Towards ending maternal and infant preventable deaths: omics tools to support vaccine development against *Plasmodium falciparum* malaria and *Streptococcus agalactiae* disease

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Julian Rothen

aus Guggisberg, BE

Basel, 2019

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Marcel Tanner, Prof. Dr. Claudia Daubenberger und PD Dr. Adrian Egli.

Basel, den 13. November 2018

Prof. Dr. Martin Spiess
Dekan

Summary

The United Nations sustainable developmental goal 3 “good health and well-being” includes the aim to significantly reduce global maternal mortality and preventable deaths of newborns and children under 5 years of age until the year 2030. Two major contributors to global maternal and infant morbidity and mortality are *Plasmodium falciparum* severe malaria and Group B *Streptococcus* (GBS) invasive disease. The central aspect in the WHO strategy towards the elimination of these two diseases is the development of effective malaria and GBS vaccines. In the case of malaria, the immunization with radiation-attenuated *P. falciparum* sporozoites (PfSPZ) has been shown to convey protective immunity against controlled human malaria infection (CHMI), making this a promising vaccination approach. However, the molecular mechanisms underlying protective anti-malarial immune responses as well as the reasons for the poor immunogenicity of the PfSPZ vaccine in malaria-experienced individuals compared to malaria-naïve volunteers, remain poorly understood. Emerging system analysis approaches, including genome-wide accession of gene expression using RNA-Sequencing (RNA-Seq) provide valuable insight into post-vaccination systemic molecular dynamics and can help to identify immunological correlates of protection.

In the case of GBS, multivalent glycoconjugate vaccines, targeting selected GBS capsular polysaccharide types, are currently under clinical trial evaluation. With demonstrated good safety and immunogenicity profiles, the licensure of such vaccines is foreseeable. Large-scale monitoring of vaccine recipients for GBS carriage and assessment of vaccine impact on vaginal colonization, potential serotype replacement and emergence of escape strains will be an important aspect of post-licensure epidemiological studies. Matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF MS), has emerged as the method of choice for high-throughput microbial species identification in clinical microbiology and has been suggested for strain-level typing of bacteria.

The overall aims of this thesis therefore included to (i) evaluate the safety and protective efficacy against CHMI of PfSPZ vaccination in Tanzanian volunteers and (ii) elucidate gene expression dynamics in unvaccinated Tanzanian volunteers following CHMI and (iii) to establish a MALDI-TOF MS typing method for GBS for rapid screening of circulating and emerging genotypes.

Building on these objectives, the here presented thesis is structured around five manuscripts:

Manuscript 1: Safety, Immunogenicity, and Protective Efficacy against Controlled Human Malaria Infection of *Plasmodium falciparum* Sporozoites Vaccine in Tanzanian Adults

In this study, we used controlled human malaria infection (CHMI) by direct venous inoculation (DVI) of cryopreserved, infectious *Plasmodium falciparum* (Pf) sporozoites (SPZ) to assess for the first time the safety, immunogenicity and protective efficacy of vaccination by radiation-attenuated PfSPZ in malaria-experienced subjects. In previous studies, immunization with five doses at 0, 4, 8, 12, and 20 weeks of 2.7×10^5 PfSPZ gave 65% vaccine efficacy (VE) at 24 weeks against mosquito bite CHMI in U.S. adults and 52% (time to event) or 29% (proportional) VE over 24 weeks against naturally transmitted Pf in Malian adults. We assessed two vaccine regimens (5 doses of 1.35×10^5 PfSPZ and 5 doses of 2.7×10^5 PfSPZ) in Tanzanians for VE against CHMI. Twenty- to thirty-year-old men were randomized to receive five doses normal saline or PfSPZ vaccine in a double-blind trial. Vaccine efficacy was assessed 3 and 24 weeks later. Adverse events were similar in vaccinees and controls. Antibody responses to Pf circumsporozoite protein were significantly lower than in malaria-naïve Americans, but significantly higher than in Malians. All 18 saline and infectivity controls developed Pf parasitemia after CHMI. In the low dose group, one of 20 (5%) vaccines remained uninfected after 3 week CHMI. In the high dose group, four of 20 (20%) vaccinees remained uninfected after 3 week CHMI and all four (100%) were uninfected after repeat 24 week CHMI. PfSPZ vaccine was safe, well tolerated, and induced durable VE in four subjects in the higher dose group, indicating PfSPZ dose effect. VE testing using homologous CHMI by DVI appeared more stringent over 24 weeks than mosquito bite CHMI in United States or natural exposure in Malian adults, thereby providing a rigorous test of VE in Africa.

Manuscript 2: Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period

We investigated global gene expression changes following CHMI using RNA sequencing (RNA-Seq). Peripheral whole blood samples were collected in Bagamoyo, Tanzania, from ten adults injected intra-dermally (ID) with 2.5×10^4 aseptic, purified, cryopreserved PfSPZ. At 5, 9 and 28 days following CHMI, a total of 2,758 genes were identified as differentially expressed. Transcriptional changes were most pronounced on day 5 after inoculation, during the clinically silent liver phase. A secondary analysis, grouping the volunteers according to

their asexual blood stage prepatent period duration, identified 265 genes whose expression levels were linked both positive and negative to time of parasitemia detection measured by qPCR. Gene modules associated with these 265 genes were linked to regulation of transcription, cell cycle, phosphatidylinositol signaling and erythrocyte development. Our study showed that in malaria pre-exposed volunteers, differences in prepatent period – possibly reflecting the size of the liver to blood inoculum of the parasite – can be linked to changes observed in the peripheral blood transcriptome.

Manuscript 3: Subspecies typing of *Streptococcus agalactiae* based on ribosomal subunit protein mass variation by MALDI-TOF MS

A ribosomal subunit protein (rsp) profiling based on matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS) was developed for fast sub-species level typing of *Streptococcus agalactiae* (Group B *Streptococcus*, GBS), a major cause of neonatal sepsis and meningitis. A total of 796 GBS whole genome sequences (WGS), mirroring the genetic diversity of the global GBS population, were used to identify molecular mass variability of 28 rsp in the molecular weight range 4,425 to 19,293 Da. We identified 62 unique rsp mass combinations, termed "rsp-profiles" which can be distinguished by MALDI-TOF MS. The majority (>80%) of these GBS strains were found to display one of the six defined rsp-profiles 1-6. Importantly, these dominant rsp-profiles classify GBS strains in high concordance with the core-genome based phylogenetic clustering. Validation of our approach by MALDI-TOF MS analysis of 248 in-house GBS isolates showed that the 28 rsp were detected in the mass spectra, allowing fast, robust and reliable assignment of GBS clinical isolates to rsp-profiles at high sensitivity (99%) and specificity (97%). Our approach distinguishes the major phylogenic GBS genotypes, identifies hyper-virulent strains, predicts probable capsular serotype and surface protein variants and distinguishes between GBS genotypes of human and animal origin. In summary, we propose an elegant method combining the advantages of the information depth generated by WGS with the highly cost efficient, rapid and robust MALDI-TOF MS approach facilitating high-throughput, inter-laboratory, large-scale GBS epidemiological and clinical studies.

Manuscript 4: Draft Genome Sequences of Seven *Streptococcus agalactiae* Strains Isolated from *Camelus dromedarius* at the Horn of Africa

In this genome announcement, we present draft whole genome sequences of seven *Streptococcus agalactiae* strains isolated from *Camelus dromedarius* in Kenya and Somalia. These data are an extension to the Group B *Streptococcus* (GBS) pan-genome and might provide more insight into the underlying mechanisms of pathogenicity and antibiotic resistance of camel GBS.

Manuscript 5: Tracing and monitoring of emerging Group B *Streptococcus* genotypes with zoonotic potential in Hong Kong

We validated in this study our novel developed MALDI-TOF MS GBS typing method for analysis of clinical and animal derived GBS collections in Hong Kong. Importantly, we confirm here the inter-laboratory transferability of our rsp-biomarker based MALDI-TOF MS typing method and its potential for rapid and cost efficient screening of hundreds of GBS isolates for the tracing and surveillance of novel, emerging genotypes. We found that 170 GBS strains isolated from adult hospitalized patients and collected from the food markets on pig and fish specimens in Hong Kong can be readily assigned by MALDI-TOF MS into five globally dominant rsp-profiles, allowing reliable prediction regarding their genetic backbone and capsular serotype. Our method is able to discriminate between human GBS genotypes and to identify unique, potentially emerging GBS lineages circulating in fish and pig with zoonotic potential.

Acknowledgements

Over the course of my PhD I had the pleasure and privilege to collaborate with numerous people from a multitude of professional and cultural backgrounds. The successful completion of this project would not have been possible without this combined effort. I would therefore like to express my gratitude to:

My thesis supervisor Claudia Daubenberger, for providing valuable guidance throughout this project and for giving me the opportunity of conducting my research in such a fascinating international environment.

Marcel Tanner, for his much-valued support and for agreeing to be my faculty representative.

Adrian Egli, for acting as co-referee and taking the time to evaluate my PhD thesis and defense.

Stephen Hoffman, the whole Sanaria Inc. team and the IHI clinical trial team in Bagamoyo, without whom the PfSPZ study would not have been possible.

The Clinical Immunology Unit at the Swiss TPH including Tobias Rutishauser, Isabelle Zenklusen, Anneth Tumbo, Catherine Mkindi, Maximillian Mpina, Tobias Schindler, Damien Portevin, Ainhoa Arbues Arribas, Etienne Guirou, Aya Charlene Yoboue, Jean-Pierre Dangy, Yasmine Butt and Nina-Orlova Fink.

My collaborators in Seattle, including Ken Stuart, Jason Carnes, Ying Du and Atashi Anupama at the CIDR and Raphael Gottardo, Carl Murie, Phu Van, Greg Finak and the team at Fred Hutch.

Valentin Pflüger, Guido Vogel, Frédéric Foucault and Roxanne Mouchet at Mabritec AG, Aline Cuenod at the DBM and Joël Pothier at ZHAW for their crucial support in our MALDI-TOF MS projects.

Margaret Ip, Dulmini Nanayakkara, Carmen Li and all the other members at the CUHK Department of Microbiology at the Prince of Wales Hospital in Hong Kong.

My parents and brothers, for their constant support throughout my life and education.

Isabelle, for her joy of life and her every day invaluable support.

Table of Contents

List of Abbreviations

aa	Amino acid
ACT	Artemisinin-based combination therapy
Alp	Alpha-like protein
ANI	Average nucleotide identity
BCG	Bacillus Calmette–Guérin
bp	Base pair
BTM	Blood transcriptome module
CC	Clonal complex
cDNA	Complementary DNA
CFR	Case fatality rate
CHMI	Controlled human malaria infection
CPS	Capsular polysaccharide
CSP	Circumsporozoite protein
CVac	Chemoprophylaxis vaccination
Da	Dalton
DNA	Deoxyribonucleic acid
DT	Diphteria toxin
DVI	Direct venous inoculation
ELISA	Enzyme-linked immunosorbent assay
EOD	Early onset disease
FDA	Food and Drug Administration
GAP	Genetically attenuated parasite
GBS	Group B *Streptococcus*
GMP	Good manufacturing practice
GSEA	Gene set enrichment analysis
HIV	Human immunodeficiency virus
HSPG	Heparan sulfate proteoglycans
IAP	Intrapartum antibiotic prophylaxis
IL	Interleukin
INF	Interferon
iRBC	Infected RBC
Lmb	Laminin-binding protein
LOD	Late onset disease
MALDI	Matrix-assisted laser desorption ionization
MHC	Major histocompatibility complex
MLST	Multi-locus sequence typing
mRNA	messenger RNA
MS	Mass spectrometry
NGS	Next-generation sequencing
NK cell	Natural killer cell
NTS	Non-typhoidal Salmonella
PAMP	Pathogen-associated molecular pattern
PCR	Polymerase chain reaction
PD-1	Programmed cell death protein 1
PfSPZ	*Plasmodium falciparum* sporozoite
PI	Pilus island
ppm	Parts per million
PRR	Pattern recognition receptor
PVM	Parasitophorous vacuole
qPCR	Real-time quantitative PCR
RBC	Red blood cell
RIN	RNA integrity number
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPKM	Reads per kilobase million
rRNA	Ribosomal RNA
rsp	Ribosomal subunit protein
ST	Sequence type
SLV	Single-locus variant
TBV	Transmission-blocking vaccine
TNF	Tumor necrosis factor
TOF	Time-of-flight
TPM	Transcripts per million
tRNA	Transfer RNA
TT	Tetanus toxoid
VE	Vaccine efficacy
WGS	Whole-genome sequence
WHO	World Health Organization

Chapter 1

Introduction

1.1 Malaria

1.1.1 Global disease burden

Despite combined efforts, malaria remains one of deadliest infectious diseases worldwide. Six protozoan parasite species of the *Plasmodium* genus have been described as cause of malaria disease in humans. Historically, no other pathogen has probably had a comparable selective pressure on human evolution than *P. falciparum*, which is the main cause of malaria severe disease and deaths [1]. *P. vivax* is the second most important *Plasmodium* species, which was long underestimated but increasingly acknowledged to contribute significantly to severe malaria [2]. The remaining four *Plasmodium* species *P. ovale curtisi*, *P. ovale wallikeri*, *P. malariae* and the simian parasite *P. knowlesi*, are seldom associated with severe malaria disease [1]. Although malaria incidence rates have decreased 18% globally between 2010-2016, an annual total of 216 million malaria cases and 445,000 malaria-related deaths were still reported in 2016, the majority of which affecting children under the age of five. Most malaria cases (90%) occurred in the World Health Organization (WHO) African Region, with 99% of these cases attributed to *P. falciparum*. Similarly, the vast majority (91%) of annual malaria-related deaths occurred in the WHO African Region [3].

1.1.2 *Plasmodium* spp. life cycle

Transmission of *Plasmodium* parasites to the human host occurs via the bite of a female *Anopheles* mosquito. During the blood meal, sporozoites, the motile, infectious stage of *Plasmodium*, are injected into the human dermis, from where they travel within 30-60 minutes through the skin to the blood vessels and the lymphatic system, and finally to the liver (Fig. 1) [4]. From there, sporozoites traverse the sinusoidal barrier via endothelial cells or Kupffer cells [5] and subsequently invade hepatocytes. A crucial factor for this invasion step is the circumsporozoite protein (CSP), which forms a dense coat on the parasites surface and mediates cell invasion by binding to heparan sulfate proteoglycans (HSPGs) [1]. After infection of the liver cells, the parasitophorous vacuole membrane (PVM) builds up and further transforms (~10 days *P. falciparum*, ~12 days *P. vivax*) until the formed merosome ruptures and up to 40,000 merozoites are released into the blood stream [1,4]. Subsequent encounter with an erythrocyte and invasion marks the starting point of the asexual blood stage

cycle. In the following ~48 hours cell division and development of the parasite from ring stage to trophozoite and schizont occurs, resulting in the release of 16-32 merozoites from each schizont which individually invade new red blood cells (RBCs). At some point during the asexual cell division cycle, a proportion of the parasites undergo a developmental switch, which leads to sexual commitment and the development of parasites via multiple stages into male and female gametocytes [1]. During the maturation of this parasite stage, the gametocytes leave the blood circulation to sequester in the bone barrow, in order to avoid spleenic clearance [6]. Upon re-entry to the bloodstream, gametocytes (stage V) are taken up by a blood feeding female mosquito again and undergo further differentiation into male microgametes and female macrogametes in the mosquito midgut. This is followed by the formation of a zygote, which migrates through the midgut epithelium and forms an oocyst in the basal lamina, in which the sporogonic cycle occurs. With the final release of sporozoites from the oocyst and their migration to the mosquito salivary glands, the parasite is ready to be transferred to the next host during blood meal, thereby closing the life cycle [1].

1.1.3 Malaria clinical symptoms and treatment

First clinical symptoms of a malaria infection occur following the clinically silent liver phase, around 4-8 days after initial invasion of erythrocytes [4]. Depending on the severity of symptoms, malaria disease is commonly classified as asymptomatic, uncomplicated or severe (complicated) [7]. In asymptomatic patients, parasites are circulating in the blood but are not causing any symptoms. In the case of uncomplicated malaria, symptoms are non-specific including fever, headache, nausea or vomiting and there is no occurrence of severe organ dysfunction. Severe malaria, which is predominantly caused by *P. falciparum*, is accompanied with severe anemia, respiratory distress and coma (cerebral malaria) [4]. The characteristic fever waves during severe malaria are caused by the synchronized rupture of thousands of infected RBCs (iRBCs) and subsequent release of merozoites, which triggers an acute and excessive host inflammatory response [1]. Microvascular obstructions due to sequestration of mature parasites in the blood vessels does further contribute to the severity of malaria disease [8]. The WHO recommendations regarding malaria treatment includes diagnostic confirmation of *Plasmodium* infection via thick blood smear microscopy or rapid-diagnostic testing. In the case of uncomplicated malaria, a 3-day treatment with artemisinin-based combination therapies (ACT) is standard. In the case of severe malaria, intravenous or

intramuscular artesunate for at least 24 hours, followed by a standard 3-day ACT regimen is recommended [9].

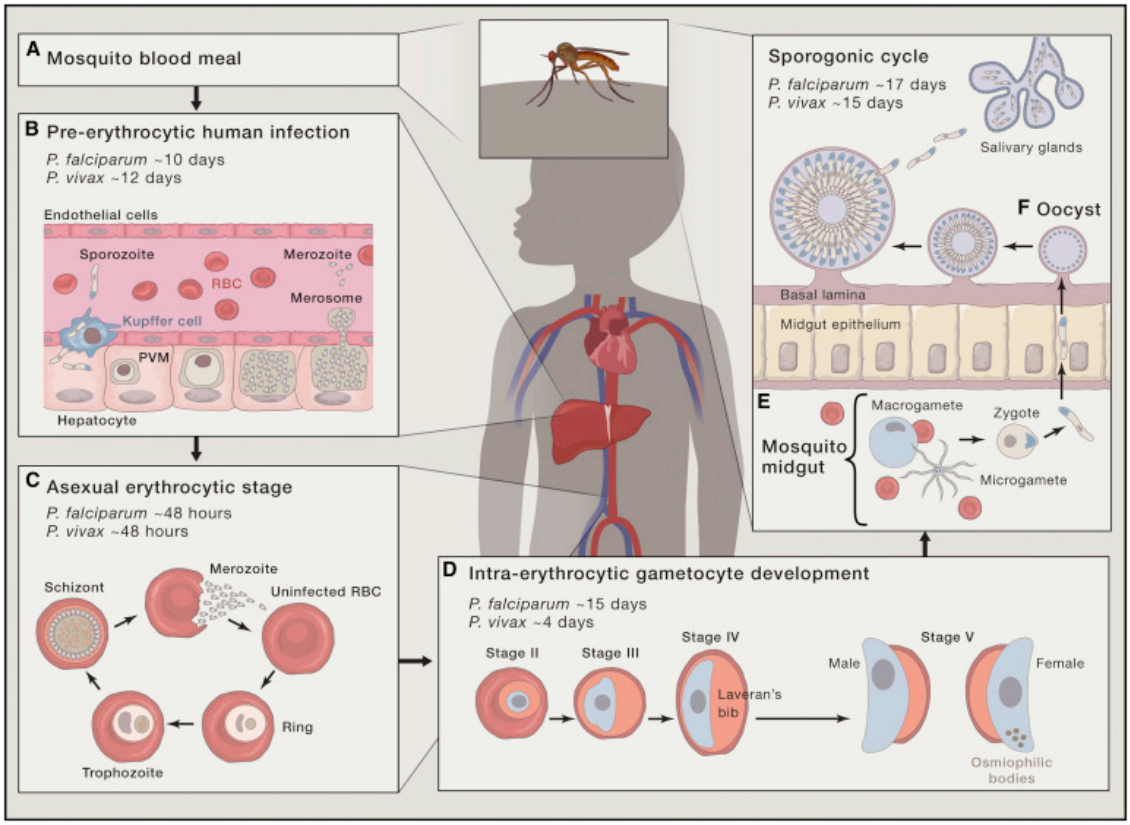


Figure 1. *Plasmodium* spp. life cycle. (A) *Plasmodium* sporozoites are injected into the host dermis during the blood meal of a female *Anopheles* mosquito. (B) During the clinically silent pre-erythrocytic infection stage, parasites migrate in the bloodstream and invade liver cells. (C) Upon release of merozoites into the bloodstream, the parasites invade erythrocytes to start the asexual blood stage. (D) A proportion of parasites undergo sexual development into male and female gametocytes (E). After uptake of the gametocytes by the mosquito, the parasites further transform, ultimately resulting in release of sporozoites to the mosquito salivary gland. Figure from Cowman *et al.* [1].

1.1.4 *P. falciparum* immune evasion and host immune response

1.1.4.1 Parasite evasion strategies and adaption to human host

The *P. falciparum* parasite poses two major challenges to the human immune system. First, its complex life cycle in the human host, during which the parasite mostly resides intracellularly in hepatocytes or RBCs and therefore remains largely protected from direct immune attack [10] The time intervals during which the merozoite stage is present in the blood stream is

extremely short and further limits the time of action for the immune system. Additionally, the duration of the liver stage phase is too short for prevention of parasite development in due time by an effective adaptive immune response [11]. The second challenge is the enormous genetic variability of the *P. falciparum* parasite. Extensive antigenic complexity of polymorphic surface antigens expressed during the different life stages allows the parasite to evade the human immune system (Fig. 2). [12].

In areas of high transmission, natural immunity against severe malaria has been suggested to be acquired after few infections already, while clinical immunity to milder forms of malaria take much longer to develop [13]. Although people living in endemic areas may develop clinical immunity to mild malaria disease over years of repeated exposure, this immunity is not sterile. Such people continue to have asymptomatic, low-level parasite malaria episodes and continue to serve as reservoir for transmission of the parasite [14,15].

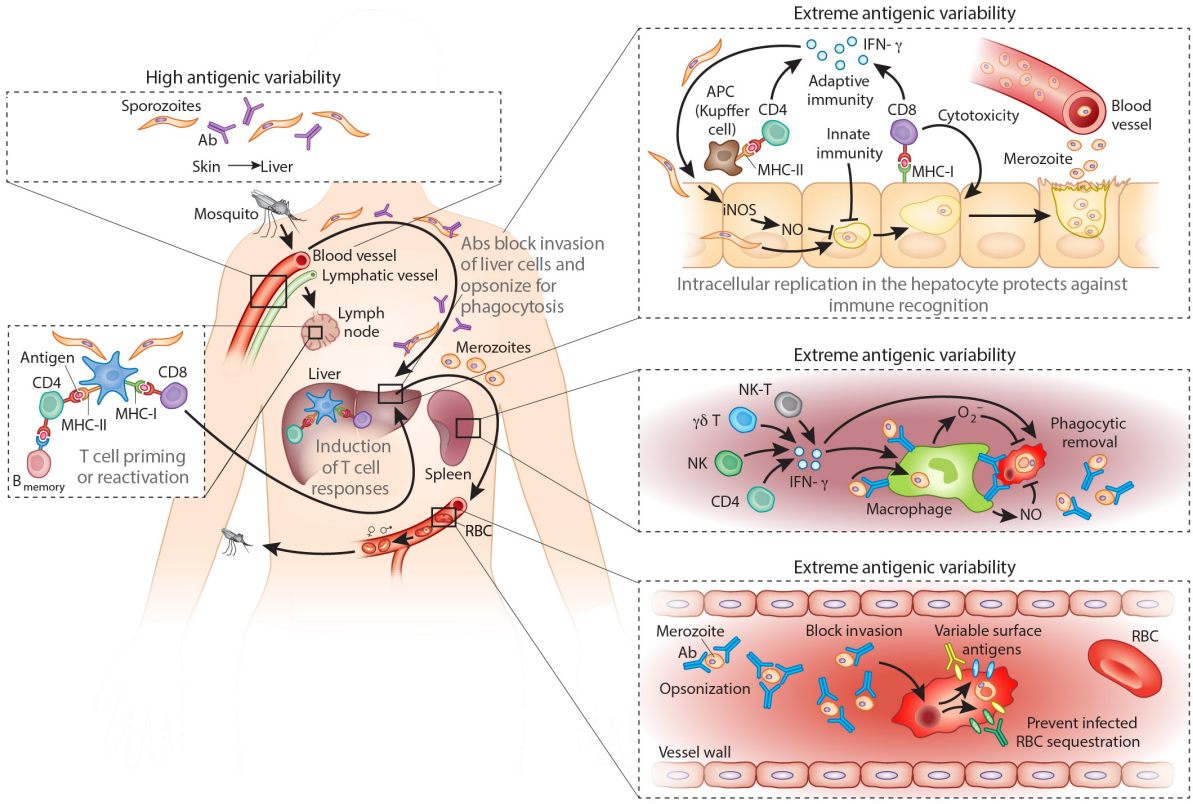


Figure 2. High antigenic variability and intracellular development protects *P. falciparum* from the human immune responses at the various stages of infection. Figure adapted from Riley *et al.* [11].

1.1.4.2 Immune response against pre-erythrocytic stage

The early *P. falciparum* infection stages in skin and liver remain clinically silent, because of a lack of systemic inflammatory responses, reflecting the generally weak innate immune mechanisms against natural *P. falciparum* infection [16]. The immuno-regulatory environment of the human skin, which includes regulatory T cells (T regs) that are thought to provide immune tolerance to sporozoites, allows the parasite to avoid complete clearance in the dermis [17,18]. In the blood stream and lymphatic system, the extent of antibody mediated opsonization and subsequent phagocytosis of sporozoites is not sufficient to prevent progression to liver invasion (Fig. 2). $CD4^+$ and $CD8^+$ T cells, which are primed in skin draining lymph nodes upon encounter with antigen-presenting dendritic cells, can induce interferon (INF)-γ responses in the liver that prevent merozoite development in infected hepatocytes. However, these responses are mostly too weak to prevent transition to blood stage [11]. In striking contrast to lack of pre-erythrocytic immunity in natural infection is the demonstrated induction of sterile immunity in humans that were injected experimentally with high numbers of radiation-attenuated sporozoites [19] or low numbers of fully infectious sporozoites under chloroquine prophylaxis [20]. Such studies demonstrated strong antibody and T cell responses against various liver stage antigens. Pluripotent effector memory T cells producing INF-γ, tumor necrosis factor (TNF)-α and interleukin (IL)-2 were identified as potential correlates of protection [20]. Similarly, high levels of interferon-gamma producing $CD8^+$ T cells targeting infected hepatocytes were found to be crucial for protection [21]. Collectively, such studies continue to contribute to a better understanding of pre-erythrocytic immunological mechanisms during early stage malaria disease.

1.1.4.3 Immune response against asexual stage

The immunological response against the *P. falciparum* blood stage parasites is characterized on the one hand by extensive production of pro-inflammatory cytokines and chemokines including IL-6, IL-10, IL-12 (p70), INF-γ and TNF [22,23]. Besides of $CD4^+$ T cells that are centrally involved in the protective immune response through production of INF-γ and interaction with antibody producing B cells, other T cell subsets including γδ T cells and natural killer T cells (NKT) are acknowledged to be involved in INF-γ production (Fig. 2). *P. falciparum* pathogen-associated molecular pattern molecules (PAMPs), including hemozoin [24] and AT-rich DNA motifs [25], which are recognized by pattern recognition receptors (PRR) on immune cells, are a main trigger for systemic inflammation [16]. While these

extensive inflammatory responses contribute to killing of the parasite, for instance through natural killer (NK) cells induced INF-γ production that promotes destruction of iRBCs by activated macrophages [26], the excess systemic inflammation can also contribute to severe clinical symptoms [27].

The second crucial mechanism to cope with blood-stage malaria is antibody-mediated immunity, through direct opsonization of merozoites, blocking of RBC infection and preventing of iRBC sequestration (Fig. 2). Many questions regarding which of the manifold parasite antigens display targets for protective antibodies remain unanswered [16]. As opposed to other pathogens, antibodies produced against *P. falciparum* are rapidly lost after 3-9 months, providing no long-term protection [28,29]. It has been suggested that *P. falciparum* directly impairs $CD4^+$ T helper cell and B cell function. For instance it was shown that the programmed cell death protein 1 (PD-1), a regulator of T cell exhaustion, is up-regulated on $CD4^+$ T cells in children following *P. falciparum* infection [30].

1.1.5 Vaccine development against malaria

Vector control measures including the large-scale distribution of insecticide-treated bed nets and indoor residual spraying have led to a significant decrease of malaria burden in sub-Saharan Africa [31]. Lately, gene editing technologies like CRISPR-Cas9 promise a new level of vector control by direct manipulation and alteration of the mosquito population via gene drive systems [32,33]. Nevertheless, there is a common consensus that in order to move towards malaria elimination, an effective anti-malaria vaccine is necessary.

There are numerous potential malaria vaccines in the global pipeline (Fig. 3). Either in translational projects (Phase 1a, 2a and 1b) or in more advanced clinical evaluation as vaccine candidates (Phase 2b and 3). The majority of these candidates either target the pre-erythrocytic or the blood-stage of the *Plasmodium* parasite. A third, less abundant class of transmission-blocking vaccine (TBV) candidates aim to prevent development of the malaria parasite during the sporogonic cycle, e.g. after blood meal ingestion of gametocytes by the mosquito and therefore provide no direct protection to the vaccinee [34].

The currently most advanced pre-erythrocytic stage malaria vaccine candidate is RTS,S/AS01, also referred to by its trade name Mosquirix. RTS,S is a recombinant protein vaccine targeting the *P. falciparum* CSP, which is the dominant protein expressed on the parasite surface during the pre-erythrocytic stage. RTS,S co-expresses part of the circumsporozoite sequence together with fused and free hepatitis B surface antigen and

formulated with the AS01 adjuvant as viral like particles. Clinical evaluation in a phase 3 trial showed that vaccination with RTS,S reduced the number of malaria cases by roughly 36% in children between 5-7 months at first time of vaccination over a follow-up period of 48 months. In children aged 6-12 weeks at first time of vaccination, the vaccine efficacy was lower at roughly 26% over a follow-up period of 38 months [35]. Following WHO recommendation, RTS,S will be introduced 2019 in pilot studies in three African countries to evaluate its use as a complementary malaria control tool [36]. A broad range of other recombinant-protein based vaccines, targeting novel surface proteins of the malaria parasite are in developmental pipelines and novel multi-stage or multi-component formulations for increased vaccine efficacy are being investigated (reviewed in Draper *et al.* [37]).

An alternative approach to malaria the subunit-based vaccines is the immunization of human volunteers with live-attenuated *P. falciparum* sporozoites. Early studies in the 1970s demonstrated that inoculation of malaria-naïve patients with radiation-attenuated sporozoites via mosquito bite induces protective immunity against *P. falciparum* malaria [38,39]. A significant step marked the implementation of an in-house pipeline, enabling isolation and cryopreservation of *P. falciparum* sporozoites (PfSPZ), which allows to move away from immunization via mosquito bite to direct inoculation of volunteers [40,41]. This cleared the path not only for a series of clinical PfSPZ vaccine trials but also for the implementation of mosquito-independent controlled human malaria infection (CHMI), during which volunteers are injected with fully infectious PfSPZ. CHMI has become an indispensable tool to investigate anti-malarial drug efficacy [42,43] and importantly to assess protective vaccine efficacy (VE) in malaria vaccine trials without the need to have large field studies with sufficiently high malaria transmission. An early radiation-attenuated PfSPZ vaccine study with subcutaneous immunization of malaria-naïve volunteers was found to be safe but did only induce limited immunogenicity and protection from subsequent homologous CHMI [44] (NCT01001650). It was later found, that intravenous application of PfSPZ is required to induce potent immune responses in humans and to induce protection from malaria in malaria-naïve U.S. volunteers [45] (NCT01441167). A recent study conducted in Mali showed that intravenous PfSPZ immunization leads to partial protection (52% (time to event) or 29% (proportional) VE over 24 weeks) from naturally acquired malaria disease in malaria-experienced individuals [46] (NCT01988636). Immune responses in patients from Mali were found to be weaker compared to immune responses to the same vaccine regiment in malaria-naive volunteers. This finding was supported by a recent clinical trial conducted on Bioko

Island in Equatorial Guinea, where PfSPZ vaccine was found to be safe and well-tolerated, with induced immune responses comparable to what has been observed in the volunteers of the Mali trial [47] (NCT02418962). The outcome of a phase I clinical PfSPZ vaccine trial (NCT02132299) conducted in Bagamoyo, Tanzania that investigated the safety, efficacy and protective efficacy against homologous CHMI, is presented in *Chapter 2*.

Besides the development and refinement of the radiation-attenuated PfSPZ vaccine method, two alternative and promising approaches to weaken the parasite prior to immunization have emerged over the past years. Immunization with fully infectious *P. falciparum* sporozoites and simultaneous treatment with chloroquine as anti-malarial chemoprophylaxis (PfSPZ-CVac), has shown to induce sterile protection in malaria-naïve volunteers (NCT02115516) [48]. A second approach using genetically attenuated parasites (GAP), that had three genes deleted, rendering the parasite unable to successfully develop in the hepatocyte, indicated a good safety and immunogenicity profile in human volunteers [49].

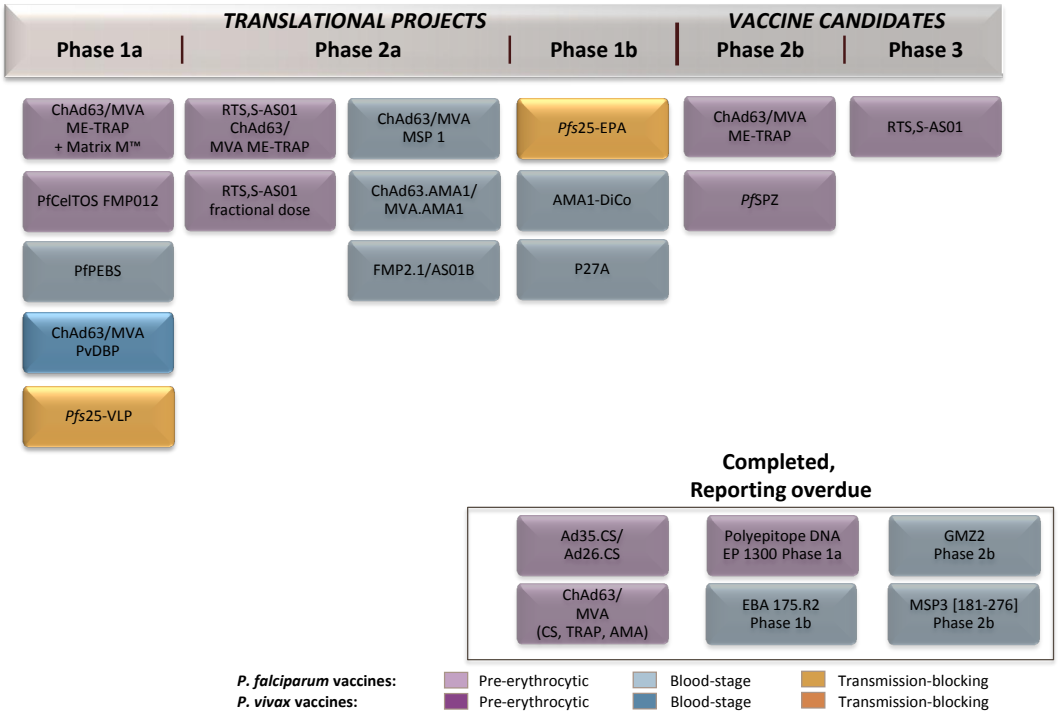


Figure 3. WHO global malaria vaccine pipeline.

1.2 RNA-Seq method for transcriptome analysis

1.2.1 Transcriptomics: definition and history

The transcriptome is defined as the complete set of RNA transcripts present in a single cell or in a population of cells at a specific time point [50]. Unlike the genome, the transcriptome is a highly dynamic system, influenced by both environmental and developmental factors. While most research involving the transcriptome has traditionally focused on the protein encoding messenger RNA (mRNA), the complexity of the transcriptome is becoming increasingly evident as novel non-coding RNA (ncRNA) species are continuously discovered. Thought to be involved in translation, splicing or gene expression regulation at post-transcriptional level, the most prominent ncRNA species include transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA), microRNA (miRNA) and long non-coding RNA (lncRNA) [51]. The key aim of transcriptomics is to index the collective abundance and quantity of RNA species and to observe their dynamics across changing physiological conditions in order to understand development and disease [50,51].

Initial low-throughput gene expression methods like northern blots and qPCR only allowed analysis of single transcripts. With the introduction of high-throughput technologies, genome-wide interrogation of transcript abundances and quantities became feasible [51]. The first of these transcriptomic technologies was the hybridization-based microarray method, where fluorescently labeled complementary DNA (cDNA) are incubated on DNA oligonucleotide chips [52]. Despite of allowing high-throughput gene expression analysis at relatively low costs, microarrays have the major limitations that they are dependent on existing knowledge of the analyzed sequences, display high levels of background noise due to cross-hybridization of highly similar sequences and cannot quantify changes of gene expression levels with a good dynamic range [50]. With the recent advance of high-throughput sequence technology, a new way of investigating the transcriptome has emerged. RNA sequencing (RNA-Seq) is superior to conventional tools and is revolutionizing the way eukaryotic transcriptomes are analyzed [50,51].

1.2.2 RNA-Seq workflow

In the first step of the RNA-Seq workflow (Fig. 4a), total RNA is extracted from the biological sample of choice and the RNA quality, a crucial prerequisite for all following analyses, assessed via the RNA integrity number (RIN). The RNA isolation is followed by the RNA-Seq library preparation, a multi-step process which needs to be customized depending on the research question at hand. Typically, the RNA species of interest are first isolated from the total RNA. A common protocol is the direct depletion of rRNA, which makes up the majority of the total cellular RNA and would therefore suppress the signals from other less abundant RNA species. Most protocols also focus on the enrichment of mRNA with the use of poly-T oligos, which are covalently linked to magnetic beads and target the poly-A tail of mRNA molecules. Another popular approach is the enrichment of miRNA by size selection using gel electrophoresis. Next and universal to all preparation protocols, is the conversion of RNA to cDNA through reverse transcription. After sequencing adaptors are ligated to the ends of the cDNA fragments and subsequent amplification by PCR, the RNA-Seq library is ready for sequencing. Depending on the experiment, further considerations regarding the sequencing steps typically include single-end vs. paired-end sequencing, short read (50-100 bp) vs. long read (>1000 bp) and type of sequencing platform, with Illumina HiSeq representing the most popular NGS technology used for RNA-Seq [51].

1.2.3 RNA-Seq data analysis

During an RNA-Seq experiment, millions of raw sequence reads are generated. This poses significant computational challenges for the downstream analyses, which can be roughly divided into the four steps: quality control, alignment, quantification and differential expression analysis (Fig. 4b). In a first step, the sequence reads are checked for various quality parameters including base call scores, guanine-cytosine (GC) content, overrepresented k-mers, number of duplicated reads, sequencing errors and contaminations [53]. A popular tool for streamlined quality control of raw reads is FastQC [54]. The raw reads that passed quality control are then mapped against a reference genome or transcriptome or subjected to *de novo* assembly. Aligning RNA reads against a reference genome is more complicated than mapping of DNA reads, given that the reads generated in RNA-Seq will align across splice junctions [51]. Two ‘splice-aware’ alignment tools developed for RNA-Seq analysis are TopHat [55] and STAR [56]. The overall percentage of mapped reads is a central quality

measure and is expected to range between 70-90 % [53]. RNA-Seq is most commonly used to determine expression levels of genes or transcripts. As with other parts of the RNA-Seq workflow, there is a wealth of different algorithms available for gene expression quantification. Some applications like HTSeq-count [57] summate the raw counts of reads mapping to given genomic coordinates that are indicative of specific genes or exons. Multi-mapping reads are often discarded in such gene-level quantification approaches. Since read raw counts are directly dependent on the feature's length and the overall sequencing depth, they cannot be compared between samples. Commonly used as expression values in RNA-Seq are therefore the metrics RPKM (reads per kilobase of exon model per million reads) or TPM (transcript per million), which normalize the read counts based on transcript length and library size [53]. There is a range of advanced algorithms used for transcript-level expression quantification. Such algorithms, including Cufflinks [58], RSEM [59] or Sailfish [60], take into account the fact that highly similar transcripts share many of their reads and allocates such multi-mapping reads between transcripts [53]. Differential expression analysis, e.g. the identification of genes and pathways that are collectively up- or down-regulated in response to varying conditions, marks the primary end-point of the RNA-Seq analysis. Detection of such genes or transcripts can be achieved by a diverse set of algorithms. Statistical models including edgeR [61] and DESeq2 [62], that assume negative binomial distribution of RNA-Seq read data have been shown to perform best in detecting differential expression [51].

1.2.4 Application of RNA-Seq in malaria vaccine research

We still remain with many unanswered questions on how exactly protective immunity against malaria is mediated and why malaria-experienced people display overall weaker immune responses compared to malaria-naïve people undergoing the same PfSPZ vaccine regimen [46]. With the entrance into the next-generation sequencing (NGS) era and the rapidly evolving technical possibilities in generating vast amount of biological data, the field of systems biology emerged [63]. As opposed to conventional immunological or molecular assays, that target a specific component of the biological system, the systems biology approach tries to compile a comprehensive picture of multiple molecular parameters and interpret them as one coherent biological network. Several omics-technologies are thereby exploited, including transcriptomics, e.g. the assessment of global gene expression changes using RNA-Seq [63]. In principle, there are two major application of systems biology in vaccinology. Namely, the prediction of vaccine immunogenicity and efficacy, and novel

scientific discoveries pertaining to innate and adaptive immunological mechanisms [63]. A prominent example for the utility of systems vaccinology approach is the yellow fever vaccine, for which early molecular signatures were identified that prospectively predict the vaccine immunogenicity in humans [64,65]. In malaria research, a systems biological approach has been used to investigate the innate and adaptive protective immune response to RTS,S malaria vaccination in humans. The results provided important insights on protective molecular signatures that occur after vaccination, and that can potentially be used in the future as biomarkers of protective efficacy of vaccine-induced immunity against malaria [66].

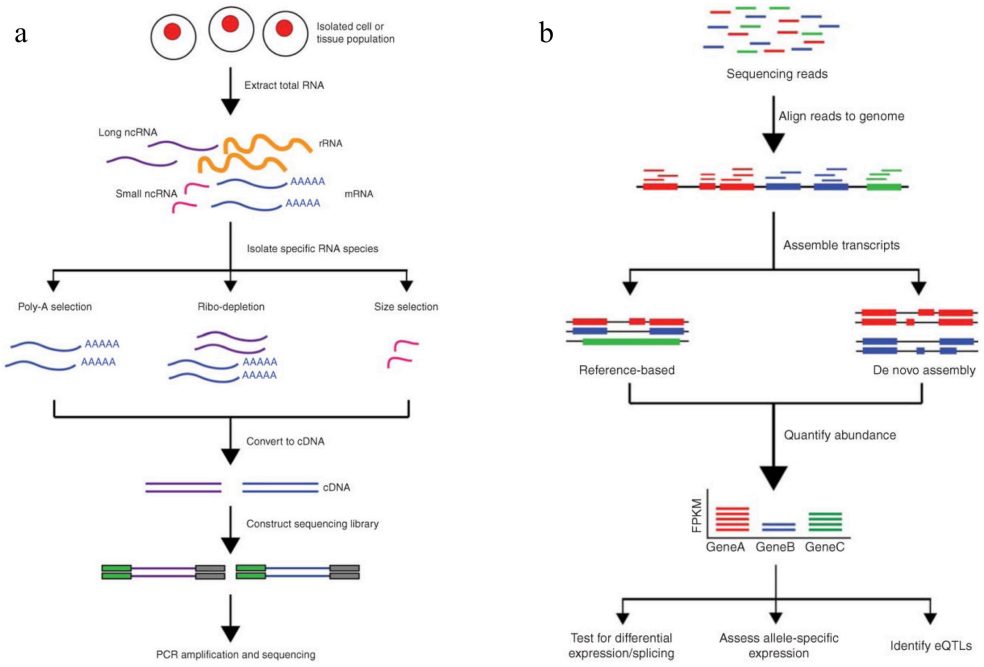


Figure 4. Schematic overview of RNA-Seq workflow from (a) RNA extraction to cDNA conversion and library preparation towards (b) mapping of sequence reads, quantification of transcript abundance and differential expression analysis [51].

1.3 Group B *Streptococcus* / *Streptococcus agalactiae*

1.3.1 GBS disease history

Group B *Streptococcus* (GBS; *Streptococcus agalactiae*) is a beta-hemolytic, gram-positive bacterium, organized in pairs (*diplococci*) or short chains (Fig. 5a) and is a frequent, asymptomatic commensal of the human gastrointestinal and genitourinary tracts [67]. *S. agalactiae* has a broad host spectrum, including mammals, reptiles, amphibians and fish [68]. In 1887, GBS was for the first time identified as a pathogen in animals, causing bovine mastitis [69]. The first reported case of fatal human GBS infection was reported in 1938 [70], but it was not until the 1970s, when GBS emerged in the United States as one of the leading causes of sepsis and meningitis in neonates and infants aged under 3 months [67].

1.3.2 Global GBS disease burden

GBS is worldwide a common colonizer of pregnant women. A recent comprehensive meta-analysis that integrated available data from over 85 countries reported an adjusted estimate for maternal GBS colonization worldwide of 18%. The prevalence rates are subjected to considerable regional variation: From 11% to 13% in Southern and Eastern Asia, to 19.5% in Western Europe (18.7% in Switzerland) to as high as 35% in the Caribbean [71]. Transition from asymptomatic colonization to GBS invasive disease is declared when *S. agalactiae* is isolated from normally sterile body sites, such as blood or cerebrospinal fluid accompanied by clinical symptoms [72].

GBS is mostly recognized as a leading cause of neonatal and infant sepsis and meningitis. Depending on the time frame of first manifestations of symptoms, GBS neonatal and infant disease is categorized as early-onset disease (EOD), which occurs between days 0-6 of life through vertical GBS transmission during delivery or late-onset disease (LOD), which occurs between days 7-89 through vertical or horizontal GBS transmission. Globally, the combined incidence risk for infant GBS disease is 0.49 per 1,000 live births, with case fatality rates (CFR) varying between 4.7% in developed regions and 18.9% in Africa (global average CFR of 8.4%). The estimated incidence risk for EOD is 0.41 per 1,000 live births, with an average CFR of 10% (5% in developed regions, 27% in Africa). EOD manifests itself mainly as sepsis (78%) or meningitis (16%). The incidence risk for LOD is 0.26 per 1,000 live births, with an

average CFR of 7% (4% in developed regions, 12% in Africa). LOD manifests itself mainly as sepsis (53%) and meningitis (43 %) [72]. The overall disease burden of EOD and LOD is further increased by the fact that an estimated 18% of survivors of infant GBS meningitis display moderate to severe neurodevelopmental impairments [73].

While research has traditionally focused on GBS neonatal disease, the broader spectrum of GBS disease is increasingly acknowledged [74]. *S. agalactiae* is a significant contributor to the annual global burden of 2.6 million stillbirths [75]. An estimated 4% of stillbirths in Africa (1.1 million stillbirths/year) and 1% of stillbirths in developed regions are thought to be associated with GBS colonization, suggesting that stillbirth might exceed GBS associated neonatal disease [75]. Also, evidence is accumulating that GBS maternal colonization is associated with preterm delivery [76]. GBS disease occurs also in adults, like in maternal disease during pregnancy or post-partum. Data on how likely GBS colonization is resulting in maternal sepsis remains scarce. An incidence rate of 0.38 per 1,000 pregnancies has been reported in developed countries [77]. Although the risk of maternal mortality or morbidity is low, maternal disease poses a further risk for the neonate [77]. Cases of GBS invasive diseases in non-pregnant adults are increasing with mostly elderly and immunocompromised patients suffering from underlying conditions like cancer, diabetes or HIV found to be more susceptible [78,79]. The clinical manifestations of GBS disease in adults include bacteremia without focus, skin and soft-tissue infection and pneumonia [79]. Adult invasive GBS disease is mostly due to endogenous infection, where GBS switches from an asymptomatic commensal to an invasive pathogen [80]. However, reports of food-borne GBS infection, i.e. through the consumption of contaminated raw fish [81,82] or raw milk [83] highlight the zoonotic potential of GBS.

1.3.3 GBS diagnosis and treatment in pregnant women

Clinical studies in the 1980s documented that intravenous administration of ampicilin or penicillin to women at risk of GBS transmission during labor greatly reduced neonatal early-onset disease [84,85]. Considering the threat of emerging antibiotic resistances and the unknown long-term effect of antibiotic use on the neonatal microbiome, a universal intrapartum antibiotic prophylaxis (IAP) approach for all pregnant women is regarded as unfavorable. Two strategies are therefore commonly employed to ensure a targeted use of IAP. The screen-based IAP approach employs culture-based screening of women late in pregnancy (week 35-37 of gestation) for GBS colonization [86]. The consideration of both

vaginal and rectal swabs for GBS carriage screening is essential as it significantly increases testing sensitivity. Swabs are either inoculated into a selective enrichment broth with subsequent transfer of colonies to selective agar or directly inoculated into selective agar. Phenotypic characteristics including GBS pigment production or the beta-hemolytic properties (Fig. 5b) can be detected using Granada agar and blood agar, respectively. In most cases, additional confirmatory tests including CAMP test, latex agglutination test or MALDI-TOF MS [87] are recommended (reviewed in [88]). While a positive finding is taken as basis for subsequent IAP, culture-negative women are further assessed based on a second, risk-based strategy. The risk-based IAP approach can vary between countries but commonly considers known risk factors including maternal fever, prolonged rupture of membranes, preterm delivery and previous birth to an infant with invasive GBS disease as precondition for antibiotic administration. Both strategies, but specifically the screen-based approach are well realizable in developed countries but are difficult to routinely implement in low and middle-income countries, where the required laboratory structures are often lacking [86]. Nevertheless, the combined use of the IAP strategies have successfully reduced incidence of EOD. It is estimated that the risk of EOD among GBS-colonized women is 1.1% (without IAP), and linearly decreases with increased IAP coverage (0.3% risk if IAP at 80%) [89]. Case prevalence for LOD however, which is possibly acquired postpartum horizontally either as nosocomial infection or through breast milk, remains unaffected by IAP guidelines [90]. A further concern pertains to emerging antimicrobial resistance. Presently, GBS remains susceptible to beta lactams but there have been reports of clinical isolates displaying decreased susceptibility to this group of antibiotics [91,92].

1.3.4 GBS virulence factors

In order to cause invasive disease, GBS has to go through three consecutive steps: (i) The successful colonization of the genitourinary tract, (ii) with subsequent crossing of either placental and epithelial barriers, including the blood-brain barrier in case of meningitis and (iii) evasion of the immune system, mainly by avoiding phagocytic clearance [93]. A range of GBS virulence factors involved in these steps of colonization, adhesion, invasion and evasion have been identified: The main determinant and best-studied GBS virulence factor are the capsular polysaccharides (CPS), each consisting of a combination of four monosaccharides (glucose, galactose, n-acetylglucosamine, sialic acid) and together forming a dense coat on the surface of *S. agalactiae* [94]. The sialic acid component of the GBS capsule is identical to a

sugar epitope found on all mammalian cells, thereby contributing to reduced host cell activation and dampened antibacterial immune response [95]. The main way GBS sialic acid interferes with the human innate immune system is by impairing the complement system, e.g. the deposition of C3 on the bacterial surface. This is accompanied by a decreased production of the complement-derived chemoattractant C5a, which is central for mobilization of neutrophils phagocytosing and intracellular killing. Combined, these mechanisms protect GBS from opsonization, phagocytosis and intracellular killing [94]. To date, ten GBS-specific CPS or serotypes (Ia, Ib, II, III, IV, V, VI, VII, VIII, IX) have been described [96]. These capsular serotypes display varying degrees of virulence, with the hyper-virulent serotype III being the most clinically relevant. It is estimated that the majority of global cases of EOD (47%) and LOD (73%) can be attributed to serotype III, followed by serotypes Ia, Ib and V (22.8%, 8% and 10.6%, respectively, in EOD and 14.2%, 5.3% and 4.0%, respectively, in LOD) [72].

An important feature for GBS cell adhesion are pili, which are cell-wall anchored, filamentous structures extending from the bacterial surface. They are thought to play a central role in epithelial colonization, biofilm formation, translocation and invasion [80]. Two pilus islands (PI) have been described: pilus island 1 (PI-1) and pilus island 2 (PI-2), the latter further divided into two variants PI-2a and PI-2b. PI-1 pili have been shown to be crucial for evasion of macrophage-mediated phagocytosis but show less contribution to GBS epithelial cell adhesion [50]. PI-2a pili play a central role in adherence and biofilm formation [98,99]. PI-2b pili contribute to increased intracellular survival in macrophages [100] and increased strain invasiveness [80]. GBS strains either harbor a variant of PI-2 or a combination of a PI-2 variant and PI-1 for which no variants have been described [101]. The specific pilus combinations thereby seem to be indicative of host specificity and disease presentation: For instance invasive GBS strains generally display a combination of PI-1 and a PI-2 variant, which stands in contrast to maternal colonizing strains [80]. Strains with bovine origin have been shown to carry a PI-2b variant only, which renders them distinct from the human strains [101]. Some other prominent surface proteins that are involved in GBS pathogenesis are the fibrinogen-binding proteins including FbsA [102] and FbsB [103], the laminin-binding protein (Lmb) and the streptococcal fibronectin-binding protein A (SfbA) [104,105], the group B streptococcal C5a peptidase (ScpB) [106], the GBS immunogenic bacterial adhesin (BibA) [107] and the alpha-like protein (Alp) family, including Alpha-C, the prototype alp protein involved in colonization, invasion and translocation, and its other variants Alp1, Alp2, Alp3, Rib and Alp4 [108].

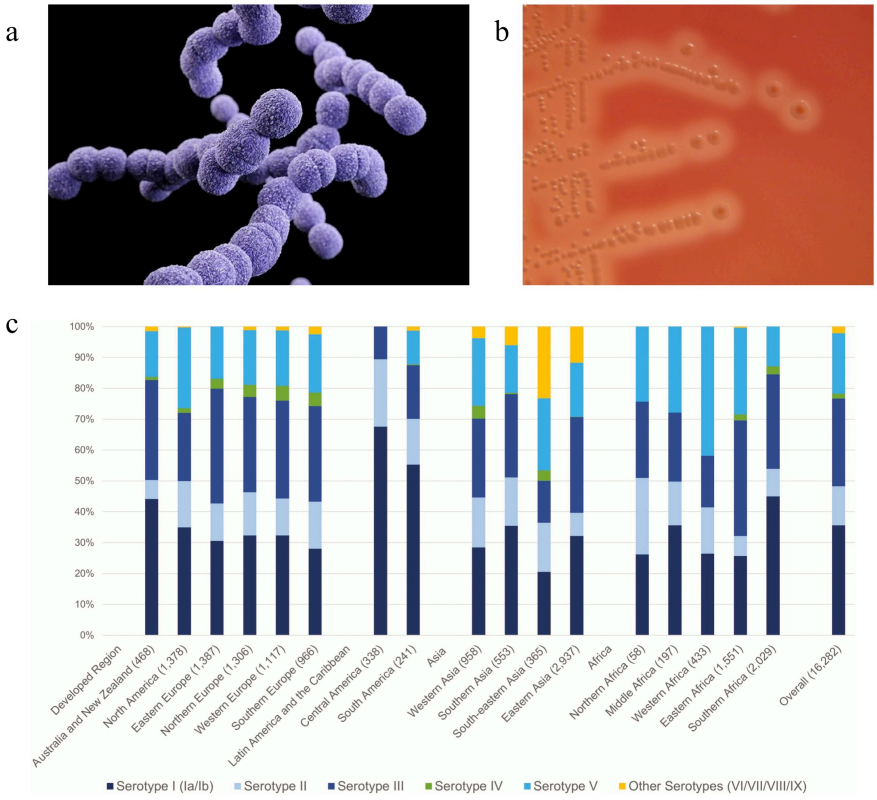


Figure 5: (a) Molecular structure of *S. agalactiae* (source: CDC). (b) Beta-hemolytic characteristics of GBS when grown on blood agar. (c) Global geographical distribution of maternal colonizing GBS serotypes [71].

1.3.5 GBS global epidemiology

GBS strain identification is commonly carried-out based on the ten CPS. The traditional serological method to determine the GBS serotype is latex agglutination [109], but more recently, molecular methods based on PCR [110,111] or *in silico* serotype prediction from GBS whole genome sequences [112] have emerged as viable alternatives.

There are geographical differences in the prevalence of the ten different GBS serotypes. A recent systematic review, incorporating serotype data from more than 16,000 maternal samples provided a comprehensive view on global GBS serotype distribution. Serotypes Ia, Ib, II, III and V can be found as colonizer in women in all global regions, together accounting for 98% of serotypes globally (Fig. 5c). Regional differences in the frequencies of specific serotypes are evident. For instance lower prevalence of serotype III in Central America,

South-eastern Asia and Western Africa and on average higher incidence of serotype V in Western Africa. Interestingly, serotypes VI, VII, VIII and IX, which are uncommon on a global scale, were found to be more frequent in Southern, South-eastern and Eastern Asia [71]. Capsular serotype switching as a result of DNA exchange through homologous recombination has been reported for GBS, and is thought to be a key contributor to clonal diversification and emergence of novel serotypes [113–115].

The current method of choice for epidemiological studies is based on multi-locus sequence typing (MLST) [116], which has the advantage that it can be easily shared and compared across different laboratories. MLST indexes nucleotide variation in the seven GBS housekeeping genes alcohol dehydrogenase (adhP), phenylalanyl tRNA synthetase (pheS), amino acid transporter (atr), glutamine synthetase (glnA), serine dehydratase (sdhA), glucose kinase (glcK), and transketolase (tkt) and subsequently assigns an isolate to a corresponding sequence type (ST). The ST can be further clustered into clonal complexes (CC), each consisting of a single founder ST and its descending single-locus variants (SLV) [116]. Although there are presently over 1,300 different ST described in the global repository for GBS isolates and sequence types (PubMLST) [117], the vast majority of circulating GBS isolates can be attributed to one of the five major lineages CC1, CC10, CC17, CC19 and CC23 [116]. The grouping of GBS strains into these lineages was found to be indicative of their capsular serotype as well as their pathogenicity potential. For instance CC17, which was found to harbor hypervirulent clones [116,118] and is strongly associated with the emergence of GBS neonatal disease [119]. Subsequent studies reported the occurrence of CC67, a further dominant lineage, consisting of GBS strains with obligate bovine origin [119]. The reason for the global occurrence of conserved lineages has been a subject of intensive research. On the basis of MLST, it was first speculated that the hypervirulent CC17 clone emerged from a bovine ancestor [120]. More recent studies that employed comparative genomics analyses indicated that GBS genomes are shaped by transfers of large DNA segments [121] and that such large recombinatorial events are the driving force for the evolutionary emergence of the dominant lineages [119]. It was proposed that GBS consists of a genetically highly diverse core population, that displays no clear serotype-genotype correlations and possesses an almost infinite gene pool, according to the concept of the bacterial pan-genome [122]. From this core population, only few clones would from time to time successfully spread and form dominant, globally-established lineages [119].

An important work by Da Cunha *et al*. published in 2014, further expanded on the latter studies and proposed that extensive use of the broad-spectrum antibiotic tetracycline from 1948 onwards led to the selection and subsequent emergence of few tetracycline resistant clones. This evolutionary bottleneck driven by tetracycline usage resulted in the disappearance of human GBS population diversity and the emergence of few dominant lineages [123]. Among these, CC17 is thought to have emerged most recently since it was found to be genetically most homogenous with relatively lower recombination rates compared to the other CC and almost exclusively displays capsular serotype III [123]. The GBS strains circulating in animal hosts have been shown to display significantly higher degrees of genetic variability compared to strains belonging to the dominant human lineages. For instance the increased genetic diversity within the bovine cluster CC67 [119], the genetically distinct strains isolated from fish [124], or the novel GBS ST circulating in camels [125], all together suggest that animals remain an under-researched reservoir of GBS strain diversity with zoonotic potential.

1.3.6 Vaccine Development against GBS

The main mode of protection against GBS invasive disease is thought to be based on opsonization and subsequent phagocytosis of the bacterium. The opsonization step requires the deposition of complement components like C3b with or without specific antibody binding to the bacterial surface [86]. Neonates have an immature immune system and therefore an impaired ability to produce antibodies, especially against polysaccharide antigens [126]. Newborns therefore rely entirely on maternal antibody transfer for protection from early GBS infection. Hence, the goal of GBS vaccination is to achieve high transplacental transfer of CPS specific antibodies from the immunized mother, leading to protection of the newborn during the early phase of life [86].

Already in the 1930s experiments by Lancefield in animal models indicated CPS-specific antibody mediated protection from GBS infection [127,128]. In 1976 Baker *et al*. demonstrated that low levels of maternal anti CPS type III antibody levels correlated with increased neonatal susceptibility to GBS EOD and LOD and that transplacental transfer of IgG immunoglobulins conferred protection to newborns [129]. The first generation of GBS vaccines consequently consisted of purified type-III CPS. Although these vaccines were safe and well tolerated in clinical trial assessment, they displayed only limited immunogenicity [130]. In addition, these unconjugated polysaccharide vaccines failed to induce T cell

dependent B cell activation and B cell memory response [67]. The GBS vaccine development then slowed down, mainly because of the high effectiveness of IAP treatment in reducing cases of EOD and due to concerns regarding the acceptance and liability coverages for maternal immunization [86].

Following the success of glycoconjugate vaccines against *Neisseria meningitidis* and *Streptococcus pneumoniae* [131], the second generation of GBS vaccines constituted of CPS antigens that were linked to highly immunogenic proteins. This conjugation elicits a long-lasting adaptive immune response against the polysaccharide, inducing B and T cell memory, B cell proliferation and antibody class switching [132]. Initially starting from a GBS type III CPS / tetanus-toxoid (TT) glycoconjugate [133], the design of the vaccines have been further adapted in order to achieve broader coverage against more GBS serotypes. The currently most advanced glycoconjugate vaccine candidate is a trivalent vaccine (serotypes Ia, Ib and III). Unlike its precursors, this vaccine is conjugated to CRM_{197}, a non-toxic mutant of diphteria toxin (DT), which is also used as carrier in already licensed glycoconjugate vaccines against *N. meningitidis* and *S. pneumoniae* [134]. The trivalent GBS polysaccharide-CRM vaccine has already undergone clinical evaluation in a phase I trial in non-pregnant women [135], a phase II trial in pregnant-women [136] and a phase II study in pregnant, HIV-positive women [137]. Results from the first two trials attest the trivalent CPS-CRM vaccine a good safety profile and immunogenicity against all serotypes and successful transfer of antibodies to the newborns. In the HIV-positive cohort, the vaccine was found to be safe, but less immunogenic which could have implications regarding the protective efficacy of the vaccine [137]. However, current GBS glycoconjugate vaccines have limitations in that they provide serotype-specific immunity only, fail to cover non-serotypeable strains and are vulnerable against capsular switching and replacement. A third generation of broad-coverage protein-based GBS vaccines is attempting to overcome this hurdle. Through application of whole-genome bioinformatics analyses of the rapidly increasing number of publicly available GBS whole genome sequences (WGS), a range of novel protein vaccine candidates have been identified in the last decades. GlaxoSmithKline (GSK) are investigating structural components of the pili proteins as vaccine target and the company Minervax (www.minervax.com) are conducting clinical evaluation of a vaccine targeting the alpha-like protein family, reported to cover close to 100% of the GBS population [67]. Phase I trial results indicate good safety profile and high immunogenicity of the vaccine, with induction of IgG and IgA antibody production. The latter is of special interest given the possible IgA antibody transfer via milk could protect newborns long after birth [55].

