Development and validation of serological markers for detecting recent Plasmodium vivax infection

Rhea J Longley, Michael T White, Eizo Takashima, Jessica Brewster, Masayuki Morita, Fumie Matsuura, Thomas Obadia, Zoe S J Liu, Connie S N Li-Wai-Suen, Wai-Hong Tham, et al.

1. Extended Data

Complete the Inventory below for all Extended Data figures.

Figure #	Figure title One sentence only	Filename This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: *Smith_ED_Fig1.jpg*	Figure Legend If you are citing a reference for the first time in these legends, please include all new references in the Online Methods References section, and carry on the numbering from the main References section of the paper.
Extended Data Fig. 1	Study design and follow-up schedule	Ext_Data_Fig1_study_design.tif	Study design and follow-up schedule. (A) Thai and Brazilian patients were enrolled following a clinical episode of *P. vivax* and treated according to the relevant National Guidelines, with directly observed treatment (DOT) to ensure compliance and reduced risk of relapse. Volunteers were followed for nine months after enrolment, with finger-prick blood samples collected at enrolment and week 1, then every two weeks for six months, then every month. Antibody levels were measured in a subset of 32 Thai and 33 Brazilian volunteers who were confirmed to be free of blood-stage *Plasmodium* parasites by analysing all samples by light microscopy and qPCR. (B) 999 participants from Thailand, 1274 participants from Brazil, and 860 participants from the Solomon Islands were followed longitudinally for 12 months with active surveillance visits every month. For the analysis in the validation phase antibody levels were measured in plasma samples from the last visit. For the analysis in the application phase, antibody levels were measured in plasma samples from the first visit.
Extended Data Fig. 2	Measured antibody responses to 60 proteins on the Luminex® platform, stratified by geographical location and time since last PCR detected infection	Ext_Data_Fig2_all_AB.tif	Measured antibody responses to 60 proteins on the Luminex® platform, stratified by geographical location and time since last PCR detected infection.
Extended Data Fig. 3	Association between background reactivity in non-malaria exposed controls and ranking of candidate SEMs	Ext_Data_Fig3_AUC_background.tif	Association between background reactivity in non-malaria exposed controls and ranking of candidate SEMs by area under the curve (AUC). Mean relative antibody units (RAU) detected in malaria-naïve control panels from Melbourne, Australia (n=202), Bangkok, Thailand (n=72) and Rio de Janeiro Brazil (n = 96) compared to the AUC of the 60 candidate P.vivax proteins generated during the validation phase. WGCF expressed

			proteins are in black and E. coli or Baculovirus expressed proteins are in blue. RBP2b161-1454 (E. coli) is in red and RBP2b1986-2653 is in orange.
Extended Data Fig. 4	Detailed breakdown of classification performance for the target of 80% sensitivity and 80% specificity	Ext_Data_Fig4_classification_breakdown.tif	Breakdown of the classification of the Random Forests algorithm with target sensitivity and specificity of 80%. The size of each rectangle is proportional to the number of samples in each category (See Extended Data Table 1 of accompanying manuscript). The coloured area represents the proportion correctly classified, and the shaded area represents the proportion mis-classified.
Extended Data Fig. 5	Cross-validated receiver operating characteristic (ROC) curve for the composite classification algorithm	Ext_Data_Fig5_cross-val.tif	Receiver operating characteristic (ROC) curve for the composite classification algorithm. All curves presented are the median of 1000 repeat cross-validations.
Extended Data Fig. 6			
Extended Data Fig. 7			
Extended Data Fig. 8			
Extended Data Fig. 9			
Extended Data Fig. 10			

Delete rows as needed to accommodate the number of figures (10 is the maximum allowed).

2. Supplementary Information:

A. Flat Files

Complete the Inventory below for all additional textual information and any additional Supplementary Figures, which should be supplied in one combined PDF file.

Item	Present?	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The	A brief, numerical description of file contents. i.e.: *Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.*

		extension must be .pdf	
Supplementary Information	Choose an item.	SEM_Supp_Info.pdf	**Supplementary Methods and Results,** Supplementary Figures 1-12
Reporting Summary	Choose an item.	nr-reporting-summary_NMED-A93738A	

B. Additional Supplementary Files

Complete the Inventory below for all additional Supplementary Files that cannot be submitted as part of the Combined PDF.

Type	Number If there are multiple files of the same type this should be the numerical indicator. i.e. "1" for Video 1, "2" for Video 2, etc.	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_Supplementary_Video_1.mov*	Legend or Descriptive Caption Describe the contents of the file
Choose an item.			
Choose an item.			
Choose an item.			
Choose an item.			
Choose an item.			
Choose an item.			

Add rows as needed to accommodate the number of files.

3. Source Data

Complete the Inventory below for all Source Data files.

Figure	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_SourceData_Fig1.xls*, or *Smith_Unmodified_Gels_Fig1.pdf*	Data description i.e.: Unprocessed Western Blots and/or gels, Statistical Source Data, etc.
Source Data Fig. 1	All source data is available at: https://github.com/MWhite-InstitutPasteur/Pvivax_sero_dx	Sources and analyses scripts required to reproduce all figures and tables.
Source Data Fig. 2		

Source Data Fig. 3		
Source Data Fig. 4		
Source Data Fig. 5		
Source Data Fig. 6		
Source Data Fig. 7		
Source Data Fig. 8		
Source Data Extended Data Fig. 1		
Source Data Extended Data Fig. 2		
Source Data Extended Data Fig. 3		
Source Data Extended Data Fig. 4		
Source Data Extended Data Fig. 5		
Source Data Extended Data Fig. 6		
Source Data Extended Data Fig. 7		
Source Data Extended Data Fig. 8		
Source Data Extended Data Fig. 9		
Source Data Extended Data Fig. 10		

Title: Development and validation of serological markers for detecting recent *Plasmodium vivax* infection

Short title: Serological exposure markers for *Plasmodium vivax*

Authors: Rhea J. Longley[1,2,3]†, Michael T. White[4]†, Eizo Takashima[5], Jessica Brewster[1], Masayuki Morita[5], Matthias Harbers[6,19], Thomas Obadia[4,20], Leanne J. Robinson[1,2,7], Fumie Matsuura[6], Zoe SJ Liu[1,2], Connie S. N. Li-Wai-Suen[1,2], Wai-Hong Tham[2,8], Julie Healer[2,8], Christele Huon[9], Chetan E. Chitnis[9], Wang Nguitragool[10], Wuelton Monteiro[11,12], Carla Proietti[13,14], Denise L. Doolan[13,14], Andre M. Siqueira[15], Xavier C. Ding[16], Iveth J. Gonzalez[16], James Kazura[17], Marcus Lacerda[11,18], Jetsumon Sattabongkot[3], Takafumi Tsuboi[5], Ivo Mueller[1,2,4]*

Affiliations:

[1]Population Health and Immunity Division, Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia.

[2] Department of Medical Biology, University of Melbourne, Melbourne, Australia.

[3]Mahidol Vivax Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand.
[4]Unité Malaria: Parasites et Hôtes, Département Parasites et Insectes Vecteurs, Institut Pasteur, Paris, France.
[5]Division of Malaria Research, Proteo-Science Center, Ehime University, Matsuyama, Japan.
[6]CellFree Sciences Co., Ltd., Yokohama, Japan.
[7]Burnet Institute, Melbourne, Australia.
[8]Infection and Immunity Division, Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia
[9]Malaria Parasite Biology and Vaccines, Department of Parasites & Insect Vectors, Institut Pasteur, Paris, France.
[10]Department of Molecular Tropical Medicine and Genetics, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand.
[11]Fundacão de Medicina Tropical Dr. Heitor Vieira Dourado, Manaus, Brazil.
[12]Universidade do Estado do Amazonas, Manaus, Brazil.
[13]Centre for Molecular Therapeutics, Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, Australia.
[14]QIMR Berghofer Medical Research Institute, Brisbane, Australia.
[15]Instituto Nacional de Infectologia Evandro Chagas – Fiocruz, Rio de Janeiro, Brazil
[16]Foundation for Innovative New Diagnostics, Geneva, Switzerland.
[17]Center for Global Health and Diseases, Case Western Reserve University, Cleveland, United States.
[18]Instituto Leônidas & Maria Deane (Fiocruz), Manaus, Brazil.
[19]RIKEN Center for Integrated Medical Sciences (IMS), Yokohama, Japan.
[20]Hub de Bioinformatique et Biostatistique, Département Biologie Computationnelle, Institut Pasteur, USR 3756 CNRS, Paris, France.

*To whom correspondence should be addressed: mueller@wehi.edu.au.
†These authors contributed equally.

Abstract

A major gap in the *Plasmodium vivax* elimination toolkit is the identification of individuals carrying clinically silent and undetectable liver-stage parasites, called hypnozoites. This study developed a panel of serological exposure markers capable of classifying individuals with recent *P. vivax* infections who have a high likelihood of harboring hypnozoites. We measured IgG antibody responses to 342 *P. vivax* proteins in longitudinal clinical cohorts conducted in Thailand and Brazil and identified candidate serological markers of exposure. Candidate markers were validated using samples from yearlong observational cohorts conducted in Thailand, Brazil and the Solomon Islands, and antibody responses to eight *P. vivax* proteins classified *P. vivax* infections in the prior 9-months with 80% sensitivity and specificity. Mathematical models demonstrate that a serological testing and treatment strategy could reduce *P. vivax* prevalence by 59% – 69%. These eight antibody responses can serve as a biomarker identifying individuals who should be targeted with anti-hypnozoite therapy.

Introduction

Elimination of malaria by 2030 is now the explicit goal of all malaria endemic countries in the Asia-Pacific and the Americas. While impressive progress towards this goal has been made, global funding for malaria control has remained unchanged since 2010 and progress has stalled in many parts of the world[1]. New interventions and tools for better targeting of limited resources are urgently needed.

A major hurdle for elimination is the increasing proportion of malaria cases caused by *Plasmodium vivax* as malaria endemicity declines[1,2]. *P. vivax* has unique biological features that make its control difficult[3], including high prevalence of low density, asymptomatic infections[4] and a liver-stage that can reactivate weeks to months after initial infection resulting in relapses that cause morbidity and sustain transmission. These hypnozoites, undetectable with current diagnostics, can be responsible for >80% of all blood-stage infections[5]. Identifying and targeting individuals with hypnozoites is thus essential for accelerating and achieving malaria elimination. In addition, as endemicity decreases, malaria transmission becomes increasingly fragmented and highly seasonal[6,7] rendering blanket approaches to malaria control and elimination inefficient. This requires new, innovative tools and approaches specifically designed to assist in targeting interventions to the changing malaria epidemiological landscape[8].

National malaria control programs rely almost exclusively on microscopy or rapid diagnostic tests for routine detection of malaria cases at health clinics, and for surveillance using mass blood surveys[9]. These tools have limited sensitivity for detecting individuals with low density asymptomatic infections[10], making it difficult for control programs to efficiently identify areas of low and high *P. vivax* transmission and target their resources accordingly. Molecular techniques such as PCR have greater sensitivity[10] but are rarely implemented by control programs due to high cost and the need for specialised laboratories. All of these methods can only detect

individuals with a current blood-stage infection, rendering mass screening and treatment (MSAT) approaches ineffective for reducing *P. vivax* transmission because they fail to treat individuals that only carry hidden *P. vivax* liver-stage infections[5,11] with no circulating blood-stage parasitemia. Mass drug administration (MDA) is predicted to be a highly effective control tool but only if it includes a drug that targets *P. vivax* hypnozoites[5]. Unfortunately, primaquine and tafenoquine, the only current drugs able to eliminate hypnozoites, have toxic side effects in glucose-6-phosphate dehydrogenase (G6PD) deficient individuals[12], limiting their acceptability for MDA campaigns, particularly in low transmission areas where >90% recipients will have no direct benefit from the treatment. There is thus a diagnostic gap, with MSAT ineffective due to undertreatment and MDA unacceptable due to overtreatment. A strategy that efficiently targets individuals with current blood-stage infections and those with hypnozoites is required.

Blood-stage *P. vivax* infections induce robust IgG antibody responses to a broad range of *P. vivax* proteins, even following low-density asymptomatic infections[13]. These responses can be long-lived, even after clearance of blood-stage infection[14,15]. Hence, antibodies are markers of past exposure as well as current infection[15]. Serological exposure markers (SEM) have been used for surveillance of malaria and a number of other infectious diseases (i.e. leishmaniasis, influenza, trachoma), and antibody responses can be simply and cheaply measured in point-of-care tests[16-18]. SEMs may also be used for risk stratification and as guidance for targeted malaria control and elimination interventions[19]. For *P. vivax* there is an important additional, individual-level application for SEMs: to target treatment to people at-risk of carrying clinically silent hypnozoites. While it is not possible to directly detect hypnozoites with current technology, all tropical and sub-tropical *P. vivax* strains cause a primary infection followed by a first relapse after no more than six to nine months[20-22]. Therefore, any individual with a blood-stage infection within the previous nine months, who has not received anti-liver-stage drugs, is likely to be a hypnozoite carrier. A carefully selected panel of proteins inducing antibodies that signify

exposure within the previous nine months could provide an accurate measure of current transmission and also highlight individuals likely to harbour hypnozoites who could be targeted for treatment with anti-hypnozoite drugs after serological testing ("serological testing and treatment", seroTAT).

Here we screened 342 *P. vivax* proteins for their ability to detect recent *P. vivax* infections, and validated their use in malaria-endemic regions of Thailand, Brazil and the Solomon Islands.

Results

Antigen discovery phase

Data on the magnitude and longevity of IgG responses to 342 *P. vivax* proteins following symptomatic *P. vivax* infections[23] were used to identify suitable proteins for detecting recent exposure. IgG was measured in four longitudinal plasma samples collected over a 9 month period from individuals in Thailand (n=32) and Brazil (n=33) to enable estimation of antibody half-life[23] (Table S1).

A down-selection pipeline was developed to identify candidate serological markers using this data (Fig 1). *P. vivax* proteins were first prioritised, with selection of those that had similar estimated IgG half-lives in both antigen discovery cohorts (Thailand and Brazil) and were highly immunogenic at the time of infection (≥50% seropositive individuals) (Fig 1A). Using these, 142 of the 342 *P. vivax* proteins were prioritised as having suitable characteristics for candidate SEMs.

Fig 1. Flow diagram of antigen down-selection pipeline. The studies and samples used are listed to the left in (A) and (B), whereas the pipeline and number of proteins included at each stage are listed to the right.

A statistical model for estimating time since infection was used to test the ability of the 142 prioritised proteins, individually or in combination, to predict time since last infection (see Supplementary Methods and Supplementary Figs S1-3). Fig 2A-B shows examples of antibody kinetics to five proteins in two representative individuals from Thailand and Brazil. Fig 2C-D shows the estimated time since last infection with uncertainty, for antibody responses measured to the five proteins in Fig 2A-B at six months after infection, for the same two representative individuals. Increasing the number of proteins to at least five resulted in higher accuracy compared to using a single protein alone, with only small incremental improvements in the accuracy beyond 20 proteins (Fig S4). A simulated annealing algorithm was employed to determine optimal combinations of proteins that maximise the likelihood of correctly estimating the time since last infection (Fig 2E). Whilst some proteins had 100% probability of being included in a successful panel of a set size (Fig 2E), there was redundancy in choice of additional proteins and hence all were ranked based on their probability of inclusion in a 20-protein panel (given the limited improvements beyond 20 proteins). The 24 highest ranked proteins were selected for further testing.

