

RESEARCH ARTICLE

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

Lou S. Herman¹, Kimberly Fornace¹, Jody Phelan¹, Matthew J. Grigg^{2,3}, Nicholas M. Anstey^{2,3}, Timothy William^{3,4,5}, Robert W. Moon¹, Michael J. Blackman^{1,6}, Chris J. Drakeley¹, 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

Copyright: © 2018 Herman et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

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: *PkSERA3* antigens 1 and 2, *PkSPP2/TRAP* and *PkTSERA2* 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 *PkSERA3* 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 *PkSERA3* 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 (*PkAMA1*) [14] and *PkSPATR* (secreted protein with altered thrombospondin repeat) [15] 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 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] (*PkAMA1* (PKNH_0931500), *PkMSP1* (PKNH_0728900), *PkSSP2/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–

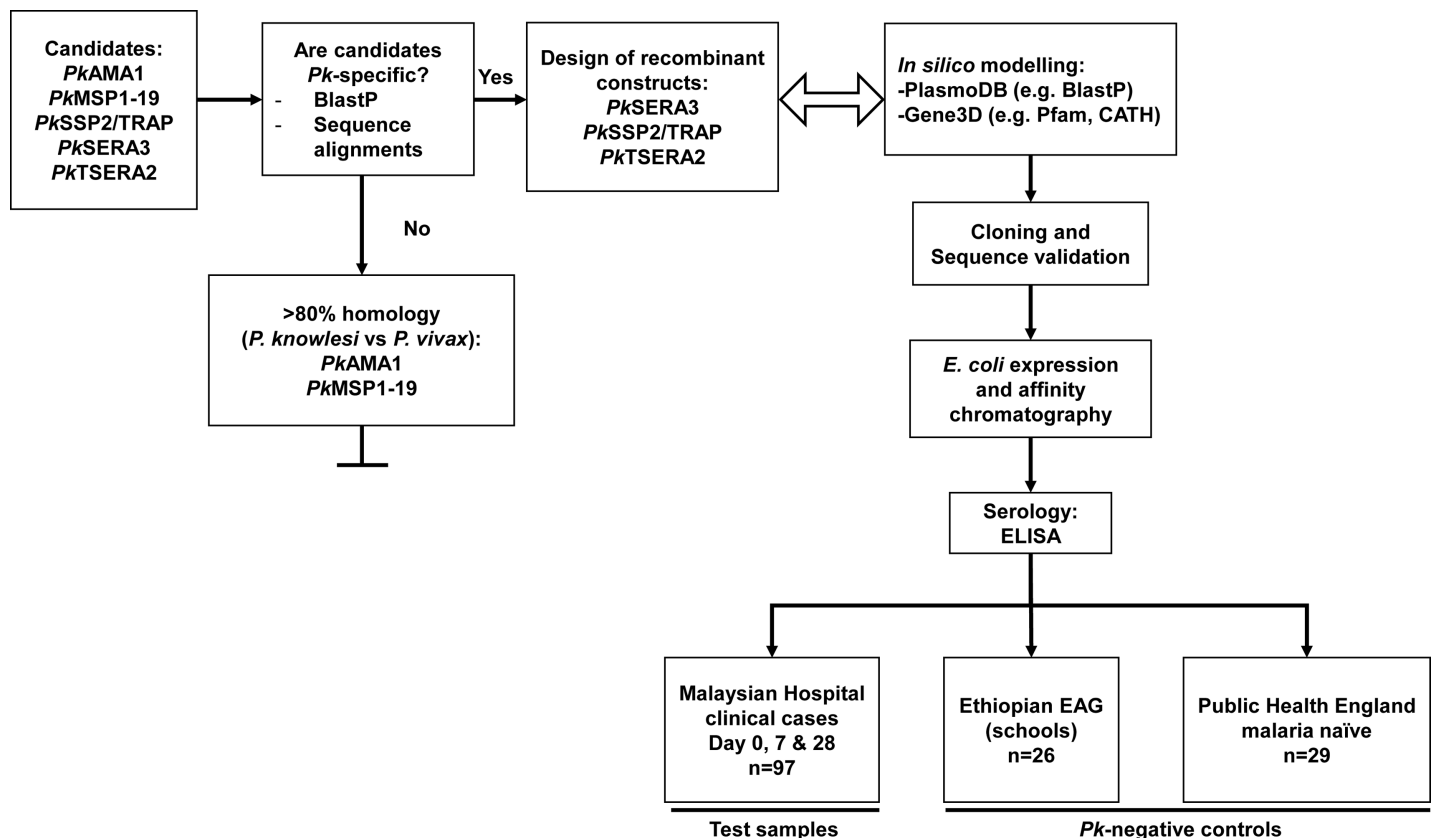


Fig 1. Flowchart summarising the experimental strategy used in the identification and validation of the *P. knowlesi*-specific candidates.