An increased attention for the need of GBS vaccine development by governments and health authorities can be observed in recent times. The WHO has released a technology roadmap in 2017 listing priority activities to achieve the strategic goal of development and licensure of a safe, effective and affordable GBS vaccine for use in pregnant mothers in high, middle and low income countries [139]. Looking ahead, an introduction of a licensed GBS vaccine will require large-scale monitoring of vaccine recipients for GBS carriage and assessment of vaccine impact on vaginal colonization. Building on our experience from glycoconjugate vaccine introduction against *S. pneumoni*ae [140,141], potential serotype replacement and emergence of GBS escape strains will need to be surveyed closely [142].

1.4 MALDI-TOF MS in clinical microbiology

1.4.1 Microbial species identification in clinical routine

The clinical characterization and identification of microbial species has undergone fundamental shifts during the past decades, moving from time-consuming phenotyping tools to molecular methods and lately to high-throughput proteomic typing systems [143]. Conventional methods initially classified microorganisms based on their phenotypic properties, either based on morphological features like gram- staining that could be assessed via microscopy, or based on biochemical methods, with microbes grown on selective media [144]. Subsequent immunological, antibody-based assays like ELISA or agglutination assays only partially achieved to overcome limitations of preceding methods [143]. A significant improvement in both speed and accuracy of microbial characterization was achieved with the introduction of a variety of molecular methods, including real-time PCR [145] and fluorescence *in situ* hybridization (FISH) [146]. The gold standard for highly discriminatory molecular microbe identification is based on 16S rRNA (for bacteria) or 18S rRNA (for fungi) gene amplification and sequencing [147]. Although a powerful diagnostic tool, this method is expensive, demanding in both infrastructure and technical knowledge and therefore of limited use in clinical routine [143]. In parallel to the nucleotide sequencing methods, matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS), a method initially employed in the field of chemical sciences [143], evolved into a tool for characterization of the microbial proteome. With its demonstrated high accuracy and propensity for inexpensive and high-throughput microbial sample identification, MALDI-TOF MS has emerged as a now widely accepted method in clinical routine microbiology and epidemiology [143,144].

1.4.2 Principle of MALDI-TOF MS technology

In order to investigate the proteomic makeup of a microbial cell using MALDI-TOF MS, a so-called soft-ionization technique is employed, allowing the analysis of high molecular weight molecules. A microbial sample is, together with a low-mass organic solution (the matrix) applied to a steel target plate (Fig. 6a). For microbiological applications of MALDI-TOF MS, three matrices are commonly used including α-cyano-4-hydroxycinnamic acid

(CHCA), 2,5-dihydroxy benzoic acid (DHB), and 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) [143]. The co-crystallized matrix-analyte mixture is subjected to a series of short UV laser beams, leading to the direct release of molecules into the gas phase and the formation of both sample and matrix ions. These ionized proteins are accelerated in a TOF mass analyzer, racing through a linear flight tube at the end of which they collide with a detector. The thereby recorded time-of-flight depends on the proteins mass-to-charge ratio (m/z) and is subsequently used to determine the protein molecular weight. The combined TOF information of thousands of measured proteins is summarized in a mass spectrum, representing a unique protein mass fingerprint of the investigated microbe (Fig. 6b) [143,144].

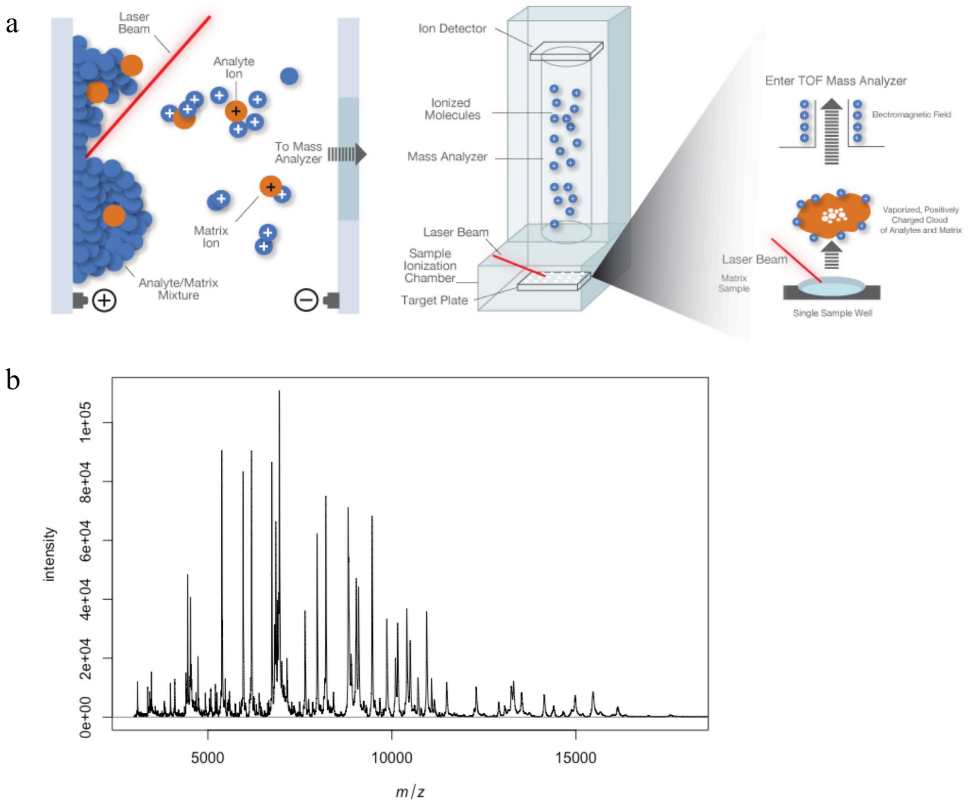


Figure 6: (a) Overview of MALDI-TOF MS technique (source: https://infograph.venngage.com/p/171868/maldi-tof-ms). (b) Bacterial (GBS) mass fingerprint generated by MALDI-TOF MS.

The quality of the mass spectrum, which can vary greatly depending on sample starting material or chemical properties of the microorganism to be analyzed, is essential for

successful microbial species assignment. Many microbes encountered in routine diagnostics, including most gram-negative bacteria, are analyzed via direct cellular profiling, e.g. the direct deposition of microbial colonies on the MALDI target plate together with matrix solution [148,149]. This method has the obvious advantage of minimal pre-analysis sample processing required, which facilitates high-throughput analysis in routine settings. For the analysis of gram-positive bacteria, whose composition of the cell wall poses a challenge for adequate ionization of the intracellular proteins, a pre MALDI-TOF MS analysis extraction protocol using formic acid is often carried out for improved species identification [144,150]. Other pre-analysis sample processing methods that can be employed to increase spectral quality for more demanding microbes, include treatment of microbial cells with ethanol, acetonitrile and formic acid [151], protein precipitation using chloroform/methanol mixture [152] or mechanical rupture of bacterial cells with bead beating [153].

1.4.3 MALDI-TOF MS in microbiological diagnostics

Based on its MALDI-TOF MS mass spectrum, genus or species identity of a microbe can be reliably determined and in some cases even closely related strains can be distinguished [148,154]. For this classification step, a generated protein mass spectrum is matched against a set of representative spectra of known microbes deposited in a reference database. In the early phase of MALDI-TOF MS development in the microbiology laboratory, mass spectra were classified by comparison against in-house reference records. Given their highly customized nature, such in-house databases can possess a high discriminatory capacity, making them a suitable solution for targeted research purposes [144]. More common in routine diagnostic laboratories, which need to cover a broad range of microorganisms, is the use of commercial MALDI-TOF MS databases. Such databases contain thousands of reference spectra of microbes commonly encountered in clinical settings and are permanently increasing in size, as reference spectra of novel microbial species are added by the manufacturers. The two best-established MALDI-TOF MS databases for microbiological identification are the MALDI-Biotyper database, which was developed by Bruker Daltonics and the Vitek-MS database, which was formerly known as SARAMIS and was jointly developed by Shimadzu and BioMérieux [143]. Both databases allow the user to manually add new reference mass spectra and to develop custom databases, which is important for more targeted discriminatory analyses that go beyond species level identification of organisms [144]. A hallmark step for the application of MALDI-TOF MS in clinical microbiology was the recent regulatory

approval of the MALDI-Biotyper and Vitek-MS platforms for the routine identification of bacteria and yeasts, by the Food and Drug Administration (FDA) [155].

MALDI-TOF MS has rapidly developed into an extremely versatile tool in routine microbiological diagnostics. In bacteriology, MALDI-TOF MS has been applied to identify a broad range of species, isolated from blood, cerebrospinal fluids, stool and urine samples, with an overall speed and accuracy that surpasses conventional diagnostic methods [143]. MALDI-TOF MS was also successfully applied for identifying food- or water-borne bacteria, in the characterization of environmental bacterial species [156] and for the detection of agents of biological warfare [143]. A further and highly relevant field of application is the rapid detection of antimicrobial resistances using MALDI-TOF MS. Examples include the discrimination of lineages and strains of methicillin-resistant *Staphylococcus aureus* [157,158] and the identification of vancomycin-resistant enterococci [159,160]. MALDI-TOF MS has been well described in the context of GBS identification and is considered a highly reliable method for GBS confirmatory testing following bacterial cultivation on agar medium [87]. There are studies that investigated strain-level typing of GBS employing MALDI-TOF MS. Several protein masses that are specific for the hypervirulent ST17 and the emerging ST1 GBS clones were described [161–163].

1.4.4 Ribosomal subunit protein biomarkers

Although each of the common commercial databases uses their own algorithms and interpretive criteria for microbial identification, they all operate on a pattern-recognition or fingerprinting based approach [144]. Novel biomarker-based MALDI-TOF MS approaches for the characterization of bacteria have been shown to perform excellent in terms of overall sensitivity and discrimination of organisms below species level [164,165]. The numerous proteins constituting the two ribosomal subunits, have been reported as ideal biomarker candidates for MALDI-TOF MS by several authors [156,164–167]. Proteins with molecular weights between 2-20 kDa, which is the primarily considered mass range for microbial identification by MALDI-TOF MS, make up an estimated 60-70% of a bacterial cells dry weight. The vast majority of these proteins are ribosomal subunit proteins (rsp), thereby accounting for most mass peaks contained in the MALDI spectra, along with other housekeeping proteins like DNA-binding proteins, RNA chaperones and other proteins involved in cell division and metabolism [143,168]. Given the fundamental role of the ribosome within the bacterial cell, rsp are generally highly conserved, but nevertheless

subjected to some inter-strain variability, making them an ideal target to deduce long-term phylogenetic relationships and to distinguish between closely related strains [165,169]. Using rsp as biomarkers requires access to WGS, e.g. knowledge of the *in silico* predicted rsp molecular masses. Acquiring the WGS of thousands of bacterial species and strains may not have been feasible twenty years ago but is now realistic, given the wealth of genomic data becoming widely available [170]. Targeted, rsp-based MALDI-TOF MS therefore has a huge potential for large-scale and high-throughput strain-level typing of organisms in the context of epidemiological studies.

1.5 Aims of this thesis

The United Nations sustainable developmental goal 3 (good health and well-being) includes the goals to significantly reduce global maternal mortality and preventable deaths of newborns and children under 5 years of age until the year 2030 [171]. We have contributed with the work presented in this thesis to vaccine development efforts targeting two major infectious diseases of infants and children in sub-Saharan Africa, namely *P. falciparum* malaria and *S. agalactiae* invasive disease.

In the first part of this thesis, we aimed to:

i) Investigate the safety, immunogenicity and efficacy of a cryopreserved, radiation-attenuated, whole *P. falciparum* sporozoite vaccine in Tanzanian volunteers followed by homologous controlled human malaria infection (*Chapter 2*).

ii) Elucidate gene expression changes in peripheral blood upon controlled human malaria infection in malaria pre-exposed, unvaccinated volunteers using RNA-Seq (*Chapter 3*).

In the second part of this thesis, we aimed to:

iii) Develop a MALDI-TOF MS based typing method that allows sub-species level typing of Group B *Streptococcus* for monitoring of vaccination impact on population structure (*Chapters 4 and 5*).

iv) Confirm the inter-laboratory transferability of the novel MALDI-TOF MS typing method for rapid screening and detection of GBS genotypes (*Chapter 6*).

Chapter 2

Safety, Immunogenicity, and Protective Efficacy against Controlled Human Malaria Infection of *Plasmodium falciparum* Sporozoites Vaccine in Tanzanian Adults

---

This chapter contains the following publication:

Said A. Jongo, Seif A. Shekalage, L.W. Preston Church, Adam J. Ruben, Tobias Schindler, Isabelle Zenklusen, Tobias Rutishauser, **Julian Rothen**, Anneth Tumbo, Catherine Mkindi, Maximilian Mpina, Ali T. Mtoro, Andrew S. Ishizuka, Kamaka Ramadhani Kassim, Florence A. Milando, Munira Qassim, Omar A. Juma, Solomon Mwakasungula, Beatus Simon, Eric R. James, Yonas Abebe, Natasha KC, Sumana Chakravarty, Elizabeth Saverino, Bakari M. Bakari, Peter F. Billingsley, Robert A. Seder, Claudia Daubenberger, B. Kim Lee Sim, Thomas L. Richie, Marcel Tanner, Salim Abdulla and Stephen L. Hoffman. "Safety, Immunogenicity and Protective Efficacy against Controlled Human Malaria Infection of *Plasmodium falciparum* Sporozoites Vaccine in Tanzanian Adults". 2018. *Am J Trop Med Hyg.*

Am. J. Trop. Med. Hyg., 99(2), 2018, pp. 338–349
doi:10.4269/ajtmh.17-1014

Safety, Immunogenicity, and Protective Efficacy against Controlled Human Malaria Infection of *Plasmodium falciparum* Sporozoite Vaccine in Tanzanian Adults

Said A. Jongo,[1]† Seif A. Shekalaghe,[1]† L. W. Preston Church,[2] Adam J. Ruben,[2] Tobias Schindler,[3,4] Isabelle Zenklusen,[3,4] Tobias Rutishauser,[3,4] Julian Rothen,[3,4] Anneth Tumbo,[1] Catherine Mkindi,[1] Maximillian Mpina,[1] Ali T. Mtoro,[1] Andrew S. Ishizuka,[5] Kamaka Ramadhani Kassim,[1] Florence A. Milando,[1] Munira Qassim,[1] Omar A. Juma,[1] Solomon Mwakasungula,[1] Beatus Simon,[1] Eric R. James,[2] Yonas Abebe,[2] Natasha KC,[2] Sumana Chakravarty,[2] Elizabeth Saverino,[2] Bakari M. Bakari,[1] Peter F. Billingsley,[2] Robert A. Seder,[5] Claudia Daubenberger,[3,4] B. Kim Lee Sim,[2,6] Thomas L. Richie,[2] Marcel Tanner,[3,4] Salim Abdulla,[1]† and Stephen L. Hoffman[2]*†

[1]*Bagamoyo Research and Training Centre, Ifakara Health Institute, Bagamoyo, Tanzania;* [2]*Sanaria Inc., Rockville, Maryland;* [3]*Swiss Tropical and Public Health Institute (Swiss TPH), Basel, Switzerland;* [4]*University of Basel, Basel, Switzerland;* [5]*Vaccine Research Center (VRC), National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland;* [6]*Protein Potential LLC, Rockville, Maryland*

Abstract. We are using controlled human malaria infection (CHMI) by direct venous inoculation (DVI) of cryopreserved, infectious *Plasmodium falciparum* (Pf) sporozoites (SPZ) (PfSPZ Challenge) to try to reduce time and costs of developing PfSPZ Vaccine to prevent malaria in Africa. Immunization with five doses at 0, 4, 8, 12, and 20 weeks of 2.7×10^5 PfSPZ of PfSPZ Vaccine gave 65% vaccine efficacy (VE) at 24 weeks against mosquito bite CHMI in U.S. adults and 52% (time to event) or 29% (proportional) VE over 24 weeks against naturally transmitted Pf in Malian adults. We assessed the identical regimen in Tanzanians for VE against PfSPZ Challenge. Twenty- to thirty-year-old men were randomized to receive five doses normal saline or PfSPZ Vaccine in a double-blind trial. Vaccine efficacy was assessed 3 and 24 weeks later. Adverse events were similar in vaccinees and controls. Antibody responses to Pf circumsporozoite protein were significantly lower than in malaria-naïve Americans, but significantly higher than in Malians. All 18 controls developed Pf parasitemia after CHMI. Four of 20 (20%) vaccinees remained uninfected after 3 week CHMI ($P = 0.015$ by time to event, $P = 0.543$ by proportional analysis) and all four (100%) were uninfected after repeat 24 week CHMI ($P = 0.005$ by proportional, $P = 0.004$ by time to event analysis). *Plasmodium falciparum* SPZ Vaccine was safe, well tolerated, and induced durable VE in four subjects. Controlled human malaria infection by DVI of PfSPZ Challenge appeared more stringent over 24 weeks than mosquito bite CHMI in United States or natural exposure in Malian adults, thereby providing a rigorous test of VE in Africa.

INTRODUCTION

In 2015 and in 2016, there were an estimated 429,000–730,500 deaths caused by malaria.[1–3] *Plasmodium falciparum* (Pf) is the cause of > 98% of malaria deaths and > 80% of malaria cases in sub-Saharan Africa. Our goal is to field a vaccine that will prevent infection with Pf and thereby prevent all manifestations of Pf malaria and parasite transmission from humans to mosquitoes.[4]

Plasmodium falciparum sporozoites (SPZ) are the only immunogens that have ever prevented Pf infection in > 90% of subjects.[5–7] Sanaria® PfSPZ Vaccine (Sanaria Inc., Rockville, MD) is composed of radiation-attenuated, aseptic, purified, cryopreserved PfSPZ.[8,9] When administered by rapid intravenous injection, PfSPZ Vaccine protected 100% (6/6) of malaria-naïve subjects in the United States against mosquito bite–controlled human malaria infection (CHMI) with Pf parasites similar to those in the vaccine (homologous) 3 weeks after the last immunization,[10] and 65% at 24 weeks.[11] Protection was durable against homologous mosquito bite CHMI for at least 59 weeks[12] and heterologous (parasites different than in vaccine) mosquito bite CHMI for at least 33 weeks.[13] PfSPZ Vaccine also prevented naturally transmitted heterogeneous Pf in adults in Mali for at least 24 weeks (vaccine efficacy [VE] 52% by time to event and 29% by proportional analysis).[14]

We used the same dosage regimen as in the United States and Mali to evaluate the tolerability, safety, immunogenicity, and VE of PfSPZ Vaccine in young adult male Tanzanians. Previously, we had conducted the first modern CHMI in Africa and showed that injection of aseptic, purified, cryopreserved PfSPZ, Sanaria® PfSPZ Challenge, consistently infected Tanzanian volunteers and subsequently repeated in multiple other countries.[15–21] In this study, we took advantage of this capability to assess VE of PfSPZ Vaccine by CHMI with PfSPZ Challenge (NF54). The same PfSPZ Vaccine dosage regimen was less immunogenic and protective in Tanzanians than in Americans,[11] and VE against homologous CHMI in Tanzania was lower (or similar) to VE against intense field exposure to heterogeneous Pf parasites in Mali.[14]

MATERIAL AND METHODS

Study design and population. This double-blind, randomized, controlled trial was conducted in Bagamoyo, Tanzania, between April 2014 and August 2015. Sixty-seven healthy male volunteers of 18–35 years of age were recruited from higher learning institutions in Dar es Salaam. After initial screening, prospective volunteers were invited to the Bagamoyo Clinical Trial Unit of the Ifakara Health Institute (IHI) to complete informed consent and screening.

All had to complete a 20-question assessment of trial understanding with a 100% correct response rate on the first or second attempt (Supplemental Table 1) to be eligible. Volunteers were screened using predetermined inclusion and exclusion criteria (Supplemental Tables 2 and 3). History of malaria in the previous 5 years or antibodies to Pf exported protein 1 (PfEXP1) by an enzyme-linked immunosorbent assay (ELISA) above a level associated with a single, recent Pf infection by CHMI[19] (see the Antibody assays section) were the exclusion criteria. Hematology, biochemistry, and parasitology testing, including malaria thick blood smear (TBS), stool,

* Address correspondence to Stephen L. Hoffman, Sanaria Inc., 9800 Medical Center Dr., Suite A209, Rockville, MD 20850. E-mail: slhoffman@sanaria.com
† These authors contributed equally to this study.

and urine by microscopy was carried out. Tests for human immunodeficiency virus and hepatitis B and C were performed after counseling; volunteers were excluded if positive and referred for evaluation and management by appropriate local physicians. Volunteers were excluded if they had significant abnormalities on electrocardiograms.

The trial was performed in accordance with Good Clinical Practices. The protocol was approved by institutional review boards (IRBs) of the IHI (Ref. No. IHI/IRB/No:02-2014), the National Institute for Medical Research Tanzania (NIMR/HQ/R.8a/Vol.IX/1691), the Ethikkommission Nordwest-und Zentralschweiz, Basel, Switzerland (reference number 261/13), and by the Tanzania Food and Drug Authority (Ref. No. TFDA 13/CTR/0003); registered at Clinical Trials.gov (NCT02132299); and conducted under U.S. FDA IND application.

Investigational products (IPs). The IPs were Sanaria® PfSPZ Vaccine[8–14] and Sanaria® PfSPZ Challenge.[15–20] PfSPZ Vaccine consists of aseptic, purified, vialed, metabolically active, nonreplicating (radiation attenuated), cryopreserved PfSPZ (NF54 strain). It was stored, thawed, diluted, and administered by direct venous inoculation (DVI) in 0.5 mL through a 25-gauge needle.[11,14,18,20] PfSPZ Challenge is identical to PfSPZ Vaccine except it is not radiation attenuated. It was handled and administered like PfSPZ Vaccine. Preparation of IPs was supervised by the study pharmacist. After labeling the syringe, the pharmacist handed it to the clinical team through a window.

Allocation and randomization. Volunteers were allocated to five groups (Table 1; Figure 1). Forty-nine received PfSPZ Vaccine and eight normal saline (NS). Ten were additional infectivity controls. The clinical team and volunteers were blinded to assignment to vaccine or NS until study end.

Group 1. Three volunteers received consecutive doses of 3×10^4, 1.35×10^5, and 2.7×10^5 PfSPZ of PfSPZ Vaccine at 4-week intervals to assess safety (Group 1).

Groups 2 and 3. Volunteers were randomized to receive 1.35×10^5 PfSPZ of PfSPZ Vaccine ($N = 20$) or NS ($N = 4$) (Group 2), or 2.7×10^5 PfSPZ of PfSPZ Vaccine ($N = 20$) or NS ($N = 4$) (Group 3) at 0, 4, 8, 12, and 20 weeks.

Group 4. Six volunteers were immunized with 2.7×10^5 PfSPZ of PfSPZ Vaccine on the same schedule as Group 3.

Group 5. Ten volunteers served as unblinded infectivity controls during CHMIs (see in the following paragraph): two with CHMI #1, two with CHMI #2, and six with CHMI #3.

Vaccine efficacy. *Controlled human malaria infection.* Vaccine efficacy was assessed by CHMI by DVI of 3.2×10^3 PfSPZ of PfSPZ Challenge. Controlled human malaria infection #1 was 3 weeks after the last immunization in Group 2. Controlled human malaria infection #2 was 3 weeks after the last immunization in Group 3. Controlled human malaria infection #3 was 24 weeks after the last immunization in Groups 3 and 4 and included the four volunteers in Group 3 who did not develop parasitemia after CHMI #2 and the six Group 4 volunteers. Volunteers were inpatients from day 9 after PfSPZ Challenge injection for observation until diagnosed and treated for malaria or until day 21; daily outpatient monitoring for TBS-negative volunteers continued until day 28. Thick blood smears were obtained every 12 hours on days 9–14 after CHMI and daily on days 15–21 until positive or until day 21. Thick blood smears could be performed more frequently, if volunteers had symptoms/signs consistent with malaria. After initiation of treatment, TBSs were assessed until two consecutive daily TBSs were negative and on day 28.

Detection of Pf parasites and parasite DNA. Slide preparation and reading for TBSs were performed as described.[19] Sensitivity was 2 parasites/μL blood unless the volunteer was symptomatic, in which case four times as many fields were read. Parasitemia was also determined by quantitative polymerase chain reaction (qPCR) with sensitivity of 0.1 parasites/μL blood based on a multiplex assay detecting *Plasmodium* spp. 18S genes and the human RNaseP gene as endogenous control.[22] A second, more sensitive qPCR assay with a sensitivity of 0.05 parasites/μL blood and targeting the Pf-specific telomere-associated repetitive element 2[23] was used to reanalyze all samples that were negative by 18S-based qPCR. After the start of CHMI, the time of first blood sample positivity by qPCR was used to determine infection status and for the calculation of prepatent period. Volunteers were continuously monitored by qPCR until malaria treatment based on TBS positivity. The World Health Organization International Standard for Pf DNA Nucleic Acid Amplification Techniques (NIBSC, Hertfordshire, United Kingdom) was used as standard for calculation of parasite densities. DNA was extracted from 100 μL whole blood and eluted with 50 μL Elution Buffer using Quick-gDNA Blood MicroPrep Kit (Zymo Research, Irvine, CA). Blood samples were analyzed retrospectively by qPCR after storing at –80°C after the conclusion of CHMIs. To exclude field strain infections, parasite genotyping was performed on samples randomly chosen as described.[24] In all cases in which TBS was negative and qPCR was considered positive, two consecutive samples were positive by qPCR.

Adverse events (AEs). Volunteers were observed as inpatients for 48 hours after administration of IP and discharged with diaries and thermometers for recording AEs and temperatures and followed with daily telephone calls. Symptoms and signs (solicited and unsolicited) were recorded and graded by physicians: mild (easily tolerated), moderate (interfere with normal activity), severe (prevents normal activity),

TABLE 1
Demographic characteristics of volunteers

	Vaccinees	Normal saline controls	Infectivity controls
Number of volunteers	49	8	10
Percentage males	100%	100%	100%
Mean age in years (range)	24 (20, 30)	23 (20, 28)	25 (21, 28)
Percentage Africans	100%	100%	100%
Mean body mass index (range)	22.33 (18.00, 29.70)	21.91 (19.00, 24.20)	21.68 (18.40, 24.30)
Number (%) heterozygous for alpha thalassemia	22 (44.9%)	4 (50%)	5 (50%)
Number (%) with LTBI* (QuantiFERON positive)	17 (34.7%)	3 (36.5%)	1 (10%)
Number (%) positive on screening of urine or stool for parasitic infection	0 (0%)	1 (12.5%)	0 (0%)
Number (%) students	49 (100%)	8 (100%)	10 (100%)

* Latent tuberculosis infection.

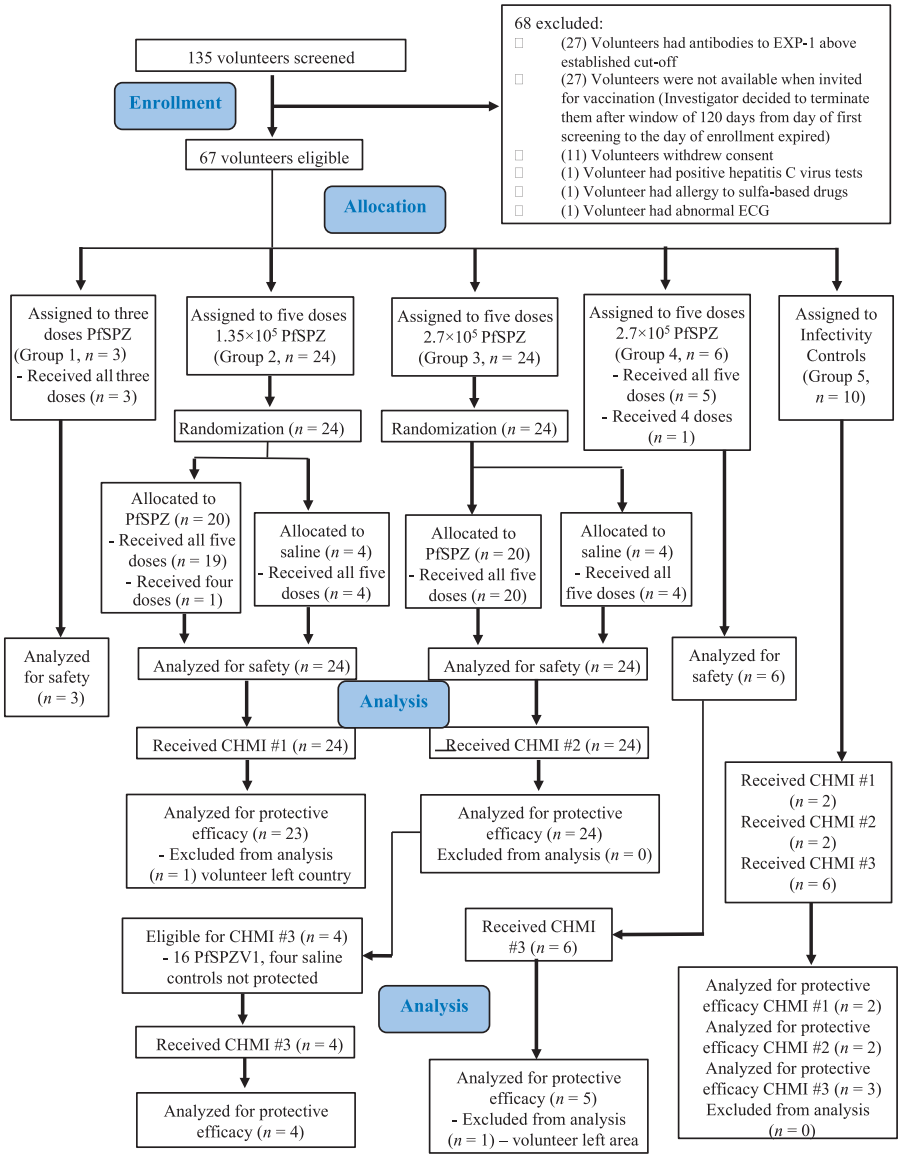


FIGURE 1. Volunteer participation (CONSORT 2010 diagram). This figure appears in color at www.ajtmh.org.

or life threatening. Axillary temperature was grade 1 (> 37.5–38.0°C), grade 2 (> 38.0–39.0°C), grade 3 (> 39.0–40.0°C), or grade 4 (> 40.0°C). Hematological and biochemical abnormalities were also assessed using standard clinical assays.

During the first 7 days after injection of IPs, prespecified local (site of injection) and systemic AEs were solicited. Open-ended questioning was used to identify unsolicited AEs through day 28 (Supplemental Table 4). All AEs were assessed for severity and relatedness to IP administration. Adverse events were classified as definitely related, probably related, possibly related, unlikely to be related, and not related. Definitely, probably, and possibly were considered to be related. Unlikely to be related and not related were considered to be unrelated. For CHMIs, volunteers returned on day 9 for admission to the ward for diagnosis and treatment of malaria. Events during the 8–28 day period were assessed for relationship to Pf infection and considered related if the event was within 3 days before and 7 days after TBS was first positive.

Antibody assays. Sera were assessed for antibodies by ELISA, immunofluorescence assay (aIFA), and inhibition of sporozoite invasion (ISI) assay as described (see Supplemental Table 5).[25] For ELISAs, the results are reported as the serum dilution at which the optical density (OD) was 1.0. Enzyme-linked immunosorbent assay for PfEXP1 was used to screen volunteers for possible malaria exposure (Supplemental Table 6). Any subject with an OD 1.0 of ≥ 600 was excluded. This was because we had previously determined in Tanzanians who underwent CHMI[19] that antibodies to PfEXP1 at this level were a sensitive indicator of recent Pf infection (unpublished).

T-cell assays. T-cell responses in cryopreserved peripheral blood mononuclear cells (PBMC) were measured by flow cytometry in a single batch after the study as described.[12] After stimulation, cells were stained as described.[26] The staining panels are in Supplemental Table 7 and antibody clones and manufacturers are in Supplemental Table 8. All antigen-specific frequencies are reported after background

subtraction of identical gates from the same sample incubated with control antigen. Data were analyzed with FlowJo v9.9.3 (TreeStar, Ashland, OR) and graphed in Prism v7.0a (GraphPad, San Diego, CA).

Statistical analysis. Comparisons of categorical variables between groups were analyzed using 2-tailed Fisher's exact test. Comparisons of continuous variables between groups were analyzed by 2-tailed nonparametric tests. For multiple group comparisons, the Kruskal–Wallis test was used. Time to event was assessed by the Kaplan–Meier curves and log-rank test. Vaccine efficacy by time to event was quantified using Cox proportional hazards ratios. Time to event data were analyzed from CHMI injection until positive TBS result or positive qPCR result. Controlled human malaria infection follow-up period lasted until day 28 after CHMI injection. Analyses of immunological data are described with the data.

Role of the funding source. The funders were involved in study design, study management, data collection, data analysis, data interpretation, and writing the report. Salim Abdulla and Stephen L. Hoffman had full access to all data in the study and final responsibility for decision to submit for publication.

RESULTS

Study population and experience with DVI. Fifty-seven Tanzanian men (Table 1; Figure 1) met the criteria (Supplemental Tables 2 and 3) and received PfSPZ Vaccine (*N* = 49) or NS (*N* = 8). All volunteers had AA hemoglobin and normal G6PD activity. Thirty-one volunteers (46%) were heterozygous for α-thalassemia; 21 had evidence of latent tuberculosis infection by Quantiferon testing, but showed no evidence of active tuberculosis. One volunteer (group 2, NS) had *Strongyloides stercoralis* on screening and was successfully treated before vaccination (Table 1).

Of 237 immunizations with PfSPZ Vaccine, 234 were completed with a single injection (98.7%). Two hundred and thirty injections (97.0%) were considered painless by the volunteer. For NS subjects, 39 of 40 immunizations (97.5%) were completed in a single injection and 39 of 40 (97.5%) considered painless by the volunteer. The nurse performing immunizations considered the procedure to be simple in 265 of 273 single injections (97.1%).

One subject in Group 2 received four immunizations. The third immunization was withheld while the subject was evaluated for what was diagnosed as benign ethnic neutropenia.[27,28] One subject in Group 4 missed his second immunization when he left town. All other subjects (other than Group 1 and added infectivity controls) received five immunizations.

Safety. Among 49 volunteers who received 237 doses of PfSPZ Vaccine, there were 17 solicited AEs possibly related to IP (17/237 = 7.2%) in 10 of the 49 vaccinees (20.4%) (Table 2). Among eight volunteers who received 40 doses of NS, there were two solicited AEs possibly related to IP (2/40 = 5.0%) in one of the eight controls (12.5%) (Table 2). There were no AEs considered by the clinicians to be probably or definitely related to IP. There were no local or serious AEs. One episode each of headache and fever were grade 2; all other solicited AEs were grade 1. None of the comparisons of AEs between vaccinees and controls or between Group 2 (1.35×10^5 PfSPZ) and Groups 3 and 4 (2.7×10^5 PfSPZ) showed statistically significant differences (Table 2). Twenty-six of 49 vaccinees (53.1%) experienced 43 unsolicited AEs (0.88/individual) in the 28 days following injections #1–#4 and the 21 days before CHMI after injection #5. Seven of eight controls (87.5%) experienced 14 unsolicited AEs (2/individual) during this period. None of these unsolicited AEs recorded within 28 days of an immunization was considered related to IP.

Laboratory abnormalities occurred at roughly equal rates comparing PfSPZ Vaccine recipients and controls, except for leukocytosis and eosinophilia, which were more frequent in controls (Table 3). There was no apparent explanation for these differences. A cyclic variation in total bilirubin following

TABLE 2

Solicited AEs by group considered possibly* related to administration of the investigational product during the first 7 days post immunization

	Group 1 (dose escalation)	Group 2 (1.35×10^5 PfSPZ)	Group 3 (2.7×10^5 PfSPZ)	Group 4 (2.7×10^5 PfSPZ)	Total PfSPZ vaccine	NS controls
Number of volunteers	3	20	20	6	49	8
Total number of injections	9	99	100	29	237	40
Number of local AEs	0	0	0	0	0	0
Numbers of systemic AEs (% of total immunizations)						
All	1 (11%)	10 (10.1%)	6 (6%)	0	17 (7.2%)	2 (5.0%)
Headache*	1 (11%)	7 (7%)†	2 (2%)	0	10 (4.2%)	1 (2.5%)
Abdominal pain	0	2 (2%)	1 (1%)	0	3 (1.3%)	0
Chills	0	0	1 (1%)	0	1 (0.4%)	0
Fever	0	0	2 (2%)	0	2 (0.8%)	0
Diarrhea	0	0	0	0	0	1 (2.5%)
Chest pain	0	1 (1%)	0	0	1 (0.4%)	0
Other	0	0	0	0	0	0
Systemic AEs - no. volunteers with ≥ 1 event (% of volunteers)						
Any	1 (33%)	7 (35%)	2 (10%)	0	10 (20.4%)	1 (13%)
Headache	1 (33%)	6 (30%)	2 (10%)	0	9 (18.4%)	1 (13%)
Abdominal pain	0	2 (10%)	1 (5%)	0	3 (6.1%)	0
Chills	0	0	1 (5%)	0	1 (2.0%)	0
Fever	0	0	2 (10%)	0	2 (4.1%)	0
Diarrhea	0	0	0	0	0	1 (13%)
Chest pain	0	1 (5%)	0	0	1 (2.0%)	0
All other	0	0	0	0	0	0

AEs = adverse events; PfSPZ = *Plasmodium falciparum* sporozoites. There were no significant differences between vaccinees as compared with normal saline (NS) controls for any or all AEs. All AEs were grade 1, except one headache and one fever. Local solicited AEs: injection site pain, tenderness, erythema, swelling, or induration. Systemic solicited AEs: allergic reaction (rash, pruritus, wheezing, shortness of breath, bronchospasm, allergy-related edema/angioedema, hypotension, and anaphylaxis), abdominal pain, arthralgia, chest pain/discomfort, chills, diarrhea, fatigue, fever, headache, malaise, myalgia, nausea, pain (other), palpitations, shortness of breath, and vomiting.

* All AEs were considered possibly related. None were considered probably or definitely related.

† 4/7 episodes of headache occurred after the third vaccine dose and did not recur with fourth or fifth doses. No factor was identified to account for this apparent clustering of headache.

each immunization was observed equally in volunteers receiving vaccine or NS that was attributed to enriched diet, as the volunteers were transported to Bagamoyo from Dar es Salaam during the periods of immunization and CHMI and were amply fed (see Supplemental Figure 1). In Dar es Salaam, malaria transmission is low. No volunteer had malaria during screening or during the trial other than from CHMI.

Tolerability, safety, and VE during CHMI. Forty-six vaccinees, eight NS controls, and 10 added infectivity controls underwent homologous CHMI. All subjects were negative by TBS and qPCR for Pf infection on the day of CHMI. Two volunteers were excluded from primary analysis—a Group 2 volunteer who left the area 2 days after administration of PfSPZ Challenge and a Group 4 volunteer who left 9 days after. Both volunteers were located and treated preemptively.

Tolerability and safety of administration of PfSPZ challenge. Controlled human malaria infection was well tolerated with no local solicited AEs and three systemic solicited AEs (grade 1 headache in Group 3, grade 2 headache in Group 4, and grade 1 arthralgia in an infectivity control) in the 7 days post-administration of PfSPZ Challenge.

Parasitemia. Controls. The 18 NS and infectivity controls developed Pf infection after CHMI (16 TBS and qPCR positive and two TBS negative and qPCR positive) (Figure 2A–D and Supplemental Table 9). These included four NS and two infectivity controls in CHMI #1, the same in CHMI #2, and six infectivity controls in CHMI #3. All received the same lot of PfSPZ Challenge. One isolate of those positive from CHMI #1, one from CHMI #2, and four from CHMI #3 were genotyped,[24] and all parasites tested were PfNF54. Vaccine efficacy was calculated based on the results of qPCR assays from the six controls in CHMI #1, CHMI #2, and CHMI #3 individually (Figure 2D).

Group 2 (1.35×10^5 PfSPZ). Seventeen of 18 volunteers who received five doses and 1/1 volunteer who received four doses developed parasitemia (Figure 2A), 15 positive by TBS and qPCR, and 3 by qPCR only (CHMI #1) (Supplemental Table 10). One volunteer was negative through day 28 by TBS and qPCR. Vaccine efficacy by proportional analysis was 5.56% (95% confidence interval [CI]: 3.61%, 14.73%; $P > 0.99$, Fisher's exact test, 2-tailed). There was no significant delay in parasitemia by qPCR in the vaccinees as compared with controls ($P = 0.4481$ by log rank).

Group 3 (2.7×10^5 PfSPZ). First CHMI at 3 weeks (CHMI #2): 16/20 volunteers who received five doses developed parasitemia (Figure 2B), all positive by TBS and qPCR; four volunteers were negative through day 28 by TBS and qPCR. Vaccine efficacy by proportional analysis was 20% (95% CI: 4.62%, 35.38%; $P = 0.543$). There was a delay in the onset of parasitemia in vaccinees as compared with controls ($P = 0.015$ by log rank).

Second CHMI at 24 weeks (CHMI #3): The four uninfected volunteers from the first CHMI underwent a second CHMI 24 weeks after the last vaccine dose (Figure 2C). Three were negative by TBS and qPCR through day 28 day. The fourth volunteer, who was asymptomatic, was reported to have a positive TBS on day 12 and treated. The sample with positive TBS was negative by retrospective qPCR. Reevaluation of the TBS indicated an error in slide reading (false-positive). Vaccine efficacy by proportional analysis at this time point was 100% (for 3/3 and 4/4 protected: 95% CI: 43.8%, 100%, and 51.01%, 100%; $P = 0.012$ and 0.005, respectively). However, given the 20% VE at 3 weeks by proportional analysis, overall VE by proportional analysis was considered to be 20%.

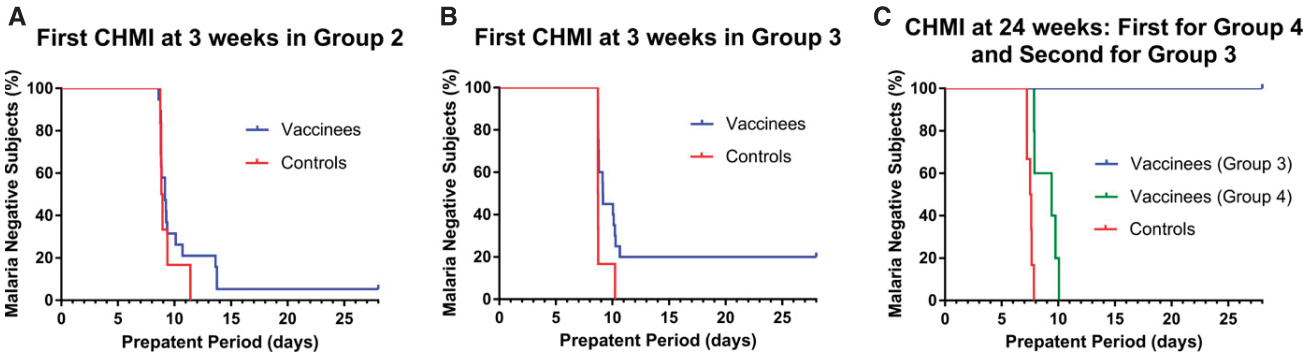
Group 4 (2.7×10^5 PfSPZ). First CHMI at 24 weeks after the last vaccine dose (CHMI #3): 4/5 vaccinees developed parasitemia by TBS and qPCR. The fifth was negative by TBS, but positive by qPCR (see Supplemental Table 10). There was one excluded volunteer (see the previous paragraph). Vaccine efficacy by proportional analysis was 0% ($P > 0.99$%). There was a significant delay in the onset of parasitemia by qPCR in vaccinees as compared with controls ($P = 0.001$ by log rank).

α-thalassemia. Volunteers heterozygous for α-thalassemia were no more likely to be TBS negative and qPCR positive than volunteers without α-thalassemia (three of 27 versus three of 34, $P = 1.0$). Protection from CHMI did not correlate with α-thalassemia status; 3/37 with normal hemoglobin and 2/29 heterozygous for α-thalassemia were protected.

TABLE 3
Summary of abnormal laboratory values and severity grades

Laboratory parameter	Vaccinees in Group 2 (1.35×10^5 PfSPZ) ($N = 20$)		Vaccinees in groups 3 and 4 (2.7×10^5 PfSPZ) ($N = 26$)		NS controls ($N = 8$)		*P* values: vaccinees ($N = 46$) vs. controls ($N = 8$)
	No.	%	No.	%	No.	%	
Leukocytosis	1	5	2	7.7	3	37.5	0.0358
Leukopenia	6	30	7	27	1	12.5	> 0.05
Neutropenia	6	30	5	19	2	25	> 0.05
Lymphopenia	3	15	3	11.5	2	25	> 0.05
Eosinophilia	0	0	2	7.7	3	37.5	0.0194
Decreased hemoglobin	1	5	0	0	0	0	> 0.05
Thrombocytopenia	1	5	0	0	0	0	> 0.05
Elevated creatinine	2	10	4	15.4	2	25	> 0.05
Low total bilirubin	4	20	7	27	1	12.5	> 0.05
Elevated total bilirubin	2	10	2	7.7	2	25	> 0.05
Elevated alkaline phosphatase	1	5	2	7.7	0	0	> 0.05
Elevated alanine aminotransferase	3	15	5	19	2	25	> 0.05
Elevated aspartate aminotransferase	0	0	3	11.5	0	0	> 0.05

PfSPZ = *Plasmodium falciparum* sporozoites. *P* values calculated using Fisher's exact test (2-tailed). One volunteer who received saline developed Grade 3 eosinophilia attributed to *Strongyloides stercoralis* infection, which improved with anthelminthic therapy. This volunteer had a baseline of mild eosinophilia, which persisted throughout the clinical trial. All other laboratory abnormalities were Grade 2 or less. There was no association between laboratory abnormalities and time after a dose or increasing number of doses. Three abnormalities during immunization were deemed clinically significant or Grade 3. One was diagnosed as benign ethnic neutropenia, one was lymphopenia associated with an infected foot laceration, and one was eosinophilia associated with *Fasciolopsis buski* and *S. stercoralis* infection. Lymphopenia and eosinophilia resolved with treatment. Two Group 4 volunteers had asymptomatic hookworm infections diagnosed before controlled human malaria infection; one was coinfected with *Enterobius vermicularis*.



D

Group	Timing of CHMI	Number protected/ Number underwent CHMI	Vaccine efficacy by proportional analysis	Vaccine efficacy by proportional analysis 95% CI	Vaccine efficacy by proportional analysis *P* value (Fisher's Exact Test, 2-tailed)	Median prepatent period by qPCR of positive vaccinees (days)****	Median prepatent period by qPCR of controls (days)****	Log rank test
2 (1.35×10^5 PfSPZ)	3 weeks after fifth immunization	1/18	5.56%	(-3.61%, 14.73%)	> 0.999	9.2	8.9	0.45
3 (2.7×10^5 PfSPZ)	3 weeks after fifth immunization	4/20	20%	(4.62%, 35.38%)	0.543	9.0	8.7	**0.02**
3 (2.7×10^5 PfSPZ)	24 weeks after fifth immunization*	3/3 or 4/4**	100% (20% overall)	(43.85%, 100%) or (51.01%, 100%)***	**0.012 or 0.005****	not applicable	7.6	**0.01 or 0.004****
4 (2.7×10^5 PfSPZ)	24 weeks after fifth immunization	0/5	0%	(0.00%, 25.88%)***	> 0.999	8.7	7.6	**< 0.001**

FIGURE 2. Kaplan–Meier survival curves in immunized volunteers vs. controls as assessed by quantitative polymerase chain reaction (qPCR). Kaplan–Meier curves in volunteers undergoing controlled human malaria infection (CHMI) 3 weeks after the last of five doses with 1.35×10^5 (Group 2) (**A**) or 2.7×10^5 (Group 3) (**B**) *Plasmodium falciparum* Sporozoites (PfSPZ) of PfSPZ Vaccine. Panel (**C**) volunteers undergoing either first (Group 4) or second (Group 3) CHMI 24 weeks after the fifth immunization with 2.7×10^5 PfSPZ of PfSPZ Vaccine. (**D**) Vaccine efficacy and prepatent period results. *This was the second CHMI for the 4 volunteers in Group 3 who were protected after the first CHMI at 3 weeks. **One volunteer was inappropriately treated on day 13 for a false positive TBS. Without this volunteer, 3/3 protected. With this volunteer 4/4 were protected. ***Confidence intervals were calculated using Wilson's score interval. ****Volunteers in CHMI #1 and #2 (3 week CHMI in Groups 2 and 3) had specimens first acquired on day9. Volunteers in CHMI #3 (24 week CHMI in Groups 3 and 4) had specimens first acquired on day 8. This figure appears in color at www.ajtmh.org.

Prepatent periods and parasite densities. Although the median prepatent periods by TBS in controls in each CHMI group (12.5, 13.0, and 12.0, respectively) were shorter than in the vaccinees in Groups 2–4 (14.0, 14.0, and 15.3 days, respectively), these differences did not reach the level of statistical significance ($P = 0.486$, $P = 0.491$, and $P = 0.333$, respectively) (Supplemental Table 9). The prepatent periods by qPCR in vaccineees in Group 3 (3 and 24 week CHMIs) and Group 4 (24 week CHMI) were significantly longer than in the respective controls (Figure 2D). The parasite densities by qPCR and TBS at the time of diagnosis for each individual are in Supplemental Table 10. The median parasite density in controls versus vaccinees at the time of first positivity were 0.5 versus 0.4 parasites/μL for qPCR ($P = 0.5714$) and 11.2 versus 15.0 parasites/μL for TBS ($P = 0.1492$).

Tolerability and safety of parasitemia during CHMI. Controls. Sixteen controls developed parasitemia by TBS; 9 (56%) never had symptoms (Supplemental Table 11). Headache occurred in 7/7 symptomatic individuals. One of two control volunteers only positive by qPCR did not have any symptoms; the second had headache 8 days after qPCR spontaneously reverted to negative. No volunteer had symptoms at the time of first positive qPCR.

Vaccinees. Thirty-five immunized volunteers developed parasitemia by TBS; 20 (57%) never had symptoms. Three volunteers had temperature > 39.0°C; all other clinical manifestations were grade 1 or 2. Fever (28.6%) and headache (31.4%) were most common. Compared with controls, elevated temperature was more common in vaccinees with positive TBSs (9/35 versus 0/16, $P = 0.043$). There was no significant difference in the frequency of headache between controls and vaccinees. In the three volunteers in Group 2 who were qPCR positive and TBS negative, one developed headache 3 days after qPCR positivity. No volunteer had symptoms at the time of first positive qPCR.

Clinical laboratories. No unexpected changes were observed following CHMI. Declines in lymphocyte counts were observed in TBS positive controls and vaccinees (mean decline 1,110 ± 720 cells/μL and 1,180 ± 680 cells/μL, respectively) on day of first positive TBS. Absolute lymphocyte

counts less than 1,000 cells/μL were observed in 8/16 and 16/35 TBS positive controls and vaccinees. All lymphocyte counts returned to the baseline by day 28. There were mild decreases in platelet counts in TBS positive subjects, but all platelet counts were > 100 × 10^3 cells/μL.

Treatment. Volunteers with positive TBSs were treated with either atovaquone/proguanil (*N* = 43) or artemether/lumefantrine (*N* = 8) within 24 hours of first positive TBS. Normal saline and infectivity controls who were TBS negative (*N* = 2) were treated at day 28.

Immunogenicity. *Antibody responses.* Pf circumsporozoite protein (PfCSP) and PfSPZ. Antibodies against PfCSP by ELISA 1), PfSPZ by aIFA 2), and PfSPZ by ISI 3) in sera taken 2 weeks after the last vaccine dose and just before CHMI (20–23 days after the last dose) for Groups 2 (CHMI #1) and 3 (CHMI #2) are in Figure 3A–C. The median responses and those uninfected and infected by qPCR are shown.

For all three assays, median antibody responses before first CHMI were higher in uninfected than in infected vaccinees. There was a significant difference in median net aIFA

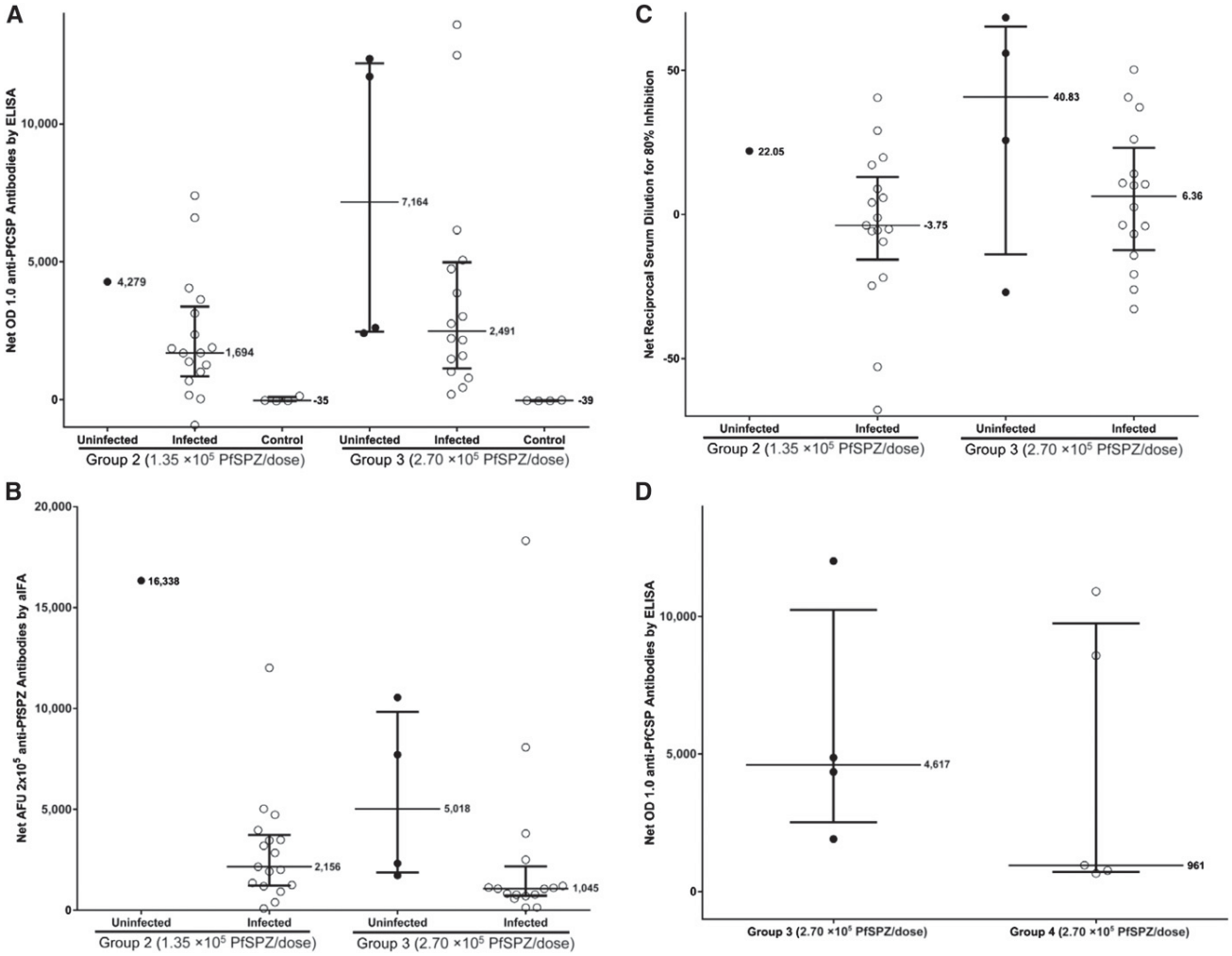


FIGURE 3. Antibody responses to *Plasmodium falciparum* Sporozoites (PfSPZ) and PfCSP before controlled human malaria infection (CHMI). For all assays, uninfected subjects are shown as filled (black) circles and infected subjects are open circles. For each of the defined subject groups, the interquartile ranges and the median values of response of subjects in each group are shown. Assessment of antibodies was performed in sera from subjects before immunization and before CHMI #1 (~2 weeks after the last dose of PfSPZ Vaccine or normal saline [NS]) and/or CHMI #2 (~24 weeks after last dose of PfSPZ or NS) (**A**, **D**). Antibodies to PfCSP by ELISA are reported as net optical density (OD) 1.0 (the difference in OD 1.0 between pre-CHMI and preimmunization sera). (**B**, **E**) Antibodies to PfSPZ by aIFA are reported as net AFU 2 × 10^5, the reciprocal serum dilution at which the fluorescent units were 2 × 10^5 (AFU 2 × 10^5) in pre-CHMI minus preimmunization sera. (**C**, **F**) Results of inhibition of sporozoite invasion (ISI) assay are reported as serum dilution at which there was 80% reduction of the number of PfSPZ that invaded a human hepatocyte line (HC-04) in the presence of pre-CHMI as compared with preimmunization sera from the same subject. Panels **A–C** show groups 2 (five doses of 1.35 × 10^5 PfSPZ) and 3 (five doses of 2.7 × 10^5 PfSPZ) before short-term CHMI (2 weeks after the last dose of PfSPZ or NS) and panels **D–F** show those volunteers in Groups 3 (five doses of 2.7 × 10^5 PfSPZ) and 4 (five doses of 2.7 × 10^5 PfSPZ) who underwent long-term CHMI (24 weeks after the last dose of PfSPZ). Panel **G** shows net optical density (OD) 1.0 anti-PfCSP antibodies by an enzyme-linked immunosorbent assay (ELISA) comparing vaccinated Tanzanian volunteers to volunteers in other trials receiving the same regimen. After five doses of 2.70 × 10^5 PfSPZ/dose, volunteers in bagamoyo sporozoite vaccine 1 (BSPZV1) (*N* = 25) had a 4.3-fold lower median net OD 1.0 than those in the U.S.-based clinical trial Walter Reed Army Institute of Research (WRAIR) 2080 (*N* = 26) but a 6.6-fold higher median OD 1.0 than volunteers in 14-I-N010 in Bamako, Mali (*N* = 42), where malaria transmission rates are higher. There was a significant difference between the results for WRAIR 2080 vs. BSPZV1 (*P* = 0.0012), WRAIR 2080 vs. 14-I-N010 (*P* < 0.0001), and even 14-I-N010 vs. BSPZV1 (*P* = 0.002) (two-tailed *t*-test). AFU = arbitrary fluorescence units; aIFA = antibodies by immunofluorescence assay.

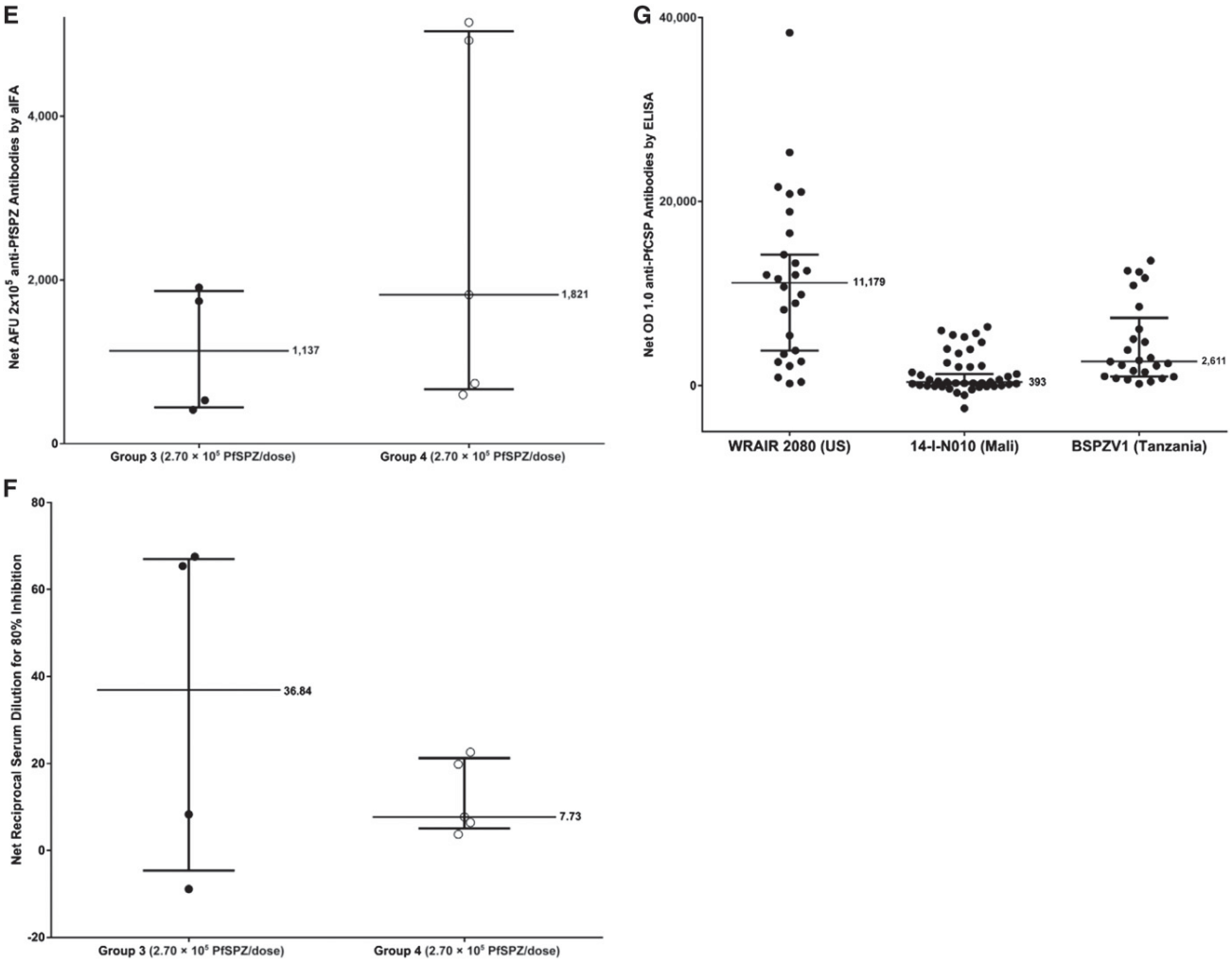


FIGURE 3. (*Continued*)

responses between infected and uninfected volunteers in Group 3 before CHMI #1 ($P = 0.0499$, Wilcoxon Rank-Sum Test), but not PfCSP ($P = 0.290$) or for ISI ($P = 0.249$).

In sera collected before CHMI #3 (170–171 days after the last vaccine dose), antibodies by the three assays for Group 4 and for the four volunteers in Group 3 uninfected in CHMI #1 who underwent CHMI #2 are in Figure 3D–F. All data appear in Supplemental Table 12.

After the fifth dose, in the PfCSP ELISA, volunteers were considered to have seroconverted if their net OD 1.0 and OD 1.0 ratio calculated, respectively, by subtracting or dividing by the prevaccination antibody OD 1.0, were ≥ 50 and ≥ 3.0. By these criteria, 15/18 volunteers (83%) in Group 2, 20/20 (100%) in Group 3, and 5/5 (100%) in Group 4 seroconverted, median net OD 1.0 of positives of 1,189, 2,685, and 961, and median OD 1.0 ratio of positives of 11.50, 21.15, and 37.83, respectively (Supplemental Table 13). In the aIFA, volunteers with a net arbitrary fluorescence unit (AFU) 2×10^5 of ≥ 150 and a ratio of post- to pre-AFU 2×10^5 of ≥ 3.0 were considered positive (Supplemental Table 13). By these criteria, 17/18 volunteers (94%) in Group 2, 18/20 (90%) in Group 3, and 5/5 (100%) in Group 4 seroconverted, median net OD 1.0 of positives of 2,844, 1,165, and 1,820, and median OD 1.0 ratio of positives of 1,193.00, 552.88, and 224.86, respectively (Supplemental Table 13). For the ISI, volunteers with a net ISI activity of ≥ 10% and a ratio of post- to pre-ISI activity of ≥ 3.0 were considered positive. By these criteria, 3/18 volunteers (17%) in Group 2, 8/20 (40%) in Group 3, and 3/5 (60%) in Group 4 were positive, median net OD 1.0 of positives of 22.05, 38.92, and 12.44, and median OD 1.0 ratio of positives of 19.79, 12.53, and 13.44, respectively (Supplemental Table 13).

