Reverse immunodynamics: a new method for identifying targets of protective immunity

Katrina J Spensley[a*], Paul S Wikramaratna[b*], Bridget S Penman[c], Andrew Walker[c], Adrian L Smith[c], Oliver G Pybus[c], Létitia Jean[d], Sunetra Gupta[c] & José Lourenço[c**]

a. Imperial College London, London W2 1PG
b. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT
c. Department of Zoology, University of Oxford, Oxford OX1 3PS
d. The Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE

*these authors contributed equally to the manuscript

**Correspondence to: jose.lourenco@zoo.ox.ac.uk

Abstract

Despite a dramatic increase in our ability to catalogue variation among pathogen genomes, we have made far fewer advances in using this information to identify targets of protective immunity. We propose a novel methodology that combines predictions from epidemiological models with phylogenetic and structural analyses to identify such targets. Epidemiological models predict that strong immune selection can cause antigenic variants to exist in non-overlapping combinations. A corollary of this theory is that targets of immunity may be identified by searching for non-overlapping associations among antigenic variants. We applied this concept to the AMA-1 protein of the malaria parasite *Plasmodium falciparum* and found strong signatures of immune selection among certain regions of low variability which could render them ideal vaccine candidates.

Introduction

Apical membrane antigen 1 (AMA-1) is a trans-membrane protein common to all *Plasmodium* species with a putative role in invasion at erythrocytic and pre-erythrocytic stages of the parasite life-cycle. Antibodies to AMA-1 appear to inhibit erythrocyte invasion and are thus thought to play a role in clinical protection against both *P. falciparum* [1-7] and *P. vivax* [8]. Cellular responses to AMA-1 may provide immunity against both erythrocytic [9] and pre-erythrocytic forms of *P. falciparum* [10]. The potential of AMA-1 as a component of a multi-stage malaria vaccine is compromised by the polymorphism of many of the targets of natural immunity [11,12] as well as by difficulties in inducing lasting immune responses through vaccination [13-16]. Identifying variable regions of AMA-1 that are under strong immune selection would afford the possibility of focusing vaccine-induced immune responses on these regions and ensure full coverage of variability at these critical epitopes.

The extracellular region of AMA-1 can be structurally resolved into three domains: DI, DII and DIII. The highest level of polymorphism is seen among residues in DI and DII surrounding a conserved hydrophobic trough [17,18] which plays a vital role in the invasion process [19-21]. Antibodies against this region have been shown to block invasion in a strain-specific manner; however, it is also clear that these antibodies constitute only a small proportion of the total inhibitory humoral response [12,22,23]. *Pf*AMA1 also contains a number of less variable residues across the three domains, and our aim was to identify which of these were also located within epitopes eliciting protective immune responses.

55 We attempted to identify these additional epitopes under strong immune selection by
56 using a novel approach based on the predictions of a multilocus mathematical model
57 of pathogen evolution [24-26]. The central premise is that regions or epitopes under
58 strong immune selection will self-organise into non-overlapping combinations of
59 variants [24-26] and supported by empirical observations within a variety of host-
60 pathogen systems [27-32]. We inverted this result to search for signatures of immune
61 selection among polymorphic loci by identifying those pairs which exhibit a high
62 degree of 'non-overlap' but excluding those that may have arisen as a consequence of
63 shared ancestry or due to structural interactions. Using this "reverse immunodynamic"
64 method, we found strong signatures of immune selection among a small number of
65 dimorphic residues in Domain III of *Pf*AMA-1, suggesting that epitopes containing
66 these residues are targets of protective immunity.
67

68 ## Results

69 **A limited number of site-pairs exhibit high non-overlap**

70 We implemented the reverse immunodynamic method by performing pairwise
71 comparisons of all dimorphic sites among 1198 unique *Pf*AMA-1 sequences obtained
72 from GenBank, to determine whether any of these pairs exhibited non-overlapping
73 associations between amino acid residues. We used an Information Theory statistic
74 known as Mutual Information (MI), scaled to its maximum value, to measure the
75 degree of non-overlap in amino acid combinations between pairs of sites. Most pairs
76 of sites showed low non-overlap, with only 10 pairs exhibiting a scaled MI in excess
77 of 0.3 (Tables 1 & S1). Of these, pairs 404/405, 448/451 and 283/285 were
78 eliminated on grounds of structural proximity, leaving 7 pairs as shown by the bars in
79 Figure 1A.

