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

José Lourenço[1*], Katrina J Spensley[1*], Paul S Wikramaratna[2*], Bridget S Penman[1], Andrew Walker[1], Adrian L Smith[1], Oliver G Pybus[1], Létitia Jean[3] & Sunetra Gupta[1**]

1. Department of Zoology, University of Oxford, Oxford OX1 3PS
2. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT
3. The Sir William Dunn School of Pathology, South Parks Road, Oxford OX1 3RE

*these authors contributed equally to the manuscript

**Correspondence to: sunetra.gupta@zoo.ox.ac.uk (SG)

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. Mathematical 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.

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.

We attempted to identify these additional epitopes under strong immune selection by using a novel approach based on the predictions of a multilocus mathematical model

of pathogen evolution [24-26]. The central premise is that regions or epitopes under strong immune selection will self-organise into non-overlapping combinations of variants [24-26] and supported by empirical observations within a variety of host-pathogen systems [27-32]. We inverted this result to search for signatures of immune selection among polymorphic loci by identifying those pairs which exhibit a high degree of 'non-overlap' but excluding those that may have arisen as a consequence of shared ancestry or due to structural interactions. Using this "reverse immunodynamic" method, we found strong signatures of immune selection among a small number of dimorphic residues in Domain III of *Pf*AMA-1, suggesting that epitopes containing these residues are targets of protective immunity.

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

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

81 number of mutations (in this case, back and forth between the two possible amino
82 acids at each site) required to explain the observed ancestral relationships. A new
83 statistic S, the product of MI and the geometric mean of PS for each site, was found to
84 preserve the same order as the scaled MI scores of the 7 sites under consideration,
85 except in the case of the 206/225 (Table S2). To obtain an appropriate null
86 distribution of S for hypothesis testing, we simulated 1000 sets of sequences which
87 were (a) consistent with the empirical phylogenetic tree and (b) had the same PS score
88 at each dimorphic site as the empirical data. The null distribution of the top-scoring
89 measures of S (Fig 1C) shows that neutral evolution can generate sites with high MI
90 or high PS values, but not both. In contrast, the pairs of sites indicated in red in Fig
91 1A lie well outside the null distribution of high-scoring site-pairs (Fig 1C). Only site-
92 pair 206/225 (shown in green) did not meet this criterion. Reasons for this pattern
93 become evident upon inspection of the distribution of the amino acid variation on the
94 phylogenetic tree (Fig 2). The dominant non-overlapping pairs of amino acids at
95 206/225 are highly segregated on the phylogeny and not improbable under the null
96 hypothesis, while those of 451/485 exhibit a more interdigitated pattern, indicating
97 multiple independent and convergent paired amino acid changes that are
98 exceptionally unlikely in the absence of immune selection.