Fig 2. Antigen discovery phase: estimating time since last *P. vivax* infection. (A, B) Examples of antibody kinetics to five proteins in individuals from Thailand and Brazil. Points represent measured antibody responses in triplicate on the AlphaScreen™ platform, and solid lines depict the fit of mixed-effects linear regression models. Each colour corresponds to one of the five proteins (PVX_097625, PVX_097680, PVX_112655, PVX_003905, PVX_123685-03) selected to provide optimal classification performance. (C, D) Using data on measured antibody responses six

months after infection, the black curve shows the model estimated time since infection using the antibody responses shown in (A, B). The square point denotes the most likely estimate and dashed vertical lines denote the 95% confidence interval. In both cases the confidence intervals span six months. (E) Results of simulated annealing algorithm for estimating the probability that combinations of antigens predict time since last infection in the antigen discovery data sets. Colour is representative of the inclusion probability (red 100%, white 0%). The five proteins with data shown in A-D are highlighted in colour in panel E. These results came from the second round of down-selection on 104 *P. vivax* proteins (see Appendix section 1 for further details).

The wheat germ cell free (WGCF) system was selected for expression of the proteins due to its eukaryotic nature and past success in expressing *Plasmodium* proteins[24]. Despite all proteins being expressed in this system previously as crude proteins[23] not all could be produced at high yield and high purity, and thus an additional round of down-selection using ranking from the simulated annealing algorithm was performed to increase the number of proteins retained at this point in the pipeline. The algorithm was run on 104 of the 142 prioritised proteins, excluding the top 24 already selected (to provide more discriminatory power for selecting additional proteins) and 14 proteins known to have low yields or form aggregates from our previous work. An additional 31 proteins were selected in the second round using the same methodology, with 55 *P. vivax* proteins in total identified as suitable candidate SEMs. Of these, 40 were successfully produced at high yield and purity using the WGCF system (Table S2).

An additional 20 proteins known to be highly immunogenic were added to these 40 proteins[25-28]. This included four proteins already within the down-selected panel of 40 (different constructs or produced by a different laboratory and protein expression systems), and three distinct constructs of PVX_110810 (Duffy binding protein, each representing a different region of the protein or sequence from a different strain) (Table S2).

Validation phase

The 60 candidate SEMs were validated in larger geographically diverse cohorts with a known history of malaria infections in the preceding 12 months, using plasma samples from three yearlong observational cohort studies in Thailand, Brazil and the Solomon Islands (see online Methods and Extended Data Figure 1). Individuals in these cohorts were assessed monthly for *Plasmodium spp.* by qPCR, with continuous concurrent passive case detection at local malaria clinics and hospitals. Plasma samples from the last visit of these cohorts (n=829, 928 and 754 in Thailand, Brazil and the Solomon Islands, respectively) were used to measure IgG responses in a multiplex Luminex® assay to the 60 *P. vivax* proteins. 158/2511 individuals in these cohorts had a concurrent *P. vivax* infection, detected by qPCR, at the time the plasma was collected; IgG antibody levels were strongly associated with current infection status for each of the 60 proteins in the Thai cohort (OR 2·1-7·7, p<0·05) and most proteins in the Brazilian cohort (57/60, OR 1·5-7·1, p<0·05) (Table S3). This association was not as clear for the paediatric Solomon Islands cohort, with IgG levels to only 29 out of 60 proteins significantly associated with current *P. vivax* infections (OR 1·7-6·3, p<0·05). Overall, there was a pattern of decreasing IgG magnitude with increasing time since last *P. vivax* infection (Fig 3A-H, Extended Data Figure 2), with minimal reactivity in malaria naïve negative control individuals from Bangkok, Rio de Janeiro and Melbourne.

Fig 3. Validation phase. (A-H) Antibody responses to eight antigens measured in n = 2,281 biologically independent samples on the Luminex® platform, stratified by geographical location and time since last PCR detected infection. Boxplots denote median and interquartile ranges (IQR) of the data, with whiskers denoting the median ± 1.5*IQR. (I) Receiver operating characteristic (ROC) curve for classifying individuals as infected within the last nine months using a threshold antibody titre to a single antigen. Coloured curves correspond to those proteins in panels A-H. (J) Spearman correlation between antibody titres measured in 2,281 samples. Correlation coefficients are indicating by the colour scale, with red being 100% correlation and dark green 0% correlation. (K) 2-dimensional distribution of anti-PVX_099980 and anti-PVX_094255 antibody titres.

Recent infections were defined as PCR-confirmed infections occurring within the past nine months. Measurements of IgG antibodies against a number of individual proteins were able to classify individuals as infected with *P. vivax* within the past nine months or not, with differing degrees of accuracy (Fig 3I). Protein PVX_094255 (reticulocyte binding protein 2b, RBP2b) reached 75% sensitivity and 75% specificity when used alone. IgG levels to the 60 candidate SEMs were correlated (Fig 3J) with different distributions evident between the three geographic regions (Fig 3K). A linear discriminant analysis (LDA) classification algorithm was used to identify combinations of the 60 candidate SEMs that could accurately predict recent infections. Redundancy was found with multiple combinations of *P. vivax* proteins able to accurately predict recent infection when included in panels of up to eight proteins (Fig S5). The top eight most frequently identified proteins when used in combination were: PVX_094255 (RBP2b[29]), PVX_087885 (rhoptry associated membrane antigen, RAMA[27], putative), PVX_099980 (merozoite surface protein 1, MSP1[30]), PVX_096995 (tryptophan-rich antigen (Pv-fam-a), PvTRAg_2[31]), PVX_097625 (MSP8[32], putative), PVX_112670 (unspecified product previously annotated as a tryptophan-rich antigen (PvTRAg_28[31]), KMZ83376.1 (erythrocyte binding protein II, EBPII[33]), and PVX_097720 (MSP3.10[34]) (see Fig 3I for ROC curves, Table 1 for protein details and Fig S6 for network analysis). Table S2 shows the individual ranking of all 60 proteins by AUC, note that it is not necessarily the proteins with the highest individual AUC that work best in combination.

Table 1. Top *P. vivax* proteins identified in the validation phase and their individual performance. Area under the curve (AUC) from the single antigen classifier is shown; proteins are listed in order of best individual performance. Nine of 60 proteins are shown here; the top 8 identified as best used in combination, plus the alternate construct of RBP2b. See Table S2 for the extended and complete Table. References listed are for the protein production and purification method.

Protein ID[a]	Gene Annotation[a]	Protein length, aa	Construct, aa (size)	Expression System	Purification Method	AUC
PVX_094255B	reticulocyte binding protein 2b (RBP2b)	2806	161-1454 (1294)	*E. coli*	2x affinity + size exclusion[26,29,35]	0.816
PVX_094255A	reticulocyte binding protein 2b (RBP2b)	2806	1986-2653 (667)	WGCF	One-step Ni column	0.748
PVX_087885B	rhoptry-associated membrane antigen, putative	730	462-730 (269)	WGCF	One-step Ni column[27]	0.728
PVX_099980	merozoite surface protein 1 (MSP1-19)	1751	1622-1729 (108)	WGCF	One-step Ni column	0.713
PVX_112670	unspecified product	335	34-end (302)	WGCF	One-step Ni column	0.698
PVX_096995	tryptophan-rich antigen (Pv-fam-a)	480	61-end (420)	WGCF	One-step Ni column	0.689
KMZ83376.1[b]	erythrocyte binding protein II (PvEBPII)	786	109-432 (324)	*E. coli*	Ni, ion exchange, gel filtration[25,36]	0.684
PVX_097625	merozoite surface protein 8 (MSP8), putative	487	24-463 (440)	WGCF	One-step Ni column	0.681
PVX_097720	merozoite surface protein 3 (MSP3.10)	852	25-end (828)	WGCF	One-step Ni column	0.670

[a]PlasmoDB release 36 (http://plasmodb.org/plasmo/), [b]GenBank

As shown in Table 1, for the top protein PVX_094255 (RBP2b), there were two separate constructs mapping to different regions of the protein ($RBP2b_{161-1454}$ and $RBP2b_{1986-2653}$), expressed and purified using different methods[29,37]. Antibody levels to these two constructs ($RBP2b_{161-1454}$ and $RBP2b_{1986-2653}$) were highly correlated (spearman r=0·72, p<0·0001, all cohorts combined) and for this reason the construct that provided lower levels of classification accuracy ($RBP2b_{1986-2653}$) was excluded from the top eight. $RBP2b_{161-1454}$ induced very low levels of antibody reactivity in the malaria naïve control panels. Across all candidate SEMs, a significant negative association was observed between the mean antibody levels detected in the malaria naïve negative control panels (n=274 individuals, see Methods) and the AUC values of the 60 SEMs determined from their individual receiver operating characteristic (ROC) curves (spearman r=-0.5, p<0.0001, Extended Data Figure 3. Removing data from these malaria naïve control participants did not cause any substantial reductions in classification accuracy for the top eight proteins (Fig S7).

Classification performance of an eight-protein SEM panel

Fig 4A-D presents ROC curves for assessing the classification performance of the top panel of eight SEMs for identifying individuals with exposure to *P. vivax* within the prior nine months. There are only incremental improvements in classification performance as the number of proteins is increased, with a plateau of 80% sensitivity and 80% specificity reached in all three geographic regions with five proteins. The algorithm correctly classifies more than 97% of malaria naïve negative controls, but classification performance was poor for individuals who had their last blood-stage *P. vivax* infection 9-12 months ago, with 40-66% of samples misclassified (Extended Data Fig 4). Incorporating an individual's age into the classification algorithm did not result in substantial improvements in classification accuracy (Fig S9). The cohort studies were conducted in regions co-endemic for *P. vivax* and *P. falciparum*, although the total number of individuals experiencing a *P. falciparum* infection during the study period was low (n=19, 31, 22

for Thailand, Brazil and the Solomon Islands, respectively). No associations were observed between recent *P. falciparum* infections and antibody level to the top eight proteins (Fig S10).

We assessed the potential performance of targeted treatment using a seroTAT approach (with 80% sensitivity and specificity) compared to both MDA and MSAT with PCR (Fig 4E-L). In an MDA campaign, all hypnozoite carriers are targeted (except those ineligible due to G6PD deficiency), but >80% of the population receive unnecessary primaquine (Fig 4H). In contrast for MSAT with PCR, no individual was over-treated (it is assumed that an individual with blood-stage *P. vivax* is a likely hypnozoite carrier). However, only 20%-40% of all likely hypnozoite carriers are targeted. Using seroTAT, at least 80% of individuals with hypnozoites are targeted (outperforming MSAT with PCR) (Fig 5H), with <20% of the population treated unnecessarily (substantially less than an MDA approach). We also assessed the potential performance of seroTAT using two alternate strategies: 50% sensitivity/98% specificity and 98% sensitivity/50% specificity. The high specificity approach only slightly outperforms MSAT in identifying hypnozoites carriers (45% missed) and is thus likely not feasible for seroTAT. The high sensitivity approach identifies nearly all likely hypnozoites carriers and reduces overtreatment to approximately 50%, compared to >80% with MDA. The high sensitivity approach identifies nearly all likely hypnozoites carriers and reduces overtreatment to approximately 50%, compared to >80% with MDA.

Fig 4. Performance of the top eight *P. vivax* proteins for classifying recent infection and modelled ability to identify hypnozoites carriers. Cross-validated ROC curves generated from the composite algorithm classifier for identifying individuals with PCR detected infection in the last nine months in (A) Thailand; (B) Brazil; (C) The Solomon Islands; and (D) all regions combined. The coloured triangles denote three different sensitivity and specificity targets for

serological screening and treatment (SeroTAT) strategies. (E-H) Proportion of individuals targeted for primaquine treatment under a range of mass treatment strategies. (I-L) Proportion of individuals who had no exposure to *P. vivax* during the previous nine months who would have been administered primaquine.

Capability of an eight-protein SEM panel to identify individuals at high risk of recurrent infection

IgG antibody responses were measured in plasma samples from the first visit of the three yearlong cohort studies against the top eight proteins using our established Luminex® assay (the validation phase used samples from the last visit). Antibody responses against these proteins were input into the classification algorithm to classify individuals as either sero-negative (i.e. not recently exposed to *P. vivax*) or sero-positive (exposed to *P. vivax* during the past 9-months) at the baseline visit. As shown in Fig 5A-C, sero-positive individuals at baseline subsequently developed PCR-detectable infections at far higher rates than sero-negative individuals, with statistically significant hazards ratios in the range of 3.23 – 8.55. This confirms the ability of the eight-panel SEM to identify individuals at high risk of developing recurrent infections.

The potential public health impact of a seroTAT strategy was assessed using a mathematical model of *P. vivax* transmission[38] (Fig 5D-F). PCR prevalence was based on monthly data from our 2013-2014 yearlong observational cohort studies and assumed to provide baseline *P. vivax* transmission levels for population-level treatment strategies beginning in 2020. The public health impact of three treatment strategies was simulated: MDA, MSAT with light microscopy, and seroTAT with 80% sensitivity and specificity. Each strategy was modelled to be implemented at 80% population coverage for two rounds with a primaquine regimen assumed to clear all hypnozoites. Two yearly rounds of MDA resulted in an estimated 70% - 84% reduction in *P. vivax* PCR prevalence across all three study sites whilst MSAT with LM was predicted to

have substantially less impact with a 33% - 45% reduction. SeroTAT was not as effective as MDA but resulted in an estimated 59% - 69% reduction in PCR prevalence.

Fig. 5. Pilot application phase. (A-C) Kaplan-Meier analyses of time to first *P. vivax* blood-stage infection detected by PCR. Using measured antibody responses at the first time point, participants were classified as positive (blood-stage infection within the past 9 months) or negative. n = 985 biologically independent samples in Thailand, n = 1,207 in Brazil; n = 754 in Solomons. Hazards ratio (HR) were calculated using Cox proportional hazards, and were statistically significant with $P < 10^{-10}$ in all cases. Solid lines denote the proportion uninfected by PCR, and shaded regions denote the 95% confidence intervals. (D-F) Prediction from a mathematical model of *P. vivax* transmission of the effect of population-level treatment strategies with primaquine at 80% coverage. Solid lines denote model predictions and shaded regions denote 95% uncertainty intervals due to stochasticity. The model was calibrated to longitudinal data on PCR prevalence, and the potential impact of two rounds of treatment in 2020 and 2021 were simulated. The percentage reduction in transmission was calculated as the estimated prevalence in June 2021 compared to June 2019. Red points denote measured PCR prevalence with 95% confidence intervals. Estimates of PCR prevalence based on n = 12,829 measured samples in Thailand, n = 13,973 samples in Brazil, and n = 8,276 samples in the Solomons.

Discussion

New tools and strategies to directly target *P. vivax* are urgently needed if malaria elimination in the Asia-Pacific and Americas is to be achieved by 2030[39]. *P. vivax* presents a unique challenge to elimination due to the presence of undetectable hypnozoites that contribute to maintaining residual transmission. *P. vivax* relapses are expected to occur at a frequency of every one to two months[21,22]. Apart from temperate 'hibernans' strains that are now restricted to the Korean

Peninsula and have no primary blood-stage infections[20] virtually all individuals who carry hypnozoites will have had a *P. vivax* blood-stage infection within the prior nine months[20-22]. Herein, a novel panel of candidate serological exposure markers ("SEMs") were identified and validated that allow detection of recent exposure to *P. vivax* within the prior nine-months. To our knowledge, these SEMs represent the first method that can indirectly identify likely hypnozoites carriers that could be targeted for treatment with liver-stage drugs.

