Published in final edited form as:
Infect Genet Evol. 2012 July ; 12(5): 1037–1045. doi:10.1016/j.meegid.2012.02.019.

The *SLC4A1* gene is under differential selective pressure in primates infected by *Plasmodium falciparum* and related parasites

Michael E. Steiper[a,b,c,d,*], **Fiona Walsh**[a], and **Julia M. Zichello**[b,d]

Michael E. Steiper: msteiper@hunter.cuny.edu; Fiona Walsh: fwalsh@hunter.cuny.edu; Julia M. Zichello: jzichello@gc.cuny.edu

[a]Department of Anthropology, Hunter College of the City University of New York (CUNY), 695 Park Avenue, NY, NY 10065, USA

[b]Program in Anthropology, The Graduate Center, CUNY, 365 5th Avenue, NY, NY 10016, USA

[c]Program in Biology, The Graduate Center, CUNY, 365 5th Avenue, NY, NY 10016, USA

[d]New York Consortium in Evolutionary Primatology (NYCEP), NY, USA

Abstract

Malaria is a disease caused by *Plasmodium* parasites and is responsible for high mortality in humans. This disease is caused by four different species of *Plasmodium* though the main source of mortality is *Plasmodium falciparum.* Humans have a number of genetic adaptations that act to combat *Plasmodium*. One adaptation is a deletion in the *SLC4A1* gene that leads to Southeast Asian ovalocytosis (SAO). There is evidence that SAO erythrocytes are resistant to multiple *Plasmodium* species. Here we analyze *SLC4A1* in 23 primates and mammals to test for differential selective pressures among different primate lineages. Because primates are infected with both human *Plasmodium* parasites and their relatives, this analysis can be used to test which human *Plasmodium* parasite is the likely target of SAO. A significantly different pattern of molecular evolution was found in humans and African apes, species that are infected by *P. falciparum* and its relatives. This effect was restricted to the cytosolic domain of the *SLC4A1* gene. The evidence is consistent with a different selective regime operating on this gene domain in humans and African apes, when compared to other primates and mammals. Alternatively, this pattern is consistent with a relaxation of selection or weak adaptive evolution operating on a small number of amino acids. The adaptive interpretation of the results is consistent with the SAO allele of the *SLC4A1* gene interacting with *P. falciparum* in humans, rather than other *Plasmodium* parasites. However, additional investigation of the relationship between *SLC4A1* variants and *Plasmodium* in humans and African apes is required to test whether the different selective regime in humans and African apes is due to natural selection or relaxed constraint.

[*]Corresponding author at: Department of Anthropology, Hunter College of the City University of New York (CUNY), 695 Park Avenue, NY, NY 10065, USA. Tel.: +1 12127725418.

Keywords

Malaria; Evolution; Adaptation; African ape; Hominoid; *Plasmodium*

1. Introduction

Malaria is a parasitic disease responsible for approximately 1 million human deaths annually (WHO, 2008). This disease is caused by four different species of *Plasmodium* parasite: *P. falciparum*, *P. vivax*, *P. malariae*, and *P. ovale*. The main source of human mortality is the *P. falciparum* parasite, though *P. vivax* infections are also responsible for considerable levels of morbidity (Mendis et al., 2001). A number of genetic traits have evolved by natural selection to mitigate this strong selective agent in humans (Kwiatkowski, 2005). One malarial adaptation is apparently related to Southeast Asian Ovalocytosis (SAO) (Baer et al., 1976). SAO is a condition caused by a mutation in the *solute carrier, anion exchanger, member 1* gene, *SLC4A1*, which encodes the band 3 protein, a major component of erythrocytes (Liu et al., 1990). Band 3 has two main functional domains (Fig. 1). The N-terminal intracellular region plays a role in the structure of the erythrocyte (Jay, 1996) and the C-terminal transmembrane region plays a role in anion exchange (Tanner, 1993). The SAO-causing mutation is a 27 base pair deletion in *SLC4A1* that removes amino acids 400–408, a region at the junction of the functional domains of the protein. This deletion causes an increased rigidity of erythrocytes (Jarolim et al., 1991; Mohandas et al., 1992). The prevalence of the SAO allele can be as high as 35% in malarial areas (Mgone et al., 1996), suggesting that there is an extremely high selective benefit for the heterozygote carriers of the deletion allele, given that homozygosity is apparently lethal (Liu et al., 1994). The evidence is clear that SAO chiefly protects individuals against severe cerebral malaria (Allen et al., 1999; Genton et al., 1995), a form of the disease caused almost exclusively by *P. falciparum*. Although it was originally thought that SAO worked via a reduction in cytoadherence (Allen et al., 1999), Cortés et al. (2005) suggested that the protective mechanism of SAO is actually an increased cytoadherence of infected erythrocytes through the CD36 protein.

