2 GENERAL METHODS

Several general methodologies were key to the completion of this thesis. In this chapter, I will describe those methods. Publication note: the methods described in sections 2.3.1, 2.3.3, 2.4.1, 2.4.3, 2.5.2, 2.6.1, 2.6.2, and 2.7.1.3 were slightly modified from a previously published manuscript (Hostetler et al., 2015). I drafted the text described in these sections, which was edited by co-authors prior to publication. I am solely responsible for the work described in sections 2.1-2.7.1, under the supervision of my PhD supervisors, Rick Fairhurst and Julian Rayner, except where noted in the text. The experiments in section 2.7.2 were performed in close collaboration with Camila Franca as specified in the text. Camila Franca performed all experiments in section 2.7.3. The results from sections 2.7.2-2.7.3 were recently published (Franca et al., 2016).

2.1 *P. vivax* schizont transcriptome sequencing

2.1.1 Schizont enrichment of *P. falciparum* samples for testing

To compare RNA extraction methods for isolating RNA from *P. vivax* field samples, mock samples of *P. falciparum* laboratory strains were used. *P. falciparum* cultures (provided by Jennifer Volz) were enriched for schizonts to compare different RNA extraction methods. Five and 10 ml cultures were pelleted (800 RCF, 4ºC, 5 min), resuspended in 10 ml of HI-FBS complete media (Table 2.4), pelleted and washed a second time. The pellet was resuspended in HI-FBS complete media to 10% hematocrit, and separated by centrifugation (1500 RCF, 4ºC, 20 min, no brakes) on a 60% Percoll® gradient (12.5 ml solution/1 ml erythrocyte pellet at 4ºC). The interface containing

schizonts was collected into an RNase-free tube with 20 ml PBS per 1 ml initial pellet. Schizonts were pelleted (800 g, 4ºC, 5 min, no brakes), washed with PBS (20 ml/1 ml starting pellet), and pelleted. RNAlater® (Ambion) was added to isolated schizonts (at 2:1 initial pellet size), and samples were stored at 4ºC overnight followed by storage at -20ºC until samples were combined for RNA extraction experiments.

2.1.2 Comparison of RNA extraction methods

Schizont-enriched *P. falciparum* parasites in RNAlater® (Ambion) were combined and divided equally (1.5 ml each) to compare 3 RNA extraction methods. All samples were initially pelleted (8000 RCF, 1 min) and RNAlater® (Ambion) supernatant was discarded. Samples 1 and 2 were extracted using the RNeasy Plus Mini kit (Qiagen) or the RiboPure Blood kit (Ambion), respectively, according to manufacturer instructions. Sample 3 was extracted using a TRIzol (Invitrogen) extraction protocol modified from Kyes, et al (Kyes et al., 2000). The schizont pellet was lysed by adding TRIzol (Invitrogen, 500 µl incubated at 37ºC, 5 min), followed by extraction with chloroform (added 250 µl, mixed well, incubated at RT, 3 min). The sample was centrifuged (8000 rpm, 4ºC, 30 min) and the aqueous layer was removed to a new tube, mixed with isopropanol (250 µl), and incubated (4ºC, 2 h). RNA was pelleted (14000 rpm, 4ºC, 30 min), and the pellet was then washed with 75% ethanol (500 µl), centrifuged (14000 rpm, 4ºC, 5 min), and the supernatant was discarded. The sample was air dried (5 min), resuspended in DEPC-H_2O (Ambion), and heated to solubilize the RNA (65ºC, 5 min). Extracted samples were all subjected to DNA digestion using the DNA-*free* kit (Ambion) according to manufacturer instructions and analyzed by Bioanalyzer® (Agilent Technologies, Inc.) using an RNA Nano Chip for RNA quantity and quality.

2.1.3 Field isolate collection and enrichment for schizonts

The collection of *P. vivax* clinical isolates from patients was performed as part of an on-going NIAID-approved *P. vivax* protocol (ClinicalTrials.gov Identifier: NCT00663546). Chanaki Amaratunga, a staff scientist in the Fairhurst laboratory, determined which samples were best for subjecting to *ex vivo* culture based on the following exclusion criteria: patients who took antimalarials less than 1 month prior to sample collection and parasitemia <0.1%. She performed all sample processing in the field described in this subsection according to a previously published *P. vivax ex vivo* culture protocol (Russell

et al., 2011). *P. vivax* clinical isolates were processed (PV0417-3, PV0563, PV0565, PV0568) from Cambodian patients with malaria. Patient blood samples (16 ml for ages <18 years and 32 ml for >18 years collected by venipuncture into sodium heparin Vacutainers) were centrifuged (2000 rpm for 5 min) and plasma was removed. Samples were diluted with PBS (up to 64 ml, mixed by inverting tube), depleted of white blood cells and platelets (pass over 8 autoclaved, pre-wet CF11 columns packed to the 5.5-ml mark in a 10-ml syringe, and collect flow through), washed twice (centrifuge at 2000 rpm for 5 min, wash with 1x PBS, repeat 1 time), and re-suspend (packed cells at a 10% hematocrit in modified McCoys 5A complete media with 25% AB serum, and culture at 37ºC, 5% CO_2) until the parasites matured to schizonts. After maturation, cultures were pelleted (centrifuge at 2000 rpm for 5 min), the supernatant was removed, and cells were resuspended (to 50% hematocrit in 1x PBS). To prevent rosetting, cells were treated with trypsin (7.5 ml of 500 mg/l Trypsin-Versene) and incubated (15 min at 37ºC). To stop digestion, samples were diluted (add 2x volume of 1x PBS) and centrifuged (2000 rpm for 5 min). The recovered pellet was incubated with AB serum (6 ml for 5 min at RT), diluted with PBS (up to 30 ml) and separated on a 45% isotonic Percoll® gradient (5 ml of suspension overlaid on six 15-ml tubes containing 5 ml 45% isotonic Percoll® each). The suspensions were centrifuged (1200 g for 15 min) and the fine band of concentrated schizonts on the Percoll® interface was removed, centrifuged (2000 rpm for 5 min), and resuspended (5 ml of 1x PBS). Smears (made by pelleting ~200 µl) were counted with a hemocytometer. The remaining sample was pelleted, mixed with up to 10x RNAlater® (1 ml) and divided into 2 cryovials (500 µl each). Samples were stored at 4ºC overnight and transferred to liquid nitrogen*. Samples were shipped to the WTSI on dry ice.

*To note. -20ºC storage and transport would have been sufficient and recommended by RNAlater®. Storage in liquid nitrogen appeared to lead to some brown cloudiness (likely lysis), which made pelleting difficult, and multiple fractions were extracted. This did not appear to impact RNA quality and yields were very high.

2.1.4 RNA extraction of field isolates

RNA was isolated from the 4 field isolates (PV0417-3, PV0563, PV0565, PV0568) using the RiboPure Blood kit (Ambion) according to manufacturer instructions and subjected to 2 rounds of DNA digestion using the DNA-free kit (Ambion) according to manufacturer

instructions. Samples were analyzed by Bioanalyzer® (Agilent Technologies, Inc.) using an RNA Nano Chip to test quantity and quality.

2.1.5 cDNA synthesis and PCR

To test for genomic DNA contamination, 1 µl of extracted RNA from parasite isolates (PV0417-3, PV0563, PV0565, PV0568) was used to make cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to manufacturer instructions. Using both RNA and cDNA samples, a region surrounding an intron in the PvDBP gene PVX_110810 (5'AAACCGCTCTTTATTTGTTCTCC, 3' TTCCTCACTTCTTCTTTCATT) was amplified by PCR. Reaction volumes were as follows: 2.5 µl buffer, 2 µl dNTP mix (10 µM), 0.1 µl Platinum Pfx DNA Polymerase (Invitrogen), 16.9 µl water, 1 µl of each primer (10 µM), 1 µl cDNA or extracted RNA. Thermocycler conditions were as follows: incubation at 95ºC for 15 min, 35 cycles consisting of denaturation at 95ºC for 40 seconds, annealing at 55ºC for 40 seconds, elongation at 68ºC for 1 min, followed by a final extension at 65ºC for 5 min.

