

**No evidence for positive selection at two potential targets for malaria transmission-blocking vaccines in *Anopheles gambiae* s.s.**

Jacob E. Crawford, Susan M. Rottschaefer, Boubacar Coulibaly, Madjou Sacko, Oumou Niare, Michelle M. Riehle, Sekou F. Traore, Kenneth D Vernick, Brian P. Lazzaro

► **To cite this version:**

Jacob E. Crawford, Susan M. Rottschaefer, Boubacar Coulibaly, Madjou Sacko, Oumou Niare, et al.. No evidence for positive selection at two potential targets for malaria transmission-blocking vaccines in *Anopheles gambiae* s.s.. *Infection, Genetics and Evolution*, Elsevier, 2013, 16, pp.87–92. 10.1016/j.meegid.2013.01.006 . pasteur-02008327

**HAL Id: pasteur-02008327**

**<https://hal-pasteur.archives-ouvertes.fr/pasteur-02008327>**

Submitted on 5 Feb 2019

**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.



Published in final edited form as:

*Infect Genet Evol.* 2013 June ; 16: 87–92. doi:10.1016/j.meegid.2013.01.006.

## No evidence for positive selection at two potential targets for malaria transmission-blocking vaccines in *Anopheles gambiae*

S.S

Jacob E. Crawford<sup>a,\*</sup>, Susan M. Rottschaefer<sup>a</sup>, Boubacar Coulibaly<sup>b</sup>, Madjou Sacko<sup>b</sup>, Oumou Niaré<sup>b</sup>, Michelle M. Riehle<sup>c</sup>, Sékou F. Traore<sup>b</sup>, Kenneth D. Vernick<sup>c,d</sup>, and Brian P. Lazzaro<sup>a</sup>

<sup>a</sup>Department of Entomology, Cornell University, Ithaca, NY, USA

<sup>b</sup>Malaria Research and Training Center, University of Bamako, Bamako, Mali

<sup>c</sup>Department of Microbiology, University of Minnesota, St. Paul, MN, USA

<sup>d</sup>Unit for Genetics and Genomics of Insect Vectors, Institut Pasteur, Paris, France

### Abstract

Human malaria causes nearly a million deaths in sub-Saharan Africa each year. The evolution of drug resistance in the parasite and insecticide resistance in the mosquito vector has complicated control measures and made the need for new control strategies more urgent. *Anopheles gambiae* s.s. is one of the primary vectors of human malaria in Africa, and parasite-transmission-blocking vaccines targeting *Anopheles* proteins have been proposed as a possible strategy to control the spread of the disease. However, the success of these hypothetical technologies would depend on the successful ability to broadly target mosquito populations that may be genetically heterogeneous. Understanding the evolutionary pressures shaping genetic variation among candidate target molecules offers a first step towards evaluating the prospects of successfully deploying such technologies. We studied the population genetics of genes encoding two candidate target proteins, the salivary gland protein saglin and the basal lamina structural protein laminin, in wild populations of the M and S molecular forms of *A. gambiae* in Mali. Through analysis of intraspecific genetic variation and interspecific comparisons, we found no evidence of positive natural selection at the genes encoding these proteins. On the contrary, we found evidence for particularly strong purifying selection at the laminin gene. These results provide insight into the patterns of genetic diversity of saglin and laminin, and we discuss these findings in relation to the potential development of these molecules as vaccine targets.

### Keywords

Malaria; *Anopheles gambiae*; Transmission-blocking vaccine; *Plasmodium falciparum*; Natural selection; Population genetics

## 1. Introduction

Mosquito-stage, transmission blocking vaccines have been proposed as an alternative novel technology for blocking transmission of human malaria parasites (Carter and Chen, 1976;

Gwadz, 1976; Barreau et al., 1995; Brennan et al., 2000). The logic of this technology is that mosquito proteins essential for parasite development could be targeted, for example by human derived antibodies, and blocked such that transmission is halted within the mosquito. In the case of mosquito midgut receptors exploited by the ookinete stage of the parasite, anti-receptor anti-bodies delivered in the bloodmeal could directly compete with and inhibit parasite development, while anti-salivary gland receptor anti-bodies will have to be delivered to the mosquito through bloodmeals subsequent to the initial exposure (Brennan et al., 2000; Lavazec and Bourgouin, 2008). At the heart of this hypothetical technology is the assumption that the vaccine target (i.e. mosquito proteins) can be easily identified and blocked by antibodies, but genetic variation segregating in natural populations could result in a heterogeneous molecule population that is difficult to target. Such genetic variation could be neutral with respect to natural selection or it could adaptively evolving, including through selection to resist parasite establishment. Substantial traction has been made in identifying potential target molecules within the mosquito, but most of that work has relied on genetically inbred lab strains of both mosquito and the parasite (Brennan et al., 2000; Arrighi et al., 2005; Saul, 2007; Dinglasan and Jacobs-Lorena, 2008), providing little information on potential genetic heterogeneity and its consequences in nature. Understanding the evolutionary pressures shaping genetic variation among candidate target molecules offers a first step towards evaluating the prospects of successfully deploying such technologies.