Other antigens. Two weeks after the fifth dose in Group 2 (1.35×10^5 PfSPZ) and groups 3 and 4 (2.7×10^5 PfSPZ), there were antibodies to PfCSP in 15/18 and 25/25 subjects, respectively. Ten of 25 volunteers immunized with 2.7×10^5 PfSPZ made antibodies to Pf apical membrane antigen 1 and 4–16% responded to PfCelTOS, PfMSP5, PfMSP1, or Pf erythrocyte binding antigen 175 (PfEBA175) (Supplemental Table 14). The presence of antibodies, albeit at low incidence, against proteins first expressed in late liver stages (PfMSP1 and PfEBA175) was unexpected; results were confirmed by repeating the assays. No antibody responses were associated with protection.

T-cell responses. T cells against liver-stage malaria parasites in mice and nonhuman primates immunized with

radiation-attenuated SPZ mediate protection[9,29–31] and it is likely this is the case in humans.[12] CD8 and CD4 T-cell responses generally peak after the first vaccination with PfSPZ Vaccine.[13] In this trial, T-cell responses were measured before immunization, 2 weeks after the first and 2 weeks after the final immunization in Group 2 (1.35×10^5 PfSPZ). For technical reasons (loss of viability), the other groups could not be studied.

After the first vaccination, the percent of Pf red blood cell (PfRBC)-specific and PfSPZ-specific cytokine-producing memory CD4 T-cell responses increased by 0.25 ± 0.06 (mean ± SEM) and 0.24 ± 0.04, respectively (Figure 4A, B). Throughout, "naïve T cell" refers to cells that co-express CCR7 and CD45RA, and "memory T cell" refers to all other T cells. After the final vaccination, at week 22, the CD4 T-cell responses were above prevaccine responses by 0.17 ± 0.05 and 0.18 ± 0.05% points, respectively. These responses were lower than after the same immunization regimen in malaria-naïve U.S. adults.[10]

PfRBC-specific CD8 T cells were not significantly above the prevaccine levels, and PfSPZ-specific CD8 T cells were slightly above background (Figure 4C, D); responses were lower than in U.S. adults.[10,12]

In contrast to other PfSPZ Vaccine trials,[10,12–14] there was negligible change in the frequency of circulating γδ T cells (Figure 4E) or activation as measured by change in expression of the activation markers HLA-DR and CD38 following immunization (Figure 4F). To identify potential explanations for lower cellular immune responses in Tanzanians, we examined frequency of T regulatory (Treg) cells ($CD4^+Foxp3^+CD25^+CD127^-$) expressing the activation marker CD137 (also known as 4-1BB)[32] after stimulation with PfRBC. There was no difference in prevaccine frequency of PfRBC-specific Tregs in the Tanzanians as compared with Americans[10] (Figure 4G). Consistent with CD4 and CD8 T-cell responses, PfRBC-specific Tregs were highest after first immunization (Figure 4H). Last, the prevaccine frequency of total memory T cells relative to total naïve T cells was significantly higher than in Americans (Figure 4I).

DISCUSSION

To our knowledge, this was the first assessment of the VE of a malaria vaccine in Africa against CHMI. *Plasmodium falciparum* SPZ Vaccine was well tolerated and safe but less immunogenic and protective in Tanzanian men than in U.S. volunteers.

In our studies, all 18 controls became infected. Four of 20 (20%) recipients of five doses of 2.7×10^5 PfSPZ did not become infected after homologous CHMI by DVI 3 weeks after the last immunization. By contrast, 12/13 (92.3%) volunteers in the United States who received five doses of 2.7×10^5 PfSPZ were protected after homologous CHMI by mosquito bite 3 weeks after the last vaccine dose.[11] When the four uninfected Tanzanian volunteers underwent repeat homologous CHMI at 24 weeks after the last dose, all four (100%) were protected. In the United States, Seven of 10 previously protected volunteers were protected when they underwent homologous CHMI at 24 weeks[11] and all five volunteers in the United States who were protected at 21 weeks after the last immunization (four doses of 2.7×10^5 PfSPZ) were protected against repeat mosquito-administered CHMI at 59 weeks.[12] This could be due to boosting by the small numbers of PfSPZ administered during the CHMI, or is more likely due to the fact that in these protected individuals, the protective immune responses induced by immunization were sustained.

The same exact immunization regimen was assessed for VE against intense field transmission of heterogeneous Pf in Mali. Vaccine efficacy against infection with Pf on TBS was 52% by time to event and 29% by proportional analysis during 24 weeks after the last vaccine dose.[14] This was higher than the VE by proportional analysis against homologous CHMI in Tanzania. In Tanzania, there was a significant delay in the onset of parasitemia after CHMI at 3 and 24 weeks in subjects who received five doses of 2.7×10^5 PfSPZ and were not fully protected (Figure 2B–D). Nonetheless, the proportional analysis suggests that homologous CHMI by DVI of a 100% infectious dose of homologous PfSPZ Challenge is at least as rigorous as a test of VE and potentially more rigorous than intense field transmission of heterogeneous Pf.

Vaccine-induced antibody and T-cell responses in the Tanzanians were lower than in malaria-naïve Americans who received the exact same dosage regimen. Two weeks after the last dose, the median antibody responses to PfCSP, the major protein on the surface of PfSPZ, were 4.3 times lower in the Tanzanians than those in Americans ($P = 0.0012$, Student's *t*-test, 2-tailed),[11] but significantly higher than in Malians who received the same immunization regimen ($P = 0.002$)[14] (Figure 3G).

The T-cell responses were also lower than in Americans[10,12] (Figure 4), but this could only be assessed in PBMCs from individuals who received the lower dose (five doses of 1.35×10^5 PfSPZ), not in the individuals who received the higher dose (five doses of 2.7×10^5 PfSPZ), the group that had sustained protection for 24 weeks. Thus, it is possible that had PBMCs from the higher dose group been assessed, responses would have been comparable to the responses in nonimmune Americans. The Tanzanians who were assessed had a significantly higher proportion of total memory T cells compared with total naïve T cells at the baseline than did the Americans. This higher frequency of memory cells compared with naïve cells may explain the lower immunogenicity due to less available naïve cells for expansion during the vaccinations. Moreover, the greater frequency of non-Pf–specific memory T cells may compete for infected cell contacts during pathogen surveillance.[33] These data suggest that PfSPZ Vaccine immunogenicity may be dependent on cumulative history of Pf exposure. Another explanation is that an activated immune microenvironment in the Tanzanians as compared with the Americans reduced immune responses.[34] Helminth infections have been associated with reduced immune responses to malaria[35]; however, the paucity of helminth infections in this population does not support helminth infection as a cause of the reduced immune responses.

There were no differences between vaccine and NS placebo recipients in regard to vaccine tolerability or AEs; 97.1% of the DVI administrations were rated painless and no volunteer experienced any local AE. Systemic AEs, most commonly headache, were mild, infrequent, and of short duration, with a similar frequency in NS controls as in vaccinees (no statistically significant differences in rates).

Among the controls, 16 of 18 were positive for Pf by TBS after CHMI. However, all 18 were positive by qPCR. This is consistent with findings in Gabon after CHMI.[21] It is likely that preexisting asexual blood stage immunity limits Pf replication in some individuals. Thus, they never reach the threshold for detection by

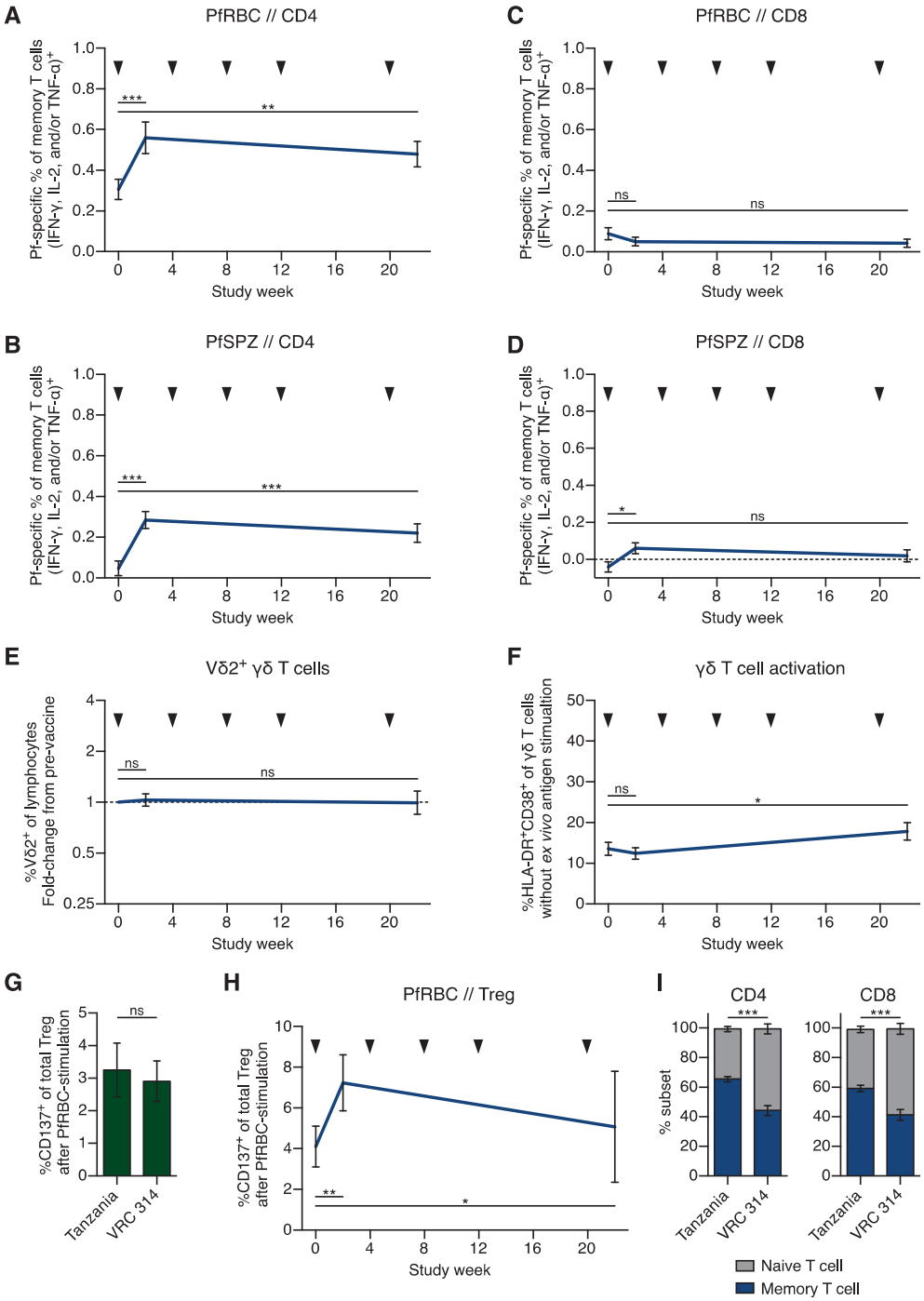


FIGURE 4. *Plasmodium falciparum* Sporozoites (PfSPZ)–specific T-cell responses in vaccine recipients receiving 1.35×10^5 PfSPZ. (**A–D**) PfSPZ-specific T-cell responses. Frequency of cytokine-producing memory CD4 T cells responding to (**A**) PfRBC or (**B**) PfSPZ. Throughout, "naïve T cell" refers to cells that co-express CCR7 and CD45RA, and "memory T cell" refers to all other T cells. Frequency of cytokine-producing memory CD8 T cells responding to (**C**) PfRBC or (**D**) PfSPZ. Results are the percentage of memory T cells producing interferon gamma, interleukin 2, and/or tumor necrosis factor alpha following stimulation minus the percentage of cells following control stimulation. (**E**) Frequency of the $V\delta2^+$ subfamily of γδ T cells of total lymphocytes. Results are expressed as fold-change from the prevaccine frequency. (**F**) γδ T-cell activation in vivo. Data are the percentage of memory γδ T cells expressing HLA-DR and CD38 as measured on PBMCs following incubation with control stimulation (vaccine diluent). (**G**) Prevaccine frequency of PfRBC-specific Tregs in Tanzania compared with malaria-naïve U.S. subjects from the Vaccine Research Center (VRC) 314 study. (**H**) Frequency of PfRBC-specific Treg. Results are the percentage of $CD4^+Foxp3^+CD25^+CD127^-$ T cells expressing CD137 (also known as 4-1BB) after stimulation with Pf red blood cell (PfRBC) minus the percentage of cells following stimulation with uninfected RBC. (**I**) Percentage of total CD4 (left) or CD8 (right) T cells that are naïve (gray bar; $CCR7^+CD45RA^+$) or memory (blue bar; not $CCR7^+CD45RA^+$) phenotype assessed prevaccination in all 48 subjects vaccinated in Tanzania or in 14 healthy U.S. subjects from the VRC 314 study.[13] For **A–F** and **H**, $N = 24$, and statistical difference was measured by using the Wilcoxon matched-pairs signed rank test. For **G** and **I**, statistical difference was measured by using the Mann–Whitney *U* test. *P* values are reported as not significant (ns), < 0.05 (*), < 0.01 (**), or < 0.001 (***). Data are mean ± SEM. Time points are prevaccine, 2 weeks after the first vaccination, and 2 weeks after the final vaccination. Black arrowhead designates PfSPZ Vaccine administration. This figure appears in color at www.ajtmh.org.

TBS. In our CHMI studies in Bagamoyo, we now use qPCR to confirm positive TBS, and retrospectively or in real time, assess parasitemia in all volunteers by qPCR.

We propose that increasing the numbers of PfSPZ per dose and altering intervals between doses will lead to overcoming the downregulation of humoral and cell-mediated immunity most likely because of previous exposure to Pf and thereby increase immune responses to PfSPZ Vaccine and VE. We also hypothesize that immune responses in younger, less malaria-exposed individuals will be of greater magnitude than those in adults.

Received December 27, 2017. Accepted for publication April 19, 2018.

Published online June 25, 2018.

Note: Supplemental tables and figure appear at www.ajtmh.org.

Acknowledgments: We would like to thank the study volunteers for their participation in the study. We also thank the entire study team at the Bagamoyo branch of the Ifakara Health Institute and the manufacturing, quality control, regulatory, clinical, and statistical teams at Sanaria, Inc. for their contributions to the conduct of this trial. We would also like to thank the members of the Safety Monitoring Committee, Kent Kester (chair), Peter Kisenge (safety monitor), and Alan Cross, for their thoughtful oversight. Clinicaltrials.gov Registration: NCT02132299.

Financial support: This clinical trial was supported by the Tanzanian Commission on Science and Technology (COSTECH), the Ifakara Health Institute, and the Swiss Tropical Public Health Institute. The development, manufacturing, and quality control release and stability studies of PfSPZ Vaccine and PfSPZ Challenge were supported in part by National Institute of Allergy and Infectious Diseases Small Business Innovation Research grant 5R44AI055229. Sanaria supported transport of PfSPZ Vaccine and PfSPZ Challenge to the study site and syringe preparation.

Disclosures: Sanaria Inc. manufactured PfSPZ Vaccine and PfSPZ Challenge, and Protein Potential LLC is affiliated with Sanaria. Thus, all authors associated with Sanaria or Protein Potential have potential conflicts of interest. There are no other conflicts of interest.

Authors' addresses: Said A. Jongo, Anneth Tumbo, Catherine Mkindi, Maxmillian Mpina, Ali T. Mtoro, Kamaka Ramadhani Kassim, Florence A. Milando, Munira Qassim, Omar A. Juma, Solomon Mwakasungula, Beatus Simon, Bakari M. Bakari, and Salim Abdulla, Bagamoyo Research and Training Centre, Ifakara Health Institute, Bagamoyo, Tanzania, E-mails: sjongo@ihi.or.tz, atumbo@ihi.or.tz, cmkindi@ihi.or.tz, mmpina@ihi.or.tz, amtoro@ihi.or.tz, kramadhani@ihi.or.tz, fmilando@ihi.or.tz, mqassim@ihi.or.tz, ojuma@ihi.or.tz, smwakasungula@ihi.or.tz, bbongole@ihi.or.tz, bbakari@ihi.or.tz, and sabdulla@ihi.or.tz. Seif A. Shekalaghe, District Commissioner, Maswa District, Simiyu, Tanzania, E-mail: sshekalaghe@yahoo.com. L. W. Preston Church, Adam J. Ruben, Eric R. James, Yonas Abebe, Natasha KC, Sumana Chakravarty, Elizabeth Saverino, Peter F. Billingsley, Thomas L. Richie, and Stephen L. Hoffman, Sanaria Inc., Rockville, MD, E-mails: lwpchurch@sanaria.com, aruben@sanaria.com, ejames@sanaria.com, yabebe@sanaria.com, nkc@sanaria.com, schakravarty@sanaria.com, esaverino@sanaria.com, pbillingsley@sanaria.com, trichie@sanaria.com, and slhoffman@sanaria.com. Tobias Schindler, Isabelle Zenklusen, Tobias Rutishauser, Julian Rothen, Claudia Daubenberger, and Marcel Tanner, Swiss Tropical and Public Health Institute (Swiss TPH), Basel, Switzerland, and University of Basel, Basel, Switzerland, E-mails: tobias.schindler@unibas.ch, isabelle.zenklusen@gmail.com, tobias.rutishauser@unibas.ch, julian.rothen@unibas.ch, claudia.daubenberger@unibas.ch, and marcel.tanner@unibas.ch. Andrew S. Ishizuka and Robert A. Seder, Vaccine Research Center (VRC), National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, E-mails: andrew.ishizuka@nih.gov and rseder@mail.nih.gov. B. Kim Lee Sim, Sanaria Inc., Rockville, MD, and Protein Potential LLC, Rockville, MD, E-mail: ksim@protpot.com.

REFERENCES

1. WHO, 2017. *World Malaria Report 2017*. Geneva, Switzerland: World Health Organization.
2. GBD 2015 Mortality and Causes of Death Collaborators, 2016. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet 388:* 1459–1544.
3. Gething PW et al., 2016. Mapping *Plasmodium falciparum* mortality in Africa between 1990 and 2015. *N Engl J Med 375:* 2435–2445.
4. Richie TL et al., 2015. Progress with *Plasmodium falciparum* sporozoite (PfSPZ)-based malaria vaccines. *Vaccine 33:* 7452–7461.
5. Clyde DF, Most H, McCarthy VC, Vanderberg JP, 1973. Immunization of man against sporozoite-induced falciparum malaria. *Am J Med Sci 266:* 169–177.
6. Rieckmann KH, Carson PE, Beaudoin RL, Cassells JS, Sell KW, 1974. Sporozoite induced immunity in man against an Ethiopian strain of *Plasmodium falciparum*. *Trans R Soc Trop Med Hyg 68:* 258–259.
7. Hoffman SL et al., 2002. Protection of humans against malaria by immunization with radiation-attenuated *Plasmodium falciparum* sporozoites. *J Infect Dis 185:* 1155–1164.
8. Hoffman SL et al., 2010. Development of a metabolically active, non-replicating sporozoite vaccine to prevent *Plasmodium falciparum* malaria. *Hum Vaccin 6:* 97–106.
9. Epstein JE et al., 2011. Live attenuated malaria vaccine designed to protect through hepatic $CD8^+$ T cell immunity. *Science 334:* 475–480.
10. Seder RA et al.; VRC 312 Study Team, 2013. Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. *Science 341:* 1359–1365.
11. Epstein JE et al., 2017. Protection against *Plasmodium falciparum* malaria by PfSPZ vaccine. *JCI Insight 2:* e89154.
12. Ishizuka AS et al., 2016. Protection against malaria at 1 year and immune correlates following PfSPZ vaccination. *Nat Med 22:* 614–623.
13. Lyke KE et al., 2017. Attenuated PfSPZ vaccine induces strain-transcending T cells and durable protection against heterologous controlled human malaria infection. *Proc Natl Acad Sci USA 114:* 2711–2716.
14. Sissoko MS et al., 2017. Safety and efficacy of PfSPZ vaccine against *Plasmodium falciparum* via direct venous inoculation in healthy malaria-exposed adults in Mali: a randomised, double-blind phase 1 trial. *Lancet Infect Dis 17:* 498–509.
15. Roestenberg M et al., 2013. Controlled human malaria infections by intradermal injection of cryopreserved *Plasmodium falciparum* sporozoites. *Am J Trop Med Hyg 88:* 5–13.
16. Sheehy SH et al., 2013. Optimising controlled human malaria infection studies using cryopreserved *P. falciparum* parasites administered by needle and syringe. *PLoS One 8:* e65960.
17. Hodgson SH et al., 2014. Evaluating controlled human malaria infection in Kenyan adults with varying degrees of prior exposure to *Plasmodium falciparum* using sporozoites administered by intramuscular injection. *Front Microbiol 5:* 686.
18. Mordmüller B et al., 2015. Direct venous inoculation of *Plasmodium falciparum* sporozoites for controlled human malaria infection: a dose-finding trial in two centres. *Malar J 14:* 117.
19. Shekalaghe S et al., 2014. Controlled human malaria infection of Tanzanians by intradermal injection of aseptic, purified, cryopreserved *Plasmodium falciparum* sporozoites. *Am J Trop Med Hyg 91:* 471–480.
20. Gomez-Perez GP et al., 2015. Controlled human malaria infection by intramuscular and direct venous inoculation of cryopreserved *Plasmodium falciparum* sporozoites in malaria-naive volunteers: effect of injection volume and dose on infectivity rates. *Malar J 14:* 306.
21. Lell B et al., 2018. Impact of sickle cell trait and naturally acquired immunity on uncomplicated malaria after controlled human malaria infection in adults in Gabon. *Am J Trop Med Hyg 98:* 508–515.
22. Kamau E, Alemayehu S, Feghali KC, Saunders D, Ockenhouse CF, 2013. Multiplex qPCR for detection and absolute quantification of malaria. *PLoS One 8:* e71539.

23. Hofmann N, Mwingira F, Shekalaghe S, Robinson LJ, Mueller I, Felger I, 2015. Ultra-sensitive detection of *Plasmodium falciparum* by amplification of multi-copy subtelomeric targets. *PLoS Med 12:* e1001788.
24. Anderson TJ, Su XZ, Bockarie M, Lagog M, Day KP, 1999. Twelve microsatellite markers for characterization of *Plasmodium falciparum* from finger-prick blood samples. *Parasitology 119:* 113–125.
25. Mordmuller B et al., 2017. Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. *Nature 542:* 445–449.
26. Lamoreaux L, Roederer M, Koup R, 2006. Intracellular cytokine optimization and standard operating procedure. *Nat Protoc 1:* 1507–1516.
27. Haddy TB, Rana SR, Castro O, 1999. Benign ethnic neutropenia: what is a normal absolute neutrophil count? *J Lab Clin Med 133:* 15–22.
28. Hammerschmidt DE, 1999. It's as simple as black and white! Race and ethnicity as categorical variables. *J Lab Clin Med 133:* 10–12.
29. Schofield L, Villaquiran J, Ferreira A, Schellekens H, Nussenzweig RS, Nussenzweig V, 1987. Gamma interferon, CD8+ T cells and antibodies required for immunity to malaria sporozoites. *Nature 330:* 664–666.
30. Weiss WR, Sedegah M, Beaudoin RL, Miller LH, Good MF, 1988. CD8+ T cells (cytotoxic/suppressors) are required for protection in mice immunized with malaria sporozoites. *Proc Natl Acad Sci USA 85:* 573–576.
31. Weiss WR, Jiang CG, 2012. Protective CD8+ T lymphocytes in primates immunized with malaria sporozoites. *PLoS One 7:* e31247.
32. Schoenbrunn A et al., 2012. A converse 4-1BB and CD40 ligand expression pattern delineates activated regulatory T cells (Treg) and conventional T cells enabling direct isolation of alloantigen-reactive natural Foxp3+ Treg. *J Immunol 189:* 5985–5994.
33. Fernandez-Ruiz D et al., 2016. Liver-resident memory CD8+ T cells form a front-line defense against malaria liver-stage infection. *Immunity 45:* 889–902.
34. Muyanja E et al., 2014. Immune activation alters cellular and humoral responses to yellow fever 17D vaccine. *J Clin Invest 124:* 3147–3158.
35. Hartgers FC, Yazdanbakhsh M, 2006. Co-infection of helminths and malaria: modulation of the immune responses to malaria. *Parasite Immunol 28:* 497–506.
36. Purkins L, Love ER, Eve MD, Wooldridge CL, Cowan C, Smart TS, Johnson PJ, Rapeport WG, 2004. The influence of diet upon liver function tests and serum lipids in healthy male volunteers resident in a phase I unit. *Br J Clin Pharmacol 57:* 199–208.

Chapter 3

Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period

---

This chapter contains the following publication:

Julian Rothen, Carl Murie, Jason Carnes, Atashi Anupama, Salim Abdulla, Mwajuma Chemba, Maxmillian Mpina, Marcel Tanner, B. Kim Lee Sim, Stephen L Hoffman, Raphael Gottardo, Claudia Daubenberger and Ken Stuart. "Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period". 2018. *PLOS One*.

PLOS ONE

RESEARCH ARTICLE

Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period

Julian Rothen[1,2]*, Carl Murie[3], Jason Carnes[4], Atashi Anupama[4], Salim Abdulla[5], Mwajuma Chemba[5], Maxmillian Mpina[1,2,5], Marcel Tanner[1,2], B. Kim Lee Sim[6], Stephen L. Hoffman[6], Raphael Gottardo[3☯], Claudia Daubenberger[1,2☯], Ken Stuart[4☯]

1 Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland, **2** University of Basel, Basel, Switzerland, **3** Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **4** Center for Infectious Disease Research, Seattle, Washington, United States of America, **5** Bagamoyo Research and Training Centre, Ifakara Health Institute, Bagamoyo, Tanzania, **6** Sanaria Inc., Rockville, Maryland, United States of America

☯ These authors contributed equally to this work.
* julian.rothen@unibas.ch

Citation: Rothen J, Murie C, Carnes J, Anupama A, Abdulla S, Chemba M, et al. (2018) Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period. PLoS ONE 13(6): e0199392. https://doi.org/10.1371/journal.pone.0199392

Editor: Delmiro Fernandez-Reyes, University College London, UNITED KINGDOM

Received: January 9, 2018

Accepted: May 29, 2018

Published: June 19, 2018

Data Availability Statement: The Illumina read and normalized count data generated in this study were released to the Gene Expression Omnibus (GEO) database under accession number GSE97158.

Funding: This work was supported by the National Institutes of Health - Human Immunology Project Consortium grant (number: U19AI089986) to KS. Authors KLS and SH are affiliated with the commercial company Sanaria Inc., the manufacturer of PfSPZ Challenge. The funders

Abstract

Malaria continues to be one of mankind's most devastating diseases despite the many and varied efforts to combat it. Indispensable for malaria elimination and eventual eradication is the development of effective vaccines. Controlled human malaria infection (CHMI) is an invaluable tool for vaccine efficacy assessment and investigation of early immunological and molecular responses against *Plasmodium falciparum* infection. Here, we investigated gene expression changes following CHMI using RNA-Seq. Peripheral blood samples were collected in Bagamoyo, Tanzania, from ten adults who were injected intradermally (ID) with $2.5x10^4$ aseptic, purified, cryopreserved *P. falciparum* sporozoites (Sanaria® PfSPZ Challenge). A total of 2,758 genes were identified as differentially expressed following CHMI. Transcriptional changes were most pronounced on day 5 after inoculation, during the clinically silent liver phase. A secondary analysis, grouping the volunteers according to their prepatent period duration, identified 265 genes whose expression levels were linked to time of blood stage parasitemia detection. Gene modules associated with these 265 genes were linked to regulation of transcription, cell cycle, phosphatidylinositol signaling and erythrocyte development. Our study showed that in malaria pre-exposed volunteers, parasite prepatent period in each individual is linked to magnitude and timing of early gene expression changes after ID CHMI.

provided support in the form of salaries for authors KLS, SH and KS, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: Authors KLS and SH are affiliated with the commercial company Sanaria Inc., the manufacturer of PfSPZ Challenge. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Introduction

Malaria caused by *Plasmodium falciparum* continues to be one of mankind's most devastating infectious diseases despite the many and varied efforts to combat it. It has been eliminated in certain areas of the world by combination of treatment with effective drugs, e.g. chloroquine, and by large scale vector control programs, e.g. through insecticide spraying and insecticide-treated nets, only to resurge as a result of drug and insecticide resistance. In 2016, there was an estimated number of 445,000 deaths related to malaria, the overwhelming majority (90%) occurring in the WHO African Region [1].

An effective malaria vaccine would be a powerful tool for regional elimination and eventual eradication of malaria. Currently the most advanced malaria vaccine candidate is RTS,S/AS01, for which large-scale clinical evaluation in African countries has demonstrated vaccine efficacy against clinical malaria of 34% during the 20 months following dose 1 in children aged 5–17 months [2]. Experimental vaccines comprised of live attenuated *P. falciparum* sporozoites have gained increased attention because they are highly effective in providing sterile immunity, i.e. immunity to infection [3–13]. Such vaccines primarily targeting the pre-erythrocytic stage are safe because development of the parasite is arrested before, during or shortly after the liver stage, hence prior to the blood stage during which malaria disease symptoms occur. Several approaches aiming to determine the optimal design and administration mode of such a vaccine are being pursued. Promising results have been obtained in studies using radiation-attenuated sporozoites that were administered by either direct intravenous inoculation [3,5–8] or mosquito bite [12,13], genetically attenuated sporozoites [10,11], or inoculation of volunteers with fully infectious sporozoites under coverage with an anti-malarial drug [4,9]. Besides their application as potential anti-malaria vaccine candidates, aseptic, purified, cryopreserved, whole infectious sporozoites are useful in controlled human malaria infection (CHMI) studies. Targeted infection of volunteers in a controlled environment enables the clear and efficient assessment of vaccine efficacy [14–16], aids the development of anti-malarial drugs [17], and is useful for studying human immune responses to malaria infection [18]. The latter is of particular importance given that we still lack a detailed understanding of the host responses to early stages of *P. falciparum* infection.

To overcome aforementioned gaps, high-throughput transcriptome analyses employing microarray and/or RNA-Seq can be valuable. Both technologies have already been used for gene expression profiling of malaria-naïve subjects undergoing anti-malaria vaccination and/or CHMI [19–24], malaria pre-exposed subjects undergoing natural *P. falciparum* infection [21,22,25] and the *Plasmodium* parasite itself [25,26]. Collectively, such studies contribute to a more comprehensive understanding of molecular patterns and cell signatures involved in the interaction of the human host with malaria.

Here, we aimed to investigate human transcriptional dynamics during *P. falciparum* liver and early asexual blood stage with data from a CHMI study conducted in Bagamoyo, Tanzania in 2014, as described by Shekalage et al. [27]. We investigated the transcriptional responses by RNA-Seq analysis of whole blood from 10 adults from malaria endemic regions following CHMI by intradermal inoculation of PfSPZ Challenge, the first such CHMI ever carried out in malaria pre-exposed adults. Our results add insights into gene pathways and associated molecular functions elicited by the *P. falciparum* parasite in malaria-experienced subjects as well as important findings regarding the interplay between differential expression magnitude and malaria asexual parasite prepatent period at an individual level.

Results

DE genes shared among subjects after CHMI

Limma linear modeling was applied to normalized and voom transformed sequence count data to assess temporal gene expression level changes in response to infection with sporozoites. Pairwise comparison of samples collected at baseline, day 5, day 9 and day 28 post CHMI, allowed us to assess the direction and extent of expression changes at the different study visits (Fig 1A). Setting the baseline transcriptional level as a comparator, a multitude of genes were differentially expressed in the blood at day 5 (5/0) and day 28 (28/0) after CHMI. Remarkably, gene expression levels recorded at day 9 post CHMI (9/0) did not differ significantly from the baseline levels. However, extended pairwise comparative analyses revealed substantial numbers of DE genes on day 9 and day 28 relative to day 5 (9/5 and 28/5) and day 28 relative to day 9 (28/9). Most of the DE genes at day 5 (749) were expressed at lower levels in the blood relative to baseline with fewer genes (226) expressed at relatively higher levels. The opposite is true at day 28, when more genes had higher (378) rather than lower levels of expression (88) relative to baseline. The greatest number of DE genes was observed at comparison 9/5 (1,536 genes up, 421 down). Similarly, albeit to a lesser extent, on day 28, 893 genes had increased and 128 genes had decreased expression levels relative to day 5 and 209 genes had increased and 97 had decreased expression levels relative to day 9. Not surprisingly, a significant number of genes were differentially expressed in multiple comparisons. For example, there was a large overlap between the DE genes determined for comparisons 5/0 and 9/5 (Fig 1B). Many of the up-regulated DE genes at 5/0 were down-regulated at 9/5 (Fig 1C) and similarly, the majority of down-regulated DE genes at 5/0 were up-regulated at 9/5 (Fig 1D). Combined, the six pairwise comparisons identified a total of 2,758 unique genes or 16.7% of the total 16,473 genes contained in the data set that were differentially expressed. A list containing the DE genes and their direction of change for each tested contrast is provided in the supplementary section of this manuscript (S1 File).

GSEA detects DE trends across all genes

Gene set enrichment analysis (GSEA) generated a picture of progression of differential expression over 28 days following CHMI. This analysis incorporated all 16,473 genes in the dataset and ranked the genes in terms of differential expression. GSEA accounts for subtle fold expression changes and simultaneous increased and decreased DE genes in a given gene module [28]. This allowed us to also identify gene dynamics for the 9/0 contrast, despite the absence of DE genes at $>$ 1.5 fold expression changes for this comparison.

GSEA identified several blood transcriptome modules (BTMs) [29] whose expression levels were decreased at comparison 5/0 (Fig 2). These were linked to modules for ubiquitination (M138), transcription factors (M213), and inositol phosphate pathways (M101, M129) as well as cell cycle and intracellular transport (M143, M144, M147, M230, M237). Among BTMs that appeared up-regulated for contrast 5/0 were modules linked to the CORO1A-DEF6 network (M32.2, M32.4), platelet activation (M32.0, M32.1), regulation of localization (M63), signaling events (M100, M215) as well as processes in translation (M245) and transcription (M32.3, M234).

Interestingly, DE modules identified for contrasts 9/5 and 28/5 had a largely reciprocal pattern of expression compared to contrast 5/0. Two BTMs with decreased relative expression at contrast 9/5 that were not detected as DE at contrast 5/0 were associated with blood coagulation (M11.1) and cytoskeletal remodeling (M32.8). The latter BTM was also down-regulated at contrast 28/5. BTMs with no significant change at contrast 5/0 that appeared increased at both

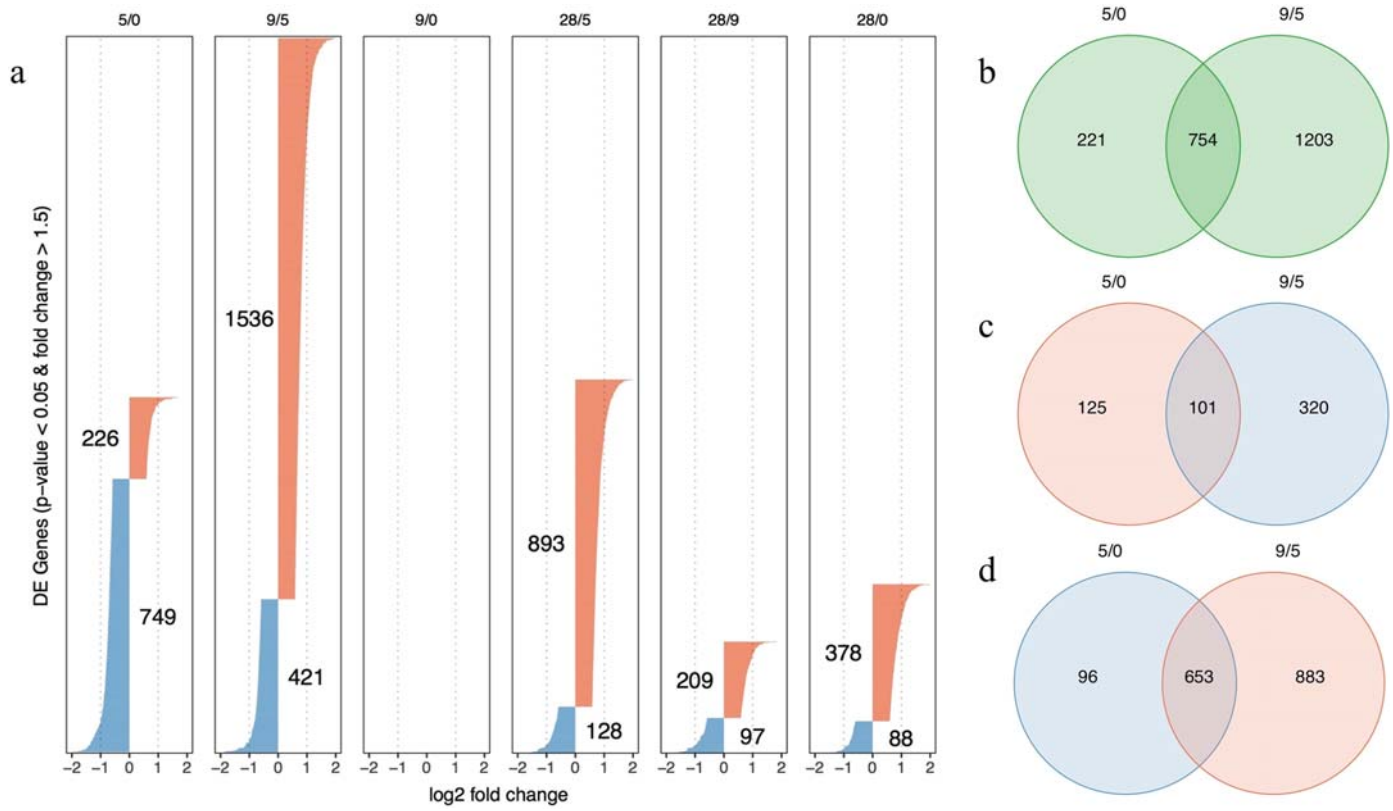


Fig 1. DE genes determined by limma pairwise visit comparison. DE was pronounced at a BH-adjusted p-value < 0.05 and >1.5 fold expression change. (a) DE genes (red: up-regulated genes, blue: down-regulated genes) identified for each tested contrast are visualized as bars. The number of DE genes per contrast is indicated and additionally emphasized by the length of the bars. The bar width / x-axis indicates the log2 fold expression change of each DE gene. The Venn diagrams display the overlaps between (b) all DE genes of contrasts 5/0 and 9/5, (c) the 5/0 up-regulated and 9/5 down-regulated DE genes and (d) the 5/0 down-regulated and 9/5 up-regulated DE genes.

https://doi.org/10.1371/journal.pone.0199392.g001

contrasts 9/5 and 28/5 were linked to the proteasome (M226) and mitosis transcription factors (M169).

Amongst other BTMs with higher expression levels at contrast 28/5 were modules linked to erythrocyte differentiation and heme biosynthesis (M173, M171, M222). A similar trend in increased expression of red blood cell (RBC) related BTMs was observed for contrasts 9/0, 28/9 and 28/0 as well. Similarly, modules linked to the mitochondrial electron transport chain (M216, M219, M231, M238) and translation and transcriptional processes (M234, M245) were up-regulated with increasing magnitude at contrasts 9/0, 28/9 and 28/0.

Despite not having detected any DE genes for the 9/0 contrast in the first round of our analysis, GSEA revealed a variety of BTMs being differentially expressed at day 9 relative to baseline. In addition to the aforementioned DE modules at contrast 9/0, BTMs linked to ubiquitination (M138), cell cycle (M144), mitosis (M169) and most pronounced, to the proteasome (M226) were identified as up-regulated compared to baseline. The down-regulated modules at 9/0 comparison largely corresponded to the negatively enriched BTMs of the 9/5 and 28/5 comparisons, with the exception of one module linked to cell junction (M4.13). Lastly, contrast 28/9 showed, with exception to the already mentioned modules, DE patterns similar to contrast 5/0. BTMs linked to blood coagulation, cytoskeletal remodeling and cell junction were found to be positively enriched exclusively for the 28/9 contrast.

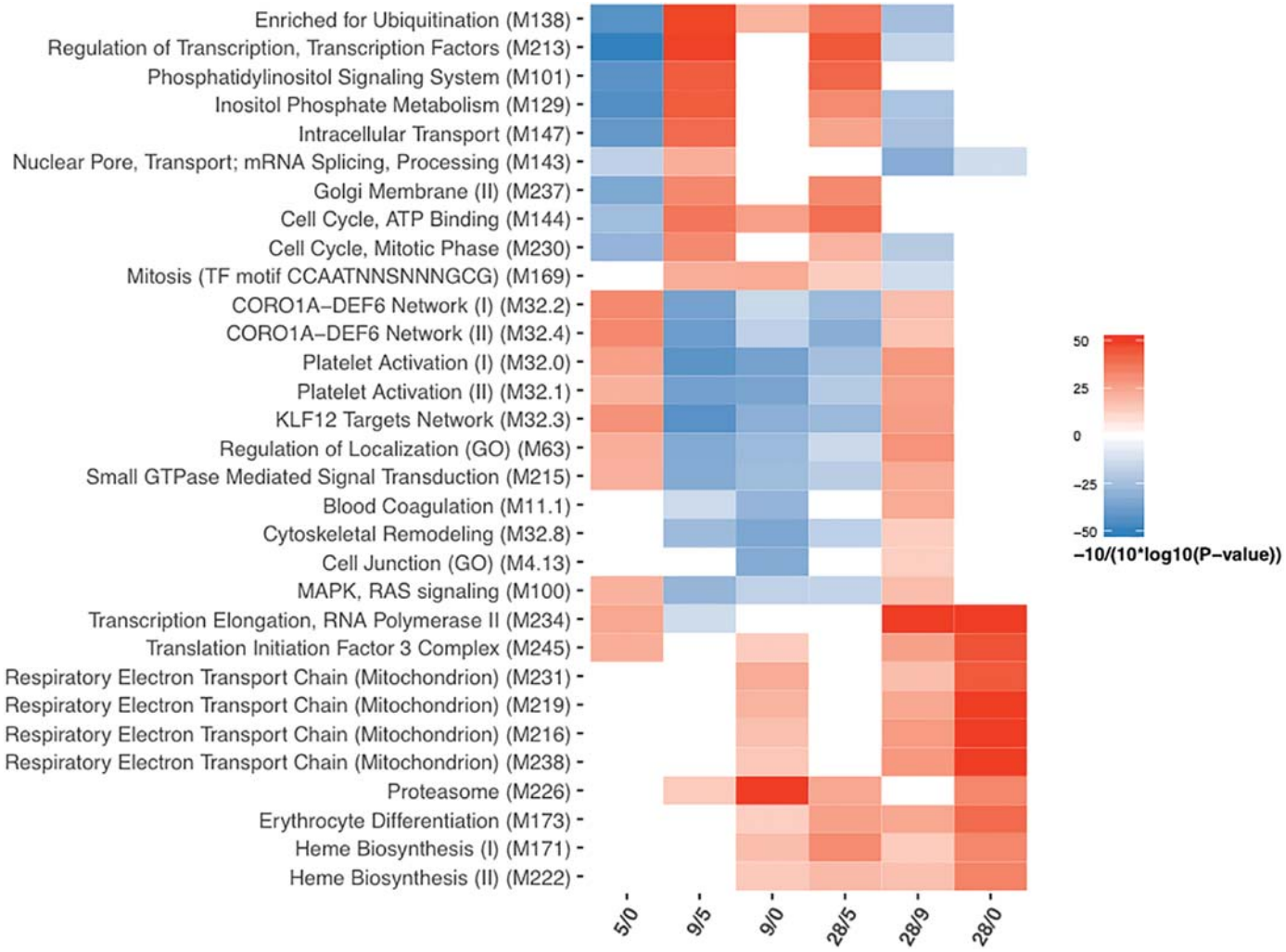


Fig 2. GSEA using camera (limma), visualized as heatmap. Statistical significance is pronounced at a p-value and FDR < 0.05. Red: up-regulated, blue: down-regulated.

https://doi.org/10.1371/journal.pone.0199392.g002

As a supporting analysis, we repeated the competitive GSEA, using gene sets designed by Chaussabel et al. [30] that incorporate larger numbers of genes per set when compared to the BTMs (S1 Fig). As an additional ancillary analysis, we applied hypergeometric gene set testing, testing for overlaps between the DE genes and BTMs or Chaussabel defined sets (S2 and S3 Figs). For both analyses the identified sets were largely congruent with our initial results using GSEA and BTMs. Additional Chaussabel sets detected were linked to the myeloid lineage and monocyte development (down-regulated at 9/5). In addition, gene sets linked to CD4 cell division and cell cycle (up-regulated at 9/5) and NK cell development and cytotoxicity (down-regulated 5/0) were seen.

DE gene dynamics and linkage to blood stage parasitemia

Examining the expression dynamics of the 2,758 DE genes determined in the first part of the study, it became evident that the expression patterns varied not only between different visits but also greatly between volunteers. These expression dynamics are visualized as heatmap in

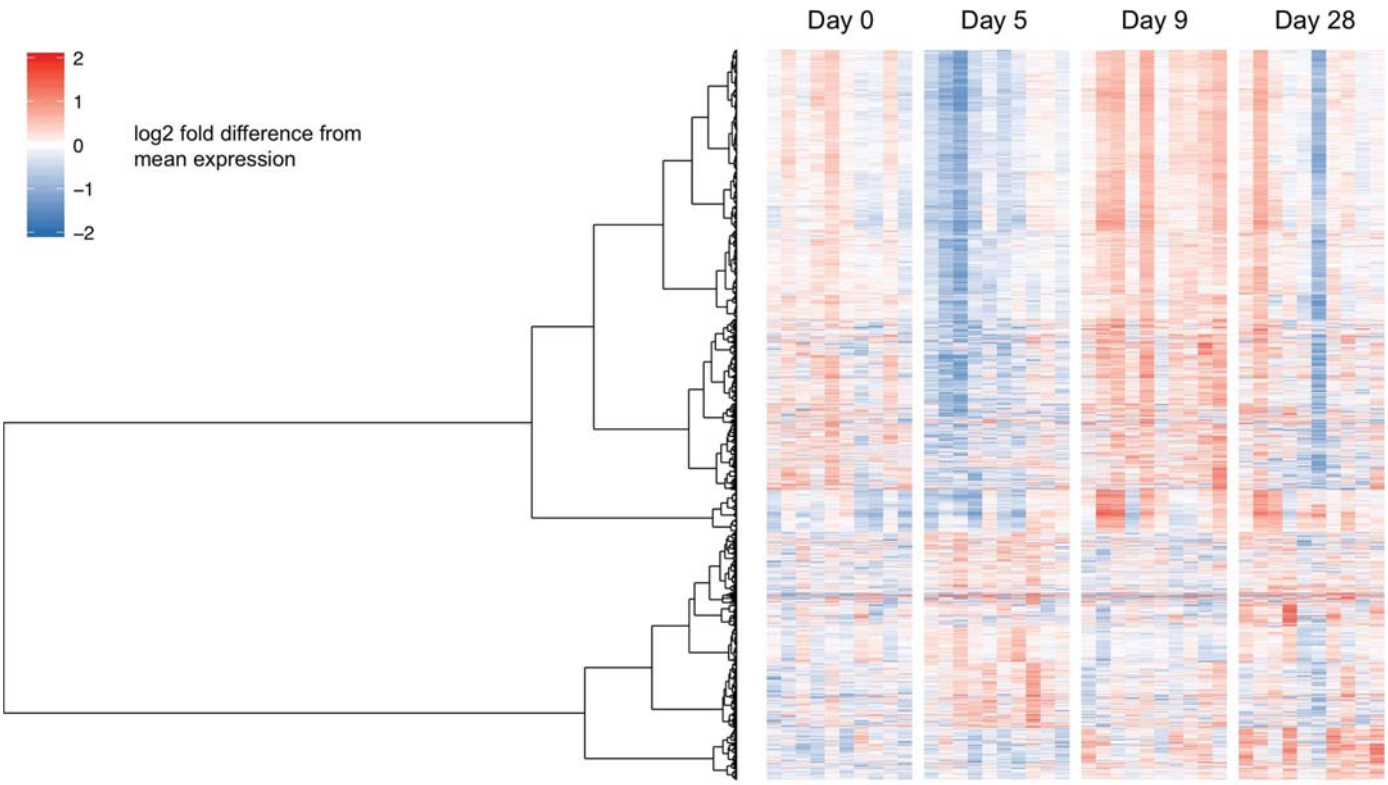


Fig 3. Gene expression patterns in relation to time to detection of blood stage parasitemia. The heatmap displays expression levels of 2,758 DE genes (rows), as determined by limma pairwise comparison of all visits. Subjects (columns) are ordered by increasing pre-patent periods of blood stage parasitemia. Log2 transformed raw counts were centered gene-wise by subtracting the corresponding mean expression values. For visualization purposes, the expression values were limited to 2 and -2. The dendrogram indicates hierarchical clustering of the DE genes based on the Ward method. Two major clusters among the DE genes were identified.

https://doi.org/10.1371/journal.pone.0199392.g003

Fig 3, alongside a dendrogram grouping the DE genes in two major clusters. Ordering the heatmap columns based on increasing individual prepatent period, indicated that the majority of DE genes located in the larger cluster seemed to follow a distinct pattern regarding magnitude and direction of expression changes. Primarily at day 5, subjects with a short prepatent period displayed an overall stronger down-regulation of DE genes than subjects with a moderate or long prepatent period. The pronounced down-regulation of genes from the larger cluster at day 28 in one of the subjects was most likely a technical artifact (RIN score of 5.2). This might be a quality issue but nevertheless did not affect the statistical analyses conducted here. We performed both limma linear modeling and competitive GSEA on a reduced set of samples, removing all four samples of two volunteers with the low RIN score samples. Not surprisingly, the number of DE genes determined for the different contrasts were slightly changed, with the ratio between up- and down-regulated DE genes remaining stable. Importantly, this did not influence the GSEA outcome, with identical gene sets being identified as before when analyzing the complete sample set.

Next, we grouped all subjects according to early (9, 9.5, 10), average (11, 11, 11, 11, 11.5, 12) and late (16 days) appearance of blood stage parasitemia measured by qPCR (S4 Fig). Using limma, we performed an F-test to test for differences in temporal expression changes across the three groups. This analysis identified a group of 265 genes linked to parasitemia (S2 File).

Hypergeometric testing revealed significant overlaps of the 265 DE genes with BTMs linked to regulation of transcription factors (M179, M213), phosphatidylinositol signaling (M101),

cell cycle (M144), intracellular transport (M147), ubiquitination (M138) as well as Chaussabel gene sets linked to erythrocyte development (M2.3) and inflammatory processes (M138) (S2 File). Among these BTMs and Chaussabel gene sets, the magnitude of DE gene change was most strongly affected by time to blood stage parasitemia for the 5/0 comparison (Figs 4 and 5). The three subjects (early group) that were within a time window of 4–5 days between day 5 blood collection and parasite detection displayed the strongest down-regulation of genes. Time window differences of 6–7 days (average group; six subjects) or 11 days (one late subject) between day 5 and blood stage parasitemia detection, respectively, correlated with reduced changes to gene expression. Similarly, many genes of the erythrocyte development (M2.3) set displayed increased expression levels in two of the three early subjects already at day 5. By day 9, all other subjects displayed uniform up-regulation of these genes, with the late subject showing the least dynamics.

We assumed that the early and average groups with more robust sample sizes of 3 and 6 volunteers were the main drivers for the here reported results and that the late group with only 1 volunteer had only a weak effect in the statistical model. In order to confirm this, we repeated the limma linear modeling without the late parasitemia subject, comparing only early vs. average subjects. This analysis produced similar results as before, showing an even higher overall number of DE genes (365). Subsequent hypergeometric testing produced the same significant DE gene sets linked to parasitemia. Taken together, this confirms our reported results are not driven by the single late subject but by comparison of the early and average volunteers.

Gene expression changes in relation to leukocyte population frequencies

In order to rule out that the observed transcriptional dynamics were driven by proportional changes in major cell populations, we also integrated hematology data generated by Shekalage et al. [27] into our analysis. We could indeed observe changes of the leukocyte populations in our study population. This is in line with a recent study by Wolfswinkel et al. that reports changes in total and differential leukocyte counts during the clinically silent liver phase in a controlled human malaria infection in malaria-naïve Dutch volunteers [31]. In our study, a statistically significant increase of lymphocytes, neutrophils and monocytes was observed during the early liver phase of the infection at day 5 (S5 Fig). The increase of neutrophils was even more pronounced at day of parasitemia (e.g. time point of first positive microscopy thick smear). In contrast, lymphocytes numbers were reduced at day of parasitemia. We integrated these WBC dynamics with our limma linear model, investigating whether the magnitude of increase in the leukocyte populations at day 5 correlated with the magnitude of gene expression changes at the same time point. We found no statistical evidence linking the magnitude of cell change to the individual magnitude of gene expression change. WBC changes may certainly influence gene expression patterns but were not the driving force for the subject to subject differences in transcriptional dynamics reported in this study.

Discussion

We report here for the first time whole blood transcriptome changes over 28 days following intradermal CHMI with aseptic, purified, cryopreserved, infectious PfSPZ in malaria-experienced subjects. Transcriptional changes of hundreds of genes that had increased or decreased relative to their expression levels between days 0 (day of infection), 5, 9 and 28 were identified. Unlike comparable studies that investigated transcriptional responses following vaccination and/or mosquito bite challenge [19–23] this study examined subjects who were infected by intradermal injection with malaria parasites. It also focused on malaria experienced Tanzanian adults who over the course of their life had been repeatedly exposed to *Plasmodium* parasites.

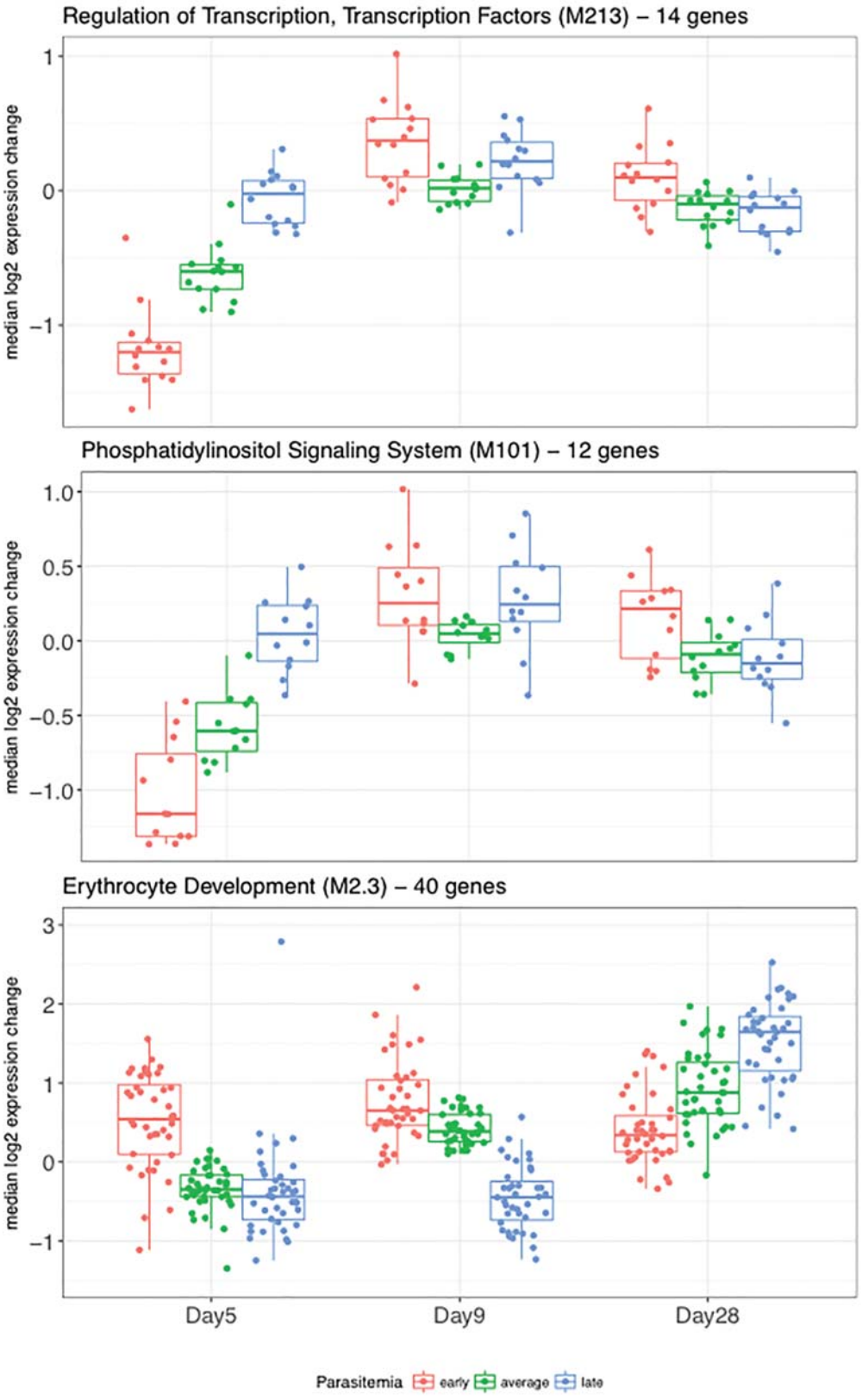

Regulation of Transcription, Transcription Factors (M213) – 14 genes
median log2 expression change
1
0
−1
Phosphatidylinositol Signaling System (M101) – 12 genes
median log2 expression change
1.0
0.5
0.0
−0.5
−1.0
Erythrocyte Development (M2.3) – 40 genes
median log2 expression change
3
2
1
0
−1
Day5
Day9
Day28
Parasitemia
early
average
late

Fig 4. Volunteer gene expression trends visualized as boxplots. Gene expression trends are shown for two differentially expressed BTMs and one gene set linked to parasitemia. Boxplots with gene-wise baseline-subtracted expression values are shown separately for subjects with early (red), average (green) and late (blue) detection of blood stage parasitemia.

https://doi.org/10.1371/journal.pone.0199392.g004

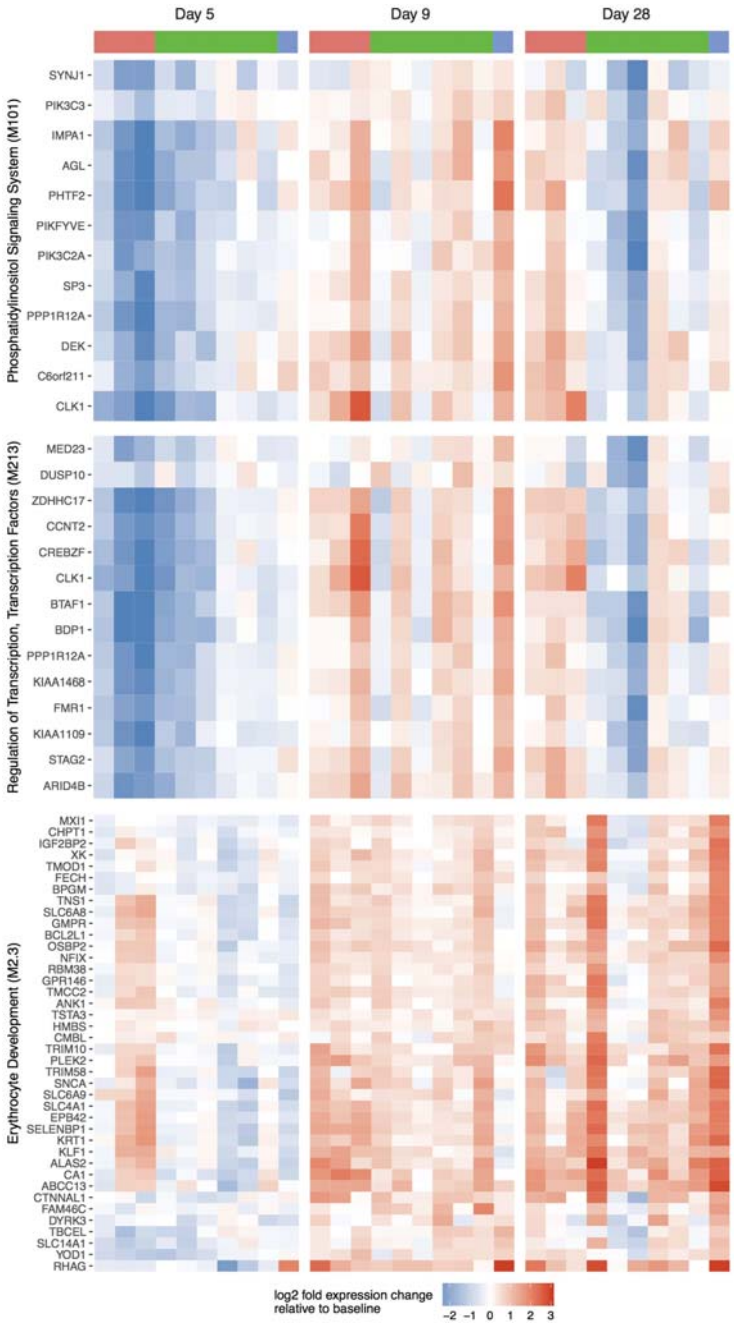


Fig 5. Volunteer gene expression trends visualized as heatmaps. Gene expression trends are shown for two differentially expressed BTMs and one gene set linked to parasitemia. Gene-wise expression levels of the Fig 4 DE modules are visualized as heatmap. Column color bars indicate grouping of subjects into early (red), average (green) and late (blue) group, based on time point of parasitemia detection. Each row corresponds to one gene.

https://doi.org/10.1371/journal.pone.0199392.g005

It has been shown that such individuals' immunological responses to CHMI are different from those of malaria-naïve subjects. In Tanzanians, a stronger humoral immune response was recalled after CHMI when compared to Dutch volunteers undergoing identical challenge conditions [32]. Naturally acquired immunity in Kenyans strongly impacts on parasite multiplication rate observed after CHMI, calling for qPCR based molecular monitoring tools in addition to blood slide microscopy for parasite detection [33].

Our study aimed to provide comprehensive insights into early host transcriptional responses occurring during the pre-erythrocytic developmental stage. Since this stage is clinically silent, it is only possible to be studied during a CHMI with parasite strain and infectious doses defined. Therefore, we collected whole blood samples at days 0, 5 and 9 after intradermal CHMI covering this under-researched, early infection period.

A surprising finding of our study was the modest transcriptional changes recorded at day 9 relative to baseline. Although several DE gene sets were later identified by GSEA for this contrast (Fig 2), the absolute expression level changes on gene level remained all below the DEG threshold (1.5 fold change) and were therefore not picked up by the initial limma pairwise analysis. This was unexpected since day 9 is the time point when parasite transition from the liver to the blood occurs in the first volunteers. This finding supports the hypothesis that the timing of whole blood collection as well as the inclusion of early time points (before day 9) during the clinically silent liver stage needs to be targeted in order to optimally capture transcriptional signals. Also noteworthy are the 400 DE genes identified at day 28 post CHMI relative to baseline (Fig 1A). These changes in expression levels cannot be attributed to the effect of CHMI only. We assume that these changes are the combined result of infection and treatment resulting in parasite clearance and development of cellular immune responses.