We undertook a relatively unbiased approach to choosing the best markers by screening a large panel of 342 *P. vivax* proteins, and strategically selecting those that can predict recent infection based on immunogenicity and antibody half-lives. The final panel, which has a sensitivity and specificity of 80% at identifying individuals with PCR-detectable blood-stage infection in the last nine months in three geographically distinct regions, incorporates antibody responses to eight *P. vivax* proteins: PVX_094255 ($RBP2b_{161-1454}$), PVX_087885 (RAMA, putative), PVX_099980 (MSP1), PVX_096995 (tryptophan-rich antigen (Pv-fam-a)), PVX_097625 (MSP8, putative), PVX_112670 (unspecified product), KMZ83376.1 (EBPII), and PVX_097720 (MSP3.10). Most of these *P. vivax* proteins are not well described in the literature and only MSP1 and RAMA have previously been used, or suggested as, markers of exposure[27,40-42]. Our strategy was thus successful in identifying novel exposure markers.

Individually, PVX_094255 ($RBP2b_{161-1454}$) was able to classify exposure with 75% sensitivity and specificity. This may be sufficient for sero-surveillance and community-level targeting of preventative measures such as intensified vector control or focal test and treat[19]. Greater classification accuracy will be required for individual-level targeting for treatment with liver-stage drugs, for example in an elimination campaign using "serological testing and treatment" (seroTAT), where the goal is to treat all individuals who have hypnozoites or blood-stage *P. vivax* parasites. The final panel of eight proteins obtained 80% sensitivity and specificity across

three geographic regions. Using antibody responses against these eight proteins, we show that individuals classified as sero-positive at baseline in our yearlong observational cohort studies subsequently develop PCR-detectable *P. vivax* infections at a far higher rate than their sero-negative counterparts. Whilst in these cohorts we cannot distinguish whether these *P. vivax* infections are due to relapses or new mosquito bite-derived primary infections, this data clearly demonstrates the ability of our SEMs to identify individuals at-risk for recurrent infections.

Many factors affect anti-malarial antibody responses, most notably time since last infection, intensity of infection, and age. Antibody responses may also be associated with other factors such as human host or parasite genetic variation that are challenging to account for. Key properties of the antibody response data are the high level of individual variation and the high degree of correlation. Our analytic pipeline accounts for these data properties, and the possibility that combinations of antibody responses may best identify recent infections. However, rather than past infections being identified by complex antibody signatures and sophisticated algorithms, a number of simple factors contributed to good classification performance: (i) *P. vivax* proteins that can identify recent infections when used individually also do well in combinations; (ii) only small incremental improvements in classification performance occur as new *P. vivax* proteins are added to combinations; (iii) there are diminishing returns to algorithmic complexity: simple algorithms such as logistic regression capture most signal in the data, with more complex classifiers such as random forests providing only incremental improvements; (iv) there is no single best combination of antigens, there is redundancy; (v) there was no substantial advantage to including an individual's age once three or more antigens were included; and (vi) algorithms and combinations of antigens that provided good classification in one region also performed well in another region.

Although the performance of the pilot marker panel may seem modest at 80% sensitivity and specificity, several factors need to be taken into account when evaluating the performance of the SEMs. Firstly, given that individual *P. vivax* blood-stage infections can be relatively short (i.e. <4 - 6 weeks)[43], some blood-stage infections may have been missed in the validation cohorts as they had only monthly active follow-up. In the analyses, such cases would be classified as false positive by the SEMs, when in reality the SEMs would have accurately detected these infections. Therefore, the real specificity of the test may well be higher. Similarly, a number of people (5-20%) with PCR-positive infections at the time of antibody measurement were missed. This may represent individuals with very recent infections who have not yet generated a significant rise in antibody titres[44]. These missed, concurrent infections could however be detected using an ultra-sensitive antigen capture test. A combination of antibody detection and antigen-capture assays would further increase the sensitivity of the diagnostic approach. Lastly, the SEMs are designed for application in low-transmission malaria-endemic regions progressing towards elimination. Our data has previously shown that antibody responses are longer lasting in areas with higher transmission[23]. In elimination contexts, overall (population) levels of immune responses to malarial antigens are low[23,45], potentially increasing the difference between recently exposed and non-recently exposed individuals.

In the three settings tested, classification performance was greatest in Thailand, the region with lowest transmission. The association of antibody levels with current *P. vivax* infections was less clear in the Solomon Islands compared to Thailand and Brazil, and the magnitude of these responses were also generally lower. In contrast to the Thai and Brazilian cohorts that included individuals of all ages, the Solomon Islands was a paediatric cohort. Antibody levels tend to increase with age in malaria-endemic regions[13,28], and so the lower responses in the Solomon Islands may be due to the design of the cohort including only children. However, the difference in antibody responses could also be influenced by other factors such as genetic diversity of *P.*

511 *vivax* parasite proteins. Importantly, despite the lower magnitude of antibody responses, the top
512 eight SEMs could still accurately classify recent infections in the Solomon Islands.
513
514 To further improve classification performance, modifications of the *P. vivax* protein constructs
515 resulting in reduced background in malaria naïve control samples would be advantageous, as
516 proteins with lower levels of antibody reactivity in these controls had greater classification
517 performance. Addition of further purification steps, testing of other protein expression systems,
518 design of smaller protein fragments, investigations of antigenic diversity and strain-specificity of
519 antibody responses, could all be considered. Further information could potentially be gained by
520 looking at antibody subclasses and/or IgM. Whilst antibody subclass responses will likely
521 correlate with total IgG[46], the relationship with IgM is expected to be weaker[46] and acquisition
522 and breadth of responses to malaria proteins has been shown to differ for IgG and IgM[47]. Such
523 optimisation and improvement of our assay may allow selection of a smaller panel of three to
524 five proteins with comparable classification accuracy to our current eight protein panel. This
525 would reduce the cost of running our SEMs in their current format (Luminex® assay) and make
526 a simpler point-of-care test feasible.
527
528 The SEMs' current level of performance and their application in potential public health
529 interventions such as seroTAT needs to be seen in the context of the alternatives. MDA (with
530 anti-hypnozoite drugs[5]) is a blanket approach that while effectively targeting all hypnozoite
531 carriers, results in substantial overtreatment. In low transmission settings, up to 90% of people
532 treated do not carry hypnozoites (i.e. a specificity of only 6.5-23.3%, Fig 5) and do not derive
533 any direct benefit from a treatment that may carry risks. Therefore, few *P. vivax* endemic
534 countries are considering MDA with primaquine. Conversely MSAT interventions, even with a
535 highly sensitive molecular test, miss 64.9-80.5% of hypnozoites carriers (Fig 5) [5,11]. Testing and
536 treatment with SEMs thus results in less overtreatment compared to MDA, whilst still targeting a

higher proportion of likely hypnozoites carriers compared to MSAT. Through modelling we show that implementation of seroTAT could result in 59% - 69% reductions in *P. vivax* prevalence in our cohort settings. While preliminary, these modelling results indicate that seroTAT has potential to be a highly promising new intervention to assist control programs in accelerating towards *P. vivax* elimination.

Similar efforts to develop SEMs are currently underway for *P. falciparum*[15] and efforts to develop SEMs for both species should be coordinated where possible. In addition, there is evidence of antibody cross-reactivity between *P. falciparum* and *P. vivax* [48]. Five of the top eight proteins have *P. falciparum* orthologs but extensive cross-reactivity is unlikely given the relatively low sequence homology (ranging from 20.7-37.4%). Indeed, no association was detected between recent *P. falciparum* infections and antibody levels to *P. vivax* SEMs in our validation. This may not hold for *P. ovale*, *P. malariae* or *P. knowlesi,* which are more closely related to *P. vivax* (protein sequence identity 13.7-83.4%). On a programmatic level, limited cross-reactivity between human malaria species may not be an issue in areas outside Africa as rates of *P. vivax* relapse after *P. falciparum* infections are very high[49,50], sometimes as high as following a *P. vivax* infection. Given the marked difference in patterns of exposure to the zoonotic *P. knowlesi*[51], cross-reactivity between *P. vivax* and *P. knowlesi* would be significantly more problematic[52].

In summary, the panel of eight novel SEMs can accurately predict recent *P. vivax* infection. As almost all *P. vivax* relapses occur within nine months of a previous blood-stage infection[53], these markers can indirectly identify individuals with the highest risk of harbouring hypnozoites. With our markers able to identify individuals at-risk of future recurrent infections, we demonstrate that this is indeed the case. We have shown that a carefully selected panel of SEMs can specifically detect recent exposure and could be used in a programmatic setting

563 ("surveillance as an intervention"[54]). Application of our SEMs for seroTAT holds promise for an
564 effective elimination strategy using primaquine or tafenoquine to target dormant hypnozoites.
565 Given the risk of haemolysis in G6PD-deficient individuals treated with primaquine or
566 tafenoquine, our tool has the potential to ensure such elimination strategies are targeted and
567 therefore a safer, more acceptable and more effective option in malaria-endemic communities.
568
569

Online Methods

Study design

The goal of the study was to identify and validate suitable candidate *P. vivax* proteins for use as serological markers of recent exposure to *P. vivax* infections. This study was conducted in three parts: i) an antigen discovery phase utilising samples from two longitudinal cohorts that followed symptomatic *P. vivax* malaria patients over nine months, ii) a validation phase utilising samples from three one-year long observational cohort studies, and iii) a pilot application phase to demonstrate the ability of SEMs to identify at-risk individuals. The sample size was predefined by availability of suitable plasma samples with matching epidemiological and molecular data.

Study populations: antigen discovery phase

Patients with confirmed *P. vivax* malaria (by qPCR, see below) were enrolled from local malaria clinics and hospitals in two sites in 2014: Tha Song Yang, Tak Province, Thailand, and Manaus, Amazonas State, Brazil. After receiving anti-malarial drug treatment according to respective National Treatment Guidelines, and providing written informed consent and/or assent, blood samples were taken over a period of nine months as previously described[23] (Extended Data Fig 1a, Table S4). Presence or absence of *Plasmodium spp.* parasites during follow-up was determined by both microscopy and quantitative PCR. Blood samples were collected at enrolment and week 1, then every two weeks for six months, then every month until the end of the study. A subset of enrolled participants who had no evidence of recurrent parasitaemia during follow-up was selected for antibody measurements (n=32 Thai patients, n=33 Brazilian patients).

Study populations: validation phase

Year-long longitudinal observational cohort studies were conducted over 2013-2014 in three sites: Kanchanaburi/Ratchaburi provinces, Thailand[55], Manaus, Brazil (Monteiro et al., in preparation), and the island of Ngella, Solomon Islands[56] (Extended Data Fig 1B, Table S4). 999 volunteers were enrolled from Thailand and sampled every month over the yearlong cohort, with 14 active case detection visits performed in total. 829 volunteers attended the final visit. 1,274 residents of all age groups were enrolled from Brazil and sampled every month over the year-long period, with 13 active case detection visits performed in total. 928 volunteers attended the final visit with plasma from 925 available. 1,111 children were enrolled from the Solomon Islands, with 860 used for final analysis of the cohort (after exclusion of children who withdrew, had inconsistent attendance, or failed to meet other inclusion criteria). The children were sampled monthly, with 11 active case detection visits in total. Of the 860 children, 754 attended the final visit. For all three cohorts, at each visit volunteers completed a brief survey outlining their health over the past month (to determine the possibility of missed malarial infections), in addition to travel history and bed net usage. A finger-prick blood sample was also taken and axillary temperature recorded. Passive case detection was performed throughout the yearlong period. All cohort participants provided individual consent for both participation in study and the future use of samples for the study of antimalarial immune responses. In cases of children, parental consent and (where appropriate) assent was obtained.

Study populations: pilot application phase

To assess the ability of our SEM to identify individuals at risk of future recurrent *P. vivax* infections, we utilised plasma samples from the first visit of the three year-long observational cohorts used in the validation phase (Extended Data Fig 1B). Antibody measurements were

made in 992 plasma samples from the Thai cohort, 1207 samples from the Brazilian cohort, and 655 samples from the Solomon Islands cohort.

Study populations: malaria naïve control panels

Four panels of malaria naïve control plasma samples were collected from individuals with no known recent exposure to malaria (Table S4). Samples were as follows: 102 individuals from the Volunteer Blood Donor Registry (VBDR) in Melbourne, Australia; 100 individuals from the Australian Red Cross (ARC), Melbourne, Australia; 72 individuals from the Thai Red Cross (TRC), Bangkok, Thailand; and 96 individuals from the Rio de Janeiro State Blood Bank (RBB), Rio de Janeiro, Brazil, an area non-endemic for malaria since the 1960s. Standard TRC and RBB screening procedures exclude individuals from donating blood if they have had a confirmed malaria infection within the previous three years or have travelled to malaria-endemic regions within the past year.

Ethical approvals

The Ethics Committee of the Faculty of Tropical Medicine, Mahidol University, Thailand, approved the Thai antigen discovery study (MUTM 2014-025-01 and 02) and the Thai yearlong cohort study (MUTM 2013-027-01). The Ethics Review Board of the *Fundação de Medicina Tropical Dr Heitor Vieira Dourado* (FMT-HVD) approved the Brazilian antigen discovery study ((957.875/2014) and the Brazilian yearlong cohort study (349.211/2013). The Brazilian yearlong cohort study was also approved by the Brazilian National Committee of Ethics (CONEP) and by the Ethics Committee of the Hospital Clínic, Barcelona, Spain (2012/7306). The National Health Research and Ethics Committee of the Solomon Islands Ministry of Health and Medical Services (HRC12/022) approved the Solomon Islands yearlong cohort study. The Human Research Ethics Committee (HREC) at the Walter and Eliza Hall Institute of Medical Research

(WEHI) approved samples for use in Melbourne (#14/02), and also approved use and collection of the control panel samples (#14/02).

Procedures

Blood samples were collected by finger prick into EDTA tubes. Blood was separated into packed red cells and plasma at the field site. Packed red cells were stored at -20°C and plasma at -80°C prior to use in molecular and serological assays, respectively. *Plasmodium spp.* parasites were detected by qPCR as previously described[57,58]. The limit of detection of the molecular methods was 1-3 copy numbers/µl[57].

Antibody measurements

The majority of *P. vivax* malaria proteins (Sal1 strain) were expressed using the wheat germ cell-free (WGCF) system (CellFree Sciences, Yokohama, Japan) at either Ehime University or CellFree Sciences. For the antigen discovery phase, 342 *P. vivax* protein constructs were expressed at a small-scale with a biotin tag and antibodies measured using the AlphaScreen™ assay, as previously described[23]. 307 of these proteins were selected as previously described[23], and are known or expected to be immunogenic based on protein features such as the presence of signal peptides, transmembrane domains, or orthology with immunogenic *P. falciparum* proteins. An additional 35 proteins were added to enrich the largely blood-stage expressed set of proteins with pre-erythrocytic and sexual stage proteins. Plasma samples from weeks zero (enrolment), 12, 24 and 36 in the subset of volunteers described above were used (n=32 Thai patients, n=33 Brazilian patients). Raw AlphaScreen™ data was converted to relative antibody units based on plate-specific standard curves, with final units ranging from 0-400 (seropositivity was defined as a relative antibody unit greater than 0).

