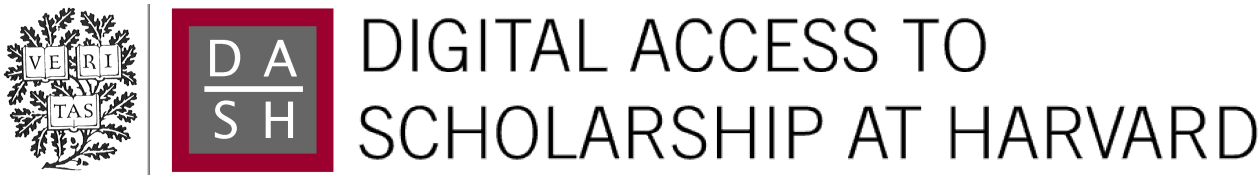

Glycophorin B is the Erythrocyte Receptor of Plasmodium Falciparum Erythrocyte-Binding Ligand, EBL-1

Supporting Information

Mayer et al. 10.1073/pnas.0900878106

SI Text

This supporting information (SI) considers some issues of population genetics in regard to the hypothesis that the null allele of glycophorin B confers a selective advantage because of reduced susceptibility to *Plasmodium falciparum* malaria. The estimated frequency of the null allele in the Ituri forest pygmies is 0.59. Are there plausible scenarios of selection that could account for the null allele attaining such a high frequency in the relevant time frame? The answer depends on the time available, the initial null allele frequency in the population, the fitness advantage of homozygous null genotypes, and the degree of dominance affecting the fitness of null allele heterozygotes.

The time frame is likely bracketed by 10,000–100,000 years. The older date is an estimate of when *P. falciparum* malaria was spread around the world, and the more recent date corresponds to an expansion of infection with the introduction of swidden agriculture into Africa and the diversification of the *Anopheles gambiae* complex of mosquitoes. Both dates are based on the estimated times to coalescence of mitochondrial DNA lineages in the parasite (1). Assuming 20 years per human generation, 10,000–100,000 years corresponds to 500–5,000 generations. The initial allele frequency depends on the size of the population because it is reasonable to assume that the null allele was originally present in a single copy. We will assume population sizes of 1,000–10,000, yielding initial allele frequencies of $1/(2 \times 1{,}000) = 0.0005$ or $1/(2 \times 10{,}000) = 0.00005$, respectively. If the actual population size of the Ituri forest pygmies was outside this range when selection began, the actual size was likely to be smaller than 1,000, which makes the following calculations conservative.

Two models of selection need to be considered: directional selection (homozygote superiority) and overdominance (heterozygote superiority). We first consider directional selection with genotypes +/+, +/*null*, and *null*/*null* having relative fitnesses $1{:}1 + hs{:}1 + s$. In this symbolism, s is the selection coefficient favoring null homozygotes and h is the degree of dominance of the null allele with respect to fitness. Assuming random mating in a large population, and ignoring mutation and migration, the allele frequency of the null allele changes as $q' = [q^2(1 + s) + q(1 - q)(1 + hs)]/[q^2(1 + s) + 2q(1 - q)(1 + hs) + (1 - q)^2]$, where q is the allele frequency of the null allele in any generation and q' is the allele frequency a generation later (2). This recursion can easily be explored by iteration. Table S1 shows the parameter values in which selection drives the allele frequency from $q_0 = 0.0005$ or $q_0 = 0.00005$ to approximately $q = 0.59$ in 500–5,000 generations.

Most favorable for the hypothesis is the additive case (Table S1), which means that the fitness of heterozygous genotypes is exactly intermediate between those of the homozygous genotypes. In the additive case, the fitness advantage of the homozygous null that needs to be postulated is 3–4% for a time of 500 generations and 0.3–0.4% for a time of 5,000 generations.

It is probably more plausible to assume some degree of dominance less than additivity. The upper part of Table S1 assumes a 10% selective advantage for the homozygous null, based on the protective effect in sickle-cell heterozygotes (3) and indicates that values of $h = 0.12$–0.17 are consistent with a current null allele frequency of 0.59 after 500 generations. With partial dominance, most of the initial selection occurs in heterozygous genotypes, and the values of hs imply a selective advantage in heterozygotes of only 1–2%. For the longer time frame, the selective advantage can be 1 order of magnitude smaller (Table S1 assumes $s = 1\%$).

We have also explored models in which $h = 0$ (data not shown), which implies that the fitness advantage of the null homozygote is completely recessive. We find that models with completely recessive effects can be excluded because they require highly implausible values for the selective advantage of the homozygous null (e.g., $s \approx 4.0$).