2.1.6 Strand-specific RNA library production

Lia Chappell developed a strand-specific Illumina library protocol for *P. falciparum* and *P. knowlesi* RNA (Chappell, et al, manuscript in preparation), and her protocol was followed with slight modifications for producing strand-specific Illumina libraries for the 4 *P. vivax* RNA samples (PV0417-3, PV0563, PV0565, PV0568). The work in this section used filter tips, RNAse-free tubes (0.2 ml), water, and other reagents throughout, and the workspace was regularly cleaned with RNAzap/70% EtOH. Tubes were placed in magnetic racks for any step in which supernatants were removed. Aliquots were taken throughout the process to check for the quantity and quality of material made after each step. I performed all laboratory steps through to the USER treatment of the libraries. Lia Chappell performed the PCR reactions and final clean-up prior to sending samples to the core WTSI sequencing facility.

2.1.6.1 Isolate mRNA using oligo(dT) magnetic beads

Aliquots of 5 µg of total RNA were made for each sample and diluted to a final volume of 50 µl in nuclease-free water. Dynabead Oligo(dT) beads (20 µl) were washed twice (100 µl of 2x RNA Binding Buffer), resuspended (50 µl of 2X RNA Binding Buffer), and combined with the total RNA. RNA was denatured to facilitate mRNA binding to the

beads by heating (65°C for 5 min), cooling (to 4°C), and incubating (RT, 5 min). Beads+mRNA were separated from the solution by placing tubes on the magnetic rack and incubating (RT, 2 min), and the supernatant was stored (-80°C). Beads+mRNA were washed twice (200 µl of Wash Buffer, with thorough pipette mixing 6 times). Beads+mRNA were separated from the solution by placing tubes on the magnetic rack and incubating (RT, 2 min), and the wash buffer was discarded. mRNA was eluted from the beads by removing tubes from the magnetic rack, mixing thoroughly with elution buffer (50 µl) and incubating samples (80°C, 2 min, then hold at 25°C). mRNA isolation was then repeated to further purify the sample. mRNA was bound to beads by adding 2X RNA buffer (50 µl), mixing thoroughly, and incubating (RT, 5 min). Beads+mRNA were separated from the solution by placing tubes on the magnetic rack and incubating (RT, 2 min), and the supernatant was discarded. Beads+mRNA were washed twice (200 µl of Wash Buffer, with thorough pipette mixing 6 times). Beads+mRNA were separated from the solution by placing tubes on the magnetic rack and incubating (RT, 2 min), and the wash buffer was discarded. mRNA was eluted from the beads by removing tubes from the magnetic rack, mixing thoroughly with elution buffer (17 µl) and incubating samples (80°C, 2 min), placing tubes on magnetic rack and collecting the purified mRNA by transferring the supernatant to a clean nuclease-free PCR Tube on ice. Purified mRNA yield and size distribution were assessed using 2 µl on a Bioanalyzer® (Agilent Technologies, Inc.) using a RNA Nano Chip.

2.1.6.2 mRNA fragmentation by Covaris

Samples were diluted (to 120 µl, nuclease-free water), heated (65°C, 5 min) and cooled (on ice, 5 min) to minimize secondary structure. Samples were transferred to Covaris tubes and fragmented using the following settings: Duty cycle 10%, Intensity 5, Cycles per burst 200, Time 60s (more for total RNA), water bath 4-8°C. Fragmentation success was assessed immediately using a 2 µl aliquot on a Bioanalyzer® (Agilent Technologies, Inc.) using a RNA Nano Chip. Fragmented RNA (118 µl) was precipitated by adding the sample to a 1.5-ml eppendorf tube and adding 2.5x the volume of RNase-free 96-100% EtOH (325 µl), 1/10 volume 3M Sodium Acetate (13 µl), glycogen (20 µg) and incubating (-80°C, overnight). Precipitated RNA was centrifuged (14,000 rpm, 4°C, 30 min), with supernatant discarded, washed twice (500 µl, 75% EtOH), and centrifuged (14,000 rpm, 4°C, 5 min). Supernatant was discarded by pipette and pellet was air dried

in a covered box (RT, 2-4 min), resuspended (12 µl, nuclease-free water), and stored on ice.

2.1.6.3 Reverse transcription/first strand cDNA synthesis and modified second-strand synthesis

Reverse transcription/first strand cDNA synthesis was performed with Superscript II according to manufacturer instructions for step 1 to 6 for a 20-µl reaction volume using random hexamers and 10 µl of resuspended fragmented mRNA. Following step 6, the RNA-DNA hybrid was cleaned using 1.8x reaction volume with Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.) (36 µl beads to 20 µl starting volume). Samples were mixed (vortex mixer), incubated (RT, 5 min), and separated from supernatant by magnetic stand until solution was clear (~5 min). The supernatant was discarded and the beads were washed twice (180 µl, 80% EtOH, incubated for 30 seconds). The EtOH was discarded and the beads were air-dried (10 min while the tube was on the magnetic stand). The RNA-DNA hybrid was eluted from the beads (25 µl water, mixed with vortex mixer), and placed on the magnetic stand until the solution was clear. The supernatant (23 µl) was transferred to a clean tube. Second-strand synthesis incorporated dUTPs instead of dTTPs by mixing together the following: 22.6 µl cleaned RNA-DNA, 3 µl Buffer 2 (NEB), 2 µl of 10 mM dNTP mix (using dUTP instead of dTTP), 0.4 µl RNAse H (Invitrogen), and 2.0 µl DNA pol I (Invitrogen). Reaction conditions were as follows: 16°C for 2 h, hold at 4°C. Samples were cleaned using Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.) and eluted as above but with 39.5 µl elution buffer.

2.1.6.4 End repair of cDNA libraries

To end repair the cDNA, the following reagents (all from NEB, except cDNA) were mixed together: 37.5 µl cDNA, 2 µl dNTP mix, 2.5 µl T4 DNA polymerase, 0.5 µl *E. coli* DNA polymerase 1, 2.5 µl T4 Polynucleotide Kinase, and 5 µl 10x phosphorylation buffer. The reactions were heated (20°C for 30 min), cleaned using Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.), and eluted as above but with 34 µl elution buffer.

2.1.6.5 dA-Tailing of cDNA libraries

Performed dA-Tailing of cDNA library using the NEBNext DNA modules with the following reaction volumes: 32 µl purified end-repaired cDNA, 5 µl NEBNext dA-Tailing Reaction Buffer, 10 µl 1 mM deoxyadenosine 5'-triphosphate, 3 µl Klenow fragment. Reactions were incubated (37°C for 30 min), cleaned using Agencourt

RNAClean XP Spri-beads (Beckman Coulter, Inc.), and eluted as above but with 22 µl elution buffer.

2.1.6.6 Ligation of PCR adapters

PCR adapters were added to the libraries using standard Illumina adapters. Reaction volumes were as follows: 20 µl end-repaired, dA-tailed cDNA, 25 µl DNA ligase buffer, 1 µl 10 µM no-PCR IDT P.E., and 4 µl Quick DNA ligase (NEB). Reactions were incubated (25°C for 15 min), cleaned using Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.), and eluted as above but with 1x Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.) and 22 µl elution buffer. The eluted samples were transferred to clean 1.5-mL LoBind tubes.

2.1.6.7 Digestion of dUTP strand and PCR

The strand containing dUTPs was digested using a standard protocol by mixing together the following: 20 µl cDNA and 2 µl USER enzyme (NEB). Reactions were incubated (37°C for 15 min), heated (95°C for 10 min), and cooled (held at 4°C). The libraries were then amplified by PCR. Reaction volumes were as follows: 20 µl USER treated cDNA, 5 µl water, 25 µl 2x KAPA HIFI HS Master mix, 1 µl of 10 mM PE 1 Illumina primer, and 1 µl of 10 mM Illumina index primer (matching the index of the ligated PCR-free adapters). Reaction conditions were as follows: 95°C for 5 min followed by 4 cycles of 95°C for 20 s, 60°C for 15 s, 72°C for 60, and a final extension of 72°C for 5 min. PCR reactions were then cleaned using Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.) and eluted as above but with 1x Agencourt RNAClean XP Spri-beads (Beckman Coulter, Inc.) and 22 µl elution buffer. The eluted samples were transferred to clean 1.5-mL LoBind tubes.

