



HAL
open science

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.

► To cite this version:

Rhea J Longley, Michael T White, Eizo Takashima, Jessica Brewster, Masayuki Morita, et al.. Development and validation of serological markers for detecting recent *Plasmodium vivax* infection. *Nature Medicine*, 2020, 26 (5), pp.741-749. 10.1038/s41591-020-0841-4 . pasteur-02612720

HAL Id: pasteur-02612720

<https://pasteur.hal.science/pasteur-02612720v1>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1
2
3
4
5
6

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.: <i>Smith_ED_Fig1.jpg</i>	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 <i>P. vivax</i> 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 <i>Plasmodium</i> 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 <i>P.vivax</i> 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 misclassified.
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			

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

8 **2. Supplementary Information:**

9

10 **A. Flat Files**

11

12 **Complete the Inventory below for all additional textual information and**

13 **any additional Supplementary Figures, which should be supplied in one**

14 **combined PDF file.**

15

16

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.: <i>Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.</i>
------	----------	--	---

		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	

17
18
19
20
21
22
23
24

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.: <i>Smith_Supplementary_Video_1.mov</i>	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.			

25 *Add rows as needed to accommodate the number of files.*
26

3. Source Data

27
28
29
30

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.: <i>Smith_SourceData_Fig1.xls, or Smith_Unmodified_Gels_Fig1.pdf</i>	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_serodx	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		

31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

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 Takashima⁵, Jessica Brewster¹, Masayuki Morita⁵, Matthias Harbers^{6,19}, Thomas Obadia^{4,20}, Leanne J. Robinson^{1,2,7}, Fumie Matsuura⁶, Zoe SJ Liu^{1,2}, Connie S. N. Li-Wai-Suen^{1,2}, Wai-Hong Tham^{2,8}, Julie Healer^{2,8}, Christele Huon⁹, Chetan E. Chitnis⁹, Wang Nguitragool¹⁰, Wuelton Monteiro^{11,12}, Carla Proietti^{13,14}, Denise L. Doolan^{13,14}, Andre M. Siqueira¹⁵, Xavier C. Ding¹⁶, Iveth J. Gonzalez¹⁶, James Kazura¹⁷, Marcus Lacerda^{11,18}, Jetsumon Sattabongkot³, Takafumi Tsuboi⁵, Ivo Mueller^{1,2,4*}

Affiliations:

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

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

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

79

80

81 *To whom correspondence should be addressed: mueller@wehi.edu.au.

82 †These authors contributed equally.

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130

131

132

133 **Abstract**

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

146

147 **Introduction**

148 Elimination of malaria by 2030 is now the explicit goal of all malaria endemic countries in the
149 Asia-Pacific and the Americas. While impressive progress towards this goal has been made,
150 global funding for malaria control has remained unchanged since 2010 and progress has stalled
151 in many parts of the world¹. New interventions and tools for better targeting of limited resources
152 are urgently needed.

153

154 A major hurdle for elimination is the increasing proportion of malaria cases caused by
155 *Plasmodium vivax* as malaria endemicity declines^{1,2}. *P. vivax* has unique biological features that
156 make its control difficult³, including high prevalence of low density, asymptomatic infections⁴ and
157 a liver-stage that can reactivate weeks to months after initial infection resulting in relapses that
158 cause morbidity and sustain transmission. These hypnozoites, undetectable with current
159 diagnostics, can be responsible for >80% of all blood-stage infections⁵. Identifying and targeting
160 individuals with hypnozoites is thus essential for accelerating and achieving malaria elimination.
161 In addition, as endemicity decreases, malaria transmission becomes increasingly fragmented
162 and highly seasonal^{6,7} rendering blanket approaches to malaria control and elimination
163 inefficient. This requires new, innovative tools and approaches specifically designed to assist in
164 targeting interventions to the changing malaria epidemiological landscape⁸.

165

166 National malaria control programs rely almost exclusively on microscopy or rapid diagnostic
167 tests for routine detection of malaria cases at health clinics, and for surveillance using mass
168 blood surveys⁹. These tools have limited sensitivity for detecting individuals with low density
169 asymptomatic infections¹⁰, making it difficult for control programs to efficiently identify areas of
170 low and high *P. vivax* transmission and target their resources accordingly. Molecular techniques
171 such as PCR have greater sensitivity¹⁰ but are rarely implemented by control programs due to
172 high cost and the need for specialised laboratories. All of these methods can only detect

173 individuals with a current blood-stage infection, rendering mass screening and treatment
174 (MSAT) approaches ineffective for reducing *P. vivax* transmission because they fail to treat
175 individuals that only carry hidden *P. vivax* liver-stage infections^{5,11} with no circulating blood-
176 stage parasitemia. Mass drug administration (MDA) is predicted to be a highly effective control
177 tool but only if it includes a drug that targets *P. vivax* hypnozoites⁵. Unfortunately, primaquine
178 and tafenoquine, the only current drugs able to eliminate hypnozoites, have toxic side effects in
179 glucose-6-phosphate dehydrogenase (G6PD) deficient individuals¹², limiting their acceptability
180 for MDA campaigns, particularly in low transmission areas where >90% recipients will have no
181 direct benefit from the treatment. There is thus a diagnostic gap, with MSAT ineffective due to
182 undertreatment and MDA unacceptable due to overtreatment. A strategy that efficiently targets
183 individuals with current blood-stage infections and those with hypnozoites is required.

184
185 Blood-stage *P. vivax* infections induce robust IgG antibody responses to a broad range of *P.*
186 *vivax* proteins, even following low-density asymptomatic infections¹³. These responses can be
187 long-lived, even after clearance of blood-stage infection^{14,15}. Hence, antibodies are markers of
188 past exposure as well as current infection¹⁵. Serological exposure markers (SEM) have been
189 used for surveillance of malaria and a number of other infectious diseases (i.e. leishmaniasis,
190 influenza, trachoma), and antibody responses can be simply and cheaply measured in point-of-
191 care tests¹⁶⁻¹⁸. SEMs may also be used for risk stratification and as guidance for targeted
192 malaria control and elimination interventions¹⁹. For *P. vivax* there is an important additional,
193 individual-level application for SEMs: to target treatment to people at-risk of carrying clinically
194 silent hypnozoites. While it is not possible to directly detect hypnozoites with current technology,
195 all tropical and sub-tropical *P. vivax* strains cause a primary infection followed by a first relapse
196 after no more than six to nine months²⁰⁻²². Therefore, any individual with a blood-stage infection
197 within the previous nine months, who has not received anti-liver-stage drugs, is likely to be a
198 hypnozoite carrier. A carefully selected panel of proteins inducing antibodies that signify

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

203

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

206

207 **Results**

208

209 *Antigen discovery phase*

210 Data on the magnitude and longevity of IgG responses to 342 *P. vivax* proteins following
211 symptomatic *P. vivax* infections²³ were used to identify suitable proteins for detecting recent
212 exposure. IgG was measured in four longitudinal plasma samples collected over a 9 month
213 period from individuals in Thailand (n=32) and Brazil (n=33) to enable estimation of antibody
214 half-life²³ (Table S1).

215

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

222

223

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

226
227 A statistical model for estimating time since infection was used to test the ability of the 142
228 prioritised proteins, individually or in combination, to predict time since last infection (see
229 Supplementary Methods and Supplementary Figs S1-3). Fig 2A-B shows examples of antibody
230 kinetics to five proteins in two representative individuals from Thailand and Brazil. Fig 2C-D
231 shows the estimated time since last infection with uncertainty, for antibody responses measured
232 to the five proteins in Fig 2A-B at six months after infection, for the same two representative
233 individuals. Increasing the number of proteins to at least five resulted in higher accuracy
234 compared to using a single protein alone, with only small incremental improvements in the
235 accuracy beyond 20 proteins (Fig S4). A simulated annealing algorithm was employed to
236 determine optimal combinations of proteins that maximise the likelihood of correctly estimating
237 the time since last infection (Fig 2E). Whilst some proteins had 100% probability of being
238 included in a successful panel of a set size (Fig 2E), there was redundancy in choice of
239 additional proteins and hence all were ranked based on their probability of inclusion in a 20-
240 protein panel (given the limited improvements beyond 20 proteins). The 24 highest ranked
241 proteins were selected for further testing.

242
243
244
245 **Fig 2.** Antigen discovery phase: estimating time since last *P. vivax* infection. (A, B) Examples of antibody kinetics to
246 five proteins in individuals from Thailand and Brazil. Points represent measured antibody responses in triplicate on
247 the AlphaScreen™ platform, and solid lines depict the fit of mixed-effects linear regression models. Each colour
248 corresponds to one of the five proteins (PVX_097625, PVX_097680, PVX_112655, PVX_003905, PVX_123685-03)
249 selected to provide optimal classification performance. (C, D) Using data on measured antibody responses six
250

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

258

259 The wheat germ cell free (WGCF) system was selected for expression of the proteins due to its
260 eukaryotic nature and past success in expressing *Plasmodium* proteins²⁴. Despite all proteins
261 being expressed in this system previously as crude proteins²³ not all could be produced at high
262 yield and high purity, and thus an additional round of down-selection using ranking from the
263 simulated annealing algorithm was performed to increase the number of proteins retained at this
264 point in the pipeline. The algorithm was run on 104 of the 142 prioritised proteins, excluding the
265 top 24 already selected (to provide more discriminatory power for selecting additional proteins)
266 and 14 proteins known to have low yields or form aggregates from our previous work. An
267 additional 31 proteins were selected in the second round using the same methodology, with 55
268 *P. vivax* proteins in total identified as suitable candidate SEMs. Of these, 40 were successfully
269 produced at high yield and purity using the WGCF system (Table S2).

270

271 An additional 20 proteins known to be highly immunogenic were added to these 40 proteins²⁵⁻²⁸.
272 This included four proteins already within the down-selected panel of 40 (different constructs or
273 produced by a different laboratory and protein expression systems), and three distinct
274 constructs of PVX_110810 (Duffy binding protein, each representing a different region of the
275 protein or sequence from a different strain) (Table S2).

276

277 *Validation phase*

278 The 60 candidate SEMs were validated in larger geographically diverse cohorts with a known
279 history of malaria infections in the preceding 12 months, using plasma samples from three
280 yearlong observational cohort studies in Thailand, Brazil and the Solomon Islands (see online
281 Methods and Extended Data Figure 1). Individuals in these cohorts were assessed monthly for
282 *Plasmodium spp.* by qPCR, with continuous concurrent passive case detection at local malaria
283 clinics and hospitals. Plasma samples from the last visit of these cohorts (n=829, 928 and 754
284 in Thailand, Brazil and the Solomon Islands, respectively) were used to measure IgG responses
285 in a multiplex Luminex® assay to the 60 *P. vivax* proteins. 158/2511 individuals in these cohorts
286 had a concurrent *P. vivax* infection, detected by qPCR, at the time the plasma was collected;
287 IgG antibody levels were strongly associated with current infection status for each of the 60
288 proteins in the Thai cohort (OR 2.1-7.7, $p < 0.05$) and most proteins in the Brazilian cohort
289 (57/60, OR 1.5-7.1, $p < 0.05$) (Table S3). This association was not as clear for the paediatric
290 Solomon Islands cohort, with IgG levels to only 29 out of 60 proteins significantly associated
291 with current *P. vivax* infections (OR 1.7-6.3, $p < 0.05$). Overall, there was a pattern of decreasing
292 IgG magnitude with increasing time since last *P. vivax* infection (Fig 3A-H, Extended Data
293 Figure 2), with minimal reactivity in malaria naïve negative control individuals from Bangkok, Rio
294 de Janeiro and Melbourne.

295
296 **Fig 3.** Validation phase. (A-H) Antibody responses to eight antigens measured in n = 2,281 biologically independent
297 samples on the Luminex® platform, stratified by geographical location and time since last PCR detected infection.
298 Boxplots denote median and interquartile ranges (IQR) of the data, with whiskers denoting the median $\pm 1.5 \times$ IQR. (I)
299 Receiver operating characteristic (ROC) curve for classifying individuals as infected within the last nine months using
300 a threshold antibody titre to a single antigen. Coloured curves correspond to those proteins in panels A-H. (J)
301 Spearman correlation between antibody titres measured in 2,281 samples. Correlation coefficients are indicating by
302 the colour scale, with red being 100% correlation and dark green 0% correlation. (K) 2-dimensional distribution of
303 anti-PVX_099980 and anti-PVX_094255 antibody titres.

