Development of *Plasmodium knowlesi* species specific reagents to help characterise antibody isotype profiles in endemic human populations

Lou Salomé Rose Tainá Herman

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the
University of London

September 2018

Department of Immunology & Infection

Faculty of Infectious & Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

No funding received

Supervisor: Dr Kevin Tetteh, Faculty of Infectious and Tropical Diseases

1

DECLARATION

I, Lou Salomé Rose Tainá Herman, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed

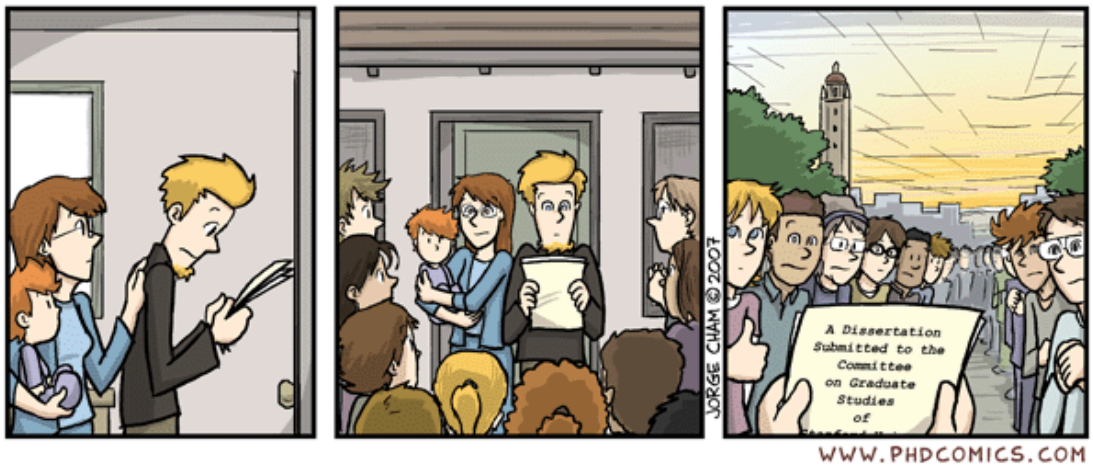
September 2018

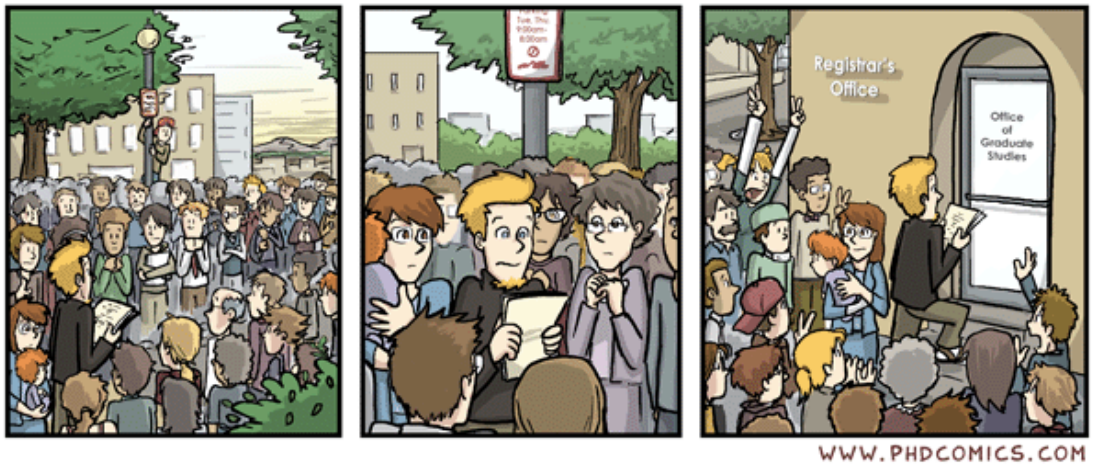
ABSTRACT

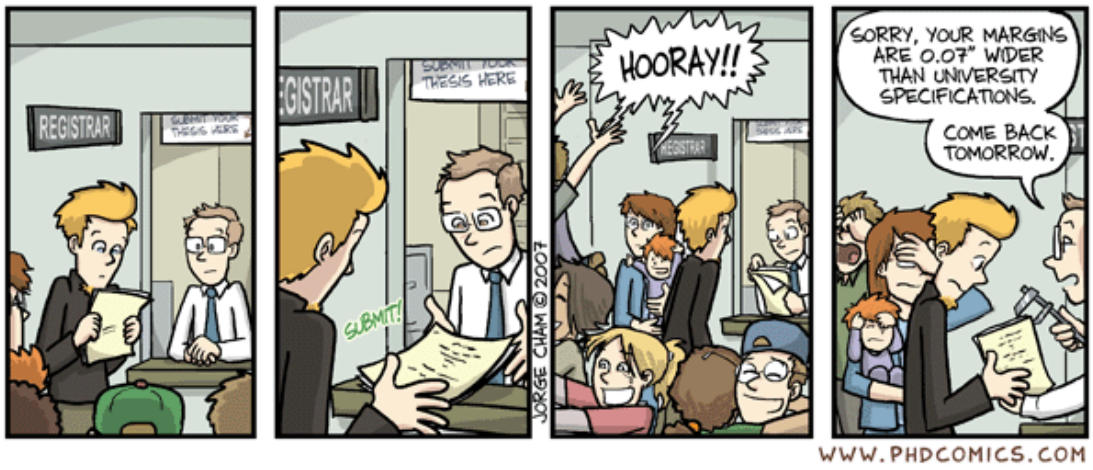
Plasmodium knowlesi is the most common cause of malaria in Malaysian Borneo, with reporting limited to clinical cases presenting to health facilities and scarce data on the true extent of transmission. Serological estimations of transmission have been used with other malaria species to garner information about epidemiological patterns. However, there are a distinct lack of suitable serosurveillance tools for this neglected disease. Using *in silico* tools, we designed and expressed a panel of *P. knowlesi* protein products to address the distinct lack of suitable serosurveillance tools. Antibody prevalence to these antigens was determined by ELISA, Multiplex Bead Assay (MBA) and Protein Microarray for three time-points post-treatment from a hospital-based clinical treatment trial in Sabah, Malaysia (n=110 individuals; 298 total samples for all time points), a small community survey in Sabah, Malaysia (~2000), and Palawan, The Philippines (~550), and a large cross sectional survey in Sabah, Malaysia (~10,000).

Both ELISA and MBA showed similar results, with higher responses observed for the *Pk*SERA3 antigen 2, both at the clinical level across all time points and at the community level. This antigen is suggested to be a short term marker of exposure as reactivity to it tended to decline by day 28 of diagnosis. It was possible to determine that *P. knowlesi* is prevalent in Sabah, Malaysia (37.8%, 3,827/10,125). Seropositivity was found to be associated with an increase of age ($p<0.0001$) as well as seeing macaques ($p<0.0001$). The protein microarray, which was used to determine isotype reactivity profiles (IgM, IgG and IgA) obtained results concordant with the ELISA and MBA results, with *Pk*SERA3 antigen 2 eliciting the highest response for all isotypes. IgM decreased significantly across time while IgG tended to increase across time. We find it necessary to further develop and expand our current panel of *P. knowlesi* antigens in order to better dissect the epidemiology of *P. knowlesi* in Southeast Asia.

ACKNOWLEDGEMENTS







So many people have helped me and guided me during this crazy project! Where shall I start? I suppose I should thank my supervisor Dr. Kevin Tetteh to start with. Throughout this PhD, we have established a love-hate relationship. I picked on him and made fun of him and he gave me more work. Now that I got that out of the way, the real thanks! I would like to thank Kevin, my uber super duper ultra mega supervisor for guiding me in this epic journey and being incredibly patient with me. Thank you for helping me design this project as a self-funded PhD student and for guiding me through it, Master. This Padawan is done! May the Force be with you.

MAM 2016

I would like to give and enormous thank you to Professor Chris Drakeley for giving me the opportunity of having a project in his group and providing me with the space and reagents needed for the entirety of the project. I would like to thank the London School for awarding me the Research Degree Travelling Scholarship. I also want to thank the Sir Richard Stapley Educational Trust for awarding me a grant. A massive thank you to Tate Oulton for teaching and helping me run the microarrays and to Katie Glass, who helped me with the analysis of the microarray data and listened to my numerous rants. I would like to thank everyone else in the Drakeley group, past and present members and Lab 234, for everything they have taught me that helped

make my project stronger. Thank you for the emotional support you gave me over drinks or lab banter. I would like to thank everyone on the 4th floor open plan office for all the advice and tips, as well as the emotional support they have given me. You know who you are! I will always cherish my time at LSHTM and every single person I met and shared a moment with there.

Last, but never least, I would like to thank my family. Ariella Herman, for supporting my education as per my grandfather's wishes. You made sure his words towards his granddaughters were kept, "Education is the most important thing you can have in your life". Thank you for believing in me as well and helping me make this possible.

I would like to thank Sam Willcocks. You gave me your ears to listen to all the good times and bad times. You gave me your shoulders to cry on when I felt like I could not cope anymore. You showed me there was light at the end of the tunnel and that all I needed to do was continue persevering and I would get there. I got there! Thank you Ananda and Gabi. Voces sao chatos mas amo voces e obrigada.

Obrigada mãe, Laura Santana. The strongest, fiercest, most independent woman I know. Your strength is what inspires me to be strong and never give up on the things I want for my life. I still have a long journey to go but I know that your guidance will always help me go up the right path. You have shown me how to cope with thunderstorms in our lives. Eu te amo! Thank you to my late father, Daniel Herman. I know you would have supported me in this journey until the very end and would have helped guide me in the right direction. It is amazing I ended up doing tropical medicine science, as you did. In a way, I followed in your footsteps without realising it. Thank you to my late father Bobby "Bee" Fowler. You were the father I grew up with and I believe the one who introduced me to science. I remember being a 6 year old asking for stories and you telling me about atoms, cells and the Big Bang. I wish you could see me now!

I dedicate this thesis to my mother and my two fathers. Your constant love and support my entire life is what helped me get this far. I love you all ♥

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

No figures

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 7

No figures

APPENDIX I

Publication

APPENDIX II

PUBLICATION

APPENDIX III

PROTOCOL

APPENDIX IV

Appendix V

PROTOCOL

APPENDIX VI

PROTOCOL

APPENDIX VII

APPENDIX VIII

LIST OF TABLES

ABBREVIATIONS

AIA	Afro-Immuno Assay
AIM	Auto Induction Media
AMA	Apical Membrane Antigen
amp	ampicillin
BLAST	Basic Local Alignment Search Tool
BS	Blood spot
BSA	Bovine Serum Albumin
CQ	Chloroquine
CSP	Circumsporozoite protein
CTRP	Circumsporozoite protein and thrombospondin-related adhesive protein [TRAP]-related protein
d H_2O	distilled water
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-Linked ImmunoSorbent Assay
GLURP	Glutamate Rich Protein
GST	Glutathione Sepharose Transferase
H2O	water
HP	Hypothetical Protein
i.v.	intravenous
IFAT	Immunofluorescence Antibody Test
IgA	Immunoglobulin A
IgD	Immunoglobulin D
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Bertani
MBA	Multiplex Bead Assay
MFI	median fluorescent intensity
mRNA	messenger RNA
MSP	Merozoite Surface Protein
MSP1-19	Merozoite Surface Protein 1, C-terminal 19-kilodaldon region
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NHP	Non-Human Primates
o/n	overnight
OD	Optical Density
OR	Odds ratio
PBS	Phosphate Buffered Saline
PBS/T	Phosphate Buffered Saline and Tween 20
PCR	Polymerase Chain Reaction
PE	phycoerythrin
PHE	Public Health England

Qdot	Quantum dot
RBC	Red Blood Cells
RT	room temperature
RT-	Reverse Transcriptase without oligo d(T)20 primers
RT+	Reverse Transcriptase with oligo d(T)20 primers
SERA	Serine Repeat Antigen
SLIC	Sequence and Ligation Independent Cloning strategy
SOC	Super Optimal Broth
SP	Signal Peptide
SPATR	secreted protein with altered thrombospondin type I repeat
SSP2	Sporozoite Surface Protein 2
TM	Transmembrane
TRAP	thrombospondin-related adhesion protein
TSERA	Truncated Serine Repeat Antigen
WHO	World Health Organisation
X-Gal	5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside
XSS	Sabah cross-sectional survey

PUBLICATIONS

Herman, L. S., K. Fornace, J. Phelan, M. J. Grigg, N. M. Anstey, T. William, R. W. Moon, M. J. Blackman, C. J. Drakeley and K. K. A. Tetteh. "Identification and Validation of a Novel Panel of Plasmodium Knowlesi Biomarkers of Serological Exposure." PLoS Negl Trop Dis 12, no. 6 (2018): e0006457.

Fornace, K. M., **L. S. Herman**, T. R. Abidin, T. H. Chua, S. Daim, P. J. Lorenzo, L. Grignard, N. A. Nuin, L. T. Ying, M. J. Grigg, T. William, F. Espino, J. Cox, K. K. A. Tetteh and C. J. Drakeley. "Exposure and Infection to Plasmodium Knowlesi in Case Study Communities in Northern Sabah, Malaysia and Palawan, the Philippines." PLoS Negl Trop Dis 12, no. 6 (2018): e0006432.

TABLE OF CONTENTS

Chapter 1 : INTRODUCTION

1.1. MALARIA AND GLOBAL IMPACT

The term “malaria” comes from Middle Age Italian “mala aria”, literally meaning “bad air”, as it was believed the disease came from swampy marshy areas with bad smelling air[1], related to the miasma theory (diseases caused by noxious polluting air). Malaria is a mosquito borne protozoal disease caused by members of the genus *Plasmodium*. The disease caused imposes a huge public healthcare burden on endemic populations, inhibiting social and economic development in endemic regions around the world[2,3]. In humans, the disease is caused by a number of different *Plasmodium* spp. of which five are the best characterised, *Plasmodium falciparum*, *P. vivax*, *P. ovale (*sub-species *P. o. wallikeri and P. o. curtisi[4])*, *P. malariae* and, most recently confirmed, the zoonotic infection *P. knowlesi*[2,3,5]. There has been some evidence of *P. cynomolgi* infecting humans[6-8] but it has yet to be acknowledged as an established human pathogen. It is genetically similar to *P. knowlesi* and *P. vivax*[9] and frequently confused with *P. vivax* morphologically[10]. Malaria affects up to 200 million individuals worldwide[2,3] (**Figure *1.1***).

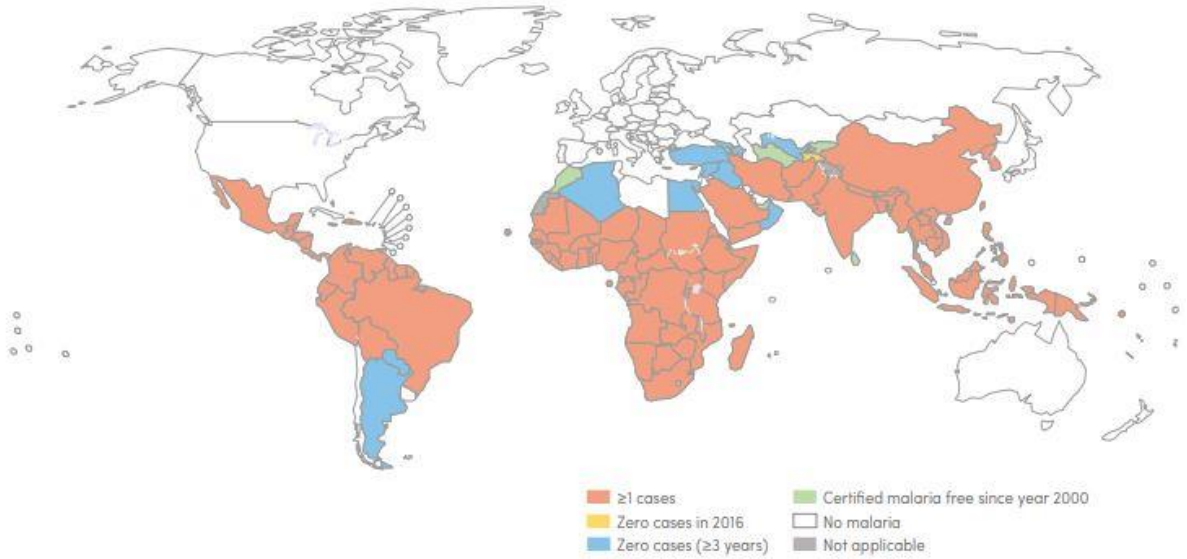


Figure 1.1 Indigenous malaria cases from 2000 and status of countries and territories by 2016. Image from published source[3].

Mortality from malaria in 2016 was approximately 445,000 deaths, with a large proportion of infections due to *P. falciparum* with *P. vivax* not far behind, having the highest rates of mortality in Southeast Asia[2,3,11]. The most prevalent malaria parasite in sub-Saharan Africa is *P. falciparum*, which is responsible for 99% of estimated cases reported in 2016[3] (**Figure *1.2***).

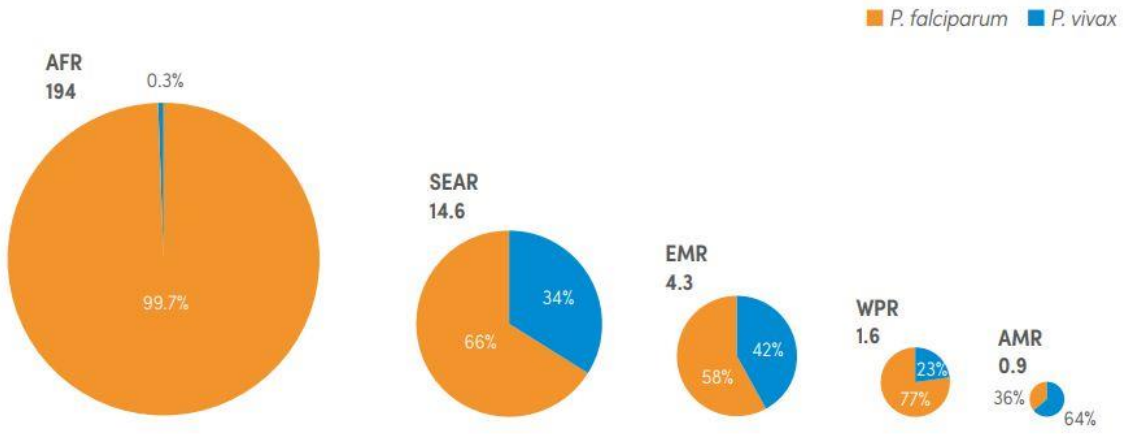


Figure 1.2 Estimated malaria cases in 2016 (millions) by WHO regions. Circle area is proportional to number of estimated cases per region. Image from published source[3].

The greatest burden of disease is largely borne by children under 5 and pregnant women, primarily women undergoing their first pregnancy (primagravidae)[11]. Severe anaemia, organ failure, and rarely, cerebral malaria are symptoms caused by the blood stage of infection and it is at this point that the disease can be fatal. Recent research suggests that the burden of disease caused by *P. vivax* infection has been grossly under reported[12,13]. *P. vivax* is now regarded as a significant public health challenge responsible for at least 75% of all infections in East Africa, Oceania, Central, South and Southeast Asia, the Middle East, Central and South America[14-16] (**Figure *1.2***). *P. knowlesi* is a simian parasite with a close genetic relationship to *P. vivax* and has recently been shown to cause human malaria in Southeast Asia[17], including Thailand, Cambodia, Myanmar, Vietnam, the Philippines and Singapore[18-27] (**Figure *1.3***).

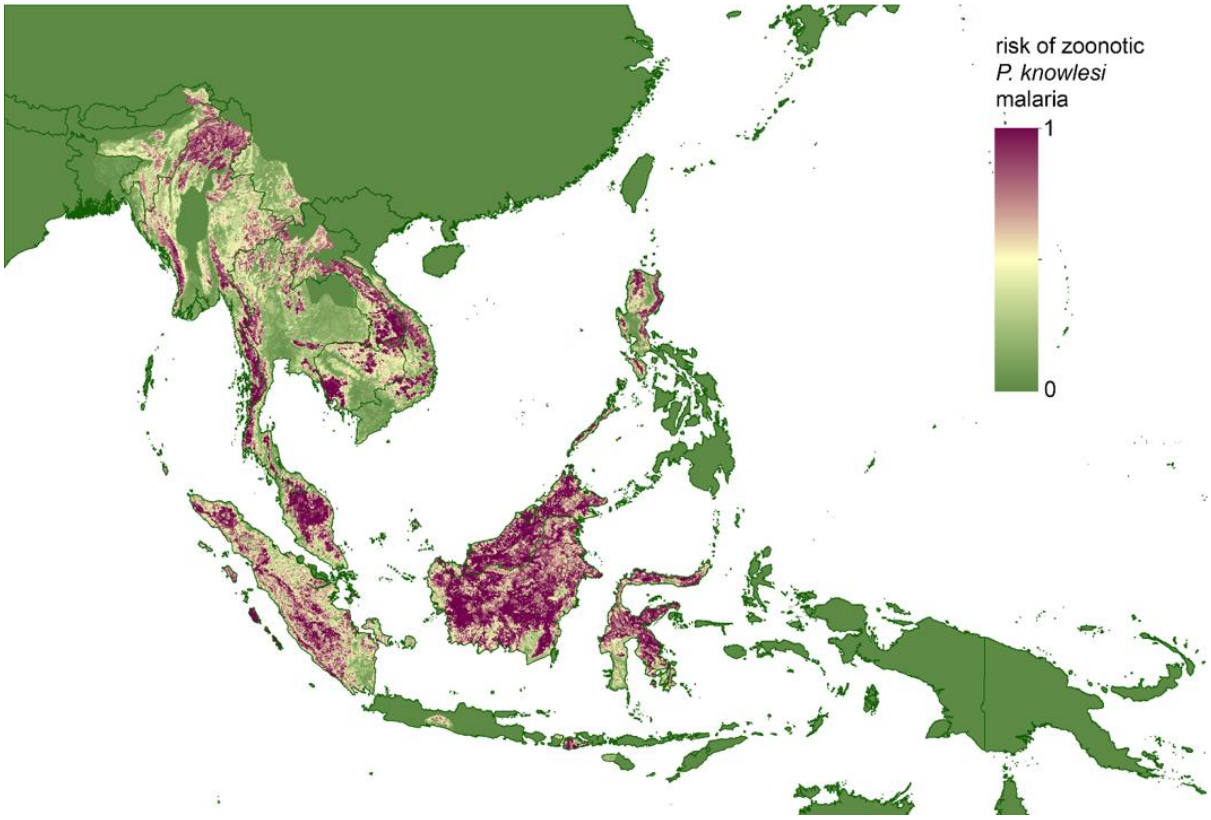


Figure 1.3 Prediction of *P. knowlesi* malaria risk in Southeast Asia. The coloured scale from 0 to 1 indicates risk of *P. knowlesi* malaria. The colours indicates regions with higher risk of *P. knowlesi* malaria, labelled as dark red, and regions with lower risk or no known risk, labelled in green. Image from published source[28].

From the total number of malaria notifications in Sabah, Malaysia, in 2013 (**Figure *1.4***), 62% were caused by *P. knowlesi* and/or *P. malariae*[29]. The World Health Organisation (WHO) Southeast Asia region accounted for 6% of all malaria deaths in 2016, although it was also the region which recorded the largest reduction in mortality rates since 2010 (44% reduction)[3]. Recent evidence suggests that human *P. knowlesi* infections are a growing public health threat in South East Asia, particularly in Malaysia[17,29-32]. It is now the most common cause of clinical malaria in the Malaysian state of Sabah, particularly the Kudat district, where in 2009 87% of 157 patients admitted with malaria infection in Kudat District Hospital were caused by *P. knowlesi*[33]. In 2012, the most common species diagnosed was *P. knowlesi*, with 130 out of 295 (44%) patients recorded as positive[34]. Between 2014 and 2015, 82.9% of malaria diagnosed cases were in fact *P. knowlesi* in Sarawak (http://jknsarawak.moh.gov.my/en/).

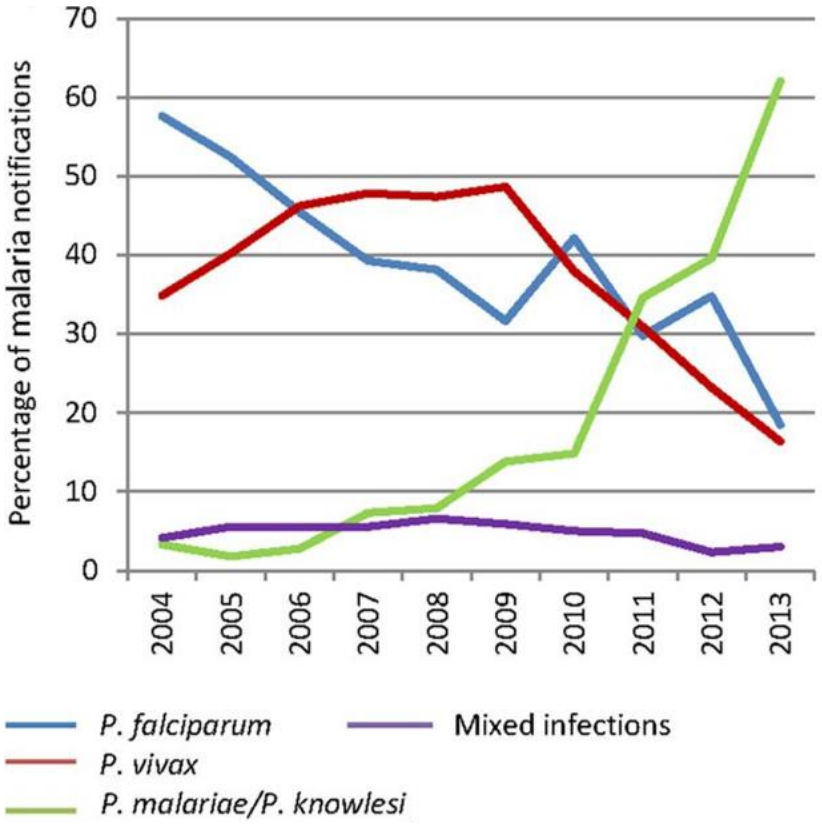


Figure 1.4 Malaria notifications by *Plasmodium* species between 2004 and 2013 in Sabah, Malaysia. Percentage of total malaria notifications for *P. falciparum* (blue), *P. vivax* (red), *P. knowlesi/P. malariae* (green) and mixed infections (purple) in Sabah. Image adapted from published source[29].

1.2. HISTORICAL CONTEXT OF *PLASMODIUM KNOWLESI*

P. knowlesi was first described in 1927, as an infection showing morphological similarity to *P. malariae*[35]. A few years later this same observation was made by Dr. H. Campbell, yet it was Dr. Napier who inoculated 3 different macaques with this unknown *Plasmodium* spp. to scrutinize the effects[36]. It was observed that one of the inoculated macaques developed a severe infection, an event that caught the attention of Dr. Biraj Das Gupta, assistant to Dr. Robert Knowles at the time. The original strain was passaged in Rhesus macaques by Dr. Gupta, who also demonstrated the ability of the parasite to establish infection in humans and observed different blood forms of the parasite[37]. The parasite was named in honour of Dr. Knowles, by Colonel Sinton and Dr. Mulligan who described the 24 hour shizogonic cycle, an observation that classified this organism as a new species of *Plasmodium*[38,39].

Up until 1955, *P. knowlesi* was used in the treatment of neurosyphilis, the most severe form of which is general paralysis of the insane. This treatment used to be called malariotherapy or Malaria Fever Therapy[40]. The high temperatures caused by the malaria fever seemed to supress and kill the spirochete *Treponema pallidum*, the causative agent of syphilis. This method was the standard method of treatment up until the discovery of penicillin and, even then, this therapy was used in combination with penicillin, as penicillin was not widely accepted[41,42]. At the time *P. knowlesi* required no treatment, but after 170 transfers from patient to patient, the infection became nearly as virulent as neurosyphilis and this type of treatment had to be terminated[43,44]. The earliest reported case of a natural infection with *P. knowlesi* was in 1965 when an American traveller returned to the USA with an infection after visiting peninsular Malaysia[45]. This incident indicated that *P. knowlesi* could be transmitted to a human host from mosquitoes.

P. knowlesi is commonly misdiagnosed as *P. malariae*[10,46], but also as *P. falciparum* and *P. vivax*[31,34,47,48] by microscopy due to morphological similarities, which can cause a delay in treatment and mismanagement of the disease, potentially leading to case fatalities[48-50]. *P. knowlesi* has the potential to cause severe disease in Malaysian Borneo and other endemic regions[49], and is now the most common cause of clinical malaria in Malaysia[33,51].

In a study started by Singh *et. al.*[52] in the year 2000 it was shown that more than half of all malaria cases were misdiagnosed microscopically as *P. malariae*. In the state of Sarawak, Malaysian Borneo, thick blood films previously identified by microscopy as *P. malariae* were tested by nested PCR assay and 35 out of 36 slides were actually *P. knowlesi*[53]. However, it has been shown that the primers used for this Polymerase Chain Reaction (PCR) assay randomly cross react with *P. vivax*, creating opportunities for an increase in false positive results[54,55].

Moreover, Lucchi *et. al.*[55] have developed a single-step PCR assay that accurately detects *P. knowlesi* sequences with 100% specificity, though it requires a specific set of skills and infrastructure making it less field applicable.

To date, no reliable and cost effective serosurveillance tools have been established to identify and monitor *P. knowlesi* infections in endemic regions. The development of species-specific tools are a necessity to better distinguish human serological responses between *Plasmodium* spp. to improve our understanding of immunity to these infections and help define the geographical boundaries of infection. Serosurveillance is a powerful approach that utilizes species-specific antibodies as tools to improve the monitoring of exposure, transmission and immunity towards diseases. Such reagents are currently lacking for *P. knowlesi.* The few recombinant reagents that do exist are not truly species-specific and as a result are unable to distinguish between the two infections due to the high level of sequence homology at the protein level. The development of well validated species-specific tools would represent a vital resource that would be of great importance to the malaria research

community. These protein reagents will also serve as an important tool to help identify potential correlates of immunity.

1.3. *PLASMODIUM KNOWLESI* HOSTS AND VECTORS

Transmission of the parasite is by the *Anopheles leucosphyrus* group of mosquito vectors, which is attracted to both monkeys and humans, meaning the geographic boundaries of *P. knowlesi* is restricted by the distribution of these vectors and their macaque hosts[30] (**Figure *1.5***). In the 1960s *A. hackeri* was the first mosquito to be identified as a vector for transmission of *P. knowlesi* in macaques[56], though it is not attracted to humans. More recently, *A. latens*[57,58] in Sarawak*, A. cracens*[59,60] in peninsular Malaysia and *A. balabacensis* were found to be transmission vectors, with the later found to be the predominant vector species for *P. knowlesi* in Sabah, Malaysia, in mosquito collections using human landing catches[61,62].

The langur (*Presbytis melalophos*) and the Old World pig-tailed (*Macaca nemestrina*) and long-tailed, or crab-eating, macaques (*M. fascicularis*) are the natural hosts for *P. knowlesi*.

P. knowlesi can also be maintained in one of its experimental host, with usually fatal outcomes, the Rhesus macaque (*M. mullata*)[10], which is hypothesised to have diverged from a long-tailed macaque ancestor[63]. These macaques are also the natural host for four other *Plasmodium* species: *P. cynomolgi*, *P. fieldi*, *P. coatneyi* and *P. inui*[10,64].

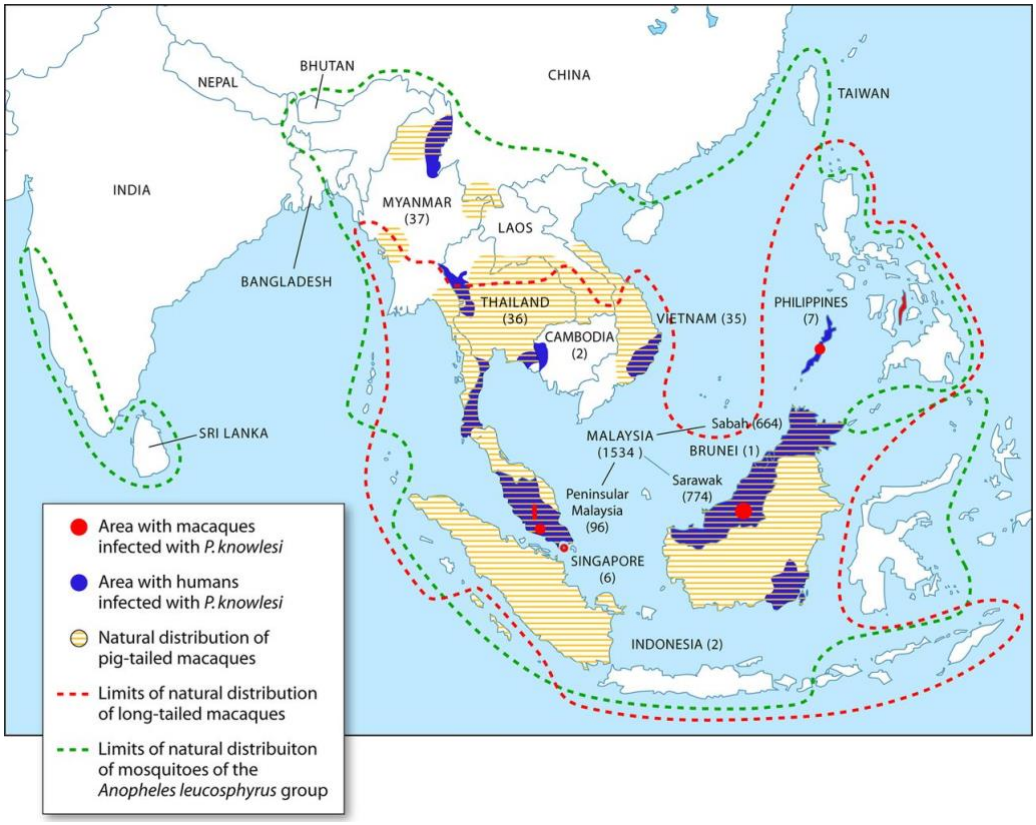


Figure 1.5 Distribution of *P. knowlesi* natural hosts and vectors in Southeast Asia. Green lines delineate limits of natural distribution of the mosquito vectors, red lines delineate limits of natural distribution of long-tailed macaques, yellow areas represent the natural distribution of pig-tailed macaques, blue areas represent areas with humans infected with *P. knowlesi* and red dots represent areas with macaques infected with *P. knowlesi*. Image from published source[31].

Compared to any other non-human primates (NHP) genus, *Macaca spp*. occupy a range of different habitats under different climates, making these the most widely distributed NHPs[65]. The preference for these different habitats is different for each species, contributing to species diversification. Macaques' ability to coexist with humans surpasses that of all other non-human primates. As a result, much of the data on macaque behaviour and ecology come from settings in which the monkeys derive large parts of their food either directly or indirectly from humans[65]. Macaques occupy a wider range of habitats and climates than any other NHP genus. Even more, the habitat preferences and foraging strategies for individual species vary in many ways, a factor that has contributed to the diversity of species. The ecological differences among macaques have been most clearly documented for the pig-tailed and the long-tailed macaques[65]. The pig-tailed macaques have a preference for

living in upland and hilly environments, travelling more on the ground and frequently feeding in trees with mostly quadrupedal walking and running[65]. The long-tailed macaques, on the other hand, prefer living in lowland secondary forests near rivers and are usually found in the lower levels of the main canopy as they are mainly an arboreal species, where they feed and travel, although they use all tree levels including the ground[65]. Interestingly, the long-tailed macaque was shown to easily control *P. knowlesi* infections, unlike the closely related Rhesus[66], which requires several infections cycles before giving any indication of parasitemia control[67].

The most common laboratory primates are macaques, particularly the Rhesus macaque but others as well, due to them being thoroughly studied throughout the years on many different diseases[68] such as human immunodeficiency virus (HIV)[69], cytomegalovirus (CMV)[70] and tuberculosis (TB)[71], just to name a few. The physiology of these animals, as well as their captured behaviour, are well known, enabling researchers to study pathogenic mechanisms, new treatments and vaccines. With *P. knowlesi* being a simian malaria, many studies used rhesus macaques as experimental animal models to gain more insight about this particular malaria parasite such as antigenic variation and protective immunity[67,72-75]. Studies have predicted that simian and human malaria species share similar regulatory mechanisms[76-78].

1.4. *PLASMODIUM* LIFE CYCLE AND CLINICAL PRESENTATION

Plasmodium spp. parasites are obligate intracellular organisms. The life cycle of *P. knowlesi*[64] is similar to other *Plasmodium* infections with notable exceptions. In contrast to *P. falciparum, P. vivax and P. ovale,* which follow a 48 hour erythrocytic cycle, and *P. malariae* a 72 hour cycle, *P. knowlesi* follows a 24 hour quotidian cycle, meaning a mild infection could quickly progress to a complicated or severe disease which could lead to increased mortality[38,39,64,79]. The parasites life cycle begin when the female *Anopheles* mosquito takes a blood meal from one of the vertebrate hosts,

a human or a primate, injecting sporozoites into the blood stream which then takes the parasites into the liver, beginning the liver stage of the cycle[64] (**Figure 1.6**). The sporozoites invade the hepatocytes, where they undergo asexual multiplication and mature into liver stage schizonts. Mature hepatic schizonts will rupture, releasing merozoites into the blood stream where they invade reticulocytes, young red blood cells (RBC), initiating the blood stage of the life cycle. *P. knowlesi* produces fewer, larger merozoites per schizonts (10 to 16), compared to *P. falciparum* (20 to 30)[10]. The merozoite within the RBC develops into a trophozoite which matures and undergoes asexual multiplication, forming a blood stage schizont which then erupts, releasing numerous merozoites. Some merozoites will invade new RBCs and repeat the cycle while others will develop into male or female gametocytes, within the RBCs. Once these gametocytes are taken up in a blood meal by a mosquito vector, they continue their development and undergo sexual multiplication (**Figure 1.6**).

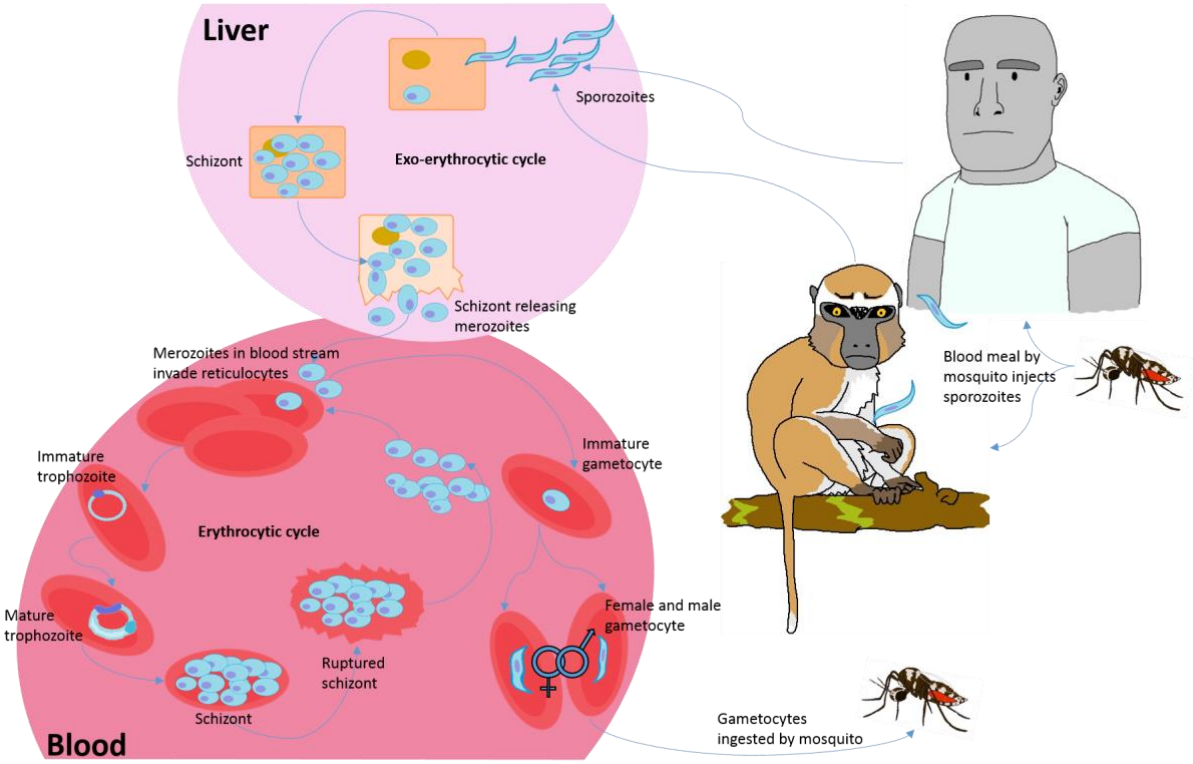


Figure 1.6 ***Plasmodium knowlesi*** **life cycle in human and macaque vectors.** Sporozoites are innoculated into the vertebrate host from an infected mosquito via the skin during a blood meal. The invasive sporozoites then migrate to the liver via the blood stream, invading hepatocytes developing into mature liver stage schizonts. The merozoites enter the blood stream following rupture of the liver stage schizont going on to invade

reticulocytes and initiating the blood stage of infection. A proportion of the invading merozoites develop into male and female gametocytes that are then taken up during a mosquito blood meal. Verbal consent was obtained from Gabriel Fowler to use and adapt single images for this figure.

The rapid replication rate means *P. knowlesi* can reach high parasite densities (>5000 parasites/µL of blood) in a short amount of time[52,64]. The high parasitemia may be due to *P. knowlesi* preferring to invade young RBCs (reticulocytes), but can adapt to invade RBCs of any age[80]. Clinical signs and symptoms are absent during the liver stage of the parasite life cycle and only start to manifest during the blood stage, which can vary in duration[64]. If left untreated, parasitemia burden will increase with each replicative cycle.

Severe disease is caused by high parasite densities, similarly to *P. falciparum*[51], and parasite adaptation allows an increase in proliferation in human hosts and in virulence of infections. In the natural hosts, *P. knowlesi* infection causes a mild and ephemeral disease with chronically low parasitemia levels[10,36]. Antigenic variation was first shown to occur in *P. knowlesi*[72], indicating that antigenic variation is associated with parasite virulence, underlining chronic infections. It is an essential process within the parasite life cycle, securing transmission and avoidance of the host immune system. However, in the human host, the disease can cause a range of different symptoms including fever patterns associated with the release of merozoite, fatigue and hyperparasitemia[34,50,51]. Paradoxically, anaemia and cerebral malaria have so far not been observed[50], although *P. knowlesi* parasites have been found in the brain[81].

1.5. IMPACT OF DEFORESTATION ON *P. KNOWLESI* INCIDENCE

It is hypothesized that the dominance of *P. knowlesi* in the Borneo region is caused by a change in the behavioural patterns between humans and macaques due to deforestation[82]. Another hypothesis is the reduction of competition with *P. vivax* and *P. falciparum*, or the cross-species protection from these resulting from the reduction

in the prevalence of these species. It has been previously suggested that the main driver for this emergence of *P. knowlesi* in Malaysia are the changes in land usage and deforestation for the expansion of agriculture[31,82] (**Figure *1.7***), which affects vector populations globally[83].

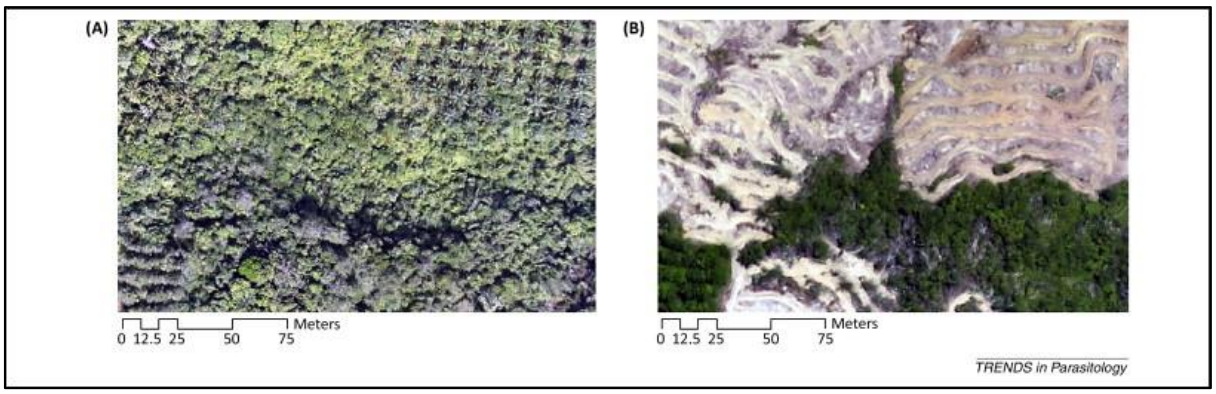


Figure 1.7 Mapping of land coverage changes in Sabah, Malaysia. Study site in February 2014 (A) and May 2014 (B). Image from published source[84].

One of the reasons for the lack of a large scale *P. knowlesi* epidemic is because transmission is restricted to the *A. leucosphyrus* mosquito vector, which is attracted to both monkeys and humans, but prefers to feed in the forest and forest fringe after dusk[57,62]. This shows that *P. knowlesi* can be transmitted from monkeys to humans if they share the same environment, which is being propelled by the increase in urbanisation and deforestation, bringing humans and macaques closer together[85]. In a study led by Barber *et. al.*[34], farmers and plantation workers had higher parasitemia than patients with other occupations. Another study led by Fornace *et.*

al.[82] in Sabah, Malaysia, also strongly suggests an association between environmental changes and an increase in *P. knowlesi* incidence (**Figure *1.8***).

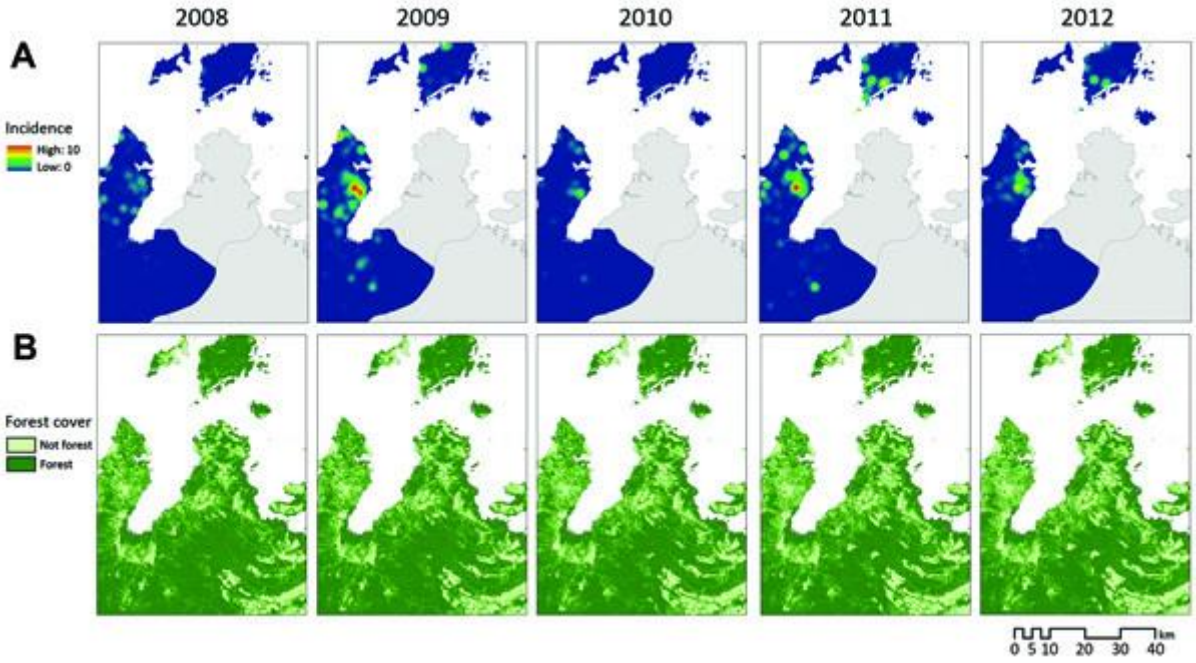


Figure 1.8 *Plasmodium knowlesi* infection incidence (A) compared to forest coverage (B) between 2008 and 2012 in Sabah, Malaysia. The incidence is shown as the mean of estimated annual parasite incidence, expressed as cases per 1,000 person per year. Image from published source[82].

These environmental changes, in addition to the change in distribution of people, have a great impact on infectious disease risks and have been proposed as a driver of infectious disease emergence and transmission[86-89]. Total annual notifications of *P. knowlesi* infections in Sabah, Malaysia, have increased from 1% to 35% between 1992 and 2013, acknowledging and increase in *P. knowlesi* incidence[29,90]. It is still open to question if these values are due to an increase in the accuracy of identification or an actual emergence of the species. This increase in *P. knowlesi* could compromise malaria elimination strategies in several countries if not attended to as soon as possible.

1.6. SEROLOGY AND ANTIBODIES

Serology has been extensively used as an epidemiological tool in malaria research by measuring antibody reactivity at the community level in endemic countries[91-93]. Recent and historical changes in malaria transmission and disease exposure can be measured using different serological methods of identifying antibody responses[94-100]. Natural antimalarial antibody acquisition and maintenance is dependent on exposure to malarial infection and these antibodies can be used as markers to measure transmission intensity[95,101]. A slow and gradual loss of antibody levels indicates different levels of exposure to infection and that memory antibody response exists, though these responses are not yet established in children as they are in adults[102,103]. Sensitive and specific serosurveillance tools are a necessity to measure and monitor malaria incidence in areas of declining transmission[104,105], particularly for *P. knowlesi*, as none currently exist. It is important to know how appropriate the antigenic target is, ensuring it detects the correct species for instance, as well as the sensitivity of the antibody detecting assay.

Serosurveillance tests in epidemiological studies have specific requirements and the surveys done to predict levels of transmission and disease exposure are generally very large. The serological test should be high throughput compatible, quick, reproducible in the field and preferably low cost. Multiple tests have been developed to detect malarial antibodies in epidemiological studies, such as the Complement Fixation Test (CFT)[106,107], which was one of the earliest developed, the Indirect Haemagglutination Assay (IHA)[108], the Immunofluorescence Antibody Test (IFAT), the Enzyme-Linked Immunosorbent Assay (ELISA)[109], the protein microarray[110,111] and the Multiplex Bead Assay (MBA)[112-115].

For the purpose of this thesis we will focus more on the Enzyme-Linked Immunosorbent Assay (ELISA), protein microarray and cytometric bead assays (MBA). The latter two are high throughput methods, which are currently being

successfully adapted to study antibody responses to malarial antigens. Most studies have used the ELISA method to determine malaria endemicity[99,101,116,117]. The ELISA was first used to detect malarial antibodies in a study led by Voller *et. al.* in 1975[109]. Crude *P. knowlesi* antigens were successfully used to detect antibody in *P. falciparum* and *P. vivax* positive individuals. Protein microarrays and MBAs have a very similar technique to the ELISA but can screen hundreds of recombinants in a single assay and are more sensitive assays. Both methods are fairly new as serosurveillance platforms but because these are easily multiplexed, there is a possibility of looking at a much wider range of serological responses using minimal resources.

Until recently, most epidemiological studies on malaria have used the ELISA to look at total immunoglobulin G (IgG) reactivity in endemic populations[101,116,118-121]. Immunoglobulins (Igs) are produced by B cells and have an important role in the human immune response as these recognise and bind to potentially threatening antigens to help destroy them[122].

Igs play an important role against malaria as repeated exposure to malaria parasites induces the development of antibodies against the parasites. Antibodies that target blood-stage malarial antigens have a number of possible functions and specificities such the inhibition of RBC invasion by merozoites, inhibition of rosette formation, reducing hepatocyte invasion and reducing infectivity after being ingested by mosquitoes, just to name a few[123].

Igs are glycoproteins made up of four polypeptide chains that consist of two identical light chains (L) and two identical heavy chains (H) which are connected by disulfide bridges. Both L and H chains contain a variable domain, forming the antigen-binding site or paratope, and constant domain, which determines the antibody functional class and location in the body (**Figure *1.9***). The different types of Igs can be

distinguished by the type of H chain they contain, grouping the Igs into specific classes, or isotypes[124].

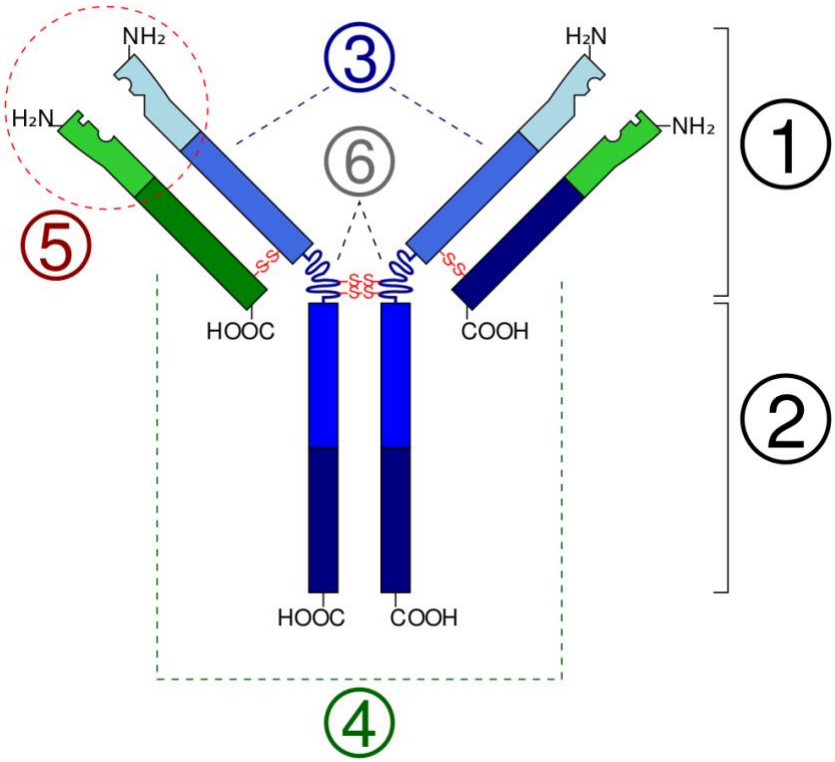


Figure 1.9 Immunoglobulin basic unit characteristics. 1) Fab region; 2) Fc region; 3) Heavy chain with one variable domain followed by one constant domain and two more constant domains; 4) Light chain with one variable domain followed by one constant domain; 5) Paratope; 6) hinge regions. SS in red represent disulfide bridges. Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.

The immune response changes according to the antigens being presented to the immune system, therefore knowing different isotype levels can help interpret the immune response.

In humans there are 5 different isotypes, or classes, each with their own structure, function and ability to recognise different antigens[125]. These are differentiated by differences in the H chains: IgG contain H chains known as γ-chains, IgM contain μ-

chains, IgA contain α-chains, IgE contain ε-chains and IgD contain δ-chains (**Figure *1.10***).

Property	Immunoglobulin Class				
	IgG	IgM	IgA	IgE	IgD
Number of monomers	1	5	2	1	1
Antibody in serum (%)	80	5–8	10–15	<1	<1
Activates complement	Yes	Yes	No	No	No
Crosses placenta	Yes	No	No	No	No
Neutralizes bacterial toxins	Yes	Yes	Yes	No	No
Binds to phagocytes by Fc	Yes	No	No	No	No
Binds to basophils and mast cells	No	No	No	Yes	No
Additional properties	Principal antibody of secondary antibody response	First antibody formed in a primary antibody response	Monomer in serum; dimer secreted onto epithelial surfaces	Role in allergic reactions; effective against parasitic worms	Receptor on B cell surface

Figure 1.10 Immunoglobulin isotypes and functional characteristics. Image from published source[124].

The major class of Igs is IgG, a monomer, having 75% abundance in serum in relation to total Igs present[125] and has the longest serum half-life of all isotypes[126]. It is the only Ig with the ability to cross the placenta and penetrate extravascular areas and it is also the main antibody of secondary responses[125]. IgG is thought to be the main antibody to mediate protection in new-borns during the first months of life as it is transferred to the foetus during the last trimester of pregnancy[127-130]. There are 4 different IgG subclasses in human, with the differences found in the structure of the H chains and are ranked in the order of abundance in serum (IgG1>IgG2>IgG3>IgG4)[126,131]. IgG and its subclasses are the most studied antibodies in malaria and have been associated with protection against malaria infection[132-137].

IgM, usually presented as a pentamer and more rarely as an hexamer, is the first antibody during an immune response and it is the first Ig to be produced by neonates[125]. It is responsible for agglutination and is involved in complement activation. A single IgM molecule can activate the classical complement system,

whilst several IgG molecules are needed for this[125]. IgM can have high avidity due to the multimeric interactions between the pentameric molecule and the antigen[126]. Because IgM is the first Ig to be expressed during B cell development, the H chains associate with variable components that have not undergone much somatic mutation, meaning they are more immature and have lower affinity, tending to be more poly-reactive and stickier than other isotypes[126]. Merozoite Surface Protein 1-specific (MSP1) IgM expressing memory B cells were found to be early responders to malarial re-infections, followed by IgG expressing memory B cells[138]. In a separate study, IgM antibody responses to a number of *P. falciparum* antigens were shown to be associated with protection to malaria in Ghanaian children[139]. It was also observed that IgM responses were significantly higher in a malaria protected ethnic group compared to the other[140,141]. IgM is potentially an important antibody in the study of malarial infections and there has been an increased interest in this isotype. IgA is the third major class of Ig, representing approximately 15% of all Igs in serum, more than IgM and considerably less than IgG[126] and it is generally presented as a monomer[142]. In mucosal tissue and in secretions (secretory IgA), including saliva and breast milk, IgA is presented as a dimer and levels are much higher than IgG[143], contributing up to at least 50% of protein in colostrum. Secretory IgA in breastmilk has been shown to inhibit parasite growth in vitro[144]. Most of IgA is secretory. In humans there are two known IgA subclasses, IgA1, which is the most dominant in serum (over 85% of IgA) eliciting a good immune response to protein antigens, and IgA2, which is the most dominant in mucosal surfaces (60% of IgA) eliciting a good immune response to polysaccharide and lipopolysaccharide antigens[125,145]. The main role of IgA is to protect mucosal surfaces from invading pathogens and toxins. IgA responses appear to follow IgM and IgG responses in malaria infections[146]. With most information about IgA coming from malaria-caused nephropathy cases[147,148], the role of this antibody in malaria infection is yet to be described.

IgE, presented as a monomer, is the least common Ig having the lowest concentration in serum with a short half-life and is associated with hypersensitivity, allergies and worm parasite infections[126]. It is a very potent Ig due to its high affinity and is mostly present on the surface of mast cells, Langerhans cells, eosinophils and basophils[125,126]. When cell bound IgE encounter multivalent antigens, these stimulate the degranulation of cells and the release of inflammatory mediators such as histamine and prostaglandins, causing the typical symptoms of allergies such as mucus secretion and sneezing for the expulsion of remaining allergens[125,149]. The measurement of IgE levels in individuals with allergic conditions can be used for diagnostic purposes. In a study led by Dottorini *et. al.*[150] an association was found between IgE reactivity profiles and atopic diseases (i.e. asthma, eczema) indicating that particular patterns of reactivity were related to asthmatic individuals. IgE antibody levels, contrary to IgG, have been shown to be higher in individuals suffering from severe malaria when compared to uncomplicated malaria and this was positively correlated with high parasitemia levels, suggesting a pathogenic role for this isotype[151-153]. IgE antibody-antigen complexes induce the production of tumour necrosis factor (TNF) cytokines[153-155], which are involved in *P. falciparum* disease severity when excessively produced[156,157]. However, the opposite has also been seen, with low IgE levels seen in individuals suffering from severe malaria, although not statistically significant[158]. Higher levels of this antibody were seen in asymptomatic individuals in a study in Tanzania[159] and the same was observed in a malaria protected tribe (Fulani tribe) when compared to the less protected neighbouring tribe (Dogon tribe) in Mali[160]. The role of this antigen in malaria infections is still unclear.

Lastly, IgD, presented as monomers, are also the least common Ig with a short half-life[161], mostly due to its sensitivity to proteolysis[126]. The function of this Ig is largely unknown. IgD was has been identified in bony cartilaginous fishes from a time when the adaptive immune system first evolved[162], indicating that this antibody is highly

preserved. This implies that this antibody has important functions in the survival of the host[163]. These antibodies are usually found on the surface of B cells and participate in B cell receptor signalling alongside membrane IgM, although the quantity of IgM is much lower than IgD[125,126]. Geisberger *et. al.* have proposed that this Ig could be associated with the regulation of B cells at specific developmental stages[164]. Research has suggested that IgD is part of an immune surveillance strategy at mucosal sites of antigen entry such as nasopharyngeal cavities[161,165-167]. One study reported that chronic malaria infections raised the concentration of serum IgD[168], although it was also reported that changes in IgD levels were not very evident in children from The Gambia with acute *P. falciparum* infection[169]. Another study around the same time period reported no relationship between IgD and malaria infection, although serum levels appeared to slowly rise throughout childhood[170]. Most of the available research on IgD has been about membrane bound IgD[161]. Children in Kenya from malaria endemic regions were found to have suppressed circulating IgD+CD27+ memory B cells compared to children living in areas with unstable malaria transmission, indicating that early exposure to *P. falciparum* could dysregulate the development of B cell subsets in children[171].

The importance and activity of the different isotypes and their relationship to one another in malaria has only been studied for IgG and IgM, with the other isotypes poorly studied. The differing isotype profiles in infections may be related to disease progression and severity, therefore it is important to measure and interpret these in order to better understand host immune responses towards malaria infection.

The full genome of *P. knowlesi* (H strain, Pk1 (A+) clone) has recently been sequenced by whole-genome shotgun sequencing. The size of the *P. knowlesi* genome is approximately 23.5 Mb across 14 chromosomes ranging from 0.84 Mb and 3.16 Mb[172]. With this data available to interrogate, it is now possible to develop novel sensitive and specific *P. knowlesi* tools tailored for the serosurveillance of malaria caused by this organism.

REFERENCES

1 Hempelmann, E. & Krafts, K. Bad air, amulets and mosquitoes: 2,000 years of changing perspectives on malaria. *Malaria journal* **12**, 232, doi:10.1186/1475-2875-12-232 (2013).

2 World Health Organization. World Malaria report: 2016. (2016).

3 World Health Organization. World Malaria report: 2017. (2017).

4 Oguike, M. C. *et al.* Plasmodium ovale curtisi and Plasmodium ovale wallikeri circulate simultaneously in African communities. *Int J Parasitol* **41**, 677-683, doi:10.1016/j.ijpara.2011.01.004 (2011).

5 Cullen, K. A., Arguin, P. M., Division of Parasitic, D., Malaria, C. f. G. H. C. f. D. C. & Prevention. Malaria surveillance--United States, 2011. *Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C. : 2002)* **62**, 1-17 (2013).

6 Eyles, D. E., Coatney, G. R. & Getz, M. E. Vivax-type malaria parasite of macaques transmissible to man. *Science* **131**, 1812-1813 (1960).

7 Contacos, P. G., Elder, H. A., Coatney, G. R. & Genther, C. Man to man transfer of two strains of Plasmodium cynomolgi by mosquito bite. *Am J Trop Med Hyg* **11**, 186-193 (1962).

8 Ta, T. H. *et al.* First case of a naturally acquired human infection with Plasmodium cynomolgi. *Malaria journal* **13**, 68, doi:10.1186/1475-2875-13-68 (2014).

9 Ansari, H. R. *et al.* Genome-scale comparison of expanded gene families in Plasmodium ovale wallikeri and Plasmodium ovale curtisi with Plasmodium malariae and with other Plasmodium species. *Int J Parasitol* **46**, 685-696, doi:10.1016/j.ijpara.2016.05.009 (2016).

10 Coatney, G. R., Collins, W. E. & Contacos, P. G. *The Primate malarias*. (U.S. National Institute of Allergy and Infectious Diseases, 1971).

11 World Health Organization. World Malaria report: 2014. (2014).

12 Price, R. N., Douglas, N. M. & Anstey, N. M. New developments in Plasmodium vivax malaria: severe disease and the rise of chloroquine resistance. *Current opinion in infectious diseases* **22**, 430-435, doi:10.1097/QCO.0b013e32832f14c1 (2009).

13 Price, R. N. *et al.* Vivax malaria: neglected and not benign. *Am J Trop Med Hyg* **77**, 79-87 (2007).

14 Carlton, J. M., Das, A. & Escalante, A. A. Genomics, population genetics and evolutionary history of Plasmodium vivax. *Advances in parasitology* **81**, 203-222, doi:10.1016/b978-0-12-407826-0.00005-9 (2013).

15 Guerra, C. A. *et al.* The international limits and population at risk of Plasmodium vivax transmission in 2009. *PLoS neglected tropical diseases* **4**, e774, doi:10.1371/journal.pntd.0000774 (2010).

16 Mueller, I. *et al.* Key gaps in the knowledge of Plasmodium vivax, a neglected human malaria parasite. *The Lancet infectious diseases* **9**, 555-566, doi:10.1016/s1473-3099(09)70177-x (2009).

17 White, N. J. Plasmodium knowlesi: the fifth human malaria parasite. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **46**, 172-173, doi:10.1086/524889 (2008).

18 Orth, H. *et al.* Plasmodium knowlesi infection imported to Germany, January 2013. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **18** (2013).

19 Jongwutiwes, S. *et al.* Plasmodium knowlesi Malaria in humans and macaques, Thailand. *Emerging infectious diseases* **17**, 1799-1806, doi:10.3201/eid1710.110349 (2011).

20 Khim, N. *et al.* Plasmodium knowlesi infection in humans, Cambodia, 2007-2010. *Emerging infectious diseases* **17**, 1900-1902, doi:10.3201/eid1710.110355 (2011).
21 Seilmaier, M. *et al.* Severe Plasmodium knowlesi infection with multi-organ failure imported to Germany from Thailand/Myanmar. *Malaria journal* **13**, 422, doi:10.1186/1475-2875-13-422 (2014).
22 Jiang, N. *et al.* Co-infections with Plasmodium knowlesi and other malaria parasites, Myanmar. *Emerging infectious diseases* **16**, 1476-1478, doi:10.3201/eid1609.100339 (2010).
23 Luchavez, J. *et al.* Human Infections with Plasmodium knowlesi, the Philippines. *Emerging infectious diseases* **14**, 811-813, doi:10.3201/eid1405.071407 (2008).
24 Ng, O. T. *et al.* Naturally acquired human Plasmodium knowlesi infection, Singapore. *Emerging infectious diseases* **14**, 814-816, doi:10.3201/eid1405.070863 (2008).
25 Cox-Singh, J. Knowlesi malaria in Vietnam. *Malaria journal* **8**, 269, doi:10.1186/1475-2875-8-269 (2009).
26 Van den Eede, P. *et al.* Human Plasmodium knowlesi infections in young children in central Vietnam. *Malaria journal* **8**, 249, doi:10.1186/1475-2875-8-249 (2009).
27 Nakaviroj, S., Kobasa, T., Teeranaipong, P., Putaporntip, C. & Jongwutiwes, S. An Autochthonous Case of Severe Plasmodium knowlesi Malaria in Thailand. *Am J Trop Med Hyg*, doi:10.4269/ajtmh.14-0610 (2014).
28 Shearer, F. M. *et al.* Estimating Geographical Variation in the Risk of Zoonotic Plasmodium knowlesi Infection in Countries Eliminating Malaria. *PLoS neglected tropical diseases* **10**, e0004915, doi:10.1371/journal.pntd.0004915 (2016).
29 William, T. *et al.* Changing epidemiology of malaria in Sabah, Malaysia: increasing incidence of Plasmodium knowlesi. *Malaria journal* **13**, 390, doi:10.1186/1475-2875-13-390 (2014).
30 Moyes, C. L. *et al.* Defining the geographical range of the Plasmodium knowlesi reservoir. *PLoS neglected tropical diseases* **8**, e2780, doi:10.1371/journal.pntd.0002780 (2014).
31 Singh, B. & Daneshvar, C. Human infections and detection of Plasmodium knowlesi. *Clin Microbiol Rev* **26**, 165-184, doi:10.1128/cmr.00079-12 (2013).
32 Lubis, I. N. *et al.* Contribution of Plasmodium knowlesi to multi-species human malaria infections in North Sumatera, Indonesia. *The Journal of infectious diseases*, doi:10.1093/infdis/jix091 (2017).
33 Barber, B. E. *et al.* Plasmodium knowlesi malaria in children. *Emerging infectious diseases* **17**, 814-820, doi:10.3201/eid1705.101489 (2011).
34 Barber, B. E. *et al.* A prospective comparative study of knowlesi, falciparum, and vivax malaria in Sabah, Malaysia: high proportion with severe disease from Plasmodium knowlesi and Plasmodium vivax but no mortality with early referral and artesunate therapy. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **56**, 383-397, doi:10.1093/cid/cis902 (2013).
35 Franchini, G. Su di un plasmodio pigmentato di una scimmia. *Archivio Italiano di Scienze Mediche Coloniali*, 187–190 (1927).
36 Napier, I. E. & Campbell, H. G. M. Observations on a Plasmodium infection which causes haemoglobinuria in certain species of monkey. *Indian Medical Gazette*, 151-160 (1932).
37 Knowles, R. & Das Gupta, B. M. A study of monkey-malaria and its experimental transmission to man. *Indian Medical Gazette*, 301-320 (1932).
38 Sinton, J. A. & Mulligan, H. W. A critical review of the literature relating to the identification of the malaria parasites recorded from monkeys of the families Cercopithecidae and Colobidae. *Records of Malaria Survey of India III*, 357-380 (1932).

39 Sinton, J. A. & Mulligan, H. W. A critical review of the literature relating to the identification of the malaria parasites recorded from monkeys of the families Cercopithecidae and Colobidae. *Records of Malaria Survey of India III*, 381-443 (1933).

40 Karamanou, M., Liappas, I., Antoniou, C., Androutsos, G. & Lykouras, E. Julius Wagner-Jauregg (1857-1940): Introducing fever therapy in the treatment of neurosyphilis. *Psychiatrike = Psychiatriki* **24**, 208-212 (2013).

41 Kaplan, R. M. Syphilis, sex and psychiatry, 1789-1925: Part 2. *Australasian psychiatry : bulletin of Royal Australian and New Zealand College of Psychiatrists* **18**, 22-27, doi:10.3109/10398560903295717 (2010).

42 Albert, M. R. Fever therapy for general paresis. *International journal of dermatology* **38**, 633-637 (1999).

43 Ciuca, M. *et al.* Contribution à l' étude de la virulence du Pl. knowlesi chez l'homme. Caractères de la maladie et biologie du parasitie. *Archives Roumaines de Pathologie Experimentale et de Microbiologie*, 5-28 (1937).

44 Ciuca, M., Chelarescu, M., Sofletea, A., Constantineascu, P. & Terineanu, E. Contribution expèrimentale a l'ètude de l'immunitè dans le paludisme. *Edition de l'Academique de la Republique Populare. Roumaine*, 1-108 (1955).

45 Chin, W., Contacos, P. G., Coatney, G. R. & Kimball, H. R. A NATURALLY ACQUITED QUOTIDIAN-TYPE MALARIA IN MAN TRANSFERABLE TO MONKEYS. *Science* **149**, 865 (1965).

46 Lee, K. S., Cox-Singh, J. & Singh, B. Morphological features and differential counts of Plasmodium knowlesi parasites in naturally acquired human infections. *Malaria journal* **8**, 73, doi:10.1186/1475-2875-8-73 (2009).

47 Barber, B. E., William, T., Grigg, M. J., Yeo, T. W. & Anstey, N. M. Limitations of microscopy to differentiate Plasmodium species in a region co-endemic for Plasmodium falciparum, Plasmodium vivax and Plasmodium knowlesi. *Malaria journal* **12**, 8, doi:10.1186/1475-2875-12-8 (2013).

48 Rajahram, G. S. *et al.* Deaths due to Plasmodium knowlesi malaria in Sabah, Malaysia: association with reporting as Plasmodium malariae and delayed parenteral artesunate. *Malaria journal* **11**, 284, doi:10.1186/1475-2875-11-284 (2012).

49 Cox-Singh, J. *et al.* Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **46**, 165-171, doi:10.1086/524888 (2008).

50 William, T. *et al.* Severe Plasmodium knowlesi malaria in a tertiary care hospital, Sabah, Malaysia. *Emerging infectious diseases* **17**, 1248-1255, doi:10.3201/eid1707.101017 (2011).

51 Daneshvar, C. *et al.* Clinical and laboratory features of human Plasmodium knowlesi infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **49**, 852-860, doi:10.1086/605439 (2009).

52 Singh, B. *et al.* A large focus of naturally acquired Plasmodium knowlesi infections in human beings. *Lancet* **363**, 1017-1024, doi:10.1016/s0140-6736(04)15836-4 (2004).

53 Lee, K. S., Cox-Singh, J., Brooke, G., Matusop, A. & Singh, B. Plasmodium knowlesi from archival blood films: further evidence that human infections are widely distributed and not newly emergent in Malaysian Borneo. *Int J Parasitol* **39**, 1125-1128, doi:10.1016/j.ijpara.2009.03.003 (2009).

54 Imwong, M. *et al.* Spurious amplification of a Plasmodium vivax small-subunit RNA gene by use of primers currently used to detect P. knowlesi. *Journal of clinical microbiology* **47**, 4173-4175, doi:10.1128/jcm.00811-09 (2009).

55 Lucchi, N. W. *et al.* A new single-step PCR assay for the detection of the zoonotic malaria parasite Plasmodium knowlesi. *PLoS One* **7**, e31848, doi:10.1371/journal.pone.0031848 (2012).
56 Wharton, R. H. & Eyles, D. E. Anopheles hackeri, a vector of Plasmodium knowlesi in Malaya. *Science* **134**, 279-280 (1961).
57 Vythilingam, I. *et al.* Natural transmission of Plasmodium knowlesi to humans by Anopheles latens in Sarawak, Malaysia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **100**, 1087-1088, doi:10.1016/j.trstmh.2006.02.006 (2006).
58 Tan, C. H., Vythilingam, I., Matusop, A., Chan, S. T. & Singh, B. Bionomics of Anopheles latens in Kapit, Sarawak, Malaysian Borneo in relation to the transmission of zoonotic simian malaria parasite Plasmodium knowlesi. *Malaria journal* **7**, 52, doi:10.1186/1475-2875-7-52 (2008).
59 Vythilingam, I. *et al.* Plasmodium knowlesi in humans, macaques and mosquitoes in peninsular Malaysia. *Parasites & vectors* **1**, 26-26, doi:10.1186/1756-3305-1-26 (2008).
60 Jiram, A. I. *et al.* Entomologic investigation of Plasmodium knowlesi vectors in Kuala Lipis, Pahang, Malaysia. *Malaria journal* **11**, 213, doi:10.1186/1475-2875-11-213 (2012).
61 Brant, H. L. *et al.* Vertical stratification of adult mosquitoes (Diptera: Culicidae) within a tropical rainforest in Sabah, Malaysia. *Malaria journal* **15**, 370, doi:10.1186/s12936-016-1416-1 (2016).
62 Collins, W. E., Contacos, P. G. & Guinn, E. G. Studies on the Transmission of Simian Malarias II. Transmission of the H Strain of Plasmodium knowlesi by Anopheles balabacensis balabacensis. *The Journal of Parasitology* **53**, 841-844, doi:10.2307/3276783 (1967).
63 Fooden, J. Rhesus and Crab-Eating Macaques: Intergradation in Thailand. *Science* **143**, 363-364, doi:10.1126/science.143.3604.363 (1964).
64 Garnham, P. C. C. *Malaria parasites and other haemosporidia*. (Blackwell Scientific, 1966).
65 Fleagle, J. G. *Primate Adaptation & Evolution*. 165-168 (Academic Press, Inc., 1988).
66 Butcher, G. A., Mitchell, G. H. & Cohen, S. Plasmodium knowlesi infections in a small number of non-immune natural hosts (Macaca fascicularis) and in rhesus monkeys (M. mulatta). *Transactions of the Royal Society of Tropical Medicine and Hygiene* **104**, 75-77, doi:10.1016/j.trstmh.2009.05.017 (2010).
67 Butcher, G. A. & Cohen, S. Antigenic variation and protective immunity in Plasmodium knowlesi malaria. *Immunology* **23**, 503-521 (1972).
68 Gardner, M. B. & Luciw, P. A. Macaque Models of Human Infectious Disease. *ILAR Journal* **49**, 220-255, doi:10.1093/ilar.49.2.220 (2008).
69 Kumar, N., Chahroudi, A. & Silvestri, G. Animal models to achieve an HIV cure. *Current opinion in HIV and AIDS* **11**, 432-441, doi:10.1097/coh.0000000000000290 (2016).
70 Itell, H. L., Kaur, A., Deere, J. D., Barry, P. A. & Permar, S. R. Rhesus monkeys for a nonhuman primate model of cytomegalovirus infections. *Current opinion in virology* **25**, 126-133, doi:10.1016/j.coviro.2017.08.005 (2017).
71 Pena, J. C. & Ho, W. Z. Non-Human Primate Models of Tuberculosis. *Microbiology spectrum* **4**, doi:10.1128/microbiolspec.TBTB2-0007-2016 (2016).
72 Brown, K. N. & Brown, I. N. Immunity to Malaria: Antigenic Variation in Chronic Infections of Plasmodium knowlesi. *Nature* **208**, 1286-1288 (1965).
73 Cohen, S. & Butcher, G. A. Properties of protective malarial antibody. *Immunology* **19**, 369-383 (1970).

74 Cohen, S. & Butcher, G. A. Serum antibody in acquired malarial immunity. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **65**, 125-135 (1971).

75 Galinski, M. R. *et al.* Plasmodium knowlesi: a superb in vivo nonhuman primate model of antigenic variation in malaria. *Parasitology*, 1-16, doi:10.1017/s0031182017001135 (2017).

76 Barnwell, J. W., Howard, R. J. & Miller, L. H. Influence of the spleen on the expression of surface antigens on parasitized erythrocytes. *Ciba Foundation symposium* **94**, 117-136 (1983).

77 Howard, R. J. Antigenic variation of bloodstage malaria parasites. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **307**, 141-158 (1984).

78 Korir, C. C. & Galinski, M. R. Proteomic studies of Plasmodium knowlesi SICA variant antigens demonstrate their relationship with P. falciparum EMP1. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **6**, 75-79, doi:10.1016/j.meegid.2005.01.003 (2006).

79 Jiang, G. *et al.* Sterile protection against Plasmodium knowlesi in rhesus monkeys from a malaria vaccine: comparison of heterologous prime boost strategies. *PLoS One* **4**, e6559, doi:10.1371/journal.pone.0006559 (2009).

80 Lim, C. *et al.* Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nat Commun* **4**, 1638 (2013).

81 Cox-Singh, J. *et al.* Severe malaria - a case of fatal Plasmodium knowlesi infection with post-mortem findings: a case report. *Malaria journal* **9**, 10, doi:10.1186/1475-2875-9-10 (2010).

82 Fornace, K. M. *et al.* Association between Landscape Factors and Spatial Patterns of Plasmodium knowlesi Infections in Sabah, Malaysia. *Emerging infectious diseases* **22**, 201-209, doi:10.3201/eid2202.150656 (2016).

83 Guerra, C. A., Snow, R. W. & Hay, S. I. A global assessment of closed forests, deforestation and malaria risk. *Ann Trop Med Parasitol* **100**, 189-204, doi:10.1179/136485906x91512 (2006).

84 Fornace, K. M., Drakeley, C. J., William, T., Espino, F. & Cox, J. Mapping infectious disease landscapes: unmanned aerial vehicles and epidemiology. *Trends in Parasitology* **30**, 514-519, doi:https://doi.org/10.1016/j.pt.2014.09.001 (2014).

85 Ramasamy, R. Zoonotic malaria - global overview and research and policy needs. *Frontiers in public health* **2**, 123, doi:10.3389/fpubh.2014.00123 (2014).

86 Gottdenker, N. L., Streicker, D. G., Faust, C. L. & Carroll, C. R. Anthropogenic land use change and infectious diseases: a review of the evidence. *EcoHealth* **11**, 619-632, doi:10.1007/s10393-014-0941-z (2014).

87 Kilpatrick, A. M. & Randolph, S. E. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet* **380**, 1946-1955, doi:10.1016/s0140-6736(12)61151-9 (2012).

88 Lambin, E. F., Tran, A., Vanwambeke, S. O., Linard, C. & Soti, V. Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts. *International journal of health geographics* **9**, 54, doi:10.1186/1476-072x-9-54 (2010).

89 Patz, J. A. *et al.* Unhealthy landscapes: Policy recommendations on land use change and infectious disease emergence. *Environmental health perspectives* **112**, 1092-1098 (2004).

90 William, T. *et al.* Increasing incidence of Plasmodium knowlesi malaria following control of P. falciparum and P. vivax Malaria in Sabah, Malaysia. *PLoS neglected tropical diseases* **7**, e2026, doi:10.1371/journal.pntd.0002026 (2013).

91 van den Hoogen, L. L. *et al.* Serology describes a profile of declining malaria transmission in Farafenni, The Gambia. *Malaria journal* **14**, 416, doi:10.1186/s12936-015-0939-1 (2015).

92 Oduro, A. R. *et al.* Seroepidemiological and parasitological evaluation of the heterogeneity of malaria infection in the Gambia. *Malaria journal* **12**, 222, doi:10.1186/1475-2875-12-222 (2013).

93 Arnold, B. F. *et al.* Serological measures of malaria transmission in Haiti: comparison of longitudinal and cross-sectional methods. *PLoS One* **9**, e93684, doi:10.1371/journal.pone.0093684 (2014).

94 Bruce-Chwatt, L. J., Draper, C. C., Dodge, J. S., Topley, E. & Voller, A. Originally published as Volume 1, Issue 7749 SERO-EPIDEMIOLOGICAL STUDIES ON POPULATION GROUPS PREVIOUSLY EXPOSED TO MALARIA. *The Lancet* **299**, 512-515, doi:http://dx.doi.org/10.1016/S0140-6736(72)90176-6 (1972).

95 Corran, P., Coleman, P., Riley, E. & Drakeley, C. Serology: a robust indicator of malaria transmission intensity? *Trends in Parasitology* **23**, 575-582, doi:http://dx.doi.org/10.1016/j.pt.2007.08.023 (2007).

96 Drakeley, C. & Cook, J. in *Advances in parasitology* Vol. 69 299-352 (Academic Press, 2009).

97 Medina Costa, R., de Sousa, K. P., Atouguia, J., Tavira, L. T. & Silva, M. S. Prevalence and Level of Antibodies Anti-Plasmodium spp. in Travellers with Clinical History of Imported Malaria. *Journal of Parasitology Research* **2013**, 247273, doi:10.1155/2013/247273 (2013).

98 Ramasamy, R., Nagendran, K. & Ramasamy, M. S. Antibodies to epitopes on merozoite and sporozoite surface antigens as serologic markers of malaria transmission: studies at a site in the dry zone of Sri Lanka. *Am J Trop Med Hyg* **50**, 537-547 (1994).

99 Helb, D. A. *et al.* Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E4438-4447, doi:10.1073/pnas.1501705112 (2015).

100 Meraldi, V. *et al.* Natural antibody response to Plasmodium falciparum Exp-1, MSP-3 and GLURP long synthetic peptides and association with protection. *Parasite immunology* **26**, 265-272, doi:10.1111/j.0141-9838.2004.00705.x (2004).

101 Drakeley, C. J. *et al.* Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5108-5113, doi:10.1073/pnas.0408725102 (2005).

102 Akpogheneta, O. J. *et al.* Duration of naturally acquired antibody responses to blood-stage Plasmodium falciparum is age dependent and antigen specific. *Infection and immunity* **76**, 1748-1755, doi:10.1128/iai.01333-07 (2008).

103 Stanisic, D. I. *et al.* Acquisition of antibodies against Plasmodium falciparum merozoites and malaria immunity in young children and the influence of age, force of infection, and magnitude of response. *Infection and immunity* **83**, 646-660, doi:10.1128/iai.02398-14 (2015).

104 Greenwood, B. M. Control to elimination: implications for malaria research. *Trends Parasitol* **24**, 449-454, doi:10.1016/j.pt.2008.07.002 (2008).

105 Hay, S. I., Smith, D. L. & Snow, R. W. Measuring malaria endemicity from intense to interrupted transmission. *The Lancet infectious diseases* **8**, 369-378, doi:10.1016/s1473-3099(08)70069-0 (2008).

106 Thomson, J. G. PRELIMINARY NOTE ON THE COMPLEMENT DEVIATION IN CASES OF MALARIA: A NEW AID TO DIAGNOSIS. *British Medical Journal* **2**, 628-629 (1918).

107 Coggeshall, L. T. & Eaton, M. D. THE COMPLEMENT FIXATION REACTION IN MONKEY MALARIA. *The Journal of experimental medicine* **67**, 871-882 (1938).

108 Wilson, M., Fife, E. H., Jr., Mathews, H. M. & Sulzer, A. J. Comparison of the complement fixation, indirect immunofluorescence, and indirect hemagglutination tests for malaria. *Am J Trop Med Hyg* **24**, 755-759 (1975).

109 Voller, A., Huldt, G., Thors, C. & Engvall, E. New serological test for malaria antibodies. *Br Med J* **1**, 659-661 (1975).

110 Gupta, S., Manubhai, K. P., Kulkarni, V. & Srivastava, S. An overview of innovations and industrial solutions in Protein Microarray Technology. *Proteomics* **16**, 1297-1308, doi:10.1002/pmic.201500429 (2016).

111 Uplekar, S. *et al.* Characterizing Antibody Responses to Plasmodium vivax and Plasmodium falciparum Antigens in India Using Genome-Scale Protein Microarrays. *PLoS neglected tropical diseases* **11**, e0005323, doi:10.1371/journal.pntd.0005323 (2017).

112 Seideman, J. & Peritt, D. A novel monoclonal antibody screening method using the Luminex-100 microsphere system. *Journal of immunological methods* **267**, 165-171 (2002).

113 Mandy, F. F., Nakamura, T., Bergeron, M. & Sekiguchi, K. Overview and application of suspension array technology. *Clinics in laboratory medicine* **21**, 713-729, vii (2001).

114 Nolan, J. P. & Sklar, L. A. Suspension array technology: evolution of the flat-array paradigm. *Trends in biotechnology* **20**, 9-12 (2002).

115 Rogier, E. *et al.* Evaluation of Immunoglobulin G Responses to Plasmodium falciparum and Plasmodium vivax in Malian School Children Using Multiplex Bead Assay. *Am J Trop Med Hyg* **96**, 312-318, doi:10.4269/ajtmh.16-0476 (2017).

116 Tadesse, F. G. *et al.* The shape of the iceberg: quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia. *Malaria journal* **16**, 99, doi:10.1186/s12936-017-1749-4 (2017).

117 Esposito, F. *et al.* Evaluation of an ELISA kit for epidemiological detection of antibodies to Plasmodium falciparum sporozoites in human sera and bloodspot eluates. *Acta tropica* **47**, 1-10 (1990).

118 Ondigo, B. N. *et al.* Estimation of recent and long-term malaria transmission in a population by antibody testing to multiple Plasmodium falciparum antigens. *The Journal of infectious diseases* **210**, 1123-1132, doi:10.1093/infdis/jiu225 (2014).

119 Bousema, T. *et al.* Serologic markers for detecting malaria in areas of low endemicity, Somalia, 2008. *Emerging infectious diseases* **16**, 392-399, doi:10.3201/eid1603.090732 (2010).

120 Cook, J. *et al.* Serological Markers Suggest Heterogeneity of Effectiveness of Malaria Control Interventions on Bioko Island, Equatorial Guinea. *PLOS ONE* **6**, e25137, doi:10.1371/journal.pone.0025137 (2011).

121 Corran, P. H. *et al.* Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. *Malaria journal* **7**, 195, doi:10.1186/1475-2875-7-195 (2008).

122 Pommerville, J. C. in *Alcamo's fundamentals of microbiology* Ch. 21 Resistance and the Immune System: Acquired Immunity, (Jones and Bartlett, 2011).

123 Teo, A., Feng, G., Brown, G. V., Beeson, J. G. & Rogerson, S. J. Functional Antibodies and Protection against Blood-stage Malaria. *Trends in Parasitology* **32**, 887-898, doi:https://doi.org/10.1016/j.pt.2016.07.003 (2016).

124 Pommerville, J. C. in *Alcamo's fundamentals of microbiology* Ch. 21 Resistance and the Immune System: Acquired Immunity, (Jones and Bartlett, 2011).

125 Nezlin, R. *The Immunoglobulins: Structure and Function*. Vol. 10 (1998).
126 Schroeder, H. W. & Cavacini, L. Structure and Function of Immunoglobulins. *The Journal of allergy and clinical immunology* **125**, S41-S52, doi:10.1016/j.jaci.2009.09.046 (2010).
127 Apinjoh, T. O. *et al.* Determinants of infant susceptibility to malaria during the first year of life in South Western cameroon. *Open forum infectious diseases* **2**, ofv012, doi:10.1093/ofid/ofv012 (2015).
128 Riley, E. M., Wagner, G. E., Akanmori, B. D. & Koram, K. A. Do maternally acquired antibodies protect infants from malaria infection? *Parasite immunology* **23**, 51-59 (2001).
129 Amaratunga, C. *et al.* A role for fetal hemoglobin and maternal immune IgG in infant resistance to Plasmodium falciparum malaria. *PLoS One* **6**, e14798, doi:10.1371/journal.pone.0014798 (2011).
130 Dobbs, K. R. & Dent, A. E. Plasmodium malaria and antimalarial antibodies in the first year of life. *Parasitology* **143**, 129-138, doi:10.1017/S0031182015001626 (2016).
131 Vidarsson, G., Dekkers, G. & Rispens, T. IgG subclasses and allotypes: from structure to effector functions. *Frontiers in immunology* **5**, 520, doi:10.3389/fimmu.2014.00520 (2014).
132 Soe, S., Theisen, M., Roussilhon, C., Aye, K. S. & Druilhe, P. Association between protection against clinical malaria and antibodies to merozoite surface antigens in an area of hyperendemicity in Myanmar: complementarity between responses to merozoite surface protein 3 and the 220-kilodalton glutamate-rich protein. *Infection and immunity* **72**, 247-252 (2004).
133 Oeuvray, C. *et al.* Cytophilic immunoglobulin responses to Plasmodium falciparum glutamate-rich protein are correlated with protection against clinical malaria in Dielmo, Senegal. *Infection and immunity* **68**, 2617-2620 (2000).
134 Dodoo, D. *et al.* Naturally acquired antibodies to the glutamate-rich protein are associated with protection against Plasmodium falciparum malaria. *The Journal of infectious diseases* **181**, 1202-1205, doi:10.1086/315341 (2000).
135 Kana, I. H. *et al.* Naturally Acquired Antibodies Target the Glutamate-Rich Protein on Intact Merozoites and Predict Protection Against Febrile Malaria. *The Journal of infectious diseases* **215**, 623-630, doi:10.1093/infdis/jiw617 (2017).
136 Osier, F. H. *et al.* Breadth and magnitude of antibody responses to multiple Plasmodium falciparum merozoite antigens are associated with protection from clinical malaria. *Infection and immunity* **76**, 2240-2248, doi:10.1128/iai.01585-07 (2008).
137 Nebie, I. *et al.* Humoral Responses to Plasmodium falciparum Blood-Stage Antigens and Association with Incidence of Clinical Malaria in Children Living in an Area of Seasonal Malaria Transmission in Burkina Faso, West Africa. *Infection and immunity* **76**, 759-766, doi:10.1128/IAI.01147-07 (2008).
138 Krishnamurty, A. T. *et al.* Somatically Hypermutated Plasmodium-Specific IgM(+) Memory B Cells Are Rapid, Plastic, Early Responders upon Malaria Rechallenge. *Immunity* **45**, 402-414, doi:10.1016/j.immuni.2016.06.014 (2016).
139 Dodoo, D. *et al.* Cohort study of the association of antibody levels to AMA1, MSP119, MSP3 and GLURP with protection from clinical malaria in Ghanaian children. *Malaria journal* **7**, 142, doi:10.1186/1475-2875-7-142 (2008).
140 Arama, C. *et al.* Genetic Resistance to Malaria Is Associated With Greater Enhancement of Immunoglobulin (Ig)M Than IgG Responses to a Broad Array of Plasmodium falciparum Antigens. *Open forum infectious diseases* **2**, ofv118, doi:10.1093/ofid/ofv118 (2015).

141 Bolad, A. *et al.* Distinct interethnic differences in immunoglobulin G class/subclass and immunoglobulin M antibody responses to malaria antigens but not in immunoglobulin G responses to nonmalarial antigens in sympatric tribes living in West Africa. *Scandinavian journal of immunology* **61**, 380-386, doi:10.1111/j.1365-3083.2005.01587.x (2005).

142 Lomax-Smith, J. D., Zabrowarny, L. A., Howarth, G. S., Seymour, A. E. & Woodroffe, A. J. The immunochemical characterization of mesangial IgA deposits. *Am J Pathol* **113**, 359-364 (1983).

143 Woof, J. M. & Mestecky, J. Mucosal immunoglobulins. *Immunological reviews* **206**, 64-82, doi:10.1111/j.0105-2896.2005.00290.x (2005).

144 Kassim, O. O. *et al.* Inhibitory factors in breastmilk, maternal and infant sera against in vitro growth of Plasmodium falciparum malaria parasite. *Journal of tropical pediatrics* **46**, 92-96, doi:10.1093/tropej/46.2.92 (2000).

145 Ladjeva, I., Peterman, J. H. & Mestecky, J. IgA subclasses of human colostral antibodies specific for microbial and food antigens. *Clinical and Experimental Immunology* **78**, 85-90 (1989).

146 Duah, N. O., Miles, D. J., Whittle, H. C. & Conway, D. J. Acquisition of antibody isotypes against Plasmodium falciparum blood stage antigens in a birth cohort. *Parasite immunology* **32**, 125-134, doi:10.1111/j.1365-3024.2009.01165.x (2010).

147 Rafieian-Kopaei, M., Nasri, H., Alizadeh, F., Ataei, B. & Baradaran, A. Immunoglobulin A Nephropathy and Malaria falciparum Infection; a Rare Association. *Iranian journal of public health* **42**, 529-533 (2013).

148 Yoo, D. E. *et al.* Immunoglobulin A nephropathy associated with Plasmodium falciparum malaria. *Journal of Korean medical science* **27**, 446-449, doi:10.3346/jkms.2012.27.4.446 (2012).

149 eBioscience. *Antibody Isotyping Guide*, <https://assets.thermofisher.com/TFS-Assets/LSG/brochures/Q112007e-Ab_Isotyping_Guide.pdf> (

150 Dottorini, T. *et al.* Serum IgE Reactivity Profiling in an Asthma Affected Cohort. *PLOS ONE* **6**, e22319, doi:10.1371/journal.pone.0022319 (2011).

151 Perlmann, P. *et al.* Contrasting functions of IgG and IgE antimalarial antibodies in uncomplicated and severe Plasmodium falciparum malaria. *Am J Trop Med Hyg* **62**, 373-377 (2000).

152 Perlmann, H. *et al.* IgE elevation and IgE anti-malarial antibodies in Plasmodium falciparum malaria: association of high IgE levels with cerebral malaria. *Clinical and Experimental Immunology* **97**, 284-292 (1994).

153 Perlmann, P. *et al.* Immunoglobulin E, a pathogenic factor in Plasmodium falciparum malaria. *Infection and immunity* **65**, 116-121 (1997).

154 Dugas, B., Mossalayi, M. D., Damais, C. & Kolb, J. P. Nitric oxide production by human monocytes: evidence for a role of CD23. *Immunology today* **16**, 574-580, doi:10.1016/0167-5699(95)80080-8 (1995).

155 Perlmann, P., Perlmann, H., ElGhazali, G. & Blomberg, M. T. IgE and tumor necrosis factor in malaria infection. *Immunology letters* **65**, 29-33 (1999).

156 Grau, G. E. *et al.* Tumor necrosis factor and disease severity in children with falciparum malaria. *The New England journal of medicine* **320**, 1586-1591, doi:10.1056/nejm198906153202404 (1989).

157 Kwiatkowski, D. *et al.* TNF concentration in fatal cerebral, non-fatal cerebral, and uncomplicated Plasmodium falciparum malaria. *Lancet* **336**, 1201-1204 (1990).

158 Tangteerawatana, P. *et al.* Differential regulation of IgG subclasses and IgE antimalarial antibody responses in complicated and uncomplicated Plasmodium falciparum malaria. *Parasite immunology* **29**, 475-483, doi:10.1111/j.1365-3024.2007.00965.x (2007).

159 Bereczky, S. *et al.* Elevated anti-malarial IgE in asymptomatic individuals is associated with reduced risk for subsequent clinical malaria. *Int J Parasitol* **34**, 935-942, doi:10.1016/j.ijpara.2004.04.007 (2004).
160 Dolo, A. *et al.* Difference in susceptibility to malaria between two sympatric ethnic groups in Mali. *Am J Trop Med Hyg* **72**, 243-248 (2005).
161 Vladutiu, A. O. Immunoglobulin D: properties, measurement, and clinical relevance. *Clin Diagn Lab Immunol* **7**, 131-140 (2000).
162 Ohta, Y. & Flajnik, M. IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 10723-10728, doi:10.1073/pnas.0601407103 (2006).
163 Flajnik, M. F. Comparative analyses of immunoglobulin genes: surprises and portents. *Nature reviews. Immunology* **2**, 688-698, doi:10.1038/nri889 (2002).
164 Geisberger, R., Lamers, M. & Achatz, G. The riddle of the dual expression of IgM and IgD. *Immunology* **118**, 429-437, doi:10.1111/j.1365-2567.2006.02386.x (2006).
165 Gutzeit, C., Chen, K. & Cerutti, A. The enigmatic function of IgD: some answers at last. *European journal of immunology* **48**, 1101-1113, doi:10.1002/eji.201646547 (2018).
166 Chen, K. *et al.* Immunoglobulin D enhances immune surveillance by activating antimicrobial, proinflammatory and B cell-stimulating programs in basophils. *Nature immunology* **10**, 889-898, doi:10.1038/ni.1748 (2009).
167 Forsgren, A. *et al.* Isolation and characterization of a novel IgD-binding protein from Moraxella catarrhalis. *Journal of immunology (Baltimore, Md. : 1950)* **167**, 2112-2120 (2001).
168 Colwell, E. J., Bernier, G. M. & Fife, E. H., Jr. Serum immunoglobulin D and malaria antibodies in South Vietnamese residents. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **65**, 310-314 (1971).
169 Targett, G. A. Antibody response to Plasmodium falciparum malaria. Comparisons of immunoglobulin concentrations, antibody titres and the antigenicity of different asexual forms of the parasite. *Clin Exp Immunol* **7**, 501-517 (1970).
170 McGregor, I. A., Rowe, D. S., Wilson, M. E. & Billewicz, W. Z. Plasma immunoglobulin concentrations in an African (Gambian) community in relation to season, malaria and other infections and pregnancy. *Clin Exp Immunol* **7**, 51-74 (1970).
171 Asito, A. S. *et al.* Suppression of circulating IgD+CD27+ memory B cells in infants living in a malaria-endemic region of Kenya. *Malaria journal* **10**, 362, doi:10.1186/1475-2875-10-362 (2011).
172 Pain, A. *et al.* The genome of the simian and human malaria parasite Plasmodium knowlesi. *Nature* **455**, 799-803, doi:10.1038/nature07306 (2008).

Chapter 2 : HYPOTHESIS, AIMS AND OBJECTIVES

The following chapters will discuss the development of novel serological tools for the detection of *P. knowlesi* and the subsequent antibody profiling using high throughput multiplex platforms. My central hypothesis was based on the premise that paralogous proteins previously described as targets of immunity in other *Plasmodium spp.,* will also be of immunogenic importance in *P. knowlesi.* I intend to demonstrate that *P. knowlesi*-specific proteins may be useful as serosurveillance tools in Southeast Asia and that these could be used to dissect the antibody response profiles within endemic populations.

Overall aim:

The overall aim of this PhD project was to develop *P. knowlesi*-specific serological tools to help dissect and improve our current understanding of human and macaque responses to *P. knowlesi* infections. I envisage that this information could be used to build a foundation for describing the geographical boundaries of *P. knowlesi* infections.

Objectives:

1. To develop and validate a small panel of novel *P. knowlesi*-specific recombinant antigen reagents, based on paralogous genes as a proof-of-concept towards the development of potential serosurveillance tools for *P. knowlesi* with greater specificity than those currently available, based on an *in silico* rational design approach.

2. To expand the panel of available reagents using our recently developed in silico approach to include non-paralogous targets present in *P. knowlesi* alone and determine their utility as potential serosurveillance tools.
3. To characterise the antibody isotype responses (IgM, IgG and IgA responses) to the panel of *P. knowlesi*-specific recombinant antigens in clinical samples from an endemic setting, and use the data to contribute to our current understanding of immunity to *P. knowlesi*.

Chapter 3 : MATERIALS AND METHODS

3.1. SDS-PAGE AND WESTERN BLOT

Affinity purified proteins were electrophoresed on a Mini-PROTEAN® TGX™ Precast Gel (Bio-Rad, UK) using TGx buffer (3.03 g/L Trisma base, 14.44 g/L glycine and 0.1% SDS, all at pH 8.3) at 200V for 35 min under non-denaturing and denaturing conditions at approximately 0.4 mg/ml and visualised after an o/n incubation using Coomassie blue stain (BioSafe, BioRad USA) or Pierce® Silver staining as per manufacturer's instructions (ThermoFisher Scientific, USA). Briefly, for the Silver staining, the gel was washed twice in ultrapure H_2O for 5 minutes each wash and fixed twice in 30% ethanol:10% acetic acid solution for 15 minutes each time. The gel was washed twice in 10% ethanol for 5 minutes each wash and again twice in ultrapure H_2O for 5 minutes each. Sensitizer working solution was prepared as indicated in the kit instructions and the gel incubated for 1 minute before being washed twice in ultrapure H_2O for 1 minute each wash. Stain working solution was prepared as indicated in the kit instructions and the gel was incubated for 30 minutes before being washed twice in ultrapure H_2O for 20 seconds each wash. Developer working solution was prepared as indicated in the kit instructions and the gel was incubated until protein bands started to appear (2-3 minutes maximum) before being replaced by 10% acetic acid stop solution and incubated for 10 minutes. The gels were dried using a gel drying solution (30% methanol and 5% glycerol in H_2O) and preserved in gel drying film (Promega). The empirical sizes of each protein were calculated with the PageRuler prestained marker (Fermentas) using the ImageLab (BioRad) software.

For the Western blots, a gel containing expressed antigens was transferred onto nitrocellulose using a Hoeffer semi-dry blotter and processed using a BioRad general protocol for Western blotting. Briefly, the transferred membrane blocked in

5% (w/v) skimmed milk in 1x Phosphate Buffered Saline, 0.05% Tween 20 (PBS/T) solution and either left overnight (o/n) at 4^0 C or incubated at room temperature (RT) on a rolling platform for 1 h. Once blocked, it was washed 3-5 x in 1x PBS/T then probed with a 1/1000 dilution of pooled *Pk* human sera prepared in 1x PBS/T and incubated at RT for 1 h or o/n. Once the incubating was finished, the blot was washed as before and probed with a 1/10,000 dilution of secondary antibody, horse-radish peroxidase (HRP) conjugate polyclonal rabbit anti-human IgG antibody (Dako, Denmark), also prepared in 1x PBS/T and incubated at RT for 1 h and washed again before being developed. Further blots were generated using single well SDS-PAGE for Slots blots and were prepared following the Western protocol, except the transferred membranes were assembled in a Miniblotter® 28 (Immunetics, USA) slot blot manifold. All other steps are as described above. Both blots were developed using a Pierce™ ECL western blotting substrate kit (ThermoFisher Scientific) and incubated at RT for 5 minutes before being exposed to an X-ray film (Amersham hyperfilm™ ECL chemoluminescence film) or read using the Azure c600 (Azure Biosystems, USA). Exposure times varied between 15 secs, 30 secs, 1 min and 10 min before being developed using an appropriate developing solution and fixative. Blots using the Miniblotter were read using the Azure c600 and were not included here as the figure was unreadable due to lack of expertise with the equipment resulting in inappropriate exposure.

3.2. SNP AND PHYLOGENETIC ANALYSIS

Full-length sequence data from PlasmoDB and construct-specific truncated sequences generated in-house using Sanger sequencing were mapped to an in-house reference sequence strain (*P. knowlesi* H strain) using the Burrows-Wheeler Aligner (BWA) software (v0.7.5a-r405)[1]. Single nucleotide polymorphisms (SNPs) were called using the SAMtools (v1.3) (Sequence Alignment/Map) software using default settings[2] and were filtered to increase stringency and target only high quality

variants (missingness<10%, mixed calls<10%). Overlap between these SNPs and the gene candidates was then found using custom perl scripts. SNP variants were annotated using snpEFF (v4.3i) (http://snpeff.sourceforge.net/)[3] to retrieve the amino acid position and type effect of the variant. Maximum likelihood phylogenetic trees were constructed from protein sequences using RAxML[4] with a fixed empirical substitution matrix and 200 bootstraps and was visualised using iTOL (http://itol.embl.de)[5]. This work was done with Dr. Jody Phelan.

3.3. HUMAN TEST SAMPLES AND ETHICS APPROVAL

3.3.1. SABAH HOSPITAL CASE SAMPLES

Written informed consent was obtained from all study participants[6,7]. Samples were collected as part of a hospital-based clinical trial in Malaysia, Sabah, details of which have been described elsewhere[7,8] (www.clinicaltrials.gov: #NCT01708876). The study was approved by the human research ethics committees of Malaysia (MREC) (#NMRR-12-537-12568), the Menzies School of Health and Research, Australia (#HREC-2012-1814), and the London School of Hygiene and Tropical Medicine, UK (#6244). Serum and dried blood spot (BS) samples were collected from malaria diagnosed individuals at day 0 (n=154; *P. knowlesi* positive n=102), 7 (n=120; *P. knowlesi* positive n=76) and 28 (n=122; *P. knowlesi* positive n=81) post hospital admission and bloodspot samples at 1 (n=46; *P. knowlesi* positive n=37) and 2 (n=3; *P. knowlesi* positive n=2) year follow ups after that, with drug treatment also taking place at day 0 (***Table* 3.1**). Only samples from day 0, 7 and 28 were tested using the ELISA and the protein microarray platforms. The Luminex multiplex bead assay (MBA) platform was used to test all time points.

Table 3.1 Hospital case control sample number by species, time point, sample type and serological platform used

	P. knowlesi					*P. vivax*					*P. falciparum*					*P. malariae*				
Time point	0	7	28	Y1	Y2	0	7	28	Y1	Y2	0	7	28	Y1	Y2	0	7	28	Y1	Y2
Serum	92	72	77	0	0	31	28	27	0	0	14	14	13	0	0	0	0	0	0	0
BS	10	4	4	37	2	4	2	0	5	0	1	0	0	3	0	1	0	0	1	0
Total	102	76	81	37	2	35	30	27	5	0	15	14	13	3	0	1	0	0	1	0
ELISA	92	72	77	0	0	31	28	27	0	0	14	14	13	0	0	0	0	0	0	0
Luminex	102	76	81	37	2	35	30	27	5	0	15	14	13	3	0	1	0	0	1	0
Microarray	12	12	12	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0

3.3.2. SABAH AND PALAWAN CASE STUDY COMMUNITIES SURVEY SAMPLES

Written informed consent was obtained from all participants in this study[9]. A survey of case study communities in Malaysian Borneo and Palawan, Philippines, was approved by the Medical Research Sub-Committee of the Malaysian Ministry of Health (#NMRR-14-713-21117), the institutional Review Board of the Research Institute for Tropical Medicine (The Philippines), and the London School of Hygiene and Tropical Medicine (UK) (#8340) Research Ethics Committee. The survey was performed between October 2014 and January 2015 in Matunggong, Kudat (n=1162) and Limbuak, Pulau Banggi Island (n=795), Malaysian Sabah, and in September 2014 in Bacungan (n=546), Palawan, The Philippines; regions endemic for *P. knowlesi* (***Figure* 3.1**). The study sites were selected based on previously reported *P. knowlesi* clinical cases. Each site had a specific characteristic: largely deforested (Kudat, Malaysia), patches of secondary forest bordering oil palm plantations (Pulau Banggi, Malaysia) and intact primary and secondary forest (Bacungan, The Philippines). The total sample size used from this project was ~2503 individuals with a median age of 24 years (age range 3 months – 99 years) and similar proportions of sampled men and women in all study sites. Community wide surveys in these intensive study areas describe age specific patterns of exposure to *P. knowlesi* and other malaria species. Individuals less than 3 months of age were excluded from this study as well as individuals who had not resided in the

area of study for the past month. Filter paper bloodspot samples were collected and processed as previously described[10].

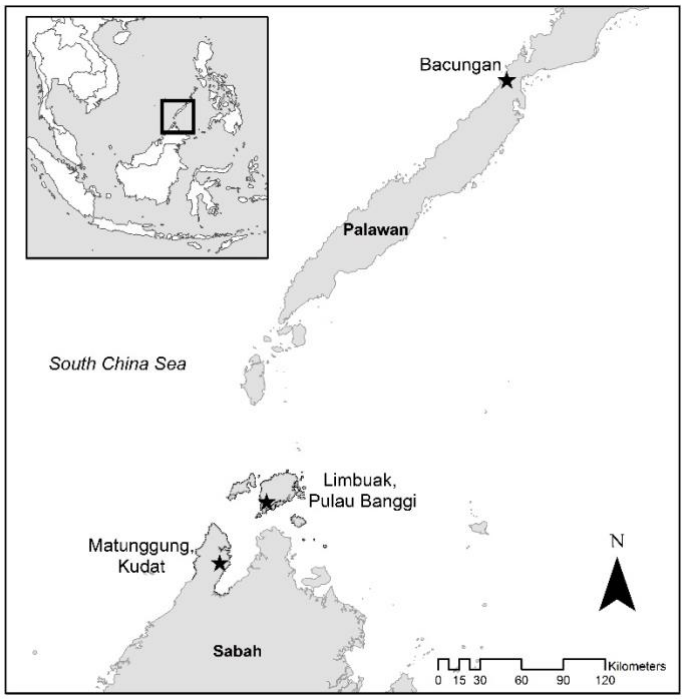


Figure 3.1 Map of community survey collection sites in Sabah, Malaysia. Stars indicate villages where samples were collected: Matunggong in Kudat and Limbuak in Pulau Banggi, Sabah, Malaysia, and Bacugan in Palawan, The Philippines. Figure generated by Kim Fornace and from published source[9].

