www.sciencetranslationalmedicine.org/cgi/content/full/1/2/2ra5/DC1

# Supplementary Materials for

## Extreme Polymorphism in a Vaccine Antigen and Risk of Clinical Malaria: Implications for Vaccine Development

Shannon L. Takala, Drissa Coulibaly, Mahamadou A. Thera, Adrian H. Batchelor, Michael P. Cummings, Ananias A. Escalante, Amed Ouattara, Karim Traoré, Amadou Niangaly, Abdoulaye A. Djimdé, Ogobara K. Doumbo, Christopher V. Plowe*

*To whom correspondence should be addressed. E-mail: cplowe@medicine.umaryland.edu

Published 14 October 2009, *Sci. Transl. Med.* **2**, 2ra5 (2009)
DOI: 10.1126/scitranslmed.3000257

**This PDF file includes:**



**Other Supplementary Material for this manuscript includes the following:**
(available at www.sciencetranslationalmedicine.org/cgi/content/full/1/2/2ra5/DC1)

Fig. S1 (.txt format)

## Results and Discussion

### Application of population structure algorithms to infer AMA-1 haplotype groups

Clustering algorithms implemented in STRUCTURE and Structurama were used to categorize the Malian AMA-1 sequences into groups of similar AMA-1 haplotypes. Based on 10 replicate Structurama runs, the 506 AMA-1 sequences from single- or predominant-clone infections (comprising 214 haplotypes) were sorted into 16 haplotype groups (range, 15 to 18 groups), with 12.6% of infections belonging to the group that includes the 3D7 vaccine strain of *P. falciparum* (fig. S3). This number of haplotype groups is more than twice that observed in a previous study (*S1*) that examined fewer sequences (107 haplotypes, including some from a different vaccine testing site in Mali), suggesting that the number of groups might continue to increase if yet more AMA-1 sequences were to be examined.

To address this question, Structurama was used to group 824 AMA-1 ectodomain sequences, including sequences from Bandiagara and published sequences from GenBank, into haplotype groups. Based on 10 replicates, the 824 ectodomain sequences (335 haplotypes) were sorted into 17 haplotype groups (range, 14 to 19 groups), with members of the haplotype group that includes the 3D7 strain of *P. falciparum* having a prevalence of 20.1% and that corresponding to the FVO strain having a prevalence of 3.0%. All 17 haplotype groups were observed in Africa ($n$ = 658), 10 in Asia ($n$ = 102), and 6 in South America ($n$ = 58), and the distribution of haplotype groups differed according to geography (fig. S4).

As was done with changes within AMA-1 domains and clusters, logistic regression was used to estimate whether a change in AMA-1 haplotype group was also associated with the development of clinical symptoms in individuals' consecutive infections. Change in group was weakly associated with the development of symptoms (OR, 3.61; 95% CI, 0.90-14.5; $P$ = 0.070). The biological significance of these AMA-1 haplotype groups will be further examined in a phase 2 trial of an AMA-1 vaccine that was recently completed at the Bandiagara site.

It is a matter of debate whether it is valid to apply these algorithms, which were designed to infer population structure with data from multiple, unlinked, neutral loci to group sequences containing potentially linked polymorphisms in a single gene under balancing selection. It is possible that violation of the assumptions of the model could lead to overestimation of the number of haplotype groups and/or underestimation of the degree of uncertainty in ancestry estimates and will depend on the extent of recombination across the gene. Evidence for intragenic recombination within the *ama-1* gene is mixed with some studies finding evidence for recombination (*S2*) and others finding no such evidence. Evaluation of the Bandiagara AMA-1 sequences with other clustering methods, such as

network trees and principal components analysis, did not show any distinct clustering, which suggests that overestimation of the number of groups and the confidence in those groups may be a problem.

# Supplementary Material

## Supplementary Figures

**Fig. S1**. **Apical membrane antigen-1 (AMA-1) haplotypes from Bandiagara, Mali.**
See the separate supplemental material file “figS1.txt”.