Aside from a link between SAO and malaria, there is also evidence that band 3 has a more general role in erythrocyte invasion by *Plasmodium*. Though the major erythrocyte receptor for *P. falciparum* invasion is glycophorin A (GYPA) (Pasvol et al., 1982; Sim et al., 1994), the parasite is not solely dependent on this protein (Okoyeh et al., 1999). Indeed, *P. falciparum* uses a suite of surface receptors for erythrocyte invasion, some of which are yet to be fully characterized (reviewed in Baum et al., 2005). The band 3 protein has been forwarded as an alternate receptor for *Plasmodium* that is sialic acid-independent (Goel et al., 2003; Kariuki et al., 2005; Li et al., 2009, 2008, 2004; Okoye and Bennett, 1985). Finally, aside from its role in the erythrocyte, a shorter isoform of *SLC4A1* is expressed in the kidney where it plays a role in acid secretion; mutations in the gene can lead to distal renal tubular acidosis (reviewed in Williamson and Toye, 2008).

While SAO clearly targets *P. falciparum*-caused severe malaria (Allen et al., 1999; Genton et al., 1995), there is some evidence that the SAO allele may protect against other *Plasmodium* parasites. SAO erythrocytes were shown to be relatively resistant to

parasitemia or invasion by multiple *Plasmodium* parasites (Cattani et al., 1987; Foo et al., 1992; Hadley et al., 1983; Kidson et al., 1981; Serjeantson et al., 1977). However, these findings may have been artifactual due to methodological concerns in the handling of SAO cells (Bruce et al., 1999; Dluzewski et al., 1992). It also is possible that SAO may only be effective against erythrocyte invasion in particular strains of *P. falciparum* (Cortés et al., 2004), suggesting the efficacy of SAO is restricted even within this species. However, Allen et al. (1999) found that SAO did not effect erythrocyte invasion by *P. vivax* or *P. falciparum*, but instead caused resistance to cerebral malaria. Though cerebral malaria is most closely associated with *P. falciparum*, *P. vivax* can cause cerebral and other forms of severe malaria (Kochar et al., 2005; Rogerson and Carter, 2008). Given this, it is possible that SAO may be an adaptation to cerebral malaria, which can be caused by different *Plasmodium* species. A general relationship between *SLC4A1* and *Plasmodium* is supported by the conservation of the extracellular portion of the *Plasmodium* PfSPP protein, which may interact with band 3 (Li et al., 2009).

Here we examine the pattern of molecular evolution at *SLC4A1* in humans, primates, and other mammals to determine whether this gene is evolving adaptively in primates that harbor *P. falciparum* related parasites or primates that harbor *Plasmodium* parasites generally. A pattern of differential selection at the *SLC4A1* gene in primates with particular *Plasmodium* infections will help to ascertain which malarial parasites are being targeted by this locus. Primates are useful models for testing between these hypotheses because they are infected with many different *Plasmodium* species (Table 1). Critically, it is now well established that wild African apes (chimpanzees, bonobos, and gorillas) are naturally infected with the human parasite *P. falciparum* and a series of its close phylogenetic relatives (Krief et al., 2010; Liu et al., 2010; Ollomo et al., 2009; Prugnolle et al., 2010; Rich et al., 2009). Phylogenetic relatives of the other human parasites (*P. vivax*, *P. malariae*, and *P. ovale*), on the other hand, infect a larger and more diverse set of primates (Hayakawa et al., 2008; Perkins and Schall, 2002). For this study, it is especially noteworthy that the *P. falciparum* group infects a relatively small set of primates and these parasites are only distantly related to the other human *Plasmodium* parasites (Hayakawa et al., 2008).