3.3.3. SABAH CROSS-SECTIONAL SURVEY (XSS) STUDY

All samples used came from the Monkeybar project collected by groups led by Timothy William in Malaysia. Briefly, the Monkeybar project is a *P. knowlesi*-focused, multidisciplinary collaborative research project on the environmental characterisation of infection of both the hosts and vectors, determining risk factors for exposure and implementing control measures. Written and informed consent was obtained from all participants in this study and procedures were performed according to relevant guidelines. As for the community survey study mentioned previously, this study was approved by the Medical Research Sub-Committee of the Malaysian Ministry of Health (NMRR-14-713-21117) and the Research Ethics Committee of the London

School of Hygiene and Tropical Medicine (8340). The Sabah cross-sectional survey (XSS) was performed in 4 districts, Kudat, Kota Marudu (KM), Pitas and Ranau, in northern Sabah, Malaysia (***Figure* 3.2**). These districts represent high-risk areas for *P. knowlesi* transmission and the purpose of this larger survey was to estimate the transmission intensity of *P. knowlesi* across the different environmental conditions within those areas.

Figure 3.2 Collection sites for Sabah cross-sectional survey (XSS). Samples were collected from communities in Ranau, Kota Marudu, Pitas, Kudat and Pulau Banggi in Sabah, Malaysia. Map generated by Kim Fornace (personal communication).

3.3.4. MICROSCOPY AND PCR DIAGNOSIS OF HUMAN SAMPLES

Finger-prick blood samples were collected from all individuals enrolled in the different studies; these included approximately 200 µl of whole-blood specimens collected in tubes containing EDTA (Ethylenediaminetetraacetic acid) (Becton-Dickinson, Franklin Lakes, USA), three 20 µl blood spots stored on filter paper

(3MM, Whatman, UK), dried and stored with desiccant at 4°C and blood smears for microscopy detection of malaria parasites. A number of the clinical samples were collected on a thinner type of filter paper for which we do not have any information on. These were stored in the same way as described above. The thin filter paper spot samples that were also collected as serum were the only ones analysed (n=26). All clinical whole blood samples were PCR screened for species specificity as described by Grigg et. al.[7] in order to validate the microscopy results and to detect low parasitemia infections that would not be detected by standard microscopy. Similarly to the clinical samples, samples from the community surveys were all PCR screened from whole blood for plasmodium species infection as by Singh *et. al.* [11] and Fornace *et. al.* [9,12]. All products were visualised on a 2% agarose gel. The teams led by Kim Fornace both collected all samples and performed this work at the Universiti Sabah Malaysia in Malaysia and at the Research Institute for Tropical Medicine in the Philippines.

REFERENCES

1 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

2 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

3 Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80-92, doi:10.4161/fly.19695 (2012).

4 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

5 Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research* **44**, W242-245, doi:10.1093/nar/gkw290 (2016).

6 Tadesse, F. G. *et al.* The shape of the iceberg: quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia. *Malaria journal* **16**, 99, doi:10.1186/s12936-017-1749-4 (2017).

7 Grigg, M. J. *et al.* Artesunate-mefloquine versus chloroquine for treatment of uncomplicated Plasmodium knowlesi malaria in Malaysia (ACT KNOW): an open-label, randomised controlled trial. *The Lancet infectious diseases*, doi:10.1016/s1473-3099(15)00415-6 (2016).

8 Grigg, M. J. *et al.* A study protocol for a randomised open-label clinical trial of artesunate-mefloquine versus chloroquine in patients with non-severe Plasmodium knowlesi malaria in Sabah, Malaysia (ACT KNOW trial). *BMJ open* **4**, e006005, doi:10.1136/bmjopen-2014-006005 (2014).

9 Fornace, K. M. *et al.* Exposure and infection to Plasmodium knowlesi in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. *PLoS neglected tropical diseases* **12**, e0006432, doi:10.1371/journal.pntd.0006432 (2018).

10 Corran, P. H. *et al.* Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. *Malaria journal* **7**, 195, doi:10.1186/1475-2875-7-195 (2008).

11 Singh, B. *et al.* A genus- and species-specific nested polymerase chain reaction malaria detection assay for epidemiologic studies. *Am J Trop Med Hyg* **60**, 687-692 (1999).

12 Fornace, K. M. *et al.* Asymptomatic and submicroscopic carriage of Plasmodium knowlesi malaria in household and community members of clinical cases in Sabah, Malaysia. *The Journal of infectious diseases*, doi:10.1093/infdis/jiv475 (2015).

Chapter 4 : IDENTIFICATION AND VALIDATION OF A NOVEL PANEL OF *PLASMODIUM KNOWLESI* SEROLOGICAL BIOMARKERS OF EXPOSURE, AND THEIR SUBSEQUENT USE IN THE MEASUREMENT OF EXPOSURE AND INFECTION IN MALAYSIA AND THE PHILIPPINES

The work presented in this chapter is adapted and extended from Herman *et. al.* 2018[1] (Appendix I) with additional contributions from Fornace *et. al.* 2018[2] (Appendix II), both with respective research paper cover sheets.

4.1. INTRODUCTION

P. knowlesi has the potential to cause severe disease in endemic regions[3], and is now the most common cause of clinical malaria in Malaysia[4-6]. Due to morphological similarities to *P. malariae*[7,8], historical misdiagnoses of *P. knowlesi* infections as *P. malariae*[9] have been found to be very common. As previously mentioned in Chapter 1, misdiagnosis of *P. knowlesi* as *P. vivax* and *P. falciparum*[10-12] is also fairly common, with potential delay of appropriate treatment associated with case fatalities[3,13,14].

Studies have shown that antibodies to some *Plasmodium* spp. proteins persist for long periods[15-17], even in the context of limited exposure or absence of infection and have been shown to be correlated to malarial infection[18-21]. Such antibodies can be utilised in serological assays, accurately estimating both recent and historical[18,21] incidence and exposure to *Plasmodium* parasites[22-26].

Currently, the incidence of *P. falciparum* and *P. vivax* is reducing due to the progress in malaria elimination programmes in endemic areas[27,28]. However, *P. knowlesi* poses a challenge to the malaria elimination programme due to increased agricultural activities, deforestation and changes in the macaque population. An increase of *P. knowlesi* incidence has recently been acknowledged[29] and very little is known about the true extent of *P. knowlesi* exposure, infection and geographical boundaries. Associations have been shown between increases in *P. knowlesi* incidence in villages and deforestation[30], however, as this was only shown for clinical cases, this is not necessarily reflected in the broader population. Additionally, studies have identified numerous asymptomatic *P. knowlesi* infections, including individuals who will not definitely report to local health facilities[31-33], making it difficult to determine if those associations can be applied to the broader community.

It is famously difficult to detect malaria parasites in both infected humans[31,34] and mosquitos[35,36] in areas of low endemicity. A useful tool to detecting long lasting

antibody responses[37] are serologic biomarkers of exposure. Serology is more attractive than other methods due to the cheaper and simpler laboratory procedures involved, allowing the processing and testing of more samples in less time, and its sturdiness in determining short-term fluctuations in transmission[23,38]. Different antigens elicit a range of antibody responses in individual hosts, showing different magnitudes and kinetics, which can give us a more detailed analysis of a population's exposure. This said, it would be suggested that biomarkers for exposure should group several different specific antigens to analyse malarial infections. Epidemiology has long used serology as a way of measuring malaria incidence and disease exposure in endemic countries [39-41]. Previous studies have shown that the prevalence of malaria can be determined using species-specific recombinant antigens as serosurveillance tools, determining the boundaries of infection as well as the progress of elimination interventions [23,38,42].

These serological markers can detect antibodies against malaria parasites, determining exposure to current infections as well as previous infections[43]. allowing the identification of areas or groups within a community with higher transmission[23] as well as lower transmission[37], where detection of an infection is normally very low. There are several antigens currently in use as serological markers, such as MSP1-19 (merozoite surface protein 1-19) and AMA1 (apical membrane antigen 1) for *P. falciparum* and *P. vivax*[21,44-46], as described later. Previous infections can be reflected by the presence of certain antibodies in the system, which can determine the level of transmission in a certain region[37]. In a study led by Drakeley *et. al.*[45], it was shown that altitude and rainfall were correlated with parasite prevalence. Recombinant proteins do exist for *P. knowlesi*, however of these, the amino acid identity between *P. knowlesi* and other *Plasmodium* spp. is high and therefore these are not applicable to identifying species-specific antibody responses.

Previously known as PK66 due to its size (66kDa), *Pk*AMA-1 is highly conserved amongst *Plasmodium* spp.[47]. AMA-1 is expressed in the micronemes of the

sporozoite (invasive pre-erythrocytic stage) and merozoite (invasive asexual blood stage) forms of the parasite[48].

In *P. falciparum,* AMA-1 (previously PF83)[49] is recognised after exposure to malarial infection[48,50-52]. Similarly, macaques that were immunized with *Pk*AMA-1 developed antibodies inhibiting red blood cell (RBC) invasion by merozoites, demonstrating some degree of immunity towards being challenged[50,53]. The *Pf*AMA1 antigen is routinely used as a serological tool in epidemiology studies alongside *Pv*AMA1[38,39,54].

The SPATR (secreted protein with altered thrombospondin type I repeat) domain was first identified in chromosome 2 of *P. falciparum*[55], and it was suggested that, since it is on the surface of the sporozoite form[56], it is involved in the parasite invasion of host cells, therefore a potential *P. falciparum* vaccine candidate[57]. It is one of the few identified *P. falciparum* protein antigens to be expressed during multiple stages of the parasite life cycle[57], alongside AMA1 and several others (MAEBL etc)[48]. These multi-stage antigens are important in inducing protective immune responses to all parasite life stages. A gene encoding a novel *P. knowlesi* multistage secreted protein was characterised and named *Pk*SPATR due to its homology to *Pf*SPATR[58] (61% amino acid identity). It was suggested by Mahajan *et. al*. that *Pk*SPATR is involved in sporozoite invasion of liver cells as it was found to bind to a human hepatoma cell line[58]. *Pk*AMA-1 and *Pk*SPATR share 86% and 85% amino acid identity respectively with *P. vivax* (https://is.gd/MzISez), potentially making it difficult to distinguish between the two species where infections are co-endemic.

The 195kDa merozoite surface protein 1 (MSP-1), a major surface protein on the surface of the merozoite[59,60], is a well described blood stage vaccine candidate[61-64] and serosurveillance tool[52,54,60]. The result of a series of proteolytic cleavage steps of this protein is MSP1-19[65] which has been shown to be highly immunogenic, both in *P. falciparum*[66] and *P. vivax*[67], with antibodies targeting this region associated with protection[52,54,60,68].

Although *Pk*MSP1-19 is not an adequate antigenic candidate *for P. knowlesi* specific serosurveillance due to high amino acid identity with other *Plasmodium* spp. (appendix IV), it is still worth looking at *P. knowlesi*-unique sections of the *Pk*MSP-1 sequence.

Sporozoite surface protein 2 (SSP2), also known as thrombospondin-related adhesion protein (TRAP), is expressed in the pre-erythrocytic stages of the parasite life cycle. This protein is located on micronemes and the sporozoite cell-surface of *P. falciparum* and contributes to cell invasion or hepatocyte recognition[56,69-71].

SSP2/TRAP is composed by a short cytoplasmic region (~40 amino acids) with an acidic carboxy-terminus and a single trans-membrane region[69]. It was found that *P. falciparum* TRAP contributes to immunity against sporozoites, inhibiting hepatocyte and HepG2 cell invasion by sporozoites[56,70,72]. SSP2/TRAP antibodies were produced by malaria exposed individuals showing age dependent differences and seasonal fluctuations[72]. Similarly, evidence for SSP2/TRAP suggested an immunogenic antigen involved in protection from disease in mice[73].

The serine repeat antigen (SERA) family had previously attracted attention as a source of both drug and vaccine candidates[74]. The SERA antigens are asexual antigens, members of the SERA multigene family [75] that are expressed during the trophozoite and schizont stages[76,77]. *Pf*SERA5 is the most abundant parasitophorous vacuole protein and is essential to blood stage growth of the parasite [78-80], with antibodies against this antigen thought to inhibit parasite growth [81,82] and to induce protection against blood stage *in vivo* infection[83]. Although possessing a papain-like enzymatic domain, recent evidence suggests that the protein plays a non-enzymatic role[80]. *Pf*SERA3 has also been shown to be a highly immunogenic antigen with an important, although not essential role in the erythrocytic cycle[84] and has also been implicated as having a role in liver stage merozoite release in *P. berghei*[85]. *Pf*SERA5 was also shown to be a vaccine immunogen[86], therefore a strong vaccine candidate (SE36)[87,88]. To date, the number of SERA genes is different in each species, but all

Plasmodium species seem to have at least one gene from the SERA group and they show gene synteny conservation of the family cluster [89-92].

Pain *et. al.* [93] have recently sequenced the full genome of *P. knowlesi* (H strain, Pk1 (A+) clone) by whole-genome shotgun sequencing and the genome coverage has been improved since [94], giving us the opportunity to further characterise this organism.

The 2011 WHO consultation panel on the public health importance of *P. knowlesi* recommended the urgent development of *P. knowlesi*-specific diagnostic tools[95]. Key to achieving this goal would be the development of sensitive and accurate tools to help monitor the transmission of infection. In this chapter, we describe the development and evaluation of a panel of recombinant antigens based on *P. knowlesi*-specific amino acid sequences, using publically available *in silico* tools. This is the first such description of species-specific tools against *P. knowlesi*. We also measure the reactivity of one of these tools in a community sample set from Malaysian Borneo and The Philippines in order to determine the level of exposure to *P. knowlesi* in those regions. The development of such well-validated species-specific tools represent a potentially important serosurveillance tool for the monitoring of historical and current *P. knowlesi* infections in endemic areas. These reagents will also serve as an important set of tools to help identify correlates of immunity to *P. knowlesi*.

4.2. METHODS

4.2.1. IDENTIFICATION OF CANDIDATE SEQUENCES AND PRIMER DESIGN

A panel of 18 genes were originally selected based on available evidence from existing literature on *P. falciparum (*MSP1[52,54,60], apical membrane antigen 1 (AMA1)[48,51,54], thrombospondin-related adhesion protein/sporozoite surface 2 (TRAP/SSP2)[56,96], serine repeat antigen (SERA)[78,79,87-90]), *P. vivax* (MSP1[54], AMA1[54], SERA[89,90,97]) and *P. knowlesi* antigens (MSP1[98-103], AMA-1[50,53,104-107], circumsporosoite protein (CSP)[108,109], SERA[89,90], hypothetical proteins[93]), together with *in silico* mining of existing *P. knowlesi* sequence data (www.plasmodb.org). Each candidate gene was further screened against other *Plasmodium* spp. corresponding to *P. knowlesi* AMA-1, MSP1, SERA3, secreted protein with altered thrombospondin type I repeat (SPATR) and TRAP/SSP2 candidate, as well as *P. knowlesi* hypothetical proteins (HP) (**Table 4.1**) in order to identify *P. knowlesi*-specific regions that would form the basis of each construct. From the original panel, 10 antigens were successfully cloned and expressed. Gene accession codes are found in **Table 4.2**. Full-length protein sequences for each gene were initially screened using a number of packages in the DNASTAR Lasergene software suite (DNASTAR), NCBI protein BLAST (Basic Local Alignment Search Tool) (https://is.gd/MzISez) and PlasmoDB (https://is.gd/XOs7vd[110]), identifying any regions of shared identity between the human infecting *Plasmodium* spp. Sequence alignments at the amino-acid level were generated for each gene target using Multiple Sequence Comparison by Log-Expectation (MUSCLE) (http://www.ebi.ac.uk/Tools/msa/muscle/)[111]. Any full-length sequences and sequence regions with a high level of shared identity with other human *Plasmodium* spp. were excluded from further analysis and the selected *P. knowlesi*-specific truncated regions were screened again in the same manner to ensure specificity (**Table 4.1**).

Table 4.1 Gene and candidate homology between *P. knowlesi* and other *Plasmodium* spp. For the PlasmoDB BLASTa analysis, values are shown as amino acid identities between the P. *knowlesi* candidates and other Plasmodium spp. followed by the respective percentage. For the NCBI BLASTp analysis the sequence coverage is shown as a percentage, followed by the amino acid identities, also shown as a percentage.

P. knowlesi gene candidate	Protein length	Homology (%)											
		P. vivax			*P. falciparum*			*P. malariae*			*P. ovale*		
		PlasmoDB	NCBI		PlasmoDB	NCBI		PlasmoDB	NCBI		PlasmoDB	NCBI	
		Identities (%)	query cover	Ident	Identities (%)	query cover	Ident	Identities (%)	query cover	Ident	Identities (%)	query cover	Ident
AMA-1	563	480/563 (85%)	100	86	307/525 (58%)	92	59	388/564 (69%)	100	69	410/565 (73%)	100	73
MSP1	1821	1167/1861(63%)	98	63	428/1081(40%)	99	45	444/974 (46%)	52	46	860/1888(46%)	99	46
- MSP1-19	92	71/86 (83%)	96	83	47/94 (50%)	96	50	50/89 (56%)	100	57	50/89 (56%)	95	56
- MSP1 ag2	88	-	-	-	-	-	-	-	-	-	-	-	-
SERA3	1079	623/1080 (58%)	100	58	400/989 (40%)	96	39	428/877(49%)	98	47	348/760 (46%)	89	47
- SERA3 ag1	116	23/58 (40%)	-	-	16/70 (23%)	48	30	32/115 (28%)	55	25	27/104 (26%)	24	39
- SERA3 ag2	173	42/173 (24%)	-	-	21/73 (29%)	-	-	19/101 (19%)	-	-	26/100 (26%)	-	-
TRAP/SSP2	580	399/581 (69%)	100	69	243/593 (41%)	100	40	287/636 (45%)	100	45	281/587 (48%)	100	50
- SSP2 frag	120	40/93 (43%)	77	45	-	47	49	-	-	-	37/118 (31%)	85	31
TSERA2	526	185/277 (67%)	100	62	144/433 (33%)	99	34	93/231 (40%)	85	40	94/218 (43%)	100	38
- TSERA2 ag1	191	108/178 (61%)	-	-	51/106 (48%)	52	48	57/172 (33%)	89	36	47/113 (42%)	51	43
- TSERA2 ag2	46	35/44 (80%)	95	80	19/44 (43%)	95	43	16/44 (36%)	95	36	18/44 (41%)	95	39
PKH_021580	208	-	-	-	15/33 (45%)	-	-	-	-	-	21/63 (33%)	-	-
- PKH_021580 frag	155	19/52 (37%)	-	-	15/33 (45%)	-	-	-	-	-	21/63 (33%)	-	-
PKH_031930	212	16/40 (40%)	-	-	14/33 (42%)	-	-	28/98 (29%)	-	-	18/51 (35%)	-	-
- PKH_031930 ag1	93	-	65	23	19/56 (34%)	-	-	-	-	-	-	-	-
- PKH_031930 ag2	112	16/40 (40%)	39	34	14/33 (42%)	-	-	17/41 (41%)	-	-	12/32 (38%)	-	-
PKH_080030	406	98/247 (40%)	55	41	-	-	-	-	-	-	48/160 (30%)	53	23
- PKH_080030 frag	269	56/130 (43%)	44	43	-	-	-	23/68 (34%)	-	-	30/81 (37%)	-	-

P. knowlesi gene candidate	Protein length	Homology (%)											
		P. simiovale			*P. coatneyi*			*P. cynomolgi*			*P. inui*		
		PlasmoDB	NCBI		PlasmoDB	NCBI		PlasmoDB	NCBI		PlasmoDB	NCBI	
		Identities (%)	query cover	Ident	Identities (%)	query cover	Ident	Identities (%)	query cover	Ident	Identities (%)	query cover	Ident
AMA-1	563	-	99	85	492/563 (87%)	100	88	481/563 (85%)	100	86	441/563 (78%)	100	78
MSP1	1821	-	98	67	1265/1905 (66%)	98	67	517/695 (74%)	98	69	469/710 (66%)	53	59
- MSP1-19	92	-	95	76	70/89 (79%)	95	79	72/89 (81%)	96	82	65/87 (75%)	96	77
- MSP1 ag2	88	-	-	-	-	-	-	15/28 (54%)	-	-	-	-	-
SERA3	1079	-	89	66	628/867 (72%)	100	72	668/1080 (62%)	100	62	437/1078 (41%)	100	60
- SERA3 ag1	116	-	-	-	43/91 (47%)	77	47	30/110 (27%)	50	35	23/75 (31%)	-	-
- SERA3 ag2	173	-	-	-	30/158 (19%)	-	-	21/51 (41%)	-	-	20/73 (27%)	-	-
TRAP/SSP2	580	-	-	-	452/582 (78%)	100	79	437/582 (75%)	100	75	340/629 (54%)	100	54
- SSP2 frag	120	-	-	-	52/88 (59%)	96	61	62/111 (56%)	95	57	34/69 (49%)	57	49
TSERA2	526	-	81	41	343/517 (66%)	100	67	217/364 (60%)	100	60	104/288 (36%)	95	36
- TSERA2 ag1	191	-	83	64	126/191 (66%)	99	66	115/173 (66%)	89	66	61/191 (32%)	92	32
- TSERA2 ag2	46	-	95	89	37/44 (84%)	95	84	36/44 (82%)	95	82	18/44 (41%)	95	41
PKH_021580	208	-	-	-	-	-	-	-	-	-	-	-	-
- PKH_021580 frag	155	-	-	-	-	-	-	-	-	-	16/35 (46%)	-	-
PKH_031930	212	-	-	-	24/77 (31%)	-	-	28/129 (22%)	-	-	22/81 (27%)	-	-
- PKH_031930 ag1	93	-	-	-	-	-	-	15/62 (24%)	-	-	-	-	-
- PKH_031930 ag2	112	-	-	-	21/93 (23%)	-	-	-	-	-	15/30 (50%)	-	-
PKH_080030	406	-	-	-	139/278 (50%)	68	50	92/233 (39%)	-	-	-	-	-
- PKH_080030 frag	269	-	-	-	74/137 (54%)	50	54	47/125 (38%)	-	-	-	-	-

Table 4.2 *Plasmodium spp*. PlasmoDB accession codes for candidate genes.

Species	MSP1	AMA1	SERA	TSERA2	SSP2/TRAP
P. knowlesi	PKNH_0728900	PKNH_0931500	PKNH_0413400	PKNH_0413500	PKNH_1265400
P. vivax	PVX_099980	PVX_092275	PVX_003810	PVX_003805	PVX_082735
P. falciparum	PF3D7_0930300	PF3D7_1133400	PF3D7_0207600	PF3D7_0207800	PF3D7_1335900
P. malariae	PmUG01_07042000	PmUG01_09042600	PmUG01_04024800	PmUG01_04024900	PmUG01_12028900
P. ovale curtisi	PocGH01_07037900	PocGH01_09039800	PocGH01_04022100	Pocgh01_04022100	PocGH01_12027200
P. coatneyi	PCOAH_00016080	PCOAH_00026700	PCOAH_00008610	PCOAH_00008610	PCOAH_00042390
P. cynomolgi	PCYB_073770	PCYB_093930	PCYB_042270	PCYB_042280	PCYB_122680
P. inui	C922_01162	C922_02330	C922_02820	C922_02816	C922_01849

At this time, the constructs were screened for the presence or absence of signal peptides (SP) and any transmembrane (TM) domains using the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) and designed to exclude these features which, in addition to the Glutathione Sepharose Transferase (GST) solubility tag, aid in the expression of soluble proteins [112,113]. It has been previously shown by Vedadi *et. al.*[114], and from previous experience from our group, that protein expression and solubility can be hindered by the presence of SP and TM domains. The target genes were further analysed using domain prediction software to determine putative domain boundaries (http://gene3d.biochem.ucl.ac.uk/ and http://smart.embl-heidelberg.de/). These screening steps were necessary to ensure *P. knowlesi*-specificity in order to limit potential antibody cross-reactivity in regions co-endemic with other *Plasmodium* spp., as this would reduce the usefulness of these reagents as serosurveillance tools. Following this selection criteria, 10 *P. knowlesi*-specific constructs (**Table 4.3**) were designed for insertion into the pGEX-2T GST-tagged expression vector (GE Healthcare Life Sciences), a vector designed to facilitate the expression of soluble protein. Each sequence was checked for the presence of *Bam*HI, *Eco*RI and *Sma*I restriction sites on DNASTAR to ensure their absence and prevent conflict with the vector specific restriction sites. Using the same software, predicted antigenic residues were targeted based on the Jameson-Wolf antigenicity index (Protean 4.0, DNAStar), to help predict full or partial sequences within the candidates with a high antigenicity index, although it was not used as a selection criteria.

Table 4.3 Summary of *P. knowlesi*-specific recombinant antigen construct characteristics. Predicted sizes are indicated both with and without the GST tags. Values in parenthesis indicate the empirical sizes of protein aggregates.

Gene ID	Antigen	Description	Chromosome	AA position	Expression (mg/L)		Size (kDa)	
							Predicted	Empirical
PKNH_0728900	MSP1 ag2	merozoite surface protein 1	7	974 - 1061	16.32	w/ GST	34.8	44.7 (106.6)
						w/o GST	8.3	n/a
PKNH_0413400	SERA3 ag1	cysteine protease (Serine repeat-like antigen)	4	25-140	20.5	w/ GST	37.7	49.6 (118.9)
						w/o GST	11.3	n/a
PKNH_0413400	SERA3 ag2	cysteine protease (Serine repeat-like antigen)	4	826-998	15	w/ GST	44.9	68.7 (162.7)
						w/o GST	18.4	n/a
PKNH_1265400	SSP2	sporozoite surface protein 2, putative, thrombospondin-related anonymous protein (TRAP)	12	381-500	17	w/ GST	39.7	53.1 (132.5)
						w/o GST	13.2	n/a
PKNH_0413500	TSERA2 ag1	Truncated cysteine protease	4	60-251	11.9	w/ GST	46.8	59.7 (117.9)
						w/o GST	20.4	n/a
PKNH_0413500	TSERA2 ag2	Truncated cysteine protease	4	324-369	1.8	w/ GST	31.7	38.3
						w/o GST	5.2	n/a
PKNH_0216300	PKH_021580	HP	2	54-208	9.1	w/ GST	44.3	38.2
						w/o GST	17.9	n/a
PKNH_0321300	PKH_031930_ag1	HP	3	8 - 100	6.3	w/ GST	36.1	25.2
						w/o GST	9.6	n/a
PKNH_0321300	PKH_031930_ag2	HP	3	101-212	5.7	w/ GST	39.1	27.2
						w/o GST	12.6	n/a
PKNH_0800300	PKH_080030	HP	8	24-292	35.7	w/ GST	57.6	25
						w/o GST	31.1	n/a

4.2.2. DETERMINATION OF TRANSCRIPTIONAL STATUS OF *P. KNOWLESI* CANDIDATE GENES

RNA was prepared and provided by Dr. Rob Moon (PMB, LSHTM). Briefly, blood stage messenger RNA (mRNA) was isolated from culture adapted *P. knowlesi* parasites, A1-H1 strain[115] to determine the transcriptional status of each of the selected candidate genes. Based on previously described protocols[115], parasites were cultured in modified RPMI medium (Roswell Park Memorial Institute; Life Technologies) at 2% haematocrit, asynchronous parasite cultures (at approximately 5% parasitemia) were centrifuged and 10 volumes of Trizol Reagent (ThermoFisher Scientific) were added per 1 volume of parasite pellet. The modifications to the RPMI medium were: 2.3 g/L sodium bicarbonate, 4 g/L dextrose, 5.95 g/L Hepes, 0.05 g/L hypoxanthine, 5 g/L AlbuMAX II (Thermo Fisher Scientific), 0.025 g/L gentamycin sulfate, 0.292 g/L L-glutamine, and 10% serum. For routine parasite culture 10% (vol/vol) equine serum (Life Technologies) was used. RNA was purified from the sample as per the Trizol Reagent manufacturer's instructions and the resultant pellets were resuspended in DEPC (diethylpyrocarbonate) treated water. Briefly, 75 µl of dH2O, 10 µl of RDD buffer and 15 µl of DNAse (Qiagen) were added to the RNA pellet, incubated at RT for 15 minutes the purified over RNA easy mini column (Qiagen) as per manufacturer's instructions. RNA quality and quantity was measured on a DNA gel and nanodrop (NanoDrop® Spectrophotometer ND-1000; ThermoFisher Scientific). Genomic DNA was removed from the purified RNA preparation using the Turbo DNase free kit (ThermoFischer Scientific). Briefly, in a 50 µl reaction, 10 µg of purified RNA was added to 5 µl buffer and 1 µl Turbo DNase, then incubated at 37°C for 30 minutes in a thermos cycler (Thermomixer C, Eppendorf). After the incubation, 5 µl of DNase inactivation reagent, as described in the kit specifications, was added to the 50 µl reaction and incubated at room temperature (RT) for 5 minutes. The sample was briefly centrifuged (10.000 x g for

1.5 minutes) and the supernatant containing the purified and genomic DNA free RNA was removed to a fresh tube. Dr. Rob Moon performed the assay up to this point. The SuperScript® IV Reverse Transcriptase (RT) kit (Thermo Fisher Scientific) with oligo d(T)20 primers (RT+) was used for the first strand synthesis as per manufacturer's instructions, using 2 µg of the previously prepared total RNA in a 40 µl reaction. An identical reaction was set up as a negative control (RT-) where the SuperScript® IV RT was not added. For PCR analysis of the generated cDNA transcripts, RT+ and RT- samples (diluted at 1:10 in nuclease free H_2O), alongside genomic DNA controls, were used as templates for transcript specific PCR primers for each of the candidate gene sequences (***Table* 4.4**). *Pk*CTRP (circumsporozoite protein and thrombospondin-related adhesive protein [TRAP]-related protein) and *Pk*CSP (circumsporozoite protein), both previously confirmed to be pre-erythrocytic stage targets, were included in the panel as negative controls. Target specific primer pairs designed to flank introns where possible were used to distinguish amplicons from cDNA and gDNA template (***Table* 4.4**). PCRs were carried out using GoTaq Green Master Mix (Promega) using the RTPCR program conditions listed in **Table 4.5** on a thermos cycler. The amplified products were run on a 1.2% agarose gel.

Table 4.4 *P. knowlesi* gene name and ID, primer sequences, primer length, fragment size with and without intron for transcriptional analysis.

Gene	Oligo name	Sequence: (5' to 3')	length (bp)	Size with intron (bp)	Size without intron (bp)
MSP1	PKH072850R1	CCAGTGGACTACACCACGTC	20	463	-
MSP1	PKH072850F1	CCGAACAGGGGTAGAGTGAT	20		
SERA3	PKH041230F1	GCTCCAAGGGGAAAACATTA	20	401	250
SERA3	PKH041230R1	GCAAGTTTGTATTCGCCAGA	20		
SSP2	PKH121770F1	GGAAACCCATTCGAAGAAAA	20	419	-
SSP2	PKH121770R1	TTTCTGTGTGGTTTGGGGTA	20		
TSERA2	PKH041240F1	TGTGCAAGTACACAGCACGA	20	468	264
TSERA2	PKH041240R1	GTCGTATATTGGGGCTTCCT	20		
PKH_021580	PKH021580F1	AATGGTACGTTCGCACATTTT	21	401	277
PKH_021580	PKH021580R1	CATGTTGGAAGGTATCGTTTG	21		
PKH_031930	PKH031930F1	TGTCAGCAAATGACAGCAAA	20	476	239
PKH_031930	PKH031930R1	TCACCCGGCTTATAGCCTAC	20		
PKH_080030	PKH080030F1	ATCTCTTAGGAAACTTCCTCTATTCA	20	401	226
PKH_080030	PKH080030R1	GGAGCGCATTAATTTGTCCT	20		
CTRP	PKCTRPF1	TCCTCACGAGAACAATGCAG	20	436	-
CTRP	PKCTRPR1	TCACTGTCGCTTCCACTGTC	20		
CSP	PKCSPF1	GAAACAACCGAATGAAGGACA	21	467	-
CSP	PKCSPR1	CGTTTCTTCCTTGTCGTGGT	20		

Table 4.5 PCR programs and conditions.

*Elongation time varies according to fragment length. Program U75 is used for PCR products up to 0.75kb. Program O75 is used for PCR products from 0.75kb.

Program name	PCR step	Temperature (°C)	Duration	Cycles
U75*	Initial denaturation	94	2min	x1
	Denaturation	94	15s	x15
	Annealing	45 to 65	30s	
	Elongation	72	45s	
	Denaturation	94	15s	x30
	Annealing	45 to 65	30s	
	Elongation	72	45s	
	Final elongation	72	7min	x1
O75*	Initial denaturation	94	2min	x1
	Denaturation	94	15s	x15
	Annealing	45 to 65	30s	
	Elongation	72	1min	
	Denaturation	94	15s	x30
	Annealing	45 to 65	30s	
	Elongation	72	1min	
	Final elongation	72	7min	x1
RTPCR	Initial denaturation	95	2min	x1
	Denaturation	95	30s	x25
	Annealing	52	30s	
	Elongation	68	60s	
	Final elongation	68	2min	
	Hold	10	10min	x1
PCR55	Initial denaturation	94	2min	x1
	Denaturation	94	15s	x35
	Annealing	55	30s	
	Elongation	65	2min	
	Final elongation	65	7min	x1
BigDye	Initial denaturation	96	1min	x1
	Denaturation	96	30s	x25
	Annealing	50	15s	
	Elongation	60	4min	

4.2.3. CLONING STRATEGIES FOR *P. KNOWLESI*-SPECIFIC RECOMBINANT ANTIGENS

Optimal PCR amplification conditions were determined using a gradient PCR approach. Essentially, primer pairs for each potential amplicon were prepared in a reaction mix pool with *P. knowlesi* genomic DNA at 70 ng/ml and aliquoted at 5 µl per well over 12 wells in PCR strip tubes. Assay parameters for the gradient reaction were programmed into the thermocycler (***Table* 4.6**), with the annealing temperature adjusted to run between 45-65°C over the 12 wells under the appropriate programme for the amplicon size (Program U75 used for PCR products up to 0.75kb and program O75 for PCR products from 0.75kb; **Table 4.5**). The amplicons from each of the 12 wells were then analysed on a 1% agarose gel with ethidium bromide (3 µl per 100ml) (***Figure* 4.1**). The optimal conditions were identified as those that yielded the highest concentration of amplicon with no contaminating, or incorrectly sized bands. Alternatively, optimal conditions were met if the correct size amplicon is achieved with clear separation between the amplicon and any contaminating, or incorrectly sized bands. Following PCR amplification optimisation using the gradient function (G-Storm Thermal Cycler), templates for each construct were PCR amplified from *P. knowlesi* (H strain) genomic DNA using high fidelity taq (Expand High Fidelity PCR system, Roche Applied Science) (PCR conditions U75 and O75 **Table 4.5**). Vector compatible primers were designed for each completed target sequence (***Table* 4.7**). Each resultant product was gel electrophoresed on a 1% agarose gel. Amplicons were purified using a PCR purification kit (Qiagen, UK). If incorrectly sized bands were evident, then amplicons were purified using a gel extraction and purification kit (QIAquick Gel extraction kit, Qiagen, UK) and quantified (NanoDrop® Spectrophotometer ND-1000; ThermoFisher) prior to cloning. Briefly, to quantify the DNA, 2 µl of H_2O was used to calibrate and blank the machine before adding 2 µl of the DNA product to be quantified.

Table 4.6 Gradient PCR temperatures and gel positions for PCR programmes U75 (products <0.75 kb) and O75 (products >0.75 kb).

Position	Temperature (°C)
1	45.2
2	45.7
3	46.8
4	48.7
5	50.9
6	53.6
7	56.1
8	58.9
9	61.8
10	63.7
11	64.7
12	65.4

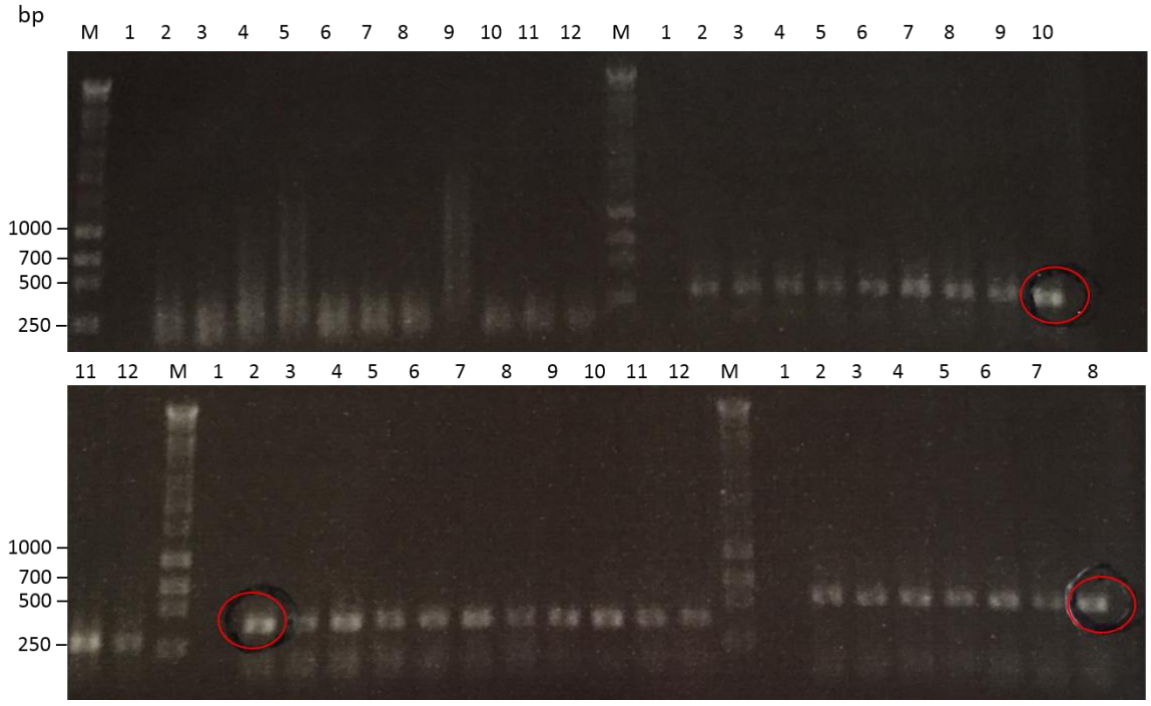


Figure 4.1 Example of a gradient PCR. Each candidate was run at 12 different temperatures in order to find the optimal PCR conditions. Top panel first set of 1-12 lanes are PkMSP1 antigen 1, second set of 1-12 lanes (lanes 11 and 12 are on the bottom panel) are PkMSP1 antigen 2, bottom panel first set of 1-12 lanes are PkSERA3 antigen 1 and second set of 1-8 are PkSERA3 antigen 2. Circles indicate temperatures chosen for that candidate. Failed candidates had to be rerun. Agarose gel at 1%.

Table 4.7 *P. knowlesi* candidate name, primer sequences and primer length. The vector portion of each primer sequence (pGEX-2T) including restriction sites are highlighted in bold and the candidate portion of the sequence in italics. Stop codons are underlined.

Antigen	Oligo name	Sequence : (5' to 3')	Length
MSP1 ag2	PKH072850_2_F1	**GTTCCGCGTGGATCC***TCGGTGACCCCTGCTGTAGTA*	36
MSP1 ag2	PKH072850_2_R1	**GTCACGATGAATTCC**TTA*AGCTTGCACGGTTACGCCTCC*	39
SERA3 ag1	PKH041230_1_F1	**GTTCCGCGTGGATCC***GAGGGAACGGCGCAATCTGGA*	36
SERA3 ag1	PKH041230_1_R1	**GTCACGATGAATTCC**TTA*AGCTTTCTGGGGGGCGACTTG*	39
SERA3 ag2	PKH041230_2_F1	**GTTCCGCGTGGATCC***AACACTGTTGAGGGTCAAGAC*	36
SERA3 ag2	PKH041230_2_R1	**GTCACGATGAATTCC**TTA*GACCATTTTTGCCTTTTGCAA*	39
SSP2	PKH121770F1	**CCAAAATCGGATCTGGTTCCGCGTGGATCC***GCGGCACAGTACCCAGAAAAT*	51
SSP2	PKH121770R1	**GCAGATCGTCAGTCAGTCACGATGAATTCC**TTA*CTCCTCAGATGAGAAGGGACC*	54
TSERA2 ag1	PKH041240F1	**GTTCCGCGTGGATCC**CGATGCACGGATACTGCTATA	36
TSERA2 ag1	PKH041240R1	**ATGAATTCCCGGGGA**TCA*TGGGGTTGGTGTCTGATCTTC*	39
TSERA2 ag2	PKH041240F2	GTTCCGCGTGGATCCAGCAATGTGGAAATGAAAAAG	36
TSERA2 ag2	PKH041240R2	ATGAATTCCCGGGGATCAAGGCATGTCGTTACTTAAGGC	39
PKH_021580	PKH021580F1_V2	**GTTCCGCGTGGATCC***GAGTTACCACTCTCTCCAGACCT*	38
PKH_021580	PKH021580R1	**ATGAATTCCCGGGGA**TCA*TTTAAATGAAGATTTCTG*	36
PKH_031930 ag1	PKH031930F1	**GTTCCGCGTGGATCC***ATATATCTATTCCTCCTGCTA*	36
PKH_031930 ag1	PKH031930R1	**ATGAATTCCCGGGGA**TCA*CAATCTACAGAAATATTGGAA*	39
PKH_031930 ag2	PKH031930F2	**GTTCCGCGTGGATCC**TGGTGCTATAGTTGTTGTACT	36
PKH_031930 ag2	PKH031930R2	**ATGAATTCCCGGGGA**TCA*TCGTTTGCTCTTCCCCTTGAG*	39
PKH_080030	PKH080030F1_V2	**CGTGGATCCCCGGG***GTAATTACACTTGGTACACTA*	35
PKH_080030	PKH080030R1	**GTCACGATGAATTCC**TTA*ATTTACTTTTCTAATTGGATA*	39

To maximise the cloning success rate several cloning strategies were employed the Sequence and Ligation Independent Cloning strategy (SLIC)[116], the Gibson cloning method[117] and conventional ligation dependant cloning (as per manufacturer's instructions) with and without using a shuttle vector (Promega pGEM-T easy).

4.2.3.1. SEQUENCE AND LIGATION INDEPENDENT CLONING (SLIC)

SLIC was performed as previously demonstrated by Mamie *et. al.*[116] with minor alterations. The single-tube reaction mix contained 100 ng of cut pGEX-2T GST tagged expression vector (GE Healthcare Life Sciences) (4.9kb), 3 fold molar excess of insert, 1 µl NEB buffer 2, 1 µl 10x BSA, 1 µl of 0.75/µl T4 DNA polymerase and nuclease free H_2O to adjust volume up to 10 µl. Reaction mix was incubated at RT for 5 minutes, 1 µl of 10 mM dGTP was added to stop the reaction and placed on ice for 1 minute. The reaction was placed on a heating block for 2 minutes at 50^0C and left to slowly cool down at RT before being placed on ice and transformed into competent *Escherichia coli* JM109 or DH5α cells as described below.

4.2.3.2. GIBSON CLONING METHOD

The Gibson method was performed as previously described [117]. The reaction mix contained 100 ng of cut pGEX-2T vector, Gibson Assembly master mix (5x ISO buffer, 10 U/µl T5 exonuclease, 2 U/µl Phusion polymerase, 40 U/µl Taq ligase and distilled water (d H_2O), all to be stored at -20^0C) and insert at a 1:1 ratio (1/3 dilution of stock was necessary to enable working volumes). The assembly reaction was incubated at 50^0C for 1 hour, placed on ice to cool down and 5 µl of the mix was transformed into 100 µl of competent *E. coli* JM109 or DH5α cells as described below.

4.2.3.3. CONVENTIONAL CLONING

Conventional cloning, or restriction site mediated cloning, was performed as per manufacturer's instructions. Briefly, each insert was subjected to a restriction digest

reaction based on the specific restriction sites employed using High-Fidelity restriction enzymes where appropriate for each candidate (New England Biolabs (NEB) BamH1-HF® and EcoR1-HF®). Simultaneously, aliquots of the GST tagged pGEX-2T expression vector was also digested with the appropriate combinations of restriction enzymes. Following incubation at 37^0C for 1 hour, the digested inserts and vector were gel purified, analysed on a 1% agarose gel and quantified using the NanoDrop® spectrophotometer ND-1000 (ThermoFisher Scientific). To concentrate the amounts of digested inserts and vectors each sample was sodium acetate precipitated at -20^0C overnight (o/n), washed in 70% ethanol and resuspended in TE buffer (1/5 original volume). The purified vector digestions and inserts were ligated at a 3:1 vector to insert ratio using T4 ligase and incubated o/n at 4^0C then transformed into competent *E. coli* JM109 or DH5α cells as described below. The shuttle vector method was performed as described previously[118] and as per manufacturer's specifications with minor modifications. Briefly, the purified inserts and linearized TA vector (pGEM-T Easy, Promega) were ligated at a 3:1 insert to vector ratio based on 25 ng of TA vector, using T4 DNA ligase and incubated o/n at 4^0C. The samples were purified then transformed into competent *E.coli* JM109 or DH5α cells as described below and positive colonies were picked to be sequence verified. Correct sequences were restriction digested with the candidate appropriate restriction enzyme, sub-cloned into the GST tagged expression vector as described above and transformed into competent *E. coli* JM109 or DH5α cells as described below.

4.2.3.4. TRANSFORMATION OF *E. COLI* COMPETENT CELLS

For transformations using competent *E. coli* JM109 cells (preparation of Mix & Go! Competent Cells Zymo Research), 5 µl of each purified plasmid DNA mix was added to 50 or 100 µl of cells and was incubated on ice for 5-10 min before being plated onto Luria Bertani (LB)/ampicillin (amp)/ X-Gal (5-Bromo-4-Chloro-3-Indolyl β-D-

Galactopyranoside)/IPTG (Isopropyl β-D-1-thiogalactopyranoside) Agar plates that were left o/n to grow at 37^0C.

For transformations using chemically competent *E. coli* DH5α cells (Invitrogen, ThermoFisher Scientific) reactions were performed as per manufacturer's instructions. Briefly, 1-5 µl of each purified plasmid were aliquoted into 50 µl of thawed DH5α cells and incubated on ice for 30 minutes before being heat-shocked at 42°C for 20 seconds and immediately placed on ice for 2 minutes. The samples were then added to 950 µl of sterile SOC medium (Super Optimal Broth) pre-warmed at 37°C, mixed and incubated at 37°C in a shaking incubator at ~200 rpm for 1 hour in order to resuscitate cells. After incubation, 100 µl of each sample was plated onto LB/amp/ X-Gal/IPTG agar and incubated o/n at 37°C. Transformation controls were performed using pUC19 plasmid DNA provided by the kits.

4.2.3.5. SCREENING FOR POSITIVE TRANSFORMANTS AND PURIFICATION

Positive transformants were identified by colony PCR of blue/white screened colonies (white colonies were indicative of insertion into the vector but colony PCRs were run for further reassurance) using pGEX-2T or pGEM-T Easy vector specific primers (Sigma-Genosys, based on sequence in the GE Healthcare pGEX-2T GST expression vector product specification sheet: pGEX5' transcription start 5' GGGCTGGCAAGCCACGTTTGGTG 3', pGEX3' transcription start 5' CCGGGAGCTGCATGTGTCAGAGG 3'; Promega, based on sequence on Promega pGEM-T Easy technical manual: T7 transcription start 5' GGGCGAATTGGGCCC GACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATT 3', SP6 transcription start 5' ATCACTAGTGAATTCGCGGCCGCCTGCAGGTCGACCATAT GGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTAT TC 3'). Briefly, 10 µl of a PCR reaction mix (5 µl Biomix Red, 0.25 µl of each vector specific primers and 4.5 µl of milli-Q H_2O) together with a sample from each colony was PCR amplified using the PCR55 amplification conditions listed in **Table 4.5**. Following completion of

the run, the PCR products were analysed on a 1% agarose gel. Colonies with positive inserts were used to seed ~5 ml of LB-amp media and incubated o/n at 37^0C in a shaking incubator (New Brunswick Scientific Innova™ 4330 Refrigerated Incubator Shaker). An aliquot of the overnight culture was used to make an interim glycerol stock (850 µl o/n culture plus 150 µl sterile glycerol) to be stored at -80^0C. The cloned plasmids from the o/n samples were then purified (QIAprep Miniprep kit, Qiagen) from each positive transformant, analysed on a 1% agarose gel and quantified by NanoDrop® Spectrophotometer ND-1000 (ThermoFisher).

4.2.3.6. SANGER SEQUENCING-BASED VALIDATION OF CONSTRUCTS

Each construct was sequence validated using Big Dye Terminator cycle sequencing chemistry (Applied Biosystems, UK) and electrophoresed on an ABi 3730x capillary sequencer (Applied Biosystems, UK) to ensure fidelity. Briefly, sequencing templates were prepared using a standard reaction mix (0.5 µl BIGDYE, 1.75 µl buffer, 3.2 µl pGEX5', pGEX3', T7 or SP6 primers, 200-500 ng template DNA and dH_2O to a final volume of 10 µl) and amplified on a Thermal Cycler using the BigDye program shown in **Table 4.5**. A BIGDYE precipitation was performed on all samples (3 µl 3 M sodium acetate, 62.5 µl 95% ethanol and 24.5 µl dH_2O for a final volume of 90 µl per well) as per product specification. The results were analysed using the EditSeq 4.00© software (DNASTAR Lasergene, 1989-1999) or CLC Sequence Viewer 7.7.

4.2.4. PROTEIN EXPRESSION AND PURIFICATION

Sequence validated clones were transformed into BL21(DE3) chemically competent cells (TransGen Biotech) and expressed in autoinduction media (AIM) based on established protocols [119,120] and manufacturer's instructions. Briefly, as performed on the DH5α cells, 1-5 µl of each purified plasmid were aliquoted into 50 µl of thawed BL21(DE3) cells and incubated on ice for 30 minutes before being heat-shocked at 42°C for 45 seconds and immediately placed on ice for 2 minutes. The samples

were then added to 500 μl of sterile SOC medium (Super Optimal Broth) pre-warmed at 37°C, mixed and incubated at 37°C in a shaking incubator at ~200 rpm for 1 hour in order to resuscitate cells. After incubation, 100 μl of each sample was plated onto LB/amp/ X-Gal/IPTG agar and incubated o/n at 37°C. Transformation controls were performed using pUC19 plasmid DNA provided by the kits. Transformed BL21(DE3) were used to prepare long term Microbank™ storage vials as per manufacturer's instructions (Pro-Lab Diagnostics). All successful *E. coli* BL21(DE3) clones were made into glycerol stocks as previously described (section 4.2.3.5 above, screening of positive transformants) and stored at -80^{0}C. Each candidate transformed cells were expressed in approximately 200 – 300 ml auto induction media (AIM)[119,120] (***Table* 4.8**). Samples were left o/n in a shaking incubator at 37^{0}C.

Table 4.8 Autoinduction media reagents, amounts and concentrations to make 1 L of growth media.

Reagent	Amount	Concentration				
			Chemical reagent	**Amount**	**Concentration**	**Notes**
ZY	928 ml	-	N-Z-amine AS	10 g	-	
			yeast extract	5 g	-	
			H_2O	925 ml	-	
$MgSO_4$	1 ml	1M	-	-	-	
5052	20 ml	50x	0.5% glycerol	250 g	-	
			H_2O	1 L	-	
			0.5% glucose	25 g	-	
			0.2% α-lactose	100 g	-	
NPS	50 ml	20x	H_2O	900 ml	-	
			$(NH_4)_2SO_4$	66 g	0.5 M	
			KH_2PO_4	136 g	1 M	
			Na_2HPO_4	142 g	1 M	
Trace metals mixture (all stocks prepared in 50 mM HCl)	1 ml	1000x	H_2O	36 ml	-	
			$FeCL_3.6H_2O$	50 ml	0.1 M	dissolved in 0.1 M HCl
			$Cacl_2$	2 ml	1 M	
			$MnCl_2.4H_2O$	1 ml	1 M	
			$ZnSO_4.7H_2O$	1 ml	1 M	
			$CoCl_2.2H_2O$	1 ml	0.2 M	
			$CuCl_2.2H_2O$	2 ml	0.1 M	
			$NiCl_2.6H_2O$	1 ml	0.2 M	
			$Na_2MoO_4.2H_2O$	2 ml	0.1 M	
			$Na_2SeO_3.5H_2O$	2 ml	0.1 M	
			H_3BO_3	2 ml	0.1 M	
ampicillin	1 ml	50 mg/ml	-		-	

Following expression, the *E. coli* were harvested following centrifugations (6000 x g for 10 min on an F10BCI-6x500y or JA-20 rotor (Avanti® JE Centrifuge, Beckman Coulter) using 500 ml (Beckman Coulter, USA) or 32 ml (Nalgene, UK) vials, respectively), resuspended in 1x Phosphate Buffered Saline (PBS) with 1/10 concentration of protease inhibitors (COmplete™ Ultra Protease inhibitor cocktail, Roche) and lysed using a cell disrupter (Stansted, UK) and processed using standard protocols (GE Healthcare). Expressed proteins to be used on the cytometric bead assay or protein microarray were not treated with protease inhibitors or it was removed by dialysis in order to avoid any potential interference. The lysates were centrifuged at 10,000 rpm for 10 minutes to remove cellular debris from the supernatant and the GST-tagged proteins were then purified by affinity chromatography (Glutathione sepharose 4B; GE Healthcare) as previously described[121]. Briefly, each clarified culture supernatant was incubated with approximately 1 ml of 50% glutathione sepharose 4B beads per ~30 ml of supernatant pre-equilibrated with 1x PBS and incubated for approximately 1 hour at RT. After incubation, the bead suspension is poured over reusable plastic columns (Qiagen, UK) assembled with a porous frit placed at the base to retain the beads, the flow through was discarded and the beads washed extensively with 1x PBS before elution with 0.5 ml of reduced glutathione elution buffer per ml of beads. Fractions from each purified protein were briefly assayed by Bradford assay (BioRad) purely to identify protein-containing fractions, pooled and finally dialysed against 1x PBS before quantifying the protein content later on.

The different proteins were dialysed in order to remove unwanted small molecular weight substances that might interfere in future experimental procedures. Samples were transferred into visking dialysis membranes (Medicell Membranes Ltd, UK) at 12-14 kDa (the GST tag on the samples is 26 kDa) previously hydrated (2% $NaHCO_3$ and 1 mM EDTA incubated at 80°C for 30 minutes) and the membranes placed in a tub filled with 2 L of 1x PBS with a mixer and left o/n at 4^0C. The

following day, the buffer was replaced by a fresh batch and left incubating at RT for approximately 2 hours before transferring the individual samples to appropriately labelled clean tubes and stored at -20°C in 50 ml aliquots.

A Bradford assay was done to determine the concentration of each sample using 10 µl of Bovine Serum Albumin (BSA) serial dilutions as a standard (2, 1, 0.5, 0.25, 0.125, 0.63, 0.31 and 0 mg/ml). To quantify each sample, a 1/5 and 1/10 dilution of the samples were done in 1x PBS or H_2O before adding 200 µl of previously diluted (1/4 in H_2O) Bradford dye. Samples were left to incubate at RT for 5 minutes before being read at 630 nm.

The total protein concentration was also determined by using the Pierce™ BCA protein kit (ThermoFisher Scientific) and Bovine Serum Albumin (BSA) at different concentrations as standard (2000x, 1500x, 1000x, 750x, 500x, 250x, 125x, 25x and 0), as per product specifications. This was done on samples at 1/5 dilution and at 1/10 dilution then read at 630 nm on the plate reader (Dynex Technologies MRX TC II). A bicinchoninic acid assay (BCA) (ThermoFisher Scientific) was also performed on the samples in a similar manner in order to quantify the samples as an alternative to the Bradford assay.

These assays were ran alternatively to the Bradford assay as the latter initially showed inconsistent values for the same batch of samples tested so it was decided to look at alternative quantifying methods. We settled on using the Bradford assay as commercial non-malarial antigens used in the group were regularly assessed using this assay. Assayed samples were stored in small aliquots at -80°C to minimise damage to the proteins from freeze/thaw cycles.

4.2.5. INDIRECT ENZYME-LINKED IMMUNOSORBENT ASSAY (ELISA)

Serum and bloodspot samples were tested for total IgG antibodies against *P. falciparum* (AMA-1[122,123], MSP1-19[124]), *P. vivax* (AMA-1, MSP1-19 (donated as a kind gift from Tony Holder)) and *P. knowlesi* (SERA3 antigens 1 and 2, TSERA2

antigen 1 and SSP2[1]) antigens as previously described[42]. Antigens used to test hospital case serum samples were *Pk*SERA3 antigens 1 and 2, *Pk*SSP2, *Pk*TSERA2 antigen 1 and *Pv*MSP1-19[1]. Antigens used to test the small case study community survey bloodspot samples were *Pk*SERA3 antigen 2, *Pv*AMA-1, *Pv*MSP1-19, *Pf*AMA-1 and *Pf*MSP1-19[2]. Briefly, 96-well Immunolon-4HBX plates were coated o/n at 4^0C at the required concentration of each recombinant antigen diluted in 50 µl of coating buffer per well. Plates were washed 3 x using PBS/T wash buffer and blocked using 1% skimmed milk powder in PBS/T at RT for 3 hours. Plates were washed 3 x using PBS/T and a 1/1,000 or 1/2,000 final dilution of sera (1/1,000 for all *P. knowlesi* antigens and MSP1-19; 1/2,000 for AMA-1) was tested on the different antigens o/n at 4^0C. The starting concentration of samples was 1/200 for both serum and bloodspot samples before being further diluted in the ELISA plate itself (10 µl of sample in 40 µl blocking solution for a 1/1,000 dilution; 5 µl of sample in 45 µl blocking solution for a 1/2,000 dilution). Briefly, bloodspot samples were cut into discs using a 3 mm leather hole puncher and placed into labelled 96-well deep-well (Costar 0.5 ml V well bottom assay block) plates before being eluted in 200 µl of reconstitution buffer (1 g sodium azide in 1 L PBS/Tween20) (appendix III). The eluted samples were sealed (96 cap sealing mat, Thermo scientific) and stored at 4°C and at -20°C for long-term storage. Serum samples were diluted from neat into 200 µl reconstitution buffer and stored under similar conditions as the bloodspot samples (at 4°C). The plates were then washed 5 x and HRP conjugate polyclonal rabbit anti-human IgG antibody was added to all wells at a dilution of 1/15,000 in PBS/T at RT for 3 hours. The plates were washed 5 x as before and antibody responses were detected after development with TMB one component HRP microwell substrate solution (Tebu-bio laboratories) for 15 minutes and stopped with 0.2 M sulphuric acid (H_2SO_4) stop solution. Plates were read at 450 nm on an ELISA plate reader (plates read in London, UK: Dynex Technologies MRX TC II; plates read in Sabah, Malaysia: MultiskanTM FC Microplate Photometer, Thermo Scientific)

and optical density (OD) values recorded. All samples were run in duplicates alongside negative and positive controls, including blank (buffer only) wells. These were used to help standardise across assay runs.

Ethiopian *P. vivax*-positive samples were selected based on positive responses to *P. vivax* AMA-1 and MSP1-19 (n=26) and were used as the *P. vivax*-positive, *P. knowlesi*-negative control group for the assays using the clinical samples[54]. In addition, malaria naïve serum samples (n=29) (anonymous donors at Public Health England (PHE); LSHTM ethics approval #11684) were used as the *P. knowlesi*-negative control group[1]. Written informed consent was obtained from all study participants. Both negative control groups were compared to the responses from the *P. knowlesi*-exposed hospital clinical case samples. OD values were corrected by subtracting the background of the blank wells average OD value.

For the small community survey samples, PHE serum samples were used as a negative control and positive controls used were a lyophilised anti-malaria patient sample (NIBSC, UK; 72/96) for *P. vivax*, a pool of hyper-immune endemic adult Tanzanian pool[38] for *P. falciparum* and pooled *P. knowlesi*-exposed hospital serum samples. Similarly to the assays using the clinical samples, OD values were corrected by subtracting the background of the blank wells average OD value. Values were normalised between plates using the appropriate standardised controls for *P. falciparum* and *P. vivax*, except for *P. knowlesi* due to the lack of a standard control. For all assays, values in excess of 1.5 CV between duplicates were considered fails and either dropped or re-ran. All samples used were anonymised and all serological analysis was performed at the Universiti Malaysia Sabah and the London School of Hygiene and Tropical Medicine.

4.2.6. DEVELOPING AND OPTIMISING A 384-WELL ELISA USING THE BIOMEK® FX^P LABORATORY AUTOMATION WORKSTATION ROBOTIC PLATFORM

A high-throughput 384-well ELISA assay was adapted from the previously established protocol for 96-well ELISA to be performed on the Biomek® FX^P Laboratory Automation Workstation. Alterations were principally made on the volumes of reagents used per plate in order to use the least serum sample available and personalised programs were designed on the Biomek® in order to perform the necessary task (appendix IV). Samples used to optimise this platform were from Bioko Island, Equatorial Guinea, collected as filter paper bloodspot samples between July and August 2012 as part of an annual Malaria Indicator Survey[125] in 18 sentinel sites. These samples were used to investigate dynamics of malaria transmission on Bioko Island. Ethics approval for this study was granted by the Equatorial Guinea Ministry of Health and Social Welfare. Written, informed consent was obtained from all heads of household prior to the survey. A number of these samples (n=160) were chosen for this optimisation assay as they had previously been tested by the standard 96-well ELISA in-house.

4.2.7. STATISTICAL ANALYSIS: SABAH HOSPITAL CASE SAMPLES

Final optical density (OD) values were obtained by subtracting blank OD values, reducing background reactivity. Cut off values for each *P. knowlesi*-specific antigen were calculated based on the average OD values of PHE malaria-naïve control serum samples ± (3xSD). Ensemble boosted regression trees were fit to determine predictive power of antibody responses for classification of *P. knowlesi* exposure. To quantify uncertainty around estimates, 100 datasets were assembled including all seronegative individuals from the malaria unexposed population and an equal number of randomly selected *P. knowlesi* seropositive individuals (from all time points). All models were fit using stratified 10-fold cross validation with model

predictive ability assessed by the area under the receiver operating curve (AUC). The learning rate was set at 0.001 and tree complexity set at 4, to allow for interactions within the dataset. Contribution of responses to each antigen to models was assessed using relative variable importance as described by Elith *et. al*.[126]. In this method, the relative importance of individual predictor variables is calculated as the number of times a variable is selected for splitting, weighted by the squared improvement to the model and averaged over all trees and scaled to 100%. Boosted regression tree analysis was completed by Kim Fornace in R statistical software (v 3.4.2) using the gbm package.

The Wilcoxon-Mann Whitney and Wilcoxon signed rank tests (run on STATA/IC 14.2) were used to generate p values for comparisons between different populations (i.e. reactivity differences between Malaysia and controls) and comparisons within a population (i.e. reactivity differences between days, 0, 7 and 28 from diagnosis and treatment). Scatter plots showing reactivity between *P. knowlesi* recombinant antigens and *P. vivax* MSP1-19 were created using STATA and dot plots showing reactivity to *P. knowlesi* recombinant antigens were created using GraphPad PRISM7.

4.2.8. STATISTICAL ANALYSIS: SABAH AND PALAWAN CASE STUDY COMMUNITY SURVEY SAMPLES

Mixture models were used to fit normalised OD values to define seropositive individuals in the case study communities survey samples, with the distribution of OD values modelled as two Gaussian distributions. In this study, cut off values calculated for *P. vivax* and *P. falciparum* were defined by the average OD of the seronegative population ± (3xSD)[2,127]. Cut off values calculated for the *P. knowlesi* antigen used was defined in a similar way ± (5xSD)[127], a more parsimonious cut off due to a lack of prior data. All OD values were corrected by subtracting the

background of the blank wells per plate. Normalisation of OD readings between plates were done for *P. vivax* and *P. falciparum* but not for *P. knowlesi* due to the lack of a standard control. Cut off values were determined for each antigen because the assays were run in different laboratories. For *P. falciparum* and *P. vivax*, individuals were considered positive if they were positive for either MSP-1 and/or AMA-1. Models were fit separately for each parasite species and site. Risk factors associated with *P. knowlesi* seropositivity were evaluated using multivariate logistic regression. Kim Fornace performed the statistical analysis for this study. The percentage of responders to each antigen from the different sample sets was determined. It was possible to stratify responders by age, gender and region through the use of statistical data analysis software, STATA (StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.) and R statistical software (RStudio Team. 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA).

4.3. RESULTS

4.3.1. IDENTIFICATION, DESIGN AN EXPRESSION OF RECOMBINANT ANTIGENS

Eighteen *P. knowlesi* gene targets were selected based on existing evidence from known serological markers in *P. falciparum* (AMA-1[48,52], MSP1[52,60], SSP2/TRAP[56] and SERA antigens[78,79]) and by interrogating the available *P. knowlesi* literature[93] and existing databases[110,128]. **Figure *4.2*** outlines the experimental strategy used in the identification of the target sequences of interest and the test samples used for this chapter. To identify *P. knowlesi*-specific targets, sequence specificity was determined by aligning the amino acid sequences for each gene targets against other *Plasmodium spp.* using the BlastP algorithm[129] in non-redundant (NCBI: https://is.gd/MzISez) and *Plasmodium* specific databases (Plasmodb: https://is.gd/XOs7vd[110]) (**Table 4.1**; appendix IV). Using this approach, regions or sequences showing high amino acid identity were avoided and regions unique to *P. knowlesi* were identified within the selected genes. These regions were once again screened using the BlastP algorithm in order to ensure *P. knowlesi* specificity.

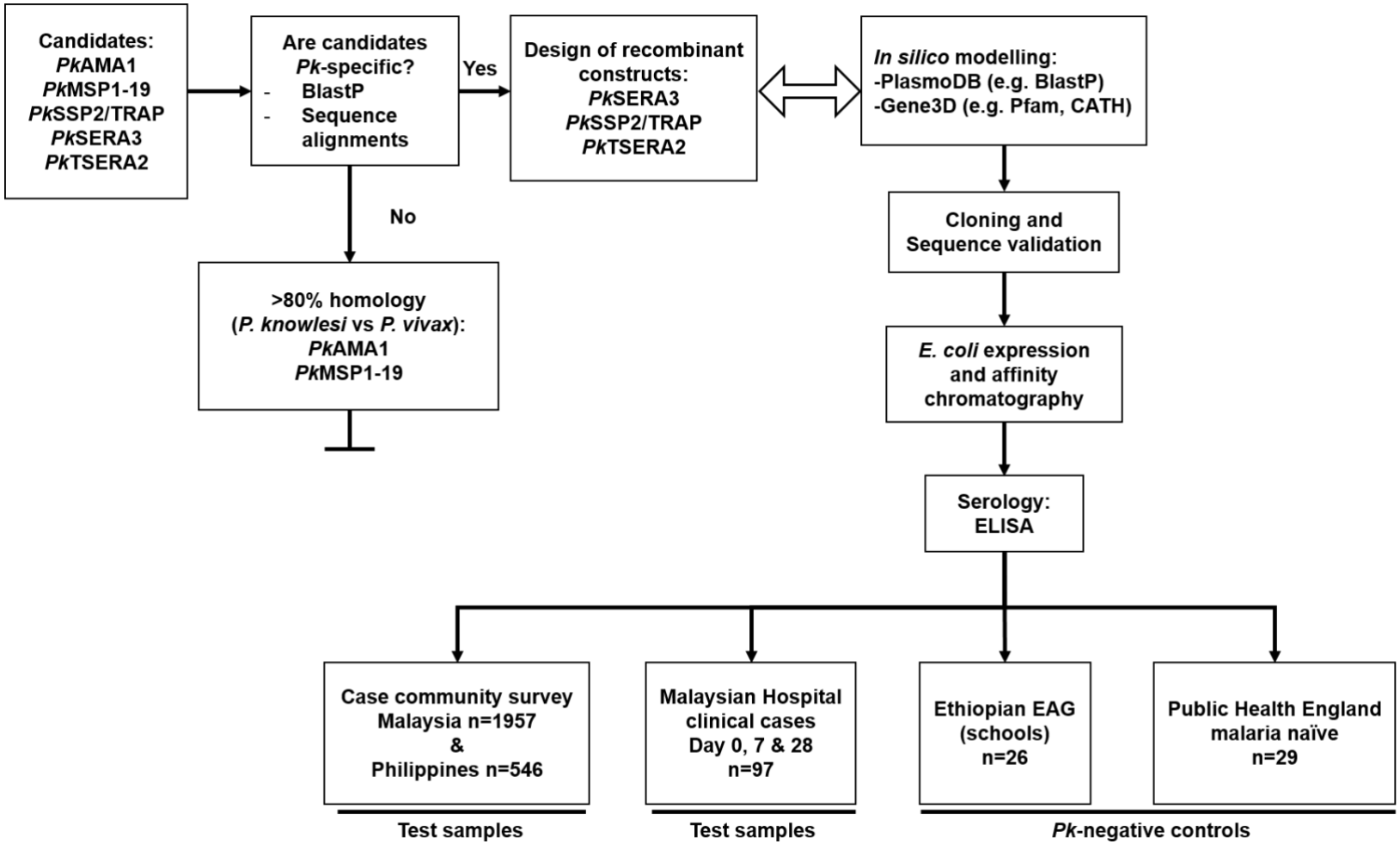


Figure 4.2. Flowchart summarising the experimental strategy used in the identification and validation of the *P. knowlesi*-specific candidates and sample sets used in this chapter. Test samples were obtained from a case community survey in Sabah, Malaysia (n=1957), Palawan, The Philippines (n=546) and from a hospital-based clinical trial in Sabah, Malaysia (n=97). The negative controls comprised of samples from an Easy Access Group (EAG) in Ethiopia (n=26) as *P. vivax*-positive *P. knowlesi*-negative samples and the malaria-naïve group were obtained from Public Health England (n=29).

In order to aid the design of each recombinant target, domain prediction software was used to help predict the sequence structure. Sites such as Gene3D [130] and SMART (http://smart.embl-heidelberg.de/) were employed to help obtain domain prediction information for each gene which helped with the design of truncated fragments (**Figure *4.3***). Although this work was done for all candidates, ***Figure 4.3*** is a representative figure for *Pk*SERA3 antigens 1 and 2, *Pk*SSP2 and *Pk*TSERA2 antigen 1.

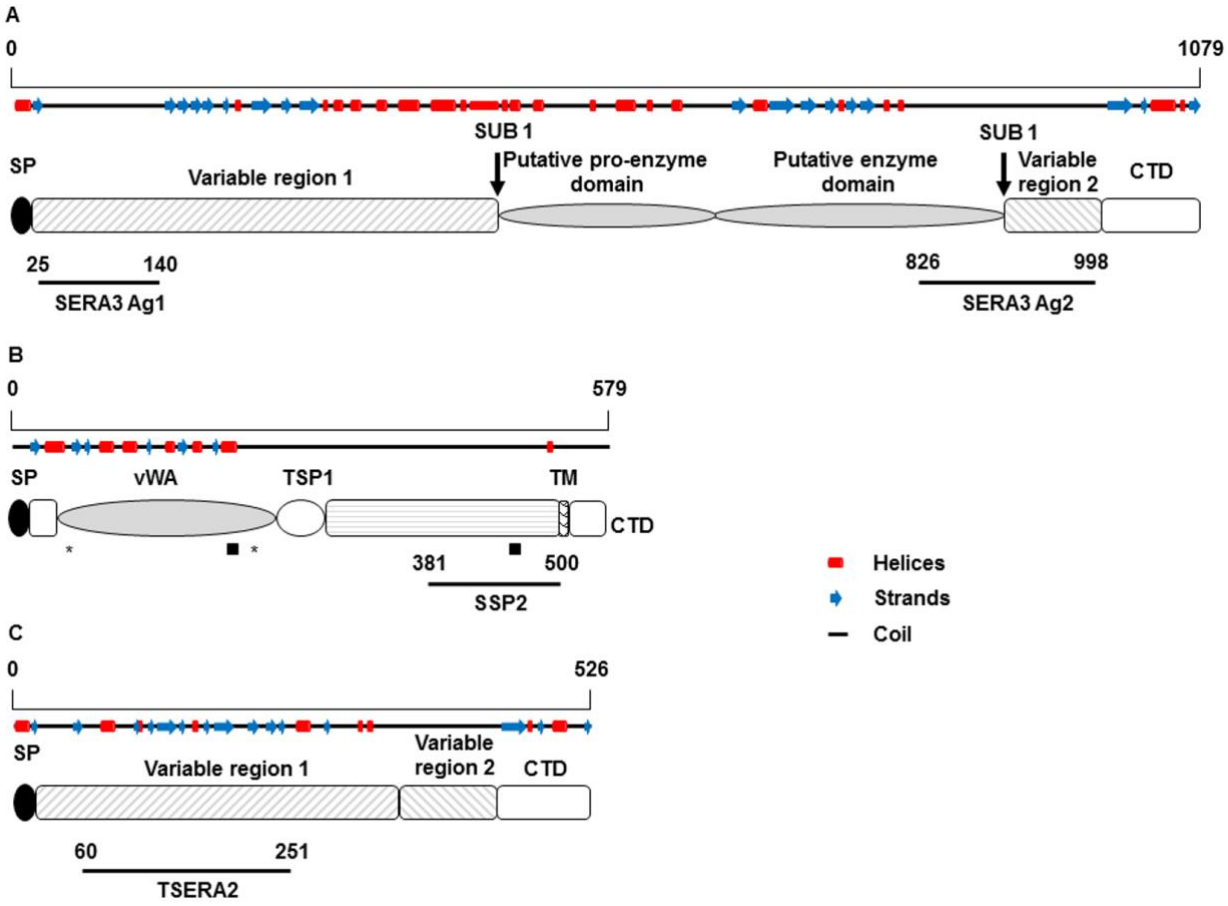


Figure 4.3 Plasmodium knowlesi-specific recombinant antigen constructs. Schematic representations for each protein are shown with key features labelled. (a) PkSERA3 shows the location of the putative pro-enzyme and enzyme domains. The predicted subtilisin (SUB) 1 cleavage sites in relation to variable regions 1 and 2 and the cytoplasmic domain[90]. (b) PkSSP2/TRAP contains a von Willebrandt A domain (vWA), thrombospondin type (TSP) 1 motif, a C-terminal transmembrane (TM) region and a cytoplasmic terminal domain (CTD). Putative T-cell and B-cell epitopes are highlighted with an asterix or black square, respectively[131]. (c) PkTSERA2 shows the lack of central enzyme domain due to truncation of the sequence[90]. Predicted secondary structures generated in I-Tasser[132,133] are shown above each scheme. Red boxes represent helices, blue arrows sheets and the black line coils. The position of recombinant proteins are highlighted below each scheme with the N- and C-terminal amino acid positions indicated. The overall length of each protein is referenced by the amino acid ruler above each secondary structure prediction. For all proteins SP refers to the signal peptide.

This information helped to avoid the unintended premature truncation of potential domains within the sequence. The hypothesis behind this approach was that premature truncation of putative domains could lead to suboptimal expression or difficulties in solubility with the recombinant construct. In addition, both signal peptides (SP) and any transmembrane (TM) regions were omitted from all of the final construct designs. The latter was done because a number of previous studies had suggested that the presence of SP and, or any TM regions could adversely affect the solubility and expression of the recombinant proteins[114].

To further characterise the recombinant sequences, Jameson-Wolf antigenic index was applied (Protean 4.0, DNAStar) to help identify potential antigenic regions. The predicted antigenic residues were targeted and, as shown in ***Figure 4.4***, although based on a prediction, it is clear that the selected sequences for PkSERA3 antigens 1 and 2 (***Figure 4.4*** B), PkSSP2 (***Figure 4.4*** C), PkTSERA2 antigens 1 and 2 (***Figure 4.4*** D) and PKH_080030 (***Figure 4.4*** H) have a high probability of being antigenic, an important characteristic for a serological marker. Other than the SERAs, TRAP/SSP2 and MSP1 antigen, the remaining candidates for expression constructs come from gene families specific to *P. knowlesi* with, as yet, undefined functions.

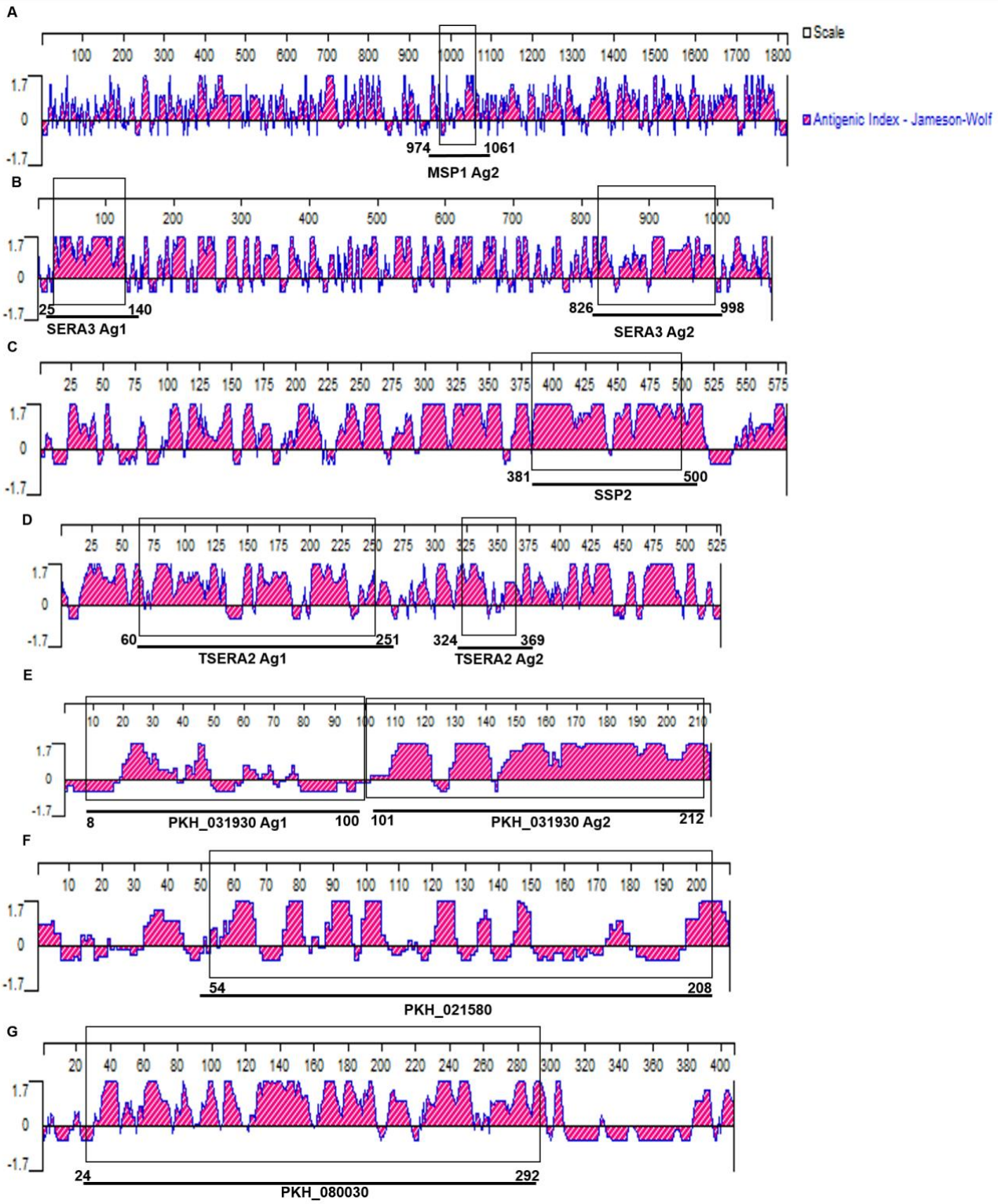


Figure 4.4 Jameson-Wolf antigenicity predictions for *P. knowlesi* candidates. Figure shows antigenicity for the full gene, highlighting the selected candidates in black rectangles: (A) *Pk*MSP1 showing *Pk*MSP1 antigen 2; (B) *Pk*SERA3 showing *Pk*SERA3 antigens 1 and 2; (C) *Pk*SSP2; (D) *Pk*TSERA2 showing *Pk*TSERA2 antigens 1 and 2; (E) PKH_021580; (F) PKH_031930 showing antigens 1 and 2; and (G) PKH_080030. The top and bottom scales indicate the amino acid number and the amino acid positions for the candidates within the gene, respectively. This figures represents 10 of the 18 antigen candidates originally selected.

Using the reference genome *P. knowlesi* H strain, maximum likelihood phylogenetic trees were assembled for the *P. knowlesi* orthologue protein candidates (AMA-1, MSP1, SSP2/TRAP, SERA3 and TSERA2), highlighting the relationship of each candidate between *Plasmodium* species (***Figure 4.5***).This was not done for the remaining candidates as these have been shown to be *P. knowlesi* specific gene candidates with no orthologues found in the other human *Plasmodium spp..* Specifically, for *Pv*AMA-1 (bootstrap value: 100%), *Pv*MSP1 (bootstrap value: 100%) and *Pv*MSP1-19 (bootstrap value: 87%), there is a strong relationship between different *Plasmodium* species, particularly between *P. knowlesi* and *P. vivax* (***Figure 4.5*** A, B and C), highlighted further by corresponding near identical amino acid alignments (appendix IV Figure 1 and 2).

Amino acid alignments were generated using available sequences for human-pathogenic *Plasmodium* spp., which clearly highlight the level of sequence identity for both genes between *P. knowlesi* and *P. vivax* (appendix IV, Figure 1 and 2). Although the bootstrap value strongly supports the grouping of *P. knowlesi* with *P. vivax* for both SERA and TRAP/SSP2 (***Figure 4.5*** D and E; bootstrap value: 100%), the alignments for TRAP/SSP2 and the SERA antigens (PKNH_0413400 and PKNH_0413500) help identify regions specific for *P. knowlesi* (**Figure *4.6***). Based on these screens, any sequences showing high amino acid sequence identity to other *Plasmodium* spp., specifically *P. ovale*, *P. malariae*, *P. falciparum* and *P. vivax*, were re-edited to focus on *P. knowlesi*-specific regions only, where possible.

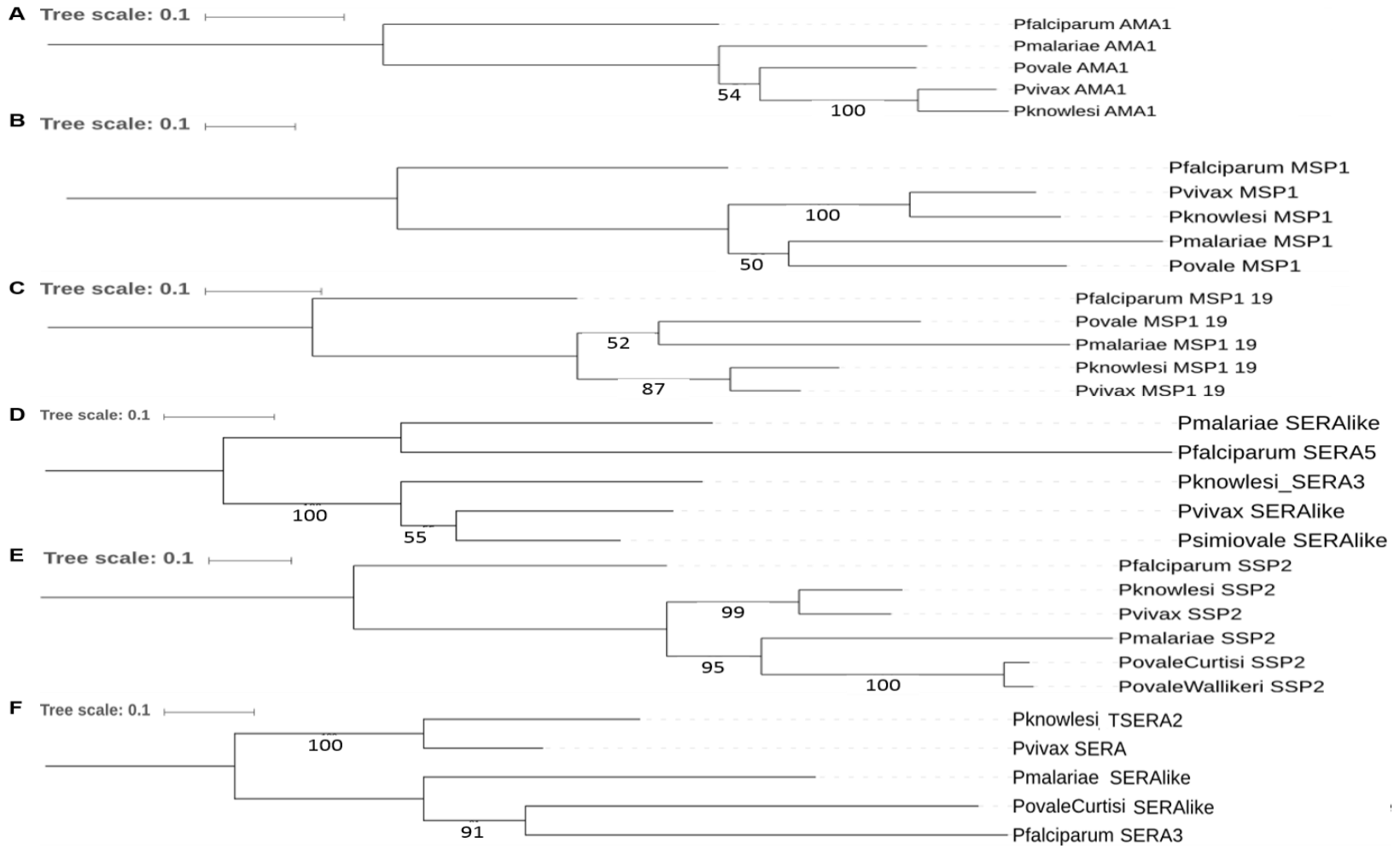


Figure 4.5 Maximum likelihood phylogenetic analysis of the amino acid sequences of AMA1 (a), MSP1 (b), MSP1-19 (c), SERA3 (d), SSP2/TRAP (e) and TSERA2 (f) and gene sequences between *P. knowlesi*, *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale/P. simiovale*. Bootstrap values are given in percentages.

Figure 4.6 Amino acid sequences alignments for SERA3 and SSP2/TRAP gene sequences between *P. knowlesi*, *P. falciparum, P. vivax*, *P. malariae* and *P. ovale/P. simiovale*. *P. knowlesi*-specific sequences selected for development as constructs are highlighted in yellow. Asterisks indicate fully conserved residues, colons indicates strong residue conservation (>0.5, Gonnet PAM 250 matrix), period indicates weak residue conservation (=<0.5, Gonnet PAM 250 matrix). Blank spaces indicate no residue conservation.

A Reverse Transcriptase-PCR (RT-PCR) assay was performed to determine the transcriptional status of our selected candidates in mixed blood stage parasite samples.

It would be ideal to use RNA extract from the different stages of the life cycle rather than a mixed stage sample, but *P. knowlesi* is known to have asynchronous cycles[134], making it difficult to culture and collect samples from each cycle stage. At the time of this experiment, synchronised parasite samples were not available. The RT-PCR confirmed that MSP1, SERA3, TSERA2 and PKH_080030 were actively transcribed in the blood stage (***Figure 4.7***).

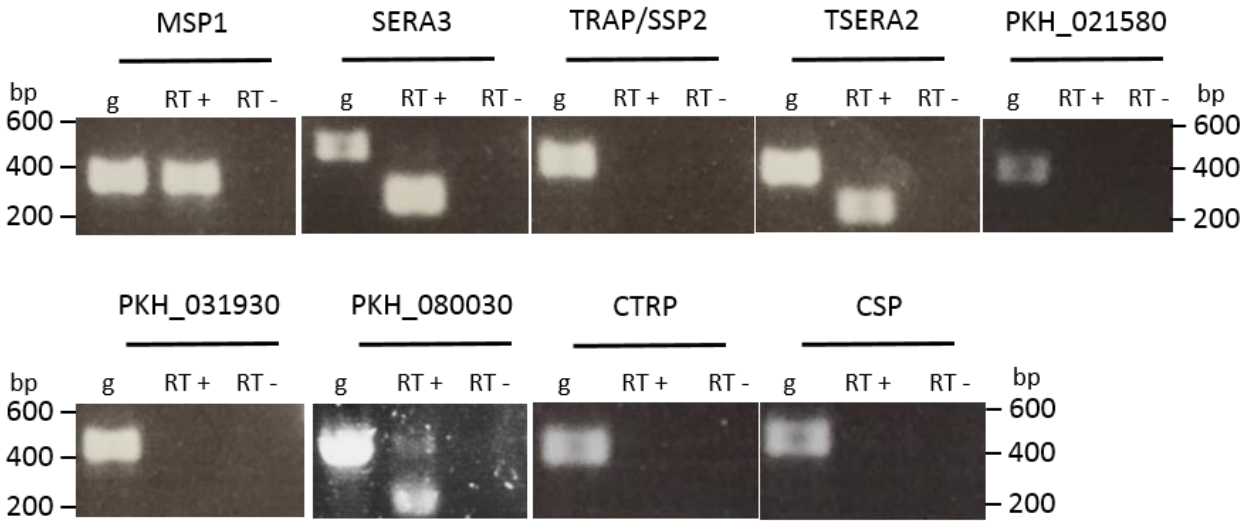


Figure 4.7 *P. knowlesi* candidate gene transcriptional status in parasite mixed blood stage. Panel 1: MSP1; panel 2: SERA3; panel 3: TRAP/SSP2; panel 4: TSERA2; panel 5: PKH_021580; panel 6: PKH_031930; panel 7: PKH_080030; panel 8: CTRP; panel 9: CSP. g refers to genomic DNA, RT+ refers to presence of RT enzyme and RT- refers to absence of RT enzyme. Samples were run on a 1.2% agarose gel. The DNA ladder is indicated in bp (Hyperladder 1Kb, Bioline).

By contrast, PKH_021580 and PKH_031930, along with TRAP/SSP2, which is a sporozoite protein, and the *Pk*CTRP and *Pk*CSP pre-erythrocityc stage controls, were negative by RT-PCR either due to the possibility of the candidate genes not being activated or simply due to the possibility of those candidates not being important in the blood stage of the parasite. In the absence of stage specific RNA

template, we could further speculate that the negative RT-PCR was due to the proteins being expressed during the liver stage or even within the mosquito vector. Thirteen of the 18 selected candidate sequences were successfully PCR amplified to be used in several cloning strategies followed by protein expression. The primary setback that arose with the remaining 5 candidates was the inability to amplify them and the reasons why these could not amplify were not clear. It was possible to exclude inappropriate temperatures as an issue due to the gradient PCR conditions running multiple temperatures simultaneously to determine the ideal amplification temperatures. We reviewed our primer design for these candidates and added 3-6 nucleotides to the sequences. Three of the 5 candidates were successfully amplified with the remaining 2 candidates being excluded from our panel due to lack of time to troubleshoot the problem any further.

Multiple strategies were employed in order to expand the cloning output in the least amount of time. Most of the candidates were successfully cloned using the SLIC cloning method. The Gibson and conventional cloning methods were applied to candidates that failed the SLIC cloning method after a series of troubleshooting steps such as changes in reagent concentrations, vector to insert ratios and incubation periods. Candidates were then transformed and positive transformants were identified by colony PCR as described in the methods (section 4.2.3.5). We determined that, although we had more setbacks using the SLIC method in obtaining any colonies, the conventional and SLIC methods produced more positive transformants than the Gibson method overall. Sixteen candidates were successfully cloned and transformed. By way of further validation, each protein construct was sequence verified by capillary sequencing to confirm each sequence and the position of the stop codons before expression.

Figure 4.8 shows us the amplified *P. knowlesi*-specific PCR products for some of these candidates. All samples fall between 0.25 and 1kb in sizes. It was important not to exceed 1kb in sizes as the increase in size could potentially impact on the

solubility of the final product[114]. The major issues found in this step were failures in sequencing or errors in the individual clone sequences. For all candidates that failed to be sequenced we repeated the sequencing step to exclude any assay errors. A number of candidates had problems due to the sequencing primers failing but it was not possible to troubleshoot this at that time. For candidates that showed errors in the clone sequences we repeated the cloning step in order to obtain clones with the appropriate sequences. It was found that some of our candidates had errors in the sequences due to the design of the original primers, therefore these were redesigned and cloned. Candidates PKH_140010, PKH_073420, PKH_000540, TSERA2 antigen 3 and two others not shown in the figure were excluded from the current panel since these did not pass the sequencing stage after a number of troubleshooting attempts and it was not possible to repeat cloning due to time constraints.

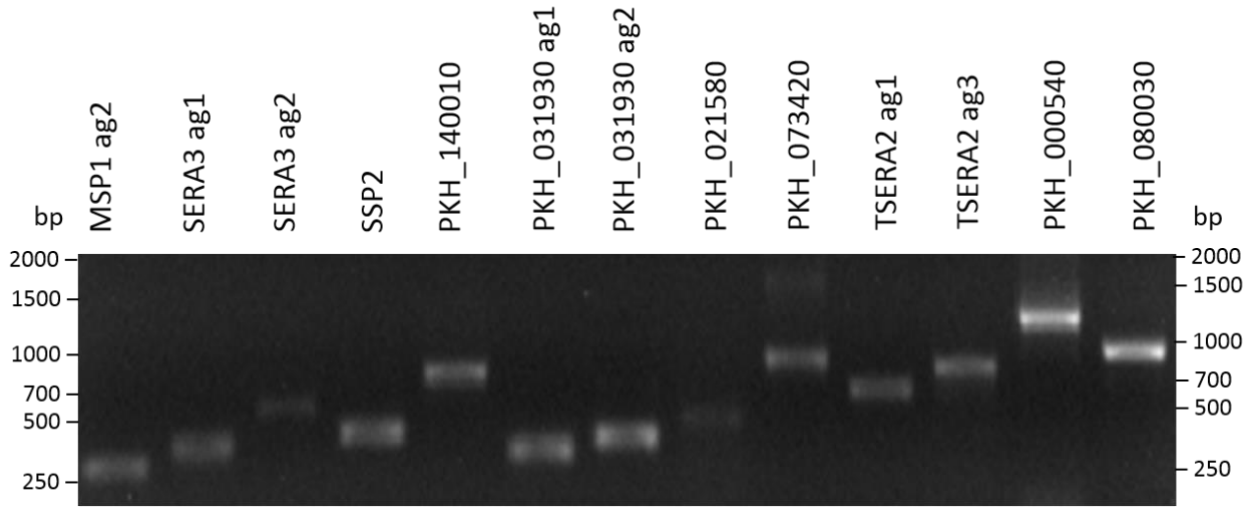


Figure 4.8 Purified *P. knowlesi*-specific PCR amplified products. Lane 1: *Pk*MSP1 ag2; lane 2: *Pk*SERA3 ag1; lane 3: *Pk*SERA3 ag2; lane 4: *Pk*SSP2; lane 5: PKH_140010; lane 6: PKH_031930 ag1; lane 7: PKH_031930 ag2; lane 8: PKH_021580; lane 9: PKH_073420; lane 10: *Pk*TSERA2 ag1; lane 11: *Pk*TSERA2 ag3; lane 12: PKH_000540; lane 13: PKH_080030. M refers to the DNA ladder (Hyperladder 1kb plus, Bioline). Products were run on a 1% agarose gel.

All the antigens were cloned into the pGEX-2T GST-tagged expression vector and expressed in BL21(DE3) *E. coli* expression cells (section 4.2.4.) as soluble products with final yields ranging from 11.9 – 20.5 mg/L (**Figure *4.9***; **Table 4.3**). The major

issue found at this step was having some of the candidates not being expressed in the allowed time frame (overnight expression). These were allowed to incubate for slightly longer periods of time in order to allow maximum cell growth with the least degradation. Candidates that failed to express with the extended time frame were transformed into new batches of expression cell lines in order to exclude cell degradation as the setback. All candidates that passed the sequencing stage were successfully expressed.

Based on their predicted molecular masses (including the GST tag), SDS-PAGE analysis of the purified proteins clearly suggested multimerisation of the purified products (both monomer and dimer; ***Figure 4.9***; **Table 4.3**). Further analysis of the protein candidates was attempted by running Western blots using the hospital case samples[135] in order to determine reactivity towards each recombinant, which was successfully observed. The Coomassie stained profile also illustrated that there is very little non-specific degradation of the recombinant proteins (***Figure 4.9***), suggesting that the proteins are stable under the conditions used. The protein sizes for each protein were larger than predicted, so called “gel shifting” when ran on SDS-PAGE, which is not uncommon[136-138].

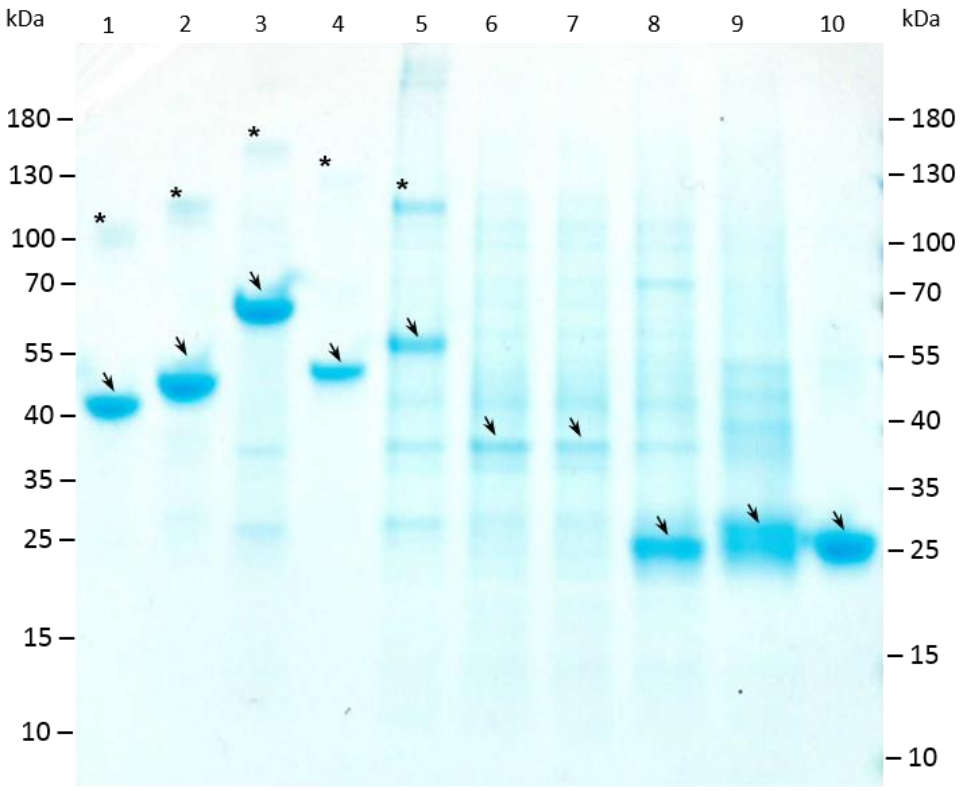


Figure 4.9 SDS-PAGE of purified recombinants in reducing conditions. Lane 1: MSP1 ag2; Lane 2: SERA3 ag1; Lane 3: SERA3 ag2; Lane 4: SSP2; Lane 5: TSERA2 ag1; Lane 6: TSERA2 ag2; Lane 7: PKH_021580; Lane 8: PKH_031930 ag1; Lane 9: PKH_031930 ag2; Lane 10: PKH_080030. Band sizes are indicated on one side in kDa (Fermentas PageRuler Prestained Protein Ladder). Samples were ran under reducing conditions at approximately 0.4 µg/ml per lane on a 4-20% gradient Tris-glycine gel (NuPAGE, BioRad) and stained with Coomassie Blue (BioSafe, BioRad). The arrows and asterisks indicate the protein monomers and aggregates, respectively.

Although not fully explained for all proteins classes, evidence suggests that the presence of acidic residues, net hydropathy or protein aggregation can reduce the effectiveness of SDS in altering the charge, and therefore the migration of proteins through the gel[139,140]. The fact that all 10 protein constructs exhibited signs of protein aggregation supports the suggestion that aggregation may affect protein migration on polyacrylamide gels.

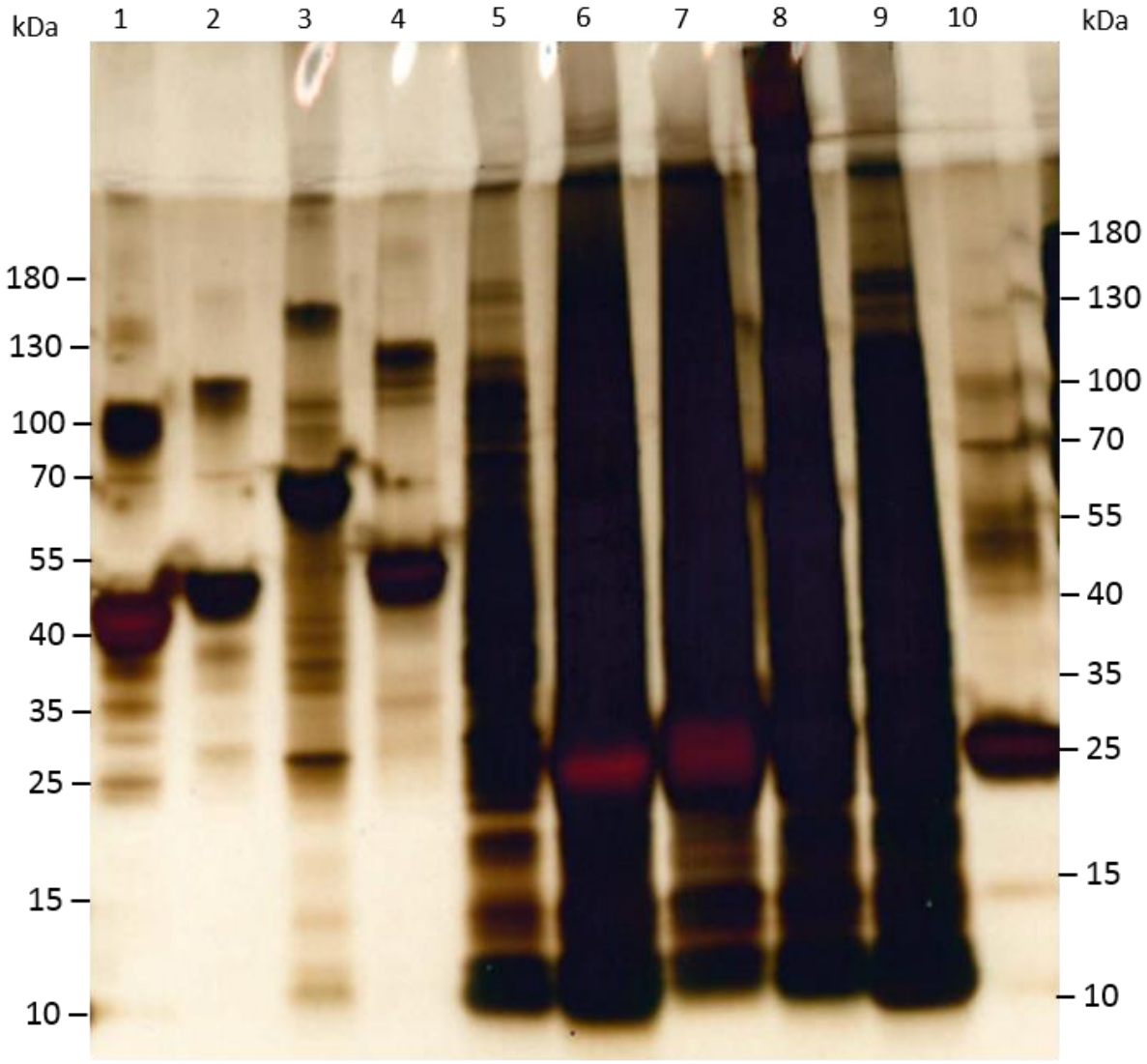


Figure 4.10 SDS-PAGE of purified recombinants. Lane 1: MSP1 ag2; Lane 2: SERA3 ag1; Lane 3: SERA3 ag2; Lane 4: SSP2; Lane 5: TSERA2 ag1; Lane 6: TSERA2 ag2; Lane 7: PKH_021580; Lane 8: PKH_031930 ag1; Lane 9: PKH_031930 ag2; Lane 10: PKH_080030. Band sizes are indicated on one side in kDa (Fermentas PageRuler Prestained Protein Ladder). Samples were ran under reducing conditions at approximately 0.4 µg /ml per lane on a 4-20% gradient Tris-glycine gel (NuPAGE, BioRad) and stained with Silver Stain (Thermo Scientific).

The Silver stained profiles (**Figure *4.10***) only reinforce what was observed in the Coomassie stained gel, confirming that the expressed candidates are highly concentrated although some appear slightly degraded. Silver stain is a much more sensitive staining assay with a detection limit of under 1 ng[141,142] (coomassie staining has a detection limit of approximately 100 ng). All products were successfully expressed but the profile of lanes 5 to 9 (TSERA2 antigen 1 and antigen 2,

PKH_02580, PKH_031930 antigens 1 and 2) suggests further engineering would be required to improve the expressed product.

4.3.2. SINGLE-NUCLEOTIDE POLYMORPHISM (SNP) ANALYSIS IN *P. KNOWLESI* TARGET GENES

P. knowlesi has recently been described as falling into three subpopulations in Malaysia, one from long-term laboratory isolates (cluster 1) and two associated with clinical human infections from separate macaque species reservoir hosts, the pig-tailed (cluster 2) and long-tailed macaques (cluster 3)[143]. To establish a solid and reliable serological tool to be used in large population studies, ideally these tools should not be biased towards a single cluster, which would limit the value of these reagents. We characterised the existence of these SNPs associated with the different clusters for each selected *P. knowlesi* specific candidate, focussing on non-synonymous (amino acid altering) positions within the *P. knowlesi*-specific truncated constructs. ***Table 4.9*** summarises the synonymous and non-synonymous SNPs associated with the three clusters for each *P. knowlesi* candidate, suggesting that the majority of SNPs are found in sequence stretches not included in the candidate antigen areas (detailed SNP files in Appendix IV, table 1). The majority of non-synonymous SNPs were found in regions not covered by the antigen design of all candidates. By omitting the majority of these cluster-specific SNPs we hoped to avoid segregation of detectable antibodies according to the defined clusters. The relevance of these genetic clusters in the context of immunity, and the potential relevance to host preferences is yet to be defined. Dr Jody Phelan's input was obtained in order to compile this data.

Table 4.9 Single-nucleotide polymorphism (SNP) frequencies of Malaysian clinical isolates sequences within *P. knowlesi* candidate genes. Full gene and candidate sequence SNPs are shown in total and for each of the three clusters. The type of SNPs are indicated.

				cluster 1	cluster 2	cluster 3	Type		
Gene ID	**Candidate name**	**Chromosome**	**SNPs**	**Lab Adapted n=8**	**Pig-tailed n=5**	**Long-tailed n=32**	**Non-synonymous**	**Synonymous**	**Stop-Gained**
PKNH_0728900	Full gene		**287**	**92**	**100**	**250**	**113**	**173**	**1**
	MSP1_Ag 2	7	38	10	13	36	30	8	0
PKNH_0413400	Full gene		**146**	**55**	**75**	**126**	**58**	**85**	**0**
	SERA3 ag1	4	4	2	1	2	4	1	0
	SERA3 ag2	4	3	1	1	3	1	2	0
PKNH_1265400	Full gene		**71**	**23**	**12**	**57**	**46**	**24**	**1**
	SSP2	12	17	6	4	14	12	5	0
PKNH_0216300	Full gene		**44**	**19**	**14**	**34**	**30**	**12**	**2**
	PKH_021580	2	31	13	7	23	20	9	2
PKNH_0321300	Full gene		**2**	**0**	**1**	**2**	**1**	**1**	**0**
	PKH_031930 ag1	3	0	0	0	0	0	0	0
	PKH_031930 ag2	3	0	0	0	0	0	0	0
PKNH_0413500	Full gene		**122**	**56**	**48**	**107**	**69**	**52**	**0**
	TSERA2 ag1	4	39	21	16	36	17	22	0
	TSERA2 ag2	4	22	9	6	20	15	7	0
PKNH_0800300	Full gene		**2**	**1**	**1**	**1**	**0**	**2**	**0**
	PKH_080030	8	2	1	1	1	0	2	0

4.3.3. OPTIMISATION OF A 384-WELL ELISA METHOD ALONGSIDE 96-WELL STANDARD ELISA METHOD

During the course of the project we were granted access to the Biomek robotic platform, which we attempted to re-task as a high throughput serological platform utilising the least reagent possible. A sample set from Bioko, Equatorial Guinea, previously tested by 96-well ELISA on *P. falciparum* antigens (AMA-1 and MSP1-19) was selected for optimising this platform. For both *Pf*AMA-1 and *Pf*MSP1-19, inconsistencies were initially found between duplicate wells in the first run compared to the second run, where duplicate wells were shown to correlate with each other (Run 1: *Pf*AMA-1 R^2=0.1166 *Pf*MSP1-19 R^2=0.5524; Run 2: *Pf*AMA-1 R^2=0.9382 *Pf*MSP1-19 R^2=0.8374; ***Figure 4.11*** A, C, B and D, respectively, and Figure ***4.12***). In order to determine if results for both platforms could be comparable, comparisons were made between the 384-well and 96-well ELISAs, showing strong correlations in both runs (run 1: *Pf*AMA-1 R^2=0.328*; Pf*MSP1-19 R^2=0.8807; run 2: *Pf*AMA-1 R^2=0.9257; *Pf*MSP1-19 R^2=0.8893; ***Figure 4.13***), with the exception of *Pf*AMA1 run 1 (R^2=0.328). Results showed a slight tendency to having higher OD values in the samples tested using 384-well ELISA compared to 96-well ELISA. Although promising, the preparation of reagents into the format required by the Biomek robotic platform proved to be more time consuming than a regular ELISA and the lack of expertise with the software reduced the possibility of further optimising the platform for our use. In addition, two high throughput multiplex platforms were purchased by the group. The suspension bead array Luminex MAGPIX® reader (Luminex corp, USA) and two protein microarray platforms (Genepix® (Molecular Devices,USA) and the Araycam® (Grace Bio-Labs, USA)).