304

305 Recent infections were defined as PCR-confirmed infections occurring within the past nine
306 months. Measurements of IgG antibodies against a number of individual proteins were able to
307 classify individuals as infected with *P. vivax* within the past nine months or not, with differing
308 degrees of accuracy (Fig 3I). Protein PVX_094255 (reticulocyte binding protein 2b, RBP2b)
309 reached 75% sensitivity and 75% specificity when used alone. IgG levels to the 60 candidate
310 SEMs were correlated (Fig 3J) with different distributions evident between the three geographic
311 regions (Fig 3K). A linear discriminant analysis (LDA) classification algorithm was used to
312 identify combinations of the 60 candidate SEMs that could accurately predict recent infections.
313 Redundancy was found with multiple combinations of *P. vivax* proteins able to accurately predict
314 recent infection when included in panels of up to eight proteins (Fig S5). The top eight most
315 frequently identified proteins when used in combination were: PVX_094255 (RBP2b²⁹),
316 PVX_087885 (rhostry associated membrane antigen, RAMA²⁷, putative), PVX_099980
317 (merozoite surface protein 1, MSP1³⁰), PVX_096995 (tryptophan-rich antigen (Pv-fam-a),
318 PvTRAg_2³¹), PVX_097625 (MSP8³², putative), PVX_112670 (unspecified product previously
319 annotated as a tryptophan-rich antigen (PvTRAg_28³¹), KMZ83376.1 (erythrocyte binding
320 protein II, EBPII³³), and PVX_097720 (MSP3.10³⁴) (see Fig 3I for ROC curves, Table 1 for
321 protein details and Fig S6 for network analysis). Table S2 shows the individual ranking of all 60
322 proteins by AUC, note that it is not necessarily the proteins with the highest individual AUC that
323 work best in combination.

324 **Table 1.** Top *P. vivax* proteins identified in the validation phase and their individual performance. Area under the curve (AUC) from
 325 the single antigen classifier is shown; proteins are listed in order of best individual performance. Nine of 60 proteins are shown here;
 326 the top 8 identified as best used in combination, plus the alternate construct of RBP2b. See Table S2 for the extended and complete
 327 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)	<i>E. coli</i>	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)	<i>E. coli</i>	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

328 ^aPlasmoDB release 36 (<http://plasmodb.org/plasmo/>), ^bGenBank

329 As shown in Table 1, for the top protein PVX_094255 (RBP2b), there were two separate
330 constructs mapping to different regions of the protein (RBP2b₁₆₁₋₁₄₅₄ and RBP2b₁₉₈₆₋₂₆₅₃),
331 expressed and purified using different methods^{29,37}. Antibody levels to these two constructs
332 (RBP2b₁₆₁₋₁₄₅₄ and RBP2b₁₉₈₆₋₂₆₅₃) were highly correlated (spearman $r=0.72$, $p<0.0001$, all
333 cohorts combined) and for this reason the construct that provided lower levels of classification
334 accuracy (RBP2b₁₉₈₆₋₂₆₅₃) was excluded from the top eight. RBP2b₁₆₁₋₁₄₅₄ induced very low levels
335 of antibody reactivity in the malaria naïve control panels. Across all candidate SEMs, a
336 significant negative association was observed between the mean antibody levels detected in the
337 malaria naïve negative control panels ($n=274$ individuals, see Methods) and the AUC values of
338 the 60 SEMs determined from their individual receiver operating characteristic (ROC) curves
339 (spearman $r=-0.5$, $p<0.0001$, Extended Data Figure 3. Removing data from these malaria naïve
340 control participants did not cause any substantial reductions in classification accuracy for the top
341 eight proteins (Fig S7).

342

343 *Classification performance of an eight-protein SEM panel*

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

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

357
358 We assessed the potential performance of targeted treatment using a seroTAT approach (with
359 80% sensitivity and specificity) compared to both MDA and MSAT with PCR (Fig 4E-L). In an
360 MDA campaign, all hypnozoite carriers are targeted (except those ineligible due to G6PD
361 deficiency), but >80% of the population receive unnecessary primaquine (Fig 4H). In contrast for
362 MSAT with PCR, no individual was over-treated (it is assumed that an individual with blood-
363 stage *P. vivax* is a likely hypnozoite carrier). However, only 20%-40% of all likely hypnozoite
364 carriers are targeted. Using seroTAT, at least 80% of individuals with hypnozoites are targeted
365 (outperforming MSAT with PCR) (Fig 5H), with <20% of the population treated unnecessarily
366 (substantially less than an MDA approach). We also assessed the potential performance of
367 seroTAT using two alternate strategies: 50% sensitivity/98% specificity and 98% sensitivity/50%
368 specificity. The high specificity approach only slightly outperforms MSAT in identifying
369 hypnozoites carriers (45% missed) and is thus likely not feasible for seroTAT. The high
370 sensitivity approach identifies nearly all likely hypnozoites carriers and reduces overtreatment to
371 approximately 50%, compared to >80% with MDA. The high sensitivity approach identifies
372 nearly all likely hypnozoites carriers and reduces overtreatment to approximately 50%,
373 compared to >80% with MDA.

374
375
376
377 **Fig 4.** Performance of the top eight *P. vivax* proteins for classifying recent infection and modelled ability to identify
378 hypnozoites carriers. Cross-validated ROC curves generated from the composite algorithm classifier for identifying
379 individuals with PCR detected infection in the last nine months in (A) Thailand; (B) Brazil; (C) The Solomon Islands;
380 and (D) all regions combined. The coloured triangles denote three different sensitivity and specificity targets for

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

384

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

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

396

397 The potential public health impact of a seroTAT strategy was assessed using a mathematical
398 model of *P. vivax* transmission³⁸ (Fig 5D-F). PCR prevalence was based on monthly data from
399 our 2013-2014 yearlong observational cohort studies and assumed to provide baseline *P. vivax*
400 transmission levels for population-level treatment strategies beginning in 2020. The public
401 health impact of three treatment strategies was simulated: MDA, MSAT with light microscopy,
402 and seroTAT with 80% sensitivity and specificity. Each strategy was modelled to be
403 implemented at 80% population coverage for two rounds with a primaquine regimen assumed to
404 clear all hypnozoites. Two yearly rounds of MDA resulted in an estimated 70% - 84% reduction
405 in *P. vivax* PCR prevalence across all three study sites whilst MSAT with LM was predicted to

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

408

409

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

423

424

425

426

427 **Discussion**

428

429 New tools and strategies to directly target *P. vivax* are urgently needed if malaria elimination in
430 the Asia-Pacific and Americas is to be achieved by 2030³⁹. *P. vivax* presents a unique challenge
431 to elimination due to the presence of undetectable hypnozoites that contribute to maintaining
432 residual transmission. *P. vivax* relapses are expected to occur at a frequency of every one to
433 two months^{21,22}. Apart from temperate 'hibernans' strains that are now restricted to the Korean

434 Peninsula and have no primary blood-stage infections²⁰ virtually all individuals who carry
435 hypnozoites will have had a *P. vivax* blood-stage infection within the prior nine months²⁰⁻²².
436 Herein, a novel panel of candidate serological exposure markers (“SEMs”) were identified and
437 validated that allow detection of recent exposure to *P. vivax* within the prior nine-months. To our
438 knowledge, these SEMs represent the first method that can indirectly identify likely hypnozoites
439 carriers that could be targeted for treatment with liver-stage drugs.

440

441 We undertook a relatively unbiased approach to choosing the best markers by screening a large
442 panel of 342 *P. vivax* proteins, and strategically selecting those that can predict recent infection
443 based on immunogenicity and antibody half-lives. The final panel, which has a sensitivity and
444 specificity of 80% at identifying individuals with PCR-detectable blood-stage infection in the last
445 nine months in three geographically distinct regions, incorporates antibody responses to eight *P.*
446 *vivax* proteins: PVX_094255 (RBP2b₁₆₁₋₁₄₅₄), PVX_087885 (RAMA, putative), PVX_099980
447 (MSP1), PVX_096995 (tryptophan-rich antigen (Pv-fam-a)), PVX_097625 (MSP8, putative),
448 PVX_112670 (unspecified product), KMZ83376.1 (EBPII), and PVX_097720 (MSP3.10). Most
449 of these *P. vivax* proteins are not well described in the literature and only MSP1 and RAMA
450 have previously been used, or suggested as, markers of exposure^{27,40-42}. Our strategy was thus
451 successful in identifying novel exposure markers.

452

453 Individually, PVX_094255 (RBP2b₁₆₁₋₁₄₅₄) was able to classify exposure with 75% sensitivity and
454 specificity. This may be sufficient for sero-surveillance and community-level targeting of
455 preventative measures such as intensified vector control or focal test and treat¹⁹. Greater
456 classification accuracy will be required for individual-level targeting for treatment with liver-stage
457 drugs, for example in an elimination campaign using “serological testing and treatment”
458 (seroTAT), where the goal is to treat all individuals who have hypnozoites or blood-stage *P.*
459 *vivax* parasites. The final panel of eight proteins obtained 80% sensitivity and specificity across

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

466
467 Many factors affect anti-malarial antibody responses, most notably time since last infection,
468 intensity of infection, and age. Antibody responses may also be associated with other factors
469 such as human host or parasite genetic variation that are challenging to account for. Key
470 properties of the antibody response data are the high level of individual variation and the high
471 degree of correlation. Our analytic pipeline accounts for these data properties, and the
472 possibility that combinations of antibody responses may best identify recent infections.
473 However, rather than past infections being identified by complex antibody signatures and
474 sophisticated algorithms, a number of simple factors contributed to good classification
475 performance: (i) *P. vivax* proteins that can identify recent infections when used individually also
476 do well in combinations; (ii) only small incremental improvements in classification performance
477 occur as new *P. vivax* proteins are added to combinations; (iii) there are diminishing returns to
478 algorithmic complexity: simple algorithms such as logistic regression capture most signal in the
479 data, with more complex classifiers such as random forests providing only incremental
480 improvements; (iv) there is no single best combination of antigens, there is redundancy; (v)
481 there was no substantial advantage to including an individual's age once three or more antigens
482 were included; and (vi) algorithms and combinations of antigens that provided good
483 classification in one region also performed well in another region.

484

485 Although the performance of the pilot marker panel may seem modest at 80% sensitivity and
486 specificity, several factors need to be taken into account when evaluating the performance of
487 the SEMs. Firstly, given that individual *P. vivax* blood-stage infections can be relatively short
488 (i.e. <4 - 6 weeks)⁴³, some blood-stage infections may have been missed in the validation
489 cohorts as they had only monthly active follow-up. In the analyses, such cases would be
490 classified as false positive by the SEMs, when in reality the SEMs would have accurately
491 detected these infections. Therefore, the real specificity of the test may well be higher. Similarly,
492 a number of people (5-20%) with PCR-positive infections at the time of antibody measurement
493 were missed. This may represent individuals with very recent infections who have not yet
494 generated a significant rise in antibody titres⁴⁴. These missed, concurrent infections could
495 however be detected using an ultra-sensitive antigen capture test. A combination of antibody
496 detection and antigen-capture assays would further increase the sensitivity of the diagnostic
497 approach. Lastly, the SEMs are designed for application in low-transmission malaria-endemic
498 regions progressing towards elimination. Our data has previously shown that antibody
499 responses are longer lasting in areas with higher transmission²³. In elimination contexts, overall
500 (population) levels of immune responses to malarial antigens are low^{23,45}, potentially increasing
501 the difference between recently exposed and non-recently exposed individuals.

502

503 In the three settings tested, classification performance was greatest in Thailand, the region with
504 lowest transmission. The association of antibody levels with current *P. vivax* infections was less
505 clear in the Solomon Islands compared to Thailand and Brazil, and the magnitude of these
506 responses were also generally lower. In contrast to the Thai and Brazilian cohorts that included
507 individuals of all ages, the Solomon Islands was a paediatric cohort. Antibody levels tend to
508 increase with age in malaria-endemic regions^{13,28}, and so the lower responses in the Solomon
509 Islands may be due to the design of the cohort including only children. However, the difference
510 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 IgG⁴⁶, the relationship with IgM is expected to be weaker⁴⁶ and acquisition
522 and breadth of responses to malaria proteins has been shown to differ for IgG and IgM⁴⁷. 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 drugs⁵) 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

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

542
543 Similar efforts to develop SEMs are currently underway for *P. falciparum*¹⁵ and efforts to
544 develop SEMs for both species should be coordinated where possible. In addition, there is
545 evidence of antibody cross-reactivity between *P. falciparum* and *P. vivax*⁴⁸. Five of the top eight
546 proteins have *P. falciparum* orthologs but extensive cross-reactivity is unlikely given the
547 relatively low sequence homology (ranging from 20.7-37.4%). Indeed, no association was
548 detected between recent *P. falciparum* infections and antibody levels to *P. vivax* SEMs in our
549 validation. This may not hold for *P. ovale*, *P. malariae* or *P. knowlesi*, which are more closely
550 related to *P. vivax* (protein sequence identity 13.7-83.4%). On a programmatic level, limited
551 cross-reactivity between human malaria species may not be an issue in areas outside Africa as
552 rates of *P. vivax* relapse after *P. falciparum* infections are very high^{49,50}, sometimes as high as
553 following a *P. vivax* infection. Given the marked difference in patterns of exposure to the
554 zoonotic *P. knowlesi*⁵¹, cross-reactivity between *P. vivax* and *P. knowlesi* would be significantly
555 more problematic⁵².