To address the conflicting evidence as to whether SAO is an adaptation to *P. falciparum* specifically or to *Plasmodium* generally, we probe the relationship between band 3 and *Plasmodium* by analyzing the coding region of *SLC4A1* in 23 mammals. The following hypotheses are tested. First, we test whether *SLC4A1* has been evolving adaptively across all 23 species. Second, we test whether patterns of molecular evolution at *SLC4A1* differ between primate species harboring *P. falciparum* related parasites (humans and African apes) and those harboring other *Plasmodium* parasites (Table 1; Fig. 2). Third, we test whether specific structural and functional domains of *SLC4A1* are evolving differently from one another.

2. Methods

2.1. Samples

The primate samples outlined in Table 1 were obtained from the Coriell Institute and the Duke Lemur Center. These were targeted for PCR and sequencing of the entire coding

region of the *SLC4A1* gene as described in the following sections. These sequences were added to existing DNA sequences from other primates and non-primate mammals from the UCSC Genome Center (Kent et al., 2002), Genbank (Benson et al., 2009), and Ensembl (Hubbard et al., 2009) (Table 1).

2.2. PCR and sequencing

The region targeted for sequencing in these primates was the region encompassed by the human *SLC4A1* gene (NG_007498.1). The entire coding region was sequenced, including all translated exons and intervening introns. In general, large overlapping regions of between 1 and 5 kb were targeted by PCR, amplicons were gel purified, these amplicons were subsequently cloned, clones were screened for the presence of the desired insert, and positive clones were sequenced in both directions. Occasionally, small fragments were sequenced directly from purified PCR products. PCRs were done using the Eppendorf Triplemaster and 5 Prime PCR Extender kits using a large number of primer sets and PCR parameters that were often species specific and fragment specific (protocol details and primer sequences available upon request). Both kits utilize a proofreading enzyme. Gel purification was done using the Eppendorf Perfectprep Gel Cleanup kit and TOPO XL Gel Purification kit. PCR purification was done using Millipore Montage Columns. Amplicons were cloned via Invitrogen TA and XL cloning kits. Plasmid DNA from multiple colonies was miniprepped using the Eppendorf FastPlasmid kit and screened for the insert via clone-test PCR and/or *Eco*RI restriction digestion. Plasmids were sequenced with the M13R and T7 sequencing primers, as well as PCR primers. DNA sequencing was performed by the Dana-Farber/Harvard Cancer Center (DF/HCC) High-Throughput DNA Sequencing Facility, Macrogen Sequencing Services, and Genewiz. Clones positive for the desired insert were sequenced in both directions for the entire sequence using multiple primers (sequencing primer sequences available upon request).

2.3. Sequence alignment and analysis

The codon sequences were aligned by eye to the human exonic sequences. This approach was satisfactory due to the relatively constrained exonic structure across the mammals studied. However, a limited number of exons were aligned using backtranslation from amino acid sequences (Bininda-Emonds, 2005). Codon sequences were analyzed for the action of positive natural selection by maximum likelihood using the *codeml* program of the *PAML* package (v. 4) (Yang, 2007). In each test, the F3x4 codon model was used, κ was estimated, and a user tree was supplied. The user tree (Fig. 2) was based on a generally accepted mammalian and primate phylogeny (Goodman et al., 1998; Murphy et al., 2004; Wildman et al., 2009). This was used to generate likelihood values under a number of different models. The convergence of the likelihood values was assessed by running each analysis multiple times. These likelihoods were compared using a series of nested likelihood ratio tests (LRTs) (Yang, 2006). In a LRT, a significant difference between models is assessed by the test statistic $2\Delta \ln L$, which follows a χ^2 distribution with degrees of freedom equal to the difference between the numbers of parameters in the competing models. When a significant difference is detected, the model with the higher likelihood is favored. If a significant difference is not detected, the simpler model is preferred, i.e. the one with fewer parameters.