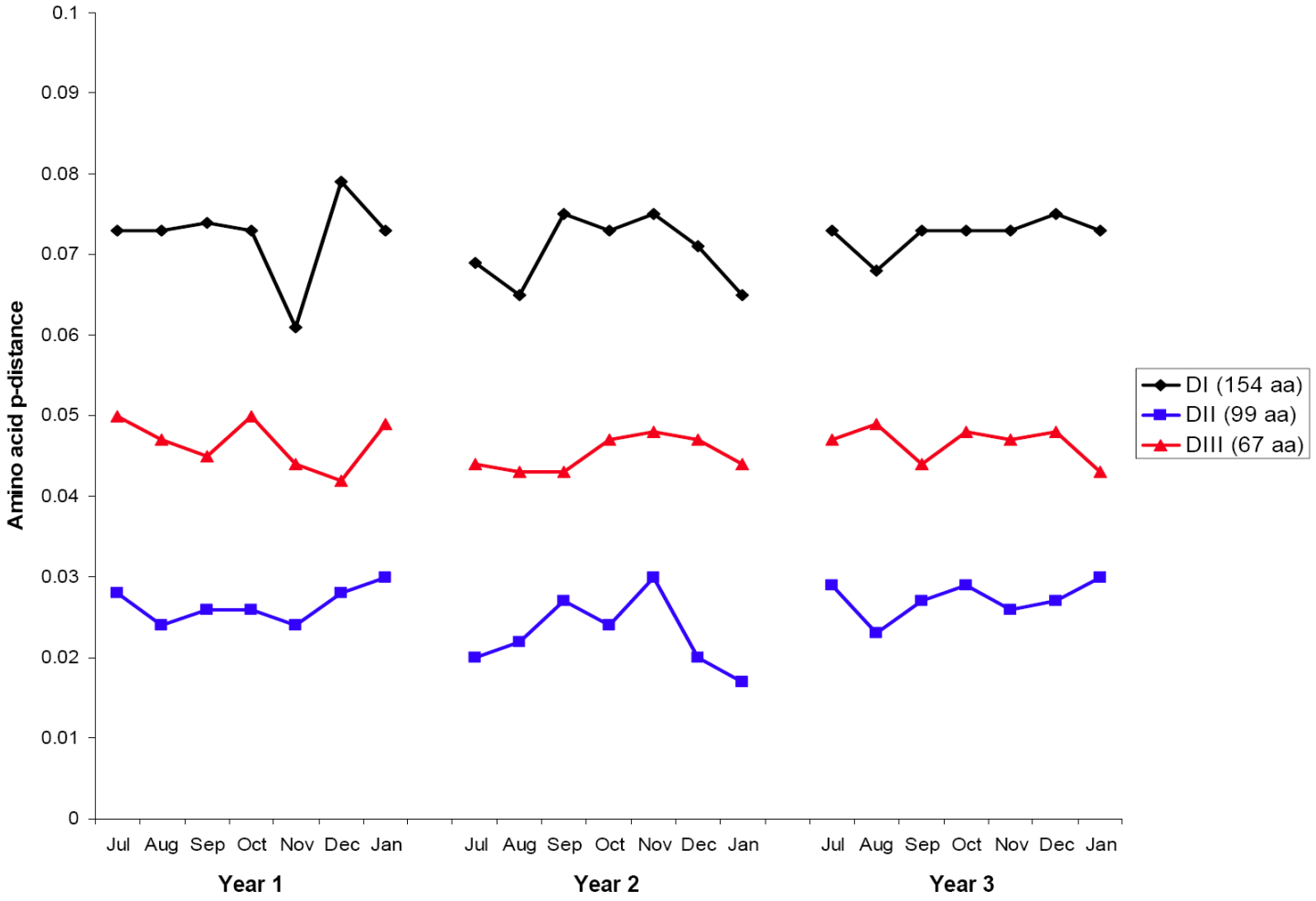


**Fig. S2. Average apical membrane antigen-1 (AMA-1) diversity over time.** Plot of the average amino acid p-distance by domain for each year of the malaria incidence study. The average amino acid p-distance is the average percent difference in amino sequence between any two randomly selected sequences from the dataset. P-distances were calculated using MEGA 4 (*S3*). As observed in the figure, domain I is the most diverse of the three major domains of AMA-1. Average diversity remained stable over the three years of the study.