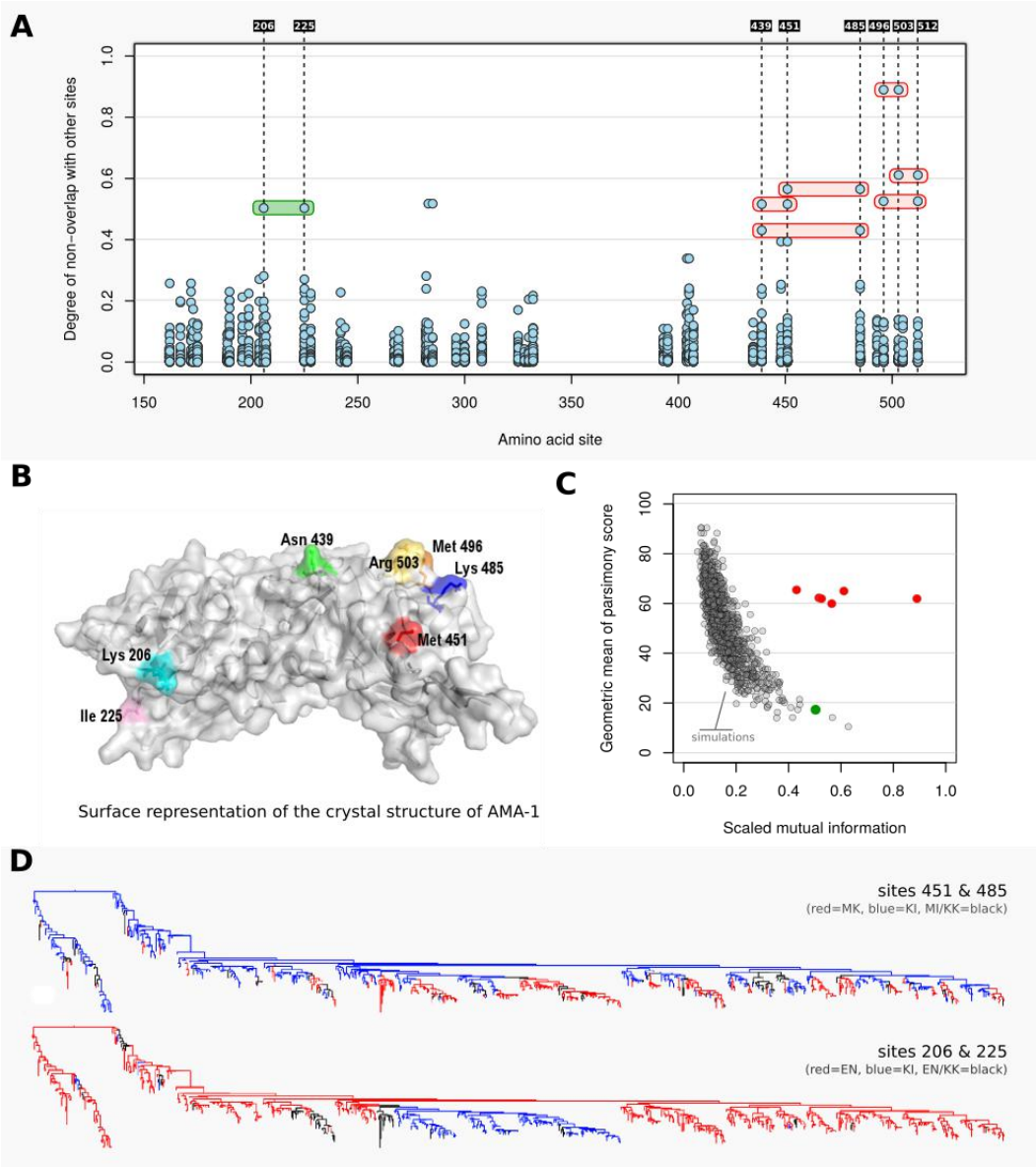
Table 1: Frequencies of amino acid associations between pairs of sites with high non-overlap (scaled MI>0.3)