Three types of analyses (Yang, 2006) were conducted to determine that pattern of adaptive molecular evolution at the *SLC4A1* gene, 'branch tests', 'site tests', and 'branch site tests'.

In the branch tests, a LRT is used to determine whether sequences across the entire tree were evolving neutrally by comparing the likelihood of a model where d_n/d_s (or ω) is fixed at 1 (the neutral model) against a model where the most likely value for ω is estimated. Subsequently, a model where ω was estimated was compared to two alternative models (Fig. 2). In both models, ω values are estimated for two sets of branches. In the *P. falciparum* model, the two sets of branches are (1) those primates infected with *P. falciparum* and *P. falciparum*-related parasites and (2) the remaining branches. In the *Plasmodium* model, the two sets of branches are (1) those primates infected with any *Plasmodium* species and (2) the remaining branches (not infected with any malarial parasite). In these models, we did not include stem branches among those being infected because some of these branches are exceptionally long in the primate phylogeny. It is not clear when along these long lineages that *Plasmodium* transferred into them. Therefore, it is more conservative not to include these lineages. Also, for the *Plasmodium* model, we infer three origins of *Plasmodium* infection in primates, one each in apes, Old World monkeys, and New World monkeys. An alternative would be to infer that primates are ubiquitously infected with *Plasmodium*, with infection being lost in *Callithrix*, *Colobus*, and lemurs and lorises. However, there is no evidence to suggest that *Plasmodium* is common in all of the main primate groups. This is especially true in lemurs, where infections have only been detected in *Eulemur* (Ellegren, 2008) and in a single specimen of *Propithecus*, despite sampling 55 samples from 6 lemur genera (Duval et al., 2010). Further information on the *Plasmodium* infections of the species sampled is presented in Table 1. In this analysis, *Papio* was included due to its infection with *Hepatocystis*, a relative of *Plasmodium* and known selective force (Tung et al., 2009). While rodents of the subfamily Murinae are infected by *Plasmodium*, *Mus musculus* and *Rattus norvegicus* were not included because these two species themselves are not infected.

Because multiple branch tests are being done (two alternative models vs. a null model) using the same data, a correction for multiple tests is necessary (Anisimova and Yang, 2007). There are a number of correction methods, which all essentially adjust the critical value (α) to reduce the type I error rate. Anisimova and Yang (2007) examined a number of such tests regarding the branch tests used here and determined that many methods were able to control error rates. Here, we used the Bonferroni correction, which is very conservative. The Bonferroni correction adjusts α by the number of tests. In the present case we are using two tests, so the adjusted α value is 0.025. These branch tests were conducted on four different functionally relevant partitions of exons. These are the entire gene (911 AA), the cytosolic domain (403 AA in humans), the transmembrane domain (508 AA), and the external sites, a subset of transmembrane amino acids that occur on the cell surface (122 AA) (from Bruce, 2006). The inferred nine amino acid insertion found in African apes was not included in the selection tests. In all cases, the number of the codon given corresponds to that from the human *SLC4A1* gene. Here, the tests of the different partitions are a refinement used to determine which regions of the gene are under selection. In this regard, the tests across these different partitions is considered a test of the robustness of the findings are do not require a correction for multiple tests.

Site tests determine if particular amino acid sites have experienced positive selection across the entire phylogeny (Yang, 2006). In these cases, three site models are compared. One is a neutral model with one ω estimated for all sites, the second is a nearly neutral model (with two ω classes, one evolving neutrally and a second evolving under purifying selection), and a third is a model with three ω classes (one neutral ($\omega = 1$), one purifying ($\omega < 1$), and one for sites evolving under positive selection ($\omega > 1$)). (These are models M0, M1, and M2 of *PAML*.) A related comparison fits 10 site classes for ω between 0 and 1, using a beta distribution (M7) and compares this to a model that also includes a site class with $\omega > 1$ (M8). LRTs that accept models allowing for a class of selected sites indicate the action of positive selection; Bayes empricial Bayes (BEB) (Yang et al., 2005) posterior probabilities are then used to determine the probabilities of those sites. These tests were conducted on the 4 different functionally relevant partitions of exons, as described above. In this regard, the tests across these different partitions is considered a test of the robustness of the findings are do not require a correction for multiple tests.