556
557 In summary, the panel of eight novel SEMs can accurately predict recent *P. vivax* infection. As
558 almost all *P. vivax* relapses occur within nine months of a previous blood-stage infection⁵³,
559 these markers can indirectly identify individuals with the highest risk of harbouring hypnozoites.
560 With our markers able to identify individuals at-risk of future recurrent infections, we
561 demonstrate that this is indeed the case. We have shown that a carefully selected panel of
562 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

570 **Online Methods**

571

572 *Study design*

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

580

581 *Study populations: antigen discovery phase*

582 Patients with confirmed *P. vivax* malaria (by qPCR, see below) were enrolled from local malaria
583 clinics and hospitals in two sites in 2014: Tha Song Yang, Tak Province, Thailand, and Manaus,
584 Amazonas State, Brazil. After receiving anti-malarial drug treatment according to respective
585 National Treatment Guidelines, and providing written informed consent and/or assent, blood
586 samples were taken over a period of nine months as previously described²³ (Extended Data Fig
587 1a, Table S4). Presence or absence of *Plasmodium spp.* parasites during follow-up was
588 determined by both microscopy and quantitative PCR. Blood samples were collected at
589 enrolment and week 1, then every two weeks for six months, then every month until the end of
590 the study. A subset of enrolled participants who had no evidence of recurrent parasitaemia
591 during follow-up was selected for antibody measurements (n=32 Thai patients, n=33 Brazilian
592 patients).

593

594

595 *Study populations: validation phase*

596 Year-long longitudinal observational cohort studies were conducted over 2013-2014 in three
597 sites: Kanchanaburi/Ratchaburi provinces, Thailand⁵⁵, Manaus, Brazil (Monteiro et al., in
598 preparation), and the island of Ngella, Solomon Islands⁵⁶ (Extended Data Fig 1B, Table S4).
599 999 volunteers were enrolled from Thailand and sampled every month over the yearlong cohort,
600 with 14 active case detection visits performed in total. 829 volunteers attended the final visit.
601 1,274 residents of all age groups were enrolled from Brazil and sampled every month over the
602 year-long period, with 13 active case detection visits performed in total. 928 volunteers attended
603 the final visit with plasma from 925 available. 1,111 children were enrolled from the Solomon
604 Islands, with 860 used for final analysis of the cohort (after exclusion of children who withdrew,
605 had inconsistent attendance, or failed to meet other inclusion criteria). The children were
606 sampled monthly, with 11 active case detection visits in total. Of the 860 children, 754 attended
607 the final visit. For all three cohorts, at each visit volunteers completed a brief survey outlining
608 their health over the past month (to determine the possibility of missed malarial infections), in
609 addition to travel history and bed net usage. A finger-prick blood sample was also taken and
610 axillary temperature recorded. Passive case detection was performed throughout the yearlong
611 period. All cohort participants provided individual consent for both participation in study and the
612 future use of samples for the study of antimalarial immune responses. In cases of children,
613 parental consent and (where appropriate) assent was obtained.

614

615 *Study populations: pilot application phase*

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

619 made in 992 plasma samples from the Thai cohort, 1207 samples from the Brazilian cohort, and
620 655 samples from the Solomon Islands cohort.

621

622 *Study populations: malaria naïve control panels*

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

632

633 *Ethical approvals*

634 The Ethics Committee of the Faculty of Tropical Medicine, Mahidol University, Thailand,
635 approved the Thai antigen discovery study (MUTM 2014-025-01 and 02) and the Thai yearlong
636 cohort study (MUTM 2013-027-01). The Ethics Review Board of the *Fundação de Medicina*
637 *Tropical Dr Heitor Vieira Dourado* (FMT-HVD) approved the Brazilian antigen discovery study
638 ((957.875/2014) and the Brazilian yearlong cohort study (349.211/2013). The Brazilian yearlong
639 cohort study was also approved by the Brazilian National Committee of Ethics (CONEP) and by
640 the Ethics Committee of the Hospital Clínic, Barcelona, Spain (2012/7306). The National Health
641 Research and Ethics Committee of the Solomon Islands Ministry of Health and Medical
642 Services (HRC12/022) approved the Solomon Islands yearlong cohort study. The Human
643 Research Ethics Committee (HREC) at the Walter and Eliza Hall Institute of Medical Research

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

646

647 *Procedures*

648 Blood samples were collected by finger prick into EDTA tubes. Blood was separated into
649 packed red cells and plasma at the field site. Packed red cells were stored at -20°C and plasma
650 at -80°C prior to use in molecular and serological assays, respectively. *Plasmodium spp.*
651 parasites were detected by qPCR as previously described^{57,58}. The limit of detection of the
652 molecular methods was 1-3 copy numbers/ μ l⁵⁷.

653

654 *Antibody measurements*

655 The majority of *P. vivax* malaria proteins (Sal1 strain) were expressed using the wheat germ
656 cell-free (WGCF) system (CellFree Sciences, Yokohama, Japan) at either Ehime University or
657 CellFree Sciences. For the antigen discovery phase, 342 *P. vivax* protein constructs were
658 expressed at a small-scale with a biotin tag and antibodies measured using the AlphaScreen™
659 assay, as previously described²³. 307 of these proteins were selected as previously described²³,
660 and are known or expected to be immunogenic based on protein features such as the presence
661 of signal peptides, transmembrane domains, or orthology with immunogenic *P. falciparum*
662 proteins. An additional 35 proteins were added to enrich the largely blood-stage expressed set
663 of proteins with pre-erythrocytic and sexual stage proteins. Plasma samples from weeks zero
664 (enrolment), 12, 24 and 36 in the subset of volunteers described above were used (n=32 Thai
665 patients, n=33 Brazilian patients). Raw AlphaScreen™ data was converted to relative antibody
666 units based on plate-specific standard curves, with final units ranging from 0-400 (seropositivity
667 was defined as a relative antibody unit greater than 0).

668

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 described¹³, 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 described²⁸. 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 described²³. Estimated half-lives for 307 *P. vivax* proteins were previously reported²³,
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

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

697

698 where α_j^0 is the geometric mean titre (GMT) at the time of infection; $\log(\alpha_{ij})$ 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 $\varepsilon_j \sim N(0, \sigma_{m,j})$
702 is the error term. It is assumed that the random effects for initial antibody titres are
703 Normally distributed: $\log(\alpha_{ij}) \sim N(0, \sigma_{A,j})$, 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

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

712

$$713 \quad \mathbf{P}(x_{ij}|t) = \frac{1}{\sqrt{2\pi(\sigma_{\alpha,j}^2 + t^2\sigma_r^2 + \sigma_m^2)}} e^{-\frac{(x_{ij} - \alpha_j^0 - r_j^0 t)^2}{2(\sigma_{\alpha,j}^2 + t^2\sigma_r^2 + \sigma_m^2)}} \quad (2)$$

714

715 The probability for the time since infection t given measured antibody titre x_{ij} can be calculated
716 by inverting equation (2) using Bayes rule⁶⁰, allowing estimation of the probability that the time
717 since last infection was less than nine months.

718

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

724

725 *Statistical modelling – validation phase*

726 Individuals in the validation phase were classified into two categories depending on whether
727 they had PCR-detectable blood-stage *P. vivax* infections within the nine months prior to
728 measurement of antibody responses (Table S4). In theory it is possible to provide quantitative
729 estimates of the time since last infection, however, in practice, a more useful outcome is
730 whether an individual has been infected within some past time period. Nine months was
731 selected due to the biology of *P. vivax* relapses (highest incidence of relapse within the first nine
732 months following mosquito bite infection²⁰), and because this threshold fell within a time period
733 for which data were available (the three longitudinal cohorts had follow-up for 12 months).

734

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

748

749 After determining the highest ranking SEMs using the LDA classifier, the top eight candidates
750 were further assessed using a range of other classification algorithms, including logistic
751 regression, quadratic discriminant analysis (QDA), decision trees, and random forests⁶².
752 Decision trees were implemented using the rpart R package. Random forests were
753 implemented using the randomForest R package. Two novel classification algorithms
754 incorporating statistical models of antibody decay over time were also developed (see Appendix
755 S1). These algorithms were combined into a composite algorithm so that the optimal algorithm
756 and selection of proteins was selected for each target of sensitivity and specificity. All
757 classification algorithms were cross-validated using 1000 randomly sampled, disjoint training
758 and testing subsets (Extended Data Figure 5 and Figure S11).

759

760 *Statistical modelling – pilot application phase*

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

768

769 The potential public health impact of seroTAT strategies with primaquine was assessed using a
770 mathematical model of *P. vivax* transmission³⁸, using the sensitivity and specificity targets
771 achieved in the validation phase. Two annual rounds of seroTAT were modelled in the Thai,
772 Brazilian and Solomon Islands populations at 80% coverage (Fig 5, Fig S12).

773

774

775

776 **Acknowledgments**

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

786

787 **Funding**

788 We acknowledge funding from the Global Health Innovative Technology Fund (T2015-142 to
789 IM), the National Institute of Allergy and Infectious Diseases (NIH grant 5R01 AI 104822 to JS
790 and 5U19AI089686-06 to JK) and the National Health and Medical Research Council Australia
791 (#1092789 and #1134989 to IM and #1143187 to W-HT). Cohort samples were derived from
792 field studies originally funded by the TransEPI consortium (supported by the Bill and Melinda
793 Gates Foundation). This work has been supported by FIND with funding from the Australian and
794 British governments. We also acknowledge support from the National Research Council of
795 Thailand. This work was made possible through Victorian State Government Operational
796 Infrastructure Support and Australian Government NHMRC IRIISS. IM is supported by an
797 NHMRC Senior Research Fellowship (1043345). DLD is supported by an NHMRC Principal
798 Research Fellowship (1023636). TT was supported in part by JSPS KAKENHI (JP15H05276,
799 JP16K15266) in Japan. W.H.T. is a Howard Hughes Medical Institute-Wellcome Trust
800 International Research Scholar (208693/Z/17/Z). R.J.L. received the Page Betheras Award from

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

803

804 **Competing interests**

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

810

811 **Author contributions**

812 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.
813 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.
814 expressed the proteins. R.J.L., E.T., M.M., J.B., and Z.S.-J.L. performed antibody
815 measurements. R.J.L., C.S.N.L.-W.-S. and M.T.W. did the data management and analysis.
816 M.T.W. and T.O. performed modelling. R.J.L., M.T.W. and I.M. wrote the draft of the report.
817 L.J.R., C.P., D.L.D., X.C.D. and I.J.G. provided expert advice. All authors contributed to data
818 interpretation and revision of the report.