Importantly, our studies show that significant changes in transcriptional patterns are already observed on day 5—a time point before parasites reach the blood. This is the time period during which an unknown proportion of the injected PfSPZ have infected liver cells and are rapidly developing into thousands of merozoites. High variation (ranging from day 9 to day 16) of parasite prepatent period measured by qPCR strongly indicates that the load of parasites egressing from the liver varied between individuals. Ultimately, this first wave of malaria parasites determines how rapidly asexual blood stage parasites amass to cross microscopy detection threshold (ranging from day 11.5 to day 19) resulting in anti-malaria treatment [34]. The time between malaria infection and microscopic detection, e.g. the prepatent period, has been shown to be associated with degree of malaria pre-exposure. Volunteers in the Tanzanian CHMI-ID trial, including the 10 subjects analyzed in our study, displayed significantly longer prepatent periods than malaria-naïve Dutch volunteers who underwent a similar CHMI study [32]. Pre-existing immunity in the Tanzanian cohort was evident, with more than 50% of the Tanzanian volunteers having a positive *P. falciparum* lysate serology at baseline. Similarly, antibody titers for the *P. falciparum* antigens CSP, LSA-1, EXP-1, and AMA-1 or preexisting *P. falciparum*-specific IFN-γ responses were reported. Importantly, none of these markers for level of pre-existing immunity was associated with the observed differences in prepatency, suggesting that other immunological parameters need to be assessed as well [32].

Based on the prepatent period, we segregated our volunteers into three groups, namely early (n = 3), average (n = 6) and late (n = 1). Interestingly, the length of the prepatent period is reflected by the extent of observed transcriptional changes in peripheral blood. From the time point of infection across all time points, we identified a total of 265 DE genes (S2 File) whose expression level dynamics correlated statistically with time to parasitemia. The majority of these genes were DE around day 5 (overlap with contrast 5/0 DE genes of 88 and 190 with

DE genes of contrast 9/5). This was expected since a majority of DE genes and associated BTMs were identified already by the pairwise visit comparisons for the day 5 contrasts.

Notably, our observation of early gene expression changes is in line with a recent study by Kazmin et al. who reported DE genes in response to mosquito bite challenge as early as day 1 and day 5 after infection [20]. Among differentially expressed gene modules correlating with time to asexual blood stage parasitemia detection at contrast 5/0 were two BTMs linked to regulation of transcription and phosphatidylinositol signaling (Figs 4 and 5, S2 File). Genes contained in these modules displayed in unison stronger down-regulation in volunteers with early to average time to blood stage parasitemia. The trends observed in these BTMs are representative of patterns seen in several other modules such as M5.1 (inflammation), M138 (enriched for ubiquitination), M144 (cell cycle, ATP binding) and M147 (intracellular transport) and M179 (enriched for TF motif PAX3). A similar, although reciprocal pattern was observed for genes belonging to gene set M2.3, linked to erythrocyte development (Figs 4 and 5, bottom panel). Genes of this set displayed increased expression levels in two of the three early subjects already at day 5. By day 9, all other subjects displayed uniform up-regulation of these genes, with the late subject showing the least dynamics.

Combined, our observations of individual´s prepatent period interlinked to the magnitude of differential expression on day 5 strongly suggest that blood collection timing is critical and should be conducted at more frequent intervals, additionally covering early time points between days 1 to 4 post CHMI. Capturing time points with the highest transcriptional expression changes might depend on the size of the parasite load multiplying in the liver. Similar studies involving malaria naïve volunteers without pre-existing immunity and with more uniform prepatent periods would shed more light on this hypothesis.

Some of the BTMs identified here as DE have been reported in similar studies that investigated transcriptional responses following controlled infection with *P. falciparum* or vaccination. We can only draw limited conclusions when comparing our results with these studies, given the differences in study participants (malaria-naïve or vaccinated vs. pre-exposed subjects), challenge model (mosquito bite vs. intradermal injection) and time point of gene expression assessment. However, there are some interesting parallels to our results: The up-regulation of genes in the proteasome module observed strongest at day 9 and significant at day 28 post CHMI has been reported in response to candidate malaria vaccines TRAP and RTS,S [19,23]. The proteasome is known to play a key function in MHC protein processing and antigen presentation [23], the genes in this module could therefore be of special interest regarding the development of adaptive immune responses against *P. falciparum*. The study by Dunachie et al. [19] further reported the antigen processing and presentation pathway and phosphotidylinositol signaling system to be key modules invoked by antigen stimulation after vaccination and the latter to be correlated with time to parasitemia in subsequent challenge by mosquito bite [19]. Interestingly, in our case of intradermal PfSPZ CHMI, we found this pathway to be negatively correlated at day 5. The up-regulation of genes in the MAPK RAS signaling module is an interesting parallel to a finding of Ockenhouse et al., who reported activation of MAP kinases by natural acquired *P. falciparum* infection. The same study reported over expression of genes linked to the GO term "protein ubiquitination"following mosquito bite challenge of malaria-naïve subjects. This is an interesting parallel to our BTM linked to ubiquitination that was found up-regulated at 9/0 and down-regulated at 5/0 [21]. Cell cycle related modules have been reported to be affected after RTS,S vaccination and homologous challenge. Interestingly, the same study reported enrichment of genes in NK and monocyte pathways following vaccination and homologous challenge [20].

Studies in malaria mouse models have revealed that liver stage infection results in accumulation of NKT and NK cells in liver tissue and that these cell subsets are involved in parasite

protective immune responses [35]. In gene modules defined by the hypergeometric overlap testing, we found that two gene modules (M7.2 and S1) associated with NK cell biology are down-regulated on day 5. These data could indicate that in humans NK cell subsets are recruited from peripheral blood into the liver during the pre-erythrocytic stage infection (S2 Fig).

It should be noted, that the classification of gene and/or module expression change in up- or down-regulated as reported in our study, might not always be the best way of describing the underlying biological or cellular dynamics. By using peripheral blood as starting material for mRNA extraction, abundance or absence of certain transcripts could either reflect general down-regulation of genes within cells or extravasation and recruitment of cells expressing the respective transcripts to other body compartments. For the sake of interpretation it might therefore be sensible to evaluate a module as changed/unchanged rather than focusing on direction of change.

We acknowledge limitations to our study: First, we did not analyze transcriptional dynamics in control subjects uninfected with sporozoites. This could be a minor concern since the samples collected at day 0 served as individual baseline for each subject. Second, the sample size of 10 volunteers limits the generalization of our findings. In our ongoing studies with Tanzanian volunteers undergoing intravenous vaccination and challenge with *P. falciparum* sporozoites, we will be able to reconcile our observations in a second, independent cohort of similar origin from Tanzania. This will include a more frequent sample collection and comparison of protected vs. non-protected subjects.

Conclusion

This study demonstrates that the wide window of parasite prepatent periods in Tanzanian volunteers, most likely due to different levels of pre-existing immunity or natural resistance, is of importance in evaluating transcriptional responses to CHMI. We found that magnitude and timing of early gene expression changes varied greatly among 10 study subjects, coinciding with the individual's parasite prepatent period. Since optimal sampling time points for each individual are difficult to establish beforehand, we suggest including frequent sampling of blood collections during early stages of infections to capture the short lived transcriptional dynamics of cell populations circulating in the peripheral blood.

Material and methods

Ethics statement

All volunteers gave written informed consent before screening and being enrolled in the study. The trial was performed in accordance with Good Clinical Practices, an Investigational New Drug (IND) application filed with the U.S. Food and Drug Administration (US FDA) (IND 14267), and an Investigational Medical Product Dossier (IMPD) filed with the Tanzanian Food and Drug Administration (TFDA). The protocol was approved by institutional review boards (IRBs) of the Ifakara Health Institute (IHI/IRB/No25) and National Institute for Medical Research Tanzania (NIMR/HQ/R.8a/Vol.IX/1217), and the Ethikkommission beider Basel (EKBB), Basel, Switzerland (EKBB 319/11). The protocol was also approved by TFDA (Ref. No. CE.57/180/04A/50), and the trial was registered at ClinicalTrials.gov (registration ID: NCT01540903, date of registration: 23/02/2012).

Clinical trial design and sample collection

Details of volunteers enrolled and study procedure are given in Shekalaghe et al., 2014 [27]. The single center, double-blind, randomized, controlled trial was conducted in Bagamoyo, Tanzania between February and August 2012. Briefly, 30 healthy male volunteers 20 to 35 years of age were recruited from institutions of higher learning in Dar es Salaam. Screening for eligibility took place at the Clinical Trial Unit of the Ifakara Health Institute in Bagamoyo. Volunteers were screened using predetermined inclusion and exclusion criteria based on clinical examinations and laboratory tests. Tests included medical history and physical examinations, standard hematology, biochemistry and test for malaria, human immunodeficiency virus, hepatitis B and C, and sickle cell disease. Volunteers were injected intradermally with 10,000 ($N = 12$) or 25,000 ($N = 12$) aseptic, purified, cryopreserved *P. falciparum* sporozoites or normal saline ($N = 6$). From day 5 after the controlled human malaria infection (CHMI), thick blood smears were obtained regularly to detect blood parasitemia. Volunteers who became microscopy smear positive, were treated with a standard 3-day regimen of arthemether/lumefantrine (Coartem). qPCR analysis for sensitive detection of blood stage parasitemia was carried out retrospectively after volunteers had been diagnosed and treated. The CHMI proved to be safe for all subjects, showing a high infectivity with 11/12 of the low dose and 10/11 of the high dose subjects developing blood parasitemia [27]. Samples for RNA-Seq were collected from the 10 subjects of the high dose (25,000 PfSPZ) group who developed blood stage parasitemia after CHMI. 2.5 ml whole blood was collected into PAXgene tubes on days 0, 5, 9 and 28 of the study, transported to the Bagamoyo research and training centre (BRTC) laboratory and stored at -80˚C.

RNA isolation and sequencing

Poly(A)$^+$ RNA was prepared from whole blood in PAXgene Blood RNA tubes that had been stored at -80˚C. Following the manufacturer's protocols, RNA was extracted using the PAXgene Blood Kit (PreAnalytiX) and quantified by spectrophotometry. A total of 1.2 μg of total RNA per sample was processed using the GLOBINclear Human kit (Ambion) in order to remove globin mRNA. The quantity and quality of the RNA was analyzed on a Bioanalyzer Eukaryote Total RNA Nano chip. The average RNA Integrity Number (RIN) score across all 40 samples was 8. Two samples collected at day 28 post CHMI (6.4 and 5.2) were below the recommended minimum RIN threshold of 7. RNAs of all samples were submitted for library preparation and sequencing (Expression Analysis Inc., NC). Sequencing libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina), 50 nt paired-end sequence reads were obtained using an Illumina HiSeq 2000 platform and captured as raw sequence data (FASTQ files). All samples were assessed for a sufficient total read count and subsequently passed quality test using FASTQC.

Data processing and statistical analysis

The reads were aligned with STAR [36] against the UCSC hg38 human reference genome and annotated with RSEM [37] (S6 and S7 Figs), applying the default parameters. Read libraries were normalized with TMM (edgeR) [38] and transformed with voom (limma) [39,40]. Following common practice [41], a total of 5,530 genes exhibiting low counts (< 0.5 counts per million) across all libraries were removed, ultimately leaving 16,473 unique genes in the dataset. For the linear modeling of differential gene expression, we performed three analyses: (1) A linear model with a moderated Bayesian variance estimator was applied to the comparisons between time points. The correlation due to repeated measures across time points for the same subjects was controlled by using subject as a blocking variable in the linear model. The analysis

used as one group the 10 high dose subjects who developed blood stage parasitemia and identified the differentially expressed (DE) genes in response to CHMI across time. DE genes were identified with comparisons between the pairwise time points of interest (day 5 vs. day 0 (5/0), 9/5, 9/0, 28/5, 28/9 and 28/0). (2) A secondary analysis was conducted where time to detected asexual blood stage parasitemia was added as a categorical variable (early, average, late) along with an interaction effect (~ parasitemia * day) to the limma linear model. DE genes were identified with an ANOVA-like comparison of all interaction effects using an F-statistics. The null hypothesis being that all interaction effects are zero, and thus time of parasitemia does not have any effect on gene expression changes over time. (3) To determine if leukocyte population frequencies had an impact on differential gene expression, we added the cell counts (reported in S5 Fig) as a continuous covariate along with an interaction effect to the limma linear model (~ day * cell_count). DE genes were identified as in the parasitemia model with an ANOVA-like comparison. The linear modeling was carried out separately for each of the 4 investigated cell populations. For all three analyses, a statistical cutoff of the Benjamini-Hochberg (BH) adjusted p-value less than 0.05 and a minimum 1.5 fold change was used to select DE genes. After the linear modeling, competitive GSEA (camera) [42] was conducted with the blood transcriptome modules (BTM) established by Li et al. [29]. Gene sets described by Chaussabel et al. [30] were used in a confirmatory competitive GSEA analysis. Hypergeometric gene testing (GeneOverlap R package) [43] was performed as an ancillary analysis to support the camera competitive GSEA findings. Given two sets of gene lists (e.g. DE genes at different contrasts and BTMs), this package calculates the overlaps between all pairs of lists from the two sets. Fisher's exact test is then used to determine the p-value and odds ratio in comparison to a genomic background (the genome size) A statistical cutoff of the BH adjusted p-value less than 0.05 was used for selecting significant modules.

Supporting information

S1 File. DE genes determined by limma pairwise visit comparison. logFC: estimate of the log2-fold-change in gene expression corresponding to the tested contrast; AveExpr: average log2 gene expression level over all visits; t: moderated t-statistic; P-value: raw p-value; adj. p-value: adjusted p-value or q value; Trend: direction of gene expression change.
(XLSX)

S2 File. DE genes and gene sets linked to parasitemia. BTMs and gene sets sharing significant overlap with 265 DE genes linked to parasite prepatent period, as determined by hypergeometric testing. FDR: false discovery rate; F: moderated F-statistics.
(XLSX)

S1 Fig. GSEA incorporating Chaussabel gene sets. Statistical significance is pronounced at a p-value & FDR < 0.05. Red: up-regulated, blue: down-regulated.
(TIF)

S2 Fig. BTM hypergeometric testing. Blood transcriptome modules (BTM) sharing significant overlap with DE genes as determined by hypergeometric overlap testing. Only significant overlaps (BH adj. p-value < 0.05) are shown.
(TIF)

S3 Fig. Chaussabel hypergeometric testing. DE Chaussabel gene sets determined by hypergeometric gene set testing. Statistical significance is pronounced at a p-value & FDR < 0.05. Each tile is labeled with the overlap size vs. overall module size.
(TIF)

S4 Fig. Volunteer parasitemia data measured by qPCR. Development of asexual blood parasitemia in 10 volunteers as reported 2014 by Shekalaghe et al. [27]. (a) PMR: parasite multiplication rate, determined applying a linear model as described by Douglas et al. [34]. (b) Development of blood parasitemia visualized as line graph. Colored bars (a) and lines (b) indicate grouping of volunteers into early (red), average (green) and late (blue) for RNA-Seq statistical analysis.
(TIF)

S5 Fig. Changes of leukocyte population frequencies following CHMI. Boxplots are shown for total leukocytes (a), lymphocytes (b), neutrophils (c) and monocytes (d). Individual volunteers are colored according to detection time point of blood stage parasitemia as early (red), average (green) or late (blue). Bars with asterisk indicate statistically significant changes between visits as determined by paired t-test (*: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.0001).
(TIF)

S6 Fig. Read mapping information. Read mapping to UCSC hg38 reference genome using STAR. Illumina sequencing yielded 58.56 to 82.54 million paired-end reads (mean 69.29 million). STAR successfully mapped an average of 87.13% (63.4% - 94.1%) reads to the human reference genome. Among these reads, 13.25% (11.31–17.5%) mapped to multiple loci (light green), with the remaining reads mapping to unique sequence stretches on the reference genome (dark green). Unmapped reads were mostly too short (97.47%, salmon) indicating impaired sequencing quality. A small fraction of unmapped reads (2.0%) were mapped to too many loci or not mapped to the reference for other reasons (0.53%, red).
(TIF)

S7 Fig. Gene count information. Distribution of log2 gene counts after RSEM read quantification. On average, 25.48 million counts were shared across 18,463 (17,513–18,877) gene symbols per sample. Half of these genes exhibit between ~30 to ~1'000 counts. Each 25% of the genes have counts below ~30 or above ~1'000 (up to 1.8 million counts per gene). Across all samples, 22,003 unique genes were covered. Samples are ordered by study day of collection (0,5,9,28) and grouped by subject. Outlier values ($>$ Q3 + 1.5xIQR) are displayed as dots.
(TIF)

Acknowledgments

We would like to thank all volunteers and staff of the Ifakara Health Institute, Bagamoyo Research and Training Centre, Sanaria Inc., Swiss Tropical and Public Health Institute who contributed to the CHMI clinical trial in 2014.

Author Contributions

Conceptualization: Julian Rothen, Salim Abdulla, Marcel Tanner, Raphael Gottardo, Claudia Daubenberger, Ken Stuart.

Data curation: Julian Rothen, Carl Murie.

Formal analysis: Julian Rothen.

Funding acquisition: Marcel Tanner, B. Kim Lee Sim, Stephen L. Hoffman, Claudia Daubenberger, Ken Stuart.

Investigation: Julian Rothen, Carl Murie, Jason Carnes, Mwajuma Chemba, Maxmillian Mpina.

Methodology: Carl Murie, Atashi Anupama, Raphael Gottardo.

Project administration: Julian Rothen, Claudia Daubenberger.

Resources: Salim Abdulla, Marcel Tanner, B. Kim Lee Sim, Stephen L. Hoffman, Raphael Gottardo, Claudia Daubenberger, Ken Stuart.

Software: Julian Rothen, Carl Murie.

Supervision: Carl Murie, Raphael Gottardo, Claudia Daubenberger, Ken Stuart.

Validation: Carl Murie, Atashi Anupama, Raphael Gottardo.

Visualization: Julian Rothen.

Writing – original draft: Julian Rothen.

Writing – review & editing: Carl Murie, Jason Carnes, Atashi Anupama, Raphael Gottardo, Claudia Daubenberger, Ken Stuart.

References

1. WHO | World malaria report 2017. In: WHO [Internet]. [cited 21 Mar 2018]. Available: http://www.who.int/malaria/publications/world-malaria-report-2017/report/en/
2. RTS,S Clinical Trials Partnership. Efficacy and safety of the RTS,S/AS01 malaria vaccine during 18 months after vaccination: a phase 3 randomized, controlled trial in children and young infants at 11 African sites. PLoS Med. 2014; 11: e1001685. https://doi.org/10.1371/journal.pmed.1001685 PMID: 25072396
3. Seder RA, Chang L-J, Enama ME, Zephir KL, Sarwar UN, Gordon IJ, et al. Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. Science. 2013; 341: 1359–1365. https://doi.org/10.1126/science.1241800 PMID: 23929949
4. Mordmüller B, Surat G, Lagler H, Chakravarty S, Ishizuka AS, Lalremruata A, et al. Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. Nature. 2017; 542: 445–449. https://doi.org/10.1038/nature21060 PMID: 28199305
5. Sissoko MS, Healy SA, Katile A, Omaswa F, Zaidi I, Gabriel EE, et al. Safety and efficacy of PfSPZ Vaccine against Plasmodium falciparum via direct venous inoculation in healthy malaria-exposed adults in Mali: a randomised, double-blind phase 1 trial. Lancet Infect Dis. 2017; 17: 498–509. https://doi.org/10.1016/S1473-3099(17)30104-4 PMID: 28216244
6. Ishizuka AS, Lyke KE, DeZure A, Berry AA, Richie TL, Mendoza FH, et al. Protection against malaria at 1 year and immune correlates following PfSPZ vaccination. Nat Med. 2016; 22: 614–623. https://doi.org/10.1038/nm.4110 PMID: 27158907
7. Epstein JE, Paolino KM, Richie TL, Sedegah M, Singer A, Ruben AJ, et al. Protection against Plasmodium falciparum malaria by PfSPZ Vaccine. JCI Insight. 2. https://doi.org/10.1172/jci.insight.89154 PMID: 28097230
8. Lyke KE, Ishizuka AS, Berry AA, Chakravarty S, DeZure A, Enama ME, et al. Attenuated PfSPZ Vaccine induces strain-transcending T cells and durable protection against heterologous controlled human malaria infection. Proc Natl Acad Sci U S A. 2017; 114: 2711–2716. https://doi.org/10.1073/pnas.1615324114 PMID: 28223498
9. Roestenberg M, McCall M, Hopman J, Wiersma J, Luty AJF, van Gemert GJ, et al. Protection against a Malaria Challenge by Sporozoite Inoculation. N Engl J Med. 2009; 361: 468–477. https://doi.org/10.1056/NEJMoa0805832 PMID: 19641203
10. Spring M, Murphy J, Nielsen R, Dowler M, Bennett JW, Zarling S, et al. First-in-human evaluation of genetically attenuated Plasmodium falciparum sporozoites administered by bite of Anopheles mosquitoes to adult volunteers. Vaccine. 2013; 31: 4975–4983. https://doi.org/10.1016/j.vaccine.2013.08.007 PMID: 24029408
11. Kublin JG, Mikolajczak SA, Sack BK, Fishbaugher ME, Seilie A, Shelton L, et al. Complete attenuation of genetically engineered Plasmodium falciparum sporozoites in human subjects. Sci Transl Med. 2017; 9: eaad9099. https://doi.org/10.1126/scitranslmed.aad9099 PMID: 28053159

12. Hoffman SL, Goh LML, Luke TC, Schneider I, Le TP, Doolan DL, et al. Protection of Humans against Malaria by Immunization with Radiation-Attenuated Plasmodium falciparum Sporozoites. J Infect Dis. 2002; 185: 1155–1164. https://doi.org/10.1086/339409 PMID: 11930326

13. Roestenberg M, Teirlinck AC, McCall MBB, Teelen K, Makamdop KN, Wiersma J, et al. Long-term protection against malaria after experimental sporozoite inoculation: an open-label follow-up study. Lancet Lond Engl. 2011; 377: 1770–1776. https://doi.org/10.1016/S0140-6736(11)60360-7

14. Roestenberg M, de Vlas SJ, Nieman A-E, Sauerwein RW, Hermsen CC. Efficacy of preerythrocytic and blood-stage malaria vaccines can be assessed in small sporozoite challenge trials in human volunteers. J Infect Dis. 2012; 206: 319–323. https://doi.org/10.1093/infdis/jis355 PMID: 22615322

15. Sauerwein RW, Roestenberg M, Moorthy VS. Experimental human challenge infections can accelerate clinical malaria vaccine development. Nat Rev Immunol. 2011; 11: 57–64. https://doi.org/10.1038/nri2902 PMID: 21179119

16. Spring M, Polhemus M, Ockenhouse C. Controlled Human Malaria Infection. J Infect Dis. 2014; 209: S40–S45. https://doi.org/10.1093/infdis/jiu063 PMID: 24872394

17. Berman JD, Nielsen R, Chulay JD, Dowler M, Kain KC, Kester KE, et al. Causal prophylactic efficacy of atovaquone-proguanil (Malarone) in a human challenge model. Trans R Soc Trop Med Hyg. 2001; 95: 429–432. PMID: 11579890

18. Teirlinck AC, McCall MBB, Roestenberg M, Scholzen A, Woestenenk R, de Mast Q, et al. Longevity and composition of cellular immune responses following experimental Plasmodium falciparum malaria infection in humans. PLoS Pathog. 2011; 7: e1002389. https://doi.org/10.1371/journal.ppat.1002389 PMID: 22144890

19. Dunachie S, Berthoud T, Hill AVS, Fletcher HA. Transcriptional changes induced by candidate malaria vaccines and correlation with protection against malaria in a human challenge model. Vaccine. 2015; 33: 5321–5331. https://doi.org/10.1016/j.vaccine.2015.07.087 PMID: 26256523

20. Kazmin D, Nakaya HI, Lee EK, Johnson MJ, Most R van der, Berg RA van den, et al. Systems analysis of protective immune responses to RTS,S malaria vaccination in humans. Proc Natl Acad Sci. 2017; 201621489. https://doi.org/10.1073/pnas.1621489114 PMID: 28193898

21. Ockenhouse CF, Hu W, Kester KE, Cummings JF, Stewart A, Heppner DG, et al. Common and Divergent Immune Response Signaling Pathways Discovered in Peripheral Blood Mononuclear Cell Gene Expression Patterns in Presymptomatic and Clinically Apparent Malaria. Infect Immun. 2006; 74: 5561–5573. https://doi.org/10.1128/IAI.00408-06 PMID: 16988231

22. Tran TM, Jones MB, Ongoiba A, Bijker EM, Schats R, Venepally P, et al. Transcriptomic evidence for modulation of host inflammatory responses during febrile Plasmodium falciparum malaria. Sci Rep. 2016; 6. https://doi.org/10.1038/srep31291 PMID: 27506615

23. Vahey MT, Wang Z, Kester KE, Cummings J, Heppner DG, Nau ME, et al. Expression of genes associated with immunoproteasome processing of major histocompatibility complex peptides is indicative of protection with adjuvanted RTS,S malaria vaccine. J Infect Dis. 2010; 201: 580–589. https://doi.org/10.1086/650310 PMID: 20078211

24. van den Berg RA, Coccia M, Ballou WR, Kester KE, Ockenhouse CF, Vekemans J, et al. Predicting RTS,S Vaccine-Mediated Protection from Transcriptomes in a Malaria-Challenge Clinical Trial. Front Immunol. 2017; 8: 557. https://doi.org/10.3389/fimmu.2017.00557 PMID: 28588574

25. Yamagishi J, Natori A, Tolba MEM, Mongan AE, Sugimoto C, Katayama T, et al. Interactive transcriptome analysis of malaria patients and infecting Plasmodium falciparum. Genome Res. 2014; 24: 1433–1444. https://doi.org/10.1101/gr.158980.113 PMID: 25091627

26. Zhu L, Mok S, Imwong M, Jaidee A, Russell B, Nosten F, et al. New insights into the *Plasmodium vivax* transcriptome using RNA-Seq. Sci Rep. 2016; 6: srep20498. https://doi.org/10.1038/srep20498 PMID: 26858037

27. Shekalaghe S, Rutaihwa M, Billingsley PF, Chemba M, Daubenberger CA, James ER, et al. Controlled Human Malaria Infection of Tanzanians by Intradermal Injection of Aseptic, Purified, Cryopreserved Plasmodium falciparum Sporozoites. Am J Trop Med Hyg. 2014; 91: 471–480. https://doi.org/10.4269/ajtmh.14-0119 PMID: 25070995

28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102: 15545–15550. https://doi.org/10.1073/pnas.0506580102 PMID: 16199517

29. Li S, Rouphael N, Duraisingham S, Romero-Steiner S, Presnell S, Davis C, et al. Molecular signatures of antibody responses derived from a systems biological study of 5 human vaccines. Nat Immunol. 2014; 15: 195–204. https://doi.org/10.1038/ni.2789 PMID: 24336226

30. Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity. 2008; 29: 150–164. https://doi.org/10.1016/j.immuni.2008.05.012 PMID: 18631455

31. van Wolfswinkel ME, Langenberg MCC, Wammes LJ, Sauerwein RW, Koelewijn R, Hermsen CC, et al. Changes in total and differential leukocyte counts during the clinically silent liver phase in a controlled human malaria infection in malaria-naïve Dutch volunteers. Malar J. 2017; 16: 457. https://doi.org/10.1186/s12936-017-2108-1 PMID: 29126422

32. Obiero JM, Shekalaghe S, Hermsen CC, Mpina M, Bijker EM, Roestenberg M, et al. Impact of Malaria Preexposure on Antiparasite Cellular and Humoral Immune Responses after Controlled Human Malaria Infection. Infect Immun. 2015; 83: 2185–2196. https://doi.org/10.1128/IAI.03069-14 PMID: 25776749

33. Hodgson SH, Juma E, Salim A, Magiri C, Kimani D, Njenga D, et al. Evaluating controlled human malaria infection in Kenyan adults with varying degrees of prior exposure to Plasmodium falciparum using sporozoites administered by intramuscular injection. Front Microbiol. 2014; 5: 686. https://doi.org/10.3389/fmicb.2014.00686 PMID: 25566206

34. Douglas AD, Edwards NJ, Duncan CJA, Thompson FM, Sheehy SH, O'Hara GA, et al. Comparison of Modeling Methods to Determine Liver-to-blood Inocula and Parasite Multiplication Rates During Controlled Human Malaria Infection. J Infect Dis. 2013; 208: 340–345. https://doi.org/10.1093/infdis/jit156 PMID: 23570846

35. Miller JL, Sack BK, Baldwin M, Vaughan AM, Kappe SHI. Interferon-Mediated Innate Immune Responses against Malaria Parasite Liver Stages. Cell Rep. 2014; 7: 436–447. https://doi.org/10.1016/j.celrep.2014.03.018 PMID: 24703850

36. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013; 29: 15–21. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886

37. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12: 323. https://doi.org/10.1186/1471-2105-12-323 PMID: 21816040

38. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26: 139–140. https://doi.org/10.1093/bioinformatics/btp616 PMID: 19910308

39. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014; 15: R29. https://doi.org/10.1186/gb-2014-15-2-r29 PMID: 24485249

40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43: e47. https://doi.org/10.1093/nar/gkv007 PMID: 25605792

41. Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. Proc Natl Acad Sci U S A. 2010; 107: 9546–9551. https://doi.org/10.1073/pnas.0914005107 PMID: 20460310

42. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. Nucleic Acids Res. 2012; gks461. https://doi.org/10.1093/nar/gks461 PMID: 22638577

43. Shen L, Mount S. GeneOverlap: Test and visualize gene overlaps [Internet]. 2013. Available: http://shenlab-sinai.github.io/shenlab-sinai/

Chapter 4

Subspecies typing of *Streptococcus agalactiae* based on ribosomal subunit protein mass variation by MALDI-TOF MS

This chapter contains the following manuscript (submitted to *Clinical Infectious Diseases*):

Julian Rothen, Joël F. Pothier, Frédéric Foucault, Jochen Blom, Dulmini Nanayakkara, Carmen Li, Margaret Ip, Marcel Tanner, Guido Vogel, Valentin Pflüger and Claudia A. Daubenberger. "Subspecies typing of *Streptococcus agalactiae* based on ribosomal subunit protein mass variation by MALDI-TOF MS". 2018. *Submitted to Clinical Infectious Diseases.*

Subspecies typing of *Streptococcus agalactiae* based on ribosomal subunit protein mass variation by MALDI-TOF MS

Short title: *Streptococcus agalactiae* subspecies typing using MALDI-TOF MS

Julian Rothen; JR [1,2*], Joël F. Pothier; JFP [3], Frédéric Foucault; FF [4], Jochen Blom; JB [5], Dulmini Nanayakkara; DN [6], Carmen Li; CL [6], Margaret Ip; MI [6], Marcel Tanner; MT [2], Guido Vogel; GV [4],Valentin Pflüger; VP [4*], Claudia A. Daubenberger; CAD [1,2]

1 Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute (Swiss TPH) Basel, 4002 Basel, Switzerland

2 University of Basel, 4002 Basel, Switzerland

3 Environmental Genomics and Systems Biology Research Group, Institute for Natural Resource Sciences, Zurich University of Applied Sciences (ZHAW), 8820 Wädenswil, Switzerland

4 Mabritec AG, 4125 Riehen, Switzerland

5 Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, 35390 Giessen, Germany

6 Department of Microbiology, The Chinese University of Hong Kong, Shatin, NT, Hong Kong

* Corresponding authors

ABSTRACT

In this study a ribosomal subunit protein (rsp) profiling based on matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) mass spectrometry (MS) was developed for fast subtyping of *Streptococcus agalactiae* (Group B *Streptococcus*, GBS), a major cause of neonatal sepsis and meningitis. A total of 796 GBS whole genome sequences, mirroring the genetic diversity of the global GBS population, were used to identify molecular mass variability of 28 rsp. We identified 62 unique rsp mass combinations, termed "rsp-profiles" which can be distinguished by MALDI-TOF MS. The majority (>80%) of GBS sequenced strains was found to display one of the six rsp-profiles 1-6. Importantly, these dominant rsp-profiles classify GBS sequenced strains in high concordance with the core-genome based phylogenetic clustering. Validation of our approach by MALDI-TOF MS analysis of 248 in-house GBS isolates showed that the 28 rsp were detected reliably in the generated mass spectra, allowing quick assignment of clinical isolates to rsp-profiles at high sensitivity (99%) and specificity (97%). Our approach distinguishes the major phylogenic GBS genotypes, identifies hyper-virulent strains, predicts probable capsular serotype and surface protein variants and distinguishes between GBS genotypes of human and animal origin. In summary, we propose an elegant method combining the advantages of the information depth generated by WGS with the highly cost efficient, rapid and robust MALDI-TOF MS approach facilitating high-throughput, inter-laboratory, large-scale GBS epidemiological and clinical studies.

KEYWORDS

Group B *Streptococcus*, MALDI-TOF MS, ribosomal subunit protein, molecular epidemiology

SIGNIFICANCE STATEMENT

The World Health Organization has recently released a technology roadmap, listing priority research activities pertaining to vaccine development against Group B *Streptococcus* (GBS), a major cause of neonatal invasive disease and responsible for 150,000 stillbirths and infant deaths every year. Large-scale and long-term GBS population monitoring has been proposed to assess vaccine impact on distribution of capsular serotypes, strain replacement and the emergence of escape strains from animal reservoirs. We present a ribosomal subunit protein based MALDI-TOF MS scheme that allows such high-throughput GBS strain-level typing. Our approach distinguishes the major phylogenic GBS genotypes, identifies hyper-virulent strains, predicts the probable serotype and distinguishes between GBS genotypes of human and animal origin.

INTRODUCTION

Streptococcus agalactiae (Group B *Streptococcus*, GBS) is a Gram-positive bacterium known to colonize the gastrointestinal and urogenital tract of around 18% of pregnant women worldwide (1). In 2015, it has been estimated that GBS caused in 205,000 and 114,000 infants early-onset (between day 0 and day 6 of age) and late-onset disease (between day 7 and day 89 of age), respectively. The reason for the emergence of GBS as an important human pathogen has been attributed to the spread of pathogenic GBS clones (2) supported by the widely use of tetracycline (3). This has led to few globally established, genetically homogeneous human GBS lineages, which stand in contrast to obligate animal GBS strains, which were not affected by the tetracycline induced evolutionary bottleneck and therefore remain an under-researched reservoir of genetically highly diverse genotypes (4, 5).

GBS carry polysaccharide capsules which are thought to be main virulence factors interfering with phagocytic clearance of the bacteria (6). Ten GBS serotypes (Ia, Ib, II, III, IV, V, VI, VII, VIII, IX) have been described and are commonly used to classify GBS and to monitor population dynamics (7). Distribution of serotypes differs globally and some serotypes are associated with higher virulent GBS isolates. In particular, serotype III has been frequently associated with infant early-onset and late-onset GBS disease and meningitis (8). Multi-locus sequence typing (MLST) has also been widely used for GBS isolate discrimination and revealed that the global population is dominated by only five major clonal complexes, namely CC1, CC10, CC17, CC19 and CC23 (9).

Vaccination of pregnant women during the second and third trimester has been proposed as a novel public health tool to prevent GBS disease in both mothers and children (10). Trivalent glycoconjugate vaccines, covering serotypes Ia, Ib and III have completed phase I and II clinical trials (11, 12) and a pentavalent vaccine including serotypes Ia, Ib, II, III and V is under development (13). GBS protein based vaccines that target surface antigens (pili and alpha-like proteins) are also under development although they will need to overcome sequence variation of the targeted proteins (14, 15). Long-term effects of vaccine introduction on circulating GBS population will need to be monitored to assess possible capsular switches, capsular replacements or novel appearance of GBS strains (13, 16).

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has become the gold standard for high-throughput microbial species identification in clinical settings (17, 18). Currently commercially available and validated MALDI-TOF MS test systems rely on detection of generic peptide patterns (peptide fingerprints), which severely limits the discriminatory power for closely related species and separation of subspecies and sub-lineages (19, 20). The conserved ribosomal subunit proteins (rsp) are cytosolic proteins of high abundance and in the molecular weight range detectable by commercial MALDI-TOF MS systems. Molecular weight variation of *in silico* predicted rsp masses can be determined by MALDI-TOF MS, thereby providing a targeted, biomarker-based approach of classifying mass spectra, which is superior to the conventional "pattern-recognition" approach (21, 22). Here, we demonstrate that measuring the allelic mass differences in 28 rsp of GBS by MALDI-TOF MS provides a highly cost efficient, rapid and robust approach that facilitates high-throughput, large-scale GBS epidemiological and clinical intervention studies.

RESULTS

Capsular serotype and sequence type distribution in whole genome GBS sequences collection

A total of 796 whole genome GBS sequences (WGS) was collated from public databases and in-house sequenced isolates from human, camel, bovine and other animal origin. A listing of the 796 GBS isolates and their corresponding metadata is provided in Supplementary Table S1. Using *in silico* MLST, 108 sequence types (ST) were included and except for ST327, all ST that are among the 28 most abundant in the global population were present in these 796 isolates (Supplementary Figure S2). In accordance with the PubMLST *S. agalatiae* isolate database (23), the global repository of MLST based ST distribution, ST17, ST1, ST23 and ST19 were the most frequent ST, accounting for 36% isolates in our collection (50% in PubMLST). ST61 and ST554, which are less abundant on a global scale, were over-represented in our collection (16% versus < 1% in PubMLST). All GBS capsular serotypes, except serotype VII, were represented as follows: Ia: n = 126, Ib: n = 52, II: n = 211, III: n = 183, IV: n = *80*, V: n=*111*, VI: n = *9*, VIII: n = 1, IX: n = 2, non-type-able: n = *21*. In summary, the 796 WGS collected constitute a global representation of the GBS population.

Genome-wide phylogenetic analysis of GBS collection

A core-genome phylogenetic analysis of the GBS WGS based on inclusion of 867 genes was conducted using EDGAR (24). Sixteen GBS strains originating from *Camelus dromedarius* were found distant from all other GBS genotypes (Figure 1a). Some strains isolated from fish, frog and cattle formed distinct, host origin specific clusters. The other phylogenetic clusters consisted predominantly of genotypes of human origin, with sporadic presence of animal associated strains. An exception to this was one very heterogeneous but distinct phylogenetic cluster, containing a range of genotypes of fish, bovine, human, rat and dog origin (Figure 1a).

The core-genome based phylogenetic clustering was subsequently compared with the classification by *in silico* MLST. One-hundred and eight ST were grouped into 15 clonal complexes (CC) of closely related isolates (CC67, CC1, CC17, CC23, CC19, CC10, CC459, CC452, CC7, CC283, CC615, CC609, CC103, CC4 and CC552) each CC consisting of a founder ST and its single-locus variants (SLV). Remaining ST that were double-locus variants (DLV) of founder ST were assigned to the corresponding CC and four ST (ST22,

ST26, ST130 and ST616) that could not be attributed to a CC were defined as stand-alone ST. There was a general high agreement between MLST classification of GBS isolates and the core-genome phylogenetic clustering (Figure 1b). As expected, core-genome based classification provides a resolution power that goes beyond CC or ST identity of an isolate, thereby further sub-grouping genotypes that appear identical by MLST. In some cases, the genetic variation of such sub-groups puts them into overall closer phylogenetic relationship with genotypes of other CC origin. This was the case for CC23, CC452, CC10, CC7, CC283 and CC1, CC4 and CC459 (Figure 1b).

Comparison of the core-genome based phylogenetic clustering with in silico assigned capsular serotypes confirmed that genotypes clustering closely together are likely to share the same serotype (Figure 1c). Similarly, variant distribution of the five GBS surface proteins (alpha-like protein (Alp) gene family, pilus islands, surface immunogenic protein (Sip), laminin-binding protein (Lmb) and Group B *Streptococcus* immunogenic bacterial adhesin (BibA) protein) investigated here was largely congruent with the core-genome phylogenetic clustering (Supplementary Figure S3).

Average nucleotide identity (ANI) analysis, unlike core-genome analysis indexing sequence variation across all genes contained in an organism, produced a grouping of the GBS strains that was in high concordance to the core-genome phylogenetic analysis (Supplementary Figure S4).

In silico prediction of ribosomal subunit protein molecular masses

WGS data of 29 GBS isolates that were also cultivated in house was used *in silico* to predict molecular masses of all known 59 rsp. MALDI-TOF MS analysis conducted with these isolates revealed that 28 out of 59 rsp were reproducibly measured in a molecular weight range between 4,425 to 19,293 Da (Figure 2a). These experiments confirmed that our novel sample preparation protocol enables us to identify the masses of 28 distinct rsp in total cell lysates of GBS as exemplified in Figure 2b with five distinct isolates measured.

Next, the 796 WGS were used to predict the molecular masses of these 28 rsp including S8-S10, S12, S13, S15-S19 and S21 of the small ribosomal subunit and L6, L13, L14, L17-L19, L21-L24, L29, L30 and L32-L36 of the large ribosomal subunit. Three rsp (L14, L29 and S15) did not show allelic mass variation across all 796 isolates. Four rsp (L22, L32, L33 and S21) showed a variant mass in 1 out of 796 isolates. Eighteen rsp (L6, L17-L19, L21, L23, L24, L30, L34, L36, S9, S10, S12, S13, S16-S19) showed low variability, with mass variants

found in fewer than 8 out of 796 isolates. The most variable rsp were L13, L35 and S8 displaying mass variation in > 100 out of 796 isolates (Figure 2c).

Definition of rsp profiles in our GBS collection

We next *in silico* predicted all possible combinations of these distinct 28 rsp masses in the 796 WGS. When taking into account the MALDI-TOF MS detection accuracy (400 ppm threshold), we identified 62 unique and distinguishable combinations, which are referred to onwards as rsp-profiles. Six dominant rsp-profiles (rsp-profiles 1-6), present in 83 to 134 GBS isolates covered 83% of the isolates (657/796). Five rsp-profiles (rsp-profiles 7-11) were present in 5 to 42 GBS representing 9% (72/796) of isolates. Rsp-profiles 12-22 existed in 2 to 4 isolates (27/796, 3%) and the remaining 40 rsp-profiles named 23-62 were singletons (40/796, 5%) (Supplementary Figure S5). These newly defined rsp-profiles classified the GBS strains in high concordance with the core-genome based phylogenetic clustering. GBS strains sharing an identical rsp-profile were located either next to each other or in the same subordinate cluster in the core-genome based phylogenetic tree (Figure 3). The exception to this were one rsp-profile 5 strain and two rsp-profile 4 strains that were grouped to strains with different rsp-profiles as well as five ST103 strains with rsp-profile 4 that formed a separate group in the core-genome analysis (Figure 3). All of these strains in fact displayed novel rsp-profiles as predicted by the *in silico* analysis, but the molecular mass differences to rsp-profile 4 and rsp-profile 5 respectively, were too small to be picked up by MALDI-TOF MS, standing exemplary for the technical limitations of our current method. The genetically distinct genotypes isolated from camels displayed a large variety of camel-specific rsp-profiles (Figure 4, Supplementary Table S1). This occurrence of unique, animal-specific rsp-profiles was also observed with ST260/ST552 isolates originating from frogs (n = 1) and fish (n = 3). Similarly, strains of the bovine-specific CC67 also displayed specific rsp-profiles: rsp-profile 1 for the majority (n = 134), but also less frequent rsp-profiles in some ST61 (n = 10), ST591 (n = 4) and ST622 (n = 2) isolates. Other GBS isolated from animal sources displayed rsp-profiles shared with human-associated GBS genotypes (Figure 4).

Association of six dominant rsp-profiles with CC, serotype and surface antigens

We next investigated how GBS strains belonging to the six dominating rsp-profiles compared with respect to the *in silico* predicted MLST based CC (Figure 5a), capsular genotype (Figure 5b) and pilus island variants (Figure 5c). Rsp-profile 1 (n = 134; 16.8%) contains only bovine

originating CC67, is dominated by capsular serotype II (*n* = 122) and indicative for pilus island PI-2b (*n* = 124). Rsp-profile 2 (*n* = 124; 15.6%) is dominated by CC17 (*n* = 101) and its recombination derivative CC452 (*n* = 23) with the capsular serotype III (*n* = 99) and IV (*n* = 25), and is indicative of pilus islands PI-1/PI-2b (*n* = 87) or PI-2b (*n* = 37). Rsp-profile 3 (*n* = 111; 13.9%) is composed of members of CC1 (*n* = 110), and one strain of CC4 (*n* = 1), while the capsular serotypes include predominantly serotype V (*n* = 80) as well as II (*n* = 17), IV (*n* = 9) and non-typeable (*n* = 5). Rsp-profile 3 is linked to pilus islands PI-1/PI-2a (*n* = 94) and PI-2a (*n* = 17). Rsp-profile 4 (*n* = 109; 13.7%) is dominated by CC23 (*n* = 71) and its recombination derivative CC452 (*n* = 11) with a larger fraction of non-dominating CC (*n* = 27), including ST22, ST26, ST130 CC103 and CC283. Capsular serotypes expressed in this profile include Ia (*n* = 78), II (*n* = 15), III (*n* = 7), V (*n* = 8) and IX (*n* = 1). Pilus islands covered by this profile are PI-2a (*n* = 96), PI-1/PI-2a (*n* = 9) and PI-2b (*n* = 4). Rsp-profile 5 (*n* = 96; 12.1%) is composed of members of CC10 (*n* = 62), CC7 (*n* = 27), CC283 (*n* = 6) and CC23 (*n* = 1), with diverse serotypes Ia (*n* = 23), Ib (*n* = 41) and II-V (*n* = 30). Equally diverse are the pilus islands, with PI1/PI-2a (*n* = 96) and PI1/PI-2b (*n* = 17) dominating. In the rsp-profile 6 (*n* = 83; 10.4%), the majority of isolates belong to CC19 (*n* = 72) and some to CC1 (*n* = 10) and CC4 (*n* = 1), with the capsular serotypes II (*n* = 17) and III (*n* = 51) and pilus island PI-1/PI-2a (*n* = 73) dominating.

Validation of *in silico* established rsp-profiles by whole-cell lysate MALDI-TOF MS analysis

A total of 248 GBS isolates were analyzed in quadruplicates using MALDI-TOF MS, resulting in 992 mass spectra. Most of the 28 rsp were detected consistently across all spectra and some of these used for internal spectra calibration at 800 ppm. Twenty-five rsp were detected in > 98% of all spectra acquired. L34 was found in 96%, L6 was detected in 91% and rsp L33_1 was detected in 80% of spectra acquired.

For the 29 GBS isolates with whole genome sequences available, 14 distinct rsp-profiles including the rsp-profiles 2-6, 11, 15, 18-22, 37 and 55 were predicted. Validation by MALDI-TOF MS demonstrated 100% sensitivity of our approach, with all 29 GBS being assigned an rsp-profile (Table 1a). All but one of the MALDI-TOF MS measured rsp-profiles corresponded to *in silico* predicted rsp-profiles, giving a specificity of 97%. The one isolate that was classified as rsp-profile 6, should display the unique rsp-profile 55 based on WGS based prediction. Rsp-profile 55 differs only minimally (unique mass allele of rsp L32) from rsp-profile 6 and we cannot exclude the possibility that lack of sequence quality led to a false *in silico* prediction of the mass allele in question.

For the remaining 219 GBS isolates, rsp-profiles were assigned to 210/219 isolates. Spectra of six isolates had to be manually updated for missing rsp but could subsequently be assigned to an rsp-profile, producing an overall sensitivity of 99% (216/219 isolates assigned to an rsp-profile). A total of 150 GBS isolates were assigned to the dominating rsp-profiles 2-6. Of the GBS isolates originating from cows, five ST591 linked strains were assigned to rsp-profile 19 and one strain (SLV ST19) was assigned to rsp-profile 6. The 65 isolates of camelid origin were assigned to seven different rsp-profiles that are, with exception of rsp-profile 22, specific for camel genotypes. Three isolates could not be classified because they displayed an rsp-profile not yet contained in our reference database that was based on the 796 WGS (Supplementary Table S1).

Confirmation of inter-laboratory reproducibility of our method

MALDI-TOF MS typing of eight GBS isolates with available WGS in an independent laboratory confirmed that our rsp-based typing method could be easily transferred to other research sites and different MALDI-TOF MS systems. Bacterial sample processing following our protocol and subsequent measurement on a Microflex MALDI-TOF MS system produced high quality mass spectra, with all 28 rsp required for classification being detected. All eight isolates were assigned to rsp-profile 5, one of the global dominant lineages. *In silico* rsp evaluation of the 28 rsp molecular masses confirmed the correct identification for all eight isolates (Table 1a).

DISCUSSION

Current GBS typing methods based on MLST, pulsed-field gel electrophoresis (PFGE) or capsular serotyping can provide insight into GBS population composition and dynamics. However, several shortcomings limit the propensity of these tools for high-throughput epidemiological analyses. All three methods are limited due to their time-consuming nature and considerable per sample processing costs. Further, a significant proportion of non-typeable strains cannot be identified by serotyping and PFGE results cannot be compared across different laboratories. All three methods are also limited in inferring evolutionary relationships between strains (25).

The rapidly increasing public availability of WGS data has supported the development of commonly available computational analysis tools, allowing in-depth comparison of whole bacterial genomes and thereby transforming our understanding of bacterial taxonomy (26). Here, we used a collection of 796 WGS that were grouped according to their core-genome and that – when compared to the conventional MLST based classification - confirmed the global presence of the major CC including CC1, CC10, CC17, CC19, CC23 and CC67 (2). The core-genome based phylogenetic analysis provided insight into the heterogeneity of genotypes classified into the same CC (Figure 1b). Distinct clustering within each CC occurs and in some cases these clusters share more similarity with members of other CC (as was the case for CC23 and CC452 strains, CC10, CC7, CC283 and CC1 strains as well as for CC4 and CC459 strains). This confirms the limitations of MLST based bacterial typing which is based on the allelic differences in seven conserved house-keeping genes in understanding the relatedness of GBS isolates (9).

MALDI-TOF MS has been previously applied to identify GBS hypervirulent ST17 and ST1 strains based on single biomarker masses of either unknown (27) or non-rsp identity (28). In the here presented study, we aimed to develop a MALDI-TOF MS based GBS typing method with increased robustness and a higher discriminatory power compared to previous approaches. This novel method targets 28 known ribosomal loci, allowing us to simultaneously detect molecular mass variation across a concatenated amino acid sequence of ~2,700 aa. We can thereby exploit subtle differences in an evolutionary highly conserved part of the GBS genome for classification of closely related genotypes according to their core-genome phylogeny.

We found that for GBS, 28 rsp between 4,425 to 19,293 Da are reliably detectable in MALDI-TOF mass spectra. This improved detection of individual rsp is fundamental to unequivocally separate *S. agalactiae* strains and requires generation of high-level spectra quality. To obtain this, we developed a refined preprocessing protocol for the bacterial cells which - despite of being more sophisticated than standard approaches like direct smear or formic acid extraction - still allows high-throughput analysis of GBS samples at low costs. A single analysis of an isolate takes less than an hour to perform and can easily be up-scaled, allowing typing of 40-60 isolates per day. Per sample costs of less than 6 USD renders our approach highly competitive to currently employed typing and GBS identification methods. Our newly developed bioinformatics analysis pipeline requires minimal bioinformatics knowledge and hands on time by the routine user. We identified six dominant rsp-profiles in the global GBS population, which can be distinguished with high specificity by MALDI-TOF MS and that classify GBS genotypes in high concordance to their corresponding core-genome phylogeny (Figure 3). Matching of these six rsp-profiles against 115,768 MALDI-TOF MS spectra from routine diagnostics, covering 3,013 bacterial species, revealed that the highest matching species following GBS was *Streptococcus pyogenes* with only 15/28 rsp having a similar size measurement like GBS, which demonstrates our methods high GBS-specificity and robustness for false positive assignment. One limitation of our method is the fact that only strains with rsp-profiles already existing in the reference database can be classified (as exemplified by the three in-house isolates with no assigned ID). This is a minor concern, given that the majority of circulating strains will display one of the already described dominant rsp-profiles. In the case of atypical GBS genotypes with an unknown rsp-profile, our method serves as a first-line screening tool. Strains flagged by our method can in a next step be subjected to whole genome sequencing and the *in silico* extracted rsp-profile subsequently be incorporated into our reference database for future direct identification of such novel genotypes. Further limitations of our approach rest with the technical capacity of MALDI-TOF MS to detect and discriminate rsp masses. Minimal molecular weight differences of less than 400 ppm are below the detection threshold of the standard MALDI-TOF MS systems. This can lead to failure in distinguishing certain rsp-profiles and therefore, false assignment of an rsp-profile. In our collection of 796 GBS strains we only observed eight isolates for which this was the case. Future advance of the MALDI-TOF MS technology, improving both accuracy and overall covered mass range, will help in overcoming these current limitations.

We propose several possible applications for the novel MALDI-TOF MS based rsp-typing: (i) Identification of GBS with rsp-profile 2 that is found to be specific for the known hyper virulent CC17 lineage and its clone ST452, which emerged from recombination between CC23 and CC17 isolates (29), could be of specific interest in clinical routine diagnostics. Rapid identification of known hyper-virulent GBS strains in pregnant women could help in the decision-making regarding the necessity of intrapartum antibiotic prophylaxis. (ii) Distinct rsp-profiles allow the prediction of probable serotype and surface protein variants, all of which are major determinants of GBS virulence (Figure 5 and Supplementary S3). The predictive value thereby varies between the different rsp-profiles, with some showing strong others less correlation. For example, rsp-profile 1 is highly indicative of serotype II and rsp-profile 2 is almost exclusively found in strains with serotype III, while rsp-profile 5 is present in strains of serotypes Ia, Ib, II, III, IV and V (Figure 5b). (iii) Tracking of GBS zoonotic events is easily possible since rsp-profiles of isolates of animal origin differ distinctly from human isolates. For example, rsp-profiles 1, 10, 12, 17, 19, 38, 39, 59, 60 and 61 are confined to bovine origin, rsp-profile 49 is unique for fish origin, rsp-profile 13 is indicative of either fish or frog origin and rsp-profiles 11, 15, 18, 20, 21 and 37 are unique to cameloid origin (Figure 4). (iv) Lastly, the here presented method would allow for large-scale monitoring of GBS vaccination impact. Currently, a trivalent CPS vaccine is most advanced in clinical development, incorporating serotypes Ia, Ib and III (10, 11, 30). These serotypes, although the most pathogenic, represent only a fraction of the GBS global population. The potential implications of a vaccine targeting only selected genotypes can be exemplified by the lessons learned from the multi-valent pneumococcal vaccine. After its introduction in Europe and North America, vaccine-type serotypes and associated invasive pneumococcal disease decreased rapidly. However, non-vaccine-type serotypes and linked disease increased in the years following vaccine introduction, indicating the importance of continued population surveillance in order to track serotype replacement (31–33). Therefore, for testing and long term follow up of the impact of maternal GBS vaccination, it will be essential to understand how the global GBS population responds to vaccine induced immune selection. Particularly, it would be interesting to understand if some of the ancient GBS lineages that carry serotypes included into the vaccine vanish, if vaccine escape strains carrying non-vaccine serotypes emerge as newly dominating lineages and if GBS with zoonotic potential fill the vaccination induced biological niches.

CONCLUSION

We present here a ribosomal subunit protein based MALDI-TOF MS typing method. Due to the evolutionary conservation and high expression level of rsp, the discriminatory power of this method allows classification of GBS isolates according to their core-genome phylogeny within minutes and with minimal bioinformatics knowledge and hands on time required. In summary, our approach is rapidly extendable by (i) including rsp-profiles or (ii) other biomarker masses from any novel GBS isolate emerging, (ii) robust against inter-laboratory/platform variation of mass spectra quality, (iii) streamlined for easy application by minimally trained users, (iv) suitable for high-throughput, large scale GBS epidemiological and clinical studies and (v) highly cost efficient with per sample analysis costs of less than 6 USD and results obtained within minutes. We propose an elegant way combining the advantages of the information depth generated by WGS with the highly cost efficient, rapid and robust MALDI-TOF MS approach for high-throughput, biomarker-based GBS microbiological research.

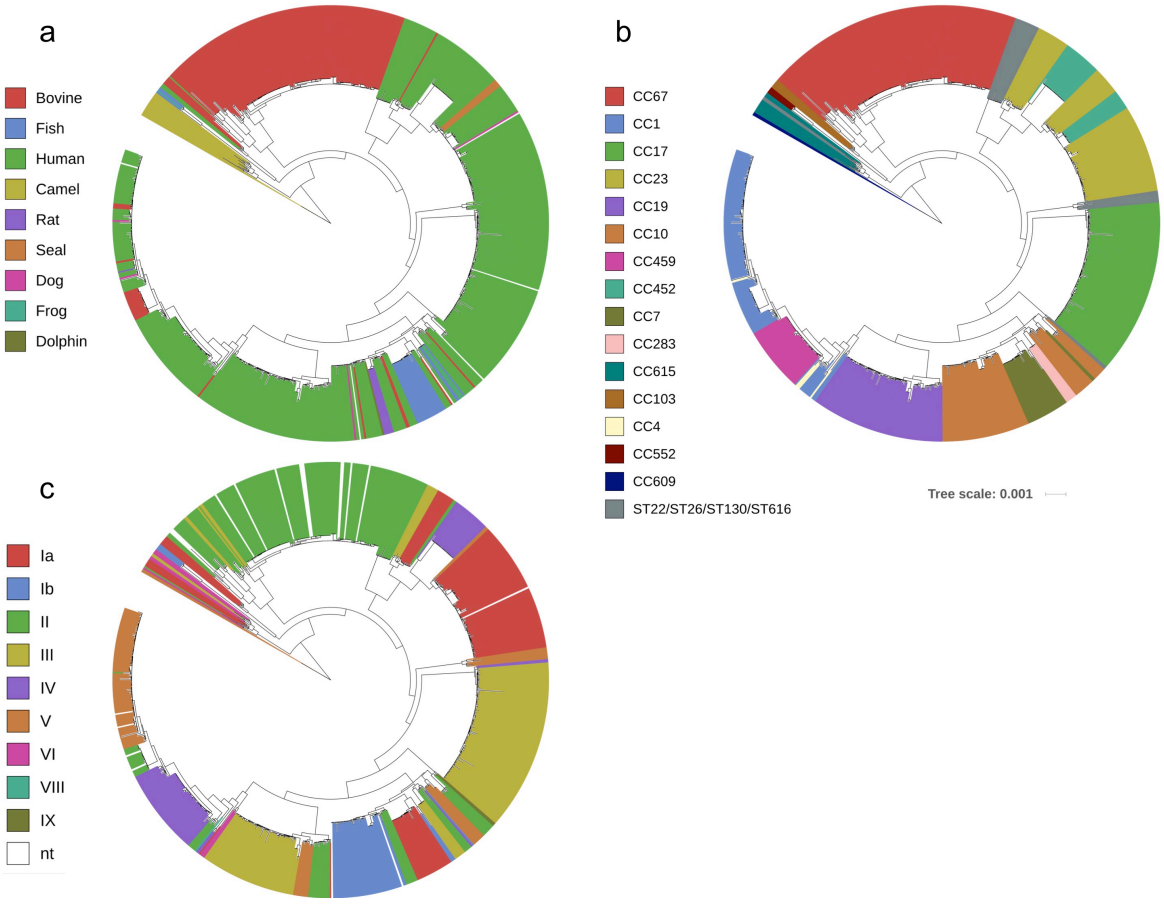


Figure 1. FastTree phylogenetic tree based on core-genome analysis of 796 Group B *Streptococcus* whole genome sequences. Individual strains are annotated with (a) host origin, (b) *in silico* predicted multi-locus sequence typing clonal complex (CC) or sequence type (ST) and (c) *in silico* predicted capsular serotype. (Scale bar: nucleotide substitutions per site).

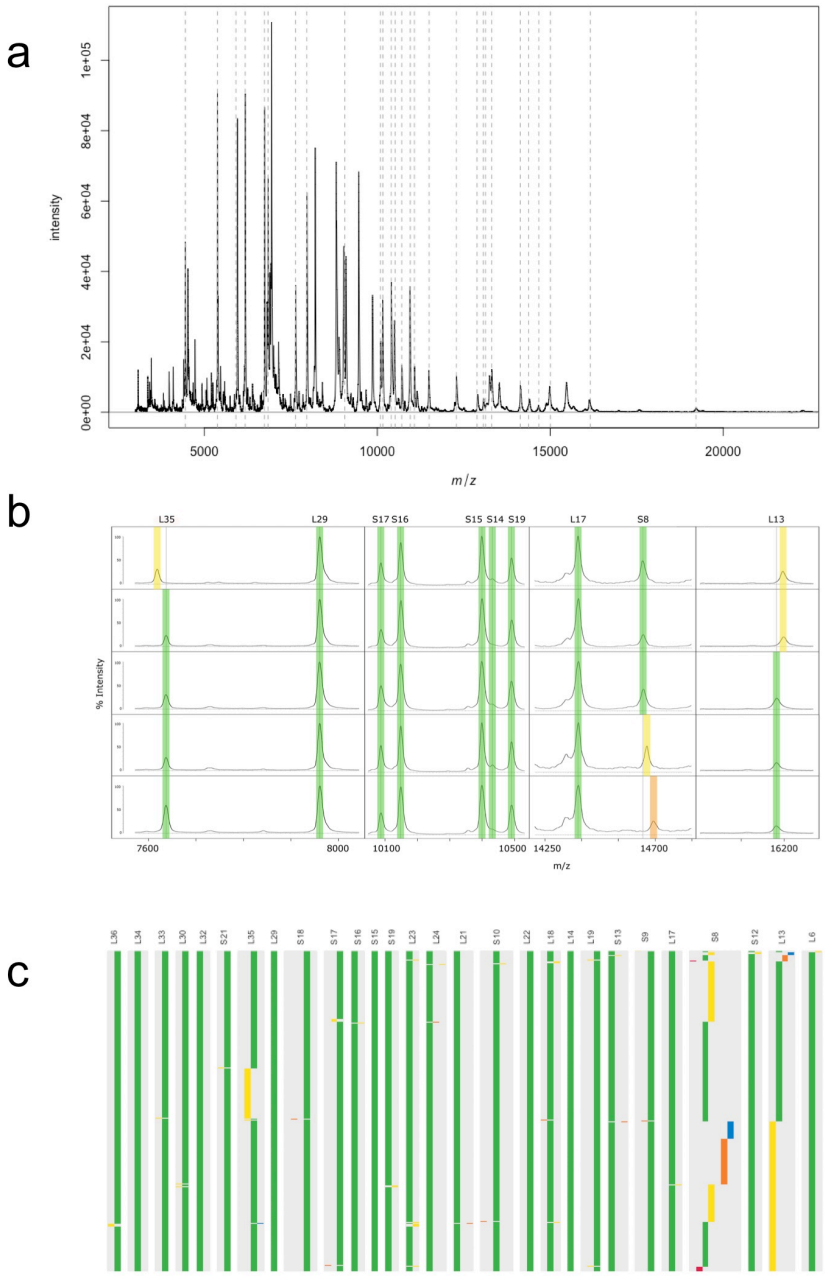


Figure 2. (a) A representative MALDI-TOF mass spectrum of a *Streptococcus agalactiae* ST7 strain, covering the mass range between 4,000 - 22,000 Da. The arbitrary intensity values of the mass peaks are given on the *y*-axis. Dashed lines indicate the position of the 28 ribosomal subunit proteins (rsp) targeted in our analyses. (b) Assessment of mass spectra belonging to five Group B *Streptococcus* (GBS) isolates confirm *in silico* predicted mass shifts in three rsp (L35, L13 and S8). Green: Major rsp mass, yellow & orange: rsp mass variants. (c) *In silico* predicted molecular mass variation of 28 rsp across 796 GBS whole genome sequences. Ribosomal subunits proteins are ordered from left to right by increasing molecular weight. Green: Most abundant rsp mass allele; yellow, orange and blue: 2nd, 3rd and 4th most abundant rsp mass allele; Red: remaining rsp mass alleles.