<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 *PkCTRP* (circumsporozoite protein and thrombospondin-related adhesive protein [TRAP]-related protein) and *PkCSP* (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 positions 1141–1500) and TSERA2 (PKNH_0413500; nucleotide positions 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-cloned 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 PvMSP1-19 (donated as a kind gift from Tony Holder) positive control antigen. Polyclonal rabbit anti-

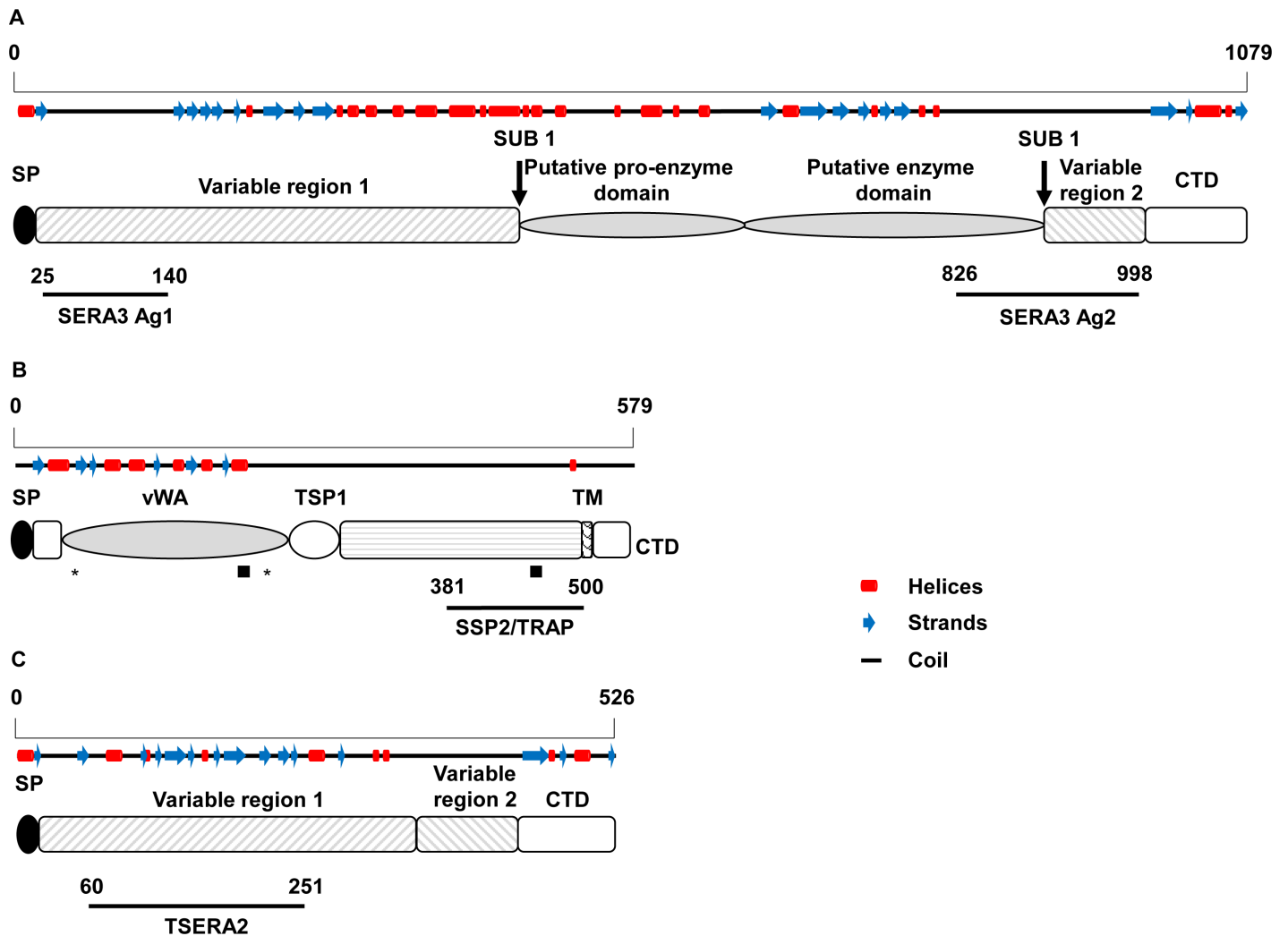


Fig 2. *Plasmodium knowlesi*-specific recombinant antigen constructs. Schematic representations for each protein are shown with key features labelled. (a) *PkSERA3* shows the location of the putative pro-enzyme and enzyme