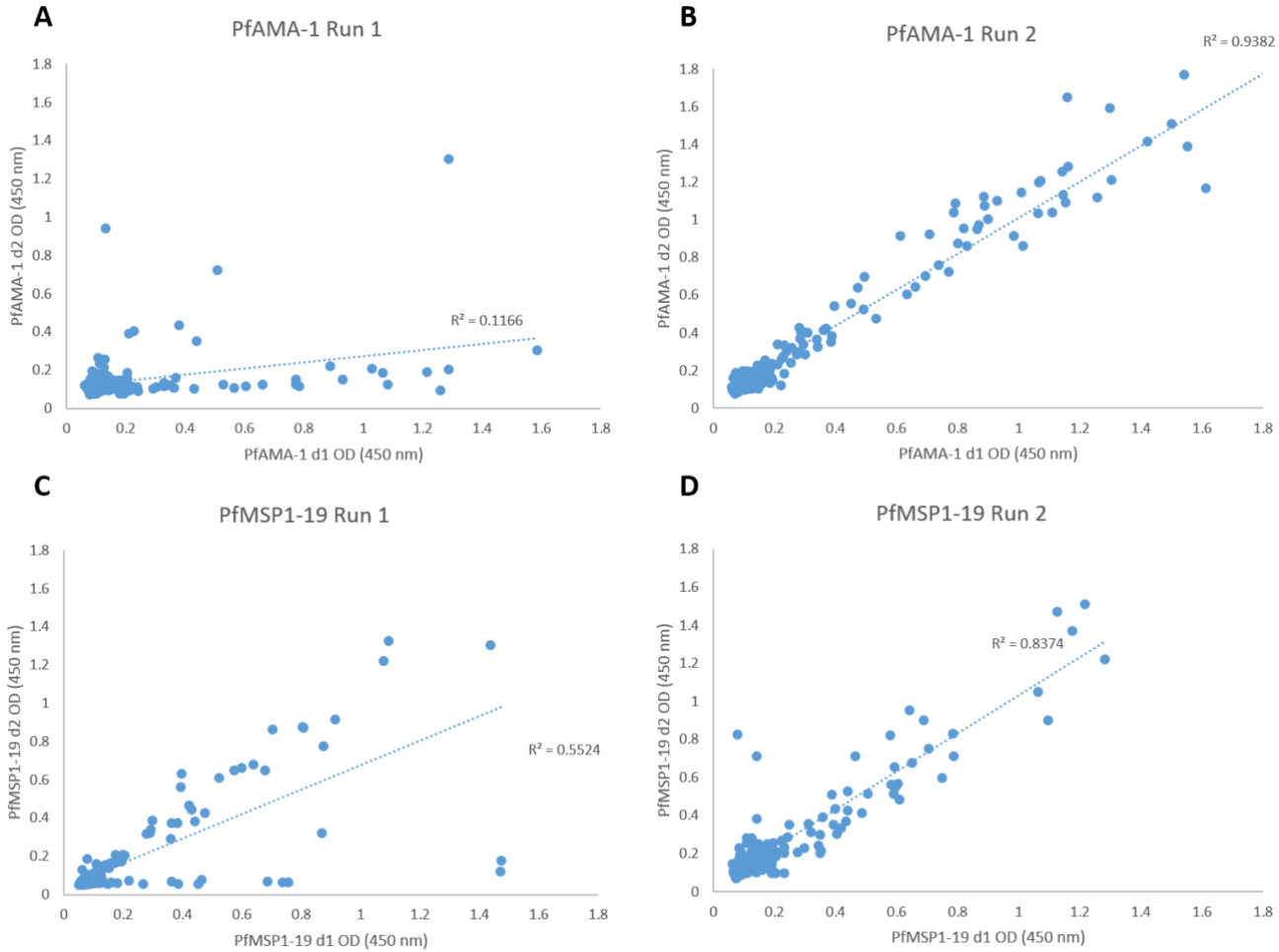


Figure 4.11 Reactivity differences between runs and duplicates using 384-well ELISA plates on the Biomek for *Pf*AMA1 and *Pf*MSP1-19. Data is plotted as raw OD values at 450 nm with R^2 values for the trendlines: *Pf*AMA1 run 1 R^2 =0.1166 (A); *Pf*AMA1 run 2 R^2 =0.9382 (B); *Pf*MSP1-19 run 1 R^2 =0.5524 (C); *Pf*MSP1-19 run 2 R^2 =0.8374 (D).

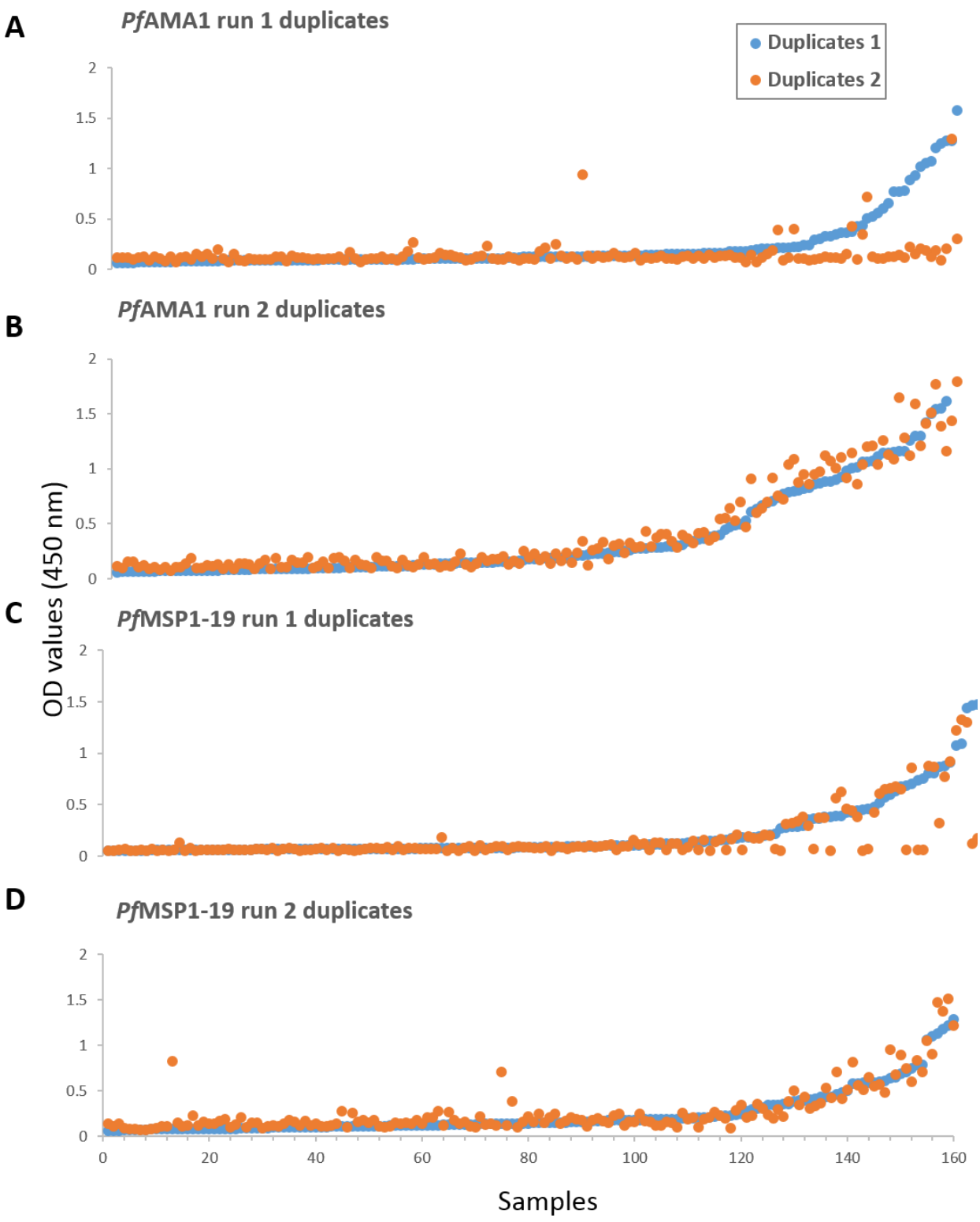


Figure 4.12 Analysis of duplicate samples per Biomek 384-well ELISA run for *Pf*AMA1 (run 1: A; and run 2: B) and *Pf*MSP1-19 (run 1: C; and run 2: D). Data is plotted as raw OD values at 450 nm showing sample duplicates from well 1 (orange) and well 2 (blue) of each sample.

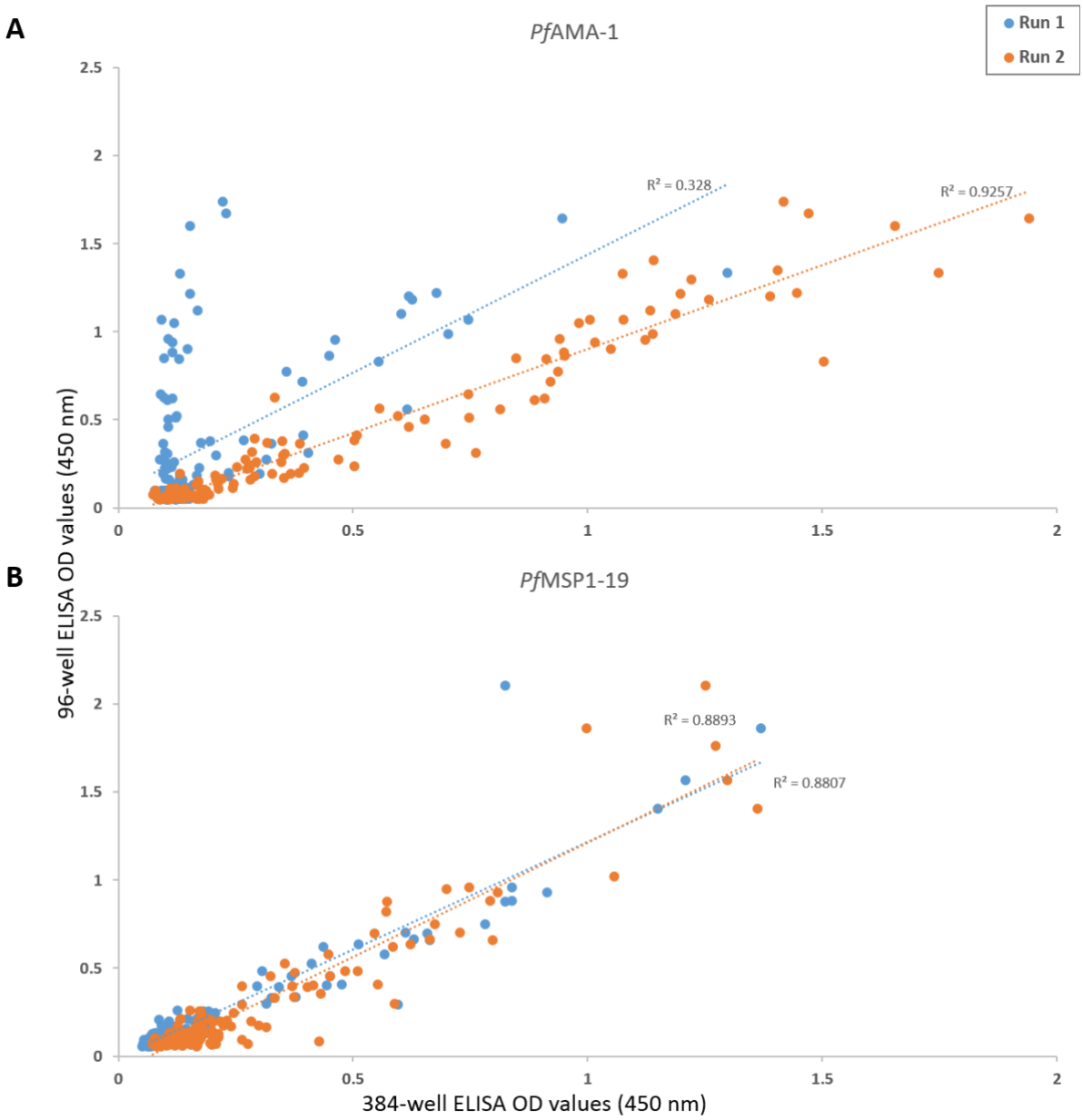


Figure 4.13 Comparisons between 96-well and 384-well ELISAs for *Pf*AMA1 and *Pf*MSP1-19. Data is plotted as raw OD values at 450 nm with R^2 values for the trendlines indicated: *Pf*AMA1 run 1 (blue) R^2=0.328 and run 2 (orange) R^2=0.9257 (A); *Pf*MSP1-19 run 1 (blue) R^2=0.8807 and run 2 (orange) R^2=0.8893 (B).

4.3.4. SERUM REACTIVITY TO *P. KNOWLESI*-SPECIFIC RECOMBINANT ANTIGENS

In Sabah, Malaysia, serum samples were collected from Malaysian adults and children hospitalised with malaria infection at different time points from day of diagnosis (day 0), 7 and 28 days after treatment. The samples were down selected to identify 97 samples which the epidemiological data identified as being infected with *P. knowlesi* and were used for this experiment.

Hospital case samples were assayed by enzyme-linked immunosorbent assay (ELISA) using four antigens from the initial *P. knowlesi*-specific protein panel (SERA3 antigens 1 and 2, SSP2 and TSERA2 antigen 1) (chapter 3, section 3.1, Table 3.3) and a *P. vivax* control antigen (MSP1-19). Ethiopian *P. vivax*-positive, *P. knowlesi*-negative malaria endemic children's sera (n=26)[54] and adult UK malaria naïve sera (n=29) were used as a *P. knowlesi*-negative control panel.
Reactivity towards *Pv*MSP1-19 was seen in the *P. knowlesi*-negative malaria endemic control samples, due to previous exposure to *P. vivax*, but not in the malaria naïve control samples, which showed no reactivity to any of the antigens tested (***Figure 4.14***, top panel) (average OD values from PHE samples: *Pk*SERA3 ag1 OD=0.124; *Pk*SERA3 ag2 OD=0.131; *Pk*SSP2 OD=0.117; *Pk*TSERA2 ag1 OD=0.118). Detectable antibody reactivity to the *P. knowlesi*-specific antigens was not seen in the control group, with the exception of one weakly positive sample to *Pk*SERA3 antigen 1 and another to *Pk*SSP2 (***Figure 4.14***, first and third graphs from the top panel).

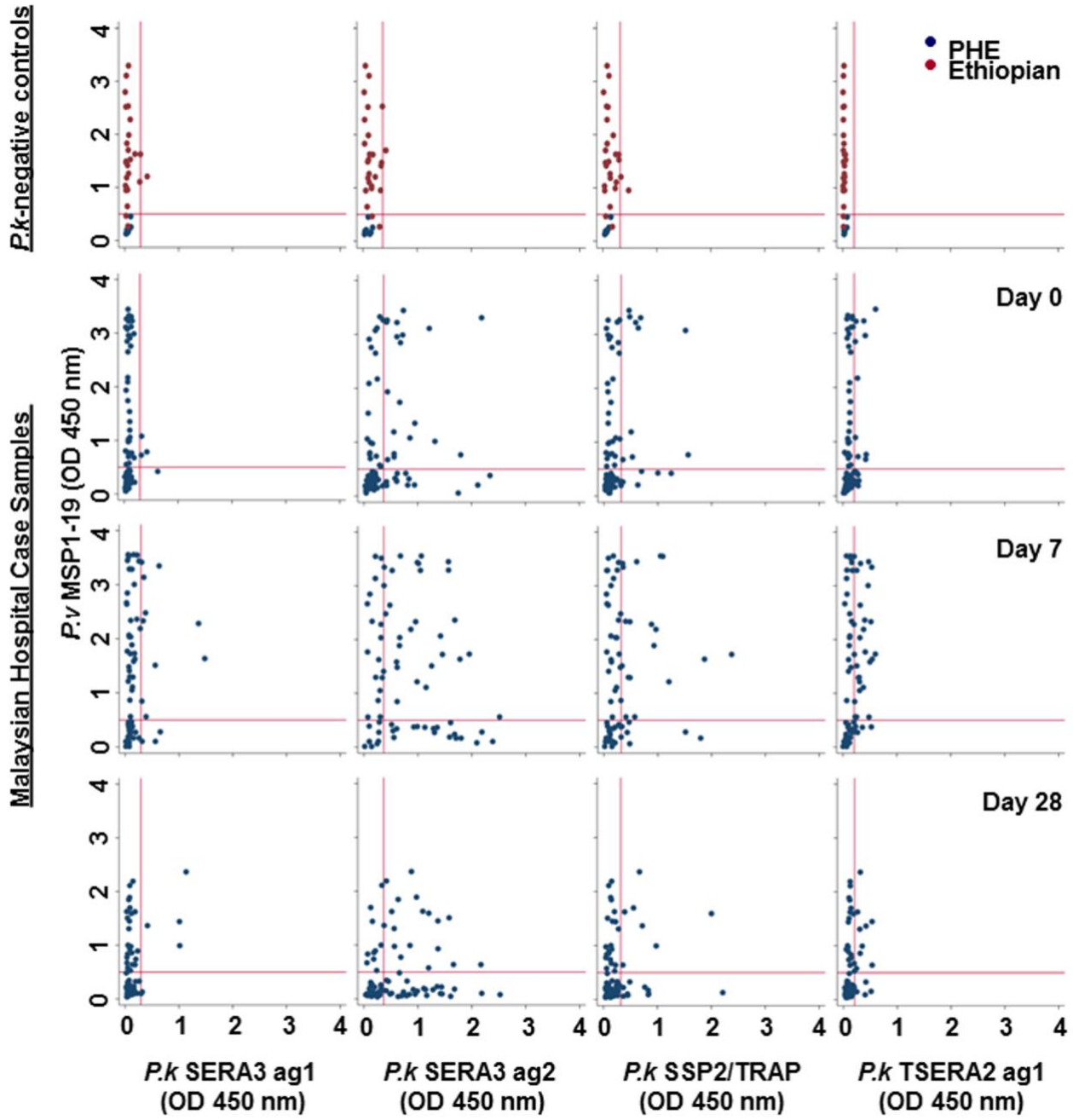


Figure 4.14 Endemic and *P. knowlesi*-negative sera reactivity to *Plasmodium knowlesi*-specific antigens. Scatter plots showing sera reactivity to*: P. vivax* MSP1-19 with *P. knowlesi* SERA3 ag1 (column 1), SERA3 ag2 (column 2), SSP2/TRAP (column 3) and TSERA2 ag1 (column 4) antigens. Sera samples from *P. knowlesi* negative controls n = 55 (row 1; PHE UK malaria naïve (blue), Ethiopian children (red)) and Malaysian hospital case sera samples from days 0 (n = 92), 7 (n = 72) and 28 (n = 77) of diagnosis (rows 2±4, respectively). The red line in each graph represent the cut off values for the respective *P. knowlesi* antigen and was calculated based on Public Health England negative control sera samples (average ODs ± (3xSD)): The vertical cut off line is based on *Pv*MSP1-19 = 0.501. The horizontal cut off line for the *P. knowlesi* antigens were based on the following values: SERA3 ag1 = 0.292; SERA3 ag2 = 0.366; SSP2/TRAP = 0.322 and TSERA2 ag1 = 0.208.

From all three time points, day 7 showed the highest antibody reactivity to all four *P. knowlesi* antigens (**Figure *4.14***, ***Figure 4.15*** and **Figure *4.16***) with *Pk*SERA3 antigen 2 showing the highest prevalence, 63.8% (46/72), in comparison to *Pk*SERA3 antigen 1 (18.1%, 13/72), *Pk*SSP2 (33.3%, 45/72) and *Pk*TSERA2 antigen 1 (43.1%, 31/72) (***Figure 4.14***, columns 1, 3 and 4).

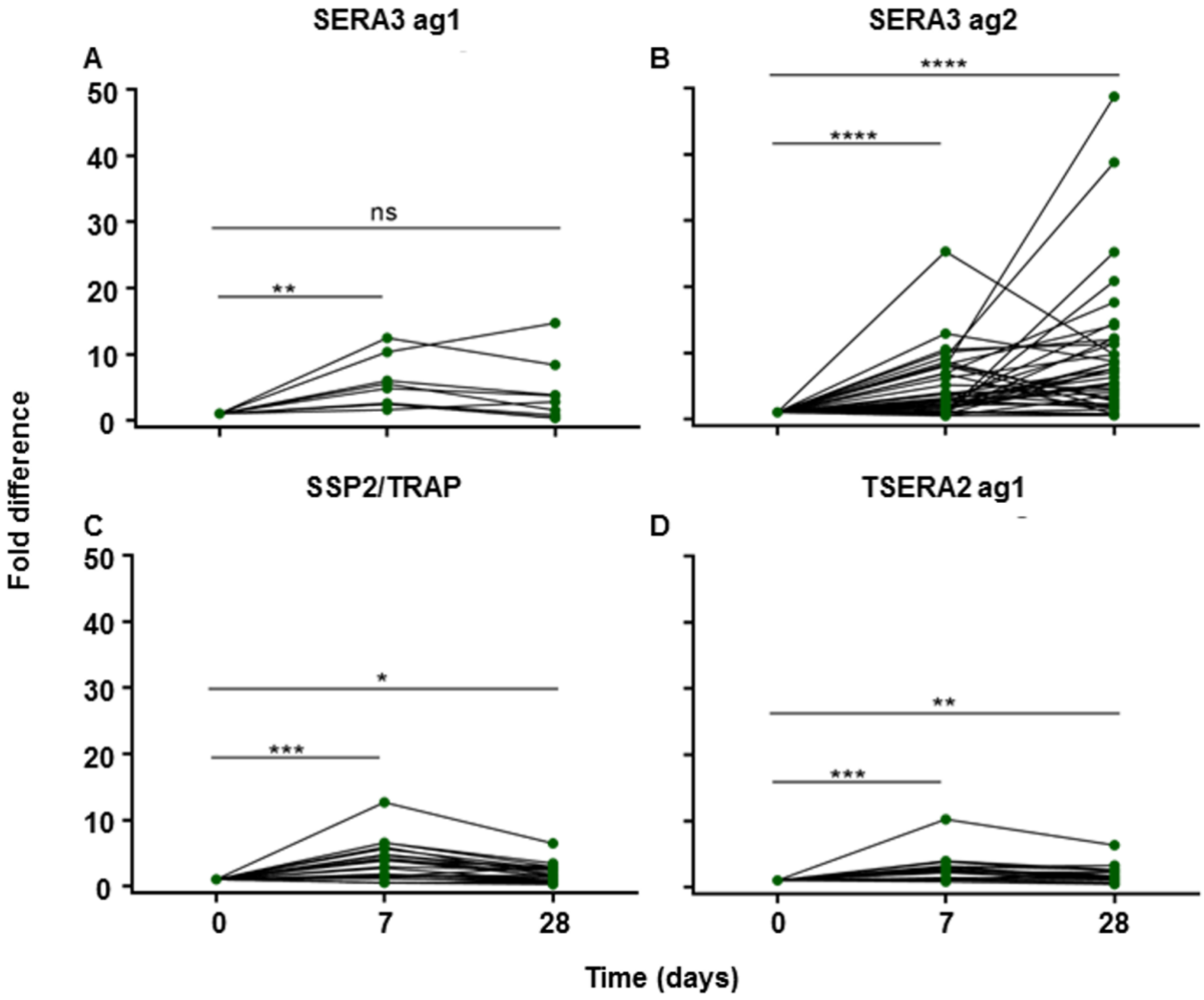


Figure 4.15 Serial fold increase in antibody reactivity for each antigen following treatment of *knowlesi* malaria. (a) SERA3 ag1, (b) SERA3 ag2, (c) SSP2/TRAP and (d) TSERA2 ag1. Asterisks indicate level of significance, ns denotes non-significant values (p0.0001; p0.001; p0.01; p0.05 and p>0.05 ns; Wilcoxon signed rank tests).

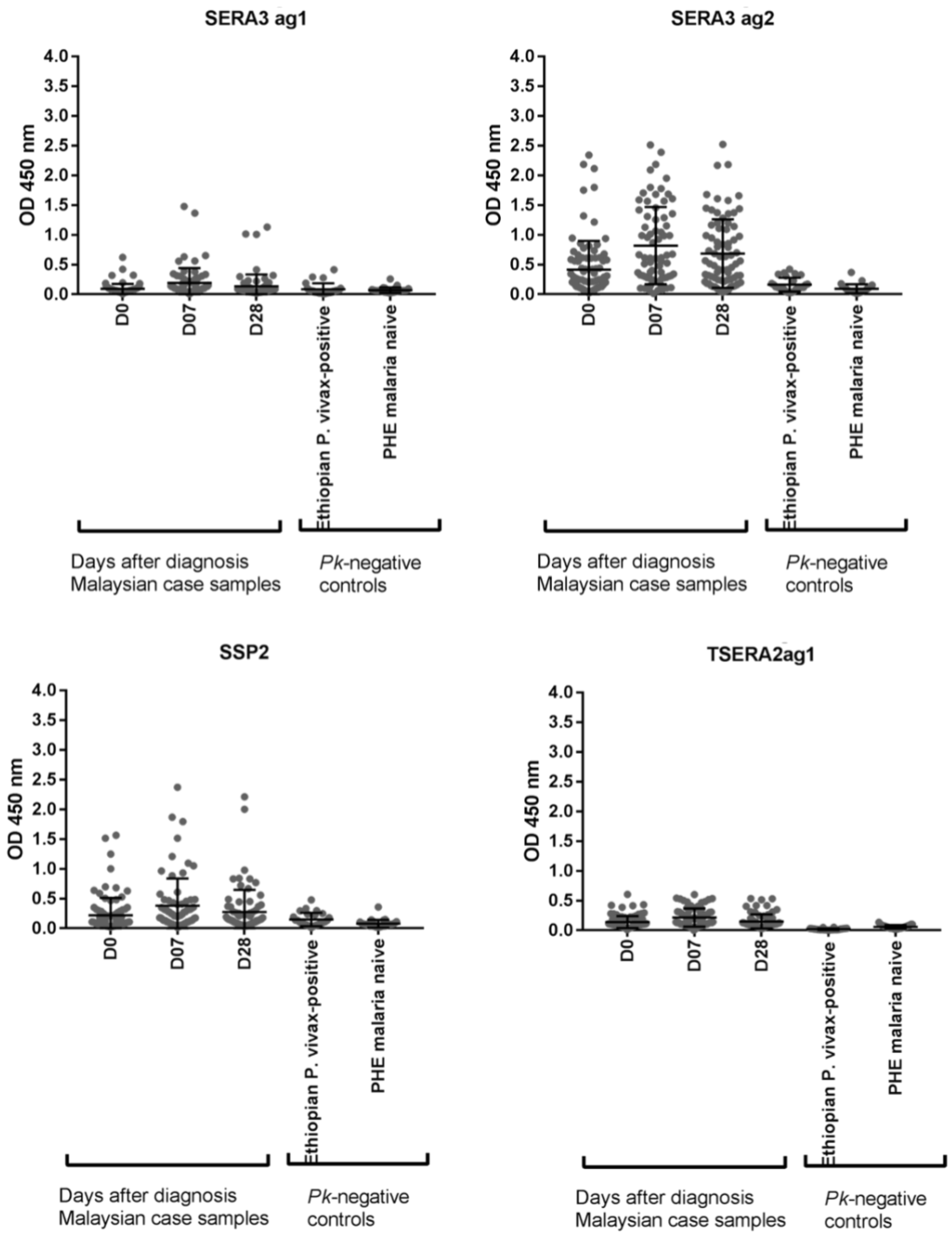


Figure 4.16 *Plasmodium knowlesi* antigen reactivity to Malaysian hospital case serum samples and negative control serum samples. Dot plot of Malaysian hospital case serum samples from days 0 (n = 92), 7 (n = 72) and 28 (n = 77) of PCR diagnosis and *P. knowlesi*-negative control serum samples (Ethiopian *Pv*-positive n = 26; PHE malaria naïve n = 29). Antibody reactivity to the *P. knowlesi*-specific antigens (a) SERA3 ag1, (b) SERA3 ag2, (c) SSP2/TRAP and (d) TSERA2 ag1 are shown.

Compared to controls, the *Pk*SERA3 antigen 2 showed a higher prevalence at all time-points (p<0.001) (***Figure 4.14*** and 4.15 B). Antibody responses measured at day 7 and 28 to *Pk*SERA3 antigen 2 demonstrated a significant increase when compared to day 0 (p<0.001 for both comparisons), with fold changes as high as 50 observed for some samples (***Figure 4.15*** B). In comparison, the fold changes observed in serum responses to the *Pk*TSERA2 antigen 1 (day 7 and 28; p=<0.001 and p=0.005 respectively), *Pk*SERA3 antigen 1 (day 7; p=0.008), and *Pk*SSP2 (day 7 and 28; p=0.001 and p=0.013), although statistically significant, had comparatively lower fold changes with a maximum of 15 (***Figure 4.15*** A, C and D).

4.3.5. IDENTIFICATION OF *P. KNOWLESI* EXPOSED INDIVIDUALS

To assess the predictive ability of responses to these antigens to identify *P. knowlesi* exposed individuals, we used boosted regression tree analysis, an ensemble modelling method combining aspects of machine learning and statistical analysis shown to have strong predictive performance and reliable identification of variable importance[126]. Similar data-adaptive statistical models are increasingly being used for classification and identification of patterns in large datasets and have previously been applied to identify predictive antigens[21]. Although the samples size is small, boosted regression trees have been used for classification with similarly small training data sets[126]. To further compensate for the small dataset, we fitted 100 models of random samples of equal numbers of seropositive and seronegative samples within this training dataset and cross-validated these model predictions. Out of the 100 models fitted for randomly sampled equal numbers of exposed and unexposed individuals, the median classification accuracy was 88.9% (IQR: 86.1-91.3%), calculated as the cross-validated area under the receiver operator curve (AUC). Relative variable importance was calculated for all models. *Pk*SERA3 antigen 2 contributed most to the models (median relative variable importance: 50.4% (IQR 43.3- 61.4%)), followed by *Pk*TSERA2 antigen 1, *Pk*SSP2 and

*Pk*SERA3 antigen 1 (***Figure 4.17***). This part of the data analysis was performed by Kim Fornace.

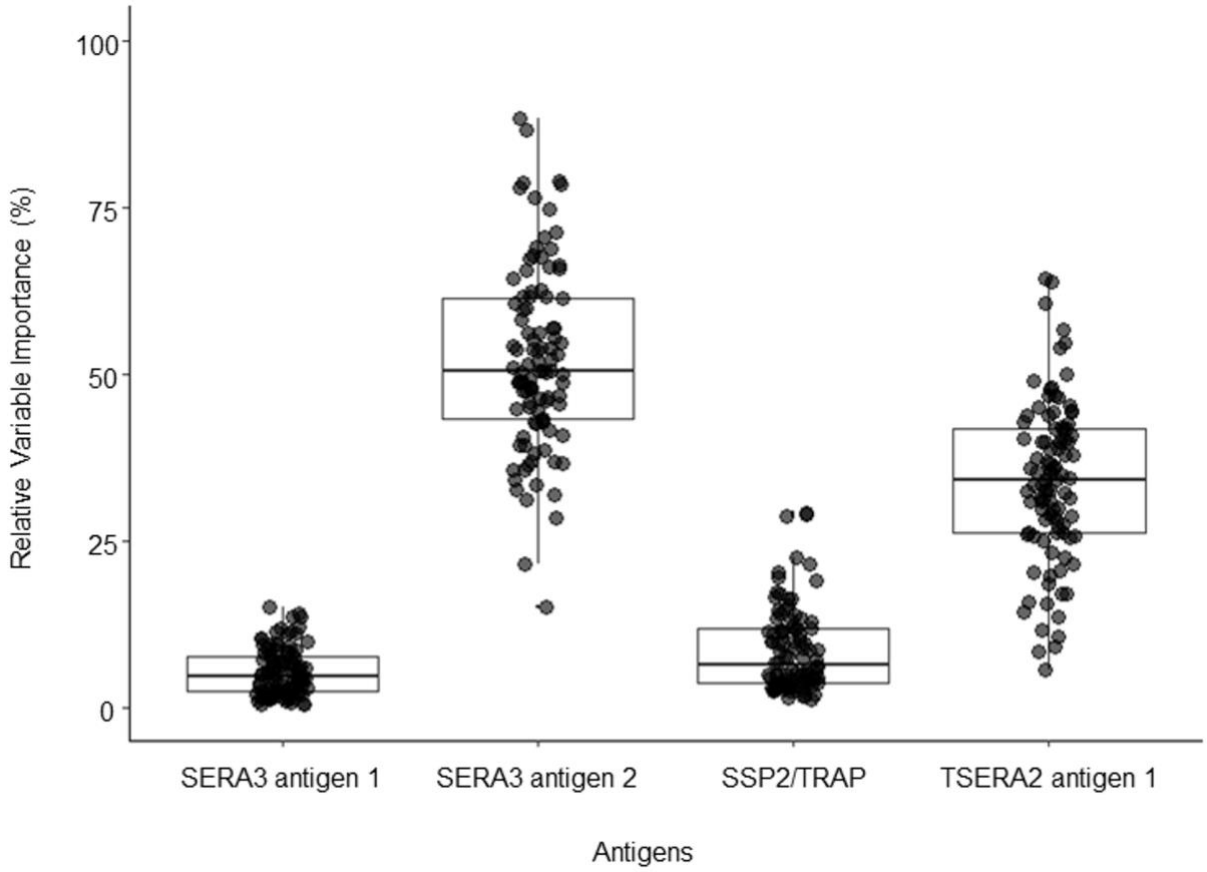


Figure 4.17 Relative variable importance of responses to each antigen from 100 boosted regression tree models predicting *P. knowlesi* seropositivity. Median values for the relative variable importance and interquartile ranges are shown for all antigens tested: *Pk*SERA3 ag 1 (4.8%; IQR 2.5±7.8%); *Pk*SERA3 ag 2 (50.4%; IQR 43.3±61.4%); *Pk*SSP2/TRAP (6.5%; IQR 3.7±11.8%) and TSERA ag 1 (34.2%; IQR 26.2±41.8%). Figure from published source[1].

4.3.6. COMMUNITY SURVEY SAMPLES DEMOGRAPHIC

Surveys and sample collection occurred from October 2014 to January 2015 in Kudat (n=1162) and Pulau Banggi (n=795) in Sabah, Malaysia and in September 2014 in Palawan (n=546), Philippines, by a team led by Kim Fornace. A total of ˜2503 individuals were samples for this study. A similar proportion of men and women were samples with ages ranging between 3 months and 99 years (median age was 24 years) and household engagement in some agricultural activities was reported by the majority of individuals (74%; 1846/2503). Forest coverage around houses (within 1 Km) were 39% in Matunggong, 55% in Bacungan and 82% in Limbuak, with Matunggong being the most fragmented[2].

4.3.7. *P. KNOWLESI* EXPOSURE ASSOCIATION TO AGE AND GENDER IN A SMALL COMMUNITY SURVEY

To determine exposure to *P. knowlesi* in the sampled regions, *Pk*SERA3 antigen 2 was selected, alongside *P. vivax* and *P. falciparum* AMA-1 and MSP1-19, as it was previously shown to be the *P. knowlesi* antigen showing most reactivity to clinical blood samples from exposed individuals[1]. Only one *P. knowlesi* antigen was selected at this occasion due to the limited volume of blood samples, limiting the ELISA method to only testing 5 antigens.

Exposure to *P. knowlesi* varied between the regions studied, with the highest antibody prevalence of 11.7% (93/795) detected in Pulau Banggi, Malaysia, followed by 6.8% (79/1162) in Kudat, Malaysia, and finally 1.1% (6/546) in Palawan, The Philippines, the region with the lowest seroprevalence from this study (**Figure 4.18**). For *P. falciparum* and *P. vivax* antigens, exposure was higher than for *P. knowlesi* in all study sites, with Matunggong, Kudat, at 16.9% (196/1162) and 6.9% (80/1162), Limbuak, Pulau Banggi, at 13.5% (107/795) and 16.7% (133/795), and Bacungan, Palawan, at 10.4% (61/587) and 9.7% (57/587), respectively.

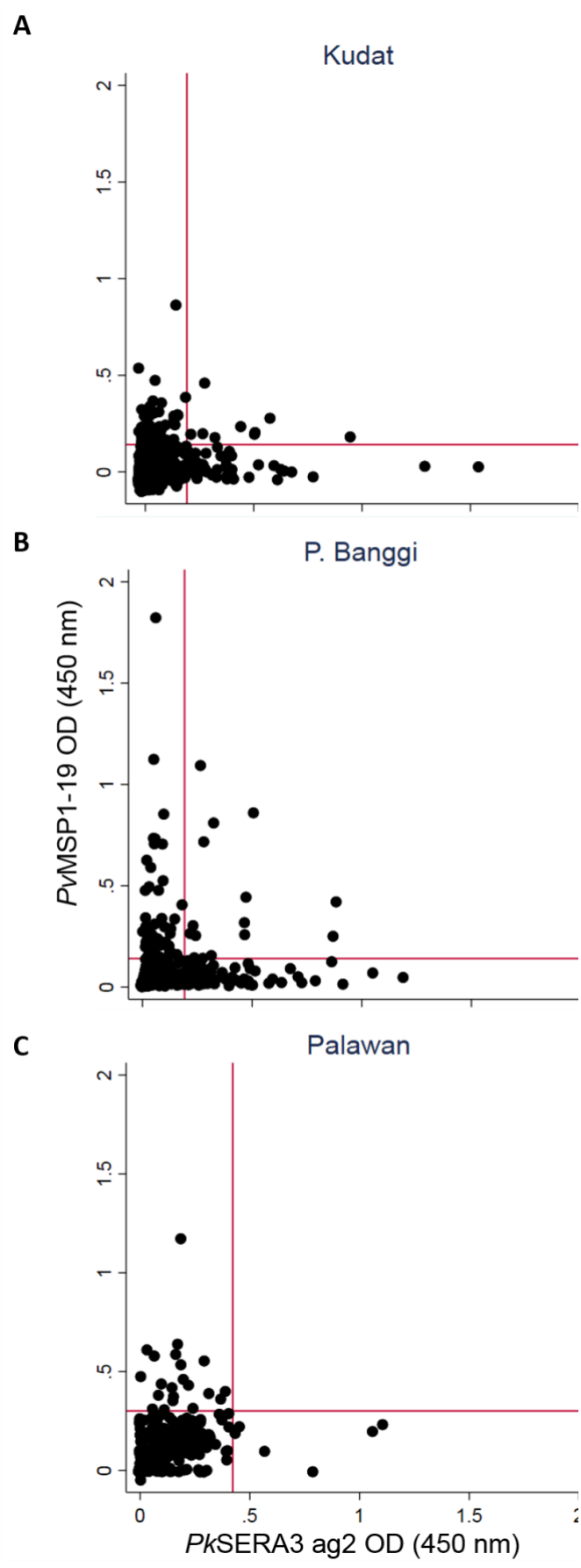


Figure 4.18 Sera reactivity to *Pk*SERA3 antigen 2 in endemic populations of Malaysia and The Philippines. Scatter plots showing sera reactivity to P. knowlesi SERA3 ag2 in Kudat (A), Pulau Banggi (B) and Palawan (C). The red lines in each graph represent the cut off values for the P. knowlesi antigen, which was calculated based on Public Health England negative control sera samples (average ODs ± (5xSD)) and *Pv*MSP1-19 (average ODs: ± (3xSD)): The vertical cut off line for Malaysia (Kudat (A) and P. Banggi (B)) is based on *Pk*SERA3 ag2=0.192 and the horizontal cut off line is based on *Pv*MSP1-19=0.141. The vertical cut off line for The Philippines (Palawan (C)) is based on *Pk*SERA3 ag2=0.422 and the horizontal cut off line is based on *Pv*MSP1-19=0.301.

From all samples surveyed, 7.1% (178/2503) of the total population were seropositive for *P. knowlesi,* 16.1% (364/2266) for both *P. falciparum* antigens and 12.6% (270/2141) for both *P. vivax* antigens. Results for all antigens were not available for all individuals due to non-systematic errors in sample labelling and insufficient sample volume.

There was no significant difference observed in reactivity to *P. knowlesi* between men and women (***Figure 4.19***), and there was no significant association between gender and seropositivity to *P. knowlesi*, with a similar proportion of men and women seropositive to *P. knowlesi* (7.1% (88/178); 7.2% (90/178) respectively; OR (odds ratio): 0.99, 95% CI: 0.71±1.37, p=0.95). For all sites, antibody reactivity to *Pk*SERA3 antigen 2 was detected in 4.2% (39/921) of individuals under 15 years of age and in 9.4% (25/265) of individuals over 60 years (***Figure 4.20***), though the highest reactivity was found in individuals between 45-60 years (11.6%; 43/370). For *P. falciparum* and *P. vivax* antigens, antibody reactivity was detected in 3.5% (29/821) and 2.9% (23/792) of individuals under 15 years and in 32.9% (78/237) and 28.1% (64/228) of individuals over 60 years, respectively. For all antigens tested, seropositivity was associated with an increase in age, despite seroreactivity being detected in the youngest age groups. Agricultural work was found to be positively associated with *P. knowlesi* seropositivity as well as age and region. Of individuals under 15 years, the youngest were most likely to react to *P. knowlesi* than to the non-zoonotic malaria species.

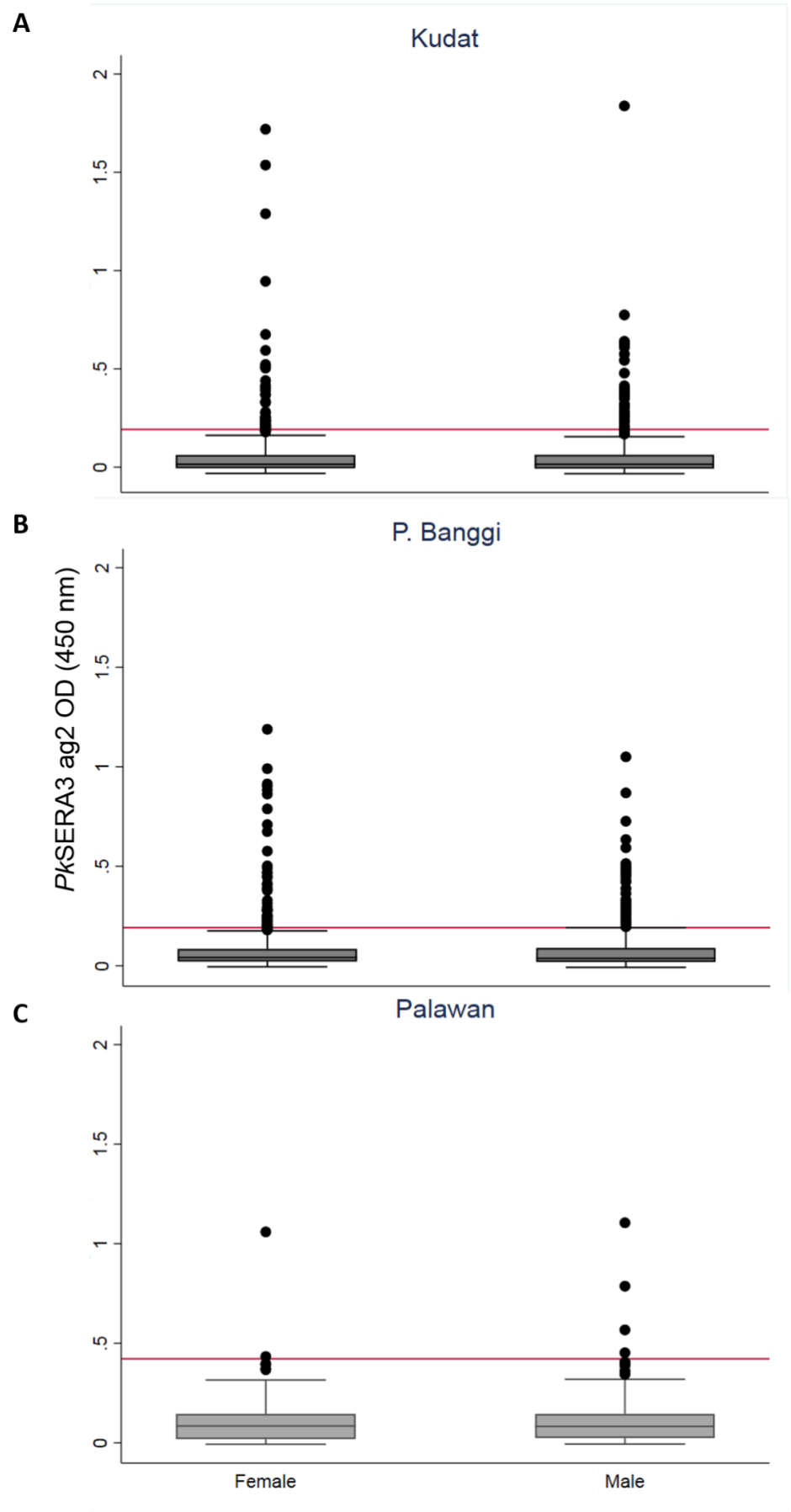


Figure 4.19 Sera reactivity in men and women to *Pk*SERA3 antigen 2 in endemic populations of Malaysia and The Philippines. Plots showing sera reactivity in men and women to *P. knowlesi* SERA3 ag2 in Kudat (A), Pulau Banggi (B) and Palawan (C). The horizontal cut off line for Kudat (A) and P. Banggi (B) is based PkSERA3 ag2=0.192. The horizontal cut off line for Palawan (C) is based on PkSERA3 ag2=0.422.

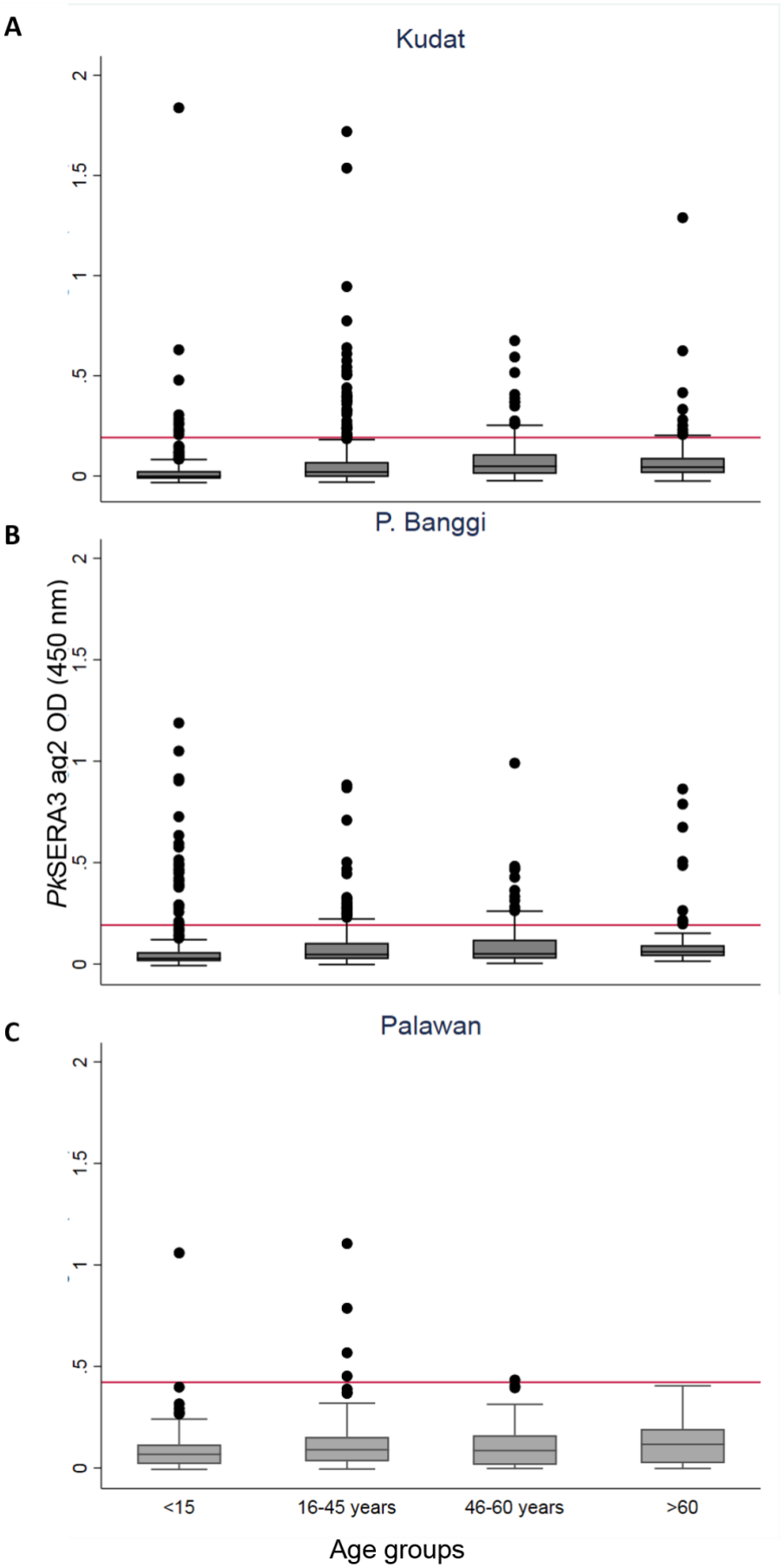


Figure 4.20 Sera reactivity by age groups to *Pk*SERA3 antigen 2 in endemic populations of Malaysia and The Philippines. Plots showing sera reactivity to *P. knowlesi* SERA3 ag2 in Kudat (A), Pulau Banggi (B) and Palawan (C). The horizontal cut off line for Kudat (A) and P. Banggi (B) is based PkSERA3 ag2=0.192. The horizontal cut off line for Palawan (C) is based on PkSERA3 ag2=0.422.

4.4. DISCUSSION

P. knowlesi is a naturally occurring infection of long-tailed and pig-tailed macaques, historically associated with forested areas of Southeast Asia[8]. Increased deforestation of their natural habitat is thought to have led to increased interaction between macaques and the human population in endemic areas[144]. Changes in village level forest cover and historical forest loss has been associated with an increase in *P. knowlesi* clinical cases in Sabah[30], with malaria caused by *P. knowlesi* increasingly reported in Southeast Asia[12]. Conversely, there has also been a steady decline in the prevalence of *P. falciparum* and *P. vivax* infections in the same region[145].

The recent efforts of the malaria community towards achieving malaria elimination means that tools to help monitor the impact and effectiveness of intervention strategies are an urgent requirement[146]. The development of species-specific tools for *P. knowlesi* would allow accurate assessment of the levels and geographical limits of infection with this zoonotic species[147]. There is an urgent need to develop a comprehensive discovery strategy to help identify *P. knowlesi* unique antigenic markers of exposure in order to further characterise this organism and develop stronger and better identification methods.

Currently, there are no specifically designed biomarkers for the serosurveillance of *P. knowlesi* infections. Recombinant proteins are available [*Pk*CSP[148], *Pk*AMA1[149], *Pk*DBP[150], *Pk*SPATR[58], *Pk*LDH[151], *Pk*1-Cys peroxiredoxin[152], *Pk* knowpains[153], *Pk*MSP1-42[99], *Pk*MSP1-33[100], PkMSP1-19[101], *Pk* tryptophan-rich antigens (PkTrags)[154], *Pk*MSP3[155] and *Pk*SBP1[156], but are limited in number and are generally not species-specific. As a result, their utility as serological diagnostic tools is generally secondary to their original design. The reported level of amino acid sequence conservation to other *Plasmodium* spp. in some currently available *P. knowlesi* proteins is > 60% across large stretches of continuous sequence. Such

reagents could not be specific to *P. knowlesi*[99,102,157,158] and would be unable to reliably discriminate between antibody responses to different parasite species in co-endemic settings.

We have successfully developed a panel of 10 *P. knowlesi*-specific recombinant antigens using an *in silico* approach. Although this project was successful, the development pipeline of this panel of antigens suffered challenges at every major step of the process. Recombinant protein development and expression requires a lot of time and effort. There were difficulties in the candidate amplification step, with some candidate sequences having to be altered slightly in order to amplify, others requiring primer design to be reviewed and others simply being excluded from the panel entirely. Three different cloning methods were applied in order to maximise the cloning output. This decision was taken due to the inefficiency of the SLIC method. Multiple troubleshooting attempts were made including changing vector to insert ratios, changes in individual reagent concentrations, using new batches of reagents and cells as well as changing incubation periods. At the time it was not possible to determine the reasons for the failed cloning attempts, therefore alternative cloning methods were sought after, the Gibson and the conventional methods. With more time, we would have liked to determine the causes of the inefficient cloning using the SLIC method and better compare this method to the others used in terms of number of transformants obtained, number of successfully cloned transformants and value for money in reagents per method. Another limitation was found during the antigen expression phase, with some candidates failing to be expressed and others expressing very little. By extending the expression time of some of the candidates it was possible to overcome this difficulty but this is a factor that would need to be reassessed more closely for individual candidates. Determining the optimal expression conditions for individual antigen candidates would have required more time and resources than was available. In future, it would be beneficial to determine

what conditions to express individual antigen candidates or groups of candidates in order to obtain substantial amounts with the least protein degradation possible.
High levels of amino acid identity (83%) between *Pv*MSP1-19 and *Pk*MSP1-19, meant we were unable to use these reagents to dissect the species-specific immune responses due to the inevitable cross-reactive antibody responses. This is consistent with a proportion (48.9% (45/92)) of the confirmed *P. knowlesi*-exposed clinical samples in this study reacting with *Pv*MSP1-19 at day 0, although it is unknown whether these participants had previously been exposed to *P. vivax*. However, this limitation simply reflects the paucity of available serological reagents for use in assessing exposure to infection, a deficit this study aims to address. The data presented here not only show us that we have successfully developed *P. knowlesi* tools to determine exposure in endemic populations, but also that specific tools for *P. vivax* and *P. falciparum* need to be developed for population studies where these three are co-endemic. One of the imitations of the assays using clinical and community samples was the lack of appropriate negative control serum samples, ideally endemic individuals that had never been infected with any of the malaria species. The lack of species-specific control antigens for *P. falciparum* and *P. vivax* was also a limitation in this study, as this would have proved helpful in determining the specificity of the *P. knowlesi* reagents as well as determining whether the individuals tested had prior exposure to these other species or not. Although the small number of clinical case samples do not give sufficient statistical power to assess either the duration of antibody responses to the panel of antigens or population-level exposure, the *P. knowlesi* clinical case samples represent a unique dataset with which to validate the immunogenicity of our antigen panel.
The use of the boosted regression tree model was able to discriminate between *P. knowlesi* exposed and unexposed individuals for the purposes of classification of seropositivity rather than to assess individual-level risk factors. While this clinical dataset is sufficient for classification as exposed or unexposed, it is not sufficiently

large enough to stratify by age, gender or previously reported malaria status. In order for us to assess these types of risk factors, we first need to apply an approach (using known negatives, mixture or probability models) to classify antibody responses as seropositive or seronegative and then assess risk factors within the population, which we discuss later on.

The panel of reagents developed for this study focussed on immunologically relevant orthologous targets previously described in *P. falciparum*. The serine repeat antigen (SERA) family had previously attracted attention as a source of both drug and vaccine candidates[74]. In *P. falciparum*, SERA 5 is the most abundant parasitophorous vacuole protein and is essential to blood stage growth of the parasite[80], with antibodies against this antigen thought to inhibit parasite growth[81]. Although possessing a papain-like enzymatic domain, recent evidence suggests that the protein plays a non-enzymatic role[80]. SERA 3 has also been shown to be a highly immunogenic antigen with an important, although not essential role in the erythrocytic cycle[84] and has also been implicated as having a role in liver stage merozoite release in *P. berghei*[85]. Similarly, evidence for the sporozoite surface protein 2 (SSP2) suggested an immunogenic antigen involved in protection from disease in mice[73]. Although we were unable to confirm active transcription of SSP2 due to the lack of available material, we were able to validate active transcription of both the SERA3 and TSERA2 candidate genes. Collectively, the evidence provided by studies on *Plasmodium* supports the design of seroepidemiology tools based on these targets. Despite the targeted approach used in designing the recombinant constructs, the SERA3 antigen 2 construct was by far the most promising candidate. The differences in the performances of the antigens could be due to a number of factors: (1) variation in the inherent immunogenicity of the regions selected, (2) variations in the expression status of the *P. knowlesi* antigens compared to *P. falciparum* or (3) the loss of immunoreactive epitopes due to the truncation of the protein and potential loss of conformational epitopes.

There are a number of potential limitations of the serology study. The small sample size of the clinical samples used prevented detailed analysis of the samples, such as monitoring the impact of factors such as age, on the profile of reactivity to the reagents under test. In addition, the lack of repeated samples per individual (i.e. longitudinal samples) prevented us from investigating the longevity of antibody responses to each target, across individuals and age groups. The availability of supporting biological information on *P. knowlesi*, such as functional data, transcriptional or RNA seq data would have helped with the rational selection of additional candidates for further study and the design recombinant tools. This is the first study to describe the development a panel of *P. knowlesi*-specific serological tools using freely available *in silico* software. We have demonstrated the importance of targeting species-specific reagents at the amino acid level and highlighted the potential of such proteins as serosurveillance tools. Not only have we developed a novel panel of *P. knowlesi* species specific recombinant antigens, but we have been able to measure specific immune responses to these reagents and described the change in antibody profile following treatment.

We tried optimising the Biomek robotic arm for the ELISA protocol as a high throughput serological platform to overcome the lack of reagents, as it would use the least reagent possible and test a higher number of samples per assay. Samples previously tested on a standard ELISA were used and rerun with the same antigens using the Biomek at least twice. We did not proceed to using this platform as an alternative as at first the results did not seem to correlate between platforms for the same reagents. The first and second runs did not seem to correlate very well, particularly for *Pf*AMA-1. It is possible that there was a problem during run 1 using the Biomek where samples showed poor reactivity to the *P. falciparum* antigens and duplicates did not correlate. Maybe the samples, the secondary antibody or the TMB substrate were not properly mixed by the robotic pipetting arm, or the washing steps were not properly performed, potentially due to a blockage in the pipetting arm. More

runs would need to be done to determine if it was human error or a true difference between runs generally. More runs would need to be compared using different samples to determine if it was an issue with just that run or if there genuinely are different results for every re-run of the same samples and reagents. The preparation of reagents into the format required proved to be more time consuming and the lack of expertise with the software reduced the possibility of further optimising the platform for our use. Instead, we focused on using the standard ELISA protocol instead with the selected *P. knowlesi* antigens.

The expanded panel of *P. knowlesi* antigens was developed and successfully expressed but were not run on the ELISA platform. Further analysis of these antigens is found in the next chapters, where we employ high throughput serology approaches such as the protein microarray to help identify additional important targets of immunity[21,159].

As such, we have already demonstrated the utility of the SERA3 antigen 2 reagents as a potential seroepidemiological tool and further validation of the SERA3 antigen 2 at the population level has been performed. Based on the clinical samples results, the *Pk*SERA 3 antigen 2 recombinant was used to survey ~2500 samples across three site; Limbuak, Pulau Banggi and Matunggung, Kudat, Sabah, Malaysia and Bacungan, Palawan, the Philippines[2]. One of the key elements from this study using this reagent was the indication of community level patterns of exposure, with higher levels of exposure among individuals over 60 years of age[2]. *P. knowlesi* exposure was identified in children under 5 in all sites, which suggests transmission was recent or ongoing, although at a low level, and it was found to increase with age. Clinical *P. knowlesi* is commonly reported in adult men[10,160], however, it was shown that men and women had similar antibody reactivity to *P. knowlesi* antigens in all sites. The main limitations of this study are the non-randomised population sampling approach and limited geographical scale. The results presented here highlight the utility of serological techniques to identify differences in transmission intensity in

settings where the sensitivity of parasite prevalence surveys is limited by the scarcity of infected individuals and suboptimal diagnostics.

REFERENCES

1 Herman, L. S. *et al.* Identification and validation of a novel panel of Plasmodium knowlesi biomarkers of serological exposure. *PLoS neglected tropical diseases* **12**, e0006457, doi:10.1371/journal.pntd.0006457 (2018).

2 Fornace, K. M. *et al.* Exposure and infection to Plasmodium knowlesi in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. *PLoS neglected tropical diseases* **12**, e0006432, doi:10.1371/journal.pntd.0006432 (2018).

3 Cox-Singh, J. *et al.* Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **46**, 165-171, doi:10.1086/524888 (2008).

4 Barber, B. E. *et al.* Plasmodium knowlesi malaria in children. *Emerging infectious diseases* **17**, 814-820, doi:10.3201/eid1705.101489 (2011).

5 Daneshvar, C. *et al.* Clinical and laboratory features of human Plasmodium knowlesi infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **49**, 852-860, doi:10.1086/605439 (2009).

6 Organization, W. H. Expert consultation on Plasmodium knowlesi malaria to guide malaria elimination strategies. (2017).

7 Garnham, P. C. C. *Malaria parasites and other haemosporidia*. (Blackwell Scientific, 1966).

8 Coatney, G. R., Collins, W. E. & Contacos, P. G. *The Primate malarias*. (U.S. National Institute of Allergy and Infectious Diseases, 1971).

9 Singh, B. *et al.* A large focus of naturally acquired Plasmodium knowlesi infections in human beings. *Lancet* **363**, 1017-1024, doi:10.1016/s0140-6736(04)15836-4 (2004).

10 Barber, B. E. *et al.* A prospective comparative study of knowlesi, falciparum, and vivax malaria in Sabah, Malaysia: high proportion with severe disease from Plasmodium knowlesi and Plasmodium vivax but no mortality with early referral and artesunate therapy. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **56**, 383-397, doi:10.1093/cid/cis902 (2013).

11 Barber, B. E., William, T., Grigg, M. J., Yeo, T. W. & Anstey, N. M. Limitations of microscopy to differentiate Plasmodium species in a region co-endemic for Plasmodium falciparum, Plasmodium vivax and Plasmodium knowlesi. *Malaria journal* **12**, 8, doi:10.1186/1475-2875-12-8 (2013).

12 Singh, B. & Daneshvar, C. Human infections and detection of Plasmodium knowlesi. *Clin Microbiol Rev* **26**, 165-184, doi:10.1128/cmr.00079-12 (2013).

13 William, T. *et al.* Severe Plasmodium knowlesi malaria in a tertiary care hospital, Sabah, Malaysia. *Emerging infectious diseases* **17**, 1248-1255, doi:10.3201/eid1707.101017 (2011).

14 Rajahram, G. S. *et al.* Deaths due to Plasmodium knowlesi malaria in Sabah, Malaysia: association with reporting as Plasmodium malariae and delayed parenteral artesunate. *Malaria journal* **11**, 284, doi:10.1186/1475-2875-11-284 (2012).

15 Wipasa, J. *et al.* Long-lived antibody and B Cell memory responses to the human malaria parasites, Plasmodium falciparum and Plasmodium vivax. *PLoS Pathog* **6**, e1000770, doi:10.1371/journal.ppat.1000770 (2010).

16 Akpogheneta, O. J. *et al.* Duration of naturally acquired antibody responses to blood-stage Plasmodium falciparum is age dependent and antigen specific. *Infection and immunity* **76**, 1748-1755, doi:10.1128/iai.01333-07 (2008).

17 Fowkes, F. J. *et al.* New insights into acquisition, boosting, and longevity of immunity to malaria in pregnant women. *The Journal of infectious diseases* **206**, 1612-1621, doi:10.1093/infdis/jis566 (2012).
18 Meraldi, V. *et al.* Natural antibody response to Plasmodium falciparum Exp-1, MSP-3 and GLURP long synthetic peptides and association with protection. *Parasite immunology* **26**, 265-272, doi:10.1111/j.0141-9838.2004.00705.x (2004).
19 Osier, F. H. *et al.* Breadth and magnitude of antibody responses to multiple Plasmodium falciparum merozoite antigens are associated with protection from clinical malaria. *Infection and immunity* **76**, 2240-2248, doi:10.1128/iai.01585-07 (2008).
20 Stanisic, D. I. *et al.* Acquisition of antibodies against Plasmodium falciparum merozoites and malaria immunity in young children and the influence of age, force of infection, and magnitude of response. *Infection and immunity* **83**, 646-660, doi:10.1128/iai.02398-14 (2015).
21 Helb, D. A. *et al.* Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E4438-4447, doi:10.1073/pnas.1501705112 (2015).
22 Bruce-Chwatt, L. J., Draper, C. C., Dodge, J. S., Topley, E. & Voller, A. Originally published as Volume 1, Issue 7749 SERO-EPIDEMIOLOGICAL STUDIES ON POPULATION GROUPS PREVIOUSLY EXPOSED TO MALARIA. *The Lancet* **299**, 512-515, doi:http://dx.doi.org/10.1016/S0140-6736(72)90176-6 (1972).
23 Corran, P., Coleman, P., Riley, E. & Drakeley, C. Serology: a robust indicator of malaria transmission intensity? *Trends in Parasitology* **23**, 575-582, doi:http://dx.doi.org/10.1016/j.pt.2007.08.023 (2007).
24 Drakeley, C. & Cook, J. in *Advances in parasitology* Vol. 69 299-352 (Academic Press, 2009).
25 Medina Costa, R., de Sousa, K. P., Atouguia, J., Tavira, L. T. & Silva, M. S. Prevalence and Level of Antibodies Anti-Plasmodium spp. in Travellers with Clinical History of Imported Malaria. *Journal of Parasitology Research* **2013**, 247273, doi:10.1155/2013/247273 (2013).
26 Ramasamy, R., Nagendran, K. & Ramasamy, M. S. Antibodies to epitopes on merozoite and sporozoite surface antigens as serologic markers of malaria transmission: studies at a site in the dry zone of Sri Lanka. *Am J Trop Med Hyg* **50**, 537-547 (1994).
27 Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526**, 207-211, doi:10.1038/nature15535 (2015).
28 Dewasurendra, R. L. *et al.* Effectiveness of a serological tool to predict malaria transmission intensity in an elimination setting. *BMC Infect Dis* **17**, 49, doi:10.1186/s12879-016-2164-0 (2017).
29 William, T. *et al.* Increasing incidence of Plasmodium knowlesi malaria following control of P. falciparum and P. vivax Malaria in Sabah, Malaysia. *PLoS neglected tropical diseases* **7**, e2026, doi:10.1371/journal.pntd.0002026 (2013).
30 Fornace, K. M. *et al.* Association between Landscape Factors and Spatial Patterns of Plasmodium knowlesi Infections in Sabah, Malaysia. *Emerging infectious diseases* **22**, 201-209, doi:10.3201/eid2202.150656 (2016).
31 Fornace, K. M. *et al.* Asymptomatic and submicroscopic carriage of Plasmodium knowlesi malaria in household and community members of clinical cases in Sabah, Malaysia. *The Journal of infectious diseases*, doi:10.1093/infdis/jiv475 (2015).
32 Marchand, R. P., Culleton, R., Maeno, Y., Quang, N. T. & Nakazawa, S. Co-infections of Plasmodium knowlesi, P. falciparum, and P. vivax among Humans and Anopheles

dirus Mosquitoes, Southern Vietnam. *Emerging infectious diseases* **17**, 1232-1239, doi:10.3201/eid1707.101551 (2011).

33 Van den Eede, P. *et al.* Human Plasmodium knowlesi infections in young children in central Vietnam. *Malaria journal* **8**, 249, doi:10.1186/1475-2875-8-249 (2009).

34 Hay, S. I., Smith, D. L. & Snow, R. W. Measuring malaria endemicity from intense to interrupted transmission. *The Lancet infectious diseases* **8**, 369-378, doi:10.1016/s1473-3099(08)70069-0 (2008).

35 Beier, J. C., Killeen, G. F. & Githure, J. I. Short report: entomologic inoculation rates and Plasmodium falciparum malaria prevalence in Africa. *Am J Trop Med Hyg* **61**, 109-113 (1999).

36 Oesterholt, M. J. *et al.* Spatial and temporal variation in malaria transmission in a low endemicity area in northern Tanzania. *Malaria journal* **5**, 98, doi:10.1186/1475-2875-5-98 (2006).

37 Bousema, T. *et al.* Serologic markers for detecting malaria in areas of low endemicity, Somalia, 2008. *Emerging infectious diseases* **16**, 392-399, doi:10.3201/eid1603.090732 (2010).

38 Drakeley, C. J. *et al.* Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5108-5113, doi:10.1073/pnas.0408725102 (2005).

39 van den Hoogen, L. L. *et al.* Serology describes a profile of declining malaria transmission in Farafenni, The Gambia. *Malaria journal* **14**, 416, doi:10.1186/s12936-015-0939-1 (2015).

40 Oduro, A. R. *et al.* Seroepidemiological and parasitological evaluation of the heterogeneity of malaria infection in the Gambia. *Malaria journal* **12**, 222, doi:10.1186/1475-2875-12-222 (2013).

41 Arnold, B. F. *et al.* Serological measures of malaria transmission in Haiti: comparison of longitudinal and cross-sectional methods. *PLoS One* **9**, e93684, doi:10.1371/journal.pone.0093684 (2014).

42 Corran, P. H. *et al.* Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. *Malaria journal* **7**, 195, doi:10.1186/1475-2875-7-195 (2008).

43 Voller, A. & Bruce-Chwatt, L. J. Serological malaria surveys in Nigeria. *Bulletin of the World Health Organization* **39**, 883-897 (1968).

44 Cook, J. *et al.* Serological Markers Suggest Heterogeneity of Effectiveness of Malaria Control Interventions on Bioko Island, Equatorial Guinea. *PLOS ONE* **6**, e25137, doi:10.1371/journal.pone.0025137 (2011).

45 Drakeley, C. J. *et al.* Altitude-dependent and -independent variations in Plasmodium falciparum prevalence in northeastern Tanzania. *The Journal of infectious diseases* **191**, 1589-1598, doi:10.1086/429669 (2005).

46 Ghinai, I. *et al.* Malaria epidemiology in central Myanmar: identification of a multi-species asymptomatic reservoir of infection. *Malaria journal* **16**, 16, doi:10.1186/s12936-016-1651-5 (2017).

47 Waters, A. P. *et al.* A merozoite receptor protein from Plasmodium knowlesi is highly conserved and distributed throughout Plasmodium. *J Biol Chem* **265**, 17974-17979 (1990).

48 Yang, A. S. *et al.* AMA1 and MAEBL are important for Plasmodium falciparum sporozoite infection of the liver. *Cell Microbiol*, doi:10.1111/cmi.12745 (2017).

49 Peterson, M. G. *et al.* Integral membrane protein located in the apical complex of Plasmodium falciparum. *Molecular and Cellular Biology* **9**, 3151-3154 (1989).

50 Thomas, A. W. *et al.* Aspects of immunity for the AMA-1 family of molecules in humans and non-human primates malarias. *Memorias do Instituto Oswaldo Cruz* **89 Suppl 2**, 67-70 (1994).

51 Laurens, M. B. *et al.* Strain-specific Plasmodium falciparum growth inhibition among Malian children immunized with a blood-stage malaria vaccine. *PLoS One* **12**, e0173294, doi:10.1371/journal.pone.0173294 (2017).

52 Ssewanyana, I. *et al.* Avidity of anti-malarial antibodies inversely related to transmission intensity at three sites in Uganda. *Malaria journal* **16**, 67, doi:10.1186/s12936-017-1721-3 (2017).

53 Deans, J. A. *et al.* Vaccination trials in rhesus monkeys with a minor, invariant, Plasmodium knowlesi 66 kD merozoite antigen. *Parasite immunology* **10**, 535-552 (1988).

54 Tadesse, F. G. *et al.* The shape of the iceberg: quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia. *Malaria journal* **16**, 99, doi:10.1186/s12936-017-1749-4 (2017).

55 Gardner, M. J. *et al.* Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum. *Science* **282**, 1126-1132 (1998).

56 Rogers, W. O. *et al.* Characterization of Plasmodium falciparum sporozoite surface protein 2. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 9176-9180 (1992).

57 Chattopadhyay, R. *et al.* PfSPATR, a Plasmodium falciparum protein containing an altered thrombospondin type I repeat domain is expressed at several stages of the parasite life cycle and is the target of inhibitory antibodies. *J Biol Chem* **278**, 25977-25981, doi:10.1074/jbc.M300865200 (2003).

58 Mahajan, B. *et al.* Identification, cloning, expression, and characterization of the gene for Plasmodium knowlesi surface protein containing an altered thrombospondin repeat domain. *Infection and immunity* **73**, 5402-5409, doi:10.1128/iai.73.9.5402-5409.2005 (2005).

59 Holder, A. & Freeman, R. R. Biosynthesis and processing of a Plasmodium falciparum schizont antigen recognized by immune serum and a monoclonal antibody. *The Journal of experimental medicine* **156**, 1528-1538 (1982).

60 Moss, D. K. *et al.* Plasmodium falciparum 19-kilodalton merozoite surface protein 1 (MSP1)-specific antibodies that interfere with parasite growth in vitro can inhibit MSP1 processing, merozoite invasion, and intracellular parasite development. *Infection and immunity* **80**, 1280-1287, doi:10.1128/iai.05887-11 (2012).

61 Bisseye, C. *et al.* An engineered Plasmodium falciparum C-terminal 19-kilodalton merozoite surface protein 1 vaccine candidate induces high levels of interferon-gamma production associated with cellular immune responses to specific peptide sequences in Gambian adults naturally exposed to malaria. *Clin Exp Immunol* **166**, 366-373, doi:10.1111/j.1365-2249.2011.04467.x (2011).

62 Pusic, K., Clements, D., Kobuch, S. & Hui, G. Antibody and T cell responses in reciprocal prime-boost studies with full-length and truncated merozoite surface protein 1-42 vaccines. *PLoS One* **8**, e75939, doi:10.1371/journal.pone.0075939 (2013).

63 Ogutu, B. R. *et al.* Blood Stage Malaria Vaccine Eliciting High Antigen-Specific Antibody Concentrations Confers No Protection to Young Children in Western Kenya. *PLOS ONE* **4**, e4708, doi:10.1371/journal.pone.0004708 (2009).

64 Arama, C. & Troye-Blomberg, M. The path of malaria vaccine development: challenges and perspectives. *Journal of internal medicine* **275**, 456-466, doi:10.1111/joim.12223 (2014).

65 Guevara Patino, J. A., Holder, A. A., McBride, J. S. & Blackman, M. J. Antibodies that inhibit malaria merozoite surface protein-1 processing and erythrocyte invasion are blocked by naturally acquired human antibodies. *The Journal of experimental medicine* **186**, 1689-1699 (1997).

66 O'Donnell, R. A. *et al.* Antibodies against Merozoite Surface Protein (Msp)-1(19) Are a Major Component of the Invasion-Inhibitory Response in Individuals Immune to Malaria. *The Journal of experimental medicine* **193**, 1403-1412 (2001).

67 Soares, I. S., Levitus, G., Souza, J. M., Del Portillo, H. A. & Rodrigues, M. M. Acquired immune responses to the N- and C-terminal regions of Plasmodium vivax merozoite surface protein 1 in individuals exposed to malaria. *Infection and immunity* **65**, 1606-1614 (1997).

68 Blackman, M. J., Heidrich, H. G., Donachie, S., McBride, J. S. & Holder, A. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion- inhibiting antibodies. *The Journal of experimental medicine* **172**, 379-382 (1990).

69 Templeton, T. J. & Kaslow, D. C. Cloning and cross-species comparison of the thrombospondin-related anonymous protein (TRAP) gene from Plasmodium knowlesi, Plasmodium vivax and Plasmodium gallinaceum. *Molecular and biochemical parasitology* **84**, 13-24 (1997).

70 Muller, H. M. *et al.* Thrombospondin related anonymous protein (TRAP) of Plasmodium falciparum binds specifically to sulfated glycoconjugates and to HepG2 hepatoma cells suggesting a role for this molecule in sporozoite invasion of hepatocytes. *The EMBO journal* **12**, 2881-2889 (1993).

71 Robson, K. J. *et al.* Thrombospondin-related adhesive protein (TRAP) of Plasmodium falciparum: expression during sporozoite ontogeny and binding to human hepatocytes. *The EMBO journal* **14**, 3883-3894 (1995).

72 Scarselli, E. *et al.* Analysis of the human antibody response to thrombospondin-related anonymous protein of Plasmodium falciparum. *Infection and immunity* **61**, 3490-3495 (1993).

73 Khusmith, S. *et al.* Protection against malaria by vaccination with sporozoite surface protein 2 plus CS protein. *Science* **252**, 715-718 (1991).

74 McCoubrie, J. E. *et al.* Evidence for a common role for the serine-type Plasmodium falciparum serine repeat antigen proteases: implications for vaccine and drug design. *Infection and immunity* **75**, 5565-5574, doi:10.1128/iai.00405-07 (2007).

75 Bzik, D. J., Li, W. B., Horii, T. & Inselburg, J. Amino acid sequence of the serine-repeat antigen (SERA) of Plasmodium falciparum determined from cloned cDNA. *Molecular and biochemical parasitology* **30**, 279-288 (1988).

76 Fox, B. A. & Bzik, D. J. Analysis of stage-specific transcripts of the Plasmodium falciparum serine repeat antigen (SERA) gene and transcription from the SERA locus. *Molecular and biochemical parasitology* **68**, 133-144, doi:http://dx.doi.org/10.1016/0166-6851(94)00162-6 (1994).

77 Delplace, P. *et al.* Protein p126: a parasitophorous vacuole antigen associated with the release of Plasmodium falciparum merozoites. *Biology of the cell / under the auspices of the European Cell Biology Organization* **64**, 215-221 (1988).

78 Owalla, T. J., Palacpac, N. M., Shirai, H., Horii, T. & Egwang, T. G. Association of naturally acquired IgG antibodies against Plasmodium falciparum serine repeat antigen-5 with reduced placental parasitemia and normal birth weight in pregnant Ugandan women: a pilot study. *Parasitology international* **62**, 237-239, doi:10.1016/j.parint.2013.01.006 (2013).

79 Yagi, M. *et al.* Antibody titres and boosting after natural malaria infection in BK-SE36 vaccine responders during a follow-up study in Uganda. *Sci Rep* **6**, 34363, doi:10.1038/srep34363 (2016).
80 Stallmach, R. *et al.* Plasmodium falciparum SERA5 plays a non-enzymatic role in the malarial asexual blood-stage lifecycle. *Mol Microbiol* **96**, 368-387, doi:10.1111/mmi.12941 (2015).
81 Aoki, S. *et al.* Serine repeat antigen (SERA5) is predominantly expressed among the SERA multigene family of Plasmodium falciparum, and the acquired antibody titers correlate with serum inhibition of the parasite growth. *J Biol Chem* **277**, 47533-47540, doi:10.1074/jbc.M207145200 (2002).
82 Chulay, J. D. *et al.* Monoclonal antibody characterization of Plasmodium falciparum antigens in immune complexes formed when schizonts rupture in the presence of immune serum. *Journal of immunology (Baltimore, Md. : 1950)* **139**, 2768-2774 (1987).
83 Perrin, L. H. *et al.* Antimalarial immunity in Saimiri monkeys. Immunization with surface components of asexual blood stages. *The Journal of experimental medicine* **160**, 441-451 (1984).
84 Miller, S. K. *et al.* A subset of Plasmodium falciparum SERA genes are expressed and appear to play an important role in the erythrocytic cycle. *J Biol Chem* **277**, 47524-47532, doi:10.1074/jbc.M206974200 (2002).
85 Schmidt-Christensen, A., Sturm, A., Horstmann, S. & Heussler, V. T. Expression and processing of Plasmodium berghei SERA3 during liver stages. *Cell Microbiol* **10**, 1723-1734, doi:10.1111/j.1462-5822.2008.01162.x (2008).
86 Horii, T. *et al.* Evidences of protection against blood-stage infection of Plasmodium falciparum by the novel protein vaccine SE36. *Parasitology international* **59**, 380-386, doi:10.1016/j.parint.2010.05.002 (2010).
87 Palacpac, N. M., Arisue, N., Tougan, T., Ishii, K. J. & Horii, T. Plasmodium falciparum serine repeat antigen 5 (SE36) as a malaria vaccine candidate. *Vaccine* **29**, 5837-5845, doi:10.1016/j.vaccine.2011.06.052 (2011).
88 Yagi, M. *et al.* Protective epitopes of the Plasmodium falciparum SERA5 malaria vaccine reside in intrinsically unstructured N-terminal repetitive sequences. *PLoS One* **9**, e98460, doi:10.1371/journal.pone.0098460 (2014).
89 Arisue, N. *et al.* Clues to Evolution of the SERA Multigene Family in 18 Plasmodium Species. *PLoS ONE* **6**, e17775, doi:10.1371/journal.pone.0017775 (2011).
90 Arisue, N., Hirai, M., Arai, M., Matsuoka, H. & Horii, T. Phylogeny and evolution of the SERA multigene family in the genus Plasmodium. *Journal of molecular evolution* **65**, 82-91, doi:10.1007/s00239-006-0253-1 (2007).
91 Kooij, T. W. *et al.* A Plasmodium whole-genome synteny map: indels and synteny breakpoints as foci for species-specific genes. *PLoS Pathog* **1**, e44, doi:10.1371/journal.ppat.0010044 (2005).
92 Bourgon, R. *et al.* The serine repeat antigen (SERA) gene family phylogeny in Plasmodium: the impact of GC content and reconciliation of gene and species trees. *Molecular biology and evolution* **21**, 2161-2171, doi:10.1093/molbev/msh228 (2004).
93 Pain, A. *et al.* The genome of the simian and human malaria parasite Plasmodium knowlesi. *Nature* **455**, 799-803, doi:10.1038/nature07306 (2008).
94 Benavente, E. D. *et al.* A reference genome and methylome for the Plasmodium knowlesi A1-H.1 line. *Int J Parasitol*, doi:10.1016/j.ijpara.2017.09.008 (2017).
95 World Health Organization. *Informal Consultation on the Public Health Importance of Plasmodium knowlesi*,

<http://www.wpro.who.int/mvp/documents/docs/Pknowlesi_final_report.pdf> (2011).

96 Bargieri, D. Y. *et al.* Plasmodium Merozoite TRAP Family Protein Is Essential for Vacuole Membrane Disruption and Gamete Egress from Erythrocytes. *Cell Host Microbe* **20**, 618-630, doi:10.1016/j.chom.2016.10.015 (2016).

97 Palacpac, N. M. *et al.* Plasmodium vivax serine repeat antigen (SERA) multigene family exhibits similar expression patterns in independent infections. *Molecular and biochemical parasitology* **150**, 353-358, doi:10.1016/j.molbiopara.2006.07.006 (2006).

98 Cheong, F. W., Fong, M. Y. & Lau, Y. L. Identification and characterization of epitopes on Plasmodium knowlesi merozoite surface protein-142 (MSP-142) using synthetic peptide library and phage display library. *Acta tropica* **154**, 89-94, doi:10.1016/j.actatropica.2015.11.005 (2016).

99 Cheong, F. W., Fong, M. Y., Lau, Y. L. & Mahmud, R. Immunogenicity of bacterial-expressed recombinant Plasmodium knowlesi merozoite surface protein-142 (MSP-142). *Malaria journal* **12**, 454, doi:10.1186/1475-2875-12-454 (2013).

100 Cheong, F. W., Lau, Y. L., Fong, M. Y. & Mahmud, R. Evaluation of recombinant Plasmodium knowlesi merozoite surface protein-1(33) for detection of human malaria. *Am J Trop Med Hyg* **88**, 835-840, doi:10.4269/ajtmh.12-0250 (2013).

101 Lau, Y. L. *et al.* Evaluation of codon optimized recombinant Plasmodium knowlesi Merozoite Surface Protein-119 (pkMSP-119) expressed in Pichia pastoris. *Tropical biomedicine* **31**, 749-759 (2014).

102 Sonaimuthu, P. *et al.* Detection of human malaria using recombinant Plasmodium knowlesi merozoire surface protein-1 (MSP-1(1)(9)) expressed in Escherichia coli. *Experimental parasitology* **153**, 118-122, doi:10.1016/j.exppara.2015.03.010 (2015).

103 Blackman, M. J. *et al.* Plasmodium knowlesi: secondary processing of the malaria merozoite surface protein-1. *Experimental parasitology* **83**, 229-239, doi:10.1006/expr.1996.0069 (1996).

104 Faber, B. W. *et al.* Low levels of polymorphisms and no evidence for diversifying selection on the Plasmodium knowlesi Apical Membrane Antigen 1 gene. *PLoS One* **10**, e0124400, doi:10.1371/journal.pone.0124400 (2015).

105 Triglia, T. *et al.* Apical membrane antigen 1 plays a central role in erythrocyte invasion by Plasmodium species. *Mol Microbiol* **38**, 706-718 (2000).

106 Deans, J. A., Thomas, A. W., Alderson, T. & Cohen, S. Biosynthesis of a putative protective Plasmodium knowlesi merozoite antigen. *Molecular and biochemical parasitology* **11**, 189-204, doi:http://dx.doi.org/10.1016/0166-6851(84)90065-3 (1984).

107 Deans, J. A. Protective antigens of bloodstage Plasmodium knowlesi parasites. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **307**, 159-169 (1984).

108 Fong, M. Y., Ahmed, M. A., Wong, S. S., Lau, Y. L. & Sitam, F. Genetic Diversity and Natural Selection of the Plasmodium knowlesi Circumsporozoite Protein Nonrepeat Regions. *PLoS One* **10**, e0137734, doi:10.1371/journal.pone.0137734 (2015).

109 Sharma, S., Goswami, A., Singh, N. J., Kabilan, L. & Deodhar, S. S. Immunogenicity of the nonrepetitive regions of the circumsporozoite protein of Plasmodium knowlesi. *Am J Trop Med Hyg* **55**, 635-641 (1996).

110 Aurrecoechea, C. *et al.* PlasmoDB: a functional genomic database for malaria parasites. *Nucleic acids research* **37**, D539-543, doi:10.1093/nar/gkn814 (2009).

111 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

112 Gopal, G. J. & Kumar, A. Strategies for the production of recombinant protein in Escherichia coli. *The protein journal* **32**, 419-425, doi:10.1007/s10930-013-9502-5 (2013).

113 Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Current opinion in structural biology* **18**, 581-586, doi:10.1016/j.sbi.2008.07.001 (2008).

114 Vedadi, M. *et al.* Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms. *Molecular and biochemical parasitology* **151**, 100-110, doi:10.1016/j.molbiopara.2006.10.011 (2007).

115 Moon, R. W. *et al.* Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 531-536, doi:10.1073/pnas.1216457110 (2013).

116 Li, M. Z. & Elledge, S. J. SLIC: a method for sequence- and ligation-independent cloning. *Methods in molecular biology (Clifton, N.J.)* **852**, 51-59, doi:10.1007/978-1-61779-564-0_5 (2012).

117 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth* **6**, 343-345, doi:http://www.nature.com/nmeth/journal/v6/n5/suppinfo/nmeth.1318_S1.html (2009).

118 Tetteh, K. K. *et al.* Analysis of antibodies to newly described Plasmodium falciparum merozoite antigens supports MSPDBL2 as a predicted target of naturally acquired immunity. *Infection and immunity* **81**, 3835-3842, doi:10.1128/iai.00301-13 (2013).

119 Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**, 207-234 (2005).

120 Studier, F. W. Stable expression clones and auto-induction for protein production in E. coli. *Methods in molecular biology (Clifton, N.J.)* **1091**, 17-32, doi:10.1007/978-1-62703-691-7_2 (2014).

121 Polley, S. D. *et al.* Repeat sequences in block 2 of Plasmodium falciparum merozoite surface protein 1 are targets of antibodies associated with protection from malaria. *Infection and immunity* **71**, 1833-1842 (2003).

122 Collins, C. R. *et al.* Fine mapping of an epitope recognized by an invasion-inhibitory monoclonal antibody on the malaria vaccine candidate apical membrane antigen 1. *J Biol Chem* **282**, 7431-7441, doi:10.1074/jbc.M610562200 (2007).

123 Collins, C. R., Withers-Martinez, C., Hackett, F. & Blackman, M. J. An inhibitory antibody blocks interactions between components of the malarial invasion machinery. *PLoS Pathog* **5**, e1000273, doi:10.1371/journal.ppat.1000273 (2009).

124 Burghaus, P. A. & Holder, A. A. Expression of the 19-kilodalton carboxy-terminal fragment of the Plasmodium falciparum merozoite surface protein-1 in Escherichia coli as a correctly folded protein. *Molecular and biochemical parasitology* **64**, 165-169 (1994).

125 Cook, J. *et al.* Trends in parasite prevalence following 13 years of malaria interventions on Bioko island, Equatorial Guinea: 2004–2016. *Malaria journal* **17**, 62, doi:10.1186/s12936-018-2213-9 (2018).

126 Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *The Journal of animal ecology* **77**, 802-813, doi:10.1111/j.1365-2656.2008.01390.x (2008).

127 Stewart, L. *et al.* Rapid assessment of malaria transmission using age-specific sero-conversion rates. *PLoS One* **4**, e6083, doi:10.1371/journal.pone.0006083 (2009).

128 Logan-Klumpler, F. J. *et al.* GeneDB--an annotation database for pathogens. *Nucleic acids research* **40**, D98-108, doi:10.1093/nar/gkr1032 (2012).
129 Mount, D. W. Using the Basic Local Alignment Search Tool (BLAST). *CSH protocols* **2007**, pdb.top17, doi:10.1101/pdb.top17 (2007).
130 Lewis, T. E. *et al.* Genome3D: a UK collaborative project to annotate genomic sequences with predicted 3D structures based on SCOP and CATH domains. *Nucleic acids research* **41**, D499-507, doi:10.1093/nar/gks1266 (2013).
131 Offeddu, V., Thathy, V., Marsh, K. & Matuschewski, K. Naturally acquired immune responses against Plasmodium falciparum sporozoites and liver infection. *Int J Parasitol* **42**, 535-548, doi:10.1016/j.ijpara.2012.03.011 (2012).
132 Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nature methods* **12**, 7-8, doi:10.1038/nmeth.3213 (2015).
133 Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research* **43**, W174-181, doi:10.1093/nar/gkv342 (2015).
134 Kantele, A. & Jokiranta, T. S. Review of Cases With the Emerging Fifth Human Malaria Parasite, Plasmodium knowlesi. *Clinical Infectious Diseases* **52**, 1356-1362, doi:10.1093/cid/cir180 (2011).
135 Grigg, M. J. *et al.* Artesunate-mefloquine versus chloroquine for treatment of uncomplicated Plasmodium knowlesi malaria in Malaysia (ACT KNOW): an open-label, randomised controlled trial. *The Lancet infectious diseases*, doi:10.1016/s1473-3099(15)00415-6 (2016).
136 Shirahama, K., Tsujii, K. & Takagi, T. Free-boundary electrophoresis of sodium dodecyl sulfate-protein polypeptide complexes with special reference to SDS-polyacrylamide gel electrophoresis. *Journal of biochemistry* **75**, 309-319 (1974).
137 Ibel, K. *et al.* Protein-decorated micelle structure of sodium-dodecyl-sulfate--protein complexes as determined by neutron scattering. *European journal of biochemistry* **190**, 311-318 (1990).
138 Segrest, J. P., Jackson, R. L., Andrews, E. P. & Marchesi, V. T. Human erythrocyte membrane glycoprotein: a re-evaluation of the molecular weight as determined by SDS polyacrylamide gel electrophoresis. *Biochem Biophys Res Commun* **44**, 390-395 (1971).
139 Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 1760-1765, doi:10.1073/pnas.0813167106 (2009).
140 Nadeau, V. G., Rath, A. & Deber, C. M. Sequence hydropathy dominates membrane protein response to detergent solubilization. *Biochemistry* **51**, 6228-6237, doi:10.1021/bi201853n (2012).
141 Chevallet, M., Luche, S. & Rabilloud, T. Silver staining of proteins in polyacrylamide gels. *Nature protocols* **1**, 1852-1858, doi:10.1038/nprot.2006.288 (2006).
142 Weiss, W., Weiland, F. & Gorg, A. Protein detection and quantitation technologies for gel-based proteome analysis. *Methods in molecular biology (Clifton, N.J.)* **564**, 59-82, doi:10.1007/978-1-60761-157-8_4 (2009).
143 Assefa, S. *et al.* Population genomic structure and adaptation in the zoonotic malaria parasite Plasmodium knowlesi. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1509534112 (2015).
144 Imai, N., White, M. T., Ghani, A. C. & Drakeley, C. J. Transmission and control of Plasmodium knowlesi: a mathematical modelling study. *PLoS neglected tropical diseases* **8**, e2978, doi:10.1371/journal.pntd.0002978 (2014).

145 Mohd Abd Razak, M. R. *et al.* Genetic Diversity of Plasmodium falciparum Populations in Malaria Declining Areas of Sabah, East Malaysia. *PLoS One* **11**, e0152415, doi:10.1371/journal.pone.0152415 (2016).
146 World Health Organization. World Malaria report: 2016. (2016).
147 World Health Organization. *WHO Malaria Policy Advisory Committee (MPAC) meeting* <http://www.who.int/malaria/publications/atoz/WHO-HTM-GMP-2017.8-eng.pdf?ua=1> (2017).
148 Sharma, S. & Godson, G. N. Expression of the major surface antigen of Plasmodium knowlesi sporozoites in yeast. *Science* **228**, 879-882 (1985).
149 Mahdi Abdel Hamid, M. *et al.* Vaccination with Plasmodium knowlesi AMA1 formulated in the novel adjuvant co-vaccine HT protects against blood-stage challenge in rhesus macaques. *PLoS One* **6**, e20547, doi:10.1371/journal.pone.0020547 (2011).
150 Singh, A. P., Puri, S. K. & Chitnis, C. E. Antibodies raised against receptor-binding domain of Plasmodium knowlesi Duffy binding protein inhibit erythrocyte invasion. *Molecular and biochemical parasitology* **121**, 21-31 (2002).
151 Singh, V., Kaushal, D. C., Rathaur, S., Kumar, N. & Kaushal, N. A. Cloning, overexpression, purification and characterization of Plasmodium knowlesi lactate dehydrogenase. *Protein expression and purification* **84**, 195-203, doi:10.1016/j.pep.2012.05.008 (2012).
152 Hakimi, H. *et al.* Plasmodium vivax and Plasmodium knowlesi: cloning, expression and functional analysis of 1-Cys peroxiredoxin. *Experimental parasitology* **133**, 101-105, doi:10.1016/j.exppara.2012.10.018 (2013).
153 Prasad, R., Atul, Soni, A., Puri, S. K. & Sijwali, P. S. Expression, characterization, and cellular localization of knowpains, papain-like cysteine proteases of the Plasmodium knowlesi malaria parasite. *PLoS One* **7**, e51619, doi:10.1371/journal.pone.0051619 (2012).
154 Tyagi, K. *et al.* Recognition of Human Erythrocyte Receptors by the Tryptophan-Rich Antigens of Monkey Malaria Parasite Plasmodium knowlesi. *PLoS One* **10**, e0138691, doi:10.1371/journal.pone.0138691 (2015).
155 De Silva, J. R., Lau, Y. L. & Fong, M. Y. Expression and Evaluation of Recombinant Plasmodium knowlesi Merozoite Surface Protein-3 (MSP-3) for Detection of Human Malaria. *PLoS One* **11**, e0158998, doi:10.1371/journal.pone.0158998 (2016).
156 Lucky, A. B. *et al.* Plasmodium knowlesi Skeleton-Binding Protein 1 Localizes to the 'Sinton and Mulligan' Stipplings in the Cytoplasm of Monkey and Human Erythrocytes. *PLoS One* **11**, e0164272, doi:10.1371/journal.pone.0164272 (2016).
157 Ahmed, A. M. *et al.* Disease progression in Plasmodium knowlesi malaria is linked to variation in invasion gene family members. *PLoS neglected tropical diseases* **8**, e3086, doi:10.1371/journal.pntd.0003086 (2014).
158 Rawa, M. S., Fong, M. Y. & Lau, Y. L. Genetic diversity and natural selection in the rhoptry-associated protein 1 (RAP-1) of recent Plasmodium knowlesi clinical isolates from Malaysia. *Malaria journal* **15**, 62, doi:10.1186/s12936-016-1127-7 (2016).
159 Uplekar, S. *et al.* Characterizing Antibody Responses to Plasmodium vivax and Plasmodium falciparum Antigens in India Using Genome-Scale Protein Microarrays. *PLoS neglected tropical diseases* **11**, e0005323, doi:10.1371/journal.pntd.0005323 (2017).
160 Grigg, M. J. *et al.* Individual-level factors associated with the risk of acquiring human Plasmodium knowlesi malaria in Malaysia: a case-control study. *The Lancet. Planetary health* **1**, e97-e104, doi:10.1016/s2542-5196(17)30031-1 (2017).

Chapter 5 : HIGH THROUGHPUT SEROLOGICAL SCREENING OF MALAYSIAN HOSPITAL CASE SAMPLES AND COMMUNITY SURVEY SAMPLES USING THE MULTIPLEX BEAD ASSAY

5.1. INTRODUCTION

Serological methods have been especially developed over the years to detect antibodies towards *Plasmodium* spp. proteins, identifying infection and determining disease exposure over time[1-7] and are now widely accepted. Routinely, the measurement of exposure to malaria in endemic populations is carried out using a small number of well characterised plasmodium antigens[8-11]. New tools are needed for close monitoring of malaria infection as it moves towards elimination in certain areas because submicroscopic infections are missed by routine microscopy diagnosis[12-14], which can give false reassurance of elimination. Appropriate *P. knowlesi* detection tools in serology have long been overdue and, as discussed in chapter 4, we have now developed and made available a panel of *P. knowlesi* specific antigenic markers of exposure for further characterisation of this organism and the human host immune response to it[15]. Novel proteins are being identified at an increasing rate[16-23] and developing a platform appropriate to screen such a large number of targets has become essential in order to obtain a better understanding of host-pathogen interactions and antibody-mediated immunity towards pathogens, in this case, malaria. Antibody-based detection assays such as ELISA and Western blot for example, in general, require large quantities of reagents and serum samples, the latter being the most difficult to overcome in particular settings where samples are finite, and these platforms can only look at very few antigens at a time. There has been a technological shift towards multiplex platforms in the current market in order to meet this need of detecting multiple targets in a single sample. Cytometric Bead Assay, also known as multiplex bead assay (MBA)[24-26] or bead suspension assay, is one assay that can be used as a multiplex platform for malaria serology. The MBA is a flow cytometry application that uses microspheres (beads) that are uniquely identified by fluorescent dyes and differentiated based on specific spectral addresses which can be recognized by wavelength differences[27,28].

Internally dyed or different sized[29] microspheres are distinguished by light scatter characteristics and are also detected by phycoerythrin(PE)-labelled anti-human antibodies, therefore different bead sets can be measured alongside immunoassay signals[30]. These microspheres can be coupled to a multitude of different reagents (***Figure 5.1***).

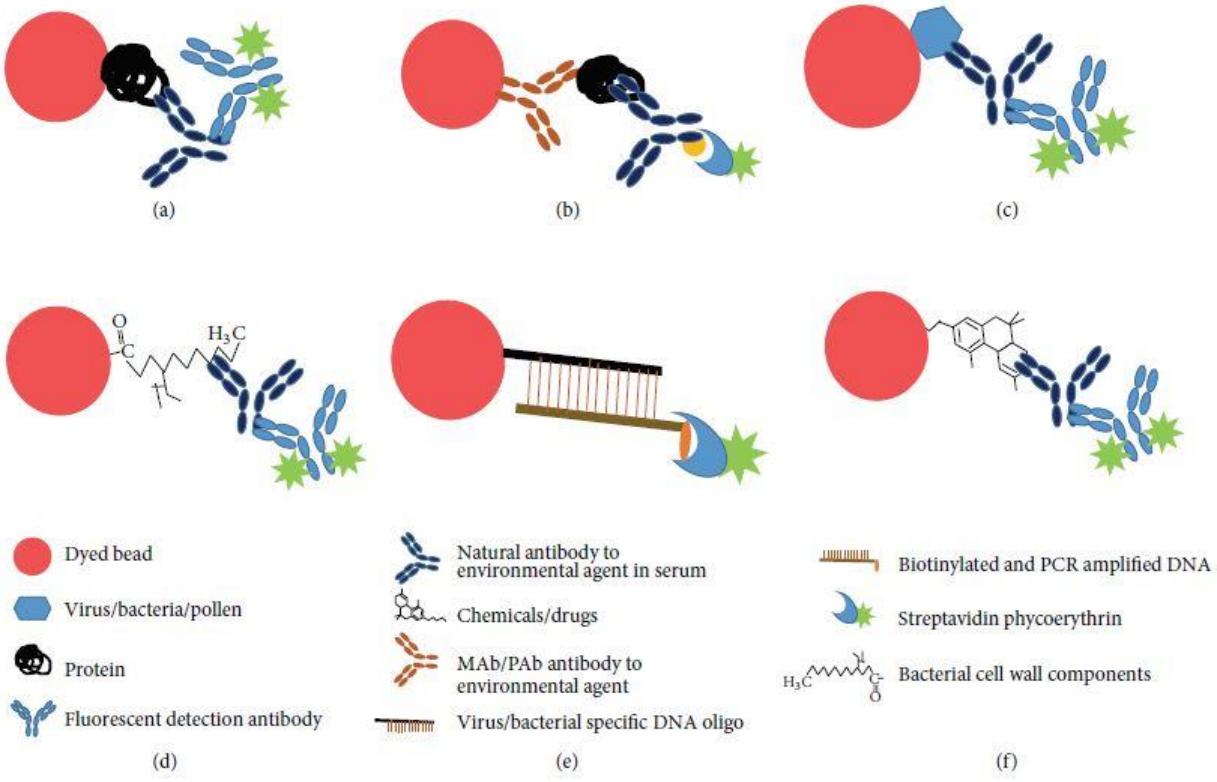


Figure 5.1 Multiplex bead assay range of applications. Beads can be covalently coupled to: (a) Protein; (b) monoclonal or polyclonal antibodies; (c) viruses, (d) bacterial cell wall components; (e) DNA from virus/bacteria; and (f) chemicals/drugs. Coupled beads can then be detected using phycoerythrin-labelled detection agents such as fluorescent antibodies or streptavidin. Image from a published source[30].

In this case, beads can be coupled to recombinant malarial protein antigens, combined to screen multiple antigens on a small amount of plasma, and then individually identified by their specific spectral address (***Figure 5.2***). This technology is a sound replacement to the ELISA method when testing multiple targets on small amounts of serum and it has been reported to have a better reproducibility, better and larger dynamic range and to give comparable results to the ELISA platform[31-35]. Some of the advantages to replacing the ELISA platform with a bead-based assay

are the high throughput aspect of this platform, the need for fewer reagents and reduced sample volume and the ability to analyse a target in the context of multiple others, all under the same experimental conditions[36]. With this technology it is possible to quantify up to 500 analytes in a single well per sample using as little as 5 µl of serum[37,38]. With this platform being increasingly used, it is now an affordable mid-high throughput multiplexing platform.

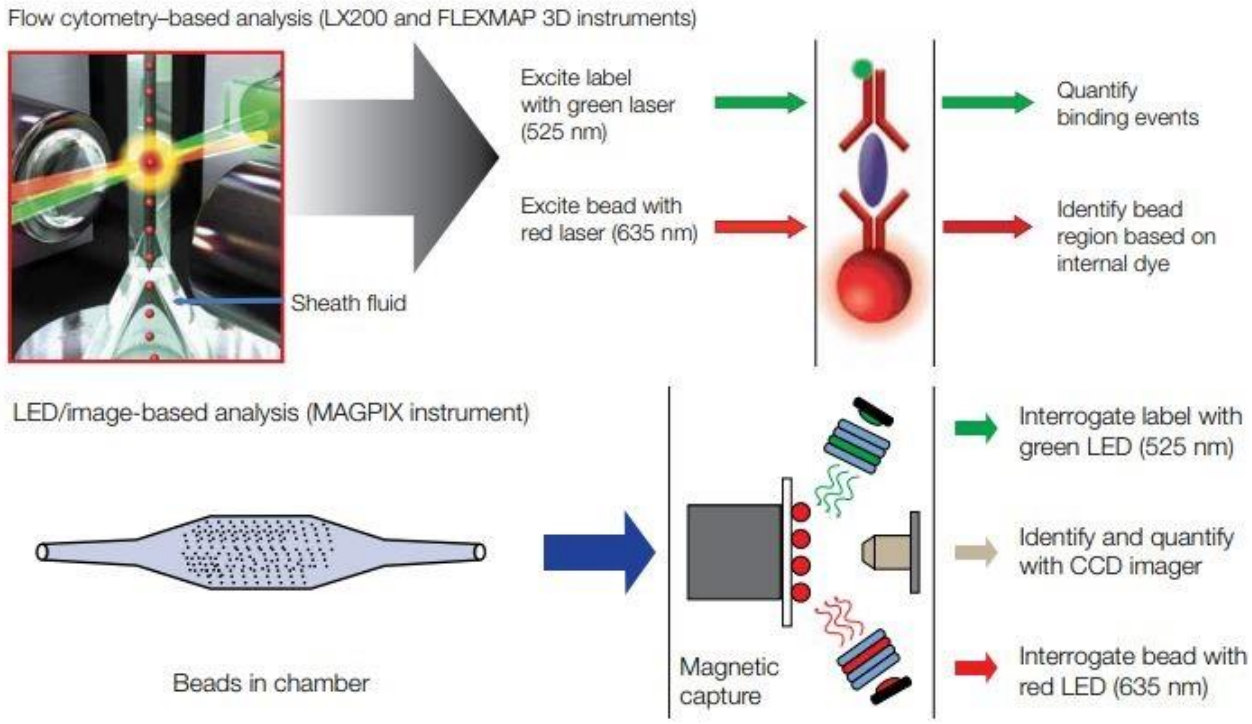


Figure 5.2 Schematic representation of a sandwich-based Luminex assay. Targets are measured by the Luminex xMAP technology using light-emitting diodes (LEDs) or lasers. Typically up to 30 targets can be measured simultaneously. The coupled beads are dispensed into an assay plate before sample addition, then a detection antibody is added. The beads are then individually excited by the appropriate laser (top panels), in the case of the LX200 and FLEXMAP 3D instruments, or the different LEDs which are then detected by the CCD imager (bottom panels), in the case of MAGPIX instrument, and results saved. Figure from ThermoFisher Scientific (https://tinyurl.com/yauzl9c7).

This technology can be used in multiple ways, for example in vaccine development[39] or pathogen characterisation[40-42] (***Figure 5.1***). It has been developed for detecting antibodies towards multiple pathogens causing diseases such as Meningitis[31], Tetanus[33], Diphtheria[33], Influenza[33], Toxoplasmosis[43-45], and now, Malaria[46-50] as well, a field where this method is gradually taking ground. Not only can the MBA

platform be used in immunoassays measuring antibodies[41,47,51] and cytokines[52-54], but also in nucleic acid assays[55-58], making it a versatile high throughput technology. This chapter describes the analysis of serum samples from Malaysian clinical cases used in the previous chapter on the MBA platform using *P. knowlesi*-specific recombinant antigens.

5.2. METHODS

5.2.1. Coupling of recombinant proteins to MicroPlex microspheres for multiplex bead assay analysis

The antigen coupling protocol was modified from a method previously described by Ondigo *et. al*.[59]. The appropriate antigen concentrations were determined and optimised by a coupling titration assay, where the antigens were coupled to beads in a range of concentrations (i.e. 0.015, 0.12, 0.94, 7.5, 60 and 480 µg/ml) and the results plotted to determine the ideal coupling concentration per antigen. Tom Hall helped to perform this part of the experiment. Briefly, for coupling a whole MicroPlex® bead stock (1 ml of beads at $12.5x10^6$/ml concentration), beads were both vortexed and sonicated for 60 seconds to homogenise the suspension and avoid clumps of beads, placed on a magnetic rack before being washed with 500 µl mili-Q water and, after being removed from the magnetic rack, resuspended in 400 µl 100 mM monobasic Sodium Phosphate buffer (Sigma-Aldrich), pH 6.2 (activation buffer). To activate the beads for chemical coupling of the antigen, 50 µl of 50 mg/ml Sulfo-NHS (N-Hydroxysulfosuccinimide sodium salt) (Sigma-Aldrich) and 50 µl 50 mg/ml EDC (N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride) (Sigma-Aldrich) were added sequentially, vortexing after each addition. The bead suspensions were incubated at RT and vortexed moderately for 10 seconds at both 10 and 20 minutes during the incubation period. The activated beads were briefly spun in a minifuge to pellet beads, placed in a magnetic rack for 1 minute and the supernatant removed. Beads were washed three times in 500 µl 1x PBS/T (Oxoid) to remove the activation buffer. After the final wash, the activated beads were coated with antigen by resuspension in 1 ml of the relevant antigen previously diluted in 1x PBS solution, at the appropriate concentration. They were then incubated at RT for 2 hours away from light on a tube rotator. After incubation, the beads were centrifuged; the supernatant removed, and washed three times with 500 µl Buffer A (1x PBS,

0.5% BSA, 0.05%Tween, 0.02% sodium azide). After the final wash, coupled beads were pelleted and resuspended in 1000 µl Buffer A with 0.02% Pefabloc (Sigma-Aldrich) (storage buffer) and counted on a haemocytometer (NanoEnTek) to determine the percentage bead recovery. Coated beads were stored at 4ºC until required.