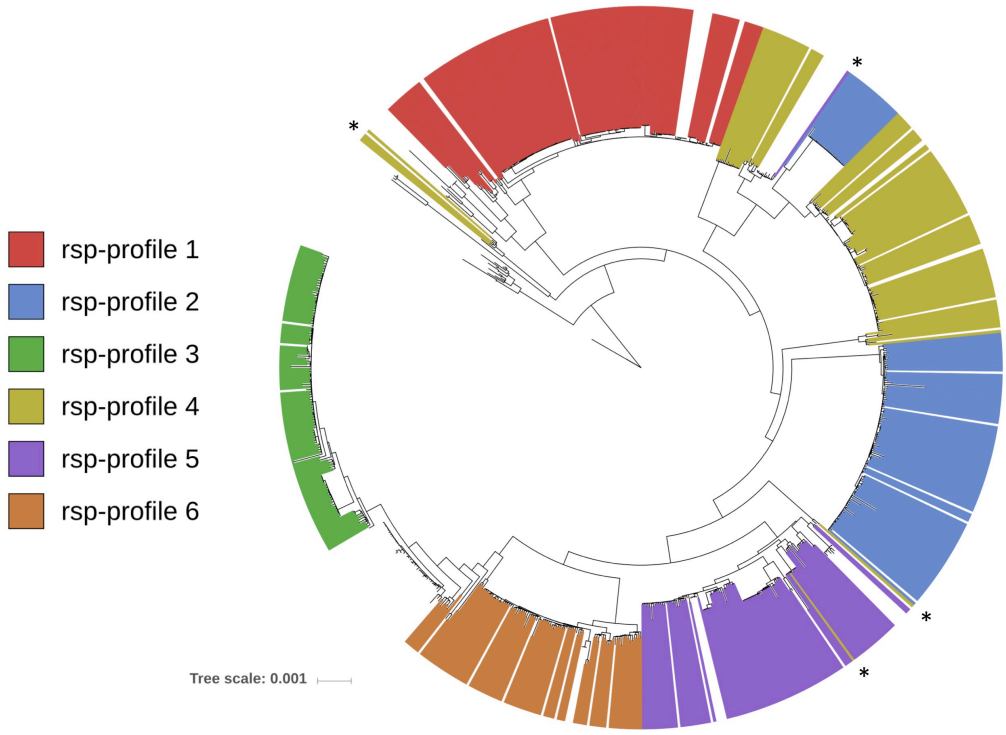


Figure 3. FastTree phylogenetic tree based on core-genome analysis of 796 Group B *Streptococcus* whole genome sequences (WGS). Individual strains are annotated with their *in silico* determined ribosomal subunit proteins (rsp)-profile. For simplicity, only the six globally dominant rsp-profiles are shown (covering 83% of isolates in our WGS collection). Marked with asterisks are eight strains whose rsp-profile was miss-assigned due to limitation of MALDI-TOF MS resolution (i.e. 400 ppm). (Scale bar: nucleotide substitutions per site).

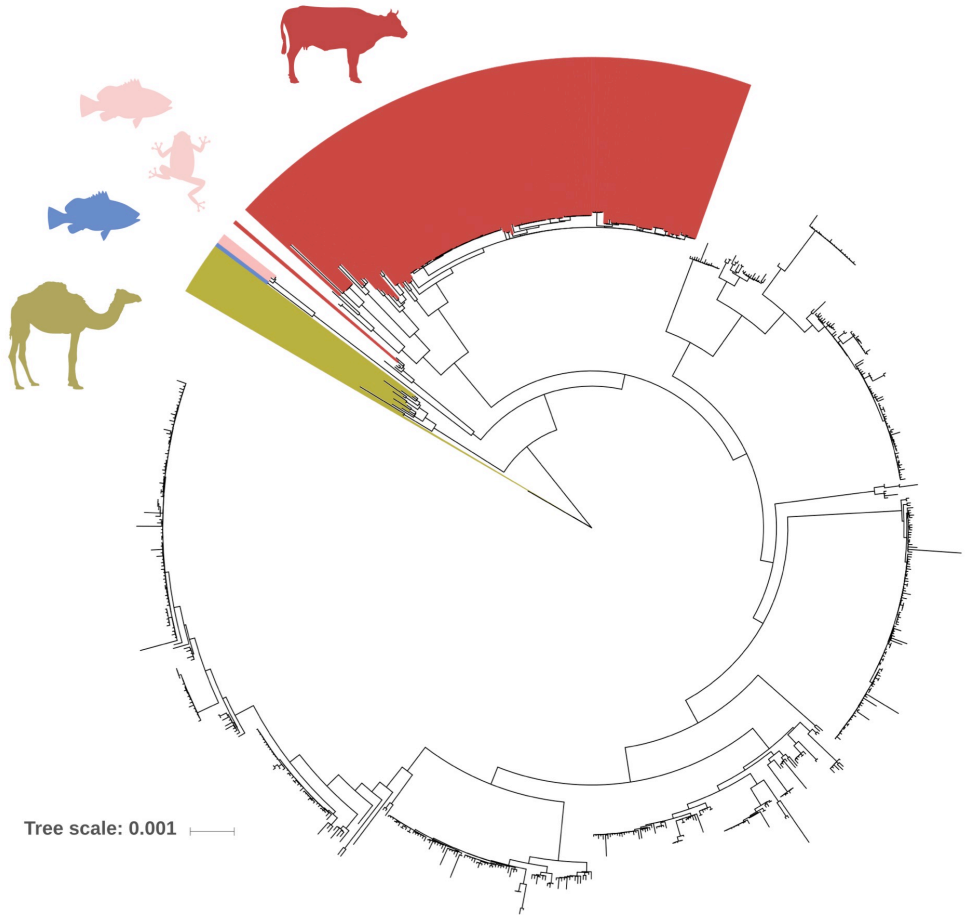


Figure 4. FastTree phylogenetic tree based on core-genome analysis of 796 Group B *Streptococcus* (GBS) whole genome sequences. Individual GBS strains are annotated with ribosomal subunit proteins (rsp)-profiles, which are distinct for GBS genotypes of obligate animal origin. Red: rsp-profiles 1, 10, 12, 17, 19, 38, 39, 59, 60 and 61 are exclusively found in bovine isolates. Blue: rsp-profile 49 is unique for fish origin. Rose: rsp-profile 13 is indicative of either fish or frog origin. Khaki: rsp-profiles 11, 15, 18, 20, 21 and 37 are exclusively found in camel isolates. (Scale bar: nucleotide substitutions per site).

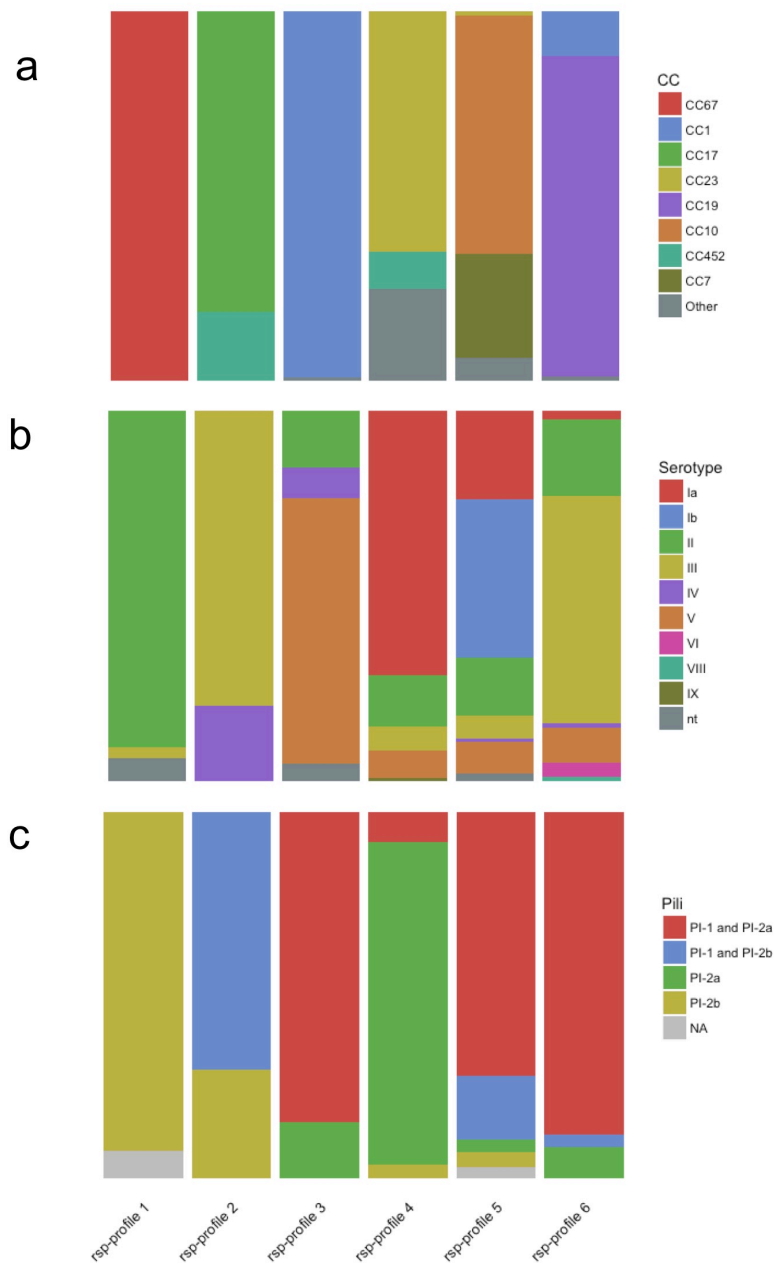


Figure 5. The global major ribosomal subunit proteins (rsp)-profiles 1-6 provide a probabilistic value regarding the Group B *Streptococcus* genotypes' associated (a) multi locus sequence typing clonal complex (CC), (b) capsular serotype and (c) pilus variants. nt: non-typeable.

Table 1. Validation of the established ribosomal subunit protein (rsp) typing scheme with MALDI-TOF MS measurements of in-house group B *Streptococcus* (GBS) isolates.

a) MALDI-TOF MS measurements of 37 isolates with available whole genome sequence (WGS) data.

n *isolates*	*source*	*MALDI assigned rsp-profile*	in silico *assigned rsp-profile*	n *correctly assigned*
3	Human	rsp-profile 2	rsp-profile 2	3
1	Human	rsp-profile 3	rsp-profile 3	1
1	Human	rsp-profile 4	rsp-profile 4	1
3	Human	rsp-profile 5	rsp-profile 5	3
1	Human	rsp-profile 6	rsp-profile 6	1
5	Camel	rsp-profile 11	rsp-profile 11	5
3	Camel	rsp-profile 15	rsp-profile 15	3
2	Camel	rsp-profile 18	rsp-profile 18	2
2	Bovine	rsp-profile 19	rsp-profile 19	2
2	Camel	rsp-profile 20	rsp-profile 20	2
2	Camel	rsp-profile 21	rsp-profile 21	2
1	Camel	rsp-profile 22	rsp-profile 22	1
1	Human	rsp-profile 22	rsp-profile 22	1
1	Camel	rsp-profile 37	rsp-profile 37	1
1	Human	rsp-profile 6	rsp-profile 55	0
8 *	Fish	rsp-profile 5	rsp-profile 5	8
37/37 assigned (100% sensitivity)			**specificity:**	**97% (36/37)**

* isolates typed in an external laboratory on a different MALDI-TOF MS system

b) MALDI-TOF MS measurements of 219 isolates without WGS available.

n *assigned*	*source*	*MALDI assigned rsp-profile*
33	Human	rsp-profile 2
16	Human	rsp-profile 3
49	Human	rsp-profile 4
22	Human	rsp-profile 5
30	Human	rsp-profile 6
1	Human	rsp-profile 7
22	Camel	rsp-profile 11
20	Camel	rsp-profile 15
17	Camel	rsp-profile 18
2	Bovine	rsp-profile 19
2	Camel	rsp-profile 20
1	Camel	rsp-profile 21
1	Camel	rsp-profile 37
0 (3)**	Human	new profile
216/219 assigned (99% sensitivity)		

** three isolates displayed novel rsp-profiles not yet contained in our reference database

MATERIAL AND METHODS

GBS WGS accession and metadata

A total of 876 WGS were obtained on 24-07-2017 from the National Center for Biotechnology Information (NCBI) genome database. Since sequence quality did not allow prediction of all 28 rsp molecular masses 98 WGS were removed from the dataset. Together with 18 in-house sequenced isolates, the final dataset consisted of 796 WGS. The dataset contained data from GBS genotypes isolated over a long time span, with collection time points dating from 1934 to 2016. The strains stem from various geographic regions in Africa, the Americas, Asia, Australia and Europe. The majority of the strains were isolated from human (n = 543) or cattle (n = 187), with the remaining genotypes isolated from fish (n = 25), camel (n = 16), rat (n = 7), seal (n = 5), dog (n = 4), frog (n = 2), dolphin (n = 1), or unknown origin (n = 6). Of the strains for which information of the host health was available, 449 were reported to be associated with disease, while 277 occurred as non-disease-causing colonizer. Comprehensive metadata of all WGS is provided in Supplementary Table S1.

GBS whole-genome sequencing

Genomic DNA was extracted using the QIAamp DSP DNA minikit (Qiagen, Hilden, Germany). A first batch of three isolates was processed as described in Rothen *et al*. (34). For a second batch of 15 isolates, paired-end libraries constructed by the Nextera XT DNA library prep kit (Illumina, San Diego, CA) were sequenced on a MiSeq system (Illumina) using a 600-cycle MiSeq reagent kit v3 (Illumina). *De novo* assemblies were created using SeqMan NGen from the Lasergene genomics package version 12.1.0 (DNAStar, Madison, WI) with standard settings. Comprehensive WGS information including accession numbers are provided in Supplementary Table S1.

Core-genome phylogenetic analysis

Automatic genome annotation of the WGS was performed with the Prokka software tool version 1.12 (35), using a *Streptococcus* genus database. The core-genome phylogenetic relationships of the WGS were obtained using EDGAR version 2.2 (24). Detailed information on the core-genome analysis is provided as Supplementary text.

In silico capsular serotyping, MLST and ANI analysis

In silico capsular typing was performed as described by Sheppard *et al*. (36) and MLST using a custom R script, accessing the query references of the seven housekeeping genes from the PubMLST database (https://pubmlst.org/). Average Nucleotide Identity (ANI) analysis was carried out using the Python module PYANI (https://github.com/widdowquinn/pyani), applying the ANIm method. ANI calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing core facility at University of Basel.

In silico typing of GBS surface protein variants

tBLASTn analyses were carried out for an *in silico* variant typing of five major GBS surface proteins. For variant typing of the laminin-binding protein (Lmb) and the surface immunogenic protein (Sip), one query sequence was used for BLAST and the identified protein variants assigned an allele number in decreasing order of frequency. Variant-specific protein sequences published by Creti *et al*. (37) were used as query files for the alpha-like protein (Alp) gene family. For the surface protein gbs2018 (BibA), variant-specific query sequences described by Springman *et al*. (38) were used. Distribution of pilus islands (PIs) types among the WGS was determined using representative sequences of the three described variants PI-1, PI-2a and PI-2b (39). A summary of the protein used as queries, their respective sequence accession numbers and the thresholds used to retain BLAST hits is provided as Supplementary Table S7.

In silico molecular weight prediction of ribosomal subunit proteins

The theoretical monoisotopic molecular weights of ribosomal subunit proteins were predicted using an in-house Python bioinformatic pipeline. Post-translational modifications, specifically N-terminal methionine loss and methylation, were taken into account. tBLASTn analyses were carried out for an *in silico* typing of the rsp in 796 GBS WGS. Based on the predicted 28 rsp masses in our collection, we assessed the variability of each mass (mass alleles) and defined unique combinations of mass alleles (rsp-profiles) across the WGS, taking into account the MALDI-TOF MS detection threshold of 400 ppm.

GBS isolates used for MALDI-TOF MS analyses

The 248 GBS isolates used in this study were obtained from four different sources: (i) 156 human isolates belonging to a set of *S. agalactiae* strains described by Huber *et al*. in 2011 (40). These both inpatient and outpatient samples were obtained and cultivated at the Aga Khan University Hospital in Nairobi, Kenya between January 2007 and June 2010; (ii) 79 samples from the International Livestock Research Institute (ILRI) isolated from camels in Kenya and Somalia (4, 34, 41); (iii) Six GBS samples from cattle, isolated during 2009 in Switzerland by Prof. J. Frey from the University of Bern (unpublished); (iv) Seven human GBS reference strains were provided by Dr. H. Tettelin from the University of Maryland (26). More comprehensive information of all analyzed GBS isolates is provided in Supplementary Table S6.

Bacteria cultivation and sample preparation

GBS bacteria were stored at -80 °C and transferred to blood agar for overnight growth prior to MALDI-TOF MS measurements. The pre-processing of the GBS bacterial samples followed a protocol established in this study. Briefly, a standardized amount of bacterial cells was first subjected to several washing steps and subsequently mechanically ruptured by bead beating. The bacterial protein cocktail was then transferred to an ultrafiltration column (Amicon®, Sigma-Aldrich), for removal of molecules with a molecular mass below 3,000 Da. The final protein solution was spotted in quadruplicates on a steel MALDI-TOF target plate and overlaid with 1 µl sinapic acid matrix solution. The detailed pre-processing protocol is provided as Supplementary text.

MALDI-TOF MS analyses

The MS measurements were carried out using a MALDI-TOF Mass Spectrometer Axima Confidence machine (Shimadzu-Biotech, Kyoto, Japan). Detailed information on instrument setup, mass spectra processing and internal calibration is provided as Supplementary text. An ascii file containing the recalibrated protein mass values and corresponding intensities was automatically generated for every GBS isolate.

Classification of mass spectra

The mass lists were classified using a custom Python script. Briefly, all mass list entries were queried against the *in silico* predicted mass alleles of 28 rsp and the thereby generated sequence of mass alleles matched against the reference library containing the 62 defined rsp-profiles. A mass list was assigned an rsp-profile identification (ID) if (i) there was one single top matching reference and (ii) if at least 24 rsp masses could be detected. An isolate was assigned a final rsp-profile ID if (i) at least two of the four technical replicate mass lists were assigned the same rsp-profile and (ii) if there was no contradicting match with a different rsp-profile in the other technical replicate mass lists. If a specific rsp was missing in all mass lists considered for the final ID of an isolate, a warning message was generated, indicating the possibility of a new rsp-profile not yet contained in the database.

Confirmation of inter-laboratory reproducibility of our method

In order to confirm the inter-site transferability and reproducibility of our method, additional MALDI-TOF MS analyses were performed in an independent laboratory. Eight GBS isolates were cultivated and pre-processed following our established protocol. The MALDI-TOF measurements were carried out on a Microflex machine (Bruker Daltonics, Bremen, Germany), with the instrument parameter settings adjusted for the use of sinapinic acid. Spectra post-processing, internal calibration, rsp prediction and classification was carried out in an automated way using our custom R and python scripts as described above. WGS data of the eight GBS isolates were available and used *in silico* to confirm the molecular masses of the 28 rsp.

ACKNOWLEDGEMENTS

Dr. H. Tettelin, Prof. J. Frey, Dr. Mario Younan and Prof. Gunturu Revathi for providing *S. agalactiae* isolates. Dr. S. D. Manning, Michigan State University for providing metadata of ~80 GBS genomes.

REFERENCES

1. Russell NJ, *et al*. (2017) Maternal Colonization With Group B *Streptococcus* and Serotype Distribution Worldwide: Systematic Review and Meta-analyses. Clin Infect Dis - 65(Suppl 2):S100–S111.

2. Sørensen UBS, Poulsen K, Ghezzo C, Margarit I, Kilian M (2010) Emergence and Global Dissemination of Host-Specific *Streptococcus agalactiae* Clones. mBio 1(3).

3. Da Cunha V, *et al*. (2014) *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline. Nat Commun 5:4544.

4. Fischer A, *et al*. (2013) Camel *Streptococcus agalactiae* populations are associated with specific disease complexes and acquired the tetracycline resistance gene tetM via a Tn916-like element. Vet Res 44:86.

5. Godoy DT, *et al*. (2013) Genetic diversity and new genotyping scheme for fish pathogenic *Streptococcus agalactiae*. Lett Appl Microbiol 57(6):476–483.

6. Chen VL, Avci FY, Kasper DL (2013) A Maternal Vaccine against Group B *Streptococcus*: Past, Present, and Future. Vaccine 31(0 4):D13–D19.

7. Slotved H-C, Kong F, Lambertsen L, Sauer S, Gilbert GL (2007) Serotype IX, a Proposed New *Streptococcus agalactiae* Serotype. J Clin Microbiol 45(9):2929–2936.

8. Fluegge K, Supper S, Siedler A, Berner R (2005) Serotype Distribution of Invasive Group B Streptococcal Isolates in Infants: Results from a Nationwide Active Laboratory Surveillance Study over 2 Years in Germany. Clin Infect Dis 40(5):760–763.

9. Jones N, *et al*. (2003) Multilocus Sequence Typing System for Group B *Streptococcus*. J Clin Microbiol 41(6):2530–2536.

10. Baker CJ, Rench MA, McInnes P (2003) Immunization of pregnant women with group B streptococcal type III capsular polysaccharide-tetanus toxoid conjugate vaccine. Vaccine 21(24):3468–3472.

11. Heyderman RS, *et al*. (2016) Group B *Streptococcus* vaccination in pregnant women with or without HIV in Africa: a non-randomised phase 2, open-label, multicentre trial. Lancet Infect Dis 16(5): 546–555.

12. Leroux-Roels G, *et al*. (2016) A randomized, observer-blind Phase Ib study to identify formulations and vaccine schedules of a trivalent Group B *Streptococcus* vaccine for use in non-pregnant and pregnant women. Vaccine 34(15):1786–1791.

13. Kobayashi M, *et al*. (2016) WHO consultation on group B *Streptococcus* vaccine development: Report from a meeting held on 27–28 April 2016. Vaccine.

14. Nuccitelli A, *et al*. (2011) Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B *Streptococcus* infections. Proc Natl Acad Sci U S A 108(25):10278–10283.

15. MinervaX announces positive data from Phase I clinical trial. Press release. January 5th, 2017. Available at: http://minervax.com/news/2017/1/5/minervax-announces-positive-data-from-phase-i-clinical-trial.html.

16. Bellais S, *et al*. (2012) Capsular switching in group B *Streptococcus* CC17 hypervirulent clone: a future challenge for polysaccharide vaccine development. J Infect Dis 206(11):1745– 1752.

17. Singhal N, Kumar M, Kanaujia PK, Virdi JS (2015) MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front Microbiol 6.

18. Seng P, *et al*. (2009) Ongoing Revolution in Bacteriology: Routine Identification of Bacteria by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. Clin Infect Dis 49(4):543–551.

19. Body BA, *et al*. (2018) Evaluation of the Vitek MS v3.0 Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry System for Identification of *Mycobacterium* and *Nocardia* Species. J Clin Microbiol 56(6).

20. van Belkum A, Welker M, Pincus D, Charrier J-P, Girard V (2017) Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry in Clinical Microbiology: What Are the Current Issues? Ann Lab Med 37(6):475–483.

21. Suarez S, *et al*. (2013) Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory. J Microbiol Methods 94(3):390–396.

22. Ziegler D, *et al*. (2015) Ribosomal protein biomarkers provide root nodule bacterial identification by MALDI-TOF MS. Appl Microbiol Biotechnol 99(13):5547–5562.

23. Jolley KA, Maiden MC (2010) BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11(1):595.

24. Blom J, *et al*. (2016) EDGAR 2.0: an enhanced software platform for comparative gene content analyses. Nucleic Acids Res 44(W1):W22-28.

25. Furfaro LL, Chang BJ, Payne MS (2018) Perinatal *Streptococcus agalactiae* Epidemiology and Surveillance Targets. Clin Microbiol Rev 31(4):e00049-18.

26. Tettelin H, *et al*. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome." Proc Natl Acad Sci U S A 102(39):13950–13955.

27. Lartigue M-F, *et al*. (2011) Rapid detection of "highly virulent" Group B *Streptococcus* ST-17 and emerging ST-1 clones by MALDI-TOF mass spectrometry. J Microbiol Methods 86(2):262–265.

28. Lin H-C, *et al*. (2017) Identification of a proteomic biomarker associated with invasive ST1, serotype VI Group B *Streptococcus* by MALDI-TOF MS. J Microbiol Immunol Infect.

29. Campisi E, *et al*. (2016) Serotype IV *Streptococcus agalactiae* ST-452 has arisen from large genomic recombination events between CC23 and the hypervirulent CC17 lineages. Sci Rep 6.

30. Donders GG g, *et al*. (2016) Maternal Immunization With an Investigational Trivalent Group B Streptococcal Vaccine: A Randomized Controlled Trial. Obstet Gynecol 127(2):213–221.

31. Weinberger DM, Malley R, Lipsitch M (2011) Serotype replacement in disease following pneumococcal vaccination: A discussion of the evidence. Lancet 378(9807):1962–1973.

32. Devine VT, *et al*. (2017) The rise and fall of pneumococcal serotypes carried in the PCV era. Vaccine 35(9):1293–1298.

33. Miller E, Andrews NJ, Waight PA, Slack MP, George RC (2011) Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. Lancet Infect Dis 11(10):760–768.

34. Rothen J, *et al*. (2017) Draft Genome Sequences of Seven *Streptococcus agalactiae* Strains Isolated from *Camelus dromedarius* at the Horn of Africa. Genome Announc 5(28):e00525-17.

35. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl 30(14):2068–2069.

36. Sheppard AE, *et al*. (2016) Capsular Typing Method for *Streptococcus agalactiae* Using Whole-Genome Sequence Data. J Clin Microbiol 54(5):1388–1390.

37. Creti R, Fabretti F, Orefici G, von Hunolstein C (2004) Multiplex PCR Assay for Direct Identification of Group B Streptococcal Alpha-Protein-Like Protein Genes. J Clin Microbiol 42(3):1326–1329.

38. Springman AC, *et al*. (2009) Selection, Recombination, and Virulence Gene Diversity among Group B Streptococcal Genotypes. J Bacteriol 191(17):5419–5427.

39. Martins ER, Andreu A, Melo-Cristino J, Ramirez M (2013) Distribution of Pilus Islands in *Streptococcus agalactiae* That Cause Human Infections: Insights into Evolution and Implication for Vaccine Development. Clin Vaccine Immunol CVI 20(2):313–316.

40. Huber CA, McOdimba F, Pflueger V, Daubenberger CA, Revathi G (2011) Characterization of Invasive and Colonizing Isolates of *Streptococcus agalactiae* in East African Adults. J Clin Microbiol 49(10):3652–3655.

41. Zubair S, *et al*. (2013) Genome Sequences of Two Pathogenic *Streptococcus agalactiae* Isolates from the One-Humped Camel *Camelus dromedarius*. Genome Announc 1(4).

Chapter 5

Draft Genome Sequences of Seven *Streptococcus agalactiae* Strains isolated from *Camelus dromedarius* at the Horn of Africa

---

This chapter contains the following publication:

Julian Rothen[1,2], Tobias Schindler[1,2], Joël F. Pothier[3], Mario Younan[4], Ulrich Certa[5], Claudia Daubenberger[1,2], Valentin Pflüger[6] and Joerg Jores[7,8]. "Draft Genome Sequences of Seven *Streptococcus agalactiae* Strains isolated from *Camelus dromedarius* at the Horn of Africa". 2017. *Genome Announc*.

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Draft Genome Sequences of Seven *Streptococcus agalactiae* Strains Isolated from *Camelus dromedarius* at the Horn of Africa

Julian Rothen,[a,b] Tobias Schindler,[a,b] Joël F. Pothier,[c] Mario Younan,[d] Ulrich Certa,[e] Claudia Daubenberger,[a,b] Valentin Pflüger,[f] Joerg Jores[g,h]

Swiss Tropical and Public Health Institute, Basel, Switzerland[a]; University of Basel, Basel, Switzerland[b]; Environmental Genomics and Systems Biology Research Group, Institute for Natural Resource Sciences, Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland[c]; Food and Agriculture Organization of the United Nations, Gaziantep, Turkey[d]; Roche Pharmaceutical Research and Early Development, Department of Pharmaceutical Sciences, Translational Technologies and Bioinformatics, Roche Innovation Center Basel, Basel, Switzerland[e]; Mabritec AG, Riehen, Switzerland[f]; International Livestock Research Institute, Nairobi, Kenya[g]; Vetsuisse-Fakultät Universität Bern, Institut für Veterinär-Bakteriologie, Bern, Switzerland[h]

ABSTRACT We present draft whole-genome sequences of seven *Streptococcus agalactiae* strains isolated from *Camelus dromedarius* in Kenya and Somalia. These data are an extension to the group B *Streptococcus* (GBS) pangenome and might provide more insight into the underlying mechanisms of pathogenicity and antibiotic resistance of camel GBS.

Received 22 May 2017 **Accepted** 25 May 2017 **Published** 13 July 2017

Citation Rothen J, Schindler T, Pothier JF, Younan M, Certa U, Daubenberger C, Pflüger V, Jores J. 2017. Draft genome sequences of seven *Streptococcus agalactiae* strains isolated from *Camelus dromedarius* at the Horn of Africa. Genome Announc 5:e00525-17. https://doi.org/10.1128/genomeA.00525-17.

Address correspondence to Julian Rothen, julian.rothen@unibas.ch.

The natural colonizer of human gastrointestinal and genitourinary tracts *Streptococcus agalactiae*, also known as Lancefield's group B *Streptococcus* (GBS), is an emerging pathogen of serious clinical concern (1). As a main causative agent of meningitis, sepsis, and respiratory diseases in neonates, GBS is strongly linked to child mortality and morbidity (2). *S. agalactiae* has also been isolated from both healthy and diseased camels in countries from the Horn of Africa (3–7). Given the fundamental role of camels for human nutrition and financial safety in these regions, GBS-associated diseases, such as mastitis or udder abscesses resulting in significant losses in milk production, can have a devastating impact (5). Here, we report the whole-genome sequences of seven GBS strains, isolated from Kenyan and Somalian camels (*Camelus dromedarius*). Previous genomic analysis of these isolates by multilocus sequence typing (MLST) indicated a detached phylogenetic relationship compared to GBS strains of human or bovine origin (5). The three isolates ILRI025, ILRI030, and ILRI067 were isolated from healthy camels, while ILRI037 (causing gingivitis), ILRI054 (causing wound infection), ILRI120 (causing chronic cough), and ILRI127 (causing periarthricular abscess) were associated with disease.

Genomic DNA was extracted from a single bacterial colony cultivated on Columbia sheep blood agar using the QIAamp DSP DNA minikit (Qiagen, Hilden, Germany). DNA was fragmented by ultrasonication using the Covaris S2 instrument (Covaris, Inc., Woburn, MA, USA). Barcoded libraries were generated with the Ion fragment library kit and Ion Xpress DNA barcode adaptors (Life Technologies, Inc., Carlsbad, CA, USA). Sequencing was performed on an Ion Torrent Personal Genome Machine (PGM) system, with the Ion PGM sequencing 400 kit and the Ion 318 Chip version 2 (Life Technologies, Inc.). After sequencing, single processing and base calling were performed using Torrent Suite 3.6 (Life Technologies, Inc.), and barcode-separated FASTQ files were generated. For *de novo* assemblies, we used MIRA version 4.0 (8). Contigs were sorted along the already published (9) GBS genomes of ILRI112 (accession no. HF952106) and

TABLE 1 List of *Streptococcus agalactiae* draft whole genomes released to GenBank

Strain	GenBank accession no.	Multilocus ST[a]	Serotype	Genome size (bp)	No. of proteins
ILRI025	NDGG00000000	610	VI	2,013,384	1,876
ILRI030	NDGF00000000	617	VI	1,999,626	1,883
ILRI037	NDGE00000000	612	Ia	2,020,002	1,895
ILRI054	NDGD00000000	615	II	2,021,031	1,867
ILRI067	NDGC00000000	614	V	1,980,469	1,812
ILRI120	NDGB00000000	618	Ia	2,049,911	1,954
ILRI127	NDGA00000000	613	Ia	1,973,342	1,875

[a]ST, sequence type.

ILRI005 (accession no. HF952105) (only for isolate ILRI067) using the Move Contigs function in Mauve version 2.3.1 (10). SeqMan Pro from the Lasergene genomics package version 12.1.0 (DNAStar, Madison, WI) was used to check and manually close gaps between contigs. Genome annotation was added using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). The seven genomes displayed an overall size between 1,973,342 and 2,049,911 bp, with 1,812 to 1,954 proteins detected (Table 1).

The draft genome sequences of cameloid GBS isolates presented here are a valuable addition to the pangenome of *S. agalactiae* (11). These genomic data provide a basis for the investigation of adaptive factors in GBS host colonization as well as underlying mechanisms of antibiotic resistance development and pathogenicity of camel *S. agalactiae*.

Accession number(s). The annotated draft whole-genome sequences of the seven *S. agalactiae* isolates were deposited in GenBank under BioProject no. PRJNA382326. The accession numbers for each isolate are shown in Table 1.

ACKNOWLEDGMENTS

This work was supported by funding from the Swiss Tropical and Public Health Institute and Mabritec AG.

REFERENCES

1. Huber CA, McOdimba F, Pflueger V, Daubenberger CA, Revathi G. 2011. Characterization of invasive and colonizing isolates of *Streptococcus agalactiae* in East African adults. J Clin Microbiol 49:3652–3655. https://doi.org/10.1128/JCM.01288-11.
2. Schrag SJ, McGee L, Verani J. 2010. Prevention of perinatal group B streptococcal disease—revised guidelines from CDC. MMWR Morb Mortal Wkly Rep 59:1–32.
3. Edelstein RM, Pegram RG. 1974. Contagious skin necrosis of Somali camels associated with *Streptococcus agalactiae*. Trop Anim Health Prod 6:255–256. https://doi.org/10.1007/BF02383286.
4. Bekele T, Molla B. 2001. Mastitis in lactating camels (*Camelus dromedarius*) in Afar Region, north-eastern Ethiopia. Berl Münch Tierarztl Wochenschr 114:169–172.
5. Fischer A, Liljander A, Kaspar H, Muriuki C, Fuxelius HH, Bongcam-Rudloff E, de Villiers EP, Huber CA, Frey J, Daubenberger C, Bishop R, Younan M, Jores J. 2013. Camel *Streptococcus agalactiae* populations are associated with specific disease complexes and acquired the tetracycline resistance gene *tetM* via a Tn*916*-like element. Vet Res 44:86. https://doi.org/10.1186/1297-9716-44-86.
6. Younan M, Bornstein S. 2007. Lancefield group B and C streptococci in East African camels (*Camelus dromedarius*). Vet Rec 160:330–335. https://doi.org/10.1136/vr.160.10.330.
7. Abera M, Abdi O, Abunna F, Megersa B. 2010. Udder health problems and major bacterial causes of camel mastitis in Jijiga, Eastern Ethiopia: implication for impacting food security. Trop Anim Health Prod 42:341–347. https://doi.org/10.1007/s11250-009-9424-6.
8. Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer science and biology. Proceedings of the German Conference on Bioinformatics, GCB '99. GCB, Hannover, Germany.
9. Zubair S, de Villiers EP, Younan M, Andersson G, Tettelin H, Riley DR, Jores J, Bongcam-Rudloff E, Bishop RP. 2013. Genome sequences of two pathogenic *Streptococcus agalactiae* isolates from the one-humped camel *Camelus dromedarius*. Genome Announc 1(4):e00515-13. https://doi.org/10.1128/genomeA.00515-13.
10. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704.
11. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome." Proc Natl Acad Sci U S A 102:13950–13955. https://doi.org/10.1073/pnas.0506758102.

Chapter 6

Tracing and monitoring of emerging Group B *Streptococcus* genotypes with zoonotic potential in Hong Kong

---

This chapter contains the following working manuscript:

Julian Rothen, Dulmini Nanayakkara, Carmen Li, Frédéric Foucault, Valentin Pflüger, Claudia Daubenberger and Margaret Ip. "Tracing and monitoring of emerging Group B *Streptococcus* genotypes with zoonotic potential in Hong Kong". 2018.

Tracing and monitoring of emerging group B *Streptococcus* genotypes with zoonotic potential in Hong Kong

Julian Rothen; JR [1,2*], Dulmini Nanayakkara; DN [3*], Carmen Li; CL [3], Frédéric Foucault; FF [4], Valentin Pflüger; VP [4*], Claudia A. Daubenberger; CAD [1,2], Margaret Ip; MI [3]

1 Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute (Swiss TPH) Basel, 4002 Basel, Switzerland

2 University of Basel, 4002 Basel, Switzerland

3 Department of Microbiology, The Chinese University of Hong Kong (CUHK), Shatin, NT, Hong Kong

4 Mabritec AG, 4125 Riehen, Switzerland

* these authors contributed equally

ABSTRACT

Streptococcus agalactiae (Group B *Streptococcus*, GBS), a frequent colonizer of the human gastrointestinal and genital tracts, is a leading cause of neonatal meningitis and an emerging infectious pathogen in non-pregnant adults. GBS possesses a broad animal host spectrum, including pigs and fish and there is increasing evidence that human invasive disease caused by atypical GBS genotypes can occur through animal sources, e.g. through food-borne, zoonotic infections. In this study, we made use of a previously described matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) typing method, which, based on molecular variations of 28 ribosomal subunit proteins (rsp), classifies GBS genotypes into major phylogenetic lineages. We found that 170 GBS strains isolated from adult hospital patients in Hong Kong, can be readily assigned by MALDI-TOF MS into five globally dominant rsp-profiles, allowing reliable prediction regarding their evolutionary background and capsular serotype. We further demonstrate that MALDI-TOF MS allows for high-throughput screening and detection of novel GBS genotypes, which we found to predominantly arise from the under-researched pig and fish host reservoirs. In conclusion, we confirm here the inter-laboratory transferability of an rsp-biomarker based MALDI-TOF MS typing method, its capability in discriminating between GBS genotypes of the major global phylogenetic lineages and its potential for rapid screening of hundreds of GBS isolates for the surveillance of emerging GBS genotypes.

INTRODUCTION

Streptococcus agalactiae (Group B *Streptococcus*, GBS), a beta-hemolytic and gram-positive bacterium, is a frequent colonizer of the human gastrointestinal and genitourinary tracts [1]. GBS also possesses a broad animal host spectrum including cattle [2], pigs [3], camel [4] and various freshwater fish species [5, 6]. Besides of its status as a leading cause of neonatal meningitis and sepsis [7], GBS is an emerging infectious disease in non-pregnant adults, especially in the elderly and immuno-compromised individuals [1, 8]. GBS disease in non-pregnant adults prerequisites the switch of GBS from a harmless commensal to an invasive pathogen, a mechanism that remains poorly understood [9]. However, there is increasing evidence that infection can also occur through nosocomial or food-borne e.g. zoonotic infection [10, 11].

The GBS clone ST283 impressively exemplifies the threat of zoonotic GBS infection in adults. Between 1993-2012, this clone accounted for 27.4% of serotype III caused cases of invasive disease in non-pregnant adults in Hong Kong [12, 13]. Interestingly, ST283 was never found as colonizer in women, but was described as disease-causing strain in farmed freshwater fish [6], suggesting the potential zoonotic hazard of raw fish caused by this genotype. Conclusive proof followed in 2015, when ST283 was linked to a massive outbreak of severe adult GBS disease in Singapore, unequivocally linked to consumption of raw fish [10]. Genomic analysis of human and fish ST283 strains later confirmed freshwater fish as reservoir of ST283, declaring this zoonotic clone a major infectious disease threat [14]. Large-scale epidemiological monitoring studies will be essential to gain insight into GBS transmission dynamics, in particular regarding the significance of animal reservoirs for emerging hyper-virulent GBS clones.

Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) has developed into a widely used method for high-throughput species identification in routine diagnostics [15, 16]. Classifying microbial species based on their highly specific whole-cell peptide fingerprint, MALDI-TOF MS can be used to accurately discriminate between thousands of bacteria, including GBS [17, 18]. The genetic diversity of GBS was found to be concurrent with variations in the protein patterns measured by MALDI-TOF MS [19], which can be exploited for sub-species level discrimination of GBS strains [20]. We recently expanded on these findings and showed that the highly conserved ribosomal subunit proteins (rsp) serve as ideal biomarkers for strain level typing of GBS (Rothen *et al.* manuscript submitted). Specifically, indexing the mass variations of pre-defined 28 rsp using

MALDI-TOF MS allows for classification of GBS isolates into the major phylogenetic lineages, detection of hypervirulent CC17 strains and identification of obligate animal associated lineages (Rothen *et al.* manuscript submitted).

In the study presented here, we aimed to apply the rsp-based MALDI-TOF MS approach to (i) identify the major GBS genotypes circulating among patients with admission to the Prince of Wales Hospital in Hong Kong between 2010-2018 and (ii) to identify the major GBS genotypes found in fish and pig meat samples collected from Hong Kong wet markets. This analyses will provide insights on whether (a) GBS genotypes circulating in hospitalized humans and food animals in the Hong Kong area differ and (b) MALDI-TOF MS can be used as a high-throughput and cost efficient screening tool for monitoring of emerging, potential zoonotic GBS clones. Here, we have analyzed a collection of 249 GBS isolates using our novel sample preparation and bioinformatics pipeline and demonstrate that our typing approach based on measurement of 28 rsp can be easily transferred and deployed in collaborating microbiological laboratories.

MATERIAL AND METHODS

Origin of Group B *Streptococcus* strains analyzed in this study

For this study, a total of 249 GBS samples were analyzed that originated from various sources. Sixty-three GBS strains stem from tilapias (*Oreochromis mossambicus*) and big head carps (*Hypophthalmichthys nobilis*). Fish specimens were collected between 2016-2017 from twelve wet markets across the Hong Kong areas New Territories, Kowloon and Hong Kong Island. Bacteria were isolated from various body parts including the heart, liver, pancreas, spleen, gills, flesh, minced meat and the skin. Fifteen GBS strains were isolated from domestic pigs (*Sus scrofa domesticus*). Pig meat specimens were collected in 2018 from four wet markets located in the Hong Kong areas New Territories and Hong Kong Island. Bacteria were isolated from the tongue, the small intestine and from minced meat. One hundred seventy-one GBS human strains were collected between January 2018 to May 2018 from adult patients at the Prince of Wales Hospital in Shatin, Hong Kong.

Isolation of bacteria from animal and human samples

From the collected animal and human samples, bacteria were inoculated into Todd-Hewitt enrichment broth with a subsequent transfer to chromogenic selective medium. Colonies thereby identified as GBS were cultured on Blood agar (OXOID, Basingstoke, Britain) at 37 °C with 5 % CO^2 and single colonies stored at -80 °C until further use.

Capsular serotyping by multiplex-PCR

DNA extraction was done by emulsifying two to four bacterial colonies in 200 µl lysis buffer (0.25 % sodium dodecyl sulfate, 0.05 N NaOH) at 95 °C for 5 minutes, followed by centrifugation at 16,000 g for 5 minutes [21]. The supernatants were directly used as template DNA and were stored at -20 °C until further analysis. Serotyping was performed by a multiplex PCR method described by Imperi *et al.* [22]. Briefly, PCR products were subjected to gel electrophoresis, performed on 1.5 % agarose gels in 0.5 X TBE (45 mM Tris-HCl, 45 mM boric acid, 1 mM EDTA) buffer at 150 V for 50 minutes and bands were visualized using Gel Doc (BioRad Laboratories, London, UK). Isolates that failed to be assigned to a serotype by visual assessment of the PCR bands were grouped as non-typeable.

MALDI-TOF MS analyses

Sample pre-processing

For each GBS strain, single colonies were used to prepare agar slants for interim storage prior to MALDI-TOF MS analysis. On the day before mass spectrometry measurements, bacteria were transferred to blood agar for overnight growth. The pre-processing of the GBS bacterial samples followed the protocol established by Rothen *et al.* (manuscript submitted). Briefly, a standardized amount of bacterial cells was first subjected to several washing steps and subsequently mechanically ruptured by bead beating. The bacterial protein cocktail was then transferred to an ultrafiltration column (Amicon®, Sigma-Aldrich) for removal of molecules with a molecular mass below 3,000 Da. The final protein solution was spotted in quadruplicates on a steel MALDI-TOF MS target plate and overlaid with 1 µl sinapic acid matrix solution.

Microflex instrument setup

The MS measurements were carried out on a microflex LT MALDI-TOF MS system (Bruker Daltonics, Bremen, Germany) with detection in the linear mode, allowing the interrogation of high molecular weight samples. The analysis was carried out in the mass range between 3,000 and 25,000 Da. The instrument parameter settings were adjusted to be on par with the use of sinapic acid matrix. The linear detector voltage was set to 1,943 V, with a laser frequency of 66.7 Hz, initial laser power of 70 %, maximal laser power of 90 % and laser attenuation offset of 60 %. For each spectrum, 1,000 laser shots in 100 shot steps were acquired (flexControl software 2.0, Bruker Daltonics) in a random walk movement, thereby ensuring an even measurement covering the entire area of the sample spot. Each target plate was externally calibrated using the reference spectra of *Escherichia coli* strain DH5α.

Spectra post-processing and internal calibration

Post-processing of the raw mass spectra (fid files) was carried out using a custom R script, building on the R package "MALDIquant" [23]. Briefly, peak intensities were square root transformed and smoothed using the "SavitzkyGolay" method. Baseline removal was done using the "SNIP" baseline estimation method. Peak detection was carried out using the "MAD" noise estimation method and a signal-to-noise ratio of two. Internal calibration with 800 ppm was carried out using ten GBS rsp masses (three mass alleles of L6, two mass alleles

of L36 and S12, one mass allele of L14, L29 and S15) that altogether display molecular weights distributed over a wide mass range (4,425 to 19,293 Dalton). An ascii file containing the recalibrated protein mass values and corresponding intensities was automatically generated for every mass spectrum.

Classification of mass spectra according to rsp-profile

The generated ascii files were used as input for a custom Python bioinformatics analysis pipeline established by Rothen *et al.* (manuscript submitted). This pipeline allows for automated identification of rsp mass variants contained in the ascii files, matching of rsp mass variant combinations with 62 rsp-profiles deposited in a reference database and final assignment of a rsp-profile to the mass spectrum. We found that the low molecular weight rsp L36 (4,452 Da) and L34 (5,378 Da) were often subjected to a peak shift of a few Dalton, and as a result were missed by our script. We therefore increased the allowed detection mass range for these rsp from 400 ppm to 1500 ppm.

The classification decision steps were defined as follows: (1a) A single mass spectrum was assigned an rsp-profile if there was one single top matching reference. (1b) No rsp-profile was assigned if there were multiple top matching references or if less than 20 rsp were detected in the mass spectrum. (2a) An isolate was assigned a final rsp-profile ID if there was one top matching rsp-profile in one or more of the four replicate spectra. (2b) An isolate was assigned no rsp-profile ID if all four replicate spectra were assigned no ID due to low mass counts or if there were contradicting top matching rsp-profiles among the four spectra. (3) If a specific rsp was missing in all mass spectra considered for the final rsp-profile ID of an isolate, a warning message was generated, indicating the possibility of a new rsp-profile not yet contained in the reference database.

Whole-genome sequencing of GBS strains

Genomic DNA from the GBS strains was extracted with the Wizard® Genomic DNA Purification Kit according to the manufacturer's protocol for gram-positive bacteria (Qiagen, Limburg, Netherlands). Library preparation was done using the Illumina Nextera XT library preparation kit and whole genome sequencing was carried on an Illumina Nextseq 500 system. Genomes were assembled using the metAMOS pipeline (version 1.5rc3). The draft genomes were not yet deposited in the NCBI genome database.

In silico molecular weight prediction of ribosomal subunit proteins

The ten generated GBS whole-genome sequences were used for *in silico* extraction of the nucleotide sequences coding for 28 rsp using tBLASTn analysis. The most frequent post-translational modifications [24], specifically N-terminal methionine loss and methylation, were taken into account for subsequent prediction of the monoisotopic molecular rsp weights.

Average Nucleotide Identity (ANI) analysis

The phylogenetic relationship between the whole-genome sequenced GBS strains of this study and a collection of publicly available GBS WGS was assessed by Average Nucleotide Identity (ANI) analysis [25]. ANI was carried out using the Python module PYANI (https://github.com/widdowquinn/pyani), applying the ANIm method. ANI calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing core facility at University of Basel. Euclidian distance matrix calculation and UPGMA hierarchical cluster analysis was performed using the R stats base package. Phylogenetic trees were edited and visualized using the interactive tree of life (iTOL) website [26].

RESULTS

Capsular Serotyping

PCR analysis revealed that there were six different capsular serotypes contained in our collection of 249 GBS isolates. The human isolates were composed of serotype Ia (*n*=35), Ib (*n*=22), II (*n*=3), III (*n*=62), IV (*n*=9) and V (*n*=12). There were 17 non-typeable (11%) human strains and another 10 strains for which serotyping was not yet carried out. The GBS isolates from fish were composed of serotypes Ia (*n*=32), III (*n*=1), IV (*n*=1) and V (*n*=4). A large proportion of fish GBS strains (40%) were non-typeable by PCR (*n*=25). All pig isolates were found to carry capsular serotype III (*n*=15).

MALDI-TOF MS analyses

Identification of 28 predefined ribosomal subunit proteins in mass spectra

A total of 249 GBS isolates were measured in quadruplicates, totaling to 996 single spectra. Of the twenty-eight ribosomal subunit proteins, which we previously found to be reliably measurable by MALDI-TOF MS (Rothen *et al*. manuscript submitted), twenty-six could be detected on average per spectrum. Of note, there were few low-quality mass spectra (*n*=51 from 41 isolates) with exceptionally small rsp counts (between 7 to 19 rsp). Hence, the median rsp count per measurement was 27 out of the 28 rsp to be detected. With the exception of L19 (found in 50% of spectra), all rsp were found at high levels across the spectra. L13, L14, L17, L18, L23, L29, L30, L33, L35, S16, S19, S21 and S8 were found in between 92-95% of spectra. L21, L22, L24, L32, L34, L36, L6, S10, S12, S13, S15, S17, S18 and S9 were found in more than 95% of spectra. Visual inspection of some mass spectra with missing L19 indicated, that this protein mass was in fact present in the spectra, but its mass peak was rather diffuse and therefore not passing the signal-to-noise threshold set by us.

Assignment of isolates to known rsp-profiles

A first batch of mass spectra generated from 174 isolates allowed the assessment of how well GBS isolates are assigned to known rsp-profiles by MALDI-TOF MS. Classification according to rsp-profile identity initially failed for nine isolates. Upon visual inspection of the mass spectra and manual identification of missing rsp, five of the nine isolates could be assigned to an rsp-profile. In total, 170/174 isolates (98%) were successfully assigned to rsp-

profiles contained in the reference library. The rsp-profiles 2-6 which stand representative for the globally dominant GBS phylogenetic genotype clusters were also most abundant in our collection (Fig. 1a). Of the 170 classified isolates, 38% (*n*=65) were assigned to rsp-profile 5, 26% (*n*=44) to rsp-profile 6, 19% (*n*=32) to rsp-profile 4, 13% (*n*=22) to rsp-profile 2, 5% (*n*=8) to rsp-profile 3 and one isolate each were assigned to rsp-profile 7 and rsp-profile 31. With regards to the isolation source of the 170 GBS isolates, a clear pattern regarding assigned rsp-profile was seen. While isolates of human origin were found to stem from all seven rsp-profiles covered here, fish isolates almost exclusively fell into rsp-profile 5, with one isolate displaying rsp-profile 4 and one with rsp-profile 6, respectively. The sole isolate from pig origin was assigned to rsp-profile 4 (Fig. 1a).

Assignment of isolates to novel rsp-profiles

In all four replicate spectra of 31 isolates, one or more rsp were found to display an rsp mass variant not yet contained in our reference database, indicating the occurrence of novel rsp-profile. Visual inspection of the mass spectra led to the assignment of two isolates to the already known rsp-profiles 4 and 5, respectively. The remaining 29 isolates (12% of the 249 isolates) were found to either display previously unknown rsp mass variants (*n*=27) or new combinations of known rsp mass variants (*n*=2). The novel rsp-profiles contained in these 29 isolates were cross-compared and the distinct profiles termed novel rsp (nrsp)-profile 1 to 12 (Fig. 1b). The most abundant novel profile was nrsp-profile 8 (*n*=13), which was exclusively found in GBS strains isolated from pigs. Nrsp-profile 7 (*n*=5) and nrsp-profile 5 (*n*=2), the only other novel profiles found in multiple isolates, were found to be specific for strains from fish and human origin, respectively. All remaining nrsp-profiles were observed just once in GBS strains isolated from human samples.

Classification of isolates with missing L19 mass variant

The overall low measurability of rsp L19 had implications for 44 GBS isolates, in which L19 was missing in all replicate spectra considered for rsp-profile classification. For 13 isolates, the absence of L19 led to the assignment of a double rsp-profile ID (Fig. 1c). Precisely, rsp-profile 4 and rsp-profile 31, which aside from L19 share an identical combination of rsp variants, were identified as closest match, with 27 of 28 rsp detected. For the remaining 31 isolates, the absence of L19 was not an issue, since the detected rsp mass variants allowed unambiguous assignment to a distinct rsp-profile.

In silico confirmation of rsp-profiles

From a total of ten GBS isolates from fish, WGS data was available and was used for *in silico* prediction of rsp molecular weights and assignment of MLST identity through the *S. agalactiae* PubMLST website [27]. For all ten isolates, the rsp-profile assigned by MALDI-TOF MS was confirmed by the WGS data (Table 1). Eight isolates were assigned to rsp-profile 5 by both MALDI-TOF MS analysis and *in silico* typing. Of these eight isolates, one strain was identified as MLST single-locus variant of ST7 and the remaining seven isolates as ST7. One isolate was assigned to rsp-profile 4 by MALDI-TOF MS analysis. Due to insufficient sequence quality, only 26/28 rsp masses and no ST could be predicted *in silico* for this isolate. The 26 predicted rsp masses all corresponded to the mass alleles of rsp-profile 4, thus supporting that this is the true rsp-profile of this GBS strain. One remaining GBS isolate was not assigned to any rsp-profile by MALDI-TOF MS, due to rsp L18 which displayed a mass variant not yet contained in our reference database in all quadruplicate spectra, indicating the occurrence of a novel rsp-profile (nrsp-profile 7). This was confirmed by *in silico* analysis, which confirmed a previously unknown molecular rsp mass variant of L18 at 12,867 Da, corresponding to the peak seen the mass spectra. Interestingly, the ST of this particular strain was identified as ST931, a single-locus variant of the bovine ST591.

Correlation between MALDI-TOF MS assigned rsp-profiles and capsular serotype

The 168 GBS isolates, which were serotyped and assigned by MALDI-TOF MS to one of the dominant rsp-profiles 2-6 in this study, were used to assess the correlation between strain rsp-profile and capsular serotype (Fig. 2a). We found that strains assigned to rsp-profile 2 (n=21) and rsp-profile 6 (n=40) were mostly displaying capsular serotype III (90% and 70%, respectively). Strains assigned to rsp-profile 4 (n=26) were found to be mostly linked to serotypes Ia (65%) or III (12%). Strains assigned to rsp-profile 5 (n=62) were either linked to serotypes Ia (34%) and Ib (26%) or were non-typeable (31%). Strains assigned to rsp-profile 3 (n=8) were most heterogeneous in terms of associated capsular serotype. One strain each displayed serotype Ia, Ib, II, III and IV, two strains displayed serotype V and one strain was non-typeable.

The rsp-profile-serotype patterns observed in the GBS isolates from Hong Kong correspond largely to what can be observed in a global collection of 523 GBS strains analyzed by our group (Fig. 2b). In accordance to the here presented results, rsp-profiles 2 and 6 are commonly linked to strains with serotype III and rsp-profile 4 is linked to serotype Ia in the

global collection. Rsp-profile 5, in addition to be strongly linked to serotypes Ia, Ib, shows a stronger link to serotype II strains globally than in our study. Lastly, rsp-profile 3 has been reported to be strongly linked to serotype V and to a lesser degree to serotypes II and IV in the global collection. This pattern could not be convincingly supported here based on the low number (n=8) of isolates in this profile.

Average Nucleotide Identity analysis of whole genome sequenced GBS strains

After demonstrating that the rsp-profiles assigned to the ten fish GBS isolates by MALDI-TOF MS correspond to the *in silico* typed rsp-profiles (Table 1), average nucleotide identity (ANI) analysis was performed in order to confirm that genotypes displaying identical rsp-profiles share the same phylogenetic background. ANI analysis was carried out using WGS data of the ten fish isolates and combined with publicly available WGS data of 43 GBS strains. These strains represent the six globally dominant rsp-profiles 1-6 which we previously found to correspond to the major GBS phylogenetic lineages (Rothen *et al*. manuscript submitted). ANI analysis grouped the total 53 strains based on genome-wide assessment of inter-strain similarity (Fig. 3). The eight fish isolates (A1, A11, A12, A23, A31, A41, A60, A63) displaying rsp-profile 5 were located within the rsp-profile 5 cluster in the UPGMA phylogenetic tree. This phylogenetic cluster has been shown to harbor GBS genotypes from a broad range of hosts including human and fish species (Rothen *et al*. manuscript submitted).
The sole fish isolate (A26) assigned to rsp-profile 4 was grouped together with rsp-profile 4 genotypes of MLST clonal complex (CC) 103 background. These CC103 genotypes are of special interest because they were shown to form a genetically highly distinct lineage (Rothen *et al*. manuscript submitted). Lastly, the ST931 isolate (A49), displaying the novel nrsp-profile 7, was found to cluster closest to genotypes of the obligate bovine rsp-profile 1 lineage.

DISCUSSION

The potential use of MALDI-TOF MS for Group B *Streptococcus* strain-level typing has previously been demonstrated by the successful identification of hypervirulent GBS genotypes ST17 and ST1 based on detection of genotype-specific protein masses [20, 28]. We have recently expanded on these findings and proposed a MALDI-TOF MS method that classifies GBS genotypes based on molecular weight variations of 28 ribosomal subunit proteins (rsp). Unlike previous studies, this approach builds on beforehand *in silico* calculated protein molecular masses, moving away from the traditional 'pattern-recognition' approach towards targeted, biomarker-based MALDI-TOF MS microbiological identification. The thereby gained resolution power allows assignment of GBS strains to distinct rsp-profiles, which ultimately allows classification of strains according to their core-genome phylogenetic backbone and provides a predictive value regarding probable capsular serotype, virulence capacity or host origin (Rothen *et al*. manuscript submitted). Here, we have applied this method to analyze a collection of 249 GBS strains isolated from human and animal sources in Hong Kong. We aimed to (a) confirm the transferability of this method between different laboratories and MALDI-TOF MS platforms and (b) investigate the usefulness of this method for high-throughput screening of novel GBS genotypes that we hypothesized can be often found in the diverse animal host reservoirs.

Our MALDI-TOF MS analyses demonstrate that translation of our previously established bacterial sample processing protocol to a different laboratory and different MALDI-TOF MS platform allows for the generation of high quality mass spectra. All but one of the 28 rsp were reliably measured (27 rsp above 92% and 14 rsp above 95%) in the 996 mass spectra generated in this study. The overall measurability of the rsp is likely even higher, given that there were 29 isolates displaying novel rsp mass variants, which were not yet present in our reference database and therefore not picked up in the analysis. The only current drawback pertains to the low measurability of the rsp L19, which was found in only half of the spectra. We do not think that this is a general technical limitation of the Microflex platform, given that L19 was still abundant in the other half of the spectra and that by experience a simple repetition of MALDI-TOF MS measurement resulted in the detection of missed L19. Visual inspection of isolates A23, A31 and A60 mass spectra with missing L19 revealed that the main mass variant was present but the peak not distinct enough to pass the signal-to-noise threshold. *In silico* prediction of L19 in these isolates confirmed this observation and can likely be extrapolated to the other isolates with missing L19. Of note, L19 is a highly

conserved rsp, mostly present as the main mass variant and for the majority of genotypes, missing of L19 does not interfere with the correct assignment of the rsp-profile. The 13 isolates that were assigned with a double-ID due to the missing L19, all matched to both rsp-profile 4 and the rsp-profile 31, which is not a frequently seen profile (1 out of 796 WGS). Furthermore, it has been attributed to a GBS genotype that is located in the same phylogenetic cluster as the majority of rsp-profile 4 genotypes. Hence, the double ID would in this case still allow the assignment of a genotype to its correct evolutionary relationship (Rothen *et al*. manuscript submitted).

The classification of the 170 GBS isolates with known rsp-profile confirmed that the same major GBS phylogenetic lineages which are dominant on global scale are also present highest in our collection (Fig. 1a). Of the six dominant lineages described previously (Rothen *et al*. manuscript submitted), only isolates with rsp-profile 1 were not contained in our collection. This was to be expected since GBS genotypes of rsp-profile 1 were found to be strictly associated with bovine origin, a host that we have not yet covered in our dataset. As opposed to PCR serotyping which failed to type 40% of fish isolates, MALDI-TOF MS based typing was very sensitive, with 98% (170/174) of isolates being assigned to a distinct rsp-profile. Comparison of 168 GBS strains from Hong Kong and 523 strains from the NCBI genome database, revealed high concordance between the two collections with regards to association of rsp-profile and capsular serotype (Fig 2). This further supports the value of the MALDI-TOF MS assigned rsp-profile in providing a predictive measure of likely associated capsular serotype.

A central outcome relates to the 29 genotypes (12% of total collection) that were found to display a novel rsp-profile not previously contained in our reference database. Our data supports that rsp-based MALDI-TOF MS can be used to reliably screen for such genotypes, which we found to predominantly arise from the under-researched animal hosts. While two of the pig GBS isolates fell into known rsp-profiles (rsp-profile 4 and/or rsp-profile 31), all remaining 13 isolates (87% of all pig isolates) displayed an identical and novel rsp-profile (Fig. 1b), raising the possibility that these genotypes represent a distinct, pig-associated phylogenetic lineage.

Of 62 fish GBS isolates investigated here, 87% (n=54), including eight of the ten whole-genome sequenced isolates (Table 1) fell into the already known rsp-profile 5. However, there were five fish isolates (8%) that were, upon visual inspection of the mass spectra, found to

share a novel rsp-profile (nrsp-profile 7). WGS based *in silico* confirmation of the rsp-profile in one of these isolates (A49) confirmed the new fish specific lineage nrsp-profile 7 (Table 1). Integration of nrsp-profile 7 into our bioinformatics pipeline followed by repeated measurement of the other four fish isolates confirmed the presence of nrsp-profile 7 in these strains. This example stands representative of how the rsp-based MALDI-TOF MS method can be used for rapid screening of hundreds of isolates, flagging of potential novel genotypes which are then subjected to WGS. The WGS are subsequently used for *in silico* confirmation of the measured 28 rsp followed by the final integration of newly discovered rsp-profiles to the reference database. This approach of collecting animal samples from markets and integration of novel rsp-profiles to the reference database, allows to rapidly screen for possible occurrence of such GBS genotypes in human patients in the future. The ANI phylogenomic analyses conducted here exemplified that the expression of an atypical rsp-profile can be a reliable indicator that its GBS genotype belongs to a very distinct phylogenetic lineage. Two fish isolates (A26, A49) not belonging to the dominant rsp-profile 5 were found to display a profoundly different phylogenetic background (Fig. 3).

In conclusion, we confirm here the inter-laboratory transferability of our rsp-biomarker based MALDI-TOF MS typing method, its capability to discriminate between GBS genotypes of the major global phylogenetic lineages and the power for rapid screening of hundreds of GBS isolates for surveillance of GBS populations circulating and potential identification of novel emerging genotypes.

ONGOING ANALYSES & OUTLOOK

As part of this ongoing collaboration, we will next evaluate whether single repetition of MALDI-TOF MS measurements will be sufficient to resolve the problem of low rsp L19 measurability. If necessary, technical adjustments of the Microflex machine or the bioinformatics pipeline will be considered. Further, collection of more GBS strains isolated from fish, pig and cow samples from Hong Kong wet markets is currently conducted. Together with a large collection of more than 1,000 human GBS isolates, which have been collected at the Prince of Wales Hospital during recent years, these samples will be analysed by MALDI-TOF MS, thereby significantly adding to our collection of currently 249 isolates. In addition, whole-genome sequencing of all here reported GBS strains isolated from pig and fish samples is currently under way. The WGS data will allow us to *in silico* confirm the rsp-profiles assigned by MALDI-TOF MS. Importantly, the sequencing data will also allow confirmation and integration of the novel rsp-profiles into our reference database. Lastly, we will extend the genome-wide phylogenetic analyses by incorporating all newly sequenced strains into our collection of 796 GBS strains described in our previous work (Rothen *et al*. manuscript submitted). By performing additional phylogenomic analyses, including core-genome analysis using EDGAR [29], we will gain a better picture on the evolutionary background of the novel GBS genotypes retrieved from pig and fish sources in this study.