domains. The predicted subtilisin (SUB) 1 cleavage sites in relation to variable regions 1 and 2 and the cytoplasmic domain [29]. (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 asterisk or black square, respectively [30]. (c) *PkTSERA2* 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

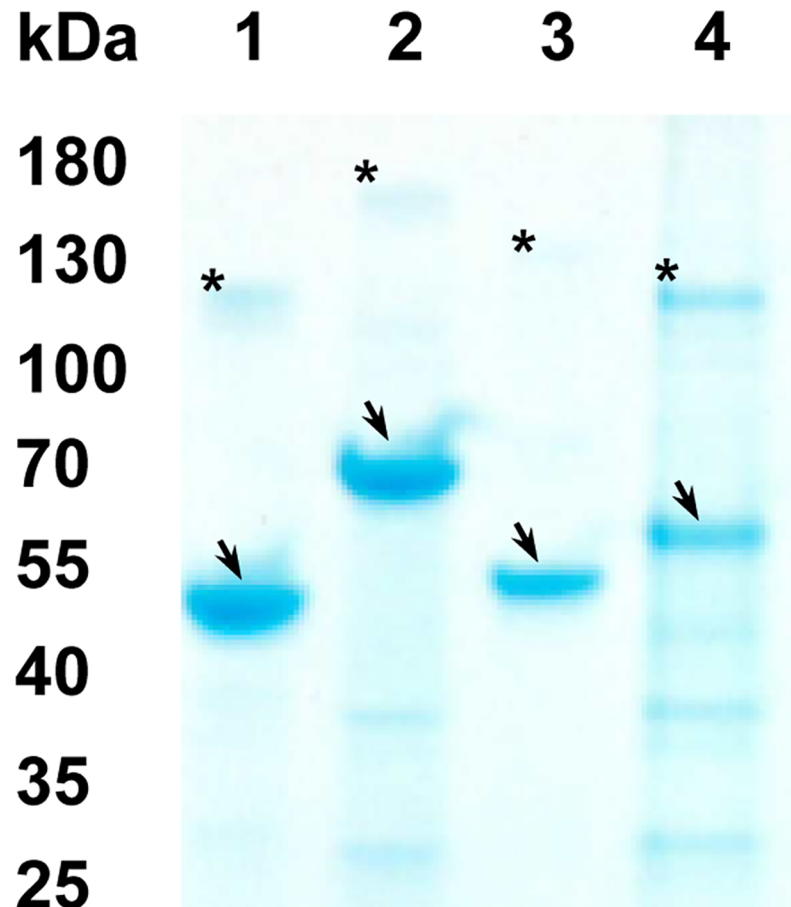


Fig 3. SDS-PAGE of purified recombinants. Lane 1: SERA3 ag1; Lane 2: SERA3 ag2; Lane 3: SSP2/TRAP and Lane 4: TSERA2 ag1. Band sizes are indicated on one side in kDa (Fermentas PageRuler Prestained Protein Ladder). Samples were ran under reducing conditions at approximately 0.4 mg/ml per lane on a 4–20% 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 PvAMA1 and PvMSP1-19 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