The suggestions that pathogen-related selection pressures are likely to drive host evolution date back 100 years (Biffen, 1905; Haldane, 1949; Lederberg, 1999). As one of the most permissive and common vectors of the human malaria parasite *Plasmodium falciparum*, *Anopheles gambiae sensu stricto* is a good candidate for experiencing such selection. Several non-immune mosquito proteins directly interact with the developing malaria parasite and are in some cases required for successful parasite tissue invasion and development. One of the first *Anopheles* proteins shown to interact directly with both ookinete and oocyst stages of *Plasmodium* parasites is laminin, a component of basal laminae in the mosquito, including the one surrounding the midgut (Adini and Warburg, 1999; Vlachou et al., 2001; Arrighi and Hurd, 2002; Dessens et al., 2003). Subsequent studies of laminin suggest that this protein may act as a trigger for the transition from ookinete to oocyst development, or even as a protective coating that masks the parasite from immune detection (Arrighi et al., 2005; Warburg et al., 2007). Further evidence for the intimate and perhaps protective role of laminin was provided by the observations that laminin becomes localized within the oocysts and sporozoites, and it is incorporated into the oocyst capsule (Nacer et al., 2008).

A second host protein, saglin, plays a crucial role in parasite localization and invasion of the salivary glands through a receptor-ligand interaction with the *Plasmodium* TRAP protein (Brennan et al., 2000; Korochkina et al., 2006; Okulate et al., 2007; Ghosh et al., 2009). *P. falciparum* invasion of the salivary gland is inhibited when saglin is blocked with either antibody interference or receptor saturation by SM1, a short peptide whose physical conformation resembles TRAP (Brennan et al., 2000; Ghosh et al., 2001, 2009). Moreover, point mutations in TRAP completely abrogate gland invasion (Matuschewski et al., 2002). Population genetic analysis of *Plasmodium falciparum* and *Plasmodium vivax* suggests adaptive maintenance of variation, especially in the A-domain that binds to saglin (Weedall et al., 2007; Barry et al., 2009).

Based on the evidence that laminin and saglin mediate *Plasmodium* infection in *Anopheles*, we hypothesized that these proteins may be under pathogen-related selection pressure. It should be noted that, while we hypothesize that *Plasmodium* may be a selectively driving force, it is in practice difficult to identify the proximal agent of selection. Nonetheless, the potential effects of selection on potential vaccine targets has important implications

regardless of the agent of selection. To elucidate the selective history of the *saglin* and *laminin* genes, we sequenced alleles sampled from wild populations of the two incipient species of *A. gambiae*, the M and S molecular forms from Mali. We analyzed patterns of intraspecific polymorphism and divergence at these loci but found no significant evidence for adaptive evolution at these loci in either population.

## 2. Materials and methods

### 2.1. Mosquito samples

*Anopheles gambiae* individuals were collected inside dwellings from the villages of Bancoumana and N'gabakoro Droit outside the Malian capital city, Bamako (12°39'N 8°0'W), and an additional collection was drawn from Toumani-Oulena, Mali (10°83'N 7°81'W). In Mali and much of West Africa, *A. gambiae* is largely composed of two populations that are defined by fixed differences in rDNA on the X chromosome (Lehmann and Diabate, 2008). The M/S molecular form of each individual mosquito was determined using the PCR diagnostic developed by Favia et al. (2001). Of the mosquitoes sampled from Bancoumana, four were M form and 11 were S form. All mosquitoes sampled from N'gabakoro Droit were M form ( $n = 10$ ), and all Toumani-Oulena individuals were S form ( $n = 7$ ). *Anopheles merus* DNA from mosquitoes of the OPHAN-SI colony was obtained from MR4.

### 2.2. DNA extraction, PCR and sequencing

DNA was extracted from the mosquitoes using DNeasy kits (Qiagen) under slight modifications to the manufacturers' suggested protocols. PCR primers were designed based on the published *A. gambiae* genome sequence (Holt et al., 2002). Each gene was amplified from genomic DNA using iProof high fidelity DNA Polymerase (BioRad). PCR products were run out on a 1% agarose gel and the product fragments were excised and purified using the PureLink gel extraction kit (Invitrogen). Adenosine tails were added to the purified products by incubating for 20 min at 72 °C with PCR buffer, dATP and Taq polymerase. Products were then cloned using the TOPO XL cloning kit (Invitrogen). Colonies to be sequenced were grown overnight at 37 °C in liquid Luria-Bertani broth supplemented with 20 mg/ml kanamycin, and the plasmids were isolated using the Qiaprep spin miniprep kit (Qiagen). The products were then sequenced directly from the plasmids using the BigDye Terminator Cycle Sequencing Kit v3.1(ABI). The sequences were assembled using Sequencher (Gene Codes Corp.) and CodonCode Aligner (CodonCode Corp.). Only one of the two alleles at each gene was sequenced from any given mosquito in the study. All sequences have been deposited into Genbank under accession numbers KC438305–KC438359.

To control for sequencing error, all singleton polymorphisms were verified by re-amplification and direct sequencing of heterozygous PCR products. The entire gene was amplified from genomic DNA using iProof high fidelity DNA Polymerase (BioRad) and this full-length amplicon was then used as template in a secondary PCR using internally nested primers to robustly amplify the gene region containing the singleton to be validated. Unincorporated primers and dNTPs were inactivated from these secondary amplification products by incubation with ExoI and SAP (both manufactured by USB), and amplification products were then sequenced using the BigDye Terminator Cycle Sequencing Kit v3.1(ABI). To further avoid errors stemming from homopolymers, we deleted all homopolymer sequences from the alignment.