Association	Amino Acid Combinations and Frequencies				Measure of non-overlap (scaled MI)
496 vs 503	**IN**	**IR**	**MN**	**MR**	0.89
	602	3	14	561	
503 vs 512	**NK**	**NR**	**RK**	**RR**	0.61
	565	52	40	524	
451 vs 485	**KI**	**KK**	**MI**	**MK**	0.57
	604	103	15	476	
496 vs 512	**IK**	**IR**	**MK**	**MR**	0.53
	552	68	54	523	
283 vs 285	**LE**	**LQ**	**SE**	**SQ**	0.52
	242	71	1	883	
439 vs 451	**HK**	**HM**	**NK**	**NM**	0.5
	660	72	46	390	
206 vs 225	**EI**	**EN**	**KI**	**KN**	0.5
	96	806	290	6	
439 vs 485	**HI**	**HK**	**NI**	**NK**	0.41
	588	144	30	406	
448 vs 451	**DK**	**DM**	**NK**	**NM**	0.39
	705	196	2	295	
404 vs 405	**RE**	**RK**	**TE**	**TK**	0.34
	108	635	367	84	

High non-overlap is unlikely to have arisen by neutral evolutionary processes

To assess whether high non-overlap between sites could have arisen by neutral evolutionary processes, we estimated a phylogenetic tree from the sequences and calculated the parsimony score (PS) for each site, which indicates the minimum number of mutations (in this case, back and forth between the two possible amino acids at each site) required to explain the observed ancestral relationships. A new statistic S, the product of MI and the geometric mean of PS for each site, was found to preserve the same order as the scaled MI scores of the 7 sites under consideration, except in the case of the 206/225 (Table S2). To obtain an appropriate null distribution of S for hypothesis testing, we simulated 1000 sets of sequences which were (a) consistent with the empirical phylogenetic tree and (b) had the same PS score at each dimorphic site as the empirical data. The null distribution of the top-scoring measures of S (Figure 1C) shows that neutral evolution can generate sites with high

MI or high PS values, but not both. In contrast, the pairs of sites indicated in red in Figure 1A lie well outside the null distribution of high-scoring site-pairs (Figure 1C). Only site-pair 206/225 (shown in green) did not meet this criterion. Reasons for this pattern become evident upon inspection of the distribution of the amino acid variation on the phylogenetic tree (Figure 1D). The dominant non-overlapping pairs of amino acids at 206/225 are highly segregated on the phylogeny and not improbable under the null hypothesis, while those of 451/485, for example, exhibit a more interdigitated pattern, indicating multiple independent and convergent paired amino acid changes that are exceptionally unlikely in the absence of immune selection.



108 **Figure 1:** (A) Degree of non-overlap in associations between amino acid residues at pairs of dimorphic
109 sites. Each site on the x-axis is compared with all other dimorphic sites: those with reciprocally high
110 scores are shown by bars spanning both sites. Red bars show significant associations. Green bars
111 represent associations that do not achieve significance when phylogenetic relationships are taken into
112 account. (B) Surface representation of the crystal structure of AMA-1 with specific amino acids
113 highlighted: Lysine (Lys) 206 = Cyan; Isoleucine (Ile) 225 = Magenta; Asparagine (Asn) 439 = Green;
114 Methionine (Met) 451 = Red, Lys 485 = Blue, Met 496 = Orange, Arginine (Arg) 503 = Yellow. (C) Pairs
115 of sites shown by red and green bars in Fig1A are shown in relation to a null distribution of site-pairs
116 from 1000 simulations that scored most highly on the product of scaled Mutual Information (measuring
117 non-overlap) and Geometric Mean of Parsimony Scores (indicating the minimum number of mutations
118 required to explain the observed ancestral relationships). (D) Phylogenetic trees of AMA-1
119 with branches coloured on the basis of observed combination of amino acid variants found at tips for
120 sites 451 and 485 (red = MK, blue = KI, grey = MI/KK) and sites 206 and 225 (red = EN, blue = KI, grey
121 = EI/KN) using "Annotate Nodes from Tips" in FigTree. Black branches indicate nodes whose annotation
122 cannot be resolved by the algorithm within FigTree.
123

124 **Analysis of tertiary structure identifies those sites that are most likely to be**
125 **under immune selection**

126 The group of sites which scored most highly on the conservative test (described above)
127 comprised 496, 503 and 512. However, mapping these onto the tertiary structure of
128 AMA-1 (Figure 1B) shows that they are in very close proximity of each other and
129 therefore we were unable to exclude biochemical interactions as the primary cause of
130 this striking association. It is still possible that variation at these sites is held in strong
131 non-overlap through immune selection. Antibodies specific to epitopes containing
132 amino acids I/N/K may cross-react with all other epitopic variants except those
133 containing M/R/R, thereby promoting the co-circulation of these non-overlapping
134 combinations of residues at these sites; however, because of their structural adjacency,
135 it is impossible to privilege this explanation over intrinsic steric advantages to these
136 combinations.
137