819

820 **Data and code availability**

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

- 824 1. Sutanto, I., *et al.* Negligible Impact of Mass Screening and Treatment on Mesoendemic Malaria
825 Transmission at West Timor in Eastern Indonesia: A Cluster-Randomized Trial. *Clin Infect Dis* **67**, 1364-
826 1372 (2018).
- 827 2. Waltmann, A., *et al.* High Rates of Asymptomatic, Sub-microscopic *Plasmodium vivax* Infection and
828 Disappearing *Plasmodium falciparum* Malaria in an Area of Low Transmission in Solomon Islands. *PLoS*
829 *Negl Trop Dis* **9**, e0003758 (2015).
- 830 3. Mueller, I., *et al.* Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite.
831 *Lancet Infect Dis* **9**, 555-566 (2009).
- 832 4. Moreira, C.M., Abo-Shehada, M., Price, R.N. & Drakeley, C.J. A systematic review of sub-microscopic
833 *Plasmodium vivax* infection. *Malar J* **14**, 360 (2015).
- 834 5. Robinson, L.J., *et al.* Strategies for understanding and reducing the *Plasmodium vivax* and *Plasmodium*
835 *ovale* hypnozoite reservoir in Papua New Guinean children: a randomised placebo-controlled trial and
836 mathematical model. *PLoS Med* **12**, e1001891 (2015).
- 837 6. Parker, D.M., *et al.* Microgeography and molecular epidemiology of malaria at the Thailand-Myanmar
838 border in the malaria pre-elimination phase. *Malar J* **14**, 198 (2015).
- 839 7. Mogeni, P., *et al.* Detecting Malaria Hotspots: A Comparison of Rapid Diagnostic Test, Microscopy, and
840 Polymerase Chain Reaction. *J Infect Dis* **216**, 1091-1098 (2017).
- 841 8. Ding, X.C., *et al.* Defining the next generation of *Plasmodium vivax* diagnostic tests for control and
842 elimination: Target product profiles. *PLoS Negl Trop Dis* **11**, e0005516 (2017).
- 843 9. Malaria surveillance, monitoring & evaluation: a reference manual. (2018).
- 844 10. Harris, I., *et al.* A large proportion of asymptomatic *Plasmodium* infections with low and sub-microscopic
845 parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for
846 malaria diagnostics in an elimination setting. *Malar J* **9**, 254 (2010).
- 847 11. Sutanto, I., *et al.* Negligible Impact of Mass Screening and Treatment on Meso-endemic Malaria
848 Transmission at West Timor in Eastern Indonesia: A Cluster-Randomised Trial. *Clin Infect Dis* (2018).
- 849 12. Ong, K.I.C., *et al.* Systematic review of the clinical manifestations of glucose-6-phosphate dehydrogenase
850 deficiency in the Greater Mekong Subregion: implications for malaria elimination and beyond. *BMJ Glob*
851 *Health* **2**, e000415 (2017).
- 852 13. Longley, R.J., *et al.* Asymptomatic *Plasmodium vivax* infections induce robust IgG responses to multiple
853 blood-stage proteins in a low-transmission region of western Thailand. *Malar J* **16**, 178 (2017).
- 854 14. Longley, R.J., *et al.* Acquisition and Longevity of Antibodies to *Plasmodium vivax* Preerythrocytic
855 Antigens in Western Thailand. *Clin Vaccine Immunol* **23**, 117-124 (2016).
- 856 15. Helb, D.A., *et al.* Novel serologic biomarkers provide accurate estimates of recent *Plasmodium falciparum*
857 exposure for individuals and communities. *Proc Natl Acad Sci U S A* **112**, E4438-4447 (2015).
- 858 16. Welch, R.J., Anderson, B.L. & Litwin, C.M. Rapid immunochromatographic strip test for detection of anti-
859 K39 immunoglobulin G antibodies for diagnosis of visceral leishmaniasis. *Clin Vaccine Immunol* **15**, 1483-
860 1484 (2008).
- 861 17. Gwyn, S., *et al.* Lateral flow-based antibody testing for *Chlamydia trachomatis*. *J Immunol Methods* **435**,
862 27-31 (2016).
- 863 18. Landry, M.L. Diagnostic tests for influenza infection. *Curr Opin Pediatr* **23**, 91-97 (2011).
- 864 19. Greenhouse, B., Smith, D.L., Rodriguez-Barraquer, I., Mueller, I. & Drakeley, C. Taking sharper pictures
865 of malaria with CAMERAs: Combined Antibodies to Measure Exposure Recency Assays. *bioRxiv* (2018).
- 866 20. White, N.J. Determinants of relapse periodicity in *Plasmodium vivax* malaria. *Malar J* **10**, 297 (2011).
- 867 21. Battle, K.E., *et al.* Geographical variation in *Plasmodium vivax* relapse. *Malar J* **13**, 144 (2014).
- 868 22. White, M.T., Shirreff, G., Karl, S., Ghani, A.C. & Mueller, I. Variation in relapse frequency and the
869 transmission potential of *Plasmodium vivax* malaria. *Proc Biol Sci* **283**, 20160048 (2016).
- 870 23. Longley, R.J., *et al.* Naturally acquired antibody responses to more than 300 *Plasmodium vivax* proteins in
871 three geographic regions. *PLoS Negl Trop Dis* **11**, e0005888 (2017).
- 872 24. Tsuboi, T., *et al.* Wheat germ cell-free system-based production of malaria proteins for discovery of novel
873 vaccine candidates. *Infect Immun* **76**, 1702-1708 (2008).
- 874 25. Franca, C.T., *et al.* Identification of highly-protective combinations of *Plasmodium vivax* recombinant
875 proteins for vaccine development. *Elife* **6**(2017).
- 876 26. Franca, C.T., *et al.* *Plasmodium vivax* Reticulocyte Binding Proteins Are Key Targets of Naturally
877 Acquired Immunity in Young Papua New Guinean Children. *PLoS Negl Trop Dis* **10**, e0005014 (2016).

- 878 27. Lu, F., *et al.* Profiling the humoral immune responses to Plasmodium vivax infection and identification of
879 candidate immunogenic rhoptry-associated membrane antigen (RAMA). *J Proteomics* **102**, 66-82 (2014).
- 880 28. Franca, C.T., *et al.* An Antibody Screen of a Plasmodium vivax Antigen Library Identifies Novel
881 Merozoite Proteins Associated with Clinical Protection. *PLoS Negl Trop Dis* **10**, e0004639 (2016).
- 882 29. Gruszczuk, J., *et al.* Transferrin receptor 1 is a reticulocyte-specific receptor for Plasmodium vivax. *Science*
883 **359**, 48-55 (2018).
- 884 30. Soares, I.S., *et al.* A Plasmodium vivax vaccine candidate displays limited allele polymorphism, which
885 does not restrict recognition by antibodies. *Mol Med* **5**, 459-470 (1999).
- 886 31. Wang, B., *et al.* Immunoprofiling of the tryptophan-rich antigen family in Plasmodium vivax. *Infect Immun*
887 **83**, 3083-3095 (2015).
- 888 32. Cheng, Y., *et al.* Naturally acquired humoral and cellular immune responses to Plasmodium vivax
889 merozoite surface protein 8 in patients with P. vivax infection. *Malar J* **16**, 211 (2017).
- 890 33. Hester, J., *et al.* De novo assembly of a field isolate genome reveals novel Plasmodium vivax erythrocyte
891 invasion genes. *PLoS Negl Trop Dis* **7**, e2569 (2013).
- 892 34. Jiang, J., Barnwell, J.W., Meyer, E.V. & Galinski, M.R. Plasmodium vivax merozoite surface protein-3
893 (PvMSP3): expression of an 11 member multigene family in blood-stage parasites. *PLoS One* **8**, e63888
894 (2013).
- 895 35. Hietanen, J., *et al.* Gene Models, Expression Repertoire, and Immune Response of Plasmodium vivax
896 Reticulocyte Binding Proteins. *Infect Immun* **84**, 677-685 (2015).
- 897 36. Cole-Tobian, J.L., *et al.* Strain-specific duffy binding protein antibodies correlate with protection against
898 infection with homologous compared to heterologous plasmodium vivax strains in Papua New Guinean
899 children. *Infect Immun* **77**, 4009-4017 (2009).
- 900 37. Ntege, E.H., *et al.* Blood-stage malaria vaccines: post-genome strategies for the identification of novel
901 vaccine candidates. *Expert Rev Vaccines* **16**, 769-779 (2017).
- 902 38. White, M.T., *et al.* Mathematical modelling of the impact of expanding levels of malaria control
903 interventions on Plasmodium vivax. *Nature communications* **9**, 3300 (2018).
- 904 39. Rabinovich, R.N., *et al.* malERA: An updated research agenda for malaria elimination and eradication.
905 *PLoS Med* **14**, e1002456 (2017).
- 906 40. Cook, J., *et al.* Using serological measures to monitor changes in malaria transmission in Vanuatu. *Malar J*
907 **9**, 169 (2010).
- 908 41. Ladeia-Andrade, S., *et al.* Naturally acquired antibodies to merozoite surface protein (MSP)-1(19) and
909 cumulative exposure to Plasmodium falciparum and Plasmodium vivax in remote populations of the
910 Amazon Basin of Brazil. *Mem Inst Oswaldo Cruz* **102**, 943-951 (2007).
- 911 42. Cook, J., *et al.* Sero-epidemiological evaluation of changes in Plasmodium falciparum and Plasmodium
912 vivax transmission patterns over the rainy season in Cambodia. *Malar J* **11**, 86 (2012).
- 913 43. White, M.T., *et al.* Plasmodium vivax and Plasmodium falciparum infection dynamics: re-infections,
914 recrudescences and relapses. *Malar J* **17**, 170 (2018).
- 915 44. White, M.T., *et al.* Dynamics of the antibody response to Plasmodium falciparum infection in African
916 children. *J Infect Dis* **210**, 1115-1122 (2014).
- 917 45. Boyle, M.J., Reiling, L., Osier, F.H. & Fowkes, F.J. Recent insights into humoral immunity targeting
918 Plasmodium falciparum and Plasmodium vivax malaria. *Int J Parasitol* **47**, 99-104 (2017).
- 919 46. Richards, J.S., *et al.* Association between naturally acquired antibodies to erythrocyte-binding antigens of
920 Plasmodium falciparum and protection from malaria and high-density parasitemia. *Clin Infect Dis* **51**, e50-
921 60 (2010).
- 922 47. Arama, C., *et al.* Genetic Resistance to Malaria Is Associated With Greater Enhancement of
923 Immunoglobulin (Ig)M Than IgG Responses to a Broad Array of Plasmodium falciparum Antigens. *Open*
924 *Forum Infect Dis* **2**, ofv118 (2015).
- 925 48. Woodberry, T., *et al.* Antibodies to Plasmodium falciparum and Plasmodium vivax merozoite surface
926 protein 5 in Indonesia: species-specific and cross-reactive responses. *J Infect Dis* **198**, 134-142 (2008).
- 927 49. Douglas, N.M., *et al.* Plasmodium vivax recurrence following falciparum and mixed species malaria: risk
928 factors and effect of antimalarial kinetics. *Clin Infect Dis* **52**, 612-620 (2011).
- 929 50. Looareesuwan, S., White, N.J., Chittamas, S., Bunnag, D. & Harinasuta, T. High rate of Plasmodium vivax
930 relapse following treatment of falciparum malaria in Thailand. *Lancet* **2**, 1052-1055 (1987).
- 931 51. Fornace, K.M., *et al.* Exposure and infection to Plasmodium knowlesi in case study communities in
932 Northern Sabah, Malaysia and Palawan, The Philippines. *PLoS Negl Trop Dis* **12**, e0006432 (2018).

- 933 52. Herman, L.S., *et al.* Identification and validation of a novel panel of Plasmodium knowlesi biomarkers of
934 serological exposure. *PLoS Negl Trop Dis* **12**, e0006457 (2018).
- 935 53. White, N.J. & Imwong, M. Relapse. *Adv Parasitol* **80**, 113-150 (2012).
- 936 54. Alonso, P.L., *et al.* A research agenda to underpin malaria eradication. *PLoS Med* **8**, e1000406 (2011).
- 937 55. Nguitragool, W., *et al.* Highly heterogeneous residual malaria risk in western Thailand. *Int J Parasitol*
938 (2019).
- 939 56. Quah, Y.W., *et al.* Molecular epidemiology of residual Plasmodium vivax transmission in a paediatric
940 cohort in Solomon Islands. *Malar J* **18**, 106 (2019).
- 941 57. Wampfler, R., *et al.* Strategies for detection of Plasmodium species gametocytes. *PLoS One* **8**, e76316
942 (2013).
- 943 58. Rosanas-Urgell, A., *et al.* Comparison of diagnostic methods for the detection and quantification of the four
944 sympatric Plasmodium species in field samples from Papua New Guinea. *Malar J* **9**, 361 (2010).
- 945 59. Healer, J., *et al.* Vaccination with conserved regions of erythrocyte-binding antigens induces neutralizing
946 antibodies against multiple strains of Plasmodium falciparum. *PLoS One* **8**, e72504 (2013).
- 947 60. Borremans, B., Hens, N., Beutels, P., Leirs, H. & Reijnders, J. Estimating Time of Infection Using Prior
948 Serological and Individual Information Can Greatly Improve Incidence Estimation of Human and Wildlife
949 Infections. *PLoS Comput Biol* **12**, e1004882 (2016).
- 950 61. Kirkpatrick, S., Gelatt, C.D., Jr. & Vecchi, M.P. Optimization by simulated annealing. *Science* **220**, 671-
951 680 (1983).
- 952 62. Hastie, T., Tibshirani, R. & Friedman, J. *The elements of statistical learning: data mining, inference, and*
953 *prediction*, (Springer, New York, 2009).
- 954

955

956

957

958

959

960 **Extended Data**

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

962 **Extended Data Figure 2.** Antibody responses to 60 antigens measured in n = 2,281 biologically independent
963 samples on the Luminex® platform, stratified by geographical location and time since last PCR detected infection.

964 Boxplots denote median and interquartile ranges (IQR) of the data, with whiskers denoting the median \pm 1.5*IQR.

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

967 **Extended Data Figure 4.** Detailed breakdown of classification performance for the target of
968 80% sensitivity and 80% specificity.

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

971

972 **Supporting Information**

973 **Appendix S1.** Detailed statistical appendix, including supporting figures.

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

975 **Fig S2.** Kinetics of multiple antibodies.

976 **Fig S3.** Estimates of (A) antibody half-life, and (B) the variance of each antibody response for
977 104 *P. vivax* proteins in the antigen discovery phase.

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

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

982 **Fig S6.** Network visualization of antigens.

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

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

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

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

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

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

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

996 **Table S2.** Purified *P. vivax* proteins used in the validation phase and their individual

997 performance, complete list.

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

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

1000 phases.

1001

1002

1003

1004

1005

1006

1007 **Extended Data Fig 1.** Study design and follow-up schedule. (A) Thai and Brazilian patients were enrolled following a
1008 clinical episode of *P. vivax* and treated according to the relevant National Guidelines, with directly observed treatment
1009 (DOT) to ensure compliance and reduced risk of relapse. Volunteers were followed for nine months after enrolment,
1010 with finger-prick blood samples collected at enrolment and week 1, then every two weeks for six months, then every
1011 month. Antibody levels were measured in a subset of 32 Thai and 33 Brazilian volunteers who were confirmed to be
1012 free of blood-stage *Plasmodium* parasites by analysing all samples by light microscopy and qPCR. (B) 999
1013 participants from Thailand, 1274 participants from Brazil, and 860 participants from the Solomon Islands were
1014 followed longitudinally for 12 months with active surveillance visits every month. For the analysis in the validation
1015 phase antibody levels were measured in plasma samples from the last visit. For the analysis in the application phase,
1016 antibody levels were measured in plasma samples from the first visit.