5.2.2. Measuring total IgG responses to antigen coupled microsphere beads using the MagPix (Luminex) suspension bead array

Antigen coupled bead stocks were combined and diluted into buffer A to make a 20 microspheres/µl solution. 50 µl of the microsphere mixture was transferred to each well of a 96-well Bioplex pro microtiter plate (Bio-Rad); this resulted in 1000 beads/region/well. Each bead set has a specific spectral address making it possible to track each coupled antigen despite using a single secondary antibody reagent. The beads were washed once by magnetically pelleting the beads, vigorously dispensing the waste liquid into a sink and adding 100 µl of PBS/T. The previous day, serum samples and controls were diluted 1/400 in buffer B (0.1% casein, 0.5% PVA (Poly(vinyl alcohol)), 0.5% PVP (Polyvinylpyrrolidone) and *E. coli* extract at 15.25 µg/ml in buffer A, with the liquid plastic solutions of PVA and PVP used to prevent non-specific binding to the beads), 50 µl were added to the plate containing beads in singlet for serum samples and in duplicate for controls, and incubated for 1 hour and 30 minutes at RT in the dark on a microplate shaker set to 700 rpm.
The *P. knowlesi*-negative control samples used were based on a hyper-immune endemic Tanzanian pool (*P. falciparum*)[60], lyophilised WHO Reference Reagent for anti-malaria human serum (NIBSC, UK; 10/198 (*P. falciparum*)), lyophilised non WHO reference material for anti-malaria human plasma (NIBSC, UK; 72/096 (*P. vivax*)) and the malaria naïve negative control samples used were a pool of Public Health England (PHE; LSHTM ethics approval #11684) serum samples. Plates were washed three times using 100 µl of 1x PBS/T as described above and 50 µl of goat anti-human IgG (gamma- chain specific, F(ab`)2 fragment-R-phycoerythrin (Jackson

immunoResearch) secondary antibody diluted 1/200 in buffer A was added to each well. The plates were incubated for 1 hour and 30 minutes as above. Plates were washed three times as previously, 50 µl of buffer A was added to all wells and the plates were incubated for 30 minutes at RT in the dark on a microplate shaker set to 700 rpm. Plates were washed once using 100 µl of 1x PBS/T, the beads were resuspended in 100 µl of 1x PBS and analysed on the MAGPIX® reader (Luminex corp, USA). The results were expressed as median fluorescent intensity (MFI).

5.2.3. STATISTICAL ANALYSIS

Final net median fluorescent intensity (MFI) was obtained by subtracting background MFI values, reducing background reactivity. Cut off values were defined separately for each antigen, *Plasmodium* spp. and relevant study. Although 37 bead sets were used to screen these samples, only up to 5 bead sets were analysed for the purpose of this chapter: two antigens from the *P. knowlesi* panel (*Pk*SERA3 antigen 2 and *Pk*SSP2 as these were the most reactive candidates from the panel at the time) alongside the relevant control antigens (*Pv*MSP1-19, *Pv*AMA1 and *Pf*MSP1-19). For the hospital-based clinical treatment trial samples, seropositivity cut offs were calculated based on Public Health England negative control sera samples (average MFI ± (3xSD)). The seropositivity cut offs were based on the following values: *Pk*SERA3 ag2=4318.31; *Pk*SSP2/TRAP=2126.35 and *Pv*MSP1-19=729.5. To define seropositive individuals in the Sabah cross sectional survey, mixture models were fit for normalised MFI. Cut off values to define antigen seropositivity for *P. vivax* and *P. falciparum* were defined as the mean MFI of the seronegative population plus 3 standard deviations. For a more parsimonious cut off value for *P. knowlesi*, seropositivity cut offs were defined as the mean MFI of the seronegative population plus 5 standard deviations. The seropositivity cut offs were based on the following values: *Pk*SERA3 ag2=892.93; *Pk*SSP2=929.35; *Pv*MSP1-19=602.05, *Pv*AMA1=147.8 and *Pf*MSP1-19=466.9. Risk factors associated with *P. knowlesi* seropositivity were evaluated using multivariate logistic regression. Analysis and

graphic representation of serological data was carried out using PRISM (GraphPad PRISM 7) and STATA/IC 14.2 (StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.).

5.3. RESULTS

5.3.1. CLINICAL SAMPLE SERUM REACTIVITY TO *P. KNOWLESI*-SPECIFIC RECOMBINANT ANTIGENS USING THE CYTOMETRIC BEAD ASSAY PLATFORM

The Luminex assay used the full set of Malaysian serum and bloodspot samples collected from adults and children hospitalised with malaria at different time-points (day 0, 7 and 28, as well as 1 and 2 year follow up; Table 3.1; chapter 3, section 3.3) post-treatment, as described above but including *P. vivax* (n=36), *P. falciparum* (n=16) and *P. malariae* (n=1) as well as *P. knowlesi* (n=110) diagnosed individuals. The experimental strategy for this chapter is outlined in **Table *5.1***, showing the identification of the target sequences of interest and, highlighted in green, the test samples used.

All samples were assayed using the full Luminex bead panel (37 bead sets), including two antigens from the panel of *P. knowlesi*-specific proteins (*Pk*SERA3 antigen 2 and *Pk*SSP2) (Table 4.3; chapter 4, section 4.2.1.) alongside *Pv*MSP1-19 as a control antigen. For the purpose of this thesis, we only analysed the *P. knowlesi* antigens alongside relevant control antigens. Although we only show preliminary analysis of this here using only the *P. knowlesi* antigens, the full panel was included for the full epidemiological analysis of the Sabah cross sectional survey (unpublished data). The 2 year follow up analysis was not included here due to a very low number of samples being available (n=3).

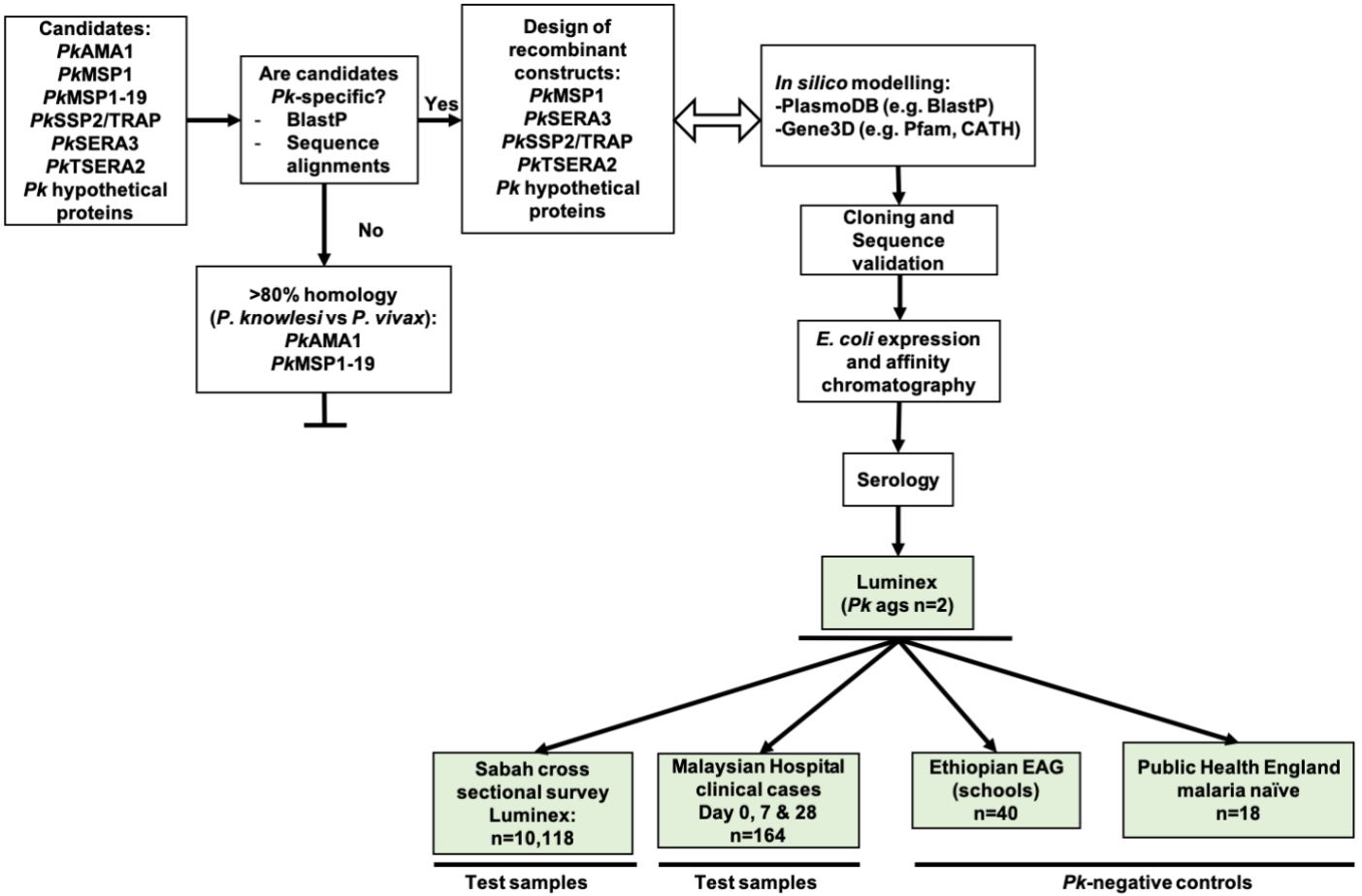


Figure 5.3 Flowchart summarising the experimental strategy used in the identification and validation of the *P. knowlesi*-specific candidates and sample sets used in this chapter. Test samples were obtained from a cross sectional survey (n=10,118) and from a hospital-based clinical treatment trial (n=164), both in Sabah, Malaysia. The negative controls comprised of samples from an Easy Access Group (EAG) in Ethiopia as *P. vivax*-positive *P. knowlesi*-negative samples (n=40) and the malaria-naïve group were obtained from Public Health England (n=18). Cross sectional survey samples and the clinical samples were tested on the Luminex cytometric bead assay platform alongside the *P. knowlesi*-negative controls. Highlighted in green are the sections relevant to this chapter.

Similar to the ELISA results, day 7 showed the highest antibody reactivity in *P. knowlesi* diagnosed individuals to both antigens with *Pk*SERA3 antigen 2 showing the highest prevalence, 57.9% (44/76) (***Table 5.1***; ***Figure 5.4***). *P. falciparum* diagnosed individuals showed the highest prevalence for *Pk*SERA3 antigen 2 at day 0, 66.7% (10/15). *P. vivax* diagnosed individuals showed the highest prevalence for *Pk*SERA3 antigen 2 at day 7, 100% (30/30) and appeared to track with the *P. knowlesi* diagnosed individuals. Because these were clinical samples from regions where both *P. vivax* and *P. falciparum* are also endemic alongside *P. knowlesi*, there is no way of guaranteeing that these individuals had not been previously infected with other *Plasmodium* spp., therefore we may see *P. knowlesi* antigen reactivity in individuals diagnosed with non-*P. knowlesi* species. Although there were fewer samples, individuals that were followed up 1 year later showed much lower reactivity towards the *P. knowlesi* antigens compared to the previous time points (***Table 5.1***; ***Figure 5.4***). From all individuals tested, independently of diagnostic status, *Pk*SERA3 antigen 2 showed the highest prevalence at day 7, with 66.7% (80/120) of individuals above the cut off value (***Table 5.1***; ***Figure 5.4***). Detectable antibody reactivity to the *P. knowlesi*-specific antigens was largely absent in the Ethiopian samples to both *P. knowlesi* antigens, with 82.5% (33/40) and 77.5% (31/40) of individuals non-reactive to *Pk*SSP2 and *Pk*SERA3 antigen 2, respectively (***Figure 5.4***).

Table 5.1 Hospital case sample seropositivity by day, antigen tested and diagnostic status.

Data is presented as number of people seropositive and percentages in parenthesis. Seropositivity was calculated based on Public Health England negative control sera samples (average net MFI ± (3xSD)). Seropositivity (pos) cut offs for the *P. knowlesi* antigens (ags) were based on the following values: *Pk*SERA3 ag2=4318.31 and *Pk*SSP2/TRAP=2126.35.

	Day 0 **n=154** *Pk* n=102 *Pf* n=15 *Pv* n=35		**Day 7** **n=120** *Pk* n=76 *Pf* n=14 *Pv* n=30		**Day 28** **n=120** *Pk* n=80 *Pf* n=13 *Pv* n=27		**Year 1** **n=46** *Pk* n=37 *Pf* n=3 *Pv* n=5	
	SSP2/TRAP	**SERA3 ag2**	**SSP2/TRAP**	**SERA3 ag2**	**SSP2/TRAP**	**SERA3 ag2**	**SSP2/TRAP**	**SERA3 ag2**
Total pos (%)	47 (30.52)	69 (44.81)	74 (61.67)	80 (66.67)	48 (40)	69 (57.5)	3 (6.52)	4 (8.7)
***Pk* diagnosed (%)**	24 (23.53)	32 (31.37)	43 (56.58)	44 (57.89)	31 (38.75)	40 (50)	1 (2.7)	3 (8.11)
***Pf* diagnosed (%)**	9 (60)	10 (66.67)	10 (71.43)	6 (42.86)	5 (38.46)	5 (38.46)	1 (33.33)	0 (0)
***Pv* diagnosed (%)**	14 (40)	27 (77.14)	21 (70)	30 (100)	12 (44.44)	24 (88.89)	1 (20)	1 (20)
Total *Pk* ag pos (%)	84 (54.54)		103 (85.83)		90 (75)		6 (13.04)	

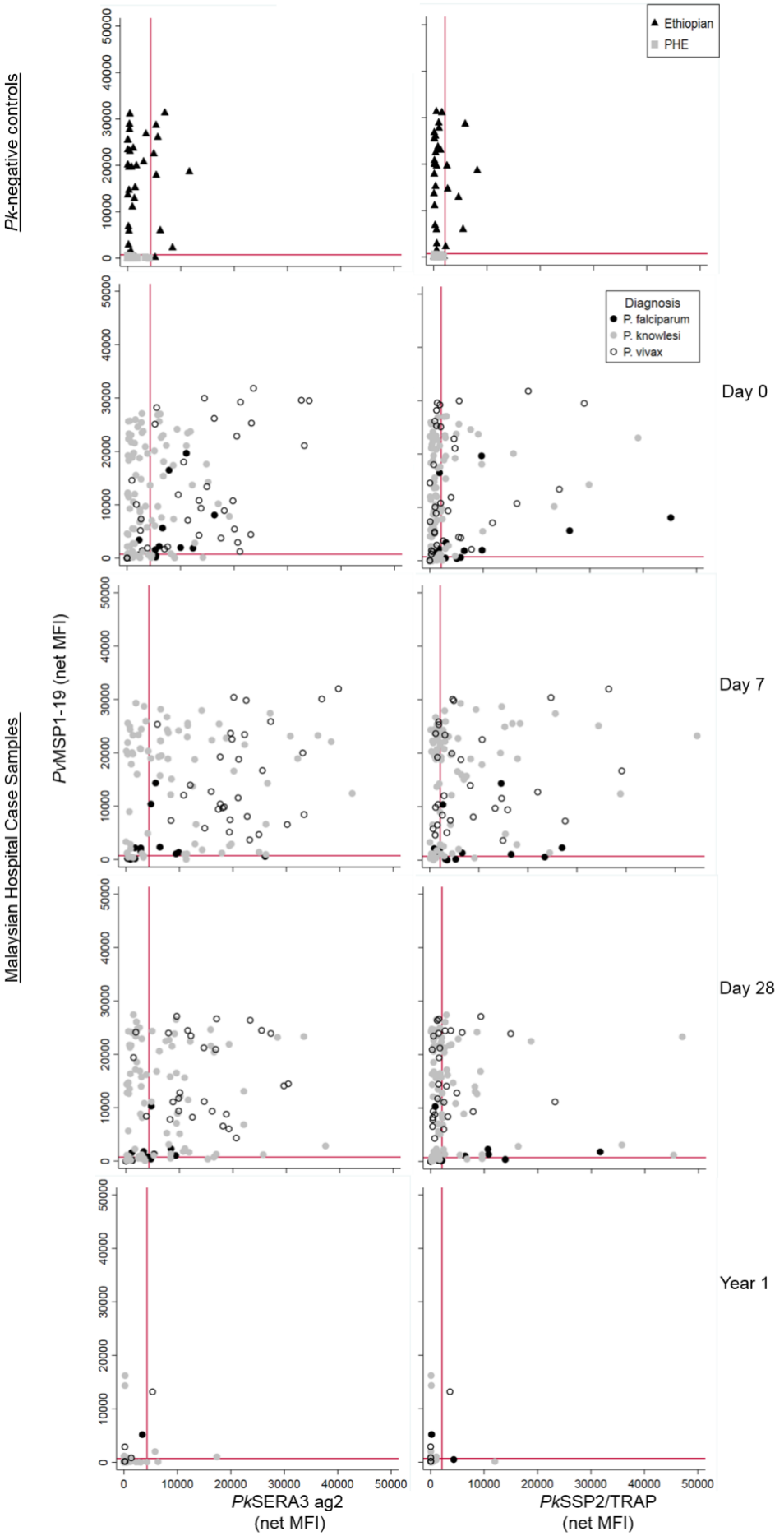

Pk-negative controls
Malaysian Hospital Case Samples
PvMSP1-19 (net MFI)
Ethiopian
PHE
Diagnosis
P. falciparum
P. knowlesi
P. vivax
Day 0
Day 7
Day 28
Year 1
0
10000
20000
30000
40000
50000
PkSERA3 ag2
(net MFI)
PkSSP2/TRAP
(net MFI)

Figure 5.4 Endemic and *P. knowlesi*-negative sera reactivity to *Plasmodium knowlesi*-specific antigens. Scatter plots showing sera reactivity to*: P. vivax* MSP1-19 with *P. knowlesi* SERA3 ag2 (column 1) and SSP2/TRAP (column 2) antigens. Sera samples from *P. knowlesi* negative controls n=58 (row 1; PHE UK malaria naïve (grey squares), Ethiopian children (black triangles)) and Malaysian hospital case sera samples from days 0 (*Pk* n=102; *Pv* n=35; *Pf* n=15), 7 (*Pk* n=76; *Pv* n=30; *Pf* n=14), 28 (*Pk* n=80; *Pv* n=27; *Pf* n=13) and year 1 (*Pk* n=37; *Pv* n=5; *Pf* n=3) of diagnosis (rows 2±5, respectively). Results were divided by diagnostic status of individuals with *P. knowlesi* diagnosed samples represented as grey circles, *P. falciparum* as black circles and *P. vivax* as clear circles. The vertical red line in each graph represents the cut off values for seropositivity to the respective *P. knowlesi* antigen and was calculated based on Public Health England negative control sera samples (average MFI ± (3xSD)). The horizontal cut off line for the *P. vivax* antigen is based on *Pv*MSP1-19=729.5. The vertical cut off line for the *P. knowlesi* antigens were based on the following values: *Pk*SERA3 ag2=4318.31 and *Pk*SSP2/TRAP=2126.35.

Since these samples had been previously tested using the ELISA method, we used this opportunity to compare the results between the two platforms to evaluate reproducibility and sensitivity. We compared only the samples that had been tested on both platforms (total n=144; *P. knowlesi* n=97; *P. vivax* n=31; *P. falciparum* n=16). The Luminex platform showed more sensitivity compared to the ELISA platform, especially *Pk*SSP2 reactivity (***Figure 5.5***). Luminex as the more sensitive assay might explain what is seen in the *P. knowlesi*-negative controls reactivity towards the *P. knowlesi* antigens (***Figure 5.5***).

A small number of the total samples tested on the Luminex platform that were collected as serum samples were also collected as blood spot samples (n=26), making it possible to compare the different types of samples in terms of reactivity. The serum samples clearly showed stronger reactivity profiles compared to the BS samples (***Figure 5.6***). The BS samples had been collected on different, thinner, filter paper instead of the usual thicker type (3MM, Whatman, UK), as described previously (chapter 2, section 2.7.4.), which likely explains why these were so poor in comparison to serum. The technique is reliable when samples are collected in the appropriate manner as previously described by Corran *et. al.*[61]. ***Figure 5.7*** is an example of this analysis using a higher number of samples.

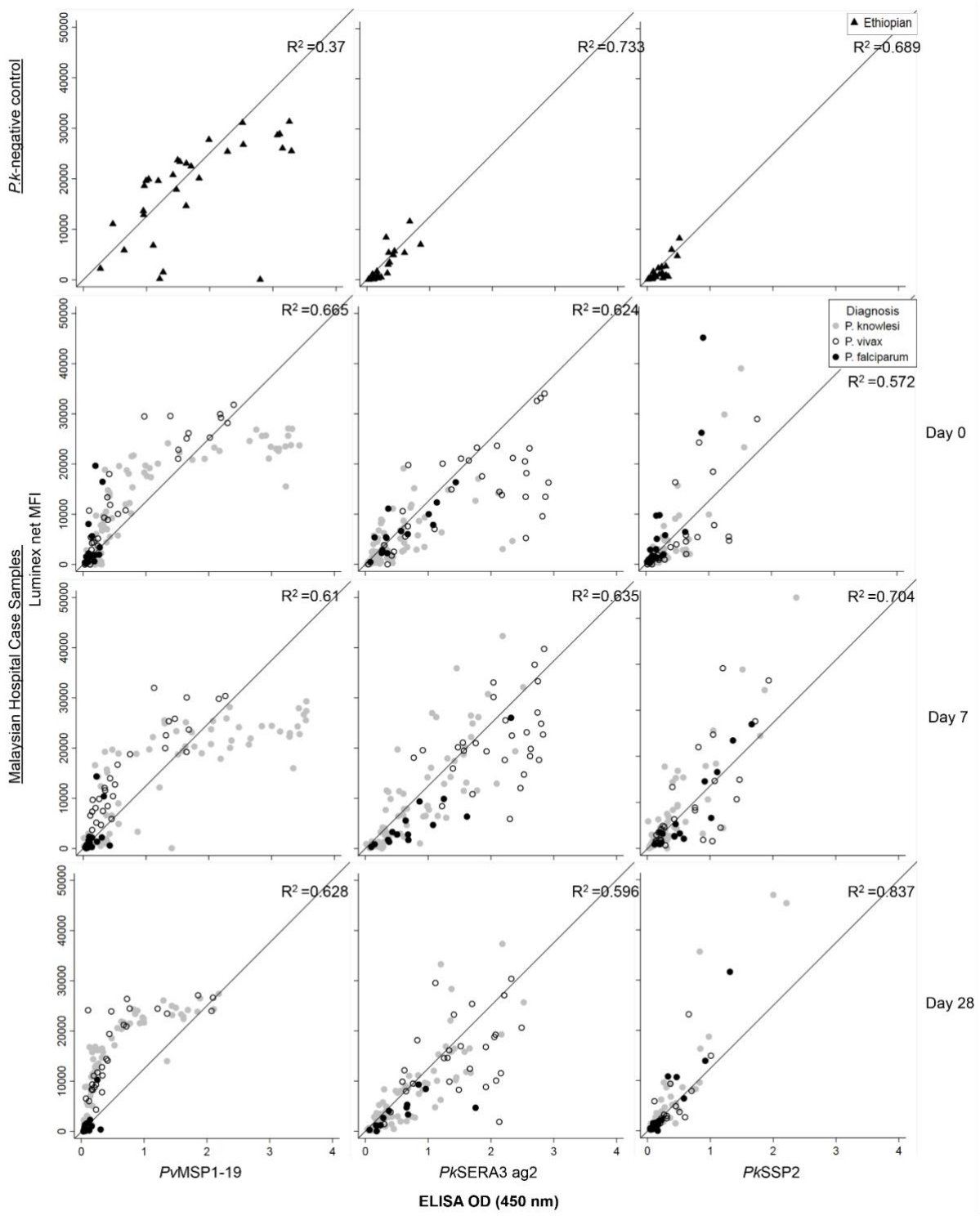


Figure 5.5 Analysis of endemic and *P. knowlesi*-negative sera reactivity to *Plasmodium knowlesi*-specific antigens using the Luminex and ELISA platforms. Scatter plots show sera reactivity between the Luminex (y axis as net MFI) and ELISA (x axis as OD (450 nm)) platforms towards*: Pv*MSP1-19 (column 1), *Pk*SERA3 ag2 (column 2) and *Pk*SSP2/TRAP (column 3) antigens. Sera samples plotted were from *P. knowlesi* negative controls n=40 (row 1; Ethiopian children shown as black triangles) and from Malaysian hospital case sera samples from days 0 (*Pk* n=92; *Pv* n=31; *Pf* n=14), 7 (*Pk* n=72; *Pv* n=28; *Pf* n=14) and 28 (*Pk* n=77; *Pv* n=27; *Pf* n=13) of diagnosis (rows 2±4, respectively). Results were divided by diagnostic status of individuals with *P. knowlesi* diagnosed samples represented as grey circles, *P. falciparum* as black circles and *P. vivax* as clear circles. Trendline values for R^2 are indicated on each plot.

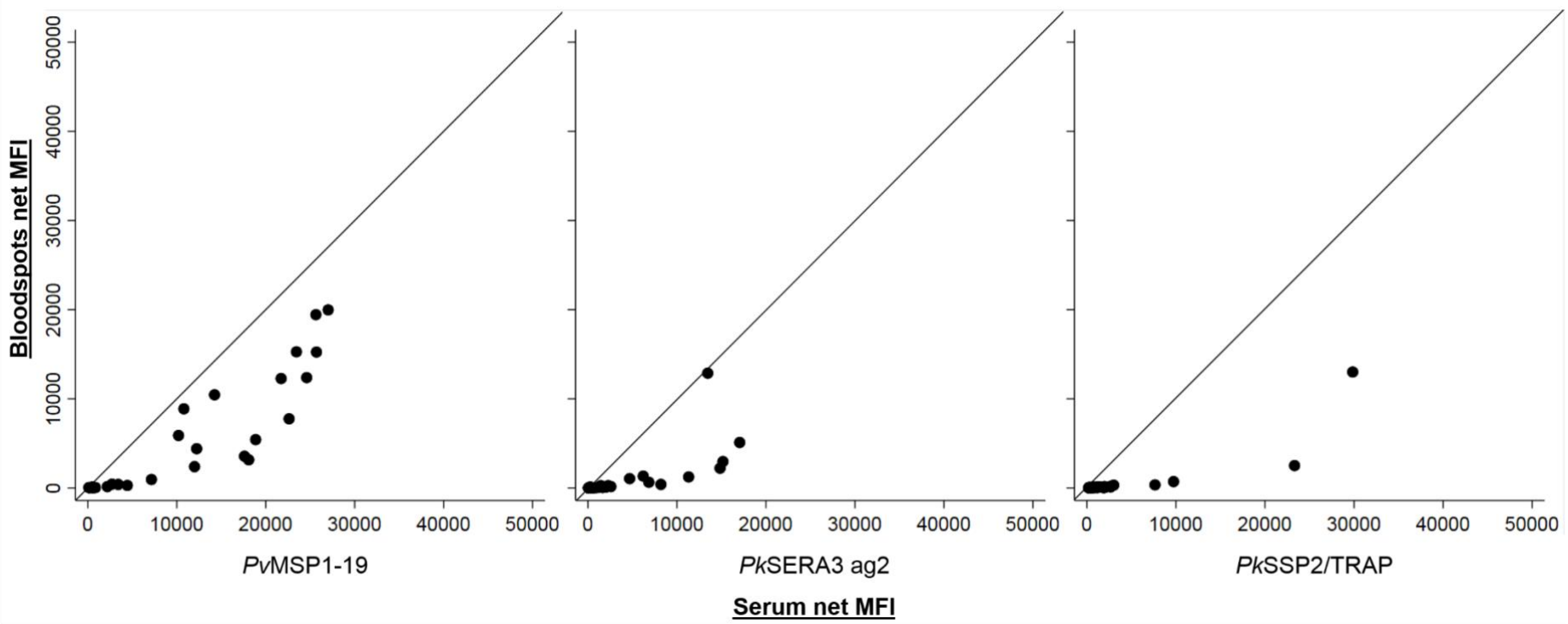


Figure 5.6 Analysis of bloodspot and serum sample reactivity differences to *Plasmodium knowlesi*-specific antigens. Scatter plots show sera reactivity between bloodspot (y axis as net MFI) and serum (x axis as net MFI) samples towards*: Pv*MSP1-19 (column 1), *Pk*SERA3 ag2 (column 2) and *Pk*SSP2/TRAP (column 3) antigens. Sera samples plotted were from Malaysian hospital case samples from day 0 (*Pk* n=24; *Pv* n=1; *Pf* n=1) of diagnosis.

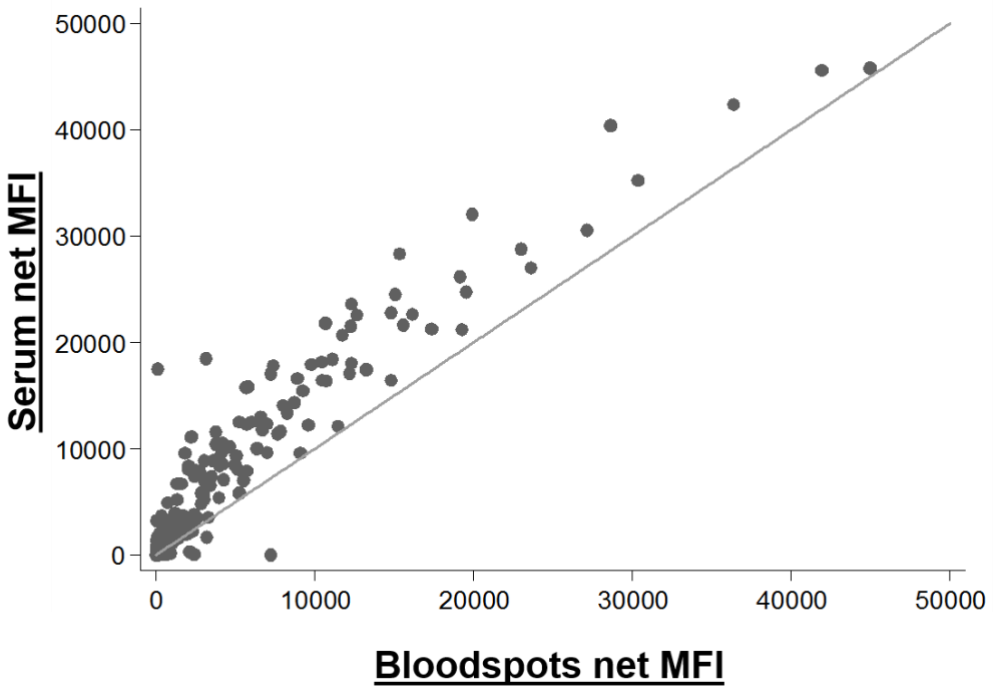


Figure 5.7 Example analysis of bloodspot and serum sample reactivity differences to a Plasmodium-specific antigen. Scatter plots show sera reactivity between bloodspot (x axis as net MFI) and serum (y axis as net MFI) samples towards *a plasmodium* antigen. This plot was used with the permission of Lotus van Den Hoogen as an example comparing bloodspots and serum for a larger sample set.

5.3.2. SABAH COMMUNITY CROSS SECTIONAL SURVEY REACTIVITY TO *P. KNOWLESI*-SPECIFIC RECOMBINANT ANTIGENS USING THE CYTOMETRIC BEAD ASSAY PLATFORM

As with the clinical samples, the Luminex assay was used to screen community samples from a cross sectional survey in Sabah, Malaysia (n=10,118; **Figure *5.3***). The median age of participants was 25 years (age range 2 months - 105 years) and similar proportions of men and women were sampled. Reactivity was seen towards both *P. knowlesi* antigens included in the 37-strong antigen panel, *Pk*SERA3 antigen 2 and *Pk*SSP2 (***Figure 5.8***).

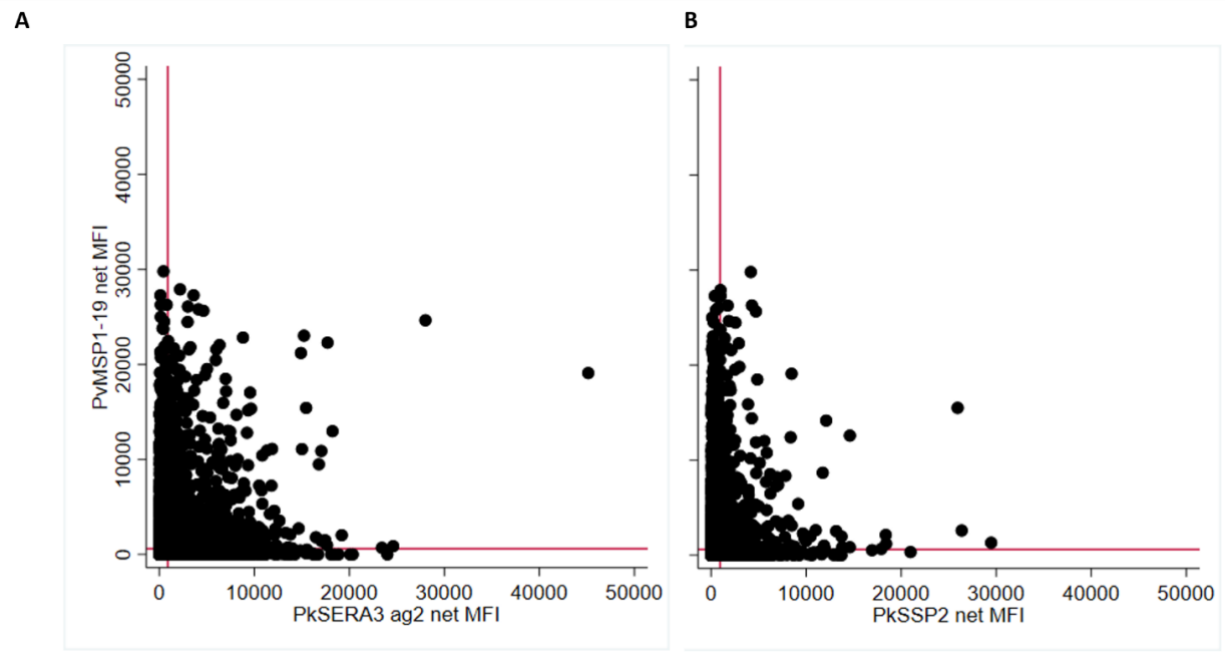


Figure 5.8 Sera reactivity to *Pk*SERA3 antigen 2 and *Pk*SSP2 in the endemic population of Sabah, Malaysia. Scatter plots showing population (n=10,118) sera reactivity to *P. vivax* MSP1-19 with *P. knowlesi* SERA3 ag2 (A) and SSP2 (B). The red lines in each graph represent the cut off values for the *P. knowlesi* and *P. vivax* antigens and were calculated based on the negative population (*P. knowlesi:* average net MFI ± (5xSD); *P. vivax*: average net MFI ± (5xSD)): The vertical cut off lines are based on *Pk*SERA3 ag2=892.93 and *Pk*SSP2=929.35. The horizontal cut off line is based on *Pv*MSP1-19=602.05.

*Pk*AMA1 had been excluded from our *P. knowlesi* antigen panel due to high amino acid sequence homology with *P. vivax* AMA1 (85%; Table 3.1; chapter 3, section 3.1), therefore we did not clone nor express this antigen. However, Dr Bart Faber (Biomedical Primate Research Centre (BPRC), The Netherlands) kindly donated *Pk*AMA1 antigen for use in the Luminex panel. We used this opportunity to look at the reactivity correlations between *Pk*AMA1 and *Pv*AMA1, which showed evidence of some correlation (R^2= 0.5650; **Figure *5.9***). From all samples surveyed, 31% (3139/10125) of the total population were seropositive for *P. knowlesi* SERA3 antigen 2, 12.7% (1286/10124) for SSP2 and 23.09% (2330/10092) for *P. vivax* MSP1-19 (***Figure 5.8***). For both *P. knowlesi* antigens combined, 37.8% (3,827/10,125) of the population were seropositive.

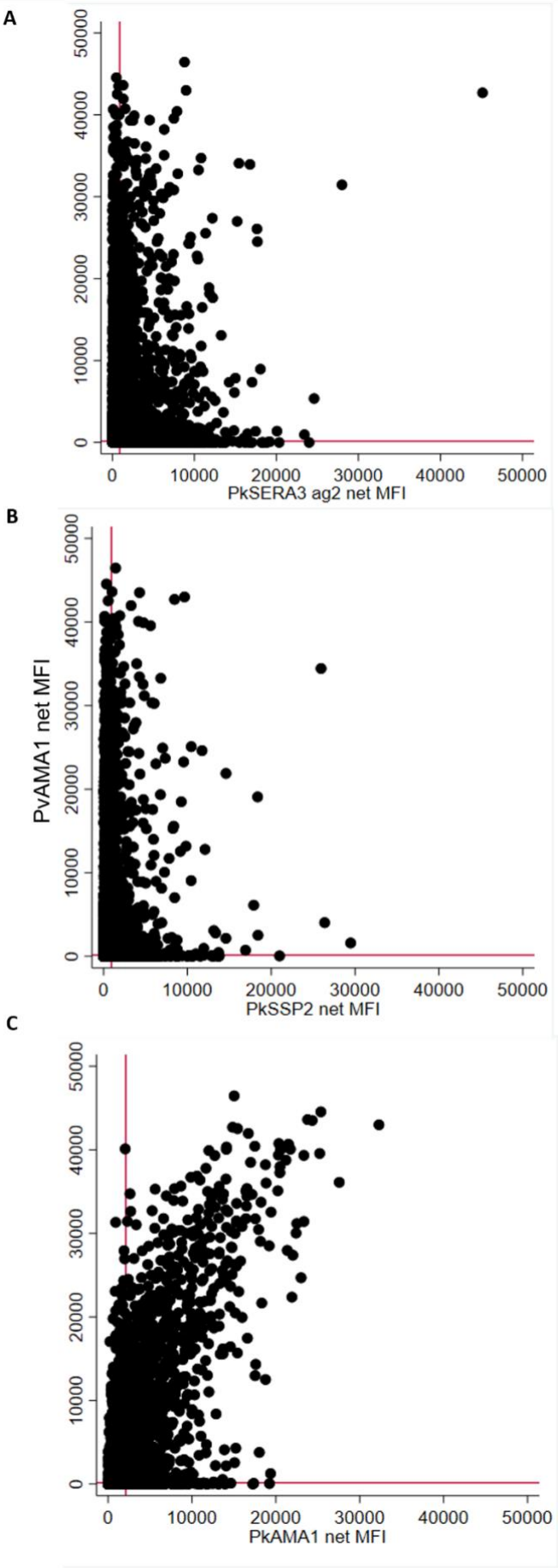


Figure 5.9 Sera reactivity to *Pk*SERA3 antigen 2, *Pk*SSP2 and *Pk*AMA1 in the endemic population of Sabah, Malaysia. Scatter plots showing population (n=10,118) sera reactivity to *P. vivax* AMA1 with *P. knowlesi* SERA3 ag2 (A), SSP2 (B) and AMA1 (C). The red lines in each graph represent the cut off values for the *P. knowlesi* and *P. vivax* antigens and were calculated based on the negative population (*P. knowlesi:* average net MFI ± (5xSD); *P. vivax*: average net MFI ± (5xSD)): The vertical and horizontal cut off lines are based on *Pk*SERA3 ag2=892.93; *Pk*SSP2=929.35 and *Pk*AMA1=2104.63. The horizontal cut off line is based on *Pv*AMA1=147.8.

A similar proportion of men and women seropositive to any *P. knowlesi* antigen (47.24% (1,808/3,827); 52.65% (2,015/3,827) respectively; ***Figure 5.10***). There was no significant association between sex and *P. knowlesi* malaria seropositivity (OR (odds ratio): 1.00, 95% CI: 0.92±1.09, p=0.92).

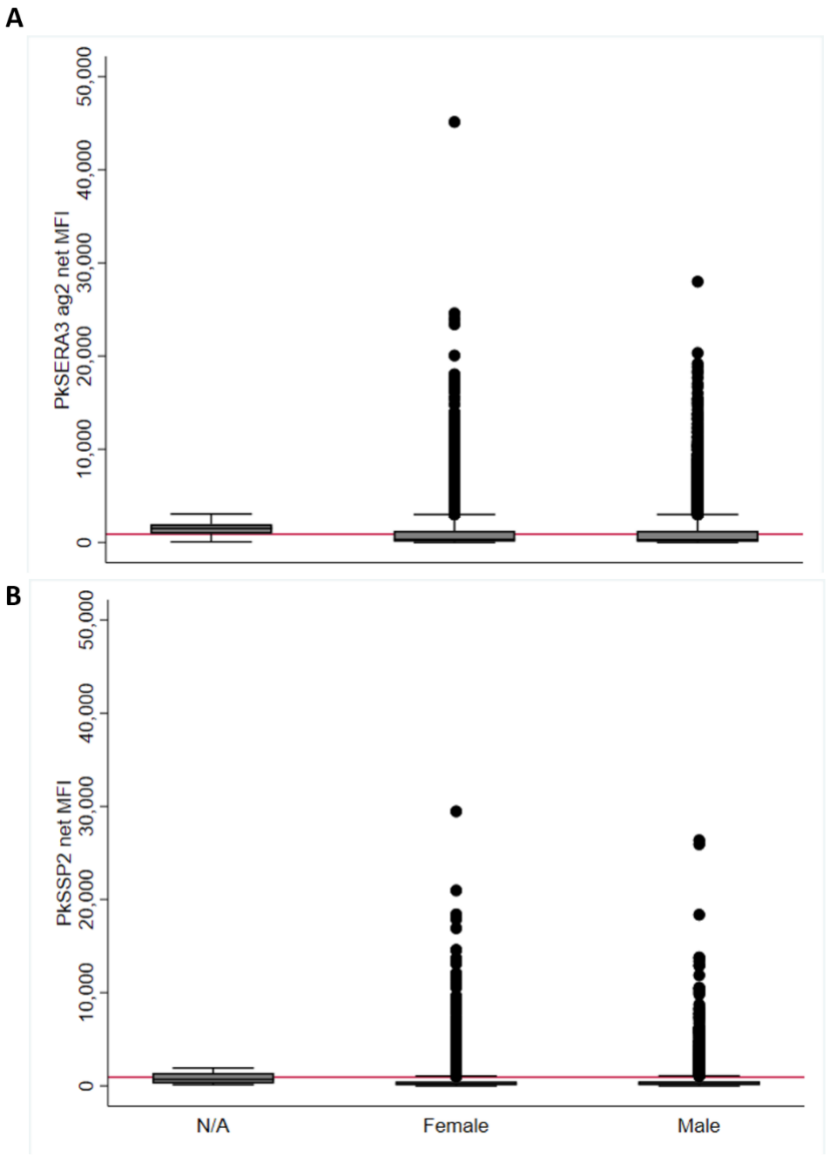


Figure 5.10 Sera reactivity in men and women to *Pk*SERA3 antigen 2 and *Pk*SSP2 in the endemic population of Sabah, Malaysia. Plots showing population (n=10,118) sera reactivity in men and women to *P. knowlesi* SERA3 ag2 (A) and SSP2 (B). Samples without sex information were labelled as not applicable (N/A). The red lines in each graph represent the cut off values for the *P. knowlesi* antigens and were calculated based on the negative population (average net MFI ± (5xSD)): The horizontal cut off lines are based on *Pk*SERA3 ag2=892.93 and *Pk*SSP2=929.35. Risk factors were determined by multivariate logistic regression using STATA.

Antibody reactivity to *Pk*SERA3 antigen 2 was detected in 2.3% (72/3,128) of individuals under 5 years of age and 23.27% (728/3,128) in individuals aged over 55

years (***Figure 5.11*** A), though the highest reactivity was found in individuals between 30-54 (36.22%; 1,133/3,128). Antibody reactivity to *Pk*SSP2 was detected in 2.73% (35/1,284) of individuals under 5 years of age and 27.10% (348/1,284) in individuals aged over 55 years (***Figure 5.11*** B), though the highest reactivity was found in individuals between 30-54 (37.69%; 484/1,284). Age was found to be associated with *P. knowlesi* malaria seropositivity (p<0.0001).

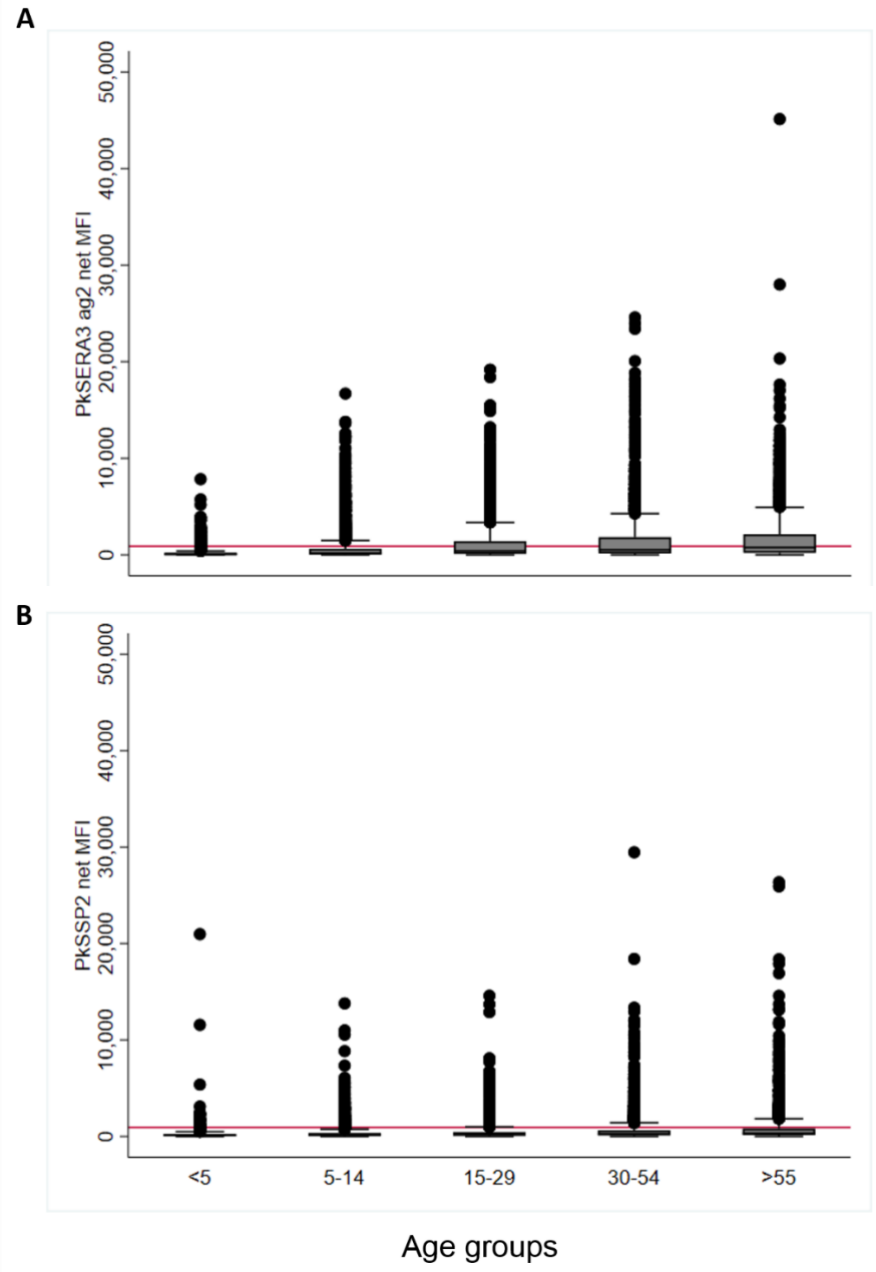


Figure 5.11 Sera reactivity by age groups in the endemic population of Sabah, Malaysia. Plots showing population (n=10,118) sera reactivity to *Pk*SERA3 ag2 (A), *Pk*SSP2 (B), *Pv*MSP1-19 (C) and *Pf*MSP1-19 (D). The red lines in each graph represent the cut off values for the *P. knowlesi* antigens and were calculated based on the

negative population (average net MFI ± (5xSD)): The horizontal cut off lines are based on *Pk*SERA3 ag2=892.93 and *Pk*SSP2=929.35. Risk factors were determined by multivariate logistic regression using STATA.

Of the population that had responded positively to having slept in the forest, 29.41% (10/34) and 20% (4/20) were seropositive to *P. knowlesi* SERA3 antigen 2 and SSP2, respectively (***Figure 5.12*** A and B). *P. vivax* MSP1-19 had 41.38 % (12/29) people responding yes and *P. falciparum* 28.21% (11/39; ***Figure 5.12*** C and D, respectively). It was not possible to determine associations with seropositivity due to the low number of individuals reporting (n=80). People were also questioned about having seen macaques or not during their daily routine. A significant association was observed between seeing macaques and *P. knowlesi* seropositivity (P<0.0001).

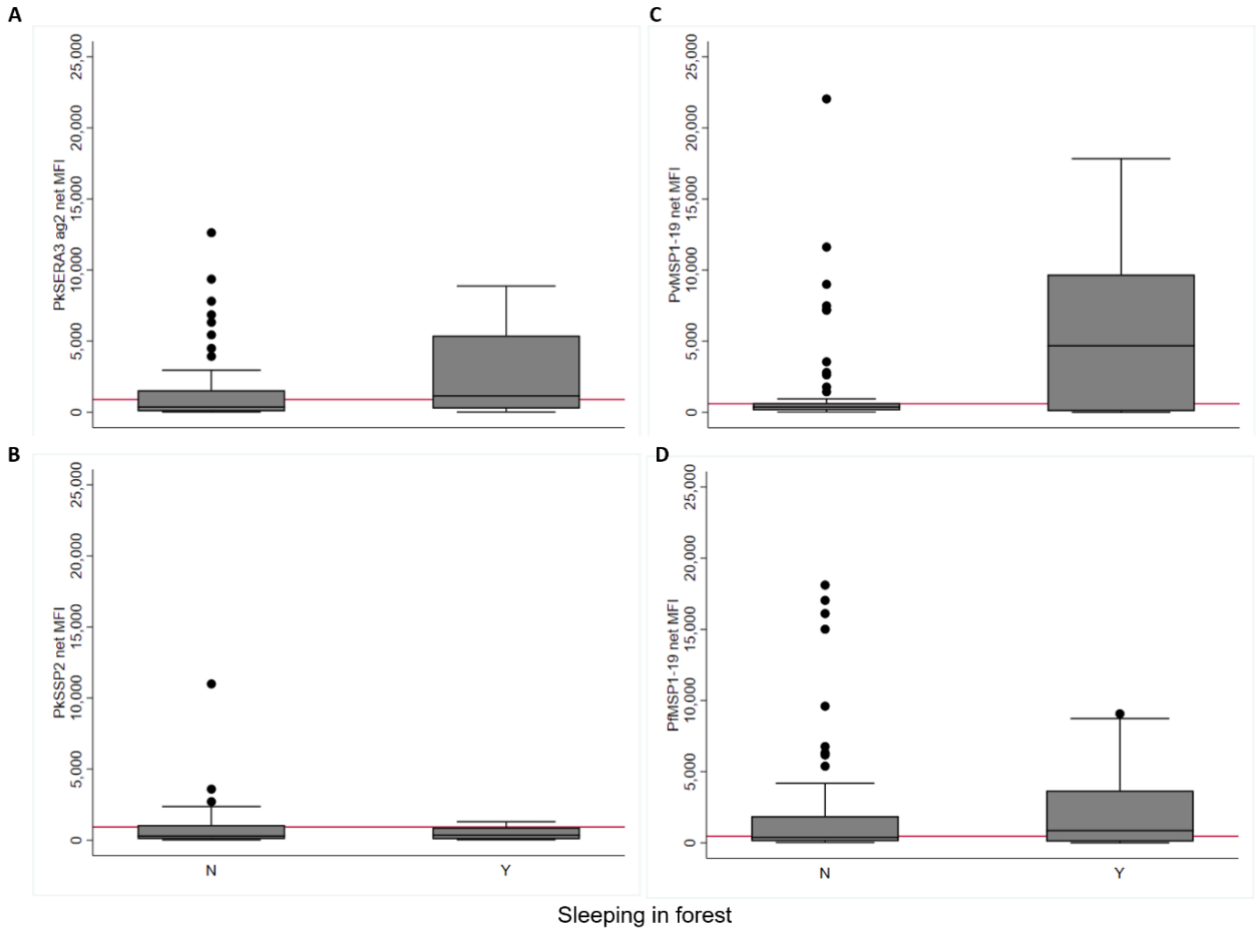


Figure 5.12 Sera reactivity in the endemic population of Sabah, Malaysia, by questionnaire responses towards having slept in the forest or not. Plots show sera reactivity to *Pk*SERA3 ag2 (A), *Pk*SSP2 (B), *Pv*MSP1-19 (C) and *Pf*MSP1-19 (D). Individuals questioned reported not having slept in the forest (N) or having slept in the forest (Y). The red lines in each graph represent the cut off values for the *P. knowlesi, P. vivax* and *P. falciparum* antigens and were calculated based on the negative population (*P. knowlesi:* average net MFI ± (5xSD); *P. vivax* and *P. falciparum*: average net MFI ± (5xSD)): The horizontal cut off lines are based on *Pk*SERA3 ag2=892.93;

*Pk*SSP2=929.35; *Pv*MSP1-19=602.05 and *Pf*MSP1-19=466.9. Risk factors were determined by multivariate logistic regression using STATA.

5.4. DISCUSSION

The results presented here provide important insights in the use of mid-high throughput technology in *P. knowlesi* serology. We have shown for the first time proof of principle that multiplexing of *P. knowlesi* antigens can be used for serological assessment of human samples on a MBA platform.

We found that *Pk*SERA3 antigen 2 was the most highly reactive antigen, which agreed with ELISA results. This was observed in both clinical and community samples. Here, we describe the kinetics of what is potentially a short-term marker of exposure, which would be useful to determine recent infection in the population. The development of such tools would help to answer questions relating to population exposure and our understanding of the geographical boundaries of infection.

Protein-based approaches have been developed in order to determine the presence and functionality of proteins using high throughout platforms[62-64].

Interest in high throughput multiplex technology is rapidly increasing, particularly in the serology field. With the help of MBA technology we can now detect multiple targets simultaneously in a single sample while utilising the least reagent possible. This will enable us to analyse targets in the context of multiple other targets, giving us a better overview of how these might affect how the immune system responds to pathogens.

From the full panel of *P. knowlesi* antigens, only two were selected for the MBA alongside 35 other antigens from other malaria species and pathogens, as these had shown the most promising results by ELISA (Chapter 4). There were difficulties in optimising the remaining *P. knowlesi* antigens for use on the MBA platform due to insufficient controls.

The MBA platform was shown to be more sensitive than the ELISA (***Figure 5.5***) in detecting antibody reactivity towards *Plasmodium* spp. antigens, with more samples showing higher reactivity on the MBA than the ELISA. This indicates that the MBA could potentially be a better platform to analyse samples from areas where submicroscopic malaria can be missed. Although the data presented here fits with the geographic distribution of *P. knowlesi*, the *P. knowlesi*-specific antigens SERA3 antigen 2 and SSP2 showed some level of cross-reactivity in the control samples from Ethiopian school children, although reasonably low (***Figure 5.4***). This had not been observed on the ELISA platform, which again indicates the higher sensitivity of the MBA platform. One of the reasons for this could be due to the less stringent cut off (mean+[3xSD]) used for the analysis of the MBA results, as was done for the initial ELISA analysis (mean+[3xSD])[15]. A more stringent analysis (mean+[5xSD]) could be applied due to the lack of prior data on *P. knowlesi* reactivity analysis, as was done on the small community survey study[65]. It is possible that mimetopes are prevalent in the Ethiopian sample set and that serum antibodies specific to these mimetopes are reacting to the *P. knowlesi* recombinant antigens. Mimetopes are peptide sequences that mimic the immunological response from an antigen's epitope, although not sequence homologous. In order to verify this, it would be necessary to dissect the antigens where we see this reactivity by creating overlapping short sequences, or tiled epitopes, and screening each of these against the specific reactive individuals, identifying the epitope potentially eliciting this response. Another reason could be that the low homology that was found between the sequences of these candidate antigens and *P. vivax* (identities: SERA3 antigen 1 23/58 (40%); SSP2 fragment 40/93 (43%)) is in fact enough to cause some cross-reactivity to be seen in a *P. knowlesi* non-endemic setting (Table 4.1, chapter 4, section 4.2.1). A larger number of a more diverse non-endemic sample set would be needed to better understand what was observed here. The MBA data has shown

that specific tools for *P. vivax* and *P. falciparum* need to be developed for population studies where these are co-endemic with, in this case, *P. knowlesi*, as this would have helped in ensuring specificity. The lack of appropriate endemic negative controls was a limitation to the clinical study.

In the large Sabah community survey *P. knowlesi* seroprevalence was higher than *P. vivax* seroprevalence. *P. knowlesi* risk was associated with an increase in age as well as having seen macaques during an individual's normal routine. Although a preliminary analysis was presented in this chapter, extensive epidemiological analysis of this cross-sectional survey is needed to identify ecological and behavioural risk factors across a broader geographic scale.

ELISA data has previously shown that IgG reactivity towards *Pk*SERA3 ag2 decreases slightly by day 28 (reference[15] and chapter 4), as was shown by MBA analysis, with barely any reactivity by 1 year post-diagnosis (***Figure 5.4***). For a more in-depth analysis of antigens across time, more time points would be required to be collected and analysed for each individual in order to have a better idea of the antigen kinetics and the potential associations with protection for instance. Studies have shown that IgG antibodies against *P. falciparum* antigens are important for the development of clinical immunity[66,67], something that could potentially be seen for *P. knowlesi* infections in endemic populations. Larger sample sets from endemic regions would need to be screened for isotype reactivity in order to determine if these or other correlations exist for *P. knowlesi* as well. There has been much interest in recent years in trying to establish whether serological markers after challenge can be useful predictors of protection particularly in the field of vaccinology[68]. The relationship between antibody titres and protection is currently not quite clear.

In future, it will be necessary to determine bead stability for each antigen under different transport and storage conditions and antigen-specific kinetics in different endemic settings will also need to be considered when using this platform. Standard

methods to determine seropositive cut off thresholds for novel antigens will also need to be established, ensuring reproducibility and comparability.

Here we have shown the application of high throughput MBA technology in the serological screening of *P. knowlesi* endemic populations, both at the clinical and the community level. The application of this method in epidemiological analysis will not only reduce the need of high sample volumes when it is hard to obtain, but will also enable the measurement of all analytes deemed important in a single assay, reducing time and costs.

This only marks the beginning of what these advancements will enable us to do, such as screening for reactivity patterns towards drug treatments and vaccines, potentially making it possible to produce personalised drugs, vaccines and treatments targeted towards certain populations or subsets, specific age groups, genders.

REFERENCES

1 Wipasa, J. *et al.* Long-lived antibody and B Cell memory responses to the human malaria parasites, Plasmodium falciparum and Plasmodium vivax. *PLoS Pathog* **6**, e1000770, doi:10.1371/journal.ppat.1000770 (2010).

2 Akpogheneta, O. J. *et al.* Duration of naturally acquired antibody responses to blood-stage Plasmodium falciparum is age dependent and antigen specific. *Infection and immunity* **76**, 1748-1755, doi:10.1128/iai.01333-07 (2008).

3 Fowkes, F. J. *et al.* New insights into acquisition, boosting, and longevity of immunity to malaria in pregnant women. *The Journal of infectious diseases* **206**, 1612-1621, doi:10.1093/infdis/jis566 (2012).

4 Meraldi, V. *et al.* Natural antibody response to Plasmodium falciparum Exp-1, MSP-3 and GLURP long synthetic peptides and association with protection. *Parasite immunology* **26**, 265-272, doi:10.1111/j.0141-9838.2004.00705.x (2004).

5 Osier, F. H. *et al.* Breadth and magnitude of antibody responses to multiple Plasmodium falciparum merozoite antigens are associated with protection from clinical malaria. *Infection and immunity* **76**, 2240-2248, doi:10.1128/iai.01585-07 (2008).

6 Stanisic, D. I. *et al.* Acquisition of antibodies against Plasmodium falciparum merozoites and malaria immunity in young children and the influence of age, force of infection, and magnitude of response. *Infection and immunity* **83**, 646-660, doi:10.1128/iai.02398-14 (2015).

7 Helb, D. A. *et al.* Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E4438-4447, doi:10.1073/pnas.1501705112 (2015).

8 Cook, J. *et al.* Trends in parasite prevalence following 13 years of malaria interventions on Bioko island, Equatorial Guinea: 2004–2016. *Malaria journal* **17**, 62, doi:10.1186/s12936-018-2213-9 (2018).

9 Tadesse, F. G. *et al.* The shape of the iceberg: quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia. *Malaria journal* **16**, 99, doi:10.1186/s12936-017-1749-4 (2017).

10 van den Hoogen, L. L. *et al.* Serology describes a profile of declining malaria transmission in Farafenni, The Gambia. *Malaria journal* **14**, 416, doi:10.1186/s12936-015-0939-1 (2015).

11 Dewasurendra, R. L. *et al.* Effectiveness of a serological tool to predict malaria transmission intensity in an elimination setting. *BMC Infect Dis* **17**, 49, doi:10.1186/s12879-016-2164-0 (2017).

12 Rek, J. *et al.* Characterizing microscopic and submicroscopic malaria parasitaemia at three sites with varied transmission intensity in Uganda. *Malaria journal* **15**, 470, doi:10.1186/s12936-016-1519-8 (2016).

13 Katrak, S. *et al.* Clinical consequences of submicroscopic malaria parasitaemia in Uganda. *Malaria journal* **17**, 67, doi:10.1186/s12936-018-2221-9 (2018).

14 Fornace, K. M. *et al.* Asymptomatic and submicroscopic carriage of Plasmodium knowlesi malaria in household and community members of clinical cases in Sabah, Malaysia. *The Journal of infectious diseases*, doi:10.1093/infdis/jiv475 (2015).

15 Herman, L. S. *et al.* Identification and validation of a novel panel of Plasmodium knowlesi biomarkers of serological exposure. *PLoS neglected tropical diseases* **12**, e0006457, doi:10.1371/journal.pntd.0006457 (2018).

16 Liu, F. *et al.* Characterization of Plasmodium berghei Pbg37 as Both Pre- and Post-Fertilization Antigen with Transmission-Blocking Potential. *Infection and immunity*, doi:10.1128/iai.00785-17 (2018).
17 Muh, F. *et al.* Identification of a novel merozoite surface antigen of Plasmodium vivax, PvMSA180. *Malaria journal* **16**, 133, doi:10.1186/s12936-017-1760-9 (2017).
18 Zheng, W. *et al.* Functional characterization of Plasmodium berghei PSOP25 during ookinete development and as a malaria transmission-blocking vaccine candidate. *Parasites & vectors* **10**, 8, doi:10.1186/s13071-016-1932-4 (2017).
19 Aguiar, J. C. *et al.* Discovery of Novel Plasmodium falciparum Pre-Erythrocytic Antigens for Vaccine Development. *PLoS One* **10**, e0136109, doi:10.1371/journal.pone.0136109 (2015).
20 Cheng, Y. *et al.* Antigenicity and immunogenicity of PvRALP1, a novel Plasmodium vivax rhoptry neck protein. *Malaria journal* **14**, 186, doi:10.1186/s12936-015-0698-z (2015).
21 Han, J. H. *et al.* Identification of Immunodominant B-cell Epitope Regions of Reticulocyte Binding Proteins in Plasmodium vivax by Protein Microarray Based Immunoscreening. *The Korean journal of parasitology* **53**, 403-411, doi:10.3347/kjp.2015.53.4.403 (2015).
22 Hostetler, J. B. *et al.* A Library of Plasmodium vivax Recombinant Merozoite Proteins Reveals New Vaccine Candidates and Protein-Protein Interactions. *PLoS neglected tropical diseases* **9**, e0004264, doi:10.1371/journal.pntd.0004264 (2015).
23 Saetear, P. *et al.* Quantification of Plasmodium-host protein interactions on intact, unmodified erythrocytes by back-scattering interferometry. *Malaria journal* **14**, 88, doi:10.1186/s12936-015-0553-2 (2015).
24 Seideman, J. & Peritt, D. A novel monoclonal antibody screening method using the Luminex-100 microsphere system. *Journal of immunological methods* **267**, 165-171 (2002).
25 Mandy, F. F., Nakamura, T., Bergeron, M. & Sekiguchi, K. Overview and application of suspension array technology. *Clinics in laboratory medicine* **21**, 713-729, vii (2001).
26 Nolan, J. P. & Sklar, L. A. Suspension array technology: evolution of the flat-array paradigm. *Trends in biotechnology* **20**, 9-12 (2002).
27 Waterboer, T., Sehr, P. & Pawlita, M. Suppression of non-specific binding in serological Luminex assays. *Journal of immunological methods* **309**, 200-204, doi:10.1016/j.jim.2005.11.008 (2006).
28 Lammie, P. J. *et al.* Development of a new platform for neglected tropical disease surveillance. *Int J Parasitol* **42**, 797-800, doi:10.1016/j.ijpara.2012.07.002 (2012).
29 Collins, D. P., Luebering, B. J. & Shaut, D. M. T-lymphocyte functionality assessed by analysis of cytokine receptor expression, intracellular cytokine expression, and femtomolar detection of cytokine secretion by quantitative flow cytometry. *Cytometry* **33**, 249-255 (1998).
30 Purohit, S., Sharma, A. & She, J. X. Luminex and other multiplex high throughput technologies for the identification of, and host response to, environmental triggers of type 1 diabetes. *Biomed Res Int* **2015**, 326918, doi:10.1155/2015/326918 (2015).
31 Lal, G., Balmer, P., Joseph, H., Dawson, M. & Borrow, R. Development and Evaluation of a Tetraplex Flow Cytometric Assay for Quantitation of Serum Antibodies to Neisseria meningitidis Serogroups A, C, Y, and W-135. *Clinical and Diagnostic Laboratory Immunology* **11**, 272-279, doi:10.1128/CDLI.11.2.272-279.2004 (2004).

32 Lal, G. *et al.* Development and validation of a nonaplex assay for the simultaneous quantitation of antibodies to nine Streptococcus pneumoniae serotypes. *Journal of immunological methods* **296**, 135-147, doi:10.1016/j.jim.2004.11.006 (2005).

33 Pickering, J. W., Martins, T. B., Schroder, M. C. & Hill, H. R. Comparison of a Multiplex Flow Cytometric Assay with Enzyme-Linked Immunosorbent Assay for Quantitation of Antibodies to Tetanus, Diphtheria, and Haemophilus influenzae Type b. *Clinical and Diagnostic Laboratory Immunology* **9**, 872-876, doi:10.1128/CDLI.9.4.872-876.2002 (2002).

34 Biagini, R. E. *et al.* Comparison of a multiplexed fluorescent covalent microsphere immunoassay and an enzyme-linked immunosorbent assay for measurement of human immunoglobulin G antibodies to anthrax toxins. *Clin Diagn Lab Immunol* **11**, 50-55 (2004).

35 Dasso, J., Lee, J., Bach, H. & Mage, R. G. A comparison of ELISA and flow microsphere-based assays for quantification of immunoglobulins. *Journal of immunological methods* **263**, 23-33 (2002).

36 Leng, S. X. *et al.* ELISA and multiplex technologies for cytokine measurement in inflammation and aging research. *The journals of gerontology. Series A, Biological sciences and medical sciences* **63**, 879-884 (2008).

37 Basile, A. J. *et al.* Multiplex microsphere immunoassays for the detection of IgM and IgG to arboviral diseases. *PLoS One* **8**, e75670, doi:10.1371/journal.pone.0075670 (2013).

38 Perraut, R. *et al.* Comparative analysis of IgG responses to Plasmodium falciparum MSP1p19 and PF13-DBL1alpha1 using ELISA and a magnetic bead-based duplex assay (MAGPIX(R)-Luminex) in a Senegalese meso-endemic community. *Malaria journal* **13**, 410, doi:10.1186/1475-2875-13-410 (2014).

39 Yu, J., Lin, J., Kim, K.-H., Benjamin, W. H. & Nahm, M. H. Development of an Automated and Multiplexed Serotyping Assay for Streptococcus pneumoniae. *Clinical and vaccine immunology : CVI* **18**, 1900-1907, doi:10.1128/CVI.05312-11 (2011).

40 Croft, H. *et al.* Use of Luminex xMAP-derived Bio-Plex bead-based suspension array for specific detection of PPV W and characterization of epitopes on the coat protein of the virus. *Journal of virological methods* **153**, 203-213, doi:10.1016/j.jviromet.2008.07.016 (2008).

41 Opalka, D. *et al.* Simultaneous Quantitation of Antibodies to Neutralizing Epitopes on Virus-Like Particles for Human Papillomavirus Types 6, 11, 16, and 18 by a Multiplexed Luminex Assay. *Clinical and Diagnostic Laboratory Immunology* **10**, 108-115, doi:10.1128/CDLI.10.1.108-115.2003 (2003).

42 Anderson, S., Wakeley, P., Wibberley, G., Webster, K. & Sawyer, J. Development and evaluation of a Luminex multiplex serology assay to detect antibodies to bovine herpes virus 1, parainfluenza 3 virus, bovine viral diarrhoea virus, and bovine respiratory syncytial virus, with comparison to existing ELISA detection methods. *Journal of immunological methods* **366**, 79-88, doi:10.1016/j.jim.2011.01.010 (2011).

43 Liang, L. *et al.* Identification of potential serodiagnostic and subunit vaccine antigens by antibody profiling of toxoplasmosis cases in Turkey. *Molecular & cellular proteomics : MCP* **10**, M110.006916, doi:10.1074/mcp.M110.006916 (2011).

44 Felgner, J. *et al.* Identification of Toxoplasma gondii antigens associated with different types of infection by serum antibody profiling. *Parasitology* **142**, 827-838, doi:10.1017/s0031182014001978 (2015).

45 Holec-Gasior, L. Toxoplasma gondii recombinant antigens as tools for serodiagnosis of human toxoplasmosis: current status of studies. *Clinical and vaccine immunology : CVI* **20**, 1343-1351, doi:10.1128/cvi.00117-13 (2013).

46 Fouda, G. G. *et al.* Multiplex assay for simultaneous measurement of antibodies to multiple Plasmodium falciparum antigens. *Clinical and vaccine immunology : CVI* **13**, 1307-1313, doi:10.1128/cvi.00183-06 (2006).

47 Rogier, E. *et al.* Evaluation of Immunoglobulin G Responses to Plasmodium falciparum and Plasmodium vivax in Malian School Children Using Multiplex Bead Assay. *Am J Trop Med Hyg* **96**, 312-318, doi:10.4269/ajtmh.16-0476 (2017).

48 Ambrosino, E. *et al.* A multiplex assay for the simultaneous detection of antibodies against 15 Plasmodium falciparum and Anopheles gambiae saliva antigens. *Malaria journal* **9**, 317, doi:10.1186/1475-2875-9-317 (2010).

49 Ubillos, I., Campo, J. J., Jimenez, A. & Dobano, C. Development of a high-throughput flexible quantitative suspension array assay for IgG against multiple Plasmodium falciparum antigens. *Malaria journal* **17**, 216, doi:10.1186/s12936-018-2365-7 (2018).

50 Valmaseda, A. *et al.* Identifying Immune Correlates of Protection Against Plasmodium falciparum Through a Novel Approach to Account for Heterogeneity in Malaria Exposure. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **66**, 586-593, doi:10.1093/cid/cix837 (2018).

51 Waterboer, T. *et al.* Multiplex Human Papillomavirus Serology Based on In Situ–Purified Glutathione S-Transferase Fusion Proteins. *Clinical chemistry* **51**, 1845, doi:10.1373/clinchem.2005.052381 (2005).

52 Carson, R. T. & Vignali, D. A. A. Simultaneous quantitation of 15 cytokines using a multiplexed flow cytometric assay. *Journal of immunological methods* **227**, 41-52, doi:https://doi.org/10.1016/S0022-1759(99)00069-1 (1999).

53 Balne, P. K. *et al.* Bead Based Multiplex Assay for Analysis of Tear Cytokine Profiles. *Journal of visualized experiments : JoVE*, doi:10.3791/55993 (2017).

54 Patro, A. R. K. *et al.* Cytokine Signature Associated with Disease Severity in Dengue. *Viruses* **11**, doi:10.3390/v11010034 (2019).

55 Armstrong, B., Stewart, M. & Mazumder, A. Suspension arrays for high throughput, multiplexed single nucleotide polymorphism genotyping. *Cytometry* **40**, 102-108 (2000).

56 Cowan, L. S., Diem, L., Brake, M. C. & Crawford, J. T. Transfer of a Mycobacterium tuberculosis Genotyping Method, Spoligotyping, from a Reverse Line-Blot Hybridization, Membrane-Based Assay to the Luminex Multianalyte Profiling System. *Journal of clinical microbiology* **42**, 474, doi:10.1128/JCM.42.1.474-477.2004 (2004).

57 Ye, F. *et al.* Fluorescent microsphere-based readout technology for multiplexed human single nucleotide polymorphism analysis and bacterial identification. *Human mutation* **17**, 305-316, doi:10.1002/humu.28 (2001).

58 Grignard, L. *et al.* Bead-based assays to simultaneously detect multiple human inherited blood disorders associated with malaria. *Malaria journal* **18**, 14, doi:10.1186/s12936-019-2648-7 (2019).

59 Ondigo, B. N. *et al.* Standardization and validation of a cytometric bead assay to assess antibodies to multiple Plasmodium falciparum recombinant antigens. *Malaria journal* **11**, 427, doi:10.1186/1475-2875-11-427 (2012).

60 Drakeley, C. J. *et al.* Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5108-5113, doi:10.1073/pnas.0408725102 (2005).

61 Corran, P. H. *et al.* Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. *Malaria journal* **7**, 195, doi:10.1186/1475-2875-7-195 (2008).
62 Zhu, H. & Snyder, M. Protein arrays and microarrays. *Current opinion in chemical biology* **5**, 40-45 (2001).
63 Kopf, E. & Zharhary, D. Antibody arrays--an emerging tool in cancer proteomics. *Int J Biochem Cell Biol* **39**, 1305-1317, doi:10.1016/j.biocel.2007.04.029 (2007).
64 Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623-627, doi:10.1038/35001009 (2000).
65 Fornace, K. M. *et al.* Exposure and infection to Plasmodium knowlesi in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. *PLoS neglected tropical diseases* **12**, e0006432, doi:10.1371/journal.pntd.0006432 (2018).
66 Murungi, L. M. *et al.* Targets and Mechanisms Associated with Protection from Severe Plasmodium falciparum Malaria in Kenyan Children. *Infection and immunity* **84**, 950-963, doi:10.1128/iai.01120-15 (2016).
67 Adamou, R. *et al.* Plasmodium falciparum infection and age influence parasite growth inhibition mediated by IgG in Beninese infants. *Acta tropica* **159**, 111-119, doi:10.1016/j.actatropica.2016.03.020 (2016).
68 Plotkin, S. A. Correlates of Protection Induced by Vaccination. *Clinical and vaccine immunology : CVI* **17**, 1055-1065, doi:10.1128/CVI.00131-10 (2010).

Chapter 6 : HIGH THROUGHPUT SEROLOGICAL PROFILING OF ANTIBODY ISOTYPES IN MALAYSIAN HOSPITAL CASE SAMPLES USING A BESPOKE PROTEIN MICROARRAY PLATFORM

6.1. INTRODUCTION

Over the years, the detection of antibodies towards *Plasmodium* spp. has been performed using especially developed serological methods in order to determine disease exposure over time[1-7]. These methods are currently widely accepted, however, focus has been on *P. falciparum* and *P. vivax* parasites, mainly looking at total IgG antibodies[4,7-12]. Currently there is a lack of data regarding IgM and IgA isotype responses in malaria research, particularly for *P. knowlesi.* Antibody and isotype data is important to characterise the human host immune response towards malarial infection and to make important associations of risk and protection.

IgG antibody responses to malaria infection have been extensively studied by a multitude of research groups[5,13-17], establishing that reactivity towards this isotype is associated with protection against malaria. In a longitudinal clinical and parasitological study in Myanmar, protection against clinical *P. falciparum* malaria infection was found to correlate with the level of IgG3 responses elicited by *Pf*MSP3 and *Pf*GLURP (Glutamate Rich Protein), whilst IgG4 antibodies increased in number with every malaria attack[13], with an attack being defined as elevated temperature with high parasitemia compared to asymptomatic individuals and separated by >72 hours. Similar results had been seen in Senegal[14], Burkina Faso[17] and more recently in Ghana[15,16], suggesting that IgG subclass responses to *Pf*GLURP potentially promote a moderate development of protection in endemic areas in Asia and Africa. It was demonstrated by Osier *et. al*.[5] that high levels of naturally acquired IgG antibodies to *P. falciparum* AMA1, MSP1 block 2, MSP2 and MSP3 antigens in Kenyan children were related to the decreased probability of developing malaria[5]. Osier *et. al*.[5] also showed that malaria risk was inversely associated with a high number of antigen targets recognised by antibodies. It has also been possible to determine recent and historic exposure[6,7,18] to malaria through the measurement of IgG reactivity, as well as transmission intensity[9,19-22] and age associations[2,22,23] such

as the increase of antibody responses with age[6]. The measurement of antibodies can be used as proxies for these other measurements, acting as biomarkers of malaria exposure.

Although less extensively studied, IgM was also found to be associated with clinical immunity to malaria infection. Using standardised Afro-Immuno Assay (AIA) network ELISA procedures[17,24], it was found that IgM and IgG antibody responses to *Pf*MSP1-19 and *Pf*MSP3 were associated with protection to malaria in Ghanaian children[24]. Similar IgG responses were seen in children from Burkina Faso which were associated with reduced risk of malaria as well as age, indicating cumulative exposure over time[17]. The AIA network is a project that has developed immunoassays which ensure that the reagents, protocols and statistical methods used to assess associations with protection are the same, making these types of studies more comparable[17,24]. Studies have also looked at IgM response differences between different ethnic groups in Mali, enabling distinction of these groups from one another based on these responses[25,26]. IgM responses were significantly higher in the Fulani people compared to the Dogon[25,26]. It was suggested that host genetic polymorphisms yielding stronger IgM responses in certain ethnic groups was the result of selective pressure due to malaria. Antibody isotype and subclass responses to malaria infection were shown to be genetically regulated, with some isotypes shown to have higher heritability than others in children[27], making it possible to eventually identify the genes responsible for up or downregulating certain isotypes. Overall, IgG is regarded as one of the most important isotypes when it comes to protective immunity towards malaria infection, but recently IgM has shown increasing significance and interest.

In contrast to IgG and IgM, IgA is one of the least studied isotypes in the field of malaria, with most information coming from malaria-caused nephropathy cases[28,29]. One of the complications in *P. falciparum* infections is acute renal failure[30], with glomerulopathy being a less common form of renal failure[29,31]. The most common

form of glomerulonephritis is IgA nephropathy, associated with glomerular IgA1 subclass deposits[32,33], suggested to be caused by aberrantly glycosylated polymeric IgA1 complexes[34].

There have been few non-nephropathy related studies. One study showed that IgA responses appear to follow IgM and IgG responses[35]. Another looked at *Pf*MSP1-19-specific recombinant IgA activity *in vitro* and *in vivo*, showing that these recognised parasites in infected RBC *in vitro,* although protection was not observed in human FcαRI transgenic mice[36] (specific cell surface receptor which binds to Fc region of specific antibodies). A similar study showed that IgG1 was protective in human FcγRI transgenic mice after challenge with lethal doses of *Pf*MSP1-19-transgenic *P. berghei*, demonstrating the importance of such receptors in triggering protection[37]. A separate study had observed that IgA1, IgA2, IgG2 and IgG4 could not activate antibody-dependant cellular inhibition of *P. falciparum*, with only IgG1 and IgG3 showing this ability[38]. It was also observed that IgA-associated immune complexes appeared to be protective in children with cerebral malaria[39]. IgA serum levels towards a *P. falciparum* antigen extract were found to increase with age in seropositive individuals from an endemic population, and it was shown to inhibit *P. falciparum* growth *in vitro*[40]. Very little is known about the importance and activity of IgA antibodies in malaria; contradicting data have been published on this topic. It is important to measure and interpret the levels of these different components and their relationships to one another as these are the foundations of a good understanding of the immune system in different health states. These can help monitor the course of certain infectious diseases in order to characterise the immune response towards the disease and the outcome of various treatments.

Routinely, the measurement of exposure to malaria in endemic populations is carried out using a small number of well characterised plasmodium antigens[20,21,41,42], though until now, there have not been any species-specific antigens for *P. knowlesi*. With the newly available *P. knowlesi* antigens we have developed, it will be possible to

further characterise this organism and the human host response to it[43]. Novel proteins are being identified at an increasing rate[44-51] and developing a platform appropriate to screen multiple targets has become essential in order to obtain a better understanding of host-pathogen interactions and antibody-mediated immunity towards pathogens. As mentioned in the previous chapter, there are several antibody-based detection assays but most require large quantities of reagents. There has been a technological shift towards multiplex platforms in the current market in order to meet this need of detecting multiple targets in a single sample. Much like the Multiplex Bead Assay (MBA) discussed in the previous chapter, protein microarrays make it possible to test thousands of different analytes from different pathogens in a single assay[10,52,53]. In addition to being able to screen multiple targets per sample (e.g. antigens) like the MBA, it has the ability to multiplex the biomarkers being measured per spot (e.g. different antibody isotypes)[54]. This can enable us to further characterise how multiple infections shape the immune response as well as potentially determine how different pathogen combinations shape the immune response.

A microarray is the miniaturisation of an assay onto a slide/chip, also called a planar array and these were originally developed for high-throughput measuring of DNA and RNA expression[55,56]. Because it is not possible to learn about the functionality of proteins purely based on gene transcription and expression levels[57], particularly since the presence of a protein cannot always be inferred directly from gene transcription, protein-based approaches have been developed for use on the microarray platform[55,56].

Protein microarray assays can be divided into three categories: 1) analytical, 2) functional and 3) reverse-phase protein microarrays (RPPM; ***Figure 6.1***; reviewed in[58]).

Briefly, an analytical protein microarray consists of immobilising a capture antibody on a chip which, when used in an assay, captures protein from the serum or protein

solution being tested[59,60] (***Figure 6.1***). This is detected using a reporter or detection antibody, much like a sandwich ELISA[61], and has been successfully used to detect protein levels of carcinoma cells after treatment[62].

A functional microarray consists of immobilising purified full-length functional proteins on a chip in order to study biochemical properties of proteins when these interact with DNA, lipids, drugs, peptides and other proteins, as well as enzyme-substrate relationships (***Figure 6.1***)[61,63]. A functional microarray was first used by Zhu *et. al*. where substrate specificity of protein kinases was determined in yeast[64]. Finally, a RPPM, related to analytical microarrays, consists of immobilising small amounts of tissue or cell lysates on a chip[58,60,65] (***Figure 6.1***). Antibodies against the target proteins are used to probe the array and can be detected using for example fluorescently tagged secondary antibodies. Although the lack of commercially available specific antibodies may limit the application of this method, we have chosen this method for the next two chapters and have used the few reagents that are available.

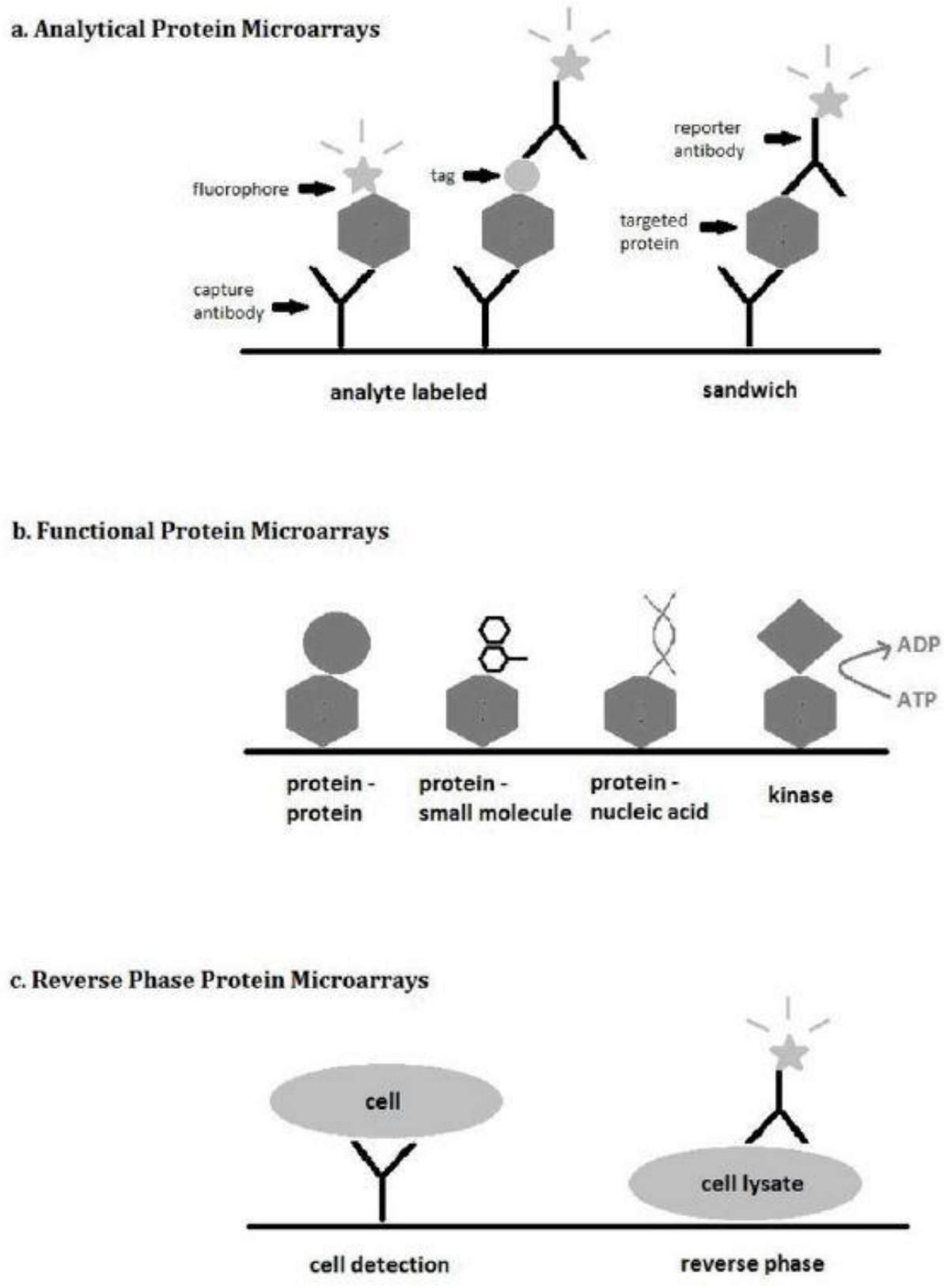


Figure 6.1 Protein microarray categories. (A) Analytical protein microarrays immobilise antibodies on a chip and capture targeted proteins in solution which can then be detected using fluorophore tags or labelled reporter antibodies. (B) Functional protein microarrays immobilise full-length functional proteins which can interact with other proteins, small molecules, nucleic acid and even enzyme-substrate relationships. These are used to study biochemical interactions. (C) Reverse Phase Protein Microarrays (RPPM) immobilise tissue or cell lysates and can detect antibody presence. Figure from a published source[58].

Malaria prevalence has previously been assessed using protein microarrays. Baum *et. al.* assessed the epidemiology of asymptomatic and submicroscopic *Plasmodium* spp. in Thailand, concluding that transmission is still ongoing although a lot of it goes undetected by traditional methods of detection[66,67]. Similarly, Uplekar *et. al.* looked at prevalence in India identifying antigens associated with asymptomatic malaria[68].

A principle benefit of large-scale proteomic approaches using microarray is that they can capture antibody isotype and subclass reactivity patterns towards a multitude of antigens. In some instances, these have been used to distinguish symptomatic from asymptomatic disease patterns in atopic diseases such as asthma and eczema by measuring IgE reactivity profiles[69]; and to identify markers of host immunity to malaria by measuring IgG reactivity. A microarray immunoassay was developed using recombinant antigens from *P. falciparum* blood-stage vaccine candidates and tested on serum from Gambian children[70]. It was reported that protection against clinical malaria was significantly associated with serum IgG reactivity towards combined antigen recognition whilst single antigen reactivity did not correlate with immune status[70]. In a separate study a protein microarray was used to identify biomarkers that accurately determined recent malaria exposure[7].

This chapter describes the analysis of serum samples from Malaysian clinical cases on the RPPM protein microarray platform using the expanded panel of *P. knowlesi*-specific recombinant antigens. More specifically, we dissect the antigen reactivity seen towards three different immunoglobulin isotypes: IgM, IgG and IgA.

6.2. METHODS

6.2.1. PRINTING OF RECOMBINANT PROTEINS ON MICROARRAY SLIDES

The selected purified and PBS-dialysed recombinant antigens were printed onto nitrocellulose-coated slides (ONCYTE® AVID, Grace Bio-Labs, USA) using the Arrayjet Marathon Classic printer (Arrayjet, UK). A total of 194 antigens and controls were printed within each array, including the full *P. knowlesi* panel of antigens. Reference spots (0.4 mg/ml) were included as fiducial markers to help the GenePix® software (GenePix Pro 7.2.29.002) and ArrayCam® Multiplex Microarray Imaging System (Grace Bio-Labs, USA) correctly assign the data extraction grid to each array during the imaging process of the arrays. A standard control curve of each purified immunoglobulin in a serial dilution (200; 100; 50; 25; 12.5; and 6.25 µg/ml) and blanks were also included on the array. The material selected to be reference spots was chosen on the basis that it would fluoresce independently of sample reactivity, in this case, purified human IgG (Sigma-Aldrich) as per standard protocol. The serial dilution immunoglobulin control curves of purified human IgG, IgM, IgA (I4506, I8260, I4036 respectively; Sigma-Aldrich), IgE and IgD (ab91022, ab65866 respectively; Abcam) were used to verify that the secondary antibodies were binding with the correct specificity as well as to derive an estimate of antibody binding in test samples. Blanks spots were one part glycerol printing buffer (1x PBS-T and 60% glycerol bidistilled) and one part 1x PBS and these were used to control for non-specific binding and background reactivity.

The array layout is prepared as a *.csv* file in MS Excel as described in appendix VI. The layout is inputted to the Command Centre for Marathon v1.4.2 software (Arrayjet) in order to generate a 384-well plate map file, indicating a specific loading pattern of antigen to obtain the desired printed layout. The recombinant antigens were diluted to 0.2 mg/ml (where stock antigen concentration allowed) in 1x PBS in order to print spots at 0.1 mg/ml (or half stock concentration, where required) on the

array. These were prepared in a 384-microwell plate with glycerol printing buffer (1x PBS-T and 60% glycerol bidistilled) to a final well volume of 20 µl following the plate map prepared previously, mixing thoroughly by pipetting and finally centrifuging the plate down at 1800 x g for 5 minutes to eliminate any bubbles. A septa (Applied Biosystems) was fitted to the plate in order to prevent evaporative loss from the diluted protein samples.

The inside of the printer is wiped clean from any dust and contaminants before starting the printing process and both the printer and atmospheric controller turned on to allow conditions inside the printer to equilibrate. Once a series of system checks are performed, printing parameters determined (as per appendix VI) and blank slides are loaded into the printer, the microplate is loaded for printing to commence. In this instance, the recombinant antigen spots were printed in singlets. The slides remain in the printer for at least 6 hours after printing with the atmospheric controller still on before being stored in slide boxes at 4°C in a vacuum desiccator. Tate Oulton performed this part of the experiment.

6.2.2. CONJUGATION OF SECONDARY ANTIBODY TO FLUORESCENT MARKERS FOR TWO DIFFERENT MICROARRAY SCANNERS TO EXPAND PANEL OF AVAILABLE DETECTION REAGENTS

As described in the following sections, two different microarray scanner machines were used to test human and macaque serum samples for immunoglobulin antibodies against the full microarray panel of recombinant antigens (*P. falciparum* (AMA1, MSP1-19, etc.), *P. vivax* (AMA1, MSP1-19, etc.) and *P. knowlesi* (MSP1 antigen 2, SERA3 antigens 1 and 2, SSP2, TSERA2 antigens 1 and 2, PKH_021580, PKH_031930 antigens 1 and 2 and PKH_080030).

6.2.2.1. Antibody isotypes compatible with ARRAYCAM® microarray scanner

Quantum dots, or Qdots®, are tools used as fluorescent probes in the field of biology due to their improved optical stability, greater fluorescence brightness, and narrow emission bands compared to molecular dyes[71]. Qdots® are nanocrystals that, according to the size of the crystal, emit light at different wavelengths. These nanocrystals are made of semiconductor materials. Bound Qdot® reagents are excited by a diode laser within the imager and fluorescent images are captured in high resolution using a digital camera. Compared to organic dyes, Qdots® are brighter and much more stable to photobleaching[71]. A single Qdot® has the fluorescent intensity of approximately 20 rhodamine molecules[71]. Qdot® technology enables multiplex detection of reactivity on a single array using colour discrimination. Individual band pass filters are used to collect the signal, in this case, from the 800nm, 655nm, and 585nm channels (***Figure* 6.2**).

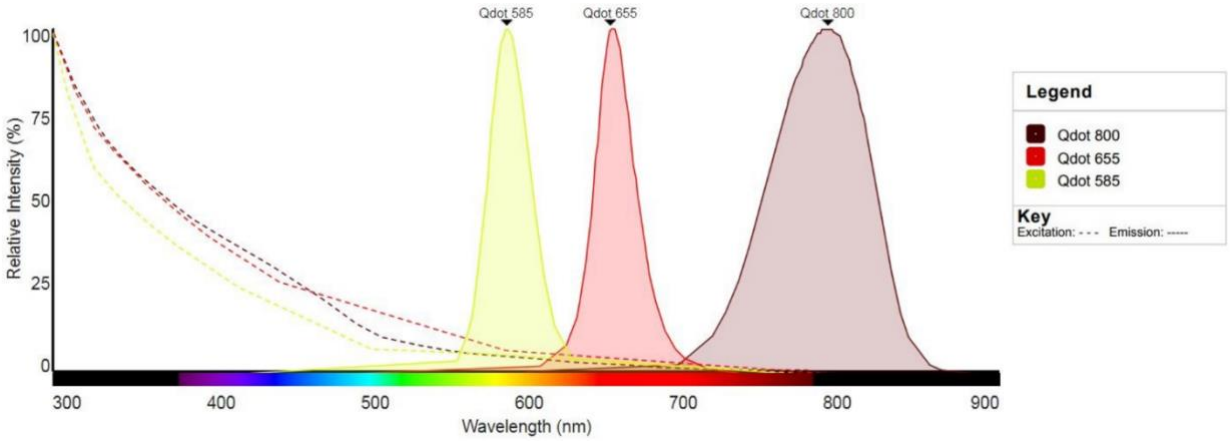


Figure 6.2 Fluorescence SpectraViewer simulation image for Qdot® 585, 655 and 800. The image shows the relative intensity percentages for the different fluorophore emission plots, shown in solid lines, and excitation plots, shown as dashed lines (Qdot 585 yellow, Qdot 655 red, Qdot 800 mauve). This image was generated using the ThermoFisher Scientific online SpectraViewer simulator https://bit.ly/2uuXBmZ.

Based on previously published data[54], anti-immunoglobulin antibodies (affinity-purified goat anti-Human IgG Qdot® 800 (catalog number 110610), goat anti-Human IgM Qdot® 655 (catalog number 110630) and goat anti-Human IgA Qdot® 585 (catalog number 110620), Grace Bio-Labs, USA) were selected for their availability

and compatibility to the ArrayCam® Multiplex Microarray Imaging System (Grace Bio-Labs, USA). These did not require conjugation steps and were ready to use reagents.

6.2.2.2. Conjugation of antibody isotypes compatible with GenePix microarray scanner

Anti-immunoglobulin antibodies were conjugated to fluorescent dyes chosen for their compatibility to the GenePix® 4000B microarray scanner (Molecular Devices, USA) as described below. As the protein microarray has yet to be fully explored, currently there are very few appropriately labelled reagents available. We therefore obtained labelling kits and reagents tailored to the experiments we wanted to perform.

An Alexa Fluor™ 532 antibody labelling kit was used to label α-human/macaque IgG (0.5-1 mg/ml stock; #555784 BD Sciences), α-human IgD (1 mg/ml stock; #2030-01 2B Scientific) and α-monkey IgA (1 mg/ml stock; #617-101-006 2B Scientific), which were all verified as Non-Human Primate (NHP) approved reagents by the NHP Reagent Resource (http://www.nhpreagents.org). Conjugations were performed as per manufacturer's instructions. Briefly, the antibodies were diluted to 1 mg/ml and 1M sodium bicarbonate buffer was added at 1/10th of the volume of the diluted antibodies, then 100 µl was transferred to the vial of reactive dye and gently inverted to dissolve the dye. The solution was incubated at RT for 1 hour and gently inverted every 10 minutes to increase labelling efficiency. During the incubation time, the kit purification columns were prepared. This step ensures the removal of unbound dye from the dye-conjugated protein. The columns have two frits inserted at the bottom before adding purification resin. The resin must be stirred before being added to the column and, once added up to 1 ml, it must be allowed to settle by gravity before more resin is added to a total of approximately 1.5 ml. The column buffer is allowed to drain from the column by gravity before being centrifuged at 1100 x g for 3 minutes. Once this is done, the antibody-dye reaction is loaded dropwise onto the

spin column and allowed to absorb into the resin bed before being centrifuged into a collection tube at 1100 x g for 5 minutes.

An Alexa Fluor™ 633 Protein labelling kit was used to label Goat α-human IgG (H+L) Cross-Adsorbed Secondary Antibody (2 mg/ml stock; # A-21091 ThermoFisher Scientific),α-human IgE (1 mg lyophilised stock rehydrated in 2mM glacial acetic acid; 01-10-04 Insight) and α-human IgM (2 mg/ml stock; # 054900 ThermoFisher Scientific), which were also all verified as NHP approved reagents by the NHP Reagent Resource (http://www.nhpreagents.org). Conjugations were performed as per manufacturer's instructions. Briefly, the antibodies were diluted to 2 mg/ml in 1x PBS and 50 µl of 1M sodium bicarbonate buffer was added to 500 µl of the diluted antibodies, then the solution was transferred to the vial of reactive dye, gently inverted to dissolve the dye and incubated at RT for 1 hour. The kit purification columns were prepared as instructed before adding purification resin. The resin must be properly stirred and added to the column until the resin is approximately 3 cm from the top of the column, allowing the excess buffer to drain away. The reaction mix is loaded onto the column and allowed to enter the resin before being rinsed with 100 µl of elution buffer, taking care not to disturb the resin bed. As the solution is run through the column, two coloured bands become visible, indicating the separation of incorporated dye and unincorporated dye. The first coloured band contains dye-incorporated protein, which should be collected. Elution buffer is added as necessary and retained in fractions for approximately 30 minutes in order to elute as much labelled protein as possible.

The protein concentration and the degree of labelling for each antibody isotype was determined as per manufacturer's instructions. Briefly, absorbance was measured using the NanoDrop® Spectrophotometer ND-1000 (ThermoFisher Scientific) at 280 nm and 530 nm for samples labelled with Alexa Fluor™ 532nm and at 632 nm for samples labelled with Alexa Fluor™ 633 nm. The following formulas were used to determine protein concentration for label 532 nm and for label 633 nm: ([A_{280}-

(A_{530}*0.09)]*5df)/2,030,000 and ([A_{280}-(A_{632}*0.55)]*5df)/2,030,000 respectively. The following formulas were used to determine the degree of labelling for label 532 nm and for label 633 nm: (A_{530}*5df)/(810,000*protein concentration) and (A_{632}*5df)/(1,000,000*protein concentration) respectively. The labelled protein was stored at 4°C in the dark and for longer storage, at -20°C in small aliquots.

6.2.3. TITRATION OF THE DIFFERENT SECONDARY ANTIBODIES

In order to run the experiment, it was necessary to determine the optimal working concentrations of the reagents in order to obtain realistic and comparable results. The serum sample dilution was selected based on the established in house protocol developed by Tate Oulton, which indicates that 1/400 dilution is ideal. A series of secondary antibody dilutions were run in order to determine the ideal working dilutions for both the Genepix® (Molecular Devices,USA) and the Araycam® (Grace Bio-Labs, USA) experiments. The dilutions were 1/125, 1/250, 1/500 and 1/1000 for the Genepix and 1/100, 1/200, 1/400, 1/800 and 1/1000 for the Arraycam. The rest of the experiment was conducted as described below and part of it was performed by Tate Oulton.

6.2.4. PROCESSING OF SERUM SAMPLES ON MICROARRAY SLIDES

The human and macaque serum samples were diluted the previous day at 1/400 in array 0.25x blocking buffer (Arrayit 1x Blockit buffer in 1x PBS). The array slides were securely placed into a hybridisation cassette (HC) (Arrayit), 100 µl of array blocking buffer was added to each array and these were left to incubate at RT for 1 hour on a rotary platform at 100 rpm after properly sealing the HC. The slides were washed three times with 200 µl 1x PBS/T (1x PBS and 0.01% Tween 20) array wash buffer before 100 µl of diluted human or macaque samples were added to each array and incubated at RT for 1 hour as described earlier. The slides were washed three times as previously described and 100 µl of diluted secondary antibody (secondary antibodies conjugated to Qdots® at 1/200 dilution) in array blocking

buffer was added to each array, then incubated in the dark at RT for 1 hour as described earlier. The microarray slides were then washed three times with 200 µl 1x PBS/T array wash buffer and once with 200 µl mili-Q water before being centrifuged at 1800 x g for 5 minutes in 50 ml tubes. It was ensured that slides were properly dried before being stored in slide boxes at 4°C in a vacuum desiccator for later scanning. The slides were scanned by Tate Oulton using GenePix 4000B microarray scanner (Molecular Devices, USA) or ArrayCam (Grace Bio-Labs, USA) as appropriate, and the results expressed as median fluorescent intensity (MFI).

6.2.5. DATA PROCESSING

All microarray data was processed using R scripts written by Dr. Nuno Sepulveda, Dr. William Stone and Dr. Katherine Glass (these scripts were run on RStudio, Version 1.1.453 – © 2009-2018 RStudio, Inc.). Briefly, background corrections were made using the normexp method from the limma package[72] with offset = 50. Targets which were printed immediately following 'high intensity spots' (e.g. top standard, reference spots) were excluded, due to a potential carryover effect during printing. Buffer spots were used to quality check the printing process, allowing the identification of deviant slides, pads and samples and excluding samples with buffer means greater than the total buffer mean plus three standard deviations. Data was normalised as a ratio of signal for each spot to the mean buffer signal within each sample (Target MFI/ Buffer MFI) and log transformed (base 2). GST signal was subtracted from the normalised data for GST-tagged targets before being analysed. Negative normalised values (values below buffer mean) were set to zero. Seropositivity cutoffs were calculated independently for each sample as the mean MFI of the buffer spots plus three standard deviations. More information on the different statistical analyses performed on the different data sets is found the respective chapters.

6.2.6. STATISTICAL ANALYSIS

For the hospital samples, scatter plot matrices and correlation plots were created using STATA/IC 14.2 (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.) and RStudio (Version 1.1.453 – © 2009-2018 RStudio, Inc.), showing the presence or absence of collinearity between *P. knowlesi* recombinant antigens. Dot plots, bar graphs and heat maps were created using PRISM (GraphPad PRISM 7). Statistically significant differences in antibody responses between time points per antigen were determined using Friedman's test followed by Dunn's post hoc test using PRISM (*P<0.05; **P<0.01; ***P<0.001). The same test was applied to the antigen breadth data but no statistical significances were recorded.

6.3. RESULTS

6.3.1. *P. knowlesi*-specific recombinant antigen serum reactivity towards a subset of Malaysian clinical samples using the ArrayCam protein microarray scanner

In the previous chapter we used the MBA to analyse reactivity towards multiple *P. knowlesi* antigens in clinical and community samples, looking at IgG responses. In this chapter we apply a similar methodology on the protein microarray platform which, unlike the MBA, can detect multiple antibody isotypes in a single assay.

The experimental strategy for this chapter is outlined in **Figure *6.3***, showing the identification of the target sequences of interest and, highlighted in green, the test samples used on the protein microarray platform. The protein microarray assay used a subset of the Malaysian serum samples collected from adults and children hospitalised with malaria at different time-points (day 0, 7 and 28; Table 3.1; chapter 3, section 3.3) post-treatment diagnosed as *P. knowlesi* (n=12), *P. vivax* (n=1) or *P. falciparum* (n=1).

This work was attempted for the GenePix microarray scanner and, although it was possible to produce custom-made reagents for this platform, the optimisation of such reagents proved to be time consuming and the reagent signals were found to cross-react. Cross-reactivity was observed between the different antibodies, which could potentially lead to epitope conflict further down the line. An alternative route to analysing isotype profiles on microarray was pursued, using Qdot® fluorescent probes and the ArrayCam scanner. Qdot® fluorescent probes bound to secondary anti-immunoglobulin antibodies IgM, IgG and IgA were procured (Grace Bio-Labs, USA) to be used on the ArrayCam microarray imaging system (Grace Bio Labs, Oregon, USA) as these had already been previously optimised[54]. Qdot® probes are a relatively new reagent which are brighter and much more stable to photobleaching than other organic dyes[71,73,74].

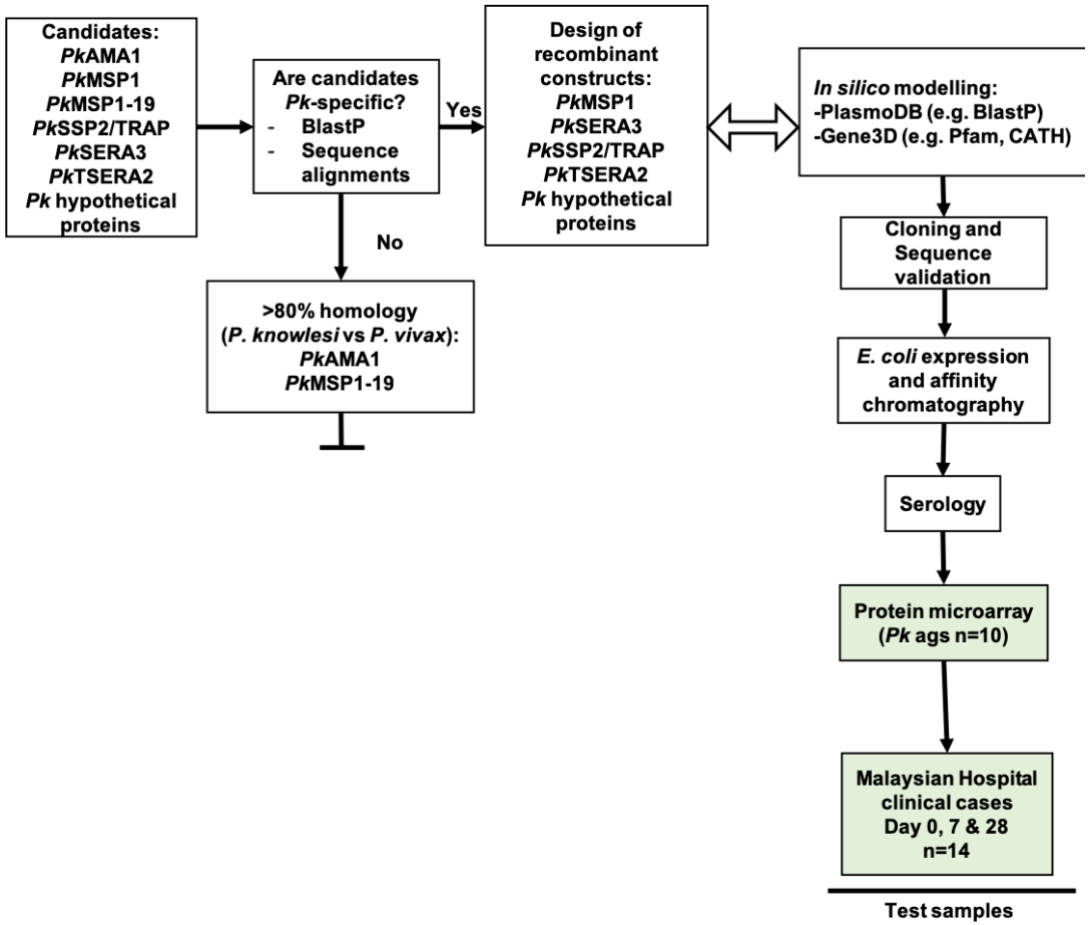


Figure 6.3 Flowchart summarising the experimental strategy used in the identification and validation of the *P. knowlesi*-specific candidates and sample sets used in this chapter. Test samples were obtained from a hospital-based clinical treatment trial in Sabah, Malaysia (n=164), and were collected at different time points post-diagnosis and treatment (day 0, 7 and 28). A subset of the clinical samples (n=14) were tested on the protein microarray platform. Highlighted in green are the sections relevant to this chapter.

Qdot® probes are a relatively new reagent which are brighter and much more stable to photobleaching than other organic dyes[71,73,74]. These can be excited with a wider range of wavelengths than organic dyes and have narrower emission spectra, enabling the use of multiple Qdot® probes in a single reading[75]. Each of the 3 antibodies are bound to a specific Qdot® fluorescent probe that is excited by a diode laser at a specific wavelength (IgM 655 nm, IgG 800 nm and IgA 585 nm; **Figure 6.2**), making it possible to determine reactivity towards all 3 isotypes in a single run. For continuity, only samples that had data recorded for all three time points (day 0, day 7 and day 28 after diagnosis and treatment) were selected from the Malaysian clinical samples for analysis. A total of 14 individuals with all three time points from the clinical samples were selected to be tested using the microarray platform, with 12 of the individuals having been PCR diagnosed with *P. knowlesi*, 1 with *P. vivax* and 1 with *P. falciparum*. One of the samples from day 0 (CCK069) was not tested in this assay as it was originally collected as a bloodspot instead of serum sample, as a result the material was not sufficiently concentrated to be able to achieve the appropriate concentration for this assay. However, the other time points for this individual (days 7 and 28) were tested and analysed. Originally 1 year follow up samples were to be included in this analysis but these were also excluded due to insufficient sample volume for use on the microarray platform at the appropriate dilution. These samples were probed for reactivity towards the full panel of *P. knowles*i recombinant antigens. ***Figure 6.4*** is a representative figure showing the different fluorescence captured within the same sector of a slide measuring all three isotypes simultaneously at their respective wavelengths.