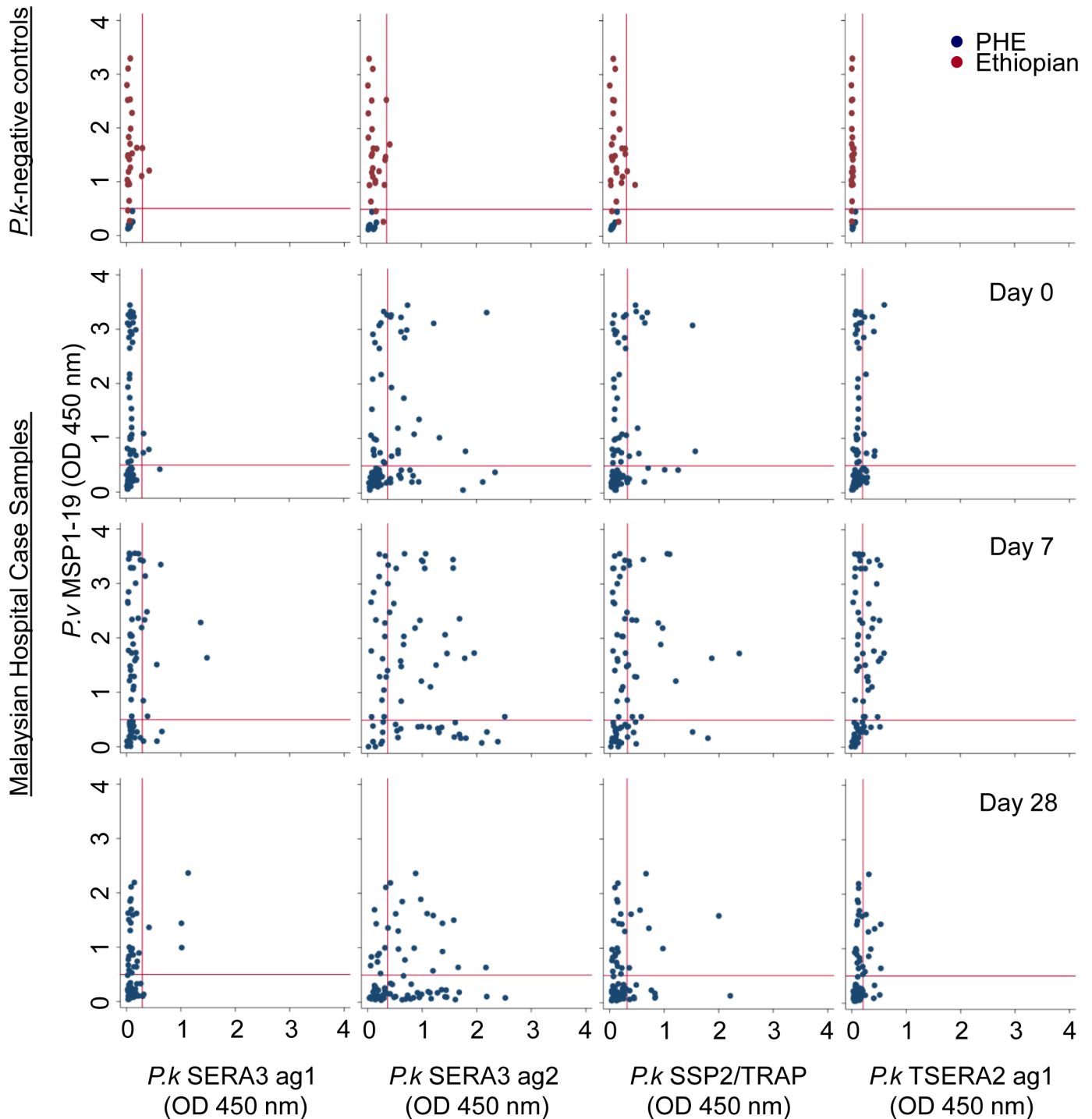


Fig 4. Endemic and *P. knowlesi*-negative sera reactivity to *Plasmodium knowlesi*-specific antigens. Scatter plots showing sera reactivity to *P. vivax* MSP1-19 with *P. knowlesi* SERA3 ag1 (column 1), SERA3 ag2 (column 2), SSP2/TRAP (column 3) and TSERA2 ag1 (column 4) antigens. Sera samples from *P. 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 \pm (3xSD)): The vertical cut off line is based on PvMSP1-19 = 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 \pm (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-erythrocytic 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 PvAMA1 (bootstrap value: 100%) and PvMSP1-19 (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