669 For the validation and pilot application phases, down-selected proteins were produced at a high
670 yield using dialysis-based refeeding WGCF methods and purified on an affinity matrix using a
671 His-tag. The purified proteins were stored and shipped in the following buffer: 20 mM Na-
672 phosphate pH 7·5, 0·3 M NaCl, 500 mM imidazole and 10% (v/v) glycerol. Protein yields and
673 purity were determined using SDS-PAGE followed by Coomassie Brilliant Blue staining using
674 standard laboratory methods. An additional 20 *P. vivax* proteins known to be highly
675 immunogenic were also included at the validation phase and were produced using previously
676 described methods[25-27,29,35,36,59] (see Table S2). All expressed proteins were based on the Sal1
677 strain unless otherwise stated in Table S2. The *P. vivax* proteins were coupled to COOH
678 microspheres as previously described[13], with the optimal amount of protein to be coupled
679 determined experimentally in order to achieve a log-linear standard curve with our positive
680 control pool. Protein-specific IgG antibody levels were measured in a multiplexed Luminex®
681 assay as previously described[28]. Plasma samples from both the first and last visits of the three
682 year-long observational cohort studies were used. Median fluorescent intensity (MFI) values
683 from the Luminex®-200 were converted to arbitrary relative antibody units based on a standard
684 curve generated with a positive control plasma pool from highly immune adults from Papua New
685 Guinea (PNG)[28].

686 *Statistical modelling – antigen discovery phase*

687 The change in measured antibody responses following *P. vivax* infection in patients in the
688 antigen discovery phase was analysed using mixed-effects linear regression models, as
689 previously described[23]. Estimated half-lives for 307 *P. vivax* proteins were previously reported[23],
690 with an additional 35 proteins included for this study. Estimated half-lives are provided in Table
691 S1. Antibody responses were measured at zero, three, six and nine months using the
692 AlphaScreen™ assay, and the model was only fitted to individuals who were seropositive at
693 baseline. Denote A_{ijk} to be the antibody titre in participant i to protein j at time t_k which can be
694 described by the following linear model:

695

696 $$\log(A_{ijk}) \sim \log(\alpha_j^o) + \log(\alpha_{ij}) + r_j^0 + r_{ij} t_k + \varepsilon_j \qquad (1)$$

697

698 where α_j^0 is the geometric mean titre (GMT) at the time of infection; log(α_{ij}) is a random effect
699 for the difference between participant i's initial antibody titre and the population-level GMT; r_j^0 is
700 the average rate of decay of antibodies to protein j in the population; r_{ij} is a random effect for the
701 difference between the decay rate of individual i with the population-level average; and $\varepsilon_j \sim$ N(0,
702 $\sigma_{m,j}$) is the error term. It is assumed that the random effects for initial antibody titres are
703 Normally distributed: log(α_{ij}) ~ N(0, $\sigma_{A,j}$), and that the random effects for the variation in decay
704 rates are also Normally distributed: $r_{ij} \sim$ N(0, $\sigma_{r,j}$).

705

706 The linear regression model for the decay of antibody titres in equation (1) has three sources of
707 variation: (i) variation in initial antibody response following infection; (ii) between individual
708 variation in antibody decay rate; and (iii) measurement error. Notably, all these sources of
709 variations are described by Normal distributions so their combined variation is also described by

a Normal distribution. Therefore, x_{ij} = log(A_{ij}), the expected log antibody titre in individual *i* to protein *j* at time *t,* can be described by the following distribution:

$$\mathbf{P}(x_{ij}|t) = \frac{1}{\sqrt{2\pi\left(\sigma_{\alpha,j}^2+t^2\sigma_{r,j}^2+\sigma_{m,j}^2\right)}} e^{-\frac{\left(x_{ij}-\alpha_j^0-r_j^0 t\right)^2}{2\left(\sigma_{\alpha,j}^2+t^2\sigma_{r,j}^2+\sigma_{m,j}^2\right)}} \qquad (2)$$

The probability for the time since infection *t* given measured antibody titre x_{ij} can be calculated by inverting equation (2) using Bayes rule[60], allowing estimation of the probability that the time since last infection was less than nine months.

A simulated annealing algorithm was used to explore the high-dimensional space arising from all combinations of the 142 prioritised SEMs in the antigen discovery phase, and select optimal combinations of proteins[25,61]. Combinations of proteins were chosen to optimise the likelihood that sampled antibody responses were correctly classified as having occurred within nine months of infection.

Statistical modelling – validation phase

Individuals in the validation phase were classified into two categories depending on whether they had PCR-detectable blood-stage *P. vivax* infections within the nine months prior to measurement of antibody responses (Table S4). In theory it is possible to provide quantitative estimates of the time since last infection, however, in practice, a more useful outcome is whether an individual has been infected within some past time period. Nine months was selected due to the biology of *P. vivax* relapses (highest incidence of relapse within the first nine months following mosquito bite infection[20]), and because this threshold fell within a time period for which data were available (the three longitudinal cohorts had follow-up for 12 months).

Initial classification performance of antibody responses to the 60 down-selected SEMs to classify recent *P. vivax* infections was assessed using a linear discriminant analysis (LDA). There were 1,770 ways of choosing two out of 60 proteins, 34,220 ways of choosing three proteins, 487,635 ways of choosing four proteins, and 5,461,512 ways of choosing five proteins. All combinations of up to four proteins were exhaustively evaluated to optimise classification performance for three sensitivity and specificity targets (80% for both; 50% sensitivity and 98% specificity; and 98% sensitivity and 50% specificity). To investigate combinations of five proteins, the 100 best combinations of four proteins were identified for each target. All possible remaining proteins were added to each of these 300 combinations, and the classification performance of all of these combinations of size five was assessed. A similar procedure was implemented to investigate classification performance of combinations of proteins up to size eight. Combinations beyond size eight were not tested given the diminishing returns in classification performance beyond eight proteins.

After determining the highest ranking SEMs using the LDA classifier, the top eight candidates were further assessed using a range of other classification algorithms, including logistic regression, quadratic discriminant analysis (QDA), decision trees, and random forests[62]. Decision trees were implemented using the rpart R package. Random forests were implemented using the randomForest R package. Two novel classification algorithms incorporating statistical models of antibody decay over time were also developed (see Appendix S1). These algorithms were combined into a composite algorithm so that the optimal algorithm and selection of proteins was selected for each target of sensitivity and specificity. All classification algorithms were cross-validated using 1000 randomly sampled, disjoint training and testing subsets (Extended Data Figure 5 and Figure S11).

Statistical modelling – pilot application phase

Antibody levels to the top 8 proteins were measured in samples from the first time point in the longitudinal cohorts used in the validation phase. At the start of longitudinal follow-up, individuals were classified as being exposed to blood-stage *P. vivax* infection within the past 9 months using the measured antibody levels and the composite classification algorithm. To test the hypothesis that infection in the past 9 months is associated with increased risk of future blood-stage infection (possibly relapses), the time to first PCR-positive *P. vivax* infection was analysed using Cox proportional hazards to estimate the hazard ratio (HR).

The potential public health impact of seroTAT strategies with primaquine was assessed using a mathematical model of *P. vivax* transmission[38], using the sensitivity and specificity targets achieved in the validation phase. Two annual rounds of seroTAT were modelled in the Thai, Brazilian and Solomon Islands populations at 80% coverage (Fig 5, Fig S12).

Acknowledgments

We gratefully acknowledge the extensive field teams that contributed to sample collection and qPCR assays, including Andrea Kuehn, Yi Wan Quah, Piyarat Sripoorote, and Andrea Waltmann. We thank all the individuals who participated in each of the studies, and thank the Australian and Thai Red Cross and the Rio de Janeiro State Blood Bank, Rio de Janeiro, Brazil for donation of plasma samples. We thank the Volunteer Blood Donor Registry at WEHI for donation of plasma samples, and Lina Laskos and Jenni Harris for their collection and advice. We thank Christopher King (Case Western Reserve University) for provision of the PNG control plasma pool. Melanie Bahlo (Walter and Eliza Hall Institute) is thanked for her advice on algorithm development.

Funding

We acknowledge funding from the Global Health Innovative Technology Fund (T2015-142 to IM), the National Institute of Allergy and Infectious Diseases (NIH grant 5R01 AI 104822 to JS and 5U19AI089686-06 to JK) and the National Health and Medical Research Council Australia (#1092789 and #1134989 to IM and #1143187 to W-HT). Cohort samples were derived from field studies originally funded by the TransEPI consortium (supported by the Bill and Melinda Gates Foundation). This work has been supported by FIND with funding from the Australian and British governments. We also acknowledge support from the National Research Council of Thailand. This work was made possible through Victorian State Government Operational Infrastructure Support and Australian Government NHMRC IRIISS. IM is supported by an NHMRC Senior Research Fellowship (1043345). DLD is supported by an NHMRC Principal Research Fellowship (1023636). TT was supported in part by JSPS KAKENHI (JP15H05276, JP16K15266) in Japan. W.H.T. is a Howard Hughes Medical Institute-Wellcome Trust International Research Scholar (208693/Z/17/Z). RJL received the Page Betheras Award from

WEHI to provide funding for technical support for this project during her parental leave. MVGL and WMM are CNPq fellows.

Competing interests

FIND contributed to early funding of this work and had a role in data interpretation, writing of the report and the decision to submit. No other funders of this study had any role in study design, data collection, data analysis, data interpretation, writing of the report, and the decision to submit. RJL, MW, TT and IM are inventors on patent PCT/US17/67926 on a system, method, apparatus and diagnostic test for *Plasmodium vivax*.

Author contributions

R.J.L., M.T.W., T.T. and I.M. designed the study. W.N., W.M., J.K., M.L., J.S. and I.M. conducted the cohort studies. E.T., M.M., M.H., F.M., W.-H.T., J.H., C.H., C.E.C. and T.T. expressed the proteins. R.J.L., E.T., M.M., J.B., and Z.S.-J.L. performed antibody measurements. R.J.L., C.S.N.L.-W.-S. and M.T.W. did the data management and analysis. M.T.W. and T.O. performed modelling. R.J.L., M.T.W. and I.M. wrote the draft of the report. L.J.R., C.P., D.L.D., X.C.D. and I.J.G. provided expert advice. All authors contributed to data interpretation and revision of the report.

Data and code availability

All data and code for reproducing this analysis are available online at https://github.com/MWhite-InstitutPasteur/Pvivax_sero_dx.

References

1. Sutanto, I., *et al.* Negligible Impact of Mass Screening and Treatment on Mesoendemic Malaria Transmission at West Timor in Eastern Indonesia: A Cluster-Randomized Trial. *Clin Infect Dis* **67**, 1364-1372 (2018).
2. Waltmann, A., *et al.* High Rates of Asymptomatic, Sub-microscopic Plasmodium vivax Infection and Disappearing Plasmodium falciparum Malaria in an Area of Low Transmission in Solomon Islands. *PLoS Negl Trop Dis* **9**, e0003758 (2015).
3. Mueller, I., *et al.* Key gaps in the knowledge of Plasmodium vivax, a neglected human malaria parasite. *Lancet Infect Dis* **9**, 555-566 (2009).
4. Moreira, C.M., Abo-Shehada, M., Price, R.N. & Drakeley, C.J. A systematic review of sub-microscopic Plasmodium vivax infection. *Malar J* **14**, 360 (2015).
5. Robinson, L.J., *et al.* Strategies for understanding and reducing the Plasmodium vivax and Plasmodium ovale hypnozoite reservoir in Papua New Guinean children: a randomised placebo-controlled trial and mathematical model. *PLoS Med* **12**, e1001891 (2015).
6. Parker, D.M., *et al.* Microgeography and molecular epidemiology of malaria at the Thailand-Myanmar border in the malaria pre-elimination phase. *Malar J* **14**, 198 (2015).
7. Mogeni, P., *et al.* Detecting Malaria Hotspots: A Comparison of Rapid Diagnostic Test, Microscopy, and Polymerase Chain Reaction. *J Infect Dis* **216**, 1091-1098 (2017).
8. Ding, X.C., *et al.* Defining the next generation of Plasmodium vivax diagnostic tests for control and elimination: Target product profiles. *PLoS Negl Trop Dis* **11**, e0005516 (2017).
9. Malaria surveillance, monitoring & evaluation: a reference manual. (2018).
10. Harris, I., *et al.* A large proportion of asymptomatic Plasmodium infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. *Malar J* **9**, 254 (2010).
11. Sutanto, I., *et al.* Negligible Impact of Mass Screening and Treatment on Meso-endemic Malaria Transmission at West Timor in Eastern Indonesia: A Cluster-Randomised Trial. *Clin Infect Dis* (2018).
12. Ong, K.I.C., *et al.* Systematic review of the clinical manifestations of glucose-6-phosphate dehydrogenase deficiency in the Greater Mekong Subregion: implications for malaria elimination and beyond. *BMJ Glob Health* **2**, e000415 (2017).
13. Longley, R.J., *et al.* Asymptomatic Plasmodium vivax infections induce robust IgG responses to multiple blood-stage proteins in a low-transmission region of western Thailand. *Malar J* **16**, 178 (2017).
14. Longley, R.J., *et al.* Acquisition and Longevity of Antibodies to Plasmodium vivax Preerythrocytic Antigens in Western Thailand. *Clin Vaccine Immunol* **23**, 117-124 (2016).
15. Helb, D.A., *et al.* Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. *Proc Natl Acad Sci U S A* **112**, E4438-4447 (2015).
16. Welch, R.J., Anderson, B.L. & Litwin, C.M. Rapid immunochromatographic strip test for detection of anti-K39 immunoglobulin G antibodies for diagnosis of visceral leishmaniasis. *Clin Vaccine Immunol* **15**, 1483-1484 (2008).
17. Gwyn, S., *et al.* Lateral flow-based antibody testing for Chlamydia trachomatis. *J Immunol Methods* **435**, 27-31 (2016).
18. Landry, M.L. Diagnostic tests for influenza infection. *Curr Opin Pediatr* **23**, 91-97 (2011).
19. Greenhouse, B., Smith, D.L., Rodriguez-Barraquer, I., Mueller, I. & Drakeley, C. Taking sharper pictures of malaria with CAMERAs: Combined Antibodies to Measure Exposure Recency Assays. *bioRxiv* (2018).
20. White, N.J. Determinants of relapse periodicity in Plasmodium vivax malaria. *Malar J* **10**, 297 (2011).
21. Battle, K.E., *et al.* Geographical variation in Plasmodium vivax relapse. *Malar J* **13**, 144 (2014).
22. White, M.T., Shirreff, G., Karl, S., Ghani, A.C. & Mueller, I. Variation in relapse frequency and the transmission potential of Plasmodium vivax malaria. *Proc Biol Sci* **283**, 20160048 (2016).
23. Longley, R.J., *et al.* Naturally acquired antibody responses to more than 300 Plasmodium vivax proteins in three geographic regions. *PLoS Negl Trop Dis* **11**, e0005888 (2017).
24. Tsuboi, T., *et al.* Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. *Infect Immun* **76**, 1702-1708 (2008).
25. Franca, C.T., *et al.* Identification of highly-protective combinations of Plasmodium vivax recombinant proteins for vaccine development. *Elife* **6**(2017).
26. Franca, C.T., *et al.* Plasmodium vivax Reticulocyte Binding Proteins Are Key Targets of Naturally Acquired Immunity in Young Papua New Guinean Children. *PLoS Negl Trop Dis* **10**, e0005014 (2016).