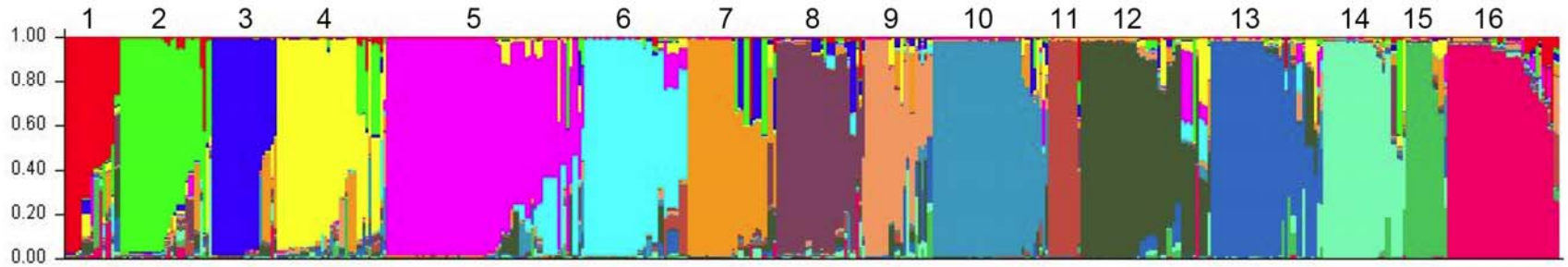


**Fig. S3. Apical membrane antigen-1 (AMA-1) haplotype groups as determined by population structure algorithm.** Sixteen haplotype groups were inferred from 506 AMA-1 sequences from Bandiagara, Mali. The graph was produced as part of the output of STRUCTURE 2.2. Each AMA-1 sequence is represented by a thin vertical line, where each color represents a different haplotype group proportional to the membership of that sequence in the group. Group 5 corresponds to the 3D7 vaccine strain of *P. falciparum.*