2.2 *P. vivax* RNA sequence analysis

2.2.1 Sequence mapping and quality control

Illumina HiSeq RNA sequences for 4 clinical isolates were mapped to the following reference genomes:

P. vivax Sal 1 (Carlton et al., 2008): http://plasmodb.org/common/downloads/release-9.3/PvivaxSaI1/fasta/data/PlasmoDB-9.3_PvivaxSaI1_Genome.fasta

P. vivax P01 (Auburn, Manuscript in preparation): ftp://ftp.sanger.ac.uk/pub/project/pathogens/gff3/CURRENT

P. falciparum 3D7 (Gardner et al., 2002): http://plasmodb.org/common/downloads/release-9.3/Pfalciparum3D7/fasta/data/PlasmoDB-9.3_Pfalciparum3D7_Genome.fasta

Homo sapiens (Lander et al., 2001): NCBI build V37

TopHat (version 2.0.14) (Kim et al., 2013) was used with the following settings: tophat -g 1 -I 2000. The max intron size was set to 2000 (-I 2000). The '-g 1' option ensured only uniquely mapping reads would be aligned (allows a single alignment to the reference, placed at the top scoring position if more than 1 alignment was found). Samtools (Li et al., 2009, Li, 2011) was used to report the numbers of total reads (samtools flagstat) and reads mapped uniquely (samtools view -c -q 1) to the *P. vivax* Sal 1, *P. vivax* P01, *P. falciparum* 3D7, and *Homo sapiens* reference genomes. Reads mapped to each genome were extracted from the bam files (samtools view -F4 file.bam | cut -f1 |sort | uniq > file_reads.txt) and compared to look for reads mapping to both *P. vivax* and *P. falciparum*, *P. vivax* and *Homo sapiens* (comm -12 file1.txt file2.txt | wc). When comparing the numbers of reads mapping to *P. vivax* Sal 1 and *P. vivax* P01, only the 14 main chromosomes were used (samtools idxstats file.bam | cut -f 1,3).

The *P. vivax* P01 reference genome annotation was divided into 3 categories (Exons, Introns and the remainder or "other"), the process of which is reviewed here: http://davetang.org/muse/2013/01/18/defining-genomic-regions/. This involves extracting the regions from the *P. vivax* P01 gtf file into to a bed file using unix and Bedtools (2.17.0) (Quinlan and Hall, 2010) as follows:

Exons: cat file.gtf | awk 'BEGIN{OFS="\t";} $3=="exon" {print $1,$4-1,$5}' | sortBed | mergeBed -i - > file.exon.bed[a]

Introns: cat file.gtf | awk 'BEGIN{OFS="\t";} $3=="transcript" {print $1,$4-1,$5}' | sortBed | subtractBed -a stdin -b file.exon.bed > file.intron.bed[a]

Remaining "other" regions:. cat file.gtf | awk 'BEGIN{OFS="\t";} $3=="transcript" {print $1,$4-1,$5}' | sortBed | complementBed -i stdin -g file.txt* > file.other.bed[a]

*create tab delimited file.txt listing chromosome ids and sizes

[a]Apicoplast and mitochondrial records removed from the files before coverage analysis

Coverage histograms were generated for the above genomic region bed files using Bedtools (coverageBed -split -hist -abam file.bam -b file.exon.bed > file.exons.bed.coverage), and aggregate statistics were calculated using Excel spread sheets or R. Coverage counts for rRNA regions using Bedtools (coverageBed -abam file.bam -b rRNA_locations.bed* > file.rRNA.coverage.counts). *created based on list compiled from the *P. vivax* P01 reference genome annotation (http://www.genedb.org/Homepage/PvivaxP01).

2.2.2 RNA-Seq expression analysis

Expression results were generated from the TopHat mapping assemblies using Cufflinks version 2.2.1) (Trapnell et al., 2012, Trapnell et al., 2010) with the following settings: cufflinks -o out.dir -b ref.fasta -u -q -g ref.gtf file.bam. The '-b' with a provided reference fasta file (ref.fasta) uses a bias detection and correction algorithm to improve accuracy of estimated transcript abundancy. The '-u' option enables an "initial estimation procedure to more accurately weight reads mapping to multiple locations in the genome," (http://cole-trapnell-lab.github.io/cufflinks/cufflinks/#transcriptome-assembly-transcriptsgtf). The '-g' option provides reference genome annotation (file.gtf) to guide the assembly. Using this option, the output file contained both the reference transcripts and any novel transcripts identified. The '-o' directs to the output directory (out.dir) and the '–q' suppresses messages other than warnings/errors. See the following for more details about options: http://cole-trapnell-lab.github.io/cufflinks/cufflinks/#transcriptome-assembly-transcriptsgtf

The results of Cufflinks computed a normalized expression level for each gene called **F**ragments **P**er **K**ilobase of transcript per **M**illion mapped reads or FKPMs. The FPKMs for each isolate were combined using unix and a custom perl script (Thomas Otto, Parasite Genomics group, WTSI). The results from the 'genes.fpkm_tracking' files output from Cufflinks were sorted, with gene_id and FPKM columns copied to a new file for each isolate in a new 'Stats' directory form the original 'cuff .SampleX" directories:

```
for x in `ls -d cuff. Sample1 cuff.Sample2 cuff. Sample3 cuff. Sample4` ; do n=$(echo $x | sed 's/cuff.//g' ); echo -e "gene_id\t$n" > Stats/$n.txt; cut -f 1,10 $x/genes.fpkm_tracking | grep -v FPKM | sort >> Stats/$n.txt; done
```

The 4 isolates FKPMs were then combined into a single file (Results.txt in the Stats directory): paste Sample1.txt Sample2.txt Sample3.txt Sample4.txt | cut -f 1,2,4,6,8,10,12 | sed 's/_R/_r/g ' > Results.txt

The product descriptions were added to the combined FPKM columns in Results.txt with a perl script and product description file supplied by Thomas Otto (Berriman Parasite Genomics Group, WTSI).

Multiple comparisons of FPKM results from Cufflinks were computed using R or Microsoft Excel. In order to validate that our samples best reflected the schizont stage of the *P. vivax* life cycle, the generated data were compared to published *P. falciparum* RNA-Seq data and microarray data, and *P. vivax* microarray data (Otto et al., 2010, Bozdech et al., 2003, Bozdech et al., 2008). Correlation plots (using the R "corrplot" library) for 1:1 orthologs between the raw FPKM values of the clinical isolates and either the RPKM data from *P. falciparum* RNA-Seq data or the microarray enrichment values for *P. falciparum* and *P. vivax*.

The similarity of expression between the isolates was investigated using R in pair-wise comparisons plotting the log(FPKM) of each sample to log(FPKM) of every other sample. Pearson correlations coefficients were computed for each pair-wise comparison (untransformed data). R (version 3.2.2, htpp://cran.r-project.org) was used to generate histograms and smooth density plots (density function) of the log(FPKM) for each isolate and plotted using the 'sm' library.

2.2.3 Expression variability between isolates

Expression variability between isolates was investigated in 3 ways using Microsoft Excel. First the 'max fold-change' was computed, by first calculating the mean FPKM from the 4 clinical isolates (PV0563, PV0565, PV0568, PV0417-3) for each gene. The fold-change from the mean for each isolate was calculated (isolate expression divided by the mean), and the maximum fold-change from the mean was recorded. Next, the Coefficient of Variation (c_v) was calculated for each gene as the ratio of standard deviation (σ) of the FPKMs from the 4 isolates to the mean (μ) FPKM:

$$c_v = \frac{\sigma}{\mu}$$

Lastly, the variance-to-mean ratio (VMR) also called index of dispersion (D) was calculated as a way to understand the spread of the expression results. This was computed as a ratio of the variance (σ^2) of the FPKMs from the 4 isolates to the mean (μ) FPKM:

$$D = \frac{\sigma^2}{\mu}.$$

The intersection of the top 300 ranked genes for each method was visualized in a Venn diagram.