138 High non-overlap was also observed among sites 439/451, 439/485 and 451/485
139 which were sufficiently far enough apart in the tertiary structure to represent potential

140 examples of immune selection forcing non-structurally related residues to co-
141 segregate. We cannot, of course, completely eliminate long-range physical
142 interactions as a cause of these associations, and whether it is indeed immune
143 selection that is responsible for these patterns can only be established by functional
144 studies. It is clear that there is no fundamental structural impediment to the existence
145 of intermediate forms since they do occur at low frequencies, but whether they are
146 disadvantaged due to lower intrinsic fitness or immune-mediated selection remains to
147 be resolved by further experimental work.

148

149 **Non-overlapping associations between sites under strong immune selection are**
150 **observed at the level of local populations**

151 We next investigated the associations between positions 439, 451 and 485 on a
152 smaller geographical scale to confirm that these patterns did not arise due to certain
153 combinations of alleles dominating by chance in particular areas. In all countries
154 studied, both of the major sequence variant associations were present, with the
155 secondary variants present at a lower frequency (Table 2).

156

157 **Table 2: Country-specific frequency of amino acid associations between positions 439, 451**
158 **and 485 within unique sequences obtained from The Gambia, Kenya and Mali**
159

Association	Country	Amino Acid Combinations and Frequencies				Scaled MI
439 vs 451		HK	HM	NK	NM	
	Gambia	46	3	0	31	0.701
	Kenya	46	3	9	27	0.402
	Mali	395	25	11	211	0.637
439 vs 485		HI	HK	NI	NK	
	Gambia	39	7	1	30	0.549
	Kenya	45	4	1	35	0.682
	Mali	359	61	17	205	0.453
451 vs 485		KI	KK	MI	MK	
	Gambia	39	4	1	35	0.673

	Kenya	46	9	0	31	0.515
	Mali	366	40	10	244	0.608

A dataset of 660 unique sequences was compiled for Mali and showed high non-overlap for each pair of associations. A smaller dataset comprising 79 sequences from the Gambia also yielded MI values in excess of those obtained from the global data set. Slightly weaker associations were found in 86 unique sequences from Kenya; however these improved significantly when the data were analysed as a collection of isolates rather than sequences: i.e. without excluding identical sequences present in different individuals (Table 3).

Table 3: Frequency of amino acid associations between positions 439, 451 and 485 in a population study (duplicate sequences not excluded) in Kilifi, Kenya

Association	Amino Acid Combinations and Frequencies				Scaled MI
439 vs 451	**HK**	**HM**	**NK**	**NM**	0.564
	61	3	9	55	
439 vs 485	**HI**	**HK**	**NI**	**NK**	0.772
	60	4	1	63	
451 vs 485	**KI**	**KK**	**MI**	**MK**	0.644
	61	9	0	59	

As the model predictions pertain to population level frequencies of allele combinations, rather than their presence in a collection of unique sequences, we believe that Table 3 presents a more valid test (than when only unique sequences are included) of the model's capabilities to pick out regions under strong immune selection; it is reassuring that these patterns are also evident under the more stringent conditions where only unique sequences are included.

