2013

Bioinformatic Analysis and in Vitro Expression of Malaria Parasite Translocon and Ribonuclease Binding-Like Rhoptry Genes

Moses Z. Timta
Cleveland State University

BIOINFORMATIC ANALYSIS AND IN VITRO EXPRESSION OF MALARIA PARASITE TRANSLOCON AND RIBONUCLEASE BINDING–LIKE RHOPTRY GENES

MOSES Z. TIMTA

Bachelors of Science in Medicine and Surgery

University of Maiduguri, Borno, Nigeria

January 1991

United States Medical Licensing Examinations

Educational Commission for Foreign Medical Graduates

Philadelphia, PA

October 2005

Submitted in partial fulfillment of the requirements for

The degree of

MASTER OF SCIENCE IN BIOLOGY

at the

CLEVELAND STATE UNIVERSITY

May 2013

We hereby approve this thesis of

MOSES Z. TIMTA

Candidate for the Master of Science in Biology degree for the

Department of Biological, Geological, and Environmental Sciences

and the CLEVELAND STATE UNIVERSITY

College of Graduate Studies

Tobili Y. Sam-Yellowe, Ph.D., MPH / Date
Thesis Committee Chairperson
Department of Biological, Geological, and Environmental Sciences

---

Girish C. Shukla, Ph.D. / Date
Department of Biological, Geological, and Environmental Sciences

---

Harry VanKeulen, Ph.D. / Date

Department of Biological, Geological, and Environmental Sciences

DEDICATION

This work is dedicated to my daughter for her patience and understanding during the long hours of absence while trying to bring this work to a fruitful end.

ACKNOWLEDGEMENT

My most sincere and grateful appreciation to my adviser, Dr. Tobili Y. Sam-Yellowe, who let me use her lab and gave me the necessary advice and tools that I needed while carrying out this research. I also thank the members of my committee, Dr. Girish S. Shukla, and Dr. Harry Van Kuelen, who both offered objective and helpful criticisms.

I also thank Raghav Yadavalli and Ruth Kebede and friends in the department who also offered their valuable help during the course of this research.

BIOINFORMATIC ANALYSIS AND IN VITRO EXPRESSION OF MALARIA PARASITE TRANSLOCON AND RIBONUCLEASE BINDING-LIKE RHOPTRY GENES

MOSES Z. TIMTA

ABSTRACT

Malaria caused by the parasite *Plasmodium*, still remains a significant public health problem worldwide, due to lack of a vaccine and emerging drug and insecticide resistance, among malaria parasites and mosquito vectors, respectively. Rhoptry proteins of *Plasmodium* enable merozoite invasion of host erythrocytes. However, only a few of these proteins have been characterized. Thirty-six *P. yoelii* merozoite rhoptry proteins were identified as putative rhoptry proteins by proteome analysis. Some of these proteins have been characterized while others still remain an intense area of active research. Molecular characterization and understanding of these novel proteins may assist in vaccine development, design of diagnostic assays and better control of malaria disease. This study was aimed at characterizing two *Plasmodium* rhoptry genes; Translocon and Ribonuclease binding-like (RNB-like) genes using bioinformatics analysis and in vitro cell free expression. Bioinformatics analysis was performed using the databases: PlasmoDB, ExPaSy, PSORTb, SWISSPROT-workspace, GeneDB, National Center for Biotechnology Information (NCBI) and COBALT: Multiple Alignment Tool. Both genes were characterized for features such as conservation profiles, domain architecture and alignment of sequences, both within *Plasmodium* species and among members of the

phylum apicomplexa. The RNB protein domains are generally conserved across *Plasmodium* species but protein identity across species is 30%. The amino acid identity is about 40% across species for the Translocon protein. This study revealed that these genes are expressed early upon merozoite invasion of the host erythrocytes. The expressed translocon protein that is annotated as hypothetical or putative has been shown to be part of a transport complex and the Ribonuclease binding-like (RNB) gene expresses a putative RNB-like protein found in all species of *Plasmodium*. The translocon of *Plasmodium falciparum* was successfully PCR amplified, cloned and a 23 kDa protein was expressed in vitro. Expression was confirmed with rhoptry specific antibodies.

TABLE OF CONTENTS

LIST OF TABLES

LISTS OF FIGURES

ABBREVIATIONS

AAA	ATPases associated with various cellular activities
AMA-1	Apical membrane antigen-1
BSA	Bovine serum albumin
CSD	Cold shock domain
CR1	Complement receptor-1
DARC	Duffy antigen receptor for chemokines
DBP	Duffy binding protein
EBA	Erythrocyte binding antigen
EBL	Erythrocyte binding like proteins
EDTA	Ethylenediamine tetra acetic acid
EGF	Epithelium growth factor
ER	Endoplasmic reticulum
EXP2	Exported protein 2
fMAST	f*alciparum* merozoite assemblage of subpellicular microtubules
GAMP	Glideosome associated membrane protein
GPA	Glycophorin A

GPB ----------	Glycophorin B
GPC ----------	Glycophorin C
GPI ----------	Glycosylphosphatidyl inositol
HT ----------	Host targeting
IMC ----------	Inner membrane complex
KAHRP ----------	Knob-associated histidine-rich proteins
NCBI ----------	National center for biotechnology information
MAEBL ----------	Membrane-associated erythrocyte binding-like
MAHRP1----------	Membrane-associated histidine-rich protein1
MCS ----------	Multiple cloning sites
MEME ----------	Multiple expectation maximization for motif elicitatioin
HMMTOP----------	Hidden markoff model topology prediction
MSP ----------	Membrane surface protein
MSRP ----------	Membrane surface receptor protein

MTRAP ----------	Merozoite thrombospondin related amorphous protein
MudPIT ----------	Multidimensional protein identitification technology
NCP ----------	Nitrocellulose paper
NEB ----------	New England Biolabs
NRS ----------	Norman rabbit serum
OD ----------	Optical density
ORF ----------	Open reading frame
PAN ----------	Apple domains, which are a subset of the plasminogen, apple, nematode
PCR ----------	Polymerase chain reaction
PfEMP-1---------- -	*Plasmodium falciparum* erythrocyte membrane protein-1
PEXEL ----------	*Plasmodium* protein export element
PfRH ----------	*Plasmodium falciparum* reticulocyte-binding Protein homolog

PfMyoA ----------	*Plasmodium falciparum* apical myosin factor
PSORTb ----------	Precise subcellular organelle targeting in bacteria
PTRAMP ----------	*Plasmodium* thrombospondin-related apical merozoite protein
PTEX ----------	*Plasmodium* translocon of exported proteins
PV ----------	Parasitophorous vacuole
RBC ----------	Red blood cell
RBL ----------	Reticulocyte binding ligand
RESA ----------	Ring-infected erythrocyte surface antigen
RON ----------	Rhoptry neck proteins
ROM ----------	Rhomboid proteases
RNB ----------	Ribonuclease-like Binding
RIFIN ----------	Repetitive interspersed family
SDS-PAGE ----------	Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis
SERA ----------	Serine rich antigen

STREVOR	----------	Subtelomeric variable open reading frame
TAE	----------	Tris-acetate-EDTA
TJ	----------	Tight junction
VTS	----------	Vacuolar transport sequences

CHAPTER I

INTRODUCTION

Plasmodium sp. are obligate intracellular protozoan parasites of humans and animals, and are the causative agents of malaria. Malaria still remains a significant economic and public health disease worldwide. Transmission of parasites to humans occurs via the female *Anopheles* mosquito as vector. The geographic distribution of *Plasmodium sp*. in endemic regions puts almost half of the world's population at risk of contracting malaria. This disease is a major cause of morbidity and mortality worldwide. About 219 million people are affected and an estimated 660,000 deaths occurred in 2010 (WHO World malaria report 2011).

The genus *Plasmodium* was first described in 1885 by Ettore Marchiafava and Angelo Celli (Garnham P., 1966). Currently, over 200 species are recognized and new species continue to be described (Iyer J. et al. 2006, 2007). Humans are hosts to four main species, although they can occasionally be infected by other species from nonhuman primates. Recently, the simian malaria species *P. knowlesi* infection has been reported to also infect humans (Luchavez et al. 2008). Most *Plasmodium* species are confined to tropical and subtropical regions depending on the distribution of their insect vectors

(Chipawaza et al. 2012; Sullivan 2010). On a global scale, the overall incidence of malaria is attributed to various species that infect humans as follows: Approximately 50% is due to *P. vivax*, 40% due to *P. falciparum*, 10% is due to *P. malariae*, and less than 1% to *P. ovale*. The parasite always has two hosts in its life cycle; a vector, usually a mosquito and a vertebrate host (Mari et al. 2006)

Classification of *Plasmodium*

Plasmodium belongs to the protozoan family Plasmodiidae, the order Haemosporidia and the phylum Apicomplexa (Telford et al. 1988, 1996). The genus *Plasmodium* was divided by Garnham in 1966 into nine genera according to morphology and their predominant hosts. (Garnham 1966). Three occurred in mammals, four in birds, and two in lizards. But since then, more genera have been recognized and new species continue to be described and re-classified using DNA molecular techniques (Kimura et al. 2006).

Species infecting humans

Two species in the subgenus *Laverania* are currently recognised: *P. falciparum* and *P. reichenowi*. The species that infect humans includes *P. falciparum, P. vivax, P. malariae, P. ovale* and *P. knowlesi.* The subgenus laverania is spread worldwide and infects both humans and chimpanzees. In addition to the cresentic shape, the gametocytes of the *laverania* are characterized by the unique distribution of the malaria pigment which is perinuclear and the cytoplasm is clear elsewhere in the cell. Pigment distribution in all other species of *Plasmodium* are irregular. Three additional species: *P. billbrayi, P. billcollinsi* and *P. gaboni* - may also exist (based on molecular data) but a full description of these species has not yet been published (Mari et al. 2006).

Rodent infecting species: Parasites infecting other mammals including lower primates (mouse, lemurs and others) are classified in the subgenus *Vinckeia. P. vinckei (berghei), P. yoelii* and *P. chabaudi* (Gupta 2011). Species of malaria parasites that infect rodents have long been used as models for malaria disease research. The whole-genome shotgun sequence of *P. yoelii,* and comparative studies with the genome of the human malaria parasite *P. falciparum* (strain 3D7) and a synteny (the physical co-localization of genetic loci on the same chromosome within an individual or species) map of 2,212 *P. yoelii* contiguous DNA sequences aligned to 14 *P. falciparum* chromosomes. The alignment revealed a marked conservation of gene synteny within the body of each chromosome (Florens et al. 2002). Of about 5,300 *P. falciparum* genes, more than 3,300 *P. y. yoelii* orthologues of predominantly metabolic function were identified. Over 800 copies of a variant antigen gene located in sub-telomeric regions were found. This was the first genome sequence of a model eukaryotic parasite, and it provides insight into the use of such systems in the modelling of *Plasmodium* biology and disease (Florens et al. 2002).

Biology of *Plasmodium*.

The genome of four *Plasmodium* species, *P. falciparum, P. knowlesi, P. vivax,* and *P. yoelii* have all been sequenced and many more species are being sequenced. All these species have genomes of about 25 megabases organized into 14 chromosomes, consistent with earlier estimates. The chromosomes vary in length from 500 kilobases to 3.5 megabases and it is presumed that this is the pattern throughout the genus. *Plasmodium* contains a degenerated chloroplast called an apicoplast (Iyer et al. 2005). While several species of *Plasmodium* cause disease in humans, *P. falciparum* is by far the deadliest

(World Malaria Report 2011). The strategies aimed at effective prevention, control, proper diagnosis and treatment has remained the main focus of active research.

Life cycle

The life cycle of *Plasmodium* (Fig. 1), is complex and takes place in two distinct hosts; the vertebrate host and arthropod vector during which the parasite undergoes multiple developmental changes. Both sexual and asexual stages of development occur during the life cycle of the parasite. The different stages of development of the parasite includes the ring stage (trophozoite), schizont, sporozoite and the merozoite that are marked by specific genomic, transcriptomic, proteomic and metabolic states. It is generally believed that a full understanding of the pathogenesis of *P. falciparum* depends on the full understanding of the genomic, transcriptomic, proteomic and metabolic states of *P. falciparum.* My research will be focused on the unique similarities and differences that exist in two genes and their products in *Plasmodium*, other species and within the apicomplexan phylum.

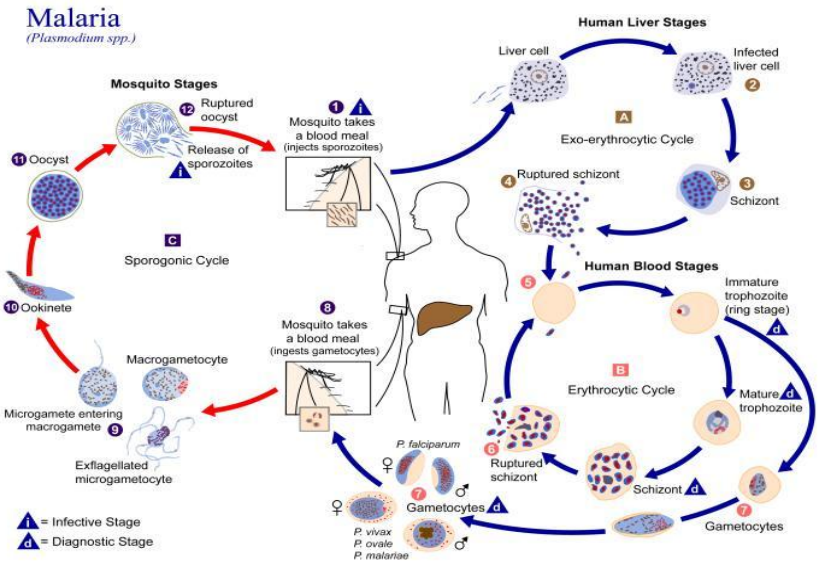


Figure 1. *Plasmodium falciparum.* Life cycle obtained from www.cdc.gov/malaria

Mosquito stage

Prior to transmission, *P. falciparum* resides within the salivary glands of the mosquito as sporozoites. As the mosquito takes its blood meal, it injects the sporozoites into the vertebrate, often a human host (Fig. 1).

Exo-erythrocytic stage

Upon inoculation of the infective sporozoite stage by the *Anopheles* mosquito into humans, the sporozoites takes only a matter of minutes before it reaches the bloodstream and infects the hepatocyes. After the sporozoites enter hepatocytes, the parasite loses its apical complex and surface coat, and transforms into schizonts. Within the parasitophorous vacuole (PV) of the hepatocyte, *Plasmodium* undergoes schizogonic development. In this stage, the nucleus divides multiple times with a concomitant increase in cell size, but without cell segmentation. This exoerythrocytic schizogony

stage of *Plasmodium* has a minimum duration of roughly 5.5 days (Karunaweera et al. 1992). This is followed by parasite cell segmentation. After segmentation, the parasite cells are differentiated into merozoites.

Erythrocytic Stage

Merozoites released from the hepatocytes enter the erythrocytic portion of their life-cycle to continue a different phase and further differentiation. Merozoites were thought to not re-infect hepatocytes, until the hypnozoite form of *P. ovale* and *P. vivax* was recognized (Markus 1978; Krotoski et al. 1980; 2011). Infection of the erythrocyte induces a series of changes in the host cell upon parasite contact with the host cell. These changes depend on the stage of infection. After release from the hepatocytes, the merozoites enter the host erythrocytes using the apical complex organelles to recognize and enter the host erythrocyte.

Trophozoite and Schizoint stages

After invading the erythrocyte, the parasite loses its specific invasion organelles and differentiates into a round trophozoite located within a PV in the red blood cell cytoplasm. The young trophozoite grows substantially before undergoing schizogonic division. The parasite's presence within the erythrocyte induces changes in the properties of the host cell. Some of the changes include the PV and the formation of a surrounding membrane that is continuous with a stack of cisternae and vesicles similar to the Golgi apparatus located beneath the erythrocyte membrane (Wickert et al. 2004; Wickert and Krohne 2007). These structures are known as Maurer’s clefts (MC) and participate in the trafficking and sorting of proteins destined for delivery to the infected erythrocyte cytosol

and membrane (Sam-Yellowe 2009; Haeggstrom et al. 2004; Haldar 1998; Lanzer et al. 2006; Trager et al. 1966). Virulence proteins such as *Plasmodium falciparum* erythrocyte membrane protein-1(PfEMP-1) transported through the clefts are associated with histidine-rich protein (HRP) and Knob-associated histidine-rich proteins (KAHRP) to facilitate cyto-adherance and sequestration of infected erythrocytes within the vascular beds, contributing to the pathogenesis of blood stage malaria (Kirchgatter and Del Portillo 2005; Lanzer et al. 2006). Another major structural alteration of the host erythrocyte is the appearance of electron-dense protrusions, or 'knobs', on the erythrocyte membrane of *P. falciparum*-infected cells. The knobs are induced by the parasite and several parasite proteins that are associated with the knobs (Deitsch and Wellems 1996). Two proteins which might participate in knob formation or affect the host erythrocyte submembrane cytoskeleton and indirectly induce KAHRP and erythrocyte PfEMP2 (also called MESA) are histidine-rich protein-1 (HRP-1) and erythrocyte membrane protein-1 (EMP-1) (Kilegian 1979). Neither KAHRP nor PfEMP2 are exposed on the outer surface of the erythrocyte, but are localized to the cytoplasmic face of the host membrane. Their exact roles in knob formation are not known, but may involve reorganizing the sub-membrane cytoskeleton. (Deitsch and Wellems 1996).

Within the red blood cell, parasite metabolism depends greatly on the digestion of hemoglobin. A set of enzymes known as plasmepsins which are aspartic acid proteases are used to degrade hemoglobin (Boyle et al. 2010). Erythrocyte invasion and growth leads to activation of several distinct anion channels and a non-selective Ca^{2+} -permeable cation channel (Francis et al. 1997). The blood stage of *P. falciparum* occurs in a highly synchronous fashion, with roughly all of the parasites throughout the blood in the same

stage of development. This precise clocking mechanism has been shown to be dependent on the human host's own circadian rhythm (Karunaweera et al. 1992). Specifically, human body temperature changes as a result of the circadian rhythm, seem to play a role in the development of *P. falciparum* within the erythrocytic stage (World Health Organization 2011).

Gametogenesis

During the erythrocytic stage, some merozoites develop into male and female gametocytes. This process is called gametogenesis (gametocytogenesis). The gametocytes remain within the erythrocytes until taken up by the mosquito host. Once taken up by the female *Anopheles* mosquito during a blood meal from an infected human, the gametocytes differentiate into male and female gametes in the gut of the mosquito. The male microgametes separate from the parasite cell in a process referred to as ex-flagellation. Fertilization of the female gametes by the male gamete occurs rapidly after gametogenesis. The fertilization event produces a zygote in the gut of the mosquito which develops into an ookinete which is the motile zygote. These ookinetes soon penetrate the wall of the gut of the mosquito and become oocysts. The oocyst later releases sporozoites which migrate into the salivary gland of the Anopheles mosquito which will be injected into the susceptible host. The zygotes are the only diploid stages of *Plasmodium.* See figure 1 above. The illustration of the mechanism of host cell invasion by *Plasmodium* parasites is outlined below (Fig. 2).

Mechanism of erythrocyte invasion by *Plasmodium* parasite

Erythrocyte invasion by malaria parasites has been studied by live cell imaging (electron microscopy) and super-resolution optical imaging and it indicates that the process is a dynamic one, involving both passive and active processes. The mechanism of erythrocyte invasion by *Plasmodium* parasites involves a number of proteins and series of steps (Chitnis et al. 2000; Iyer et al. 2007; Gilson and Crabb 2009; Angrisano et al. 2012). There are four distinct steps namely: (1) Adhesion, (2) Re-orientation (3) Tight-junction formation and (4) Ingress (Iyer et al. 2007; Gilson and Crabb. 2009; Farrow et al. 2011).