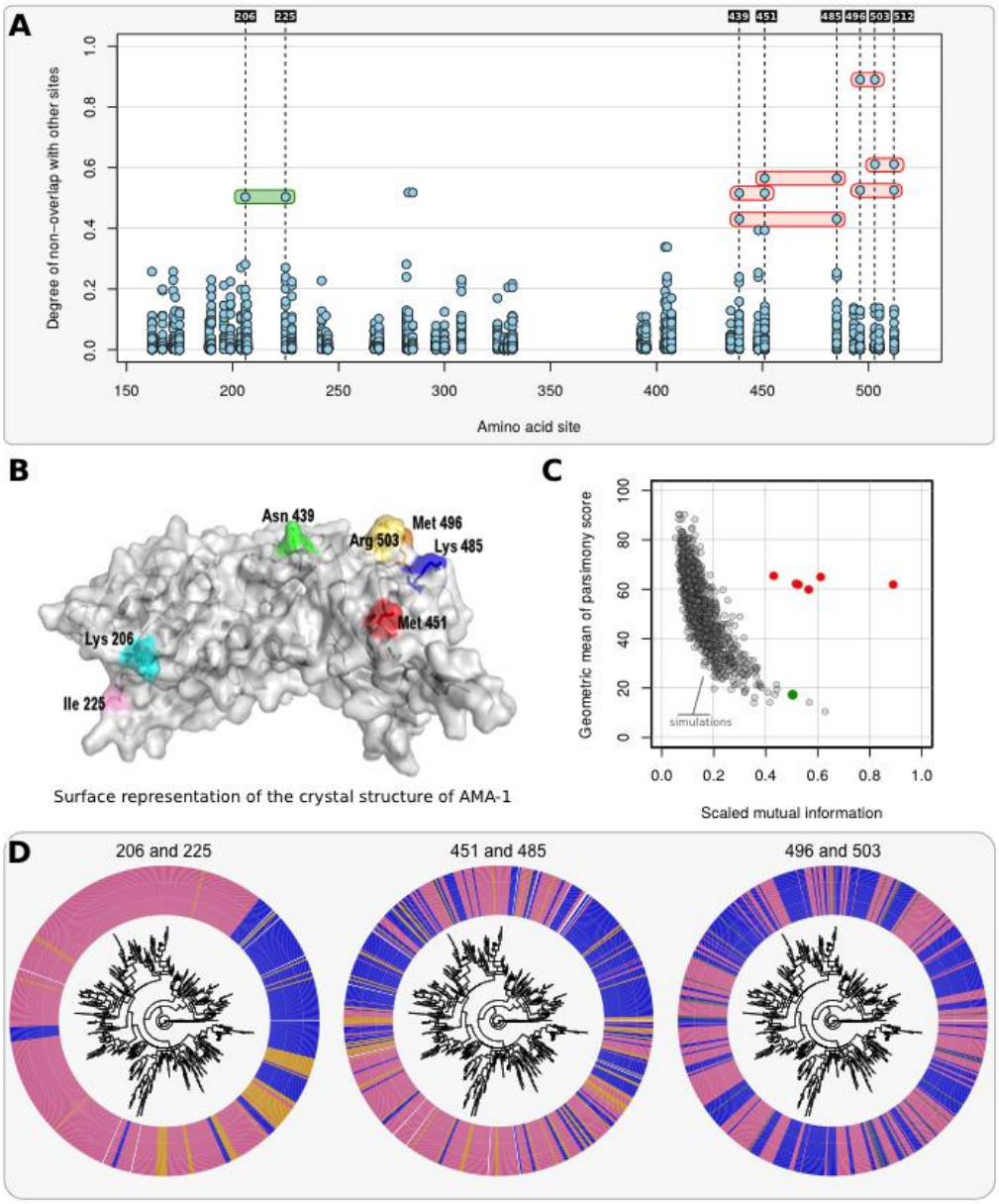

Figure 1: (A) Degree of non-overlap in associations between amino acid residues at pairs of dimorphic sites. Each site on the x-axis is compared with all other dimorphic sites: those with reciprocally high scores are shown by bars spanning both sites. Red bars show significant associations. Green bars represent associations that do not achieve significance when phylogenetic relationships are taken into account (B) Surface representation of the crystal structure of AMA-1 with specific amino acids highlighted [Lysine (Lys) 206 = Cyan; Isoleucine (Ile) 225 = Magenta; Asparagine (Asn) 439 = Green; Methionine (Met) 451 = Red, Lys 485 = Blue, Met 496 = Orange, Arginine (Arg) 503 = Yellow]. (C) Pairs of sites shown by red and green bars in Fig1A are shown in relation to a null distribution of site-pairs from 1000 simulations that scored most highly on the product of scaled Mutual Information (measuring non-overlap) and Geometric Mean of Parsimony Scores (indicating the minimum number of mutations required to explain the observed ancestral relationships). (D) Phylogenetic tree of AMA-1 with outside ring coloured on the basis of observed combination of amino acid variants found at tree tips for (A) sites 206 and 225 (pink = NE, blue = IK, orange = IE), (B) sites 451 and 485 (pink = IK, blue = KM, orange = KK), and (C) sites 496 and 503 (pink = NI, blue = RM, grey = rare variants).