In order to understand if any types of genes were enriched in the top variably-expressed gene set, the Gene Ontology (GO) terms for the one-to-one homologs with *P. falciparum* were investigated. The list of 77 *P. falciparum* one-to-one homologs was uploaded to panther.db. Analysis type: PANTHER Overrepresentation Test (release 20150430). Annotation version and release date: GO Ontology database, released 2015-08-06. Reported results only from published experimental data.

2.3 Production of recombinant *P. vivax* ectodomain library

2.3.1 *P. vivax* candidate selection

Expression plasmids corresponding to the entire ectodomains of secreted and membrane-embedded *P. vivax* merozoite proteins were designed and constructed essentially as described (Bushell et al., 2008, Crosnier et al., 2013). The entire ectodomain protein sequences were identified by removing the endogenous signal peptide, transmembrane domain, and glycosylphosphatidylinositol (GPI) anchor sequences (if present), and the corresponding nucleic acid sequences were codon-optimized for expression in human cells. N-linked glycosylation sequons were mutated from NXS/T to NXA (where X is any amino acid except proline) to prevent glycosylation when proteins were expressed in human cells. The final constructs were chemically synthesized and sub-cloned into a derivative of the pTT3 expression vector (Durocher et al., 2002), which contains an N-terminal signal peptide, a C-terminal rat CD4 domain 3 and 4 (Cd4d3+d4) tag, and 17 amino acid biotinylatable peptide using flanking NotI and AscI restriction sites (Geneart AG, Germany). In the interaction screens described below (section 2.5.2), these biotinylated proteins are referred to as “baits.” All biotinylatable bait expression plasmids are available from the non-profit plasmid repository, Addgene (www.addgene.org).

2.3.2 Subcloning *P. vivax* recombinant library

β-lactamase-tagged "prey"–expressing plasmids were produced by subcloning each ectodomain into a plasmid containing a pentamerization domain conjugated to the β-lactamase tag as described (Crosnier et al., 2013, Bushell et al., 2008). A subset of ectodomains was subcloned into a plasmid containing a six-histidine tag for producing purified proteins. All plasmids with target backbones were provided by Cecile Crosnier-Wright in the Wright laboratory. Expression plasmid backbones summarized in Table 2.1.

Table 2.1: Expression plasmid backbones

Name	Characteristics
Bio	Biotinylatable 'bait' pTT3-based expression vector (Durocher et al., 2002) with rat CD4d3+d4 tag (Brown and Barclay, 1994), 17 amino acid substrate for the *E. coli* biotin ligase BirA (Bushell et al., 2008, Brown et al., 1998)
βlac	β-lactamase-tagged 'prey' pTT3-based expression vector (Durocher et al., 2002) with rat CD4d3+d4 tag (Brown and Barclay, 1994), pentamerization domain of the rat cartilaginous oligomeric matrix protein (COMP) (Tomschy et al., 1996)with an ampicillin resistance protein (TEM) β-lactamase (Crosnier et al., 2010, Bushell et al., 2008)
Hexa- His	6-His-tagged pTT3-based expression vector (Durocher et al., 2002) with rat Cd4d3+d4 tag (Brown and Barclay, 1994) (Bushell et al., 2008)

Inserts and target vector backbones were cut using the following reaction components and conditions: 30 µl plasmid (1 mg/ml), 5 µl Buffer 4 (10x), 1 µl BSA (100x), 2 µl Not1 (New England BioLabs, NEB), 4 µl Asc1 (NEB), and 8 µl MilliQ water incubated at 37ºC for 3 h or overnight. Cut plasmids were then separated using agarose gel electrophoresis with 1% agarose gels in TAE at 100-150V for 1 h or more to achieve good separation between insert and vector backbone. Inserts or vector backbones were then cut from the gels and extracted using the QIAquick® Gel Extraction Kit (Qiagen) according to manufacturer's instructions, and DNA concentration was then quantified using a NanoDrop (Thermo Scientific).

Inserts and vector backbones were then ligated in a 3:1 ratio with a target of 60 ng insert to 20 ng vector backbone. The reaction volumes were as follows: 1-2 µl insert DNA (60 ng), 0.5-1 µl vector backbone DNA (20 ng), 1 µl T4 ligase (NEB), 1 µl T4 ligase buffer (NEB), and MilliQ water up to 10 µl. The ligations were incubated at RT for 3 h.

Ligations were stored at 4ºC for 1-2 days prior to transformation and plating or frozen for use at a later time.

Ligations were then used in transformation and plating. Chemically competent *E. coli* cells (25 µl One Shot® Top10, Invitrogen) were thawed on ice and combined with 3 µl of ligation reaction and incubated on ice for 5-30 min, heated to 42ºC for 45 seconds, cooled on ice for 2 min, and then mixed with 100-200 µl of SOC media at RT. The transformed cells were then spread using sterile techniques on LB-agar plates containing 100 µg/ml of ampicillin (Media Team at WTSI or Richard Eastman at LMVR) and grown overnight at 37ºC.

A selection of single colonies from positive plates was then checked using polymerase chain reaction (PCR). Single colonies were picked into tubes containing 12.5 µl GoTaq® Green Master Mix (Promega), 1 µl Primer 3610 (100 mM), 1 µl Primer 4006 (100 mM), 10 µl MilliQ water, and run with the following Thermocycler conditions: incubation at 95ºC for 10 min, 30 cycles consisting of denaturation at 95ºC for 30 seconds, annealing at 60ºC for 30 seconds, elongation at 68ºC for 1 min, followed by a final extension at 68ºC for 10 min. Primers for PCR and subsequent sequencing confirmation summarized in Table 2.2.

Table 2.2: Primers

Primer name	Primer sequence	Location/use
OL3610	GCCACCATGGAGTTTCAGACCCAGGTACT CATGTCCCTGCTGCTCTGCATGTCTGGTGC	5’ from insert sequence and Not1 site, in exogenous signal peptide
OL4006	TCCCTGCAGGCTTTCCTCTCCAAGGTTGAG	3’ from insert sequence and Asc1 site, in the Cd4d3+d4 tag
OL497	TGAGATCCAGCTGTTGGGGT	5’ from insert sequence, in vector backbone
OL498	AGAAGGGGCAGAGATGTCGT	3’ from insert sequence, in vector backbone

Half of the total PCR reaction was then separated using agarose gel electrophoresis with 1% agarose gels in TAE run at 100-150V for 30-60 min and checked for size agreement with the known insert size.

Single colonies with inserts of the correct size were then picked into 50 ml of LB-AMP (100 μg/ml) and grown shaking at 37ºC overnight. Plasmid DNA was then isolated using PureLink® HiPure Plasmid Filter Maxiprep Kit (Invitrogen) according to manufacturer's instructions. DNA was resuspended in TBE, and the concentration was adjusted to 1 mg/ml using a NanoDrop (Thermo Scientific).

To ensure that the final plasmid contained the expected insert and backbone, all sub-cloned plasmids were subjected to DNA sequencing using the primers summarized in Table 2.2 above, either by sending plasmid and primer aliquots to the in-house WTSI Capillary Sequencing Team, Macrogen, Inc. (USA), or by performing the reactions and running them on the LMVR 96-well capillary sequencing machine, 3730xl DNA Analyzer (Applied Biosystems).

All sequences were then assembled and aligned to the expected insert sequences and vector backbones using SeqMan Pro (version 12, DNASTAR Lasergene software package) or Sequencher (version 5.3, Gene Codes Corporation).