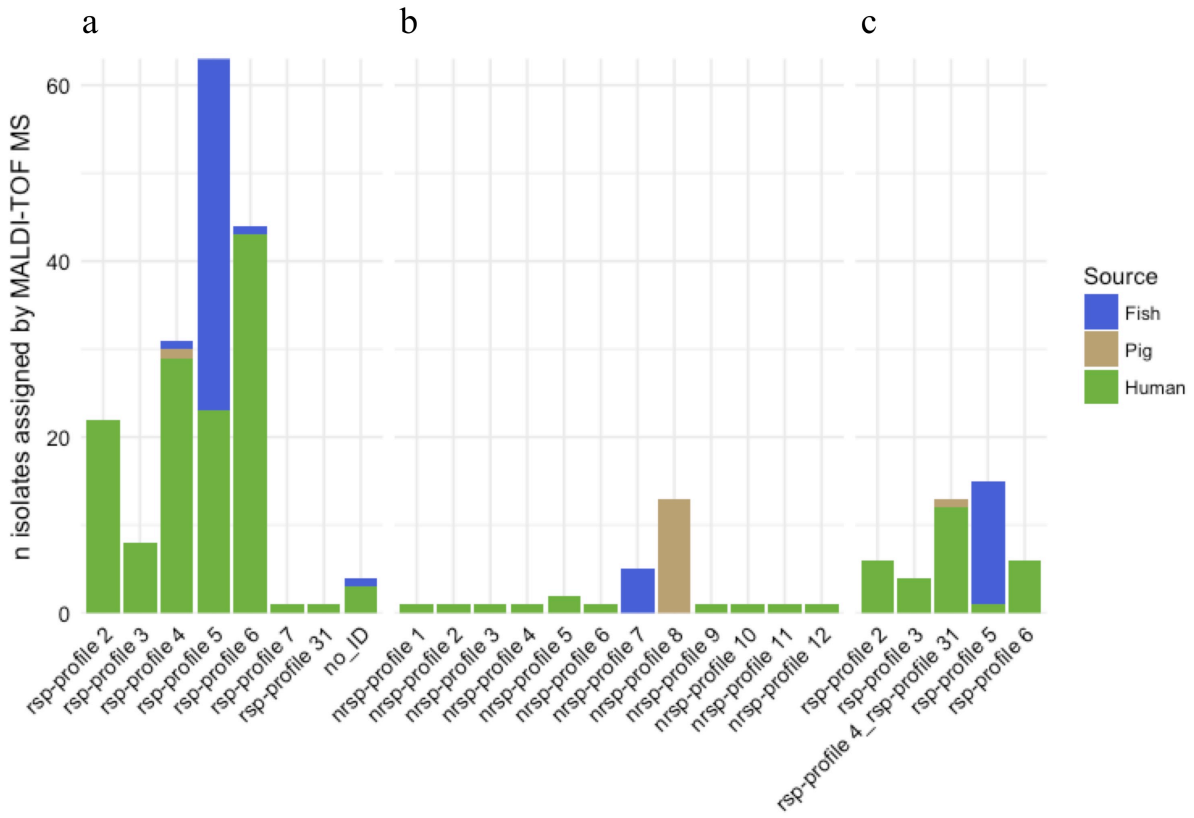


Figure 1: Classification of Group B *Streptococcus* (GBS) isolates based on ribosomal subunit protein (rsp) based MALDI-TOF MS analysis. (a) Assignment of 174 GBS isolates to rsp-profiles already contained in the reference database. (b) Assignment of 29 GBS isolates to novel rsp (nrsp)-profiles, which were not yet incorporated into our reference database. (c) Assignment to known rsp-profiles of 44 GBS isolates for which rsp L19 was not detected in the mass spectrum. Colour coding indicates isolation source of GBS strains (green: human; blue: fish; brown: pig).

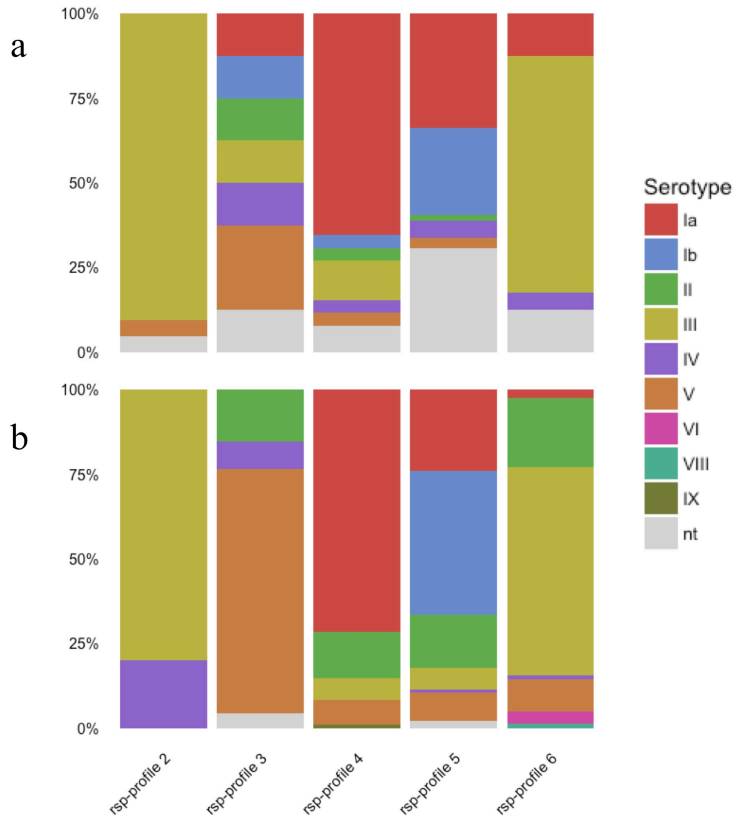


Figure 2: Barplots visualizing associated capsular serotypes of Group B *Streptococcus* (GBS) strains with ribosomal subunit protein (rsp)-profiles 2 to 6. (a) Collection of 168 GBS strains assigned to rsp-profiles by MALDI-TOF MS in this study. (b) Global collection of 523 GBS strains from the NCBI genome database (Rothen *et al*. manuscript submitted). nt: non-typeable strains.

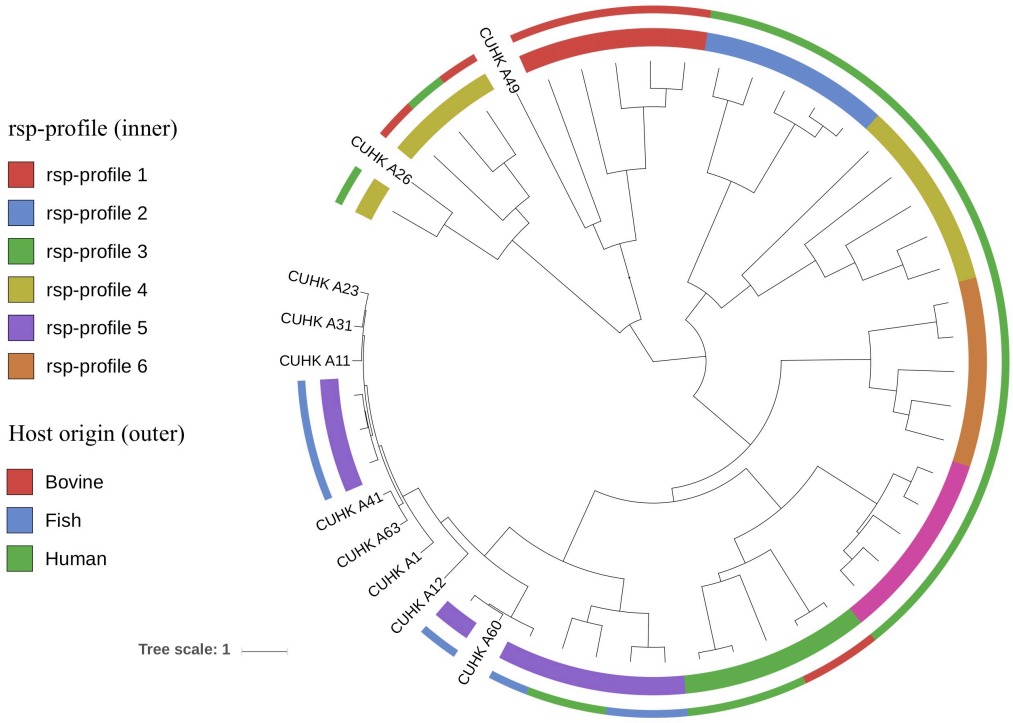


Figure 3. UPGMA phylogenetic tree of 53 Group B *Streptococcus* (GBS) strains based on genome-wide average nucleotide identity (ANI) analysis. The set includes ten strains that were whole genome sequenced in this study and 43 reference strains that represent major GBS phylogenetic lineages. Inner circle: rsp-profile identity of reference strains. Outer circle: Strain host origin. (Scale bar: nucleotide substitutions per site).

Table 1. *In silico* confirmation of MALDI-TOF mass spectrometry (MS) assigned ribosomal subunit protein (rsp)-profile of ten Group B *Streptococcus* strains. SLV: single-locus variant, nt: non-typeable, nrsp-profile: novel ribosomal subunit protein profile.

Isolate ID	*Isolation Source*	*Sequence type*	*MS assigned rsp-profile*	in silico *assigned rsp-profile*
A1	Fish	SLV 7	rsp-profile 5	rsp-profile 5
A11	Fish	7	rsp-profile 5	rsp-profile 5
A12	Fish	7	rsp-profile 5	rsp-profile 5
A23	Fish	7	rsp-profile 5	rsp-profile 5
A31	Fish	7	rsp-profile 5	rsp-profile 5
A41	Fish	7	rsp-profile 5	rsp-profile 5
A60	Fish	7	rsp-profile 5	rsp-profile 5
A63	Fish	7	rsp-profile 5	rsp-profile 5
A26	Fish	nt	rsp-profile 4	rsp-profile 4*
A49	Fish	931	nrsp-profile 7	nrsp-profile 7

*26/28 rsp predicted corresponded to rsp-profile 4 and 2/28 rsp could not be predicted due to sequence quality

REFERENCES

1. Farley MM. Group B streptococcal disease in nonpregnant adults. Clin Infect Dis. 2001;33:556–61.

2. Oliveira ICM, de Mattos MC, Pinto TA, Ferreira-Carvalho BT, Benchetrit LC, Whiting AA, *et al*. Genetic relatedness between group B streptococci originating from bovine mastitis and a human group B *Streptococcus* type V cluster displaying an identical pulsed-field gel electrophoresis pattern. Clin Microbiol. 2006;12:887–93.

3. Wibawan IW, Lämmler C, Smola J. Properties and type antigen patterns of group B streptococcal isolates from pigs and nutrias. J Clin Microbiol. 1993;31:762–4.

4. Rothen J, Schindler T, Pothier JF, Younan M, Certa U, Daubenberger C, *et al*. Draft Genome Sequences of Seven *Streptococcus agalactiae* Strains Isolated from *Camelus dromedarius* at the Horn of Africa. Genome Announc. 2017;5:e00525-17.

5. Evans JJ, Klesius PH, Pasnik DJ, Bohnsack JF. Human *Streptococcus agalactiae* isolate in Nile tilapia (*Oreochromis niloticus*). Emerg Infect Dis. 2009;15:774–6.

6. Delannoy CMJ, Crumlish M, Fontaine MC, Pollock J, Foster G, Dagleish MP, *et al*. Human *Streptococcus agalactiae* strains in aquatic mammals and fish. BMC Microbiol. 2013;13:41.

7. Madrid L, Seale AC, Kohli-Lynch M, Edmond KM, Lawn JE, Heath PT, *et al*. Infant Group B Streptococcal Disease Incidence and Serotypes Worldwide: Systematic Review and Meta-analyses. Clin Infect. 2017;65 Suppl 2:S160–72.

8. Skoff TH, Farley MM, Petit S, Craig AS, Schaffner W, Gershman K, *et al*. Increasing Burden of Invasive Group B Streptococcal Disease in Nonpregnant Adults, 1990–2007. Clin Infect Dis. 2009;49:85–92.

9. Shabayek S, Spellerberg B. Group B Streptococcal Colonization, Molecular Characteristics, and Epidemiology. Front Microbiol. 2018;9.

10. Tan S, Lin Y, Foo K, Koh HF, Tow C, Zhang Y, *et al*. Group B *Streptococcus* Serotype III Sequence Type 283 Bacteremia Associated with Consumption of Raw Fish, Singapore. Emerg Infect Dis. 2016;22:1970–3.

11. Kim HJ, Kim SY, Seo WH, Choi BM, Yoo Y, Lee KH, *et al*. Outbreak of Late-onset Group B Streptococcal Infections in Healthy Newborn Infants after Discharge from a Maternity Hospital: A Case Report. J Korean Med Sci. 2006;21:347–50.

12. Ip M, Cheuk ESC, Tsui MHY, Kong F, Leung TN, Gilbert GL. Identification of a *Streptococcus agalactiae* Serotype III Subtype 4 Clone in Association with Adult Invasive Disease in Hong Kong. J Clin Microbiol. 2006;44:4252–4.

13. Ip M, Ang I, Fung K, Liyanapathirana V, Luo MJ, Lai R. Hypervirulent Clone of Group B *Streptococcus* Serotype III Sequence Type 283, Hong Kong, 1993–2012. Emerg Infect Dis. 2016;22:1800–3.

14. Kalimuddin S, Chen SL, Lim CTK, Koh TH, Tan TY, Kam M, *et al*. 2015 Epidemic of Severe *Streptococcus agalactiae* Sequence Type 283 Infections in Singapore Associated With the Consumption of Raw Freshwater Fish: A Detailed Analysis of Clinical, Epidemiological, and Bacterial Sequencing Data. Clin Infect Dis. 2017;64 suppl_2:S145–52.

15. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, *et al*. Ongoing Revolution in Bacteriology: Routine Identification of Bacteria by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. Clin Infect Dis. 2009;49:543–51.

16. Singhal N, Kumar M, Kanaujia PK, Virdi JS. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front Microbiol. 2015;6:791.

17. Binghuai L, Yanli S, Shuchen Z, Fengxia Z, Dong L, Yanchao C. Use of MALDI-TOF mass spectrometry for rapid identification of group B *Streptococcus* on chromID Strepto B agar. Int J Infect Dis IJID. 2014;27:44–8.

18. Cherkaoui A, Emonet S, Fernandez J, Schorderet D, Schrenzel J. Evaluation of matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of Beta-hemolytic streptococci. J Clin Microbiol. 2011;49:3004–5.

19. Lartigue M-F, Héry-Arnaud G, Haguenoer E, Domelier A-S, Schmit P-O, van der Mee-Marquet N, *et al*. Identification of *Streptococcus agalactiae* Isolates from Various Phylogenetic Lineages by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. J Clin Microbiol. 2009;47:2284–7.

20. Lartigue M-F, Kostrzewa M, Salloum M, Haguenoer E, Héry-Arnaud G, Domelier A-S, *et al*. Rapid detection of "highly virulent" Group B *Streptococcus* ST-17 and emerging ST-1 clones by MALDI-TOF mass spectrometry. J Microbiol Methods. 2011;86:262–5.

21. Yang J-L, Wang M-S, Cheng A-C, Pan K-C, Li C-F, Deng S-X. A simple and rapid method for extracting bacterial DNA from intestinal microflora for ERIC-PCR detection. World J Gastroenterol. 2008;14:2872–6.

22. Imperi M, Pataracchia M, Alfarone G, Baldassarri L, Orefici G, Creti R. A multiplex PCR assay for the direct identification of the capsular type (Ia to IX) of *Streptococcus agalactiae*. J Microbiol Methods. 2010;80:212–4.

23. Gibb S, Strimmer K. MALDIquant: a versatile R package for the analysis of mass spectrometry data. Bioinforma Oxf Engl. 2012;28:2270–1.

24. Arnold RJ, Reilly JP. Observation of *Escherichia coli* Ribosomal Proteins and Their Posttranslational Modifications by Mass Spectrometry. Anal Biochem. 1999;269:105–12.

25. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A. 2005;102:2567–72.

26. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242-245.

27. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595.

28. Lin H-C, Lu J-J, Lin L-C, Ho C-M, Hwang K-P, Liu Y-C, *et al*. Identification of a proteomic biomarker associated with invasive ST1, serotype VI Group B *Streptococcus* by MALDI-TOF MS. J Microbiol Immunol Infect. 2017.

29. Blom J, Kreis J, Spänig S, Juhre T, Bertelli C, Ernst C, et al. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. Nucleic Acids Res. 2016;44:W22-28.

Chapter 7

General Discussion

The content of this P hD thesis is centered around two pathogens, namely *Plasmodium falciparum,* the main cause of severe malaria and a major driver of maternal and early childhood mortality in sub-Saharan Africa [3], and Group B *Streptococcus*, a global leading cause of neonatal meningitis and sepsis [72]. The pursuit of combating these diseases through vaccine development, in recent years increasingly aided by the introduction of high-throughput omics technologies, represents the overarching theme of the five manuscripts presented here. The key findings for each of the presented chapters are summarized below.

Chapter 2: Clinical evaluation of a radiation-attenuated *P. falciparum* whole sporozoite vaccine revealed impaired immunogenicity and protective efficacy in Tanzanian volunteers compared to malaria-naïve subjects undergoing an identical vaccination regimen.

Chapter 3: RNA-Seq whole blood gene expression analysis provides valuable insights into molecular dynamics following controlled human malaria infection patterns in malaria pre-exposed volunteers. Magnitude and timing of early transcriptional signatures is dependent on parasite pre-patent period.

Chapter 4: A ribosomal subunit protein (rsp) based MALDI-TOF MS typing method classifies GBS genotypes based on core-genome phylogenetic lineages, detects hypervirulent strains and allows prediction of serotype, CC and host origin.

Chapter 5: Whole genome sequences of under-researched GBS genotypes isolated from camels provide the data basis for phylogenomic assessment of these strains and incorporation into our rsp-database.

Chapter 6: The inter-laboratory reproducibility and the utility of the rsp-based MALDI-TOF MS method for screening and monitoring of emerging animal GBS genotypes is confirmed.

7.1 Evaluation of malaria vaccines in different populations using controlled human malaria infection: Chapters 2 and 3

Chapter 2: Safety, Immunogenicity, and Protective Efficacy against Controlled Human Malaria Infection of* Plasmodium falciparum *Sporozoite Vaccine in Tanzanian Adults

Chapter 3: Whole blood transcriptome changes following controlled human malaria infection in malaria pre-exposed volunteers correlate with parasite prepatent period

Since its use as treatment for neurosyphilis in the early 1900s [172], controlled human malaria infection (CHMI) has evolved into a highly useful tool for the evaluation of anti-malarial drugs [42,43], diagnostic tools [173] and assessment of malaria vaccine protective efficacy, as presented in *Chapter 2* of this thesis. CHMI allows the evaluation of novel drug and vaccine candidates in a well-controlled and safe environment through the injection of individuals with malaria parasite stages from either the liver or blood stage (reviewed in [174]). In the work presented here (*Chapters 2 and 3*), liver stage parasites, e.g. fully infectious, live sporozoites were used for CHMI. Traditionally, CHMI using sporozoites has been carried out by exposure of individuals to repeated, infectious mosquito-bites [175,176]. A major milestone was achieved when the biotechnology company Sanaria Inc. (www.sanaria.com) managed to produce aseptic, purified, cryopreserved *P. falciparum* sporozoites (PfSPZ), that are infectious *in vivo* and can be shipped globally to centers conducting CHMI [40,41]. These sporozoites are injected via needle and syringe either intradermally (as described in *Chapter 3*) or through direct venous inoculation (DVI) (as described in *Chapter 2*).

Both, the mosquito-bite delivery and direct inoculation approaches have their specific benefits and disadvantages when used for the assessment of vaccine efficacy. The mosquito-bite approach has the advantage that it mimics closest the natural way of malaria infection. The number of sporozoites inoculated via controlled mosquito-bite is comparable to parasites injected during natural exposure in the field. Further, the mosquito-bite inoculated sporozoites are going through the skin stage of infection, while this is not the case during direct venous inoculation of the parasites [174]. The limitations of the mosquito-bite challenge model include a high demand in onsite infrastructure (insectary and laboratory structure), the need for entomological expertise and great variability in the number of injected sporozoites between individuals [175]. In contrast, intradermal or direct venous injection of PfSPZ

enables the standardization of the number of parasites injected per individual, allowing the assessment of different and defined dosages in vaccine trials. Direct inoculation of individuals with PfSPZ for vaccine efficacy assessment circumvents the need for high natural malaria exposure which is essential in clinical studies assessing vaccine efficacy through field exposure [46]. At the same time, the high doses (3.2 x 10^3 PfSPZ in Tanzania trial presented in *Chapter 2*) used for DVI CHMI might be too rigorous and not sufficiently reflect to what extent the vaccine induced immunity would contain the malaria disease under natural conditions (Stephen Hoffman, personal communication).

The clinical trial results presented in *Chapter 2* are of great interest and relevance to the malaria vaccinology field, given that this was the first efficacy assessment of any malaria vaccine by CHMI in Africa. While confirming the safety and tolerability of the PfSPZ vaccine, our results also demonstrate significant differences between malaria-experienced Tanzanian volunteers and malaria-naïve U.S. volunteers both in terms of induced vaccine immunogenicity and protective efficacy. Among U.S. volunteers undergoing the identical immunization regimen, 12 of 13 subjects (92.3%) were protected from homologous mosquito-bite CHMI [177]. In our high-dose group of individuals receiving five doses of 2.7 × 10^5 PfSPZ, the protective vaccine efficacy (VE) was 20% (4/20) as assessed by DVI CHMI. In a third study, Malian volunteers received five doses of 2.7 × 10^5 PfSPZ and protection against natural *P. falciparum* infection was 52% by time to event or 29% by proportional analysis [46]. Although the three mentioned studies are not directly comparable due to the different types of CHMI used to assess protective efficacy, the VE reported in the Tanzanian cohort is in closer range to what has been reported in malaria-experienced individuals from Mali. While the protective efficacy of the PfSPZ vaccine was 20 % in the high dose group, only one volunteer was protected from homologous CHMI in the Tanzanian low dose group. This is an indication for a dose-effect and suggests that an even higher PfSPZ dose could potentially induce higher immunogenicity and protective efficacy in malaria-experienced volunteers.

The fact that vaccine induced immune responses and conferred protective efficacy can be subjected to significant geographical variation, especially between industrialized nations and low and middle-income countries is well-known [178]. Vaccine development including formulation, vaccination schedule and evaluation for efficacy are assessed in cohorts that are not necessarily comparable to cohorts where the vaccine will be introduced upon licensure (reviewed in [178]). One of the earliest and most extensively studied examples for regional variation in vaccine efficacy is bacille Calmette-Guérin (BCG), for which protective efficacy

was found to vary, amongst others, between subjects from the UK and Mali [179]. Other examples of regional differences in vaccine immunogenicity include the hepatitis B and *Haemophilus influenzae* type b vaccines [180] as well as several oral vaccines against cholera, polio or *Shigella* [181]. What exactly determines the varying degrees of individual responses to an identical vaccine remains subject of debate and research. While environmental factors certainly have a main influence, as for instance shown in the case of BCG [179,182], it is likely not the only determinant, given the documented cases of differing responses among individuals from the same environment [178]. The higher levels of memory T cells at baseline in the Tanzanian volunteers of our study, as compared to levels in U.S. volunteers, indicate that pre-existing immunity against malaria could be one explanation for the reduced vaccine efficacy. Co-infection with parasites, including soil-transmitted helminths, were found to have not contributed to the reduced immune responses in our study, but have been shown as potential modulator of the human immune response to malaria [183]. Similarly, the potential profound impact of the gut microbiome on vaccine effectiveness is increasingly acknowledged [184].

Collectively, the results of this first PfSPZ clinical trial in Tanzania underline the need to understand the molecular mechanisms involved in anti-malarial protective immunity to explain why and how malaria-experienced individuals appear to respond weaker to the same vaccination regimen as malaria-naïve individuals of same age. Besides of this, a standardization of the CHMI methodology and trial protocols conduct between centers is desirable in order to facilitate inter-site comparison of PfSPZ study outcomes. This will include the use of the same type of CHMI for vaccine efficacy assessment and distinction between homologous (vaccine-strain) and heterologous (non-vaccine strain) CHMI [185]. Regarding the post-CHMI follow-up period and the treatment of individuals that display blood-stage parasitemia, a decision on whether using thick blood smear or the more sensitive qPCR as decision point to initiate treatment will need to be made [186]. An ideal scenario for the conduct of harmonized PfPSZ clinical vaccine trials would be the establishment of specific, regional CHMI systems, that are tailored to account for population characteristics including genetic predispositions like hemoglobinopathies or known pre-existing immunity to malaria [174].

With the work presented in *Chapter 3*, we made a first step towards better understanding the molecular mechanisms occurring among Tanzanian volunteers following intradermal CHMI without previous vaccination. We used RNA-Seq in order to investigate whole blood

transcriptional changes at day 5, 9 and 28 following PfSPZ infection in comparison to baseline. A first and important finding was the fact that most genes with differential expression levels compared to pre-CHMI baseline were found at day 5 post CHMI, during the clinically silent liver phase. This stood in contrast to the modest transcriptional changes recorded at day 9 at the time point when it is commonly assumed that parasites transition from the liver to the blood. This finding supports that the timing of whole blood collection as well as the inclusion of early time points (before day 9) during the clinically silent liver stage needs to be targeted to capture changes in transcriptional signals in whole blood.

In order to better interpret the functional relationships of the hundreds of genes that were found differentially expressed during the liver stage, we made use of the established blood transcriptome modules (BTM). The BTM framework was established by Li *et al.*, who used more than 30,000 human blood transcriptomes from over 500 studies to define transcriptional networks of genes based on observed co-expression patterns. This allowed the final definition of 334 BTM, each standing representative of specific biological functions [187]. Such BTMs can facilitate the interpretation of whole blood transcriptomics data, as was the case in our study, where we used the BTMs as basis for gene set enrichment analyses (GSEA).

The BTMs that were found differentially expressed in our study provide first time insight into the broad responses in malaria-experienced subjects following intradermal CHMI. The up-regulation of genes in the proteasome module observed strongest at day 9 and significant at day 28 post CHMI are interesting given that the proteasome is known to play a key function in MHC protein processing and antigen presentation [188]. The genes in this module could therefore be of special interest regarding the development of adaptive immune responses against *P. falciparum*. The antigen processing and presentation pathway and phosphotidylinositol signaling system which we found to be strongest down-regulated at day 5 in subjects with low time to blood parasitemia indicates a possible importance of these networks in the early innate immune response acting on the parasite during the pre-erythrocytic stage of infection. The up-regulation of genes in the MAPK RAS signaling and ubiquitination modules is an interesting parallel to a finding of Ockenhouse *et al.* [189], who reported activation of MAP kinases by natural acquired *P. falciparum* infection. The same study reported over expression of genes linked to the genet ontology term “protein ubiquitination” following mosquito bite challenge of malaria-naïve subjects [189].

Among the BTMs characterized in our work, several have been reported in similar studies that investigated transcriptional responses following controlled infection with *P. falciparum* or vaccination [66]. We can for now only draw limited conclusions when comparing our

results with these studies, given the differences in study participants (malaria-naïve or vaccinated vs. pre-exposed subjects), challenge model (mosquito bite vs. intradermal injection) and time point of gene expression assessment. An interesting question will be to address, whether a generic immune response to malaria infection or vaccination even exists, or if the molecular networks affected vary depending on geographical location, *Plasmodium* genotype or individual genetic predispositions.

Important insights regarding this question were made by Li *et al.* who in the pursuit of determining a generic molecular signature for human vaccine response, performed BTM aided transcriptional analyses of recipients of five different human vaccines. This study confirmed that early BTM signals are predictive of later antibody response and can therefore be used to decrypt immune responses to vaccination. Although vaccines of the same general type (viral, polysaccharide or protein based) showed similar activation and deactivation of BTMs, the responses elicited by each vaccine were found to be type-specific. For example in the case of meningococcal conjugate-polysaccharide vaccines, it was found that distinct BTMs are induced by different components of the same vaccine [187]. The latter has possible implications for expected immunological responses against the subunit vaccine RTS,S and the whole parasite PfSPZ vaccine.

A further central finding of our study pertained to the individual magnitude and timing of early gene expression changes, which we found to be associated with the duration of parasite prepatent period. We speculate that this reflects the varying degrees of previous exposure and pre-existing immunity among the ten subjects. Capturing time points with the highest transcriptional expression changes might depend on the size of the parasite load multiplying in the liver. This observation strongly suggest that blood collection timing is critical and should be conducted at more frequent intervals, additionally covering early time points between days 1 to 4 post CHMI.

Building up on the data presented in *Chapter 3*, continuous RNA-Seq analyses are currently being conducted. The sample data set will include volunteers from the PfSPZ Tanzania trial (presented in *Chapter 2*) who underwent both vaccination and CHMI, as well as malaria-naïve volunteers from a mosquito-bite challenge trial (NCT01994525). In addition to expanding the RNA-Seq analyses to other PfSPZ cohorts, ongoing technological advances are contributing to a better understanding of the complex transcriptomic data generated from such studies. Presently, a fundamental challenge when performing bulk RNA-Seq is the heterogeneity of the source material, e.g. the whole blood. Such a global transcriptome

analysis provides insight into the average gene expression changes but does not reveal which specific cell subsets in the peripheral blood drive the observed gene expression dynamics. A variety of computational methods have therefore been developed for inferring cell subset-specific information via *in silico* deconvolution from heterogeneous gene expression data (reviewed in [190]). As a promising alternative, single-cell RNA-Seq (scRNA-Seq) is emerging as a powerful method allowing the assessment of gene expression changes on cell-level. Due to decreasing sequencing costs and constant evolvement of technology, scRNA-Seq will have a huge impact on our capacity to study innate and adaptive immune responses on cellular level and aid in the discovery of novel immune cell subtypes [191].

As the burden of severe, clinical malaria in sub-Saharan Africa is steadily decreasing, the visibility of malaria co-occurring infectious diseases increases. Important insights regarding malaria co-morbidities stem from research focusing on invasive non-Typhoid *Salmonella* (NTS), which is commonly associated with *P. falciparum* severe disease [192]. It was found that the induction of heme oxygenase 1 (HO-1), a protective mechanism in the human host to cope with the cytotoxic effect of heme, which is released during malarial hemolysis, also leads to decreased neutrophil function and a decreased production of bactericidal reactive oxygen species (ROS) [193]. Hence, a molecular mechanism conveying tolerance against one pathogen (*P. falciparum*) impairs resistance to another (NTS). Importantly, this detrimental effect is not restricted to severe malaria cases only. In Burkina Faso, children with subclinical malaria were found to display sustained hemolysis and induction of HO-1. This observation suggests that asymptomatic malaria, although not leading to apparent clinical symptoms nevertheless does harm to the individual, e.g. through increased susceptibility to invasive bacterial disease [194]. Besides of NTS a range of other bacterial species are likely linked to clinical and asymptomatic malaria co-morbidity burden. As a common disease-causing pathogen in sub-Saharan Africa [195], Group B *Streptococcus* should be considered as a major contributor to malaria facilitated bacterial invasive disease.

7.2 MALDI-TOF MS as a phyloproteomic tool for post-vaccination monitoring of GBS genotype landscape and screening for emerging strains: Chapters 4, 5 and 6

Chapter 4: Subspecies typing of* Streptococcus agalactiae *based on ribosomal subunit protein mass variation by MALDI-TOF MS

Chapter 5: Draft Genome Sequences of Seven* Streptococcus agalactiae *Strains Isolated from* Camelus dromedarius *at the Horn of Africa

Chapter 6: Tracing and monitoring of emerging Group B* Streptococcus *genotypes with zoonotic potential in Hong Kong

In the second part of this thesis, we have demonstrated how a combinatorial approach, that exploits the increasingly available wealth of bacterial genomic data and the propensity of MALDI-TOF MS for high-throughput protein analysis, can be used to design a phyloproteomic typing method for Group *B Streptococcus*. Our results underline the potential of this tool as a method for large-scale monitoring of vaccine impact on the GBS genotype composition and emergence of escape strains from various host reservoirs.

In the work presented in *Chapter 4* we show the path from conceptualizing the idea of a ribosomal subunit protein (rsp)-based MALDI-TOF MS GBS typing method to the design and validation of such a scheme. Analysis of whole-cells by MALDI-TOF MS can rapidly generate a highly characteristic snapshot of the proteomic makeup of the bacterial strain studied. However, distinction of closely related microbes, for instance strains or serotypes of the same species, that share highly similar protein profiles is challenging [196]. To the inherent complexity of MALDI-TOF mass spectra, which can contain hundreds of protein mass data points, an additional layer of complexity is added by differences in spectral makeup depending on confounding experimental factors like bacterial growth time and processing prior to analysis, the matrix used or MALDI-TOF MS instrument parameters. Collectively, such factors limit the inter-laboratory comparison of MALDI-TOF mass spectra [50,197]. In *Chapter 4,* we describe how we moved away from the conventional approach of pattern-recognition based classification, e.g. matching of full spectra against a reference database, towards a targeted, marker-based identification method. Besides of developing a standardized

bacterial sample processing protocol for improved spectral quality, we were able to overcome the aforementioned challenges by exploiting the wealth of publicly available GBS whole genome data. For our study, sequence data of 796 GBS strains was used to *in silico* assess the suitability of rsp as biomarker masses for targeted MALDI-TOF MS typing of GBS strains. This novel method targets 28 known ribosomal loci, allowing us to simultaneously detect molecular mass variation across a concatenated amino acid sequence of ~ 2,700 aa. We can thereby exploit subtle differences in an evolutionary highly conserved part of the GBS genome for a phyloproteomic classification of closely related GBS genotypes.

Our results revealed that the majority of GBS strains circulating in the human host can be assigned by MALDI-TOF MS analysis to one of the five rsp-profiles 2, 3, 4, 5 or 6. Our proteomic classification thereby largely corresponds to the classification by MLST, which groups GBS genotypes based on nucleotide variation at seven housekeeping gene loci into the five well-described, globally dominant clonal complexes CC17, CC1, CC23, CC10 and CC19 [116]. Core-genome phylogenetic analysis of the 796 GBS whole genome sequences that were used to establish the rsp-profile scheme, confirmed that the assigned rsp-profiles classify GBS genotypes in high concordance to their core-genome phylogenetic relationship. Although our MALDI-TOF MS assigned rsp-profile identity fails in some cases to distinguish between genotypes of different CCs, it does provide a highly reliable indication on the stable evolutionary backbone of the isolate. This was exemplified by genotypes belonging to CC17 and genotype ST452, which both display rsp-profile 2 and therefore cannot be distinguished by our typing method. GBS clone ST452 has been described to have emerged from massive genetic recombination events between CC17 and CC23 lineages [198]. Hence, the assigned rsp-profile 2 does provide an indication on the ancient origin/backbone of this newly emerging strain, raising the possibility that our rsp-based grouping might be more reflective of the "true" phylogeny of this particular clone.

Importantly, there seems to be a strong correlation between most of the MALDI-TOF MS distinguishable rsp-profiles and the capsular serotype of such genotypes. For example, rsp-profile 1 is indicative of serotype II, rsp-profile 2 is strongly linked to serotype III and rsp-profile 3 is predominantly found in genotypes carrying serotype V. Currently, a trivalent CPS vaccine is most advanced in clinical development, incorporating the globally dominant serotypes Ia, Ib and III [135–137]. These serotypes, although the most pathogenic, represent only a fraction of the GBS global population. The potential implications of a vaccine targeting only selected genotypes can be exemplified by the lessons learned from the multi-valent

pneumococcal vaccine. After its introduction in Europe and North America, vaccine-type serotypes and associated invasive pneumococcal disease decreased rapidly. However, non-vaccine-type serotypes and linked disease increased in the years following vaccine introduction, indicating the importance of continued population surveillance in order to track serotype replacement [140,141]. Long-term follow-up studies will be essential to understand how the global GBS population responds to vaccine induced immune selection. Particularly, it would be interesting to understand if some of the ancient GBS lineages that carry serotypes included into the vaccine vanish, if escape strains carrying non-vaccine serotypes emerge as newly dominating lineages and if GBS with zoonotic potential fill the vaccination induced biological niches.

A central aspect of our phyloproteomic typing tool established in *Chapter 4* is its capacity for inter-laboratory transferability to a different MALDI-TOF MS platform, which is fundamental for the future application of this method in multi-center epidemiological studies. A universal typing method that can be applied in different laboratories is a prerequisite for large-scale epidemiological studies. As demonstrated by our results in *Chapter 6*, translation of our method to a microflex MALDI-TOF MS instrument, one of the most widely used commercial platforms in routine microbiology [144], was unproblematic. We found that on average 27 of the 28 ribosomal subunit biomarker masses required for our GBS typing approach can be detected per mass spectrum, which is more than sufficient for classification of isolates. We believe a central aspect that will facilitate wider implementation of our novel method is the fact that MALDI-TOF MS instruments are already standard laboratory equipment at many clinical diagnostic sites undertaking microbiological routine diagnostics or research projects. We could therefore build on existing infrastructure and implement our novel method and sample processing workflow without imposing significant additional costs.

Animal hosts act as a potential reservoir for emerging GBS strains that can cause invasive disease in human. Especially in the context of raw milk [83] or raw fish [81,82] consumption, the zoonotic potential of *S. agalactiae* is increasingly acknowledged. Based on the results reported in *Chapter 4*, it became soon evident that there are some rsp-profiles that are indicative of the GBS host origin. Specifically, GBS genotypes with rsp-profile 1 were exclusively found to be of bovine origin, while rsp-profiles 49 and 13, which were much less abundant in our collection were indicative for fish and fish or frog origin, respectively. An interesting addition to these animal GBS genotypes, many of which share very few similarity

on MLST level with the known human clones, were the 16 WGS from *Camelus dromedarius*, seven of which were released in the genome announcement presented in *Chapter 5*. A manifold of rsp-profiles (11, 15, 18, 20, 21, 37) found in these isolates were not described in other hosts yet. This was also reflected by the core-genome phylogenetic analysis, which places these camel-derived GBS genotypes distinct from all other isolates within the global collection. These findings confirm that a rsp-based MALDI-TOF MS method can reliably distinguish between the human GBS genotypes and the novel camel-associated strains and would therefore serve as an ideal diagnostic tool for monitoring GBS transmission dynamics between these two hosts. Such epidemiological studies would be of high relevance. Camels kept as livestock are reportedly increasing in numbers, especially in regions at the Horn of Africa and the Middle East. The consumption of raw camel milk has gained increased popularity and is known to be a possible source of infection with zoonotic pathogens [125]. It remains unclear to what extend GBS strains are circulating in these micro-environments and whether the respective camels and/or humans serve as reservoirs for genetic recombination which might lead to the emergence of novel, highly pathogenic clones, similar to ST452 [198] or ST485, an emerging invasive CC103 clone in China [199,200].

Our results from *Chapter 6* pertaining to the novel GBS genotypes that we found at higher frequency in the animal hosts further underline that our method is highly adequate to rapidly screen for such genotypes among hundreds of isolates. Among the GBS isolates retrieved from fish samples were five strains that displayed the novel nrsp-profile 7. Subsequent whole genome sequencing of one strain confirmed the distinct rsp-profile and identified this strain as ST931. This sequence type has been described for the first time in a recent study in 2018, where ST931 was found to be linked to human invasive disease in the Guangxi region in Southern China [200]. Interestingly, the authors also demonstrated that this genotype genetically falls within the bovine CC67 cluster, a finding that was confirmed by our ANI phylogenetic analysis. The fact, that we now detected this bovine-associated strain on fish samples in Hong Kong is highly intriguing and raises important questions regarding the host specificity and transmission dynamics of this genotype. Our MALDI-TOF MS tool will serve as an ideal tool for continuous screening of nrsp-profile7/ST931 strains among fish samples in the Hong Kong region.

One fish GBS strain displaying rsp-profile 4, which is atypical for so far analyzed fish-derived isolates that predominantly fall within rsp-profile 5 (54 of 62 isolates), further exemplifies the utility of our method in screening for rapidly emerging known GBS

genotypes. The MLST identity of this strain could not be determined due to limited WGS quality and coverage. However, ANI phylogenetic analysis revealed that this strain belongs to the CC103 genetic lineage, a group of genotypes of central interest in terms of GBS with zoonotic potential. ST103 and its closely related clones were initially described as strains circulating in cows. Both in Europe [201] and China [202], the emerging ST103 strains were found to replace CC67 as the dominant genotype linked to bovine mastitis. The expansion of CC103 is not restricted to the bovine host only, with increasing prevalence of CC103 strains reported in human [200] and fish [124]. Although we cannot distinguish ST103/CC103 from other ST like ST22 and ST23 that also display rsp-profile 4, we can rapidly identify such genotypes among fish isolates, that usually fall within rsp-profile 5, or from classical bovine CC67 strain that display rsp-profile 1. This provides us with a mean to track further the spread of this genotype.

An intriguing finding of potentially high relevance pertains to the characteristics of the GBS isolates collected from pig specimens in Hong Kong wet markets. Our preliminary results demonstrate, that the majority of these strains (13/15) display a novel rsp-profile, hinting towards the genetic distinctiveness of these genotypes. Little is known about the prevalence of GBS colonization and disease in *Sus scrofa domesticus*. Only few studies report the isolation or characterization of GBS strains from pigs [203,204] and similarly no conclusive data on human to pig or vice versa transmission of GBS has been described [205,206]. This is surprising, given the immense economical relevance of the pig industry in many countries. The recent expansion of global pork trade, namely between the USA and China has raised concerns regarding the threat of facilitated spread of pathogens [207]. The neglect of appropriate bio-safety measures combined with dense farming conditions in China, with close contact of humans, poultry and pigs, have been demonstrated to facilitate the recombination of pathogens and the emergence of novel strains. This does not only lead to substantial agricultural losses but can also pose significant public health threats, exemplified by reported human infection through pig-reservoired influenza A viruses in China [208].

A similar scenario of increased genotype variety and emergence of hypervirulent strains through human-pig transmission is also conceivable for GBS. In the context of the post GBS vaccination phase, it can be expected that vanishing GBS genotypes which are targeted by the multivalent conjugate vaccine will open biological niches, which can accelerate the emergence of novel, vaccine escape GBS genotypes. Accordingly, large-scale GBS surveillance studies will be needed to monitor such genotype transmission dynamics. As

demonstrated by our results from *Chapter 6*, our MALDI-TOF MS method would meet all the requirements and provide the required discriminatory power for such a task.

Chapter 8

Outlook

Within the framework of the WHO malaria vaccine implementation programme (MVIP), Ghana, Kenya and Malawi will introduce the RTS,S vaccine in the routine immunization systems of selected geographical areas in 2019. Given the limited vaccine efficacy, RTS,S will be implemented only as a supporting measure in parallel to conventional vector control measures. This will provide valuable information regarding the public health usefulness and feasibility of large-scale deployment of the RTS,S vaccine [36].

Meanwhile, the pursuit for a more effective malaria vaccine will continue. The promising whole sporozoite based malaria vaccines are under active development and results from first clinical trials with radiation-attenuated PfSPZ conducted in countries throughout sub-Saharan Africa including Mali [46], Equatorial Guinea [47] and Tanzania (as presented in *Chapter 2*) are now available. The results from the PfSPZ Tanzania trial nurtures hopes that an increased dose of administered PfSPZ can help increasing the vaccine immunogenicity and confer better protective efficacy in volunteers. Accordingly, continuing clinical trials are being conducted in Tanzanian cohorts. One trial (NCT02613520) assesses how administration of a higher dose (1.8×10^6 PfSPZ) in adults undergoing three immunizations effects immunogenicity and protective efficacy of the vaccine. Hypothesizing that the magnitude of PfSPZ vaccine induced immune responses will be higher in younger, less malaria-exposed individuals, the same trial also assesses vaccination in younger cohorts, including age groups from 11-17 years, 6-10 years, 1-5 years and infants of 6-11 months of age at time of first vaccination (Jongo *et al.*, manuscript submitted). A follow-up study (NCT03420053), evaluating the safety, immunogenicity and efficacy of the PfSPZ vaccine in HIV-positive, Tanzanian volunteers has been completed in September 2018 (manuscript in preparation).

Given the promising results of the radiation-attenuated PfSPZ vaccine, conferring high protection from homologous CHMI in malaria-naïve individuals, the licensure of this vaccine candidate is on the horizon. In a recent report from the U.S. FDA, the possibility of PfSPZ vaccine licensure for use in travelers was discussed [209]. An according pre-licensure phase 3 clinical trial is therefore planned for the assessment of the vaccine in a large malaria-naïve cohort (Stephen Hoffman, personal communication). A central question pertains to the possible implications of an area-wide introduction of the PfSPZ vaccine. Given the multitude of genetically differing *P. falciparum* strains circulating in sub-Saharan Africa [210,211], it is likely that the emergence of vaccine escape strains will be observed following large-scale vaccine introduction. The clinical trial in Mali has shown that PfSPZ vaccination can confer some protection against naturally acquired malaria infection through field strains [46]. Potential cross-protection of the PfSPZ vaccine against a non-vaccine strain has also been

assessed by heterologous CHMI in malaria-naïve individuals. One study reported protection of 4/5 individuals 3 weeks and protection of just 1/10 individuals 24 weeks after last immunization (5 doses of 2.7 x 10^5) [177]. Although a second study reported increased protection of 5/6 individuals at 33 weeks after last immunization (3 doses of 9 x 10^5) [185], it will require further assessment, if long-lasting protection against non-vaccine strains can be achieved with the radiation-attenuated PfSPZ vaccine.

A possible way to circumvent this problem could be the development of multi-strain vaccines, as was done with *Theileria parva*, an apicomplexan parasite sharing a similar life cycle like *P. falciparum*, and known as the cause of East Coast fever. The trivalent, live sporozoite vaccine against *T. parva* consists of a cocktail of three genetically distinct theilerial strains and has been demonstrated to successfully immunize cattle under chemoprophylaxis against East African theileriosis [212]. It will need to be assessed in future studies if a re-formulated PfSPZ vaccine, containing multiple strains of *P. falciparum*, can confer cross-protective immunity against field strains.

Besides of the radiation-attenuated PfSPZ vaccine, the genetically attenuated *P. falciparum* parasites (GAP) are developed. Despite of being a promising candidate, this vaccine will need to overcome the administration route of live parasites via mosquito bite, which is impracticable for mass administration [49]. GMP manufacturing of a cryopreserved *P. falciparum* GAP vaccine is therefore planned in collaboration with the biotech company Sanaria Inc. This will allow clinical trial assessment in the USA and Africa of GAP direct venous inoculation [213]. Immunization by administration of fully infectious sporozoites in individuals that are under simultaneous chemoprophylaxis with chloroquine (PfSPZ-CVac) is also of great future interest. Given that such sporozoites can complete the liver stage before being killed off in the bloodstream by chloroquine, expectations are that the immune system is exposed to a broader repertoire of parasite antigens and might be able to more efficiently mount protective and long-term immune responses against the liver stages of the parasite [20,48].

Looking ahead, further advances of the MALDI-TOF MS technology in the near future will potentially allow the interrogation of even broader molecular weight ranges at decreased measurement error rates. This would significantly enhance the here presented rsp-method, by increasing the number of GBS rsp that can be detected and by allowing distinction of molecular mass differences below 400ppm. In addition, novel proteins besides of the rsp might be identified and integrated in the present scheme as additional biomarkers.

Collectively, these advances could also facilitate the direct typing of GBS from the source sample without the need for previous microbial culture, which would significantly increase the overall speed of the method.

Despite of high initial acquisition costs and requirement for regular maintenance by certified technicians, MALDI-TOF MS is not restricted to industrialized countries only. Reports from studies employing MALDI-TOF MS for microbial analysis of environmental or clinical samples in South Africa [214,215] confirm the suitability of the technology in resource-limited settings. In a recent study reporting the implementation of a bioMérieux VITEK MS MALDI-TOF platform for hospital routine identification of bacteria and fungi in Dakar, Senegal, a comprehensive sample set consisting of close to 2,500 isolates was identified with high accuracy (94.2 %) at the species level [216]. Interestingly, this study reported *S. agalactiae* to occur at higher frequency in Dakar compared to GBS prevalence in France, reminding us of the persisting lack of GBS epidemiological data from Africa. These preliminary studies emphasize that the excellent performance of MALDI-TOF MS in terms of sample turnover has the potential to substantially improve hospital care for patients suffering from infectious diseases in tropical Africa.

In October 2018, an Axima MALDI-TOF MS platform has been established in joint collaboration between the Swiss TPH and the Ministry of Health and Social Welfare on Bioko Island, Equatorial Guinea. This novel equipment will allow the implementation of the technology in routine clinical diagnostics in collaboration with local hospitals on the island. With regards to GBS epidemiology studies, it will enable the extension of our rsp-based genotype monitoring approach to a geographical region with no pre-existing knowledge on GBS transmission and genotype dynamics. The potential application of MALDI-TOF MS, especially in tropical Africa, is highly versatile and not restrained to bacterial identification and typing. Identification of tick species, common vectors of various pathogens in sub-Saharan Africa [217,218] by MALDI-TOF MS [219] (presented in the Appendix of this thesis) and the screening of *Anopheles* mosquitoes for *P. falciparum* infection [220] are two intended MALDI-TOF MS applications on Bioko Island.

The application of MALDI-TOF MS for the detection of bovine mastitis pathogens, including *S. agalactiae*, from milk samples has been reported [221]. The here presented rsp-based typing scheme could build on such workflows, allowing for genotype-level resolution screening of GBS strains. It would be highly relevant to also employ rsp-based MALDI-TOF

MS in the rapidly growing camel industry, to routinely monitor circulation of GBS strains in raw camel milk.

Lastly, in the ongoing study with our collaborators from Hong Kong, the rsp-based MALDI-TOF MS method will be used for continuous screening for emerging zoonotic GBS strains. The current GBS isolate pool (n=249) will be significantly increased, with the addition of more than 1,000 human GBS isolates and more genotypes isolated from animal food samples which are currently being collected from wet markets (Margaret Ip, personal communication).

Chapter 9

References

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

2. Naing C, Whittaker MA, Wai VN, Mak JW. Is *Plasmodium vivax* Malaria a Severe Malaria? : A Systematic Review and Meta-Analysis. PLoS Negl Trop Dis. 2014;8: e3071.

3. WHO | World malaria report 2017. In: WHO [Internet]. [cited 2 Oct 2018]. Available at: http://www.who.int/malaria/publications/world-malaria-report-2017/report/en/

4. Phillips MA, Burrows JN, Manyando C, Huijsduijnen RH van, Voorhis WCV, Wells TNC. Malaria. Nat Rev Dis Primer. 2017;3: 17050.

5. Tavares J, Formaglio P, Thiberge S, Mordelet E, Van Rooijen N, Medvinsky A, *et al*. Role of host cell traversal by the malaria sporozoite during liver infection. J Exp Med. 2013;210: 905–915.

6. Joice R, Nilsson SK, Montgomery J, Dankwa S, Egan E, Morahan B, *et al*. *Plasmodium falciparum* transmission stages accumulate in the human bone marrow. Sci Transl Med. 2014;6: 244re5.

7. World Health Organization. Severe Malaria. Trop Med Int Health. 2014;19: 7–131.

8. Wassmer SC, Grau GER. Severe malaria: what's new on the pathogenesis front? Int J Parasitol. 2017;47: 145–152.

9. WHO | Guidelines for the treatment of malaria. Third edition. In: WHO [Internet]. [cited 3 Oct 2018]. Available at: http://www.who.int/malaria/publications/atoz/9789241549127/en/

10. Mackinnon MJ, Marsh K. The selection landscape of malaria parasites. Science. 2010;328: 866–871.

11. Riley EM, Stewart VA. Immune mechanisms in malaria: new insights in vaccine development. Nat Med. 2013;19: 168–178.

12. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic variation in *Plasmodium falciparum*. Annu Rev Microbiol. 2008;62: 445–470.

13. Gupta S, Snow RW, Donnelly CA, Marsh K, Newbold C. Immunity to non-cerebral severe malaria is acquired after one or two infections. Nat Med. 1999;5: 340–343.

14. Okell LC, Ghani AC, Lyons E, Drakeley CJ. Submicroscopic infection in *Plasmodium falciparum*-endemic populations: a systematic review and meta-analysis. J Infect Dis. 2009;200: 1509–1517.

15. Tran TM, Li S, Doumbo S, Doumtabe D, Huang C-Y, Dia S, *et al*. An intensive longitudinal cohort study of Malian children and adults reveals no evidence of acquired immunity to *Plasmodium falciparum* infection. Clin Infect Dis. 2013;57: 40–47.

16. Crompton PD, Moebius J, Portugal S, Waisberg M, Hart G, Garver LS, *et al*. Malaria immunity in man and mosquito: insights into unsolved mysteries of a deadly infectious disease. Annu Rev Immunol. 2014;32: 157–187.

17. Honda T, Miyachi Y, Kabashima K. Regulatory T cells in cutaneous immune responses. J Dermatol Sci. 2011;63: 75–82.

18. Guilbride DL, Gawlinski P, Guilbride PDL. Why functional pre-erythrocytic and bloodstage malaria vaccines fail: a meta-analysis of fully protective immunizations and novel immunological model. PloS One. 2010;5: e10685.

19. Hoffman SL, Goh LML, Luke TC, Schneider I, Le TP, Doolan DL, *et al*. Protection of humans against malaria by immunization with radiation-attenuated Plasmodium falciparum sporozoites. J Infect Dis. 2002;185: 1155–1164.

20. Roestenberg M, McCall M, Hopman J, Wiersma J, Luty AJF, van Gemert GJ, *et al*. Protection against a malaria challenge by sporozoite inoculation. N Engl J Med. 2009;361: 468–477.

21. Schmidt NW, Butler NS, Badovinac VP, Harty JT. Extreme CD8 T cell requirements for anti-malarial liver-stage immunity following immunization with radiation attenuated sporozoites. PLoS Pathog. 2010;6: e1000998.

22. Day NP, Hien TT, Schollaardt T, Loc PP, Chuong LV, Chau TT, *et al*. The prognostic and pathophysiologic role of pro- and antiinflammatory cytokines in severe malaria. J Infect Dis. 1999;180:

23. Walther M, Woodruff J, Edele F, Jeffries D, Tongren JE, King E, *et al*. Innate immune responses to human malaria: heterogeneous cytokine responses to blood-stage *Plasmodium falciparum* correlate with parasitological and clinical outcomes. J Immunol Baltim Md 1950. 2006;177: 5736–5745.

24. Shio MT, Tiemi Shio M, Eisenbarth SC, Savaria M, Vinet AF, Bellemare M-J, *et al*. Malarial hemozoin activates the NLRP3 inflammasome through Lyn and Syk kinases. PLoS Pathog. 2009;5: e1000559.

25. Sharma S, DeOliveira RB, Kalantari P, Parroche P, Goutagny N, Jiang Z, *et al*. Innate immune recognition of an AT-rich stem-loop DNA motif in the *Plasmodium falciparum* genome. Immunity. 2011;35: 194–207.

26. Horowitz A, Newman KC, Evans JH, Korbel DS, Davis DM, Riley EM. Cross-talk between T cells and NK cells generates rapid effector responses to *Plasmodium falciparum*-infected erythrocytes. J Immunol Baltim Md 1950. 2010;184: 6043–6052.

27. Grau GE, Taylor TE, Molyneux ME, Wirima JJ, Vassalli P, Hommel M, *et al*. Tumor necrosis factor and disease severity in children with *falciparum* malaria. N Engl J Med. 1989;320: 1586–1591.

28. Crompton PD, Kayala MA, Traore B, Kayentao K, Ongoiba A, Weiss GE, *et al*. A prospective analysis of the Ab response to *Plasmodium falciparum* before and after a malaria season by protein microarray. Proc Natl Acad Sci U S A. 2010;107: 6958–6963.

29. Portugal S, Pierce SK, Crompton PD. Young lives lost as B cells falter: what we are learning about antibody responses in malaria. J Immunol Baltim Md 1950. 2013;190: 3039–3046.

30. Butler NS, Moebius J, Pewe LL, Traore B, Doumbo OK, Tygrett LT, *et al*. Therapeutic blockade of PD-L1 and LAG-3 rapidly clears established blood-stage *Plasmodium* infection. Nat Immunol. 2011;13: 188–195.

31. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, *et al*. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. 2015;526: 207–211.

32. Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, *et al*. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature. 2011;473: 212–215.

33. Hammond A, Galizi R, Kyrou K, Simoni A, Siniscalchi C, Katsanos D, *et al*. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. Nat Biotechnol. 2016;34: 78–83.

34. Chichester JA, Green BJ, Jones RM, Shoji Y, Miura K, Long CA, *et al*. Safety and immunogenicity of a plant-produced Pfs25 virus-like particle as a transmission blocking vaccine against malaria: A Phase 1 dose-escalation study in healthy adults. Vaccine. 2018;36: 5865–5871.

35. RTS,S Clinical Trials Partnership. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. Lancet Lond Engl. 2015;386: 31–45.

36. WHO | Malaria Vaccine Implementation Programme (MVIP). In: WHO [Internet]. [cited 5 Oct 2018]. Available at: http://www.who.int/immunization/diseases/malaria/malaria_vaccine_implementation_programme/en/

37. Draper SJ, Angov E, Horii T, Miller LH, Srinivasan P, Theisen M, *et al*. Recent advances in recombinant protein-based malaria vaccines. Vaccine. 2015;33: 7433–7443.

38. Clyde DF. Immunization of man against *falciparum* and *vivax* malaria by use of attenuated sporozoites. Am J Trop Med Hyg. 1975;24: 397–401.

39. Rieckmann KH, Beaudoin RL, Cassells JS, Sell KW. Use of attenuated sporozoites in the immunization of human volunteers against *falciparum* malaria. Bull World Health Organ. 1979;57 Suppl 1: 261–265.

40. Roestenberg M, Bijker EM, Sim BKL, Billingsley PF, James ER, Bastiaens GJH, *et al*. Controlled Human Malaria Infections by Intradermal Injection of Cryopreserved *Plasmodium falciparum* Sporozoites. Am J Trop Med Hyg. 2013;88: 5–13.

41. Hoffman SL, Billingsley PF, James E, Richman A, Loyevsky M, Li T, *et al*. Development of a metabolically active, non-replicating sporozoite vaccine to prevent *Plasmodium falciparum* malaria. Hum Vaccin. 2010;6: 97–106.

42. Nyunt MM, Hendrix CW, Bakshi RP, Kumar N, Shapiro TA. Phase I/II Evaluation of the Prophylactic Antimalarial Activity of Pafuramidine in Healthy Volunteers Challenged with *Plasmodium falciparum* Sporozoites. Am J Trop Med Hyg. 2009;80: 528–535.

43. McCarthy JS, Baker M, O'Rourke P, Marquart L, Griffin P, Hooft van Huijsduijnen R, *et al*. Efficacy of OZ439 (artefenomel) against early *Plasmodium falciparum* blood-stage malaria infection in healthy volunteers. J Antimicrob Chemother. 2016;71: 2620–2627.

44. Epstein JE, Tewari K, Lyke KE, Sim BKL, Billingsley PF, Laurens MB, *et al*. Live attenuated malaria vaccine designed to protect through hepatic $CD8^+$ T cell immunity. Science. 2011;334: 475–480.

45. Seder RA, Chang L-J, Enama ME, Zephir KL, Sarwar UN, Gordon IJ, *et al*. Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. Science. 2013;341: 1359–1365.

46. Sissoko MS, Healy SA, Katile A, Omaswa F, Zaidi I, Gabriel EE, *et al*. Safety and efficacy of PfSPZ Vaccine against *Plasmodium falciparum* via direct venous inoculation in healthy malaria-exposed adults in Mali: a randomised, double-blind phase 1 trial. Lancet Infect Dis. 2017;17: 498–509.

47. Olotu A, Urbano V, Hamad A, Eka M, Chemba M, Nyakarungu E, *et al*. Advancing Global Health through Development and Clinical Trials Partnerships: A Randomized, Placebo-Controlled, Double-Blind Assessment of Safety, Tolerability, and Immunogenicity of PfSPZ Vaccine for Malaria in Healthy Equatoguinean Men. Am J Trop Med Hyg. 2018;98: 308–318.

48. Mordmüller B, Surat G, Lagler H, Chakravarty S, Ishizuka AS, Lalremruata A, *et al*. Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. Nature. 2017;542: 445–449.

49. Kublin JG, Mikolajczak SA, Sack BK, Fishbaugher ME, Seilie A, Shelton L, *et al*. Complete attenuation of genetically engineered *Plasmodium falciparum* sporozoites in human subjects. Sci Transl Med. 2017;9: eaad9099.

50. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10: 57–63.

51. Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. Cold Spring Harb Protoc. 2015: 951–969.

52. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol. 2011;9: 34.

53. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, *et al*. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17.

54. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. [cited 21 Sep 2018]. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

55. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinforma Oxf Engl. 2009;25: 1105–1111.

56. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al*. STAR: ultrafast universal RNA-seq aligner. Bioinforma Oxf Engl. 2013;29: 15–21.

57. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinforma Oxf Engl. 2015;31: 166–169.

58. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, *et al*. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28: 511–515.

59. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12: 323.

60. Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat Biotechnol. 2014;32: 462–464.

61. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma Oxf Engl. 2010;26: 139–140.

62. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15.

63. Pulendran B, Li S, Nakaya HI. Systems Vaccinology. Immunity. 2010;33: 516–529.

64. Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, *et al*. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nat Immunol. 2009;10: 116–125.

65. Gaucher D, Therrien R, Kettaf N, Angermann BR, Boucher G, Filali-Mouhim A, *et al*. Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses. J Exp Med. 2008;205: 3119–3131.

66. Kazmin D, Nakaya HI, Lee EK, Johnson MJ, van der Most R, van den Berg RA, *et al*. Systems analysis of protective immune responses to RTS,S malaria vaccination in humans. Proc Natl Acad Sci U S A. 2017;114: 2425–2430.

67. Nuccitelli A, Rinaudo CD, Maione D. Group B *Streptococcus* vaccine: state of the art. Ther Adv Vaccines. 2015;3: 76–90.

68. Evans JJ, Klesius PH, Pasnik DJ, Bohnsack JF. Human *Streptococcus agalactiae* isolate in Nile tilapia (*Oreochromis niloticus*). Emerg Infect Dis. 2009;15: 774–776.

69. Nocard N, Mollereau R. Sur une mammite contagieuse des vaches laitieres. Ann Inst Pasteur. 1887;

70. Fry RM. Fatal Infections by Haemolytic *Streptococcus* Group B. The Lancet. 1938;231: 199–201.

71. Russell NJ, Seale AC, O'Driscoll M, O'Sullivan C, Bianchi-Jassir F, Gonzalez-Guarin J, *et al*. Maternal Colonization With Group B *Streptococcus* and Serotype Distribution Worldwide: Systematic Review and Meta-analyses. Clin Infect Dis. 2017;65: S100–S111.

72. Madrid L, Seale AC, Kohli-Lynch M, Edmond KM, Lawn JE, Heath PT, *et al*. Infant Group B Streptococcal Disease Incidence and Serotypes Worldwide: Systematic Review and Meta-analyses. Clin Infect Dis. 2017;65: S160–S172.