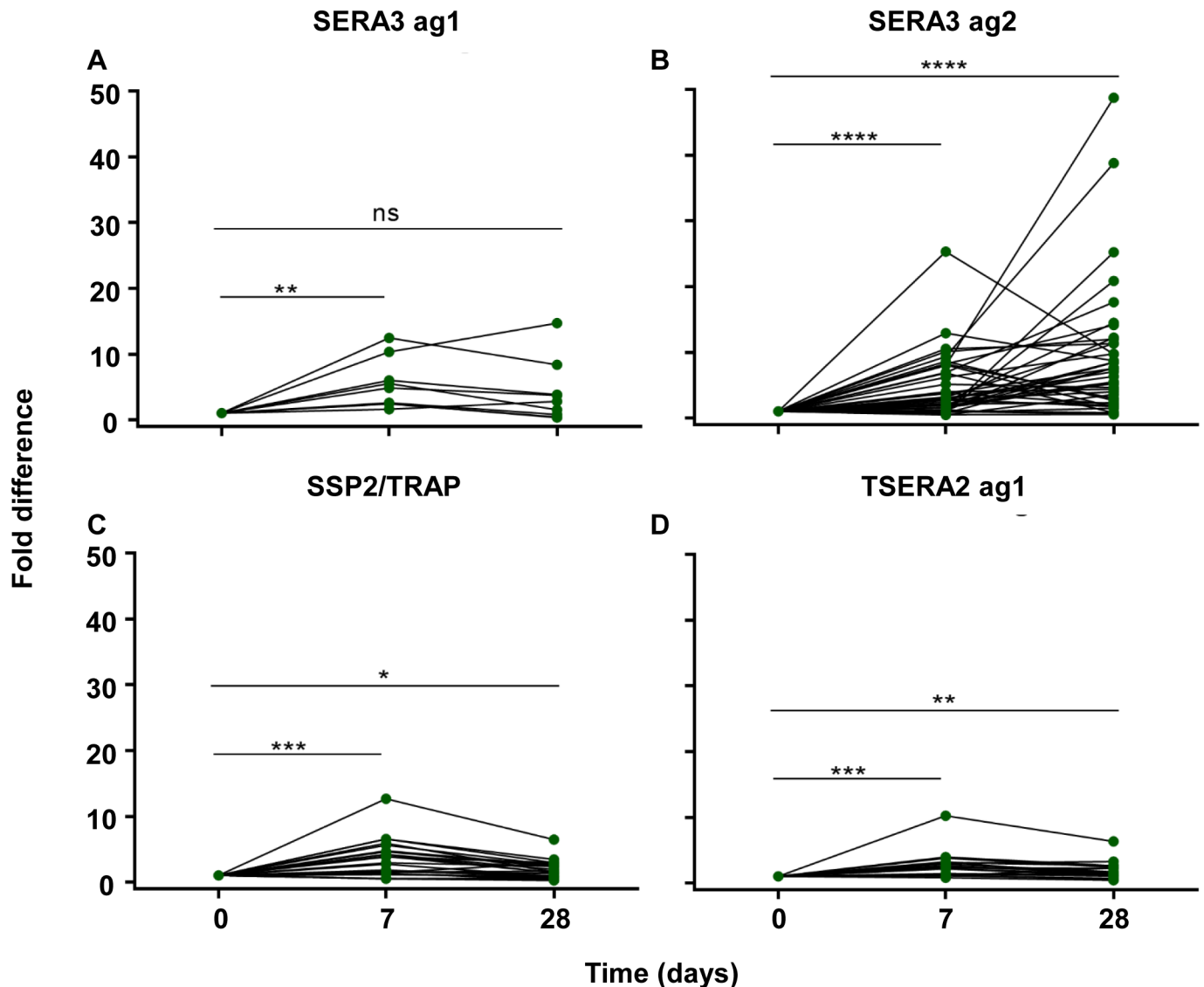


Fig 5. Serial fold increase in antibody reactivity for each antigen following treatment of knowlesi malaria. (a) SERA3 ag1, (b) SERA3 ag2, (c) SSP2/TRAP and (d) TSERA2 ag1. Asterisks indicate level of significance, ns denotes non-significant values ($p \leq 0.0001$ ****; $p \leq 0.001$ ***; $p \leq 0.01$ **; $p \leq 0.05$ * 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 hydrophathy 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 *PkCTRP* and *PkCSP* 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 *PvMSP1-19* 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, *PkSSP2* 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 *PkSERA3* 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