1017
1018
1019 **Extended Data Fig 2.** Measured antibody responses to 60 proteins on the Luminex® platform, stratified by
1020 geographical location and time since last PCR detected infection.

1021
1022
1023
1024 **Extended Data Fig 3.** Association between background reactivity in non-malaria exposed controls and ranking of
1025 candidate SEMs by area under the curve (AUC). Mean relative antibody units (RAU) detected in malaria-naïve
1026 control panels from Melbourne, Australia (n=202), Bangkok, Thailand (n=72) and Rio de Janeiro Brazil (n = 96)
1027 compared to the AUC of the 60 candidate *P.vivax* proteins generated during the validation phase. WGCF expressed
1028 proteins are in black and *E. coli* or Baculovirus expressed proteins are in blue. RBP2b161-1454 (*E. coli*) is in red and
1029 RBP2b1986-2653 is in orange.

1030
1031
1032 **Extended Data Fig 4.** Breakdown of the classification of the Random Forests algorithm with target sensitivity and
1033 specificity of 80%. The size of each rectangle is proportional to the number of samples in each category (See Table
1034 S4 of accompanying manuscript). The coloured area represents the proportion correctly classified, and the shaded
1035 area represents the proportion mis-classified.

1036
1037
1038 **Extended Data Fig 5.** Receiver operating characteristic (ROC) curve for the composite classification algorithm. All
1039 curves presented are the median of 1000 repeat cross-validations.

1040

(A) Antigen discovery phase

samples: Thailand (32); Brazil (33)
assay: AlphaScreen (342 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

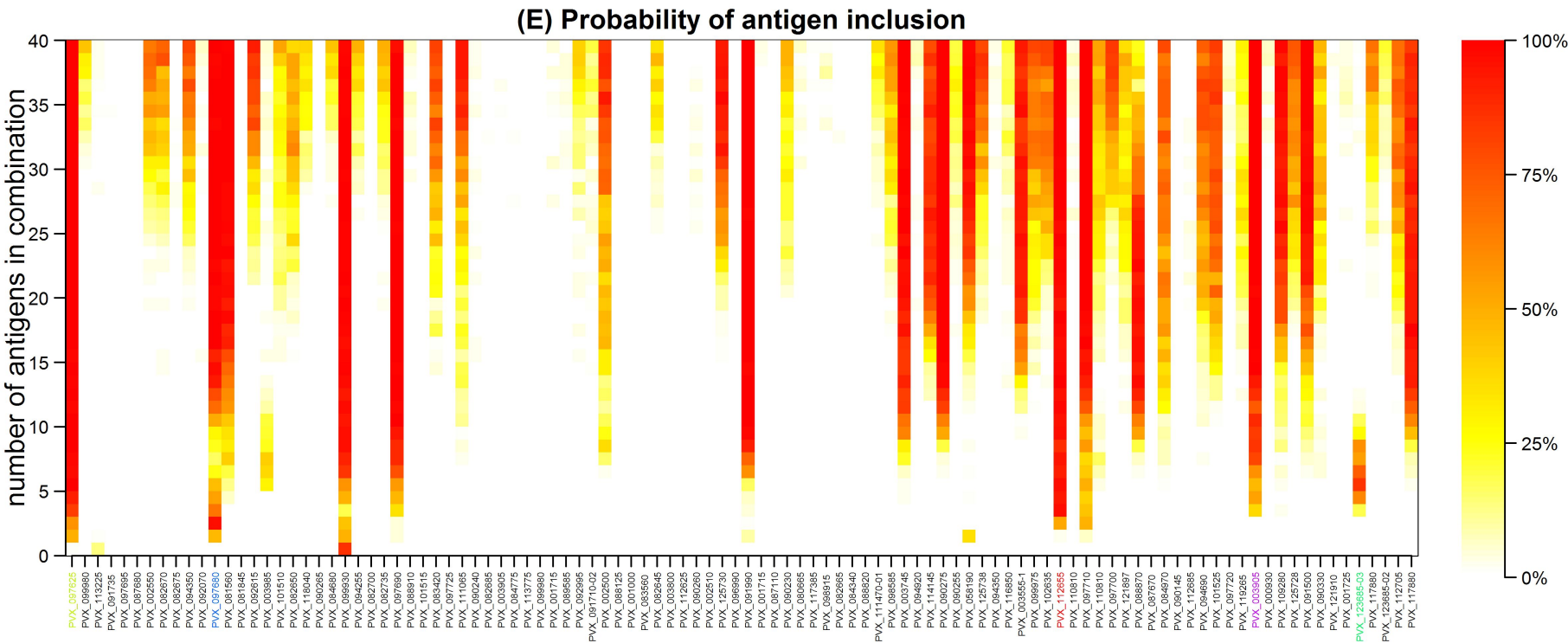
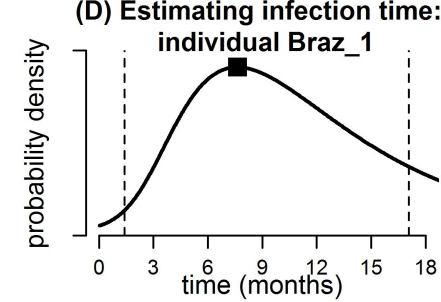
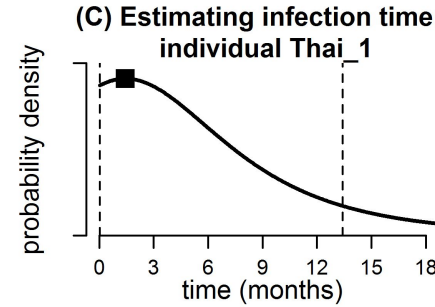
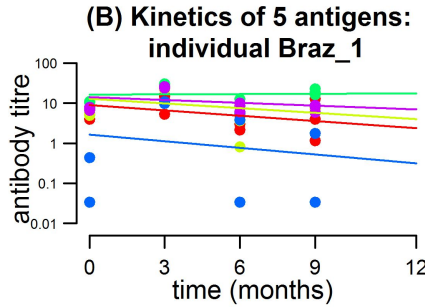
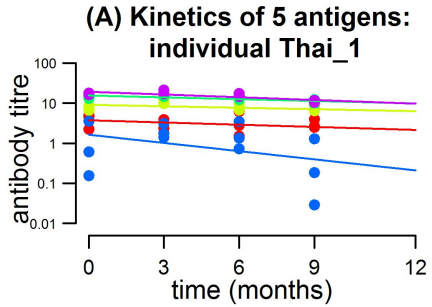
Model-based down selection (2)

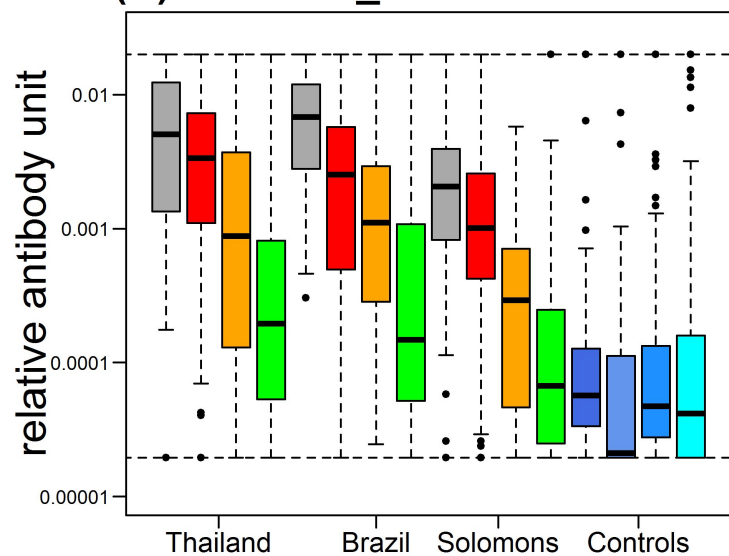
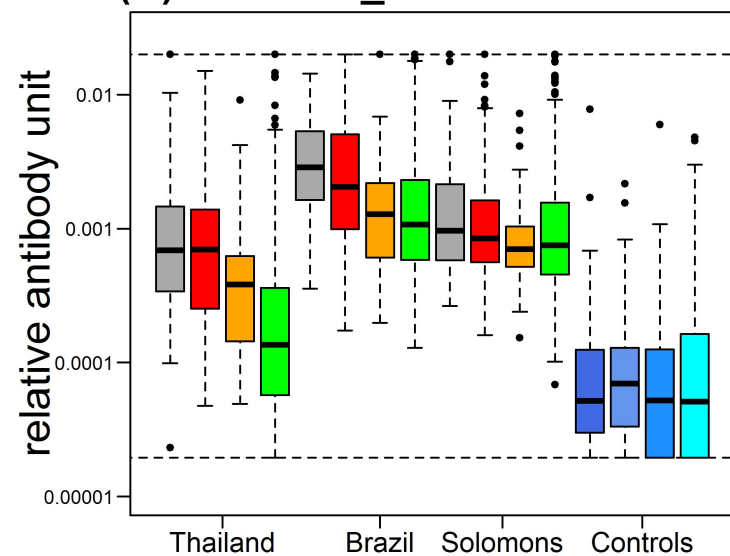
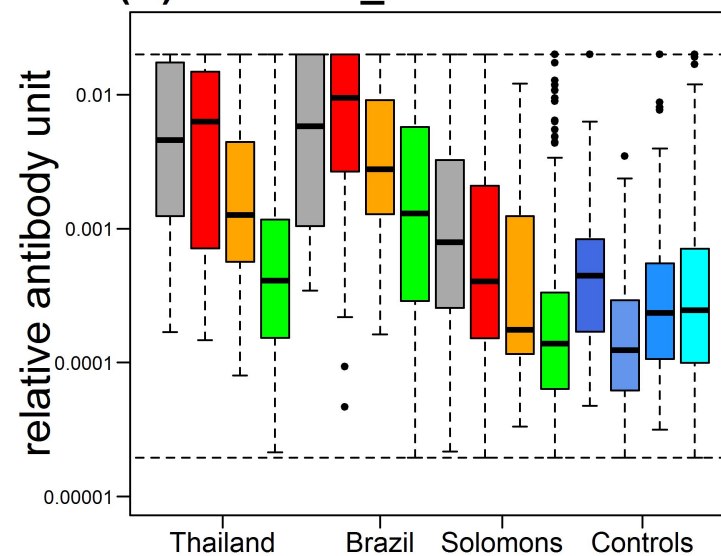
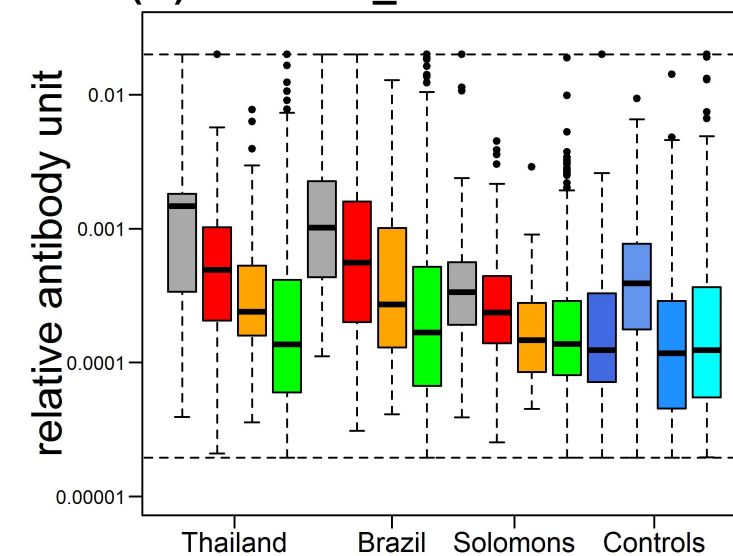
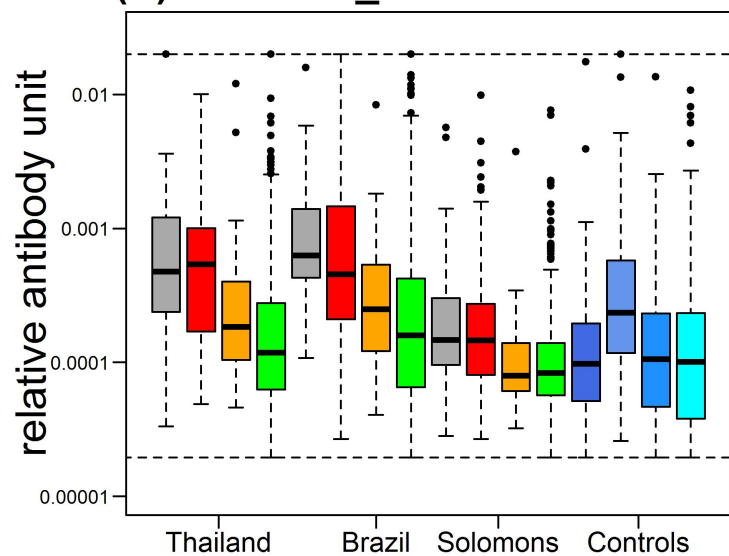
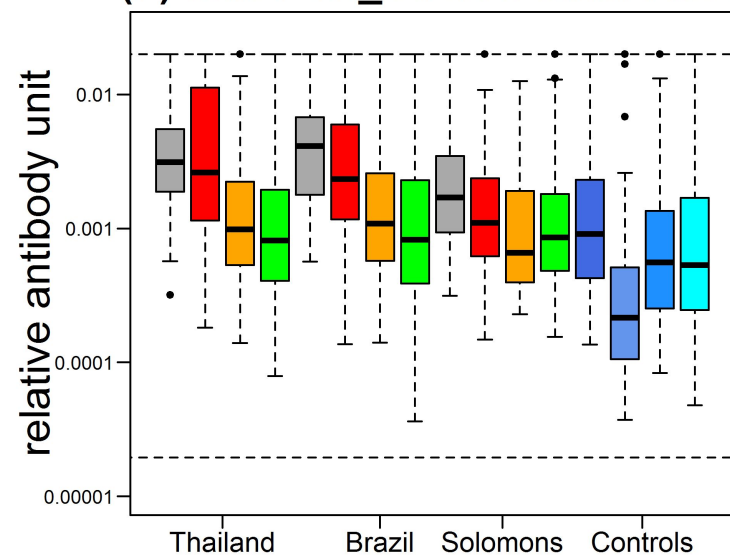
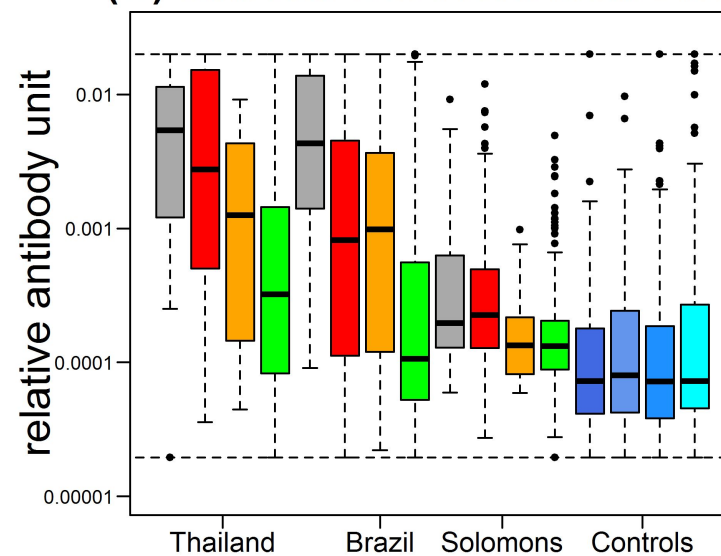
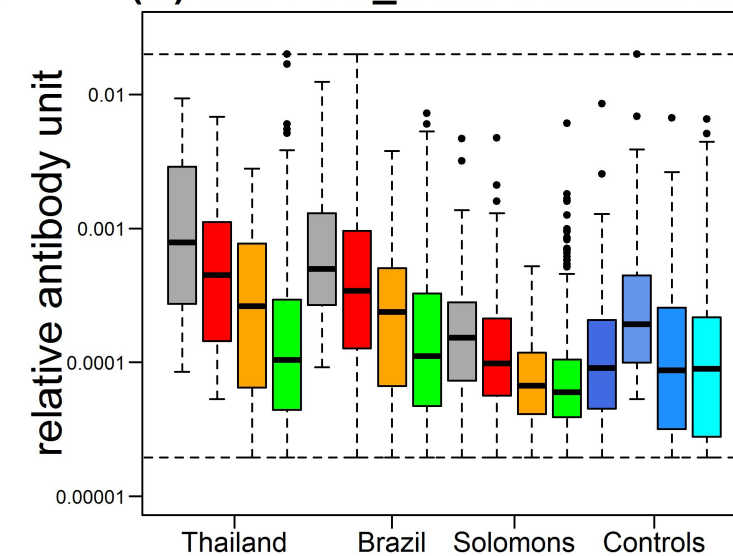
- combinations analyzed with classification algorithms
- selection based on prediction within last 9 months