27. Lu, F., *et al.* Profiling the humoral immune responses to Plasmodium vivax infection and identification of candidate immunogenic rhoptry-associated membrane antigen (RAMA). *J Proteomics* **102**, 66-82 (2014).
28. Franca, C.T., *et al.* An Antibody Screen of a Plasmodium vivax Antigen Library Identifies Novel Merozoite Proteins Associated with Clinical Protection. *PLoS Negl Trop Dis* **10**, e0004639 (2016).
29. Gruszczyk, J., *et al.* Transferrin receptor 1 is a reticulocyte-specific receptor for Plasmodium vivax. *Science* **359**, 48-55 (2018).
30. Soares, I.S., *et al.* A Plasmodium vivax vaccine candidate displays limited allele polymorphism, which does not restrict recognition by antibodies. *Mol Med* **5**, 459-470 (1999).
31. Wang, B., *et al.* Immunoprofiling of the tryptophan-rich antigen family in Plasmodium vivax. *Infect Immun* **83**, 3083-3095 (2015).
32. Cheng, Y., *et al.* Naturally acquired humoral and cellular immune responses to Plasmodium vivax merozoite surface protein 8 in patients with P. vivax infection. *Malar J* **16**, 211 (2017).
33. Hester, J., *et al.* De novo assembly of a field isolate genome reveals novel Plasmodium vivax erythrocyte invasion genes. *PLoS Negl Trop Dis* **7**, e2569 (2013).
34. Jiang, J., Barnwell, J.W., Meyer, E.V. & Galinski, M.R. Plasmodium vivax merozoite surface protein-3 (PvMSP3): expression of an 11 member multigene family in blood-stage parasites. *PLoS One* **8**, e63888 (2013).
35. Hietanen, J., *et al.* Gene Models, Expression Repertoire, and Immune Response of Plasmodium vivax Reticulocyte Binding Proteins. *Infect Immun* **84**, 677-685 (2015).
36. Cole-Tobian, J.L., *et al.* Strain-specific duffy binding protein antibodies correlate with protection against infection with homologous compared to heterologous plasmodium vivax strains in Papua New Guinean children. *Infect Immun* **77**, 4009-4017 (2009).
37. Ntege, E.H., *et al.* Blood-stage malaria vaccines: post-genome strategies for the identification of novel vaccine candidates. *Expert Rev Vaccines* **16**, 769-779 (2017).
38. White, M.T., *et al.* Mathematical modelling of the impact of expanding levels of malaria control interventions on Plasmodium vivax. *Nature communications* **9**, 3300 (2018).
39. Rabinovich, R.N., *et al.* malERA: An updated research agenda for malaria elimination and eradication. *PLoS Med* **14**, e1002456 (2017).
40. Cook, J., *et al.* Using serological measures to monitor changes in malaria transmission in Vanuatu. *Malar J* **9**, 169 (2010).
41. Ladeia-Andrade, S., *et al.* Naturally acquired antibodies to merozoite surface protein (MSP)-1(19) and cumulative exposure to Plasmodium falciparum and Plasmodium vivax in remote populations of the Amazon Basin of Brazil. *Mem Inst Oswaldo Cruz* **102**, 943-951 (2007).
42. Cook, J., *et al.* Sero-epidemiological evaluation of changes in Plasmodium falciparum and Plasmodium vivax transmission patterns over the rainy season in Cambodia. *Malar J* **11**, 86 (2012).
43. White, M.T., *et al.* Plasmodium vivax and Plasmodium falciparum infection dynamics: re-infections, recrudescences and relapses. *Malar J* **17**, 170 (2018).
44. White, M.T., *et al.* Dynamics of the antibody response to Plasmodium falciparum infection in African children. *J Infect Dis* **210**, 1115-1122 (2014).
45. Boyle, M.J., Reiling, L., Osier, F.H. & Fowkes, F.J. Recent insights into humoral immunity targeting Plasmodium falciparum and Plasmodium vivax malaria. *Int J Parasitol* **47**, 99-104 (2017).
46. Richards, J.S., *et al.* Association between naturally acquired antibodies to erythrocyte-binding antigens of Plasmodium falciparum and protection from malaria and high-density parasitemia. *Clin Infect Dis* **51**, e50-60 (2010).
47. Arama, C., *et al.* Genetic Resistance to Malaria Is Associated With Greater Enhancement of Immunoglobulin (Ig)M Than IgG Responses to a Broad Array of Plasmodium falciparum Antigens. *Open Forum Infect Dis* **2**, ofv118 (2015).
48. Woodberry, T., *et al.* Antibodies to Plasmodium falciparum and Plasmodium vivax merozoite surface protein 5 in Indonesia: species-specific and cross-reactive responses. *J Infect Dis* **198**, 134-142 (2008).
49. Douglas, N.M., *et al.* Plasmodium vivax recurrence following falciparum and mixed species malaria: risk factors and effect of antimalarial kinetics. *Clin Infect Dis* **52**, 612-620 (2011).
50. Looareesuwan, S., White, N.J., Chittamas, S., Bunnag, D. & Harinasuta, T. High rate of Plasmodium vivax relapse following treatment of falciparum malaria in Thailand. *Lancet* **2**, 1052-1055 (1987).
51. Fornace, K.M., *et al.* Exposure and infection to Plasmodium knowlesi in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. *PLoS Negl Trop Dis* **12**, e0006432 (2018).

52. Herman, L.S., *et al.* Identification and validation of a novel panel of Plasmodium knowlesi biomarkers of serological exposure. *PLoS Negl Trop Dis* **12**, e0006457 (2018).
53. White, N.J. & Imwong, M. Relapse. *Adv Parasitol* **80**, 113-150 (2012).
54. Alonso, P.L., *et al.* A research agenda to underpin malaria eradication. *PLoS Med* **8**, e1000406 (2011).
55. Nguitragool, W., *et al.* Highly heterogeneous residual malaria risk in western Thailand. *Int J Parasitol* (2019).
56. Quah, Y.W., *et al.* Molecular epidemiology of residual Plasmodium vivax transmission in a paediatric cohort in Solomon Islands. *Malar J* **18**, 106 (2019).
57. Wampfler, R., *et al.* Strategies for detection of Plasmodium species gametocytes. *PLoS One* **8**, e76316 (2013).
58. Rosanas-Urgell, A., *et al.* Comparison of diagnostic methods for the detection and quantification of the four sympatric Plasmodium species in field samples from Papua New Guinea. *Malar J* **9**, 361 (2010).
59. Healer, J., *et al.* Vaccination with conserved regions of erythrocyte-binding antigens induces neutralizing antibodies against multiple strains of Plasmodium falciparum. *PLoS One* **8**, e72504 (2013).
60. Borremans, B., Hens, N., Beutels, P., Leirs, H. & Reijniers, J. Estimating Time of Infection Using Prior Serological and Individual Information Can Greatly Improve Incidence Estimation of Human and Wildlife Infections. *PLoS Comput Biol* **12**, e1004882 (2016).
61. Kirkpatrick, S., Gelatt, C.D., Jr. & Vecchi, M.P. Optimization by simulated annealing. *Science* **220**, 671-680 (1983).
62. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning: data mining, inference, and prediction*, (Springer, New York, 2009).

Extended Data

Extended Data Figure 1. Study design and follow-up schedule.

Extended Data Figure 2. Antibody responses to 60 antigens measured in n = 2,281 biologically independent samples on the Luminex® platform, stratified by geographical location and time since last PCR detected infection. Boxplots denote median and interquartile ranges (IQR) of the data, with whiskers denoting the median ± 1.5*IQR.

Extended Data Figure 3. Association between background reactivity in non-malaria exposed controls and ranking of candidate SEMs.

Extended Data Figure 4. Detailed breakdown of classification performance for the target of 80% sensitivity and 80% specificity.

Extended Data Figure 5. Cross-validated receiver operating characteristic (ROC) curve for the composite classification algorithm.

Supporting Information

Appendix S1. Detailed statistical appendix, including supporting figures.

Fig S1. Association between measured antibody titre and time since infection.

Fig S2. Kinetics of multiple antibodies.

Fig S3. Estimates of (A) antibody half-life, and (B) the variance of each antibody response for 104 *P. vivax* proteins in the antigen discovery phase.

Fig S4. Increasing the maximum number of proteins allowed in a panel leads to diminishing increases in likelihood and classification performance.

Fig S5. Performance of LDA classifier for combinations of proteins up to size eight for identifying individuals with blood-stage *P. vivax* infection within the last 9 months.

Fig S6. Network visualization of antigens.

Fig S7. Assessment of the role of removing data from negative control participants on classification performance using a Random Forests classifier.

Fig S8. Variable importance plot generated by a random forests algorithm for identify antigens that contribute to accurate classification.

Fig S9. Assessment of incorporating information on an individual's age into a Random Forests classifier.

Fig S10. Association between measurements of our top eight *P. vivax* proteins and time since last PCR detected blood-stage *P. falciparum* infection.

Fig S11. Receiver operating characteristic (ROC) curves depicting comparison of cross-validated classification performance for the seven classification algorithms considered.

Fig S12. Pilot application analysis with removal of individuals who were *P. vivax* PCR positive at the first sampling point.

Table S1. Estimated antibody half-lives from antigen discovery phase.

Table S2. Purified *P. vivax* proteins used in the validation phase and their individual performance, complete list.

Table S3. Association of antibody level with current *P. vivax* infection.

Table S4. Epidemiological overview of data sets used in antigen discovery and validation phases.

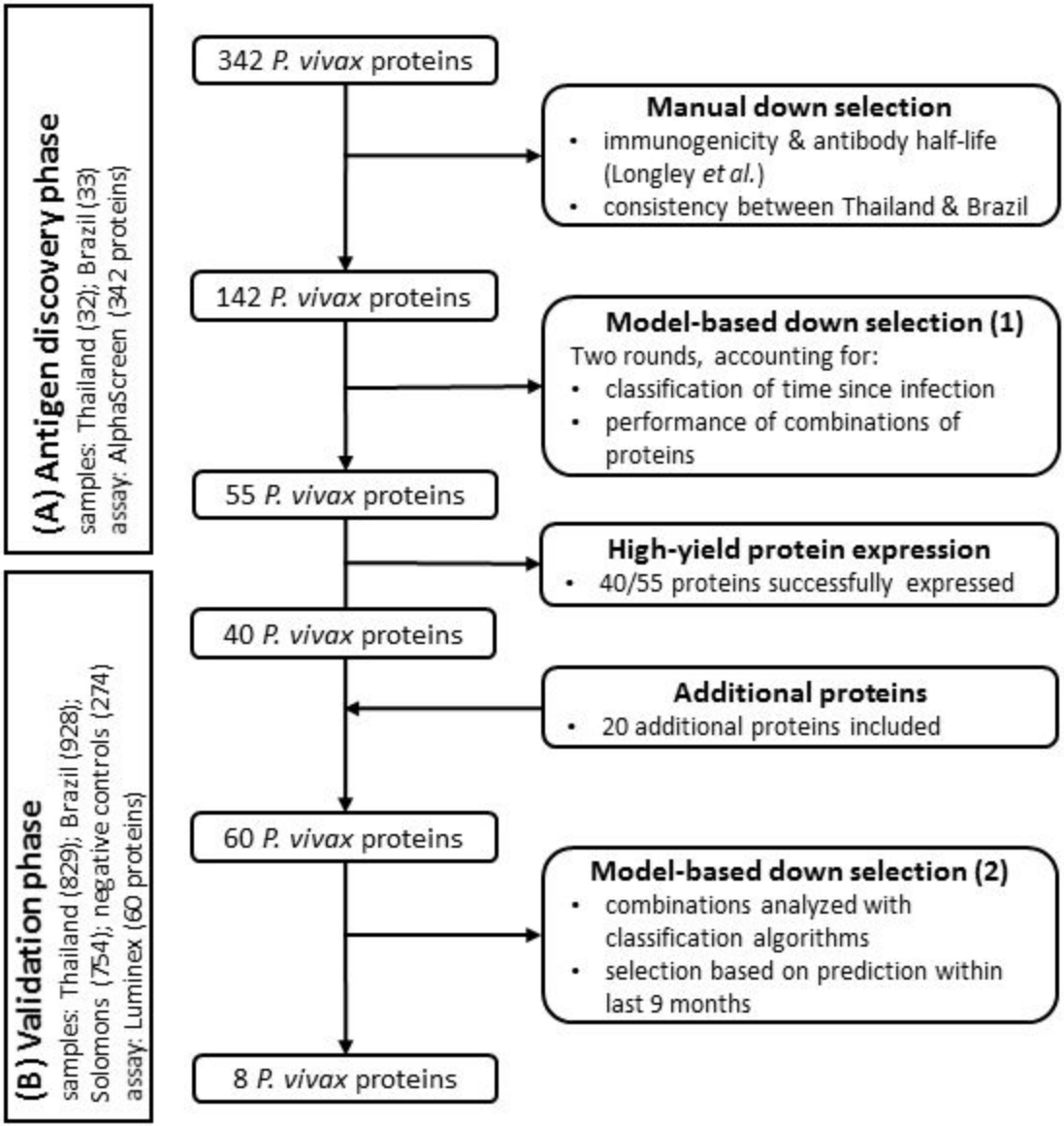
Extended Data Fig 1. Study design and follow-up schedule. (A) Thai and Brazilian patients were enrolled following a clinical episode of *P. vivax* and treated according to the relevant National Guidelines, with directly observed treatment (DOT) to ensure compliance and reduced risk of relapse. Volunteers were followed for nine months after enrolment, with finger-prick blood samples collected at enrolment and week 1, then every two weeks for six months, then every month. Antibody levels were measured in a subset of 32 Thai and 33 Brazilian volunteers who were confirmed to be free of blood-stage *Plasmodium* parasites by analysing all samples by light microscopy and qPCR. (B) 999 participants from Thailand, 1274 participants from Brazil, and 860 participants from the Solomon Islands were followed longitudinally for 12 months with active surveillance visits every month. For the analysis in the validation phase antibody levels were measured in plasma samples from the last visit. For the analysis in the application phase, antibody levels were measured in plasma samples from the first visit.