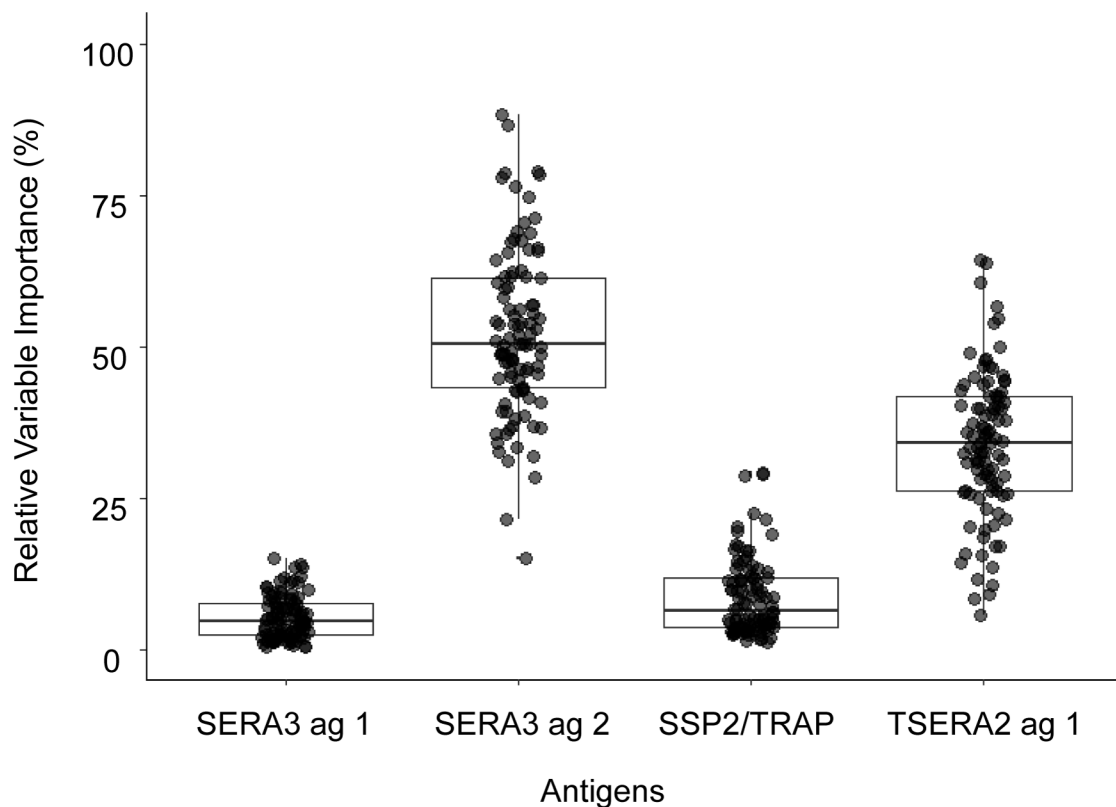


Fig 6. Relative variable importance of responses to each antigen from 100 boosted regression tree models predicting *P. knowlesi* seropositivity. Median values for the relative variable importance and interquartile ranges are shown for all antigens tested: SERA3 ag 1 (4.8%; IQR 2.5–7.8%); SERA3 ag 2 (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 [*PkCSP* [58], *PkAMA1* [59], *PkDBP* [60], *PkSPATR* [15], *PkLDH* [61], *Pk1-Cys* peroxiredoxin [62], *Pk* knowpains [63], *PkMSP1-42* [64], *PkMSP1-33* [65], *PkMSP1-19* [66], *Pk* tryptophan-rich antigens (*PkTrags*) [67], *PkMSP3* [68] and *PkSBP1* [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 *PvMSP1-19* and *PkMSP1-19*, 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 *PvMSP1-19* 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 *PkSERA 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; panel 3: TSERA2; panel 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 (Hyper-ladder 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 SERA3 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: PMC3967999.
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: PMC3967999.

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: PMC3321776.
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](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: PMC3623376.
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: PMC3381373.
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: PMC3472242.
11. Wipasa J, Suphavitai 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: PMC2824751.
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](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: PMC1231135.
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: PMC5335517.
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: PMC5104695.
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. Aurrecochea 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, Faily 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: 17609844.
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](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: PMC5301436.
 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: PMC3294643.
 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: PMC50088.
 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: PMC5050508 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: PMC3245030.
 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: PMC2656934.
 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: PMC3531217.
 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: PMC2644111.
 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: PMC4538641.
 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: PMC4109903.
 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: PMC4811561.
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: PMC3105089.
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: PMC3520923.
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: PMC3878241.
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: PMC3752745.
66. Lau YL, Cheong FW, Chin LC, Mahmud R, Chen Y, Fong MY. Evaluation of codon optimized recombinant Plasmodium knowlesi Merozoite Surface Protein-119 (pMSP-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: PMC4579084.
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: PMC4938616.
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: PMC45061513.
70. Sonaimuthu P, Cheong FW, Chin LC, Mahmud R, Fong MY, Lau YL. Detection of human malaria using recombinant Plasmodium knowlesi merozoite 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: PMC2168336.
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: PMC4671257.
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: PMC2613260.
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: PMC5291533 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, AcScl Health. The other authors have declared that no competing interests exist.