The other selective scenario is heterozygous superiority. With $q = 0.59$ in contemporary populations, it is not clear whether this is the equilibrium value or only a point on the trajectory toward the equilibrium value. For concreteness, we shall assume that $q = 0.59$ is the equilibrium value. To this end, suppose that the relative fitnesses of +/+, +/*null*, and *null*/*null* as $1 - s{:}1{:}1 - t$. With random mating in a large population without mutation or migration, the equilibrium frequency of the null allele is given by $s/(s + t)$ (2), which, for $q = 0.59$, implies that $t = 0.695s$. For the relevant values of q_0 and time, the required magnitudes of s (and therefore of t) are given in the lower part of Table S1. These values are again quite modest, which reinforces intuition in recognizing that, with heterozygote superiority, the initial selection depends on the fitness advantage of the heterozygous genotypes.

Several issues also arise in relation to genetic changes in the parasite postulated to have arisen in response to the increasing allele frequency of the null allele of glycophorin B. The first is the haploid state of the erythrocyte stages, which allows selection to occur with maximal efficiency. The second is the short sexual generation time of the parasite, estimated as approximately 6 generations per year, which affords ample time for evolutionary changes to occur.

In regard to the specific run of *T*s in the parasite gene *ebl-1* resulting in a shifted reading frame that abrogates glycophorin B binding, replication slippage in microsatellites is a prominent feature of mutation in *P. falciparum* (4). Although the relatively high frequency of the frameshift allele could be accounted for by random genetic drift, especially in view of the population bottleneck that the parasite experienced ≈10,000 years ago (1, 4), positive selection cannot be ruled out. In addition, the high rate of new gene duplications observed in eukaryotic genomes (5) would easily provide the raw material for the diversification of the parasite *DBL-EBP* gene family by duplication and divergence.

1. Joy DA, et al. (2003) Early origin and recent expansion of *Plasmodium falciparum*. *Science* 300:318–321.
2. Hartl DL, Clark AG (2007) *Principles of Population Genetics* (Sinauer, Sunderland, MA).
3. Allison AC (1954) Protection afforded by sickle-cell trait against subtertian malarial infection. *Br Med J* 1:290–294.
4. Volkman SK, et al. (2001) Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* 293:482–484.
5. Lynch M (2007) *The Origins of Genome Architecture* (Sinauer, Sunderland, MA).

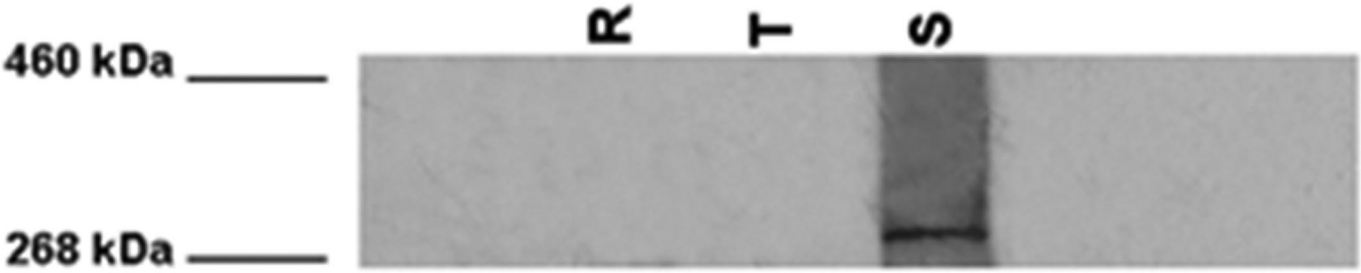

Fig. S1. Expression profile of EBL-1. EBL-1 is expressed only in the schizont stages of intraerythrocytic development. R, ring; T, trophozoite; S, schizont.

Table S1. Parameter values for directional selection and overdominance

Parameter values for directional selection		
Time = 500 generations		
Initial frequency	Additive case (h = 0.50)	Partial dominance (s = 0.10)
0.0005	s = 0.032	h = 0.12
0.00005	s = 0.041	h = 0.17
Time = 5,000 generations		
Initial frequency	Additive case (h = 0.50)	Partial dominance (s = 0.01)
0.0005	s = 0.0032	h = 0.12
0.00005	s = 0.0041	h = 0.17
Parameter values for overdominance (t = 0.695 s).		
Initial frequency	Time = 500 generations	Time = 5,000 generations
0.0005	s = 0.027	s = 0.0032
0.00005	s = 0.035	s = 0.0040