Branch site tests determine whether particular sites are evolving adaptively on particular lineages (Yang, 2006). Here, partitions of branches are the same as in the branch tests (Fig. 2). For the branch site test, the model estimates parameters for these two sets of branches. For the non-*Plasmodium* affected branches, two site classes are estimated, those under neutral evolution ($\omega = 1$) and purifying selection ($\omega < 1$). The lineages harboring *Plasmodium* parasites have these two site classes, and importantly also have an additional site class of positively selected sites. The LRT is between a model where this final site class is neutral ($\omega = 1$) and one under positive selection ($\omega \geq 1$). These tests were conducted on the 4 different functionally relevant partitions of exons, as described above. In this regard, the tests across these different partitions is considered a test of the robustness of the findings are do not require a correction for multiple tests. However, the tests between the two sets of branches do require a correction for multiple tests, implemented as described above.

3. Results

3.1. Alignment

We collected *SLC4A1* DNA sequences from 12 primates (Table 1). These sequences were added to publicly available DNA sequences from a range of primate and non-primate mammals. For two primates, we collected additional DNA sequences to cover unknown regions in the available genome sequences. In total, the alignment of *SLC4A1* exons comprised 23 species and was 2898 basepairs in length. One noteworthy aspect of the alignment was a potentially exonic 27 basepair insertion in the African apes, but not in humans (Fig. 3). This insertion is in frame and it is likely to be coding due to its pattern of intron–exon splice sites. The African ape insertion obliterates one splice site but has added another. It is also unusual to find an insertion with this pattern because it suggests that this insertion was gained in the African ape ancestor and then subsequently lost in humans because of the phylogenetic relationships of humans, chimpanzees, and gorillas. Furthermore, the insertion and subsequent deletion was at the same position and length; also very unlikely. A second interpretation is that insertion occurred once in the ancestor of humans, chimpanzees, and gorillas. Subsequently, due to ancestral polymorphism and

incomplete lineage sorting the derived condition is found in chimpanzees and gorillas, while the ancestral condition is found in humans. Because incomplete lineage sorting is a known problem within humans and African apes (Ruvolo, 1997), we find the second interpretation to be more likely for this short region. When the entire alignment is analyzed phylogenetically, the accepted phylogeny for apes is obtained, suggesting that the incomplete lineage sorting only affects this small region. Therefore, we used the accepted primate phylogeny in subsequent analyses requiring a user-generated tree. The alignment is available upon request.

3.2. Analyses of adaptive evolution across mammals

Before examining patterns of evolution specifically within primates harboring different *Plasmodium* species, the pattern of adaptive evolution over the entire mammalian phylogeny was examined. Multiple data partitions were analyzed: the entire gene (911 AA), the cytosolic domain (403 AA in humans), the transmembrane domain (508 AA), and the external sites, a subset of transmembrane amino acids that occur on the cell surface (122 AA). As expected for a functional gene, none of the partitions were evolving neutrally ($d_n/d_s = \omega = 1$); all had signatures of evolution consistent with purifying selection (Table 2) (ω ranged from 0.12 to 0.28). This level of purifying selection is consistent with average genome-wide estimates for ω in mammals (Ellegren, 2008). Comparing a model with a uniform ω across sites to a 'nearly neutral model' with two ω values for different sites (1 and <1) recovered a significant difference for all data partitions, showing that the amino acid sites are not evolving under a uniform neutral selective regime ('Site tests'; Table 3). This 'nearly neutral model' was tested against a model that incorporates an additional class of sites under positive selection, and thus is able to detect particular amino acids under selection. The model incorporating positive selection was a significantly better fit for both the entire gene and the transmembrane partition. In these analyses, nearly 1% of the amino acids were found to be evolving with a ω of approximately 3, suggesting that a limited number of sites are evolving positively in *SLC4A1*. Two amino acids were identified as evolving under statistically significant positive selection (367,658). A related comparison, where ω is modeled using a beta distribution, found evidence for positively selected sites in multiple comparisons, including codon 658 and additional sites (Table 4). Overall, *SLC4A1* is evolving in a manner consistent with a functional gene: the gene is under purifying selection and a limited number of sites are under positive selection.