73. Kohli-Lynch M, Russell NJ, Seale AC, Dangor Z, Tann CJ, Baker CJ, *et al*. Neurodevelopmental Impairment in Children After Group B Streptococcal Disease Worldwide: Systematic Review and Meta-analyses. Clin Infect Dis. 2017;65: S190–S199.

74. Lawn JE, Bianchi-Jassir F, Russell NJ, Kohli-Lynch M, Tann CJ, Hall J, *et al*. Group B Streptococcal Disease Worldwide for Pregnant Women, Stillbirths, and Children: Why, What, and How to Undertake Estimates? Clin Infect Dis. 2017;65: S89–S99.

75. Seale AC, Blencowe H, Bianchi-Jassir F, Embleton N, Bassat Q, Ordi J, *et al*. Stillbirth With Group B *Streptococcus* Disease Worldwide: Systematic Review and Meta-analyses. Clin Infect Dis. 2017;65: S125–S132.

76. Bianchi-Jassir F, Seale AC, Kohli-Lynch M, Lawn JE, Baker CJ, Bartlett L, *et al*. Preterm Birth Associated With Group B *Streptococcus* Maternal Colonization Worldwide: Systematic Review and Meta-analyses. Clin Infect Dis. 2017;65: S133–S142.

77. Hall J, Adams NH, Bartlett L, Seale AC, Lamagni T, Bianchi-Jassir F, *et al*. Maternal Disease With Group B *Streptococcus* and Serotype Distribution Worldwide: Systematic Review and Meta-analyses. Clin Infect Dis. 2017;65: S112–S124.

78. Farley MM, Strasbaugh LJ. Group B Streptococcal Disease in Nonpregnant Adults. Clin Infect Dis. 2001;33: 556–561.

79. Skoff TH, Farley MM, Petit S, Craig AS, Schaffner W, Gershman K, *et al*. Increasing Burden of Invasive Group B Streptococcal Disease in Nonpregnant Adults, 1990–2007. Clin Infect Dis. 2009;49: 85–92.

80. Shabayek S, Spellerberg B. Group B Streptococcal Colonization, Molecular Characteristics, and Epidemiology. Front Microbiol. 2018;9.

81. Rajendram P, Mar Kyaw W, Leo YS, Ho H, Chen WK, Lin R, *et al*. Group B *Streptococcus* Sequence Type 283 Disease Linked to Consumption of Raw Fish, Singapore. Emerg Infect Dis. 2016;22: 1974–1977.

82. Tan S, Lin Y, Foo K, Koh HF, Tow C, Zhang Y, *et al*. Group B *Streptococcus* Serotype III Sequence Type 283 Bacteremia Associated with Consumption of Raw Fish, Singapore. Emerg Infect Dis. 2016;22: 1970–1973.

83. Lyhs U, Kulkas L, Katholm J, Waller KP, Saha K, Tomusk RJ, *et al*. *Streptococcus agalactiae* Serotype IV in Humans and Cattle, Northern Europe1. Emerg Infect Dis. 2016;22: 2097–2103.

84. Boyer KM, Gotoff SP. Prevention of early-onset neonatal group B streptococcal disease with selective intrapartum chemoprophylaxis. N Engl J Med. 1986;314: 1665–1669.

85. Tuppurainen N, Hallman M. Prevention of neonatal group B streptococcal disease: intrapartum detection and chemoprophylaxis of heavily colonized parturients. Obstet Gynecol. 1989;73: 583–587.

86. Kobayashi M, Vekemans J, Baker CJ, Ratner AJ, Le Doare K, Schrag SJ. Group B *Streptococcus* vaccine development: present status and future considerations, with emphasis on perspectives for low and middle income countries. F1000Research. 2016;5.

87. Binghuai L, Yanli S, Shuchen Z, Fengxia Z, Dong L, Yanchao C. Use of MALDI-TOF mass spectrometry for rapid identification of group B *Streptococcus* on chromID Strepto B agar. Int J Infect Dis IJID. 2014;27: 44–48.

88. Rosa-Fraile M, Spellerberg B. Reliable Detection of Group B *Streptococcus* in the Clinical Laboratory. J Clin Microbiol. 2017;55: 2590–2598.

89. Russell NJ, Seale AC, O'Sullivan C, Le Doare K, Heath PT, Lawn JE, *et al*. Risk of Early-Onset Neonatal Group B Streptococcal Disease With Maternal Colonization Worldwide: Systematic Review and Meta-analyses. Clin Infect Dis. 2017;65: S152–S159.

90. Jordan HT, Farley MM, Craig A, Mohle-Boetani J, Harrison LH, Petit S, *et al*. Revisiting the need for vaccine prevention of late-onset neonatal group B streptococcal disease: a multistate, population-based analysis. Pediatr Infect Dis J. 2008;27: 1057–1064.

91. Dahesh S, Hensler ME, Van Sorge NM, Gertz RE, Schrag S, Nizet V, *et al*. Point mutation in the group B streptococcal pbp2x gene conferring decreased susceptibility to beta-lactam antibiotics. Antimicrob Agents Chemother. 2008;52: 2915–2918.

92. Kimura K, Suzuki S, Wachino J, Kurokawa H, Yamane K, Shibata N, *et al*. First Molecular Characterization of Group B *Streptococci* with Reduced Penicillin Susceptibility. Antimicrob Agents Chemother. 2008;52: 2890–2897.

93. Landwehr-Kenzel S, Henneke P. Interaction of *Streptococcus agalactiae* and Cellular Innate Immunity in Colonization and Disease. Front Immunol. 2014;5.

94. Maisey HC, Doran KS, Nizet V. Recent advances in understanding the molecular basis of group B *Streptococcus* virulence. Expert Rev Mol Med. 2008;10: e27.

95. Carlin AF, Lewis AL, Varki A, Nizet V. Group B streptococcal capsular sialic acids interact with siglecs (immunoglobulin-like lectins) on human leukocytes. J Bacteriol. 2007;189: 1231–1237.

96. Slotved H-C, Kong F, Lambertsen L, Sauer S, Gilbert GL. Serotype IX, a Proposed New *Streptococcus agalactiae* Serotype. J Clin Microbiol. 2007;45: 2929–2936.

97. Jiang S, Park SE, Yadav P, Paoletti LC, Wessels MR. Regulation and function of pilus island 1 in group B *streptococcus*. J Bacteriol. 2012;194: 2479–2490.

98. Rinaudo CD, Rosini R, Galeotti CL, Berti F, Necchi F, Reguzzi V, *et al*. Specific involvement of pilus type 2a in biofilm formation in group B *Streptococcus*. PloS One. 2010;5: e9216.

99. Konto-Ghiorghi Y, Mairey E, Mallet A, Duménil G, Caliot E, Trieu-Cuot P, *et al*. Dual Role for Pilus in Adherence to Epithelial Cells and Biofilm Formation in *Streptococcus agalactiae*. PLOS Pathog. 2009;5: e1000422.

100. Chattopadhyay D, Carey AJ, Caliot E, Webb RI, Layton JR, Wang Y, *et al*. Phylogenetic lineage and pilus protein Spb1/SAN1518 affect opsonin-independent phagocytosis and intracellular survival of Group B *Streptococcus*. Microbes Infect. 2011;13: 369–382.

101. Springman AC, Lacher DW, Waymire EA, Wengert SL, Singh P, Zadoks RN, *et al*. Pilus distribution among lineages of group b *streptococcus*: an evolutionary and clinical perspective. BMC Microbiol. 2014;14: 159.

102. Schubert A, Zakikhany K, Pietrocola G, Meinke A, Speziale P, Eikmanns BJ, *et al*. The fibrinogen receptor FbsA promotes adherence of *Streptococcus agalactiae* to human epithelial cells. Infect Immun. 2004;72: 6197–6205.

103. Gutekunst H, Eikmanns BJ, Reinscheid DJ. The novel fibrinogen-binding protein FbsB promotes *Streptococcus agalactiae* invasion into epithelial cells. Infect Immun. 2004;72: 3495–3504.

104. Tenenbaum T, Spellerberg B, Adam R, Vogel M, Kim KS, Schroten H. *Streptococcus agalactiae* invasion of human brain microvascular endothelial cells is promoted by the laminin-binding protein Lmb. Microbes Infect. 2007;9: 714–720.

105. Mu R, Kim BJ, Paco C, Del Rosario Y, Courtney HS, Doran KS. Identification of a group B streptococcal fibronectin binding protein, SfbA, that contributes to invasion of brain endothelium and development of meningitis. Infect Immun. 2014;82: 2276–2286.

106. Cheng Q, Stafslien D, Purushothaman SS, Cleary P. The Group B Streptococcal C5a Peptidase Is Both a Specific Protease and an Invasin. Infect Immun. 2002;70: 2408–2413.

107. Santi I, Scarselli M, Mariani M, Pezzicoli A, Masignani V, Taddei A, *et al*. BibA: a novel immunogenic bacterial adhesin contributing to group B *Streptococcus* survival in human blood. Mol Microbiol. 2007;63: 754–767.

108. Maeland JA, Afset JE, Lyng RV, Radtke A. Survey of Immunological Features of the Alpha-Like Proteins of *Streptococcus agalactiae*. Clin Vaccine Immunol CVI. 2015;22: 153–159.

109. Zuerlein TJ, Christensen B, Hall RT. Latex agglutination detection of group-B streptococcal inoculum in urine. Diagn Microbiol Infect Dis. 1991;14: 191–194.

110. Breeding KM, Ragipani B, Lee K-UD, Malik M, Randis TM, Ratner AJ. Real-time PCR-based serotyping of *Streptococcus agalactiae*. Sci Rep. 2016;6: 38523.

111. Yao K, Poulsen K, Maione D, Rinaudo CD, Baldassarri L, Telford JL, *et al*. Capsular gene typing of *Streptococcus agalactiae* compared to serotyping by latex agglutination. J Clin Microbiol. 2013;51: 503–507.

112. Sheppard AE, Vaughan A, Jones N, Turner P, Turner C, Efstratiou A, *et al*. Capsular Typing Method for *Streptococcus agalactiae* Using Whole-Genome Sequence Data. J Clin Microbiol. 2016;54: 1388–1390.

113. Bellais S, Six A, Fouet A, Longo M, Dmytruk N, Glaser P, *et al*. Capsular Switching in Group B *Streptococcus* CC17 Hypervirulent Clone: A Future Challenge for Polysaccharide Vaccine Development. J Infect Dis. 2012;206: 1745–1752.

114. Martins ER, Melo-Cristino J, Ramirez M. Evidence for Rare Capsular Switching in *Streptococcus agalactiae*. J Bacteriol. 2010;192: 1361–1369.

115. Neemuchwala A, Teatero S, Athey TBT, McGeer A, Fittipaldi N. Capsular Switching and Other Large-Scale Recombination Events in Invasive Sequence Type 1 Group B *Streptococcus*. Emerg Infect Dis. 2016;22: 1941–1944.

116. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan M-S, Kunst F, *et al*. Multilocus Sequence Typing System for Group B *Streptococcus*. J Clin Microbiol. 2003;41: 2530–2536.

117. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11: 595.

118. Tazi A, Disson O, Bellais S, Bouaboud A, Dmytruk N, Dramsi S, *et al*. The surface protein HvgA mediates group B *streptococcus* hypervirulence and meningeal tropism in neonates. J Exp Med. 2010;207: 2313–2322.

119. Sørensen UBS, Poulsen K, Ghezzo C, Margarit I, Kilian M. Emergence and Global Dissemination of Host-Specific *Streptococcus agalactiae* Clones. mBio. 2010;1.

120. Bisharat N, Crook DW, Leigh J, Harding RM, Ward PN, Coffey TJ, *et al*. Hyperinvasive neonatal group B *streptococcus* has arisen from a bovine ancestor. J Clin Microbiol. 2004;42: 2161–2167.

121. Brochet M, Rusniok C, Couvé E, Dramsi S, Poyart C, Trieu-Cuot P, *et al*. Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. Proc Natl Acad Sci U S A. 2008;105: 15961.

122. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, *et al*. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome." Proc Natl Acad Sci. 2005;102: 13950–13955.

123. Da Cunha V, Davies MR, Douarre P-E, Rosinski-Chupin I, Margarit I, Spinali S, *et al*. *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline. Nat Commun. 2014;5: 4544.

124. Godoy DT, Carvalho-Castro GA, Leal C a. G, Pereira UP, Leite RC, Figueiredo HCP. Genetic diversity and new genotyping scheme for fish pathogenic *Streptococcus agalactiae*. Lett Appl Microbiol. 2013;57: 476–483.

125. Fischer A, Liljander A, Kaspar H, Muriuki C, Fuxelius H-H, Bongcam-Rudloff E, *et al*. Camel *Streptococcus agalactiae* populations are associated with specific disease complexes and acquired the tetracycline resistance gene tetM via a Tn916-like element. Vet Res. 2013;44: 86.

126. Siegrist C-A, Aspinall R. B-cell responses to vaccination at the extremes of age. Nat Rev Immunol. 2009;9: 185–194.

127. Lancefield RC, Hare R. The Serological Differentiation of Pathogenic and non-Pathogenic Strains of Hemolytic *Streptococci* from Parturient Women. J Exp Med. 1935;61: 335–349.

128. Lancefield RC. Two Serological Types of Group B Hemolytic *Streptococci* with Related, but not Identical, Type-Specific Substances. J Exp Med. 1938;67: 25–40.

129. Baker CJ, Kasper DL. Correlation of maternal antibody deficiency with susceptibility to neonatal group B streptococcal infection. N Engl J Med. 1976;294: 753–756.

130. Baker CJ, Rench MA, Edwards MS, Carpenter RJ, Hays BM, Kasper DL. Immunization of Pregnant Women with a Polysaccharide Vaccine of Group B *Streptococcus*. N Engl J Med. 1988;319: 1180–1185.

131. Finn A, Heath P. Conjugate vaccines. Arch Dis Child. 2005;90: 667–669.

132. Avci FY, Kasper DL. How Bacterial Carbohydrates Influence the Adaptive Immune System. Annu Rev Immunol. 2010;28: 107–130.

133. Kasper DL, Paoletti LC, Wessels MR, Guttormsen HK, Carey VJ, Jennings HJ, *et al*. Immune response to type III group B streptococcal polysaccharide-tetanus toxoid conjugate vaccine. J Clin Invest. 1996;98: 2308–2314.

134. Pichichero ME. Protein carriers of conjugate vaccines: characteristics, development, and clinical trials. Hum Vaccines Immunother. 2013;9: 2505–2523.

135. Leroux-Roels G, Maes C, Willekens J, De Boever F, de Rooij R, Martell L, *et al*. A randomized, observer-blind Phase Ib study to identify formulations and vaccine schedules of a trivalent Group B *Streptococcus* vaccine for use in non-pregnant and pregnant women. Vaccine. 2016;34: 1786–1791.

136. Donders GGG, Halperin SA, Devlieger R, Baker S, Forte P, Wittke F, *et al*. Maternal Immunization With an Investigational Trivalent Group B Streptococcal Vaccine: A Randomized Controlled Trial. Obstet Gynecol. 2016;127: 213–221.

137. Heyderman RS, Madhi SA, French N, Cutland C, Ngwira B, Kayambo D, *et al*. Group B *streptococcus* vaccination in pregnant women with or without HIV in Africa: a non-randomised phase 2, open-label, multicentre trial. Lancet Infect Dis. 2016;16: 546–555.

138. MinervaX announces positive data from Phase I clinical trial. [Press release]. January 5th, 2017. [Internet]. Available at: http://minervax.com/news/2017/1/5/minervax-announces-positive-data-from-phase-i-clinical-trial.html

139. WHO | GBS vaccine research and development technical roadmap and WHO Preferred Product Characteristics. In: WHO [Internet]. [cited 23 Aug 2018]. Available at: http://www.who.int/immunization/research/development/ppc_groupb_strepvaccines/en/

140. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. Lancet Lond Engl. 2011;378: 1962–1973.

141. Miller E, Andrews NJ, Waight PA, Slack MP, George RC. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. Lancet Infect Dis. 2011;11: 760–768.

142. Maternal immunization against Group B *streptococcus*: World Health Organization research and development technological roadmap and preferred product characteristics. Vaccine. 2018;

143. Singhal N, Kumar M, Kanaujia PK, Virdi JS. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front Microbiol. 2015;6: 791.

144. Clark AE, Kaleta EJ, Arora A, Wolk DM. Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry: a Fundamental Shift in the Routine Practice of Clinical Microbiology. Clin Microbiol Rev. 2013;26: 547–603.

145. Wassenberg MWM, Kluytmans J a. JW, Box ATA, Bosboom RW, Buiting AGM, van Elzakker EPM, *et al*. Rapid screening of methicillin-resistant *Staphylococcus aureus* using PCR and chromogenic agar: a prospective study to evaluate costs and effects. Clin Microbiol Infect. 2010;16: 1754–1761.

146. Forrest GN. PNA FISH: present and future impact on patient management. Expert Rev Mol Diagn. 2007;7: 231–236.

147. Clarridge JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. 2004;17: 840–862.

148. Stephan R, Cernela N, Ziegler D, Pflüger V, Tonolla M, Ravasi D, *et al*. Rapid species specific identification and subtyping of *Yersinia enterocolitica* by MALDI-TOF Mass spectrometry. J Microbiol Methods. 2011;87: 150–153.

149. Ilina EN, Borovskaya AD, Malakhova MM, Vereshchagin VA, Kubanova AA, Kruglov AN, *et al*. Direct Bacterial Profiling by Matrix-Assisted Laser Desorption−Ionization Time-of-Flight Mass Spectrometry for Identification of Pathogenic *Neisseria*. J Mol Diagn. 2009;11: 75–86.

150. Schulthess B, Brodner K, Bloemberg GV, Zbinden R, Böttger EC, Hombach M. Identification of Gram-Positive *Cocci* by Use of Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry: Comparison of Different Preparation Methods and Implementation of a Practical Algorithm for Routine Diagnostics. J Clin Microbiol. 2013;51: 1834–1840.

151. Alatoom AA, Cunningham SA, Ihde SM, Mandrekar J, Patel R. Comparison of Direct Colony Method versus Extraction Method for Identification of Gram-Positive *Cocci* by Use of Bruker Biotyper Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry ▿. J Clin Microbiol. 2011;49: 2868–2873.

152. Liu H, Du Z, Wang J, Yang R. Universal Sample Preparation Method for Characterization of Bacteria by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. Appl Env Microbiol. 2007;73: 1899–1907.

153. Adams LL, Salee P, Dionne K, Carroll K, Parrish N. A novel protein extraction method for identification of *mycobacteria* using MALDI-ToF MS. J Microbiol Methods. 2015;119: 1–3.

154. Hettick JM, Kashon ML, Slaven JE, Ma Y, Simpson JP, Siegel PD, *et al*. Discrimination of intact *mycobacteria* at the strain level: a combined MALDI-TOF MS and biostatistical analysis. Proteomics. 2006;6: 6416–6425.

155. Patel R. MALDI-TOF MS for the diagnosis of infectious diseases. Clin Chem. 2015;61: 100–111.

156. Ziegler D, Pothier JF, Ardley J, Fossou RK, Pflüger V, de Meyer S, *et al*. Ribosomal protein biomarkers provide root nodule bacterial identification by MALDI-TOF MS. Appl Microbiol Biotechnol. 2015;99: 5547–5562.

157. Wolters M, Rohde H, Maier T, Belmar-Campos C, Franke G, Scherpe S, *et al*. MALDI-TOF MS fingerprinting allows for discrimination of major methicillin-resistant *Staphylococcus aureus* lineages. Int J Med Microbiol IJMM. 2011;301: 64–68.

158. Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev. 2012;36: 380–407.

159. Griffin PM, Price GR, Schooneveldt JM, Schlebusch S, Tilse MH, Urbanski T, *et al*. Use of matrix-assisted laser desorption ionization-time of flight mass spectrometry to identify vancomycin-resistant *enterococci* and investigate the epidemiology of an outbreak. J Clin Microbiol. 2012;50: 2918–2931.

160. Nakano S, Matsumura Y, Kato K, Yunoki T, Hotta G, Noguchi T, *et al*. Differentiation of vanA-positive *Enterococcus faecium* from vanA-negative *E. faecium* by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Int J Antimicrob Agents. 2014;44: 256–259.

161. Lartigue M-F, Kostrzewa M, Salloum M, Haguenoer E, Héry-Arnaud G, Domelier A-S, *et al*. Rapid detection of "highly virulent" Group B *Streptococcus* ST-17 and emerging ST-1 clones by MALDI-TOF mass spectrometry. J Microbiol Methods. 2011;86: 262–265.

162. Lin H-C, Lu J-J, Lin L-C, Ho C-M, Hwang K-P, Liu Y-C, *et al*. Identification of a proteomic biomarker associated with invasive ST1, serotype VI Group B *Streptococcus* by MALDI-TOF MS. J Microbiol Immunol Infect. 2017;

163. Lanotte P, Perivier M, Haguenoer E, Mereghetti L, Burucoa C, Claverol S, *et al*. Proteomic biomarkers associated with *Streptococcus agalactiae* invasive genogroups. PloS One. 2013;8: e54393.

164. Kassim A, Pflüger V, Premji Z, Daubenberger C, Revathi G. Comparison of biomarker based Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS) and conventional methods in the identification of clinically relevant bacteria and yeast. BMC Microbiol. 2017;17.

165. Suarez S, Ferroni A, Lotz A, Jolley KA, Guérin P, Leto J, *et al*. Ribosomal proteins as biomarkers for bacterial identification by mass spectrometry in the clinical microbiology laboratory. J Microbiol Methods. 2013;94: 390–396.

166. Teramoto K, Sato H, Sun L, Torimura M, Tao H, Yoshikawa H, *et al*. Phylogenetic classification of *Pseudomonas putida* strains by MALDI-MS using ribosomal subunit proteins as biomarkers. Anal Chem. 2007;79: 8712–8719.

167. Sato H, Teramoto K, Ishii Y, Watanabe K, Benno Y. Ribosomal protein profiling by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for phylogenety-based subspecies resolution of *Bifidobacterium longum*. Syst Appl Microbiol. 2011;34: 76–80.

168. Murray PR. What Is New in Clinical Microbiology—Microbial Identification by MALDI-TOF Mass Spectrometry. J Mol Diagn JMD. 2012;14: 419–423.

169. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, *et al*. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiol Read Engl. 2012;158: 1005–1015.

170. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, *et al*. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics. 2015;15: 141–161.

171. The Sustainable Development Goals Report 2018 | Multimedia Library - United Nations Department of Economic and Social Affairs [Internet]. [cited 15 Oct 2018]. Available at: https://www.un.org/development/desa/publications/the-sustainable-development-goals-report-2018.html

172. Austin SC, Stolley PD, Lasky T. The History of Malariotherapy for Neurosyphilis: Modern Parallels. JAMA. 1992;268: 516–519.

173. Beadle C, Long GW, Weiss WR, McElroy PD, Maret SM, Oloo AJ, *et al*. Diagnosis of malaria by detection of *Plasmodium falciparum* HRP-2 antigen with a rapid dipstick antigen-capture assay. Lancet Lond Engl. 1994;343: 564–568.

174. Stanisic DI, McCarthy JS, Good MF. Controlled Human Malaria Infection: Applications, Advances, and Challenges. Infect Immun. 2018;86: e00479-17.

175. Laurens MB, Billingsley P, Richman A, Eappen AG, Adams M, Li T, *et al*. Successful Human Infection with *P. falciparum* Using Three Aseptic Anopheles stephensi Mosquitoes: A New Model for Controlled Human Malaria Infection. PLOS ONE. 2013;8: e68969.

176. Laurens MB, Duncan CJ, Epstein JE, Hill AV, Komisar JL, Lyke KE, *et al*. A consultation on the optimization of controlled human malaria infection by mosquito bite for evaluation of candidate malaria vaccines. Vaccine. 2012;30: 5302–5304.

177. Epstein JE, Paolino KM, Richie TL, Sedegah M, Singer A, Ruben AJ, *et al*. Protection against *Plasmodium falciparum* malaria by PfSPZ Vaccine. JCI Insight. 2017;2: e89154.

178. Kollmann TR. Variation between Populations in the Innate Immune Response to Vaccine Adjuvants. Front Immunol. 2013;4.

179. Lalor MK, Ben-Smith A, Gorak-Stolinska P, Weir RE, Floyd S, Blitz R, *et al*. Population differences in immune responses to Bacille Calmette-Guérin vaccination in infancy. J Infect Dis. 2009;199: 795–800.

180. Asturias EJ, Mayorga C, Caffaro C, Ramirez P, Ram M, Verstraeten T, *et al*. Differences in the immune response to hepatitis B and *Haemophilus influenzae* type b vaccines in Guatemalan infants by ethnic group and nutritional status. Vaccine. 2009;27: 3650–3654.

181. Levine MM. Immunogenicity and efficacy of oral vaccines in developing countries: lessons from a live cholera vaccine. BMC Biol. 2010;8: 129.

182. Fine PE. Variation in protection by BCG: implications of and for heterologous immunity. Lancet Lond Engl. 1995;346: 1339–1345.

183. Hartgers FC, Yazdanbakhsh M. Co-infection of helminths and malaria: modulation of the immune responses to malaria. Parasite Immunol. 2006;28: 497–506.

184. Valdez Y, Brown EM, Finlay BB. Influence of the microbiota on vaccine effectiveness. Trends Immunol. 2014;35: 526–537.

185. Lyke KE, Ishizuka AS, Berry AA, Chakravarty S, DeZure A, Enama ME, *et al*. Attenuated PfSPZ Vaccine induces strain-transcending T cells and durable protection against heterologous controlled human malaria infection. Proc Natl Acad Sci U S A. 2017;114: 2711–2716.

186. Walk J, Schats R, Langenberg MCC, Reuling IJ, Teelen K, Roestenberg M, *et al*. Diagnosis and treatment based on quantitative PCR after controlled human malaria infection. Malar J. 2016;15.

187. Li S, Rouphael N, Duraisingham S, Romero-Steiner S, Presnell S, Davis C, *et al*. Molecular signatures of antibody responses derived from a systems biological study of 5 human vaccines. Nat Immunol. 2014;15: 195–204.

188. Vahey MT, Wang Z, Kester KE, Cummings J, Heppner DG, Nau ME, *et al*. Expression of genes associated with immunoproteasome processing of major histocompatibility complex peptides is indicative of protection with adjuvanted RTS,S malaria vaccine. J Infect Dis. 2010;201: 580–589.

189. Ockenhouse CF, Hu W, Kester KE, Cummings JF, Stewart A, Heppner DG, *et al*. Common and divergent immune response signaling pathways discovered in peripheral blood mononuclear cell gene expression patterns in presymptomatic and clinically apparent malaria. Infect Immun. 2006;74: 5561–5573.

190. Shen-Orr SS, Gaujoux R. Computational Deconvolution: Extracting Cell Type-Specific Information from Heterogeneous Samples. Curr Opin Immunol. 2013;25.

191. Stubbington MJT, Rozenblatt-Rosen O, Regev A, Teichmann SA. Single cell transcriptomics to explore the immune system in health and disease. Science. 2017;358: 58–63.

192. Berkley JA, Bejon P, Mwangi T, Gwer S, Maitland K, Williams TN, *et al*. HIV infection, malnutrition, and invasive bacterial infection among children with severe malaria. Clin Infect Dis. 2009;49: 336–343.

193. Cunnington AJ, de Souza JB, Walther R-M, Riley EM. Malaria impairs resistance to *Salmonella* through heme- and heme oxygenase-dependent dysfunctional granulocyte mobilization. Nat Med. 2011;18: 120–127.

194. Mooney JP, Barry A, Gonçalves BP, Tiono AB, Awandu SS, Grignard L, *et al*. Haemolysis and haem oxygenase-1 induction during persistent “asymptomatic” malaria infection in Burkinabé children. Malar J. 2018;17.

195. Sinha A, Russell LB, Tomczyk S, Verani JR, Schrag SJ, Berkley JA, *et al*. Disease Burden of Group B *Streptococcus* Among Infants in Sub-Saharan Africa: A Systematic Literature Review and Meta-analysis. Pediatr Infect Dis J. 2016;35: 933–942.

196. van Belkum A, Welker M, Pincus D, Charrier J-P, Girard V. Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry in Clinical Microbiology: What Are the Current Issues? Ann Lab Med. 2017;37: 475–483.

197. Valentine N, Wunschel S, Wunschel D, Petersen C, Wahl K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. Appl Environ Microbiol. 2005;71: 58–64.

198. Campisi E, Rinaudo CD, Donati C, Barucco M, Torricelli G, Edwards MS, *et al*. Serotype IV *Streptococcus agalactiae* ST-452 has arisen from large genomic recombination events between CC23 and the hypervirulent CC17 lineages. Sci Rep. 2016;6.

199. Wang R, Li L, Huang T, Huang Y, Huang W, Yang X, *et al*. Phylogenetic, comparative genomic and structural analyses of human *Streptococcus agalactiae* ST485 in China. BMC Genomics. 2018;19.

200. Li L, Wang R, Huang Y, Huang T, Luo F, Huang W, *et al*. High Incidence of Pathogenic *Streptococcus agalactiae* ST485 Strain in Pregnant/Puerperal Women and Isolation of Hyper-Virulent Human CC67 Strain. Front Microbiol. 2018;9.

201. Zadoks RN, Middleton JR, McDougall S, Katholm J, Schukken YH. Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans. J Mammary Gland Biol Neoplasia. 2011;16: 357–372.

202. Yang Y, Liu Y, Ding Y, Yi L, Ma Z, Fan H, *et al*. Molecular characterization of *Streptococcus agalactiae* isolated from bovine mastitis in Eastern China. PloS One. 2013;8: e67755.

203. Wibawan IW, Lämmler C, Smola J. Properties and type antigen patterns of group B streptococcal isolates from pigs and nutrias. J Clin Microbiol. 1993;31: 762–764.

204. O'Sullivan T, Friendship R, Blackwell T, Pearl D, McEwen B, Carman S, *et al*. Microbiological identification and analysis of swine tonsils collected from carcasses at slaughter. Can J Vet Res. 2011;75: 106–111.

205. Manning SD, Springman AC, Million AD, Milton NR, McNamara SE, Somsel PA, *et al*. Association of Group B *Streptococcus* Colonization and Bovine Exposure: A Prospective Cross-Sectional Cohort Study. PLOS ONE. 2010;5: e8795.

206. Manning SD, Neighbors K, Tallman PA, Gillespie B, Marrs CF, Borchardt SM, *et al*. Prevalence of Group B *Streptococcus* Colonization and Potential for Transmission by Casual Contact in Healthy Young Men and Women. Clin Infect Dis. 2004;39: 380–388.

207. Gray GC, Merchant JA. Pigs, pathogens, and public health. Lancet Infect Dis. 2018;18: 372–373.

208. Ma M-J, Wang G-L, Anderson BD, Bi Z-Q, Lu B, Wang X-J, *et al*. Evidence for Cross-species Influenza A Virus Transmission Within Swine Farms, China: A One Health, Prospective Cohort Study. Clin Infect Dis. 2018;66: 533–540.

209. Chattopadhyay R, Pratt D. Role of controlled human malaria infection (CHMI) in malaria vaccine development: A U.S. food & drug administration (FDA) perspective. Vaccine. 2017;35: 2767–2769.

210. Nabet C, Doumbo S, Jeddi F, Konaté S, Manciulli T, Fofana B, *et al*. Genetic diversity of *Plasmodium falciparum* in human malaria cases in Mali. Malar J. 2016;15: 353.

211. Zhong D, Afrane Y, Githeko A, Yang Z, Cui L, Menge DM, *et al*. *Plasmodium falciparum* Genetic Diversity in Western Kenya Highlands. Am J Trop Med Hyg. 2007;77: 1043–1050.

212. Radley DE, Brown CGD, Cunningham MP, Kimber CD, Musisi FL, Payne RC, *et al*. East coast fever: 3. Chemoprophylactic immunization of cattle using oxytetracycline and a combination of theilerial strains. Vet Parasitol. 1975;1: 51–60.

213. Coelho CH, Doritchamou JYA, Zaidi I, Duffy PE. Advances in malaria vaccine development: report from the 2017 malaria vaccine symposium. Npj Vaccines. 2017;2: 34.

214. Gemmell ME, Schmidt S. Is the microbiological quality of the Msunduzi River (KwaZulu-Natal, South Africa) suitable for domestic, recreational, and agricultural purposes? Environ Sci Pollut Res. 2013;20: 6551–6562.

215. Bulane A, Hoosen A. Use of matrix-assisted laser desorption/ionisation-time of flight mass spectrometry analyser in a diagnostic microbiology laboratory in a developing country. Afr J Lab Med. 2017;6: 1–6.

216. Fall B, Lo CI, Samb-Ba B, Perrot N, Diawara S, Gueye MW, *et al*. The Ongoing Revolution of MALDI-TOF Mass Spectrometry for Microbiology Reaches Tropical Africa. Am J Trop Med Hyg. 2015;92: 641–647.

217. Ehounoud CB, Yao KP, Dahmani M, Achi YL, Amanzougaghene N, N'Douba AK, *et al*. Multiple Pathogens Including Potential New Species in Tick Vectors in Côte d'Ivoire. PLoS Negl Trop Dis. 2016;10: e0004367.

218. Parola P. Rickettsioses in sub-Saharan Africa. Ann N Y Acad Sci. 2006;1078: 42–47.

219. Rothen J, Githaka N, Kanduma EG, Olds C, Pflüger V, Mwaura S, *et al*. Matrix-assisted laser desorption/ionization time of flight mass spectrometry for comprehensive indexing of East African ixodid tick species. Parasit Vectors. 2016;9: 151.

220. Laroche M, Almeras L, Pecchi E, Bechah Y, Raoult D, Viola A, *et al*. MALDI-TOF MS as an innovative tool for detection of Plasmodium parasites in Anopheles mosquitoes. Malar J. 2017;16.

221. Barreiro JR, Gonçalves JL, Grenfell R, Leite RF, Juliano L, Santos MV. Direct identification of bovine mastitis pathogens by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in pre-incubated milk. Braz J Microbiol. 2018;49: 801–807.

Appendix

A: Supplementary information for *Chapter 4*

Subspecies typing of *Streptococcus agalactiae* based on ribosomal subunit protein mass variation by MALDI-TOF MS

Supplementary Information for

Subspecies typing of *Streptococcus agalactiae* based on ribosomal subunit protein mass variation by MALDI-TOF MS

Julian Rothen; JR*, Joël F. Pothier; JFP, Frédéric Foucault; FF, Jochen Blom; JB, Dulmini Nanayakkara; DN, Carmen Li; CL, Margaret Ip; MI, Marcel Tanner; MT, Guido Vogel; GV, Valentin Pflüger; VP*, Claudia A. Daubenberger; CAD

*Corresponding authors:
Julian Rothen
Email: julian.rothen@unibas.ch
Valentin Pflüger
Email: valentin.pflueger@mabritec.com

This PDF file includes:

Supplementary text
Figs. S2, S3, S4 and S5
Captions for Tables S1, S6 and S7
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Tables S1, S6 and S7 (one file).

Supplementary Information Text

Methods

Core-genome phylogenetic analysis

Automatic genome annotation of the WGS was performed with the Prokka software tool version 1.12 (1), using a *Streptococcus* genus database. The core-genome phylogenetic relationships of the WGS were obtained using EDGAR version 2.2 (2). Briefly, the core-genome was defined by iterative pairwise comparison of the gene content of each of the selected genomes using the bidirectional best hits (BBH) with score ratio values as orthology criterion (2). For all calculations protein BLAST (BLASTp) was used with BLOSUM62 as similarity matrix (3, 4). Multiple alignments of each of the 867 orthologous gene set of the core genome were calculated using the MUSCLE software (5), which equaled 690,132 genes in total. The resulting alignments were concatenated to one huge alignment (6), which consisted of 212,086,240 amino acid residues, 266,440 per genome. This alignment was used to construct a FastTree phylogeny (7). Phylogenetic trees were visualized and edited using the interactive tree of life (iTOL) website (8).

Bacteria cultivation and sample preparation

GBS isolates were stored at -80 °C prior to cultivation. After thawing the isolates on ice, bacterial material was plated on Columbia Sheep Blood Agar. The plates were then stored at 37 °C in the incubator for overnight cultivation. Single colonies were transferred to a new agar plate using the four-quadrant streak method. After repeated overnight cultivation at 37 °C, *S. agalactiae* colonies were harvested for sample preparation. The bacteria material was washed repeatedly in TMA buffer (10mM Tris-HCl (pH 7.8), 30 mM NH_4Cl, 10 mM $MgCl_2$, and 6 mM 2-mercaptoethanol). In a next step, the bacterial cells were disrupted using a FastPrep FP120 bead beater in order to lay open intracellular proteins. To that end, the washed cells were transferred together with 0.1 mm glass beads to a micro tube. The mixture was agitated for multiple short time intervals (20 seconds) at maximum speed, interrupted by cooling intervals (1 minute) on ice. In a last step, protein fragments smaller than 3,000 Dalton (Da) were removed by filtering of the bacterial extract with Amicon^{TM} Ultra centrifugal devices. Lastly, the concentrated sample was mixed with the tenfold volume of ddH_2O and 1 µl of the dilution was applied in quadruplicates on a MALDI-TOF steel target plate. The spotted samples were left to air dry at room temperature and consequently overlaid with a matrix consisting of a saturated solution of 10 mg sinapic acid in 60% acetonitrile, 40 % ddH_2O and 1% TFA.

MALDI-TOF MS analyses

Instrument setup: The MS measurements were carried out using a MALDI-TOF Mass Spectrometer Axima Confidence machine (Shimadzu-Biotech, Kyoto, Japan) with detection in the linear positive mode, allowing the interrogation of high molecular weight samples. The acceleration voltage was set by default to 20 kV with an extraction delay time of 200 ns and a laser frequency of 50 Hz. The analysis was carried out in the mass range between 4,000 and 25,000 Da. To ensure an even measurement covering the entire area of the sample spot, a netlike pattern of 100 equally distributed locations was defined. At each of these profiles ten consecutive laser shots were applied, adding up to 1,000 laser shots per sample spot. The ion gate was set at 3,950 Da and the pulsed extraction optimized at 20,000 Da. Each target plate was externally calibrated using the reference spectra of Escherichia coli strain DH5α.

Mass spectra processing and internal calibration:

The individual mass fingerprints were averaged and the spectra further processed with the Launchpad 2.8 software (Shimadzu-Biotech, Kyoto, Japan). The advanced scenario setting was chosen for peak processing, with a defined peak width of 80 chans, smoothing filter width of 50 chans and baseline filter width of 500 chans. An adaptive voltage threshold, which roughly followed the signal noise level, was defined and the threshold offset and threshold response set to 0.008 and 1.000 respectively. Internal calibration with 800 ppm was carried out with MALDIquant (9), using 10 rsp masses (3 mass alleles of L6, 2 mass alleles of L36 and S12, 1 mass allele of L14, L29 and S15) that altogether display mass values distributed over a wide mass range (4,425 to 19,293 Da). An ascii file containing the recalibrated protein mass values and corresponding intensities was automatically generated for every mass spectrum.

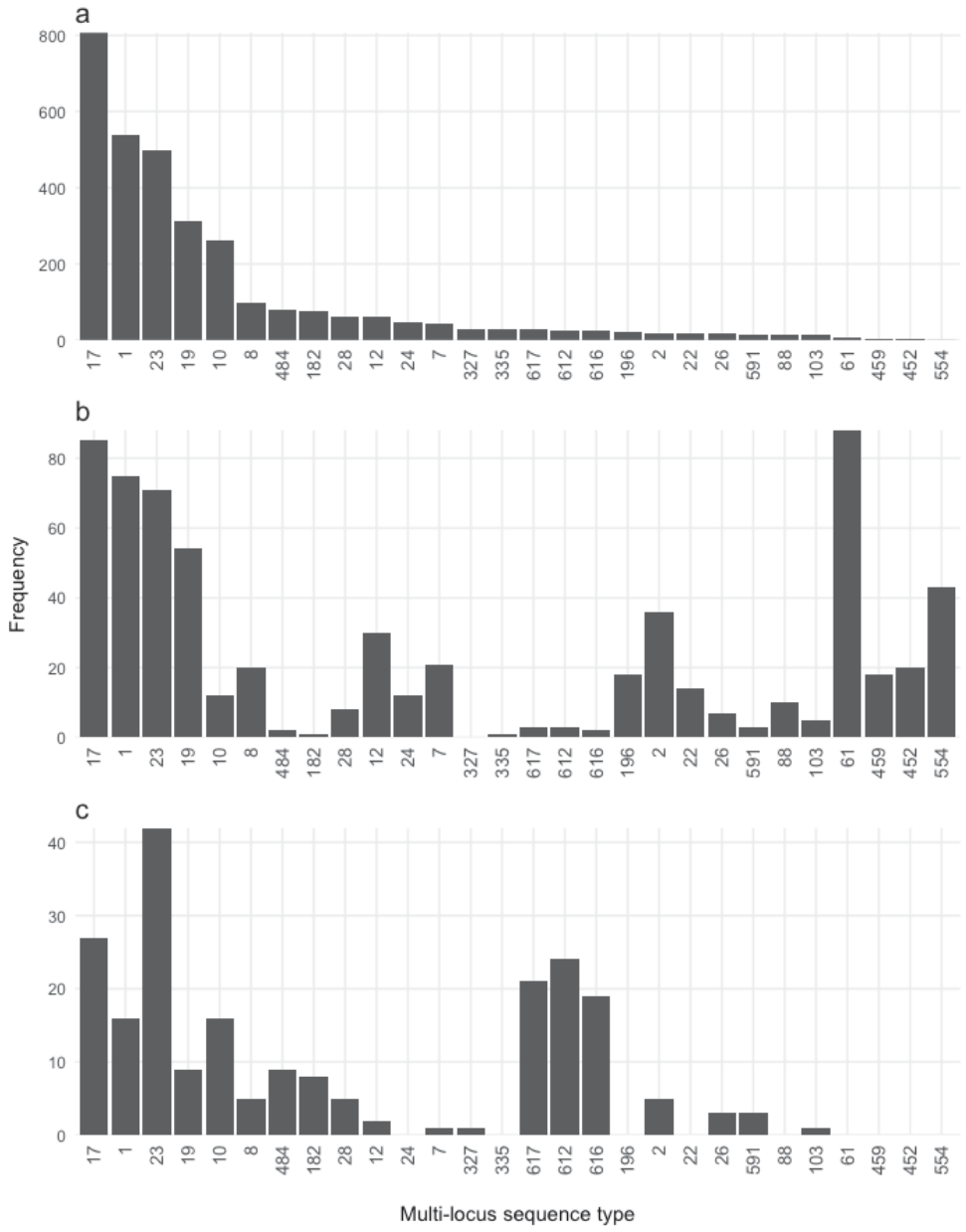


Fig. S2. Frequencies of 28 multi-locus sequence types in: (a) global Group B *Streptococcus* (GBS) population as assessed on 05-10-2017 on PubMLST (10); (b) 796 GBS whole genome sequences analyzed in this study; (c) 248 GBS isolates analyzed by MALDI-TOF MS in this study.

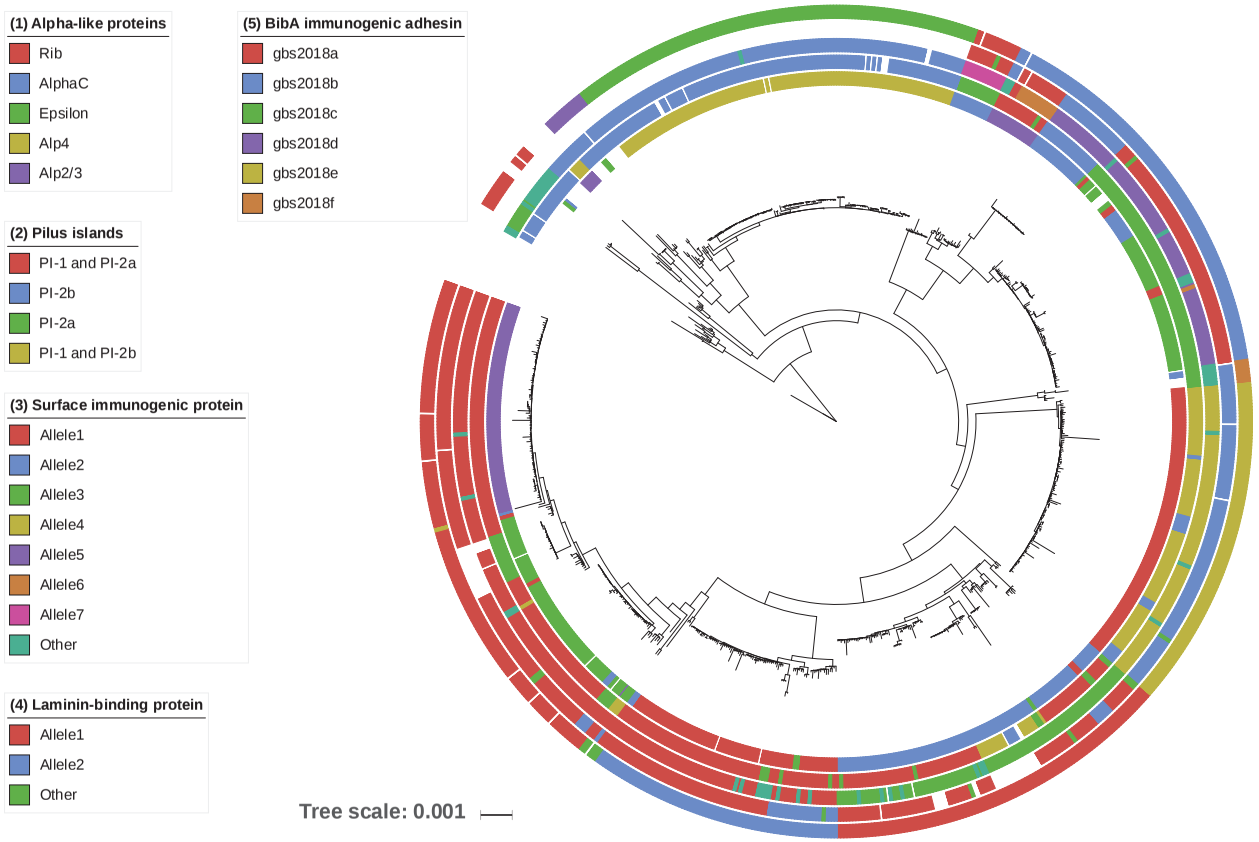


Fig. S3. Phyolgenetic tree **(**FastTree) based on core-genome analysis of 796 Group B *Streptococcus* (GBS) whole genome sequences. Each GBS strain is annotated with its *in silico* determined surface protein allelic variants. From inner circle (1) to outer circle (5): Alpha-like protein (Alp) family, pilus islands, surface immunogenic protein (Sip), laminin-binding protein (Lmb) and Group B *Streptococcus* immunogenic bacterial adhesin (BibA). (Scale bar: nucleotide substitutions per site).

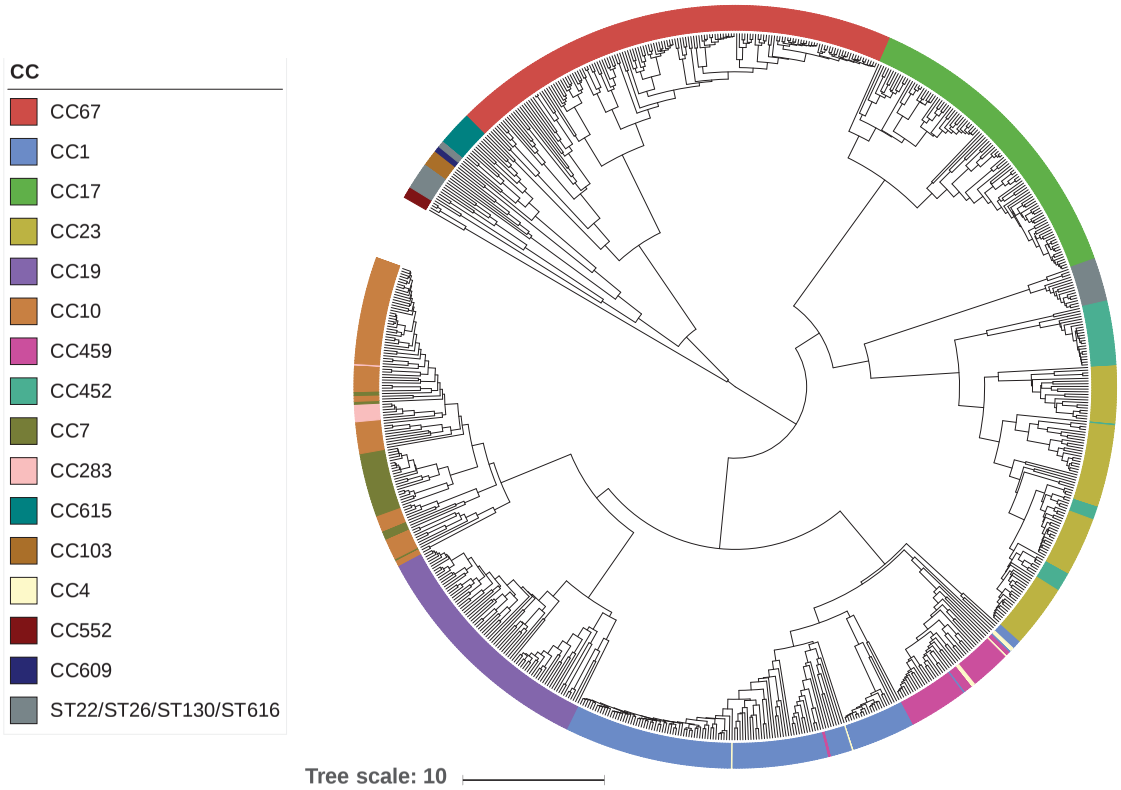


Fig. S4. UPGMA phylogenetic tree based on average nucleotide identity analysis (ANI) of 796 Group B *Streptococcus* (GBS) whole genome sequences. GBS strains are color-coded based on multi-locus sequence typing (MLST) clonal complex (CC) identity. (Scale bar: nucleotide substitutions per site).

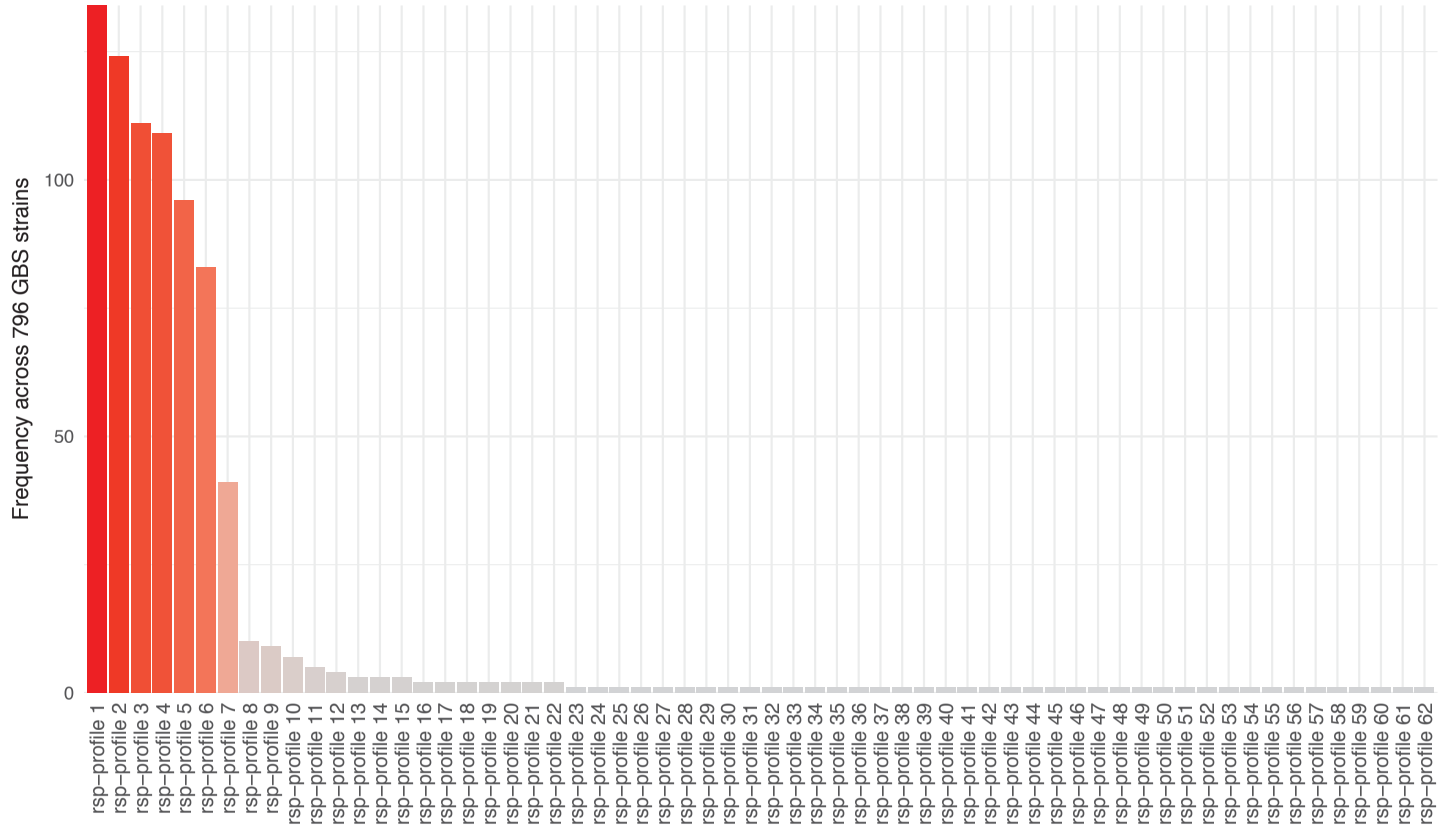


Fig. S5. Frequencies of ribosomal subunit protein (rsp) profiles across 796 Group B *Streptococcus* (GBS) strains. *In silico* prediction of molecular mass variability of 28 rsp revealed the occurrence of six dominat rsp-profiles (1-6) in the global GBS population.

Additional data table S1 (separate file)

Metadata of 796 whole genome sequenced Group B *Streptococcus* (GBS) strains used in this study. ST: multi-locus sequence type; SLV: single-locus variant; CC: clonal complex; Alp: alpha-like protein; Sip: surface immunogenic protein; Lmb: laminin-binding protein; BibA: group B *Streptococcus* immunogenic bacterial adhesin protein.

Additional data table S6 (separate file)

Metadata of 248 Group B *Streptococcus* (GBS) in-house isolates and 8 GBS isolates from an external laboratory that were subtyped by MALDI-TOF MS in this study.

Additional data table S7 (separate file)

Protein queries used for *in silico* identification of major Group B *Streptococcus* surface protein variants.

References

1. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinforma Oxf Engl* 30(14):2068–2069.

2. Blom J, et al. (2016) EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res* 44(W1):W22-28.

3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.

4. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* 89(22):10915–10919.

5. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.

6. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56(4):564–577.

7. Price MN, Dehal PS, Arkin AP (2009) FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol Biol Evol* 26(7):1641–1650.

8. Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44(W1):W242-245.

9. Gibb S, Strimmer K (2012) MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinforma Oxf Engl* 28(17):2270–2271.

10. Jolley KA, Maiden MC (2010) BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11(1):595.

B: Matrix-assisted laser desorption/ionization time of flight mass spectrometry for comprehensive indexing of East African ixodid tick species

---

This appendix section contains the following publication:

Julian Rothen, Naftaly Githaka, Esther Kanduma, Cassandra Olds, Valentin Pflüger, Stephen Mwaura, Richard Bishop, Claudia Daubenberger. "Matrix-assisted laser desorption/ionization time of flight mass spectrometry for comprehensive indexing of East African ixodid tick species" 2016. *Parasites & Vectors*.

Rothen et al. Parasites & Vectors (2016) 9:151
DOI 10.1186/s13071-016-1424-6

Parasites & Vectors

RESEARCH Open Access

Matrix-assisted laser desorption/ionization time of flight mass spectrometry for comprehensive indexing of East African ixodid tick species

Julian Rothen[2,3*†], Naftaly Githaka[1†], Esther G. Kanduma[6,7], Cassandra Olds[5], Valentin Pflüger[4], Stephen Mwaura[1], Richard P. Bishop[1] and Claudia Daubenberger[2,3]

Abstract

Background: The tick population of Africa includes several important genera belonging to the family Ixodidae. Many of these ticks are vectors of protozoan and rickettsial pathogens including *Theileria parva* that causes East Coast fever, a debilitating cattle disease endemic to eastern, central and southern Africa. Effective surveillance of tick-borne pathogens depends on accurate identification and mapping of their tick vectors. A simple and reproducible technique for rapid and reliable differentiation of large numbers of closely related field-collected ticks, which are often difficult and tedious to discriminate purely by morphology, will be an essential component of this strategy. Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) is increasingly becoming a useful tool in arthropod identification and has the potential to overcome the limitations of classical morphology-based species identification. In this study, we applied MALDI-TOF MS to a collection of laboratory and field ticks found in Eastern Africa. The objective was to determine the utility of this proteomic tool for reliable species identification of closely related afrotropical ticks.

Methods: A total of 398 ixodid ticks from laboratory maintained colonies, extracted from the hides of animals or systematically collected from vegetation in Kenya, Sudan and Zimbabwe were analyzed in the present investigation. The cytochrome *c* oxidase I (COI) genes from 33 specimens were sequenced to confirm the tentatively assigned specimen taxa identity on the basis of morphological analyses. Subsequently, the legs of ticks were homogenized and analyzed by MALDI-TOF MS. A collection of reference mass spectra, based on the mass profiles of four individual ticks per species, was developed and deposited in the spectral database SARAMIS™. The ability of these superspectra (SSp.) to identify and reliably validate a set of ticks was demonstrated using the remaining individual 333 ticks.

(Continued on next page)

* Correspondence: julian.rothen@unibas.ch
†Equal contributors
[2]Department of Medical Parasitology and Infection Biology, Clinical Immunology Unit, Swiss Tropical and Public Health Institute (Swiss TPH), Socinstr. 57, CH 4002 Basel, Switzerland
[3]University of Basel, Petersplatz 1, CH 4003 Basel, Switzerland
Full list of author information is available at the end of the article

(Continued from previous page)

Results: Ultimately, ten different tick species within the genera *Amblyomma*, *Hyalomma*, *Rhipicephalus* and *Rhipicephalus (Boophilus)* based on molecular COI typing and morphology were included into the study analysis. The robustness of the 12 distinct SSp. developed here proved to be very high, with 319 out of 333 ticks used for validation identified correctly at species level. Moreover, these novel SSp. allowed for diagnostic specificity of 99.7 %. The failure of species identification for 14 ticks was directly linked to low quality mass spectra, most likely due to poor specimen quality that was received in the laboratory before sample preparation.

Conclusions: Our results are consistent with earlier studies demonstrating the potential of MALDI-TOF MS as a reliable tool for differentiating ticks originating from the field, especially females that are difficult to identify after blood feeding. This work provides further evidence of the utility of MALDI-TOF MS to identify morphologically and genetically highly similar tick species and indicates the potential of this tool for large-scale monitoring of tick populations, species distributions and host preferences.

Keywords: MALDI-TOF MS, Ticks, Species identification, Vector epidemiology, COI, *Amblyomma*, *Boophilus*, *Hyalomma*, *Rhipicephalus*

Background

As obligate hematophagous organisms, ticks can acquire and transmit pathogenic microorganisms such as eukaryotic parasites, bacteria, viruses and fungi both through vertical transmission or when feeding on their hosts [1]. Tick-borne diseases (TBDs) cause significant economic losses to the cattle industry in tropical and subtropical regions of the world [2]. Some tick species are capable of building up focally highly dense populations, causing additional production losses in farm animals from irritation, skin damage and accompanying chronic inflammation, blood loss and in some cases, secondary infections [1, 3]. In most of Eastern Africa, several ixodid tick species share overlapping habitats and multiple tick infestations in livestock is frequently observed [4]. The control of TBDs can be improved and targeted appropriately by accurate monitoring of tick vectors. This has traditionally been done by examining morphological features using a light microscope, and with the aid of taxonomical descriptions and illustrations [4, 5]. Unfortunately, the expert knowledge required for this task is rare in most settings where TBDs are endemic [4]. In addition, damaged or immature tick stages, or replete female ticks are often difficult to identify accurately based on morphological features alone [6]. Molecular approaches like sequencing of the cytochrome *c* oxidase subunit I (COI), the 12S rDNA or the internal transcribed spacer 2 (ITS2) can overcome the limitations of conventional tick taxonomy. Due to the labor, time and costs involved in DNA extraction, PCR amplification, purification and nucleotide sequencing, this approach is typically limited to well-equipped laboratories [7, 8]. When COI, 12S or ITS reference sequences are scarce or missing from public nucleotide databases, it is difficult to conclusively resolve the species level thus non-identical sequences may remain unidentifiable [9]. Additionally, public databases are known to sometimes contain misidentified species, and sequences showing errors or obtained from contaminated samples resulting in inaccurate classification [10]. Hence, a marker-based identification system is useful to supplement morphological species identification and support taxonomy, either as corroborating evidence for existing hypotheses or as a starting point for further testing using additional techniques [11].

Matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) is emerging as an alternative to both morphology and PCR-based typing for identifying disease vectors such as mosquitoes [12, 13], tsetse fly [14] and ticks [15, 16]. MALDI-TOF MS makes use of a small quantity of whole organism material, and thus can identify damaged tick specimens or immature stages [16]. As a diagnostic technique, MALDI-TOF MS is both cost-effective and rapid, can be performed without in depth technical knowledge and the data results have been found to be highly reproducible [17]. In disease endemic areas, MALDI-TOF MS could assist in resolving questions that are difficult to answer with traditional morphological or current molecular based typing methods. For example, the unclear epidemiological status of the two closely related tick species *Rhipicephalus simus* and *Rhipicephalus praetextatus* in Kenya, where COI sequencing of field samples strongly indicates occurrence of *R. simus*, although this species is currently thought to be confined to Southern Africa [4]. It is possible that ixodid species other than *Rhipicephalus appendiculatus* may transmit *Theileria parva* in Eastern Africa, because number of *Rhipicephalus* species on domestic animals is greater than previously thought (E. Kanduma and R. Bishop, unpublished data), but more in-depth, higher resolution analyses of both tick vectors and the pathogens they transmit are necessary to confirm this. Moreover, the vector of *Theileria* sp. (buffalo) a species that is infective to cattle

at livestock-wildlife interface with unknown consequences in respect of pathology, especially in the co-infection situation is currently unknown [18]. These and similar questions require resolution especially in the context of the epidemiology of theileriosis at the livestock-wildlife interface [19, 20]. Methods endowed with higher resolution and throughput ideally for both the tick vectors and the pathogens that they transmit will be required in future to follow tick borne disease epidemiology, particularly in times of rapid climate changes in these regions [18]. The objective of this study was to extend the application of MALDI-TOF as a high-throughput tick typing method [15, 16] to a collection of Afrotropical ixodid ticks obtained from Eastern Africa. We envisage that the data from our tick collection will serve as a reference for indexing the multiple ixodid tick species that frequently occur sympatrically in Africa.

Methods

Laboratory reared ticks

A total of 398 adult ticks built the basis of this study. One hundred fifty six ticks were obtained from colonies that had been bred and maintained as closed genetic stocks at the International Livestock Research Institute (ILRI) Tick Unit. These were reared and managed as described by Bailey [21] and Irvin and Brocklesby [22]. With the exception of a *Hyalomma* sp. whose identity was uncertain until recently, the history and identity of all other tick species kept at the unit were well documented (Table 1). The laboratory maintained colonies ticks consisted of *Ambylomma variegatum* (21), *Hyalomma* sp. (16), *R. appendiculatus* (40 (Muguga colony) and 9 (Kiambu colony)), *Rhipicephalus* (*Boophilus*) *decoloratus* (22), *Rhipicephalus* (*Boophilus*) *microplus* (21) and *Rhipicephalus evertsi evertsi* (27). Ticks were transferred to 70 % ethanol and shipped at room temperature to Basel for the MALDI-TOF MS analysis and genomic DNA extraction.