8 *P. vivax* proteins

(B) Validation phase

samples: Thailand (829); Brazil (928);
Solomons (754); negative controls (274)
assay: Luminex (60 proteins)



(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

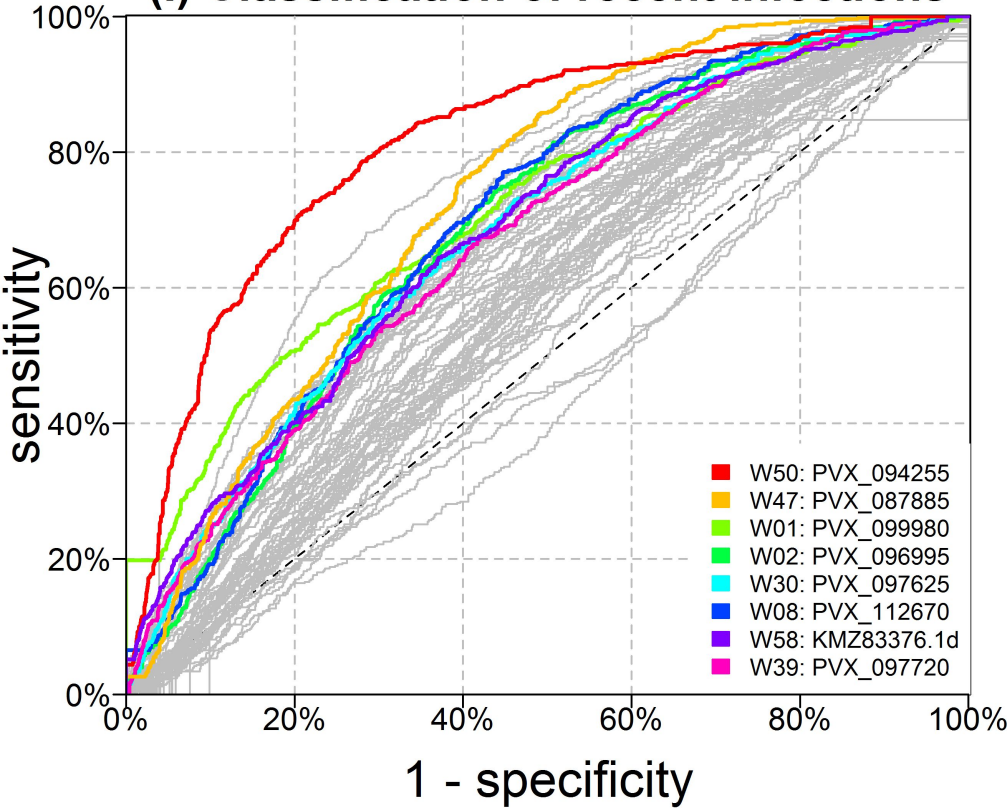
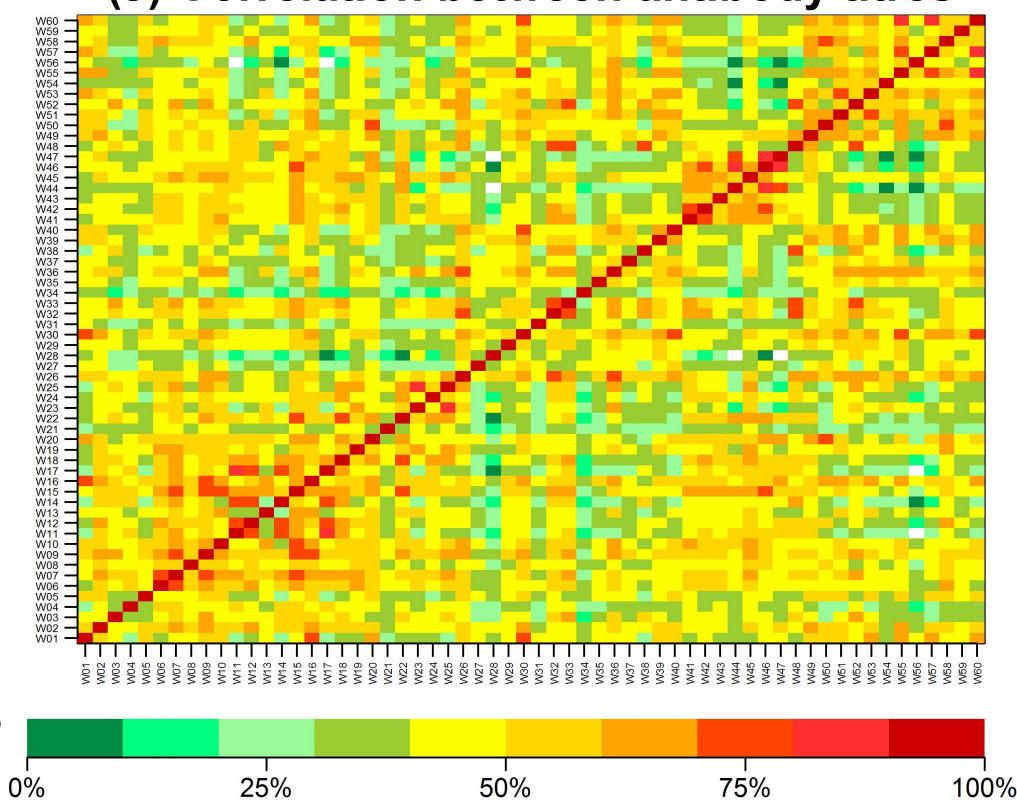
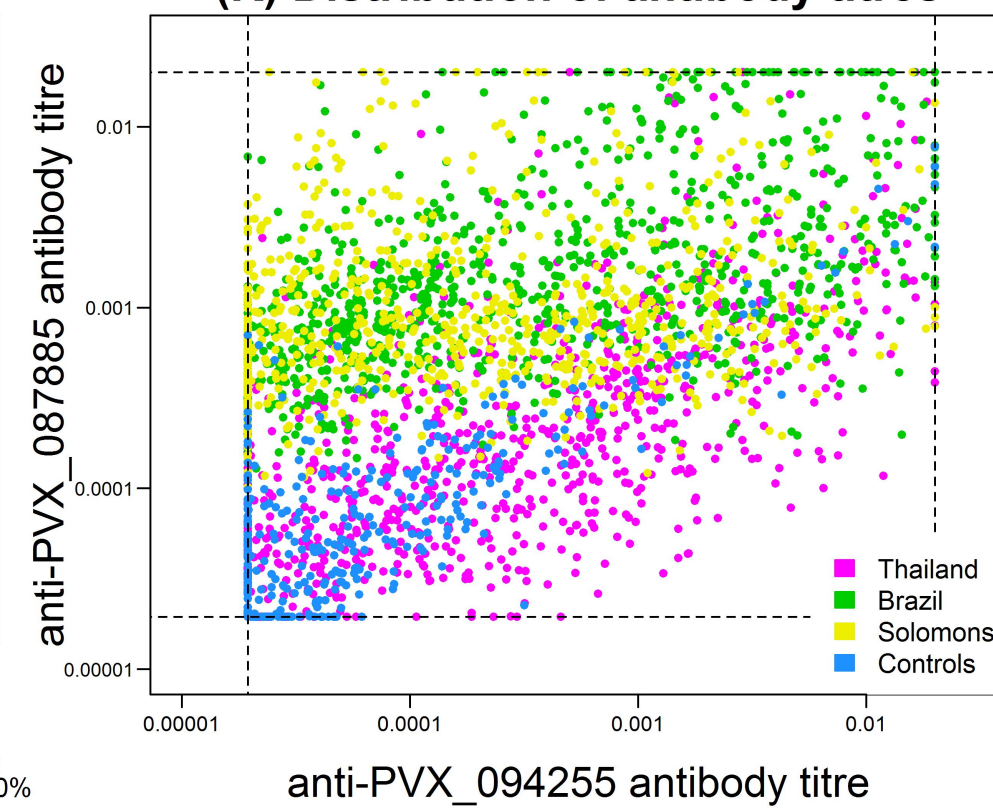
■ no detected infection