2.3.3 *P. vivax* library expression in the HEK293E system

P. vivax biotinylated bait and β-lactamase-tagged prey recombinant proteins were expressed in HEK293E cells as described (Bushell et al., 2008, Crosnier et al., 2013). HEK293E cells were maintained with shaking in Freestyle™ 293 Expression Medium (Invitrogen, USA) supplemented with G418 (50 mg/l), pen/strep (10000 units/l), and HI-FBS (1%) at 37ºC in a 5% CO_2 atmosphere. Bait and prey plasmids were transiently transfected as described (Durocher et al., 2002, Bushell et al., 2008). Cells were split into 50 ml of fresh media ($2.5x10^5$ cells/ml), incubated for 24 h, and inoculated with a transfection mix consisting of 25 μl of expression plasmid (at 1 mg/ml), 50 μl of linear PEI (at 1 mg/ml), preincubated for 10 min at RT in 2 ml of non-supplemented Freestyle media or using the 293fectin™ (Invitrogen, USA) according to manufacturer's instructions. Bait proteins were enzymatically biotinylated during synthesis using media supplemented with D-biotin (100 μM) and were co-transfected with a biotin ligase expression plasmid expressing a secreted BirA protein, along with the bait construct in a 1:10 ratio (2.5 μl of a plasmid at 1 mg/ml). After 3-6 days, cultures were centrifuged (4000 rpm for 15 min) and supernatants filtered (0.2-μm filter). Bait proteins were dialyzed in Snakeskin dialysis tubing (10 kDa MWCO, Thermo Scientific) using 5 L of

HBS (5-7 buffer changes over a 48-h period) to remove excess D-biotin. Cultures were stored in 10 mM sodium azide at 4ºC until use.

2.4 Confirmation and assessment of protein expression

2.4.1 Confirmation of protein expression by ELISA

Biotinylated bait proteins were immobilized on streptavidin-coated plates (Nunc, USA), and the dilution required to saturate all biotin-binding sites was determined by ELISA as described (Kerr and Wright, 2012). Proteins were serially diluted in HBS/1% BSA for 1 h, and plates were washed 3 times in HBS/0.1% Tween, incubated with mouse anti-rat CD4 antibody (OX68*, 1:1000) in HBS/1% BSA for 1 h, washed 3 times in HBS/0.1% Tween, incubated with mouse anti-rat alkaline phosphatase antibody (1:5000, Sigma, USA) in HBS/1% BSA for 1 h, washed 3 times in HBS/0.1% Tween, and washed once in HBS. Proteins were then incubated with phosphate substrate (1 mg/ml, Sigma) in either diethanolamine buffer (10% diethanolamine, 0.5 mM magnesium chloride, pH 9.2) or coating buffer (0.015 M sodium carbonate, 0.035 M sodium bicarbonate, pH 9.6) and detected by measuring OD at 405 nm. Each protein was concentrated or diluted using Vivaspin® 30-kDa spin columns (Sartorius, Germany) or HBS/1% BSA, respectively. As an approximate guide, the biotin binding sites are saturated at a biotinylated protein concentration of 0.3-0.5 μg/ml (Kerr and Wright, 2012, Osier et al., 2014). Based on this, and as a very rough approximation, protein expression was categorized “high” (likely above 5 μg/ml) for those proteins that could be diluted more than 1:10 and still saturate biotin binding sites, “medium” (approximately 0.5-5 μg/ml) for those proteins using 1:1 through 1:10 dilutions, or “low” (likely under 0.5 μg/ml) for those proteins that showed some signal, but failed to saturate the biotin binding sites over a range of dilutions prior to concentrating. These categories should be treated as a guide only since, due to experimental variation in transient transfections, we observed significant batch-to-batch variation.

*OX68 was isolated from hybridomas by Cecile Crosnier-Wright in the Wright laboratory and aliquoted for regular use.

2.4.2 Normalisation of β-lactamase tagged membrane protein ectodomains

Prey proteins were normalized using the β-lactamase tag activity as a proxy of protein concentration by the rate of nitrocefin, essentially as described (Kerr and Wright, 2012). Proteins were serially diluted (neat, 1: 5, and either 1:10 or 1:50) in HBS/1% BSA, and 20 µl of each dilution was transferred to a 96-well plate. Nitrocefin solution (Calbiochem) was added to each well (60 µl at 125 µg/ml), and absorbance at 485 nm was immediately read every minute for 20 minutes at RT. For each protein dilution, Absorbance at 485 nm was plotted against time and preys were concentrated or diluted as needed to achieve saturation of signal at approximately 10 min (approximating a threshold activity of ~2 nmol/min turnover of nitrocefin) using Vivaspin® 20- or 30-kDa spin columns (Sartorius, Germany) or HBS/1% BSA (Kerr and Wright, 2012).

2.4.3 SDS-PAGE, Western blotting and NativePAGE

Western blotting was used to confirm protein sizes. *P. vivax* biotinylated bait proteins were reduced with NuPAGE® Sample Reducing Agent (Invitrogen), heated at 70ºC for 10 min, fractionated by SDS-PAGE, and transferred to nitrocellulose membranes. After blocking in HBS/0.1% Tween/2% BSA at 4ºC overnight, membranes were incubated with streptavidin horseradish peroxidase (1:2000, Cell Signaling Technology, USA) in HBS/0.1% Tween/2% BSA at RT for 1-2 h, developed using the chemiluminescence substrate *Amersham ECL^{TM} Prime* (GE Healthcare, USA), and exposed to X-ray film.

In order to determine whether several proteins formed dimers or oligomers, non-reducing conditions were tested. Recombinant 6-His-tagged *P. vivax* P12, P41, and MSP7.1 were fractionated by SDS-PAGE under reducing conditions as above [NuPAGE® Sample Reducing Agent (Invitrogen), heated at 70ºC for 10 min] and non-reducing conditions, and stained using SimplyBlue SafeStain (Invitrogen). Samples were also run under native conditions using the NativePAGE Bis-Tris Gel System (Thermo Fisher Scientific), stained with Coomassie Brilliant Blue R-250 (ThermoScientific).

2.5 High-throughput functional screens

2.5.1 Erythrocyte/reticulocyte binding experiments by flow cytometry

P. vivax recombinant protein binding to whole erythrocytes and reticulocytes was investigated using a flow cytometry-based assay developed by Madushi Wanaguru as

described (Crosnier et al., 2013) based on a previously described method (Brown, 2002). The biotinylated proteins were immobilized on streptavidin-coated fluorescent beads (Nile red) and presented to erythrocytes, with binding events detected as a shift in fluorescence intensity by flow cytometry. A bead-saturation ELISA similar to the protocol in section 2.4.1 was first necessary to determine the optimal protein dilution for saturating the biotin binding sites on a uniform number of beads while leaving a minimal amount of unbound protein.

Biotinylated proteins were serially diluted in HBS/1% BSA for 1 h and 100 µl of each dilution was transferred to a 96-well microtitre plate. Proteins were either pre-incubated for 1 h at 4ºC, shaking, either without or with 4 µl of streptavidin-coated Nile Red fluorescent 0.4–0.6 µm microbeads (Spherotech Inc.). Prior to use, beads were washed twice and resuspended in HBS/1% BSA with sonication (5 min) after all washes using a bath sonicator at 4ºC. The pre-incubated dilution series were then transferred to streptavidin-coated plates (Nunc, USA), where protein not bound to beads would subsequently bind to plates. The ELISA then proceeded as described in section 2.4.1. Protein dilutions pre-incubated with beads indicating saturation of biotin sites on beads, with low remaining quantities of unbound protein (as indicated by low OD values by ELISA) were selected.

For the binding assay, proteins (100 µl) were then immobilized on 4 µl streptavidin-coated Nile Red fluorescent 0.4–0.6 µm microbeads (Spherotech Inc.) (washed and sonicated as above) by incubation with gentle shaking for 1 h at 4°C. Human erythrocytes or hematopoietic stem cell-derived reticulocytes (provided by NHS Blood and Transplant) were washed twice in RPMI and aliquoted into a 96-well flat bottomed plate at $3x10^5$ cells/well in 30 µl of RPMI, creating a monolayer of cells available to the protein-coated beads. Plates containing protein-coated beads were sonicated for 20 min at 4°C and transferred (100 µl protein-bead mixture resulting in approximately 120 beads/cell) to erythrocyte-containing plates. The erythrocyte-bead mixture was centrifuged (1000 rpm for 20 min) and incubated with gentle shaking for 1 h at 4°C. The cells were analysed using flow cytometry on an LSRII cytometer (BD Biosciences) using the BD FACS Diva software. Nile Red was excited by a blue laser and detected with a 575/26 filter. Voltages for FCS (430 V) and SSC (300 V) and a threshold of 26,100 on FSC were applied when analysing. The results were analysed using Flow Jo v7.5.3 software (Tree Star, Inc.).