### 2.3. Loci analyzed

We analyzed the *SAG* locus (AGAP000610), which is X-linked, and *LANB2* (AGAP007629), which is found on the distal tip of chromosome 2L. Both loci were surveyed in the Bancoumana and N'gabakoro Droit populations, but only *SAG* sequences were generated from the Toumani-Oulena population. The original predicted exon structure of *SAG* was reannotated in a subsequent genome release as reflected in Ensemble, removing a predicted intron leaving a single coding sequence (Brennan et al., 2000). However, to be certain, we obtained cDNA from live *A. gambiae* females and directly sequenced *SAG* transcripts to identify prospective intron boundaries. We used the same PCR, cloning, and sequencing conditions and reagents described above, using first-strand cDNA from whole mosquitoes as template in the PCR reaction. After aligning the cDNA sequence to the Agam PEST reference sequence, we identified a single 175 basepair (bp) intron beginning at position 1007 of the coding sequence, consistent with the original Ensemble annotation prior to the re-annotation based on short peptide mapping (Brennan et al., 2000). According to our sequencing results, the final saglin protein is predicted to be 374 amino acids in length, and for the analyses here, we assumed that the putative 175 bp intron is non-coding and analyzed the sequence accordingly. For *LANB2*, we analyzed the genomic region according to the exon structure annotated in Vectorbase ([www.vectorbase.org](http://www.vectorbase.org)).

### 2.4. Population genetic analysis

Measures of nucleotide diversity estimated from the average number of differences between haplotypes ( $\pi$ ) and the number of polymorphic sites ( $\theta_W$ ) were calculated on synonymous and nonsynonymous sites alone as well as on all sites combined using DnaSP version 5 (Librado and Rozas, 2009). Three neutrality tests that emphasize different features of the data including Tajima's *D* (Tajima, 1989), a normalized version of Fay and Wu's *H* (Fay and Wu, 2000; Zeng et al., 2006), and Ewens-Watterson's haplotype homozygosity statistic (*EW*) were calculated using a program provided by K. Zeng. Tajima's *D* detects significant excesses of rare SNPs or intermediate frequency SNPs, normalized *H* detects excesses of high-frequency derived SNPs, and *EW* detects excess homozygosity between haplotypes, all of which are deviations from the neutral equilibrium model expected under a hitch-hiking model of positive selection. Three compound statistics (*DH*, *HEW*, *DHEW*) that combine the *p*-values of these neutrality tests were also calculated. These compound statistics were developed to take advantage of each of the features from the different neutrality tests and are more robust to confounding factors such as demography and background selection (Zeng et al., 2006, 2007a, b). The statistical significance of the neutrality tests and compound statistics was evaluated through comparisons to  $10^5$  neutral coalescent simulations without recombination conditioned on the sample size and  $\theta_W$  estimated from the data conducted using the program from K. Zeng. The assumption of no recombination is conservative for tests based on the site frequency spectrum and has little impact on the compound test statistics (Zeng et al., 2007a).

We also used tests to identify patterns of potentially adaptive rates of evolution at nonsynonymous sites as would be expected under a model of repeated episodes of historical positive selection at these genes. First, we applied the McDonald-Kreitman test (McDonald and Kreitman, 1991) within DnaSP using *A. merus* as an outgroup and evaluating significance using Fisher's exact test. For *SAG*, we applied this test first to *SAG* alone then to all three genes in the sequenced region (*SGI-2*, *SAG*, *gSG1a*). We also calculated the ratio of evolutionary rates at nonsynonymous to synonymous sites ( $K_a/K_s$ ) within DnaSP using *A. merus* for estimates of divergence. This ratio is expected to equal one under a completely neutral model and to be greater than one when positive selection has caused repeated adaptive evolution. However, purifying selection often acts to conserve amino acid state such that the rate of substitutions at nonsynonymous sites tends to be much lower than

that of synonymous sites. As such,  $K_d/K_s$  is often less than one when considering an entire gene or gene region due to the widespread effects of purifying selection at most positions. Thus, the  $K_d/K_s$  test is conservative. To circumvent the obscuring effects of purifying selection, we used a sliding window approach to localize any potential cluster of rapidly evolving sites. For both datasets, we used a window size of 1200 basepairs (bp) with a step size of 500 bp.

### 3. Results

#### 3.1. Saglin

To test the hypothesis that saglin (*SAG*) has experienced recent positive selection, we sequenced alleles of this gene from the M and S molecular forms of *A. gambiae*. We sequenced a 3.7 kilobase (kb) region that includes the *SAG* gene as well as 1.6 kb of sequence upstream and 0.8 kb of sequence downstream of the *SAG* coding region, and used population genetic analyses to probe patterns of inter- and intra-species genetic variation. Saglin is one of several members of the SG1 protein family of salivary gland proteins arrayed on the X chromosome (Arcá et al., 1999, 2005; Lanfrancotti et al., 2002), and in addition to *SAG*, we captured the complete coding sequence of *SGI-2* (upstream of *SAG*) and partial coding sequence of the downstream paralog *gSGIa* within our 3.7 kb sequenced region. Across the entire region, nucleotide diversity is low in both populations, consistent with the expectations of low diversity on the X chromosome in *A. gambiae* (Cohuet et al., 2008). Analysis of each paralog and the intergenic regions separately revealed that diversity is approximately equal across the sequence except at *SAG* and the upstream intergenic region (data not shown). Nonsynonymous diversity is particularly rare at the *SAG* gene in these populations. The ratio of nonsynonymous to synonymous nucleotide diversity ( $\pi_d/\pi_s$ ) is 0.059 in the M form population and 0.096 in the S form population. In contrast, this ratio equals approximately 0.5 in both populations for *SGI-2*, the paralog only 300 bp upstream. If we assume that the mutation rate is not especially low at *SAG* since diversity at synonymous sites is comparable across the region, the relatively low  $\pi_d/\pi_s$  at this locus implies particularly strong effects of purifying selection acting on this protein.