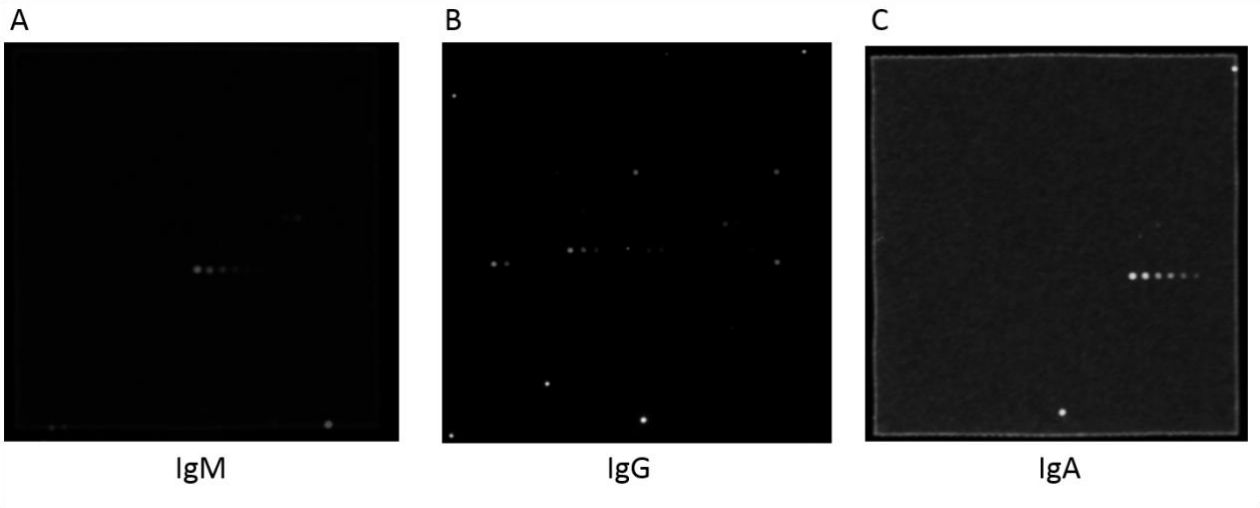


Figure 6.4 Protein microarray example slide showing fluorescent reactivity within the same slide sector towards the different isotype wavelengths. (A) Anti-human IgM is seen at 655 nm, (B) anti-human IgG is seen at 800 nm and (C) anti-human IgA is seen at 585 nm on the ArrayCam slide scanner for the same serum sample.

For the following sections, different graphical analyses were applied to the data. ***Figure 6.5*** to ***Figure 6.7*** show overall serum reactivity correlation matrices towards all *P. knowlesi* antigens at days 0, 7 and 28 after diagnosis and treatment for each of the immunoglobulins (IgM; IgG and IgA; ***Figure 6.5******Figure 6.6******Figure 6.7***, respectively).

Reactivity towards IgM occurs primarily at day 0 (***Figure 6.8***; appendix VII, Figure 1 A), with the highest reactivity seen towards *Pk*SERA3 antigen 2. Very little collinearity of responses were observed between the different *P. knowlesi*-specific antigens towards IgM at any time point, with only TSERA2 antigens and PKH_031930 antigens showing significant correlations. The only antigens to show a significant negative correlation were SSP2 and PKH_021580 at day 7 (***Figure 6.5*** B; r=-0.62).

In ***Figure 6.6***, which shows IgG reactivity correlations, the *Pk*TSERA2 antigens show significant collinearity with the PKH_031930 antigens, particularly on day 7. PKH_031930 antigen 1 and 2 show a significant and strong response collinearity at all time points, which is not surprising as these are fragments within the same gene sequence. Similarly to IgM, the strongest reactivity is seen towards *Pk*SERA3

antigen 2, however this is seen at day 28 (***Figure 6.8***; appendix VII, Figure 2 C). Antigen *Pk*TSERA2 antigen 1 also showed strong reactivity towards IgG at day 28 after diagnosis and treatment (***Figure 6.8***).

IgA was the isotype to show the least collinearity between antigens (***Figure 6.7***). Reactivity towards IgA appears to be the weakest of all three antibody isotypes, however, *Pk*SERA3 antigen 2 still seems to be the *P. knowlesi*-specific antigen to show strongest reactivity (***Figure 6.8***; appendix VII, Figure 3).

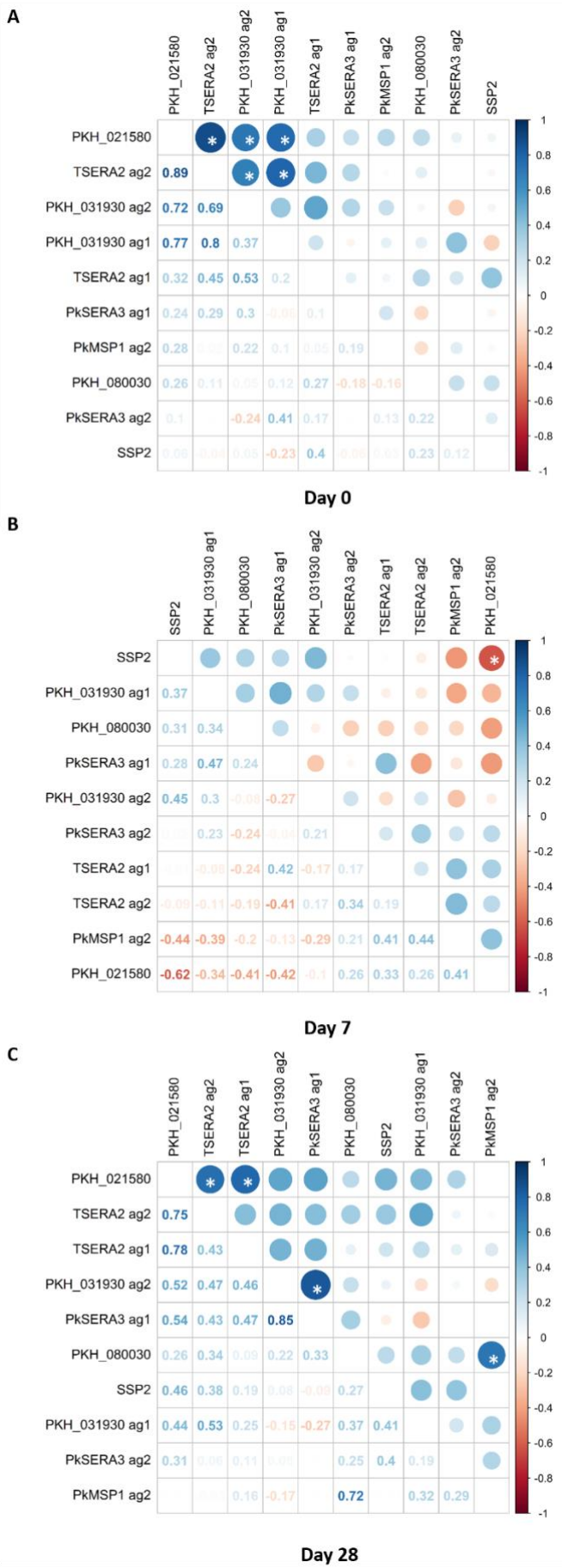


Figure 6.5 Correlation plots showing IgM isotype reactivity towards *P. knowlesi*-specific antigens in Malaysian clinical samples at days 0, 7 and 28 after diagnosis and treatment (n=12). Plots indicate positive or negative associations. The upper diagonal half shows the correlation by colour and circle size. The lower diagonal half shows the correlation values. The data is sorted from highest to lowest correlation. Correlation is shown as highest in dark blue and large circles to very low or no reactivity in red and smaller circles. Asterisks signify significant difference in correlation between antigens using Pearson's correlation test using p=0.05.

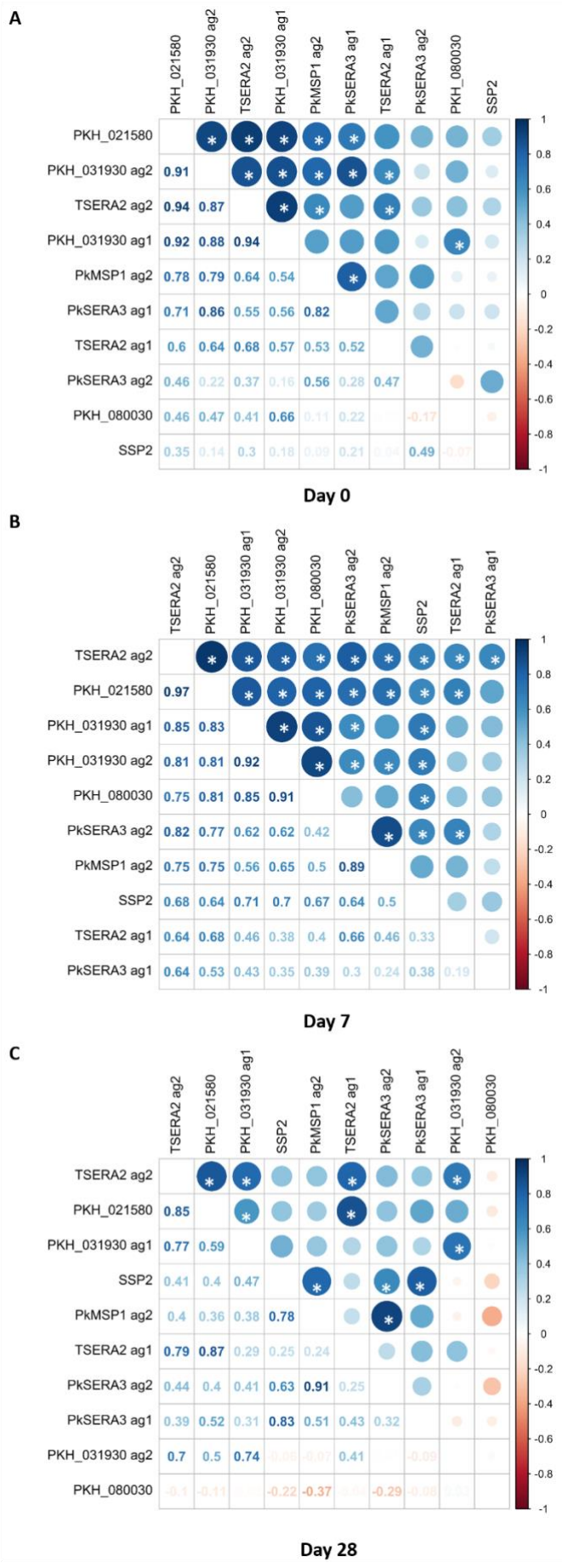


Figure 6.6 Correlation plots showing IgG isotype reactivity towards *P. knowlesi*-specific antigens in Malaysian clinical samples at days 0, 7 and 28 after diagnosis and treatment (n=12). Plots indicate positive or negative associations. The upper diagonal half shows the correlation by colour and circle size. The lower diagonal half shows the correlation values. The data is sorted from highest to lowest correlation. Correlation is shown as highest in dark blue and large circles to very low or no reactivity in red and smaller circles. Asterisks signify significant difference in correlation between antigens using Pearson's correlation test using p=0.05.

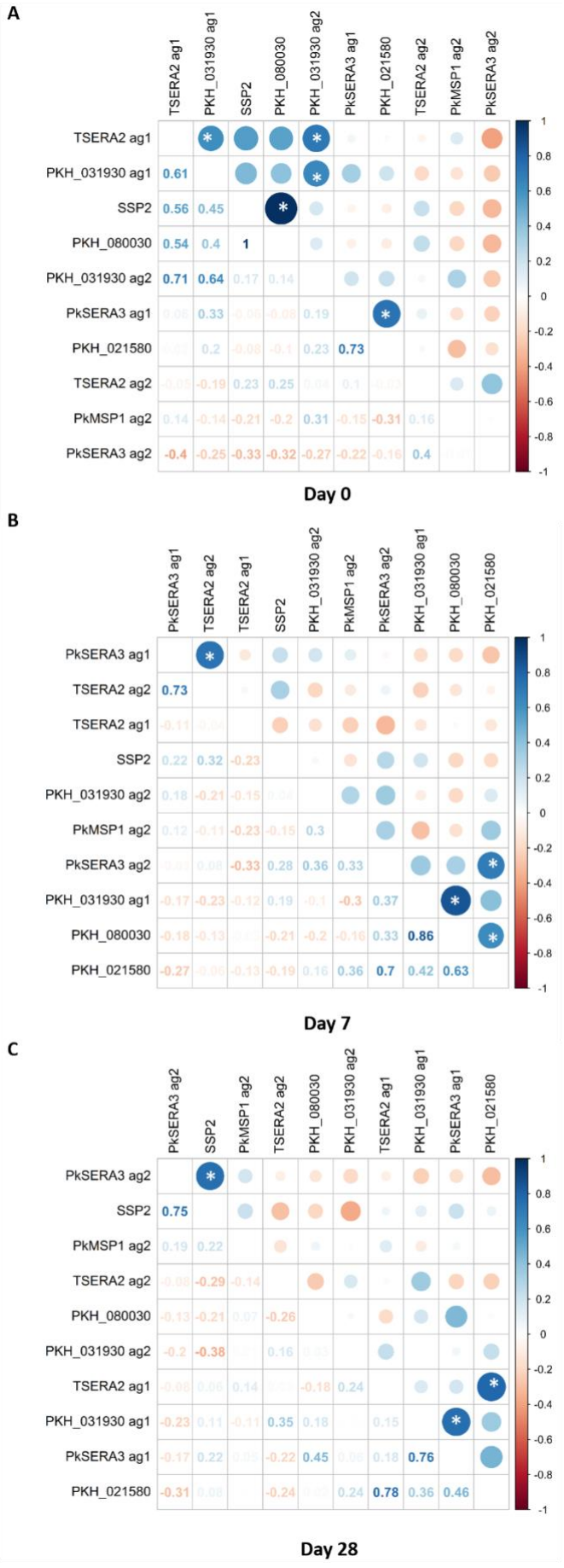


Figure 6.7 Correlation plots showing IgA isotype reactivity towards *P. knowlesi*-specific antigens in Malaysian clinical samples at days 0, 7 and 28 after diagnosis and treatment (n=12). Plots indicate positive or negative associations. The upper diagonal half shows the correlation by colour and circle size. The lower diagonal half shows the correlation values. The data is sorted from highest to lowest correlation. Correlation is shown as highest in dark blue and large circles to very low or no reactivity in red and smaller circles. Asterisks signify significant difference in correlation between antigens using Pearson's correlation test using p=0.05.

6.3.2. Isotype reactivity profiles to each *P. knowlesi*-specific antigen

Much like the results obtained from the ELISA analysis of these same samples (chapter 4, section 4.2.5.), It was clear that for these Malaysian clinical samples *Pk*SERA3 antigen 2 showed the strongest reactivity towards all isotypes, followed by *Pk*TSERA2 antigen 1 for IgM and IgG (**Figure *6.8*Figure 6.9**).

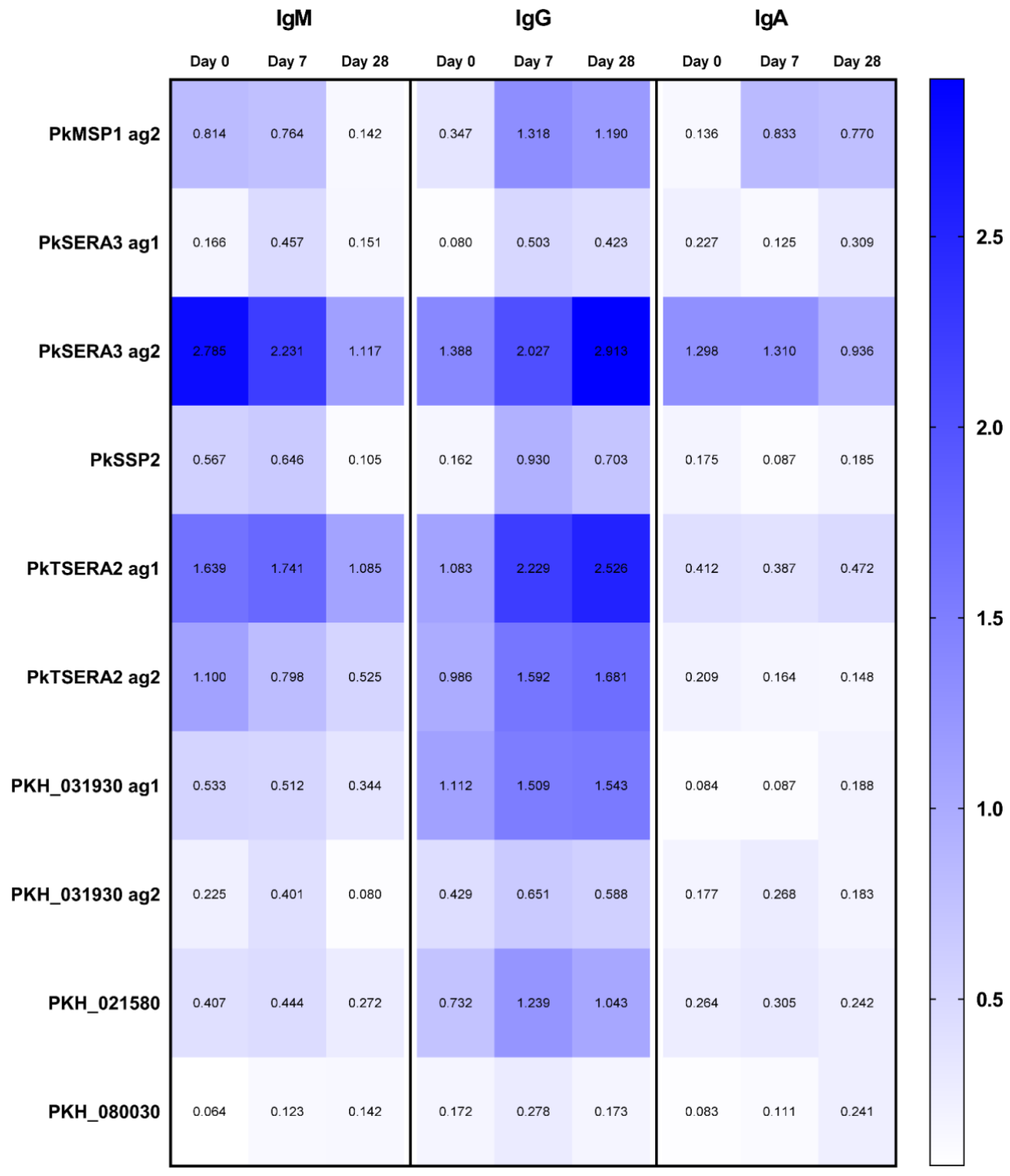


Figure 6.8 Heat map of IgM, IgG and IgA reactivity towards *P. knowlesi*-specific antigens at days 0, 7 and 28 after diagnosis and treatment of Malaysian clinical samples (n=12). The colour scale represents the antigen reactivity geometric mean log2 MFI ratio. Reactivity is shown as highest in dark blue to very low or no reactivity in white.

It was observed that *P. knowlesi*-specific antigen reactivity towards anti-IgM antibody showed a decreasing trend over time for all antigens without exception, with *Pk*SERA3 antigen 2, *Pk*SSP2 and *Pk*TSERA2 antigen 1 showing a significant decrease ($P<0.001$, $P<0.05$ and $P<0.05$, respectively; ***Figure 6.9*** A).

Unlike IgM, there is an increasing trend in reactivity towards IgG across time for all antigens with a peak in reactivity at day 7 and the occasional decrease by day 28, with PkSSP2 showing a significant increase at day 7 ($P<0.05$; ***Figure 6.9*** B). Only 4 antigens indicated an increasing trend in reactivity towards IgG across time, peaking at day 28 (*Pk*SERA3 antigen 2, *Pk*TSERA2 antigens 1 and 2 and PKH_031930 antigen 1; ***Figure 6.9***).

The isotype to show the least reactivity of all three was IgA, with irregular peaks and falls across time for the different *P. knowlesi*-specific antigens (***Figure 6.9*** C). As mentioned previously, the highest responding antigen was *Pk*SERA3 antigen 2, and unlike for the other isotypes, followed by *Pk*MSP1 antigen 2, which showed reversed reactivity towards IgM across time (***Figure 6.8***; ***Figure 6.9*** A and C).

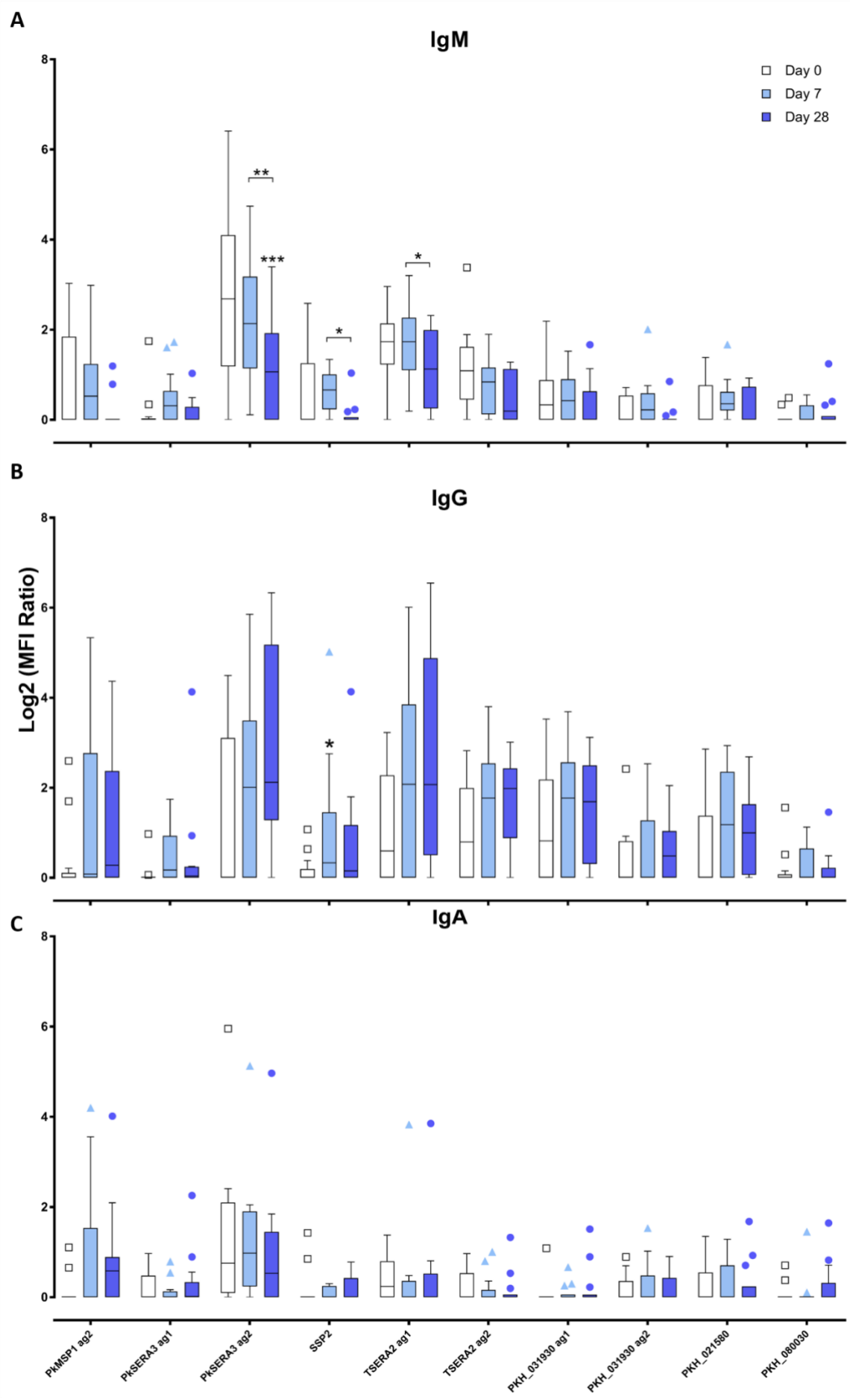

A
IgM
Day 0
Day 7
Day 28
**
***
*
*
B
IgG
Log2 (MFI Ratio)
*
C
IgA
PkMSP1 ag2
PkSERA3 ag1
PkSERA3 ag2
SSP2
TSERA2 ag1
TSERA2 ag2
PKH_031930 ag1
PKH_031930 ag2
PKH_021580
PKH_080030

Figure 6.9 Tukey box and whisker plots of Immunoglobulin isotype reactivity to the *P. knowlesi*-specific antigens for the Malaysian clinical sera (n=12). Reactivity to each time point is grouped by antigen: day 0 (white), 7 (light blue) and 28 (dark blue). The data is plotted for (A) IgM, (B) IgG and (C) IgA. The median is based on Log2 of the MFI ratio for each antigen and the box indicates values for the lower and upper quartile. The error bars shown are the lower and upper quartiles ±1.5xIQR. Level of significance between day 0 and the other days are indicated by asterisks, level of significance between day 7 and 28 are indicated by a line with asterisks, non-significant values were not plotted (*P<0.05; **P<0.01; ***P<0.001; Friedman test with Dunn post hoc comparisons of reactivity in the three time points).

6.3.3. Breadth of antibody response towards *P. knowlesi* antigens in clinical samples

In the following section we determine the number of individuals reactive to a specific antigen and the number of antigens eliciting an antibody response in a single person (breadth of response per isotype). *Pk*SERA3 antigen 2 was the antigen to elicit the most IgM, IgG and IgA reactivity (***Figure 6.10***). For IgM, this was followed by *Pk*TSERA2 antigens 1 and 2 (***Figure 6.10*** A). Independently of time points, the majority of individuals tested showed IgG reactivity towards *Pk*SERA3 antigen 2, followed by *Pk*TSERA2 antigen 2 and PKH_031930 antigen 1 (***Figure 6.10*** B). For *Pk*SERA3 antigen 2, 10 out of 12 individuals were reactive to this antigen at day 28. IgG was the isotype with most seropositive antigens independently of time point. IgA reactivity towards *Pk*SERA3 antigen 2 was highest at day 7 (***Figure 6.10*** C).

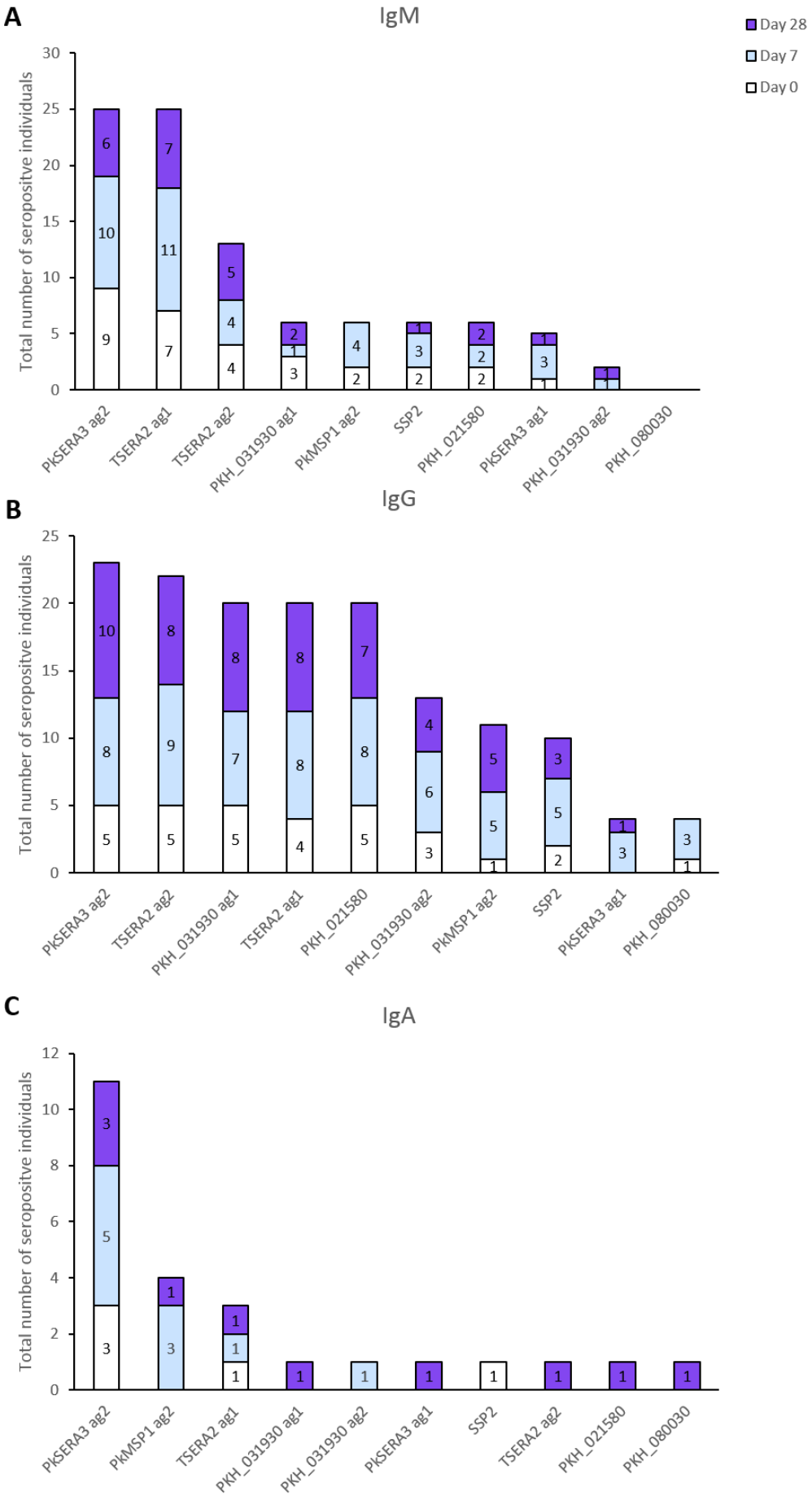

A
IgM
Day 28
Day 7
Day 0
Total number of seropositve individuals
PkSERA3 ag2
TSERA2 ag1
TSERA2 ag2
PKH_031930 ag1
PkMSP1 ag2
SSP2
PKH_021580
PkSERA3 ag1
PKH_031930 ag2
PKH_080030
B
IgG
Total number of seropositve individuals
PkSERA3 ag2
TSERA2 ag2
PKH_031930 ag1
TSERA2 ag1
PKH_021580
PKH_031930 ag2
PkMSP1 ag2
SSP2
PkSERA3 ag1
PKH_080030
C
IgA
Total number of seropositve individuals
PkSERA3 ag2
PkMSP1 ag2
TSERA2 ag1
PKH_031930 ag1
PKH_031930 ag2
PkSERA3 ag1
SSP2
TSERA2 ag2
PKH_021580
PKH_080030

Figure 6.10 Total number of seropositive individuals per *P. knowlesi* antigen for IgM, IgG and IgA at each time point. Malaysian clinical serum samples (n=12) were used to screen a panel of antigens using the protein microarray platform. Cut offs were calculated based on each sample array buffer spots, identifying cut offs per individual tested. Graph shows the number of seropositive people for each antigen at days 0 (white), 7 (blue) and 28 (purple).

Figure 6.11 shows the number of antigens each individual showed reactivity towards each isotype across the different time points. Antigen breadth eliciting an IgM response was mostly scattered at day 0, with every individual reacting to a different number of antigens (***Figure 6.11*** A). At day 7 most individuals showed reactivity towards 3 antigens, which went down to 1 antigen by day 28 (***Figure 6.11*** A). For IgG, the highest number of antigens seen per person was during day 7 post-diagnosis and treatment, where all 10 antigens elicited reactivity in at least one individual (***Figure 6.11*** B). IgA showed the lowest antigen breadth, although a higher antigen breadth was seen for some individuals later in time (***Figure 6.11*** C).

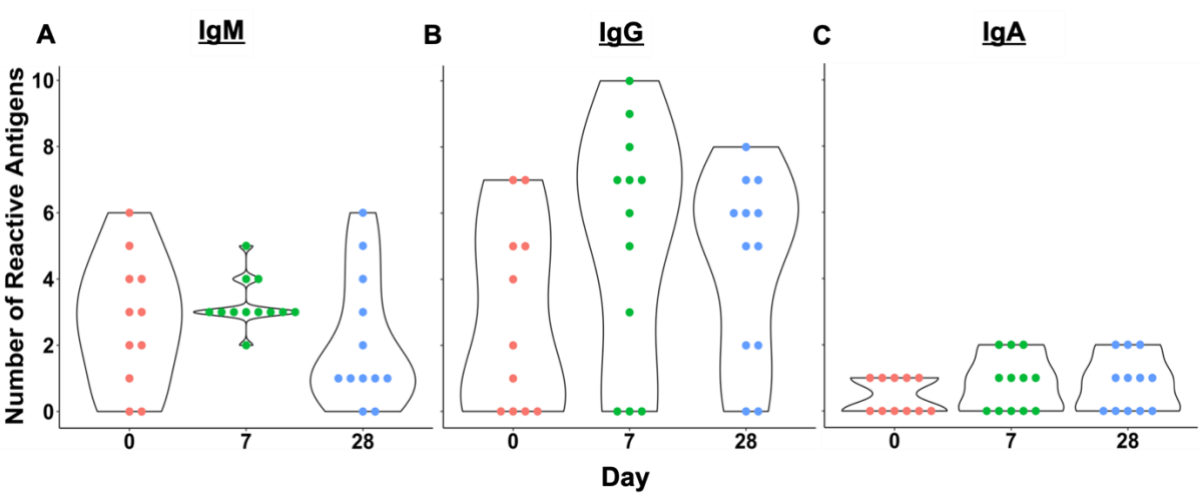


Figure 6.11 Number of seropositive *P. knowlesi* antigens for each individual across the different time points for IgM, IgG and IgA. Malaysian clinical serum samples (n=12) were used to screen a panel of antigens using the protein microarray platform. Figure represents the number of reactive antigens detected by a single person at a specific time point. The violin plots are proportional to the number of people. The different time points are indicated as follows: day 0 (red), day 7 (green) and day 28 (blue). Differences in antibody responses between time points per antigen were determined using Friedman's test followed by Dunn's post hoc test using PRISM. Non-significant results were not plotted.

6.4. DISCUSSION

The results presented here not only have shown the usefulness of high throughput technology in serology, but also how this technology can be used to measure multiple components of the immune response simultaneously. We have shown proof of principle that multiplexing of *P. knowlesi* antigens can be used for serological assessment of human samples using the protein microarray platform. Additionally, responses to IgM, IgG, and IgA were measured simultaneously on the microarray platform for a subset of *P. knowlesi* infected endemic samples.

We found that *Pk*SERA3 antigen 2 was the most highly reactive antigen, which agreed with ELISA and Multiplex Bead Assay (MBA) results. This was observed in clinical samples, and responses were both time and isotype specific. We have described the kinetics of a potentially short-term marker of *P. knowlesi* exposure, which would be useful to determine recent infection in the population. It might be possible to infer that IgM^{high}, IgG^{low} reactivity could be indicative of very recent exposure, while an IgM^{low}, IgG^{high} response could indicate an older infection. The differing isotype profiles in infections may be related to disease progression and severity – further work is need in this area.

The large scale analysis of genes, gene expression and function has long existed[76-79]. Although it is possible to determine gene function by measuring levels of mRNA transcription[80], gene function is more accurately measured by the activity of the translated proteins these encode. However, there are limitations to the transcriptomic approaches that can be overcome by proteomic tools. For example, levels of mRNA do not always reflect translation into proteins and protein levels[57] and it is not possible to measure post-translational modifications in mRNA. For instance, a lack of mRNA may lead to the false assumption that the protein is not present, particularly stable proteins that do not require constant replenishment by de novo synthesis. Protein-based approaches have been developed in order to

determine the presence and functionality of proteins using high throughout platforms[55,56,81].

Detecting multiple targets in a single sample and analysing these in the context of multiple others as well as the host immune system has long been pursued. This is now possible with MBA and protein microarray technology, which not only enable multiplexing, but also have a low sample requirement, making it possible to screen thousands of samples for thousands of targets.

IgG reactivity towards *Pk*SERA3 ag2 was previously shown by ELISA to decrease slightly by day 28 (reference[43] and chapter 4), as was shown by MBA analysis, with barely any reactivity by 1 year post-diagnosis (**Error! Reference source not found.**). In this chapter, the protein microarray data shows an increasing trend in IgG reactivity to this antigen (***Figure 6.9*** B). Although not statistically significant, this could potentially be due to a smaller sample size used for the protein microarray analysis (n=12) and in future, larger sample sets would be used for this type of analysis. Unfortunately, the reason for a smaller sample set was that not enough samples had all time points available. For a more in-depth analysis of antigens across time, more time points would be required to be collected and analysed for each individual in order to have a better idea of the antigen kinetics and the potential associations with protection for instance. Studies have shown that IgG antibodies against *P. falciparum* antigens are important for the development of clinical immunity[23,82], something that could potentially be seen for *P. knowlesi* infections in endemic populations. Further optimisation of the assay analysis would be required and larger sample sets from endemic regions would need to be screened for isotype reactivity in order to determine any correlations. It would also be necessary to develop the assay in a way that would make it possible to directly compare the different isotype reactivity with one another.

There has been much interest in trying to establish whether serological markers can be used as predictors of protection in the field of vaccinology[83,84], however, the relationship between antibody titres and protection is currently not quite clear. Associations were made between time after diagnosis and antibody isotype increase (IgG) or decrease (IgM), although only the decrease in IgM across time was statistically significant. The data presented here is consistent with previously published data showing that IgM antibody expression happens as an early response to malaria infections, followed by IgG expression[85]. The increase and decrease of IgG and IgM respectively occur within 30 days of diagnosis and in the data presented here, seem to cross at day 7. The overall lower IgA response can potentially be a result of first time infection, a theory which was previously indicated in mice[86], although not translatable to this study as the isotype profiling was done on a small subset of human samples. The high IgA response towards *Pk*MSP1 ag2 is reminiscent of a study where *Pf*MSP1-19-specific recombinant IgA was able to recognise parasites in infected RBC[36]. Reversed IgM responses were observed towards *Pk*MSP1 ag2 (***Figure 6.8***, ***Figure 6.9*** A and C). Biswas *et. al*. published similar results where a negative correlation was seen between IgA and IgM positivity looking at reactivity towards soluble extract from cultured *P. falciparum,* potentially indicating a class switch from IgM to IgA[40]. A similar observation between IgM and IgA was seen with PKH_031930 ag1 (***Figure 6.8***, ***Figure 6.9*** A and C). This would require further investigation as IgA has not been studied extensively in malaria research.

One of the limitations of protein microarray is the commercial unavailability of specific tagged antibodies, making it more difficult to procure specific reagents and coupling these in-house. Not only is this more time-consuming, it also reduces the comparability of results between research groups. Providing specific tools for antibody profiling are made readily available, it would be possible to screen samples

not only for isotype reactivity, but also for subclass reactivity in a high throughput manner, making it possible to dissect malaria exposed and unexposed populations through reactivity patterns, as has been done for asthmatic patients[69]. Here we have shown the application of high throughput protein microarray technology in the serological screening of *P. knowlesi* endemic populations, at the clinical level, not only looking at reactivity towards different antigens but also looking at individual isotype reactivity to those antigens. The simultaneous measurement of analytes and biomarkers in a single individual makes it possible to determine the relationships these have with one another and how these might shape immunity, or lack thereof, towards particular pathogens or a group of them.

This only marks the beginning of what these advancements will enable us to do, such as screening for reactivity patterns towards drug treatments and vaccines, potentially making it possible to produce personalised drugs, vaccines and treatments targeted towards certain populations or subsets, specific age groups, genders.

REFERENCES

1 Wipasa, J. *et al.* Long-lived antibody and B Cell memory responses to the human malaria parasites, Plasmodium falciparum and Plasmodium vivax. *PLoS Pathog* **6**, e1000770, doi:10.1371/journal.ppat.1000770 (2010).

2 Akpogheneta, O. J. *et al.* Duration of naturally acquired antibody responses to blood-stage Plasmodium falciparum is age dependent and antigen specific. *Infection and immunity* **76**, 1748-1755, doi:10.1128/iai.01333-07 (2008).

3 Fowkes, F. J. *et al.* New insights into acquisition, boosting, and longevity of immunity to malaria in pregnant women. *The Journal of infectious diseases* **206**, 1612-1621, doi:10.1093/infdis/jis566 (2012).

4 Meraldi, V. *et al.* Natural antibody response to Plasmodium falciparum Exp-1, MSP-3 and GLURP long synthetic peptides and association with protection. *Parasite immunology* **26**, 265-272, doi:10.1111/j.0141-9838.2004.00705.x (2004).

5 Osier, F. H. *et al.* Breadth and magnitude of antibody responses to multiple Plasmodium falciparum merozoite antigens are associated with protection from clinical malaria. *Infection and immunity* **76**, 2240-2248, doi:10.1128/iai.01585-07 (2008).

6 Stanisic, D. I. *et al.* Acquisition of antibodies against Plasmodium falciparum merozoites and malaria immunity in young children and the influence of age, force of infection, and magnitude of response. *Infection and immunity* **83**, 646-660, doi:10.1128/iai.02398-14 (2015).

7 Helb, D. A. *et al.* Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E4438-4447, doi:10.1073/pnas.1501705112 (2015).

8 Bruce-Chwatt, L. J., Draper, C. C., Dodge, J. S., Topley, E. & Voller, A. Originally published as Volume 1, Issue 7749 SERO-EPIDEMIOLOGICAL STUDIES ON POPULATION GROUPS PREVIOUSLY EXPOSED TO MALARIA. *The Lancet* **299**, 512-515, doi:http://dx.doi.org/10.1016/S0140-6736(72)90176-6 (1972).

9 Corran, P., Coleman, P., Riley, E. & Drakeley, C. Serology: a robust indicator of malaria transmission intensity? *Trends in Parasitology* **23**, 575-582, doi:http://dx.doi.org/10.1016/j.pt.2007.08.023 (2007).

10 Drakeley, C. & Cook, J. in *Advances in parasitology* Vol. 69 299-352 (Academic Press, 2009).

11 Medina Costa, R., de Sousa, K. P., Atouguia, J., Tavira, L. T. & Silva, M. S. Prevalence and Level of Antibodies Anti-Plasmodium spp. in Travellers with Clinical History of Imported Malaria. *Journal of Parasitology Research* **2013**, 247273, doi:10.1155/2013/247273 (2013).

12 Ramasamy, R., Nagendran, K. & Ramasamy, M. S. Antibodies to epitopes on merozoite and sporozoite surface antigens as serologic markers of malaria transmission: studies at a site in the dry zone of Sri Lanka. *Am J Trop Med Hyg* **50**, 537-547 (1994).

13 Soe, S., Theisen, M., Roussilhon, C., Aye, K. S. & Druilhe, P. Association between protection against clinical malaria and antibodies to merozoite surface antigens in an area of hyperendemicity in Myanmar: complementarity between responses to merozoite surface protein 3 and the 220-kilodalton glutamate-rich protein. *Infection and immunity* **72**, 247-252 (2004).

14 Oeuvray, C. *et al.* Cytophilic immunoglobulin responses to Plasmodium falciparum glutamate-rich protein are correlated with protection against clinical malaria in Dielmo, Senegal. *Infection and immunity* **68**, 2617-2620 (2000).

15 Dodoo, D. *et al.* Naturally acquired antibodies to the glutamate-rich protein are associated with protection against Plasmodium falciparum malaria. *The Journal of infectious diseases* **181**, 1202-1205, doi:10.1086/315341 (2000).
16 Kana, I. H. *et al.* Naturally Acquired Antibodies Target the Glutamate-Rich Protein on Intact Merozoites and Predict Protection Against Febrile Malaria. *The Journal of infectious diseases* **215**, 623-630, doi:10.1093/infdis/jiw617 (2017).
17 Nebie, I. *et al.* Humoral Responses to Plasmodium falciparum Blood-Stage Antigens and Association with Incidence of Clinical Malaria in Children Living in an Area of Seasonal Malaria Transmission in Burkina Faso, West Africa. *Infection and immunity* **76**, 759-766, doi:10.1128/IAI.01147-07 (2008).
18 Voller, A. & Bruce-Chwatt, L. J. Serological malaria surveys in Nigeria. *Bulletin of the World Health Organization* **39**, 883-897 (1968).
19 Drakeley, C. J. *et al.* Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5108-5113, doi:10.1073/pnas.0408725102 (2005).
20 Dewasurendra, R. L. *et al.* Effectiveness of a serological tool to predict malaria transmission intensity in an elimination setting. *BMC Infect Dis* **17**, 49, doi:10.1186/s12879-016-2164-0 (2017).
21 van den Hoogen, L. L. *et al.* Serology describes a profile of declining malaria transmission in Farafenni, The Gambia. *Malaria journal* **14**, 416, doi:10.1186/s12936-015-0939-1 (2015).
22 Bousema, T. *et al.* Serologic markers for detecting malaria in areas of low endemicity, Somalia, 2008. *Emerging infectious diseases* **16**, 392-399, doi:10.3201/eid1603.090732 (2010).
23 Adamou, R. *et al.* Plasmodium falciparum infection and age influence parasite growth inhibition mediated by IgG in Beninese infants. *Acta tropica* **159**, 111-119, doi:10.1016/j.actatropica.2016.03.020 (2016).
24 Dodoo, D. *et al.* Cohort study of the association of antibody levels to AMA1, MSP119, MSP3 and GLURP with protection from clinical malaria in Ghanaian children. *Malaria journal* **7**, 142, doi:10.1186/1475-2875-7-142 (2008).
25 Arama, C. *et al.* Genetic Resistance to Malaria Is Associated With Greater Enhancement of Immunoglobulin (Ig)M Than IgG Responses to a Broad Array of Plasmodium falciparum Antigens. *Open forum infectious diseases* **2**, ofv118, doi:10.1093/ofid/ofv118 (2015).
26 Bolad, A. *et al.* Distinct interethnic differences in immunoglobulin G class/subclass and immunoglobulin M antibody responses to malaria antigens but not in immunoglobulin G responses to nonmalarial antigens in sympatric tribes living in West Africa. *Scandinavian journal of immunology* **61**, 380-386, doi:10.1111/j.1365-3083.2005.01587.x (2005).
27 Duah, N. O. *et al.* Heritability of antibody isotype and subclass responses to Plasmodium falciparum antigens. *PLoS One* **4**, e7381, doi:10.1371/journal.pone.0007381 (2009).
28 Rafieian-Kopaei, M., Nasri, H., Alizadeh, F., Ataei, B. & Baradaran, A. Immunoglobulin A Nephropathy and Malaria falciparum Infection; a Rare Association. *Iranian journal of public health* **42**, 529-533 (2013).
29 Yoo, D. E. *et al.* Immunoglobulin A nephropathy associated with Plasmodium falciparum malaria. *Journal of Korean medical science* **27**, 446-449, doi:10.3346/jkms.2012.27.4.446 (2012).
30 Weber, M. W., Boker, K., Horstmann, R. D. & Ehrich, J. H. Renal failure is a common complication in non-immune Europeans with Plasmodium falciparum malaria.

Tropical medicine and parasitology : official organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ) **42**, 115-118 (1991).

31 Martin, B. & Smith, R. J. H. in *GeneReviews((R))* (eds M. P. Adam *et al.*) (University of Washington, Seattle

University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., 1993).

32 Donadio, J. V. & Grande, J. P. IgA nephropathy. *The New England journal of medicine* **347**, 738-748, doi:10.1056/NEJMra020109 (2002).

33 Lomax-Smith, J. D., Zabrowarny, L. A., Howarth, G. S., Seymour, A. E. & Woodroffe, A. J. The immunochemical characterization of mesangial IgA deposits. *Am J Pathol* **113**, 359-364 (1983).

34 Coppo, R. & Amore, A. Aberrant glycosylation in IgA nephropathy (IgAN). *Kidney international* **65**, 1544-1547, doi:10.1111/j.1523-1755.2004.05407.x (2004).

35 Duah, N. O., Miles, D. J., Whittle, H. C. & Conway, D. J. Acquisition of antibody isotypes against Plasmodium falciparum blood stage antigens in a birth cohort. *Parasite immunology* **32**, 125-134, doi:10.1111/j.1365-3024.2009.01165.x (2010).

36 Shi, J. *et al.* The generation and evaluation of recombinant human IgA specific for Plasmodium falciparum merozoite surface protein 1-19 (PfMSP1 19). *BMC biotechnology* **11**, 77, doi:10.1186/1472-6750-11-77 (2011).

37 McIntosh, R. S. *et al.* The importance of human FcgammaRI in mediating protection to malaria. *PLoS Pathog* **3**, e72, doi:10.1371/journal.ppat.0030072 (2007).

38 Jafarshad, A. *et al.* A novel antibody-dependent cellular cytotoxicity mechanism involved in defense against malaria requires costimulation of monocytes FcgammaRII and FcgammaRIII. *Journal of immunology (Baltimore, Md. : 1950)* **178**, 3099-3106 (2007).

39 Mibei, E. K., Otieno, W. O., Orago, A. S. & Stoute, J. A. Distinct pattern of class and subclass antibodies in immune complexes of children with cerebral malaria and severe malarial anaemia. *Parasite immunology* **30**, 334-341, doi:10.1111/j.1365-3024.2008.01030.x (2008).

40 Biswas, S., Saxena, Q. B., Roy, A. & Kabilan, L. Naturally occurring plasmodium-specific IgA antibody in humans from a malaria endemic area. *Journal of Biosciences* **20**, 453-460, doi:10.1007/BF02703849 (1995).

41 Cook, J. *et al.* Trends in parasite prevalence following 13 years of malaria interventions on Bioko island, Equatorial Guinea: 2004–2016. *Malaria journal* **17**, 62, doi:10.1186/s12936-018-2213-9 (2018).

42 Tadesse, F. G. *et al.* The shape of the iceberg: quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia. *Malaria journal* **16**, 99, doi:10.1186/s12936-017-1749-4 (2017).

43 Herman, L. S. *et al.* Identification and validation of a novel panel of Plasmodium knowlesi biomarkers of serological exposure. *PLoS neglected tropical diseases* **12**, e0006457, doi:10.1371/journal.pntd.0006457 (2018).

44 Liu, F. *et al.* Characterization of Plasmodium berghei Pbg37 as Both Pre- and Post-Fertilization Antigen with Transmission-Blocking Potential. *Infection and immunity*, doi:10.1128/iai.00785-17 (2018).

45 Muh, F. *et al.* Identification of a novel merozoite surface antigen of Plasmodium vivax, PvMSA180. *Malaria journal* **16**, 133, doi:10.1186/s12936-017-1760-9 (2017).

46 Zheng, W. *et al.* Functional characterization of Plasmodium berghei PSOP25 during ookinete development and as a malaria transmission-blocking vaccine candidate. *Parasites & vectors* **10**, 8, doi:10.1186/s13071-016-1932-4 (2017).

47 Aguiar, J. C. *et al.* Discovery of Novel Plasmodium falciparum Pre-Erythrocytic Antigens for Vaccine Development. *PLoS One* **10**, e0136109, doi:10.1371/journal.pone.0136109 (2015).

48 Cheng, Y. *et al.* Antigenicity and immunogenicity of PvRALP1, a novel Plasmodium vivax rhoptry neck protein. *Malaria journal* **14**, 186, doi:10.1186/s12936-015-0698-z (2015).

49 Han, J. H. *et al.* Identification of Immunodominant B-cell Epitope Regions of Reticulocyte Binding Proteins in Plasmodium vivax by Protein Microarray Based Immunoscreening. *The Korean journal of parasitology* **53**, 403-411, doi:10.3347/kjp.2015.53.4.403 (2015).

50 Hostetler, J. B. *et al.* A Library of Plasmodium vivax Recombinant Merozoite Proteins Reveals New Vaccine Candidates and Protein-Protein Interactions. *PLoS neglected tropical diseases* **9**, e0004264, doi:10.1371/journal.pntd.0004264 (2015).

51 Saetear, P. *et al.* Quantification of Plasmodium-host protein interactions on intact, unmodified erythrocytes by back-scattering interferometry. *Malaria journal* **14**, 88, doi:10.1186/s12936-015-0553-2 (2015).

52 Gupta, S., Manubhai, K. P., Kulkarni, V. & Srivastava, S. An overview of innovations and industrial solutions in Protein Microarray Technology. *Proteomics* **16**, 1297-1308, doi:10.1002/pmic.201500429 (2016).

53 Chen, Z., Dodig-Crnković, T., Schwenk, J. M. & Tao, S.-c. Current applications of antibody microarrays. *Clinical Proteomics* **15**, 7, doi:10.1186/s12014-018-9184-2 (2018).

54 Taghavian, O. *et al.* Antibody Profiling by Proteome Microarray with Multiplex Isotype Detection Reveals Overlap between Human and Aotus nancymaae Controlled Malaria Infections. *Proteomics*, doi:10.1002/pmic.201700277 (2017).

55 Zhu, H. & Snyder, M. Protein arrays and microarrays. *Current opinion in chemical biology* **5**, 40-45 (2001).

56 Kopf, E. & Zharhary, D. Antibody arrays--an emerging tool in cancer proteomics. *Int J Biochem Cell Biol* **39**, 1305-1317, doi:10.1016/j.biocel.2007.04.029 (2007).

57 Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**, 1720-1730 (1999).

58 Sutandy, F. X., Qian, J., Chen, C. S. & Zhu, H. Overview of protein microarrays. *Current protocols in protein science* **Chapter 27**, Unit 27.21, doi:10.1002/0471140864.ps2701s72 (2013).

59 Bertone, P. & Snyder, M. Advances in functional protein microarray technology. *The FEBS journal* **272**, 5400-5411, doi:10.1111/j.1742-4658.2005.04970.x (2005).

60 Hall, D. A., Ptacek, J. & Snyder, M. Protein Microarray Technology. *Mechanisms of ageing and development* **128**, 161-167, doi:10.1016/j.mad.2006.11.021 (2007).

61 Chen, C. S. & Zhu, H. Protein microarrays. *BioTechniques* **40**, 423, 425, 427 passim, doi:10.2144/06404te01 (2006).

62 Sreekumar, A. *et al.* Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res* **61**, 7585-7593 (2001).

63 Poetz, O. *et al.* Protein microarrays: catching the proteome. *Mechanisms of ageing and development* **126**, 161-170, doi:10.1016/j.mad.2004.09.030 (2005).

64 Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101-2105, doi:10.1126/science.1062191 (2001).

65 Wachter, A., Bernhardt, S., Beissbarth, T. & Korf, U. Analysis of Reverse Phase Protein Array Data: From Experimental Design towards Targeted Biomarker Discovery. *Microarrays* **4**, 520-539, doi:10.3390/microarrays4040520 (2015).

66 Baum, E. *et al.* Submicroscopic and asymptomatic Plasmodium falciparum and Plasmodium vivax infections are common in western Thailand - molecular and

serological evidence. *Malaria journal* **14**, 95, doi:10.1186/s12936-015-0611-9 (2015).

67 Baum, E. *et al.* Common asymptomatic and submicroscopic malaria infections in Western Thailand revealed in longitudinal molecular and serological studies: a challenge to malaria elimination. *Malaria journal* **15**, 333, doi:10.1186/s12936-016-1393-4 (2016).

68 Uplekar, S. *et al.* Characterizing Antibody Responses to Plasmodium vivax and Plasmodium falciparum Antigens in India Using Genome-Scale Protein Microarrays. *PLoS neglected tropical diseases* **11**, e0005323, doi:10.1371/journal.pntd.0005323 (2017).

69 Dottorini, T. *et al.* Serum IgE Reactivity Profiling in an Asthma Affected Cohort. *PLOS ONE* **6**, e22319, doi:10.1371/journal.pone.0022319 (2011).

70 Gray, J. C. *et al.* Profiling the antibody immune response against blood stage malaria vaccine candidates. *Clinical chemistry* **53**, 1244-1253, doi:10.1373/clinchem.2006.081695 (2007).

71 Chan, W. C. & Nie, S. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science* **281**, 2016-2018 (1998).

72 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47, doi:10.1093/nar/gkv007 (2015).

73 Francis, J. E., Mason, D. & Lévy, R. Evaluation of quantum dot conjugated antibodies for immunofluorescent labelling of cellular targets. *Beilstein Journal of Nanotechnology* **8**, 1238-1249, doi:10.3762/bjnano.8.125 (2017).

74 Zhao, J. J., Chen, J., Wang, Z. P., Pan, J. & Huang, Y. H. Double labeling and comparison of fluorescence intensity and photostability between quantum dots and FITC in oral tumors. *Molecular medicine reports* **4**, 425-429, doi:10.3892/mmr.2011.457 (2011).

75 Jaiswal, J. K., Mattoussi, H., Mauro, J. M. & Simon, S. M. Long-term multiple color imaging of live cells using quantum dot bioconjugates. *Nature Biotechnology* **21**, 47, doi:10.1038/nbt767 (2002).

76 DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686 (1997).

77 Fields, S., Kohara, Y. & Lockhart, D. J. Functional genomics. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 8825-8826 (1999).

78 Winzeler, E. A. *et al.* Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194-1197 (1998).

79 Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**, 901-906 (1999).

80 Alberts, B. *et al.* Ch. 8. Manipulating Proteins, DNA, and RNA, (Garland Science, 2002).

81 Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623-627, doi:10.1038/35001009 (2000).

82 Murungi, L. M. *et al.* Targets and Mechanisms Associated with Protection from Severe Plasmodium falciparum Malaria in Kenyan Children. *Infection and immunity* **84**, 950-963, doi:10.1128/iai.01120-15 (2016).

83 Plotkin, S. A. Correlates of Protection Induced by Vaccination. *Clinical and vaccine immunology : CVI* **17**, 1055-1065, doi:10.1128/CVI.00131-10 (2010).

84 Collins, K. A., Snaith, R., Cottingham, M. G., Gilbert, S. C. & Hill, A. V. S. Enhancing protective immunity to malaria with a highly immunogenic virus-like particle vaccine. *Sci Rep* **7**, 46621, doi:10.1038/srep46621 (2017).

85 Krishnamurty, A. T. *et al.* Somatically Hypermutated Plasmodium-Specific IgM(+) Memory B Cells Are Rapid, Plastic, Early Responders upon Malaria Rechallenge. *Immunity* **45**, 402-414, doi:10.1016/j.immuni.2016.06.014 (2016).

86 Kumar, A. *et al.* Kix domain specific Immunoglobulin A can protect from adverse lung and cerebral pathology induced by Plasmodium berghei ANKA. *Biochem Biophys Res Commun* **464**, 943-948, doi:10.1016/j.bbrc.2015.07.075 (2015).

Chapter 7 : GENERAL DISCUSSION

For many years, microscopy has been the gold standard in identification of *Plasmodium* spp. infection. However, in distinguishing between species within this genus, it is limited when they have a similar morphological appearance[1-6]. There is therefore an urgent need for new techniques that can achieve greater sensitivity, specificity and depth of analysis. We have successfully developed a panel of ten *P. knowlesi* recombinant antigens. The results presented here provide important insights in the use of high throughput technology in serology and aid understanding of antibody isotype responses towards *P. knowlesi* infection in an endemic setting. We have shown for the first time proof of principle that multiplexing of *P. knowlesi* antigens can be used for serological assessment of human samples using the Multiplex bead Assay (MBA) and the protein microarray platforms. Two of these antigens have already been used as serosurveillance tools in two community surveys in Sabah, Malaysia, using the high throughput MBA technology.

Interest in high throughput multiplex technology is rapidly increasing, particularly in the serology field. Modern advances in multiplexing technology have made it possible for us to not only quantitatively assess reactivity towards multiple targets at once, but also to simultaneously measure the qualitative host response to antigen exposure in the form of antibody isotypes and subclasses[7]. Detecting multiple targets in a single sample and analysing these in the context of multiple others as well as the host immune system has long been pursued. This is now possible with MBA and protein microarray technology, which not only enable multiplexing, but also have a low sample requirement, making it possible to screen thousands of samples for thousands of targets.

Between the 1970s and 1980s, it was determined that Eukaryotic DNA could be replicated and transcribed in *Escherichia coli*[8] and this was followed by the

development of the PCR method[9], discoveries which enabled the development of recombinant gene expression technologies. Currently, genomic, proteomic and transcriptomic datasets are widely available, making it possible to use sequence-based approaches to develop recombinant antigens.

With these advances it is now possible to express specific genes of interest from a number of different organisms and many different expression platforms have since been developed[10-14]. These include yeast expression systems[10,15,16], insect cell lines combined with recombinant viruses[13], mammalian cell lines[17-19], bacterial expression systems[12,20] and cell-free systems[21,22], just to name a few, each with their advantages and disadvantages. Recombinant proteins are an important part of drug discovery[23], vaccine research[12,24] and now serology[25,26] as well. The development of recombinant antigens is an arduous task in itself, as reported in a number of reviews[27,28], and this project was no different. One of the major setbacks in developing these antigens was during the cloning phase of the production pipeline, with some candidate clones incorporating errors or producing low yields while other candidates refused to clone and transform after a series of troubleshooting attempts. In future, due to our large and increasing number of candidates, we would outsource the cloning stage of our development pipeline in order to reduce the time spent on this phase. By reducing the time spent on cloning troubleshooting, more time would be available to focus on the development and characterisation of antigens, ensuring specificity. Alternatively, we would consider applying a wider range of cloning techniques and determine the most suitable method for the project. We could potentially outsource the expression of our antigens as well, not only ensuring that they do express but also in order to obtain similar yields for each antigen and try to reduce variability between batches.

Developing recombinant antigens is one thing, determining their usefulness is another, and this varies with the field it will be intended for. In our case, we would

eventually like to determine whether any of these *P. knowlesi* recombinant antigens or antigen combinations could be useful predictors of protection, for instance. There have been many studies indicating associations between specific malarial antigen reactivity and protection. Evidence for the Sporozoite Surface Protein 2 (SSP2) suggested an immunogenic antigen involved in protection from disease in mice[29] and in human vaccine studies[30,31]. Antibodies targeting antigen Merozoite Surface Protein 1-19 (MSP1-19), shown to be highly immunogenic in both *P. falciparum*[32] and *P. vivax*[33], appeared to be associated with protection[34-37]. Antibodies against *Pf*SERA5 are thought to inhibit parasite growth [38,39] and to induce protection against blood stage *in vivo* infection[40]. On the other hand, it has also been shown that, rather than a single antigen, a panel of 19 different *P. falciparum* antigens were shown to be a potential signature associated with sporozoite-induced protective immunity in volunteers immunised with radiation attenuated *P. falciparum* sporozoites[41]. Other studies have shown similar outputs[42,43]. However, in the *P. knowlesi* field, there have not been many in depth studies such as these, and most were in macaque studies[44-48]. It is imperative that we not only develop new antigens, but also that we study their individual effects as well as their combined effects on the host immune system. Antigen recognition profiles using the protein microarray platform for instance, can not only be used to determine correlations with protection to disease but also with clinical disease stage classification[41,49], susceptibility[41] or in vitro virus neutralising activity[50] for example, all of which can determine antigens which might be associated with protection. The advantage of using high throughput technology is largely the high throughput serum analyses it enables, making it possible to screen thousands of samples for thousands of targets. It is important that we develop new antigens but we must be aware of the limitations which exist in antigen discovery, for instance that expression and folding of the recombinant proteins might be different from the native structure. Another point to be aware of is that an antigen might elicit a cellular immune response rather than a humoral

response, which means analyses of responses towards new antigens should not be restricted to antibody reactivity. Determining T cell targets can be more challenging than determining antibody targets but several approaches have been developed to tackle this, such as epitope-prediction algorithms[51] and high throughput cellular assays[52]. Cell-free in vitro transcription translation (IVTT) strategies have been used to identify targets of cellular immunity for influenza, cytomegalovirus and Epstein Barr virus in humans[53] and *Plasmodium* spp. in mice[53], measuring T cell activation. These cell-free systems are a successful alternative to producing proteins from complex organisms, such as *Plasmodium* spp., which have shown difficulties in cell-based systems[54,55]. In future, alternative approaches such as this would be considered in order to produce *P. knowlesi* recombinant antigens which had proven difficult to produce.

Given enough time, we would have liked to further expand our panel of *P. knowlesi* recombinant antigens and characterise them. We would have also liked to determine the ideal storage conditions for the individual antigens by performing an accelerated stability testing study. This would determine how stable the individual antigens are to specific conditions and how long they are stable for. Lasting high MFI values would be the ideal output for every antigen in all conditions.

As previously discussed in this thesis, the presence or absence of antibodies can be used to infer exposure to infection, in this case, *P. knowlesi* malaria. We have successfully tested Malaysian clinical samples against our panel of recombinant antigens using the MBA platform and not only found similar results to the ELISA, but we were able to confirm that the MBA platform is more sensitive than the ELISA. Because the MBA platform is more sensitive, we observed some level of cross-reactivity in the MBA analysis of the clinical samples. In hindsight it would have been ideal to have developed a number of species-specific antigens for both *P. vivax* and *P. falciparum* to test alongside the *P. knowlesi* antigens to be used as true control antigens in the assessment of cross reactivity.

After testing the clinical samples, for the first time, we examined sera at the community level using our panel of recombinant *P. knowlesi* antigens. Prior to this study it has only been possible to assess the presence of *P. knowlesi* in Malaysia from clinical research. Not only does this underestimate the presence of *P. knowlesi* at the community level, but as discussed, it is difficult to attribute the symptoms of malaria to a specific species. We were able to demonstrate surprisingly high levels of exposure to the parasite (PkSERA3 antigen 2: 31% (3139/10125) seropositive), and from an early age. Seropositivity was detected in children under 5 years and became more frequent with increase in age. This indicates that *P. knowlesi* malaria is being actively transmitted within those populations and has the potential to worsen if left unchecked. Although there was no statistically significant association between seropositivity and sleeping in the forest, a reasonably high number of people who did sleep in forests were seropositive to *P. knowlesi* (PkSERA3 antigen 2, 29.41% (10/34); PkSSP2, 20% (4/20)) and seeing macaques was associated with seropositivity ($p<0.0001$). Other studies have shown that occupations connected to forestry and farming led to higher parasitemias[5]. Not only is human behaviour connected to acquiring *P. knowlesi* infection but potentially macaque behaviour as well[56]. With the ever increasing rates of deforestation in Malaysia[4,56], humans and wild macaques are brought closer together, which could increase the risk of infection in humans[56]. This increase in risk of *P. knowlesi* infection could compromise malaria elimination strategies in several countries if not attended to as soon as possible, especially because it is zoonotic. Malaria elimination strategies have not taken into account the existence of animal reservoirs besides the insect vector, because until recently, the human disease was thought to be caused purely by an obligate human parasite. With the zoonotic *P. knowlesi* recently being classified as the fifth human malaria parasite[57], this story has changed, and now we must remember the original host to this organism, macaques. Current malaria control strategies include the use of antimalarial drugs combined with the use of bed nets and insecticides[58], and now

a recently licensed vaccine against malaria exists, although with modest efficacy[59]. There are many challenges when it comes to the eradication of malaria due to the incorporation of multiple approaches, interventions, disciplines and organisations all leading to malaria surveillance, transmission interruption, and treatment programs[58,60]. One of the challenges of elimination for instance is *P. vivax*, which unlike *P. falciparum*, has a dormant liver-stage (hypnozoite) which can relapse many months or even years after the initial infection[61,62]. Rapid diagnostic tests (RDTs), routinely used for malaria diagnosis by many health facilities, detecting *P. falciparum* histidine-rich protein 2 and 3 (*Pf*hrp2 and *Pf*hrp3) have failed to detect malaria parasites in various countries[63-67], due to a deletion of the gene encoding those proteins[65,66]. *P. knowlesi* elimination faces similar challenges purely because of the existence of the macaque animal reservoir[68].

The work presented here has potential applications for other emerging *Plasmodium* species. *P. knowlesi* is not the only monkey malaria parasite to make itself known as infecting humans. *P. cynomolgi* has recently been found to naturally infects humans as well, although it is not as common[69]. *P. cynomolgi* is morphologically similar to *P. vivax*[1], therefore diagnosis of this parasite could be as difficult as it is for *P. knowlesi.*

There has been much interest in recent years in trying to establish whether serological markers after challenge can be useful predictors of protection particularly in the field of vaccinology[70]. There have been high-profile studies such as in the development of the MVA85A vaccine, that have shown that we do not yet fully understand this field, particularly across different species[71,72]. It is becoming clear, however, that the quality of an immune response is at least as important as its magnitude, for instance the avidity of the response and the antigen or antigens it is directed towards.

As it stands, we find it necessary to further develop our current panel of *P. knowlesi* antigens and to expand the number of antigens currently on it in order to better dissect the epidemiology of *P. knowlesi* in Southeast Asia. In future projects we

would like to use the next generation sequenced genome[73] to further develop and increase the current *P. knowlesi* panel of recombinant antigens. We would also aim to raise polyclonal antibodies to each antigen in order to localise them on the parasite and further characterise their function in the parasite. Large panels of antigens have previously been used to successfully determine reactivity patterns in small and large groups of individuals[42,74] and learning about the intricacies of these antigens could help us understand these patterns as well.

Here we have described a novel panel of antigens and show the potential these antigens have in determining patterns of reactivity towards *P. knowlesi* infection in humans. Based on this, it will be possible to determine antibody patterns associated with protection and risk. Establishing associations between seropositive individuals and specific activities, occupations and/or geographic locations will enable the design of highly tailored interventions with the purpose of reducing the risk of infection and disease, leading to the eradication of *P. knowlesi* malaria in those areas. Once we have established the existence of associations and appropriate management strategies are put in place, it will be possible to achieve the goal of eradicating malaria in Southeast Asia.

THE END

REFERENCES

1 Coatney, G. R., Collins, W. E. & Contacos, P. G. *The Primate malarias*. (U.S. National Institute of Allergy and Infectious Diseases, 1971).

2 Lee, K. S., Cox-Singh, J. & Singh, B. Morphological features and differential counts of Plasmodium knowlesi parasites in naturally acquired human infections. *Malaria journal* **8**, 73, doi:10.1186/1475-2875-8-73 (2009).

3 Barber, B. E., William, T., Grigg, M. J., Yeo, T. W. & Anstey, N. M. Limitations of microscopy to differentiate Plasmodium species in a region co-endemic for Plasmodium falciparum, Plasmodium vivax and Plasmodium knowlesi. *Malaria journal* **12**, 8, doi:10.1186/1475-2875-12-8 (2013).

4 Singh, B. & Daneshvar, C. Human infections and detection of Plasmodium knowlesi. *Clin Microbiol Rev* **26**, 165-184, doi:10.1128/cmr.00079-12 (2013).

5 Barber, B. E. *et al.* A prospective comparative study of knowlesi, falciparum, and vivax malaria in Sabah, Malaysia: high proportion with severe disease from Plasmodium knowlesi and Plasmodium vivax but no mortality with early referral and artesunate therapy. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **56**, 383-397, doi:10.1093/cid/cis902 (2013).

6 Rajahram, G. S. *et al.* Deaths due to Plasmodium knowlesi malaria in Sabah, Malaysia: association with reporting as Plasmodium malariae and delayed parenteral artesunate. *Malaria journal* **11**, 284, doi:10.1186/1475-2875-11-284 (2012).

7 Taghavian, O. *et al.* Antibody Profiling by Proteome Microarray with Multiplex Isotype Detection Reveals Overlap between Human and Aotus nancymaae Controlled Malaria Infections. *Proteomics*, doi:10.1002/pmic.201700277 (2017).

8 Morrow, J. F. *et al.* Replication and transcription of eukaryotic DNA in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 1743-1747 (1974).

9 Saiki, R. K. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491 (1988).

10 Nielsen, K. H. Protein expression-yeast. *Methods in enzymology* **536**, 133-147, doi:10.1016/b978-0-12-420070-8.00012-x (2014).

11 Fernandez, F. J. & Vega, M. C. Choose a Suitable Expression Host: A Survey of Available Protein Production Platforms. *Advances in experimental medicine and biology* **896**, 15-24, doi:10.1007/978-3-319-27216-0_2 (2016).

12 Overton, T. W. Recombinant protein production in bacterial hosts. *Drug Discov Today* **19**, 590-601, doi:10.1016/j.drudis.2013.11.008 (2014).

13 Geisler, C. & Jarvis, D. L. Adventitious viruses in insect cell lines used for recombinant protein expression. *Protein expression and purification* **144**, 25-32, doi:10.1016/j.pep.2017.11.002 (2018).

14 Legastelois, I. *et al.* Non-conventional expression systems for the production of vaccine proteins and immunotherapeutic molecules. *Hum Vaccin Immunother* **13**, 947-961, doi:10.1080/21645515.2016.1260795 (2017).

15 Baghban, R. *et al.* Yeast Expression Systems: Overview and Recent Advances. *Molecular biotechnology*, doi:10.1007/s12033-019-00164-8 (2019).

16 Bathurst, I. C. Protein expression in yeast as an approach to production of recombinant malaria antigens. *Am J Trop Med Hyg* **50**, 20-26 (1994).

17 Li, F., Vijayasankaran, N., Shen, A. Y., Kiss, R. & Amanullah, A. Cell culture processes for monoclonal antibody production. *mAbs* **2**, 466-479 (2010).

18 Lin, J. *et al.* Genetic ablation of luteinizing hormone receptor improves the amyloid pathology in a mouse model of Alzheimer disease. *J Neuropathol Exp Neurol* **69**, 253-261, doi:10.1097/NEN.0b013e3181d072cf (2010).

19 Jenkins, N., Meleady, P., Tyther, R. & Murphy, L. Strategies for analysing and improving the expression and quality of recombinant proteins made in mammalian cells. *Biotechnology and applied biochemistry* **53**, 73-83, doi:10.1042/ba20080258 (2009).

20 Lee, M. J. & Kim, P. Recombinant Protein Expression System in Corynebacterium glutamicum and Its Application. *Frontiers in microbiology* **9**, 2523, doi:10.3389/fmicb.2018.02523 (2018).

21 Zaheer, R., Klima, C. L. & McAllister, T. A. Expeditious screening of candidate proteins for microbial vaccines. *Journal of microbiological methods* **116**, 53-59, doi:10.1016/j.mimet.2015.06.018 (2015).

22 Takai, K., Sawasaki, T. & Endo, Y. Practical cell-free protein synthesis system using purified wheat embryos. *Nature protocols* **5**, 227-238, doi:10.1038/nprot.2009.207 (2010).

23 Assenberg, R., Wan, P. T., Geisse, S. & Mayr, L. M. Advances in recombinant protein expression for use in pharmaceutical research. *Current opinion in structural biology* **23**, 393-402, doi:10.1016/j.sbi.2013.03.008 (2013).

24 Heguy, A., Stewart, A. A., Haley, J. D., Smith, D. E. & Foulkes, J. G. Gene expression as a target for new drug discovery. *Gene expression* **4**, 337-344 (1995).

25 Jones, S. *et al.* Naturally acquired antibody responses to recombinant Pfs230 and Pfs48/45 transmission blocking vaccine candidates. *Journal of Infection* **71**, 117-127, doi:http://dx.doi.org/10.1016/j.jinf.2015.03.007 (2015).

26 King, C. L. *et al.* Biosignatures of Exposure/Transmission and Immunity. *Am J Trop Med Hyg* **93**, 16-27, doi:10.4269/ajtmh.15-0037 (2015).

27 Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in microbiology* **5**, 172, doi:10.3389/fmicb.2014.00172 (2014).

28 Celie, P. H., Parret, A. H. & Perrakis, A. Recombinant cloning strategies for protein expression. *Current opinion in structural biology* **38**, 145-154, doi:10.1016/j.sbi.2016.06.010 (2016).

29 Khusmith, S. *et al.* Protection against malaria by vaccination with sporozoite surface protein 2 plus CS protein. *Science* **252**, 715-718 (1991).

30 Roestenberg, M. *et al.* Protection against a malaria challenge by sporozoite inoculation. *The New England journal of medicine* **361**, 468-477, doi:10.1056/NEJMoa0805832 (2009).

31 Felgner, P. L. *et al.* Pre-erythrocytic antibody profiles induced by controlled human malaria infections in healthy volunteers under chloroquine prophylaxis. *Sci Rep* **3**, 3549, doi:10.1038/srep03549 (2013).

32 O'Donnell, R. A. *et al.* Antibodies against Merozoite Surface Protein (Msp)-1(19) Are a Major Component of the Invasion-Inhibitory Response in Individuals Immune to Malaria. *The Journal of experimental medicine* **193**, 1403-1412 (2001).

33 Soares, I. S., Levitus, G., Souza, J. M., Del Portillo, H. A. & Rodrigues, M. M. Acquired immune responses to the N- and C-terminal regions of Plasmodium vivax merozoite surface protein 1 in individuals exposed to malaria. *Infection and immunity* **65**, 1606-1614 (1997).

34 Blackman, M. J., Heidrich, H. G., Donachie, S., McBride, J. S. & Holder, A. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion- inhibiting antibodies. *The Journal of experimental medicine* **172**, 379-382 (1990).

35 Tadesse, F. G. *et al.* The shape of the iceberg: quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia. *Malaria journal* **16**, 99, doi:10.1186/s12936-017-1749-4 (2017).

36 Ssewanyana, I. *et al.* Avidity of anti-malarial antibodies inversely related to transmission intensity at three sites in Uganda. *Malaria journal* **16**, 67, doi:10.1186/s12936-017-1721-3 (2017).

37 Moss, D. K. *et al.* Plasmodium falciparum 19-kilodalton merozoite surface protein 1 (MSP1)-specific antibodies that interfere with parasite growth in vitro can inhibit MSP1 processing, merozoite invasion, and intracellular parasite development. *Infection and immunity* **80**, 1280-1287, doi:10.1128/iai.05887-11 (2012).

38 Aoki, S. *et al.* Serine repeat antigen (SERA5) is predominantly expressed among the SERA multigene family of Plasmodium falciparum, and the acquired antibody titers correlate with serum inhibition of the parasite growth. *J Biol Chem* **277**, 47533-47540, doi:10.1074/jbc.M207145200 (2002).

39 Chulay, J. D. *et al.* Monoclonal antibody characterization of Plasmodium falciparum antigens in immune complexes formed when schizonts rupture in the presence of immune serum. *Journal of immunology (Baltimore, Md. : 1950)* **139**, 2768-2774 (1987).

40 Perrin, L. H. *et al.* Antimalarial immunity in Saimiri monkeys. Immunization with surface components of asexual blood stages. *The Journal of experimental medicine* **160**, 441-451 (1984).

41 Trieu, A. *et al.* Sterile protective immunity to malaria is associated with a panel of novel P. falciparum antigens. *Molecular & cellular proteomics : MCP* **10**, M111.007948, doi:10.1074/mcp.M111.007948 (2011).

42 Osier, F. H. *et al.* Breadth and magnitude of antibody responses to multiple Plasmodium falciparum merozoite antigens are associated with protection from clinical malaria. *Infection and immunity* **76**, 2240-2248, doi:10.1128/iai.01585-07 (2008).

43 John, C. C. *et al.* Correlation of high levels of antibodies to multiple pre-erythrocytic Plasmodium falciparum antigens and protection from infection. *Am J Trop Med Hyg* **73**, 222-228 (2005).

44 Schmidt-Ullrich, R. *et al.* Rhesus monkeys protected against Plasmodium knowlesi malaria produce antibodies against a 65,000-MrP. knowlesi glycoprotein at the surface of infected erythrocytes. *Infection and immunity* **34**, 519-525 (1981).

45 Deans, J. A. Protective antigens of bloodstage Plasmodium knowlesi parasites. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **307**, 159-169 (1984).

46 Rogers, W. O. *et al.* Multistage multiantigen heterologous prime boost vaccine for Plasmodium knowlesi malaria provides partial protection in rhesus macaques. *Infection and immunity* **69**, 5565-5572 (2001).

47 Rogers, W. O. *et al.* Protection of rhesus macaques against lethal Plasmodium knowlesi malaria by a heterologous DNA priming and poxvirus boosting immunization regimen. *Infection and immunity* **70**, 4329-4335 (2002).

48 Jiang, G. *et al.* Sterile Protection against Plasmodium knowlesi in Rhesus Monkeys from a Malaria Vaccine: Comparison of Heterologous Prime Boost Strategies. *PLOS ONE* **4**, e6559, doi:10.1371/journal.pone.0006559 (2009).

49 Barbour, A. G. *et al.* A genome-wide proteome array reveals a limited set of immunogens in natural infections of humans and white-footed mice with Borrelia burgdorferi. *Infection and immunity* **76**, 3374-3389, doi:10.1128/iai.00048-08 (2008).

50 Davies, D. H. *et al.* Proteome-wide analysis of the serological response to vaccinia and smallpox. *Proteomics* **7**, 1678-1686, doi:10.1002/pmic.200600926 (2007).
51 Peters, B. *et al.* The immune epitope database and analysis resource: from vision to blueprint. *PLoS biology* **3**, e91, doi:10.1371/journal.pbio.0030091 (2005).
52 Seder, R. A., Darrah, P. A. & Roederer, M. T-cell quality in memory and protection: implications for vaccine design. *Nature reviews. Immunology* **8**, 247-258, doi:10.1038/nri2274 (2008).
53 Cardoso, F. C., Roddick, J. S., Groves, P. & Doolan, D. L. Evaluation of approaches to identify the targets of cellular immunity on a proteome-wide scale. *PLoS One* **6**, e27666, doi:10.1371/journal.pone.0027666 (2011).
54 Doolan, D. L. Plasmodium immunomics. *Int J Parasitol* **41**, 3-20, doi:10.1016/j.ijpara.2010.08.002 (2011).
55 Doolan, D. L. *et al.* Profiling humoral immune responses to P. falciparum infection with protein microarrays. *Proteomics* **8**, 4680-4694, doi:10.1002/pmic.200800194 (2008).
56 Fornace, K. M. *et al.* Association between Landscape Factors and Spatial Patterns of Plasmodium knowlesi Infections in Sabah, Malaysia. *Emerging infectious diseases* **22**, 201-209, doi:10.3201/eid2202.150656 (2016).
57 White, N. J. Plasmodium knowlesi: the fifth human malaria parasite. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **46**, 172-173, doi:10.1086/524889 (2008).
58 Shretta, R. *et al.* in *Major Infectious Diseases* (eds rd *et al.*) (The International Bank for Reconstruction and Development / The World Bank

(c) 2017 International Bank for Reconstruction and Development / The World Bank., 2017).
59 Kaslow, D. C. & Biernaux, S. RTS,S: Toward a first landmark on the Malaria Vaccine Technology Roadmap. *Vaccine* **33**, 7425-7432, doi:10.1016/j.vaccine.2015.09.061 (2015).
60 Rabinovich, R. N. *et al.* malERA: An updated research agenda for malaria elimination and eradication. *PLoS medicine* **14**, e1002456, doi:10.1371/journal.pmed.1002456 (2017).
61 White, N. J. & Imwong, M. in *Advances in parasitology* Vol. 80 (eds S. I. Hay, Ric Price, & J. Kevin Baird) 113-150 (Academic Press, 2012).
62 Mueller, I. *et al.* Key gaps in the knowledge of Plasmodium vivax, a neglected human malaria parasite. *The Lancet infectious diseases* **9**, 555-566, doi:10.1016/s1473-3099(09)70177-x (2009).
63 Kumar, N. *et al.* Genetic deletion of HRP2 and HRP3 in Indian Plasmodium falciparum population and false negative malaria rapid diagnostic test. *Acta tropica* **125**, 119-121, doi:10.1016/j.actatropica.2012.09.015 (2013).
64 Koita, O. A. *et al.* False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. *Am J Trop Med Hyg* **86**, 194-198, doi:10.4269/ajtmh.2012.10-0665 (2012).
65 Gamboa, D. *et al.* A large proportion of P. falciparum isolates in the Amazon region of Peru lack pfhrp2 and pfhrp3: implications for malaria rapid diagnostic tests. *PLoS One* **5**, e8091, doi:10.1371/journal.pone.0008091 (2010).
66 Houze, S., Hubert, V., Le Pessec, G., Le Bras, J. & Clain, J. Combined deletions of pfhrp2 and pfhrp3 genes result in Plasmodium falciparum malaria false-negative rapid diagnostic test. *Journal of clinical microbiology* **49**, 2694-2696, doi:10.1128/jcm.00281-11 (2011).
67 Berhane, A. *et al.* Rapid diagnostic tests failing to detect Plasmodium falciparum infections in Eritrea: an investigation of reported false negative RDT results. *Malaria journal* **16**, 105, doi:10.1186/s12936-017-1752-9 (2017).

68 Brock, P. M. *et al.* Plasmodium knowlesi transmission: integrating quantitative approaches from epidemiology and ecology to understand malaria as a zoonosis. *Parasitology* **143**, 389-400, doi:10.1017/S0031182015001821 (2016).
69 Ta, T. H. *et al.* First case of a naturally acquired human infection with Plasmodium cynomolgi. *Malaria journal* **13**, 68, doi:10.1186/1475-2875-13-68 (2014).
70 Plotkin, S. A. Correlates of Protection Induced by Vaccination. *Clinical and vaccine immunology : CVI* **17**, 1055-1065, doi:10.1128/CVI.00131-10 (2010).
71 Bhatt, K., Verma, S., Ellner, J. J. & Salgame, P. Quest for Correlates of Protection against Tuberculosis. *Clinical and Vaccine Immunology* **22**, 258-266, doi:10.1128/cvi.00721-14 (2015).
72 Cardona, P.-J. What We Have Learned and What We Have Missed in Tuberculosis Pathophysiology for a New Vaccine Design: Searching for the “Pink Swan”. *Frontiers in immunology* **8**, 556, doi:10.3389/fimmu.2017.00556 (2017).
73 Benavente, E. D. *et al.* A reference genome and methylome for the Plasmodium knowlesi A1-H.1 line. *Int J Parasitol*, doi:10.1016/j.ijpara.2017.09.008 (2017).
74 Dottorini, T. *et al.* Serum IgE Reactivity Profiling in an Asthma Affected Cohort. *PLOS ONE* **6**, e22319, doi:10.1371/journal.pone.0022319 (2011).

APPENDIX I: PUBLICATION

RESEARCH ARTICLE

Identification and validation of a novel panel of *Plasmodium knowlesi* biomarkers of serological exposure

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Lou Salomé Herman
Principal Supervisor	Kevin Tetteh
Thesis Title	Development of Plasmodium knowlesi species specific reagents to help characterise antibody isotype profiles in endemic human populations and experimental macaque infections

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	PLOS Neglected Tropical Diseases		
When was the work published?	June, 14, 2018		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	ATTACHED

Student Signature: Lou Herman. **Date:** 03/09/2018

Supervisor Signature: _____ **Date:** 11/9/18

Section D – Multi-authored work

Author Contributions

Conceptualization: Lou S. Herman, Michael J. Blackman, Chris J. Drakeley, Kevin K. A. Tetteh.

Data curation: Lou S. Herman, Jody Phelan.

Formal analysis: Lou S. Herman, Kimberly Fornace, Jody Phelan, Kevin K. A. Tetteh.

Funding acquisition: Chris J. Drakeley.

Investigation: Lou S. Herman, Kimberly Fornace, Kevin K. A. Tetteh.

Methodology: Lou S. Herman, Kimberly Fornace, Robert W. Moon, Michael J. Blackman, Kevin K. A. Tetteh.

Project administration: Matthew J. Grigg, Nicholas M. Anstey, Timothy William, Kevin K. A. Tetteh.

Resources: Kimberly Fornace, Matthew J. Grigg, Nicholas M. Anstey, Timothy William, Robert W. Moon, Chris J. Drakeley.

Supervision: Kevin K. A. Tetteh.

Visualization: Kevin K. A. Tetteh.

Writing - original draft: Lou S. Herman, Kevin K. A. Tetteh.

Writing - review & editing: Lou S. Herman, Kimberly Fornace, Jody Phelan, Matthew J. Grigg, Chris J. Drakeley, Kevin K. A. Tetteh.

PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Identification and validation of a novel panel of *Plasmodium knowlesi* biomarkers of serological exposure

Lou S. Herman[1], Kimberly Fornace[1], Jody Phelan[1], Matthew J. Grigg[2,3], Nicholas M. Anstey[2,3], Timothy William[3,4,5], Robert W. Moon[1], Michael J. Blackman[1,6], Chris J. Drakeley[1], Kevin K. A. Tetteh[1]*

1 Department Immunology and Infection, London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Menzies School of Health Research and Charles Darwin University, Darwin, Northern Territory, Australia, **3** Infectious Diseases Society Sabah-Menzies School of Health Research Clinical Research Unit, Kota Kinabalu, Sabah, Malaysia, **4** Clinical Research Centre, Queen Elizabeth Hospital, Kota Kinabalu, Sabah, Malaysia, **5** Jesselton Medical Centre, Kota Kinabalu, Sabah, Malaysia, **6** Malaria Biochemistry Laboratory, The Francis Crick Institute, London, United Kingdom

* Kevin.tetteh@lshtm.ac.uk

OPEN ACCESS

Citation: Herman LS, Fornace K, Phelan J, Grigg MJ, Anstey NM, William T, et al. (2018) Identification and validation of a novel panel of *Plasmodium knowlesi* biomarkers of serological exposure. PLoS Negl Trop Dis 12(6): e0006457. https://doi.org/10.1371/journal.pntd.0006457

Editor: Alyssa E. Barry, Walter and Eliza Hall Institute, AUSTRALIA

Received: December 5, 2017

Accepted: April 17, 2018

Published: June 14, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The Wellcome Trust (091924/Z/10/Z) and the Medical Research Council (G1100796; https://www.mrc.ac.uk/). RWM is supported by an MRC Career Development Award (MR/M021157/1) jointly funded by the UK Medical Research Council and UK Department for International Development. The funders had no role in study design, data

Abstract

Background

Plasmodium knowlesi is the most common cause of malaria in Malaysian Borneo, with reporting limited to clinical cases presenting to health facilities and scarce data on the true extent of transmission. Serological estimations of transmission have been used with other malaria species to garner information about epidemiological patterns. However, there are a distinct lack of suitable serosurveillance tools for this neglected disease.

Methodology/Principal findings

Using *in silico* tools, we designed and expressed four novel *P. knowlesi* protein products to address the distinct lack of suitable serosurveillance tools: *Pk*SERA3 antigens 1 and 2, *Pk*SSP2/TRAP and *Pk*TSERA2 antigen 1. Antibody prevalence to these antigens was determined by ELISA for three time-points post-treatment from a hospital-based clinical treatment trial in Sabah, East Malaysia (n = 97 individuals; 241 total samples for all time points). Higher responses were observed for the *Pk*SERA3 antigen 2 (67%, 65/97) across all time-points (day 0: 36.9% 34/92; day 7: 63.8% 46/72; day 28: 58.4% 45/77) with significant differences between the clinical cases and controls (n = 55, mean plus 3 SD) (day 0 $p<0.0001$; day 7 $p<0.0001$; day 28 $p<0.0001$). Using boosted regression trees, we developed models to classify *P. knowlesi* exposure (cross-validated AUC 88.9%; IQR 86.1–91.3%) and identified the most predictive antibody responses.

Conclusions/Significance

The *Pk*SERA3 antigen 2 had the highest relative variable importance in all models. Further validation of these antigens is underway to determine the specificity of these tools in the context of multi-species infections at the population level.

collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Malaria caused by *Plasmodium knowlesi* is the most common form of the disease in Malaysia. The parasite is transmitted from monkeys to humans via the bite of an infected mosquito, with the resulting *P. knowlesi* infection potentially leading to severe symptoms and in some cases, death. Although adult males working close to areas with infected monkeys are at the greatest risk of infection, the true extent of the geographical boundaries of *P. knowlesi* transmission is as yet unknown. The ability to measure antibodies to infection is a powerful technique that would help to address this deficit. However, currently available recombinant proteins lack the required specificity for this role. Here, we have developed a panel of recombinant proteins for eventual use as serological tools, strongly supported by robust statistical methods. We envisage that these tools will complement existing approaches to identifying the geographical limits of *P. knowlesi* transmission.

Introduction

Plasmodium knowlesi is a simian parasite which can cause zoonotic malaria in humans [1]. Recent evidence suggests that human *P. knowlesi* infections are a growing public health threat in South East Asia, particularly in Malaysia [2]. *P. knowlesi* has the potential to cause severe disease in endemic regions [3], and is now the most common cause of clinical malaria in Malaysia [4]. *P. knowlesi* is morphologically similar to *P. malariae* [5], historically leading to the misdiagnosis of *P. knowlesi* infections as *P. malariae* [6]. Recent publications have also demonstrated misdiagnosis of *P. knowlesi* as *P. vivax* and *P. falciparum* [7, 8] with potential delay of appropriate treatment associated with case fatalities [3, 9, 10]. Studies have shown that antibodies to *Plasmodium* proteins persist for long periods [11], even in the context of limited exposure or absence of infection. Such antibodies can be utilised in serological assays, accurately estimating the incidence and exposure to *Plasmodium* parasites [12, 13].

One key requirement for serological studies is the identification of *Plasmodium* species-specific biomarkers, particularly in regions where multi-species infections are likely to occur. It is important to distinguish between human serological responses to different *Plasmodium* species to improve our understanding of immunity to these infections, as well as define the geographical spread of infection. Such information can also help to evaluate the impact of how control measures targeting a single species might affect the transmission and immunological profile of other co-endemic species. The few recombinant protein reagents that do exist for *P. knowlesi* have a high level of sequence homology with orthologues from other *Plasmodium* species and, as such, are not applicable to identifying species-specific antibody responses. For example, PK66 (*Pk*AMA1) [14] and *Pk*SPATR (secreted protein with altered thrombospondin repeat) [15] share 86% and 85% amino acid identity respectively with *P. vivax* (https://is.gd/MzlSez), potentially making it difficult to distinguish between the two species where infections are co-endemic.

The 2011 WHO consultation panel on the public health importance of *P. knowlesi* recommended the urgent development of *P. knowlesi*-specific diagnostic tools [16]. Key to achieving this goal would be the development of sensitive and accurate tools to help monitor the transmission of infection.

In this study, we describe the development and evaluation of a panel of novel recombinant antigens based on *P. knowlesi*-specific amino acid sequences, using publicly available *in silico* tools. The development of such well-validated species-specific tools represent a potentially

important serosurveillance tool to help monitor for historical *P. knowlesi* infections in endemic areas. To illustrate how these data can be used to identify seropositive individuals, we utilise data-adaptive statistical methods (boosted regression trees) to classify exposed individuals. By assessing relative variable importance within these models, we identify the antigen responses contributing most to model predictions and potential serological tools for use in epidemiological studies. These reagents will also serve as an important set of tools to help identify correlates of immunity to *P. knowlesi*.

Methods

Identification and screening of target sequences

Fig 1 outlines the experimental strategy used in the identification of the target sequences of interest. Known markers of seroincidence were selected based on available evidence from *P. falciparum*: AMA1 [17], MSP1 [18], SSP2/TRAP [19] and SERA [20] (*Pk*AMA1 (PKNH_0931500), *Pk*MSP1 (PKNH_0728900), *Pk*SSP2/TRAP (PKNH_1265400), SERA3 (PKNH_0413400) and TSERA2 (PKNH_0413500), respectively). Full-length protein sequences for each gene were initially screened using the BlastP search tool (Plasmodb: https://is.gd/XOs7vd [21] and NCBI: https://is.gd/MzISez). Amino acid sequences were used to generate maximum likelihood phylograms to summarise the relatedness of each gene target between species (S1A–S1E Fig). Alignments were also generated for each target using amino acid sequences from other plasmodia matching the query sequence using the MUltiple Sequence Comparison by Log-Expectation (MUSCLE) software (http://www.ebi.ac.uk/Tools/msa/muscle/) [22] (S2A–

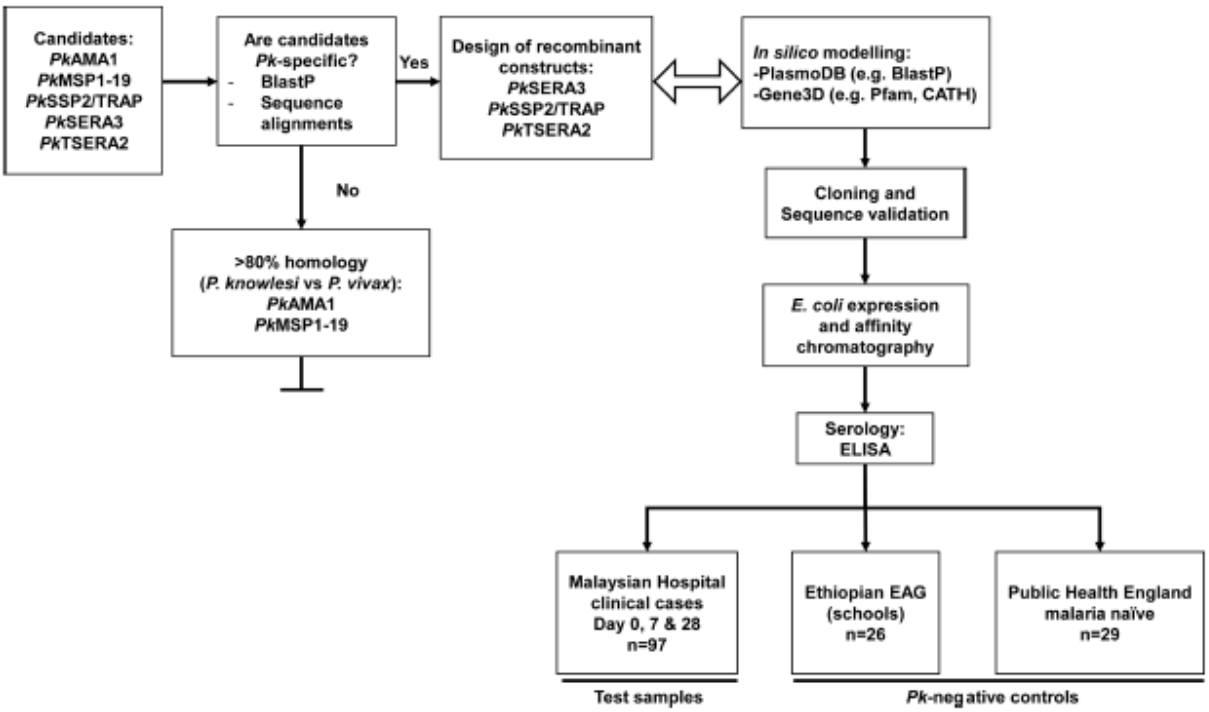


Fig 1. Flowchart summarising the experimental strategy used in the identification and validation of the *P. knowlesi*-specific candidates.

https://doi.org/10.1371/journal.pntd.0006457.g001

S2E Fig). Each alignment was then interrogated to identify regions of identity primarily with *P. vivax* and *P. falciparum* but also with *P. malariae* and *P. ovale*. Regions or entire sequences showing high levels of identity were excluded from further analysis and the *P. knowlesi*-specific truncated regions were again screened using BlastP to validate sequence specificity (Fig 1 and S1 Table). Each target sequence was analysed using domain prediction software (http://gene3d.biochem.ucl.ac.uk/ and http://smart.embl-heidelberg.de/) to help define putative domain boundaries, where possible. The aim was to limit the level of potential antibody cross-reactivity, which would limit the usefulness of the antigens as serological tools due to the high level of identical amino acids between species. A particular problem in co-endemic settings. Simultaneously, sequences were also screened using the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) to help confirm the presence, or absence, of signal peptides and transmembrane regions. Previous experience from our group and others has shown that the presence of signal peptides and/or transmembrane domains can significantly impede protein expression and solubility [23]. Based on this, each confirmed target construct was designed to exclude both the signal peptide and transmembrane domains, which in addition to the GST solubility tag was designed to aid expression of soluble proteins [24].

An additional selection criteria step was to determine the transcriptional status of the candidate genes. Blood stage messenger RNA was collected and analysed using the human red blood cell culture adapted *P. knowlesi* A1-H1 line [25], grown in human blood obtained from the United Kingdom National Blood Transfusion Service. First strand synthesis was carried out using SuperScript IV Reverse Transcriptase (RT) (Thermo Fisher Scientific) using oligo d (T)20 for priming (RT+) according to the manufacturer's instructions. As a negative control (RT-), a second identical reaction was set up in parallel without the addition of the SuperScript IV RT. For PCR analysis of cDNA transcripts, RT+ and RT- samples were used as templates for transcript specific PCR primers for the candidate gene sequences alongside genomic DNA controls. In addition, both *Pk*CTRP (circumsporozoite protein and thrombospondin-related adhesive protein [TRAP]-related protein) and *Pk*CSP (circumsporozoite protein), both shown to be pre-erythrocytic stage targets, were included in the panel as negative controls. Where possible, primer pairs were designed to flank introns so that amplicons from cDNA and gDNA could be distinguished. Sequences of primer pairs used to amplify each transcript are listed in S2 Table alongside the expected cDNA and gDNA amplicon size. Amplicons were PCR amplified using GoTaq Green Master Mix (Promega) and analysed on a 1.2% agarose gel (S3 Fig).

Cloning and expression of *Plasmodium knowlesi*-specific recombinant antigens in *Escherichia coli*

Four new constructs were designed (Table 1 and Fig 2) based on three genes. Two sequences within SERA3 (PKNH_0413400; nucleotide positions 73–419 (Antigen 1) and 2476–2994 (Antigen 2) based on the reference *P. knowlesi* H strain), SSP2/TRAP (PKNH_1265400; nucleotide positons 1141–1500) and TSERA2 (PKNH_0413500; nucleotide positons 178–751 (Antigen 1)) and were PCR amplified from *P. knowlesi* genomic DNA (H strain). Vector compatible primers were designed for each completed target sequence (S3 Table). Cloning of amplified sequences is as described previously [26]. Briefly, purified inserts were cloned into a TA vector (pGEM-T Easy, Promega) and sequence verified. Correct sequences were restriction digested and sub-doned into a GST expression vector (pGEX-2T, GE Healthcare) and sequence verified before transforming into BL21 (DE3) *Escherichia coli* expression cells (Novagen). Validated expression clones were expressed automatically using an autoinduction media based on established protocols [27]. Following expression, protein purification was as described [28].

Table 1. Summary of recombinant antigen construct characteristics.

Gene ID	Antigen	Description	Chromosome	AA position	Expression (mg/L)		Size (kDa)	
							Predicted	Empirical
PKNH_0413400	SERA3 ag1	cysteine protease (Serine repeat-like antigen)	4	25–140	20.5	w/ GST	37.7	49.6 (118.9)
						w/o GST	11.3	n/a
PKNH_0413400	SERA3 ag2	cysteine protease (Serine repeat-like antigen)	4	826–998	15	w/ GST	44.9	68.7 (162.7)
						w/o GST	18.4	n/a
PKNH_1265400	SSP2/ TRAP	sporozoite surface protein 2, putative, thrombospondin-related anonymous protein (TRAP)	12	381–500	17	w/ GST	39.7	53.1 (132.5)
						w/o GST	13.2	n/a
PKNH_0413500	TSERA2 ag1	Truncated cysteine protease	4	60–251	11.9	w/ GST	46.8	59.7 (117.9)
						w/o GST	20.4	n/a

https://doi.org/10.1371/journal.pntd.0006457.t001

Briefly, GST-tagged proteins from clarified bacterial lysate were purified by affinity chromatography (Glutathione sepharose 4B; GE Healthcare) and fractions from each protein analysed (Bradford assay reagent, BioRad) to identify protein-containing fractions. Pooled protein positive fractions were dialysed against PBS and the protein content quantified (Bicinchoninic acid assay (BCA), Thermo Fisher). The dialysed purified proteins were analysed on a 4–20% gradient gel (NuPAGE Bis-Tris acetate) under denaturing conditions and visualised using the Coomassie blue staining method (BioRad BioSafe, USA) (Fig 3).

The empirical sizes of each protein were calculated using the ImageLab (BioRad) software with the PageRuler prestained marker (Fermentas) as a reference standard (Table 1).

SNP and phylogenetic analysis

Full-length sequence data from Plasmodb and construct-specific truncated sequences generated in-house using Sanger sequencing were mapped to an in-house reference sequence strain (*Pk*-H strain) using the Burrows-Wheeler Aligner (BWA) software (v0.7.5a-r405) [32]. Single nucleotide polymorphisms (SNPs) (S4–S8 Tables) were called using the SAMtools (v1.3) (Sequence Alignment/Map) software using default settings [33] and were filtered to increase stringency and target only high quality variants (missingness<10%, mixed calls<10%). Custom Perl scripts identified overlap between these SNPs and each gene candidate. Variants were annotated using snpEFF (v4.3i) (http://snpeff.sourceforge.net/) [34] to retrieve the amino acid position and type effect of the variant. Maximum likelihood phylogenetic trees were constructed from protein sequences using RAxML [35] with a fixed empirical substitution matrix and 200 bootstraps and was visualised using iTOL (http://itol.embl.de) [36].

Enzyme-linked immunosorbent assay (ELISA) and sera collection

The indirect enzyme-linked immunosorbent assay was performed to screen for antibodies to *P. falciparum*, *P. vivax* and *P. knowlesi* antigens using previously described methods [37]. Briefly, antigens were coated at 50 ng/well and serum samples (diluted from frozen serum stocks) assayed at 1/1000 dilution for both the *P. knowlesi* recombinants and the *Pv*MSP1-19 (donated as a kind gift from Tony Holder) positive control antigen. Polyclonal rabbit anti-

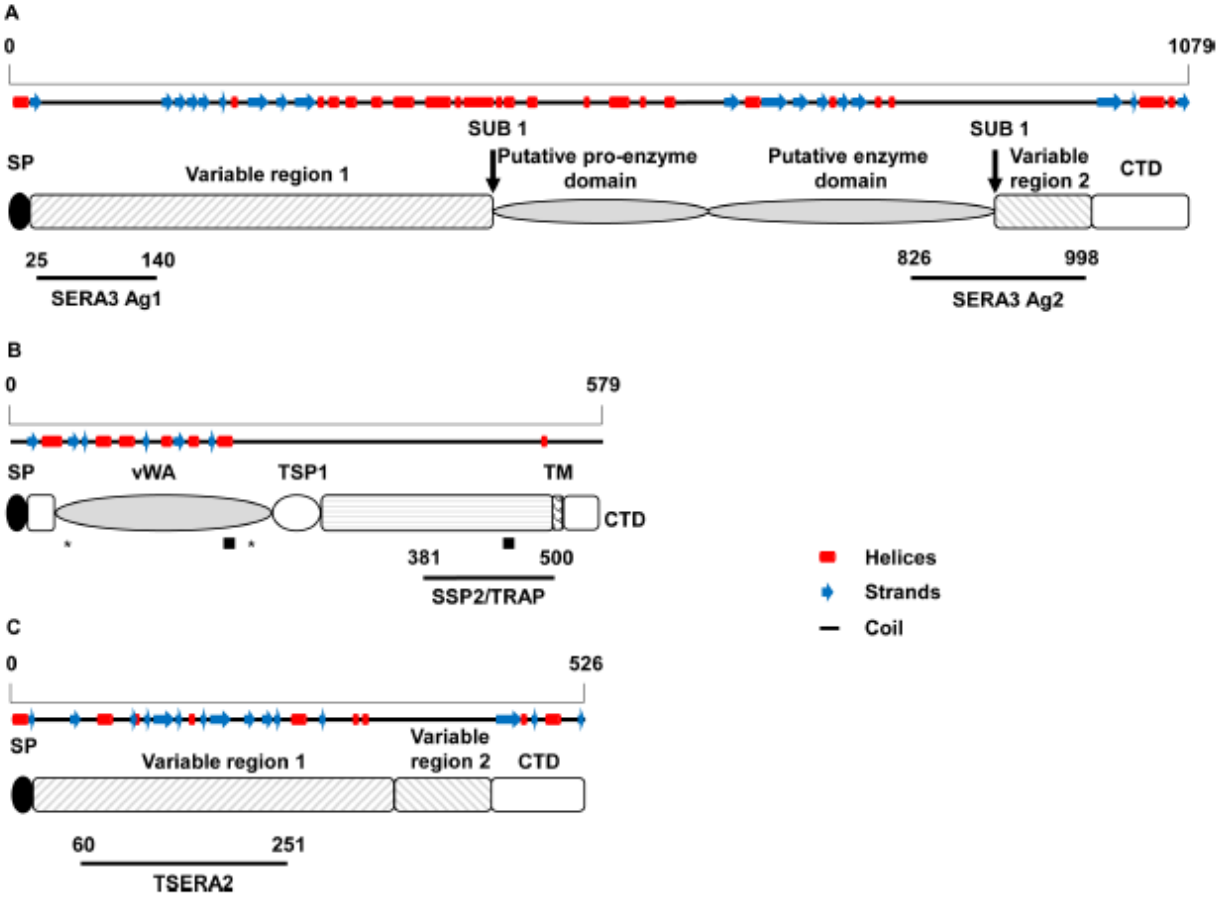


Fig 2. *Plasmodium knowlesi*-specific recombinant antigen constructs. Schematic representations for each protein are shown with key features labelled. (a) *Pk*SERA3 shows the location of the putative pro-enzyme and enzyme domains. The predicted subtilisin (SUB) 1 cleavage sites in relation to variable regions 1 and 2 and the cytoplasmic domain [29]. (b) *Pk*SSP2/TRAP contains a von Willebrandt A domain (vWA), thrombospondin type (TSP) 1 motif, a C-terminal transmembrane (TM) region and a cytoplasmic terminal domain (CTD). Putative T-cell and B-cell epitopes are highlighted with an asterix or black square, respectively [30]. (c) *Pk*TSERA2 shows the lack of central enzyme domain due to truncation of the sequence [29]. Predicted secondary structures generated in I-Tasser [31] are shown above each scheme. Red boxes represent helices, blue arrows sheets and the black line coils. The position of recombinant proteins are highlighted below each scheme with the N- and C-terminal amino acid positions indicated. The overall length of each protein is referenced by the amino acid ruler above each secondary structure prediction. For all proteins SP refers to the signal peptide.

https://doi.org/10.1371/journal.pntd.0006457.g002

human IgG-HRP (Dako, Denmark) was used at 1/15,000 dilution and plates were developed using TMB (One component HRP microwell substrate, Tebu-bio). All assays were performed in duplicate. Negative and positive controls, including blank (buffer only) wells were used to help standardise across assay runs. Values in excess of 1.5 CV between duplicates were considered fails and re-ran.