2.5.2 Protein screens using AVEXIS

P. vivax biotinylated bait and β-lactamase-tagged prey proteins were screened using AVEXIS (Bushell et al., 2008, Crosnier et al., 2013). For AVEXIS, biotin-binding sites on streptavidin-coated plates (Nunc) were saturated with biotinylated bait proteins for 1 h. Plates were washed 3 times in HBS/0.1% Tween, probed with β-lactamase-tagged prey proteins for 1 h, washed twice with HBS/0.1% Tween, washed once with HBS, and incubated with nitrocefin (60 µl at 125 µg/ml; Calbiochem, USA). Positive interactions were indicated by nitrocefin hydrolysis, which was detected by measuring absorbance at 485 nm at 90 min.

2.6 Biophysical analysis of protein interactions

2.6.1 Purification of 6-His-tagged membrane protein ectodomains

The entire ectodomains of P12, P41, and MSP7.1 were sub-cloned into a modified plasmid containing a 6-His tag (Bushell et al., 2008), expressed as above, and purified by immobilized metal ion affinity chromatography using HisTrap HP columns on an AKTA Xpress (GE Healthcare) following the manufacturer's instructions. Columns were pre-equilibrated with binding buffer (20 mM sodium phosphate, 40 mM imidazole, 0.5 M NaCl, pH 7.4) at a flow rate of 1 ml/min. Harvested supernatant (150-500 ml) was supplemented with imidazole (10 mM) and NaCl (100 mM) and passed over the column at a flow rate of 1 ml/min. The column was washed with 15 column volumes (CV) of binding buffer in order to remove non-specific adherents, and the proteins were eluted using 10 CV of elution buffer (20 mM sodium phosphate, 0.4 M imidazole, 0.5 M NaCl, pH 7.4). The eluant was collected in 0.5-ml fractions; fractions containing the highest concentration of purified protein (1-2 total), as estimated by measuring the absorbance at 280 nm, were subsequently subjected to SDS-PAGE and further purification by gel filtration.

In order to remove aggregates, purified proteins were subjected to size-exclusion chromatography (SEC) immediately before use as described (Taechalertpaisarn et al., 2012). Either a Superdex™ 200 Increase™ 10/300 GL column or Superdex™ 200 Tricorn™ 10/600 GL column using an AKTA Xpress (GE Healthcare) was pre-equilibrated with 2 CV of running buffer HBS-EP (0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v surfactant P20, pH 7.4 from GE Healthcare) at a flow rate of 0.2

ml/min. Protein samples were then injected, and the column was washed with running buffer at 0.2 ml/min, and 0.5-ml fractions were collected after passage of the void volume of the column. Absorbance at 280 nm and extinction coefficients (calculated using the protein sequence and the ProtParam tool, http://www.expasy.org/tools/protparam.html) were used to estimate the protein concentrations of peak fractions. Protein sizes were deduced from the elution volumes of the peak fractions by comparison to a standard curve generated by Josefin Bartholdson using well-defined protein standards from calibration kits (GE Healthcare).

2.6.2 Surface plasmon resonance (SPR)

SPR was used to biochemically confirm all interactions detected by AVEXIS. SPR was performed using a BIAcore T100 instrument (GE Healthcare) at 37°C in HBS-EP running buffer (0.01 M HEPES, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v surfactant P20, pH 7.4 from GE Healthcare). Biotinylated P12, P41, MSP3.10, and PVX_110945 proteins (expressed as in 2.3.3) were immobilized to the sensor chip using the Biotin CAPture Kit (GE Healthcare) in approximately molar equivalents with a negative control reference biotinylated rat CD4d3+4 (tag alone). Increasing concentrations of each soluble purified analyte was injected at low (20 µl/min) flow rates for equilibrium experiments. The chip was regenerated between each injection cycle using regeneration buffer provided by the kit. Biacore evaluation software version 2.0.3 (GE Healthcare) was used to analyze the reference-subtracted sensorgrams. Binding was investigated by plotting the maximum binding response for a range of concentrations (R_{eq}) prior to washing, and plotted as a function of analyte concentration (C) and fitted using the equation $R_{eq}=CR_{max}/(C+K_D)$, where R_{max} is the maximum binding response and K_D is the equilibrium dissociation constant.

2.7 *P. vivax* seroepidemiology

2.7.1 Seroepidemiology of recombinant *P. vivax* proteins in Cambodian patients

2.7.1.1 Collection of Cambodian *P. vivax*-exposed plasma

In Pursat Province, Western Cambodia, Sokunthea Sreng and Seila Suon oversaw the collection and preparation of plasma samples from patients with acute vivax malaria.

Adult patients or the parents of child patients provided written informed consent under a protocol approved by the National Ethics Committee for Health Research in Cambodia and the NIAID Institutional Review Board in the United States (ClinicalTrials.gov Identifier, NCT00663546). Paired samples were also collected including 'acute' plasma from clinical presentation and 'convalescent' plasma 21-28 days later. Patient plasma samples were used in ELISA experiments described below.

2.7.1.2 ELISA optimization screening biotinylated *P. vivax* recombinant proteins with Cambodian patient plasma

Plasma from Cambodian patients with *P. vivax* infections were used to optimize ELISA conditions using Cambodian patient plasma. Plasma samples were selected based on prior ELISA data showing reactivity using recombinant *P. vivax* MSP1 (data generated and provided by Daria Nikolaeva and Carole Long, LMVR, NIH).

Carole Long and Kazutoyo Miura (LMVR/NIH) advised on the initial ELISA protocols based on the standard ELISA used in the Long Laboratory (Miura et al., 2008). Ababacar Diouf provided extensive initial laboratory training. The ELISA protocol for screening *P. vivax* recombinant proteins with Cambodian *P. vivax*-infected patient plasma was optimised using a subset of 6 *P. vivax* recombinant proteins and the Cd4d3+d4 tag alone expressed as in 2.3.3. To conserve protein supernatant, biotinylated bait proteins were diluted (HBS/1% BSA) to the maximum dilution where all biotin-binding sites on streptavidin-coated plates (Nunc, USA) were saturated. Optimal dilutions were determined by ELISA as described in section 2.4.1, and diluted proteins were immobilized on streptavidin-coated plates (Nunc, USA) for 1 h and washed 3 times (HBS, 0.1% Tween). Proteins were incubated in duplicate against 42 Cambodian *P. vivax* patient plasma samples (diluted 1:200, 1:400; 1:800, and 1:1600) in HBS with 1% BSA, 2 h) and washed 3 times (HBS, 0.1% Tween). Proteins were incubated with alkaline-phosphatase conjugated goat anti-human IgG (KPL, Inc., diluted 1:1000 in HBS with 1% BSA, 2 h), washed 3 times in HBS/0.1% Tween, and washed once in HBS. Proteins were then incubated with phosphate substrate (1 mg/ml, Sigma) in coating buffer (0.015 M sodium carbonate, 0.035 M sodium bicarbonate, pH 9.6) and detected by measuring OD at 405 nm measured at 5, 10, 20 and 25 min.

An OD value of 0.1 was set as a conservative lower limit of detection based on the plate reader used. Samples with OD values under 0.1 were set to 0.1. Differences in population

antibody responses comparing the proteins and the Cd4d3+d4 tag alone were assessed using Mann-Whitney U tests. A reactivity cut-off was set at 2 standard deviations above the mean of 5 American naïve controls after correcting for background responses by subtracting the Cd4d3+d4 values. Analysis and graphing performed with GraphPad Prism (version 6, GraphPad Software, Inc.)