The process of genetic hitchhiking that affects genetic variation linked to a positively selected site in a population modifies linked nucleotides in a number of hallmark ways, including increase in the physical scale of observed linkage disequilibrium (LD) and shifts in the site-frequency spectrum at polymorphic sites (Maynard Smith and Haigh, 1974; Braverman et al., 1995; Przeworski, 2002). Statistical tests have been developed to detect such genomic footprints (Tajima, 1989; Fay and Wu, 2000; Kim and Stephan, 2002). To test for the presence of hitchhiking at *SAG*, we used a compound statistic, *HEW* (Zeng et al., 2007b), that combines a site-frequency spectrum based test, Fay and Wu's *H* (Fay and Wu, 2000), and a haplotype-based test, *EW* (Watterson, 1978). We are specifically aiming to detect scenarios where genetic hitchhiking has resulted in an excess of high-frequency derived variants and an increase in haplotype structure. When we applied this test to the entire sequenced region, we found that patterns of genetic variation in the M form were consistent with neutrality (*HEW*:  $p = 1.0$ ), but the S form harbored patterns of variation that nearly significantly departed from the neutral expectation in the composite tests (*HEW*:  $p = 0.0507$ ) and could reflect the effects of recent positive selection. This S form signal appears to be driven largely by an increase in haplotype structure that is unexpected under a neutral model, although not significantly so (*EW* = 0.2089,  $p = 0.0712$ ; Table 1). Scanning this region using a sliding-window approach to localize the signal indicated that genetic variation downstream of *SAG*, including part of *gSGIa*, harbors the most significant deviations from neutrality (*HEW*: uncorrected  $p = 0.0267$ ). The windows that include *SAG* and *SGI-2* do not reject the neutral model (*HEW*:  $p > 0.05$ ; Fig. 1).



We applied the McDonald-Kreitman test to these data to test for the signature of historical selection using *A. merus* as an outgroup species and found no evidence for an excess of nonsynonymous fixed differences in either population (M form  $p = 0.4892$ ; S form  $p = 0.2757$ ; Table 2). A comparison of the rates of substitution at nonsynonymous and synonymous sites ( $K_d/K_s$ ) yielded  $K_d/K_s = 0.250$  for the S form and  $K_d/K_s = 0.258$  for the M form. If multiple episodes of positive selection have fixed nonsynonymous sites at this locus, the  $K_d/K_s$  ratio would be expected to exceed one (Hughes and Nei, 1988). The lower rate of substitution at nonsynonymous sites relative to synonymous sites ( $K_s = 0.11$  for both M and S form) in both subpopulations, combined with the non-significant McDonald-Kreitman test, is consistent with the operation of purifying selection as the predominant mode of evolution at *SAG*. We sequenced only a small portion of the coding sequence for *gSG1a*, so more data will be required to determine whether this gene could be under positive selection.

### 3.2. Lanb2

We tested the hypothesis that *LANB2* might evolve under positive selection by analyzing both intraspecific and interspecific genetic variation. *LANB2* is a relatively large gene composed of nine exons that span approximately 8 kb of the distal tip of the left arm of chromosome 2. We sequenced an approximately 10 kb region that includes *LANB2* as well as 497 bp of sequence upstream and 789 bp of sequence downstream of *LANB2*. Consistent with other estimates of nucleotide diversity at *A. gambiae* autosomal loci (Cohuet et al., 2008), *LANB2* harbors substantial genetic diversity in both populations with per site estimates of diversity ( $\theta_W$ ) equaling 0.0180 and 0.0175 for the M and S molecular forms, respectively, for all functional classes of nucleotide sites (Table 1). The vast majority of this variation maps to synonymous sites, however, and nonsynonymous variation is nearly absent (Table 1).

We tested *LANB2* for evidence of recent positive selection. As for *SAG*, we applied the compound *HEW* test statistic (Zeng et al., 2007b) using *A. merus* as the outgroup species. When applied to the entire region, *HEW* is not significant for either population ( $p_M = 1.0$ ;  $p_S = 1.0$ ; Table 1). To rule out the possibility that a smaller region within the 10 kb sequenced region could have experienced positive selection and is being masked by neutral patterns in the larger surrounding sequence, we calculated *HEW* using a sliding-window of 1200 bp with a step size of 500 bases, but this approach failed to reveal any significant windows (all  $p > 0.05$ ). Our inability to reject a neutral equilibrium model at this locus suggests that the protein coding sequence of *LANB2* has been evolving neutrally or under purifying selection in the recent past.

We used interspecies comparative analyses to determine whether this gene has experienced repeated selective sweeps in the more distant past. First, we applied the McDonald-Kreitman test to these data for each population independently using *A. merus* as the outgroup, but we found no evidence of deviations from neutral expectations (M form  $p = 0.1367$ ; S form  $p = 0.1846$ ; Table 2). Second, we used  $K_d/K_s$  to test for accelerated rates of evolution at nonsynonymous sites. Since the  $K_d/K_s$  test is conservative when analyzed across large sequence spans, we used a sliding window approach as described above. We found that the maximum  $K_d/K_s$  value never exceeded 0.1 in either molecular form in any 1200 bp window (maximum  $K_d/K_s = 0.046$  for S and 0.053 for M,  $K_s = 0.05$  for M and S forms). Coupled with the results from intraspecific analyses, these data provide no evidence for positive selection on *LANB2*.