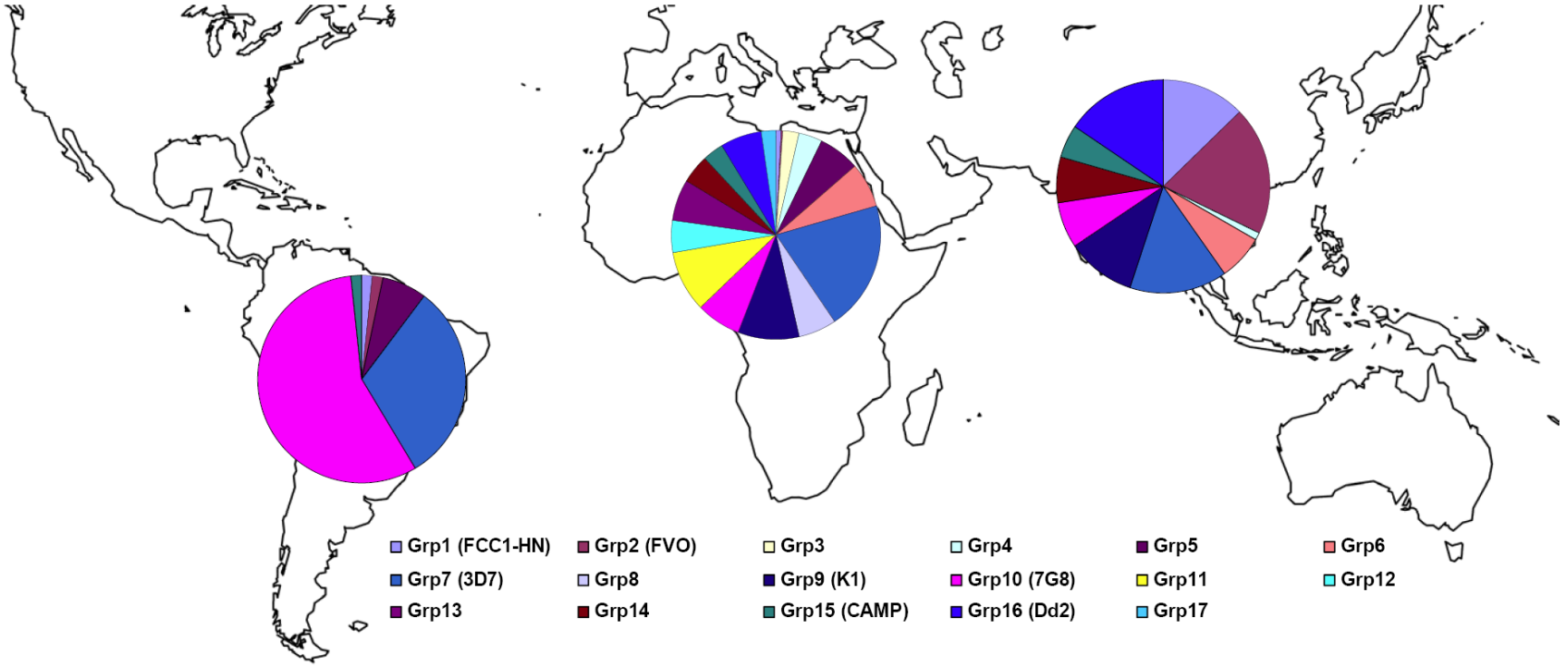


**Fig. S4. Global distribution of apical membrane antigen-1 (AMA-1) subpopulations as determined by population structure algorithm.** A clustering algorithm (as implemented in Structurama) grouped 824 complete AMA-1 ectodomain sequences (including published sequences from GenBank) into 17 subpopulations. The prevalence of the 17 subpopulations in Africa (n=658), Asia (n=102), and South America (n=58) are shown. Subpopulations corresponding to specific strains of *P. falciparum* are indicated in parentheses. All 17 subpopulations were observed in Africa, 10 were observed in Asia, and 6 were observed in South America.

**Table S1. Associations between changes in clusters of polymorphic AMA-1 residues and clinical malaria.** Odds ratios (OR), 95% confidence intervals (CI), and P values comparing the odds of an individual's next consecutive infection being symptomatic to the odds of their next infection being asymptomatic during intervals where a medium or high proportion of amino acid changes at polymorphic sites occurred, with the lowest proportion of amino acid changes as the reference (medium vs low and high vs low, respectively). The same information is also shown comparing intervals with a high proportion of amino acid changes to those with a medium proportion of changes (high vs medium). Categories of low, medium, and high proportions of change were based on cutoff points at the first and third quartiles. Estimated effects were adjusted for age and repeated measurements from the same individual, and are shown stratified by time between consecutive infections to account for a significant interaction between this variable and the amount of genetic change.