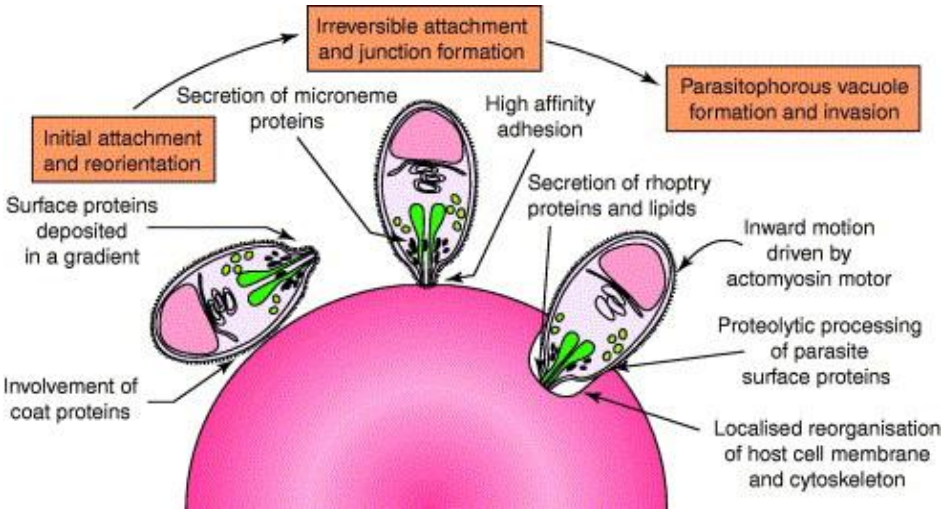


Fig. 2 Illustration of host cell invasion by *Plasmodium* (Chitnis et al. 2000)

The preferred dogma for the initial interaction between the free merozoites and the host erythrocytes is that adhesion occurs randomly when the merozoites collide with the erythrocytes soon upon release from the hepatocytes into the blood stream. Merozoites

recognize and bind to the erythrocyte plasma membrane. This is followed by re-orientation of the merozoite until its apical tip is in contact with the erythrocyte. A tight-junction complex is then formed at the point of contact through the secretion of neck proteins from micronemes and rhoptry organelles (Singh et al. 2013; Angrisano et al. 2012). After the parasite has entered the host cell, the dense granule proteins are released and the PV is formed. Ingression involves the encircling of the merozoite by the tight junction as it actively moves backwards from the apex till the merozoite is completely encapsulated by the erythrocyte membrane forming the PV inside the red cell.

The process of erythrocyte invasion is a dynamic one mediated by a set of proteins through ligand-receptor interactions. While a subset of proteins such as merozoite surface protein (MSP1) and apical membrane antigen-1(AMA1) mediate the initial merozoite-erythrocyte interaction (Gruber et al. 2012; Haussig et al. 2011; Crosnier et al. 2011; Cowman and Crabb 2006; Cowman et al. 2012), other parasite proteins are also known to be implicated in the recognition and binding of parasites to the host cells. They are broadly classified into two families; the reticulocyte binding protein homolog (*Pf*RH) and erythrocyte binding-like proteins EBL (Iyer et al. 2007). Initial work done in *P. yoelii*, *P. vivax*, and *P. knowlesi* identified RBL homologs (*Py*235kDa rhoptry proteins of *P. yoelii*) (Holder and Freeman 1981) and *P. falciparum* rhoptry neck protein homologs (*Pf*RH1, *Pf*RH2a, *Pf*RH2b, *Pf*RH3, *Pf*RH4, *Pf*RH5; (Rayner et al. 2003; Taylor et al. 2001; Triglia et al. 2001a; Cowman and Crab, 2006). EBL homologs include the Duffy-binding protein (DBP) of *P. vivax*, and *P. knowlesi.* The *P. falciparum* homologs include EBA175, EBA140, EBL1, EBA181, and EBA165 (Iyer et al. 2007). The expression of the RBL and EBL is highly polymorphic and variedly expressed in

terms of receptor recognition or invasion profiles across *Plasmodium* species in order to modulate their invasive potential. Only few sets of the RBL or EBL is exposed in any one merozoite upon contact with one erythrocyte (Taylor et al. 2002; Nery et al. 2006).

The 235kDa rhoptry proteins, located in the rhoptry organelles of the parasite, were found to be essential for erythrocyte invasion by these parasites. Ultimately, parasites are able to discriminate among the different red blood cells through receptor-ligand interactions. Analogous receptor molecules known as glycophorins A, B, C (GPA, GPB & GPC) and Band 3 are receptors for invasion by *P. falciparum* (Hadley et al. 1987; Goel et al. 2003 and Maier et al. 2003). The Duffy-binding protein (DBP) is also known as the Duffy antigen receptor for chemokines, DARC (Chitnis et al. 2008). DBP is the receptor for *P. vivax* invasion into the reticulocytes. Cells lacking DBP are refractory to *P. vivax* invasion. However, unlike *P. vivax*, no known human erythrocyte lacking an individual receptor has been found to be refractory to invasion by *P. falciparum*, which reflects redundancy in receptor utilization by *P. falciparum*, an evolutionary act by *P. falciparum* to ensure survival (Deepak and Chetan 2011). Upon contact with the erythrocyte, merozoite surface proteins (MSPs) bind to the erythrocyte and cause the parasite to adhere to the host cell. Genetic knockout experiments have demonstrated that one or more merozoite surface proteins, especially erythrocyte binding antigens (EBA) and rhoptry proteins are essential in erythrocyte adhesion both of which are known to bind to GPA on erythrocyte membrane and PfRh4 binds to complement receptor-1 (CR-1). The re-orientation of the merozoite allows for the parasite's passive movement across the erythrocyte surface such that the parasite and the erythrocyte make contact at the

point of maximum curvature, which is the apex and the erythrocyte outer rim in order to minimize the energy cost required for membrane distortion (Farrow et al. 2011).

Formation of tight junction and parasitophorous vacuole

Once contact and re-orientation has occurred, micronemal and rhoptry proteins mediate specific parasite-host interactions that lead to the formation of a tight junction. Invasion proceeds rapidly with the discharge of the contents of the micronemes, rhoptries and finally, the dense granules. It is thought that rhoptries are discharged after the micronemes and assist in the formation of the PV. To complete the process of invasion, merozoites secrete serine proteases which cleave an erythrocyte membrane protein called Band 3 and causes a localized disruption in the erythrocyte cytoskeleton, thus allowing the parasite to enter the host cell (Hanssen et al. 2008; Dowse et al. 2008). A PV forms at the junction area, creating a small annulus through which the parasite enters the host cell. This PV expands as the parasite enters the host cell. Upon completion of entry of the parasite into the host cell, the tight junction disappears and the PV and the erythrocyte membrane fuses and separates completing the entry process. The contents of the dense granules are thought to be released only after complete entry of the parasite and is implicated in the host cell modification (Zuccala and Baum 2011).

Invasion motor and moving junction

Plasmodium merozoites have a conserved molecular machinery for motility which is comprised of a central actin-myosin motor unit located in the membrane of the parasite and is linked to the adhesin on the outer surface and to the inner membrane complex (IMC). The actin filament is connected to the myosin heavy chain which is also

connected to the IMC myosin A tail via interaction with the merozoite thrombospondin-related anonymous proteins (MTRAP). As the merozoite invades the erythrocyte, it drags the erythrocyte plasma membrane over its surface until it is completely enveloped by the PV inside the erythrocyte. The parasite then forms a firm grip on the erythrocyte membrane by forming a bridge between itself and the membrane via protein-protein interactions forming a ring-like structure near the apex of the parasite. Two trans-membrane proteins namely, MTRAP and apical membrane antigens-1 (AMA-1) bridge the gap between the parasite and the erythrocyte (Deepak and Chetan 2011; Farrow et al. 2011). Associated with these proteins on the cytoplasmic side of the parasite is an enzyme aldolase, which mediates the formation of complexes with rhoptry neck (RON) proteins, specifically RONs 2, 4, 5 and 8 that are localized to the moving junction (Deepak and Chetan 2011; Farrow et al. 2011). The RON proteins form stable connections with the erythrocyte while the aldolase helps form a motor unit on the merozoite side of the complex (Besteiro et al. 2009). The aldolase helps both to form a stable connection with the tight junction and also binds a filamentous actin (F-actin) which is ATP driven (Dluzewski et al. 1989). The actin filaments associated with the tight junction complex are then pulled towards the rear end of the parasite by a myosin motor unit that is coupled to the IMC. Active migration of the tight junction complex towards the rear of the merozoite is a crictical step in host cell invasion (Farrow et al. 2011). A number of drugs such as cytochalasin D and butanedione monoxime in drug inhibitory studies have both demonstrated that acto-myosin motor activity underpins ingress into the erythrocyte (Deepak and Chetan 2011). *Plasmodium falciparum* Myosin-ATPase (*Pf*MyoA) is responsible for powering *P. falciparum* merozoite ingress into the

erythrocytes (Farrow et al. 2011). Also, the recent discovery of an actin depolymerizing factor *Pf*formin-2 found localized in the apical region of the merozoites shortly before invasion is initiated, demonstrate that G-actin might generate a unidirectional filament with pointed ends directed towards the rear of the parasite (Farrow et al. 2011). The movement and force generated by the actin-myosin motor is responsible for many forms of cell motility, including muscle contraction. The basic mechanism of force generation from actin-myosin, involves the cyclical interactions of the myosin motor domain with the actin filament coupled to ATP hydrolysis (Geeves et al. 2005)

In order to generate force and movement, *Pf*MyoA and *Pf*actin must firmly be anchored to both the fixed and moving parts within the merozoite. Since *Pf*MyoA is a motor unit, several molecules are coupled together in order to keep hold of an individual actin filament. *Pf*MyoA is attached to the IMC by a set of accessory proteins (the inner membrane complex (IMC). The IMC association of the myosin motor, PfMyoA, is maintained by its association with three proteins; PfMTIP, a myosin light chain, PfGAP45, an IMC peripheral membrane protein, and PfGAP50, an integral membrane protein of the IMC (Farrow et al. 2011). Immunofluorescence studies have shown that the merozoite inner membrane is a hetero-oligomeric complex that is associated with the proteins known as myosin tail interacting protein (MTIP) and Glideosome associated Protein (GAP), (*Pf*GAP45 and *Pf*GAP50). MTIP is responsible for targeting the *Pf*MyoA to the cortical space where it interacts with other proteins to form the IMC. The *Pf*MyoA-*Pf*MTIP-*Pf*GAP45 complex is then anchored to the *Pf*GAP50 at the IMC which is also associated with other glideosome associated trans-membrane proteins (GAPM) to form large monomeric proteins. For the motor system to work, actin filaments must be

tethered or anchored to the tight junction complex and directed towards the rear end of the merozoite. It was reported recently that actin depolymerization factor plays an important role in both *Plasmodium* invasion and motility. This actin depolymerizing factor known as Pfformin-2 localizes to the apical region of the merozoite shortly before invasion. (Sivaraj and Spudich 2009; Farrow et al. 2011).

Apical complex organelles

The Apicomplexa are a large group of protists, most of which possess a unique organelle called an apicoplast and apical complex organelles involved in host cell invasion. Apicomplexans are a diverse group including organisms such as coccidia, gregarines, piroplasms, haemogregarines, and plasmodia. Diseases caused by apicomplexan organisms include, but are not limited to: malaria (*Plasmodium*), babesiosis (*Babesia*), coccidian diseases including: cryptosporidiosis (*Cryptosporidium parvum),* cyclosporiasis (*Cyclospora cayetanensis*), isosporiasis (*Isospora belli*), and toxoplasmosis (*Toxoplasma gondii*). These organisms have a basic apical complex structure comprising the following organelles and structures namely; rhoptries, micronemes, dense granules, conoids and microtubules. The organelles secrete proteins (ligands) that facilitates sporozoite and merozoite recognition and subsequent invasion of the host cell. Table 1A-E is a summary of the apicomplexan organelle proteins used by the *Plasmodium* parasite for host cell invasion.

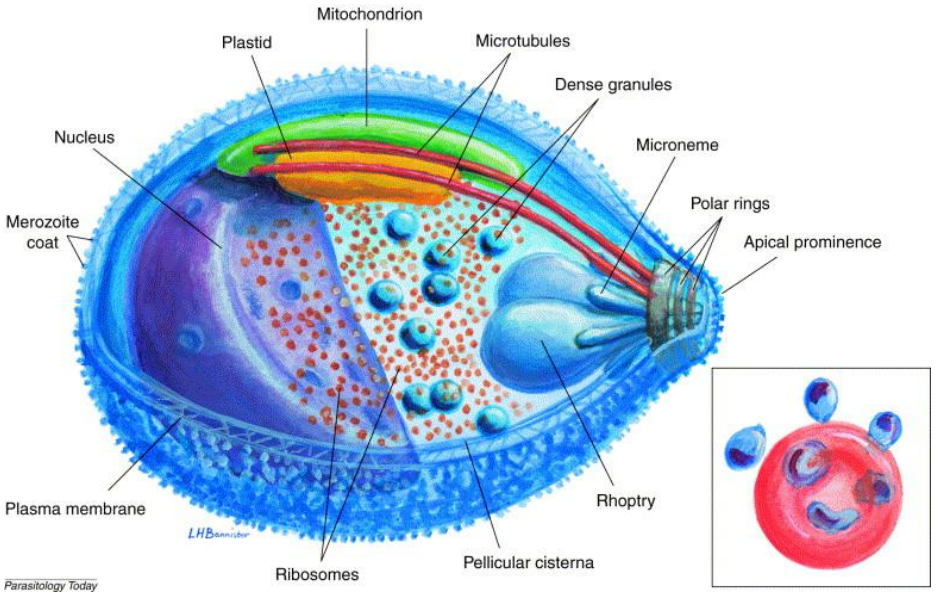


Fig. 3 Illustrated guide to Ultrastructure of *Plasmodium* merozoite. Bannister et al. 2000

Role of neck proteins in host cell invasion

The molecular organization of the moving junction is not completely understood, but recent work has shown that AMA-1 forms a complex with the rhoptry neck proteins RONs 2, 4, 5 and 8 localized to the moving junction. The RON proteins are exported by the parasite and inserted into the erythrocyte membrane. RON2 is lodged into the erythrocyte membrane and interacts with AMA-1 which is secreted to the surface of the parasite after secretion from the micronemes. The other RONs (2, 5, and 8) cross the host cell membrane and are located to the cytosolic side. AMA-1 could not be demonstrated to bind to normal erythrocytes and antibodies against AMA-1 block erythrocyte invasion by homologous *P. falciparum* strains. Limited blocking of invasion was observed in heterologous strains (Deepak and Chetan 2011).

Rhoptry Proteome

The *Plasmodium* rhoptries have been shown to participate in merozoite host cell invasion. Initial studies to elucidate the role of rhoptry proteins were carried out on rodent *Plasmodium* species such as *P. yoelii*. Several orthologues and paralogues of rhoptry genes have been identified in all *Plasmodium* species and in some other members of the phylum, apicomplexa. (Sam-Yellowe et al. 2004). Several rhoptry proteins have been implicated in the process of host cell invasion. Two of these are the genes PfRh2a and PfRh2b. PfRh2b is essential for a well-defined invasion pathway while PfRh2a is not required or sufficient for this pathway.

It appears that MSP1 binds to heparin-like molecules such as Band 3 on the surface of the erythrocyte and that binding is an essential step in the invasion process (Goel et al. 2003). The protein PfRON2 via the C-terminal as well as the central cysteine-rich domain interacts with PfAMA1 (Hossain et al. 2011; Besteiro et al. 2009). The role of PfRON2 and PfAMA1 and other rhoptry neck proteins in the formation of the membrane junction complex in host cell invasion was fully described in both *T. gondii* and *P. falciparum* using immuno-histochemical staining and inhibition studies (Aikawa and Miller 1983; Ward et al. 1993; Dluzewski et al. 1995; Besteiro et al. 2009; Gabriela et al. 2011).

Table1A. GPI-anchored membrane surface proteins (MSP) secreted during erythrocyte invasion (Source: Cowman et al. 2012)

Name	PlasmoDB accession number	Genetic knockout	Localization in merozoite before/during invasion	Potential function	Feature/structure
GPI-anchored MSPs					
MSP-1	PF3D7_0930300	N	Surface/complex shed during invasion with MSP1/19 EGF C-terminal domain retained in PV of ring stage	Putative Band 3 ligand; C-terminal double EGF domain redundant for divergent molecules: processed SUB1 and -2	Two C-terminal EGF domains: compact side by side arrangement
MSP-2	PF3D7_0206800	N	Surface	Highly polymorphic; likely structural role as surface coat	Unordered repetitive structure
MSP-5	PF3D7_0207000	N	Surface	Not known	C-terminal EGF domain
MSP-4	PF3D7_0206900.1	Y	Surface	Not known	C-terminal EGF domain
MSP-10	PF3D7_0620400	N	Surface	Not known	C-terminal EGF domain
Pf12	PF3D7_0612700	Y	Surface/shed	Potential adhesive protein	6-Cys domains
Pf38	PF3D7_0508000	Y	Surface/shed	Potential adhesive protein	6-Cys domains
Pf92	PF3D7_1364100	Y	Surface/shed	Not known	Cys-rich protein

Table1B. Peripheral proteins secreted during erythrocyte invasion (Source: Cowman et al. 2012)

Pf113	PF3D7_1420700	N	Surface/shed	Not known	No data
MSP-9 (ABRA)	PF3D7_1228600	Y	Surface/shed	Putative protease	No data
S-antigen	PF3D7_1035200	N	Secreted into PV of schizont and released on egress	Not known; potential immunomodulatory role	Highly repetitive and diverse protein
GLURP	PF3D7_1035300	Y	Secreted into PV of schizont and released on egress	Not known	Repetitive Glutamate-rich
MSP-3	PF3D7_1035400	Y	Surface/shed	Not known; binds to MSP-1	Repetitive and Glutamate-rich
MSP-6	PF3D7_1035500	Y	Surface/shed	Not known; binds to MSP-1	Leucine zipper-like C-terminal domain
H101 (MSP-11)	PF3D7_1035600	Y	Surface/shed	Not known	MSP-3 family, leucine zipper-like C-terminal domain
H103	PF3D7_1035900	Y	Surface/shed	Not known	MSP-3 family, leucine zipper-like C-terminal domain
MSP-7	PF3D7_1335100	Y	Surface/shed	Associates with MSP-1, gene knockout in *P. berghei* shows important in invasion of mature erythrocytes	No data
MSP-7-like (MSRP2)	PF3D7_1334800	Y	Surface/shed	Not known; may associate with MSP-1	MSP-7 family
MSPDBL-1	PF3D7_1036300	Y	Surface/shed	Binds to unknown receptor on red cell	Member of EBL family, DBL and leucine zipper-like domains
MSPDBL-2	PF3D7_1035700	Y	Surface/shed	Binds to unknown receptor on red cell	Member of EBL family, DBL and leucine zipper-like domains
SERA3	PF3D7_0207800	Y	Secreted into PV of schizont and released on egress	Cysteine protease domain with active site serine	Cysteine protease domain
SERA4	PF3D7_0207700	N	Most secreted into PV of schizont and released on egress	Cysteine protease domain with active site serine	Cysteine protease domain
SERA5	PF3D7_0207600	N	Secreted into PV of schizont and released on egress	Cysteine protease domain with active site serine	Cysteine protease domain
SERA6	PF3D7_0207500	N	Most secreted into PV of schizont and released on egress	Cysteine protease domain with active site cysteine	Cysteine protease domain
Pf41	PF3D7_0404900	Y	Surface/shed	Potential adhesive protein; binds Pf12 on merozoite	6-Cys domains

Table 1C. Plasma membrane proteins secreted during erythrocyte invasion (Source: Cowman et al. 2012)