182 **Discussion**

183 Our results suggest that Domain III (DIII) of the AMA-1 protein is under strong
184 immune selection, and may play an important role in protective immunity alongside
185 the more polymorphic C1L region of Domain I. At a functional level, Domain III has
186 been shown to bind to the erythrocyte membrane protein Kx [33]. Immune responses
187 that interrupt this process may, therefore, have a significant inhibitory effect on
188 parasite growth but, so far, the evidence that Domain III elicits protective antibody
189 responses is limited. Nair *et al.* [34] have shown that antibodies from a single blood
190 donor affinity purified on DIII are inhibitory. In a study by Polley *et al.* [2], it was
191 found that antibody responses against recombinant full-length ectodomains were more
192 strongly associated with protection than responses to those which did not express DIII,
193 even though naturally acquired antibodies to recombinant DIII alone were rare. A
194 virosomal formulation of loop I of DIII, comprising residues 446-490, has been
195 shown to elicit growth-inhibitory antibodies [35] and volunteers within a vaccine trial
196 containing this AMA-1 peptide exhibited both reduced blood stage growth rates and a
197 higher rate of morphological parasite abnormalities [36]. A recent analysis of allele-
198 specific efficacy of a vaccine against AMA-1 [11] showed that the relative risk ratio
199 (RRR) of re-infection with strains containing vaccine-type amino acids among
200 vaccinees vs controls was high (although not significant) at residues 439, 451 and 485.
201 Residues associated with the polymorphic sites surrounding the hydrophobic trough
202 displayed significant RRR, as might be expected; sites 308 and 332 also had high
203 scores but no evidence of sustained immune selection at these sites was available in
204 our analysis. In the same study, the likelihood of a shift from a vaccine-type to non-
205 vaccine-type residue among sequences obtained following vaccination was found to
206 be significantly increased at position 485 and non-significantly increased at position

451. Significant changes in residue composition were also found at highly polymorphic site 197 but not at any of the other polymorpic sites in the C1L region. Significant change was also observed at dimorphic site 435, but we found no non-overlapping associations in any of the pairwise comparisons between 435 and other sites. Several population genetic studies [37-41] based on site frequency spectrum analysis techniques (such as Tajima's D) strongly indicate that DIII is under balancing selection; our "reverse immunodynamic" method complements and extends these analyses by showing that we can use a measure of non-overlap between site-pairs in addition to standard population genetic measures to finely map regions under immune selection within DIII.

An interesting feature of our analyses is that while particular groups of sites exhibit strong non-overlap, the associations between these groups appear to be random. For example, the triad 439-451-485 is in complete linkage equilibrium with triad 496-503-512. Furthermore, there are no non-overlapping associations present between the highly polymorphic residues in the C1L region of Domain I and the 439-451-485 triad (although the former exhibit strong non-overlap: Table S1). If the analysis we present here is correct, then it would be sufficient to present the dominant "non-overlapping" variants of 439-451-485 (i.e. HKI and NMK) in a manner that induced strong immune responses to the epitopes containing these residues, rather than to try and cover the entire haplotypic diversity of AMA-1 [42]. Immune responses to the epitopes containing residues 439, 451 and 485 appear to develop more slowly under natural conditions than antibodies to the highly polymorphic regions shielding the hydrophobic groove [2], but should be easier to induce by vaccination than immune responses to completely conserved epitopes [43].

Our analysis of AMA-1 further highlights the role of less polymorphic epitopes in population-level immune evasion. At a more general level, our analyses emphasize that targets of natural immunity are not polarised between the completely invariant and the highly polymorphic but also include a range of epitopes of intermediate variability that have, so far, been largely overlooked. Among pathogens where protective immune responses are directed against conserved targets or epitopes, a single infection ensures lifelong immunity and the goal of vaccination is to elicit natural immunity against such conserved targets. Most infectious diseases (such as measles) against which we have effective vaccines fall into this category. By contrast, the primary targets of protective immunity in several important pathogens such as *P. falciparum*, are highly diverse and present a challenge to the development of vaccines.

One solution to this problem that is currently being explored by several research groups is the artificial boosting of immunity towards conserved regions that are not sufficiently naturally immunogenic to prevent further infection upon single exposure but do eventually confer protection after multiple exposures. Another strategy is to develop multi-allelic vaccines in which immunogenic epitopes that confer protection against a sub-population of strains are combined to provide broad coverage. However, many of the known targets of protective immunity are too diverse for this strategy to be successfully implemented. Our work in this paper highlights that between these two extremes, there exist epitopes of intermediate variability which are nonetheless potentially under strong immune selection and thus offer alternative targets of vaccination.

258 # Methods

259

260 **Genetic sequence selection**

261 *Pf*AMA-1 sequences were obtained by searching the NCBI Entrez nucleotide and
262 protein databases. This yielded 2835 protein sequences and 2926 nucleotide
263 sequences (up to and including year 2014). In cases when the GenBank nucleotide
264 record did not include the corresponding protein sequence, these were translated using
265 Biopython, an open source set of tools (biopython.org).

266

267 The analysis was restricted to sequences between 400 and 622 amino acids in length.
268 The lower limit aims to ensure that all the polymorphic regions across the three
269 domains are included; the upper limit corresponds to the entire processed protein, and
270 therefore removes any entries which contain the non-coding sequence. To avoid
271 pseudoduplication, all identical sequences were removed, leaving 1198 unique
272 sequences. These were then aligned using Clustal W [40].