Written informed consent was obtained from all study participants [18, 38]. Samples were collected as part of a hospital-based clinical trial in Malaysia, Sabah (www.clinicaltrials.gov: #NCT01708876) (Fig 1) [38]. Serum samples were collected at Day 0 (n = 92), 7 (n = 72) and 28 (n = 77) following hospital admission, with drug treatment also taking place at Day 0. The human research ethics committees of Malaysia (MREC) (#NMRR-12-537-12568), the Menzies

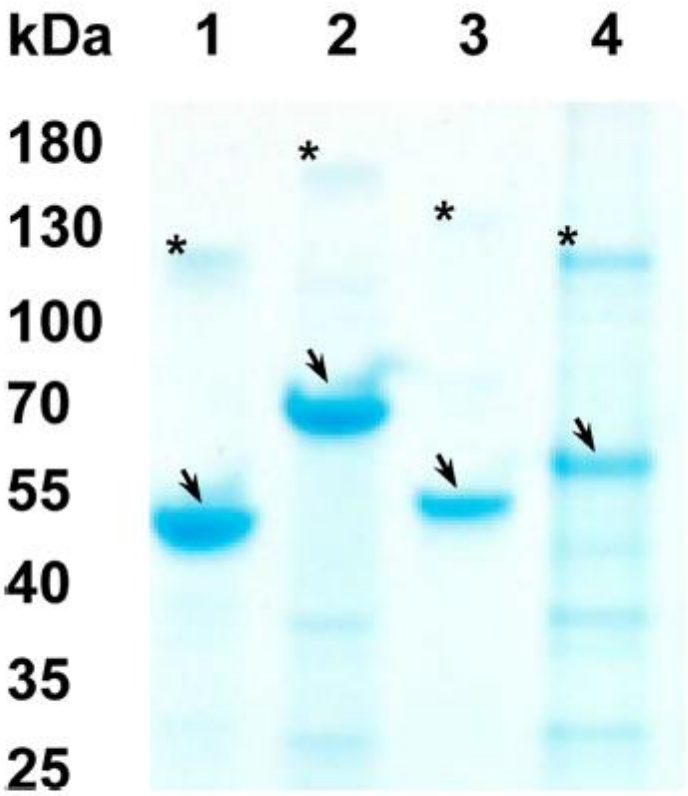


Fig 3. SDS-PAGE of purified recombinants. Lane 1: SERA3 ag1; Lane 2: SERA3 ag2; Lane 3: SSP2/TRAP and Lane 4: TSERA2 ag1. Band sizes are indicated on one side in kDa (Fermentas PageRuler Prestained Protein Ladder). Samples were ran under reducing conditions at approximately 0.4 mg/ml per lane on a 4–20% gradient gel (NuPAGE, BioRad) and stained with Coomassie Blue (BioSafe, BioRad). The arrows and asterisks indicate the protein monomers and aggregates, respectively.

https://doi.org/10.1371/journal.pntd.0006457.g003

School of Health and Research (Australia) (#HREC-2012-1814) and the London School of Hygiene and Tropical Medicine (UK) (#6244) approved the study. Twenty-six *P. vivax*-positive Ethiopian samples [18] based on positive responses to *Pv*AMA1 and *Pv*MSP1-19 were used as the *P. vivax*-positive, *P. knowlesi*-negative control group. In addition, 29 malaria naïve (Public Health England; LSHTM ethics approval #11684) serum samples were used as the *P. knowlesi*-negative control group. For the scatterplot presented in Fig 4, both negative control groups were compared to the responses from the *P. knowlesi*-exposed hospital clinical case samples. All samples used in the study were anonymised.

Statistical and sequence analysis

Descriptive analysis of serological data was performed using STATA/IC 14.2 (StataCorp LP, USA) and PRISM (GraphPad PRISM 7). P values were generated using the Wilcoxon signed rank and Wilcoxon-Mann Whitney tests (STATA/IC 14.2). Scatter plots showing reactivity between *P. knowlesi* recombinant antigens and *P. vivax* MSP1-19 were created using STATA (Fig 4) and dot plots showing reactivity to *P. knowlesi* recombinant antigens were created using GraphPad PRISM (Fig 5 and S4 Fig). Final optical density (OD) values were obtained by

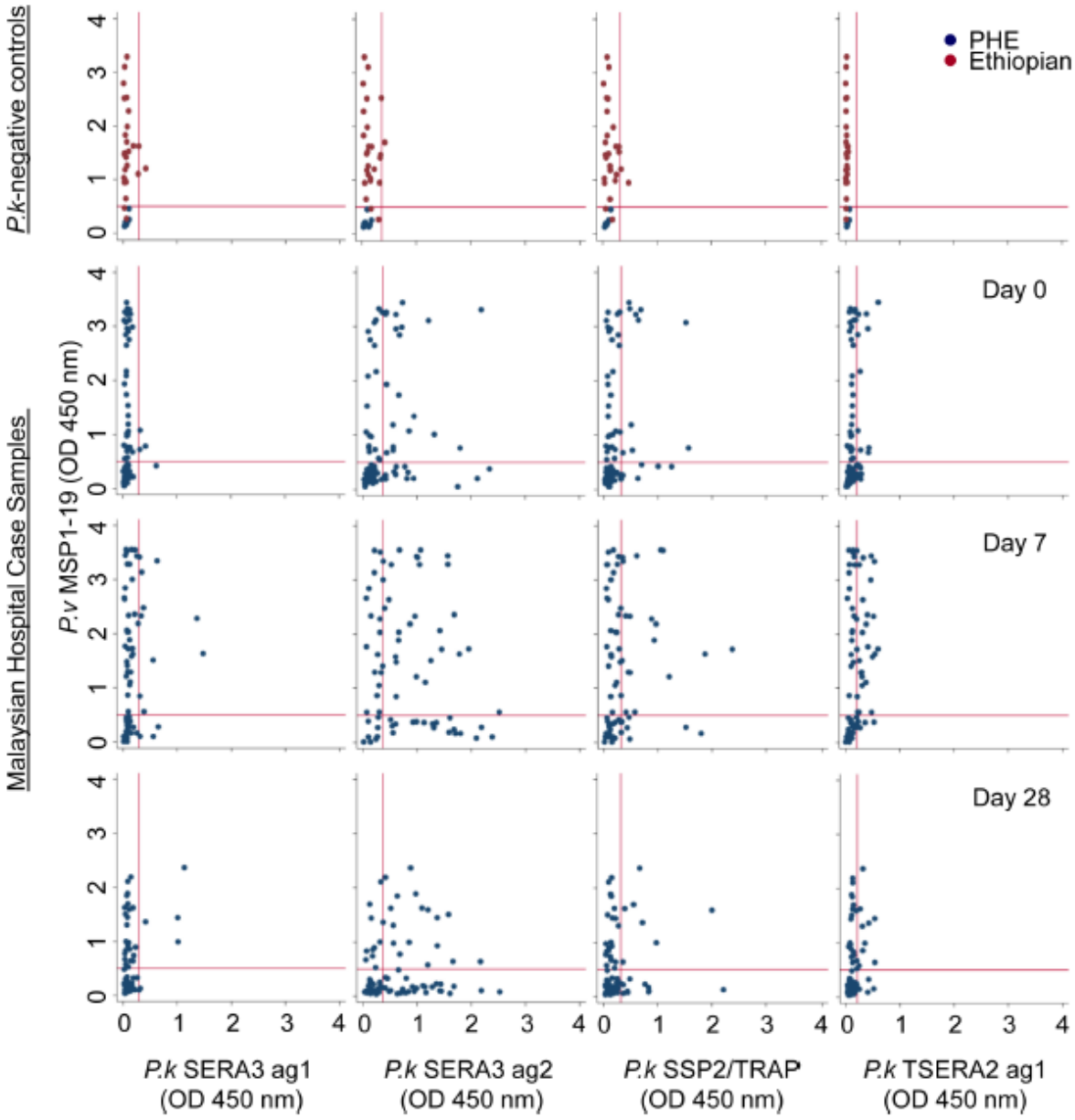


Fig 4. Endemic and *P. knowlesi*-negative sera reactivity to *Plasmodium knowlesi*-specific antigens. Scatter plots showing sera reactivity to *P. vivax* MSP1-19 with *P. knowlesi* SERA3 ag1 (column 1), SERA3 ag2 (column 2), SSP2/TRAP (column 3) and TSERA2 ag1 (column 4) antigens. Sera samples from *P. knowlesi*-negative controls n = 55 (row 1; PHE UK malaria naïve (blue), Ethiopian children (red)) and Malaysian hospital case sera samples from days 0 (n = 92), 7 (n = 72) and 28 (n = 77) of diagnosis (rows 2–4, respectively). The red line in each graph represent the cut off values for the respective *P. knowlesi* antigen and was calculated based on Public Health England negative control sera samples (average ODs ± (3xSD)): The vertical cut off line is based on *Pv*MSP1-19 = 0.501. The horizontal cut off line for the *P. knowlesi* antigens were based on the following values: SERA3 ag1 = 0.292; SERA3 ag2 = 0.366; SSP2/TRAP = 0.322 and TSERA2 ag1 = 0.208.

https://doi.org/10.1371/journal.pntd.0006457.g004

subtracting blank OD values, reducing background reactivity. Cut off values for each *P. knowlesi*-specific antigen were calculated based on the average ODs of Public Health England negative control sera samples ± (3xSD).

Ensemble boosted regression trees were fit to determine predictive power of antibody responses for classification of *P. knowlesi* exposure. To quantify uncertainty around estimates, 100 datasets were assembled including all seronegative individuals from the malaria unexposed population and an equal number of randomly selected *P. knowlesi* seropositive individuals (from all time points). All models were fit using stratified 10-fold cross validation with model predictive ability assessed by the area under the receiver operating curve (AUC). The learning rate was set at 0.001 and tree complexity set at 4, to allow for interactions within the dataset. Contribution of responses to each antigen to models was assessed using relative variable importance as described by Elith *et. al.*[39]. In this method, the relative importance of individual predictor variables is calculated as the number of times a variable is selected for splitting, weighted by the squared improvement to the model and averaged over all trees and scaled to 100%. Boosted regression tree analysis was completed in R statistical software (v 3.4.2) using the gbm package.

Amino acid sequence alignments were generated using MUltiple Sequence Comparison by Log-Expectation (MUSCLE) (http://www.ebi.ac.uk/Tools/msa/muscle/) [22].

Results

In silico identification, design and expression of target sequences

Sequences associated with known immunological markers in *P. falciparum* were selected based on existing evidence (AMA1 [17, 40], MSP1 [40, 41], SSP2/TRAP [42] and SERA antigens [20, 43]), by interrogating existing *P. knowlesi* databases[21, 44] and supporting literature [45] (Fig 1). AMA1 is expressed in the micronemes of both the merozoite (invasive asexual blood stage) and sporozoite (invasive pre-erythrocytive stage) forms [17]. MSP1 is a major protein located on the surface of the merozoite[41]. SSP2/TRAP is also expressed on the surface of the sporozoite forms [42], and the SERA antigens are soluble parasitophorous vacuole proteins [20, 43]. Each sequence was processed using available *in silico* analytical tools (Fig 1). Gene3D [46] and SMART (http://smart.embl-heidelberg.de/) were used to obtain domain prediction information for each gene which helped with the design of truncated fragments (Fig 2). This approach ensured that the design of truncated sequences properly accounted for the presence of any potential domains within each sequence, avoiding unintended truncation of domains which could impact on the solubility of the recombinant proteins. To ensure that expressed products would be specific for *P. knowlesi*, target sequences were interrogated multiple times using the BlastP algorithm [47] against both the *Plasmodium* specific (Plasmodb: https://is.gd/XOs7vd [21]) and non-redundant databases (NCBI: https://is.gd/MzISez).

Maximum likelihood phylogenetic trees were constructed using the *P. knowlesi* H reference strain, highlighting the relationship of each gene between *Plasmodium* species (S1A–S1E Fig). Specifically, for both *Pv*AMA1 (bootstrap value: 100%) and *Pv*MSP1-19 (bootstrap value: 87%), there is a strong relationship between different *Plasmodium* species, particularly between *P. knowlesi* and *P. vivax* (S1A Fig), highlighted further by corresponding near identical amino acid alignments (S2A Fig). Amino acid alignments were generated using available sequences for human-pathogenic *Plasmodium* spp., which clearly highlight the level of sequence identity for both genes between *P. knowlesi* and *P. vivax* (S2A–S2E Fig). Although the bootstrap value strongly supports the grouping of *P. knowlesi* with *P. vivax* and *P. simiovale* (*P. simiovale* was used when data for *P. ovale* was lacking) (S2C–S2E Fig; bootstrap value: 100%), the alignments for SSP2/TRAP and the SERA antigens (PKNH_0413400 and PKNH_0413500), help identify

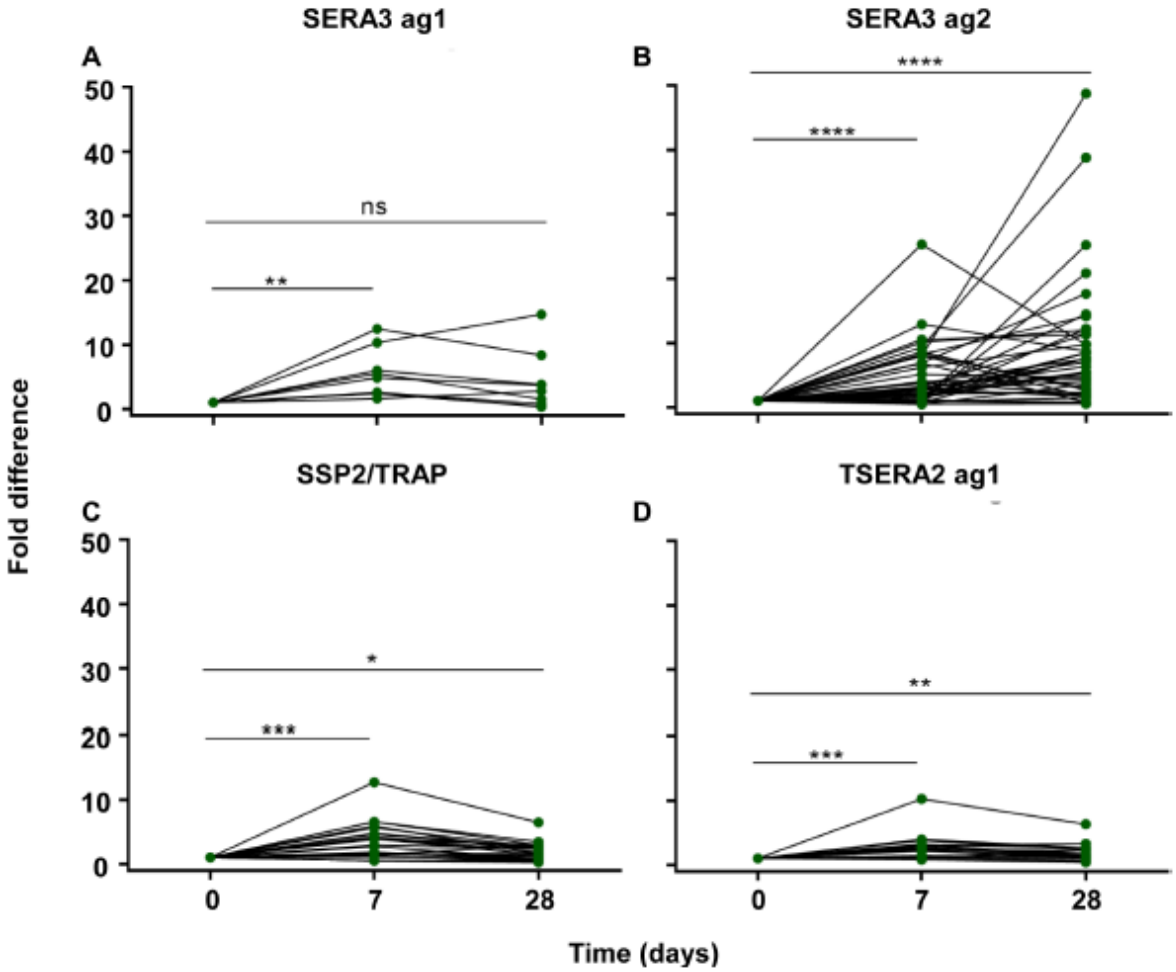


Fig 5. Serial fold increase in antibody reactivity for each antigen following treatment of knowlesi malaria. (a) SERA3 ag1, (b) SERA3 ag2, (c) SSP2/TRAP and (d) TSERA2 ag1. Asterisks indicate level of significance, ns denotes non-significant values ($p \leq 0.0001$ ****; $p \leq 0.001$ ***; $p \leq 0.01$ **; $p \leq 0.05$ * and $p > 0.05$ ns).

https://doi.org/10.1371/journal.pntd.0006457.g005

regions specific for *P. knowlesi* (S2C–S2E Fig). Based on these screens, any sequences showing high amino acid sequence identity to other *Plasmodium* spp., specifically *P. ovale*, *P. malariae*, *P. falciparum* and *P. vivax*, were re-edited to focus on *P. knowlesi*-specific regions only, where possible. All the antigens were expressed in *Escherichia coli* as soluble products with final yields ranging from 11.9–20.5 mg/L (Fig 3, Table 1).

Based on their predicted molecular masses (including the GST tag), SDS PAGE analysis of the purified proteins clearly suggested multimerisation of the purified products (both monomer and dimer) (Fig 3 and Table 1). The Coomassie stained profiles also illustrated that there is very little non-specific degradation of the recombinant proteins (Fig 3), suggesting that the proteins are stable under the conditions used. The protein sizes for each protein were larger than predicted, so called "gel shifting" when ran on SDS PAGE, which is not uncommon. All though not fully explained for all proteins classes evidence suggests that the presence of acidic

residues, net hydropathy or protein aggregation can reduce the effectiveness of SDS in altering the charge, and therefore the migration of proteins through the gel [48, 49]. The fact that all four protein constructs exhibited signs of protein aggregation supports the suggestion that aggregation may affect protein migration on polyacrylamide gels (Fig 3 and Table 1). By way of further validation each protein construct was sequence verified to confirm each sequence and the position of the stop codons to ensure that the departure from the predicted sizes was not due to sequence errors in the construct.

The results of the Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) confirmed that both the SERA3 and TSERA2 candidate genes were actively transcribed in the blood stage (S3 Fig). By contrast, SSP2/TRAP, a sporozoite stage along with the *Pk*CTRP and *Pk*CSP pre-erythrocytic stage controls, were negative by RT-PCR (S3 Fig).

SNP analysis: Capturing polymorphic epitopes in target genes

The existence of three major subpopulations of *P. knowlesi* have been recently described, two associated with clinical human infections from separate macaque species reservoir hosts and the third from long-term laboratory isolates [50]. The presence of amino acid polymorphisms biased towards a single cluster would likely limit the utility of any reagents generated to function as *P. knowlesi*-specific, for all *P. knowlesi*-strains. Therefore, we characterised the presence of SNPs associated with the clusters, focussing on non-synonymous positions within the *P. knowlesi*-specific truncated constructs. S4 Table summarises both the synonymous and non-synonymous SNPs associated with the three clusters (S5–S8 Tables shows the raw SNP data for all four constructs; SERA3 Ag1, SERA3 Ag2, SSP2/TRAP and TSERA2 respectively). For all antigens, the vast majority of the non-synonymous SNPs lie in regions not covered by the antigen design. By omitting the majority of these cluster-specific SNPs we hoped to avoid segregation of detectable antibodies according to the defined clusters. The relevance of these genetic clusters in the context of immunity, and the potential relevance to host preferences is yet to be defined.

Serum reactivity to recombinant antigen panel

Serum samples were collected from 97 Malaysian adults and children hospitalised with *P. knowlesi* malaria on day of diagnosis (day 0), 7 and 28 days post-treatment. Hospital case samples were assayed by enzyme-linked immunosorbent assay (ELISA) using the *P. knowlesi*-specific protein panel. Ethiopian non- *P. knowlesi* malaria endemic children's sera (n = 26) and adult UK malaria naïve sera (n = 29) were used as a *P. knowlesi*-negative control panel. The *P. knowlesi*-negative malaria endemic controls were all reactive with the *Pv*MSP1-19 antigen due to previous *P. vivax* exposure. The malaria naïve controls showed no reactivity to any of the antigens tested (Fig 4, top row and S4 Fig) (SERA3 ag1 OD = 0.124; SERA3 ag2 OD = 0.131; SSP2/TRAP OD = 0.117; TSERA2 ag1 OD = 0.118). With the exception of one weakly positive sample to SERA3 ag 1 and SSP2/TRAP, there was no other detectable antibody reactivity in the control group to the *P. knowlesi*-specific antigens (Fig 4). Antibody reactivity to all four antigens appeared to peak at day 7 (Figs 4 and 5 and S4 Fig), although prevalence of antibody responses to SERA3 antigen 1, *Pk*SSP2 and TSERA2 antigen 1 remained relatively low (18.1% (13/72); 33.3% (45/72) and 43.1% (31/72) respectively) (Fig 4, columns 1, 3 and 4), compared to SERA 3 antigen 2 (63.8% (46/72)). The *Pk*SERA3 antigen 2 had a higher prevalence compared to controls at all time-points ($p<0.001$) (Fig 4 and S4 Fig). Antibody responses measured at day 7 and 28 to SERA3 antigen 2 demonstrated a significant increase when compared to day 0 ($p<0.001$ for both comparisons), with fold changes as high as 50 observed for some samples (Fig 5). In comparison, the fold changes observed in serum responses to the TSERA2

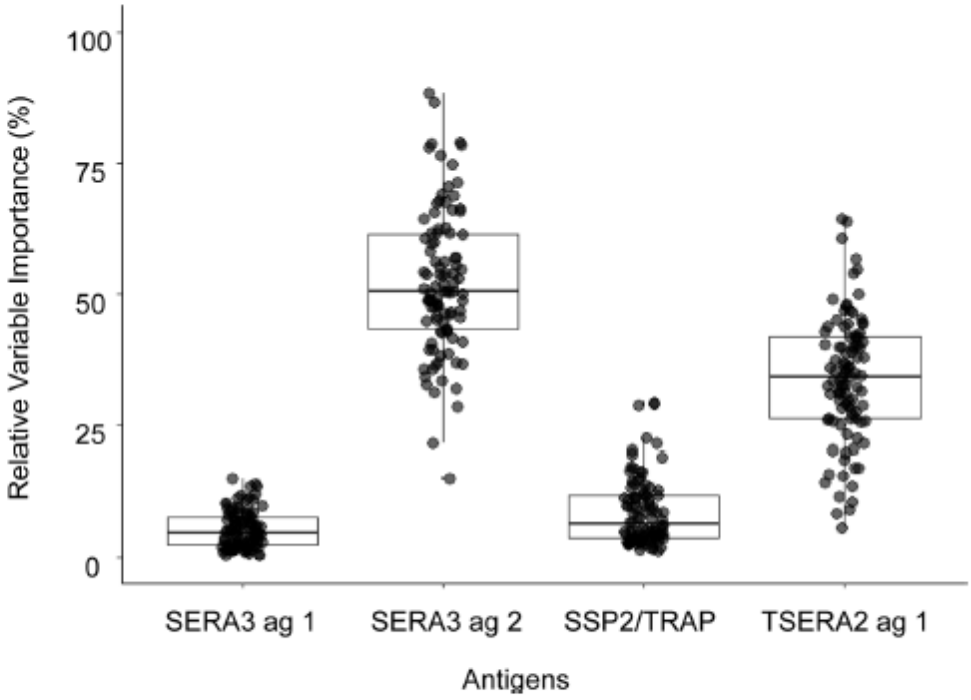


Fig 6. Relative variable importance of responses to each antigen from 100 boosted regression tree models predicting *P. knowlesi* seropositivity. Median values for the relative variable importance and interquartile ranges are shown for all antigens tested: SERA3 ag 1 (4.8%; IQR 2.5–7.8%); SERA3 ag 2 (50.4%; IQR 43.3–61.4%); *Pk*SSP2/TRAP (6.5%; IQR 3.7–11.8%) and TSERA ag 1 (34.2%; IQR 26.2–41.8%).

https://doi.org/10.1371/journal.pntd.0006457.g006

antigen 1 (day 7 and 28; p = <0.001 and p = 0.005 respectively), SERA3 antigen 1 (day 7; p = 0.008), and *Pk*SSP2 (day 7 and 28; p = 0.001 and p = 0.013), although statistically significant had comparatively lower fold changes with a maximum of 15 (Fig 5).

Identification of *P. knowlesi* exposed individuals

To assess the predictive ability of responses to these antigens to identify *P. knowlesi* exposed individuals, we used boosted regression tree analysis, an ensemble modelling method combining aspects of machine learning and statistical analysis shown to have strong predictive performance and reliable identification of variable importance [39]. Similar data-adaptive statistical models are increasingly being used for classification and identification of patterns in large datasets and have previously been applied to identify predictive antigens [51]. Although the samples size is small, boosted regression trees have been used for classification with similarly small training data sets [39]. To further compensate for the small dataset, we fitted 100 models of random samples of equal numbers of sero-positive and sero-negative samples within this training dataset and cross-validated these model predictions. Out of the 100 models fitted for randomly sampled equal numbers of exposed and unexposed individuals, the median classification accuracy was 88.9% (IQR: 86.1–91.3%), calculated as the cross-validated area under the receiver operator curve (AUC). Relative variable importance was calculated for all models. SERA3 antigen 2 contributed most to the models (median relative variable importance: 50.4% (IQR 43.3–61.4%)), followed by TSERA2 antigen 1, *Pk*SSP2/TRAP and SERA3 antigen 1 (Fig 6).

Discussion

P. knowlesi is a naturally occurring infection of long-tailed and pig-tailed macaques, historically associated with forested areas of Southeast Asia [52]. Increased deforestation of their natural habitat is thought to have led to increased interaction between macaques and the human population in endemic areas [53]. Changes in village level forest cover and historical forest loss has been associated with an increase in *P. knowlesi* clinical cases in Sabah [54], with malaria caused by *P. knowlesi* increasingly reported in Southeast Asia [8]. Conversely, there has also been a steady decline in the prevalence of *P. falciparum* and *P. vivax* infections in the same region [55].

The recent efforts of the malaria community towards achieving malaria elimination means that tools to help monitor the impact and effectiveness of intervention strategies are an urgent requirement [56]. The development of species-specific tools for *P. knowlesi* would allow accurate assessment of the levels and geographical limits of infection with this zoonotic species [57]. There is an urgent need to develop a comprehensive discovery strategy to help identify *P. knowlesi* unique antigenic markers of exposure in order to further characterise this organism and develop stronger and better identification methods.

Currently, there are no specifically designed biomarkers for the serosurveillance of *P. knowlesi* infections. Recombinant proteins are available [*Pk*CSP [58], *Pk*AMA1 [59], *Pk*DBP [60], *Pk*SPATR [15], PkLDH [61], *Pk*1-Cys peroxiredoxin [62], *Pk* knowpains [63], *Pk*MSP1-42 [64], *Pk*MSP1-33 [65], *Pk*MSP1-19 [66], Pk tryptophan-rich antigens (*Pk*Trags) [67], *Pk*MSP3 [68] and *Pk*SBP1 [69], but are limited in number and are generally not species-specific. As a result, their utility as serological diagnostic tools is generally secondary to their original design. The reported level of amino acid sequence conservation to other *Plasmodium* spp. in some currently available *P. knowlesi* proteins is > 60% across large stretches of continuous sequence. Such reagents could not be specific to *P. knowlesi* [70] and would be unable to reliably discriminate between antibody responses to different parasite species in co-endemic settings.

High levels of amino acid identity (83%) between *Pv*MSP1-19 and *Pk*MSP1-19, meant we were unable to use these reagents to dissect the species-specific immune responses due to the inevitable cross-reactive antibody responses. This is consistent with a proportion (48.9% (45/92)) of the confirmed *P. knowlesi*-exposed clinical samples in this study reacting with *Pv*MSP1-19 at day 0, although it is unknown whether these participants had previously been exposed to *P. vivax*. However, this limitation simply reflects the paucity of available serological reagents for use in assessing exposure to infection, a deficit this study aims to address. Although the small number of clinical case samples do not give sufficient statistical power to assess either the duration of antibody responses to the panel of antigens or population-level exposure, the *P. knowlesi* clinical case samples represent a unique dataset with which to validate the immunogenicity of our antigen panel.

The use of the boosted regression tree model was able to discriminate between *P. knowlesi* exposed and unexposed individuals for the purposes of classification of seropositivity rather than to assess individual-level risk factors. While this dataset is sufficient for classification as exposed or unexposed, it is not sufficiently large enough to stratify by age, gender or previously reported malaria status. In order for us to assess these types of risk factors, we would first need to apply an approach (using known negatives, mixture or probability models) to classify antibody responses as sero-positive or sero-negative and then assess risk factors within the population. Based on this result the *Pk*SERA 3 antigen 2 recombinant was used to survey ~2500 samples across three site; Limbuak, Pulau Banggi and Matunggung, Kudat, Sabah, Malaysia and Bacungan, Palawan, the Philippines [71]. One of the key elements from this study using

this reagent was the indication of community level patterns of exposure that differed markedly from reported cases, with higher levels of exposure among women and children [71].

The panel of reagents developed for this study focussed on immunologically relevant orthologous targets previously described in *P. falciparum*. The serine repeat antigen (SERA) family had previously attracted attention as a source of both drug and vaccine candidates [72]. In *P. falciparum*, SERA 5 is the most abundant parasitophorous vacuole protein and is essential to blood stage growth of the parasite [73], with antibodies against this antigen thought to inhibit parasite growth [74]. Although possessing a papain-like enzymatic domain, recent evidence suggests that the protein plays a non-enzymatic role [73]. SERA 3 has also been shown to be a highly immunogenic antigen with an important, although not essential role in the erythrocytic cycle [75] and has also been implicated as having a role in liver stage merozoite release in *P. berghei* [76]. Similarly, evidence for the sporozoite surface protein 2 (SSP2/TRAP) suggested an immunogenic antigen involved in protection from disease in mice [77]. Although we were unable to confirm active transcription of SSP2/TRAP due to the lack of available material, we were able to validate active transcription of both the SERA3 and TSERA2 candidate genes. Collectively, the evidence provided by studies on *Plasmodium* supports the design of seroepidemiology tools based on these targets. Despite the targeted approach used in designing the recombinant constructs, the SERA3 antigen 2 construct was by far the most promising candidate. The differences in the performances of the antigens could be due to a number of factors: (1) variation in the inherent immunogenicity of the regions selected, (2) variations in the expression status of the *P. knowlesi* antigens compared to *P. falciparum* or (3) the loss of immunoreactive epitopes due to the truncation of the protein.

There are a number of potential limitations of the study. The small sample size of the clinical samples used prevented detailed analysis of the samples, such as monitoring the impact of factors such as age, on the profile of reactivity to the reagents under test. In addition, the lack of repeated samples per individual (i.e. longitudinal samples) prevented us from investigating the longevity of antibody responses to each target, across individuals and age groups. The availability of supporting biological information on *P. knowlesi*, such as functional data, transcriptional or RNA seq data would have helped with the rational selection of additional candidates for further study and the design recombinant tools.

This is the first study to describe the development a panel of *P. knowlesi*-specific serological tools using freely available *in silico* software. We have demonstrated the importance of targeting species-specific reagents at the amino acid level and highlighted the potential of such proteins as serosurveillance tools. Using these tools we have been able to measure specific immune responses to these reagents and described the change in antibody profile following treatment. As such, we have already demonstrated the utility of the SERA3 antigen 2 reagents as a potential seroepidemiological tool. Studies are also currently in development to expand the existing panel of *P. knowlesi* species-specific reagents to identify additional serological tools. Beyond this we envisage employing high throughput antigen discovery approaches such as the protein microarray to help identify additional important targets of immunity [51, 78]. Further validation of the SERA3 antigen 2 at the population level has recently been performed [71]. Further studies are also planned to characterise the wider immunoglobulin responses, such as IgG subclasses and IgM, to these and future antigens.

Supporting information

S1 Checklist. PRISMA: Clinical trial in Malaysia, Sabah (P. Knowlesi Trial of Artesunate-mefloquine Versus Chloroquine; www.clinicaltrials.gov: #NCT01708876).
(PDF)

S1 Fig. Maximum likelihood phylogenetic analysis of the amino acid sequences of AMA1 (a), MSP1-19 (b), SERA3 (c), TSERA2 (d) and SSP2/TRAP (e) gene sequences between *P. knowlesi*, *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale*/*P. simiovale*. Bootstrap values are given in percentages.
(DOCX)

S2 Fig. Amino acid sequences alignments for AMA1 (a), MSP1-19 (b), SERA3 (c), SSP2/TRAP (d) and TSERA2 (e) gene sequences between *P. knowlesi*, *P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale*/*P. simiovale*. *P. knowlesi*-specific sequences selected for development as constructs are highlighted in yellow. Asterisks indicate fully conserved residues, colons indicates strong residue conservation (>0.5, Gonnet PAM 250 matrix), period indicates weak residue conservation (= <0.5, Gonnet PAM 250 matrix). Conserved cysteine residues are highlighted in green. Blank spaces indicate no residue conservation.
(DOCX)

S3 Fig. *Plasmodium knowlesi* candidate gene transcriptional status in parasite mixed blood stage. Panel 1: SERA3; panel 2: SSP2/TRAP; pand 3: TSERA2; pand 4: CTRP; panel 5: CSP. g refers to genomic DNA, RT+ refers to presence of RT enzyme and RT- refers to absence of RT enzyme. Samples were run on a 1.2% agarose gel. The DNA ladder is indicated in bp (Hyperladder 1Kb, Bioline).
(DOCX)

S4 Fig. *Plasmodium knowlesi* antigen reactivity to Malaysian hospital case serum samples and negative control serum samples. Dot plot of Malaysian hospital case serum samples from days 0 (n = 92), 7 (n = 72) and 28 (n = 77) of PCR diagnosis and *P. knowlesi*-negative control serum samples (Ethiopian *Pv*-positive n = 26; PHE malaria naïve n = 29). Antibody reactivity to the *P. knowlesi*-specific antigens (a) SERA3 ag1, (b) SERA3 ag2, (c) SSP2/TRAP and (d) TSERA2 ag1 are shown.
(DOCX)

S1 Table. Summary of the percentage amino acid identity between *P. knowlesi* and the other *Plasmodium* spp. for all five candidate sequences.
(XLSX)

S2 Table. *P. knowlesi* gene name and ID, primer sequences, primer length, fragment size with and without intron.
(XLSX)

S3 Table. *P. knowlesi* candidate name, primer sequences and primer length. The vector portion of each primer sequence (pGEX-2T) are highlighted in bold and the candidate portion of the sequence in italics. Stop codons are underlined.
(XLSX)

S4 Table. Single-nucleotide polymorphism frequencies of Malaysian clinical isolates sequences within *P. knowlesi* candidate genes.
(XLSX)

S5 Table. Synonymous and non-synonymous SNPs associated with the three *P. knowlesi* genetic clusters for SERA3 Ag1.
(XLSX)

S6 Table. Synonymous and non-synonymous SNPs associated with the three *P. knowlesi* genetic clusters for SERA 3 Ag2.
(XLSX)

S7 Table. Synonymous and non-synonymous SNPs associated with the three *P. knowlesi* genetic clusters for SSP2/TRAP.
(XLSX)

S8 Table. Synonymous and non-synonymous SNPs associated with the three *P. knowlesi* genetic clusters for TSERA2.
(XLSX)

Acknowledgments

We are grateful to all study participants for their cooperation.

Author Contributions

Conceptualization: Lou S. Herman, Michael J. Blackman, Chris J. Drakeley, Kevin K. A. Tetteh.

Data curation: Lou S. Herman, Jody Phelan.

Formal analysis: Lou S. Herman, Kimberly Fornace, Jody Phelan, Kevin K. A. Tetteh.

Funding acquisition: Chris J. Drakeley.

Investigation: Lou S. Herman, Kimberly Fornace, Kevin K. A. Tetteh.

Methodology: Lou S. Herman, Kimberly Fornace, Robert W. Moon, Michael J. Blackman, Kevin K. A. Tetteh.

Project administration: Matthew J. Grigg, Nicholas M. Anstey, Timothy William, Kevin K. A. Tetteh.

Resources: Kimberly Fornace, Matthew J. Grigg, Nicholas M. Anstey, Timothy William, Robert W. Moon, Chris J. Drakeley.

Supervision: Kevin K. A. Tetteh.

Visualization: Kevin K. A. Tetteh.

Writing – original draft: Lou S. Herman, Kevin K. A. Tetteh.

Writing – review & editing: Lou S. Herman, Kimberly Fornace, Jody Phelan, Matthew J. Grigg, Chris J. Drakeley, Kevin K. A. Tetteh.

References

1. White NJ. Plasmodium knowlesi: the fifth human malaria parasite. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2008; 46(2):172–3. Epub 2008/01/04. https://doi.org/10.1086/524889 PMID: 18171246.
2. Moyes CL, Henry AJ, Golding N, Huang Z, Singh B, Baird JK, et al. Defining the geographical range of the Plasmodium knowlesi reservoir. PLoS neglected tropical diseases. 2014; 8(3):e2780. Epub 2014/03/29. https://doi.org/10.1371/journal.pntd.0002780 PMID: 24676231; PubMed Central PMCID: PMCPMC3967999.
3. Cox-Singh J, Davis TM, Lee KS, Shamsul SS, Matusop A, Ratnam S, et al. Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2008; 46(2):165–71. Epub 2008/01/04. https://doi.org/10.1086/524888 PMID: 18171245; PubMed Central PMCID: PMCPmc2533694.

4. Barber BE, William T, Jikal M, Jilip J, Dhararaj P, Menon J, et al. Plasmodium knowlesi malaria in children. Emerging infectious diseases. 2011; 17(5):814–20. Epub 2011/05/03. https://doi.org/10.3201/eid1705.101489 PMID: 21529389; PubMed Central PMCID: PMCPmc3321776.
5. Garnham PCC. Malaria parasites and other haemosporidia: Blackwell Scientific; 1966.
6. Singh B, Kim Sung L, Matusop A, Radhakrishnan A, Shamsul SS, Cox-Singh J, et al. A large focus of naturally acquired Plasmodium knowlesi infections in human beings. Lancet. 2004; 363(9414):1017–24. Epub 2004/03/31. https://doi.org/10.1016/S0140-6736(04)15836-4 PMID: 15051281.
7. Barber BE, William T, Grigg MJ, Menon J, Auburn S, Marfurt J, et al. A prospective comparative study of knowlesi, falciparum, and vivax malaria in Sabah, Malaysia: high proportion with severe disease from Plasmodium knowlesi and Plasmodium vivax but no mortality with early referral and artesunate therapy. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America. 2013; 56(3):383–97. Epub 2012/10/23. https://doi.org/10.1093/cid/cis902 PMID: 23087389.
8. Singh B, Daneshvar C. Human infections and detection of Plasmodium knowlesi. Clin Microbiol Rev. 2013; 26(2):165–84. Epub 2013/04/05. https://doi.org/10.1128/CMR.00079-12 PMID: 23554413; PubMed Central PMCID: PMCPmc3623376.
9. William T, Menon J, Rajahram G, Chan L, Ma G, Donaldson S, et al. Severe Plasmodium knowlesi malaria in a tertiary care hospital, Sabah, Malaysia. Emerging infectious diseases. 2011; 17(7):1248–55. Epub 2011/07/19. https://doi.org/10.3201/eid1707.101017 PMID: 21762579; PubMed Central PMCID: PMCPmc3381373.
10. Rajahram GS, Barber BE, William T, Menon J, Anstey NM, Yeo TW. Deaths due to Plasmodium knowlesi malaria in Sabah, Malaysia: association with reporting as Plasmodium malariae and delayed parenteral artesunate. Malaria journal. 2012; 11:284. Epub 2012/08/22. https://doi.org/10.1186/1475-2875-11-284 PMID: 22905799; PubMed Central PMCID: PMCPMC3472242.
11. Wipasa J, Suphavilai C, Okell LC, Cook J, Corran PH, Thaikla K, et al. Long-lived antibody and B Cell memory responses to the human malaria parasites, Plasmodium falciparum and Plasmodium vivax. PLoS Pathog. 2010; 6(2):e1000770. Epub 2010/02/23. https://doi.org/10.1371/journal.ppat.1000770 PMID: 20174609; PubMed Central PMCID: PMCPMC2824751.
12. Corran P, Coleman P, Riley E, Drakeley C. Serology: a robust indicator of malaria transmission intensity? Trends in Parasitology. 2007; 23(12):575–82. https://doi.org/10.1016/j.pt.2007.08.023 PMID: 17988945
13. Drakeley C, Cook J. Chapter 5 Potential Contribution of Sero-Epidemiological Analysis for Monitoring Malaria Control and Elimination: Historical and Current Perspectives. Advances in parasitology. 69: Academic Press; 2009. p. 299–352. https://doi.org/10.1016/S0065-308X(09)69005-9 PMID: 19622411
14. Waters AP, Thomas AW, Deans JA, Mitchell GH, Hudson DE, Miller LH, et al. A merozoite receptor protein from Plasmodium knowlesi is highly conserved and distributed throughout Plasmodium. J Biol Chem. 1990; 265(29):17974–9. Epub 1990/10/15. PMID: 2211675.
15. Mahajan B, Jani D, Chattopadhyay R, Nagarkatti R, Zheng H, Majam V, et al. Identification, cloning, expression, and characterization of the gene for Plasmodium knowlesi surface protein containing an altered thrombospondin repeat domain. Infection and immunity. 2005; 73(9):5402–9. Epub 2005/08/23. https://doi.org/10.1128/IAI.73.9.5402-5409.2005 PMID: 16113256; PubMed Central PMCID: PMCPmc1231135.
16. World Health Organization. Informal Consultation on the Public Health Importance of Plasmodium knowlesi: Kota Samaharan, Sarawak, Malaysia; 2011 [3 May 2017]. Available from: http://www.wpro.who.int/mvp/documents/docs/Pknowlesi_final_report.pdf.
17. Yang AS, Lopaticki S, O'Neill MT, Erickson SM, Douglas DN, Kneteman NM, et al. AMA1 and MAEBL are important for Plasmodium falciparum sporozoite infection of the liver. Cell Microbiol. 2017. Epub 2017/04/04. https://doi.org/10.1111/cmi.12745 PMID: 28371168.
18. Tadesse FG, van den Hoogen L, Lanke K, Schildkraut J, Tetteh K, Aseffa A, et al. The shape of the iceberg: quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia. Malaria journal. 2017; 16(1):99. Epub 2017/03/04. https://doi.org/10.1186/s12936-017-1749-4 PMID: 28253867; PubMed Central PMCID: PMCPMC5335517.
19. Bargieri DY, Thiberge S, Tay CL, Carey AF, Rantz A, Hischen F, et al. Plasmodium Merozoite TRAP Family Protein Is Essential for Vacuole Membrane Disruption and Gamete Egress from Erythrocytes. Cell Host Microbe. 2016; 20(5):618–30. Epub 2016/11/11. https://doi.org/10.1016/j.chom.2016.10.015 PMID: 27832590; PubMed Central PMCID: PMCPMC5104695.
20. Owalla TJ, Palacpac NM, Shirai H, Horii T, Egwang TG. Association of naturally acquired IgG antibodies against Plasmodium falciparum serine repeat antigen-5 with reduced placental parasitemia and normal birth weight in pregnant Ugandan women: a pilot study. Parasitology international. 2013; 62(3):237–9. Epub 2013/02/12. https://doi.org/10.1016/j.parint.2013.01.006 PMID: 23395684.

21. Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, et al. PlasmoDB: a functional genomic database for malaria parasites. Nucleic acids research. 2009; 37(Database issue):D539–43. Epub 2008/10/30. https://doi.org/10.1093/nar/gkn814 PMID: 18957442; PubMed Central PMCID: PMCPMC2686598.

22. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research. 2004; 32(5):1792–7. Epub 2004/03/23. https://doi.org/10.1093/nar/gkh340 PMID: 15034147; PubMed Central PMCID: PMCPMC390337.

23. Vedadi M, Lew J, Artz J, Amani M, Zhao Y, Dong A, et al. Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms. Molecular and biochemical parasitology. 2007; 151(1):100–10. Epub 2006/11/28. https://doi.org/10.1016/j.molbiopara.2006.10.011 PMID: 17125854.

24. Gopal GJ, Kumar A. Strategies for the production of recombinant protein in Escherichia coli. The protein journal. 2013; 32(6):419–25. Epub 2013/07/31. https://doi.org/10.1007/s10930-013-9502-5 PMID: 23897421.

25. Moon RW, Hall J, Rangkuti F, Ho YS, Almond N, Mitchell GH, et al. Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110(2):531–6. Epub 2012/12/26. https://doi.org/10.1073/pnas.1216457110 PMID: 23267069; PubMed Central PMCID: PMCPmc3545754.

26. Tetteh KK, Osier FH, Salanti A, Kamuyu G, Drought L, Failly M, et al. Analysis of antibodies to newly described Plasmodium falciparum merozoite antigens supports MSPDBL2 as a predicted target of naturally acquired immunity. Infection and immunity. 2013; 81(10):3835–42. Epub 2013/07/31. https://doi.org/10.1128/IAI.00301-13 PMID: 23897617; PubMed Central PMCID: PMCPmc3811751.

27. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein expression and purification. 2005; 41(1):207–34. Epub 2005/05/26. PMID: 15915565.

28. Polley SD, Tetteh KK, Cavanagh DR, Pearce RJ, Lloyd JM, Bojang KA, et al. Repeat sequences in block 2 of Plasmodium falciparum merozoite surface protein 1 are targets of antibodies associated with protection from malaria. Infection and immunity. 2003; 71(4):1833–42. Epub 2003/03/26. https://doi.org/10.1128/IAI.71.4.1833-1842.2003 PMID: 12654798; PubMed Central PMCID: PMCPMC152097.

29. Arisue N, Hirai M, Arai M, Matsuoka H, Horii T. Phylogeny and evolution of the SERA multigene family in the genus Plasmodium. Journal of molecular evolution. 2007; 65(1):82–91. Epub 2007/07/05. https://doi.org/10.1007/s00239-006-0253-1 PMID: 17609814.

30. Offeddu V, Thathy V, Marsh K, Matuschewski K. Naturally acquired immune responses against Plasmodium falciparum sporozoites and liver infection. Int J Parasitol. 2012; 42(6):535–48. Epub 2012/05/09. https://doi.org/10.1016/j.ijpara.2012.03.011 PMID: 22561398.

31. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nature methods. 2015; 12(1):7–8. Epub 2014/12/31. https://doi.org/10.1038/nmeth.3213 PMID: 25549265; PubMed Central PMCID: PMCPMC4428668.

32. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324 PubMed PMID: PMC2705234. PMID: 19451168

33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. Epub 2009/06/10. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943; PubMed Central PMCID: PMCPmc2723002.

34. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012; 6(2):80–92. Epub 2012/06/26. https://doi.org/10.4161/fly.19695 PMID: 22728672; PubMed Central PMCID: PMCPmc3679285.

35. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9):1312–3. Epub 2014/01/24. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623; PubMed Central PMCID: PMCPMC3998144.

36. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research. 2016; 44(W1):W242–5. Epub 2016/04/21. https://doi.org/10.1093/nar/gkw290 PMID: 27095192; PubMed Central PMCID: PMCPMC4987883.

37. Corran PH, Cook J, Lynch C, Leendertse H, Manjurano A, Griffin J, et al. Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. Malaria journal. 2008; 7:195. Epub 2008/10/02. https://doi.org/10.1186/1475-2875-7-195 PMID: 18826573; PubMed Central PMCID: PMCPmc2567984.

38. Grigg MJ, William T, Menon J, Dhanaraj P, Barber BE, Wilkes CS, et al. Artesunate-mefloquine versus chloroquine for treatment of uncomplicated Plasmodium knowlesi malaria in Malaysia (ACT KNOW): an

open-label, randomised controlled trial. The Lancet infectious diseases. 2016. Epub 18 November 2015. https://doi.org/10.1016/s1473-3099(15)00415-6 PMID: 26603174.

39. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. The Journal of animal ecology. 2008; 77(4):802–13. Epub 2008/04/10. https://doi.org/10.1111/j.1365-2656.2008.01390.x PMID: 18397250.

40. Ssewanyana I, Arinaitwe E, Nankabirwa JI, Yeka A, Sullivan R, Kamya MR, et al. Avidity of anti-malarial antibodies inversely related to transmission intensity at three sites in Uganda. Malaria journal. 2017; 16 (1):67. Epub 2017/02/12. https://doi.org/10.1186/s12936-017-1721-3 PMID: 28183299; PubMed Central PMCID: PMCPMC5301436.

41. Moss DK, Remarque EJ, Faber BW, Cavanagh DR, Arnot DE, Thomas AW, et al. Plasmodium falciparum 19-kilodalton merozoite surface protein 1 (MSP1)-specific antibodies that interfere with parasite growth in vitro can inhibit MSP1 processing, merozoite invasion, and intracellular parasite development. Infection and immunity. 2012; 80(3):1280–7. Epub 2011/12/29. https://doi.org/10.1128/IAI.05887-11 PMID: 22202121; PubMed Central PMCID: PMCPMC3294643.

42. Rogers WO, Malik A, Mellouk S, Nakamura K, Rogers MD, Szarfman A, et al. Characterization of Plasmodium falciparum sporozoite surface protein 2. Proceedings of the National Academy of Sciences of the United States of America. 1992; 89(19):9176–80. Epub 1992/10/01. PMID: 1409621; PubMed Central PMCID: PMCPmc50088.

43. Yagi M, Palacpac NM, Ito K, Oishi Y, Itagaki S, Balikagala B, et al. Antibody titres and boosting after natural malaria infection in BK-SE36 vaccine responders during a follow-up study in Uganda. Sci Rep. 2016; 6:34363. Epub 2016/10/06. https://doi.org/10.1038/srep34363 PMID: 27703240; PubMed Central PMCID: PMCPMC5050508 following received renumeration from BIKEN: N.M.Q.P., B.B., E.H.N., A.Y., B.N.K., O.K., H.S. and T.G.E. M.Y., K.I., Y.O., S.I., W.F. and Y.H. declare no conflict of interest.

44. Logan-Klumpler FJ, De Silva N, Boehme U, Rogers MB, Velarde G, McQuillan JA, et al. GeneDB—an annotation database for pathogens. Nucleic acids research. 2012; 40(Database issue):D98–108. Epub 2011/11/26. https://doi.org/10.1093/nar/gkr1032 PMID: 22116062; PubMed Central PMCID: PMCPMC3245030.

45. Pain A, Bohme U, Berry AE, Mungall K, Finn RD, Jackson AP, et al. The genome of the simian and human malaria parasite Plasmodium knowlesi. Nature. 2008; 455(7214):799–803. Epub 2008/10/10. https://doi.org/10.1038/nature07306 PMID: 18843368; PubMed Central PMCID: PMCPmc2656934.

46. Lewis TE, Sillitoe I, Andreeva A, Blundell TL, Buchan DW, Chothia C, et al. Genome3D: a UK collaborative project to annotate genomic sequences with predicted 3D structures based on SCOP and CATH domains. Nucleic acids research. 2013; 41(Database issue):D499–507. Epub 2012/12/04. https://doi.org/10.1093/nar/gks1266 PMID: 23203986; PubMed Central PMCID: PMCPMC3531217.

47. Mount DW. Using the Basic Local Alignment Search Tool (BLAST). CSH protocols. 2007; 2007:pdb.top17. Epub 2007/01/01. https://doi.org/10.1101/pdb.top17 PMID: 21357135.

48. Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(6):1760–5. Epub 2009/02/03. https://doi.org/10.1073/pnas.0813167106 PMID: 19181854; PubMed Central PMCID: PMCPMC2644111.

49. Nadeau VG, Rath A, Deber CM. Sequence hydropathy dominates membrane protein response to detergent solubilization. Biochemistry. 2012; 51(31):6228–37. Epub 2012/07/12. https://doi.org/10.1021/bi201853n PMID: 22779403.

50. Assefa S, Lim C, Preston MD, Duffy CW, Nair MB, Adroub SA, et al. Population genomic structure and adaptation in the zoonotic malaria parasite Plasmodium knowlesi. Proceedings of the National Academy of Sciences. 2015. https://doi.org/10.1073/pnas.1509534112 PMID: 26438871

51. Helb DA, Tetteh KK, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, et al. Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112(32): E4438–47. Epub 2015/07/29. https://doi.org/10.1073/pnas.1501705112 PMID: 26216993; PubMed Central PMCID: PMCPMC4538641.

52. Coatney GR, Collins WE, Contacos PG. The Primate malarias. Bethesda, Md.: U.S. National Institute of Allergy and Infectious Diseases; 1971.

53. Imai N, White MT, Ghani AC, Drakeley CJ. Transmission and control of Plasmodium knowlesi: a mathematical modelling study. PLoS neglected tropical diseases. 2014; 8(7):e2978. Epub 2014/07/25. https://doi.org/10.1371/journal.pntd.0002978 PMID: 25058400; PubMed Central PMCID: PMCPmc4109903.

54. Fornace KM, Abidin TR, Alexander N, Brock P, Grigg MJ, Murphy A, et al. Association between Landscape Factors and Spatial Patterns of Plasmodium knowlesi Infections in Sabah, Malaysia. Emerging infectious diseases. 2016; 22(2):201–9. https://doi.org/10.3201/eid2202.150656 PubMed PMID: PMC4734530. PMID: 26812373

55. Mohd Abd Razak MR, Sastu UR, Norahmad NA, Abdul-Karim A, Muhammad A, Muniandy PK, et al. Genetic Diversity of Plasmodium falciparum Populations in Malaria Declining Areas of Sabah, East Malaysia. PLoS One. 2016; 11(3):e0152415. Epub 2016/03/31. https://doi.org/10.1371/journal.pone.0152415 PMID: 27023787; PubMed Central PMCID: PMCPMC4811561.

56. World Health Organization. World Malaria report: 2016. 2016.

57. World Health Organization. WHO Malaria Policy Advisory Committee (MPAC) meeting 2017 [9 May 2017]. Available from: http://www.who.int/malaria/publications/atoz/WHO-HTM-GMP-2017.8-eng.pdf?ua=1.

58. Sharma S, Godson GN. Expression of the major surface antigen of Plasmodium knowlesi sporozoites in yeast. Science. 1985; 228(4701):879–82. Epub 1985/05/17. PMID: 3890178.

59. Mahdi Abdel Hamid M, Remarque EJ, van Duivenvoorde LM, van der Werff N, Walraven V, Faber BW, et al. Vaccination with Plasmodium knowlesi AMA1 formulated in the novel adjuvant co-vaccine HT protects against blood-stage challenge in rhesus macaques. PLoS One. 2011; 6(5):e20547. Epub 2011/06/10. https://doi.org/10.1371/journal.pone.0020547 PMID: 21655233; PubMed Central PMCID: PMCPMC3105089.

60. Singh AP, Puri SK, Chitnis CE. Antibodies raised against receptor-binding domain of Plasmodium knowlesi Duffy binding protein inhibit erythrocyte invasion. Molecular and biochemical parasitology. 2002; 121(1):21–31. Epub 2002/05/03. PMID: 11985860.

61. Singh V, Kaushal DC, Rathaur S, Kumar N, Kaushal NA. Cloning, overexpression, purification and characterization of Plasmodium knowlesi lactate dehydrogenase. Protein expression and purification. 2012; 84(2):195–203. Epub 2012/06/12. https://doi.org/10.1016/j.pep.2012.05.008 PMID: 22683723.

62. Hakimi H, Asada M, Angeles JM, Kawai S, Inoue N, Kawazu S. Plasmodium vivax and Plasmodium knowlesi: cloning, expression and functional analysis of 1-Cys peroxiredoxin. Experimental parasitology. 2013; 133(1):101–5. Epub 2012/11/28. https://doi.org/10.1016/j.exppara.2012.10.018 PMID: 23178658.

63. Prasad R, Atul, Soni A, Puri SK, Sijwali PS. Expression, characterization, and cellular localization of knowpains, papain-like cysteine proteases of the Plasmodium knowlesi malaria parasite. PLoS One. 2012; 7(12):e51619. Epub 2012/12/20. https://doi.org/10.1371/journal.pone.0051619 PMID: 23251596; PubMed Central PMCID: PMCPMC3520923.

64. Cheong FW, Fong MY, Lau YL, Mahmud R. Immunogenicity of bacterial-expressed recombinant Plasmodium knowlesi merozoite surface protein-142 (MSP-142). Malaria journal. 2013; 12:454. Epub 2013/12/21. https://doi.org/10.1186/1475-2875-12-454 PMID: 24354660; PubMed Central PMCID: PMCPmc3878241.

65. Cheong FW, Lau YL, Fong MY, Mahmud R. Evaluation of recombinant Plasmodium knowlesi merozoite surface protein-1(33) for detection of human malaria. Am J Trop Med Hyg. 2013; 88(5):835–40. Epub 2013/03/20. https://doi.org/10.4269/ajtmh.12-0250 PMID: 23509118; PubMed Central PMCID: PMCPMC3752745.

66. Lau YL, Cheong FW, Chin LC, Mahmud R, Chen Y, Fong MY. Evaluation of codon optimized recombinant Plasmodium knowlesi Merozoite Surface Protein-119 (pkMSP-119) expressed in Pichia pastoris. Tropical biomedicine. 2014; 31(4):749–59. Epub 2015/03/18. PMID: 25776601.

67. Tyagi K, Gupta D, Saini E, Choudhary S, Jamwal A, Alam MS, et al. Recognition of Human Erythrocyte Receptors by the Tryptophan-Rich Antigens of Monkey Malaria Parasite Plasmodium knowlesi. PLoS One. 2015; 10(9):e0138691. Epub 2015/09/24. https://doi.org/10.1371/journal.pone.0138691 PMID: 26393350; PubMed Central PMCID: PMCPmc4579084.

68. De Silva JR, Lau YL, Fong MY. Expression and Evaluation of Recombinant Plasmodium knowlesi Merozoite Surface Protein-3 (MSP-3) for Detection of Human Malaria. PLoS One. 2016; 11(7):e0158998. Epub 2016/07/09. https://doi.org/10.1371/journal.pone.0158998 PMID: 27391270; PubMed Central PMCID: PMCPMC4938616.

69. Lucky AB, Sakaguchi M, Katakai Y, Kawai S, Yahata K, Templeton TJ, et al. Plasmodium knowlesi Skeleton-Binding Protein 1 Localizes to the 'Sinton and Mulligan' Stipplings in the Cytoplasm of Monkey and Human Erythrocytes. PLoS One. 2016; 11(10):e0164272. Epub 2016/10/13. https://doi.org/10.1371/journal.pone.0164272 PMID: 27732628; PubMed Central PMCID: PMCPMC5061513.

70. Sonaimuthu P, Cheong FW, Chin LC, Mahmud R, Fong MY, Lau YL. Detection of human malaria using recombinant Plasmodium knowlesi merozoire surface protein-1 (MSP-1(1)(9)) expressed in Escherichia coli. Experimental parasitology. 2015; 153:118–22. Epub 2015/03/31. https://doi.org/10.1016/j.exppara.2015.03.010 PMID: 25812552.

71. Fornace KM, Herman LS, Abidin TR, Chua TH, Daim S, Lorenzo PJ, et al. Exposure and infection to Plasmodium knowlesi in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. (Submitted).

72. McCoubrie JE, Miller SK, Sargeant T, Good RT, Hodder AN, Speed TP, et al. Evidence for a common role for the serine-type Plasmodium falciparum serine repeat antigen proteases: implications for vaccine and drug design. Infection and immunity. 2007; 75(12):5565–74. Epub 2007/09/26. https://doi.org/10.1128/IAI.00405-07 PMID: 17893128; PubMed Central PMCID: PMCPmc2168336.

73. Stallmach R, Kavishwar M, Withers-Martinez C, Hackett F, Collins CR, Howell SA, et al. Plasmodium falciparum SERA5 plays a non-enzymatic role in the malarial asexual blood-stage lifecycle. Mol Microbiol. 2015; 96(2):368–87. Epub 2015/01/21. https://doi.org/10.1111/mmi.12941 PMID: 25599609; PubMed Central PMCID: PMCPMC4671257.

74. Aoki S, Li J, Itagaki S, Okech BA, Egwang TG, Matsuoka H, et al. Serine repeat antigen (SERA5) is predominantly expressed among the SERA multigene family of Plasmodium falciparum, and the acquired antibody titers correlate with serum inhibition of the parasite growth. J Biol Chem. 2002; 277(49):47533–40. Epub 2002/09/24. https://doi.org/10.1074/jbc.M207145200 PMID: 12244052.

75. Miller SK, Good RT, Drew DR, Delorenzi M, Sanders PR, Hodder AN, et al. A subset of Plasmodium falciparum SERA genes are expressed and appear to play an important role in the erythrocytic cycle. J Biol Chem. 2002; 277(49):47524–32. Epub 2002/09/14. https://doi.org/10.1074/jbc.M206974200 PMID: 12228245.

76. Schmidt-Christensen A, Sturm A, Horstmann S, Heussler VT. Expression and processing of Plasmodium berghei SERA3 during liver stages. Cell Microbiol. 2008; 10(8):1723–34. Epub 2008/04/19. https://doi.org/10.1111/j.1462-5822.2008.01162.x PMID: 18419771; PubMed Central PMCID: PMCPMC2613260.

77. Khusmith S, Charoenvit Y, Kumar S, Sedegah M, Beaudoin RL, Hoffman SL. Protection against malaria by vaccination with sporozoite surface protein 2 plus CS protein. Science. 1991; 252(5006):715–8. Epub 1991/05/03. PMID: 1827210.

78. Uplekar S, Rao PN, Ramanathapuram L, Awasthi V, Verma K, Sutton P, et al. Characterizing Antibody Responses to Plasmodium vivax and Plasmodium falciparum Antigens in India Using Genome-Scale Protein Microarrays. PLoS neglected tropical diseases. 2017; 11(1):e0005323. Epub 2017/01/25. https://doi.org/10.1371/journal.pntd.0005323 PMID: 28118367; PubMed Central PMCID: PMCPMC5291533 has licensed the protein microarray technology described in this paper. The terms of this arrangement has been reviewed and approved by the University of California, Irvine, in accordance with its conflict of interest policies. PS contributed to the work while employed at New York University, and therefore there is no perceived conflict of interest with his current employer, Acsel Health. The other authors have declared that no competing interests exist.

APPENDIX II: PUBLICATION

RESEARCH ARTICLE

Exposure and infection to *Plasmodium knowlesi* in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Lou Salomé Herman
Principal Supervisor	Kevin Tetteh
Thesis Title	Development of P. knowlesi species specific reagents to help ~~detect~~ characterise antibody isotype profiles in endemic human populations and experimental macaque infections.

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	PLOS Neglected Tropical Diseases		
When was the work published?	June. 14. 2018		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	YES

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	ATTACHED

Student Signature: Lou Herman. **Date:** 03/09/2018

Supervisor Signature: **Date:** 11/9/18

Section D – Multi-authored work

Author Contributions

Conceptualization: Kimberly M. Fornace, Timothy William, Fe Espino, Jonathan Cox, Chris J. Drakeley.

Data curation: Tommy R. Abidin, Pauline J. Lorenzo.

Formal analysis: Kimberly M. Fornace.

Funding acquisition: Jonathan Cox, Chris J. Drakeley.

Investigation: Kimberly M. Fornace, Tommy R. Abidin, Tock Hing Chua, Sylvia Daim, Pauline J. Lorenzo, Lynn Grignard, Nor Afizah Nuin, Lau Tiek Ying, Matthew J. Grigg, Fe Espino, Jonathan Cox, Kevin K. A. Tetteh, Chris J. Drakeley.

Methodology: Kimberly M. Fornace, Lou S. Herman, Sylvia Daim, Pauline J. Lorenzo, Lynn Grignard, Nor Afizah Nuin, Lau Tiek Ying, Jonathan Cox, Kevin K. A. Tetteh, Chris J. Drakeley.

Supervision: Timothy William, Fe Espino, Kevin K. A. Tetteh, Chris J. Drakeley.

Writing - original draft: Kimberly M. Fornace.

Writing - review & editing: Lou S. Herman, Tommy R. Abidin, Tock Hing Chua, Sylvia Daim, Lynn Grignard, Nor Afizah Nuin, Lau Tiek Ying, Matthew J. Grigg, Timothy William, Fe Espino, Jonathan Cox, Chris J. Drakeley.

PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Exposure and infection to *Plasmodium knowlesi* in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines

Kimberly M. Fornace[1]+, Lou S. Herman[1], Tommy R. Abidin[2], Tock Hing Chua[3], Sylvia Daim[3], Pauline J. Lorenzo[4], Lynn Grignard[1], Nor Afizah Nuin[5], Lau Tiek Ying[5], Matthew J. Grigg[2,6], Timothy William[2,7], Fe Espino[4], Jonathan Cox[1], Kevin K. A. Tetteh[1], Chris J. Drakeley[1]

1 Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Infectious Diseases Society Kota Kinabalu- Menzies School of Health Research Clinical Research Unit, Kota Kinabalu, Malaysia, **3** Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia, **4** Research Institute of Tropical Medicine, Department of Health, Muntinlupa City, Philippines, **5** Biotechnology Research Institute, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia, **6** Menzies School of Health Research and Charles Darwin University, Darwin, Australia, **7** Jesselton Medical Centre, Kota Kinabalu, Malaysia

* Kimberly.Fornace@lshtm.ac.uk

OPEN ACCESS

Citation: Fornace KM, Herman LS, Abidin TR, Chua TH, Daim S, Lorenzo PJ, et al. (2018) Exposure and infection to *Plasmodium knowlesi* in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. PLoS Negl Trop Dis 12(6): e0006432. https://doi.org/10.1371/journal.pntd.0006432

Editor: Alyssa E. Barry, Walter and Eliza Hall Institute, AUSTRALIA

Received: December 20, 2017

Accepted: April 5, 2018

Published: June 14, 2018

Data Availability Statement: As the full dataset contains identifying information on the health of human participants, the data are available from the Monkeybar Project Data Repository for researchers who meet the criteria for access to confidential data. Contact details for the project are available at: http://malaria.lshtm.ac.uk/MONKEYBAR or through contact for the MONKEYBAR project data repository at monkeybar.esei@gmail.com

Abstract

Background

Primarily impacting poor, rural populations, the zoonotic malaria *Plasmodium knowlesi* is now the main cause of human malaria within Malaysian Borneo. While data is increasingly available on symptomatic cases, little is known about community-level patterns of exposure and infection. Understanding the true burden of disease and associated risk factors within endemic communities is critical for informing evidence-based control measures.

Methodology/Principal findings

We conducted comprehensive surveys in three areas where *P. knowlesi* transmission is reported: Limbuak, Pulau Banggi and Matunggung, Kudat, Sabah, Malaysia and Bacungan, Palawan, the Philippines. Infection prevalence was low with parasites detected by PCR in only 0.2% (4/2503) of the population. *P. knowlesi* PkSERA3 ag1 antibody responses were detected in 7.1% (95% CI: 6.2–8.2%) of the population, compared with 16.1% (14.6–17.7%) and 12.6% (11.2–14.1%) for *P. falciparum* and *P. vivax*. Sero-prevalence was low in individuals <10 years old for *P. falciparum* and *P. vivax* consistent with decreased transmission of non-zoonotic malaria species. Results indicated marked heterogeneity in transmission intensity between sites and *P. knowlesi* exposure was associated with agricultural work (OR 1.63; 95% CI 1.07–2.48) and higher levels of forest cover (OR 2.40; 95% CI 1.29–4.46) and clearing (OR 2.14; 95% CI 1.35–3.40) around houses. Spatial patterns of *P. knowlesi* exposure differed from exposure to non-zoonotic malaria and *P. knowlesi* exposed individuals were younger on average than individuals exposed to non-zoonotic malaria.

PLOS | NEGLECTED TROPICAL DISEASES

Funding: This study was funded by the Medical Research Council, Natural Environmental Research Council, Economic and Social Research Council and Biotechnology and Biosciences Research Council for the funding received for this project through the Environmental and Social Ecology of Human Infectious Diseases Initiative (ESEI). Grant number: G1100796The funders had no role in the study design, data collection and analysis, decision to publish or preparation of this manuscript.

Competing interests: The authors have declared that no competing interests exist.

Conclusions/Significance

This is the first study to describe serological exposure to *P. knowlesi* and associated risk factors within endemic communities. Results indicate community–level patterns of infection and exposure differ markedly from demographics of reported cases, with higher levels of exposure among women and children. Further work is needed to understand these variations in risk across a wider population and spatial scale.

Author summary

Plasmodium knowlesi is a species of malaria parasite found in wild macaque populations which is now the main cause of human malaria in Malaysian Borneo. Spread from macaques to people through infected mosquitoes, human *P. knowlesi* malaria cases have primarily been reported in adult men working in forests or plantations. However, little data is available on the extent of asymptomatic infections or people exposed to *P. knowlesi* not reporting to clinics. We conducted comprehensive surveys of three case study communities in Malaysian Borneo and Palawan, the Philippines with varying numbers of *P. knowlesi* cases reported. In addition to testing for infection, we measured species-specific antibody responses to *P. knowlesi* and other malaria species to identify exposed individuals. Few asymptomatic infections were detected and varying levels of *P. knowlesi* exposure was detected between sites. *P. knowlesi* exposure was identified in both men and women and associated with farm work and forest and clearing around the house. Spatial patterns and risk factors for *P. knowlesi* differed from other malaria species, highlighting the need for *knowlesi* specific disease control measures. Results suggest more people are exposed to *P. knowlesi* than are identified at clinics and exposure to *P. knowlesi* may occur in different demographic groups and geographic areas than previously reported.

Introduction

After the initial recognition of a large number of human cases of the zoonotic malaria *Plasmodium knowlesi* in 2004 and advent of routine diagnosis of malaria cases by molecular methods, increasing numbers of human *P. knowlesi* cases have been reported in Southeast Asia and *P. knowlesi* is now the most common cause of human malaria in Malaysian Borneo [1–3]. Although regional control programmes have reduced the incidence of other malaria species in Malaysia and the Philippines, such as *P. falciparum* and *vivax*, the emergence of *P. knowlesi* presents a challenge to malaria elimination programmes. Despite increasing amounts of data available for symptomatic malaria cases presenting at hospital facilities, little is known about patterns of *P. knowlesi* exposure and infection at a community level [4].

Effectively targeting resources to identify and control *P. knowlesi* requires a detailed understanding of environmental and social risk factors. Carried by long and pig-tailed macaques (*Macaca fasicularus* and *M. nemestrina*), environmental changes affecting contact between people, mosquito vectors and simian hosts are believed to contribute to this apparent emergence of *P. knowlesi* in people [5, 6]. *Anopheles balabacensis*, the main knowlesi vector, has been associated with forest environments but is also found in peridomestic and agricultural areas [7, 8]. Associations between deforestation and increases in village-level incidence have been shown for clinical cases but this may not fully reflect exposure in the wider community

[9]. Additionally, multiple studies have reported asymptomatic *P. knowlesi* infections, including in women and children, demographic groups comprising a minority of cases reported to facilities [10–14].

Patterns of community-level exposure can be assessed by the prevalence of specific antibodies against malaria parasites; these antibodies reflect exposure to previous infection and can be used to characterise the level of transmission and identify areas or groups with higher transmission [15]. These serological markers may be particularly useful in low transmission settings, where the probability of detecting infections is low [16]. Seroconversion rates derived from age specific sero-prevalence have also been shown to be closely correlated with more traditional measures of malaria transmission intensity, such as entomological inoculation rates or parasite prevalence, and can be used to identify differences in spatial patterns in transmission [17, 18]. Further, as these antibody responses represent exposure over time, longer term transmission patterns and temporal changes in transmission can be evaluated [19]. There are an increasing number of reagents for serological studies available for both *P. falciparum* and *P. vivax* e.g. [17, 20, 21]but antigens specific for *P.knowlesi* have only recently been described[22].

This study aimed to characterise these community level patterns of serological exposure to and prevalence of asymptomatic parasitemia of *P. knowlesi* and other malaria species in three case study communities where *P. knowlesi* transmission has been reported; a largely deforested and highly fragmented site at Matunggong, Kudat, an area with large patches of secondary forest bordering large scale clearing for an oil palm plantation in Limbuak, Pulau Banggi in Sabah, Malaysia and an area with intact secondary forest and some remaining primary forest in Bacungan, Palawan, The Philippines (Fig 1). These areas were selected as areas representative of locations were *P. knowlesi* transmission is occurring based on district hospital reports and were the sites of integrated entomology, primatology and social science studies within a wider research programme on risk factors for *P. knowlesi* (http://malaria.lshtm.ac.uk/MONKEYBAR). *P. knowlesi* is the main cause of reported human malaria in both the Matunggong and Limbuak sites while only few sporadic *P. knowlesi* cases have been reported from Bacungan [23–25]. Based on reporting of symptomatic cases to the national malaria programmes, the annual parasite incidence per 1000 people for *P. knowlesi* in 2014 was 12 for Matunggong, 2 for Limbuak and 0 for Bacungan.

Methods

Ethics approval and informed consent

This study was approved by the Medical Research Sub-Committee of the Malaysian Ministry of Health (NMRR-14-713-21117), the Institutional Review Board of the Research Institute for Tropical Medicine, Philippines and the Research Ethics Committee of the London School of Hygiene and Tropical Medicine (8340). Written informed consent was obtained from all participants or parents or guardians and assent obtained from children under 18 in this study and all methods were performed in accordance with relevant guidelines and regulations.

Sampling methods

This study involved comprehensive sampling of all individuals residing within the study areas. Study sites were selected based on the locations of previously reported clinical *P. knowlesi* cases and all households within these communities were enumerated and geo-located. All individuals were asked to participate in the study and consenting individuals were interviewed on demographic characteristics, movement patterns, malaria prevention methods and land use practices. Individuals were excluded if they were less than 3 months old, had not primarily resided in the area for the past month or could not be reached after three attempts to contact

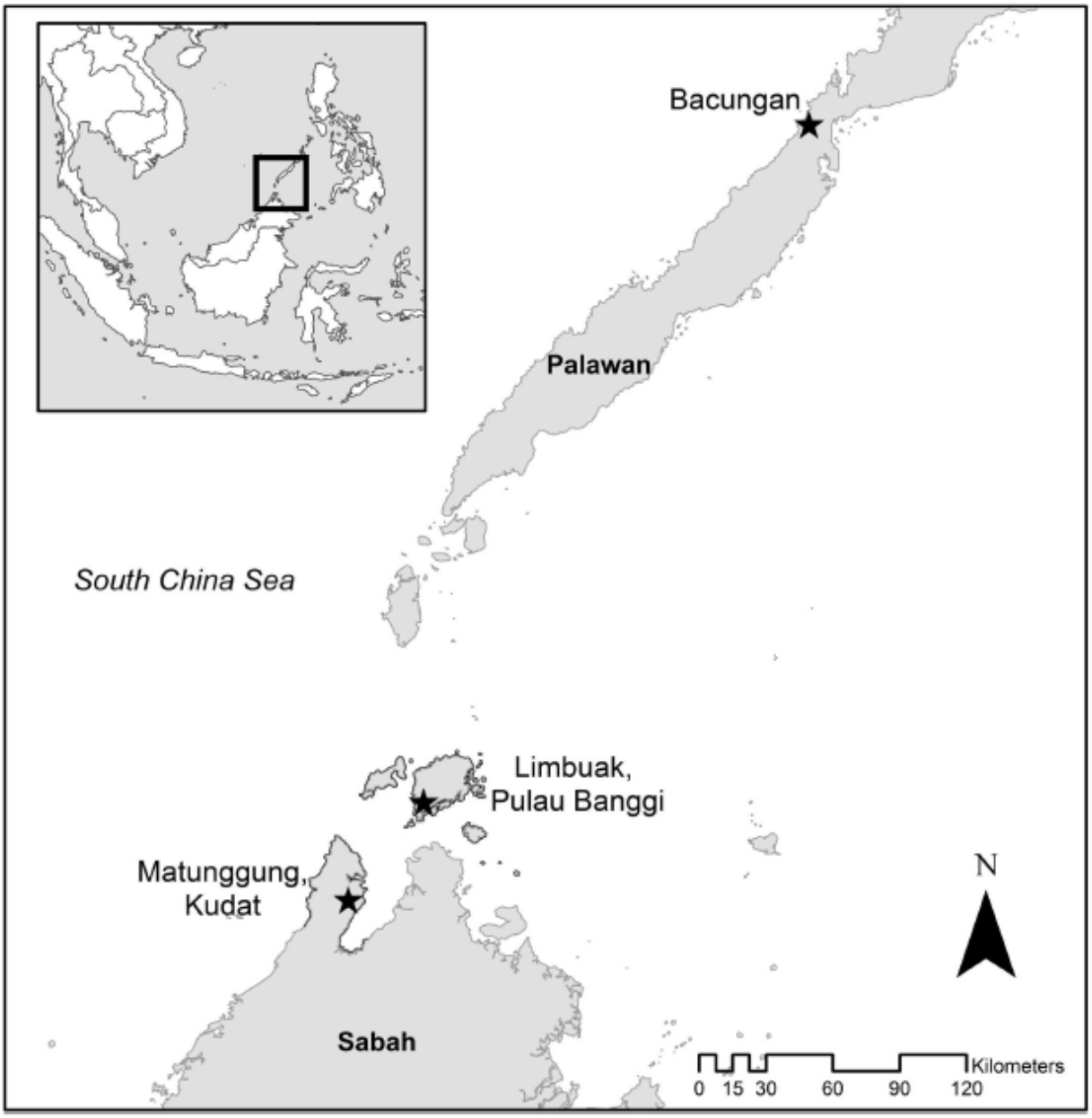


Fig 1. Study site locations in Matunggong, Kudat and Limbuak, Pulau Banggi in Sabah, Malaysia and Bacungan, Palawan, Philippines.

https://doi.org/10.1371/journal.pntd.0006432.g001

them, including during evenings and weekends. Finger-prick blood samples were collected to test for malaria infection and exposure; these included blood smears to detect malaria parasites by microscopy and approximately 200μl whole-blood specimens collected in a tube containing

EDTA (Becton-Dickinson, Franklin Lakes, New Jersey) and three 20μl spots stored on filter paper (3MM, Whatman, Maidstone, United Kingdom). Filter paper was dried and stored with desiccant at 4˚C.

Detection of malaria infection

All blood smears were examined by trained malaria microscopists. DNA was extracted from filter paper or 10 μl blood pellets using the Chelex-100 boiling method and a nested polymerase chain reaction (PCR) method targeting the *Plasmodium* small subunit ribosomal RNA (ssRNA) was used to identify malaria infected individuals, as described by [10, 26]. This assay used the genus-specific primers rPLU1 (5'-TCA AAG ATT AAG CCA TGC AAG TGA-3') and rPLU5 (5'-CCT GTT GTT GCC TTA AAC TTC-3') for nest 1 and rPLU3 (5'-TTT TTA TAA GGA TAA CTA CGG AAA AGC TGT-3') and rPLU4 (5'-TAC CCG TCA TAG CCA TGT TAG GCC AAT ACC-3') for nest 2. Thermal cycling conditions for primary and nested PCRs were 35 cycles at 94˚C, 60˚C and 72˚C. Samples positive for the *Plasmodium* genus were then screened using species specific primers targeting the ssRNA region; for *P. knowlesi* these included PkF1140 (5'-GATTCATCTATTA AAAATTTGCTTC-3') and PkR1150 (5' GAGT TCTAATCTCCGGAGAGAAA AGA 3') for 35 cycles at 50˚C, 72˚C and 94˚C. All products were visualised on a 2% agarose gel. PCR for malaria infection was performed at laboratories at the Universiti Sabah Malaysia in Malaysia and Research Institute for Tropical Medicine in the Philippines, with PCR validation of a subset of samples at the London School of Hygiene and Tropical Medicine in the UK.

Serological detection of exposure

Enzyme-linked immunosorbent assays (ELISA) were performed as previously described [27]. Briefly, 3 mm disc was excised from each dried blood spot and incubated in reconstitution buffer (PBS/tween with sodium azide) overnight at 4˚C. Antibodies were eluted from the blood spots equivalent to a 1:100 dilution of whole blood or a 1:200 dilution of serum [16]. Antibody responses were measured against apical membrane antigen-1 or the 19 kDa fragment of merozoite surface protein-1 for *P. vivax* (PvAMA-1 and PvMSP-1_{19}, respectively), *P. falciparum* (PfAMA-1 (PMID: 17192270; PMID: 19165323) and PfMSP-1_{19} (PMID: 8078519) and *P.knowlesi* SERA3 antigen 2[22]. The *Pk* serine repeat antigen (SERA) 3 antigen 2 (PKNH_0413400; chromosome 4) is a novel recombinant protein, N-terminally located between positions 826–998 aa, inclusive. SERA3 (1079 aa) belongs to a multigene family whose members encode a papain-like cysteine protease domain (ref: PMID: 21423628). In *P falciparum*, the N-terminal domain of SERA 5 is showing promise as a potential vaccine candidate (ref: PMID: 24886718, PMID: 27343834). The recombinant protein was expressed in *Escherichia coli* and affinity purified by a GST tag. Knowlesi -exposed hospital clinical case control samples showed antigen specific reactivity to the SERA3 antigen 2 recombinant when compared to responses from European malaria naïve and Ethiopian vivax-exposed serum samples (Herman et al. submitted) Eluates were tested in duplicate at a final concentration of 1:1000 for all antigens except 1:2000 for PfAMA-1. In addition, blank wells and a dilution series of the appropriate positive plasma pool were added per plate. Positive controls based on a hyper-immune endemic adult Tanzanian pool (PMID: 15792998), a lyophilised anti-malaria patient sample (NIBSC, UK; 72/96) and pooled *Pk*-exposed hospital serum samples were used to assay for *P. falciparum*, *P. vivax* and *P. knowlesi* antigens, respectively. Polyclonal rabbit anti-human IgG-HRP (Dako, Denmark) was used at 1/15,000 dilution and plates were developed using TMB (One component HRP microwell substrate, Tebu-bio). Optical density (OD) values were measured at 450 nm with a microplate reader. Values in excess of 1.5 CV between

duplicates were considered fails and re-ran. OD values were corrected by subtracting the background of the blank well per plate. For *P. falciparum* and *P. vivax* OD readings, values were normalised between plates using a standardised control. Normalisation was not done for *P. knowlesi* results due to the lack of standard control. All serological analysis was performed at the Universiti Malaysia Sabah and the London School of Hygiene and Tropical Medicine.

Environmental classification

All households and roads within these areas were geo-located using a hand-held GPS (global positioning system). Land cover maps were derived from LANDSAT 8 30m resolution satellite images [28] and supervised classification was performed using random forests [29, 30]. In order to generate training data, high- resolution aerial images of areas within study sites were produced using the Sensefly eBee unmanned aerial vehicle flown at 400 metres above ground level (UAV or drone; Sensefly, Cheseaux-sur-Lausanne, Switzerland) and processed using Postflight Terra 3D (Pix4D SA, Lausanne, Switzerland) as described by [31]. These data were manually digitised and classified as forest, agricultural land (including cropland and agroforestry such as rubber and palm oil), open areas and water bodies. Additional data on elevation, aspect and slope was extracted from the ASTER global digital elevation model [32]. All data were resampled to 30m per pixel and datasets including topographic variables, distance from roads and houses, normalised differential vegetation indices (NDVI) and Landsat satellite data were included in the initial model. Random forest models were run using 10,000 trees to ensure stability and were run iteratively with least predictive variables excluded at every run [33]. A random subset of the training data for each site was withheld to independently validate the classification; estimated classification accuracy was 88%, 97% and 85% for Matunggung, Limbuak and Palawan respectively (Fig 2).

These classified land cover maps were used to calculate distance from the household to the forest edge. The proportions of different land types surrounding all households were evaluated for 100m, 500m and 1000m buffer radii. Additionally, the level of forest fragmentation was assessed within 500m and 1000m of each household; this was represented as the ratio of forest perimeter to forest area as described by [34]. All geographic data was stored and visualised in a Geographic Information System using ArcGIS (ArcGIS, Redlands, USA) and all other analysis was performed using R statistical software (R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org).

Statistical analysis and data management

Questionnaire data was collected electronically using Pendragon Forms VI (Pendragon Software Corporation, Chicago, USA) and analysed using R statistical software. To define sero-positive individuals, mixture models were fit for normalised optical densities (ODs), with the distribution of ODs modelled as two Gaussian distributions. Cut off values to define sero-prevalence for each antigen were defined as the mean OD of the sero-negative population plus 3 standard deviations for *P. falciparum* and *P. vivax* as described by [16]. For the *P.knowlesi* antigen a more parsimonious cutoff value was defined as the mean OD plus 5 standard deviations due to a lack of prior data. Because the assays were run in different laboratories, cut off values were defined separately for each antigen, malaria species and location (Palawan and Sabah). For *P. falciparum* and *P. vivax*, individuals were considered positive if they were positive for either MSP-1 and/or AMA-1. Reversible catalytic models were fit to age sero-prevalence data using maximum likelihood methods; these models were then used to generate age sero-prevalence curves and estimate the seroconversion rate (SCR) [17]. Evidence of historical changes in transmission was explored by using profile likelihood plots. Models with two SCR were

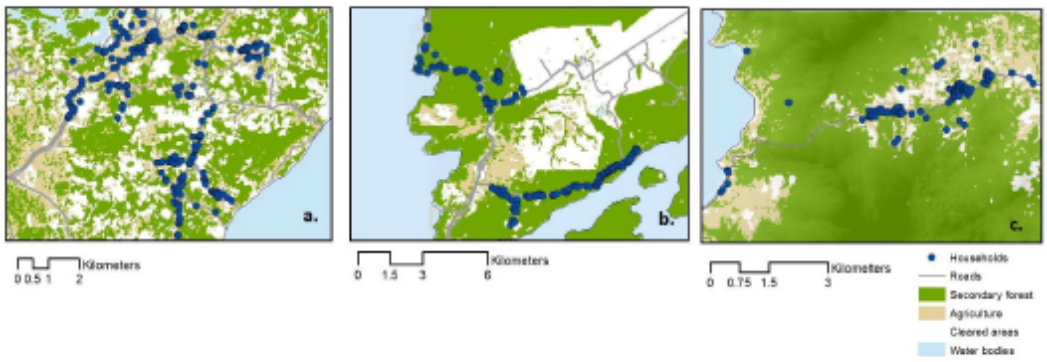


Fig 2. Land use classification of study sites. a. Highly deforested and fragmented site at Matunggong, Kudat, Sabah, Malaysia; b. Some forested area bordering large scale clearing in Limbuak, Pulau Banggi, Sabah, Malaysia; c. Mostly intact forest in Bacungan, Palawan, Philippines.

https://doi.org/10.1371/journal.pntd.0006432.g002

assessed by likelihood ratio tests and used if the fit was significantly better ($p < 0.05$) than models with a constant seroconversion rate [19]. Models were fit separately for each parasite species and site.

Risk factors associated with *P. knowlesi* sero-positivity were evaluated using multivariate logistic regression, with household included as a random effect to account for correlation between individuals from the same household. An additional model was developed to compare individuals sero-positive for *P. knowlesi* with those sero-positive for non-zoonotic malaria species. Explanatory variables included age, gender, site, individual and household level farming activities, residence in the area, elevation and distance to forest. Additionally, the proportions and configuration of different land types were extracted for each household at 100m, 500m and 1000m radii and categorised as greater or less than 30% coverage within a specific radius in the final model. Univariate analysis was conducted for all explanatory variables and variables with $p < 0.2$ were included in multivariate analyses. For highly correlated variables (such as land cover proportions at different radii), single variables were selected based on marginal increases in Akaike Information Criterion (AIC). The final adjusted models were developed by retaining all variables significant at a 0.05 level and variables were added in a forward stepwise fashion to check for interactions. Potential residual spatial autocorrelation of exposure to *P. knowlesi* was assessed separately for all sites using Moran's I.

Correlation between spatial patterns of exposure to *P. knowlesi* and nonzoonotic malaria species was explored through correlograms, plots of spatial autocorrelation with lag distances. First, ODs were log-transformed and adjusted for age by linear regression as described by [18]. For each site, cross-correlograms of antibody responses to *P.knowlesi* and each other antigen were plotted. Correlation ranges were determined by significance values ($p < 0.05$) of individual bins of lag distances of 500m. Pairwise correlation between antibody responses was determined using a simple Mantel test to test the significance of associations [35, 36].

Results

The total populations resident in the sites were 1260 in Matunggong, 1009 in Limbuak and 686 in Bacungan. Surveys were conducted from October 2014 to January 2015 in Limbuak ($n = 795$) and Matunggong ($n = 1162$) sites in Sabah and in September 2014 in Bacungan, Palawan ($n = 546$). During this time, no clinical *P. knowlesi* cases were reported from the Bacungan study site while one *P. knowlesi* case was reported in the Limbuak site and two cases were

Table 1. Demographic and environmental characteristics of included participants.

		Limbuak, Pulau Banggi (n = 795)	Matunggong, Kudat (n = 1162)	Bacungan, Palawan (n = 546)
Demographic factors				
Gender				
	Female, % (n)	52.5 (417)	51.8 (602)	43.6 (238)
	Male, % (n)	47.5 (378)	48.2 (560)	56.4 (308)
Age in years, median (IQR)		22 (9–44)	25 (10–47)	25 (11–44)
Farming or plantation work as main occupation, % (n)		14.2 (113)	30.6 (356)	12.6 (69)
Household farm activities, % (n)		68.1 (542)	88.6 (1030)	50.1 (274)
Stay overnight outside village, % (n)		8.2 (65)	13.6 (161)	29.5 (161)
Environmental factors				
Elevation (metres above sea level), median (IQR)		11 (8–15)	50 (35–75)	84 (77–114)
Distance to forest edge (metres), median (IQR)		30 (30–60)	95 (68–120)	67 (30–108)
Proportion of cleared areas around house (%), median (IQR)				
	Within 100m	0.43 (0.21–0.65)	0.63 (0.46–0.74)	0.39 (0.26–0.61)
	Within 500m	0.14 (0.10–0.24)	0.38 (0.28–0.47)	0.22 (0.16–0.26)
	Within 1000m	0.14 (0.09–0.17)	0.37 (0.31–0.39)	0.18 (0.16–0.20)
Proportion of agriculture around house (%), median (IQR)				
	Within 100m	0.14 (0.03–0.31)	0.33 (0.23–0.49)	0.43 (0.32–0.60)
	Within 500m	0.06 (0.05–0.14)	0.38 (0.28–0.48)	0.39 (0.36–0.42)
	Within 1000m	0.05 (0.03–0.10)	0.31 (0.24–0.36)	0.37 (0.29–0.39)
Proportion of forest around house (%), median (IQR)				
	Within 100m	0.31 (0.12–0.50)	0.03 (0–0.08)	0.08 (0–0.20)
	Within 500m	0.71 (0.59–0.81)	0.22 (0.13–0.34)	0.37 (0.32–0.43)
	Within 1000m	0.79 (0.76–0.86)	0.33 (0.27–0.39)	0.44 (0.40–0.52)
Forest area to perimeter ratio around house, median (IQR)				
	Within 500m	0.02 (0.01–0.02)	0.05 (0.04–0.06)	0.04 (0.03–0.04)
	Within 1000m	0.01 (0.01–0.01)	0.03 (0.03–0.04)	0.04 (0.03–0.04)

https://doi.org/10.1371/journal.pntd.0006432.t001

reported in Matunggong site. The median age of participants was 24 years (age range 3 months– 99 years) and similar proportions of men and women were sampled in all study sites. While only 22% (538/2503) of individuals reported their primary occupation as farming or plantation work, the majority of individuals (74%; 1846/2503) reported their household engaged in some agricultural activities (Table 1). The proportion of forest cover within 1km of the houses in each site ranged from 39% in Matunggong, 55% in Bacungan to 82% in Limbuak (Fig 2). The Matunggong site was the most highly fragmented, with a forest perimeter to area ratio of 0.03 compared to 0.01 in Bacungan and 0.005 in Limbuak.

Infection with malaria

Two microscopy positive individuals were identified from the Matunggong, Kudat site; these were both subsequently identified as *P. knowlesi* mono-infections by PCR. All PCR infections were re-confirmed at the laboratory in the U.K. Both of these individuals were male plantation workers (ages 21 and 25) residing in the same household. An additional two individuals in

Matunggong were microscopy negative but identified as *P. knowlesi* infected by molecular methods; these included a three year old girl and 33 year old woman residing in different villages within the study site. Only one out of these four infected individuals identified self-reported history of fever. None of the survey participants in either the Limbuak or Bacungan sites were positive by microscopy or PCR and no infections with any other malaria species were identified in any participants.

Serological assessment of exposure to *P. knowlesi*

Overall, 7.1% (178/2503) of the population surveyed was seropositive to *P. knowlesi* (Fig 3). Exposure varied substantially between study sites, with the highest *P. knowlesi* antibody prevalence detected in Limbuak, Pulau Banggi (11.7%; 93/795) followed by 6.8% (79/1162) in Matunggong Kudat. Bacungan, Palawan had the lowest sero-prevalence (1.1%; 6/546). Similar reactivity to *P. knowlesi* was observed in men (optical density (OD): med: 0.035, IQR: 0.006–0.094) and women (OD: median: 0.035, IQR: 0.007–0.089) and gender was not significantly associated with *P. knowlesi* sero-positivity (OR: 0.99, 95% CI: 0.71–1.37, p = 0.95).

Antibody prevalences to *P. falciparum* and *P. vivax* were higher in all sites, with 16.1% (364/2266) of individuals sero-positive to one or both *P. falciparum* antigens and 12.6% (270/2141) positive for one or more *P. vivax* antigens. Sero-prevalence to *P. falciparum* was 16.9% (196/1162) in Matunggong, 13.5% (107/795) in Limbuak and 10.4% (61/587) in Bacungan. In contrast, reactivity to *P. vivax* was highest in Limbuak (16.7%; 133/795) with sero-prevalences of 6.9% (80/1162) and 9.7% (57/587) in Matunggong and Bacungan respectively. Due to insufficient samples and non-systematic errors in labelling, results for all antigens were not available for all individuals. Out of individuals with complete test results for all antigens, 25.7% (499/1941) of participants were sero-positive to at least one species of malaria and 7.9% (54/1941) were sero-positive for two or more malaria species. Of individuals exposed to *P. knowlesi*, 29.7% (53/ 178) were also positive for *P. falciparum* or *P. vivax* antigens. There was no evidence of correlation between *P. knowlesi* and other antigens tested (S1 Fig).

Sero-prevalence was positively associated with increases in age for all antigens tested. However, despite this, seroreactivity, including individuals with high antibody titres (S2 Fig), was still detected in the youngest age groups and 4.2% (39/921) individuals under 15 years had antibodies to *P. knowlesi*, 3.5% (29/821) had antibodies to *P. falciparum* and 2.9% (23/792) to *P. vivax*. Changes in age sero-prevalence were more pronounced for *P. falciparum* and *P. vivax*, with 32.9% (78/237) and 28.1% (64/228) reactivity to *P. falciparum* and *P. vivax* in individuals over the age of 60 years. In contrast, antibodies for *P. knowlesi* were detected in 9.4% (25/265) of individuals over 60 years old and the highest sero-prevalence was detected in adults from 45–60 years old (11.6%; 43/370). As reactivity to *P. knowlesi* was low and not evenly

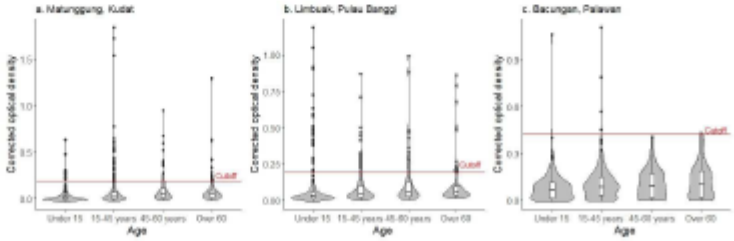


Fig 3. Violin plots of antibody density by age group: a. Matunggong, Kudat, b. Limbuak, Pulau Banggi, c. Bacungan Palawan.

https://doi.org/10.1371/journal.pntd.0006432.g003

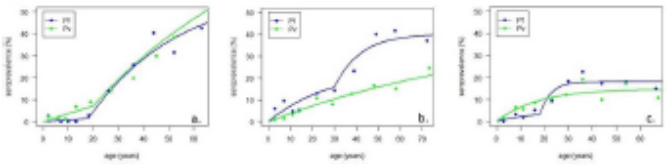


Fig 4. Seroprevalence curves for each location used to calculate SCRS (λ). a. Pulau Banggi, Sabah (Pf λ1: 0.0196 (0.0078–0.0492); Pf λ2: 0.0008 (0.0002–0.0031); Pv λ1: 0.0185 (0.0136–0.0248)) b. Matunggong, Kudat ((Pf λ1: 0.0588 (0.0198–0.1746); Pf λ2: 0.0085 (0.0063–0.0116); Pv λ: 0.0039 (0.0024–0.0064)) c. Bacungan, Palawan (Pf λ1: 0.1441 (0.0175–1.1892); Pf λ2: 0.0031 (0.0012–0.0086); Pv λ: 0.0086 (0.0044–0.0166)).

https://doi.org/10.1371/journal.pntd.0006432.g004

distributed through the population, seroconversion rates (SCR) for *P. knowlesi* could not be calculated. Historical changes in falciparum transmission intensity were apparent in all sites and SCR models fitted with two forces of infection suggest substantial reductions in *P. falciparum* transmission occurred 18–30 years ago ($p < 0.05$) (Fig 4). Strong evidence of decreased transmission intensity for *P. vivax* was only seen in Limbuak, where transmission decreased over 25-fold in the past 20 years.

Factors associated with *P. knowlesi* sero-positivity

Demographic and environmental characteristics of survey participants are summarised in (Table 1). In addition to age and site, reporting farm or plantation work as a primary occupation was positively associated with *P. knowlesi* sero-positivity (Table 2). Higher proportions of forest cover within 1km of the household and cleared areas within 500m of the house were both associated with increased odds of *P. knowlesi* positivity. While forest fragmentation, elevation and agricultural land around the house were significant within the univariate analysis, none of these variables were significant in the final multivariate model (Supplementary

Table 2. Multivariate analysis of risk factors for *P. knowlesi* seropositivity. (comparison of *P. knowlesi* exposed individuals with non-exposed individuals).

		Adjusted OR (95% CI)	P value
Age			
	Under 15 years	-	< 0.001
	15–45 years	2.05 (1.30–3.22)	
	45–60 years	2.94 (1.70–5.11)	
	Over 60 years	2.46 (1.32–4.58)	
Site			
	Palawan	-	< 0.001
	Mainland Kudat	4.30 (1.66, 11.15)	
	Pulau Banggi	10.83 (4.50, 26.10)	
Main occupation farm or plantation work			
	No	-	0.025
	Yes	1.63 (1.07, 2.48)	
Forest cover within 1km			0.004
	Less than 30%	-	
	Over 30% forest cover	2.40 (1.29, 4.46)	
Proportion of cleared/ open area within 500m of house			0.001
	Less than 30%	-	
	Over 30% cleared	2.14 (1.35, 3.40)	

https://doi.org/10.1371/journal.pntd.0006432.t002

Table 3. Multivariate analysis of risk factors for *P. knowlesi* seropositivity in malaria exposed individuals. (comparison of *P. knowlesi* exposed individuals with individuals exposed to other non-zoonotic malaria species).

		Adjusted OR (95% CI)	P value
Age			
	Under 15 years	-	0.05
	15–45 years	0.72 (0.37–1.39)	
	45–60 years	0.53 (0.26–1.06)	
	Over 60 years	0.38 (0.18–0.82)	
Site			
	Palawan	-	< 0.001
	Mainland Kudat	3.79 (1.50, 11.00)	
	Pulau Banggi	6.55 (2.88, 17.68)	
Proportion of forest within 1km of house			
	Less than 30%	-	< 0.001
	Over 30% cleared	4.86 (2.30, 11.37)	
Proportion of cleared/ open area within 500m of house			
	Less than 30%	-	0.001
	Over 30% cleared	2.70 (1.60, 4.66)	

https://doi.org/10.1371/journal.pntd.0006432.t003

information, S1 Table). Similar proportions of men and women reacted to *P. knowlesi* in all sites and gender was not associated with sero-positivity.

Individuals reacting to *P. knowlesi* were more likely to be younger than individuals sero-positive for only non-zoonotic malaria species (Table 3). Forest cover was not associated with exposure to non-zoonotic malaria and malaria positive individuals residing in areas with high forest cover around the house had 4.86 (95% CI: 2.30–11.37) the odds of being positive for *P. knowlesi*. Similarly, cleared areas around the house were also positively associated with *P. knowlesi* cases compared to other malaria species.

Based on Moran's I, there was no evidence of residual spatial autocorrelation for *P. knowlesi* antibody responses (Moran's I $p > 0.2$ for all sites). There was no significant spatial correlation detected between age-adjusted antibody responses for *P. knowlesi* and other malaria species for either Matunggong or Limbuak ($p > 0.30$ for all pairwise comparisons). Comparisons between *P. knowlesi* and other malaria species could not be evaluated for Bacungan due to the low prevalence of *P. knowlesi* sero-positivity.

Discussion

This is the first study to describe exposure to *P. knowlesi* through antigen specific antibody responses and associated risk factors and is one of few studies to assess *P. knowlesi* carriage prevalence at a community level. Results indicate spatial and temporal patterns of *P. knowlesi* transmission differ markedly from other non-zoonotic malaria species within the region. Although *P. knowlesi* sero-positivity was associated with some landscape attributes within these communities, extensive cross sectional surveys are needed to identify ecological risk factors across a broader geographic scale.

Sero-prevalence data indicate distinct heterogeneities in *P. knowlesi* transmission intensity between sites. Although formal comparisons between *P. knowlesi* infection and exposure could not be undertaken due to the low prevalence of parasite carriage, these geographical differences in transmission mirror hospital-based reporting rates in the study sites at Kudat, Pulau Banggi and Palawan [23–25]. These results also highlight the utility of serological techniques to identify differences in transmission intensity in settings where the sensitivity of parasite

prevalence surveys is limited by the scarcity of infected individuals and suboptimal diagnostics. This is the first time these knowlesi-specific antigens have been used at a population level to assess species-specific exposure to malaria. Although high levels of homology between *P. knowlesi* and *P. vivax* indicate the possibility of cross reactivity between antigens, relatively low numbers of individuals were identified as sero-positive for both knowlesi and vivax (2%; 43/2102 individuals with results for both assays) and individuals could have been plausibly exposed to both species due to the co-endemicity of these species within this region. Additional work has been done to characterise response to *P. knowlesi* in vivax-exposed individuals and validate these antigens for population-based studies [22].

Changes in seroconversion rates can also reflect temporal changes in malaria transmission. In Sabah, state-wide malaria notification records describe dramatic decreases in clinical *P. falciparum* and *P. vivax* cases within the past 20 years following the scale up of malaria control and elimination programmes [2]. The Philippines has also reported a substantial decline in the number of malaria cases reported in the past few decades, most notably for *P. falciparum*[37]. These changes are evident in seroconversion rates to non-zoonotic malaria species from the 3 sites with over 5-fold difference between current and previous SCRs. *P. knowlesi* exposure was identified in children under 5 in all sites, suggesting recent or on-going transmission albeit at a low level. Further work is needed to refine *P. knowlesi* serological analysis to allow for antigenic variation, identify further antigenic targets and assess the differential responsiveness of individuals and longevity of antibody responses [38].

Despite these similarities between existing case data and community-level exposure to *P. knowlesi*, levels of exposure between different demographic groups varied markedly from clinical case reports. While clinical *P. knowlesi* has been commonly reported in adult men, men and women had similar antibody reactivity to *P. knowlesi* antigens in all sites [23, 39]. Within Kudat district, wide age distributions and family clusters of knowlesi cases have previously been described; however, from 2012–2015, 73% (84/115) and 77% (27/35) of all clinical cases reported from Kudat and Pulau Banggi respectively were men [23]. Asymptomatic knowlesi carriage has been detected in higher proportions of women by this study and other studies; however these results are extremely limited by sampling design and the small numbers of infected individuals detected [10, 12]. As forest and agricultural activities have been identified as risk factors for clinical *P. knowlesi* infection, more men could develop clinical infections due to higher exposure or number of bites; however, this requires further research to assess [40]. Larger scale population-based cross sectional surveys are needed to determine if these patterns occur in the wider community and if *P. knowlesi* affects groups which may be underrepresented by current passive surveillance systems.

P. knowlesi exposure was also associated with landscape factors. Both the proportion of forest cover and cleared areas around the household were positively associated with knowlesi sero-positivity, potentially reflecting the higher vectorial capacity and sporozoite rates reported in secondary forest within these study sites [7]. Although plantation or farm work as a primary occupation was associated with increased exposure and previous reports have described associations between *P. knowlesi* and forest activities, data on movement into different environments was not available for all survey participants [39, 41]. Instead, to explore the potential range of spatial interactions between people and mosquito vectors, proportions of habitat were evaluated at different buffer distances around houses. The significance of both clearing and forest areas at different radii suggests the importance of edge effects, transition areas between habitats where increased overlap of human, macaque and mosquito populations may occur [9, 42, 43]. Despite this, no associations were identified between metrics of fragmentation or distance to forest edges; this may reflect the limited environmental variation within these small spatial scales. Future studies could assess the importance of these variables across different ecotypes as

well as collect more detailed data on the human movement into different environments, particularly during peak mosquito biting times.

These spatial patterns differed markedly from exposure to other non-zoonotic malaria species. Individuals with antibodies to *P. knowlesi* were more likely to reside in areas with higher proportions of forest cover; this may reflect differences in disease dynamics between species or temporal changes in transmission. Because of the longevity of antibody responses and the rapid rates of land use change within these areas, seroreactivity to non-zoonotic species is probably more likely to be associated with past rather than current environmental factors. The main mosquito vector species of *P. knowlesi*, *Anopheles balabacensis*, was historically incriminated as the main vector of other human malaria species within these same areas [44, 45]. While these vectors have been primarily associated with forest habitats, high vector densities have also been reported in small scale farms and other habitat types [7, 42, 46]. Deforestation and increased application of vector control measures may have triggered changes in vector composition and biting preferences; similarly, habitat changes and encroachment of human settlements into forest areas may have also led to changes in macaque population densities and closer contact between macaques, people and mosquito vectors [6, 47].

The main limitations of this study are the non-randomised population sampling approach and limited geographical scale. While this study describes fine scale patterns of malaria exposure and infection within these three case study communities, these results cannot be generalised to extrapolate *P. knowlesi* risks across wider populations. As this study surveyed three relatively homogenous populations, there was minimal variation in environment, ethnicity, socioeconomic status and access to healthcare within each site. Identifying environmental and population-level risk factors will require randomised sampling across a wider ecological gradient; community level data on presence and absence of exposure and infection are required to understand spatial heterogeneity of disease transmission and develop and refine predictions of disease risk [48]. Additionally, extensive surveys of parasite prevalence may allow the application of genetic approaches to track parasite diversity and transmission and explore the roles of host and parasite genetic factors.

Despite these limitations, this study describes *P. knowlesi* infection and exposure within these communities and illustrates how serologic markers can be used to describe differences in transmission intensity between malaria species in low transmission settings. Results from these surveys indicate patterns of *P. knowlesi* exposure and infection within the community may be substantially different from cases detected by passive surveillance systems. Cross sectional surveys across a broader geographical scale are needed to describe spatial variation in transmission intensity and identify associated environmental and population-based risk factors. Integration of serology into these surveys would provide vital information on rare infections for control programmes [49].

Supporting information

S1 Checklist. STROBE checklist for observational studies.
(DOCX)

S1 Fig. Comparison of normalised optical densities for *P. knowlesi* and other antigens: Optical densities and cut offs for Sabah, Malaysia.
(TIF)

S2 Fig. Comparison of normalised optical densities for *P. knowlesi* and other antigens: Optical densities and cut offs for Palawan, Philippines.
(TIF)

S3 Fig. Violin plots of antibody density for *P. falciparum* by age group and site. a. *P. falciparum* AMA-1 antibody density; b. *P. falciparum* MSP-1 antibody density.
(TIF)

S4 Fig. Violin plots of antibody density for *P. vivax* by age group and site. a. *P. vivax* AMA-1 antibody density; b. *P. vivax* MSP-1 antibody density.
(TIF)

S1 Table. Proportions and univariate analysis of risk factors for *P. knowlesi* seropositivity.
(DOCX)

Acknowledgments

The authors would like to thank Albert M Lim (Infectious Diseases Society Sabah), Judy Dorothy Madarcos, Jo Anne Bibit (Research Institute for Tropical Medicine) and the teams in Sabah and Palawan for their support for this project.

Author Contributions

Conceptualization: Kimberly M. Fornace, Timothy William, Fe Espino, Jonathan Cox, Chris J. Drakeley.

Data curation: Tommy R. Abidin, Pauline J. Lorenzo.

Formal analysis: Kimberly M. Fornace.

Funding acquisition: Jonathan Cox, Chris J. Drakeley.

Investigation: Kimberly M. Fornace, Tommy R. Abidin, Tock Hing Chua, Sylvia Daim, Pauline J. Lorenzo, Lynn Grignard, Nor Afizah Nuin, Lau Tiek Ying, Matthew J. Grigg, Fe Espino, Jonathan Cox, Kevin K. A. Tetteh, Chris J. Drakeley.

Methodology: Kimberly M. Fornace, Lou S. Herman, Sylvia Daim, Pauline J. Lorenzo, Lynn Grignard, Nor Afizah Nuin, Lau Tiek Ying, Jonathan Cox, Kevin K. A. Tetteh, Chris J. Drakeley.