Field ticks and identification by morphological characteristics

Two hundred forty two field ticks were either plucked directly from animal hosts (cattle/sheep) or were collected from pastures/vegetation. One hundred sixty of these ticks were collected 2014 at various sites in Kenya (Fig. 1). The remaining 82 specimens were collected from multiple localities within East Africa during a study investigating the population structure of *R. appendiculatus* in the field [23] (Fig. 1). Ticks were assigned to sex and species as described by Hoogstraal (1956) and Walker (2000 and 2003) [4, 5] and stored in 70 % ethanol and kept at 4 °C prior to shipping to Basel. Due to the high proportion (242/398) of field ticks and the inclusion of both male and female specimens, we expected our collection to reliably reflect intra-species physiological and molecular diversity.

Molecular COI gene typing

DNA extraction & PCR

To confirm the morphologically assigned species identities of field ticks and to check for potential molecular differences to laboratory ticks, specimens of both origins were subjected to molecular analysis. A tick was randomly chosen from the library and thoroughly rinsed with distilled water in order to remove any ethanol residues. The tick legs were detached with a scalpel and stored in 70 % ethanol for later MALDI-TOF MS analysis. The tick body was transferred to a 1.5 ml microcentrifuge tube and placed in liquid nitrogen for 5 min. Using a polypropylene pestle (Sigma-Aldrich), the frozen tick body was thoroughly grinded to powder. Whole genomic DNA was extracted using the QIAGEN® *DNeasy® Blood & Tissue Kit* (Qiagen GmbH, Germany). One hundred eighty microliter buffer ATL and 20 μl proteinase K were added to the grinded tick body and the mixture incubated overnight at 56 °C to ensure complete lysis of the tissue. The further extraction steps were carried out according to the manufacturers' protocol. The final concentration of extracted gDNA was determined with a spectrophotometer (WPA Lightwave II, Biochrom). Cytochrome *c* oxidase subunit I (COI) gene sequences of individual ticks were obtained by PCR amplification using the forward primer LCO149021 (5′-GGTCAACAAATCATA AAGATATTGG-3′) and reverse primer HC02198 (5′ TA AACTTCAGGGTGACCAAAAAATCA-3′) [24]. PCR was performed in a 50 μl reaction consisting of 5 μl 10× PCR Buffer (containing 10 mM $MgCl_2$), 1 μl dNTP mix, 1 μl $MgCl_2$, 2.5U HotStarTaq DNA Polymerase (Qiagen GmbH, Germany), 1 μl each of forward and reverse primers and 25 to 500 ng of tick gDNA as a template. The final volume of the reaction mixture was made up to 50 μl with nuclease-free water (Thermo Scientific, Germany). The PCR conditions consisted of an initial heat activation step at 95 °C for 15 min followed by 35 cycles of denaturation at 94 °C for 1 min, annealing at 45 °C for 1 min and extension at 72 °C for 1 min. The final extension was performed for 10 min at 72 °C. Per run, between 5 to 15 reactions were amplified, including two reactions where the gDNA template was omitted and compensated with nuclease-free water, serving as negative controls. The quality of the PCR products was determined by running 5 μl of stained (DNA-Dye NonTox, PanReac/AppliChem) DNA on a 1.0 % agarose gel. The bands were visualized and examined with the GelDoc™ EZ Imager (BioRAD). The amplified COI products were purified using the QIAquick PCR Purification Kit (Qiagen GmbH, Germany) following the manufacturers' protocol. DNA samples were eluted with 40 μl elution buffer (10 mM TrisCl). The COI gene was sequenced using the gene specific forward and reverse primer pair used for PCR amplification at Microsynth AG, Switzerland.

Table 1 Overview of 398 ticks that built the basis of this study

Morphologically assigned species name	Quantity	Sex	Geographical origin	Source
Amblyomma gemma	21	4 F, 15 M, 2 ND	Kenya	Vegetation & Animal
Amblyomma hebraeum	4	2 F, 1 M, 1 ND	Zimbabwe	Vegetation
Amblyomma variegatum	21	4 F, 15 M, 2 ND	Kenya	Lab colony
Amblyomma variegatum	19	5 F, 14 M	Kenya	Vegetation & Animal
Hyalomma anatolicum anatolicum	13	10 F, 1 M, 2 ND	Sudan	Vegetation
Hyalomma dromedarii	3	2 M, 1 ND	Kenya	Vegetation & Animal
Hyalomma marginatum rufipes	18	5 F, 11 M, 2 ND	Kenya	Vegetation & Animal
Hyalomma truncatum	14	5 F, 6 M, 3 ND	Kenya	Vegetation & Animal
Hyalomma sp.	16	8 F, 6 M, 2 ND	Kenya	Lab colony
Rhipicephalus (Boophilus) decoloratus	22	21 F, 1 M	Kenya	Lab colony
Rhipicephalus (Boophilus) decoloratus	19	19 F	Kenya & Sudan	Animal
Rhipicephalus (Boophilus) microplus	21	21 F	Kenya	Lab colony
Rhipicephalus (Boophilus) microplus	3	3 F	Kenya	Animal
Rhipicephalus appendiculatus	40	16 F, 23 M, 1 ND	Kenya	Lab colony (Muguga)
Rhipicephalus appendiculatus	38	15 F, 22 M, 1 ND	Kenya	Vegetation & Animal
Rhipicephalus appendiculatus	9	8 F, 1 ND	Kenya	Lab colony (Kiambu)
Rhipicephalus evertsi evertsi	27	10 F, 16 M, 1 ND	Kenya	Lab colony
Rhipicephalus evertsi evertsi	28	7 F, 20 M, 1 ND	Kenya	Vegetation & Animal
Rhipicephalus praetextatus	8	5 F, 2 M, 1 ND	Kenya	Vegetation & Animal
Rhipicephalus pulchellus	37	21 F, 14 M, 2 ND	Kenya	Vegetation & Animal
Rhipicephalus simus	17	8 F, 7 M, 2 ND	Kenya	Vegetation & Animal
Total	398			

ND sex not determined, *M* male, *F* female

Data analysis: sequence editing and multiple alignments

COI sequence chromatograms were visually inspected and manually edited using Seqtrace [25]. Using the Molecular Evolutionary Genetic Analysis (MEGA) software version 6.0 [26], consensus sequences were generated from a forward and reverse sequence for each of the COI PCR products. Species identity was confirmed by matching of the consensus sequences with reference data deposited in the NCBI GenBank [27] and/or the BOLD database [28], a barcoding database that is a component of the Tree of Life project and contains only COI nucleotide sequences. A positive match with a GenBank record was defined as more than 95 % query coverage and ≥ 97 % identity. A positive match with a record on BOLD was declared at identity values ≥ 97 %. Multiple sequence alignment analysis of all consensus sequences was performed using the MUSCLE tool in MEGA. The nucleotide sequences were trimmed to around 680 bp and the phylogenetic analyses computed based on maximum likelihood algorithm and the tree file exported to FigTree [29] for final editing.

MALDI-TOF MS analysis of ticks

Sample preparation

The sample processing protocol has been adopted from previous studies [15, 30] and modified accordingly. Specimens were removed from the library, rinsed once with distilled water and dried on absorbent paper. Depending on the size of the tick, two to eight legs were detached with a scalpel and placed in a 1.5 ml microcentrifuge tube containing 10 μl of 25 % formic acid. The samples were homogenized using a stainless steel micropestle (LLG Labware, Switzerland) powered by a portable drilling machine for 30 s. The homogenate was then centrifuged at 10,000 rpm for 1 min and 1 μl of the supernatant transferred into a microcentrifuge tube containing 8 μl of matrix solution (saturated sinapinic acid, 60 % acetonitrile, 40 % high-performance liquid chromatography (HPLC)-grade water and 0.3 % trifluoroacetic acid). After thoroughly mixing, the solution was spotted in quadruplicates (1 μl each) on a steel target plate (Mabritec AG, Basel, Switzerland). The spots were allowed to dry for several minutes until crystallization of the matrix/analyte mixture was complete

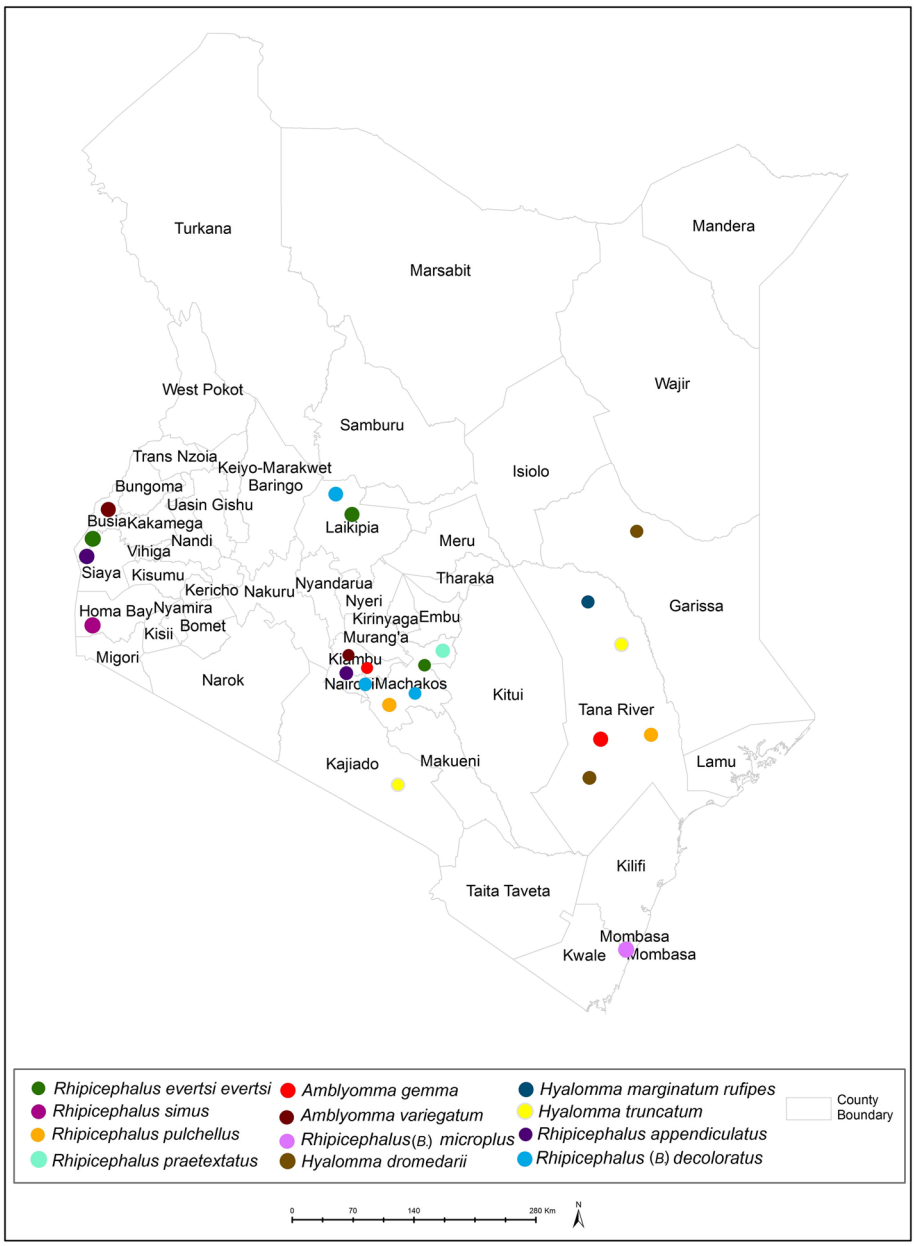


Fig. 1 Geographical origin of ticks used for MALDI-TOF MS analysis in Kenya. The collection sites of the ambiguous *Hyalomma* species and the specimens obtained from outside Kenya are not shown

and the target plate thereafter transferred to the MALDI-TOF MS instrument.

MALDI-TOF parameters

The MS measurements were carried out using a MALDI-TOF Mass Spectrometry Axima™ Confidence machine (Shimadzu-Biotech Corp., Kyoto, Japan) with detection in the linear positive mode, allowing the interrogation of high molecular weight samples. The acceleration voltage was set by default to 20 kV with an extraction delay time of 200 ns and a laser frequency of 50 Hz. The analysis was carried out in the mass range between 4000 and 20,000 Da. To ensure an even measurement covering the entire area of the sample spot, a netlike pattern of 100 equally distributed locations was defined. At 50 of these profiles, 10 consecutive laser shots were applied, adding up to 500 laser shots per sample spot. The ion gate was set at 3900 Da and the pulsed extraction optimized at 12,000 Da. The generated raw spectra were processed with the Launchpad™ version 2.9 software (Shimadzu-Biotech Corp., Kyoto, Japan) using the following settings: the advanced scenario was

chosen from the parent peak cleanup menu, peak width was set to 80 channels, smoothing filter width to 50 channels, baseline filter width to 500 channels and the threshold apex was chosen as the peak detection method. The threshold apex peak detection was set as a dynamic type and the offset was set to 0.020 mV with a response factor of 1.2. The processed spectra were exported as peak lists with m/z values for each peak and signal intensity in the ASCII format. Each target plate was externally calibrated using the reference spectra of *Escherichia coli* strain DH5α.

Spectral analysis: superspectrum design & validation

The generated mass spectra were exported to Launchpad™, quality-checked by eye and repeat measurements carried out if necessary. Mass spectra of reviewed spectra were then transferred in ASCII format to the spectral archive and microbial identification system (SARAMIS™) (AnagnosTec, Potsdam-Golm, Germany). A biomarker mass pattern, called superspectrum (SSp.) was calculated for each tick species using the SARAMIS™ SuperSpectra™ tool. To that end, the quadruplicate mass lists of four ticks were consolidated, peaks with a relative intensity below 1 % removed and average masses calculated with an error of 800 ppm. Masses of high species specificity were determined by comparison between the different tick SSp. and weighted manually.

In the validation step, using the SARAMIS™ identification tool, quadruplicate mass spectra of the remaining ticks were matched against the previously designed reference superspectra. A match between a SSp. and acquired mass spectra was regarded as positive at 75 % identity or higher. Accordingly, each mass spectrum achieved either a single match, sharing ≥75 % identity with only one SSp., a multiple match if sharing ≥ 75 % identity with more than one SSp., or no match if the 75 % identity threshold with no SSp. was reached. In case of a multiple match, the SSp. achieving the highest identity score was assumed the valid match. Subsequently, a given tick was assigned a final ID (i.e. species identification) when two criteria were met: (1) at least one of the four mass spectra matched to a SSp. and (2) assigned matches amongst the four mass spectra were not in conflict with each other.

Ethical statement

ILRI's Institutional Animal Care and Use Committee (IACUC) governed the use of cattle and rabbits for the maintenance of the lab tick colonies (approval no. 2010.1). The collection of field ticks did not involve endangered or protected species.

Results

Morphological identification of field ticks

Morphological identification grouped the ticks collected from vegetation and animals into 14 different species (Table 1). While five of these species were already represented by the laboratory colonies, nine species were exclusively covered by field ticks only. These species included *Amblyomma gemma* (21), *Amblyomma hebraeum* (4), *Hyalomma anatolicum anatolicum* (13), *Hyalomma dromedarii* (3), *Hyalomma marginatum rufipes* (18), *Hyalomma truncatum* (14), *R. praetextatus* (8), *Rhipicephalus pulchellus* (37) and *R. simus* (17).

Laboratory reared and field ticks combined, the 398 ticks grouped into 14 species within three genera, namely *Amblyomma* (3), *Hyalomma* (4) and *Rhipicephalus* (5)/ *Rhipicephalus (Boophilus)* (2). 51.3 % (204/398) of the ticks belonged to the genus *Rhipicephalus*, 16.3 % (65/398) to *Rhipicephalus* (*Boophilus*), 16.3 % (65/398) to *Amblyomma* and 16.1 % (64/398) to *Hyalomma*. Six tick species including *R. appendiculatus* (87), *R. evertsi evertsi* (55), *R.* (*B.*) *decoloratus* (41), *A. variegatum* (40) and *R. pulchellus* (37) represented 65.3 % of the collection (Table 1).

COI gene sequencing

For a total of 33 ticks, COI gene sequences were obtained (Table 2). No unspecific amplification occurred for the negative controls included in each PCR amplification run. The sequenced amplicons - with the exception of three ticks where no consensus sequence could be determined - were all approximately 700 bp in size, and the sequences were used for comparison against entries at GenBank and/or BOLD. Since there are no COI gene sequences for *A. gemma* currently deposited in GenBank, specimen identity for these ticks was assigned solely based on the BOLD entries. Two ticks morphologically identified as *A. hebraeum* matched clearly with reference records of *A. gemma* in the BOLD database (identity scores of 100 and 99.70 %). At the same time, the identity shared with NCBI reference sequences for *A. hebraeum* was only 89 %. The whole group of four specimens morphologically determined as *A. hebraeum*, were henceforth assigned as *A. gemma*. The morphologically unidentified *Hyalomma* sp. was clearly determined as *H. dromedarii*, with four COI sequenced ticks matching with high scores to the respective reference records in both databases. The species designation was adopted accordingly for the further course of this study. Two members of the morphologically identified *R. simus* ticks both matched with the *R. simus* reference sequence (AF132840.1) present in NCBI with a score of 92 % (data not shown) and with a slightly higher score (94.5 %) to *R. praetextatus* in the BOLD database. Higher molecular identity (99 % with only 67 % query cover) was achieved

Table 2 Tabular overview of 33 ticks additionally identified by COI molecular typing

			Morphological identification	BOLD identification	GenBank identification	
Tick ID	COI gene length [bp]	Origin	Species ID	Species ID (Identity)	Species ID (Accession Nr.)	Identity
154	711	Field	*A. gemma*	*A. gemma* (99.80 %)	*no reliabe ID*	
183	688	Field	*A. gemma*	*A. gemma* (100 %)	*no reliabe ID*	
32	687	Field	*A. hebraeum*	*A. gemma* (100 %)	*no reliabe ID*	
109	686	Field	*A. hebraeum*	*A. gemma* (99.70 %)	*no reliabe ID*	
36	692	Lab	*A. variegatum*	*A. variegatum* (100 %)	*A. variegatum* (GU062743.1)	97 %
86	651[a]	Field	*A. variegatum*	*A. variegatum* (99.70 %)	*A. variegatum* (GU062743.1)	99 %
242	702	Lab	*A. variegatum*	*A. variegatum* (100 %)	*A. variegatum* (GU062743.1)	97 %
27	688	Field	*H. dromedarii*	*H. dromedarii* (100 %)	*H. dromedarii* (AJ437071.1)	99 %
74	688	Field	*H. dromedarii*	*H. dromedarii* (100 %)	*H. dromedarii* (AJ437061.1)	99 %
118	680	Field	*H. dromedarii*	*H. dromedarii* (100 %)	*H. dromedarii* (AJ437071.1)	99 %
34	688	Lab	*H. sp.*	*H. dromedarii* (100 %)	*H. dromedarii* (AJ437061.1)	99 %
112	686	Lab	*H. sp.*	*H. dromedarii* (100 %)	*H. dromedarii* (AJ437061.1)	99 %
194	680	Lab	*H. sp.*	*H. dromedarii* (100 %)	*H. dromedarii* (AJ437061.1)	99 %
207	686	Lab	*H. sp.*	*H. dromedarii* (100 %)	*H. dromedarii* (AJ437061.1)	99 %
139	689	Field	*H. m. rufipes*	*H. m. rufipes* (99.84 %)	*H. m. rufipes* (AJ437100.1)	99 %
					H. truncatum (AJ437088.1)	99 %
359	688	Field	*H. m. rufipes*	*H. m. rufipes* (99.12 %)	*H. m. rufipes* (AJ437095.1)	99 %
72	684	Field	*H. truncatum*	*H. truncatum* (99 %)	*H. truncatum* (AJ437084.1)	97 %
361	555[a]	Field	*H. truncatum*	*H. truncatum* (98 %)	*H. truncatum* (AJ437084.1)	97 %
38	693	Lab (Kiambu)	*R. appendiculatus*	*R. appendiculatus* (99.50 %)	*R. appendiculatus* (AF132833.1)	97 %
121	687	Lab (Muguga)	*R. appendiculatus*	*R. appendiculatus* (99.50 %)	*R. appendiculatus* (AF132833.1)	98 %
198	679	Lab (Muguga)	*R. appendiculatus*	*R. appendiculatus* (99.50 %)	*R. appendiculatus* (AF132833.1)	98 %
225	687	Field	*R. appendiculatus*	*R. appendiculatus* (99.85 %)	*R. appendiculatus* (AF132833.1)	99 %
146	673	Field	*R. (B.) decoloratus*	*R. (B.) decoloratus* (100 %)	*R. (B.) decoloratus* (AF132826.1)	99 %
278	690	Lab	*R. (B.) decoloratus*	*R. (B.) decoloratus* (99.85 %)	*R. (B.) decoloratus* (AF132826.1)	99 %
165	585[a]	Lab	*R. (B.) microplus*	*R. (B.) microplus* (100 %)	*R. (B.) microplus* (KC503261.1)	100 %
377	689	Field	*R. (B.) microplus*	*R. (B.) microplus* (100 %)	*R. (B.) microplus* (KC503261.1)	99 %
170	694	Field	*R. evertsi evertsi*	*R. evertsi evertsi* (100 %)	*R. evertsi evertsi* (AF132835.1)	98 %
275	688	Lab	*R. evertsi evertsi*	*R. evertsi evertsi* (100 %)	*R. evertsi evertsi* (AF132835.1)	98 %
370	688	Field	*R. praetextatus*	*no reliabe ID*	*no reliabe ID*	
396	702	Field	*R. praetextatus*	*no reliabe ID*	*no reliabe ID*	
397	702	Field	*R. pulchellus*	*R. pulchellus* (98 %)	*R. pulchellus* (AY008682.1)	99 %
29	701	Field	*R. simus*	*no reliabe ID*	*no reliabe ID*	
115	690	Field	*R. simus*	*no reliabe ID*	*no reliabe ID*	

[a]no consensus sequence
Marked in bold: Ticks later used for SSp. design
No reliable ID: Identity with top match < 97 %

with deposited partial COI gene sequences (472 bp) of *Rhipicephalus muhsamae*. The two ticks morphologically identified as *R. praetextatus* matched with a relatively high identity of 96 % (data not shown) to a *R. simus* record on NCBI and with 100 % identity to unpublished *R. praetextatus* records on BOLD. Given these uncertain molecular results and the limited reference records available, the specimens of the *R. simus* and *R. praetextatus* batch were merged to one group and the species designation changed to *R. simus* group. Extraction of DNA qualitatively sufficient for COI gene sequencing failed with all specimens of *H. anatolicum anatolicum*. The morphologically assigned species identity of these ticks could therefore not be confirmed on molecular basis. The COI gene sequence of a tick (ID 139, Table 2) morphologically identified as *H. marginatum*

rufipes was identified with equal score as *H. marginatum rufipes* and *H. truncatum* on GenBank. Since the same COI sequence matched highest to *H. marginatum rufipes* record on the BOLD database, the species identity as initially determined on a morphological basis was assumed correct. Taking together the molecular results, all ticks morphologically identified as *A. hebraeum, Hyalomma* sp., *R. simus* and *R. praetextatus* were reclassified to *A. gemma*, *H. dromedarii* and *R. simus* group, respectively. The species identity of *H. anatolicum anatolicum* could not be confirmed due to insufficient quality of genomic DNA. All the remaining tick species that were assigned morphologically were confirmed by our molecular analysis.

MALDI-TOF MS analysis

Spectra quality

A total of 1592 single mass spectra were generated, corresponding to 398 ticks measured in quadruplicate. Inadequate spectral quality was assessed visually and accordingly 55 ticks were re-measured and integrated into the sample set. Seventeen ticks were excluded from the study after the first MALDI-TOF MS measurement since the poor overall state of the specimens did not allow for the generation of qualitatively adequate mass spectra. Among the excluded ticks were four ticks of the *R. simus* group batch that were partially overgrown by fungus. The entire collection (13 specimens) of *H. anatolicum anatolicum* for which also preceding DNA extraction had failed, were stored in leaky microtubes and as a result completely desiccated. The remaining 1524 mass spectra (381 ticks) used for the subsequent analyses presented a good signal-to-noise ratio and clear protein peaks, mostly distributed between 4000 to 13,000 Da (Additional file 1: Figure S1). The number of protein peaks per spectrum ranged from 14 to 145, with an average peak count of 42.5.

Intra species reproducibility of mass spectra

Visual inspection of spectral profiles revealed a generally highly similar mass fingerprint shared between individuals of the same species. Simultaneously, mass spectra were still heterogeneous on species level with missing or exclusive mass peaks only present in certain specimens or spectral profiles (Fig. 2). However, comparative analysis of the spectral profiles with SARAMIS™ confirmed that few major protein peaks remained highly conserved within a given species (Fig. 2). A significant difference between the mass spectra of male and female ticks of the same species could not be observed. This is in line with what has been reported in other studies [31, 32].

Inter species specificity of mass spectra

To assess the interspecies specificity of the mass spectra generated by MALDI-TOF MS, the spectral profiles of the previous molecularly identified 33 ticks were subjected to cluster analysis (Fig. 3). As expected, the spectra derived from the same tick e.g. the technical replicates each clustered together in closest proximity. This was not the case for the mass profiles of just one tick (specimen no. 029). Importantly, within this set of ticks, all spectra derived from specimens of the same species seemed to share distinct masses that separated them clearly from the remaining tick species.

Definition of superspectra identifying East African tick species

After COI molecular and MALDI-TOF MS analysis, our specimen collection was slightly reduced from 398 to 381 ticks, now grouping into ten different species and the ambiguous *R. simus* group. Incorporating these results, SSp. were designed from a total of 48 ticks (Table 3). The mass profiles derived from the Kiambu laboratory tick strain that has been maintained for many years at the ILRI tick unit showed consistently high deviations from the other *R. appendiculatus* profiles (indicated in Fig. 3). This lead us to define a distinct SSp. designated as *R. appendiculatus* II exclusively covering this batch of ticks. The final 12 SSp. designed in this study were based on 192 mass spectra of 48 individual ticks and consisted of 14 to 30 individual protein masses (Table 3). In addition to including 24 COI typed specimens to the SSp. design, we also incorporated ticks representative of the diversity within a given species. This led to the inclusion of both field and laboratory ticks in some cases and to the inclusion of ticks with distinct mass patterns in other cases.

Validation of defined superspectra for tick identification

After removal of the 48 ticks used to build the reference spectra, 333 ticks remained for the validation step of the generated SSp. Our approach failed to assign an ID to 13 specimens, 319 ticks were correctly identified and only one tick was assigned a wrong ID. This corresponded to an overall sensitivity of 96.1 % and a specificity of 99.7 % (Table 4). Among the correctly identified ticks, 182 or 57.1 % matched with the correct SSp. in all four acquired mass spectra (indicated as 4× CC in Table 4). Sixty-six (20.7 %) ticks matched with three mass spectra to the correct SSp. while one mass spectrum resulted in no match. Twenty-three (7.2 %) ticks matched with two mass spectra to the correct SSp. while two mass spectra achieved no match. Nineteen (6.0 %) ticks matched with three mass spectra to the correct SSp. while one mass spectrum reached multiple matches, with the correct SSp. as the top match. The remaining 29 (9.1 %) ticks were positively identified, with their mass spectra matching in other combinations (indicated as "other" in Table 4).

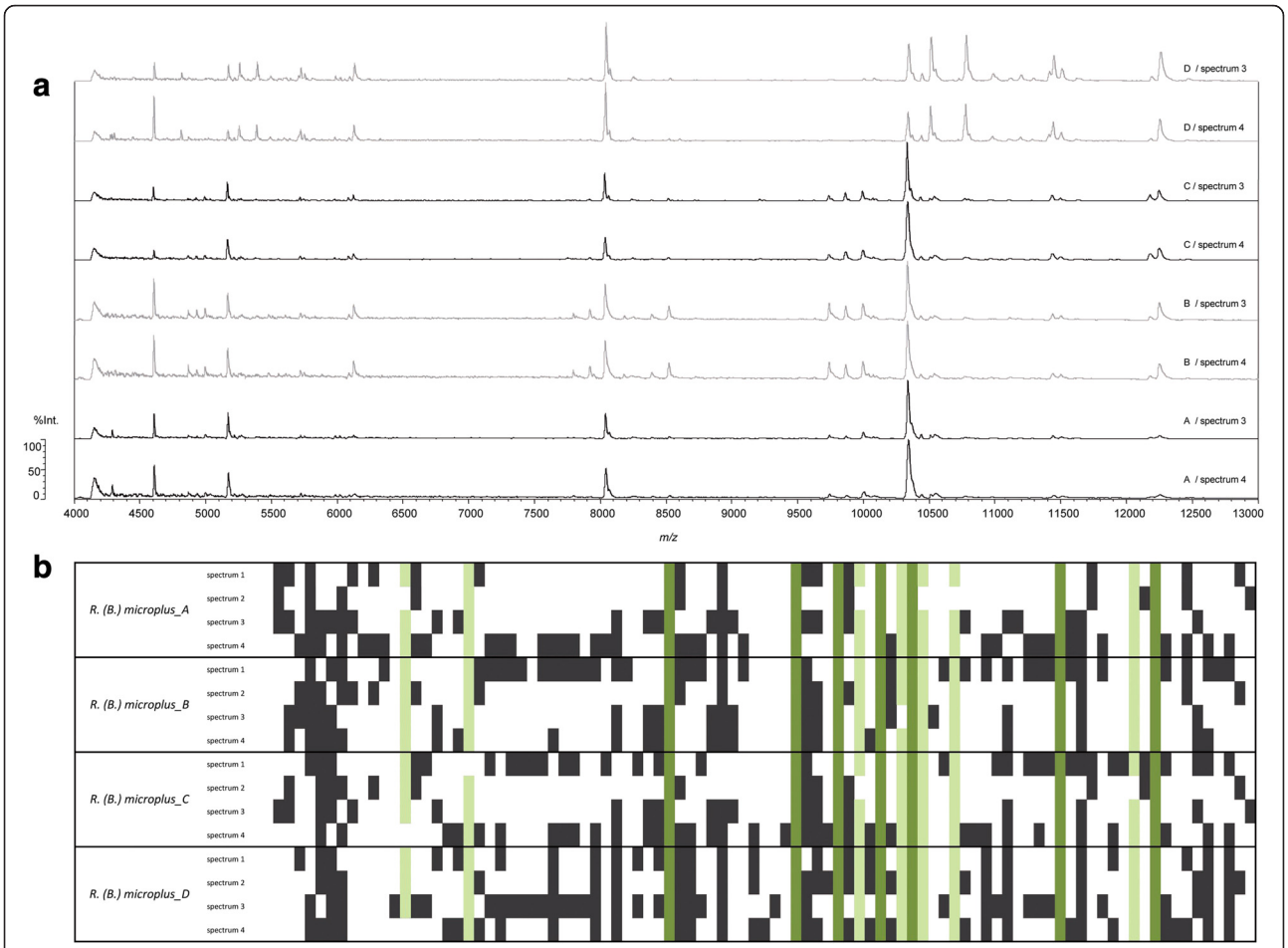


Fig. 2 Comparison of *R.* (*B*) *microplus* mass profiles to assess intraspecies reproducibility. **a** Visual comparison of spectral profiles derived from four specimens (A–D) indicates both significant similarity as well as individual differences in mass fingerprints. **b** Comparative analysis of 16 mass spectra, corresponding to ticks A–D measured in quadruplicates (spectrum 1–4). Vertically arranged are the total 93 different masses found amongst the 16 spectra. A *white field* indicates absence, a *black* or *green field* presence of the given mass in the respective spectrum. Highly conserved protein masses are marked in *dark green* (100 % abundance) or *light green* (abundant in at least 3 out of 4 technical replicates in all specimens)

Among the successfully identified ticks, the sensitivity of our approach was lowest for *A. gemma* (81 %) where a total of 21 ticks were used for validation and the *R. appendiculatus* ticks derived from the Kiambu stock (80 %), where only five ticks were used for validation. The sole tick that was assigned a false species ID, was a specimen morphologically identified as *R. evertsi evertsi.* While three of this tick's mass spectra did not achieve a match at all, one mass spectrum was marginally similar (78 %, data not shown) to the *R. simus* group SSp. The mass spectra of the 13 ticks that could not be assigned to any SSp. and the spectrum of the wrongly assigned tick were inspected visually to assess the spectral quality. It appeared that most spectra displayed alterations like distorted or shifted mass peaks. The species identity of the wrongly identified *R. evertsi evertsi* specimen could not be verified on a molecular basis, since the extracted gDNA was not qualitatively sufficient for PCR. Three ticks with no SSp. ID (1× *H. dromedarii* and 2× *A. gemma*) were subjected to molecular COI analysis. The morphologically assigned species identities of all three specimens (Fig. 4; tick no. 95, 60 and 30) were confirmed on molecular basis.

The 48 ticks initially used to build the reference SSp. were not considered for the study validation. These mass spectra were later experimentally validated against all SSp. (data not shown). All mass spectra of the ticks were, as one would expect, correctly identified with their corresponding SSp.

Discussion

Several genera of ixodid tick genera co-exist throughout Eastern Africa, including *Hyalomma*, *Amblyomma* and *Rhipicephalus*. Precise and timely data on tick population distribution and size in a given geographical area are required to model epidemiological trends of tick-

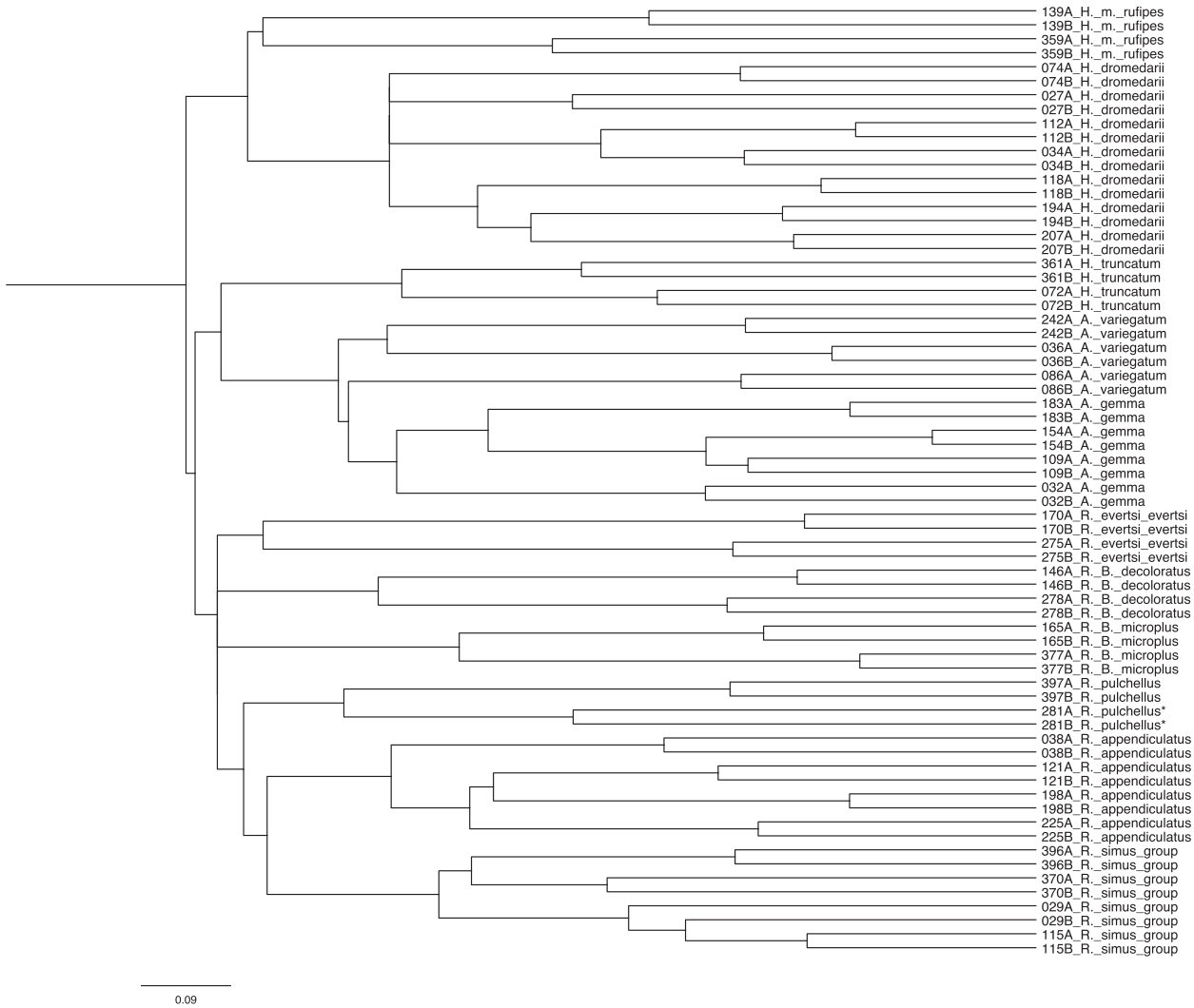


Fig. 3 Cladogram (neighbour joining algorithm) illustrating the inter species specificity of tick mass spectra. Spectral profiles of two technical replicates (A and B) of the 33 COI gene sequenced ticks were integrated to the analysis. Two spectra (*) of a solely morphologically identified *R. pulchellus* tick were added to the dataset to maintain a minimum number of two specimens per species

borne diseases and formulate effective control strategies [1, 33]. However, tick identification by morphology can be limited by a lack of expertise, the need of several male specimens, whereas immature tick stages are difficult to identify by morphology alone [6].

In this study, MALDI-TOF MS was used to investigate a collection of laboratory-bred and field-collected afrotropical ixodid ticks with the aim of confirming their identity and establishing a reference MS spectra index designated as SSp.

The quality of the spectra generated for the vast majority of the ticks included in this study corresponded to what has been observed in similar studies with a range of arthropod vectors including European tick species [15, 16], tsetse flies [14], mosquitoes [12] and midges [31].

We found that spectra quality, overall protein mass counts and the molecular weight range that can be determined mainly depend on the initial quality of the sample itself. Seventeen ticks that were improperly stored and overgrown by fungus, or that were completely desiccated needed to be removed from this study due to inadequate quality of mass spectra obtained. A less apparent factor negatively affecting the overall spectral quality seems to be long-term storage of tick specimens in ethanol [31]. This could have been a factor in the failure to correctly identify 14 ticks, where most mass profiles revealed alterations in spectral quality on close examination. Poor peak resolution, diffuse signals in the low molecular weight range, and a shift in peak patterns were the most common characteristics observed in spectra from the unidentified specimens. Additional

Table 3 Superspectra designed in this study

Name of SSp.	Condensed Mass Count	N (COI-typed)	Origin (N)
Amblyomma gemma	24	4 (2)	Field (4)
Amblyomma variegatum	30	4 (2)	Lab (3), Field (1)
Hyalomma dromedarii	29	4 (4)	Lab (4)
Hyalomma marginatum rufipes	24	4 (2)	Field (4)
Hyalomma truncatum	23	4 (2)	Field (4)
Rhipicephalus (Boophilus) decoloratus	14	4 (1)	Lab (2), Field (2)
Rhipicephalus (Boophilus) microplus	16	4 (2)	Lab (3), Field (1)
Rhipicephalus appendiculatus I	21	4 (3)	Lab Muguga (2), Field (2)
Rhipicephalus appendiculatus II	18	4 (1)	Lab Kiambu (4)
Rhipicephalus evertsi evertsi	19	4 (2)	Lab (2), Field (2)
Rhipicephalus pulchellus	26	4 (1)	Field (4)
Rhipicephalus simus group	18	4 (2)	Field (4)
Total		48 (24)	Lab (20), Field (28)

MALDI-TOF MS analysis with ticks plucked directly from animals or collected from vegetation without prior storage in ethanol could support these assumptions and reveal if spectral quality and taxonomic resolution can be enhanced significantly using freshly collected ticks.

In the majority of cases with samples sufficiently conserved, identification of ticks by matching their mass profiles against reference SSp. proved to be very robust. This was demonstrated by the high sensitivity (96.1 %) with which tick species were identified successfully. This is a significant achievement considering the large tick collection size, with some of the species represented by specimens originating from very different laboratory or field environments.

Together with COI gene sequences, a number of interesting conclusions can be inferred from the present study. The known problem of morphological tick species misidentification can be exemplified by two of our findings. (a) COI-typing of two ticks morphologically identified as *A. hebraeum*, revealed that the specimens were in fact members of the closely related *A. gemma*. This was for the most part resolved by our MALDI-TOF MS analyses, where three of the tick samples were identified clearly as *A. gemma*. The fourth tick, although showing high similarity with the *A. gemma* SSp., displayed spectral alterations and was not assigned any ID. (b) Similarly, a batch of ticks included into our collection clearly

Table 4 Validation of SSp. with 333 ticks

Tick species name	N	True ID assigned[a]					No ID assigned[a]	Wrong ID assigned[a]	Sensitivity	Specificity
		4× CC	3× CC 1× N	2× CC 2× N	3× CC 1× C	other				
Amblyomma gemma	21	4	4	5	0	4	4	0	81.00 %	100.00 %
Amblyomma variegatum	36	16	9	4	4	3	0	0	100.00 %	100.00 %
Hyalomma dromedarii	15	11	1	0	1	0	2	0	86.70 %	100.00 %
Hyalomma marginatum rufipes	14	4	4	2	0	2	2	0	85.70 %	100.00 %
Hyalomma truncatum	10	2	4	2	1	0	1	0	90.00 %	100.00 %
Rhipicephalus (Boophilus) decoloratus	37	25	7	1	2	2	0	0	100.00 %	100.00 %
Rhipicephalus (Boophilus) microplus	20	6	8	0	3	3	0	0	100.00 %	100.00 %
Rhipicephalus appendiculatus	74	53	13	3	2	2	1	0	98.60 %	100.00 %
Rhipicephalus appendiculatus (Kiambu)	5	4	0	0	0	0	1	0	80.00 %	100.00 %
Rhipicephalus evertsi evertsi	51	34	9	1	4	1	1	1	98.00 %	98.00 %
Rhipicephalus pulchellus	33	15	5	4	2	7	0	0	100.00 %	100.00 %
Rhipicephalus simus group	17	8	2	1	0	5	1	0	94.10 %	100.00 %
		182	66	23	19	29				
Total	333	319					13	1	96.10 %	99.70 %

[a]For each tick, four technical replicate mass spectra were matched against designed SSp. and a final ID assigned accordingly
CC: one correct SSp. matching; C: multiple SSp. matching, correct SSp. as top match; N: no matching SSp
other: true ID was assigned based on a different combination

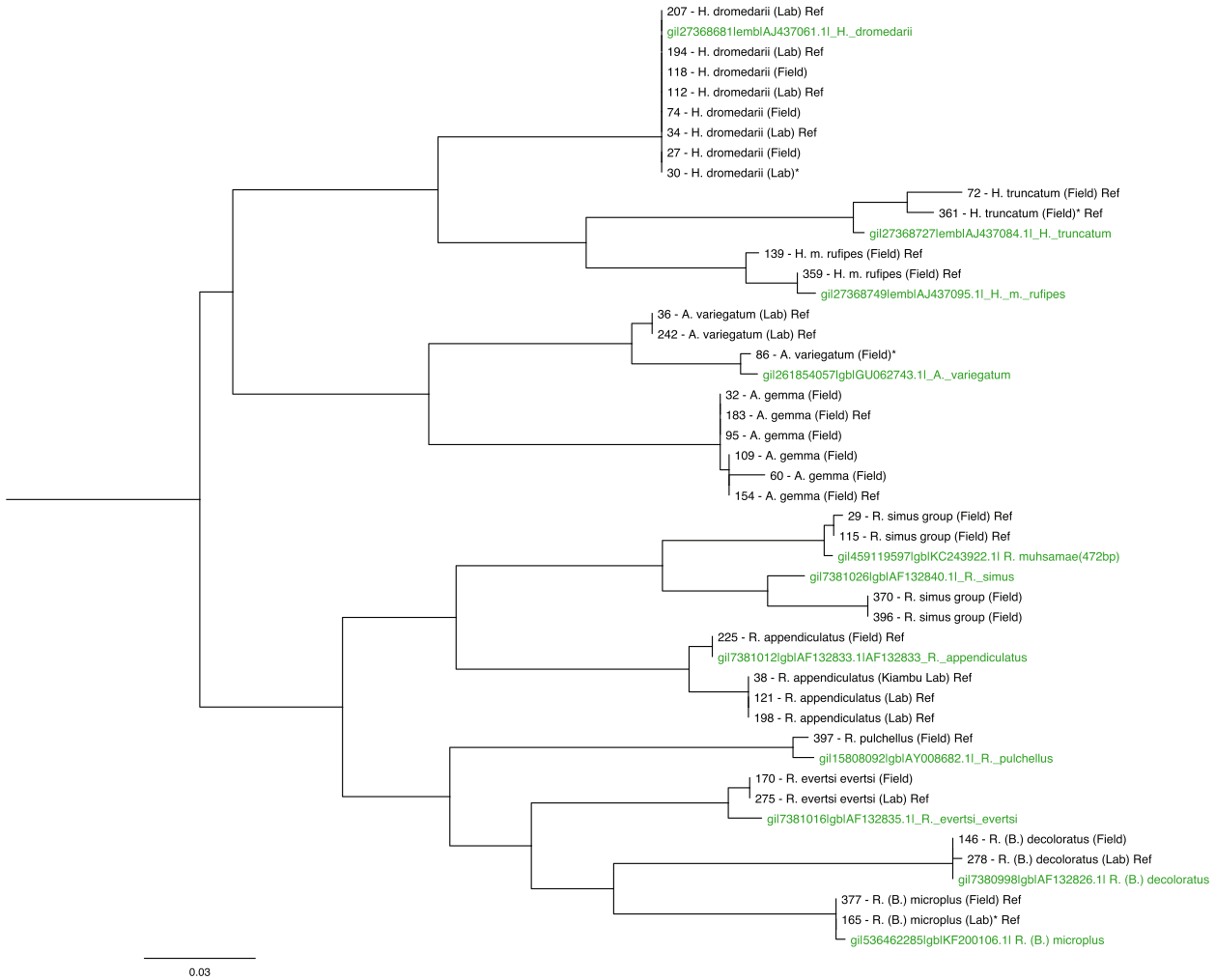


Fig. 4 Phylogenetic relationship of 11 reference (NCBI) and 36 study ticks based on their COI gene sequences, illustrated as a maximum likelihood phylogenetic tree. Accentuated in green are top matching GenBank reference sequences. *Asterisks* (*) indicate non-consensus sequences. *Scale*: The bar length corresponds to 0.03 % (20 nt) difference in nucleotide sequence. *Ref*: Reference tick used to design SSp

belonged to the genus *Hyalomma*. However, the absence of any reference specimen did not allow reliable morphological species identification. COI gene sequencing and MALDI-TOF MS both convincingly identified these ticks as *H. dromedarii*. These examples confirm the value of MALDI-TOF MS for resolution of tick taxonomic ambiguities. MALDI-TOF MS can provide improved and fast discrimination, especially when morphological examination is insufficient for a clear species designation.

The limitations of conventional tick typing are not restricted to the morphological approach but can extend to molecular techniques such as sequencing of the mitochondrial COI gene. This issue has been highlighted by the example of genetic hybridization occurring amongst members of the genus *Hyalomma* as described by Rees et al. [34]. While individuals of the species *H. truncatum*, *H. dromedarii* and *H. marginatum rufipes* are well differentiated both morphologically and genetically, sexual reproduction between members of these species can occur, resulting in hybrid offspring. Such intermediate individuals (e.g. NCBI record AJ437088.1 in Table 2) still display the distinct paternal morphological features while possessing the maternally inherited mtDNA genotype. The use of COI sequencing on its own can therefore result in misclassification of such specimens. It will be the subject of further research to establish how the mass spectra of hybrid ticks differ from the parental protein fingerprint and to what extent MALDI-TOF MS can serve as monitoring tool for following the gene flow amongst different tick species.

One unresolved issue in African tick taxonomy was highlighted by our findings regarding the ticks

morphologically assigned as *R. praetextatus* and *R. simus*. The ongoing debate, as to which of these species is distributed where in East-Africa is based largely on the fact that they can not be easily separated morphologically [5]. Defining the accurate spread of *R. simus* and *R. praetextatus* is further complicated by the co-occurrence of other, highly similar species including *R. lunulatus* and *R. muhsamae* known to be present in the same East African habitats [5, 35]. The fact, that *R. simus* has been described to be restricted to Southern Africa [4, 5] suggested early on that our *R. simus* field isolates from Kenya were mistaken with a morphologically highly similar species. This was supported by the COI gene analysis, which grouped this ticks closer to *R. praetextatus* (BOLD, 94 % identity) and *R. muhsamae* (GenBank, 99 % identity and 67 % query cover) than to *R. simus* (GenBank, 92 % identity and 99 % query cover). The sequence data was equally unclear for our *R. praetextatus* specimens, with both high matches to unpublished records of *R. praetextatus* reference sequences on BOLD (100 % identity) and *R. simus* in GenBank (96 % identity, 98 % query cover). A phylogenetic maximum-likelihood analysis of all COI nucleotide sequences derived in this study and reference records from GenBank (Fig. 4), supports these inferences regarding *R. simus* and *R. praetextatus*. Although there appears to be a clear molecular boundary between the analyzed ticks with a suggested close proximity of the *R. simus* ticks (specimen no. 29 and 115) to *R. muhsamae*, the limited reference records available do not allow a conclusive answer regarding the true identity of our ticks. We therefore merged these ticks to one group defined as *R. simus* group, enveloping the tick species *R. simus*, *R. praetextatus* and *R. muhsamae*, as previously suggested by Walker, Keirans and Horak [5].

Whether MALDI-TOF MS analysis can distinguish between these three tick species, where current COI, 12S and ITS2 molecular data is non-conclusive (E. Kanduma and R. Bishop, unpublished data), requires further investigation with representative specimens from all three species. It is however worth noting that our phylogenetic cluster analysis of four specimens designated as members of the *R. simus* group indicated potential differences between the spectral profiles (Fig. 3). Further studies will be needed to conclusively confirm the value of MALDI-TOF MS in discriminating between the species of the *R. simus* group.

The Kiambu *R. appendiculatus* specimens, where only five ticks were available for validation, and the ticks belonging to *A. gemma* were detected with the lowest sensitivity by our SSp. approach. The failure in species ID assignment for these specimens might partially be explained by the negative effect of long-term storage in ethanol as discussed before. Additionally, we hypothesized that intraspecies genetic heterogeneity could be increased in these two sets of ticks, leading to stronger diversity in spectral fingerprints. Phylogenetic analysis of the COI gene sequences does not support this theory (Fig. 4). The *A. gemma* ticks among each other, as well as the *R. appendiculatus* ticks of both Muguga and Kiambu stock shared almost identical COI nucleotide sequences. Continuative studies, incorporating freshly plucked ticks, will help to determine to what extent the overall sensitivity of a SSp. based identification approach can still be improved.

Looking ahead, the potential applications of MALDI-TOF MS as a tick species typing tool are diverse, ranging from pathogen and vector epidemiological monitoring for disease outbreak detection, to following consequences of climate change and its influence on changing patterns of tick distribution and its associated disease risks as described [36]. Furthermore, the SSp. established here provide the basis to move towards simultaneous characterization of African tick vectors and pathogens transmitted by MALDI-TOF MS, as has been shown with *Rickettsia* [37]. Another immediate use of this technique would be monitoring the spread of the invasive single host tick *R.* (*B.*) *microplus* [38]. This tick is both more adaptable to changing environments than native species like *R.* (*B.*) *decoloratus* and has greater potency in transmission of protozoan and bacterial pathogens, including *Babesia bigemina, Babesia bovis* and *Anaplasma marginale* [39].

Conclusions

In summary, our study demonstrated the applicability of MALDI-TOF MS as a suitable tool for East African tick identification. The processing steps of the ticks for MALDI-TOF MS are straightforward, with little time and human and equipment resources needed. The rapid generation of mass spectra profiles and their automated, immediate comparison against pre-designed reference SSp. allow high-throughput measurement of large numbers of samples. We identified the quality of the samples used as the main limiting factor for the MALDI-TOF MS analyses. Whenever possible, tick material collected freshly from the field should be analyzed. The negative impact of sample storage under ethanol for limited periods of time should be evaluated carefully, since this would increase applicability to large tick collections sampled across Africa. Under good conditions of sample storage, MALDI-TOF MS can generate highly distinctive mass profile patterns that will allow precise and rapid monitoring of tick populations, species movements, pathogen transmission and host feeding preferences on a large-scale.

Additional file

Additional file 1: Figure S1. Comparison of MALDI-TOF MS spectral profiles, indicating distinct mass peak patterns among the different tick genera *Amblyomma*, *Hyalomma*, *Rhipicephalus* and *Rhipicephalus* (*Boophilus*). The spectra illustrated in the figure cover a mass range between 4000 to 18,000 Da. The relative peak intensities are indicated on the y-axis. (PDF 143 kb)

Abbreviations

BOLD: Barcode of Life Database; bp: base pair; COI: cytochrome *c* oxidase I; DNA: deoxyribonucleic acid; ILRI: International Livestock Research Institute; NCBI: National Center Bioinformatics; nt: nucleotide; MALDI: matrix-assisted laser desorption/ionization; MS: mass spectrometry; PCR: polymerase chain reaction; ppm: parts per million; SSp.: superspectrum; sp.: species; TOF: time-of-flight.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

EGK, SM and NG collected tick specimens and carried out the morphological identification of samples. VP supported the MALDI-TOF MS project design and data interpretation and JR conducted the MALDI-TOF MS analysis. NG and JR carried out the molecular genetic studies, participated in the sequence alignment and drafted the manuscript. CD, CO and RPB conceived of this study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Christoph Stalder (Swiss TPH, Basel, Switzerland) for technical advice in DNA extraction from ticks and Roxanne Mouchet (Mabritec AG, Riehen, Switzerland) for assistance in pre-processing and MALDI-TOF MS analysis of ticks. Further gratitude is expressed to Susan Njenga (ILRI Research Method Group, Nairobi, Kenya) for the assistance in generating the tick distribution map, Edward Kariuki of Kenya Wildlife Service (KWS) for valuable discussion on tick identification by morphology and finally the ILRI Kapiti ranch staff for assistance during field sampling. Esther Kanduma was supported financially by Biosciences eastern and central Africa (BecA), the German Academic Exchange Service (DAAD) and the African Women in Agricultural Research and Development (AWARD).

Author details

[1]International Livestock Research Institute (ILRI), PO Box 30709-00100, Nairobi, Kenya. [2]Department of Medical Parasitology and Infection Biology, Clinical Immunology Unit, Swiss Tropical and Public Health Institute (Swiss TPH), Socinstr. 57, CH 4002 Basel, Switzerland. [3]University of Basel, Petersplatz 1, CH 4003 Basel, Switzerland. [4]Mabritec SA, Lörracherstrasse 50, CH 4125 Riehen, Switzerland. [5]Department of Veterinary Microbiology and Pathology, Washington State University, PO Box 647040, Pullman, WA 99163, USA. [6]Biosciences eastern and central Africa – International Livestock Research Institute (BecA-ILRI) Hub, PO Box 3070900100 Nairobi, Kenya. [7]Department of Biochemistry, School of Medicine, University of Nairobi, PO Box 30197, Nairobi, Kenya.

Received: 16 December 2015 Accepted: 4 March 2016
Published online: 15 March 2016

References

1. Jongejan F, Uilenberg G. The global importance of ticks. Parasitology. 2004; 129 Suppl:S3–14.
2. Kivaria FM. Estimated direct economic costs associated with tick-borne diseases on cattle in Tanzania. Trop Anim Health Prod. 2006;38:291–9.
3. Estrada-Peña A, Farkas R, Jaenson TGT, Koenen F, Madder M, Pascucci I, Salman M, Tarrés-Call J, Jongejan F. Association of environmental traits with the geographic ranges of ticks (Acari: Ixodidae) of medical and veterinary importance in the western Palearctic. A digital data set. Exp Appl Acarol. 2013;59:351–66.
4. Walker AR, Bouattour A, Camicas JL, Estrada-Pena A, Horak IG, Latif A, Pegram RG, Preston PM. Ticks of domestic animals in Africa: a guide to identification of species. Edinburgh: Bioscience Reports; 2003.
5. Walker JB, Keirans JE, Horak IG. The Genus Rhipicephalus (Acari, Ixodidae): A Guide to the Brown Ticks of the World. Cambridge University Press; 2000
6. Guglielmone AA, Robbins RG, Apanaskevich DA, Petney TN, Estrada-Peña A, Horak IG. The hard ticks of the world. Dordrecht: Springer Netherlands; 2014.
7. Araya-Anchetta A, Busch JD, Scoles GA, Wagner DM. Thirty years of tick population genetics: a comprehensive review. Infect Genet Evol. 2015;29:164–79.
8. Kanduma EG, Mwacharo JM, Mwaura S, Njuguna JN, Nzuki I, Kinyanjui PW, Githaka N, Heyne H, Hanotte O, Skilton RA, Bishop RP. Multi-locus genotyping reveals absence of genetic structure in field populations of the brown ear tick (*Rhipicephalus appendiculatus*) in Kenya. Ticks Tick-Borne Dis. 2015;7:26-35.
9. Will KW, Rubinoff D. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics. 2004;20:47–55.
10. Shen Y-Y, Chen X, Murphy RW. Assessing DNA barcoding as a tool for species identification and data quality control. PLoS One. 2013;8:e57125.
11. DeSalle R, Egan MG, Siddall M. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philos Trans R Soc B Biol Sci. 2005;360:1905–16.
12. Yssouf A, Parola P, Lindström A, Lilja T, L'Ambert G, Bondesson U, Berenger J-M, Raoult D, Almeras L. Identification of European mosquito species by MALDI-TOF MS. Parasitol Res. 2014;113:2375–8.
13. Schaffner F, Kaufmann C, Pflüger V, Mathis A. Rapid protein profiling facilitates surveillance of invasive mosquito species. Parasit Vectors. 2014;7:142.
14. Hoppenheit A, Murugaiyan J, Bauer B, Steuber S, Clausen P-H, Roesler U. Identification of Tsetse (*Glossina* spp.) using matrix-assisted laser desorption/ ionisation time of flight mass spectrometry. PLoS Negl Trop Dis. 2013;7:e2305.
15. Yssouf A, Flaudrops C, Drali R, Kernif T, Socolovschi C, Berenger J-M, Raoult D, Parola P. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid identification of tick vectors. J Clin Microbiol. 2013;51:522–8.
16. Karger A, Kampen H, Bettin B, Dautel H, Ziller M, Hoffmann B, Süss J, Klaus C. Species determination and characterization of developmental stages of ticks by whole-animal matrix-assisted laser desorption/ionization mass spectrometry. Ticks Tick-Borne Dis. 2012;3:78–89.
17. Seng P, Drancourt M, Gouriet F, La Scola B, Fournier P-E, Rolain JM, Raoult D. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis Off Publ Infect Dis Soc Am. 2009;49:543–51.
18. Bishop RP, Hemmink JD, Morrison WI, Weir W, Toye PG, Sitt T, Spooner PR, Musoke AJ, Skilton RA, Odongo DO. The African buffalo parasite *Theileria* sp. (buffalo) can infect and immortalize cattle leukocytes and encodes divergent orthologues of Theileria parva antigen genes. Int J Parasitol Parasites Wildl. 2015;4:333–42.
19. Mans BJ, Pienaar R, Latif AA. A review of *Theileria* diagnostics and epidemiology. Int J Parasitol Parasites Wildl. 2015;4:104–18 [Including Articles from "International Congress on Parasites of Wildlife", Pp. 49–158].
20. Sitt T, Poole EJ, Ndambuki G, Mwaura S, Njoroge T, Omondi GP, Mutinda M, Mathenge J, Prettejohn G, Morrison WI, Toye P. Exposure of vaccinated and naive cattle to natural challenge from buffalo-derived *Theileria parva*. Int J Parasitol Parasites Wildl. 2015;4:244–51.
21. Bailey KP: Notes on the rearing of Rhipicephalus appendiculatus and their infection with *Theileria parva* for experimental transmission. Bull Epizoot Dis Afr. 1960:33–64.
22. Irvin AD, Brocklesby DW. Rearing and maintaining *Rhipicephalus appendiculatus* in the laboratory. Ji Nan Idoon Animat Technol. 1970;20:106-12.
23. Kanduma EG, Mwacharo JM, Sunter JD, Nzuki I, Mwaura S, Kinyanjui PW, Kibe M, Heyne H, Hanotte O, Skilton RA, Bishop RP. Micro- and minisatellite-expressed sequence tag (EST) markers discriminate between populations of *Rhipicephalus appendiculatus*. Ticks Tick-Borne Dis. 2012;3:128–36.
24. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 1994;3:294–9.
25. Stucky BJ. SeqTrace: a graphical tool for rapidly processing DNA sequencing chromatograms. J Biomol Tech JBT. 2012;23:90–3.
26. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30:2725–9.
27. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Mizrachi I, Ostell J, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M,

Sirotkin K, Souvorov A, Starchenko G, Tatusova TA, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2009;37(Database issue):D5–15.
28. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes. 2007;7:355–64.
29. Rambaut A. FigTree v1.4.2. Tree figure drawing tool. 2014. Available: http://tree.bio.ed.ac.uk/software/figtree/. Accessed Sept 2015.
30. Kaufmann C, Schaffner F, Ziegler D, Pflüger V, Mathis A. Identification of field-caught *Culicoides* biting midges using matrix-assisted laser desorption/ionization time of flight mass spectrometry. Parasitology. 2012;139:248–58.
31. Kaufmann C, Ziegler D, Schaffner F, Carpenter S, Pflüger V, Mathis A. Evaluation of matrix-assisted laser desorption/ionization time of flight mass spectrometry for characterization of *Culicoides* nubeculosus biting midges. Med Vet Entomol. 2011;25:32–8.
32. Campbell PM. Species differentiation of insects and other multicellular organisms using matrix-assisted laser desorption/ionization time of flight mass spectrometry protein profiling. Syst Entomol. 2005;30:186–90.
33. George JE, Pound JM, Davey RB. Chemical control of ticks on cattle and the resistance of these parasites to acaricides. Parasitology. 2004;129(Suppl):S353–66.
34. Rees DJ, Dioli M, Kirkendall LR. Molecules and morphology: evidence for cryptic hybridization in African *Hyalomma* (Acari: Ixodidae). Mol Phylogenet Evol. 2003;27:131–42.
35. Pegram RG, Walker JB, Clifford CM, Keirans JE. Comparison of populations of the *Rhipicephalus simus* group: *R. simus*, *R. praetextatus*, and *R. muhsamae* (Acari: Ixodidae). J Med Entomol. 1987;24:666–82.
36. Gray JS, Dautel H, Estrada-Peña A, Kahl O, Lindgren E. Effects of climate change on ticks and tick-borne diseases in Europe. Interdiscip Perspect Infect Dis. 2009;2009:e593232.
37. Yssouf A, Almeras L, Terras J, Socolovschi C, Raoult D, Parola P. Detection of Rickettsia spp in ticks by MALDI-TOF MS. PLoS Negl Trop Dis. 2015;9:e0003473.
38. Chevillon C, de Garine-Wichatitsky M, Barré N, Ducornez S, de Meeûs T. Understanding the genetic, demographical and/or ecological processes at play in invasions: lessons from the southern cattle tick *Rhipicephalus microplus* (Acari: Ixodidae). Exp Appl Acarol. 2013;59:203–18.
39. Jonsson NN, Bock RE, Jorgensen WK. Productivity and health effects of anaplasmosis and babesiosis on *Bos indicus* cattle and their crosses, and the effects of differing intensity of tick control in Australia. Vet Parasitol. 2008;155:1–9.