## 4. Discussion

Structural and receptor host proteins can also be intimately involved in host-pathogen interactions, including in relation to their crucial role in tissue recognition and invasion (Brennan et al., 2000; Arrighi et al., 2005). Like immune proteins, non-immune proteins with a direct role in pathogen development within the host could be subject to pathogen-related selection pressure. One such case of non-immune pathogen-related selection has been documented in humans at the Duffy erythrocyte receptor protein, which is exploited by the malaria parasite for merozoite invasion (Hamblin and Di Rienzo, 2000). Similarly, HIV-related selection is thought to be currently driving the evolution of the chemokine CCR5 in Africa (Schliekelman et al., 2001). Several *A. gambiae* proteins have been identified as playing roles in the establishment or development of malaria parasites within the mosquito host (Brennan et al., 2000; Arrighi et al., 2005; Rodrigues et al., 2012). We tested the genes that code for two of these proteins, saglin and laminin, for the signatures of *Plasmodium*-related selection pressure. To address hypotheses of selection at these genes, we sequenced alleles of these genes in a small population sample of both the M and S molecular forms of *A. gambiae* and applied population genetic tests to identify significant deviations from expectations under a neutral model. Neither intra-specific analyses based on the site-frequency spectrum of genetic variation, nor inter-specific comparisons designed to detect accelerated rates of evolution at nonsynonymous sites, revealed any evidence for adaptive evolution at these loci. The *HEW* neutrality test failed to reject the neutral model, confirming that the number of high-frequency derived alleles and the degree of haplotype homozygosity are both consistent with expectations under a neutral model. Inter-specific comparisons with *A. merus* also revealed patterns expected under models of purifying selection and failed to provide any evidence for adaptive evolution in the protein coding sequences.

It is possible that the population genetic tests in our study may be underpowered because of the limited number of chromosomes sampled and the limited divergence between *A. gambiae* and *A. merus*. In population genetic analyses, as in any experimental setting, small sample sizes lead to increased sampling variance and limit the amount of information available for analysis (i.e. number of segregating sites), in turn potentially resulting in false negative test results (Simonsen et al., 1995). This may be of special concern for *SAG* since levels of diversity are particularly low in this gene region, partly owing to its location on the X chromosome, which is generally depauperate of diversity in *A. gambiae* (Cohuet et al., 2008; Lawniczak et al., 2010; Neafsey et al., 2010). The issue was partially mitigated by sequencing a larger physical region to capture both additional segregating sites as well as the spatial distribution of diversity. Statistical tests based on divergence from an outgroup, such as the McDonald-Kreitman test, are underpowered within the *A. gambiae* species complex because the very recent divergence of the species violates assumptions of the model upon which these tests are based (Besansky et al., 2003; Obbard et al., 2009). These violations are of most concern when comparing *A. gambiae* to very closely related species such as *Anopheles arabiensis*, however, and previous studies have suggested that divergence between *A. merus* and *A. gambiae* may be sufficient (synonymous divergence ranging  $K_s = 4-11\%$ ) to allow the proper application of divergence based tests (Obbard et al., 2007, 2009; Wang-Sattler et al., 2007; Parmakelis et al., 2008; Rottschaefer et al., 2011). Nonetheless, some caution is warranted when interpreting the results of these tests.

Our results lead us to accept the null hypothesis of neutral evolution and purifying selection at the *SAG* and *LANB2* loci, suggesting that the proteins encoded by these genes are not involved in evolutionary host-parasite conflict, or at least are not co-evolving with parasites. Both saglin and laminin could serve critical functions to the host that require the protein structures to remain conserved, consistent with the evidence of particularly strong purifying



selection at *LANB2* (Table 1). Overall, these results suggest that both proteins remain reasonable candidates for interventions such as transmission blocking vaccines (Brennan et al., 2000; Dinglasan and Jacobs-Lorena, 2008) since neither protein is rapidly evolving or adaptively maintaining functionally variable alleles. Although *SAG* harbors a substantial number of fixed differences between species, the level of nonsynonymous polymorphism currently segregating in the population is very low for both proteins. This suggests that technologies could easily be developed to target existing alleles, and those technologies could be expected to remain effective at least in the natural populations sampled here.

## Acknowledgments

We thank Chip Aquadro for helpful feedback on an earlier draft of the manuscript. We are also grateful to two anonymous reviewers for helpful comments that improved the manuscript. This work was supported by NIH grant AI062995 and a fellowship to JEC from the Cornell Center for Comparative and Population Genomics.