2.7.1.3 Cambodian reactivity screen using a library of biotinylated *P. vivax* recombinant proteins with pooled Cambodian patient plasma

A set of 17 plasma samples was obtained from 14 patients who experienced their third, fourth, or fifth episode of acute vivax malaria within 2 years of their initial episode on our clinical protocol (Table 2.3). Control sera from 5 malaria-naïve individuals were obtained from Interstate Blood Bank (Memphis, TN, USA). Equal volumes of individual plasma or serum samples were pooled for use in ELISA.

Table 2.3: Plasma samples from Cambodian patients with acute vivax malaria

Patient ID	Sex	Age (years) at first episode	Date of episode				
			1	2	3	4	5
PV0005	M	10	20080619	20080723	20080829	20081008	20081114
PV0012	M	10	20080624	20080729	20080906	20081013	20081204
PV0022	M	24	20080627	20080809	20080925	20081212	
PV0092	F	5	20080830	20081005	20081212	20091015	
PV0106	M	8	20080914	20081021	20081201	20090713	20100729
PV0128	M	10	20081009	20100709	20101020		
PV0242	M	22	20090817	20091118	20101217		
PV0340	M	15	20100705	20101116	20101226		
PV0345	M	20	20100710	20100914	20101203		
PV0352	M	39	20100717	20100829	20101015		
PV0359	M	14	20100719	20100902	20101014		
PV0361	M	10	20100720	20100915	20101028		
PV0378	F	8	20100728	20100901	20101013	20101209	
PV0412	F	17	20100814	20100925	20101112	20101226	

Plasma samples were collected from patients in Pursat Province, Western Cambodia, in 2008-2010. Plasma was pooled from 17 malaria episodes (in bold) from 14 patients who presented with their third, fourth, or fifth episode of acute vivax malaria on our clinical protocol. Table provided by Chanaki Amaratunga and modified by Jessica Hostetler. S1 Table from (Hostetler et al., 2015).

ELISAs were used to assess the seroreactivity and conformation of 37 biotinylated *P. vivax* recombinant proteins and the Cd4d3+d4 tag alone (expressed as in section 2.3.3). *P. vivax* biotinylated proteins were denatured at 80ºC for 10 min or left untreated, and then immobilized on streptavidin-coated plates (as in section 2.7.1.2). Proteins were incubated in a single assay in triplicate wells with pooled plasma (1:600 or 1:1000) from 14 vivax malaria patients from Cambodia or pooled serum from 5 malaria-naïve individuals from the United States in HBS/1% BSA. The ELISA then proceeded as described in section 2.7.1.2.

2.7.1.4 Cambodian reactivity screen using a 11 biotinylated *P. vivax* recombinant proteins with individual Cambodian patient plasma

In order to better understand IgG reactivity to recombinant *P. vivax* proteins in Cambodian patients, a subset of proteins was screened using ELISA against 48 individual Cambodian patient plasma experiencing their second, third, fourth, or fifth episode of acute vivax malaria within 2 years of their initial episode on our clinical protocol. Cambodian plasma samples (at 1:1000) and 24 sera samples from American malaria-naïve donors (at 1:1000) in HBS/1% BSA were screened in singlicate as described in 2.7.1.2, except that wash steps were performed using a plate washer machine with 4 washes in TBS, 0.1% Tween between each step, rather than HBS, 0.1% Tween.

2.7.1.5 Cambodian acute -convalescent screen using a library of biotinylated *P. vivax* recombinant proteins with Cambodian patient plasma

ELISAs were used to measure IgG responses in ‘acute’ and ‘convalescent’ Cambodian patent plasma to 10 biotinylated *P. vivax* recombinant proteins and the Cd4d3+d4 tag alone (expressed as in section 2.3.3). *P. vivax* biotinylated proteins were immobilized on streptavidin-coated plates (as in section 2.7.1.2). Proteins were incubated in a single assay in duplicate wells with ‘acute’ and ‘convalescent’ plasma (1:600) from 18 vivax malaria patients from Cambodia selected at random. Proteins were incubated individually with serum from 5 malaria-naïve individuals from the United States in HBS/1% BSA. The ELISA then proceeded as described in section 2.7.1.2, with the exception of using a plate washer machine with 4 washes in TBS, 0.1% Tween between each step, rather than HBS, 0.1% Tween.

2.7.1.6 Analysis

Differences in reactivity were tested using Mann-Whitney U tests (unpaired) and Wilcoxon rank-sum tests (paired). Analysis and graphing performed with GraphPad Prism (version 6, GraphPad Software, Inc.)

2.7.2 Seroepidemiology of recombinant *P. vivax* proteins in Solomon Islander patients

2.7.2.1 Cross-sectional study and sample selection

Blood samples were collected in a cross-sectional survey of 3501 individuals aged ≥6 months in May 2012 in Ngella, Central Island Province, SI (Waltmann et al., 2015). Samples were collected with ethical approval from the Solomon Islands National Health Research Ethics Committee and the Walter and Eliza Hall Institute, and verbal informed consent was given by all adult participants and parents or guardians of child participants (as approved by Australian and Solomon Islands' IRBs, and documented on each participant's case report form). Survey participants were grouped into a 3x3 factorial design for downstream IgG reactivity screening by ELISA. This included 3 age categories (5-9, 10-19, and 20-99 years) and 3 infection statuses as follows:

1. Not infected [negative by PCR and light microscopy (LM)]
2. PCR + (positive by PCR and negative by LM)
3. PCR + LM + (positive by PCR and positive by LM)

2.7.2.2 Solomon Islands plasma screening of *P. vivax* recombinant proteins using ELISA

Reactivity screening: ELISAs were used to measure IgG response in SI patient plasma to 34 biotinylated *P. vivax* recombinant proteins and the Cd4d3+d4 tag alone (expressed as in section 2.3.3). *P. vivax* biotinylated proteins immobilized on streptavidin-coated plates (as in section 2.7.1.2). Proteins were incubated with individual plasma samples (1:600) from 22 adolescents (10-19 years) and 24 adults (20-50 years) selected at random from the 3 infection categories described above. Proteins were incubated with pooled serum from 5 malaria-naïve individuals from Australia (1:600) in HBS/1% BSA as a negative control and plasma pooled highly-immune PNG adult donors (1:600) in HBS/1% BSA as a positive control. After 2 h of incubation, wells were washed 3 times with HBS/0.1% Tween, incubated with alkaline phosphatase-conjugated goat anti-human IgG (1:1000, KPL Inc., USA) in HBS/1% BSA for 2 h, washed 3 times with HBS/0.1% Tween, and

washed once with HBS. Seroreactivity was detected using phosphate substrate (1 mg/ml, Sigma) in coating buffer (0.015 M sodium carbonate, 0.035 M sodium bicarbonate, pH 9.6) and measuring OD at 405 nm at various times up to 30 min.

Comprehensive screening: ELISAs were repeated as above with an expanded set of plasma samples for 12 highly immunogenic proteins. In order to investigate relationships between infection status, clinical symptoms, and other factors, 144 plasma samples were selected at random from the 3 age categories and 3 infection statuses including the following: 48 children (5-9 years), 48 adolescents (10-19 years), and 48 adults (20-80 years) either without any *Plasmodium* infections (Not Infected), with a current *P. vivax* monoinfection detected by PCR (PCR+), or with a current *P. vivax* monoinfection detected by PCR and light microscopy (PCR+ LM+). Plasma samples were tested in duplicate on separate plates.

2.7.2.3 Statistical analysis

Reactivity screening: An OD value of 0.1 was set as a conservative lower limit of detection based on the plate reader used. Samples with OD values under 0.1 were set to 0.1. Differences in population antibody responses comparing the proteins and the Cd4d3+d4 tag alone, and 3 age categories were assessed using Mann-Whitney U tests. Analysis and graphing performed with GraphPad Prism (version 6, GraphPad Software, Inc.)