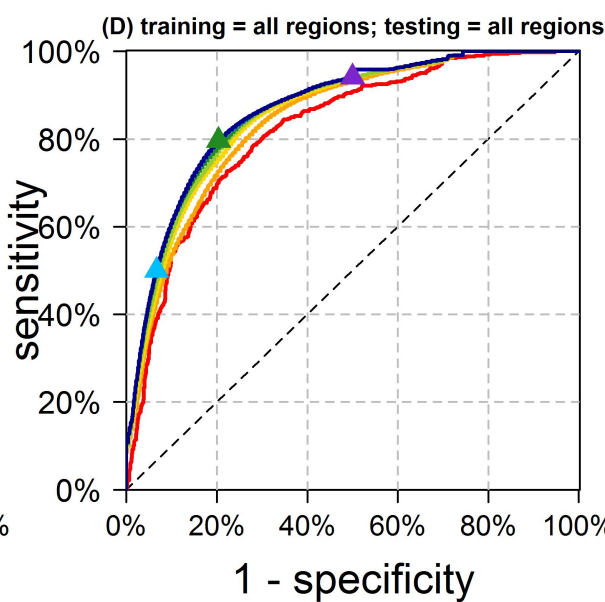
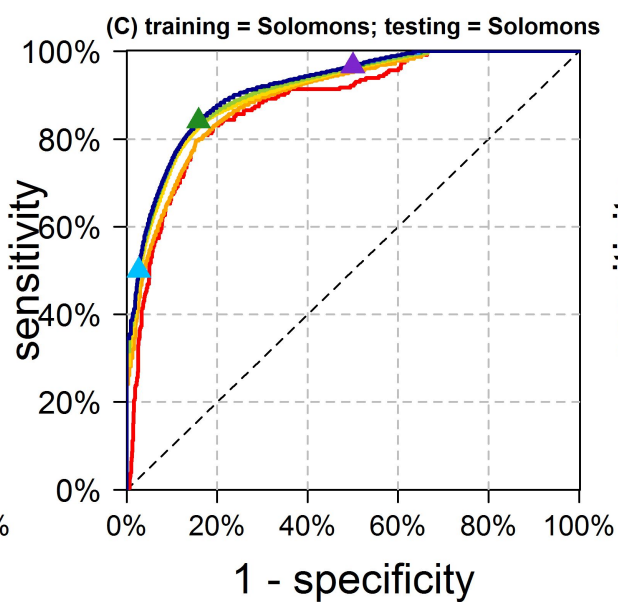
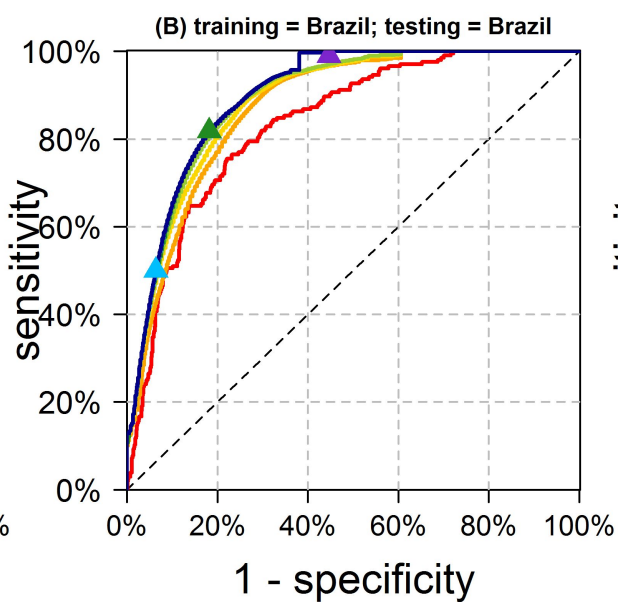
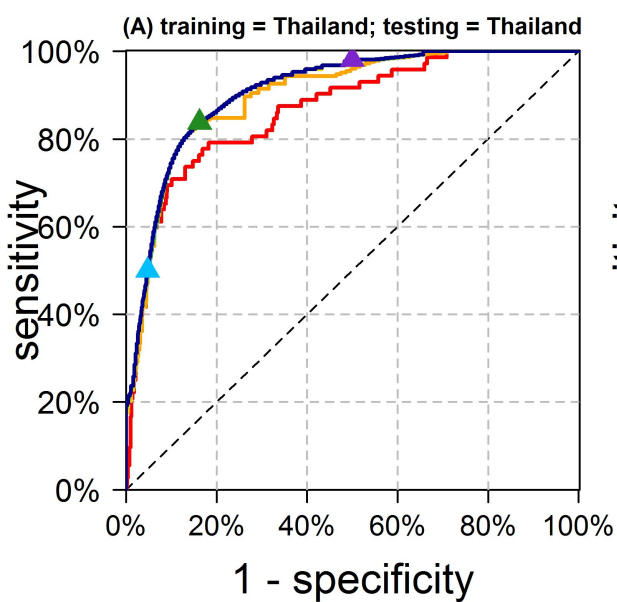
■ negative controls: Thai RC

■ negative controls: Brazil RC

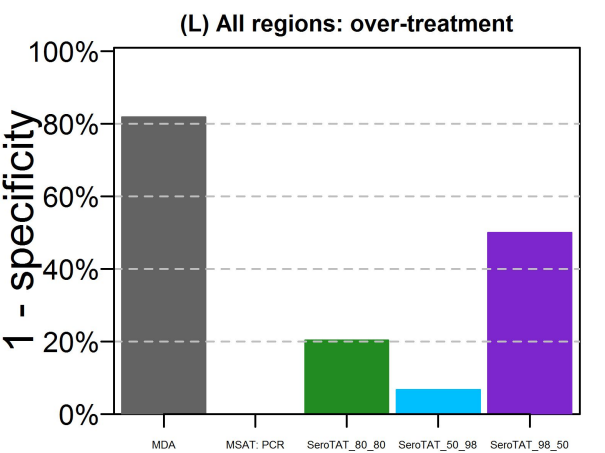
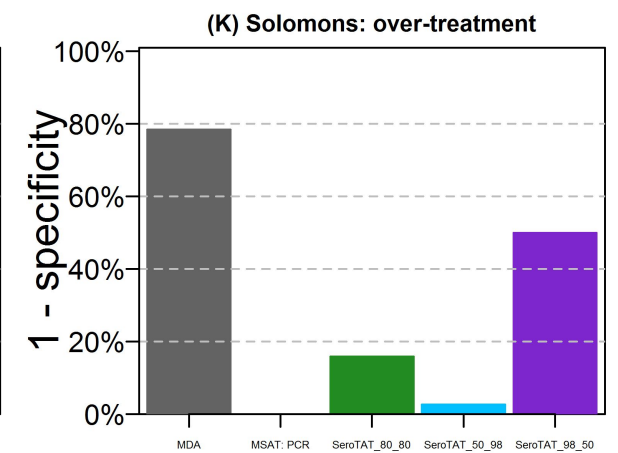
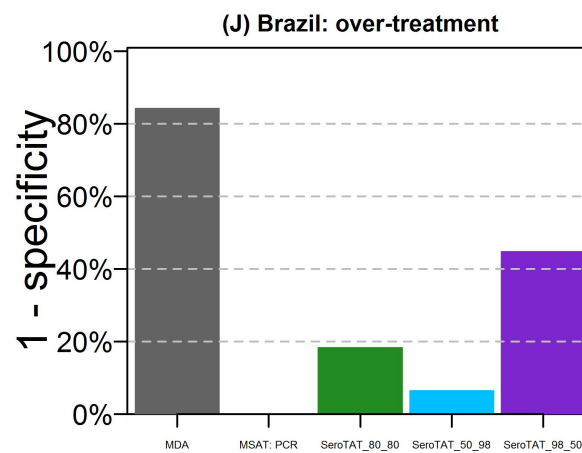
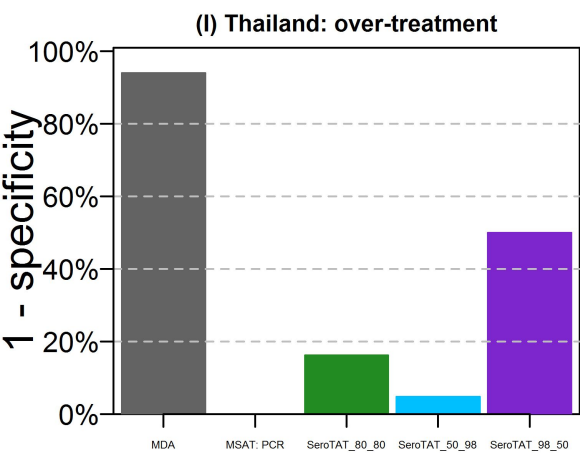
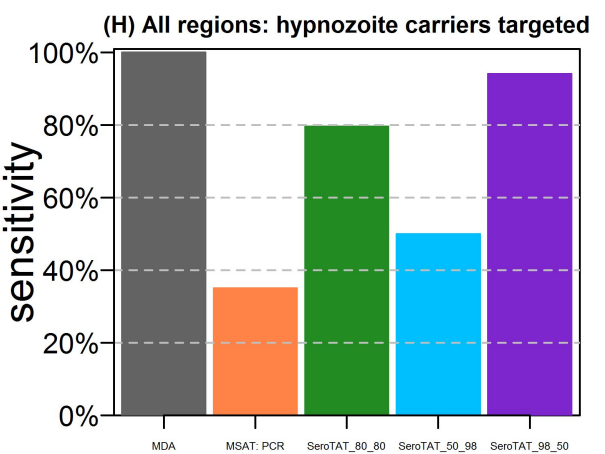
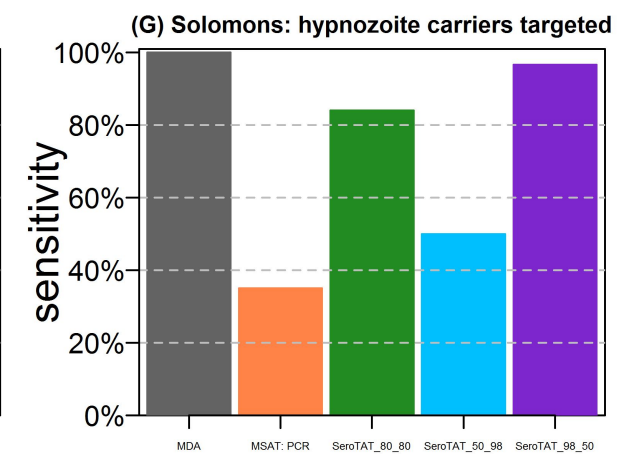
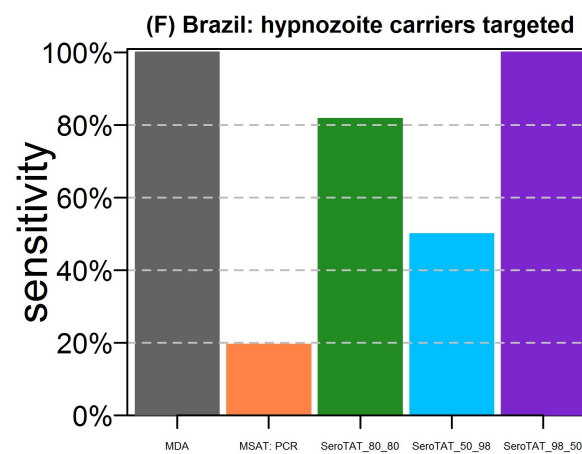
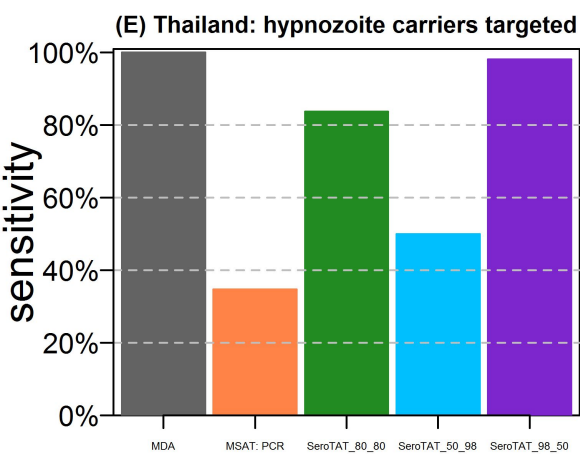
■ negative controls: Aus RC