Name	PlasmoDB accession number	Genetic knockout	Localization in merozoite before/during invasion	Potential function	Feature/structure
ROM1	PF3D7_1114100	Y	Mononeme (proposed new apical organelle) or microneme/surface	Rhomboid protease; cleaves AMA1, MAEBL, EBLs, PfRh proteins; likely role after invasion in PV formation	Multipass transmembrane protein
ROM4	PF3D7_0506900	ND	Surface/shed	Rhomboid protease; cleaves AMA1, MTRAP, EBL, and PfRh proteins in transmembrane to allow shedding during invasion	Multipass transmembrane protein

Table 1D. Microneme proteins secreted during erythrocyte invasion (Source: Cowman et al. 2012)

AMA 1	PF3D7_1133400	N	Micronemes/surface and binds to RON2 that has been inserted into red cell membrane and tracks with tight junction	Released on merozoite surface; binds RON complex; potential ligand for McLeod antigen, phosphorylation of cytoplasmic tail essential, may be involved in signaling	PAN (plasminogen, apple, nematode) motifs
EBA-175	PF3D7_0731500	Y[a]	Micronemes/surface and binds to glycophorin A	Binds to glycophorin A, likely signaling role for invasion	EBL family with DBL domains; "handshake" association between region II dimers creates groove for glycophorin A binding
EBA-181/JESEBL	PF3D7_0102500	Y	Micronemes/surface and binds to unknown receptor	Binds to unknown receptor on red cell	EBL family member with DBL domains
EBA-140/BAEBL	PF3D7_1301600	Y	Micronemes/surface and binds to glycophorin C	Binds to glycophorin C on red cell	EBL family member with DBL domains
EBL-1	PF3D7_1371600	Y	No data	Binds to glycophorin B, nonfunctional because of mutations causing truncated protein	EBL family member with DBL domains
PTRAMP	PF3D7_1218000	ND		Not known; cleaved by SUB2 on merozoite surface	Long extended structure
PfRipr	PF3D7_0323400	N	Micronemes/surface and binds to PfRh5	Binds to PfRh5	10 EGF domains, 87 cysteines
MTRAP	PF3D7_1028700	N	Micronemes/PV	Potential motor-associated protein	Thrombospondin-like domains
PTRAMP	PF3D7_1218000	N	Micronemes/surface	Potential motor-associated protein	Thrombospondin-like domains
SPATR	PF3D7_0405900	ND	Micronemes/surface	Not known for blood stages	Thrombospondin-like domains
GAMA	PF3D7_0828800	ND	Micronemes/surface	Binds to red cells; has GPI anchor	No data
SUB2	PF3D7_1136900	N	Micronemes/PV	Protease that processes MSP-1, MSP-6, MSP-7, AMA1, PTRAMP and other proteins to prime merozoite for invasion	Subtilisin-like serine protease

Table 1E. Exoneme proteins secreted during erythrocyte invasion (Source: Cowman et al. 2012)

Name	PlasmoDB accession number	Genetic knockout	Localization in merozoite before/during invasion	Potential function	Feature/structure
SUB1	PF3D7_0507500	N	Exonemes/PV	Protease that processes MSP-1, MSP-6, MSP-7, AMA1, RAP1, MSRP2 and SERAs to prime merozoite for invasion	Subtilisin-like serine protease
Rhoptry neck proteins					
PfRh1	PF3D7_0402300	Y[a]	Rhoptry neck/surface	Binds to red cells via receptor Y	PfRh family
PfRh2a	PF3D7_1335400	Y	Rhoptry neck/surface	Binds to red cells via receptor Z	PfRh family
PfRh2b	PF3D7_1335300	Y	Rhoptry neck/surface	Binds to red cells via receptor Z	PfRh family
PfRh4	PF3D7_0424200	Y	Rhoptry neck/surface	Binds to red cells via complement receptor 1	PfRh family
PfRh5	PF3D7_0424100	N	Rhoptry neck/surface forms complex with Ripr	Binds to red cells via Basigin	Classed as PfRh family but lacks homology and no transmembrane so likely functionally distinct
RON2	PF3D7_1452000	ND	Rhoptry neck/into red cell membrane	Inserted in red cell membrane at invasion, forms complex at tight junction with RON proteins and AMA-1	Multipass transmembrane protein
RON3	PF3D7_1252100	ND	Rhoptry neck/into red cell	Likely also forms complex at tight junction with other RON proteins and AMA-1	No data
RON4	PF3D7_1116000	ND	Rhoptry neck/into red cell	Injected into red cell, binds to RON2 and forms a complex at tight junction with RON proteins and AMA-1	Binds to AMA1 via hydrophobic groove
RON5	PF3D7_0817700	ND	Rhoptry neck/into red cell	Forms complex at tight junction with RON proteins and AMA-1	No data
ASP	PF3D7_0405900	ND	Rhoptry neck/surface	Not known; has putative GPI anchor	Sushi domains

N, knockout attempt unsuccessful; Y, knockout generated; ND, knockout not attempted; PV, parasitophorous vacuole; MSP, merozoite surface protein
[a]EBL and PfRh families show overlap in function and, while individually nonessential, overall are essential.

The 23-megabase nuclear genome of *P. falciparum* consists of 14 chromosomes, containing 5,300 genes. The genome is highly (A + T)-rich (Gardner 2002). Genes involved in antigenic variation are concentrated in the sub-telomeric regions of the chromosomes (Gardner et al. 2002). Compared to the genomes of free-living eukaryotic microbes, the genome of this intracellular parasite encodes fewer enzymes and

transporters, but a large proportion of genes are devoted to immune evasion and host–parasite interactions (Gardner et al. 2002). Many nuclear-encoded proteins are targeted to the apicoplast, an organelle involved in fatty-acid and isoprenoid metabolism. The genome sequence provides the foundation for future studies of this organism, and is being exploited in the search for new drugs and vaccines to fight malaria (Gardner et al. 2002).

Purpose

In this research, the main objective was to investigate the level of gene organization and conservation, structural conformation and domain organization of two rhoptry genes, within *Plasmodium* species and across the apicomplexan phylum.

Two genes, PF3D7_1436300, previously known as PF14_0344, (*P. yoelii* PY02301 Py17XNL), encoding a translocon protein in *P. falciparum* species, and PF3D7_0906000, (*P. yoelii* PY04959) previously known as PFI0295c, encoding a Ribonuclease binding-like (RNB-like) putative protein were studied. Both genes were chosen as candidates for study due to their potential role in erythrocyte invasion and potential as vaccine candidates.

Objectives

The main objectives of this research include:

1. To analyze Translocon and RNB-like proteins using bioinformatics analysis
2. Clone and express PF3D7_1436300 (translocon)
3. Test the reactivity of the expressed translocon with rhoptry-specific antisera

CHAPTER II

EXPERIMENTAL APPROACH

Bioinformatics analysis

The general information regarding the proteins encoded by the genes PF3D7_1436300 and PF3D7_0906000 were obtained from Plasmodb (Plasmodb.org). *Plasmodium falciparum* orthologues and the FASTA sequences were also obtained from Plasmodb.org. Multiple sequence alignment was done using the constraint-based multiple alignment tool (COBALT) (http://www.ncbi.nlm.nih.gov/tools/cobalt/) from GenBank. This helped in analyzing the sequence identity between proteins encoded by *P. falciparum* and *P. yoelii* genes. The expert protein analysis system (ExPASy) bioinformatics Portal software helped in identifying the possible secondary structure of both proteins. The conserved domain database (CDD) helped in identifying unique motifs on the proteins.

Parasitic cell culturing:

Plasmodium falciparum (strains 3D7 and FCR-3) were cultured in type A+ human erythrocytes (Interstate Blood Bank, Memphis TN. USA) at 5% hematocrit and 20% parasitemia was maintained in vitro according to the method of Trager and Jensen (1976) in RPMI 1640-Hepes media supplemented with 10% human serum (Interstate blood

bank). Schizont pellets were collected by treating *Plasmodium*-infected erythrocytes with 10mM Tris HCl pH 8.8 and centrifuging at 15000 rpm for 15 min (Sam-Yellowe et al. 1988). Parasite pellets were stored at -70°C after adding protease inhibitor (aprotinin) to the pellets. The parasite pellets were used for protein extraction, genomic DNA isolation and RNA isolation.

DNA isolation:

Plasmodium falciparum pellets with approximately 20% parasitemia was used. Pellets were treated with 600 µl of 10 mM Tris-HCl of pH 7.6, 50mM EDTA pH 8.0, 0.1% SDS and 1mg/ml proteinase K, incubated overnight at 50°C. Phenol:chloroform extractions were performed twice and supernatants were collected into an eppendorf tube. Chloroform extraction was performed to remove excess phenol. Precipitation of DNA was done using 0.1 vol of 5 M sodium acetate and 1 vol of isopropanol by incubating at room temperature for 15 minutes. DNA was centrifuged at 14000 rpm for 10 minutes, to collect the DNA pellet followed by a 70% ethanol wash to remove excess salt from the DNA. DNA was stored in dH_20 at -20°C.

Polymerase chain reaction (PCR), cloning and protein expression

Plasmodium falciparum genes PF3D7_1436300 (PfA0680, Pfc140344) and PF3D7_0906000 (PFI0295c) previously identified in proteome studies (Sam Yellowe et al. 2004) were amplified using the polymerase chain reaction (PCR). Amplification of the genes was performed using primers designed with appropriate restriction sites for cloning. Appropriate restriction sites were included in the primers to permit cloning of the gene sequence into the multiple cloning sites of the expression plasmid pT7CFE1-

CHis (Thermo Fischer). Genomic DNA from *P. falciparum* strain 3D7 and from FCR3 were used as templates. The PCR products were size fractionated on 1% agarose gels, DNA bands were excised from the gels and purified by Freeze and Squeeze DNA Gel Extraction Spin Columns (Bio-Rad Laboratories. Hercules CA). The pT7CFE1-CHis expression vector and gel purified DNA were then separated on a 1% agarose gel, followed by gel purification of the digested DNA. Ligation of the digested plasmid vector and PCR products were carried out at 12°C overnight (Thermolyne Thermokool, Barnstead).

Expression of proteins from genes

In vitro human cell free expression system

The Pierce® Human in vitro protein Expression Kit for DNA templates (Thermo Scientific. Rockford. IL) was employed for protein synthesis from the cloned DNA fragments using transcription and translation kit components following the manufacturer's protocols. Proteins expressed using pT7CFE1-CHis expression vector have a C-terminal His tag. Twenty microliters (20 µl) of transcription mixture containing 2µl of the recombinant plasmid DNA was incubated for 75 min at 30°C. After incubation, 2µl of transcript mixture containing transcribed mRNA was mixed with 25µl of the translation mixture followed by incubation at 30°C for 90 min. The Pierce® 1 – step Human Coupled in vitro translation (IVT) Protein Expression Kit for DNA templates (Thermo Scientific, Rockford, IL) was also employed for protein synthesis from the cloned DNA fragments, using the components provided in the kit and following manufacturer's protocols.

Table 2. List of primers used for PCR

Gene ID	Protein	Primer Sequence for PCR
PF3D7_1436300	Translocon component	(F) *5' cccggatcc* CTCGAGAATAATAACAATCATAATAATAAG (R)*5' ccccctcgag* GAATTCATTATCATCAGGTTTAGCTAATTTTC
PF3D7_0906000	RNB-like	(*F1) 5' aagctt* GGATCCATGTTAGGTCAAAAAAACACAAATA (R1)*5'cccgcggccgc* CTCGAGTGTTATTTGCTTTTTGTTTTGAAAA (F2) 5' *aagctt* ATATATATTTTTACCATAATACTATGTG (R2)*5' cccgcggccgc* ACCTATTTTCATGTCAGGAAAATAACCCTT

SDS-PAGE and Western blotting:

Translated proteins were separated on 10 % SDS-PAGE gels and transferred to nitrocellulose paper (NCP) using a semi-dry western blotting chamber for 2h. Nitrocellulose paper was blocked with 2% non-fat milk and incubated with specific antibodies. Rabbit antisera #676 (Sam-Yellowe et al. 1995) specific for *P. falciparum* merozoite rhoptries, and antiserum #685 specific for the *P. falciparum* parasitophorous vacuole protein, SERA (serine rich antigen) (Sam-Yellowe et al. 2000), were used in western blotting. Goat anti-rabbit antibody conjugated to horse radish peroxidase (HRP) was used as secondary antibody at a dilution of 1:1000. Goat anti-rabbit antibodies were

diluted 1:1000 in 2% milk in 1X blot buffer. Normal rabbit serum was used as control. *Plasmodium falciparum* schizont protein extracts were separated on the gel as control.

CHAPTER III

RESULTS

Bioinformatic analysis of PF3D7_1436300

Data mining of *P. falciparum* 3D7 gene PF3D7_1436300 in plasmodb indicated that the gene is present on Chromosome 14. The size of the gene is 2982 bp and it encodes a protein of 993 amino acids with a molecular mass of 112.4 kDa. The gene has six *Plasmodium* orthologs PBANKA_100850, PCHAS_100940, PCYB_132260, PFIT_1437500, PKH_131290, PVX_084720, PY02301 and PYYM_1010100. The gene is annotated as a translocon in *P. faciparum, P. berghei,* and *P. chabaudi,* but as a hypothetical protein in *P. vivax, P. yoelii* and in *Theileria parva muguga.* The protein has an isoelectric point of 3.99. The hidden markoff model topology prediction (HMMTOP) predicted that the protein has one transmembrane domain and a consensus signaling peptide with an RxLxE/Q/D domain (de Koning-Ward et al. 2000; Osborne et al. 2010; Chang et al. 2008). The gene is conserved within the genus *Plasmodium* in terms of the type of residues present but polymorphism is significantly high with only 30% identity.

Based on data from immunofluorescence, immunoelectron microscopy and gene knockout studies, PF3D7_1436300 is a putative translocon rhoptry gene that expresses a protein co-transporter known as PTEX150 (Plasmodb). HSP101 associates with PTEX150 to transport nascent peptides during protein biogenesis from the PV to the

erythrocyte cytoplasm to the erythrocyte membrane (de Koning-Ward et al. 2009). PTEX150 is part of a *Plasmodium* transport protein complex comprising PEXEL, PTEX88, EXP2 and HSP70 (Chang et al. 2008, 2010; Osborne et al. 2010; de Koning-Ward et al. 2009) (Fig. 4). PTEX150 has a putative endoplasmic reticulum (ER) signal sequence that is found in most species of *Plasmodium* that infect humans and in *Theileria*, another member of the apicomplexan phylum (de Koning-Ward et al. 2009).

There are numerous repetitive amino acid residues of asparagine, lysine and glutamic acids in PTEX150 detected in the predicted protein sequences and multiple alignment analysis (Appendix 3. Plasmodb). PTEX150 is synthesized and annotated as either putative or protein of unknown function in *P. yoelii* and in some species of the same phylum that expresses the gene. No conserved domain architecture was found both within the *Plasmodium* species and other members of the apicomplexa that still carry the gene (*Theileria p. muguga*). There are also four conserved motifs with a CC6+RK expression pattern in *P. berghei* and *P. berghei- ANKA*. PF3D7-1436300 is annotated to possess three phosphorylation sites, an elongation factor hand-1 (EF-Hand1) binding domain with numerous repetitive amino acid residues of asparagines, lysine and glutamic acid found throughout the protein (Appendix 3. Plasmodb).

The specific sequence of PTEX150 that interacts with the export machinery has not been described. However, through multiple expectation maximization for motif elicitation (MEME) and motif alignment search tool (MAST) on the *P. falciparum* database (Plasmodb; Hiller et al. 2004) revealed the presence of conserved motifs of 11 amino acids RxSRILAExxx in the vacuolar transport sequences (VTSs) of parasite proteins that are transported to the erythrocyte cytoplasm. The primary sequence pattern

RxSRILAExxx is observed in over 250 parasite proteins (Hiller et al. 2006). Prominent among these proteins are those that are host-membrane bound such as *P. falciparum* erythrocyte membrane proteins1 and 2 (*Pf*EMP1, *Pf*EMP2), *P. falciparum* histidine rich protein II (*Pf*HRPII), Repetitive Interspersed Family (RIFINs) and subtelomeric variable open reading frame (STREVOR) (Gardner et al. 2002).

Mass Spectroscopy evidence from Plasmodb showed that the translocon gene product is found in both the nuclear and cytoplasmic fractions of trophozoite and schizont stages of the 3D7 strain of *Plasmodium*. The protein is expressed at all stages of the parasite post-invasion.

Mechanism of Translocon protein as a cotransporter

The mechanism of protein export from the parasite PV to the erythrocyte cytosol and membrane is still unclear (Fig. 4). Recent observations by a number of investigators have shown through immunoprecipitation and fluorescence studies that PF3D7-1436300 expresses a putative protein (PTEX150) in association with HSP101 and other proteins (Bullen et al. 2011; de Koning-Ward et al. 2009 and Crabb et al. 2010).

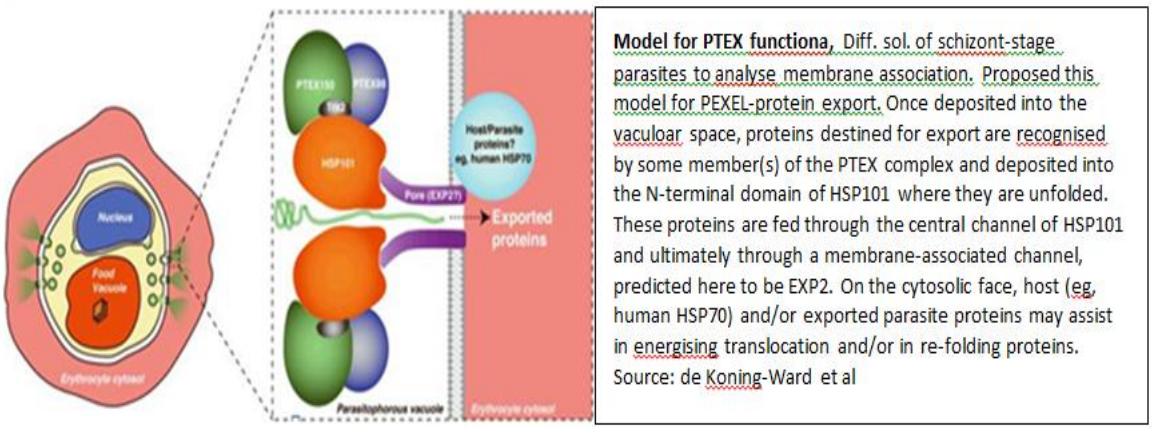


Fig. 4 Model for Translocon (PTEX150) function de Koning et al. Nature 2009

There are seven orthologs and paralogs found in Plasmodb for *P. berghei*-ANKA, *P. chabaudi*-AS, *P. chabaudi*-YB, *P. falciparum*-IT, *P. knowlesi*-H, *P. vivax* and *P. yoelii*-YM species. PTEX150 has an elongation factor hand 1 (EF-Hand1) binding domain, with accession number IPR018247 (Plasmodb). This type of domain consists of a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bi-pyramidal configuration (Plasmodb).

Protein Expression Data

Affymetrix studies using RNA isolated from gametocytes and sporozoites using temperature and sorbitol synchronization by both Derisi (2003) and Winzelar (2004) showed that the parasite expresses the translocon soon after invasion.