Comprehensive screening: An OD value of 0.1 was set as a conservative lower limit of detection based on the plate reader used. Samples with OD values under 0.1 were set to 0.1. OD values from duplicate wells were averaged, the Cd4d3+d4 values were subtracted to correct for background, and $\log_{(10)}$-transformed. Differences between mean antibody responses between age groups or infection groups were analysed using ANOVA. Positivity cut-offs were set at 2 standard deviations above the mean antibody levels of the negative controls. Negative binomial regression was used to assess differences in the breath of antibody levels by age and infection group. Analysis and graphing performed using STATA (version 12, StataCorp), GraphPad Prism (version 6, GraphPad Software Inc.), or R (version 3.2.2, htpp://cran.r-project.org).

Multivariate analysis: multivariate ANOVA models were fitted including all variables univariately associated with IgG levels. The best model was determined using backward

elimination using Wald's Chi-square tests for individual variables. Analysis performed by Camila Franca and Ivo Mueller using STATA (version 12, StataCorp).

2.7.3 Seroepidemiology of recombinant *P. vivax* proteins in Papua New Guinea patients

2.7.3.1 Longitudinal cohort study design and sample collection

In order to investigate protein responses that correlated with clinical protection, we screened a subset of proteins using plasma samples collected in a longitudinal cohort from PNG described in detail (Lin et al., 2010). Samples were collected with ethical approval from the Medical Research and Advisory Committee of the Ministry of Health in PNG and the Walter and Eliza Hall Institute, and written informed consent was given by all adult participants and parents or guardians of child participants. The study enrolled 264 children aged 1-3 years near Maprik, East Sepik Province, in March-September, 2006 and followed for 16 months. Blood samples were taken every 2 weeks and passive case detection occurred throughout the study period. Samples that were positive by PCR for *P. vivax* infection were genotyped to determine the number of genetically distinct infections acquired during the follow-up period. This molecular force of blood-stage infections, molFOB has been described in detail (Mueller et al., 2012, Koepfli et al., 2013). Samples from the 230 children who completed follow-up were used in the Luminex screen described below.

2.7.3.2 Papua New Guinea plasma screening of *P. vivax* recombinant proteins using Luminex

A subset of 6 highly immunogenic proteins from the SI 'Comprehensive screening' were selected by screening against plasma samples from the PNG longitudinal cohort. The proteins were prioritized based on reactivity in SI and Cambodia and on expression levels. The entire ectodomains of 6 *P. vivax* proteins (P12, P41, GAMA, CyRPA, ARP, PVX_081550) and the Cd4d3+d4 tag alone were sub-cloned into a modified plasmid containing a 6-His tag (Bushell et al., 2008). Sumana Sharma expressed the proteins as in section 2.3.3, and purified them by immobilized metal-ion affinity chromatography using HisTrap HP columns on an AKTA Xpress (GE Healthcare) following the manufacturer's instructions.

Camila Franca conjugated the purified proteins to Luminex Microplex microspheres (Luminex Corporation) as described (Kellar et al., 2001), with the following

concentrations per $2.5x10^6$ beads: P41, 0.5µg/ml; PVX_081550, 1.2µg/ml; P12, 0.2µg/ml; GAMA, 0.015µg/ml; ARP, 0.09µg/ml; CyRPA, 1.5µg/ml; and Cd4, 2µg/ml. Coupling efficiency was previously determined by testing for high fluorescence intensity by the reporter fluorochrome using an immune plasma pool from PNG adults with high reactivity to the antigens by ELISA.

Camila Franca performed a Luminex bead array assay as described (Piriou et al., 2009) to measure total IgG response against the proteins in singlicate to 230 PNG patient plasma samples with secondary antibody donkey F(ab')2 anti-human IgG Fc R-PE (1:100 in PBS) (Jackson ImmunoResearch). Serum from malaria-naïve individuals from Australia was used as negative controls. A dilution series of plasma pooled from highly-immune PNG adult donors was used as a positive control to standardize plate-to-plate variations.

2.7.3.3 Statistical analysis

Camila and Ivo Mueller performed the analyses described below, which were subsequently published (Franca et al., 2016).

> *To correct plate-to-plate variations, the dilutions of the highly-immune PNG positive control pool were fitted as plate-specific standard curves using a 5-parameter logistic regression model (Giraldo et al., 2002). For each plate, Luminex median fluorescence intensity (MFI) values were interpolated into relative antibody units based on the parameters estimated from the plate's standard curve. Antibody units ranged from 1.95x10-5 (i.e., equivalent to 1:51200 dilution of the immune pool) to 0.02 (1:50). To account for the background reactivity to the Cd4-tag, antibody levels were re-scaled by using linear regression to estimate the antibody levels that would be detected if reactivity to the Cd4-tag was zero, as follows:*
>
> $$log(AB\_meas) = log(AB\_true) + \beta * log(Cd4)$$
>
> *where AB_meas = measured antibody level, AB_true = true antibody level to a given antigen, and Cd4 = measured antibody level to the Cd4-tag.*
>
> *Associations between antibodies and age and exposure were assessed using Spearman rank correlation, and differences with infection using 2-tailed unpaired t-test on log_{10}-transformed values. Negative binomial GEE models with exchangeable correlation structure and semi-robust variance estimator (Stanisic*

et al., 2013) were used to analyse the relationship between IgG levels and prospective risk of P. vivax *episodes (defined as axillary temperature ≥37.5°C or history of fever in the preceding 48 hours with a current* P. vivax *parasitemia >500 parasites/μL). For this, IgG levels were classified into terciles and analyses done comparing children with low versus medium and high antibody levels. Children were considered at risk from the first day after the initial blood sample was taken. The molFOB, representing individual differences in exposure, was calculated as the number of new* P. vivax *clones acquired per year at risk, and square-root transformed for better fit (Koepfli et al., 2013). All GEE models were adjusted for seasonal trends, village of residency, age, and molFOB. In multivariate models that included all antigens univariately associated with protection, the best model was determined by backward elimination using Wald's Chi-square tests for individual variables. To investigate the effect of increasing cumulative IgG levels to the combination of antigens on the risk of* P. vivax *episodes, we assigned a score of 0, 1, and 2 to low, medium, and high antibody levels, respectively, and then added up the scores to the 6 antigens to generate a breadth score per child. The breadth score was then fitted as a continuous covariate in the GEE model described above. Analyses were performed using STATA version 12 (StataCorp) or R version 3.2.1 (htpp://cran.r-project.org).*

2.8 Commonly used buffers and solutions

The several buffers and solutions that were regularly used throughout this thesis work are summarized in Table 2.4 [Jessica Hostetler (JBH), Camila Franca (CF), Susana Campino (SC)].

Table 2.4: Buffers, media and solutions

Name	Recipe	Usage	Produced by
Coating Buffer	0.015 M sodium carbonate, 0.035 M sodium bicarbonate, pH 9.6	ELISAs	Long lab. at NIH, JBH and CF at WEHI
Diethanolamine buffer	10% diethanolamine, 0.5 mM MgCl2, pH 9.2)	ELISAs	Wright Lab. and JBH
Freestyle supplemented media	Freestyle™ 293 Expression Medium (Invitrogen, USA), 50 mg/l G418, 10000 units/l pen/strep, and 1% HI-FBS	HEK293E cell culture	Wright Lab., William Proto, Sumana Sharma, and JBH
HBS	0.14M NaCl, 5mM KCl, 2mM CaCl2, 1mM MgCl2, 10mM HEPES	*P. vivax* protein dialysis; recipe from Cecile Crosnier-Wright.	10X HBS: WTSI Media Team, JBH at NIH
HI-FBS Complete media	90% incomplete media, 10% HI-FBS, 0.2% sodium bicarbonate	Parasite culturing and schizont enrichment	JBH and SC
Incomplete media	RPMI, 25 mM HEPES, 50 µM hypoxanthine 0.2% gentamycin	Parasite culturing and schizont enrichment	JBH and SC
Modified McCoys media	McCoys media, Ascorbic acid 5ug/L, MgSO4 16ng/L, hypoxanthine 10ng/L, gentamycin 2mL/L (from 10mg/mL stock), 25% heat inactivated AB serum	*P. vivax ex vivo* parasite culturing	Chanaki Amaratunga
PBS		*P. falciparum*/*P. vivax* proteins, western blotting	10X PBS: WTSI Media Team, Invitrogen at NIH