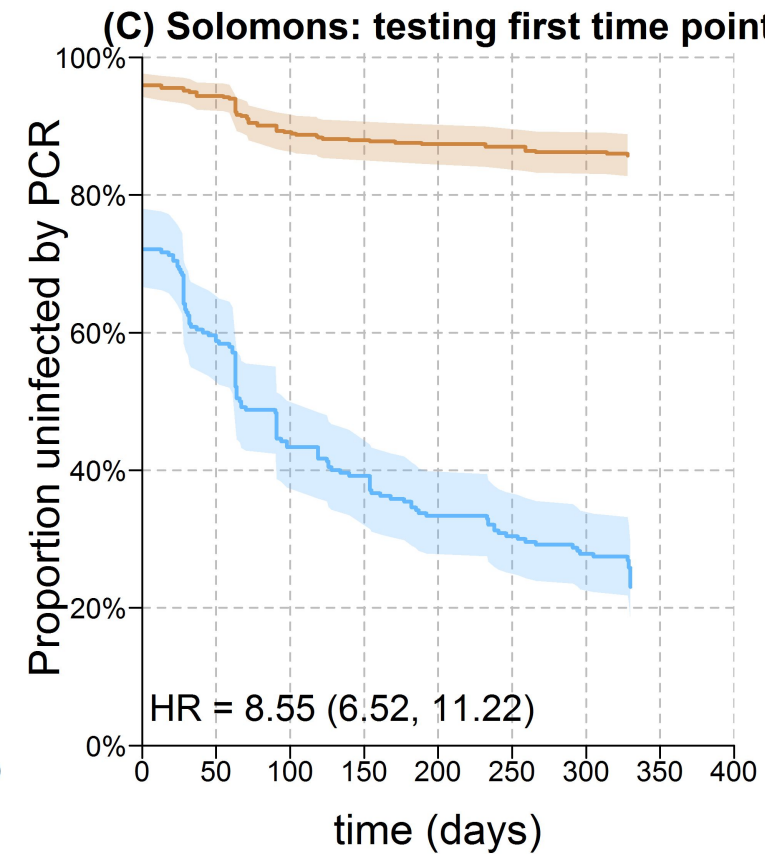
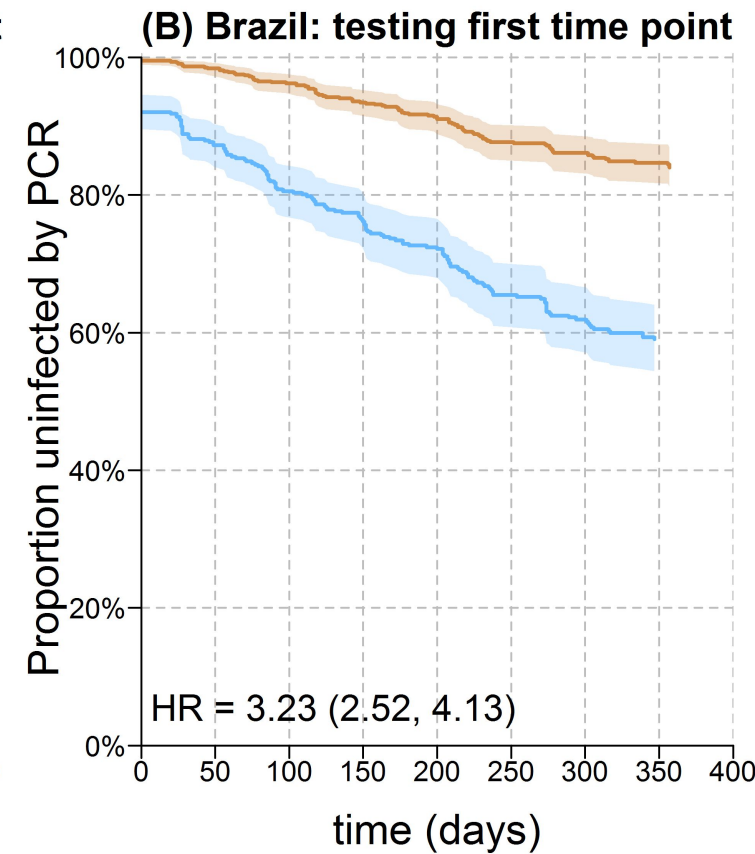
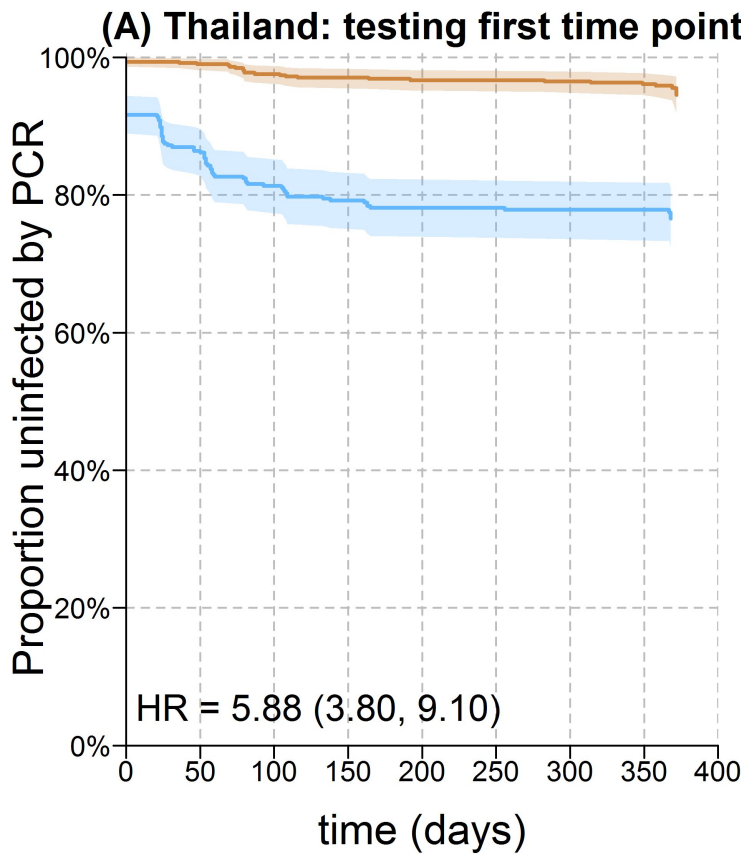
■ negative controls: VBDR

(I) Classification of recent infections**(J) Correlation between antibody titres****(K) Distribution of antibody titres**

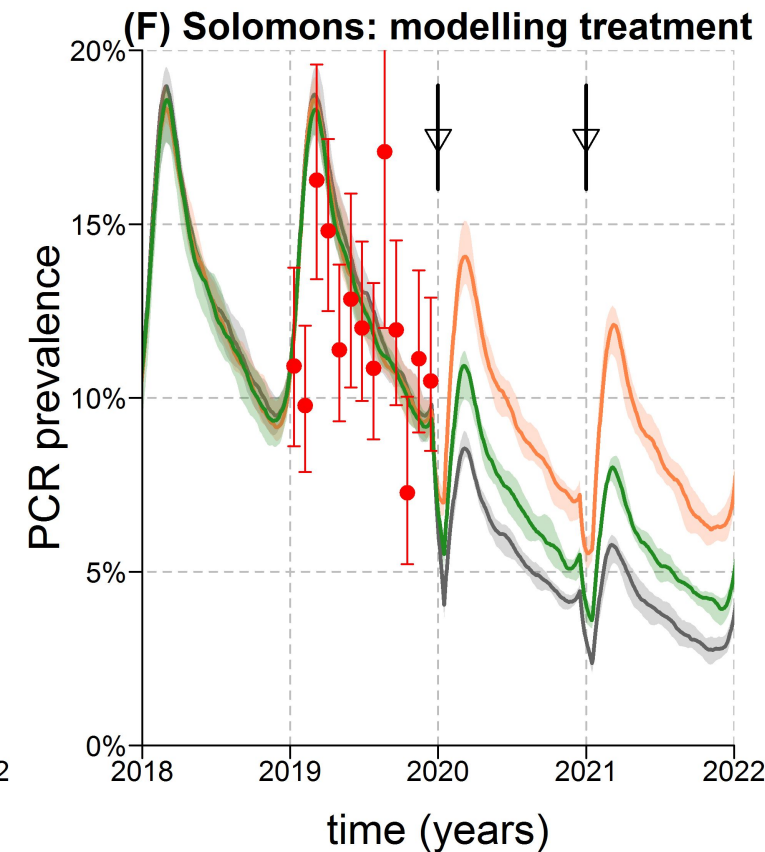
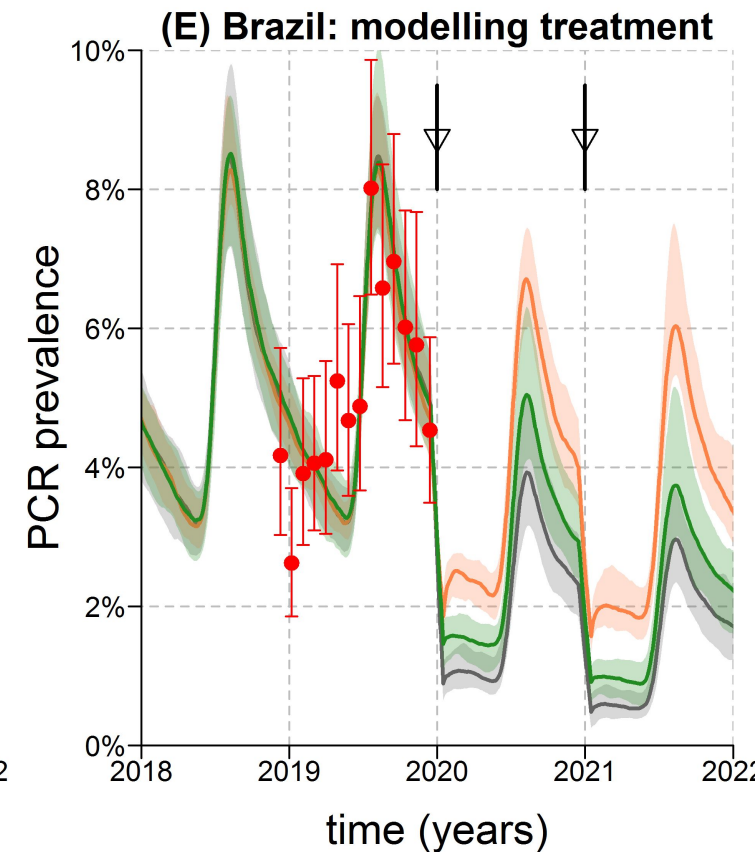
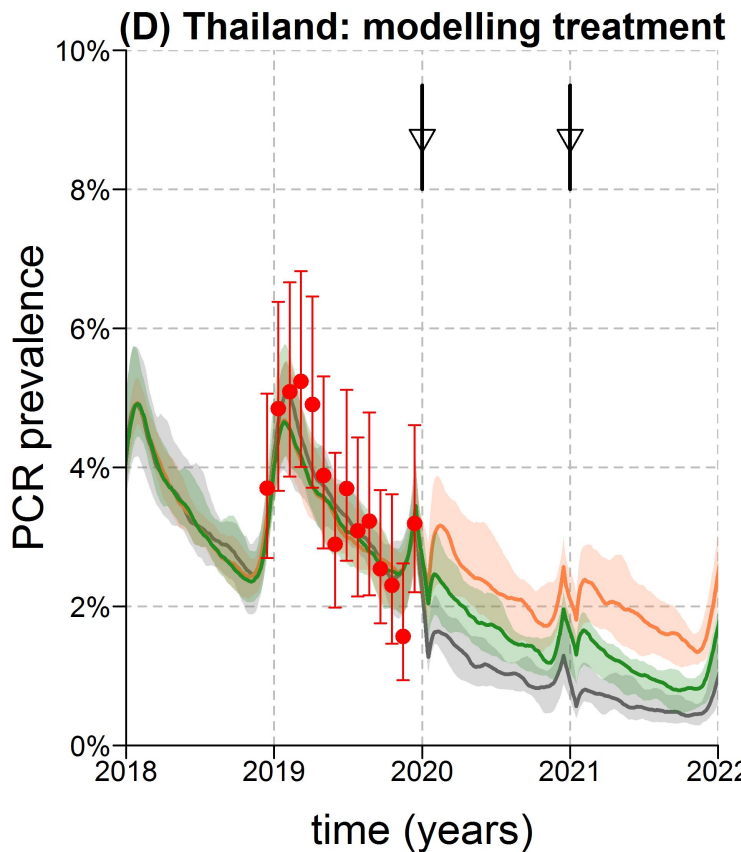


■ 1 antigen
 ■ 2 antigens
 ■ 3 antigens
 ■ 4 antigens
 ■ 5 antigens
 ■ 6 antigens
 ■ 7 antigens
 ■ 8 antigens



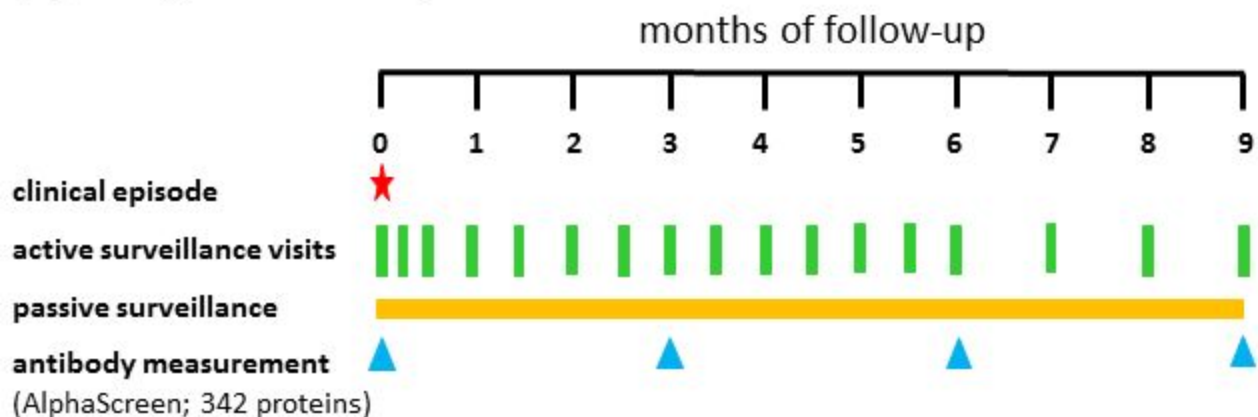


— first time point classified positive — first time point classified negative



● data: PCR prevalence ▼ population treatment — MDA — MSAT: LM — SeroTAT: 80%, 80%

(A) Antigen discovery data



(B) Validation and application data