273

274 **Measuring non-overlap among dimorphic sites**

275 A frequency analysis of all residues within the three domains was used to identify any
276 polymorphic sites. Only variants with a frequency exceeding 2.5% were included in
277 the subsequent analysis. This identified 44 dimorphic sites (Figure S1). We searched
278 for signatures of immune selection by analysing the mutual information (MI) scores
279 of all pairs of sites (Table S1).

280

281 **Geographic sub-analysis**

282 We repeated the calculation of MI for all isolates whose country of origin could be
283 identified (either from the Entrez database or with reference to published papers).

284

285 **Phylogenetic analysis**

286 We first truncated our protein alignment to the region where we had >99% coverage
287 [namely, from residue 149 to residue 534]. This alignment was then used to construct
288 a Maximum Likelihood Tree in Mega6 under the JTT model (with also: 4 discrete
289 categories of gamma-distributed rates among sites; the nearest-neighbour-interchange
290 heuristic; and a very strong branch swap filter) and this tree was used to calculate the
291 maximum parsimony score (PS) for each dimorphic site (Figure S2; Table S1). We
292 defined a new statistic S as the product of the MI and the geometric mean of the PS
293 for each site (Table S2). We used Seq-Gen to simulate 1000 sets of protein sequences
294 with the same phylogenetic characteristics, restricting variation at each of the sites to
295 two amino acids. We were able to establish that the pairs of interest (see Table 1) lay
296 significantly outside the distribution of the identically ranked (by S-score) sites from
297 each simulation (Figure S3).

298

299 **Tertiary structures**

300 From the protein data bank (PDB) the sequences and tertiary structures of
301 *Plasmodium vivax* AMA-1 (*Pv*AMA-1, PDB 1W81) and *Pf*AMA-1 domain III (PDB
302 1HN6) were downloaded. As the tertiary structure of 1HN6 was mostly unstructured,
303 it could not be used on its own. Therefore a Fold and Function Assignment System
304 (FFAS) was used to align the amino acid sequences of 1W81 and 1HN6. From this
305 alignment a 'hybrid' structural model with 94% confidence was generated using
306 SCWRL. This resulting model gave a single tertiary structure with the residues within
307 domain III from *Pv*AMA-1 (1W81) substituted by those from *Pf*AMA-1 (1HN6)

using the *P. vivax* structural coordinates. Pymol (http://www.pymol.org/) was used to visualise the 'hybrid' tertiary structure and to highlight the relevant residues.

Supplementary files

Table S1: parsimony score and MI for pairs of sites
Table S2: statistic S (product of MI and the geometric mean of the PS for each site)

Acknowledgements

We thank Dr. CRC Spensley for assistance with data manipulation and programming, and Dr. Pietro Roversi for assistance with generating the structure.

Author Contributions

JL, KJS, PSW, SG designed the study. JL, PSW, AW performed the experiments. All authors wrote and revised the manuscript. JL, KJS, PSW curated the data.