## References

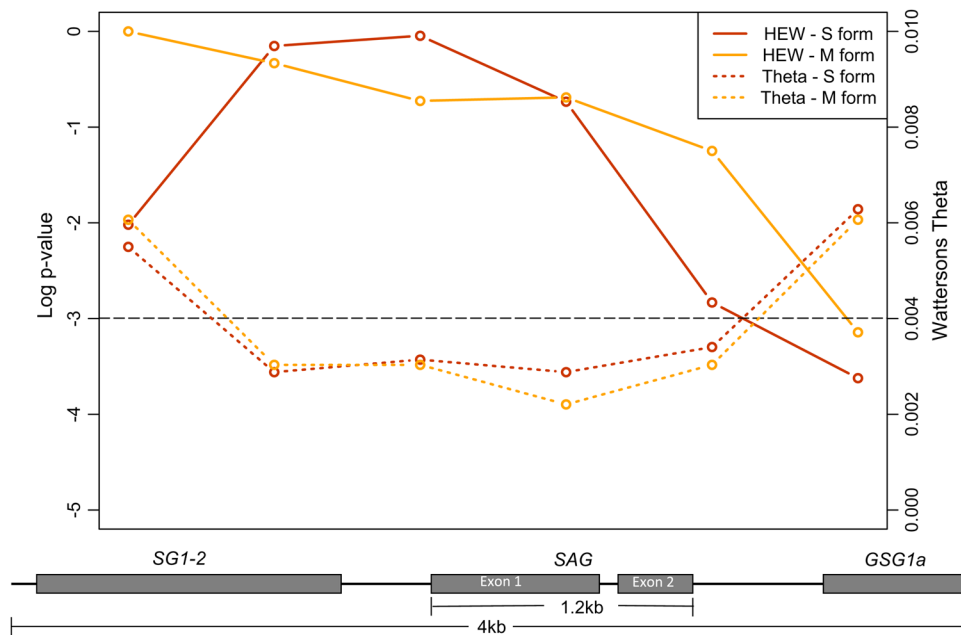
- Adini A, Warburg A. Interaction of *Plasmodium gallinaceum* ookinetes and oocysts with extracellular matrix proteins. *Parasitology*. 1999; 119 (Pt 4):331–336. [PubMed: 10581609]
- Arcá B, Lombardo F, De Lara Capurro M, Della Torre A, Dimopoulos G, James AA, Coluzzi M. Trapping cDNAs encoding secreted proteins from the salivary glands of the malaria vector *Anopheles gambiae*. *Proc Natl Acad Sci USA*. 1999; 96:1516–1521. [PubMed: 9990055]
- Arcà B, Lombardo F, Valenzuela JG, Francischetti IMB, Marinotti O, Coluzzi M, Ribeiro JMC. An updated catalogue of salivary gland transcripts in the adult female mosquito, *Anopheles gambiae*. *J Exp Biol*. 2005; 208:3971–3986. [PubMed: 16215223]
- Arrighi RBG, Hurd H. The role of *Plasmodium berghei* ookinete proteins in binding to basal lamina components and transformation into oocysts. *Int J Parasitol*. 2002; 32:91–98. [PubMed: 11796126]
- Arrighi RBG, Lycett G, Mahairaki V, Siden-Kiamos I, Louis C. Laminin and the malaria parasite's journey through the mosquito midgut. *J Exp Biol*. 2005; 208:2497–2502. [PubMed: 15961736]
- Barreau C, Touray M, Pimenta PF, Miller LH, Vernick KD. *Plasmodium gallinaceum*: sporozoite invasion of *Aedes aegypti* salivary glands is inhibited by anti-gland antibodies and by lectins. *Exp Parasitol*. 1995; 81:332–343. [PubMed: 7498430]
- Barry AE, Schultz L, Buckee CO, Reeder JC. Contrasting population structures of the genes encoding ten leading vaccine-candidate antigens of the human malaria parasite *Plasmodium falciparum*. *PLoS One*. 2009; 4:e8497. [PubMed: 20041125]
- Besansky NJ, Krzywinski J, Lehmann T, Simard F, Kern M, Mukabayire O, Fontenille D, Touré Y, Sagnon N. Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proc Natl Acad Sci USA*. 2003; 100:10818–10823. [PubMed: 12947038]
- Biffen RH. Mendel's laws of inheritance and wheat breeding. *J Agric Sci*. 1905; 1:4–48.
- Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*. 1995; 140:783–796. [PubMed: 7498754]
- Brennan JD, Kent M, Dhar R, Fujioka H, Kumar N. *Anopheles gambiae* salivary gland proteins as putative targets for blocking transmission of malaria parasites. *Proc Natl Acad Sci USA*. 2000; 97:13859–13864. [PubMed: 11087838]
- Carter, R.; Chen, DH. Malaria transmission blocked by immunisation with gametes of the malaria parasite. 1976. p. 57-60. <http://dx.doi.org/10.1038/263057a0263> (Published online: 02 September 1976)
- Cohuet A, Krishnakumar S, Simard F, Morlais I, Koutsos A, Fontenille D, Mindrinos M, Kafatos FC. SNP discovery and molecular evolution in *Anopheles gambiae*, with special emphasis on innate immune system. *BMC Genomics*. 2008; 9:227. [PubMed: 18489733]
- Dessens JT, Sidén-Kiamos I, Mendoza J, Mahairaki V, Khater E, Vlachou D, Xu XJ, Kafatos FC, Louis C, Dimopoulos G, et al. SOAP, a novel malaria ookinete protein involved in mosquito

midgut invasion and oocyst development. *Mol Microbiol.* 2003; 49:319–329. [PubMed: 12828632]