Supervision: Timothy William, Fe Espino, Kevin K. A. Tetteh, Chris J. Drakeley.

Writing – original draft: Kimberly M. Fornace.

Writing – review & editing: Lou S. Herman, Tommy R. Abidin, Tock Hing Chua, Sylvia Daim, Lynn Grignard, Nor Afizah Nuin, Lau Tiek Ying, Matthew J. Grigg, Timothy William, Fe Espino, Jonathan Cox, Chris J. Drakeley.

References

1. Singh B., et al., A large focus of naturally acquired Plasmodium knowlesi infections in human beings. Lancet, 2004. 363(9414): p. 1017–24. https://doi.org/10.1016/S0140-6736(04)15836-4 PMID: 15051281
2. William T., et al., Increasing incidence of Plasmodium knowlesi malaria following control of P. falciparum and P. vivax Malaria in Sabah, Malaysia. PLoS Negl Trop Dis, 2013. 7(1): p. e2026. https://doi.org/10.1371/journal.pntd.0002026 PMID: 23359830
3. Singh B. and Daneshvar C., Plasmodium knowlesi malaria in Malaysia. Med J Malaysia, 2010. 65(3): p. 166–72. PMID: 21939162
4. Shearer F.M., et al., Estimating Geographical Variation in the Risk of Zoonotic Plasmodium knowlesi Infection in Countries Eliminating Malaria. PLoS Negl Trop Dis, 2016. 10(8): p. e0004915. https://doi.org/10.1371/journal.pntd.0004915 PMID: 27494405

5. Singh B. and Daneshvar C., Human infections and detection of Plasmodium knowlesi. Clin Microbiol Rev, 2013. 26(2): p. 165–84. https://doi.org/10.1128/CMR.00079-12 PMID: 23554413
6. Imai N., et al., Transmission and control of Plasmodium knowlesi: a mathematical modelling study. PLoS Negl Trop Dis, 2014. 8(7): p. e2978. https://doi.org/10.1371/journal.pntd.0002978 PMID: 25058400
7. Wong M.L., et al., Seasonal and Spatial Dynamics of the Primary Vector of Plasmodium knowlesi within a Major Transmission Focus in Sabah, Malaysia. PLoS Negl Trop Dis, 2015. 9(10): p. e0004135. https://doi.org/10.1371/journal.pntd.0004135 PMID: 26448052
8. Manin B.O., et al., Investigating the Contribution of Peri-domestic Transmission to Risk of Zoonotic Malaria Infection in Humans. PLoS Negl Trop Dis, 2016. 10(10): p. e0005064. https://doi.org/10.1371/journal.pntd.0005064 PMID: 27741235
9. Fornace K.M., et al., Association between Landscape Factors and Spatial Patterns of Plasmodium knowlesi Infections in Sabah, Malaysia. Emerg Infect Dis, 2016. 22(2): p. 201–9. https://doi.org/10.3201/eid2202.150656 PMID: 26812373
10. Fornace K.M., et al., Asymptomatic and Submicroscopic Carriage of Plasmodium knowlesi Malaria in Household and Community Members of Clinical Cases in Sabah, Malaysia. J Infect Dis, 2015.
11. Marchand R.P., et al., Co-infections of Plasmodium knowlesi, P. falciparum, and P. vivax among Humans and Anopheles dirus Mosquitoes, Southern Vietnam. Emerg Infect Dis, 2011. 17(7): p. 1232–9. https://doi.org/10.3201/eid1707.101551 PMID: 21762577
12. Van den Eede P., et al., Human Plasmodium knowlesi infections in young children in central Vietnam. Malar J, 2009. 8: p. 249. https://doi.org/10.1186/1475-2875-8-249 PMID: 19878553
13. Herdiana H., et al., Malaria risk factor assessment using active and passive surveillance data from Aceh Besar, Indonesia, a low endemic, malaria elimination setting with Plasmodium knowlesi, Plasmodium vivax, and Plasmodium falciparum. Malar J, 2016. 15: p. 468. https://doi.org/10.1186/s12936-016-1523-z PMID: 27619000
14. Lubis I.N., et al., Contribution of Plasmodium knowlesi to multi-species human malaria infections in North Sumatera, Indonesia. J Infect Dis, 2017.
15. Corran P., et al., Serology: a robust indicator of malaria transmission intensity? Trends Parasitol, 2007. 23(12): p. 575–82. https://doi.org/10.1016/j.pt.2007.08.023 PMID: 17988945
16. Bousema T., et al., Serologic markers for detecting malaria in areas of low endemicity, Somalia, 2008. Emerg Infect Dis, 2010. 16(3): p. 392–9. https://doi.org/10.3201/eid1603.090732 PMID: 20202412
17. Drakeley C.J., et al., Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. Proc Natl Acad Sci U S A, 2005. 102(14): p. 5108–13. https://doi.org/10.1073/pnas.0408725102 PMID: 15792998
18. Wilson S., et al., Age-adjusted Plasmodium falciparum antibody levels in school-aged children are a stable marker of microgeographical variations in exposure to Plasmodium infection. BMC Infect Dis, 2007. 7: p. 67. https://doi.org/10.1186/1471-2334-7-67 PMID: 17603885
19. Stewart L., et al., Rapid assessment of malaria transmission using age-specific sero-conversion rates. PLoS One, 2009. 4(6): p. e6083. https://doi.org/10.1371/journal.pone.0006083 PMID: 19562032
20. Helb D.A., et al., Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. Proc Natl Acad Sci U S A, 2015. 112(32): p. E4438–47. https://doi.org/10.1073/pnas.1501705112 PMID: 26216993
21. Cook J., et al., Using serological measures to monitor changes in malaria transmission in Vanuatu. Malar J, 2010. 9: p. 169. https://doi.org/10.1186/1475-2875-9-169 PMID: 20553604
22. Herman, L.S., et al., Identification and validation of a novel panel of Plasmodium knowlesi biomarkers of serological exposure. submitted.
23. Grigg M.J., et al., Factors that are associated with the risk of acquiring Plasmodium knowlesi malaria in Sabah, Malaysia: a case-control study protocol. BMJ Open, 2014. 4(8): p. e006004. https://doi.org/10.1136/bmjopen-2014-006004 PMID: 25149186
24. Luchavez J., et al., Human Infections with Plasmodium knowlesi, the Philippines. Emerg Infect Dis, 2008. 14(5): p. 811–3. https://doi.org/10.3201/eid1405.071407 PMID: 18439369
25. Barber B.E., et al., Epidemiology of Plasmodium knowlesi malaria in north-east Sabah, Malaysia: family clusters and wide age distribution. Malar J, 2012. 11: p. 401. https://doi.org/10.1186/1475-2875-11-401 PMID: 23216947
26. Singh B., et al., A genus- and species-specific nested polymerase chain reaction malaria detection assay for epidemiologic studies. Am J Trop Med Hyg, 1999. 60(4): p. 687–92. PMID: 10348249
27. Corran P.H., et al., Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. Malar J, 2008. 7: p. 195. https://doi.org/10.1186/1475-2875-7-195 PMID: 18826573

28. Land Processes Distributed Active Archive Center (LP DAAC), Landsat 8 Operational Land Imager. 2014, NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center: Sioux Falls, South Dakota.
29. Gislason P.O., Benediktsson J.A., and Sveinsson J.R., Random forests for land cover classification. Pattern Recognition Letters, 2006. 27(4): p. 294–300.
30. Breiman L., Random Forests. Machine Learning, 2001. 45: p. 5–32.
31. Fornace K.M., et al., Mapping infectious disease landscapes: unmanned aerial vehicles and epidemiology. Trends Parasitol, 2014. 30(11): p. 514–519. https://doi.org/10.1016/j.pt.2014.09.001 PMID: 25443854
32. Land Processes Distributed Active Archive Center (LP DAAC), Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model (ASTER GDEM) Version 2. 2015, NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observatoin and Science (EROS) Center: Sioux Falls, South Dakota.
33. Diaz-Uriarte R., Gene SrF and varSelRF: a web-based tool and R package for gene selection and classification using random forest. BMC Bioinformatics, 2007. 8: p. 328. https://doi.org/10.1186/1471-2105-8-328 PMID: 17767709
34. Mcgarigal K. and Cushman S., Quantifying Terrestrial Habitat Loss and Fragmentation: A Protocol. 2005.
35. Mantel N., The detection of disease clustering and a generalized regression approach. Cancer Res, 1967. 27(2): p. 209–20. PMID: 6018555
36. Goslee S.C. and Urban D.L., The ecodist package for dissimilarity-based analysis of ecological data. J Stat Software, 2007. 22: p. 1–19.
37. DOH P., WHO, and UCSF, Eliminating malaria: case-study 6, Progress towards subnational elimination in the Philippines. 2014, World Health Organisation: Geneva.
38. Divis P.C., et al., Admixture in Humans of Two Divergent Plasmodium knowlesi Populations Associated with Different Macaque Host Species. PLoS Pathog, 2015. 11(5): p. e1004888. https://doi.org/10.1371/journal.ppat.1004888 PMID: 26020959
39. Barber B.E., et al., A prospective comparative study of knowlesi, falciparum, and vivax malaria in Sabah, Malaysia: high proportion with severe disease from Plasmodium knowlesi and Plasmodium vivax but no mortality with early referral and artesunate therapy. Clin Infect Dis, 2013. 56(3): p. 383–97. https://doi.org/10.1093/cid/cis902 PMID: 23087389
40. Grigg M.J., et al., Individual-level factors associated with the risk of acquiring human Plasmodium knowlesi malaria in Malaysia: a case control study. Lancet Planetary Health, 2017. 1: p. e97–104. https://doi.org/10.1016/S2542-5196(17)30031-1 PMID: 28758162
41. Daneshvar C., et al., Clinical and laboratory features of human Plasmodium knowlesi infection. Clin Infect Dis, 2009. 49(6): p. 852–60. https://doi.org/10.1086/605439 PMID: 19635025
42. Vythilingam I., et al., Natural transmission of Plasmodium knowlesi to humans by Anopheles latens in Sarawak, Malaysia. Trans R Soc Trop Med Hyg, 2006. 100(11): p. 1087–8. https://doi.org/10.1016/j.trstmh.2006.02.006 PMID: 16725166
43. Moyes C.L., et al., Predicting the geographical distributions of the macaque hosts and mosquito vectors of Plasmodium knowlesi malaria in forested and non-forested areas. Parasit Vectors, 2016. 9(1): p. 242.
44. Hii J. and Rueda L.M., Malaria vectors in the Greater Mekong Subregion: overview of malaria vectors and remaining challenges. Southeast Asian J Trop Med Public Health, 2013. 44 Suppl 1: p. 73–165; discussion 306–7.
45. Hii J., et al., Transmission dynamics and estimates of malarial vectorial capacity for An. balabacensis and An. flavirostris on Banggi island, Sabah, Malaysia. Ann Trop Med Parasitol, 1988. 82: p. 91–101. PMID: 3041932
46. Tan C.H., et al., Bionomics of Anopheles latens in Kapit, Sarawak, Malaysian Borneo in relation to the transmission of zoonotic simian malaria parasite Plasmodium knowlesi. Malar J, 2008. 7: p. 52. https://doi.org/10.1186/1475-2875-7-52 PMID: 18377652
47. Guerra C.A., Snow R.W., and Hay S.I., A global assessment of closed forests, deforestation and malaria risk. Ann Trop Med Parasitol, 2006. 100(3): p. 189–204. https://doi.org/10.1179/136485906X91512 PMID: 16630376
48. Brock P.M., et al., Plasmodium knowlesi transmission: integrating quantitative approaches from epidemiology and ecology to understand malaria as a zoonosis. Parasitology, 2016. 143(4): p. 389–400. https://doi.org/10.1017/S0031182015001821 PMID: 26817785
49. Metcalf C.J., et al., Use of serological surveys to generate key insights into the changing global landscape of infectious disease. Lancet, 2016. 388(10045): p. 728–30. https://doi.org/10.1016/S0140-6736(16)30164-7 PMID: 27059886

APPENDIX III: PROTOCOL

Blood spot sample preparation standard operating protocol (SOP)

1. OVERVIEW

This SOP describes the processes and preparation of filter paper blood samples for the measurement of total IgG by ELISA.

2. SAFETY

Adhere to local safety regulations. Wear appropriate personal protective equipment.

3. MATERIALS

Equipment:

a) 96 deep-well plates (Costar 0.5ml V well bottom assay block)

b) 96 cap sealing mats (Thermo scientific)

c) 8 or 12-channel 300 µl Pipette

d) Tweezers

e) Leather hole puncher (3 mm is ideal)

f) Protective nitrile gloves

g) Plate shaker

Documents:

Deep-well Plate Plan

Reagents:

All reagents should be stored according to the instructions supplied with them and disposed of at the expiry date recorded on the product.

a) Sodium Azide, BDH

b) Tween20, Sigma

c) NaH_2PO_4 (Sodium dihydrogen orthophosphate), VWR International Ltd

d) Na_2HPO_4 (di-sodium hydrogen orthophosphate), VWR International Ltd

e) NaCl (Sodium chloride), Fisher scientific

4. PREPARATION OF BUFFER SOLUTIONS

All buffer solutions should be clearly labelled with:

* Reagent name

* Expiry date

* Preparation date

* Name of person who prepared the buffer

Reagents should be stored under appropriate conditions. See Table 1 for details on preparation and storage.

Table 1: Buffer solutions - preparation

Buffer solution	Reagent/chemical	Amount/Volume
Phosphate buffered saline (PBS) 10X pH 7.2	NaH_2PO_4	5.7 g
	Na_2HPO_4	16.7 g
	NaCl	85 g
	Distilled water	Make up to 1 L
Store at room temperature, dispose of after one month		
PBS/Tween (0.05%)	PBS 10X	100 ml
	Tween 20	0.5 ml
	Distilled water	900 ml
Make up as needed daily, dispose of unused solution at the end of each day		
Reconstitution buffer	Sodium Azide	1 g
	PBS/Tween	1 L
Store at room temperature, dispose of after 6 months (or earlier if solution becomes cloudy) **NB:** Sodium Azide is **Poisonous**. Adhere to appropriate safety regulations for handling and disposal.		

NB: Fully dissolve the reagent/chemical prior to use

5. PREPARATION OF BLOODSPOTS

5.1. Label a 96 deep-well with the study ID, date, dilution 1/200 and label all deep wells with an unique ID, usually label deep well sequentially DW1, DW2, DW3, DW4 etc.

5.2. Use a paper template to record the location of the sample (Deep-well Plate Plan; example at the end of the document), indicating which bloodspots occupy which well

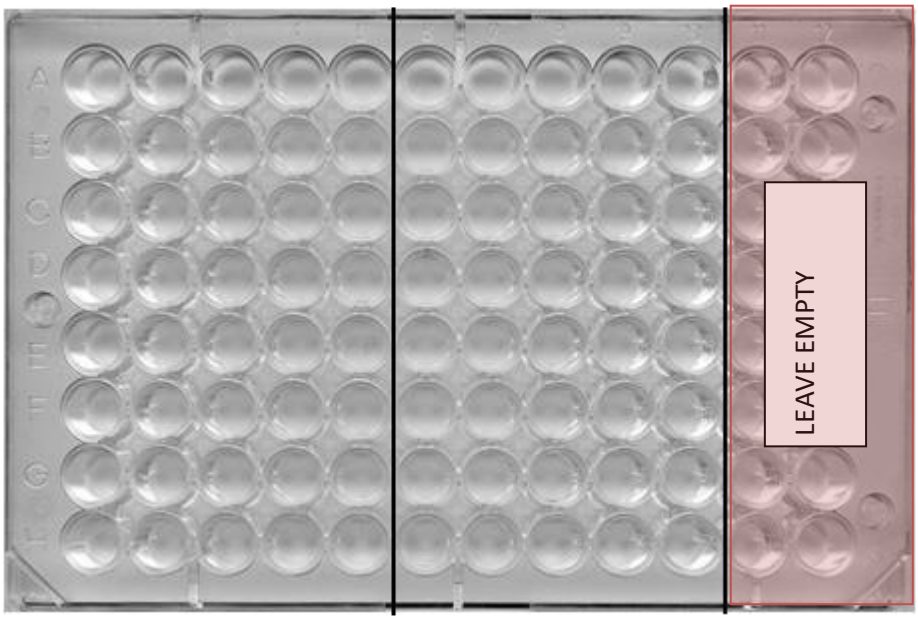


Figure 1: How to label a deep well plate.

5.3. Add samples in the direction of 1A-H, 2A-H, 3A-H etc

5.4. Use all wells except leave columns 11 & 12 empty (this will become clear why at a later date), as indicated in Figure 1.

5.5. Using the hole punch, cut a bloodspot disc from the middle of the filter paper sample (See Figure 2). It helps to have the hole of the punch facing up so you can see your positioning, especially when the spots are small (Figure 2b).

A B

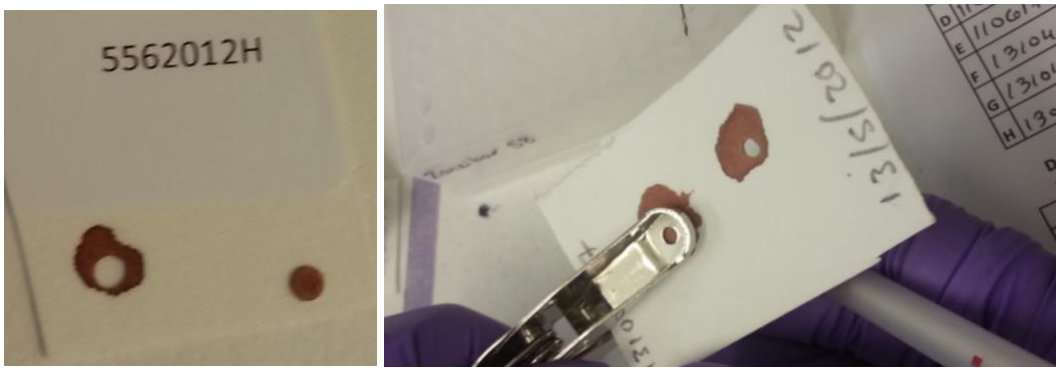


Figure 2. Example of location to cut the bloodspot from the filter paper sample. (A) Example of a cut spot. (B) Showing the suggested positioning of the hole punch over the spot

5.6. Tap the blood spot out onto the work bench (ensuring it is cleaned beforehand) (Figure 3a), use tweezers to pick up the bloodspot disc and place in the well (Figure 3b). Record the sample identity number (ID) on the plate plan.

A B

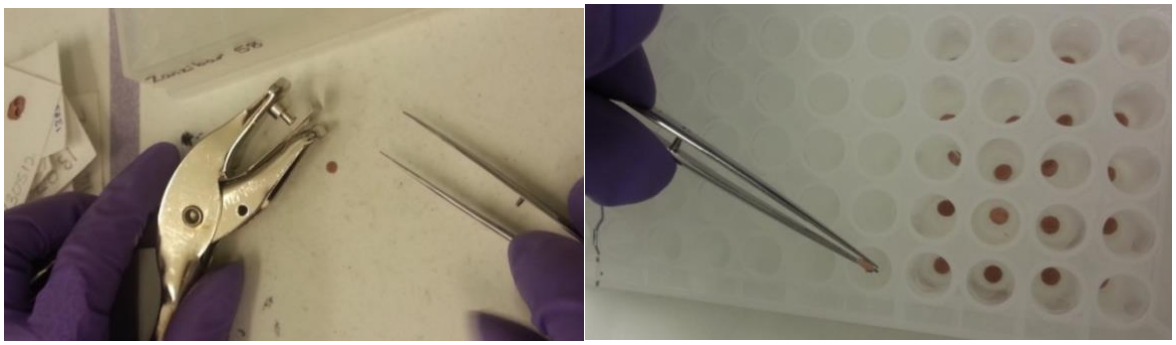


Figure 3: Handling blood spot discs once cut. (A) Tap the disc onto the work bench and then (B) pick up disc using tweezers and transfer to the deep well plate.

5.7. If handling large numbers of papers, stack in lots of 8 (1 column per stack) after cutting (Figure 4), and repackage at the end of the plate. This increases processing speed and makes it easier to keep track of where you are at (matching number of papers TO number of discs in deep well plate AND id's entered into plan).

Figure 4: Place processed papers in piles of 8 to easily keep track of where you are up to.

5.8. Depending on paper filter type & size of the spot cutter, will determine the amount of reconstitution buffer required. Optimal should be a 3 mm diameter bloodspot cutter. See **Table 2** for details.

5.9. Add the appropriate amount of the Reconstitution buffer (prepared in step 4), to the wells containing bloodspots

Table 2 –Filter paper reconstitution based on size of cutter

Spot diameter (mm)	Amount of blood (µl)	Volume to reconstitute (µl)
1.0	0.2	20
1.5	0.5	50
2.0	0.9	90
2.5	1.4	140
3.0	2.1	200
3.5	2.8	280
4.0	3.7	360
4.5	4.7	460
5.0	5.8	570
5.5	7.0	690
6.0	8.3	820

*based on serum dilution equivalent to 1/200 dilution using 3MM (Whatman) filterpaper

5.10. Place a sealing mat on the deep-well plate. Ensure it is tightly sealed

5.11. Place on plate shaker to rock gently overnight (5 - 10 revs/minute)

5.12. Next day inspect the plate: the circles of filter paper should appear white and the solution pink/red in all cases.

5.13. Reject samples where the circle is still red and the solution pale (Figure 3).

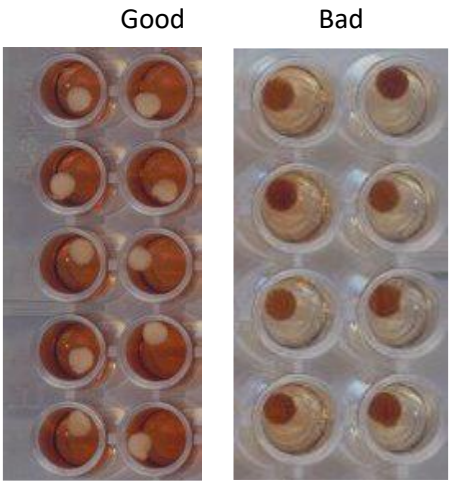


Figure 5: Bloodspot elution showing good (pink/red solution and white spot) and bad elutions (spot remains red and solution shows little colour).

5.14. Plates stored, sealed, at 4°C for up to 1 month or 1 year at - 20°C.

5.12. Type up written plate plans into an Excel spreadsheet. This allows for samples IDs to be easily merged with the processed data.

STUDY NAME:

DATE: DEEP WELL ID: NAME:

	1	2	3	4	5	6	7	8	9	10
A										
B										
C										
D										
E										
F										
G										
H										

APPENDIX IV: CHAPTER 4 SUPPLEMENTARY DATA

Supplementary figures for chapter 4: Identification and validation of a novel panel of *Plasmodium knowlesi* serological biomarkers of exposure, and their subsequent use in the measurement of exposure and infection in Malaysia and The Philippines.

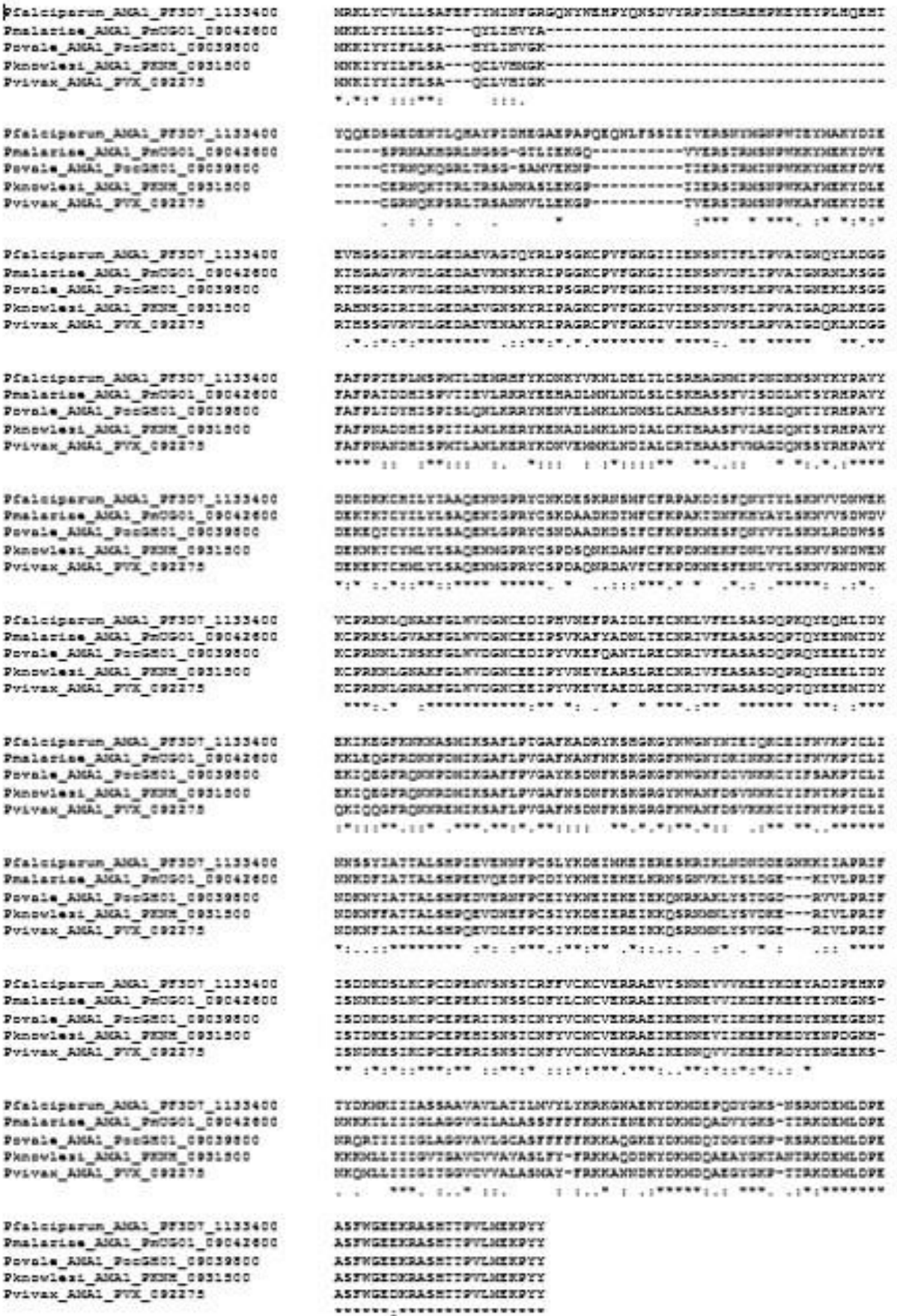
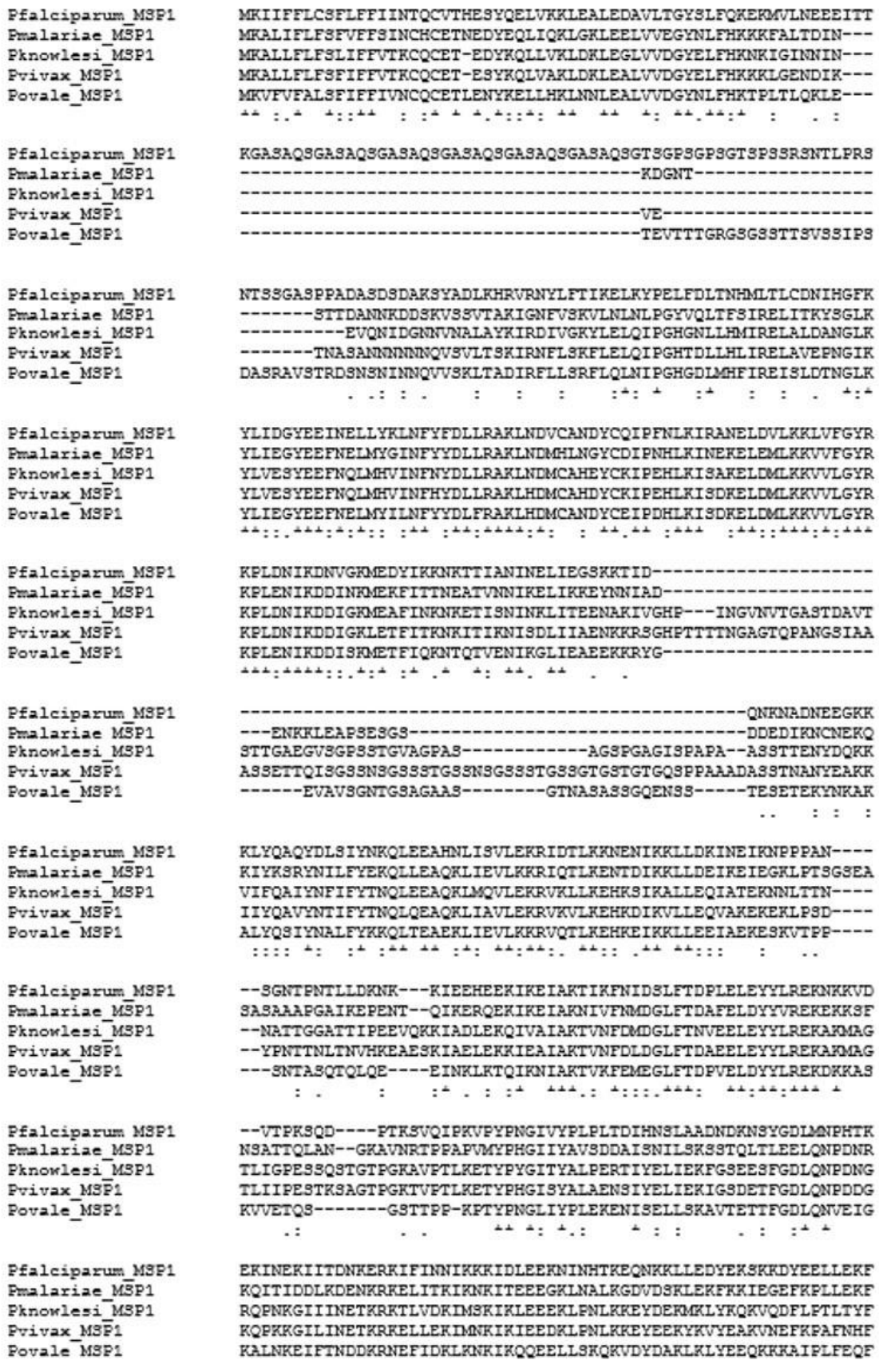
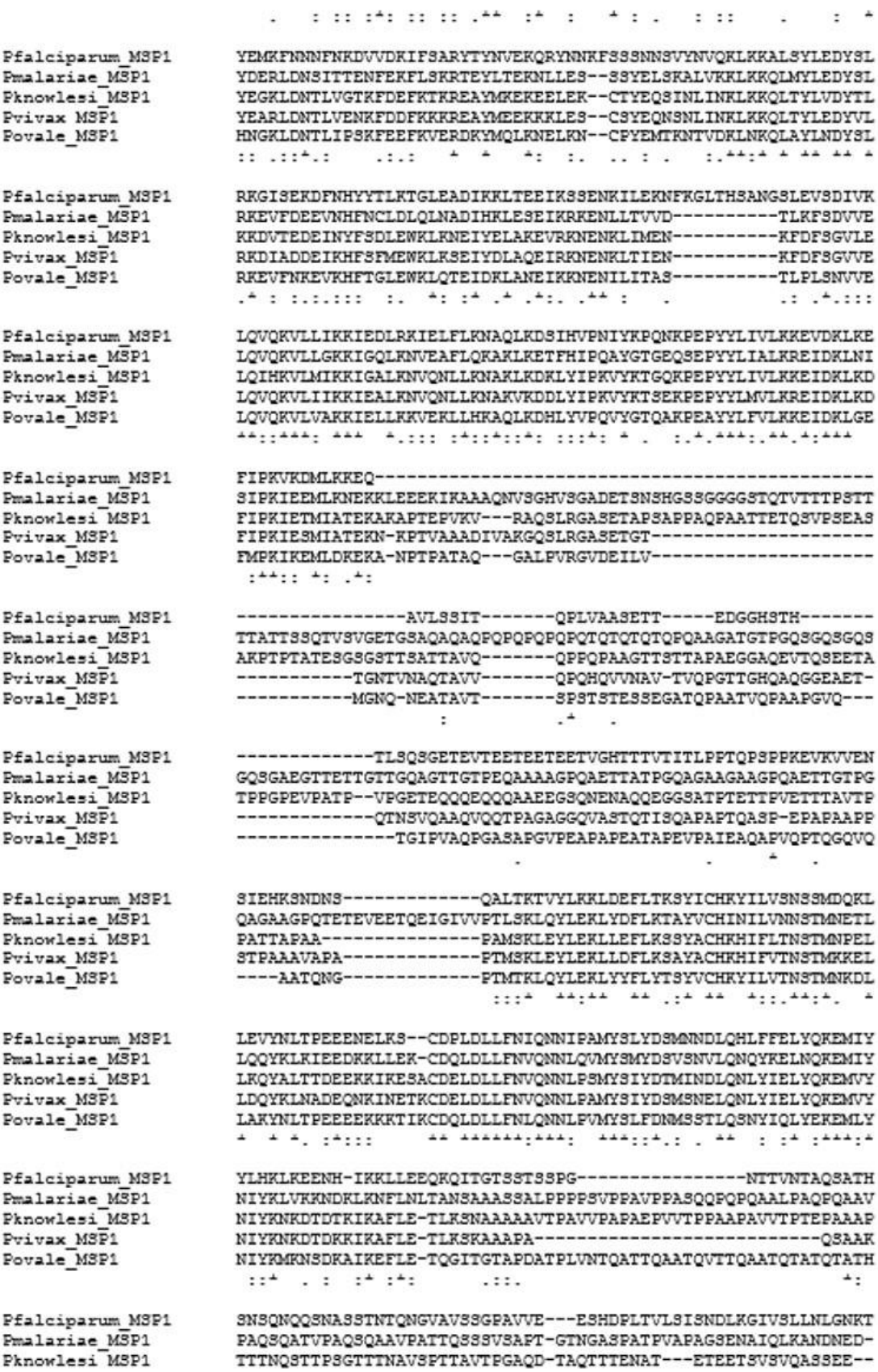

PfalciparumAMA1_PF3D7_1133400
PmalariaeAMA1_PmUG01_09042600
PovaleAMA1_PocGH01_09039800
PknowlesiAMA1_PKNH_0931300
PvivaxAMA1_PVX_092275

Figure 1 Amino acid sequences alignments for AMA1 gene sequence between P. knowlesi, P. falciparum, P. vivax, P. malariae and P. ovale/P. simiovale. Asterisks indicate fully conserved residues, colons indicates strong residue conservation (>0.5, Gonnet PAM 250 matrix), period indicates weak residue conservation (=<0.5, Gonnet PAM 250 matrix). Blank spaces indicate no residue conservation


Pfalciparum_MSP1
Pmalariae_MSP1
Pknowlesi_MSP1
Pvivax_MSP1
Povale_MSP1


Pfalciparum_MSP1
Pmalariae_MSP1
Pknowlesi_MSP1
Pvivax_MSP1
Povale_MSP1

```
Pvivax_MSP1         PSGQAGTTPVTTTA--PVTTTTVTPSPQT-SVVTSTPPTP----QAEENRRVGGNSEE--
Povale_MSP1         ASTHASTQAATQGNVPQASNDEHTPSATTVNPATTTPDKSLKESTSEGTLMTQGNADDDV
                    .                          .                          .  :

Pfalciparum_MSP1    KVPNPLTISTTEMEKFYENILKNNDTYFNDDIKQFVKSNSKVITGLTETQKNALNDEIKK
Pmalariae_MSP1      -DANELDF---DIDDIYIKYLEQVSKY-DENFKNFIESKKDIINKMSESEWKELGEEINT
Pknowlesi_MSP1      -EPETNIV---NVEKIYEKHLSQMDKY-NDYFIKFLESQKEKITSMTEEQANALGAEIEA
Pvivax_MSP1         -KPEADTA---QVEKFYEKHLSQIDKY-NDYFQKFLESQKDEITKMDETKWKALGAEIEE
Povale_MSP1         SEPEKKEI---EVEEFYKKYLEEVDKY-DDYFKAFLSSKKDAVNKMTEKDWKELEEEVKT
                      .:       ::::.:* : *.: ..* :: :  *:.*:.. :. : * . : *  *::

Pfalciparum_MSP1    LKDTLQLSFDLYNKYKLKLDRLFNKKKELGQDKMQIKKLTLLKEQLESKLNSLNNPHNVL
Pmalariae_MSP1      LKQDIQSSFDNFGKYKLKLERLLKKKNKITSSTNHIKEYSILKAQLLRKKNILNNPRHVL
Pknowlesi_MSP1      LKKKVQVSLDHYGKYKLKLERFLEKKNKISNSKEHIKKLTSLKNKLERKLNFLNNPTSVL
Pvivax_MSP1         LKKKLQVSLDHYGKYKLKLERLLKKKNKISNSKDQIKKLTSLKNKLERRQNLLNNPTSVL
Povale_MSP1         LKSRLDMSLDHYNKYKLKLGRLFKKNEKVLSSKEHIKELSILKAQLMRRQFMLNNPRHVI
                    **. :: *:* :.****** *:::*:::: ... :**: : ** :*  .   ****  *:

Pfalciparum_MSP1    QNFSVFFNKKKEAEIAETENTLENTKILLKHYKGLVKYYNGESSPLKTLSEVSIQTEDNY
Pmalariae_MSP1      AAFVVFFNKKIEAEKKEVENALKNTDIMLKYYKARTKYYISEAFPLKTITEQSLQKEINY
Pknowlesi_MSP1      KNYIVFFNKKKEAEKKEVENTLKNTEILLKYYKARAKYYIGEPFPLKTLSEESLQKEDNY
Pvivax_MSP1         KNYTAFFNKKRETEKKEVENTLKNTEILLKYYKARAKYYIGEPFPLKTLSEESMQKEDNY
Povale_MSP1         HNFRVFFNKKRETEKKEVENTLKNTDALLKYYKARVKYYNGETFPLKTISEDTLEKENNY
                      : .***** *:*  *.**:*:**. :**:**. .*** .*. ****::* :::.* **

Pfalciparum_MSP1    ANLEKFRVLSKIDGKLNDNLHLGKKKLSFLSSGLHHLITELKEVIKNKNYTGNSPSENNK
Pmalariae_MSP1      LHLEKFKVYSRLEGRIKKMLNLEKENITYLSGGLHHVLTELKEIINDKTYTGYTHTKNNE
Pknowlesi_MSP1      LNLEKFRVLSRMEGRLGNNINLEKENISYLSSGLHHVFTELKEIIKNKKYTGNDHAKNTT
Pvivax_MSP1         LNLEKFRVLSRLEGRLGKNIELEKENISYLSSGLHHVLTELKEIIKNKKYSGNDHTKNIA
Povale_MSP1         LNLEKFKLYSRLEGKLKQNINLEKENITYLSSALYHVLSELKGIIHNKKYTGNPHAANMV
                     :****.: *.::*.:  . : * *:::::**..*:*:::*** :*::*.*:*   : *

Pfalciparum_MSP1    KVNEALKSYENFLPEAK-----------------VTTVVTPPQ--PDVTPSPLSVRVSGSS
Pmalariae_MSP1      EVNKALNVYEELLPKQ------------------ISTEEQPDNALADGTE-----------
Pknowlesi_MSP1      AVKEALQAYEELLPKVATQTA-------------SLPPVAPPAVVPPVAPEAEAEAGAGTP
Pvivax_MSP1         AVKEALQAYQELIPKVTTQEGASTTA--------ATLPVTVPSAVPGGLP------GAGVP
Povale_MSP1         EVNNALNLYKDLLPKVETVASTGAATQTQGGEGASAAAAPPAALPAAPP-----------
                     *::**: *::::*:                                 .

Pfalciparum_MSP1    GSTKEET------------------QIPTSGSLLTELQQV---------------------
Pmalariae_MSP1      NATEGAE------------------VRAATAESLVQGE-----------------------
Pknowlesi_MSP1      TSAEPATADTADTADTAAPTQTSAAQPATADTAAQPATADTAAQPATADTAAQPAIAPAA
Pvivax_MSP1         GAAAGLT------------PPPPAGSVPATGPGAAAGSTEENVA-----------------
Povale_MSP1         -AAPPAA------------PPAASAAAPGTANGETATV-----------------------
                     ::                          . :.   .

Pfalciparum_MSP1    ---------VQLQNYDEEDDSLVVLPIFGESEDN--DEYLDQVVTGEAISVTMDNILSGF
Pmalariae_MSP1      ------------DEYPEEVDEVIVFPIVGKKEK---ENPLDQITKGQAETKQDDNILKPI
Pknowlesi_MSP1      VPENGETAEVKAQEYGEDYDKVITLPLFGNDEDDVEDQEEKQIITGEAENAQPENIVPEG
Pvivax_MSP1         ---------AKAQDYAEDYDKVIALPLFGNNDD---DGEEDQVTTGEAESEAPEILVPAG
Povale_MSP1         ---------AHAEDYTEDDNNVIVLPLFGKKG----THAFDQVTQGEAQDK-DDNILNEI
                                ::* *: :.::.:*:.*:.          .*:  *:*      : ::

Pfalciparum_MSP1    ENEYDVIYLKPLAGVYRSLKKQIEKNIFTFNLNLNDILNSRLKKRKYFLDVLESDLMQFK
Pmalariae_MSP1      TNEYEVLYIKPLAGVYRVLRKQIGDQIDAFNSNLTNALDTRKKKRTYFLDVLNSDLIQFK
Pknowlesi_MSP1      INEYEVVYIKPLAGMYKSIKKQLENHVAAFNTNITDMLDSRLKKRNYFLDVLDSELNPFK
Pvivax_MSP1         ISDYDVVYLKPLAGMYKTIKKQLENHVNAFNTNITDMLDSRLKKRNYFLEVLNSDLNPFK
Povale_MSP1         TNEYEVVYVKPLAGVYKTLKKQLEAHVTAFHSNVTNMLESRLKKRNYFLEVLNSDLTQYK
                     .:*:*:*:*****:*. :.**:  :: :*: *:.: *::* ***.***:**:*:*  :*

Pfalciparum_MSP1    HISSNEYIIEDSFKLLNSEQKNTLLKSYKYIKESVENDIKFAQEGISYYEKVLAKYKDDL
Pmalariae_MSP1      HATSDSYIIKDPYKLLDVDKKAKLIGSYKYIVSAIEKDITSAENGVEYYDKMTKLYKTQL
Pknowlesi_MSP1      YSSSGEYIIKDPYKLLDLEQKKKLLGSYQYIGASVDKDLITAKDGMEYYKKMGELYKKHL
Pvivax_MSP1         YSSSGEYIIKDPYKLLDLEKKKKLIGSYKYIGASIDMDLATANDGVTYYNKMGELYKTHL
Povale_MSP1         HATSDNYVIRDAYKLLDFEKKKKLLSSYKYIKDSVEKDVEIATDGIDYYEKMAALYKTYL
                    : :*..*:* *.:***: ::* .*: **:**  ::: *:  *  *:  *:*: **.*:   ** *

Pfalciparum_MSP1    ESIKKVIKEEKEKFPSSPPTTP----------PSPAKTDEQKKESKFLPFLTNIETLYNN
```

Figure 2 Amino acid sequences alignments for MSP1 gene sequence between P. knowlesi, P. falciparum, P. vivax, P. malariae and P. ovale/P. simiovale. Sequence for PkMSP1-19 is highlighted in yellow. Asterisks indicate fully conserved residues, colons indicates strong residue conservation (>0.5, Gonnet PAM 250 matrix), period indicates weak residue conservation (=<0.5, Gonnet PAM 250 matrix). Blank spaces indicate no residue conservation.


Pfalciparum_SERA3
Pvalariae_SERA3like
Pknowlesi_SERA3like
Pvivax_SERA3like
Pmalariae_SERA3like

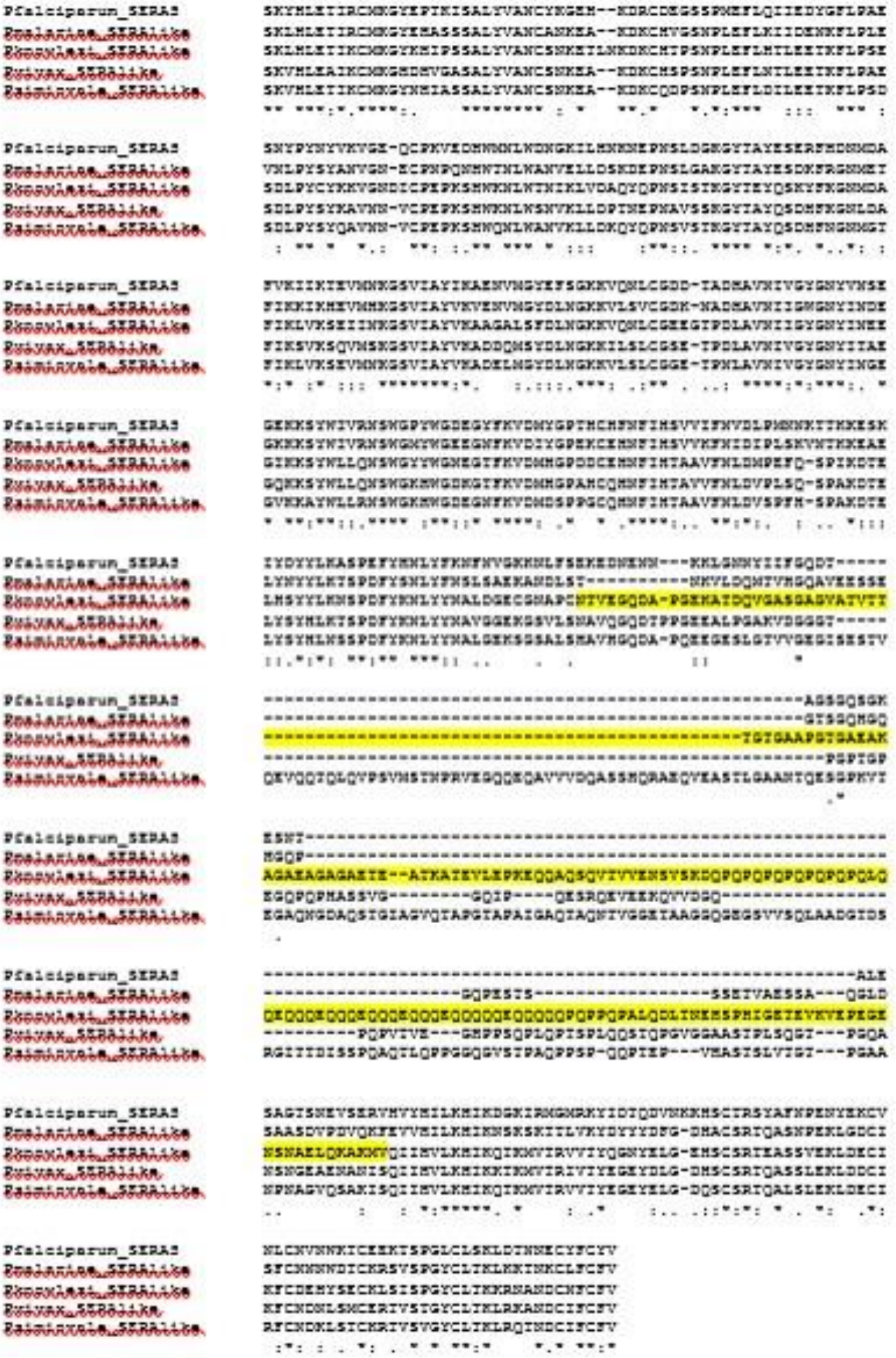


Figure 3 Amino acid sequences alignments for SERA3 gene sequence between P. knowlesi, P. falciparum, P. vivax, P. malariae and P. ovale/P. simiovale. P. knowlesi-specific sequences selected for development as constructs are highlighted in yellow (SERA3 fragment 1 and fragment 2). Asterisks indicate fully conserved residues, colons indicates strong residue conservation (>0.5, Gonnet PAM 250 matrix), period indicates weak residue conservation (=<0.5, Gonnet PAM 250 matrix). Blank spaces indicate no residue conservation.


Pfalciparum_SSP2
Pknowlesi_SSP2
Pvivax_SSP2
Pmalariae_SSP2
PovaleWallikeri_SSP2
PovaleCurtisi_SSP2

Figure 4 Amino acid sequences alignments for SSP2/TRAP gene sequence between P. knowlesi, P. falciparum, P. vivax, P. malariae and P. ovale/P. simiovale. P. knowlesi-specific sequence selected for development as constructs is highlighted in yellow. Asterisks indicate fully conserved residues, colons indicates strong residue conservation (>0.5, Gonnet PAM 250 matrix), period indicates weak residue conservation (=<0.5, Gonnet PAM 250 matrixBlank spaces indicate no residue conservation.


Pknowlesi_PKNH_0413500_TSERA2
Pvivax_PVP01_0417700_SERA
PovaleCurtisi_PocGH01_04022100_S
Pfalciparum_PF3D7_0207600_SERA3
Pmalariae_PmUG01_04024900_SERA1i

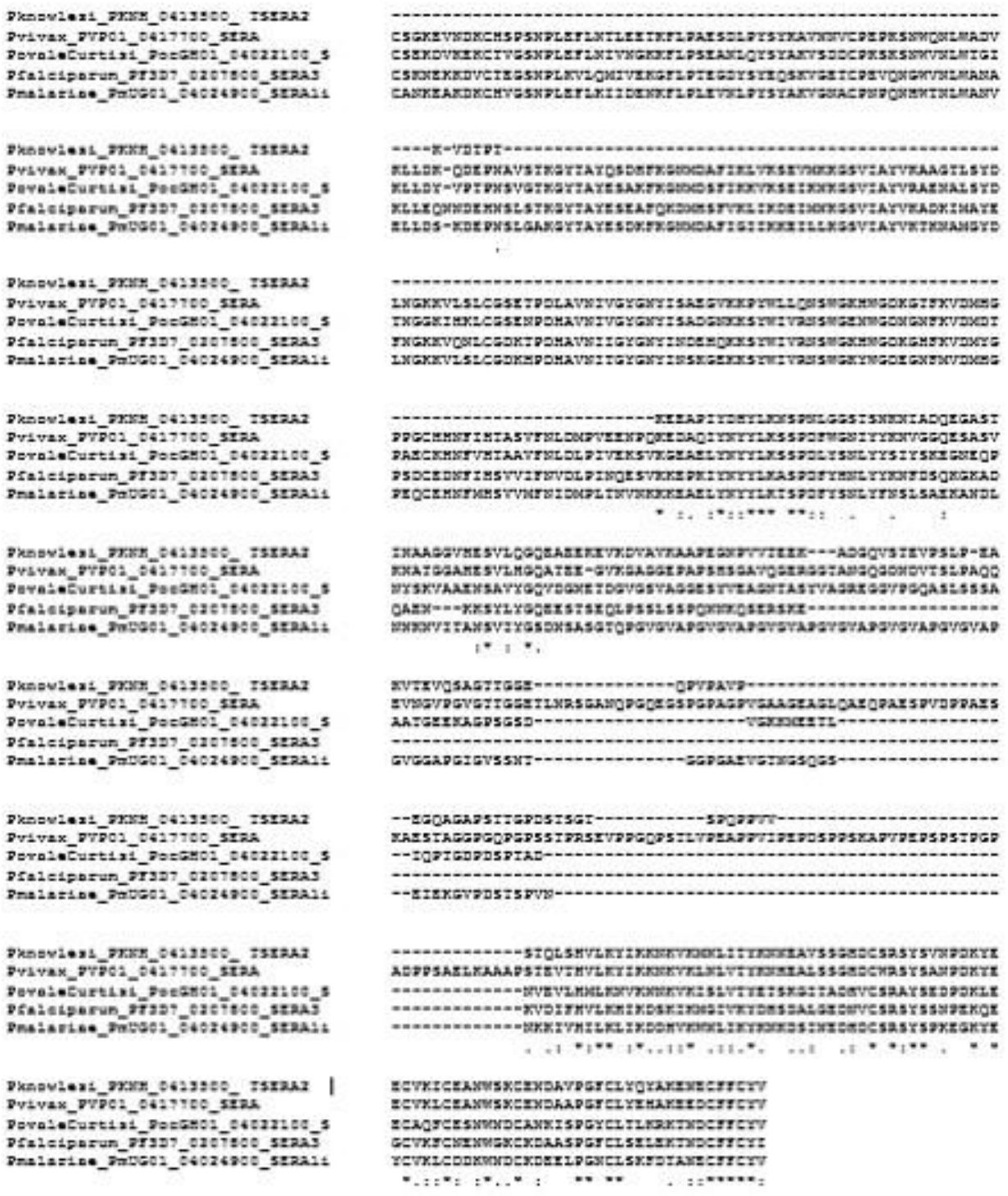


Figure 5 Amino acid sequences alignments for TSERA2 gene sequence between P. knowlesi, P. falciparum, P. vivax, P. malariae and P. ovale/P. simiovale. P. knowlesi-specific sequences selected for development as constructs (TSERA2 fragment 1 and 2) are highlighted in yellow. Asterisks indicate fully conserved residues, colons indicates strong residue conservation (>0.5, Gonnet PAM 250 matrix), period indicates weak residue conservation (=<0.5, Gonnet PAM 250 matrix). Blank spaces indicate no residue conservation.

Table 1 Single-nucleotide polymorphism frequencies of Malaysian clinical isolates sequences within P. knowlesi candidate genes.

Gene	Chromosme	Genome Position	Reference Allele	Gene Position	Number Samples	Cluster 1	Cluster 2	Cluster 3	Type
PKH_08	ordered_PKN	1366				cluster	cluster	cluster3:	synonymous
0030	H_08_v2	9	A	186	19	1:3/8	2:0/5	16/32	_variant
PKH_08	ordered_PKN	1371				cluster	cluster	cluster3:	synonymous
0030	H_08_v2	4	C	231	1	1:0/8	2:1/5	0/32	_variant
PKH_03	ordered_PKN	2116				cluster	cluster	cluster3:	synonymous
1240	H_03_v2	8	G	819	40	1:8/8	2:0/5	32/32	_variant
PKH_03	ordered_PKN	2117				cluster	cluster	cluster3:	missense_va
1240	H_03_v2	7	A	810	2	1:0/8	2:0/5	2/32	riant
PKH_03	ordered_PKN	2118				cluster	cluster	cluster3:	missense_va
1240	H_03_v2	5	G	802	1	1:0/8	2:0/5	1/32	riant
PKH_03	ordered_PKN	2119				cluster	cluster	cluster3:	missense_va
1240	H_03_v2	0	A	797	1	1:0/8	2:1/5	0/32	riant
PKH_03	ordered_PKN	2120				cluster	cluster	cluster3:	synonymous
1240	H_03_v2	1	A	786	39	1:7/8	2:0/5	32/32	_variant
PKH_03	ordered_PKN	2121				cluster	cluster	cluster3:	synonymous
1240	H_03_v2	9	A	768	4	1:3/8	2:0/5	1/32	_variant
PKH_03	ordered_PKN	2127				cluster	cluster	cluster3:	synonymous
1240	H_03_v2	3	T	714	10	1:2/8	2:0/5	8/32	_variant
PKH_03	ordered_PKN	2128				cluster	cluster	cluster3:	synonymous
1240	H_03_v2	8	G	699	27	1:3/8	2:0/5	24/32	_variant
PKH_03	ordered_PKN	2132				cluster	cluster	cluster3:	synonymous
1240	H_03_v2	1	G	666	2	1:0/8	2:0/5	2/32	_variant
PKH_03	ordered_PKN	2179				cluster	cluster	cluster3:	missense_va
1240	H_03_v2	3	T	194	14	1:0/8	2:0/5	14/32	riant
PKH_03	ordered_PKN	2232				cluster	cluster	cluster3:	missense_va
1240	H_03_v2	2	A	67	1	1:0/8	2:0/5	1/32	riant
PKH_03	ordered_PKN	2234				cluster	cluster	cluster3:	missense_va
1240	H_03_v2	1	C	48	3	1:3/8	2:0/5	0/32	riant
PKH_03	ordered_PKN	2234				cluster	cluster	cluster3:	missense_va
1240	H_03_v2	8	T	41	40	1:8/8	2:0/5	32/32	riant
PKH_14	ordered_PKN	8715				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	4	C	269	2	1:2/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	8726				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	6	A	381	26	1:0/8	2:0/5	26/32	_variant
PKH_14	ordered_PKN	8729				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	3	T	408	1	1:0/8	2:0/5	1/32	_variant
PKH_14	ordered_PKN	8733				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	2	A	447	32	1:0/8	2:0/5	32/32	_variant
PKH_14	ordered_PKN	8733				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	4	C	449	2	1:2/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	8735				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	6	A	471	8	1:8/8	2:0/5	0/32	_variant
PKH_14	ordered_PKN	8739				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	3	C	508	8	1:8/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	8739				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	9	A	514	32	1:0/8	2:0/5	32/32	riant
PKH_14	ordered_PKN	8740				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	8	A	523	32	1:0/8	2:0/5	32/32	riant
PKH_14	ordered_PKN	8741				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	9	C	534	8	1:8/8	2:0/5	0/32	_variant
PKH_14	ordered_PKN	8742				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	0	A	535	40	1:8/8	2:0/5	32/32	riant
PKH_14	ordered_PKN	8743				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	6	C	551	40	1:8/8	2:0/5	32/32	riant

PKH_14	ordered_PKN	8744				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	3	A	558	8	1:8/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	8745				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	7	T	572	40	1:8/8	2:0/5	32/32	riant
PKH_14	ordered_PKN	8746				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	4	T	579	32	1:0/8	2:0/5	32/32	riant
PKH_14	ordered_PKN	8747				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	1	A	586	32	1:0/8	2:0/5	32/32	riant
PKH_14	ordered_PKN	8747				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	3	A	588	32	1:0/8	2:0/5	32/32	riant
PKH_14	ordered_PKN	8748				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	2	C	597	1	1:0/8	2:0/5	1/32	_variant
PKH_14	ordered_PKN	8748				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	5	G	600	40	1:8/8	2:0/5	32/32	_variant
PKH_14	ordered_PKN	8748				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	8	A	603	32	1:0/8	2:0/5	32/32	_variant
PKH_14	ordered_PKN	8749				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	5	A	610	37	1:5/8	2:0/5	32/32	riant
PKH_14	ordered_PKN	8751				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	2	A	627	35	1:3/8	2:0/5	32/32	_variant
PKH_14	ordered_PKN	8754				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	1	G	656	8	1:8/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	8755				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	2	T	667	8	1:8/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	8756				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	6	A	681	8	1:8/8	2:0/5	0/32	_variant
PKH_14	ordered_PKN	8757				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	9	A	694	8	1:8/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	8758				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	8	A	703	8	1:8/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	8763				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	8	T	753	43	1:8/8	2:3/5	32/32	_variant
PKH_14	ordered_PKN	8764				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	4	C	759	10	1:8/8	2:2/5	0/32	riant
PKH_14	ordered_PKN	8767				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	1	G	786	1	1:0/8	2:0/5	1/32	_variant
PKH_14	ordered_PKN	8767				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	9	A	794	31	1:0/8	2:0/5	31/32	riant
PKH_14	ordered_PKN	8768				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	5	A	800	8	1:8/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	8769				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	9	A	814	1	1:0/8	2:0/5	1/32	riant
PKH_14	ordered_PKN	8775				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	7	G	872	4	1:0/8	2:0/5	4/32	riant
PKH_14	ordered_PKN	8776				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	7	A	882	26	1:0/8	2:0/5	26/32	_variant
PKH_14	ordered_PKN	8777				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	0	A	885	10	1:8/8	2:2/5	0/32	riant
PKH_14	ordered_PKN	8777				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	3	T	888	32	1:0/8	2:0/5	32/32	_variant
PKH_14	ordered_PKN	8781				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	1	G	926	2	1:0/8	2:0/5	2/32	riant
PKH_14	ordered_PKN	8781				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	3	T	928	10	1:8/8	2:2/5	0/32	riant
PKH_14	ordered_PKN	8782				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	3	G	938	10	1:8/8	2:2/5	0/32	riant
PKH_14	ordered_PKN	8784				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	4	A	959	10	1:8/8	2:2/5	0/32	riant
PKH_14	ordered_PKN	8785				cluster	cluster	cluster3:	missense_va
0010	H_14_v2	0	A	965	8	1:8/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	8785				cluster	cluster	cluster3:	synonymous
0010	H_14_v2	4	T	969	8	1:8/8	2:0/5	0/32	_variant

PKH_04	ordered_PKN	5908				cluster	cluster	cluster3:	
1230	H_04_v2	52	C	36	1	1:0/8	2:0/5	1/32	NA
PKH_04	ordered_PKN	5908				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	56	A	40	21	1:1/8	2:0/5	20/32	riant
PKH_04	ordered_PKN	5911				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	98	G	382	5	1:0/8	2:0/5	5/32	riant
PKH_04	ordered_PKN	5912				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	01	G	385	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5912				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	09	G	393	1	1:1/8	2:0/5	0/32	_variant
PKH_04	ordered_PKN	5912				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	10	C	394	13	1:5/8	2:0/5	8/32	riant
PKH_04	ordered_PKN	5912				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	28	C	412	4	1:0/8	2:2/5	2/32	riant
PKH_04	ordered_PKN	5912				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	51	C	435	1	1:0/8	2:1/5	0/32	_variant
PKH_04	ordered_PKN	5912				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	69	G	453	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5912				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	84	G	468	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5913				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	11	G	495	2	1:0/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5913				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	29	A	513	39	1:8/8	2:0/5	31/32	_variant
PKH_04	ordered_PKN	5913				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	38	C	522	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5913				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	41	G	525	1	1:0/8	2:1/5	0/32	_variant
PKH_04	ordered_PKN	5913				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	47	C	531	3	1:0/8	2:0/5	3/32	_variant
PKH_04	ordered_PKN	5913				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	51	G	535	9	1:0/8	2:0/5	9/32	riant
PKH_04	ordered_PKN	5913				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	54	G	538	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5913				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	66	T	550	42	1:8/8	2:2/5	32/32	_variant
PKH_04	ordered_PKN	5913				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	89	G	573	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	16	T	600	24	1:0/8	2:0/5	24/32	riant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	23	C	607	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	26	G	610	27	1:8/8	2:1/5	18/32	riant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	33	A	617	29	1:8/8	2:1/5	20/32	riant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	40	G	624	10	1:0/8	2:0/5	10/32	_variant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	44	A	628	42	1:8/8	2:3/5	31/32	riant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	47	G	631	2	1:0/8	2:2/5	0/32	riant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	53	C	637	9	1:0/8	2:1/5	8/32	riant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	56	A	640	9	1:0/8	2:1/5	8/32	riant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	64	G	648	8	1:0/8	2:0/5	8/32	_variant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	83	C	667	29	1:8/8	2:2/5	19/32	_variant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	85	G	669	1	1:0/8	2:0/5	1/32	_variant

PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	91	T	675	29	1:8/8	2:2/5	19/32	_variant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	94	C	678	6	1:0/8	2:0/5	6/32	_variant
PKH_04	ordered_PKN	5914				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	98	T	682	8	1:0/8	2:2/5	6/32	_variant
PKH_04	ordered_PKN	5915				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	00	A	684	16	1:8/8	2:2/5	6/32	_variant
PKH_04	ordered_PKN	5915				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	22	C	706	1	1:1/8	2:0/5	0/32	riant
PKH_04	ordered_PKN	5915				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	32	G	716	25	1:8/8	2:2/5	15/32	riant
PKH_04	ordered_PKN	5918				cluster	cluster	cluster3:	
1230	H_04_v2	05	A	726	23	1:1/8	2:2/5	20/32	NA
PKH_04	ordered_PKN	5918				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	12	G	733	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5918				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	15	C	736	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5918				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	41	C	762	5	1:0/8	2:1/5	4/32	_variant
PKH_04	ordered_PKN	5918				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	98	A	819	42	1:8/8	2:3/5	31/32	riant
PKH_04	ordered_PKN	5919				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	07	A	828	43	1:8/8	2:3/5	32/32	_variant
PKH_04	ordered_PKN	5919				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	13	A	834	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5919				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	43	T	864	2	1:0/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5919				cluster	cluster	cluster3:	
1230	H_04_v2	46	G	867	15	1:0/8	2:2/5	13/32	NA
PKH_04	ordered_PKN	5921				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	71	C	888	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5922				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	13	C	930	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5922				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	43	A	960	2	1:0/8	2:1/5	1/32	_variant
PKH_04	ordered_PKN	5922				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	61	T	978	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5922				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	64	C	981	1	1:1/8	2:0/5	0/32	riant
PKH_04	ordered_PKN	5922				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	70	C	987	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5922				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	92	G	1009	9	1:0/8	2:0/5	9/32	riant
PKH_04	ordered_PKN	5923				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	18	C	1035	6	1:0/8	2:2/5	4/32	_variant
PKH_04	ordered_PKN	5923				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	24	C	1041	1	1:0/8	2:1/5	0/32	_variant
PKH_04	ordered_PKN	5923				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	38	G	1055	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5923				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	39	C	1056	3	1:0/8	2:1/5	2/32	_variant
PKH_04	ordered_PKN	5923				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	64	A	1081	41	1:7/8	2:2/5	32/32	riant
PKH_04	ordered_PKN	5923				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	77	A	1094	41	1:7/8	2:2/5	32/32	riant
PKH_04	ordered_PKN	5923				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	81	A	1098	1	1:1/8	2:0/5	0/32	_variant
PKH_04	ordered_PKN	5923				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	93	A	1110	2	1:2/8	2:0/5	0/32	_variant
PKH_04	ordered_PKN	5923				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	95	G	1112	8	1:0/8	2:2/5	6/32	riant

PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	20	T	1137	5	1:0/8	2:2/5	3/32	_variant
PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	34	T	1151	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	36	G	1153	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	39	G	1156	19	1:0/8	2:0/5	19/32	riant
PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	61	G	1178	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	64	A	1181	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	71	C	1188	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	84	A	1201	3	1:0/8	2:0/5	3/32	riant
PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	85	A	1202	9	1:0/8	2:1/5	8/32	riant
PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	92	T	1209	21	1:7/8	2:3/5	11/32	_variant
PKH_04	ordered_PKN	5924				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	93	A	1210	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	07	G	1224	43	1:8/8	2:3/5	32/32	_variant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	08	G	1225	43	1:8/8	2:3/5	32/32	riant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	14	C	1231	34	1:8/8	2:2/5	24/32	riant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	19	A	1236	37	1:8/8	2:3/5	26/32	_variant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	22	G	1239	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	27	C	1244	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	37	T	1254	36	1:8/8	2:3/5	25/32	_variant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	69	A	1286	9	1:4/8	2:2/5	3/32	riant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	74	C	1291	22	1:5/8	2:3/5	14/32	riant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	77	G	1294	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	87	A	1304	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	94	C	1311	31	1:8/8	2:3/5	20/32	_variant
PKH_04	ordered_PKN	5925				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	97	G	1314	12	1:1/8	2:1/5	10/32	_variant
PKH_04	ordered_PKN	5926				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	21	C	1338	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5926				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	42	T	1359	28	1:8/8	2:2/5	18/32	_variant
PKH_04	ordered_PKN	5926				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	52	A	1369	39	1:8/8	2:2/5	29/32	riant
PKH_04	ordered_PKN	5926				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	69	G	1386	40	1:8/8	2:2/5	30/32	_variant
PKH_04	ordered_PKN	5926				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	70	C	1387	7	1:5/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5927				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	14	C	1431	3	1:0/8	2:1/5	2/32	_variant
PKH_04	ordered_PKN	5927				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	20	C	1437	34	1:8/8	2:2/5	24/32	_variant

PKH_04	ordered_PKN	5927				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	26	A	1443	33	1:8/8	2:2/5	23/32	_variant
PKH_04	ordered_PKN	5927				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	32	A	1449	41	1:8/8	2:2/5	31/32	_variant
PKH_04	ordered_PKN	5927				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	35	T	1452	33	1:8/8	2:2/5	23/32	riant
PKH_04	ordered_PKN	5927				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	38	G	1455	33	1:8/8	2:2/5	23/32	_variant
PKH_04	ordered_PKN	5927				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	59	C	1476	30	1:8/8	2:2/5	20/32	_variant
PKH_04	ordered_PKN	5927				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	63	C	1480	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5927				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	71	G	1488	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5928				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	12	C	1529	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5928				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	18	C	1535	4	1:0/8	2:0/5	4/32	riant
PKH_04	ordered_PKN	5928				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	20	T	1537	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5928				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	25	C	1542	2	1:0/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5929				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	06	A	1623	41	1:8/8	2:1/5	32/32	riant
PKH_04	ordered_PKN	5929				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	09	G	1626	15	1:1/8	2:0/5	14/32	_variant
PKH_04	ordered_PKN	5930				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	35	C	1752	7	1:0/8	2:0/5	7/32	_variant
PKH_04	ordered_PKN	5930				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	71	A	1788	2	1:0/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5931				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	26	T	1843	9	1:0/8	2:0/5	9/32	riant
PKH_04	ordered_PKN	5931				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	55	G	1872	40	1:8/8	2:1/5	31/32	_variant
PKH_04	ordered_PKN	5932				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	63	A	1980	5	1:4/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5933				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	32	A	2049	4	1:2/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5933				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	84	C	2101	3	1:0/8	2:0/5	3/32	_variant
PKH_04	ordered_PKN	5935				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	36	C	2253	2	1:0/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5935				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	69	T	2286	9	1:0/8	2:2/5	7/32	_variant
PKH_04	ordered_PKN	5936				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	11	C	2328	20	1:6/8	2:0/5	14/32	_variant
PKH_04	ordered_PKN	5936				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	14	G	2331	26	1:6/8	2:0/5	20/32	_variant
PKH_04	ordered_PKN	5936				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	41	A	2358	34	1:7/8	2:1/5	26/32	riant
PKH_04	ordered_PKN	5936				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	47	A	2364	25	1:1/8	2:1/5	23/32	_variant
PKH_04	ordered_PKN	5936				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	49	C	2366	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5936				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	50	G	2367	1	1:0/8	2:1/5	0/32	_variant
PKH_04	ordered_PKN	5936				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	56	C	2373	3	1:0/8	2:1/5	2/32	riant
PKH_04	ordered_PKN	5936				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	77	C	2394	3	1:0/8	2:1/5	2/32	_variant
PKH_04	ordered_PKN	5936				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	86	G	2403	3	1:0/8	2:0/5	3/32	_variant

PKH_04	ordered_PKN	5937				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	01	T	2418	7	1:0/8	2:1/5	6/32	_variant
PKH_04	ordered_PKN	5937				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	07	C	2424	29	1:7/8	2:0/5	22/32	_variant
PKH_04	ordered_PKN	5937				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	14	T	2431	12	1:0/8	2:1/5	11/32	_variant
PKH_04	ordered_PKN	5937				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	28	A	2445	4	1:0/8	2:0/5	4/32	_variant
PKH_04	ordered_PKN	5937				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	34	T	2451	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5937				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	36	G	2453	3	1:0/8	2:0/5	3/32	riant
PKH_04	ordered_PKN	5937				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	44	G	2461	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5942				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	40	A	2957	12	1:2/8	2:2/5	8/32	riant
PKH_04	ordered_PKN	5942				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	59	G	2976	2	1:0/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5942				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	77	C	2994	3	1:0/8	2:0/5	3/32	_variant
PKH_04	ordered_PKN	5942				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	83	T	3000	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5942				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	89	C	3006	27	1:0/8	2:1/5	26/32	_variant
PKH_04	ordered_PKN	5942				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	92	T	3009	28	1:1/8	2:1/5	26/32	_variant
PKH_04	ordered_PKN	5942				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	93	T	3010	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5943				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	25	G	3042	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5943				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	37	G	3054	2	1:0/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5943				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	41	C	3058	10	1:8/8	2:2/5	0/32	riant
PKH_04	ordered_PKN	5943				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	70	C	3087	1	1:0/8	2:1/5	0/32	_variant
PKH_04	ordered_PKN	5944				cluster	cluster	cluster3:	missense_va
1230	H_04_v2	34	C	3151	2	1:0/8	2:2/5	0/32	riant
PKH_04	ordered_PKN	5944				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	72	C	3189	5	1:0/8	2:1/5	4/32	_variant
PKH_04	ordered_PKN	5944				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	81	G	3198	43	1:8/8	2:3/5	32/32	_variant
PKH_04	ordered_PKN	5945				cluster	cluster	cluster3:	synonymous
1230	H_04_v2	11	C	3228	3	1:0/8	2:2/5	1/32	_variant
PKH_04	ordered_PKN	5956				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	18	C	13	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5956				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	27	A	22	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5958				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	68	C	39	24	1:7/8	2:2/5	15/32	_variant
PKH_04	ordered_PKN	5958				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	69	G	40	18	1:1/8	2:0/5	17/32	riant
PKH_04	ordered_PKN	5958				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	77	G	48	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5958				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	82	G	53	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5958				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	92	A	63	17	1:5/8	2:0/5	12/32	_variant
PKH_04	ordered_PKN	5958				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	98	G	69	1	1:0/8	2:1/5	0/32	_variant
PKH_04	ordered_PKN	5959				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	05	G	76	39	1:7/8	2:2/5	30/32	riant

PKH_04	ordered_PKN	5959				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	12	A	83	22	1:5/8	2:1/5	16/32	riant
PKH_04	ordered_PKN	5959				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	22	G	93	29	1:5/8	2:2/5	22/32	_variant
PKH_04	ordered_PKN	5959				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	25	T	96	22	1:5/8	2:1/5	16/32	riant
PKH_04	ordered_PKN	5959				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	63	C	134	5	1:0/8	2:0/5	5/32	riant
PKH_04	ordered_PKN	5959				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	68	G	139	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5960				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	54	A	225	11	1:0/8	2:0/5	11/32	riant
PKH_04	ordered_PKN	5960				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	76	G	247	12	1:2/8	2:1/5	9/32	riant
PKH_04	ordered_PKN	5960				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	78	T	249	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5960				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	83	A	254	13	1:2/8	2:1/5	10/32	riant
PKH_04	ordered_PKN	5960				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	97	G	268	25	1:2/8	2:0/5	23/32	riant
PKH_04	ordered_PKN	5960				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	98	C	269	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	09	G	280	27	1:3/8	2:1/5	23/32	riant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	11	C	282	9	1:0/8	2:0/5	9/32	_variant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	22	A	293	33	1:3/8	2:1/5	29/32	riant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	31	C	302	33	1:8/8	2:1/5	24/32	riant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	33	A	304	43	1:8/8	2:3/5	32/32	riant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	35	T	306	42	1:8/8	2:2/5	32/32	_variant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	47	C	318	1	1:1/8	2:0/5	0/32	_variant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	49	G	320	9	1:0/8	2:1/5	8/32	riant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	59	G	330	34	1:8/8	2:2/5	24/32	_variant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	65	A	336	29	1:8/8	2:2/5	19/32	_variant
PKH_04	ordered_PKN	5961				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	89	A	360	24	1:8/8	2:1/5	15/32	_variant
PKH_04	ordered_PKN	5962				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	07	G	378	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5962				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	58	T	429	29	1:7/8	2:0/5	22/32	_variant
PKH_04	ordered_PKN	5962				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	70	C	441	6	1:0/8	2:0/5	6/32	_variant
PKH_04	ordered_PKN	5962				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	73	G	444	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5962				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	76	C	447	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5962				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	79	T	450	13	1:2/8	2:0/5	11/32	_variant
PKH_04	ordered_PKN	5963				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	08	A	479	3	1:0/8	2:3/5	0/32	riant
PKH_04	ordered_PKN	5963				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	43	A	514	14	1:2/8	2:3/5	9/32	riant
PKH_04	ordered_PKN	5963				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	47	A	518	1	1:0/8	2:0/5	1/32	riant

PKH_04	ordered_PKN	5963				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	50	A	521	14	1:2/8	2:3/5	9/32	riant
PKH_04	ordered_PKN	5963				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	52	A	523	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5963				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	59	A	530	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5963				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	64	C	535	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5963				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	93	G	564	3	1:0/8	2:0/5	3/32	_variant
PKH_04	ordered_PKN	5964				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	11	C	582	12	1:4/8	2:0/5	8/32	_variant
PKH_04	ordered_PKN	5964				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	38	C	609	4	1:0/8	2:0/5	4/32	_variant
PKH_04	ordered_PKN	5964				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	50	G	621	19	1:6/8	2:0/5	13/32	_variant
PKH_04	ordered_PKN	5967				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	01	G	657	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5967				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	25	T	681	4	1:3/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5967				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	40	G	696	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5967				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	61	A	717	43	1:8/8	2:3/5	32/32	_variant
PKH_04	ordered_PKN	5967				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	76	C	732	39	1:8/8	2:0/5	31/32	_variant
PKH_04	ordered_PKN	5969				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	97	C	801	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5970				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	00	C	804	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5970				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	24	A	828	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5970				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	29	A	833	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5970				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	39	A	843	2	1:0/8	2:2/5	0/32	_variant
PKH_04	ordered_PKN	5970				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	48	T	852	5	1:0/8	2:2/5	3/32	_variant
PKH_04	ordered_PKN	5970				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	72	C	876	32	1:3/8	2:2/5	27/32	_variant
PKH_04	ordered_PKN	5970				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	86	A	890	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5970				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	90	C	894	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5971				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	53	C	957	11	1:6/8	2:0/5	5/32	_variant
PKH_04	ordered_PKN	5971				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	57	C	961	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5971				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	62	T	966	4	1:4/8	2:0/5	0/32	_variant
PKH_04	ordered_PKN	5971				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	71	C	975	5	1:0/8	2:3/5	2/32	_variant
PKH_04	ordered_PKN	5971				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	73	G	977	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5971				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	74	C	978	4	1:0/8	2:0/5	4/32	_variant
PKH_04	ordered_PKN	5971				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	90	A	994	43	1:8/8	2:3/5	32/32	riant
PKH_04	ordered_PKN	5971				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	92	C	996	3	1:0/8	2:3/5	0/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	05	G	1009	17	1:2/8	2:0/5	15/32	riant

PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	06	G	1010	2	1:1/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	31	T	1035	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	38	C	1042	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	40	C	1044	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	46	C	1050	10	1:0/8	2:0/5	10/32	_variant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	47	G	1051	3	1:3/8	2:0/5	0/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	52	A	1056	22	1:3/8	2:0/5	19/32	_variant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	55	G	1059	5	1:0/8	2:1/5	4/32	_variant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	63	A	1067	9	1:3/8	2:0/5	6/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	65	G	1069	8	1:2/8	2:0/5	6/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	71	G	1075	6	1:0/8	2:0/5	6/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	84	A	1088	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	90	T	1094	25	1:5/8	2:1/5	19/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	93	C	1097	10	1:3/8	2:1/5	6/32	riant
PKH_04	ordered_PKN	5972				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	97	G	1101	3	1:0/8	2:0/5	3/32	_variant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	01	G	1105	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	04	G	1108	15	1:2/8	2:2/5	11/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	05	C	1109	11	1:4/8	2:1/5	6/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	13	G	1117	30	1:8/8	2:3/5	19/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	19	C	1123	20	1:2/8	2:2/5	16/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	23	T	1127	8	1:0/8	2:0/5	8/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	29	C	1133	3	1:0/8	2:0/5	3/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	40	G	1144	34	1:8/8	2:3/5	23/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	42	A	1146	3	1:0/8	2:0/5	3/32	_variant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	43	G	1147	15	1:0/8	2:2/5	13/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	47	G	1151	3	1:0/8	2:0/5	3/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	52	G	1156	5	1:0/8	2:2/5	3/32	riant
PKH_04	ordered_PKN	5973				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	91	A	1195	1	1:0/8	2:1/5	0/32	riant
PKH_04	ordered_PKN	5974				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	01	A	1205	27	1:8/8	2:2/5	17/32	riant
PKH_04	ordered_PKN	5974				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	06	G	1210	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5974				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	19	G	1223	32	1:8/8	2:1/5	23/32	riant

PKH_04	ordered_PKN	5974				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	53	A	1257	2	1:0/8	2:1/5	1/32	_variant
PKH_04	ordered_PKN	5974				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	59	G	1263	9	1:4/8	2:0/5	5/32	_variant
PKH_04	ordered_PKN	5974				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	66	C	1270	13	1:6/8	2:1/5	6/32	riant
PKH_04	ordered_PKN	5974				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	68	G	1272	1	1:1/8	2:0/5	0/32	_variant
PKH_04	ordered_PKN	5974				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	87	A	1291	15	1:7/8	2:1/5	7/32	riant
PKH_04	ordered_PKN	5974				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	94	G	1298	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5975				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	05	C	1309	2	1:0/8	2:0/5	2/32	riant
PKH_04	ordered_PKN	5975				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	07	C	1311	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5975				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	11	C	1315	14	1:7/8	2:0/5	7/32	riant
PKH_04	ordered_PKN	5975				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	15	C	1319	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5975				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	20	G	1324	43	1:8/8	2:3/5	32/32	riant
PKH_04	ordered_PKN	5975				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	24	G	1328	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5975				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	82	G	1386	1	1:1/8	2:0/5	0/32	_variant
PKH_04	ordered_PKN	5975				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	85	T	1389	6	1:0/8	2:3/5	3/32	_variant
PKH_04	ordered_PKN	5976				cluster	cluster	cluster3:	missense_va
1240	H_04_v2	04	G	1408	1	1:0/8	2:0/5	1/32	riant
PKH_04	ordered_PKN	5976				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	09	C	1413	1	1:0/8	2:0/5	1/32	_variant
PKH_04	ordered_PKN	5977				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	29	T	1533	2	1:2/8	2:0/5	0/32	_variant
PKH_04	ordered_PKN	5977				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	35	C	1539	2	1:0/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5977				cluster	cluster	cluster3:	synonymous
1240	H_04_v2	56	G	1560	2	1:0/8	2:0/5	2/32	_variant
PKH_04	ordered_PKN	5977				cluster	cluster	cluster3:	
1240	H_04_v2	76	A	1580	3	1:3/8	2:0/5	0/32	NA
PKH_02	ordered_PKN	7634				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	08	A	623	34	1:5/8	2:1/5	28/32	riant
PKH_02	ordered_PKN	7634				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	19	A	612	42	1:8/8	2:2/5	32/32	riant
PKH_02	ordered_PKN	7634				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	23	A	608	42	1:8/8	2:2/5	32/32	riant
PKH_02	ordered_PKN	7634				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	24	C	607	42	1:8/8	2:2/5	32/32	riant
PKH_02	ordered_PKN	7634				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	25	A	606	42	1:8/8	2:2/5	32/32	_variant
PKH_02	ordered_PKN	7634				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	68	A	563	2	1:1/8	2:0/5	1/32	riant
PKH_02	ordered_PKN	7634				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	98	A	533	1	1:0/8	2:0/5	1/32	riant
PKH_02	ordered_PKN	7635				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	00	A	531	1	1:0/8	2:1/5	0/32	_variant
PKH_02	ordered_PKN	7635				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	05	G	526	1	1:0/8	2:0/5	1/32	riant
PKH_02	ordered_PKN	7635				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	08	G	523	3	1:0/8	2:0/5	3/32	riant
PKH_02	ordered_PKN	7635				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	10	C	521	1	1:0/8	2:1/5	0/32	riant

PKH_02	ordered_PKN	7635				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	24	C	507	7	1:0/8	2:0/5	7/32	_variant
PKH_02	ordered_PKN	7635				cluster	cluster	cluster3:	
1580	H_02_v2	38	C	493	18	1:2/8	2:0/5	16/32	stop_gained
PKH_02	ordered_PKN	7635				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	52	G	479	41	1:8/8	2:1/5	32/32	riant
PKH_02	ordered_PKN	7635				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	71	G	460	1	1:1/8	2:0/5	0/32	riant
PKH_02	ordered_PKN	7635				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	78	A	453	1	1:0/8	2:1/5	0/32	_variant
PKH_02	ordered_PKN	7635				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	92	A	439	41	1:8/8	2:1/5	32/32	riant
PKH_02	ordered_PKN	7636				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	14	A	417	1	1:0/8	2:0/5	1/32	_variant
PKH_02	ordered_PKN	7636				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	72	G	359	1	1:0/8	2:1/5	0/32	riant
PKH_02	ordered_PKN	7636				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	80	A	351	4	1:4/8	2:0/5	0/32	_variant
PKH_02	ordered_PKN	7637				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	16	A	315	41	1:8/8	2:1/5	32/32	_variant
PKH_02	ordered_PKN	7637				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	43	T	288	1	1:0/8	2:1/5	0/32	_variant
PKH_02	ordered_PKN	7637				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	53	G	278	5	1:0/8	2:0/5	5/32	riant
PKH_02	ordered_PKN	7637				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	55	C	276	2	1:0/8	2:0/5	2/32	_variant
PKH_02	ordered_PKN	7637				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	61	T	270	1	1:0/8	2:0/5	1/32	riant
PKH_02	ordered_PKN	7637				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	88	G	243	3	1:0/8	2:0/5	3/32	_variant
PKH_02	ordered_PKN	7637				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	98	C	233	28	1:2/8	2:0/5	26/32	riant
PKH_02	ordered_PKN	7638				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	01	A	230	28	1:2/8	2:0/5	26/32	riant
PKH_02	ordered_PKN	7638				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	02	G	229	4	1:4/8	2:0/5	0/32	riant
PKH_02	ordered_PKN	7638				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	04	G	227	1	1:0/8	2:0/5	1/32	riant
PKH_02	ordered_PKN	7638				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	27	T	204	22	1:2/8	2:0/5	20/32	_variant
PKH_02	ordered_PKN	7638				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	43	C	188	3	1:0/8	2:0/5	3/32	riant
PKH_02	ordered_PKN	7638				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	49	T	182	1	1:0/8	2:0/5	1/32	riant
PKH_02	ordered_PKN	7638				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	55	C	176	4	1:4/8	2:0/5	0/32	riant
PKH_02	ordered_PKN	7638				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	59	T	172	24	1:3/8	2:0/5	21/32	riant
PKH_02	ordered_PKN	7638				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	71	G	160	1	1:0/8	2:0/5	1/32	riant
PKH_02	ordered_PKN	7638				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	77	A	154	2	1:0/8	2:0/5	2/32	riant
PKH_02	ordered_PKN	7639				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	07	C	124	3	1:1/8	2:0/5	2/32	riant
PKH_02	ordered_PKN	7640				cluster	cluster	cluster3:	synonymous
1580	H_02_v2	62	G	93	5	1:0/8	2:0/5	5/32	_variant
PKH_02	ordered_PKN	7640				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	63	C	92	2	1:0/8	2:0/5	2/32	riant
PKH_02	ordered_PKN	7640				cluster	cluster	cluster3:	
1580	H_02_v2	84	G	71	1	1:0/8	2:0/5	1/32	stop_gained
PKH_02	ordered_PKN	7640				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	87	G	68	1	1:0/8	2:1/5	0/32	riant

PKH_02	ordered_PKN	7641				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	12	G	43	2	1:0/8	2:0/5	2/32	riant
PKH_02	ordered_PKN	7641				cluster	cluster	cluster3:	missense_va
1580	H_02_v2	48	C	7	1	1:0/8	2:0/5	1/32	riant
PKH_03	ordered_PKN	9767				cluster	cluster	cluster3:	synonymous
1930	H_03_v2	82	A	636	8	1:0/8	2:1/5	7/32	_variant
PKH_03	ordered_PKN	9770				cluster	cluster	cluster3:	missense_va
1930	H_03_v2	48	A	370	4	1:0/8	2:0/5	4/32	riant
PKH_14	ordered_PKN	1060				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	699	G	33	9	1:0/8	2:0/5	9/32	_variant
PKH_14	ordered_PKN	1060				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	700	A	34	3	1:0/8	2:0/5	3/32	riant
PKH_14	ordered_PKN	1060				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	706	G	40	4	1:0/8	2:0/5	4/32	riant
PKH_14	ordered_PKN	1060				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	713	G	47	13	1:0/8	2:0/5	13/32	riant
PKH_14	ordered_PKN	1060				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	725	T	59	1	1:0/8	2:0/5	1/32	riant
PKH_14	ordered_PKN	1060				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	738	G	72	41	1:8/8	2:2/5	31/32	_variant
PKH_14	ordered_PKN	1060				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	749	C	83	1	1:0/8	2:0/5	1/32	riant
PKH_14	ordered_PKN	1060				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	759	C	93	5	1:3/8	2:1/5	1/32	_variant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	
2200	H_14_v2	007	T	119	2	1:0/8	2:0/5	2/32	NA
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	073	C	185	1	1:0/8	2:0/5	1/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	080	A	192	37	1:8/8	2:1/5	28/32	_variant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	083	A	195	1	1:0/8	2:0/5	1/32	_variant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	094	C	206	9	1:8/8	2:0/5	1/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	117	C	229	1	1:0/8	2:0/5	1/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	127	C	239	3	1:0/8	2:0/5	3/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	162	C	274	1	1:0/8	2:0/5	1/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	207	C	319	2	1:0/8	2:0/5	2/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	210	G	322	2	1:2/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	214	T	326	1	1:1/8	2:0/5	0/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	220	A	332	1	1:0/8	2:0/5	1/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	265	C	377	2	1:0/8	2:0/5	2/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	267	G	379	3	1:2/8	2:0/5	1/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	496	C	608	1	1:0/8	2:1/5	0/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	497	A	609	2	1:0/8	2:0/5	2/32	_variant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	511	G	623	1	1:0/8	2:0/5	1/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	560	A	672	36	1:8/8	2:1/5	27/32	_variant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	694	G	806	8	1:8/8	2:0/5	0/32	riant

PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	773	A	885	1	1:0/8	2:0/5	1/32	_variant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	788	A	900	1	1:0/8	2:0/5	1/32	_variant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	806	C	918	43	1:8/8	2:3/5	32/32	_variant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	synonymous
2200	H_14_v2	822	T	934	8	1:0/8	2:0/5	8/32	_variant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	912	G	1024	7	1:0/8	2:0/5	7/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	945	A	1057	22	1:0/8	2:0/5	22/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	948	G	1060	3	1:0/8	2:0/5	3/32	riant
PKH_14	ordered_PKN	1061				cluster	cluster	cluster3:	missense_va
2200	H_14_v2	978	C	1090	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	229	G	21	2	1:0/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	359	T	151	20	1:6/8	2:2/5	12/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	
2850	H_07_v2	360	C	152	20	1:6/8	2:2/5	12/32	stop_gained
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	362	T	154	8	1:0/8	2:1/5	7/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	365	G	157	8	1:0/8	2:1/5	7/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	367	T	159	6	1:0/8	2:1/5	5/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	374	G	166	15	1:6/8	2:2/5	7/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	376	T	168	7	1:0/8	2:0/5	7/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	378	C	170	8	1:0/8	2:1/5	7/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	382	C	174	22	1:6/8	2:2/5	14/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	383	C	175	6	1:0/8	2:0/5	6/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	410	G	202	1	1:0/8	2:1/5	0/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	411	C	203	1	1:0/8	2:1/5	0/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	418	T	210	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	436	C	228	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	472	T	264	1	1:0/8	2:1/5	0/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	505	C	297	1	1:1/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	514	C	306	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	540	A	332	23	1:6/8	2:3/5	14/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	544	C	336	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	601	G	393	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	671	T	463	32	1:0/8	2:2/5	30/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	694	G	486	1	1:1/8	2:0/5	0/32	_variant

PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	730	T	522	30	1:4/8	2:3/5	23/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	739	A	531	6	1:4/8	2:1/5	1/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	755	A	547	1	1:0/8	2:1/5	0/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	780	A	572	18	1:4/8	2:3/5	11/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	784	C	576	33	1:8/8	2:3/5	22/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	790	T	582	11	1:0/8	2:0/5	11/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	791	C	583	30	1:8/8	2:3/5	19/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	802	T	594	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	804	C	596	29	1:8/8	2:3/5	18/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	809	A	601	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	813	C	605	20	1:5/8	2:1/5	14/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	821	G	613	8	1:0/8	2:0/5	8/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	824	G	616	12	1:4/8	2:0/5	8/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	830	C	622	2	1:0/8	2:1/5	1/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	850	C	642	12	1:4/8	2:0/5	8/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	852	C	644	5	1:0/8	2:0/5	5/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	856	C	648	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	859	G	651	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	976	A	768	13	1:0/8	2:0/5	13/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	978	A	770	10	1:3/8	2:0/5	7/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	979	C	771	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	983	G	775	12	1:0/8	2:0/5	12/32	riant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	985	C	777	9	1:2/8	2:0/5	7/32	_variant
PKH_07	ordered_PKN	1277				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	996	T	788	26	1:3/8	2:0/5	23/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	012	T	804	21	1:5/8	2:3/5	13/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	020	T	812	2	1:0/8	2:0/5	2/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	021	C	813	3	1:3/8	2:0/5	0/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	069	T	861	16	1:2/8	2:1/5	13/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	081	C	873	9	1:0/8	2:0/5	9/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	084	G	876	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	088	T	880	1	1:0/8	2:0/5	1/32	_variant

PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	108	C	900	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	118	T	910	33	1:8/8	2:1/5	24/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	136	A	928	32	1:8/8	2:2/5	22/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	138	C	930	31	1:8/8	2:2/5	21/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	141	G	933	26	1:8/8	2:2/5	16/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	144	G	936	31	1:8/8	2:2/5	21/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	146	A	938	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	151	C	943	31	1:8/8	2:2/5	21/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	190	A	982	11	1:0/8	2:1/5	10/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	194	C	986	11	1:0/8	2:1/5	10/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	219	C	1011	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	235	A	1027	4	1:0/8	2:0/5	4/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	243	C	1035	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	252	A	1044	8	1:6/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	255	G	1047	1	1:0/8	2:1/5	0/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	277	C	1069	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	290	T	1082	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	303	G	1095	1	1:0/8	2:1/5	0/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	318	G	1110	1	1:1/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	330	G	1122	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	339	G	1131	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	351	A	1143	1	1:1/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	369	A	1161	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	372	C	1164	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	374	C	1166	2	1:1/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	411	C	1203	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	414	A	1206	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	525	A	1317	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	537	T	1329	1	1:0/8	2:1/5	0/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	567	G	1359	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	568	G	1360	4	1:0/8	2:0/5	4/32	riant

PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	597	G	1389	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	609	A	1401	17	1:6/8	2:0/5	11/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	618	A	1410	29	1:6/8	2:3/5	20/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	627	A	1419	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	629	A	1421	1	1:0/8	2:1/5	0/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	646	G	1438	21	1:6/8	2:3/5	12/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	650	A	1442	21	1:6/8	2:3/5	12/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	651	A	1443	27	1:6/8	2:3/5	18/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	658	C	1450	20	1:6/8	2:2/5	12/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	672	T	1464	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	676	C	1468	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	678	A	1470	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	689	A	1481	1	1:0/8	2:1/5	0/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	731	G	1523	29	1:8/8	2:3/5	18/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	752	A	1544	32	1:8/8	2:3/5	21/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	756	G	1548	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	761	A	1553	29	1:8/8	2:2/5	19/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	764	C	1556	9	1:2/8	2:2/5	5/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	784	G	1576	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	798	C	1590	2	1:0/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	801	C	1593	8	1:6/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	843	A	1635	7	1:0/8	2:0/5	7/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	885	T	1677	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	888	G	1680	19	1:1/8	2:3/5	15/32	_variant
PKH_07	ordered_PKN	1278				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	915	G	1707	9	1:0/8	2:0/5	9/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	047	A	1839	25	1:0/8	2:1/5	24/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	068	A	1860	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	089	A	1881	9	1:8/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	107	T	1899	29	1:8/8	2:2/5	19/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	143	G	1935	2	1:2/8	2:0/5	0/32	riant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	170	T	1962	2	1:0/8	2:0/5	2/32	_variant

PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	176	A	1968	16	1:0/8	2:0/5	16/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	209	C	2001	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	218	G	2010	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	241	C	2033	1	1:0/8	2:1/5	0/32	riant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	254	A	2046	6	1:6/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	265	A	2057	6	1:6/8	2:0/5	0/32	riant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	269	C	2061	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	281	G	2073	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	284	C	2076	10	1:0/8	2:2/5	8/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	287	A	2079	2	1:0/8	2:1/5	1/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	304	C	2096	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	309	T	2101	7	1:4/8	2:0/5	3/32	riant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	776	A	2568	23	1:7/8	2:2/5	14/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	779	A	2571	23	1:7/8	2:2/5	14/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	782	G	2574	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	791	G	2583	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	800	C	2592	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	809	T	2601	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	830	T	2622	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	851	G	2643	39	1:8/8	2:2/5	29/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	872	A	2664	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	914	G	2706	8	1:8/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	920	G	2712	4	1:0/8	2:0/5	4/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	938	G	2730	2	1:0/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	983	C	2775	4	1:0/8	2:0/5	4/32	_variant
PKH_07	ordered_PKN	1279				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	994	A	2786	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	004	C	2796	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	020	C	2812	32	1:0/8	2:0/5	32/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	029	G	2821	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	073	A	2865	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	085	A	2877	1	1:0/8	2:0/5	1/32	_variant

PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	094	G	2886	8	1:8/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	114	A	2906	41	1:8/8	2:1/5	32/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	132	T	2924	27	1:0/8	2:1/5	26/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	141	C	2933	3	1:0/8	2:0/5	3/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	146	G	2938	2	1:0/8	2:0/5	2/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	154	A	2946	1	1:1/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	155	G	2947	16	1:0/8	2:1/5	15/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	157	A	2949	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	176	G	2968	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	179	C	2971	10	1:0/8	2:0/5	10/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	182	G	2974	10	1:0/8	2:0/5	10/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	242	A	3034	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	264	C	3056	2	1:0/8	2:0/5	2/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	278	A	3070	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	294	G	3086	14	1:0/8	2:0/5	14/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	303	C	3095	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	306	C	3098	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	316	A	3108	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	317	G	3109	15	1:0/8	2:0/5	15/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	318	G	3110	15	1:0/8	2:0/5	15/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	319	C	3111	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	320	G	3112	15	1:0/8	2:0/5	15/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	332	A	3124	2	1:0/8	2:2/5	0/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	334	T	3126	31	1:8/8	2:2/5	21/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	340	G	3132	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	341	A	3133	2	1:0/8	2:0/5	2/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	342	C	3134	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	344	A	3136	3	1:0/8	2:0/5	3/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	347	C	3139	33	1:8/8	2:2/5	23/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	350	G	3142	33	1:8/8	2:2/5	23/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	353	A	3145	29	1:8/8	2:2/5	19/32	riant

PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	365	G	3157	8	1:0/8	2:1/5	7/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	371	G	3163	27	1:8/8	2:2/5	17/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	372	G	3164	27	1:8/8	2:2/5	17/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	374	G	3166	37	1:8/8	2:2/5	27/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	378	T	3170	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	380	A	3172	19	1:8/8	2:2/5	9/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	385	G	3177	21	1:8/8	2:2/5	11/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	386	C	3178	7	1:0/8	2:0/5	7/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	391	T	3183	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	393	G	3185	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	400	A	3192	22	1:0/8	2:1/5	21/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	425	A	3217	10	1:8/8	2:1/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	433	A	3225	41	1:8/8	2:1/5	32/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	469	G	3261	16	1:0/8	2:0/5	16/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	494	T	3286	32	1:8/8	2:1/5	23/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	502	A	3294	42	1:8/8	2:2/5	32/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	519	C	3311	26	1:0/8	2:2/5	24/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	523	T	3315	25	1:0/8	2:2/5	23/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	529	T	3321	25	1:0/8	2:2/5	23/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	540	C	3332	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	553	T	3345	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	567	C	3359	1	1:1/8	2:0/5	0/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	568	A	3360	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	581	G	3373	1	1:0/8	2:1/5	0/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	637	C	3429	5	1:4/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	656	T	3448	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	682	G	3474	4	1:0/8	2:0/5	4/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	712	A	3504	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	716	C	3508	23	1:8/8	2:3/5	12/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	718	G	3510	5	1:0/8	2:0/5	5/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	725	T	3517	8	1:8/8	2:0/5	0/32	_variant

PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	733	C	3525	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	751	A	3543	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	754	T	3546	2	1:0/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	757	C	3549	24	1:0/8	2:0/5	24/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	761	A	3553	43	1:8/8	2:3/5	32/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	781	G	3573	12	1:0/8	2:0/5	12/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	823	G	3615	30	1:0/8	2:0/5	30/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	830	T	3622	8	1:8/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	892	C	3684	6	1:0/8	2:0/5	6/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	897	G	3689	15	1:0/8	2:0/5	15/32	riant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	907	A	3699	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	928	C	3720	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	934	C	3726	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1280				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	935	C	3727	2	1:0/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	051	G	3843	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	069	C	3861	9	1:0/8	2:2/5	7/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	087	C	3879	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	090	C	3882	2	1:0/8	2:1/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	111	A	3903	8	1:0/8	2:0/5	8/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	135	T	3927	2	1:0/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	156	G	3948	9	1:0/8	2:0/5	9/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	174	C	3966	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	177	A	3969	2	1:0/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	186	T	3978	43	1:8/8	2:3/5	32/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	202	G	3994	8	1:8/8	2:0/5	0/32	riant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	204	A	3996	2	1:0/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	215	T	4007	26	1:8/8	2:2/5	16/32	riant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	225	A	4017	6	1:0/8	2:0/5	6/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	231	T	4023	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	474	C	4266	11	1:0/8	2:0/5	11/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	500	T	4292	2	1:0/8	2:0/5	2/32	riant

PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	585	G	4377	6	1:0/8	2:0/5	6/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	594	G	4386	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	600	C	4392	2	1:0/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	609	A	4401	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	684	C	4476	16	1:0/8	2:0/5	16/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	696	T	4488	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	714	C	4506	2	1:0/8	2:2/5	0/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	732	A	4524	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	735	C	4527	3	1:0/8	2:0/5	3/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	747	C	4539	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	760	T	4552	8	1:8/8	2:0/5	0/32	riant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	777	C	4569	8	1:8/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	789	C	4581	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	792	A	4584	43	1:8/8	2:3/5	32/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	795	C	4587	7	1:0/8	2:0/5	7/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	798	A	4590	40	1:8/8	2:0/5	32/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	801	A	4593	5	1:0/8	2:0/5	5/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	828	C	4620	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	829	C	4621	2	1:0/8	2:2/5	0/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	873	G	4665	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	876	C	4668	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	882	G	4674	29	1:0/8	2:0/5	29/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	891	T	4683	41	1:8/8	2:1/5	32/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	906	C	4698	2	1:0/8	2:2/5	0/32	riant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	921	G	4713	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	928	C	4720	41	1:8/8	2:3/5	30/32	riant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	970	G	4762	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	982	G	4774	5	1:0/8	2:1/5	4/32	riant
PKH_07	ordered_PKN	1281				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	986	A	4778	5	1:0/8	2:1/5	4/32	riant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	003	G	4795	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	005	T	4797	9	1:8/8	2:1/5	0/32	_variant

PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	missense_va
2850	H_07_v2	006	C	4798	6	1:0/8	2:1/5	5/32	riant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	017	G	4809	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	069	C	4861	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	122	C	4914	3	1:3/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	170	A	4962	9	1:8/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	242	A	5034	43	1:8/8	2:3/5	32/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	311	A	5103	4	1:0/8	2:0/5	4/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	410	C	5202	5	1:0/8	2:1/5	4/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	440	C	5232	7	1:7/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	467	G	5259	9	1:8/8	2:1/5	0/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	470	A	5262	16	1:0/8	2:1/5	15/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	473	G	5265	16	1:0/8	2:1/5	15/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	542	G	5334	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	584	T	5376	34	1:0/8	2:2/5	32/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	608	C	5400	25	1:0/8	2:0/5	25/32	_variant
PKH_07	ordered_PKN	1282				cluster	cluster	cluster3:	synonymous
2850	H_07_v2	618	C	5410	8	1:8/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	352	A	181	2	1:1/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	synonymous
3420	H_07_v2	582	G	411	10	1:8/8	2:0/5	2/32	_variant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	615	G	444	20	1:0/8	2:0/5	20/32	riant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	646	A	475	32	1:0/8	2:0/5	32/32	riant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	655	C	484	40	1:8/8	2:0/5	32/32	riant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	synonymous
3420	H_07_v2	666	T	495	8	1:8/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	synonymous
3420	H_07_v2	714	G	543	1	1:0/8	2:1/5	0/32	_variant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	synonymous
3420	H_07_v2	720	C	549	8	1:8/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	737	T	566	32	1:0/8	2:0/5	32/32	riant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	755	G	584	8	1:8/8	2:0/5	0/32	riant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	806	T	635	8	1:8/8	2:0/5	0/32	riant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	810	A	639	8	1:8/8	2:0/5	0/32	riant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	839	T	668	1	1:0/8	2:0/5	1/32	riant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	synonymous
3420	H_07_v2	867	C	696	11	1:8/8	2:1/5	2/32	_variant
PKH_07	ordered_PKN	1489				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	981	G	810	2	1:0/8	2:0/5	2/32	riant

PKH_07	ordered_PKN	1490				cluster	cluster	cluster3:	synonymous
3420	H_07_v2	026	G	855	8	1:8/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1490				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	030	G	859	8	1:8/8	2:0/5	0/32	riant
PKH_07	ordered_PKN	1490				cluster	cluster	cluster3:	synonymous
3420	H_07_v2	065	C	894	1	1:1/8	2:0/5	0/32	_variant
PKH_07	ordered_PKN	1490				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	077	C	906	32	1:0/8	2:0/5	32/32	riant
PKH_07	ordered_PKN	1490				cluster	cluster	cluster3:	synonymous
3420	H_07_v2	092	A	921	1	1:0/8	2:0/5	1/32	_variant
PKH_07	ordered_PKN	1490				cluster	cluster	cluster3:	missense_va
3420	H_07_v2	099	G	928	21	1:0/8	2:0/5	21/32	riant
PKH_12	ordered_PKN	2142				cluster	cluster	cluster3:	missense_va
0020	H_12_v2	094	T	1253	7	1:0/8	2:0/5	7/32	riant
PKH_12	ordered_PKN	2143				cluster	cluster	cluster3:	synonymous
0020	H_12_v2	133	G	219	1	1:0/8	2:0/5	1/32	_variant
PKH_12	ordered_PKN	2143				cluster	cluster	cluster3:	synonymous
0020	H_12_v2	226	A	126	3	1:3/8	2:0/5	0/32	_variant
PKH_12	ordered_PKN	2143				cluster	cluster	cluster3:	synonymous
0020	H_12_v2	259	A	93	3	1:3/8	2:0/5	0/32	_variant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	373	C	7	1	1:0/8	2:1/5	0/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	403	G	37	2	1:0/8	2:0/5	2/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	411	C	45	1	1:0/8	2:1/5	0/32	_variant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	470	A	104	1	1:0/8	2:0/5	1/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	491	A	125	23	1:0/8	2:0/5	23/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	519	G	153	1	1:0/8	2:0/5	1/32	_variant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	679	A	313	8	1:8/8	2:0/5	0/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	703	G	337	40	1:8/8	2:0/5	32/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	704	T	338	40	1:8/8	2:0/5	32/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	710	G	344	40	1:8/8	2:0/5	32/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	731	A	365	39	1:8/8	2:0/5	31/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	741	T	375	17	1:0/8	2:0/5	17/32	_variant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	756	G	390	40	1:8/8	2:0/5	32/32	_variant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	766	T	400	40	1:8/8	2:0/5	32/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	
1770	H_12_v2	767	C	401	15	1:8/8	2:0/5	7/32	stop_gained
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	775	G	409	3	1:0/8	2:0/5	3/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	778	A	412	2	1:0/8	2:0/5	2/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	867	C	501	14	1:0/8	2:0/5	14/32	_variant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	871	G	505	1	1:0/8	2:0/5	1/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	884	C	518	10	1:0/8	2:0/5	10/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	892	G	526	10	1:0/8	2:0/5	10/32	riant

PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	933	C	567	1	1:0/8	2:0/5	1/32	_variant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	943	C	577	32	1:0/8	2:0/5	32/32	riant
PKH_12	ordered_PKN	2906				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	953	A	587	32	1:0/8	2:0/5	32/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	016	G	650	8	1:8/8	2:0/5	0/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	154	C	788	1	1:0/8	2:0/5	1/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	157	T	791	2	1:0/8	2:0/5	2/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	179	G	813	23	1:0/8	2:0/5	23/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	204	G	838	1	1:0/8	2:0/5	1/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	206	T	840	2	1:0/8	2:0/5	2/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	209	A	843	2	1:0/8	2:0/5	2/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	245	C	879	8	1:0/8	2:0/5	8/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	251	C	885	1	1:1/8	2:0/5	0/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	254	A	888	2	1:0/8	2:0/5	2/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	296	G	930	3	1:0/8	2:0/5	3/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	317	T	951	41	1:8/8	2:1/5	32/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	322	G	956	2	1:0/8	2:1/5	1/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	359	C	993	35	1:8/8	2:0/5	27/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	383	C	1017	5	1:0/8	2:0/5	5/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	385	C	1019	8	1:8/8	2:0/5	0/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	389	C	1023	2	1:0/8	2:2/5	0/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	431	C	1065	26	1:0/8	2:1/5	25/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	435	T	1069	1	1:0/8	2:0/5	1/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	442	T	1076	13	1:0/8	2:0/5	13/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	443	C	1077	1	1:0/8	2:0/5	1/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	480	A	1114	32	1:0/8	2:0/5	32/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	481	G	1115	8	1:8/8	2:0/5	0/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	497	T	1131	33	1:0/8	2:2/5	31/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	507	G	1141	33	1:0/8	2:2/5	31/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	513	C	1147	32	1:0/8	2:0/5	32/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	537	C	1171	30	1:3/8	2:1/5	26/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	559	A	1193	1	1:1/8	2:0/5	0/32	riant

PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	673	G	1307	16	1:0/8	2:0/5	16/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	686	C	1320	1	1:0/8	2:0/5	1/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	687	G	1321	1	1:0/8	2:0/5	1/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	697	C	1331	40	1:8/8	2:0/5	32/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	698	G	1332	4	1:0/8	2:0/5	4/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	704	C	1338	40	1:8/8	2:0/5	32/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	726	A	1360	1	1:0/8	2:0/5	1/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	730	A	1364	1	1:0/8	2:0/5	1/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	815	C	1449	2	1:0/8	2:0/5	2/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	817	A	1451	30	1:8/8	2:0/5	22/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	824	T	1458	30	1:8/8	2:0/5	22/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	848	T	1482	1	1:0/8	2:1/5	0/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	854	C	1488	1	1:0/8	2:1/5	0/32	_variant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	978	G	1612	1	1:1/8	2:0/5	0/32	riant
PKH_12	ordered_PKN	2907				cluster	cluster	cluster3:	missense_va
1770	H_12_v2	988	G	1622	1	1:1/8	2:0/5	0/32	riant
PKH_12	ordered_PKN	2908				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	034	G	1668	1	1:0/8	2:0/5	1/32	_variant
PKH_12	ordered_PKN	2908				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	076	C	1710	1	1:0/8	2:0/5	1/32	_variant
PKH_12	ordered_PKN	2908				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	085	C	1719	1	1:1/8	2:0/5	0/32	_variant
PKH_12	ordered_PKN	2908				cluster	cluster	cluster3:	synonymous
1770	H_12_v2	094	T	1728	35	1:0/8	2:3/5	32/32	_variant

APPENDIX V: PROTOCOL

ELISA for Sera (using TMB) – 384 well plates Biomek standard operating protocol (SOP)

1. Overview

This SOP describes the ELISA process in order to estimate levels of specific antibodies to malaria antigens, using TMB rather than OPD at the substrate for HRP.