323 # References

324

325 1. Remarque EJ, Faber BW, Kocken CHM & Thomas AW (2008) Apical membrane antigen
326 1: A malaria vaccine candidate in review. *Trends Parasitol* 24(2): 74-84.
327 2. Polley SD, et al (2004) Human antibodies to recombinant protein constructs of
328 plasmodium falciparum apical membrane antigen 1 (AMA1) and their associations with
329 protection from malaria. *Vaccine* 23(5): 718-728.
330 3. Fowkes FJI, Richards JS, Simpson JA & Beeson JG (2010) The relationship between
331 anti-merozoite antibodies and incidence of plasmodium falciparum malaria: A systematic
332 review and meta-analysis. *Plos Medicine* 7(1): e1000218.
333 4. Richards JS, et al (2013) Identification and prioritization of merozoite antigens as targets
334 of protective human immunity to plasmodium falciparum malaria for vaccine and
335 biomarker development. *Journal of Immunology* 191(2): 795-809.
336 5. Mugyenyi CK, et al (2013) Antibodies to polymorphic invasion-inhibitory and non-
337 inhibitory epitopes of plasmodium falciparum apical membrane antigen 1 in human
338 malaria. *Plos One* 8(7): e68304.
339 6. Stanisic DI, et al (2015) Acquisition of antibodies against plasmodium falciparum
340 merozoites and malaria immunity in young children and the influence of age, force of
341 infection, and magnitude of response. *Infect Immun* 83(2): 646-660.
342 7. Osier FH, et al (2014) New antigens for a multicomponent blood-stage malaria vaccine.
343 *Science Translational Medicine* 6(247)
344 8. Vicentin EC, et al (2014) Invasion-inhibitory antibodies elicited by immunization with
345 plasmodium vivax apical membrane antigen-1 expressed in pichia pastoris yeast. *Infect*
346 *Immun* 82(3): 1296-1307.
347 9. Biswas S, et al (2012) Recombinant viral-vectored vaccines expressing plasmodium
348 chabaudi AS apical membrane antigen 1: Mechanisms of vaccine-induced blood-stage
349 protection. *Journal of Immunology* 188(10): 5041-5053.
350 10. Schussek S, et al (2013) Immunization with apical membrane antigen 1 confers sterile
351 infection-blocking immunity against plasmodium sporozoite challenge in a rodent model.
352 *Infect Immun* 81(10): 3586-3599.
353 11. Ouattara A, et al (2013) Molecular basis of allele-specific efficacy of a blood-stage
354 malaria vaccine: Vaccine development implications. *J Infect Dis* 207(3): 511-519.
355 12. Drew DR, et al (2012) Defining the antigenic diversity of plasmodium falciparum apical
356 membrane antigen 1 and the requirements for a multi-allele vaccine against malaria. *Plos*
357 *One* 7(12): e51023.
358 13. Dicko A, et al (2007) Impact of a plasmodium falciparum AMA1 vaccine on antibody
359 responses in adult malians. *Plos One* 2(10): e1045.
360 14. Thera MA, et al (2008) Safety and immunogenicity of an AMA-1 malaria vaccine in
361 malian adults: Results of a phase 1 randomized controlled trial. *Plos One* 3(1): e1465.
362 15. Laurens MB, et al (2013) Extended safety, immunogenicity and efficacy of a blood-stage
363 malaria vaccine in malian children: 24-month follow-up of a randomized, double-blinded
364 phase 2 trial. *Plos One* 8(11): e79323
365 16. Biswas S, et al (2014) Assessment of humoral immune responses to blood-stage malaria
366 antigens following ChAd63-MVA immunization, controlled human malaria infection and
367 natural exposure. *Plos One* 9(9): e107903.
368 17. Bai T, et al (2005) Structure of AMA1 from plasmodium falciparum reveals a clustering
369 of polymorphisms that surround a conserved hydrophobic pocket. *Proc Natl Acad Sci U S*
370 *A* 102(36): 12736-12741.
371 18. Lim SS, et al (2014) Structure and dynamics of apical membrane antigen 1 from
372 plasmodium falciparum FVO. Biochemistry (N Y) 53(46): 7310-7320.