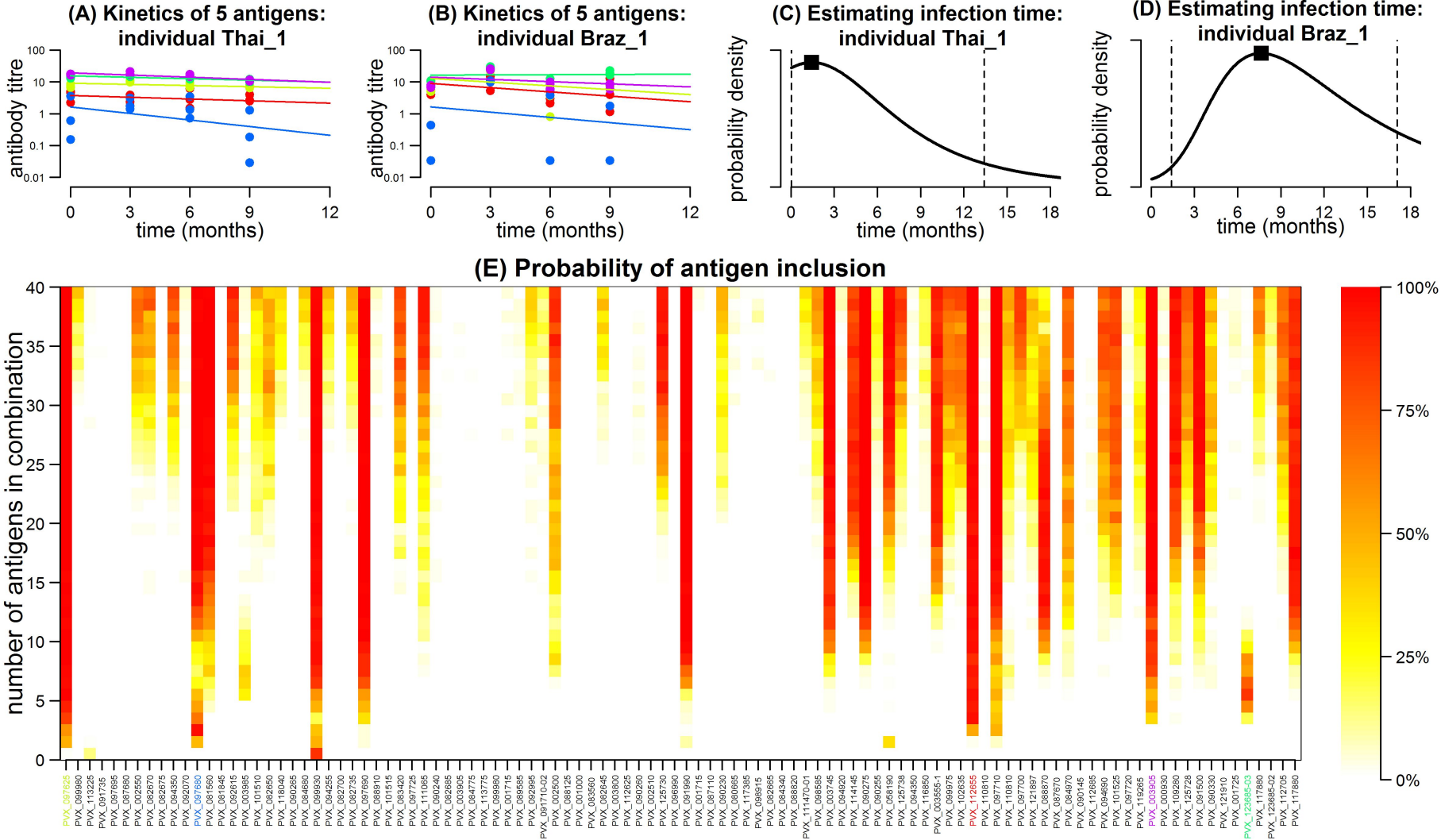
Extended Data Fig 2. Measured antibody responses to 60 proteins on the Luminex® platform, stratified by geographical location and time since last PCR detected infection.

Extended Data Fig 3. Association between background reactivity in non-malaria exposed controls and ranking of candidate SEMs by area under the curve (AUC). Mean relative antibody units (RAU) detected in malaria-naïve control panels from Melbourne, Australia (n=202), Bangkok, Thailand (n=72) and Rio de Janeiro Brazil (n = 96) compared to the AUC of the 60 candidate P.vivax proteins generated during the validation phase. WGCF expressed proteins are in black and E. coli or Baculovirus expressed proteins are in blue. RBP2b161-1454 (E. coli) is in red and RBP2b1986-2653 is in orange.

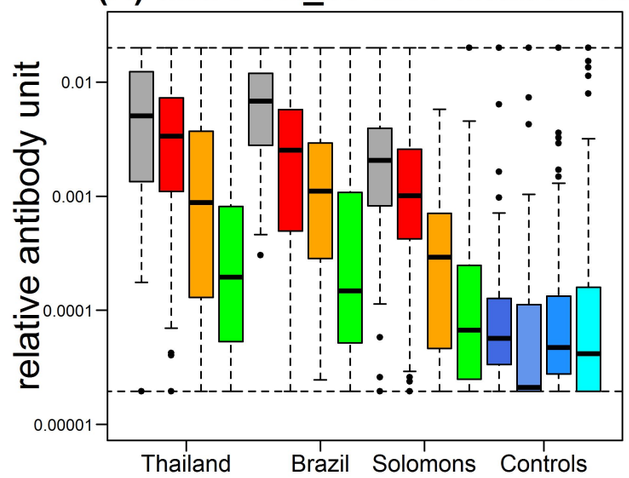
Extended Data Fig 4. Breakdown of the classification of the Random Forests algorithm with target sensitivity and specificity of 80%. The size of each rectangle is proportional to the number of samples in each category (See Table S4 of accompanying manuscript). The coloured area represents the proportion correctly classified, and the shaded area represents the proportion mis-classified.

Extended Data Fig 5. Receiver operating characteristic (ROC) curve for the composite classification algorithm. All curves presented are the median of 1000 repeat cross-validations.

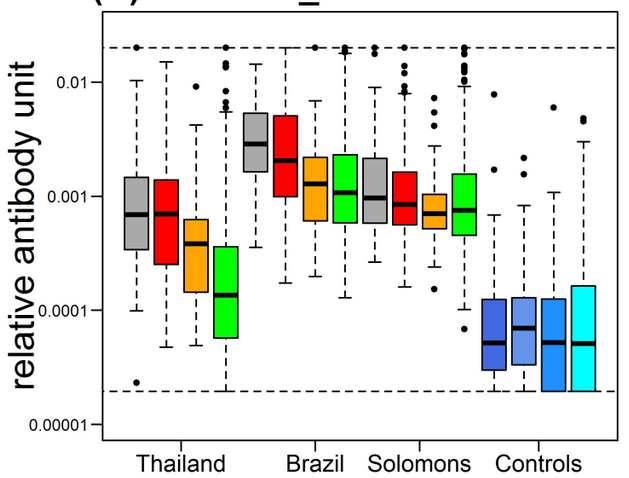

(A) Antigen discovery phase
samples: Thailand (32); Brazil (33)
assay: AlphaScreen (342 proteins)
(B) Validation phase
samples: Thailand (829); Brazil (928); Solomons (754); negative controls (274)
assay: Luminex (60 proteins)
342 P. vivax proteins
Manual down selection
immunogenicity & antibody half-life (Longley et al.)
consistency between Thailand & Brazil
142 P. vivax proteins
Model-based down selection (1)
Two rounds, accounting for:
classification of time since infection
performance of combinations of proteins
55 P. vivax proteins
High-yield protein expression
40/55 proteins successfully expressed
40 P. vivax proteins
Additional proteins
20 additional proteins included
60 P. vivax proteins
Model-based down selection (2)
combinations analyzed with classification algorithms
selection based on prediction within last 9 months
8 P. vivax proteins


(A) Kinetics of 5 antigens: individual Thai_1
antibody titre
time (months)
(B) Kinetics of 5 antigens: individual Braz_1
antibody titre
time (months)
(C) Estimating infection time: individual Thai_1
probability density
time (months)
(D) Estimating infection time: individual Braz_1
probability density
time (months)
(E) Probability of antigen inclusion
number of antigens in combination
100%
75%
50%
25%
0%

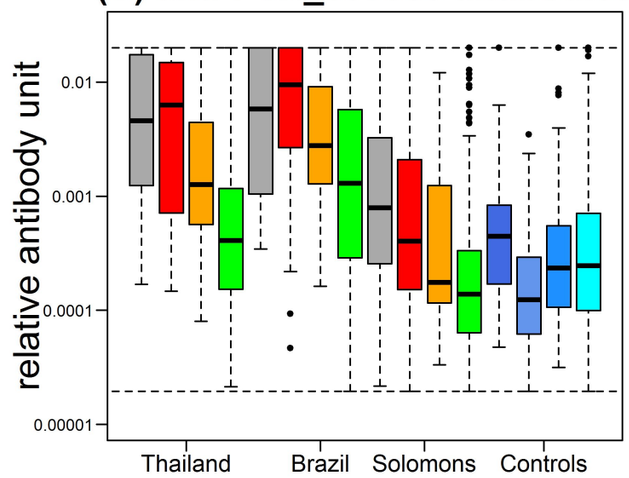
(A) anti-PVX_094255 antibodies



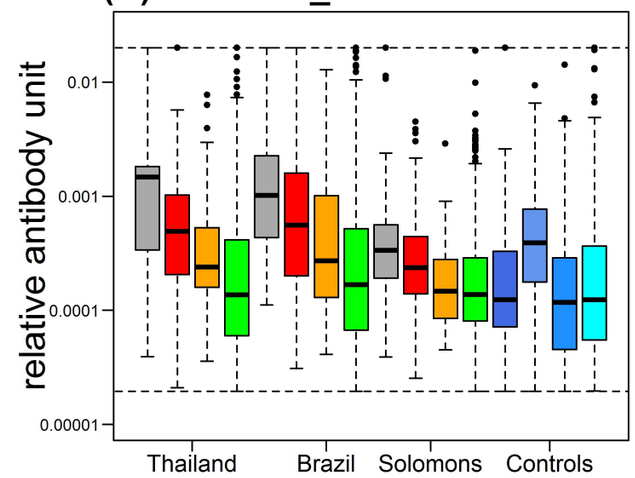
(B) anti-PVX_087885 antibodies



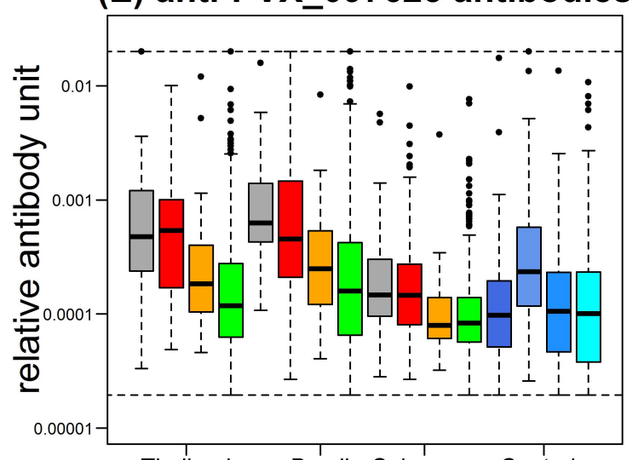
(C) anti-PVX_099980 antibodies



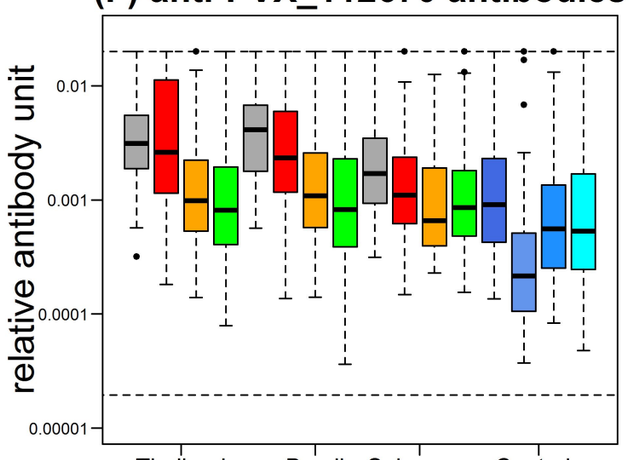
(D) anti-PVX_096995 antibodies



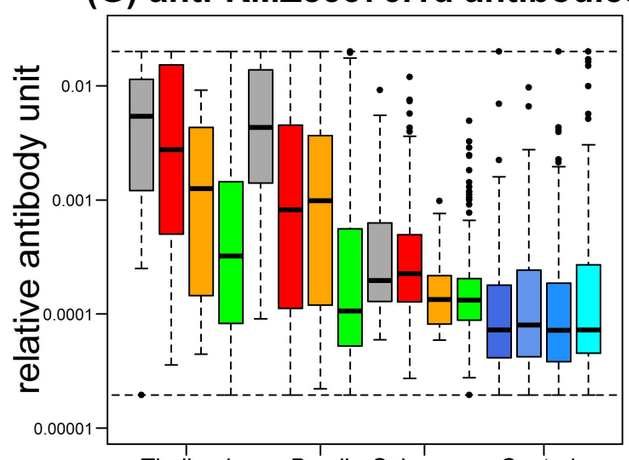
(E) anti-PVX_097625 antibodies



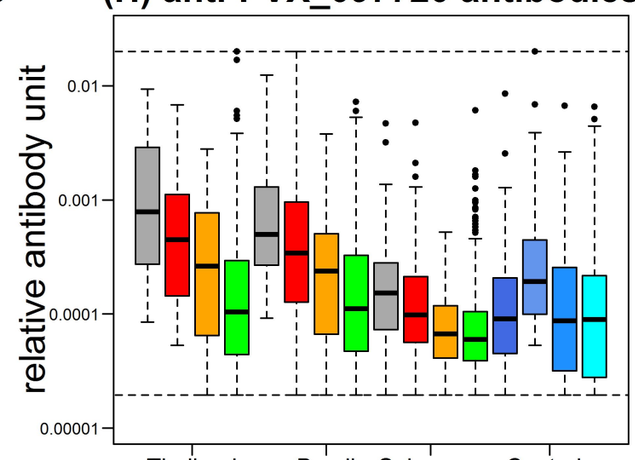
(F) anti-PVX_112670 antibodies



(G) anti-KMZ83376.1d antibodies

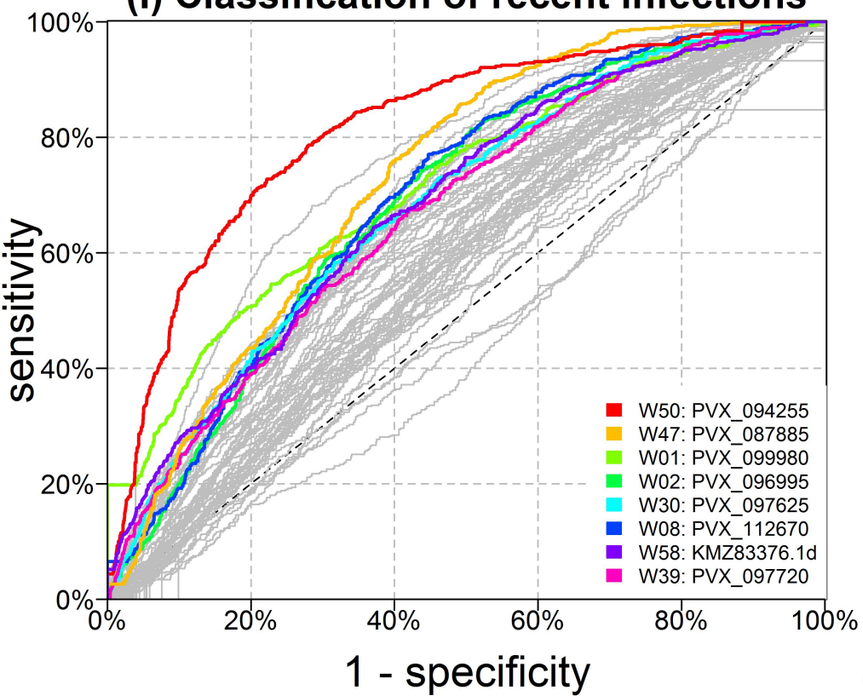


(H) anti-PVX_097720 antibodies

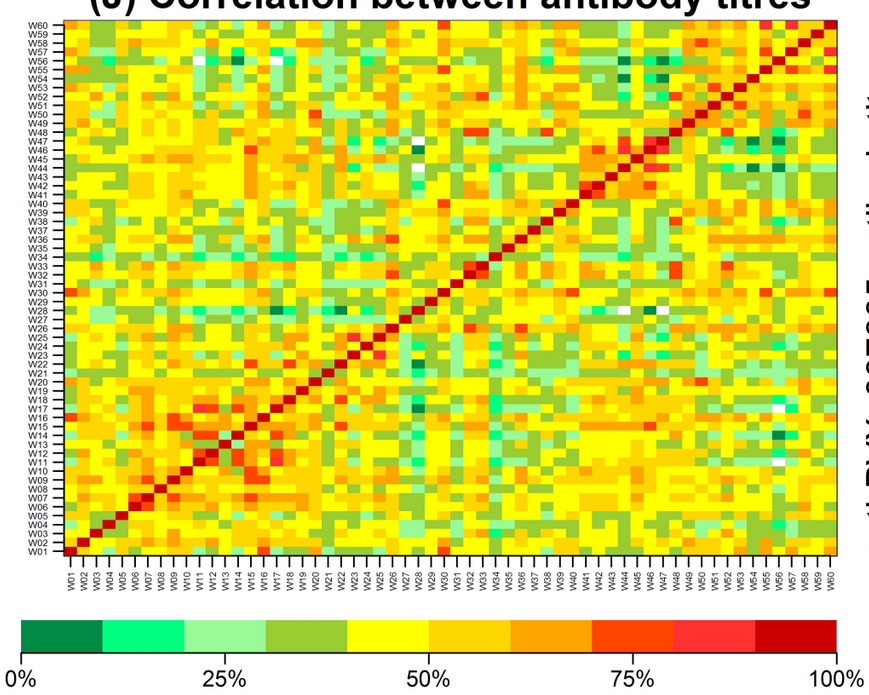


infected
infected 1-9 months
infected 9-12 months
no detected infection
negative controls: Thai RC
negative controls: Brazil RC
negative controls: Aus RC
negative controls: VBDR