- Dinglasan RR, Jacobs-Lorena M. Flipping the paradigm on malaria transmission-blocking vaccines. *Trends Parasitol.* 2008; 24:364–370. [PubMed: 18599352]
- Favia G, Lanfrancotti A, Spanos L, Sidén-Kiamos I, Louis C. Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae* s.s. *Insect Mol Biol.* 2001; 10:19–23. [PubMed: 11240633]
- Fay JC, Wu CI. Hitchhiking under positive darwinian selection. *Genetics.* 2000; 155:1405–1413. [PubMed: 10880498]
- Ghosh, AK.; Devenport, M.; Jethwaney, D.; Kalume, DE.; Pandey, A.; Anderson, VE.; Sultan, AA.; Kumar, N.; Jacobs-Lorena, M. Malaria parasite invasion of the mosquito salivary gland requires interaction between the plasmodium TRAP and the *Anopheles* saglin proteins. In: schneider, DS., editor. *PLoS Pathog.* Vol. 5. 2009. p. e1000265
- Ghosh AK, Ribolla PE, Jacobs-Lorena M. Targeting *Plasmodium* ligands on mosquito salivary glands and midgut with a phage display peptide library. *Proc Natl Acad Sci USA.* 2001; 98:13278–13281. [PubMed: 11687659]
- Gwadz RW. Successful immunization against the sexual stages of *Plasmodium Gallinaceum*. *Science.* 1976; 193:1150–1151. [PubMed: 959832]
- Haldane JBS. Disease and evolution. *La Ricerca Scientifica Supplemento A.* 1949; 19:68–76.
- Hamblin MT, Di Rienzo A. Detection of the signature of natural selection in humans: evidence from the duffy blood group locus. *Am J Hum Genet.* 2000; 66:1669–1679. [PubMed: 10762551]
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JMC, Wides R, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science.* 2002; 298:129–149. [PubMed: 12364791]
- Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature.* 1988; 335:167–170. [PubMed: 3412472]
- Kim Y, Stephan W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics.* 2002; 160:765. [PubMed: 11861577]
- Korochkina S, Barreau C, Pradel G, Jeffery E, Li J, Natarajan R, Shabanowitz J, Hunt D, Frevert U, Vernick KD. A mosquito-specific protein family includes candidate receptors for malaria sporozoite invasion of salivary glands. *Cell Microbiol.* 2006; 8:163–175. [PubMed: 16367875]
- Lanfrancotti A, Lombardo F, Santolamazza F, Veneri M, Castrignanò T, Coluzzi M, Arcà B. Novel cDNAs encoding salivary proteins from the malaria vector *Anopheles gambiae*. *FEBS Lett.* 2002; 517:67–71. [PubMed: 12062411]
- Lavazec C, Bourgouin C. Mosquito-based transmission blocking vaccines for interrupting *Plasmodium* development. *Microbes Infect.* 2008; 10:845–849. [PubMed: 18656409]
- Lawniczak MKN, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, Redmond S, Fulton L, Appelbaum E, Godfrey J, et al. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science.* 2010; 330:512–514. [PubMed: 20966253]
- Lederberg J. J.B.S. Haldane (1949) on infectious disease and evolution. *Genetics.* 1999; 153:1–3. [PubMed: 10471694]
- Lehmann T, Diabate A. The molecular forms of *Anopheles gambiae*: a phenotypic perspective. *Infect Genet Evol.* 2008; 8:737–746. [PubMed: 18640289]
- Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009; 25:1451–1452. [PubMed: 19346325]
- Matuschewski K, Nunes AC, Nussenzweig V, Ménard R. *Plasmodium* sporozoite invasion into insect and mammalian cells is directed by the same dual binding system. *EMBO J.* 2002; 21:1597–1606. [PubMed: 11927544]
- Maynard Smith J, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res.* 1974; 23:23–35. [PubMed: 4407212]
- McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature.* 1991; 351:652–654. [PubMed: 1904993]