Bioinformatic analysis of Ribonuclease binding-like (RNB-like) Gene PF3D7_0906000

The Ribonuclease binding-like (RNB) gene expresses a putative RNB-like protein found in all species of *Plasmodium*. The catalytic domain of ribonuclease (GenBank: CAX64197.1) protein and its coding gene called Pf3D7_09_v3, is located on chromosome 9. It is also found in other members of the phylum such as *Theileria, Babesia* and *Cryptosporidium.* The parasites with this gene have an RNB-like domain containing putative RNAse II enzyme activity. The RNAse II family is a group of enzymes that possesses single strand RNA exo-ribonuclease activity. The RNB-like putative activity of this protein is thought to be present in *P. falciparum*, *P. berghi*, *P. chabaudi* and *P. yoelii* (Plasmodb). RNB-like (PF3D7_0906000) gene has two exons

comprising of 2863 bp and 2527bp, and one intron of 120 bp, making a total of 5515 bps. RNB-like protein is annotated to possess four antibody epitopes (a) 8646-DIGDIVRGKDLY, (b) 73540- YDELLKRKENELF, (c) 26850-IKND (d) 44554-NKND]. The RNB-like gene has six to eight orthologs and paralogs (Plasmodb).

Protein data from GeneDB reveals that the protein has a peptide sequence length of 1797 amino acids with isoelectric point (IP) of 8.9 and molecular weight of 213.8 kDa. RNB-like protein is synthesized by both the sporozoite and merozoite stages of the parasite (Florence et al. 2002). The amino acids comprising RNB-like domain is located between 869 – 1322 residues with no predicted transmembrane domain at this region (Appendix 5. GeneDB). These domains are not located at the same position of the peptides sequence in either the *Plasmodium* species or the other members of the apicomplexa.

The RNB-like protein is still described as a hypothetical protein in several species of *Plasmodium* such as *P. vivax* and *P. chabaudi* but a putative rhoptry protein in *P. yoelii* and *P. falciparum*. The Ribonuclease binding-like (RNB) gene is thought to express an RNB-like putative protein found in most species of *Plasmodium* including *P. vivax* and *P. falciparum.* It is also found in other members of the phylum such *as* in *Theileria, Babesia* and *Cryptosporidium.* These parasites have an RNB-like domain containing putative RNAse II enzyme activity with a classification number of EC 3.1.13.1. The RNAse II family is a group of enzymes that processes single strand RNA exo-ribonuclease activity (Plasmodb). In addition to the RNB-like domain, *Cryptosporidium parvum* has an extra domain that is annotated as RNB-B which has not been characterized (Fig. 5).

Protein Expression Data

The RNB-like gene located on chromosome 9 has two domains. RNB-like gene has two exons (2863bp-exon 1 & 2527bp exon2) and one intron of 120bp making a total of 5515bps. Protein data from Genedb reveals that the encoded protein has a peptide sequence length of 1797 amino acids with an isoelectric point of 8.9 but Plasmodb annotates a PI value of 9.17 and mass of 213.8kDa respectively. The RNB domain is located at 869 – 1322 (ILGNENNFFGIIYLFLYLYKIHYYIYKIT) with no predicted TM regions at this region.

Appendix 5 shows an RNB-like segment showing amino acid alignment with *Plasmodium* species and some members of the phylum apicomplexa such as *Theileria, Babesia* and *Cryptosporidium*. The RNB-like domain has a conserved amino acid sequence of HFTSPIRRY in *P. yoelii, P. falciparum* and *E. coli* and not much is known about the function of this sequence (NCBI, CDAR). Appendix 6 shows an RNB segment of amino acid sequence homology in *Plasmodium* and *E. coli* (GeneDB).

All the enzymes included in the Ribonuclease binding-like proteins are large polypeptides sharing sequence identity and distinctive conserved motifs, suggesting the presence of functional domains. The N-terminal region of Ribonuclease-like proteins are the most variable in both length and sequence among all proteins of the family, but sequence analysis predicts that most of them contain one or more putative RNA binding domains. Sequence homology analysis reveals that the N-terminal region of *E. coli* RNase II contains a sequence similar to the cold shock domain (CSD). This domain has been described as a single-stranded nucleic acid binding domain that functions as an RNA chaperone in bacteria and is involved in regulating translation in eukaryotes

(Ambler et al. 2006). RNase II sequence found in *E. coli* is shown in appendix 6. Comparison between the RNB-like domains in *E. coli* and the phylum apicomplexa could not be made because of limited data on the nature of the RNB-like domain in the apicomplexa, especially in the *Plasmodium* species. There are no orthologs predicted among the sequences of *Plasmodium* species and *E. coli*. But, in silico analysis revealed a sequence similarity of the following residues, HFTSPIRRYPD at different positions on *Plasmodium species* and *E. coli* (Appendix 6). Based on the information obtained from Plasmodb, gene PF3D7_0906000 is annotated to express a product involved in ribosome assembly and other metabolic activities (Plasmodb). The ribosome processing activity is thought to act on some components of the ribosomes such as SnoRNP, 90S and 60S particles. PF3D7_0906000 (PfI0295c) is thought to influence the overall quality control of mRNA biogenesis (Plasmodb).

List of PF3D7_0906000 Protein Domain Architectures

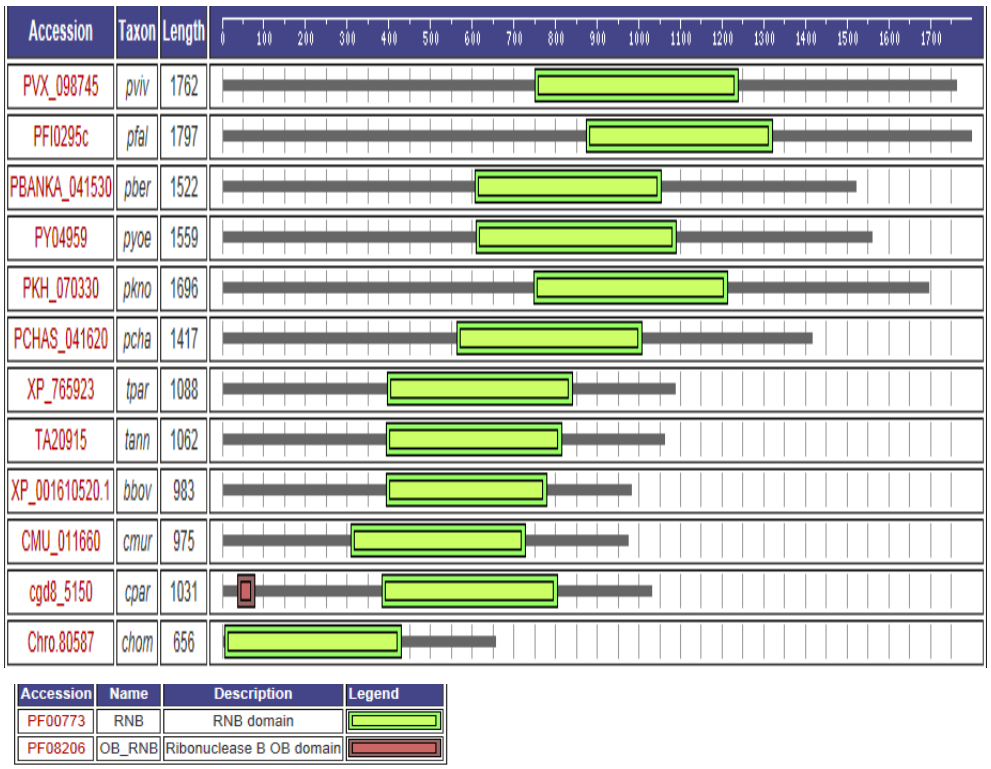


Fig. 5 An RNB domain architecture seen in *Plasmodium* and some members of the phylum, apicomplexa. Source: Plasmodb Conserved domain architecture

The RNB-like domain architecture is clustered into three broad categories based on the alignment shown above as annotated in Plasmodb (Fig. 5). There are variations both within *Plasmodium* species and within the members of the phylum. Within the members of the genus *Plasmodium*, some domains are clustered towards the N-terminus, some are in the middle while others are clustered towards the C-terminus of the protein. In *Cryptosporidium parvum, C.* hominis and in all members of the *Plasmodia* that have the gene, there are noticeable differences. The RNase domain in *Plasmodium* species has amino acid residues 350 to 1300 while in *Cryptosporidium*, the residues ranges from 10 to 700 of the RNB-like protein. *Cryptosporidium homonis* also has an extra domain which is described as ribonuclease-B (RNB- OB) domain. Although not shown in Fig. 5,

P. knowlesi (H-strain) also possess the RNB-OB domain B3L7Q0_PLAKH (OB_RNB (PF08206). The RNB domain has been described to have an important function of RNA editing. However, this enzyme has not been extensively investigated in *Plasmodium* or other members of the phylum apicomplexa. Recent research has been performed on this enzyme in *E. coli* (Ambler et al. 2006). Appendix 8 compares RNB-like protein identity between *P. yoelii* (XP_723884), *P. falciparum* (XP_ 002808916) and *E. coli* (ACI73414).

Results of Cloning and Expression of Translocon (PF3D7_1436300) gene

Genomic DNA was isolated from *P. falciparum* strains Dd2, 3D7 and FCR3, according to the standard protocol described in materials and methods. The isolated DNA from the three strains was analyzed on a 0.7% agarose gel. A high molecular weight DNA that migrated well above the standard marker of 18kb was obtained. There was minimal shearing of the DNA indicating that the extraction method was successful (Figure 6A). The isolated DNA was used as a template for PCR amplification of translocon gene PF3D7_1436300 using the designed translocon primers shown in Table 2.

The amplified gene products from Dd2 and FCR3 were size fractionated on a 1.0% agarose gel. The result showed a single band with the expected size of 600 bp and the negative control (H_2O) showed no DNA bands (Figure 6B). The 600bp product was purified from the gel and used for digestion with appropriate enzymes and cloned into the plasmid vector pT7CFE-Chis (Thermo Scientific Pierce, Rockford, IL 61105 USA). The PCR amplified product of the cloned gene is shown in figure 6C. This cloned product was then used for transcription and translation to produce the translocon protein. The transcription and translation was carried out in vitro in a cell free expression system using both the two step and the single step protocol. The in vitro translated (IVT) reaction

product was analyzed using a western blot. The expressed protein was fractionated on a 10% SDS-PAGE gel and analyzed by western blotting using two antibodies and normal rabbit serum (NRS) for negative control. Antiserum #676 is a whole rhoptry-specific antibody and #685 is a SERA-specific antibody. The translated protein fragment of molecular weight 23kDa was detected by antiserum #676 (Figures 7).

RNB-LIKE gene Amplification

The RNB-like gene was also PCR-amplified from *P. falciparum* (strains Dd2, 3D7, FCR3, HB3, K1, D10), *P. yoelii* and *P. berghei* using the *P. falciparum* primers. The genomic DNA from these strains were used as the source of template DNA for amplification of the RNB-like gene. The expected RNB-like gene fragment of 5.5 kb was detected using two sets of primers as shown in Table 3. These amplified DNAs were also gel purified according to the protocol stated in the materials and methods section. However, the DNA could not be cloned and expressed because the DNA was lost during the gel purification process. Initial attempts at amplification with a different set of *P. falciparum* primers was unsuccessful.

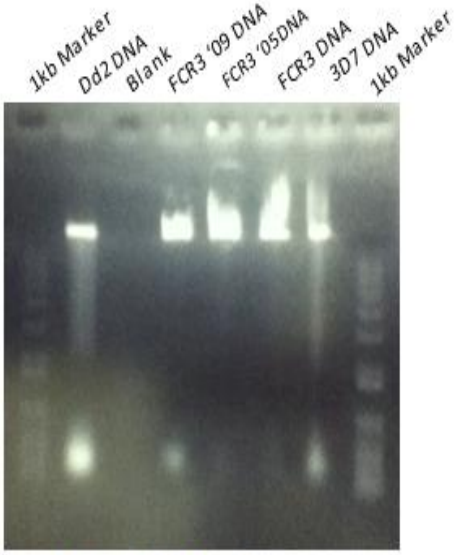


Fig. 6A Genomic DNA on 0.7% agarose

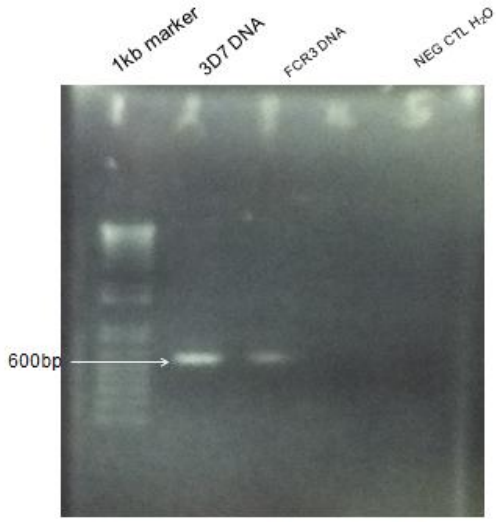


Fig. 6B Amplified 600bp translocon gene fragment from 3D7 and FCR3. 1% Agarose gel

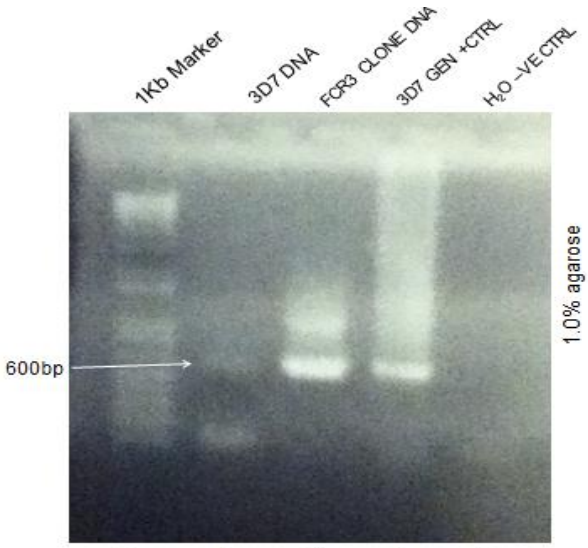


Fig. 6C Purified Clone of Dd2 and FCR3 products

Figures 6. (A) Genomic DNA isolation from three *Plasmodium falciparum* strains, Dd2, FCR3 and 3D7 with minimal shearing. **(B)** 1.0 % agarose gel of PCR amplified PF3D7_1436300 fragments. The expected 600bp fragment of DNA was amplified from

3D7 and FCR3 strains. **(C)** 1.0 % agaorse gel of a cloned fragment of 600bp from Dd2 and FCR3 in vector pT7CFE-Chis.

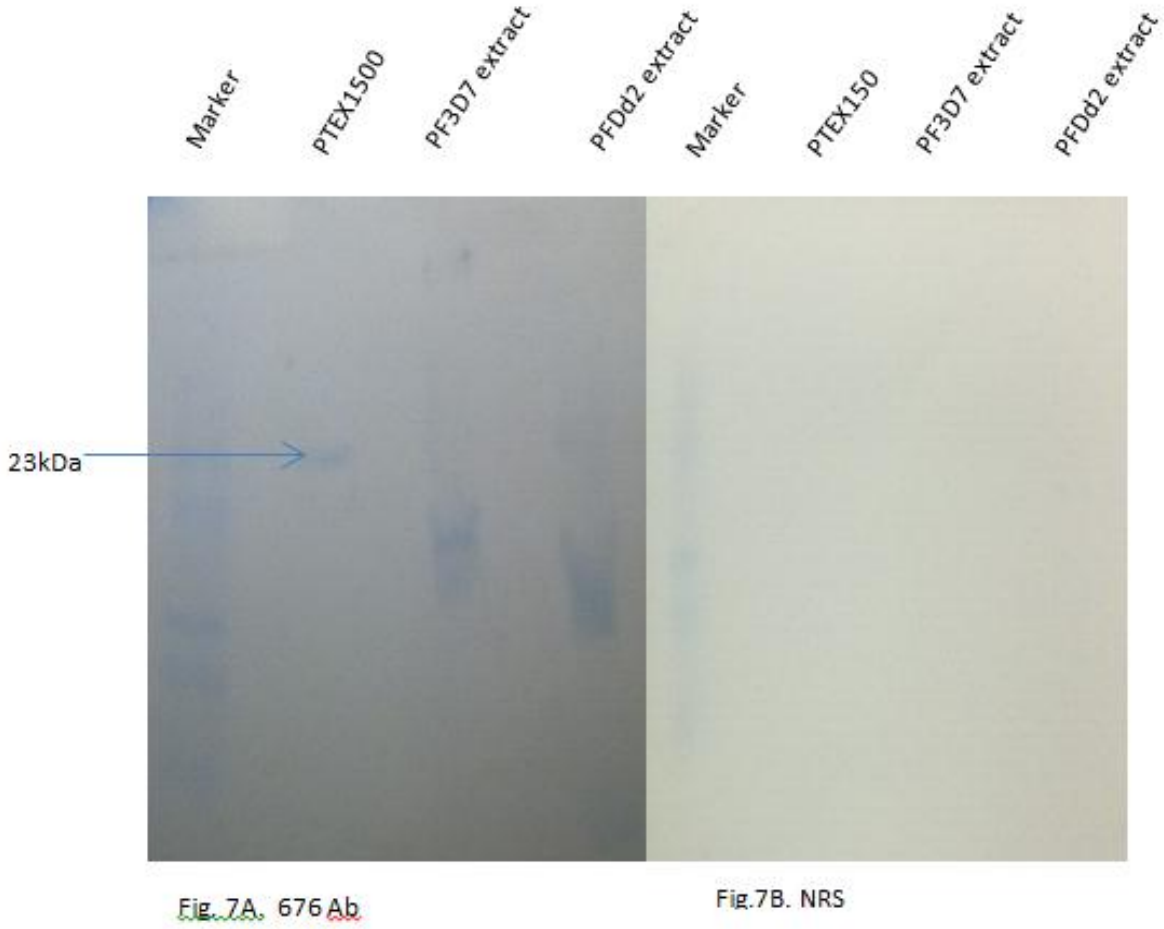


Figure 7. Western blot analysis of expressed 23kDa translocon protein detected with rhoptry specific antibody 676. The normal rabbit serum did not react with the translated protein.

CHAPTER IV

DISCUSSION

This study focuses on the *in silico* analysis and expression of translocon gene PF3D7_1436300 and RNB-like gene PF3D7_0906000. Both genes were selected from proteomics analysis of merozoite rhoptries (Sam-Yellowe et al. 2004) and were analyzed for their characteristics. A database search on *P. falciparum* genes PF3D7_1436300 and Pf3D7_0906000 was carried out using the *Plasmodium* data base (Plasmodb), PSORTb(www.psortb.org), SWISSPROT-workspace (www.expasy.org), GeneDB (www.genedb.org), National Center for Biotechnology Information (NCBI), ExPASy and COBALT software. The information obtained from searches of Plasmodb and GeneDB on PF3D7_1436300 gene indicated it has conserved sequences with its orthologs in all *Plasmodium* species. PF3D7_1436300 was described as a putative translocon protein in human parasites *P. falciparum* and *P. vivax*. PF3D7_1436300 encodes a protein described as the PTEX150 component and a cotransporter of the parasite export complex (Translocon) (de Koning-Ward et al. 2009, 2011). Data from Plasmodb also indicates that the protein expressed by PF3D7_1436300 gene has a molecular weight of 112 kDa, has six orthologs and paralogs in *P. berghei-ANKA, P. chabaudi, P. knowlesi-H, P. vivax-Sal-1, P. yoelii-17XL* and *Theileria parva-Muguga* (OrthoMCLDB OG5_161974). Despite the annotation that it is highly conserved in *Plasmodium s*pecies and some

members of the phylum apicomplexa, its protein identity is as low as 43% within the phylum. The conservation is not obvious. The amino acid similarity found in the proteins in terms of type, number and position does not seem to be the same. The conservation could be in the structural motifs or domain organization. Attempts at determining the 3D structure of the protein using PSORTb was unsuccessful.