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

125 High non-overlap was also observed among sites 439/451, 439/485 and
126 451/485 which were sufficiently far enough apart in the tertiary structure to represent
127 potential examples of immune selection forcing non-structurally related residues to
128 co-segregate. We cannot, of course, completely eliminate long-range physical
129 interactions as a cause of these associations, and whether it is indeed immune
130 selection that is responsible for these patterns can only be established by functional
131 studies. It is clear that there is no fundamental structural impediment to the existence
132 of intermediate forms since they do occur at low frequencies, but whether they are
133 disadvantaged due to lower intrinsic fitness or immune-mediated selection remains to
134 be resolved by further experimental work.

135

136 We next investigated the associations between positions 439, 451 and 485 on a
137 smaller geographical scale to confirm that these patterns did not arise due to certain
138 combinations of alleles dominating by chance in particular areas. In all countries

139 studied, both of the major sequence variant associations were present, with the
140 secondary variants present at a lower frequency (Table 2).

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

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

154

155 Our results suggest that Domain III (DIII) of the AMA-1 protein is under strong
156 immune selection, and may play an important role in protective immunity alongside
157 the more polymorphic C1L region of Domain I. At a functional level, Domain III has
158 been shown to bind to the erythrocyte membrane protein Kx [33]. Immune responses
159 that interrupt this process may, therefore, have a significant inhibitory effect on
160 parasite growth but, so far, the evidence that Domain III elicits protective antibody
161 responses is limited. Nair *et al.* [34] have shown that antibodies from a single blood
162 donor affinity purified on DIII are inhibitory. In a study by Polley *et al.* [2], it was
163 found that antibody responses against recombinant full-length ectodomains were more

164 strongly associated with protection than responses to those which did not express DIII,
165 even though naturally acquired antibodies to recombinant DIII alone were rare. A
166 virosomal formulation of loop I of DIII, comprising residues 446-490, has been
167 shown to elicit growth-inhibitory antibodies [35] and volunteers within a vaccine trial
168 containing this AMA1 peptide exhibited both reduced blood stage growth rates and a
169 higher rate of morphological parasite abnormalities [36]. A recent analysis of allele-
170 specific efficacy of a vaccine against AMA1 [11] showed that the relative risk ratio
171 (RRR) of re-infection with strains containing vaccine-type amino acids among
172 vaccinees vs controls was high (although not significant) at residues 439, 451 and 485.
173 Residues associated with the polymorphic sites surrounding the hydrophobic trough
174 displayed significant RRR, as might be expected; sites 308 and 332 also had high
175 scores but no evidence of sustained immune selection at these sites was available in
176 our analysis. In the same study, the likelihood of a shift from a vaccine-type to non-
177 vaccine-type residue among sequences obtained following vaccination was found to
178 be significantly increased at position 485 and non-significantly increased at position
179 451. Significant changes in residue composition were also found at highly
180 polymorphic site 197 but not at any of the other polymorpic sites in the C1L region.
181 Significant change was also observed at dimorphic site 435, but we found no non-
182 overlapping associations in any of the pairwise comparisons between 435 and other
183 sites. Several population genetic studies [37-41] based on site frequency spectrum
184 analysis techniques (such as Tajima's D) strongly indicate that DIII is under
185 balancing selection; our "reverse immunodynamic" method complements and extends
186 these analyses by showing that we can use a measure of non-overlap between site-
187 pairs in addition to standard population genetic measures to finely map regions under
188 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 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,

214 the primary targets of protective immunity in several important pathogens such as *P.*
215 *falciparum*, are highly diverse and present a challenge to the development of vaccines.
216 One solution to this problem that is currently being explored by several research
217 groups is the artificial boosting of immunity towards conserved regions that are not
218 sufficiently naturally immunogenic to prevent further infection upon single exposure
219 but do eventually confer protection after multiple exposures. Another strategy is to
220 develop multi-allelic vaccines in which immunogenic epitopes that confer protection
221 against a sub-population of strains are combined to provide broad coverage. However,
222 many of the known targets of protective immunity are too diverse for this strategy to
223 be successfully implemented. Our work in this paper highlights that between these
224 two extremes, there exist epitopes of intermediate variability which are nonetheless
225 potentially under strong immune selection and thus offer alternative targets of
226 vaccination.
227