2. Safety

Adhere to local safety regulations. Wear appropriate personal protective equipment.

3. Materials

Equipment

a) 384-well ELISA plates: Clear flat-bottom immuno nonsterile (4titude)
b) 96 deep well plates (Costar 0.5ml v bottom assay block)
c) 3 plastic buckets/containers for washing plates
d) Plastic buckets/containers for storing plates o/n
e) Plate Reader
f) Protective latex or nitrile gloves
g) Multi-channel pipettes (8 or 12 channel 5-50 µl, 30-300 µl)
h) Range of single channel pipettes
i) Tips

Documents

Laboratory Record Book

Reagents

All reagents should be stored according to the instructions supplied with them and disposed of at the expiry date recorded on the product.

a) Antigens

b) Controls (standard dilutions)

c) TMB one component HRP microwell substrate (#TMBW-1000-01; Tebu-bio laboratories)

d) Tween 20 (Sigma)

e) Skimmed milk powder (Supermarket)

f) Horseradish peroxidase-conjugated rabbit anti-human IgG (#P0214; Dako)

g) NaH_2PO_4 (Sodium dihydrogen orthophosphate) (VWR International Ltd)

h) Na_2HPO_4 (di-sodium hydrogen orthophosphate) (VWR International Ltd)

i) NaCl (Sodium chloride) (Fisher scientific)

j) H_2SO_4 (Sulphuric acid) (BDH)

k) Na_2CO_3 (Sodium carbonate) (Sigma)

l) $NaHCO_3$ (Sodium hydrogen carbonate) (VWR International Ltd)

4. Preparation of Buffer Solutions

All buffer solutions should be clearly labelled with:

* Reagent name

* Expiry date

* Preparation date

* Name of person who prepared the buffer

Reagents should be stored under appropriate conditions. See Table 1 for details on preparation and storage.

Table 1: Buffer solutions - preparation

Buffer solution	Reagent/chemical	Amount/Volume
Phosphate buffered saline (PBS) 10X pH 7.2	NaH_2PO_4	5.7 g
	Na_2HPO_4	16.7 g
	NaCl	85 g
	Distilled water	Make up to 1 L
Store at room temperature, dispose of after one month		
PBS/Tween (0.05%) wash solution	PBS 10X	500 ml
	Tween 20	2.5 ml
	Distilled water	4.5 L
Make up as needed daily, dispose of unused solution at the end of each day		
Coating buffer pH 9.4-9.6	Na_2CO_3	1.59 g
	$NaHCO_3$	2.93 g
	Distilled water	Make up to 1 L
Store at 4°C, dispose of after one month, **pH should be 9.5 ± 0.2**		
BLOCKING SOLUTION: 1% skimmed milk powder in PBS/Tween	Skimmed milk powder	10 g
	PBS/Tween	1 L
Make up as needed daily, dispose of unused solution at the end of each day		
0.2 M H_2SO_4	H_2SO_4 concentrate	10.7 ml
	Distilled water	Add concentrate to ~800 ml of water, then top up to 1 L
N.B. Appropriate safety regulations must be adhered to when handling H_2SO_4		
Store at room temperature, dispose of after six months		

N.B. Fully dissolve the reagent/chemical prior to use

5. Methods

Drying plates

a) Expel any excess liquid from the wash process on paper towel, but DO NOT let the plates dry out.

5.1. ELISA

5.1.1. **Day 1:** Mark out ELISA plates and record the template for each in laboratory record book. Column allocation (figure 1 and 2):

Wells 1-20 A-P -samples (tested in duplicate)

Wells 23 A-F and 24 A-F -positive control standard curve

Wells 23 G-H and 24 G-H are for negative controls (if required)

Wells 21 A-P and 22 A-P, 23 I-P and 24 I-P are blank wells

			Columns 1-5 from DW										Columns 6-10 from DW											Standards	
	384 row ID	DW row ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
DW1	A	A	sample 1		sample 9		sample 17		sample 25		sample 33		sample 41												1/10	
DW2	B	A	sample 1		sample 9		sample 17		sample 25		sample 33		sample 41												1/40	
DW1	C	B	sample 2		sample 10		sample 18		sample 26		sample 34														1/160	
DW2	D	B	sample 2		sample 10		sample 18		sample 26		sample 34														1/640	
DW1	E	C	sample 3		sample 11		sample 19		sample 27		sample 35														1/2560	
DW2	F	C	sample 3		sample 11		sample 19		sample 27		sample 35														1/10240	
DW1	G	D	sample 4		sample 12		sample 20		sample 28		sample 36														Neg controls	
DW2	H	D	sample 4		sample 12		sample 20		sample 28		sample 36												Blanks			
DW1	I	E	sample 5		sample 13		sample 21		sample 29		sample 37															
DW2	J	E	sample 5		sample 13		sample 21		sample 29		sample 37															
DW1	K	F	sample 6		sample 14		sample 22		sample 30		sample 38															
DW2	L	F	sample 6		sample 14		sample 22		sample 30		sample 38															
DW1	M	G	sample 7		sample 15		sample 23		sample 31		sample 39															
DW2	N	G	sample 7		sample 15		sample 23		sample 31		sample 39															
DW1	O	H	sample 8		sample 16		sample 24		sample 32		sample 40															
DW2	P	H	sample 8		sample 16		sample 24		sample 32		sample 40															

Figure 1 - 384-well ELISA plate layout

DW1	1	2	3	4	5	6	7	8	9	10	11	12
A	sample 1	sample 9	sample 17	sample 25	sample 33	sample 41	sample 49	sample 57	sample 65	sample 73		1/10
B	sample 2	sample10	sample 18	sample 26	sample 34	sample 42	sample 50	sample 58	sample 66	sample 74		1/160
C	sample 3	sample 11	sample 19	sample 27	sample 35	sample 43	sample 51	sample 59	sample 67	sample 75		1/2560
D	sample 4	sample 12	sample 20	sample 28	sample 36	sample 44	sample 52	sample 60	sample 68	sample 76	enough vol milk for 2 DW	Neg
E	sample 5	sample 13	sample 21	sample 29	sample 37	sample 45	sample 53	sample 61	sample 69	sample 77		
F	sample 6	sample 14	sample 22	sample 30	sample 38	sample 46	sample 54	sample 62	sample 70	sample 78		milk
G	sample 7	sample 15	sample 23	sample 31	sample 39	sample 47	sample 55	sample 63	sample 71	sample 79		
H	sample 8	sample 16	sample 24	sample 32	sample 40	sample 48	sample 56	sample 64	sample 72	sample 80		

												Standard
DW2	1	2	3	4	5	6	7	8	9	10	11	12
A												1/40
B												1/640
C												1/10240
D												Neg
E												
F												milk
G												
H												

Figure 2- Sample positions in 96-well deepwell plate

1.1.2. Prepare the antigen in coating buffer, taking into account the number of plates requiring coating and ensuring the antigen is diluted to the correct concentration (as previously determined via antigen titration).

Coating antigen must be aliquoted into a 96-well deepwell (DW) plate. Store DW at -20°C once finished adding to ELISA plates.
As shown in figure 3, 1 quad (eg. 1A, 1B, 2A, 2B) in the 384-well plate is the equivalent of 1 well (eg. 1A plate 1, 1A plate 2, 1A plate 3 and 1A plate 4) in 4 different 96-well ELISA plates.
For help calculating volumes plus 20% contingency:
25µl per well x 384 wells ≈ 10ml (total volume required for one 384-well plate)
25µl x 4 wells = 1 quad = 100µl (Total volume in 1 quad)
When preparing the coating buffer with antigen in the DW, the total volume in 1 quad is the total volume in 1 well in the DW plate.
With the 20% added contingency, we will say that the total coating volume in 1 quad is 120µl.
120µl x 96wells in DW x number of 384well plates = Total Coating buffer needed
120µl x number of 384well plates = Total volume in each well of the DW plate

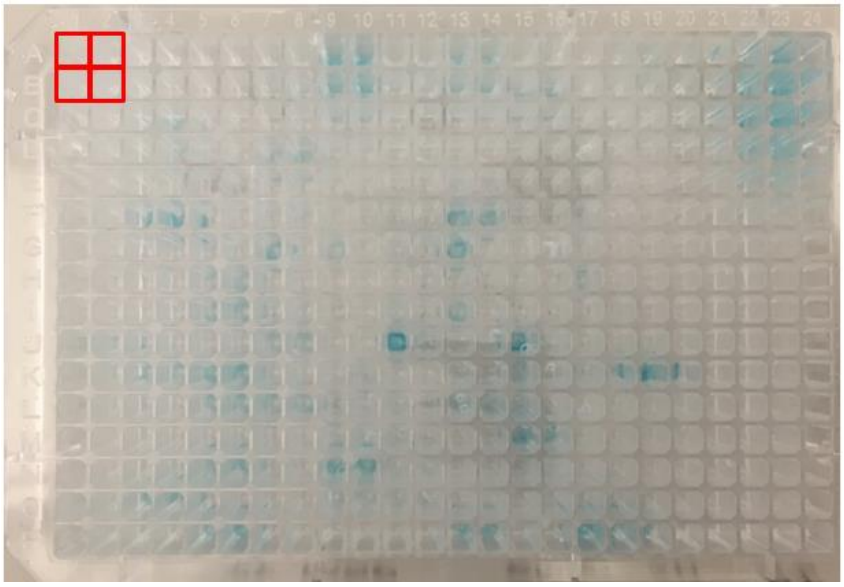


Figure 3- 384 well ELISA plate. Red squares indicate 1 quad

5.1.3. Set up the Biomek deck as indicated in figure 4. Position TL1 must always remain free, allowing the robotic arm to pick up tip boxes and switch boxes. Place p250 tips in position P3. Place the coating antigen DW in position P6. Place the 384-well ELISA plate to be coated with that antigen in position P14.

Add 25µl of this diluted antigen, prepared in 5.1.2, to all wells of the 384 well ELISA plate using program **"Lou 1 copy1 Add Coating Antibody"** .Tap the

plate to ensure the liquid covers the base of the wells or use a pipette tip to burst any bubbles. Dispose of tips.

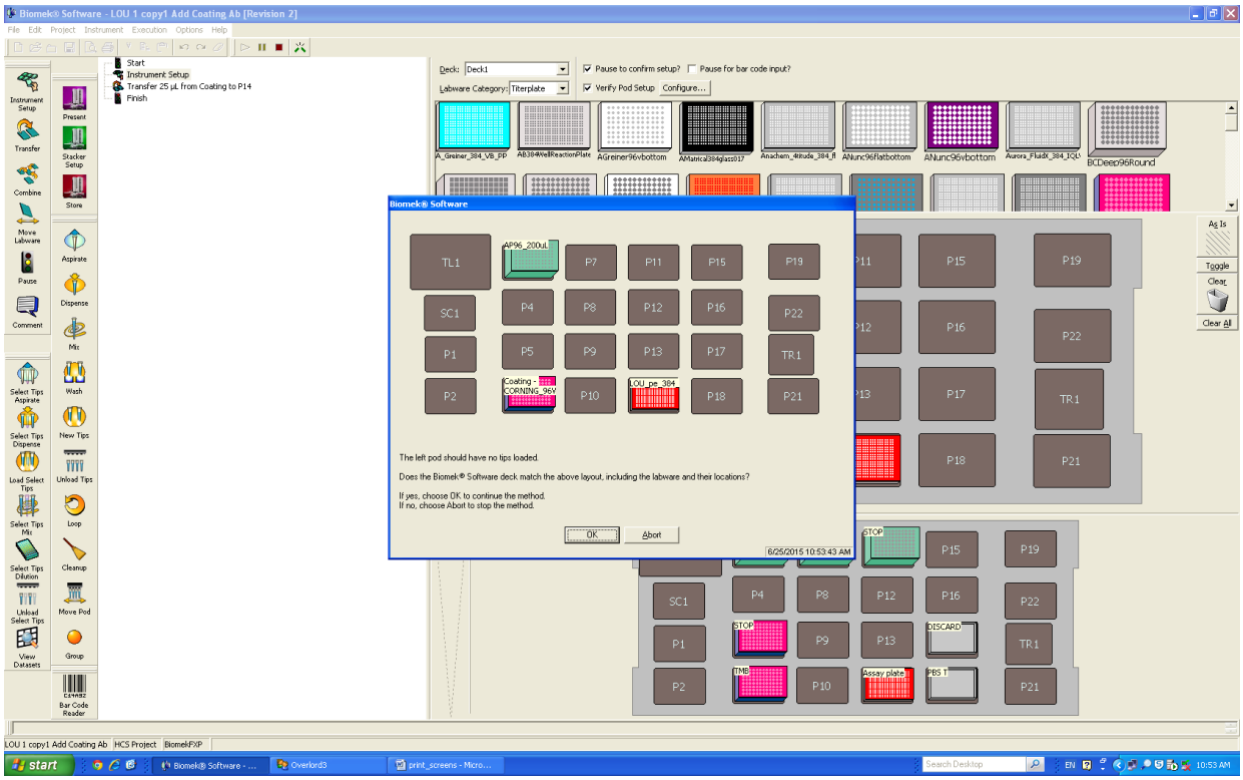


Figure 4 – Biomek deck set up for coating step

5.1.4. Incubate plates overnight at 4°C, to prevent evaporation store plates in a plastic bucket/container with wet blue towel. Consider wrapping plates in cling film if evaporation is still a problem.

NOTE: Prepare washing buffer, blocking buffer and samples the evening before and store at 4°C.

Roughly a total of 50ml of blocking buffer per 384well plate is required (300µl per quad) for both the blocking and sample addition steps.

5.1.5. **Day 2:** Set up the Biomek deck as indicated in figure 5. Place p250 tips in position P3. Place the blocking buffer container (can be a pipette tip box lid) in position P6. Place the 384-well ELISA plate in position P14. Place the discard container in position P17. Place the wash buffer container in position P18. Wash ELISA plates three times in 100µl PBS/Tween wash solution using program **"Lou 2 copy1 Wash x3 Add Block"**.

Make sure there is enough wash buffer in the respective container on the deck and make sure the waste container does not spill over. Pause the Biomek in order to refill or empty containers by either breaching the sensor curtain or by manually pausing in overlord. Once all containers have been attended to, click play and the Biomek will take over from where it was paused.

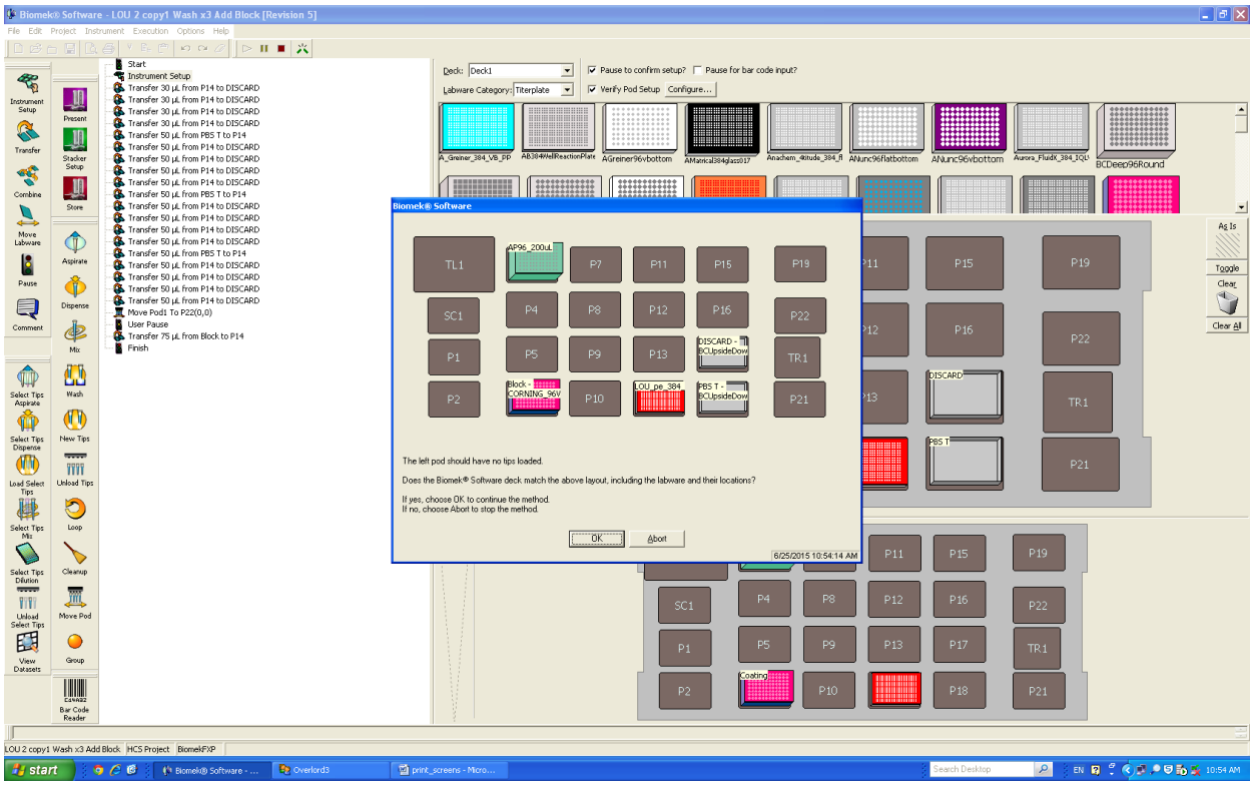


Figure 5 – Biomek deck set up for wash and blocking buffer step

5.1.6. The Biomek program will automatically pause to allow plate drying. Dry each plate using method above. Add 75µl of blocking solution to each well on the plate using the same program as above (just hit enter and the program will pick up where it stopped).

5.1.7. Incubate plates at room temperature for three hours with a cover plate on top to avoid evaporation and to keep safe from dust.

5.1.8. One hour before the end of this time remove control aliquot from freezer and allow to thaw

5.1.9. Make up standard dilutions in the samples deep well plate, refer to annex 1, ensuring sufficient volume for each plate being assayed

5.1.10. Set up the Biomek deck as indicated in figure 6. Use the same p250 tips as the wash and blocking buffer step in position P3. Place p20 tips in positions P7 and P11. Place the blocking buffer container (can be a pipette tip box lid) in position P10. Place the 384-well ELISA plate in

position P14. Place the discard container in position P17. Place the wash buffer container in position P18. Place sample DW1 in position P5 and sample DW2 in position P6. Wash ELISA plates three times in 100µl PBS/Tween wash solution using program **"Lou 3 copy1 @ 1/1000 Wash x3 Add Samples / Milk"** or **"Lou 3 copy1 @ 1/2000 Wash x3 Add Samples / Milk"**, or adjust programing to the required dilution. Make sure there is enough wash buffer in the respective container on the deck and make sure the waste container does not spill over. Pause the Biomek in order to refill or empty containers by either breaching the sensor curtain or by manually pausing in overlord. Once all containers have been attended to, click play and the Biomek will take over from where it was paused. Dispose of p20 tips. The p250 tips may still be used the next day for the washing step if they appear to be clear and unobstructed.

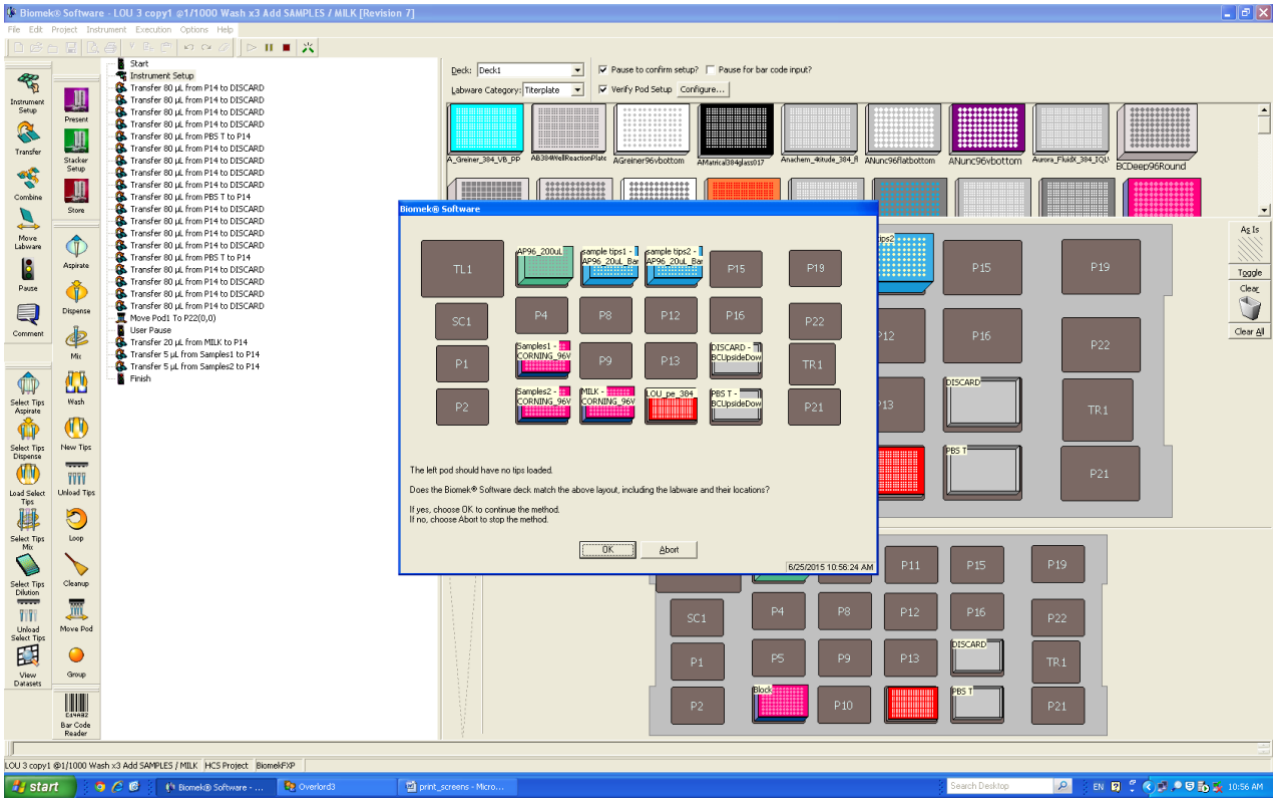


Figure 6 - Biomek deck set up for washing and sample addition step

5.1.11. The Biomek program will automatically pause to allow plate drying. Dry plates as described above. Add blocking solution, samples and controls to the ELISA plate using the same program. The volumes added depends on the final dilution of sample required (change Biomek programing for different sample dilutions!!).

Examples:
If the sample stock dilution is at 1/200 and a 1/1000 dilution is required for the ELISA, the sample stocks will need to be diluted 1 in 5 e.g. 5µl of sample stock added to 20µl blocking solution.
If the sample stock dilution is at 1/200 and a 1/2000 dilution is required for the ELISA, the sample stocks will need to be diluted 1 in 10 e.g. 2.5µl of sample stock added to 22.5µl blocking solution.
The samples will be automatically added in duplicates as shown in figure 1. Columns 1-5 of the DW (Samples) will be on the first half of the ELISA plate & columns 6-10 will be on the second half of the ELISA plate. The volume added will depend on the final dilution required.

5.1.12. Add Negative controls at the required dilution to wells 23 G-H and 24 G-H (if required) and blocking solution to 11H and 12H as added for the sample Alternatively: Add the same volume, as added for the positive control, of blocking solution to 21 A-P and 22 A-P, 23 I-P and 24 I-P (see table 1).

5.1.13. Incubate plates overnight at 4°C, to prevent evaporation store plates in a plastic bucket/container with wet blue towel. Consider wrapping plates in cling film if evaporation is still a problem.

NOTE: Prepare washing buffer, secondary antibody, TMB and Stop solution the evening before and store at 4°C.
Secondary, Stop solution and TBM should be aliquoted into a 96well DW plate. Wrap the TMB DW plate in foil to protect from light.

5.1.14. **Day 3:** Set up the Biomek deck as indicated in figure 7. Use the same p250 tips from day 2 wash step in position P3. Place the secondary DW in position P6. Place the 384-well ELISA plate in position P14. Place the discard container in position P17. Place the wash buffer container in position P18. Wash five times in PBS/Tween wash solution using program **"Lou 4 copy1 Wash x5 Add secondary"**. Make sure there is enough wash buffer in the respective container on the deck and make sure the waste container does not spill over. Pause the Biomek in order to refill or empty containers by either breaching the sensor curtain or by manually pausing in overlord. Once all containers have been attended to,

click play and the Biomek will take over from where it was paused. Dry plates

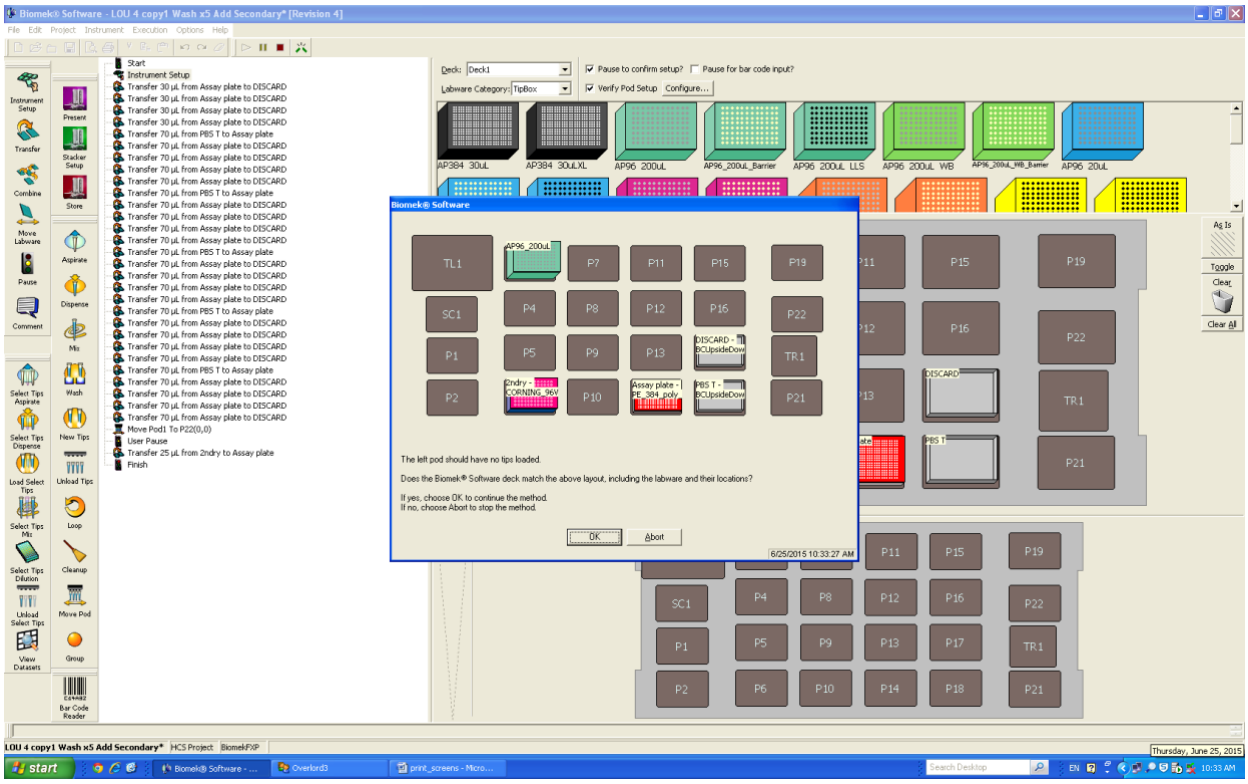


Figure 7 - Biomek deck set up for secondary addition step

5.1.15. Make the appropriate volume for the number of ELISA plates of conjugate solution Preparation: horseradish peroxidase-conjugated rabbit anti-human IgG diluted at **1/15,000**, in PBS/Tween wash solution. Refer to calculations in 5.1.2 to prepare appropriate volumes of secondary solution.

5.1.16. Add 25µl of conjugate solution to each well of the ELISA plates using the same program.

5.1.17. Incubate for three hours at room temperature with a cover plate on top to avoid evaporation and to keep safe from dust.

5.1.18. Set up the Biomek deck as indicated in figure 8. Use the same p250 tips from the wash step in position P3. Place new p250 tips in position P7 and position P11. Place the TMB DW in position P6 only when the ELISA plates are ready to receive it (keep TMB away from light during the wash step). Place the stop buffer DW in position P5. Place the 384-well ELISA plate in position P14. Place the discard container in position P17. Place the wash buffer container in position P18. Wash five times in PBS/Tween

wash solution using program **"Lou 5 copy1 Wash x5 Add TMB and STOP"**.

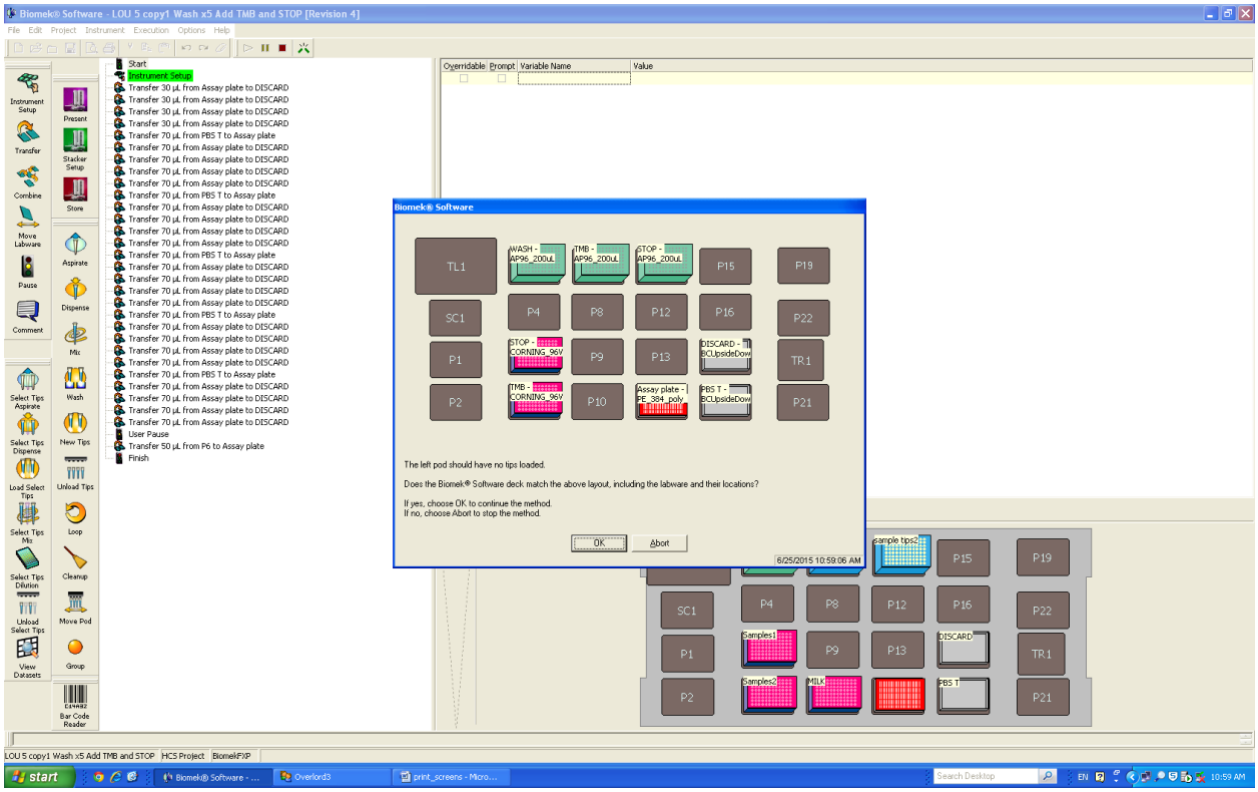


Figure 8 - Biomek deck set up for TMB step

5.1.19. Add 50µl per well of TMB substrate solution. Leave at room temperature in the dark for 15 min for the assay to develop. During the 15 min incubation period it is possible to start washing a second ELISA plate and have it washed, dried and developing before the end of the incubation time of the first plate.

5.1.20. Stop the reaction by adding 25 µl 0.2 M sulphuric acid (H_2SO_4) using program **"Lou 6 copy1 add STOP"**.

5.1.21. Read plates as soon as possible at 450 nm using the Softmax Pro program. Open reading template "My Documents/Lou". Export data to "C:/Documents and Settings/iimmhlab/Desktop/LH_ELISA_Reads" folder as text file. Softmax Pro will automatically save the plate readings in the "My Documents/Lou" but it will only be labelled with the date and time.

Annex 1

Working out control serial dilutions:

Controls must be prepared in the samples DW columns 11 and 12 when using only two DW sample plates.

Or

Controls must be prepared in a separate DW plate in columns 11 and 12.

Rather than prepare controls separately, a stock of controls that will be enough for all the ELISA plates is made. This cuts down on variability from separate dilutions.

The examples below are routinely used for MSP-1 and AMA-1 antigens. However, control preparations may vary for other antigens (refer to Figure 1 for control positions).

Transfer them to the DW used in the assay (sample transfer DW, not sample storage DW). Four fold serial dilutions are prepared at a starting dilution of 1/10:

Serial dilution: 1/10, 1/40, 1/160, 1/640, 1/2560, 1/10240. This is the stock of standard control.

To work out how much control you need, you can follow these equations:

Total needed for each plate- 10µl for 1/5 OR 5µl for 1/10

No. of plates= n

Blocking solution in positions 12A DW2, 12B DW1, 12B DW2, 12C DW1, 12C DW2 (A) = 10µl OR 5µl * n + 60µl (for excess)

Transfer volume between wells (T) = a/3 (this is for a ¼ serial dilution)

(Transfer from 12A DW1 to 12A DW2 to 12B DW1 to 12B DW2 to 12C DW1 to 12C DW2)

Total volume in top well (12A DW1) (C) = A + T

Positive control serum/plasma in top well (12A DW1) = C/10

Blocking solution in top well (12A DW1) = C/10 x 9

Add at least the same volume of blocking buffer as C in wells 12D-H.

Add the same volume of blocking buffer as C in wells 11A-H plus an extra 20 µl per well.

MSP-1:

1 in 5 dilution of the stock control (stock at 1/200) onto each ELISA plate (20 µl blocking solution & 5 µl stock control) – final dilution of sample will be 1/1000

Final serial dilution: 1/50, 1/200, 1/800, 1/3200, 1/12800 & 1/51200

AMA-1:

1 in 10 dilution of the stock control (stock at 1/200) onto each ELISA plate (22.5µl blocking solution & 2.5µl stock control) – final dilution of sample will be 1/2000

Final serial dilution: 1/100, 1/400, 1/1600, 1/6400, 1/25600, 1/102400

APPENDIX VI: PROTOCOL

Protein Microarray printing standard operating protocol (SOP)

Author:

Name: Tate Oulton

Position: Scientific Officer

Signature: Date: 21/11/2017

1 Overview

This SOP describes the methods to print protein microarrays.

2 Safety

Adhere to local safety regulations. Wear appropriate personal protective equipment.

3 Materials

Equipment

a) Oncyte NOVA microarray slides (GraceBio Labs)
b) 384-well microtiter plate(s) (ArrayJet)
c) JetGuard plate septa
d) Single/multichannel pipettes and tips
e) Centrifuge
f) 'Command Centre for Marathon' software (ArrayJet)
g) Marathon Classic microarray printer (ArrayJet)

Documents

a) Laboratory Record Book
b) Glycerol System Buffer recipe
c) Glycerol Printing Buffer recipe
d) Software Application Training
e) High Pressure Purge Procedure

Reagents

All reagents should be stored according to the instructions supplied with them and disposed of at the expiry date recorded on the product.

a) 1x PBS solution (Oxoid)
b) Ultrapure/distilled water
c) Purified proteins/antigens of interest
d) Glycerol printing buffer

4 Preparation of Buffer Solutions

All buffer solutions should be clearly labelled with:

- Reagent name
- Expiry date
- Preparation date
- Name of person who prepared the buffer

Reagents should be stored under appropriate conditions. See Table 1 for details on preparation and storage.

Table 1: Buffer solutions

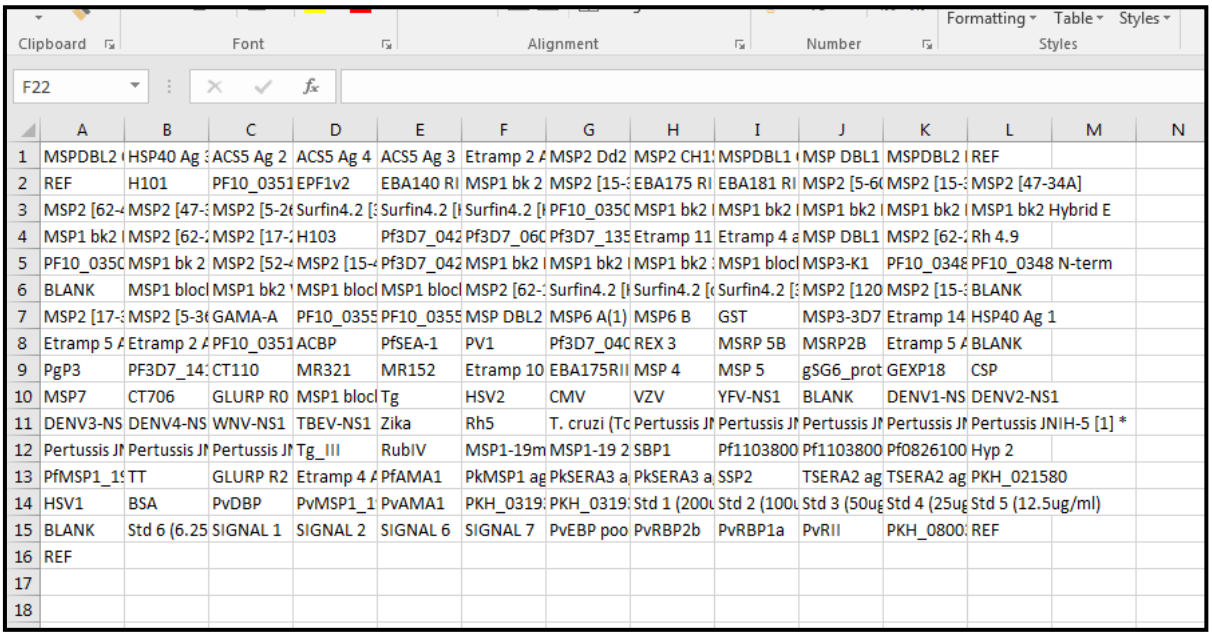
Buffer solution	Reagent/chemical	Amount/Volume
Phosphate buffered saline (PBS) 1X	PBS tablets	1
	Ultrapure/MilliQ water	100 ml
Store at room temperature, dispose of after one month		
Glycerol printing buffer	See 'Glycerol Printing Buffer recipe'	
Store at room temperature, dispose of after 3 months		
Blocking buffer 0.25X	See 'Glycerol System Buffer recipe'	
Store at room temperature, dispose of after 6 months		

5 Methods

Array design

1. Decide on a panel of proteins/antigens to be included in the arrays. It is useful to include 'reference' spots to help with finding the arrays during imaging and analysis. The material printed as reference spots should be chosen on the basis that it will show fluorescence independently of sample reactivity (e.g. purified IgG when using an anti-IgG secondary antibody, or the secondary antibody itself). You may also wish to include a standard control curve (e.g. serial dilution of purified IgG) to allow for approximation of reactivity.

2. Create a *.csv* file in MS Excel which describes the layout of the array you wish to print. For detailed instructions, refer to the 'Software Application Training' presentation, but briefly;
   2.1. The layout should start at cell A1. Each row should be **no more than** 12 cells long (Figure 1). The number of rows is limited only by the space available to print on



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	MSPDBL2 (	HSP40 Ag 3	ACS5 Ag 2	ACS5 Ag 4	ACS5 Ag 3	Etramp 2 A	MSP2 Dd2	MSP2 CH1!	MSPDBL1 (	MSP DBL1	MSPDBL2 (	REF		
2	REF	H101	PF10_0351	EPF1v2	EBA140 RI	MSP1 bk 2	MSP2 [15-3	EBA175 RI	EBA181 RI	MSP2 [5-6(	MSP2 [15-3	MSP2 [47-34A]		
3	MSP2 [62-4	MSP2 [47-3	MSP2 [5-2(	Surfin4.2 [3	Surfin4.2 [I	Surfin4.2 [I	PF10_035C	MSP1 bk2 (	MSP1 bk2 (	MSP1 bk2 (	MSP1 bk2 (	MSP1 bk2 Hybrid E		
4	MSP1 bk2 (	MSP2 [62-2	MSP2 [17-2	H103	Pf3D7_042	Pf3D7_06C	Pf3D7_135	Etramp 11	Etramp 4 a	MSP DBL1	MSP2 [62-2	Rh 4.9		
5	PF10_035C	MSP1 bk 2	MSP2 [52-4	MSP2 [15-4	Pf3D7_042	MSP1 bk2 (	MSP1 bk2 (	MSP1 bk2 :	MSP1 bloc	MSP3-K1	PF10_0348	PF10_0348 N-term		
6	BLANK	MSP1 bloc	MSP1 bk2 '	MSP1 bloc	MSP1 bloc	MSP2 [62-3	Surfin4.2 [I	Surfin4.2 [(	Surfin4.2 [3	MSP2 [120	MSP2 [15-3	BLANK		
7	MSP2 [17-3	MSP2 [5-3(	GAMA-A	PF10_0355	PF10_0355	MSP DBL2	MSP6 A(1)	MSP6 B	GST	MSP3-3D7	Etramp 14	HSP40 Ag 1		
8	Etramp 5 A	Etramp 2 A	PF10_0351	ACBP	PfSEA-1	PV1	Pf3D7_04C	REX 3	MSRP 5B	MSRP2B	Etramp 5 A	BLANK		
9	PgP3	PF3D7_14:	CT110	MR321	MR152	Etramp 10	EBA175RII	MSP 4	MSP 5	gSG6_prot	GEXP18	CSP		
10	MSP7	CT706	GLURP R0	MSP1 bloc	Tg	HSV2	CMV	VZV	YFV-NS1	BLANK	DENV1-NS	DENV2-NS1		
11	DENV3-NS	DENV4-NS	WNV-NS1	TBEV-NS1	Zika	Rh5	T. cruzi (Tc	Pertussis JI	Pertussis JI	Pertussis JI	Pertussis JI	Pertussis JNIH-5 [1] *		
12	Pertussis JI	Pertussis JI	Pertussis JI	Tg_III	RubIV	MSP1-19m	MSP1-19 2	SBP1	Pf1103800	Pf1103800	Pf0826100	Hyp 2		
13	PfMSP1_19	TT	GLURP R2	Etramp 4 A	PfAMA1	PkMSP1 ag	PkSERA3 a	PkSERA3 a	SSP2	TSERA2 ag	TSERA2 ag	PKH_021580		
14	HSV1	BSA	PvDBP	PvMSP1_1	PvAMA1	PKH_0319:	PKH_0319:	Std 1 (200(	Std 2 (100(	Std 3 (50ug	Std 4 (25ug	Std 5 (12.5ug/ml)		
15	BLANK	Std 6 (6.25	SIGNAL 1	SIGNAL 2	SIGNAL 6	SIGNAL 7	PvEBP poo	PvRBP2b	PvRBP1a	PvRII	PKH_0800:	REF		
16	REF													
17														
18														

Figure 1. A typical array layout *.csv* file. This file consists of 16 rows of 12 cells, with no row exceeding 12 cells. Note the 'REF' (reference) and 'Std…' (standard curve) spots. Also note that empty cells can be included, as in row 16, as these will be interpreted as blank 'buffer' spots by the printer.

   2.2. Note that in a printed array of 24 spots across, the first row of 24 is created by the printing of the first 2 rows of 12 in this array layout *.csv* file. Row A1-L1 is

printed from **right to left** starting in the top right corner of each array. Subsequently, row A2-L2 is printed from **right to left** to the immediate left of row A1-L1 (Figure 2a-b).

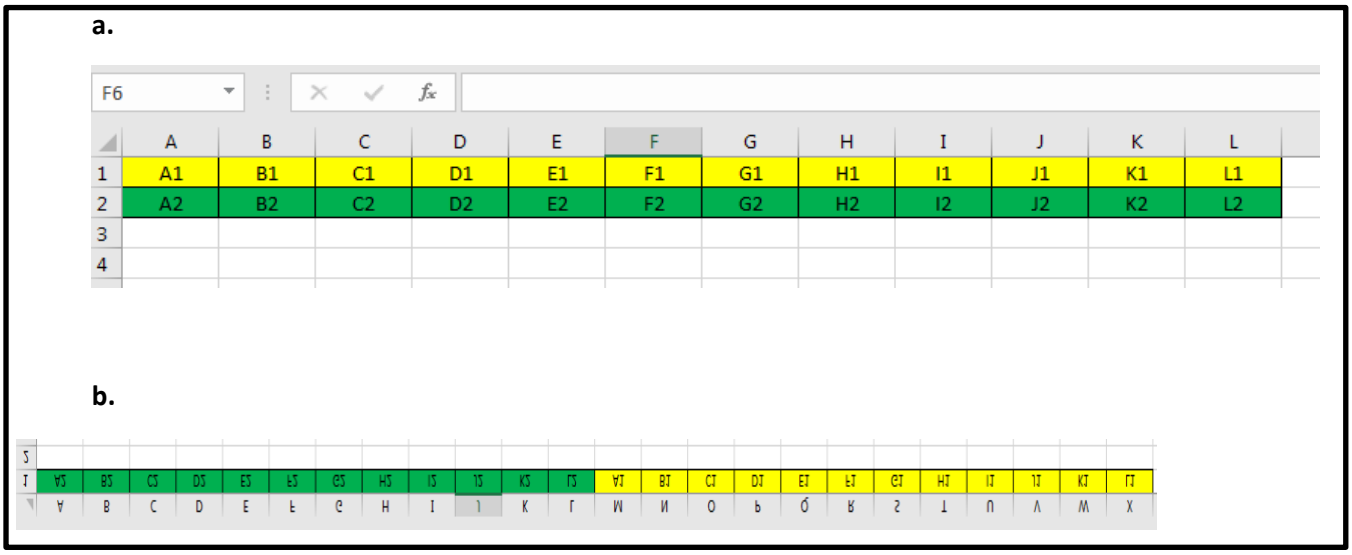


Figure 2. A comparison between **a)** The initial layout of antigens as written in the array layout *.csv* file, and **b)** The actual printed layout of spots as they would be found on the finished array.

2.3. This explains the positioning of the 'REF' spots in cells L1 and A2 (Figure 3). Using this layout, each array will be printed with a 'REF' spot in the top right and top left corner.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	MSPDBL2 (	HSP40 Ag 3	ACS5 Ag 2	ACS5 Ag 4	ACS5 Ag 3	Etramp 2 A	MSP2 Dd2	MSP2 CH1	MSPDBL1 (	MSP DBL1	MSPDBL2 (	REF	
2	REF	H101	PF10_0351	EPF1v2	EBA140 RI	MSP1 bk 2	MSP2 [15-3	EBA175 RI	EBA181 RI	MSP2 [5-6(	MSP2 [15-3	MSP2 [47-34A]	

Figure 3. The positioning of REF spots (highlighted yellow) at L1 and A2 result in a REF spot being printed in the top right and top left of each array

3. Once the array layout *.csv* file has been created, open the Command Centre for Marathon software. Go to 'File' > 'Generate Microplate Map File' > 'Browse' and select the array layout file. Click 'Generate' and a new wellplate map file (*.csv*) and an input file (*.txt*) will be created in the same folder as the array layout file. This file provides a map as to which well each antigen should be loaded in to achieve the desired printed layout.

Microplate preparation

1. Dilute each antigen/protein to **2X the desired printed concentration** in PBS (e.g. to print a spot at 0.1mg/ml, dilute to 0.2mg/ml). This allows for a 1 in 2 dilution of antigen in printing buffer.
2. Using a multichannel pipette, load each required well of the 384-microwell plate with half the desired final volume, with glycerol printing buffer (e.g. for a final well volume of 20µl, add 10µl of glycerol printing buffer.
3. Then, according to the wellplate map generated in 'Array design' step 3, add antigen/protein to each well at a volume equal to the volume of glycerol printing buffer already in the well (e.g. add 10µl of antigen to 10µl of glycerol printing buffer). In wells marked 'buffer', add one part printing buffer to one part PBS. Mix **very thoroughly** by pipetting up and down multiple times.
4. Once all of the wells have been loaded, spin the plate down at 1800 xg for 5 minutes and check that no bubbles remain in any wells. If bubbles remain spin again and check. Persistent bubbles can be popped with a pipette tip. Fit a septa to the plate to prevent evaporation.

Printing

1. Ensure that the inside of the printer is clean, free from dust and contaminants. Use a microfiber cloth to wipe down surfaces, especially areas which the print head will pass over.
2. At least half an hour before printing, switch on the printer and atmospheric controller to allow conditions inside the printer to equilibrate.
3. Open the 'Command Centre for Marathon' software and go to 'Options' > 'Initialise Instrument'. Follow on-screen instructions.
4. Check that the waste system buffer bottle is empty and that the clean system buffer bottle is sufficiently full of glycerol system buffer **(N.B. system buffer is different to the printing buffer used in the 384-microwell plate).** If emptying/refilling needs to be carried out, go to 'Options' > 'Fill/Empty System Buffer Bottles'. Follow on-screen instructions.
5. Carry out system maintenance and checks
    5.1. Perform a JetSpyder purge – Go to 'Options' > 'JetSpyder Purge'; follow on-screen instructions (installed JetSpyder is 12 capillary). Repeat for a total of 4 purges.
    5.2. Perform a print head purge x4 – Go to 'Options' > 'Print head purge' > 'x4'. Follow on-screen instructions
    5.3. Print a test slide – Go to 'Options' > 'Print Test Slide'. Follow on-screen instructions. If test slide fails, repeat print head purge x4. If test slide continues to fail, nozzle suppression may be required (see 'Command Centre for Marathon' help).
    5.4. Perfrom a 'K' test procedure – Go to 'Options' > 'K Test Procedure'. Follow on-screeen instructions. If 'K' test fails (<220.000kPa), carry out 'High Pressure Purge' procedure.
6. Create a new print run file (*.prp*). In the 'Command Centre for Marathon' software go to 'File' > 'New' > 'Print Run (slides)'

6.1. Source Microplate Properties tab

6.1.1. Select 'Standard 384' from source microplate type

6.1.2. Select the number of source microplates used

6.1.3. Select the total number of wells filled in the last plate (including buffer spots – must be a multiple of 12)

6.1.4. If required, add an additional cleaning routine, and add a plate(s) loaded with desired buffer according to the layout and instructions on-screen.

6.2. JetSpyder Properties tab

6.2.1. Ensure 'Low Volume – 12 Samples' JetSpyder is selected

6.2.2. Nozzles can be suppressed if required

6.2.3. Option for single pass printing can be selected if required

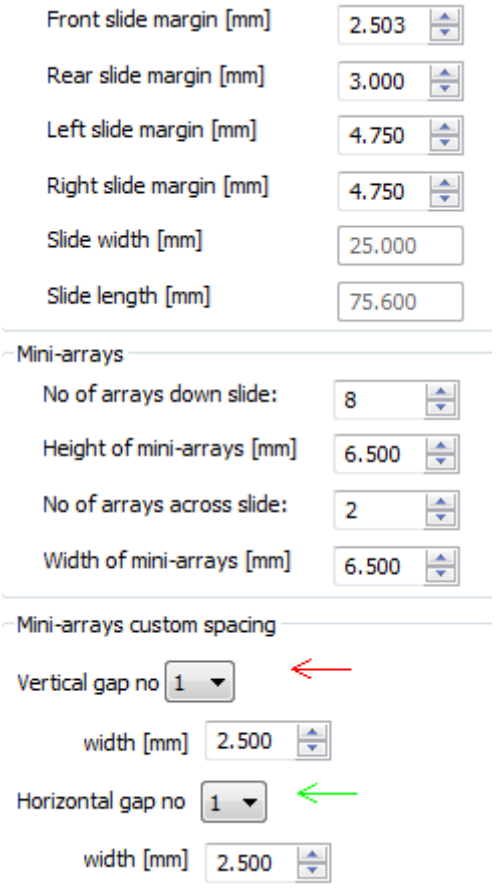
6.3. Slide Properties tab

6.3.1. Set number of trays

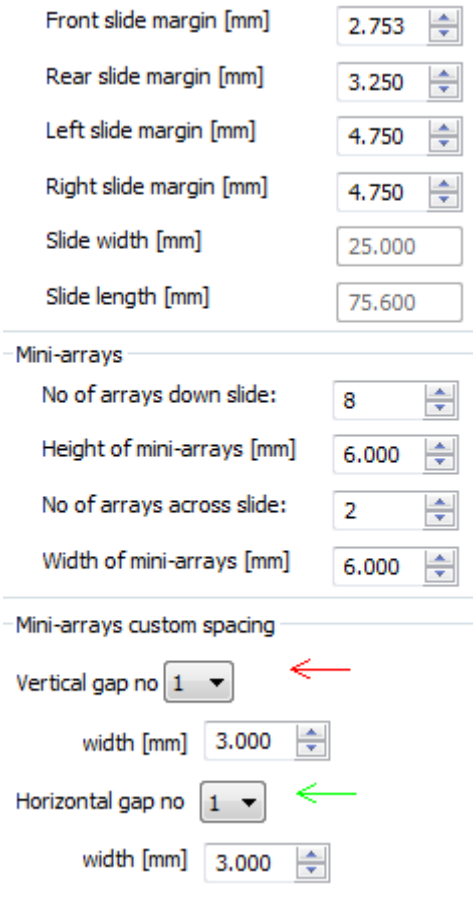
6.3.2. Set number of slides in the last tray – note that no spaces can be left between slides

6.3.3. Set up printing area parameters. For 16-pad Oncyte NOVA slides, use the following settings:

a) For pads 6.5 x 6.5mm



b) For pads 6.0 x 6.0mm



6.4. Spot Properties tab
   6.4.1. Select desired number of spots
   6.4.2. Select gap and pitch between spots. Spacing of 0.230mm is recommended
   6.4.3. Select the number of repeats, e.g. a repeat of 2 will mean every spot is printed twice in every array. Note how this will alter the layout of the printed array in the 'Repeat Layout Preview'. This can be altered by choosing different options in the 'Print mode' settings. Also note that increasing the number of repeats may require repeat pick-ups of antigen/protein depending on the number of slides and arrays to be printed, so ensure adequate volume is available.
   6.4.4. Select regular or hexagonal printing
   6.4.5. Select 'Centred' from the 'Centring' option
   6.4.6. Note that a warning will appear if your chosen settings would force spots to be printed outside the printable area.

6.5. Overview tab
   6.5.1. Check that the number of arrays and spots are as expected. Check that spots will not be printed too close to the edges of the printable area as this is not desirable.

6.6. Save the created *.prp* file

7. Load the microplate(s) and blank slides in the printer. Go to 'Options' > 'Load/Unload Microplates and Slides'. Place the microplate(s) (and wash plate(s) if selected) according to the positions indicated in the *.prp* file previously created. Remove the required number of slide trays, and load the required number of slides into sequential positions, taking care not to touch the slide surfaces. Ensure that all slides are completely flat (no raised edges) and positioned snugly in each slide recess, pushed right to the back. Use a compressed air duster to blow off any dust and contaminants on the slides before placing the loaded tray back into the printer. Close the doors and follow the on-screen instructions.
8. Load the *.prp* file through 'File' > 'Load' > 'Print Run (slides)' and make a final check of all settings. Navigate to the 'Options (Start)' tab and ensure that 'Print Run Options' is set to 'Using source microplates'. Click 'START PRINT RUN'. You will have 60 seconds to make a final check that all slides and microplates are ready and in position. Printing will then commence.
9. Once printing is complete, leave slides inside the printer for at least 6 hours with the atmospheric controller still on. After this, to shut down the printer got to 'Options' > 'Shut Down Instrument'. Remove the slides, and store them in boxes at 4°C in a vacuum dessicator.

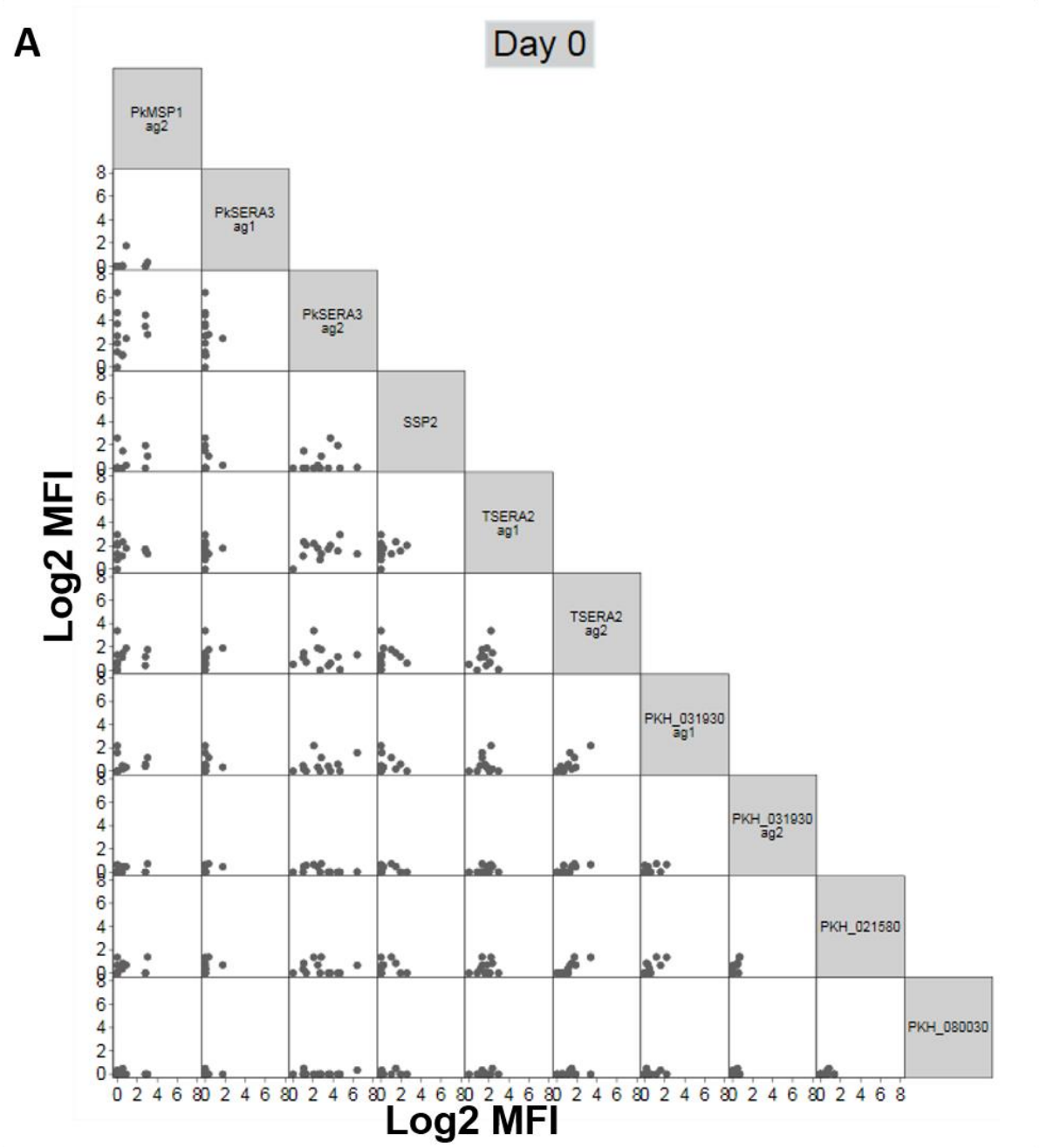
Creation of a labelled .gal file

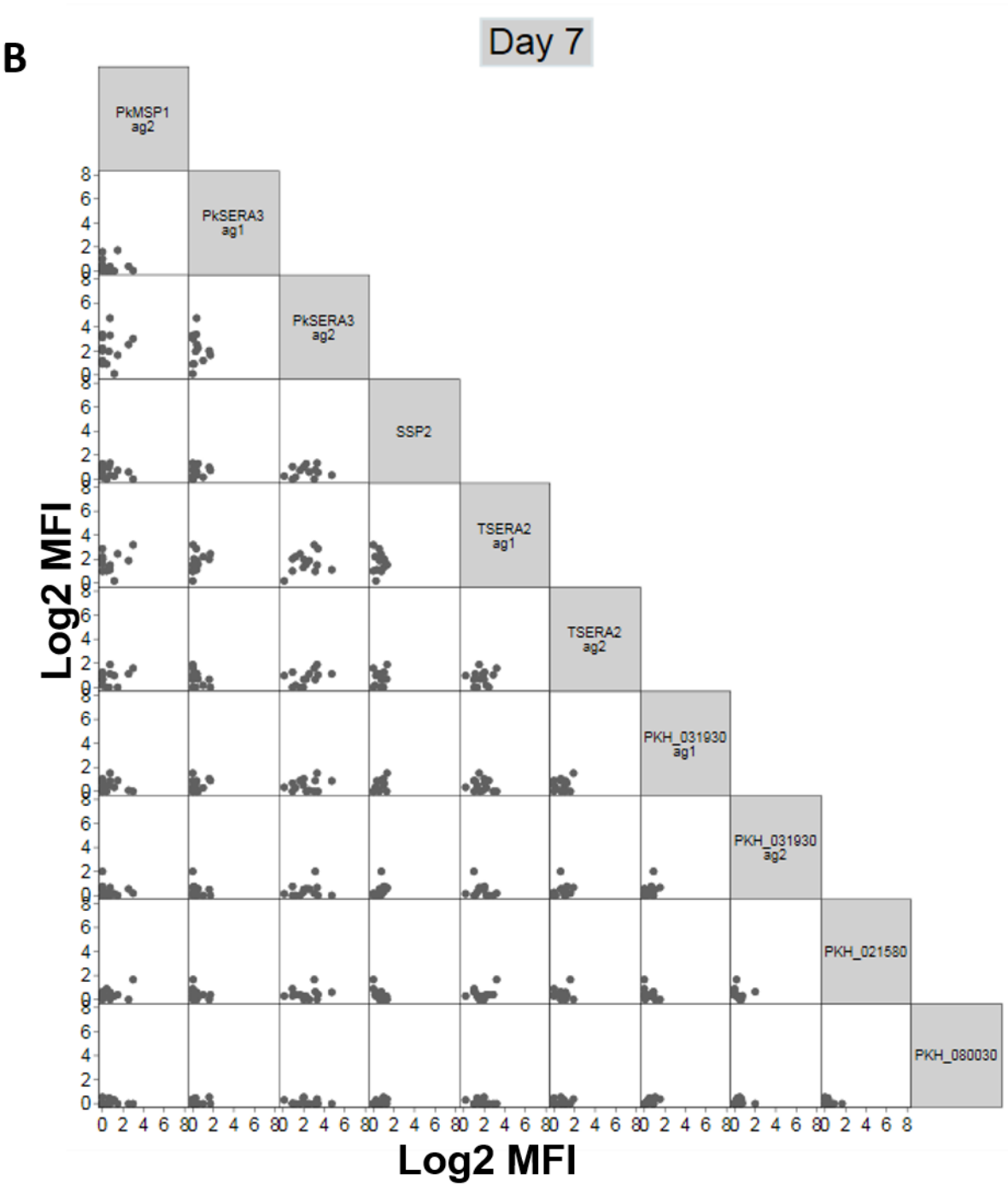
1. In the 'Command Centre for Marathon' software, go to 'File' > 'Merge Samples and GAL Files'. For the 'GAL/CSV file' option, browse to the *.gal* log file for the completed print run in C: drive > Programme Files (x86) > Arrayjet > Marathon software > Log files. For the 'Samples File' option, browse to the *.txt* file created in step 3 of Array Design methods.

2. Process the files and save the resulting labelled *.gal* file (generated in C: drive > Programme Files (x86) > Arrayjet > Marathon software > Log files) to the scanner computer for image analysis.

APPENDIX VII: CHAPTER 6 SUPPLEMENTARY DATA

Supplementary figures for Chapter 5: High throughput serological profiling of antibody isotypes in Malaysian hospital case samples using a bespoke protein microarray platform.




B
Day 7
PkMSP1 ag2
PkSERA3 ag1
PkSERA3 ag2
SSP2
TSERA2 ag1
TSERA2 ag2
PKH_031930 ag1
PKH_031930 ag2
PKH_021580
PKH_080030
Log2 MFI
Log2 MFI

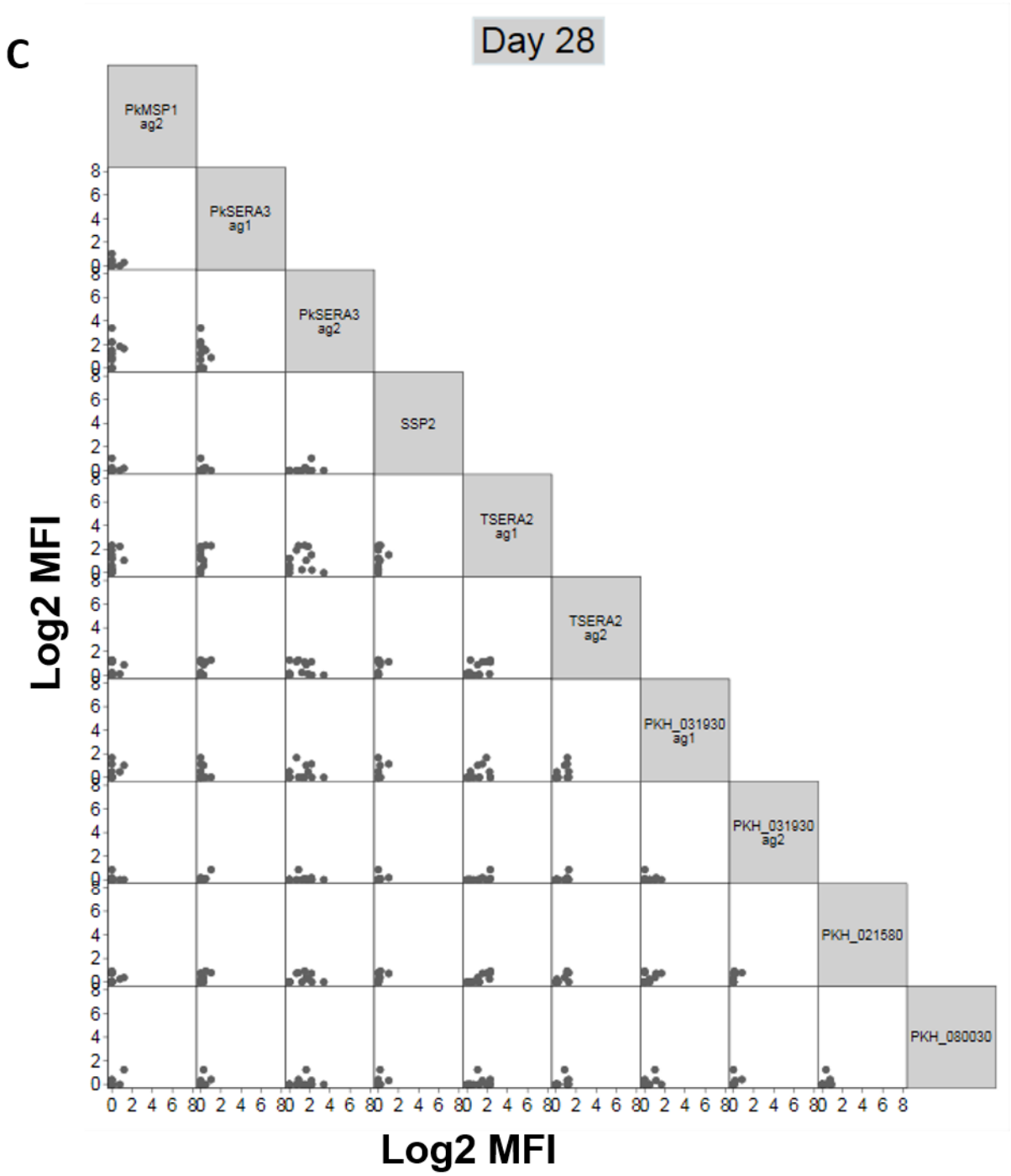
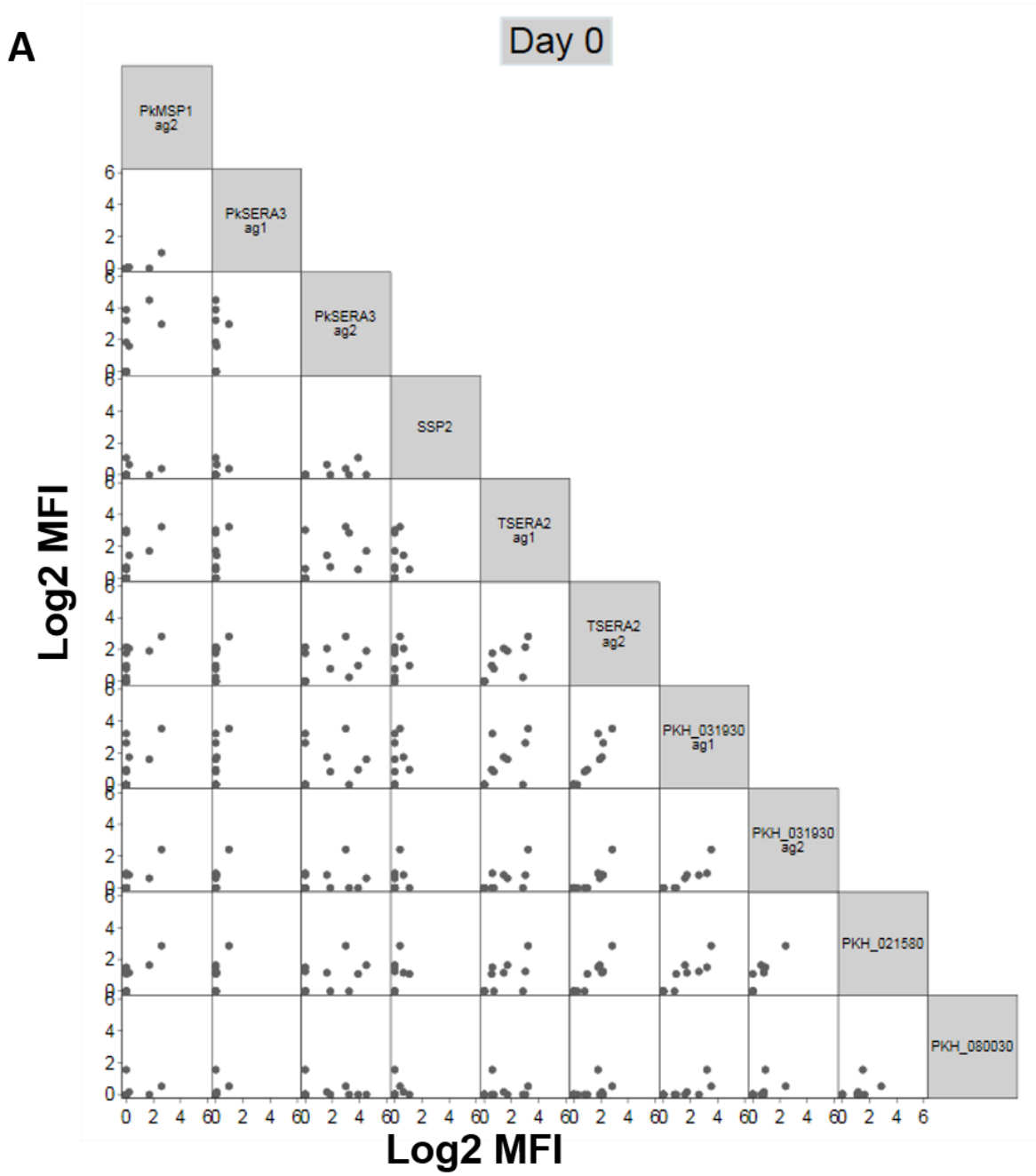
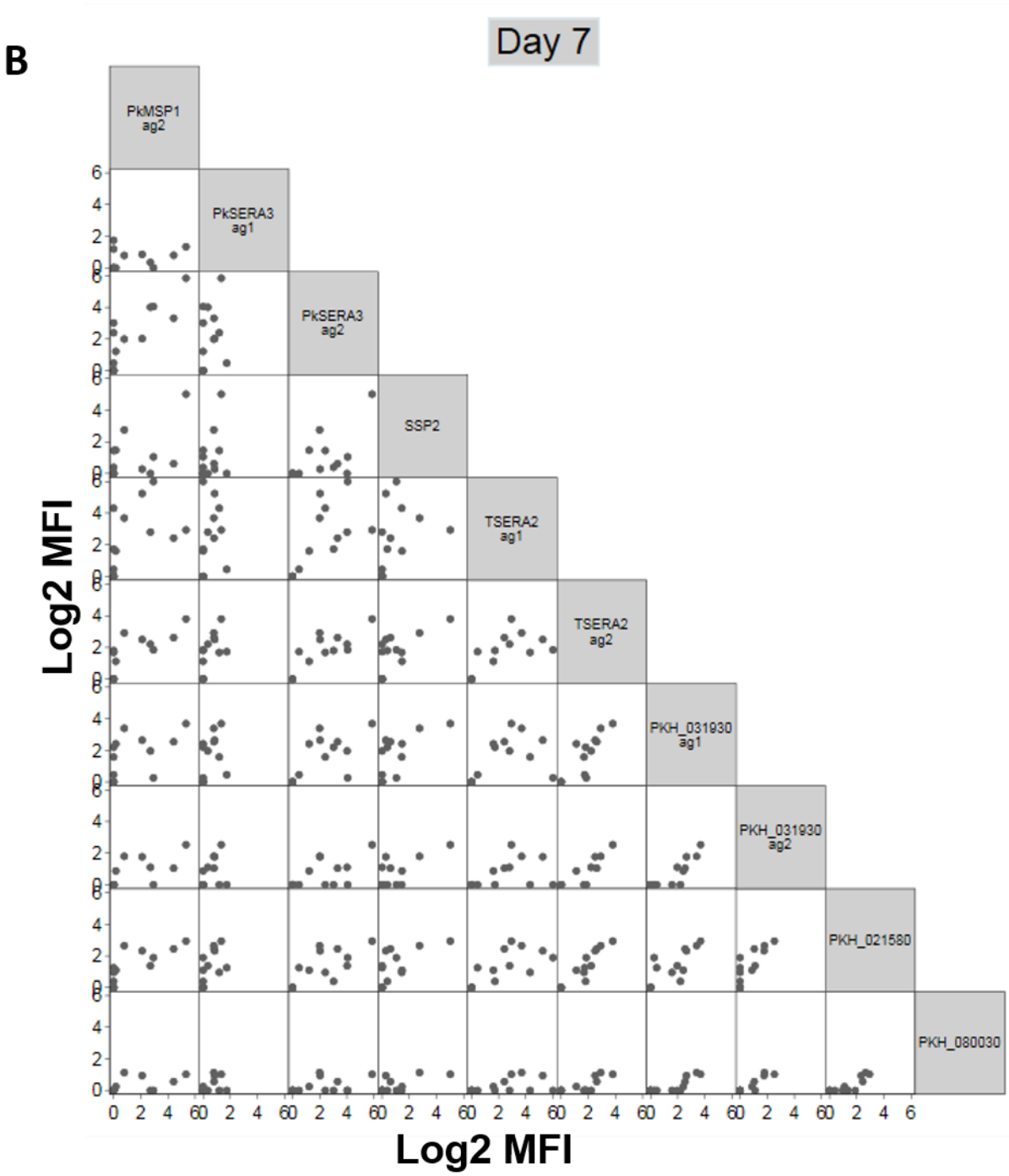


Figure 1 IgM isotype reactivity matrices towards *P. knowlesi*-specific antigens in Malaysian clinical samples at days 0, 7 and 28 after diagnosis and treatment. Data is shown as Log2 MFI ratio on both axes. For each scatterplot square, the matrix is read with the y axis showing the top antigen and the x axis showing the bottom antigen.


A
Day 0
PkMSP1 ag2
PkSERA3 ag1
PkSERA3 ag2
SSP2
TSERA2 ag1
TSERA2 ag2
PKH_031930 ag1
PKH_031930 ag2
PKH_021580
PKH_080030
Log2 MFI
Log2 MFI


B
Day 7
PkMSP1 ag2
PkSERA3 ag1
PkSERA3 ag2
SSP2
TSERA2 ag1
TSERA2 ag2
PKH_031930 ag1
PKH_031930 ag2
PKH_021580
PKH_080030
Log2 MFI
Log2 MFI

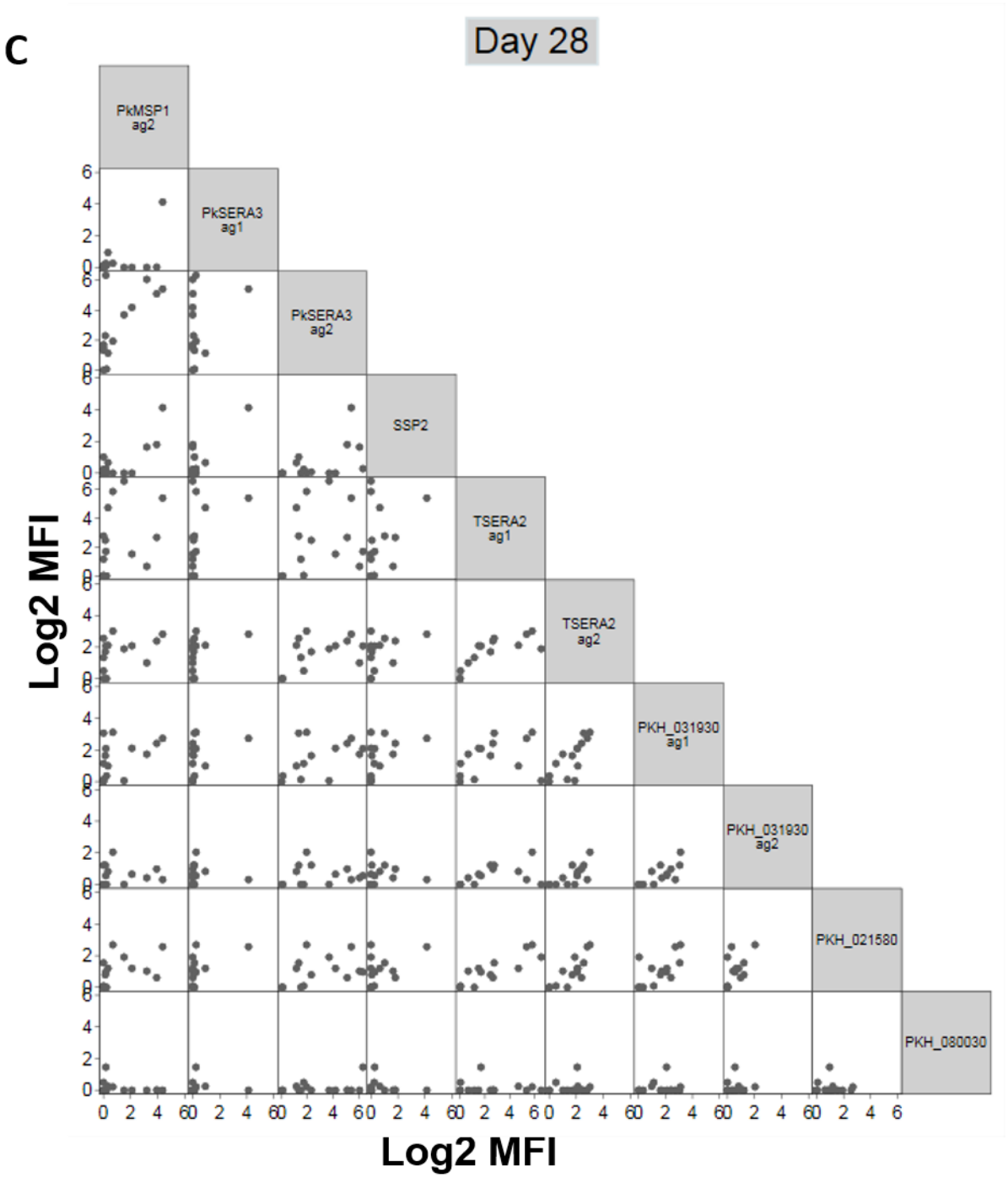
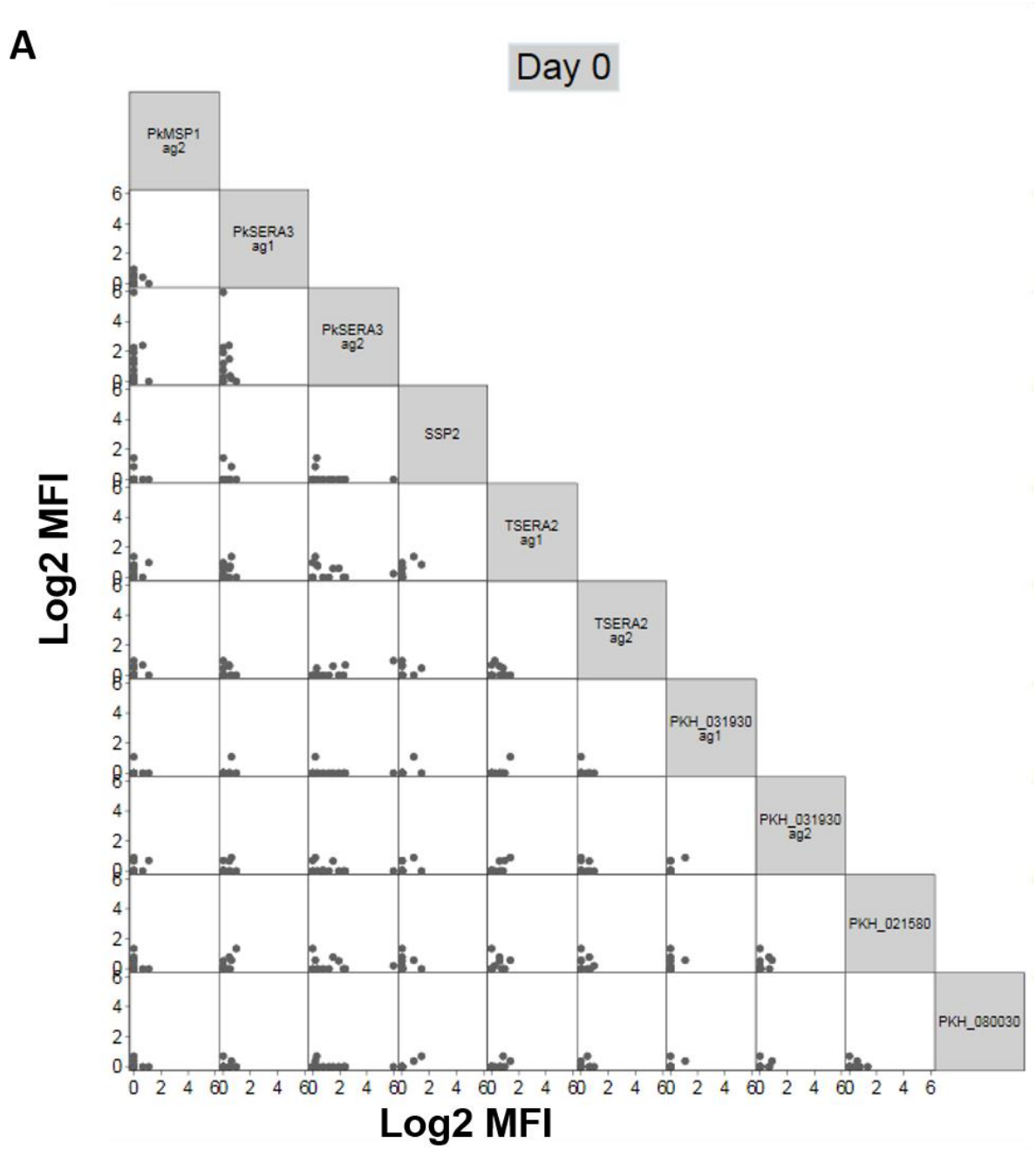
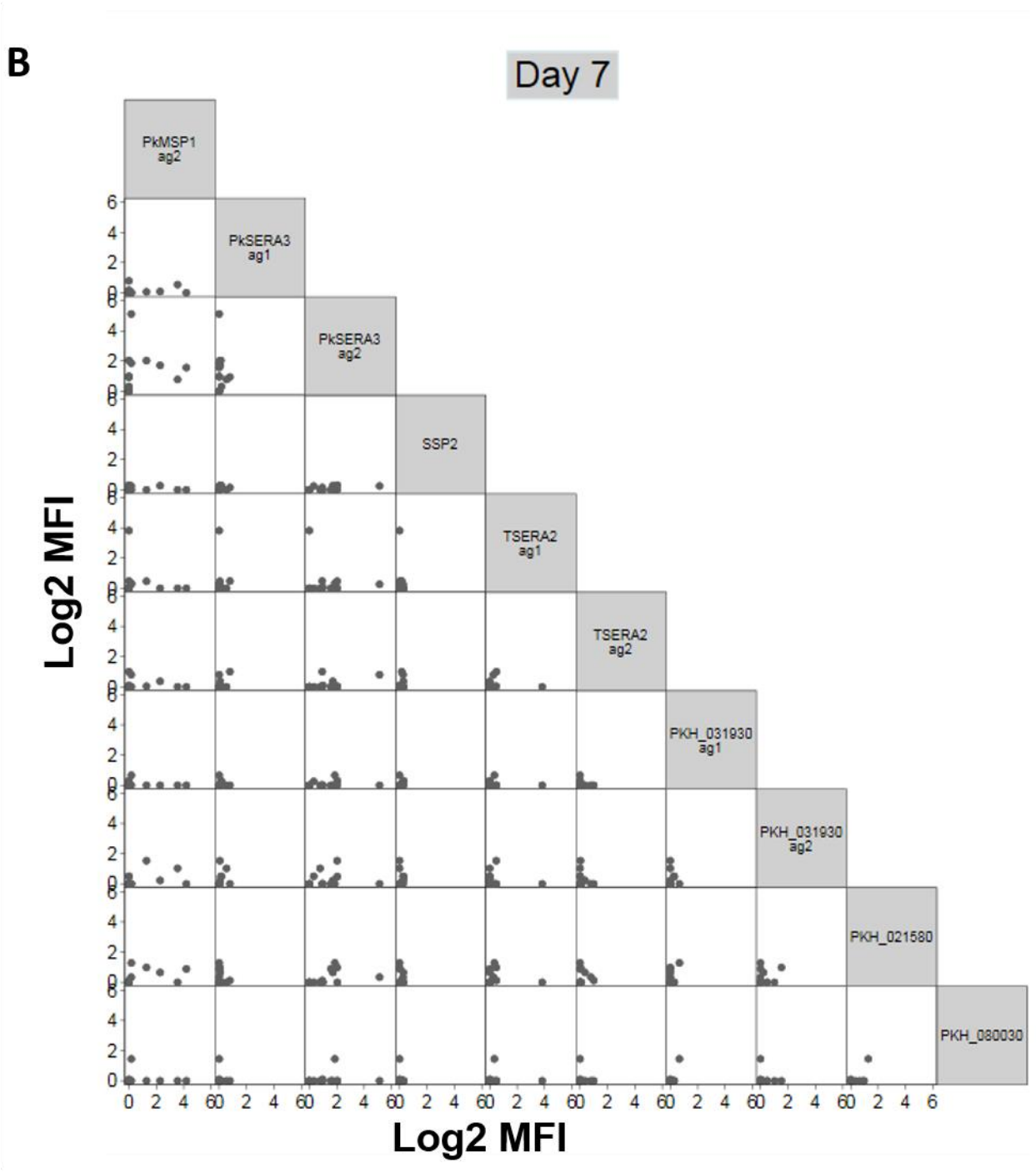


Figure 2 IgG isotype reactivity matrices towards *P. knowlesi*-specific antigens in Malaysian clinical samples at days 0, 7 and 28 after diagnosis and treatment. Data is shown as Log2 MFI ratio on both axes. For each scatterplot square, the matrix is read with the y axis showing the top antigen and the x axis showing the bottom antigen.


A
Day 0
PkMSP1 ag2
PkSERA3 ag1
PkSERA3 ag2
SSP2
TSERA2 ag1
TSERA2 ag2
PKH_031930 ag1
PKH_031930 ag2
PKH_021580
PKH_080030
Log2 MFI
Log2 MFI


B
Day 7
PkMSP1 ag2
PkSERA3 ag1
PkSERA3 ag2
SSP2
TSERA2 ag1
TSERA2 ag2
PKH_031930 ag1
PKH_031930 ag2
PKH_021580
PKH_080030
Log2 MFI
Log2 MFI

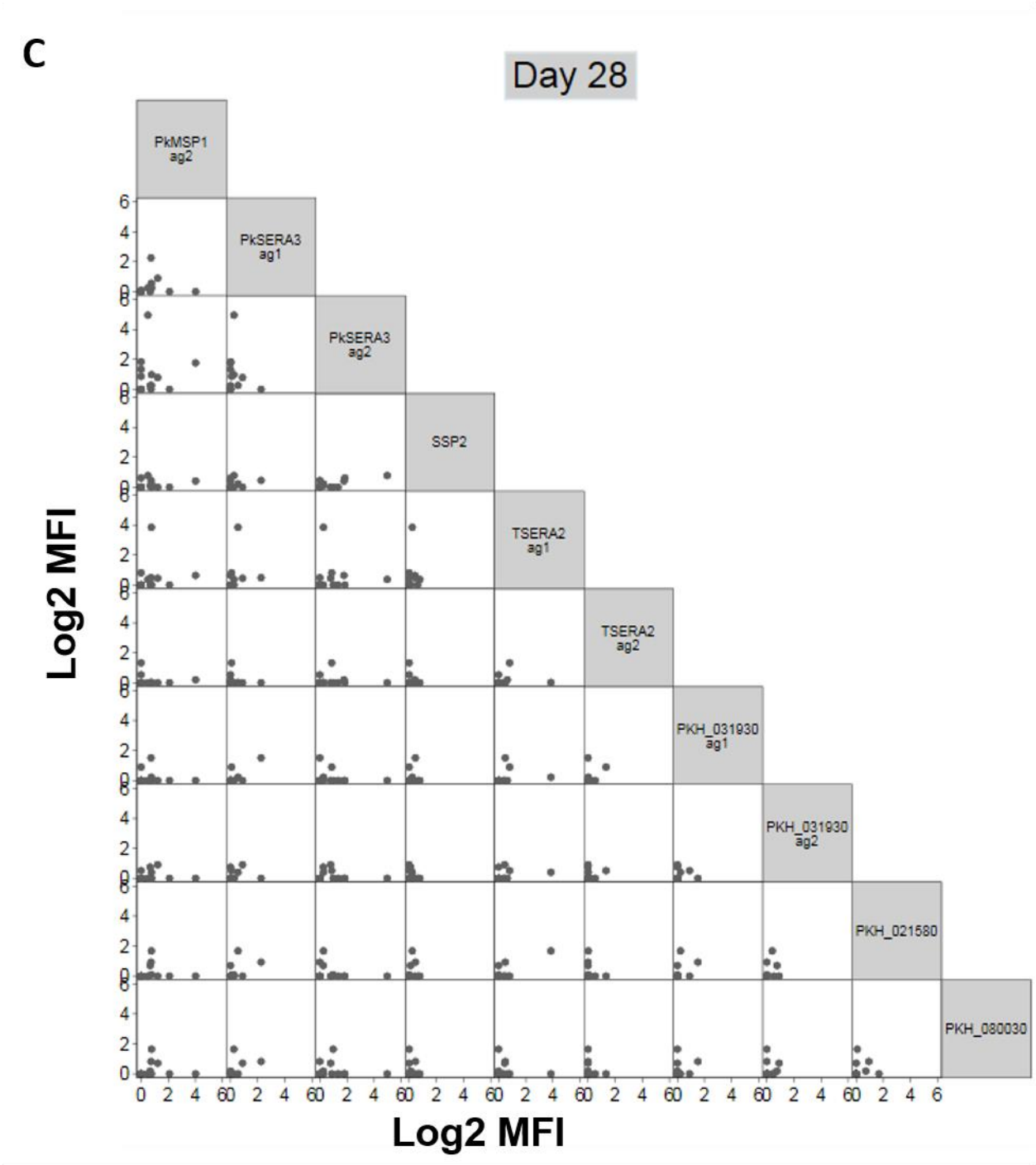


Figure 3 IgA isotype reactivity matrices towards *P. knowlesi*-specific antigens in Malaysian clinical samples at days 0, 7 and 28 after diagnosis and treatment. Data is shown as Log2 MFI ratio on both axes. For each scatterplot square, the matrix is read with the y axis showing the top antigen and the x axis showing the bottom antigen.

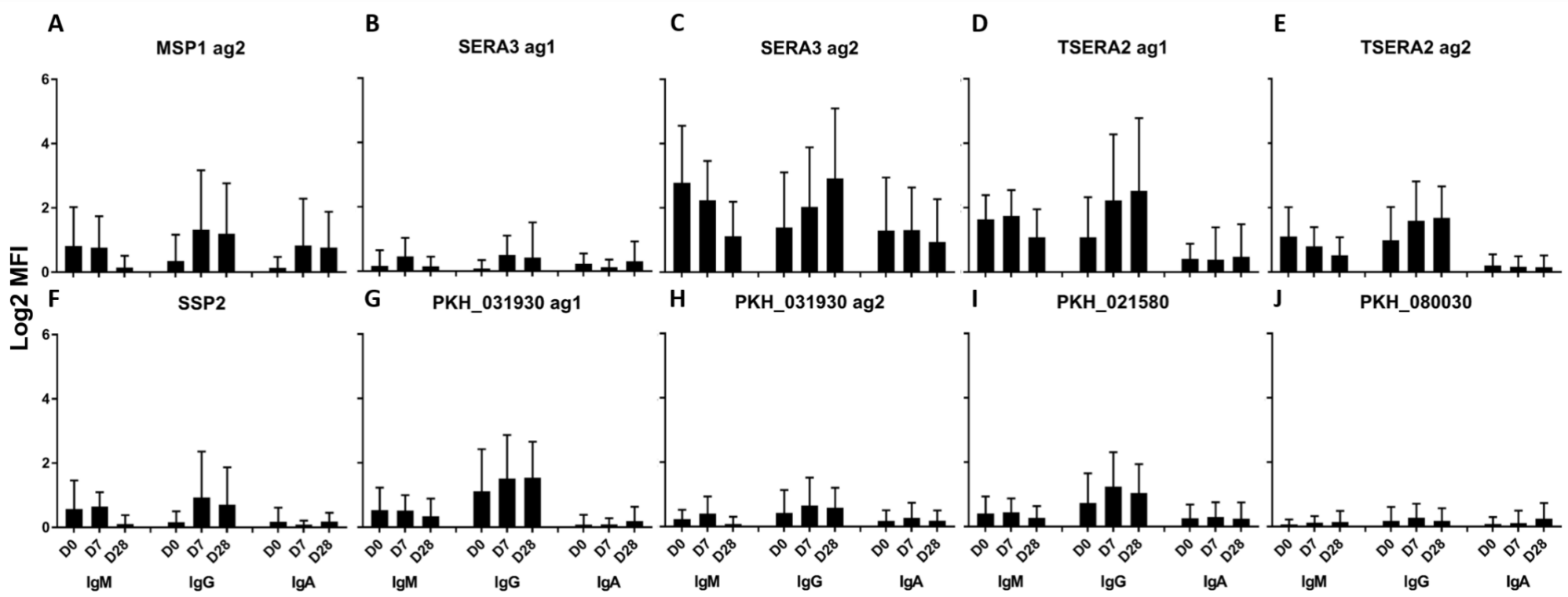


Figure 4 P. knowlesi-specific antigen reactivity towards IgM, IgG and IgA across all time points for each antigen. The data is plotted for (A) *Pk*MSP1 ag2; (B) *Pk*SERA3 ag1; (C) *Pk*SERA3 ag2; (D) *Pk*TSERA2 ag1; (E) *Pk*TSERA2 ag2; (F) *Pk*SSP2; (G) PKH_031930 ag1; (H) PKH_031930 ag2; (I) PKH_021580 and (J) PKH_080030. The results are presented as geometric mean log2 MFI ratio readings. The error bars show the standard deviation.

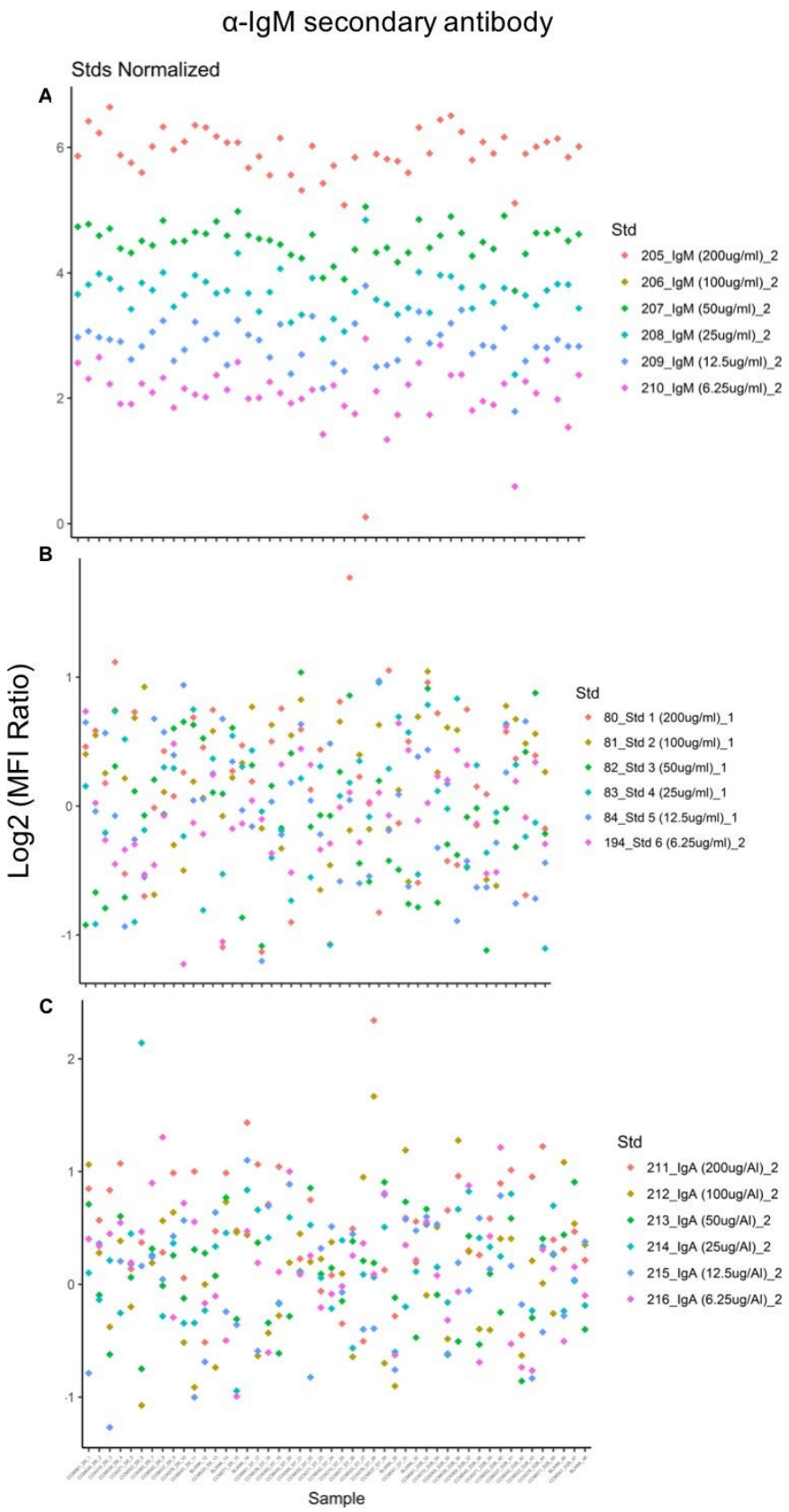


Figure 5 Specific anti-IgM secondary antibody reactivity towards printed purified isotype standards. Goat anti-Human IgM Qdot® 655 secondary antibody reactivity towards purified IgM (A), IgG (B) and IgA (C) printed at different concentrations and screened with serum from a hospital clinical treatment trial. Results are shown for every array (sample) as log2 of the MFI ratio.

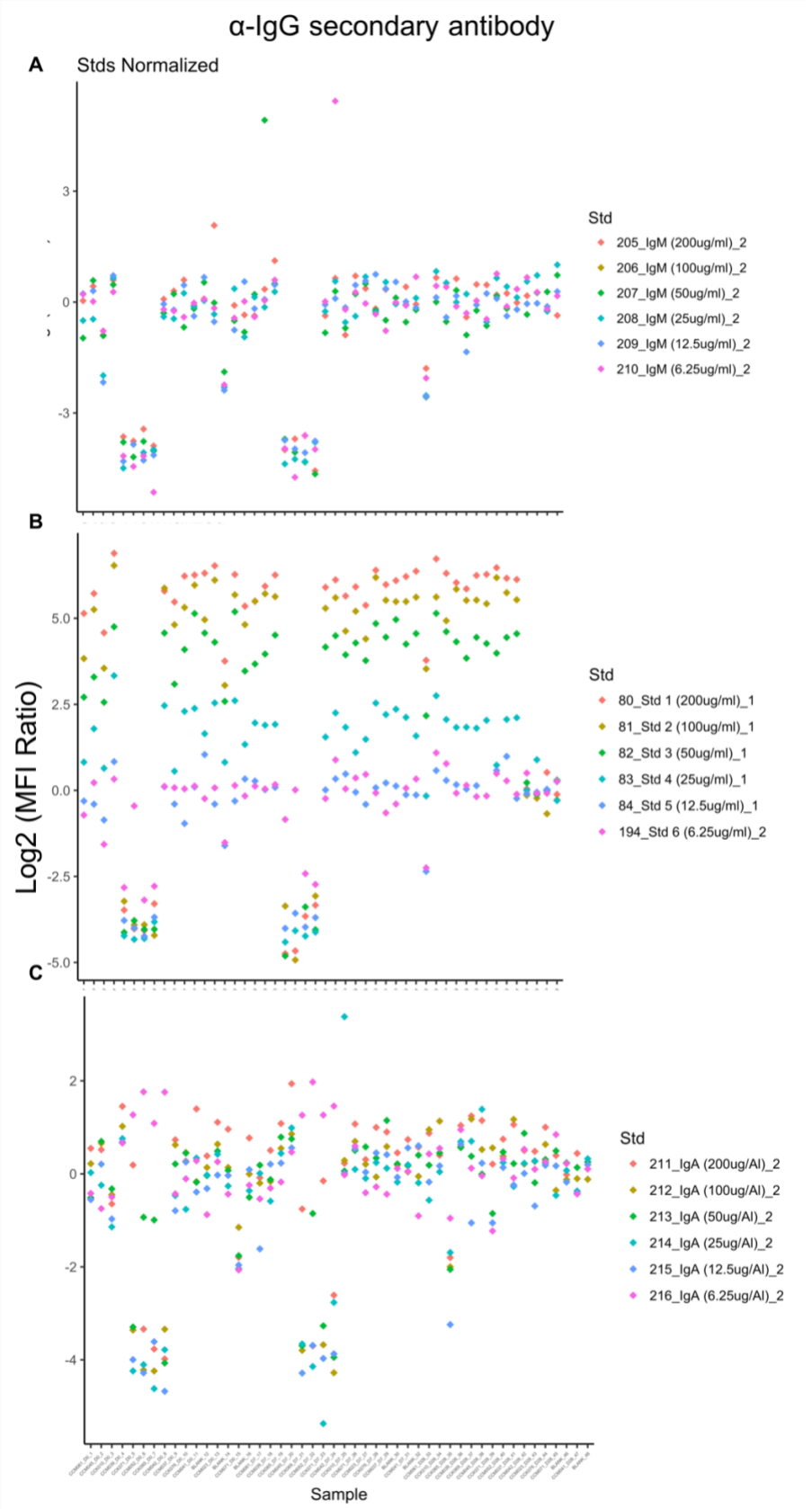


Figure 6 Specific anti-IgG secondary antibody reactivity towards printed purified isotype standards. Goat anti-Human IgM Qdot® 655 secondary antibody reactivity towards purified IgM (A), IgG (B) and IgA (C) printed at different concentrations and screened with serum from a hospital clinical treatment trial. Results are shown for every array (sample) as log2 of the MFI ratio.

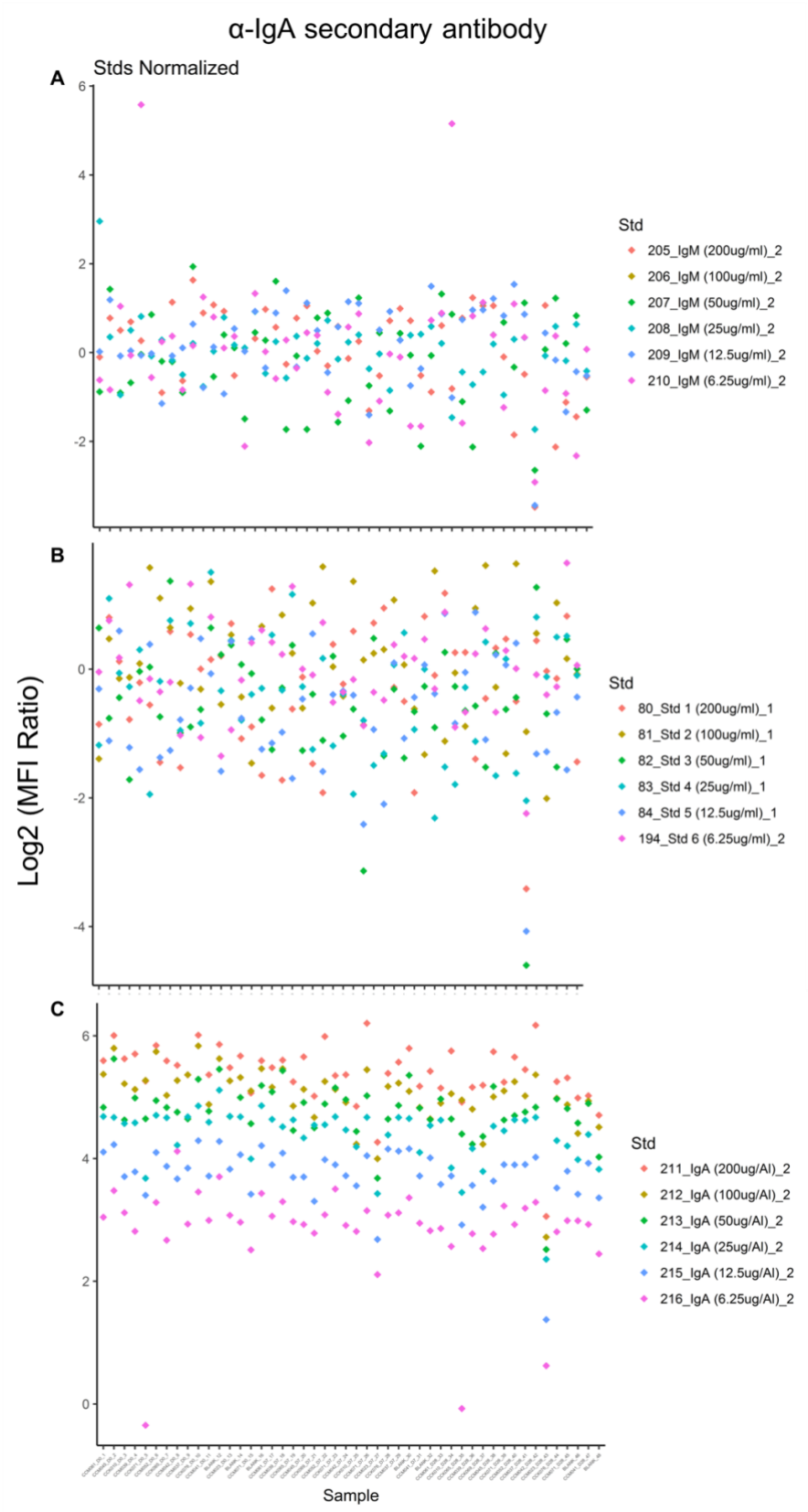


Figure 7 Specific anti-IgA secondary antibody reactivity towards printed purified isotype standards. Goat anti-Human IgM Qdot® 655 secondary antibody reactivity towards purified IgM (A), IgG (B) and IgA (C) printed at different concentrations and screened with serum from a hospital clinical treatment trial. Results are shown for every array (sample) as log2 of the MFI ratio.