Also, the exact mechanism of action of PTEX150 as a member of the translocon complex is not yet clear (Riglar et al. 2013, de Koning-Ward et al. 2009, 2012), but immunofluorescent analysis (IFA) using antibodies against recombinant PTEX150, HSP101, EXP2 or haemagglutinin (HA), showed that PTEX150, HSP101 and EXP2 are synthesized early soon upon merozoite invasion and stored in the dense granules (Riglar et al. 2013; de Koning-Ward et al. 2009, 2012). PTEX150 protein colocalizes to the PV membrane. (Hiller et al. 2004; Marti et al. 2004; de Kong-Ward et al. 2009, 2012). Also, EXP2 has already been characterized in *P. falciparum* and it is associated with the PV (de Koning-Ward et al. 2009). EXP2 has also been investigated using immune precipitation and shown to interact with the PTEX150 complex (de Koning-Ward et al 2009). Studies through microarrays, pull down studies using antibodies against HSP101 and PTEX150 (3D7-101HA and 3D7-150HA) and genetic (using transgenic parasites) methodologies in combination with predictive criteria (FUGUE Software), confocal and crystallization (Thioredoxin-2 (PfTrx-2), (Sharma et al. 2011), for components of export machinery in *P. falciparum*, revealed the presence of PTEX150 which plays a similar role as PEXEL (de Koning-Ward et al. 2012). The *P. falciparum* translocon complex consists of five discrete components within the PVM (Bullen et al. 2011; de Koning-Ward et al. 2012; Rigla et al. 2013). One of these five components is PTEX150, though its role is still

speculative. PTEX150 is postulated to have an enzymatic role of ATPase, belonging to the HSP101, a ClpA/B-like AAA+ ATPase subgroup commonly associated with translocon proteins (de Koning-Ward et al. 2012; Rigla et al. 2013). The PF3D7_1436300 (PF14_0344) gene was first described in *P. yoelii* (PY02301) (Sam-Yellowe et al. 2004) and its orthologs and paralogs were subsequently detected in all *Plasmodium* species and also described in another member of the phylum apicomplexa, *Theileria parva muguga* (Plasmodb). It is not clear if the role of PTEX150 is similar or different in all the species of *Plasmodium* and other members of the phylum. The precise transport mechanism by *Plasmodium* parasites, especially *P. falciparum* to export proteins that are essential to its survival and virulent role has been an intense area of research. Knowing the proteins participating in transport across the PVM will open up avenues for developing new antimalarial drugs, vaccines or diagnostic markers. How these translocon proteins from the Golgi reach the PVM and the erythrocyte cytosol and possibly onto the erythrocyte membrane is also an area of active research. There are still many mechanistic questions concerning the outcome of the PEXEL proteins in the ER and Golgi after their acetylation and subsequent modifications. It is still unclear how the PTEX machinery functions to divert proteins from the ER to the PVM and beyond. Current models have been proposed by de Konig-Ward et al. (2010). They proposed three possibilities: 1) The Barcode, (2) Chaperone and (3) Regional models. The barcode model proposes that the PEXEL-containing proteins simply traffic the proteins to the vacuole in bulk flow with non-exported proteins using normal endomembranous transport system. The cleaved PEXEL protein motif x/E/Q/D is specifically recognized by a member of the PTEX translocon complex which would then be unfolded and fed through

the translocon pore across into the erythrocyte cytosol. How the translocon complex recognizes the x/E/Q/D motif is still not clear (de Koning-Ward et al. 2010). The chaperone model states that a pexalase cleaves PEXEL in the ER and is closely followed by the recruitment of a chaperone-like protein or some other yet-to-be identified molecule(s) or moiety to the PEXEL protein thereby tagging it for export once secreted into the PV via the normal secretory pathway (de Koning-Ward et al. 2013). Both the barcode and the chaperone models could be mutually inclusive. The regional model proposes that once the PEXEL proteins are cleaved, they are directed and ferried in vesicles down a specific and distinct trafficking pathway that targets the protein to a specific region of the parasite membrane that is in contact or in proximity with the PTEX complex where they fuse (de Koning-Ward et al. 2013). Proteins destined for export will be deposited into a restricted vacuole close to the PTEX machine and it excludes the non-exported proteins. It is proposed in this model that recognition of the pexelase proteins by the PTEX complex is less specific. Both the PTEX complex and PEXEL cargo proteins would be expected to localize in the parasite plasma membrane apposed to the PV. (Wickham et al. 2001; de Koning-Ward 2009). However, it is yet to be determined whether the PEXEL proteins and the PTEX model overlap or not. The regional hypothesis does not exclude the possible involvement of the other export models described (de Koning-Ward et al. 2013). Proteins that bridge the contacts between the export vesicles, the parasite plasma membrane and the PVM have not yet been described. The PTEX150 translocon protein is described as an extracellular protein (PSORTb**)** and is annotated to have four kinase family motifs and 3 putative phosphorylation sites (www.prosite.org) in *P. berghei* and *P. berghei- ANKA*. Transcripts of the proteins are

seen in post invasive stages of the parasite (Derisi et al. 2003; Winzelar et al. 2004). PTEX150 protein is described to have a signal sequence of RxLxE/Q/D (Hiller et al. 2004; Marti et al. 2004) and an elongation factor hand 1 (EF- Hand1) binding domain that is not conserved (CDART) NBCI retrieval system. The presence of ubiquitous, highly conserved consensus sequences of asparagine, lysine and glutamic acid residues throughout the translocon protein in all the *Plasmodium* species and other members of the phylum apicomplexa raises some interesting questions about gene PF3D7_1436300. Why do *Plasmodium* species and some members of the phylum, apicomplexa such as *Theileria P. muguga* have the gene and others do not? The parasite depends on a set of enzymes known as plasmepsins (Plasmepsin V), (Goldberg and Cowman 2010; Russo et al. 2010), which are aspartic acid proteases that are used to degrade hemoglobin (Jortzik et al. 2011) and necessary for effective transport of products of metabolism in *Plasmodium* parasite (Boddey et al. 2013). Also, the presence of residues such as asparagine, serine and glutamic acids are crictical in post-translational modification such as glycosylation, an event that may be crictical to the parasite survival or its pathogenesis.

The fact that the PTEX150 protein is expressed early post-invasion and it is expressed extracellularly (Derisi et al. 2003; Winzelar et al. 2004) makes it a suitable target for vaccine or therapeutic studies. As the search for new, efficient and cost-effective malaria control, prevention, diagnosis and treatment continues, the significance of a clear understanding of the invasion mechanisms and how the parasite evades or mounts resistance to the host immune system cannot be over-emphasized.

Bioinformatic analysis was carried out on a second gene PF3D7_0906000, an RNB-like *Plasmodium* rhoptry gene identified from previous rhoptry proteome studies in *P. yoelii, P. falciparum* orthologs and paralogs were identified in the *Plasmodium* data base, Plasmodb. Additional searches were performed and the results showed that the *P. falciparum* protein is annotated as an RNB-like putative protein in *Plasmodium* species, *Theileria p. muguga* and *Cryptosporidium parvum.* PF3D7_0906000 gene is located on chromosome 9, is 5.5kb, has two exons and one intron. The RNB-like gene has seven orthologs and paralogs. These are found in *P. berghei*-ANKA, *P. chabaudi*-AS, *P. cynomolgi* -B, *P. falciparum*-IT, *P. knowlesi*-H, *P. vivax*-V and *P. yoelii*-YM strains of *Plasmodium.* The protein is identified as a putative or hypothetical protein in these species (Plasmodb). PF3D7_0906000 is described to have an RNB-like putative domain similar to the exoribonuclease RNAse II family in *P. berghei*, *P. chabaudi* and *P. yoelii* with an enzyme annotation number of EC 3.1.13.1. The RNB-like putative protein has a molecular weight of 213 kDa and IP of 8.9 with peptide sequence of about 1800aa and an RNB-like domain with no predicted TM domain. There are, at least three members of RNB domains, RNB R, RNB exonuclease and RNase II. No study on the characterization of *Plasmodium* RNB -like functional domains has been documented. However work done with *E. coli* RNase by Ambler et al. (2006) suggests that RNase II is an exonuclease. Sequence prediction shows domain conservation among the *Plasmodium* species and other members of the phylum apicomplexa. The nature of the conservation is not obvious. What is clear is that the domain is present in all the *Plasmodium* species and in some other members of the phylum. The presence and significance of the RNB-like domain in *Plasmodium* parasites and other members of the phylum apicomplexa will

provide some insights into how *Plasmodium* invades and remodels its host cell upon invasion. Studies performed on *E. coli* RNase-II like enzymes show the enzymes behave similarly in both prokaryotes and eukaryotes, Ambler et al. (2006). Whether the same can be said of the RNase II domain in *Plasmodium* and other members of the phylum remains to be verified. The RNB-like protein identity among members of the *Plasmodium* species and other members of the phylum apicomplexa was 33% (NBCI- CD, OrthoMCL DB). Mass spectroscopy evidence shows that the RNB-like protein is found in nuclear and cytoplasmic fractions of trophozoites and schizonts (Derisi et al. 2003 and Winzelar et al. 2004). The affymetrix studies using temperature control and sorbitol synchronization by both Derisi et al. (2003) and Winzelar et al. (2004) showed that the transcriptome appears shortly upon merozoite invasion and peaks at late trophozoite (Plasmodb).

In this study, genomic DNA was isolated from several *P. falciparum* strains (Dd2, 3D7, FCR3, HB3, K1 and D10). Two of the genomic DNA (3D7 and FCR3) samples were used for PCR amplification of translocon (PTEX150) and RNB-like genes using PCR. The PCR products were extracted and purified. The 600bp translocon gene fragment was cloned and expressed three times using the human DNA based two-step and single-step in vitro cell free expression systems. The RNB-like gene was cloned. However, expression of the protein was not successful. Western blot analysis of the tranlocon-like protein was performed and an expected 23kDa protein was recognized by a specific rabbit anti-whole rhoptry antibody, #676 (Sam-Yellowe et al. 1995) (fig. 7). Normal rabbit serum did not react with the expressed protein (Figure 7) Protein expression system

Wheat germ and rabbit reticulocyte cell free-expression systems vary in their post translational modifications, length of time required to obtain expressed protein and protein yields. These systems are labor intensive and expensive. However, these systems are currently used for expression of malaria proteins. The HeLa cell lysate based in vitro cell-free protein expression system was used in this study to express translocon (PTEX150). The in vitro human cell-free protein expression system can also be used as an alternate approach for the synthesis of malaria proteins for diagnostic and vaccine studies. This system has been used to express *Plasmodium* rhoptry proteins in previous studies (Yadavalli et al. 2012). In future studies, the RNB-like gene will be cloned and expressed. The primers will be optimized to facilitate successful amplification and cloning. The expressed product will be tested for ribonuclease activity to confirm the active status of the expressed enzyme.

SUMMARY

In Silico data mining demonstrated that, *Plasmodium* genes PF3D7_1436300 (PTEX150) and PF3D7_0906000 (RNB-like), which are highly conserved in *Plasmodium* species and other members of the apicomplexan phylum have important characteristics and properties that can be further investigated for potential as possible targets for vaccine and drug development. The early secretion of both PTEX150 and RNB-like proteins soon after merozoite invasion, as well as their persistence throughout the blood stage suggests that both proteins may play a role in the parasite's ability to invade and establish infection in the red blood cells. Taken together, both genes and their protein products could serve as useful vaccine candidates and drug targets because they are highly conserved in all *Plasmodium* species, particularly those species infecting humans.

Future studies

PTEX150 and RNB-like putative proteins will be expressed using in vitro cell-free protein expression systems. The expressed proteins will be purified and characterized. The purified proteins will also be used for immunization studies to produce antibodies for further characterization of both proteins.

REFERENCES

Adams JH, Miller LH, Aikawa M (1989) Release of Merozoite Dense Granules during Erythrocyte Invasion by *Plasmodium knowlesii*. Infect. and Immuno. 57: 3230-3233

Aikawa M, Hsiao LL, Howard RJ, Taraschi TF (1991) Modification of host cell membrane lipid composition by the intra-erythrocytic human malaria parasite *Plasmodium falciparum*. Biochem J. 274: 121–132

Angrisano F, Riglar DT, Sturm A, Volz JC, Delves MJ (2012) Spatial Localization of Actin Filaments across Developmental Stages of the Malaria Parasite. PLoS ONE 7:e32188

Ambler M, Barbas A, Fialho AM, Arriano CM (2006) Characterization of the Functional Domains of *Escherichia coli* RNase II J. Mol. Biol. 360: 921–933

Autino B, Alice N, Rosario R, Francesco C (2012) Epidemiology of Malaria in Endemic areas Mediterr J Hematol Infect Dis 4: e2012060

Berlin H, Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacci JB, Tabb DL, Geeves MA, Fedorov R, Manstein DJ (2005) Molecular mechanism of actomyosin-based motility Cell Mol Life Sci. 62: 1462-1477

Besteiro S, Lamarque M, Papoin J, Roquest M, Vulliez-Le N, Morlon-Guyot J, Dubremetz, J, Fauquenoy S, Tomavo S, Faber BW, Kocken CH, Thomas AW, Boulanger MJ, Bentley GA, Lebron M (2009) Apicomplexan Parasites. PLoS Pathog 7: e1001276

Besteiro S, Michelin A, Poncet J, Dubremetz JF, Lebrun M (2009) Export of a *Toxoplasma gondii* rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion. PLoS Pathog. 5: e1000309

Boddey JA, Carvalton TG, Hodder AN, Sargeant TJ, Sleebs BE, Marapana D, Lopaticki S, Nebl T and Cowman AF T (2013) Traffic Role of Plasmepsin V in Export of Diverse Protein Families from the *Plasmodium falciparum* Exportome 14:532-550

Bullen HE, Chanaud SC, Kalanon M, Riglar DT, Dekiwadia C, Kangwanrangsan N, Torii M, Tsuboi T, Ralph AS, Baun J, Cowman AF, de Knoning-Ward TF, Crabb, BS, Gilson PR (2011) Biosynthesis, Localization, and Macromolecular Arrangement of the *Plasmodium falciparum* Translocon of Exported Proteins (PTEX) J Biol Chem. 287:7871-7884

Calton J, Silva J, Hall N (2012) The Genome of Model Malaria Parasites, and Comparative Genomics Curr. Issues Mol. Biol. 7: 23-38

Chang HH, Falick AM, Carlton PM, Sedat JW, DeRisi JL, Marletta MA (2010) N-terminal processing of proteins exported by malaria parasites Mol Biochem Parasitol 171: 25–31

Chipawaza B, Okumu FO, Madumla EP, Mbeyela E, Lingamba G, Moore J, Ntamatungro AJ, Kavishe DR, and Moore SJ (2012) Implications of bio-efficacy and persistence of insecticides when indoor residual spraying and long-lasting insecticide nets are combined for malaria prevention Malar J 11:378

Chitnis CE, Sharma A (2008) Targeting the *Plasmodium vivax* Duffy-binding protein Rev Trends in Parasitol 24:29-34

Cowman AF, Berry D, and Baum J (2012). The cellular and molecular basis for malaria parasite invasion of the human red blood cell. J Cell Biol. 198: 961-971

Crabb BS, de Koning-Ward TF, Gilson PR (2010) Protein export in *Plasmodium* parasites: From the endoplasmic reticulum to the vacuolar export machine. Int J of Parasitol 40: 509-513

Cruz LN, Juliano MA, Budu A, Juliano L, Holder AA, Blackman MJ, Garcia CR (2012) Extracellular ATP triggers proteolysis and cytosolic Ca^{2+} rise in *Plasmodium berghei* and *Plasmodium yoelii* malaria parasites. Malar J 11: 11-69

De Koning-Ward TF, Gilson PR, Boddey JA, Rug M, Smith BJ, Papenfuss AT, Sanders PR, Lundie RJ, Maier AG, Cowman AF, Crabb BS (2011, 2012) PLoS Pathogens 7:1-14

De Koning-Ward TF, Gilson PR, Boddey JA, Rug M, Smith BJ, Papenfuss AT, Sanders PR, Lundie RJ, Maier AG, Cowman AF, Crabb BS (2009) A novel protein export machine in Malaria parasites. Nature 18: 459-949

Deepak G (2004) Parasite ligand–host receptor interactions during invasion of erythrocytes by *Plasmodium* merozoites. Int J for Parasitol 34:1413-1554

Deitsch K W and Wellems T E (1996) Membrane modifications in erythrocytes parasitized by *Plasmodium falciparum* Mol Biochem Parasitol 76 (1-2):1–10

Dluzewski AR, Zicha D, Dunn GA, Gratzer WB (1995) Origins of the parasitophorous vacuole membrane of the malaria parasite: surface area of the parasitized red cell. Eur J

Cell Biol 68: 446-449

Dowse TJ, Koussis K, Blackman MJ, Soldati-Favre D (2008) Roles of Proteases during Invasion and Egress by *Plasmodium* and *Toxoplasma* Subcell Biochem 47: 121-139

Francis SE, Sullivan DJ Jr, and Goldberg DE (1997) Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. Annu Rev Microbiol 51: 97-123

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum* Nature 419:498-511

Garnham PCC (1966). Malaria Parasites and Other Haemosporidia. Oxford: Blackwell Science Ltd. ISBN 0-632-01770-8.

Gilson PR, Crabb BS (2009) Morphology and kinetics of the three distinct phases of red blood invasion by *Plasmodium falciparum* merozoites Int J Parasitol 39:191– 196

Goel VK, Li X, Chen H, Liu SC and Christi AH (2003) Band 3 is a host receptor binding merozoite surface protein 1(MSP-1) during *Plasmodium falciparum* invasion of erythrocytes. Proc Natl Acad Sci USA 100: 5164–5169

Goldberg DE and Cowman AF (2010) Moving in and renovating: exporting protein from *Plasmodium* into host erythrocytes. Nat Rev Microbiol 8:617-621

Gruber A, Gruber G, Manimekalai MSS, Preiser PR (2012) Structural architecture and interplay of the nucleotide- and erythrocyte binding domain of the reticulocyte binding protein Py235 from *Plasmodium yoelii*. Int J of Parasitol 42: 1083–1089

Gupta DK, Gupta N, Jahan N (2011) New records of *Haemoproteus* and *Plasmodium* (Sporozoa: Haemosporida) of rock pigeon (*Columba livia*) in India. J Parasit Dis. 35: 155–168

Hadley TJ, Klotz FW, Pasvol G, Haynes JD, McGinniss MH, et al. (1987) *Falciparum* malaria parasites invade erythrocytes that lack glycophorin A and B (MkMk): Strain differences indicate receptor heterogeneity and two pathways for invasion. J Clin Invest 80: 1190–1193

Haeggström M, Fred K, Klaves B, Qijiun C, Mats W, Victor F (2004) Common trafficking pathway for variant antigens destined for the surface of the *Plasmodium falciparum-infected* erythrocyte Mol Biochem Parasitol 133: 1-14

Hanssen E, Sought R, Frankland, Deed S, Klonis N, Jennifer LS, Tilley L (2008) Electron tomography of the Maurer's cleft organelles of *Plasmodium falciparum*-infected erythrocytes reveals novel structural features. Mol. Microbiol. 67:703- 718

Haussig JM, Matuschewski K and Kooij TW (2011) Inactivation of a *Plasmodium* apicoplast

protein attenuates formation of liver merozoites Mol Microbiol 81: 1511–1525

Hiller NL, Bhattacharjee S, van Ooij C, Liolios K, Harrison T, Lopez-Estrano C, Haldar K (2004) A host-targetting signal in virulence proteins reveals a secretome in malaria infection. Science 306: 1934-1937

Hossain ME, Dhawan S, Mohmmed A (2012). The cysteine-rich regions of *Plasmodium falciparum* RON2 bind with host erythrocyte and AMA1 during merozoite invasion. Parasitol Res. 110:1711-1721

Iyer J, Gruner AC, Renia L, Snounou G, Preise PR (2007) Invasion of host cells by malaria parasites: A tale of two protein families Mol Microbiol 65: 231–249

Karunaweera ND, Wanasekara D, Chandrasekharan V, and Carter R (2007)

Plasmodium vivax: paroxysm-associated lipids mediate leukocyte aggregation
Malar J 6:62

Karunaweera ND, Grau GE, Gamage P, Carter R, Mendis KN (1992) Dynamics of fever and serum levels of tumor necrosis factor are closely associated during clinical paroxysms in *Plasmodium vivax* malaria Proc Natl Acad Sci 89:3200-3203

Kilejian A (1979). Erythrocyte Changes Induced by *P. falciparum*

Proc. Natl. Acad. Sci. 76: 4650-4653

Kimura A, Dhondt AA, Lovette IJ (2006) Phylogeographic Structuring of *Plasmodium* Lineages across The North American range of the Houses Finch (Carpodacus)
J Parasitol 92:1043-1049.