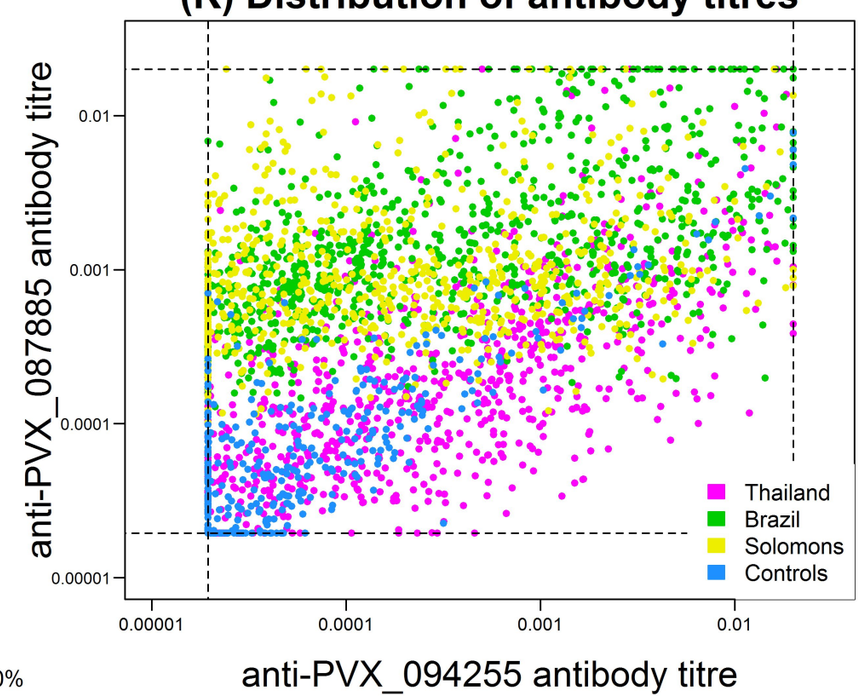
(I) Classification of recent infections

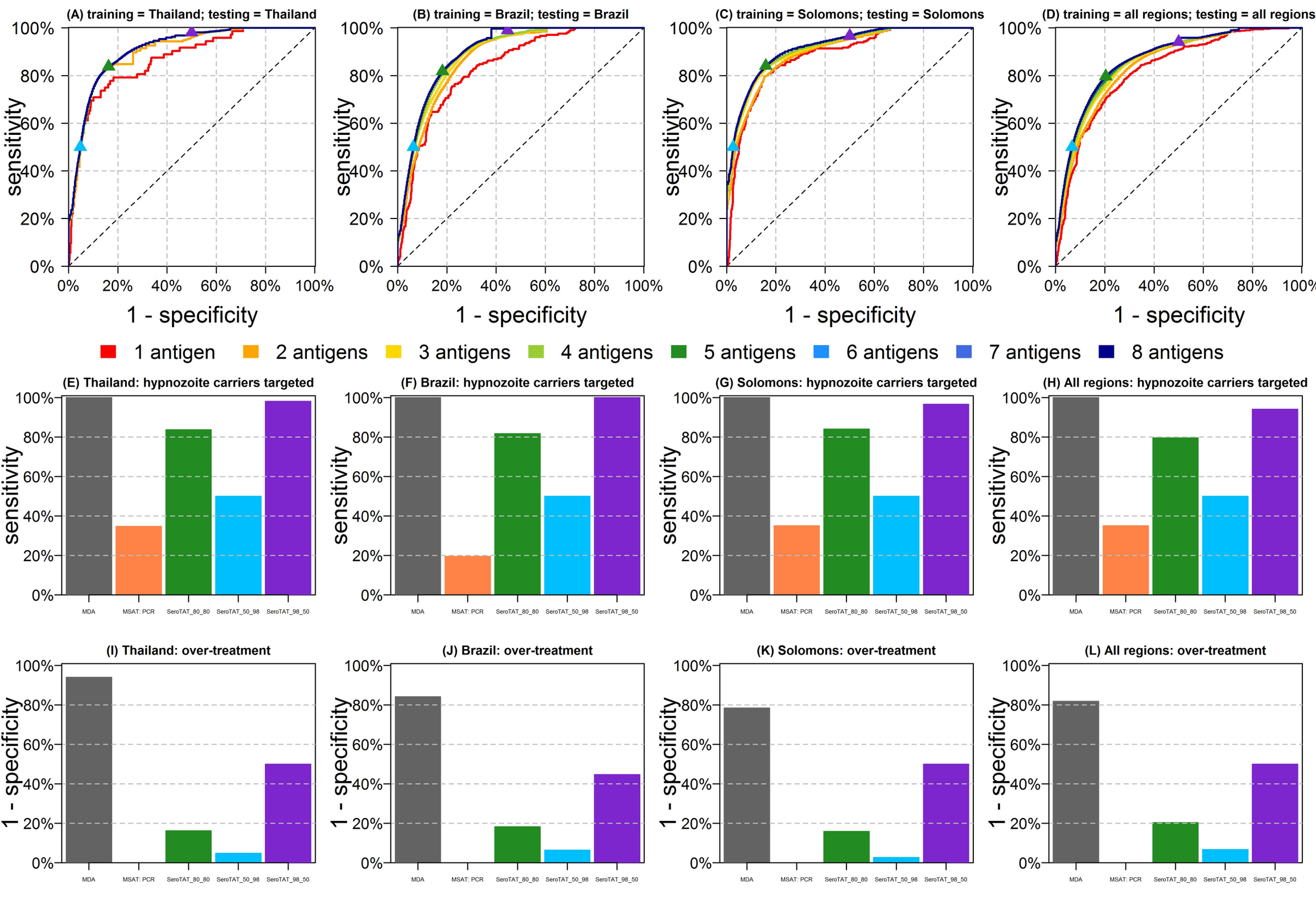


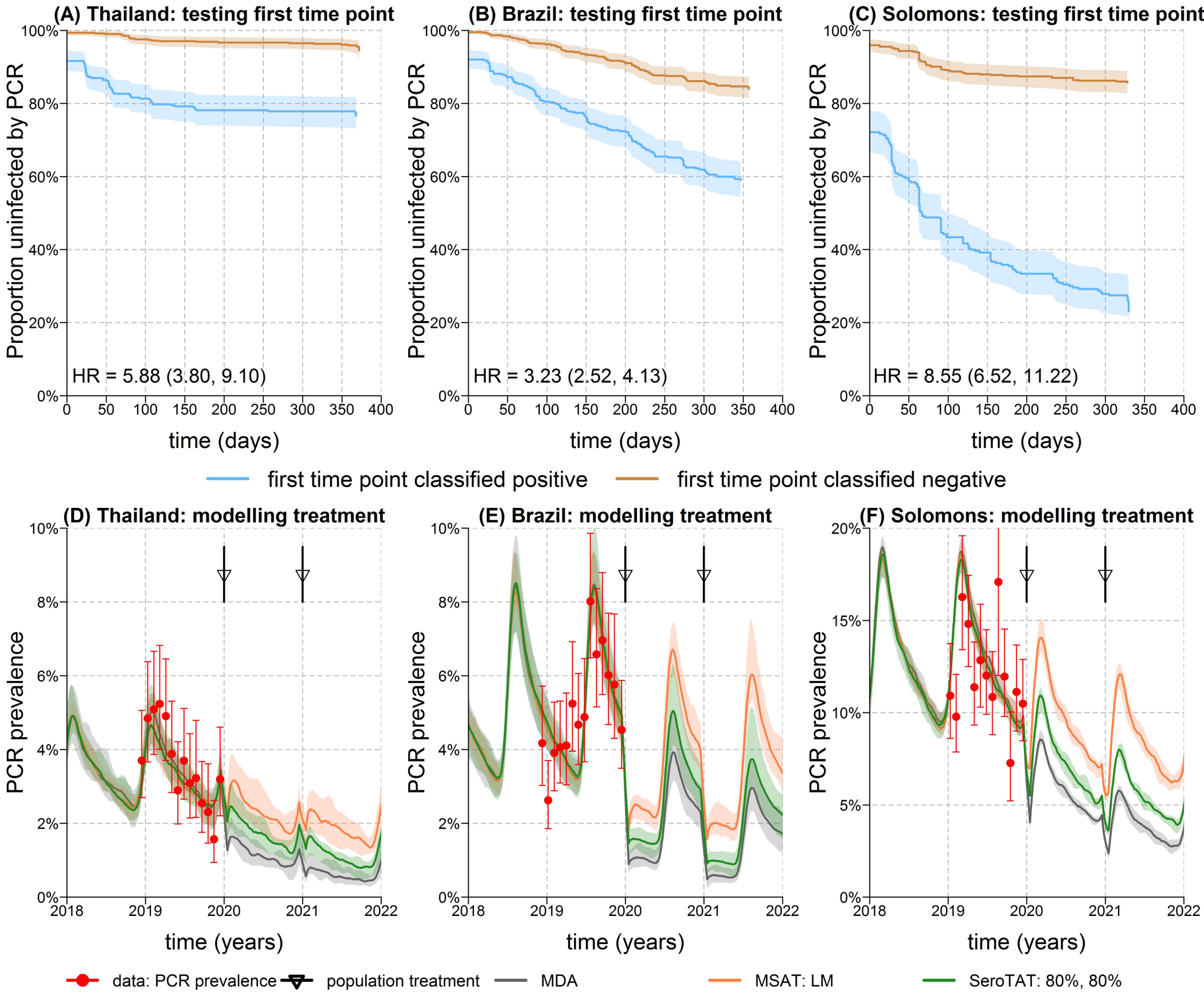
(J) Correlation between antibody titres

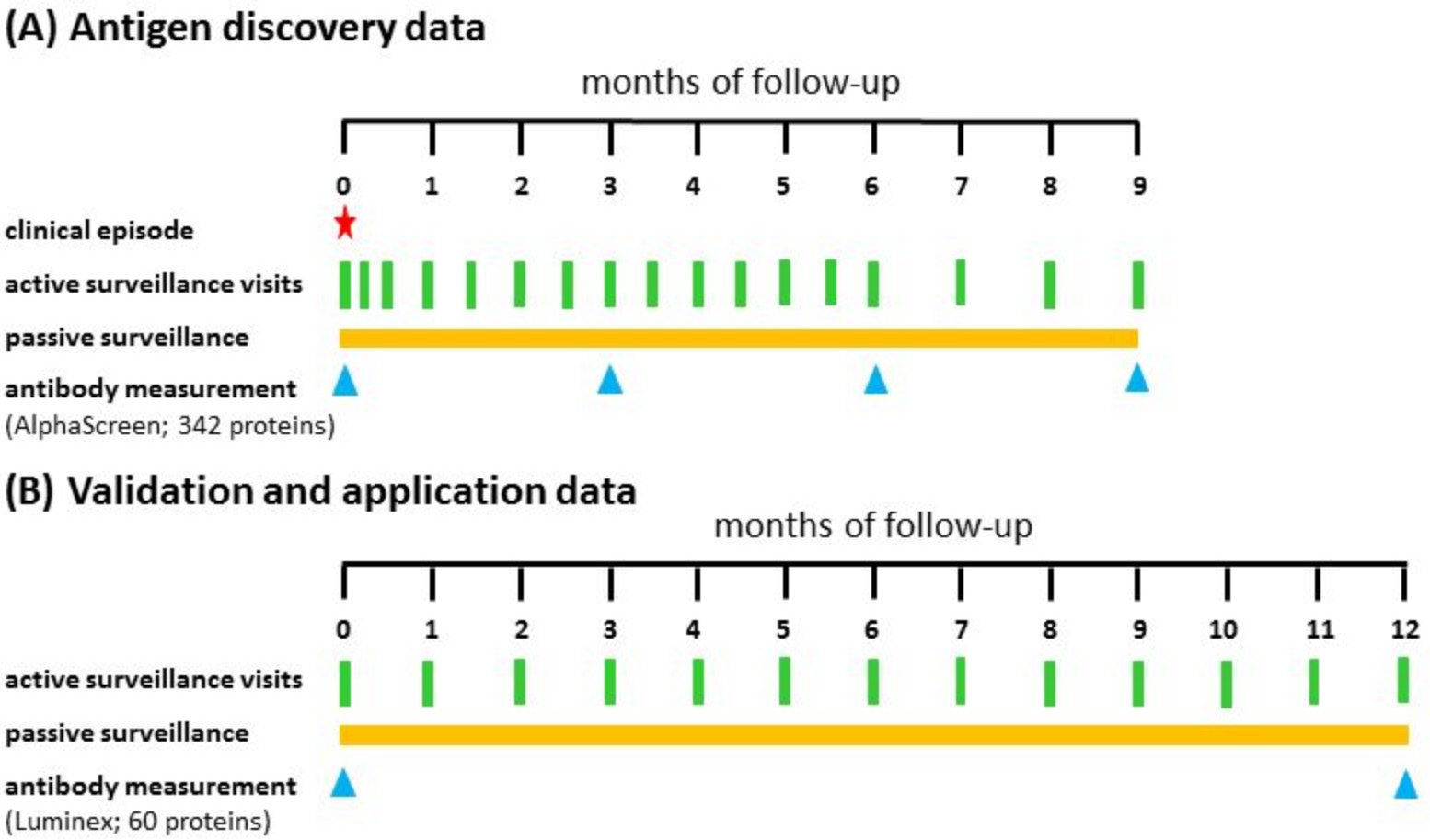


(K) Distribution of antibody titres




(A) training = Thailand; testing = Thailand
(B) training = Brazil; testing = Brazil
(C) training = Solomons; testing = Solomons
(D) training = all regions; testing = all regions
sensitivity
1 - specificity
0%
20%
40%
60%
80%
100%
1 antigen
2 antigens
3 antigens
4 antigens
5 antigens
6 antigens
7 antigens
8 antigens
(E) Thailand: hypnozoite carriers targeted
(F) Brazil: hypnozoite carriers targeted
(G) Solomons: hypnozoite carriers targeted
(H) All regions: hypnozoite carriers targeted
(I) Thailand: over-treatment
(J) Brazil: over-treatment
(K) Solomons: over-treatment
(L) All regions: over-treatment
MDA
MSAT: PCR
SeroTAT_80_80
SeroTAT_50_98
SeroTAT_98_50