19. Richard D, et al (2010) Interaction between plasmodium falciparum apical membrane antigen 1 and the rhoptry neck protein complex defines a key step in the erythrocyte invasion process of malaria parasites. *J Biol Chem* 285(19): 14815-14822.
20. Lamarque M, et al (2011) The RON2-AMA1 interaction is a critical step in moving junction-dependent invasion by apicomplexan parasites. *Plos Pathogens* 7(2): e1001276.
21. Srinivasan P, et al (2011) Binding of plasmodium merozoite proteins RON2 and AMA1 triggers commitment to invasion. *Proc Natl Acad Sci U S A* 108(32): 13275-13280.
22. Healer J, et al (2004) Allelic polymorphisms in apical membrane antigen-1 are responsible for evasion of antibody-mediated inhibition in plasmodium falciparum. *Mol Microbiol* 52(1): 159-168.
23. Cortes A, et al (2005) Allele specificity of naturally acquired antibody responses against plasmodium falciparum apical membrane antigen 1. *Infect Immun* 73(1): 422-430.
24. Gupta S, et al (1996) The maintenance of strain structure in populations of recombining infectious agents. *Nat Med* 2(4): 437-442.
25. Gupta S, Ferguson N & Anderson R (1998) Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science* 280(5365): 912-915.
26. Lourenço J, Wikramaratna SP, Gupta S (2015) MANTIS: an R package that simulates multilocus models of pathogen evolution. *BMC Bioinformatics*. 16:176.
27. Hinnebusch BJ, Barbour AG, Restrepo BI, Schwan TG. (1998) Population structure of the relapsing fever spirochete *Borrelia hermsii* as indicated by polymorphism of two multigene families that encode immunogenic outer surface lipoproteins. *Infect Immun* 66:432–40.
28. Johnson DR, Kaplan EL, VanGheem A, Facklam RR, Beall B (2006) Characterization of group A streptococci (Streptococcus pyogenes): Correlation of M-protein and emm-gene type with T-protein agglutination pattern and serum opacity factor. J Med Microbiol 55:157e164.
29. Areschoug, T., Stålhammar-Carlemalm, M., Larsson, C., & Lindahl, G. (1999). Group B Streptococcal Surface Proteins as Targets for Protective Antibodies: Identification of Two Novel Proteins in Strains of Serotype V. *Infection and Immunity*, *67*(12), 6350–6357.
30. Buckee CO, Gupta S, Kriz P, Maiden MC & Jolley KA (2010) Long-term evolution of antigen repertoires among carried meningococci. *Proc Biol Sci* 277:1635-41
31. Callaghan MJ, Buckee C, Jolley, KA, Kriz, P, Maiden MCJ , Gupta S. (2008) The Effect of Immune Selection on the Structure of the Meningococcal Opa Protein Repertoire. *PLoS Pathogens* 14;4(3):e1000020.
32. Polley SD, et al. (2007) Plasmodium falciparum merozoite surface protein 3 is a target of allele-specific immunity and alleles are maintained by natural selection. J Infect Dis 195:279e287.
33. Kato K, Mayer DCG, Singh S, Reid M & Miller LH (2005) Domain III of plasmodium falciparum apical membrane antigen 1 binds to the erythrocyte membrane protein kx. *Proc Natl Acad Sci U S A* 102(15): 5552-5557.
34. Nair M, et al (2002) Structure of domain III of the blood-stage malaria vaccine candidate, plasmodium falciparum apical membrane antigen 1 (AMA1). *J Mol Biol* 322(4): 741-753.
35. Mueller MS, et al (2003) Induction of parasite growth-inhibitory antibodies by a virosomal formulation of a peptidomimetic of loop I from domain III of plasmodium falciparum apical membrane antigen 1. *Infect Immun* 71(8): 4749-4758.
36. Thompson FM, et al (2008) Evidence of blood stage efficacy with a virosomal malaria vaccine in a phase IIa clinical trial. *Plos One* 3(1): e1493.
37. Polley SD, Conway DJ (2001) Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. Genetics 158: 1505–1512.
38. Tetteh KK, Stewart LB, Ochola LI, Amambua-Ngwa A, Thomas AW, et al. (2009) Prospective identification of malaria parasite genes under balancing selection. PLoS ONE 4: e5568 doi:10.1371/journal.pone.0005568. doi: 10.1371/journal.pone.0005568

39. Osier FH, Weedall GD, Verra F, Murungi L, Tetteh KK, et al. (2010) Allelic diversity and naturally acquired allele-specific antibody responses to *Plasmodium falciparum* apical membrane antigen 1 in Kenya. Infect Immun 78: 4625–4633. doi: 10.1128/iai.00576-10
40. Amambua-Ngwa A, Tetteh KKA, Manske M, Gomez-Escobar N, Stewart LB, Deerhake ME, et al. (2012) Population Genomic Scan for Candidate Signatures of Balancing Selection to Guide Antigen Characterization in Malaria Parasites. PLoS Genet 8(11): e1002992. doi:10.1371/journal.pgen.1002992
41. Arnott A, et al (2014) Distinct patterns of diversity, population structure and evolution in the AMA1 genes of sympatric plasmodium falciparum and plasmodium vivax populations of Papua New Guinea from an area of similarly high transmission. *Malaria Journal* 13: 233.
42. Duan J, et al (2008) Population structure of the genes encoding the polymorphic plasmodium falciparum apical membrane antigen 1: Implications for vaccine design. *Proc Natl Acad Sci U S A* 105(22): 7857-7862.
43. Dutta S, et al (2013) Overcoming antigenic diversity by enhancing the immunogenicity of conserved epitopes on the malaria vaccine candidate apical membrane antigen-1. *Plos Pathogens* 9(12): e1003840.
44. Larkin MA, et al (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23(21): 2947-2948.