Kirchgatter K and Del Portillo HA (2005) Clinical and molecular aspects of severe malaria Ann Braz Acad Sci 77: 455-475

Lanzer M, Hannes W, Georg K, Laetitia V, Catherine B B (2006) Maurer's clefts: A novel multi-functional organelle in the cytoplasm of *Plasmodium falciparum*-infected erythrocytes Intl J of Parasitol 36: 23–36

Luchavez J, Espino F, Curameng P, Espina R, Bell D, Chiodini P, Nolder D, Sutherland C, Lee KS, and Singh B (2008) Human infections with *Plasmodium knowlesi*, the Philipines. Emerg Infec Dis 14:811-813

Maier AG, Duraisingh MT, Reeder JC, Patel SS, Kazura JW, et al. (2003) *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. Nat Med 9: 87–92.

Markus, MB; Krotoski WA (2011) Malaria: Origin of the Term "Hypnozoite" J Hist of Biol 44:781-786

Marti M, Good RT, Knuepfer E, Cowman AF (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science 306:1930-1933

Miller LH (1977) Evidence for differences in erythrocyte surface receptors for the malarial parasites, *P. falciparum* and *P. knowlesi* JEM 146: 277-281

Osborne AR, Speicher KD, Tameza PA, Bhattacharjeea S, Speicher DW, Haldara K (2010)

The Host Targeting motif in exported *Plasmodium* proteins is cleaved in the parasite endoplasmic reticulum Mol Biochem Parasitol 171: 25-31

Rayner JC, Triglia T, Duraisingh MT, Ralph SA, Barnwell, JW, McFadden GI, Cowman AF (2003) Phenotypic variation of *P. falciparum* merozoite proteins directs receptor targeting for invasion of human erythrocytes The EMBO J 22:1047 - 1057

Russo I, Babbitt S, Muralidharan V, Buttler T, Oksman A, and Goldberg DE (2010) Plasmepsin V licenses *Plasmodium* proteins for export into the host erythrocyte Nature 463:632-636

Sam-Yellowe TY, Florens L, Wang T, Raine JD, Carucci DJ, Siden R, Yates JR (2004) Proteome Analysis of Rhoptry-Enriched Fractions Isolated from *Plasmodium* Merozoites J Prot Res *3*: 995–1001

Sam-Yellowe TY, Fujioka H, Aikawa M, Messineo DG (1995) *Plasmodium falciparum* rhoptry proteins of 140/130/110 kd (Rhop-H) are located in an electron lucent compartment in the neck of the rhoptries J Euk Microbiol 42:224-231

Sam-Yellowe TY (2009) The role of the Maurer's clefts in protein transport in *Plasmodium falciparum* Trends Parasitol 25:277-284

Santos JM, Soldati-Favre D (2011) Invasion factors are coupled to key signaling events leading to the establishment of infection in apicomplexan parasites. Cell Microbiol 13: 787-796

Sharma A, Ashwani A, Amita A, Dixit S (2011) Structural insights into thioredoxin-2: a component of malaria parasite protein secretion machinery Sci Rep 1:179

Singh S, More KR, Chitnis CE. (2013) Role of calcineurin and actin dynamics in regulated secretion of microneme proteins in *Plasmodium falciparum* merozoites during erythrocyte invasion DOI: 10.1111/cmi.12177

Sivaraj S and Spudich JA (2009) Coupled myosin VI motors facilitate unidirectional movement on an F-actin network J Cell Biol. 187: 53–60

Sullivan D (2010) Uncertainty in Mapping Malaria Epidemiology: Implications for Control. Epidemiol Rev. 1:175–187

Telford, S (1988). A contribution to the systematics of the reptilian malaria parasites, family Plasmodiidae (Apicomplexa: Haemosporina). Bull Fl Mus Biol Sci 34:65–96

Trager W. and Jensen JB (1976) Human Malaria Parasites in Continuous Culture Science J Parasitol 63: 883-886

Triglia T, Duraisingh MT, Good RT, Cowman AF (2005) Reticulocyte-binding protein homologue 1 is required for sialic acid-dependent invasion into human erythrocytes by *P. falciparum.* Mol Microbiol 55:162-174

Ward GE, Miller LH, Dvorak JA (1993) The origin of parasitophorous vacuole membrane lipids in malaria-infected erythrocytes. J Cell Sci 106:237-248

Wickert H, Krohne G (2006) Maurer's clefts: A novel multi-functional organelle in the cytoplasm of *Plasmodium falciparum*-infected erythrocytes. Int J Parasitol 36: 23–36

Witney AA, Wolters D, Wu Y, Gardner MJ, Holder AA, Sinden RE, Yates JR, Carucci DJ (2002) Proteomic view of the *P. falciparum* life cycle. Nature 419:520-526

Yadavalli R, Ledger C, Sam-Yellowe TY (2012) In vitro human cell-free expression system for the synthesis of malaria proteins. Parasitol Res 111:2461-2465

Zuccala ES, Baum J (2011) Cytoskeletal and membrane re-modelling during malaria parasite invasion of the human erythrocyte Brit J Haematol 154: 680–689

PlasmoDB (www.plasmodb.org)

New England BioLabs Inc (www.NEB.com)

ExPASy (www.expasy.org)

PSORT.b (www.PSORT.b.org)

swissprot-workspace (www.swissprot.org)

GeneDB (www.genedb.org),

Center for Disease Control (CDC) (www.cdc.gov/malaria)

World Health Organization. World Malaria Report 2011 summary (Report).

APPENDICES

Appendix 1. Table1A. GPI-anchored membrane surface proteins (MSP) secreted during erythrocyte invasion

Tables 1A-E: Source. Cowman et al. 2012 Invasion of erythrocytes by malaria parasites.

Name	PlasmoDB accession number	Genetic knockout	Localization in merozoite before/during invasion	Potential function	Feature/structure
GPI-anchored MSPs					
MSP-1	PF3D7_0930300	N	Surface/complex shed during invasion with MSP1/19 EGF C-terminal domain retained in PV of ring stage	Putative Band 3 ligand; C-terminal double EGF domain redundant for divergent molecules: processed SUB1 and -2	Two C-terminal EGF domains: compact side by side arrangement
MSP-2	PF3D7_0206800	N	Surface	Highly polymorphic; likely structural role as surface coat	Unordered repetitive structure
MSP-5	PF3D7_0207000	N	Surface	Not known	C-terminal EGF domain
MSP-4	PF3D7_0206900.1	Y	Surface	Not known	C-terminal EGF domain
MSP-10	PF3D7_0620400	N	Surface	Not known	C-terminal EGF domain
Pf12	PF3D7_0612700	Y	Surface/shed	Potential adhesive protein	6-Cys domains
Pf38	PF3D7_0508000	Y	Surface/shed	Potential adhesive protein	6-Cys domains
Pf92	PF3D7_1364100	Y	Surface/shed	Not known	Cys-rich protein

Table1B. Peripheral proteins secreted during erythrocyte invasion

Pf113	PF3D7_1420700	N	Surface/shed	Not known	No data
MSP-9 (ABRA)	PF3D7_1228600	Y	Surface/shed	Putative protease	No data
S-antigen	PF3D7_1035200	N	Secreted into PV of schizont and released on egress	Not known; potential immunomodulatory role	Highly repetitive and diverse protein
GLURP	PF3D7_1035300	Y	Secreted into PV of schizont and released on egress	Not known	Repetitive Glutamate-rich
MSP-3	PF3D7_1035400	Y	Surface/shed	Not known; binds to MSP-1	Repetitive and Glutamate-rich
MSP-6	PF3D7_1035500	Y	Surface/shed	Not known; binds to MSP-1	Leucine zipper-like C-terminal domain
H101 (MSP-11)	PF3D7_1035600	Y	Surface/shed	Not known	MSP-3 family, leucine zipper-like C-terminal domain
H103	PF3D7_1035900	Y	Surface/shed	Not known	MSP-3 family, leucine zipper-like C-terminal domain
MSP-7	PF3D7_1335100	Y	Surface/shed	Associates with MSP-1, gene knockout in *P. berghei* shows important in invasion of mature erythrocytes	No data
MSP-7-like (MSRP2)	PF3D7_1334800	Y	Surface/shed	Not known; may associate with MSP-1	MSP-7 family
MSPDBL-1	PF3D7_1036300	Y	Surface/shed	Binds to unknown receptor on red cell	Member of EBL family, DBL and leucine zipper-like domains
MSPDBL-2	PF3D7_1035700	Y	Surface/shed	Binds to unknown receptor on red cell	Member of EBL family, DBL and leucine zipper-like domains
SERA3	PF3D7_0207800	Y	Secreted into PV of schizont and released on egress	Cysteine protease domain with active site serine	Cysteine protease domain
SERA4	PF3D7_0207700	N	Most secreted into PV of schizont and released on egress	Cysteine protease domain with active site serine	Cysteine protease domain
SERA5	PF3D7_0207600	N	Secreted into PV of schizont and released on egress	Cysteine protease domain with active site serine	Cysteine protease domain
SERA6	PF3D7_0207500	N	Most secreted into PV of schizont and released on egress	Cysteine protease domain with active site cysteine	Cysteine protease domain
Pf41	PF3D7_0404900	Y	Surface/shed	Potential adhesive protein; binds Pf12 on merozoite	6-Cys domains

Table 1 (continued)

Table 1C. Plasma membrane proteins secreted during erythrocyte invasion

Name	PlasmoDB accession number	Genetic knockout	Localization in merozoite before/during invasion	Potential function	Feature/structure
ROM1	PF3D7_1114100	Y	Mononeme (proposed new apical organelle) or microneme/surface	Rhomboid protease; cleaves AMA1, MAEBL, EBLs, PfRh proteins; likely role after invasion in PV formation	Multipass transmembrane protein
ROM4	PF3D7_0506900	ND	Surface/shed	Rhomboid protease; cleaves AMA1, MTRAP, EBL, and PfRh proteins in transmembrane to allow shedding during invasion	Multipass transmembrane protein

Table 1D. Microneme proteins secreted during erythrocyte invasion

AMA 1	PF3D7_1133400	N	Micronemes/surface and binds to RON2 that has been inserted into red cell membrane and tracks with tight junction	Released on merozoite surface; binds RON complex; potential ligand for McLeod antigen, phosphorylation of cytoplasmic tail essential, may be involved in signaling	PAN (plasminogen, apple, nematode) motifs
EBA-175	PF3D7_0731500	Y[a]	Micronemes/surface and binds to glycophorin A	Binds to glycophorin A, likely signaling role for invasion	EBL family with DBL domains; "handshake" association between region II dimers creates groove for glycophorin A binding
EBA-181/JESEBL	PF3D7_0102500	Y	Micronemes/surface and binds to unknown receptor	Binds to unknown receptor on red cell	EBL family member with DBL domains
EBA-140/BAEBL	PF3D7_1301600	Y	Micronemes/surface and binds to glycophorin C	Binds to glycophorin C on red cell	EBL family member with DBL domains
EBL-1	PF3D7_1371600	Y	No data	Binds to glycophorin B, nonfunctional because of mutations causing truncated protein	EBL family member with DBL domains
PTRAMP	PF3D7_1218000	ND		Not known; cleaved by SUB2 on merozoite surface	Long extended structure
PfRipr	PF3D7_0323400	N	Micronemes/surface and binds to PfRh5	Binds to PfRh5	10 EGF domains, 87 cysteines
MTRAP	PF3D7_1028700	N	Micronemes/PV	Potential motor-associated protein	Thrombospondin-like domains
PTRAMP	PF3D7_1218000	N	Micronemes/surface	Potential motor-associated protein	Thrombospondin-like domains
SPATR	PF3D7_0405900	ND	Micronemes/surface	Not known for blood stages	Thrombospondin-like domains
GAMA	PF3D7_0828800	ND	Micronemes/surface	Binds to red cells; has GPI anchor	No data
SUB2	PF3D7_1136900	N	Micronemes/PV	Protease that processes MSP-1, MSP-6, MSP-7, AMA1, PTRAMP and other proteins to prime merozoite for invasion	Subtilisin-like serine protease

Table 1 (continued)

Table1E. Exoneme proteins secreted during erythrocyte invasion

Name	PlasmoDB accession number	Genetic knockout	Localization in merozoite before/during invasion	Potential function	Feature/structure
SUB1	PF3D7_0507500	N	Exonemes/PV	Protease that processes MSP-1, MSP-6, MSP-7, AMA1, RAP1, MSRP2 and SERAs to prime merozoite for invasion	Subtilisin-like serine protease
Rhoptry neck proteins					
PfRh1	PF3D7_0402300	Y[a]	Rhoptry neck/surface	Binds to red cells via receptor Y	PfRh family
PfRh2a	PF3D7_1335400	Y	Rhoptry neck/surface	Binds to red cells via receptor Z	PfRh family
PfRh2b	PF3D7_1335300	Y	Rhoptry neck/surface	Binds to red cells via receptor Z	PfRh family
PfRh4	PF3D7_0424200	Y	Rhoptry neck/surface	Binds to red cells via complement receptor 1	PfRh family
PfRh5	PF3D7_0424100	N	Rhoptry neck/surface forms complex with Ripr	Binds to red cells via Basigin	Classed as PfRh family but lacks homology and no transmembrane so likely functionally distinct
RON2	PF3D7_1452000	ND	Rhoptry neck/into red cell membrane	Inserted in red cell membrane at invasion, forms complex at tight junction with RON proteins and AMA-1	Multipass transmembrane protein
RON3	PF3D7_1252100	ND	Rhoptry neck/into red cell	Likely also forms complex at tight junction with other RON proteins and AMA-1	No data
RON4	PF3D7_1116000	ND	Rhoptry neck/into red cell	Injected into red cell, binds to RON2 and forms a complex at tight junction with RON proteins and AMA-1	Binds to AMA1 via hydrophobic groove
RON5	PF3D7_0817700	ND	Rhoptry neck/into red cell	Forms complex at tight junction with RON proteins and AMA-1	No data
ASP	PF3D7_0405900	ND	Rhoptry neck/surface	Not known; has putative GPI anchor	Sushi domains

N, knockout attempt unsuccessful; Y, knockout generated; ND, knockout not attempted; PV, parasitophorous vacuole; MSP, merozoite surface protein
[a]EBL and PfRh families show overlap in function and, while individually nonessential, overall are essential.

Appendix 2. Multiple sequence alignment of PTEX150 protein among some *Plasmodium sp.* and *Theileria* showing protein identity. Source. Plasmodb

Tpar|766139----------SLEILNPEFENIKYFRSHNLIYKPINQETFEAQLE

pcha|PCHAS_100940---
ISQENREQLENLKKNIEKALKERGINIDDLSKKYMNGNIEGKDAFIQLLKNMSQDDDIA

pber|PBANKA_10085----
MSQDNHEQLENLKKNIENALKEHGINIDELSNNFLNGNKEEGKDAFMELLKNMSQDDNII

pyoe|PY02301----------
MSQENHEQLENLKKNIENALKERGINIDELSNNFLNGNKEEGKDAFMELLKSMSQDDNLI

pfal|PF14_0344----------
NQMKQINELKDKLETMLKGAGVNVDKIKDSIKNNDLLKNKQLLKEAISKLTLDPSMM

pviv|PVX_084720------
VSPEQLNKINQLKDKLENVLKNVGVDVEQLKENMQNENIMQNKDALRDLLANLPMNPGMM

pkno|PKH_131290-----
VSPEQLNKINELKDKLENVLKKSGIDVEQLKKSMENENFMQNKDGFKDFLANMPMNPAMM

KEY: tpar = *Theileria parvum*; pcha = *Plasmodium chabaudi*; pviv = *Plasmodium vivax*; pkno = *Plasmodium knowlesi*; pf = *Plasmodium falciparum;* pber = *Plasmodium berghei*; pyoe = *Plasmodium yoelii.*

Appendix 3. Predicted protein sequence of PF3D7_1436300 (PTEX150) translocon gene.

Source Plasmodb

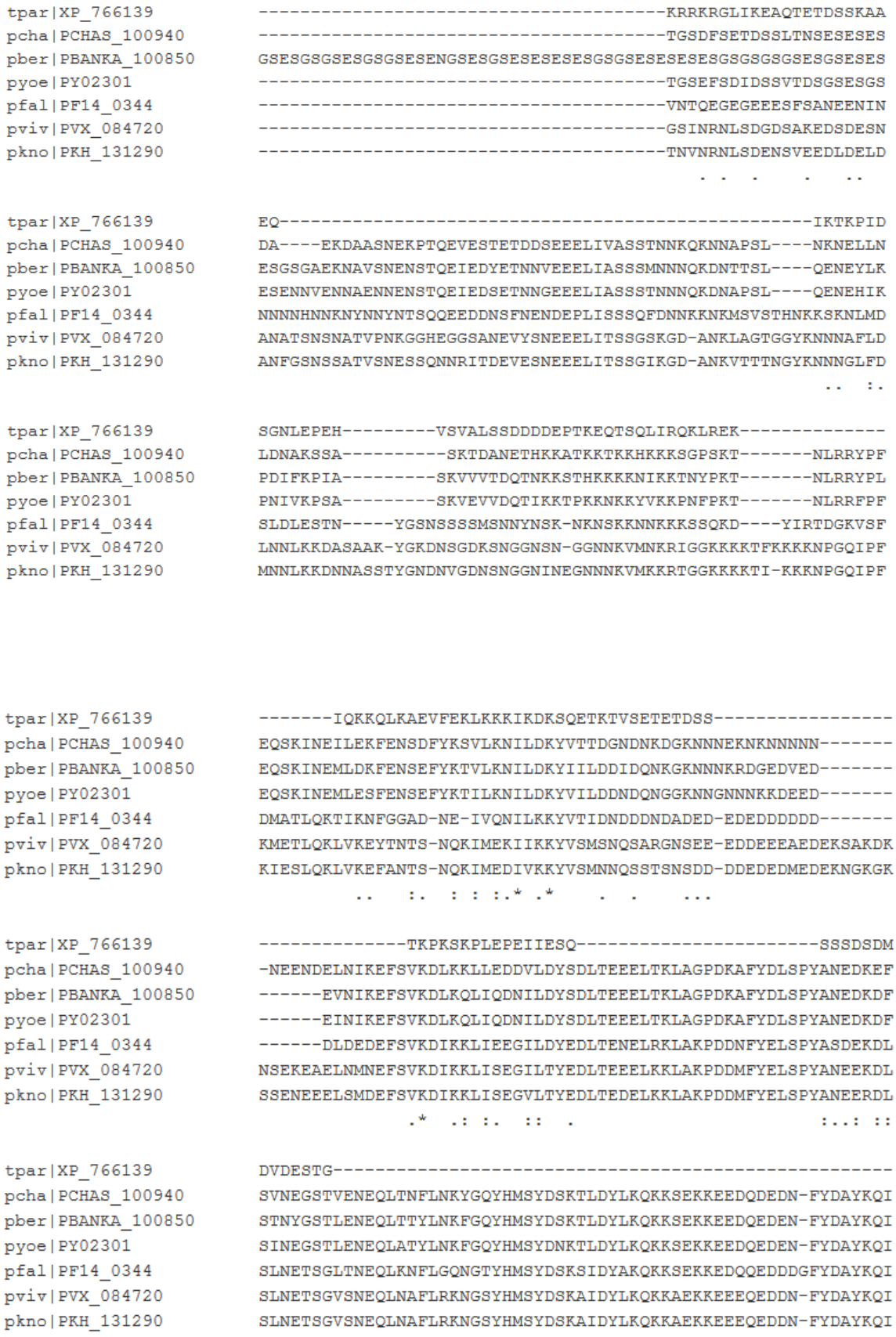
MRIIILALLIVCTIINYYCAVQNNGNKSLNVMPTCSMPGNDSDSNDNETGDVDNDKNNELGNANDNNEMNNENAESKNM
QGENSNNQEQLNENVHANDDAMYEGTPSSDNPPQENVDANNNEQEYGPPQEEPVSENNVENVEVATDDSGNDNINNNDN
FNNNDNYNDNDNFNEEPPSDDGNKNEDELTEGNQSDDKPMNEEEATINEMGKITNPFEDMLKGKVDDMDIGKMMNKDNL
QSFLSSLTGNKDGSGKNPLSDMMNIFGVPQTGKEGAEGGVNKENQMKQINELKDKLETMLKGAGVNVDKIKDSIKNNDL
LKNKQLLKEAISKLTLDPSMMNMLNNKDGANGKPFDINPDSMMKMFNALSNENGNLDDLKMKPTDGSFDSFNDGVDNNL
VPSNPKGQNNNEEDDEEGGDDDDYDDKSFVVNSKYADNSFEDKFNTFDEKDDDVKYELFGENEEAEELNNNTTTASSKG
DANNSVNTQEGEGEEESFSANEENINNNNNHNNKNYNNYNTSQQEEDDNSFNENDEPLISSSQFDNNKKNKMSVSTHNK
KSKNLMDSLDLESTNYGSNSSSSMSNNYNSKNKNSKKNNKKKSSQKDYIRTDGKVSFDMATLQKTIKNFGGADNEIVQN
ILKKYVTIDNDDDNDADEDEDEDDDDDDDLDEDEFSVKDIKKLIEEGILDYEDLTENELRKLAKPDDNFYELSPYASDE
KDLSLNETSGLTNEQLKNFLGQNGTYHMSYDSKSIDYAKQKKSEKKEDQQEDDDGFYDAYKQIKNSYDGIPNNFNHEAP
QLIGNNYVFTSIYDTKENLIKFLKKNSEYDLYDDDDKEGGNFKSPLYDKYGGKLQKFKRQRAFNILKQWRAKEKKLKEK
KKKEEMEENKEFDFSKNYNFSSKNDGGVTMFSKDQLEDMVKNFGGKPSAHVTDSFSRKENPFVPTNTKNNSNDDDDMDN
GYVTFDGKNKVSENDDDEKGNNNDDENDNDDSNDEEELDEEEDDN