(A) Thailand: testing first time point
(B) Brazil: testing first time point
(C) Solomons: testing first time point
Proportion uninfected by PCR
100%
80%
60%
40%
20%
0%
0
50
100
150
200
250
300
350
400
time (days)
HR = 5.88 (3.80, 9.10)
HR = 3.23 (2.52, 4.13)
HR = 8.55 (6.52, 11.22)
first time point classified positive
first time point classified negative
(D) Thailand: modelling treatment
(E) Brazil: modelling treatment
(F) Solomons: modelling treatment
PCR prevalence
10%
8%
6%
4%
2%
0%
20%
15%
5%
2018
2019
2020
2021
2022
time (years)
data: PCR prevalence
population treatment
MDA
MSAT: LM
SeroTAT: 80%, 80%


(A) Antigen discovery data
months of follow-up
0
1
2
3
4
5
6
7
8
9
clinical episode
active surveillance visits
passive surveillance
antibody measurement
(AlphaScreen; 342 proteins)
(B) Validation and application data
months of follow-up
0
1
2
3
4
5
6
7
8
9
10
11
12
active surveillance visits
passive surveillance
antibody measurement
(Luminex; 60 proteins)

PVX_099980 PVX_096995 PVX_088860 PVX_112680 PVX_097715 PVX_094830 PVX_112675 PVX_112670 PVX_101530 PVX_090970

PVX_084720 PVX_003770 PVX_092990 PVX_112690 PVX_091710 PVX_087885 PVX_003555 PVX_117385 PVX_082650 PVX_094255

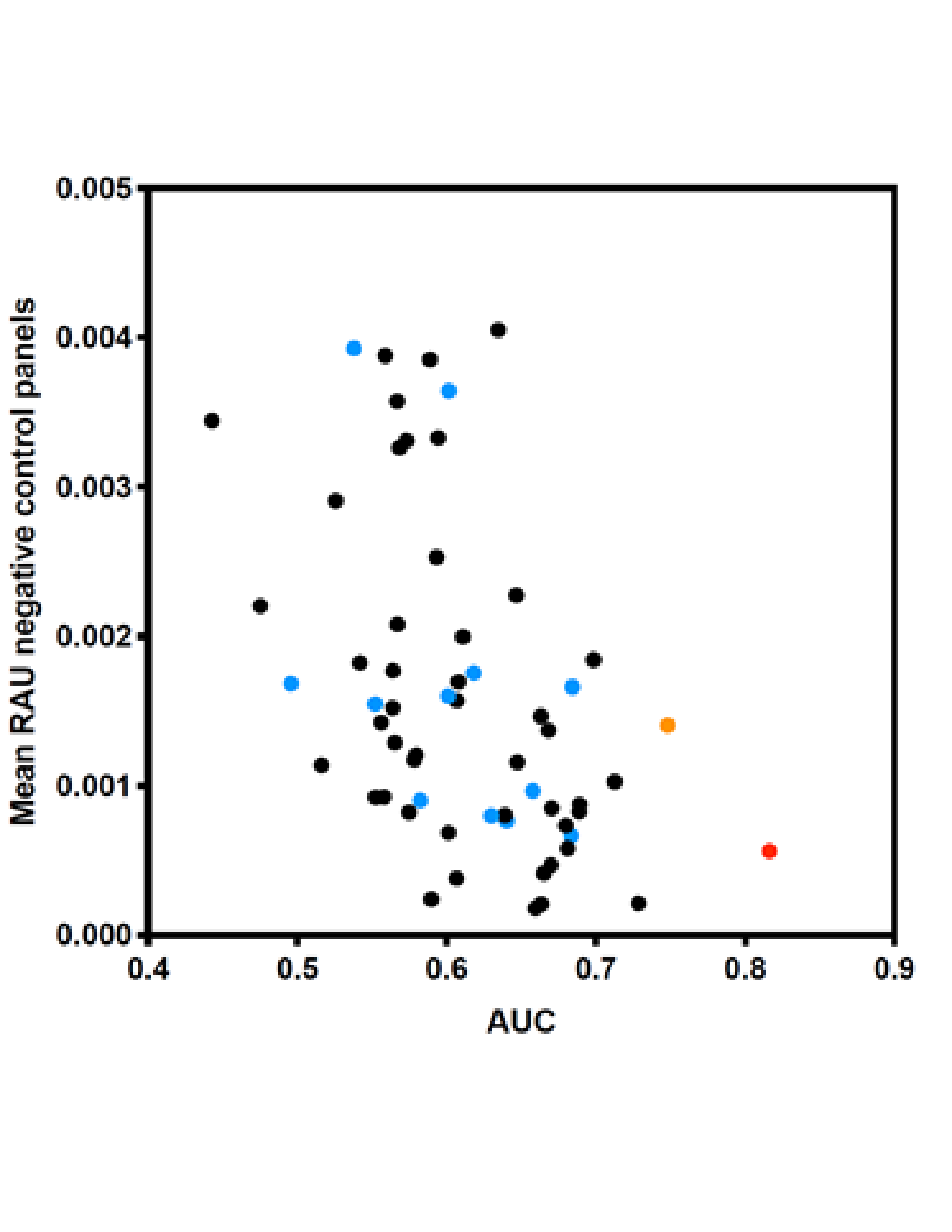
PVX_098915 PVX_088820 PVX_117880 PVX_121897 PVX_125728 PVX_090265 PVX_082700 PVX_097680 PVX_001000 PVX_097625

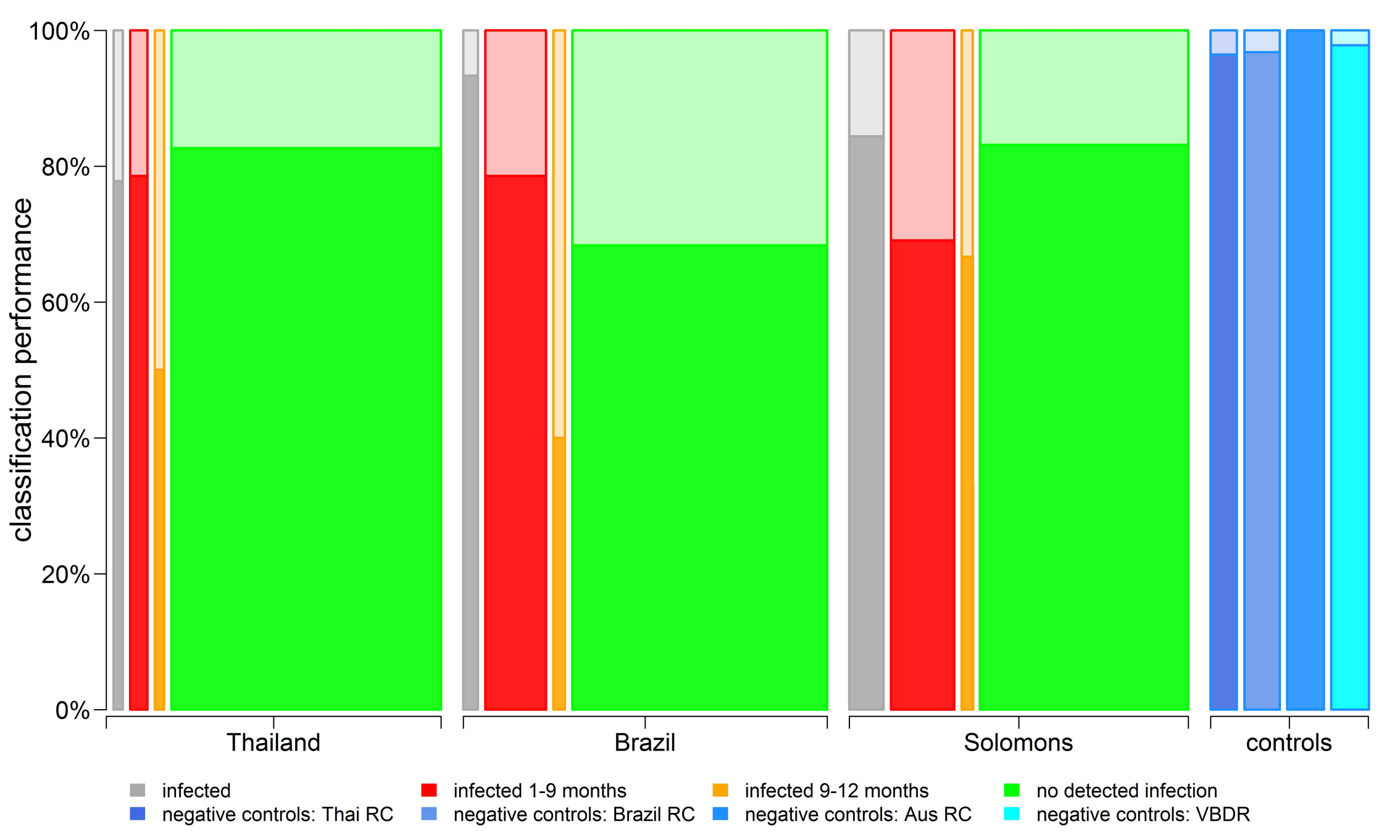
PVX_082670 PVX_099930 PVX_084340 PVX_082735 PVX_082645 PVX_090330 PVX_123685 PVX_125738 PVX_097720 PVX_000930

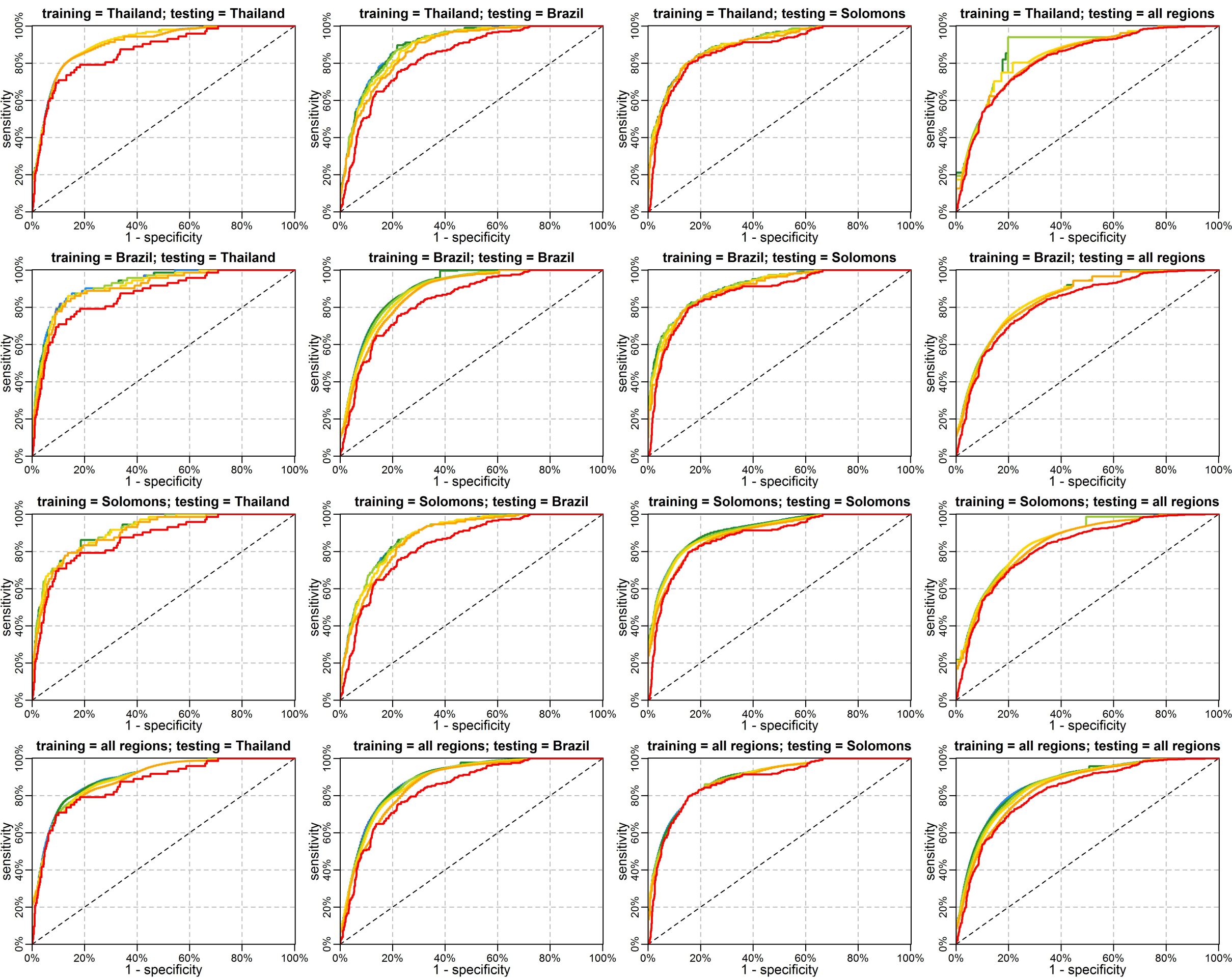
PVX_094350 PVX_099930 PVX_114330 PVX_088820 PVX_080665 PVX_092995 PVX_087885 PVX_098582 PVX_121920 PVX_094255

PVX_090325 PVX_101590 PVX_110810 PVX_088910 PVX_095055 PVX_090240 AAY34130.1d KMZ83376.1d PVX_098585 PVX_110810

infected
infected 1-9 months
infected 9-12 months
no detected infection
negative controls: Thai RC
negative controls: Brazil RC
negative controls: Aus RC
negative controls: VBDR


0.005
0.004
0.003
0.002
0.001
0.000
Mean RAU negative control panels
0.4
0.5
0.6
0.7
0.8
0.9
AUC


100%
80%
60%
40%
20%
0%
classification performance
Thailand
Brazil
Solomons
controls
infected
infected 1-9 months
infected 9-12 months
no detected infection
negative controls: Thai RC
negative controls: Brazil RC
negative controls: Aus RC
negative controls: VBDR


training = Thailand; testing = Thailand
training = Thailand; testing = Brazil
training = Thailand; testing = Solomons
training = Thailand; testing = all regions
training = Brazil; testing = Thailand
training = Brazil; testing = Brazil
training = Brazil; testing = Solomons
training = Brazil; testing = all regions
training = Solomons; testing = Thailand
training = Solomons; testing = Brazil
training = Solomons; testing = Solomons
training = Solomons; testing = all regions
training = all regions; testing = Thailand
training = all regions; testing = Brazil
training = all regions; testing = Solomons
training = all regions; testing = all regions
sensitivity
1 - specificity
0%
20%
40%
60%
80%
100%
1 antigen
2 antigens
3 antigens
4 antigens
5 antigens
6 antigens
7 antigens
8 antigens