228 ## Methods

229
230 **SEQUENCE SELECTION**: *Pf*AMA-1 sequences were obtained by searching the
231 NCBI Entrez nucleotide and protein databases. On 04/07/2014 this yielded 2835
232 protein sequences and 2926 nucleotide sequences. In cases when the GenBank
233 nucleotide record did not include the corresponding protein sequence, these were
234 translated using Biopython, an open source set of tools (biopython.org).
235

236 The analysis was restricted to sequences between 400 and 622 amino acids in length.
237 The lower limit aims to ensure that all the polymorphic regions across the three
238 domains are included; the upper limit corresponds to the entire processed protein, and
239 therefore removes any entries which contain the non-coding sequence. To avoid

240 pseudoduplication, all identical sequences were removed, leaving 1198 unique
241 sequences. These were then aligned using Clustal W [40].
242
243 **MEASURING NON-OVERLAP AMONG DIMORPHIC SITES**: A frequency
244 analysis of all residues within the three domains (q.v.) was used to identify any
245 polymorphic sites. Only variants with a frequency exceeding 2.5% were included in
246 the subsequent analysis. This identified 44 dimorphic sites (Fig S1). We searched for
247 signatures of immune selection by analysing the mutual information (MI) scores of all
248 pairs of sites (Table S1).
249
250 **GEOGRAPHIC SUB-ANALYSES**: We repeated the calculation of MI for all
251 isolates whose country of origin could be identified (either from the Entrez database
252 or with reference to published papers).
253
254 **PHYLOGENETIC ANALYSIS:** We first truncated our protein alignment to the
255 region where we had >99% coverage [namely, from residue 149 to residue 534]. This
256 alignment was then used to construct a Maximum Likelihood Tree in Mega6 under
257 the JTT model (with also: 4 discrete categories of gamma-distributed rates among
258 sites; the nearest-neighbour-interchange heuristic; and a very strong branch swap
259 filter) and this tree was used to calculate the maximum parsimony score (PS) for each
260 dimorphic site (Fig S2; Table S1). We defined a new statistic S as the product of the
261 MI and the geometric mean of the PS for each site (Table S2). We used Seq-Gen to
262 simulate 1000 sets of protein sequences with the same phylogenetic characteristics,
263 restricting variation at each of the sites to two amino acids. We were able to establish

264 that the pairs of interest (see Table 1) lay significantly outside the distribution of the
265 identically ranked (by S-score) sites from each simulation (Fig S3).
266
267 **IMAGES:** From the protein data bank (PDB) the sequences and tertiary structures of
268 *Plasmodium vivax* AMA-1 (*Pv*AMA-1, PDB 1W81) and *Pf*AMA-1 domain III (PDB
269 1HN6) were downloaded. As the tertiary structure of 1HN6 was mostly unstructured,
270 it could not be used on its own. Therefore a Fold and Function Assignment System
271 (FFAS) was used to align the amino acid sequences of 1W81 and 1HN6. From this
272 alignment a 'hybrid' structural model with 94% confidence was generated using
273 SCWRL. This resulting model gave a single tertiary structure with the residues within
274 domain III from *Pv*AMA-1 (1W81) substituted by those from *Pf*AMA-1 (1HN6)
275 using the *P. vivax* structural coordinates. Pymol (http://www.pymol.org/) was used to
276 visualise the 'hybrid' tertiary structure and to highlight the relevant residues.
277

278 Supplementary files

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

281 Acknowledgements

282 We thank Dr. CRC Spensley for assistance with data manipulation and programming,
283 and Dr. Pietro Roversi for assistance with generating the structure.
284
285