Sequence Length: 993 aa

Appendix 4. Multiple sequence alignment for PF3D7_1436300 (PTEX150) Group 0G5161974. Source: Plasmodb. The gaps represent unsequenced segments of a gene

```
tpar|XP_766139          MATLKLSHVLFTLFLYHIKIVFSN---LLDL----------------DNIA---------
pcha|PCHAS_100940       MKILTVKILAIIALLNYVSVSASNKNCSLDIKATINK-GKDDKDNNDESITPNGNEEDDK
pber|PBANKA_100850      MKILIAKVLAITALINYVSVSGSNKNCSLNIKSTINK-GKDDASNNDQPIVPNGNSNDDK
pyoe|PY02301            MKIIIPKVLAIIALINYVSVSGSNKNCSLDIKSTINK-GKDDADNKDQSGVPNGNANDDK
pfal|PF14_0344          MRIIILALLIVCTIINYYCAVQNNGNKSLNVMPTCSMPGNDSDSNDNET----GDVDNDK
pviv|PVX_084720         MRLITVGFLFLSTSFHYSHVFAANGNRNLNIKPTCHKSGKNDKANGSDNIANKGGAQHAA
pkno|PKH_131290         MRLITLGFIFLSTSLHYSHVFALNTNRNLNIKPTSHKSSKNDKANGADNITNKGDSNDAT
                        *  :    : .   : :      *    *::               :

tpar|XP_766139          ----------------------------------------------GSGYNIVQTSERG
pcha|PCHAS_100940       AGDNDETNNNVS---QDGDALNNENADS------------------NDPVQMENASEQP
pber|PBANKA_100850      TGNNGNTNNNEQNNIQNGDELNNKNANS------------------NGPIDVVNNSEQP
pyoe|PY02301            TSNNDNTNNNEQNNAQNGDALNNKNANS------------------NDPIDVVNNSEQS
pfal|PF14_0344          NNELGNANDNNEMNNENAESKNMQGENSNNQEQLNENVHANDDAMYEGTPSSDNPPQENV
pviv|PVX_084720         NGATGTPSGSSNGKKGATTTSASAGQA--------------------GASGGMAAPGMNP
pkno|PKH_131290         TGATGASNGSSNGKEGTATTSATSGQGS-----------------PGTPGGTMTPGMNP
                                                                    .         .

tpar|XP_766139          I-----------------------------------------------------------
pcha|PCHAS_100940       S-----------------------------------------------------------
pber|PBANKA_100850      I-----------------------------------------------------------
pyoe|PY02301            I-----------------------------------------------------------
pfal|PF14_0344          DANNNEQEYGPPQEEPVSENNVENVEVATDDSGNDNINNNDNFNNNDNYNDNDNFNEEPP
pviv|PVX_084720         N-----------------------------------------------------------
pkno|PKH_131290         S-----------------------------------------------------------
```

```
tpar|XP_766139          ----------------------------------------TKLMIFSTQEKKITVIYNKEK
pcha|PCHAS_100940       ------------------------------------------------IELSFDKLANGDK
pber|PBANKA_100850      ------------------------------------------------KGLSFDKLSNGDK
pyoe|PY02301            ------------------------------------------------KGLSFDKLANGEK
pfal|PF14_0344          SDDGNKNEDELTEGNQSDDKPMNEEEATINEMGKITNPFEDMLKGKVDDMDIGKMMNKDN
pviv|PVX_084720         --------------------------------FEQMMKPLNDMFKGNGEGLNIENIMNSDM
pkno|PKH_131290         --------------------------------LEEMMKPLNEIFAGNGEGLNIENIMNSES
                                                                .:  : * :

tpar|XP_766139          L------IWKG--HPGEYTKCFTIYKFEMSNIGLA-------------------------
pcha|PCHAS_100940       IKSMLDSLVEND-KNGNGNNGKDMYKVMMSNIFNG---INNGNDSTNSQNDEVNNV-TLN
pber|PBANKA_100850      IKSILDSLIENE-KNNDGNNGKDMYKVMMNNIFKG-INNSADNDTTNSQNNNVDDIGTFN
pyoe|PY02301            IKSILDSLIENE-KNDNGTNGKDMYKVMMSNIFKG-IDNGADNDTTNSQNNEVDNVGNFN
pfal|PF14_0344          LQSFLSSLTGN--KDGSG---KNPLSDMMNIFGVP-------QTGKEGAEGGVNKE----
pviv|PVX_084720         FQNFFNSLMGGNPHDGAGGGQEILFKDMLNAMNAQ----GGGAPGAAATSGGANKDPNIS
pkno|PKH_131290         FQNFFNSLMGGNPQDGSSNGQENLFKDMMSAINSQLAGATGGAAGADANNG--NNGPNIP
                        :      :  .  : .          .  :. :

tpar|XP_766139          -------------------------SLEILNPEFENIKYFRSHNLIYKPINQETFEAQLE
pcha|PCHAS_100940       ISQENREQLENLKKNIEKALKERGINIDDLSKKYMNGNI-EGKDAFIQLLKNMSQDDDIA
pber|PBANKA_100850      MSQDNHEQLENLKKNIENALKEHGINIDELSNNFLNGNKEEGKDAFMELLKNMSQDDNII
pyoe|PY02301            MSQENHEQLENLKKNIENALKERGINIDELSNNFLNGNKEEGKDAFMELLKSMSQDDNLI
pfal|PF14_0344          ---NQMKQINELKDKLETMLKGAGVNVDKIKDSIKNNDLLKNKQLLKEAISKLTLDPSMM
pviv|PVX_084720         VSPEQLNKINQLKDKLENVLKNVGVDVEQLKENMQNENIMQNKDALRDLLANLPMNPGMM
pkno|PKH_131290         VSPEQLNKINELKDKLENVLKKSGIDVEQLKKSMENENFMQNKDGFKDFLANMPMNPAMM

tpar|XP_766139          E--------------------LTKYSNEFSTTGPR-------------------------
pcha|PCHAS_100940       M--NIPDFAKNGNFPGLD---LPNIFN-LSDDEENNIDNDI-------------IAENNE
pber|PBANKA_100850      M--NFPEITKNCNFPNLD---IPNIFN-LTDAGVDSIDYDI-------------AEKNNV
pyoe|PY02301            M--NLPDFTQNGNFPNLD---IPNIFS-LSDDDLNSIDYNI-------------IEKNNE
pfal|PF14_0344          NMLNNKD-GANGKPFDINPDSMMKMFNALSNENGNLDDLKMKPTDGSFDSFNDGVDNNLV
pviv|PVX_084720         Q--NMMA-GKDGNMFNMDPNQMMNMFNQLSQGKMNMKDFGM----GDFMP----------
pkno|PKH_131290         Q--NIMG-GKDGNIFNMDSNQMMDMFNQFSQGNMNMKDFGM----GDFMSNGGFPAGMMT
                                          : .  . ::

tpar|XP_766139          ---PSGQKPPILSERIPGH---------------------------------PRPGKVPA
pcha|PCHAS_100940       S--ALNAKNENAQDNEDIGEGTSKDNKYTLFDESNSNENVN------ENAQTSNDGTPIT
pber|PBANKA_100850      N--TLNAKNDNTKENEDVG-----EDKYTLFDNSNSDENVNNEDSTIEKGQPSNDENLIS
pyoe|PY02301            N--ALNATNDNTKENEDVN-----EDKYTLFDNSNSDENVNNEDSTIENVQPSNDENLIS
pfal|PF14_0344          PSNPKGQNNNEEDDE-EGGDDDDYDDKSFVVNSKYADNSFEDKFNTFDEKD--DDVKYEL
pviv|PVX_084720         P--PVHANDQDAEDD-SRG-------KAFVTNSSNNDINFAHKLNAFEYSNGPSEGMFQL
pkno|PKH_131290         P--PVGSQDKGAEDNLMGG-------KAYVSNSNNNDINFADKLNAFEDINGSDERMFDL
                           .        .:

tpar|XP_766139          ------------------------------------TRKKRKPTSSYNYG-----------
pcha|PCHAS_100940       NAD----------------NATDSD----------SSYASESDSDINYL-----------
pber|PBANKA_100850      SGDCVIGSDIGSVSGNDSGYLTETD--SGSVSGNGSDYLTETDSGKGSGSESGSGSGSES
pyoe|PY02301            NGDSALDSDVDSVSGNDSGYLTESDNDSGYLTETDSDNMIDSDSDVNYL-----------
pfal|PF14_0344          FGE-----------NEEAEELNNNT----------TTASSKGDA--NNS-----------
pviv|PVX_084720         YGM-----------NNDDGVIDDGM----------SDSVGKNSALDVSG-----------
pkno|PKH_131290         YET-----------NDEDGVTSNSM----------SDSVEESDILEVNG-----------
```


tpar|XP_766139 ----------KRRKRGLIKEAQTETDSSKAA
pcha|PCHAS_100940 ----------TGSDFSETDSSLTNSESESES
pber|PBANKA_100850 GSESGSGSESGSGSESENGSESGSESESESESGSGSESESESESGSGSGSGSESGSESES
pyoe|PY02301 ----------TGSEFSDIDSSVTDSGSESGS
pfal|PF14_0344 ----------VNTQEGEGEEESFSANEENIN
pviv|PVX_084720 ----------GSINRNLSDGDSAKEDSDESN
pkno|PKH_131290 ----------TNVNRNLSDENSVEEDLDELD
tpar|XP_766139 EQ----------IKTKPID
pcha|PCHAS_100940 DA----EKDAASNEKPTQEVESTETDDSEEELIVASSTNNKQKNNAPSL----NKNELLN
pber|PBANKA_100850 ESGSGAEKNAVSNENSTQEIEDYETNNVEEELIASSSMNNNQKDNTTSL----QENEYLK
pyoe|PY02301 ESENNVENNAENNENSTQEIEDSETNNGEEELIASSSTNNNQKDNAPSL----QENEHIK
pfal|PF14_0344 NNNNHNNKNYNNYNTSQQEEDDNSFNENDEPLISSSQFDNNKKNKMSVSTHNKKSKNLMD
pviv|PVX_084720 ANATSNSNATVPNKGGHEGGSANEVYSNEEELITSSGSKGD-ANKLAGTGGYKNNNAFLD
pkno|PKH_131290 ANFGSNSSATVSNESSQNNRITDEVESNEEELITSSGIKGD-ANKVTTTNGYKNNNGLFD
tpar|XP_766139 SGNLEPEH---------VSVALSSDDDDEPTKEQTSQLIRQKLREK----------
pcha|PCHAS_100940 LDNAKSSA----------SKTDANETHKKATKKTKKHKKKSGPSKT-------NLRRYPF
pber|PBANKA_100850 PDIFKPIA---------SKVVVTDQTNKKSTHKKKKNIKKTNYPKT-------NLRRYPL
pyoe|PY02301 PNIVKPSA---------SKVEVVDQTIKKTPKKNKKYVKKPNFPKT-------NLRRFPF
pfal|PF14_0344 SLDLESTN-----YGSNSSSSMSNNYNSK-NKNSKKNNKKKSSQKD----YIRTDGKVSF
pviv|PVX_084720 LNNLKKDASAAK-YGKDNSGDKSNGGNSN-GGNNKVMNKRIGGKKKKTFKKKKNPGQIPF
pkno|PKH_131290 MNNLKKDNNASSTYGNDNVGDNSNGGNINEGNNNKVMKKRTGGKKKKTI-KKKNPGQIPF
tpar|XP_766139 -------IQKKQLKAEVFEKLKKKIKDKSQETKTVSETETDSS----------
pcha|PCHAS_100940 EQSKINEILEKFENSDFYKSVLKNILDKYVTTDGNDNKDGKNNNEKNKNNNNN-------
pber|PBANKA_100850 EQSKINEMLDKFENSEFYKTVLKNILDKYIILDDIDQNKGKNNNKRDGEDVED-------
pyoe|PY02301 EQSKINEMLESFENSEFYKTILKNILDKYVILDDNDQNGGKNNGNNNKKDEED-------
pfal|PF14_0344 DMATLQKTIKNFGGAD-NE-IVQNILKKYVTIDNDDDNDADED-EDEDDDDDD-------
pviv|PVX_084720 KMETLQKLVKEYTNTS-NQKIMEKIIKKYVSMSNQSARGNSEE-EDDEEEAEDEKSAKDK
pkno|PKH_131290 KIESLQKLVKEFANTS-NQKIMEDIVKKYVSMNNQSSTSNSDD-DDEDEDMEDEKNGKGK
tpar|XP_766139 ----------TKPKSKPLEPEIIESQ----------SSSDSDM
pcha|PCHAS_100940 -NEENDELNIKEFSVKDLKKLLEDDVLDYSDLTEEELTKLAGPDKAFYDLSPYANEDKEF
pber|PBANKA_100850 ------EVNIKEFSVKDLKQLIQDNILDYSDLTEEELTKLAGPDKAFYDLSPYANEDKDF
pyoe|PY02301 ------EINIKEFSVKDLKQLIQDNILDYSDLTEEELTKLAGPDKAFYDLSPYANEDKDF
pfal|PF14_0344 ------DLDEDEFSVKDIKKLIEEGILDYEDLTENELRKLAKPDDNFYELSPYASDEKDL
pviv|PVX_084720 NSEKEAELNMNEFSVKDIKKLISEGILTYEDLTEEELKKLAKPDDMFYELSPYANEEKDL
pkno|PKH_131290 SSENEEELSMDEFSVKDIKKLISEGVLTYEDLTEDELKKLAKPDDMFYELSPYANEERDL
tpar|XP_766139 DVDESTG----------
pcha|PCHAS_100940 SVNEGSTVENEQLTNFLNKYGQYHMSYDSKTLDYLKQKKSEKKEEDQDEDN-FYDAYKQI
pber|PBANKA_100850 STNYGSTLENEQLTTYLNKFGQYHMSYDSKTLDYLKQKKSEKKEEDQEDEN-FYDAYKQI
pyoe|PY02301 SINEGSTLENEQLATYLNKFGQYHMSYDNKTLDYLKQKKSEKKEEDQEDEN-FYDAYKQI
pfal|PF14_0344 SLNETSGLTNEQLKNFLGQNGTYHMSYDSKSIDYAKQKKSEKKEDQQEDDDGFYDAYKQI
pviv|PVX_084720 SLNETSGVSNEQLNAFLRKNGSYHMSYDSKAIDYLKQKKAEKKEEEQEDDN-FYDAYKQI
pkno|PKH_131290 SLNETSGVSNEQLNAFLRKNGSYHMSYDSKAIDYLKQKKAEKKEEEQEDDN-FYDAYKQI

```
tpar|XP_766139          ---------------PQVI-----------------------------------------
pcha|PCHAS_100940       QSSYSGIPSNYYHQAPQLIGDKYVFTSVYDQKKDLIKFLKKTNGKSIYVHPENMENGNGS
pber|PBANKA_100850      KSSYYGIPSNYYHQAPQLIGDKYVFTSVYDQKKDLIKFLKKTNGKSLYTHPDELGNGNTE
pyoe|PY02301            KSSYSGIPSNYYHQAPQLIGDKYVFTSVYDQKKDLIKFLKKTNGKSIYTHPDELENGNTA
pfal|PF14_0344          KNSYDGIPNNFNHEAPQLIGNNYVFTSIYDTKENLIKFLKK-----------NSEYDLYD
pviv|PVX_084720         KNSYEGIPSNYYHDAPQLIGENYVFTSVYDKKKELIDFLKRSNGA---TDSSNSSAGKDK
pkno|PKH_131290         KNSYEGIPSNYYHDAPQLIGDNYVFTSVYDKKKELIDFLKRTHGV---TDESDSTSGKDK
                                    **:*

tpar|XP_766139          ------------------------------------QSDATTQTDTHESQNSETQTV---
pcha|PCHAS_100940       GNSSEMKQF---------QKLSYYKRKNAYNILVEQRRIEKEEKEKKEKKAQESKKAIND
pber|PBANKA_100850      NKTKSDQNDKSPEIKQL-QKLSYYKRKNAYNILVEQRRIEKEEKEKREKKAQTTNTSIND
pyoe|PY02301            NKTGSDQTGKSSDIKQL-QKLSYYKRKNAYNILVEQRRIEKEEKEKMEKKAKTPNKLVND
pfal|PF14_0344          DDDKEGGNFKSPLYDKYGGKLQKFKRQRAFNILKQWRAKEKKLKEKKKKEEMEENKEF--
pviv|PVX_084720         GNSAESGTYKSKYYDKYMKKLSEYRRREAFKILKKRRAQEKKMQKKQEMQNNSSNEV---
pkno|PKH_131290         DNSAENGAYKSKYYDKYMKKLSEYRRREAYKILKKRRAQEKKLQKKQEMQNNNNNEV---
                                                    .   .   .. : :    :

tpar|XP_766139          ------------------------------------------------------------
pcha|PCHAS_100940       DYIKHLSNGGL----NKSTVLFSKDDLDKMVNGLYSE-----SGNSNGSSANINLDNMAT
pber|PBANKA_100850      DYIKHLNNGGL----NKSTVLFSKDELDKMVNGLYSQ-----SGNSNGDSVNQNFDSITA
pyoe|PY02301            DYIQHLNNGGL----NKNTVLFSKDELDKMVNGKYSQ-----SGNSNGDSANKNLDNITA
pfal|PF14_0344          DFSKNYN---FSSKNDGGVTMFSKDQLEDMVKNF---------GGKPSAHVTDSFS--RK
pviv|PVX_084720         DYSEYFKKNGFINSSNGTVKTFSKDQLDNMVKQFNSDGDDIPSSSGAGADLGDNYSGVSG
pkno|PKH_131290         DYSEYFKKNGFINSSNGTVKTFSKDQLDNMVKQFNNGGSEIFSSGGSGADVGGDYSGMGA


tpar|XP_766139          -IQTSSTETQTKT----------------------QNDDRGPSTLPIKKRPYKPD-----
pcha|PCHAS_100940       NIKNSNNGQNDNK------------------FNDSVNDDDINDNELPEDNADEDEE----
pber|PBANKA_100850      NIKNVSNEQNDNQ------SNAFV-------------NDGNSDNELLEDYPEDEDEDE--
pyoe|PY02301            NIKNVNNEQNDNQ------SNSVV------------NDDENDENELLEDYPEDEDE----
pfal|PF14_0344          ENPFVPTNTKNNSNDDDDMDNGYVTFDGKNKVSENDDDEKGNNNDDENDNDDSNDEEELD
pviv|PVX_084720         GGQFSPSGGSGNN------PSGYVTFDGQNIVGPNENEEEEPTEDVLNEDDDNADDDD--
pkno|PKH_131290         GGQFSPSSGNSNT-------TGYVTFDGQSVVGSNENEEEESNEDILNEDEDNSEDDD--
                              .  . :                          ::        :.   . :

tpar|XP_766139          ------
pcha|PCHAS_100940       ------
pber|PBANKA_100850      ------
pyoe|PY02301            ------
pfal|PF14_0344          EEEDDN
pviv|PVX_084720         ------
pkno|PKH_131290         ------
```

