*	7.70	1.400	B
FXU2-96	7.87	0.290	A
FXU2-116	8.41	0.510	A
FXU1-74	9.70	1.200	A
FXU2-6	10.40	1.200	A
FXU2-59	11.04	0.790	B
FXU2-125	11.22	0.960	A
FXU1-139	14.00	1.400	B
FXU1-75	17.00	2.300	A
FXU1-107-CO2Me	26.70	2.700	A
FXU2-1-CO2Me	38.50	0.800	A
FXU2-2	42.47	2.650	A
FXU2-37	48.60	6.500	B
FXU2-3	48.88	2.280	A
FXU2-1*	56.40	3.100	A

C.4 References

1.   Zhang B, Watts KM, Hodge D, Kemp LM, Hunstad DA, Hicks LM, et al. A second target of the antimalarial and antibacterial agent fosmidomycin revealed by cellular metabolic profiling. Biochem. 2011;50: 3570–3577.

2.   Trager W, Jensen JB. Human malaria parasites in continuous culture. Science. 1976;193: 673–675.

3.   Corbett Y, Herrera L, Gonzalez J, Cubilla L, Capson TL, Coley PD, et al. A novel DNA-based microfluorimetric method to evaluate antimalaria drug activity. Am J Trop Med Hyg. 2004;70: 119–124.