286 # References

287

288 1. Remarque EJ, Faber BW, Kocken CHM & Thomas AW (2008) Apical membrane antigen
289 1: A malaria vaccine candidate in review. *Trends Parasitol* 24(2): 74-84.
290 2. Polley SD, et al (2004) Human antibodies to recombinant protein constructs of
291 plasmodium falciparum apical membrane antigen 1 (AMA1) and their associations with
292 protection from malaria. *Vaccine* 23(5): 718-728.
293 3. Fowkes FJI, Richards JS, Simpson JA & Beeson JG (2010) The relationship between
294 anti-merozoite antibodies and incidence of plasmodium falciparum malaria: A systematic
295 review and meta-analysis. *Plos Medicine* 7(1): e1000218.
296 4. Richards JS, et al (2013) Identification and prioritization of merozoite antigens as targets
297 of protective human immunity to plasmodium falciparum malaria for vaccine and
298 biomarker development. *Journal of Immunology* 191(2): 795-809.
299 5. Mugyenyi CK, et al (2013) Antibodies to polymorphic invasion-inhibitory and non-
300 inhibitory epitopes of plasmodium falciparum apical membrane antigen 1 in human
301 malaria. *Plos One* 8(7): e68304.
302 6. Stanisic DI, et al (2015) Acquisition of antibodies against plasmodium falciparum
303 merozoites and malaria immunity in young children and the influence of age, force of
304 infection, and magnitude of response. *Infect Immun* 83(2): 646-660.
305 7. Osier FH, et al (2014) New antigens for a multicomponent blood-stage malaria vaccine.
306 *Science Translational Medicine* 6(247)
307 8. Vicentin EC, et al (2014) Invasion-inhibitory antibodies elicited by immunization with
308 plasmodium vivax apical membrane antigen-1 expressed in pichia pastoris yeast. *Infect*
309 *Immun* 82(3): 1296-1307.
310 9. Biswas S, et al (2012) Recombinant viral-vectored vaccines expressing plasmodium
311 chabaudi AS apical membrane antigen 1: Mechanisms of vaccine-induced blood-stage
312 protection. *Journal of Immunology* 188(10): 5041-5053.
313 10. Schussek S, et al (2013) Immunization with apical membrane antigen 1 confers sterile
314 infection-blocking immunity against plasmodium sporozoite challenge in a rodent model.
315 *Infect Immun* 81(10): 3586-3599.
316 11. Ouattara A, et al (2013) Molecular basis of allele-specific efficacy of a blood-stage
317 malaria vaccine: Vaccine development implications. *J Infect Dis* 207(3): 511-519.
318 12. Drew DR, et al (2012) Defining the antigenic diversity of plasmodium falciparum apical
319 membrane antigen 1 and the requirements for a multi-allele vaccine against malaria. *Plos*
320 *One* 7(12): e51023.
321 13. Dicko A, et al (2007) Impact of a plasmodium falciparum AMA1 vaccine on antibody
322 responses in adult malians. *Plos One* 2(10): e1045.
323 14. Thera MA, et al (2008) Safety and immunogenicity of an AMA-1 malaria vaccine in
324 malian adults: Results of a phase 1 randomized controlled trial. *Plos One* 3(1): e1465.
325 15. Laurens MB, et al (2013) Extended safety, immunogenicity and efficacy of a blood-stage
326 malaria vaccine in malian children: 24-month follow-up of a randomized, double-blinded
327 phase 2 trial. *Plos One* 8(11): e79323
328 16. Biswas S, et al (2014) Assessment of humoral immune responses to blood-stage malaria
329 antigens following ChAd63-MVA immunization, controlled human malaria infection and
330 natural exposure. *Plos One* 9(9): e107903.
331 17. Bai T, et al (2005) Structure of AMA1 from plasmodium falciparum reveals a clustering
332 of polymorphisms that surround a conserved hydrophobic pocket. *Proc Natl Acad Sci U S*
333 *A* 102(36): 12736-12741.
334 18. Lim SS, et al (2014) Structure and dynamics of apical membrane antigen 1 from
335 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.