- Nacer A, Walker K, Hurd H. Localisation of laminin within *Plasmodium berghei* oocysts and the midgut epithelial cells of *Anopheles stephensi*. *Parasit Vectors*. 2008; 1:33. [PubMed: 18808667]
- Neafsey DE, Lawniczak MKN, Park DJ, Redmond SN, Coulibaly MB, Traoré SF, Sagnon N, Costantini C, Johnson C, Wiegand RC, et al. SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science*. 2010; 330:514–517. [PubMed: 20966254]
- Obbard DJ, Linton YM, Jiggins FM, Yan G, Little TJ. Population genetics of *Plasmodium* resistance genes in *Anopheles gambiae*: no evidence for strong selection. *Mol Ecol*. 2007; 16:3497–3510. [PubMed: 17688548]
- Obbard DJ, Welch JJ, Little TJ. Inferring selection in the *Anopheles gambiae* species complex: an example from immune-related serine protease inhibitors. *Malar J*. 2009; 8:117. [PubMed: 19497100]
- Okulate MA, Kalume DE, Reddy R, Kristiansen T, Bhattacharyya M, Chaerkady R, Pandey A, Kumar N. Identification and molecular characterization of a novel protein Saglin as a target of monoclonal antibodies affecting salivary gland infectivity of *Plasmodium* sporozoites. *Insect Mol Biol*. 2007; 16:711–722. [PubMed: 18093000]
- Parmakelis A, Slotman MA, Marshall JC, Awono-Ambene PH, Antonio-Nkondjio C, Simard F, Caccone A, Powell JR. The molecular evolution of four anti-malarial immune genes in the *Anopheles gambiae* species complex. *BMC Evol Biol*. 2008; 8:79. [PubMed: 18325105]
- Przeworski M. The signature of positive selection at randomly chosen loci. *Genetics*. 2002; 160:1179–1189. [PubMed: 11901132]
- Rodrigues J, Oliveira GA, Kotsyfakis M, Dixit R, Molina-Cruz A, Jochim R, Barillas-Mury C. An epithelial serine protease, AgESP, is required for plasmodium invasion in the mosquito *Anopheles gambiae*. *PLoS One*. 2012; 7:e35210. [PubMed: 22509400]
- Rottschaefer SM, Riehle MM, Coulibaly B, Sacko M, Niaré O, Morlais I, Traoré SF, Vernick KD, Lazzaro BP. Exceptional diversity, maintenance of polymorphism, and recent directional selection on the APL1 malaria resistance genes of *Anopheles gambiae*. *PLoS Biol*. 2011; 9:e1000600. [PubMed: 21408087]
- Saul A. Mosquito stage, transmission blocking vaccines for malaria. *Curr Opin Infect Dis*. 2007; 20:476–481. [PubMed: 17762780]
- Schliekelman P, Garner C, Slatkin M. Natural selection and resistance to HIV. *Nature*. 2001; 411:545–546. [PubMed: 11385558]
- Simonsen KL, Churchill GA, Aquadro CF. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics*. 1995; 141:413–429. [PubMed: 8536987]
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989; 123:585. [PubMed: 2513255]
- Vlachou D, Lycett G, Sidén-Kiamos I, Blass C, Sinden RE, Louis C. *Anopheles gambiae* laminin interacts with the P25 surface protein of *Plasmodium berghei* ookinetes. *Mol Biochem Parasitol*. 2001; 112:229–237. [PubMed: 11223130]
- Wang-Sattler R, Blandin S, Ning Y, Blass C, Dolo G, Touré YT, Delle Torre A, Lanzaro GC, Steinmetz LM, Kafatos FC, et al. Mosaic genome architecture of the *Anopheles gambiae* species complex. *PLoS One*. 2007; 2:e1249. [PubMed: 18043756]
- Warburg A, Shtern A, Cohen N, Dahan N. Laminin and a *Plasmodium* ookinete surface protein inhibit melanotic encapsulation of Sephadex beads in the hemocoel of mosquitoes. *Microbes Infect*. 2007; 9:192–199. [PubMed: 17224290]
- Watterson GA. The homozygosity test of neutrality. *Genetics*. 1978; 88:405–417. [PubMed: 17248803]
- Weedall GD, Preston BMJ, Thomas AW, Sutherland CJ, Conway DJ. Differential evidence of natural selection on two leading sporozoite stage malaria vaccine candidate antigens. *Int J Parasitol*. 2007; 37:77–85. [PubMed: 17046771]
- Zeng K, Fu YX, Shi S, Wu CI. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics*. 2006; 174:1431–1439. [PubMed: 16951063]
- Zeng K, Mano S, Shi S, Wu CI. Comparisons of site- and haplotype-frequency methods for detecting positive selection. *Mol Biol Evol*. 2007a; 24:1562–1574. [PubMed: 17449894]

Zeng K, Shi S, Wu CI. Compound tests for the detection of hitchhiking under positive selection. *Mol Biol Evol.* 2007b; 24:1898–1908. [PubMed: 17557886]



**Fig. 1.** Sliding window analysis of *SAG* using *HEW* neutrality test. The compound neutrality test statistic *HEW* plotted as a function of position across sequenced region containing three genes. *HEW* was calculated in a sliding window analysis using a 1200 bp window with a 500 bp step size. The schematic below the plot indicates the location of each gene in the region. Both the uncorrected log transformed *HEW* *p*-values and Watterson's per-site estimate of diversity ( $\theta = 4 N\mu$ ) are plotted for each window. The black dotted line indicates the 0.05 significance threshold, such that any *p*-value in the *HEW* statistic below the line satisfies nominal criteria for statistical significance.

Table 1

Nucleotide diversity and neutrality tests for *SAG* and *LANB2*.

	$n^a$	$S^b$	$\theta_N^c$	$\pi^d$	$\pi_s^e$	$\pi_o^f$	$H$	$EW$	$HEW$
<i>SAG</i>									
M form	14	59	0.0051	0.0044	0.0092	0.0005	-1.0300	0.0769	1.0
S form	15	58	0.0048	0.0049	0.0081	0.0008	-0.3680	0.2089	0.0507
<i>LANB2</i>									
M form	14	480	0.0180	0.0157	0.0369	0.0002	0.4972	0.0714	1.0
S form	11	412	0.0175	0.0151	0.0358	0.0002	0.2954	0.0909	1.0

<sup>a</sup>Number of chromosomes sampled per population.

<sup>b</sup>Total number of segregating sites detected in each sample.

<sup>c</sup>Per site Watterson's Theta calculated on all segregating sites.

<sup>d</sup>Average number of pairwise differences (nucleotide diversity) calculated on all segregating sites.

<sup>e</sup>Nucleotide diversity at synonymous sites only.

<sup>f</sup>Nucleotide diversity at nonsynonymous sites only.



Table 2

McDonald-Kreitman test results for both *SAG* and *LANB2*.

	Polymorphic sites		Fixed sites		<i>P</i> -value
	Synonymous	Non-synonymous	Synonymous	Non-synonymous	
<i>S form</i>					
<i>SAG</i> <sup>a</sup>	7	2	27	25	0.2757
All <sup>b</sup>	19	17	64	77	0.4587
<i>M form</i>					
<i>SAG</i> <sup>a</sup>	7	3	27	24	0.4892
All <sup>b</sup>	17	20	66	77	1.0
<i>S form</i>					
<i>LANB2</i>	151	4	22	2	0.1846
<i>M form</i>					
<i>LANB2</i>	162	4	19	2	0.1367

Significance of contingency table evaluated using Fisher's exact test, no corrections for multiple testing applied.

<sup>a</sup>Test applied to all coding regions in the sequenced region (*SGI-2*, *SAG*, *gSGI/a*).

<sup>b</sup>Test applied to *SAG* alone.