| | | | Intervals ≤ 6 weeks (n=133) | | | Intervals ≥ 6 weeks (n=82) | | |
|---|---|---|---|---|---|---|---|---|
| **Region of protein** | **Proportion of change in region** | **Reference** | **OR** | **95% CI** | **P value** | **OR** | **95% CI** | **P value** |
| **Entire ectodomain** | Medium | Low | 2.65 | 1.20 - 5.85 | 0.0158 | 0.20 | 0.04 - 0.98 | 0.048 |
| | High | Low | 6.80 | 2.62 - 17.7 | <0.0001 | 0.32 | 0.06 - 1.74 | 0.189 |
| | High | Medium | 2.57 | 1.00- -6.58 | 0.0501 | 1.61 | 0.56 - 4.61 | 0.377 |
| **Domain I** | Medium | Low | 3.11 | 1.32 - 7.29 | 0.009 | 0.68 | 0.15 - 3.16 | 0.63 |
| | High | Low | 3.32 | 1.22 - 9.03 | 0.019 | 0.48 | 0.10 - 2.20 | 0.35 |
| | High | Medium | 1.07 | 0.37 - 3.05 | 0.901 | 0.70 | 0.23 - 2.13 | 0.53 |
| **Domain I-c1** | Medium | Low | 2.91 | 1.45 - 5.83 | 0.0027 | 0.72 | 0.19 - 2.74 | 0.63 |
| | High | Low | 6.46 | 2.68 - 15.6 | <0.0001 | 1.18 | 0.30 - 4.69 | 0.81 |
| | High | Medium | 2.22 | 0.94 - 5.25 | 0.0685 | 1.63 | 0.55 - 4.88 | 0.38 |
| **Domain I-c1L** | Medium | Low | 2.48 | 1.15 - 5.33 | 0.0204 | 1.56 | 0.44 - 5.59 | 0.49 |
| | High | Low | 5.98 | 2.58 - 13.9 | <0.0001 | 1.04 | 0.33 - 3.32 | 0.95 |
| | High | Medium | 2.41 | 1.01 - 5.79 | 0.0484 | 0.67 | 0.20 - 2.23 | 0.51 |
| **Domain I-c2** | Medium | Low | 1.84 | 0.79 - 4.32 | 0.16 | 0.44 | 0.11 - 1.80 | 0.25 |
| | High | Low | 1.24 | 0.53 - 2.87 | 0.62 | 0.68 | 0.17 - 2.71 | 0.58 |
| | High | Medium | 0.67 | 0.28 - 1.60 | 0.37 | 1.53 | 0.52 - 4.50 | 0.44 |
| **Domain I-c3** | Medium | Low | 2.29 | 0.88 - 5.96 | 0.089 | 0.28 | 0.06 - 1.22 | 0.090 |
| | High | Low | 2.64 | 1.02 - 6.80 | 0.045 | 0.56 | 0.13 - 2.49 | 0.449 |
| | High | Medium | 1.15 | 0.47 - 2.82 | 0.757 | 2.01 | 0.69 - 5.84 | 0.198 |
| **Domain II** | Medium | Low | 3.15 | 1.21 - 8.16 | 0.019 | 0.97 | 0.31 - 3.02 | 0.96 |
| | High | Low | 4.50 | 1.40 - 14.4 | 0.012 | 0.67 | 0.17 - 2.60 | 0.57 |
| | High | Medium | 1.43 | 0.58 - 3.53 | 0.438 | 0.69 | 0.16 - 3.01 | 0.62 |
| **Domain III** | Medium | Low | 1.14 | 0.42 - 3.06 | 0.802 | 0.43 | 0.12 - 1.56 | 0.20 |
| | High | Low | 3.06 | 1.32 - 7.13 | 0.009 | 0.68 | 0.16 - 2.94 | 0.61 |
| | High | Medium | 2.70 | 1.14 - 6.40 | 0.024 | 1.60 | 0.45 - 5.66 | 0.46 |

c1, cluster 1; c1L, cluster 1 loop; c2, cluster 2; c3, cluster 3; CI, confidence interval; OR, odds ratio

**Supplementary References:**

S1. J. Duan, J.Mu, M. A. Thera, D. Joy, S. L. Kosakovsky Pond, D. Diemert, C. Long, H. Zhou, K. Miura, A. Ouattara, A. Dolo, O. Doumbo, X. Z. Su, L. Miller, 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**, 7857–7862 (2008).

S2. D. P. Eisen, V. M. Marshall, H. Billman-Jacobe, R. L. Coppel, A *Plasmodium falciparum* apical membrane antigen-1 (AMA-1) gene apparently generated by intragenic recombination. *Mol. Biochem. Parasitol.* **100**, 243–246 (1999).

S3. K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).