Key: tpar = *Theileria parvum*; pcha = *Plasmodium chabaudi*; pviv = *Plasmodium vivax*; pkno = *Plasmodium knowlesi*; pf = *Plasmodium falciparum***;** pber = *Plasmodium berghei*; pyoe = *Plasmodium yoelii*. The gaps represent unsequenced segments of a gene

Appendix 5. RNB-like protein segment showing partial amino acid sequence alignment in *Plasmodium* species and some members of the phylum apicomplexa in the region of location of RNB-like residues 829-1322 (GeneDB).

bbov|XP_001610520.1-------SGVITRVIGDVTEPYSRMYSLMVFRGLNPNGFSDE

tpar|XP_765923----------SELEPKGGIIRVIGETGDPDSEMEGTLLFHGLDPKGFSEP

tann|TA20915----------SELEPKGDIIRVIGETCDPDSEMEGTMLFHGLDPKGFSEA

pviv|PVX_098745----------KKWDTHQINPSGSITTILGNEKNFLGVIYFFIYFYKIHFHI

pkno|PKH_070330 ---------KWDTHQMNPYGSITTILGNEKNFLGVIYFFIHFYKIHF

PF3D7_0906000 ---------EKKQINPIGNITTILGNENNFFGIIYLFLYLYKIHYYIYKIT

pcha|PCHAS_041620 -----EENEINPTGDIISILGNEQNFFSIIYFFLHFYKVNFHIYKKE

pber|PBANKA_041530---WEENEINPTGDIISILGNEQNFFSIIYFFLYFYKINFHIYKKN

pyoe|PY04959----------EENEINPTGDIISILGNEQNFFSIIYFFLYFYKVNFHIYKKN

cmur|CMU_011660-------STSRYIKGILEGIYGSIENIGTQMNFLMDDYNVSDHLNNSD

cpar|cgd8_5150----------CSKSVNKKLQISKFYGSCNKFETIFSSLLDCYDLGSHEKIYD

----------→1322

Key: bbov = *Babesia bovis* ; tpar = *Theileria parvum* ; tann = *Theileria annulata* ; pviv = *Plasmodium vivax* ; pkno = *Plasmodium knowlesi*; PF3D7= *Plasmodium falciparum*; pcha = *Plasmodium chabaudi* ; pber = *Plasmodium berghei*; pyoe = *Plasmodium yoelii*; cmur = *Cryptosporidium muris*; cpar = *Cryptosporidium parvum*

Appendix 6. Amino acid sequence homology in RNB-like protein in *Plasmodium* and *E. coli* (GeneDB)

PY	XP_723884	2463	NVSDIIELLKKHGIYLKVNSLAHILQFLDEKENHLKKN	2541
PF	XP_002261518	2345	ETSDIIEVLEKHGIYLPVSDLRHILKFLDEARNVIKqK	2421
EC	ACI_73414	507	AITSFRSVLAELGLELPGGNKPEPRDYAELLESVADRD	572
PY	XP_723884	2542	SFLLYTHFTSPIRRYPDILVHRIIKRIINDEHQLKEKC	2618
PF	XP_002261518	2422	SFLLYTHFTSPIRRYPDILVHRVIKKIINDENKLNGHLC	2497
EC	ACI_73414	573	ALQSYAHFTSPIRRYPDLMLHRAIKYLLAKEQGHQGNT	649

Key: PY= *Plasmodium yoelii*, PF= *Plasmodium falciparum*, EC= *Escherichia coli*

Appendix 7. RNB-like peptide sequence found in *E. coli* (Ambler et al. 2006) between residues 85-156

85-RFVGKVQGKNDRLAIVPDHPLLKDAIPCRAARGLNHEFKEGDWAVAEMRR

HPLKGDRSFYAELTQYITFGDD-156

Appendix 8. RNB-like protein identity between *P. yoelii* (XP_723884), *P. falciparum* (XP_ 002808916) and *E. coli* (ACI73414) Source: Plasmodb

```
XP_723884      1386  IEEHYGSTKQYVCIFRNKVKD[6]IPFKKNIPLIQVQNSHIKEFMKKYNVNDISNQLIYIKIFQWSPTEKFPEGKIVEIL  1468
XP_002808916   713   ------TDSRLPCFIYDSSNN  ----------------ITNMLLNYIRREKLNVYVIVKFKEWEKKQINPIGNITTIL  767
ACI73414       166   IVGRYFTEAGVGFVVPDDSR-  LSFDILIPPDQIMGAR------------MGFVVVVELTQRPTRRTKAVGKIVEVL  228

XP_723884      1469  GQNDVFHNMQNAILLNHGLNFNLKKALEDQYLKDLKNKN[4]IITDELKK-RMDLRKECVFTIDPETARDLDDAINICKI  1548
XP_002808916   768   GNENNFFGIIYLFLYLYKIHYYIYKITDMNY---LKSQI  VISDKIMNSFINRKNHMEQLLLLLNGYNKEKEYSIQEI  841
ACI73414       229   GDNMGTGMAVDIALRTHEIPYIWPQAVEQQ-VAGLKEE-  -VPEEAKAGRVDLRDLPLVTIDGEDARDFDDAVYC---  299
```

The gaps represent unsequenced segments of a gene

Appendix 9. Multiple Sequence Alignment of RNB-like (PF3D7_0906000) for Group 0G5 145111 (clusters of interest). See foot notes for keys. The gaps represent unsequenced segments of a gene.

Group 1-7 represents gene segments in the apicomplexan subsepecies with homologous base pairs. Source: Plasmodb

```
GROUP 1
bbov|XP_001610520.1      EVHHN------------------KTSIVFYIPLLNEQRSLSCETLCLEPVQVVIDG---
tpar|XP_765923           AVSESKIRKKKSKVNEMGEKRDMGKSSIVFYIPLLNEQRSVSCDSLGVK-P-VNFNFD--
tann|TA20915             VGSDSEMKKKISKKEKTGEKGDMGKSSIVFYIPLLNEQRSVSCDSLGVK-P-VNFNFD--
pviv|PVX_098745          ANHNREETAKWKFKKNPIENGKL-KNAIVFYVPLLETEKSISDSLLNLTFAFVSLSYEES
pkno|PKH_070330          ANHIREESAKCKLKKNAIENSKL-KNAVVFYVPLLETEKSISDTLLNLTFQFISVSYEES
pfal|PFI0295c            DNNENETNVETLFKKSINENNKL-KNAIVFYVPILEIEKSISDNLLSLKFHFLIISYKEK
pcha|PCHAS_041620        EEEHDEVT-KIEYTNSICDEKKM-KNAISFYVPLLETEKSVSENLLDLKFELISISSKEG
pber|PBANKA_041530       DKYFDEIT-KSEHGNCKSDEKKI-KNAISFYVPLLETEKSVSENLLNLKFELIYISFQDD
pyoe|PY04959             DRYYDDIT-KLEYANSKCDEKKI-KNAISFYVPLLETEKSVSENLLNLKFEFIYISFQED
cmur|CMU_011660          ------------------------FNSFKFSNNSINLPLSYKYHKLSNQFE-VTWRYSDE
cpar|cgd8_5150           ------------------------KNNSLVSLLKLNPSLPINIKSKDNKIE-LYWKFNQD
chom|Chro.80587          ------------------------KNNSLVSLLKLNPSLPINIKSKGNKIE-LYWKFNQD
GROUP II
bbov|XP_001610520.1      ------------------------------VTSSMSVTGHWGLALPMYLHFTSPIRRYSD
tpar|XP_765923           NSVDTSNSVNSSNSVDTSNSVNSSNPVDSVNDVGRVEVRHWGLALPVYLHFTSPIRRYPD
tann|TA20915             --------VNASNTIDTSNT--SVNSVNSLNNVGGLEVKHWGLALPVYLHFTSPIRRYPD
pviv|PVX_098745          ------------------------------YTAGEKKAHHFGLALTKYIHFTSPIRRYID
pkno|PKH_070330          ------------------------------YTEGEKKTHHFGLALNKYIHFTSPIRRYID
pfal|PFI0295c            ------------------------------TVEGQNNTTHFGLVLNKYIHFTSPIRRYID
pcha|PCHAS_041620        ------------------------------NIEKNENGHHFGLSLNNYIHFTSPIRRYID
pber|PBANKA_041530       ------------------------------NIENNENIHHFGLSLNNYIHFTSPIRRYID
pyoe|PY04959             ------------------------------NLKNNENTYHFGLNLNNYIHFTSPIRRYID
cmur|CMU_011660          ------------------------------SNSILNEAHHYALNVPFYTHFTSPIRRVAD
```

```
cpar|cgd8_5150          ------------------------------ARDSNIPTNHFALNMNLYTHFTSPIRRAAD
chom|Chro.80587         ------------------------------ARDSNVSTNHFALNMNLYTHFTSPIRRAAD
```

GROUP III

```
bbov|XP_001610520.1     PK---NAKYFGLMRFISWDENDIEPSGVITRVIGDVTEPYSRMYSLMVFRGLNPNGFSDE
tpar|XP_765923          PNSKLDKNILCINRFVEWKKSELEPKGGIIRVIGETGDPDSEMEGTLLFHGLDPKGFSEP
tann|TA20915            PNSK-DKNLLCINRFVDWKKSELEPKGDIIRVIGETCDPDSEMEGTMLFHGLDPKGFSEA
pviv|PVX_098745         -----KINLYVLVKFKKWDTHQINPSGSITTILGNEKNFLGVIYFFIYFYKIHFHIYERS
pkno|PKH_070330         -----KINLYLLVKFKKWDTHQMNPYGSITTILGNEKNFLGVIYFFIHFYKIHFHIYERS
pfal|PFI0295c           -----KLNVYVIVKFKEWEKKQINPIGNITTILGNENNFFGIIYLFLYLYKIHYYIYKIT
pcha|PCHAS_041620       -----KKNLYVLVNYRKWEENEINPTGDIISILGNEQNFFSIIYFFLHFYKVNFHIYKKE
pber|PBANKA_041530      -----KKNLYVLINFRKWEENEINPTGDIISILGNEQNFFSIIYFFLYFYKINFHIYKKN
pyoe|PY04959            -----KQNFYVLINFRKWEENEINPTGDIISILGNEQNFFSIIYFFLYFYKVNFHIYKKN
cmur|CMU_011660         KKIPIPLNILYTCNILQGTSTSRYIKGILEGIYGSIENIGTQMNFLMDDYNVSDHLNNSD
cpar|cgd8_5150          -----PCNLLFGASI-KICSKSVNKKLQISKFYGSCNKFETIFSSLLDCYDLGSHEKIYD
chom|Chro.80587         ------------------------------------------------------------
```

GROUP IV

```
bbov|XP_001610520.1     TRCLD-----GKYV-YTVGVHIADVSHY-VTEGSLVDLDARERATSVYLEHQVFPMLPQM
tpar|XP_765923          LIPTNSPDSADGGIMYRIGVHVADVSHF-VKENSLVDLDARTRATSVYMEHLVYPMLPQQ
tann|TA20915            FCQANPNDSSNGGIIYTIGVHVADVSHF-VKENSLVDLDARTRATSVYLEHLVYPMLPQQ
pviv|PVX_098745         FVGSD-TNAANEFE-YKIGVHISDVSFF-VTPDSFYDRVASKVCNTIYMDLMVIHMLPSI
pkno|PKH_070330         FVDSD-TNASREFE-YRIGVHISDVSFF-VTPDSYYDRVASKMCNTIYMDLMVIHMLPSI
pfal|PFI0295c           FIKDNKEQHELKYR-YKIGIHISDVSFF-VTPNSYYDNLASKICNTVYMDFTVFHMLPSI
pcha|PCHAS_041620       FIPPD-QFSNIKYK-YKIGVHISDVSFF-IPPDSYYDKMAATQCNTLYMDLMVFHMLPSV
pber|PBANKA_041530      FIHPD-QFSNIKYK-YKIGVHISDVSFF-ISPNSYYDRIASTLCNTLYMELMVFHMLPSI
pyoe|PY04959            FIPLD-EFSNIKYK-YKIGVHISDVSFF-VSPNSYYDKIAATLCNTLYMDLMVFHMLPPI
```

GROUP V

```
bbov|XP_001610520.1      NLQKRRAFDAQKEYKNFAFNQYLKW-------------------------ACADENRPK
tpar|XP_765923           NLQKRRAFDAQKEYKNFAFNKYLQFL-------------SFIAPKLDKAYLKTVSISQPF
tann|TA20915             NLQKRRAFDAQKEYKNFAFNKYLQFL-------------SFIAPKLDRTYLKSVSISQPF
pviv|PVX_098745          NFQKKKTDEAQIHLKNVLLNRYLVYLNEAYKGEKS---LVGNAPKAAAKGGGNLTSKDPS
pkno|PKH_070330          NVQKKKTDEAQIHLKNVLLNRYLVYLNEAYKTEKS---LVANVPKSAAKDGDNHSLKDSS
pfal|PFI0295c            NYQKKKSDEAQIHMKNYFLNKYLIYLNDIYKKVIQNHRMNGYY-SVHGNNENNIDENNDK
pcha|PCHAS_041620        NVQKKKSDDIQIHMKNFFLNKYLVCLNEEYKSKMV----KESFLKLRKKKFEKIRNDTTT
pber|PBANKA_041530       NIQKKKTDDIQIHMKNFFLNKYLVYLNEEYKHKMV----NEYSLKIRKNKNEKINRDTRI
pyoe|PY04959             NIQKKKTDDIQIHMKNFFLNKYLVYLNEEYK-------------TVSGIGNMDLKMEKYK
cmur|CMU_011660          NKKSKASKDLQREAEKMLFSYIL----------------------------KKQRMDYPN
cpar|cgd8_5150           NIKSNANKNMQRDSNNIFFSQLI----------------------------SKIHVTIPS
chom|Chro.80587          NIKSNANKNMQRDSNNIFFSQLI----------------------------SKIHVTIPS
```

GROUP VI

```
bbov|XP_001610520.1      KSVKDTDIAYEPYWDDARIHDLETE---------RPD-------LVVRGTVVIPAFATTE
tpar|XP_765923           SKKHNIEISYEEYWNDADIEELERE---------HPE-------LVVRGTVFIPAFATYE
tann|TA20915             YRKNNNEISYEEYWKEEDIEELERK---------HPE-------LVVRGTVFIPAFATSE
pviv|PVX_098745          ADKNYIDVEYEQYWGSEKIKKILEL---------QKSREKILKNKLFKGVLCVSPYDTSK
pkno|PKH_070330          ADKNYVDVEYEQYWDSEKIKKVLQL---------QKNREKIFKNKLFKGVLCVSPYDTSK
pfal|PFI0295c            VEKNNIFVEYETHWDNDKIKEMLEL---------QKD-TKILKNYLFKGVLYISPFDTNK
pcha|PCHAS_041620        NKNSYVFVKYEKYWDMDKINKVAEL---------END--SNGKHKVFKGTLFVSPFDTSK
pber|PBANKA_041530       NKNSYVHVKYEKYWSMDKINKVAEL---------ENN--ANVKNKVFKGVLFVSPFDTTK
pyoe|PY04959             NNNNYVHVKYEKYWDMDKIAELENADKNDGKNDGKNDDKNNDKNKILKGTLFVSPFDTTK
cmur|CMU_011660          -KSNETQISYIK----EKVNKYLKS---------KT--------LLYEGKVYKIYRDDNK
cpar|cgd8_5150           QKSKYTTTNINRGLNKENINT--EN---------HPE---LLAESSSQSKHPIPKYVGTI
chom|Chro.80587          ------------------------------------------------------------
```

GROUP VII

```
bbov|XP_001610520.1      LVEEMMLLANTQAAQLLSKSF-DRY-FLRVHENTSKAIKQLISSMMPPE----LKQLI--
tpar|XP_765923           LIEEMMLLANTQVAKFISEKI-DLY-FLRIHEDTSKAVKSLIAQMLPKN----LKNLI--
tann|TA20915             LIEEMMLLANTQVAKFISEKI-DLY-FLRTHEDTSKAVKSLIAQMLPKN----LKNLI--
pviv|PVX_098745          LIEEMMILTNFLVANKISQSK-KLG-ILRIHENTSEEIKNNLLHIIDHNTYSRIDALI--
pkno|PKH_070330          LIEEMMILTNFLVANKISQSK-KMG-ILRIHENTSEEIKNNLLHIIDHKTYILIDSMI--
pfal|PFI0295c            LIEEMMILTNFLVANVISINN-MLG-ILRIHEDTSEDIKKNLLKIIDYQTYNKINTMI--
pcha|PCHAS_041620        LIEEMMVLTNFLVANKICECN-NLG-ILRTHEDTSDEIKNNLLQFMDYHTYNKINKII--
pber|PBANKA_041530       LIEEMMIFTNFLVAKKICEYN-NIG-ILRIHDDTTNEIKNNLLQFIDHNTYKKINTII--
pyoe|PY04959             LIEEIMIFTNFLVAKKISEYN-NVG-ILRIHDDTTNEIKNNLLQIIDHNTYNKINKII--
cmur|CMU_011660          IIEELMIKANQLTAEYLINNL-DSKVVLRCHAEIE---KSKLSKLIKYLRGNGMGNIFGD
cpar|cgd8_5150           LIEELMLLANRVTAEFTVKNRPESGCIIRIHDEIA---NTKLYQLITYLRKHGLKHIFED
chom|Chro.80587          LIEELMLLANRVTAEFTVKNRPESGCIIRIHDEIA---NTKLYQLITYLRKHGFKHIFED
```

KEY: bbov = *Babesia bovis* ; tpar = *Theileria parvum* ; tann = *Theileria annulata* ; pviv = *Plasmodium vivax* ; pkno = *Plasmodium knowlesi*; PF3D7= *Plasmodium falciparum***;** pcha = *Plasmodium chabaudi* ; pber = *Plasmodium berghei* ; pyoe = *Plasmodium yoelii*; cmur = *Cryptosporidium muris*; cpar = *Cryptosporidium parvum*

NB: The gaps represent unsequenced segments of a gene