3.3. Analyses of adaptive evolution in primates harboring different *Plasmodium* parasites

Subsequently, we examined whether the lineages reconstructed to have harbored *P. falciparum* related parasites were evolving differently from lineages that are not reconstructed to have harbored these parasites ('branch tests'; Fig. 2). When the entire coding region is tested, a model incorporating different ω values for these two sets of lineages is a better fit to the data than a model fitting only one ω to all lineages ($P = 0.014$; significant after correction for multiple tests) (Table 2). The ω of the lineages reconstructed to harbor *P. falciparum*-related species was higher (0.37) than the lineages reconstructed not to harbor these parasites (0.19). Of the different data partitions, only the cytosolic domain recovered a significant difference. In the cytosolic domain, the pattern of difference in ω was more marked ($\omega = 0.86$ in *P. falciparum* related lineages vs. 0.27 in other lineages) than in

the entire gene ($P = 0.001$; significant after correction for multiple tests). A phylogeny showing the reconstructed numbers of nonsynonymous and synonymous changes in humans and African apes reveals that all branches have higher absolute numbers of amino acid changes, except the *P. troglodytes* branch, which had no changes (Fig. 4).

A similar test was done to test whether primates harboring any type of *Plasmodium* species were evolving differently from lineages that were not reconstructed to have harbored these parasites (Fig. 2). When the entire coding region is tested, a model incorporating different ω values for these two sets of lineages is not a better fit to the data than a model fitting only one ω to all lineages (Table 2). In the cytosolic domain there is an elevated ω ratio in the species harboring any *Plasmodium* species ($\omega = 0.38$ vs. 0.26), though this finding was not significant after correction for multiple tests. None of the other partitions yielded significant results.

Finally, we tested whether particular sites were evolving adaptively in particular lineages (i.e. the 'branch site test'). First, positive evolution was tested at sites on the lineages reconstructed to have harbored *P. falciparum* and its relatives (Table 5). These analyses showed evidence for sites under positive selection in the lineages harboring *P. falciparum* relatives in the cytosolic region, but only using a less conservative critical value for the LRT ($P < 0.05$) (Yang, 2006). In this less conservative test, one site (111) was found to be evolving positively in lineages reconstructed to have harbored *P. falciparum*. However, correction for multiple tests would render this comparison insignificant. Second, we examined the sites on the lineages reconstructed to have harbored any *Plasmodium* species. None of these comparisons recovered sites under positive natural selection.

4. Discussion

The pattern of molecular evolution at the *SLC4A1* gene offers insight into the differences in adaptations between primate species harboring different sets of *Plasmodium* parasites. Differential ω values were detected in humans and African apes—species harboring *P. falciparum* related parasites. The differential ω was most pronounced in the cytosolic domain of *SLC4A1*. Although the ω values for the cytosolic domain in humans and African apes are not greater than 1, the benchmark for positive evolution, they are three times higher than the values estimated for the rest of the tree ($\omega = 0.88$ vs 0.27) and higher than average ω values estimated from mammalian genome-wide comparisons (Ellegren, 2008). This is a significant difference, despite the conservative nature of the test employed, and it shows that a shift in selective regimes has occurred in the African apes and humans at *SLC4A1*. A cautious interpretation of this finding is that the higher ω value in African apes is reflective of a relaxation of constraint or decreased purifying selection in the cytosolic domain. Less cautiously, these findings can be interpreted to support a hypothesis where *P. falciparum* and its relatives, found exclusively in humans and African apes, are a selective force acting on the cytosolic domain of *SLC4A1*. This latter interpretation is tenable based on the relationship the SAO mutation in humans and their relationship to malaria (Baer et al., 1976), the idea that *Plasmodium* is a selective force in the evolution of primates (Fooden, 1984), the presence of *P. falciparum*-like parasites throughout the African hominoid radiation (e.g. Duval et al., 2010), and the apparent restriction of this effect to the cytosolic

domain. It may be the case that the small number of changes are under selection, a pattern of selection that is particularly difficult to detect (Yang, 2006). Furthermore, testing among the relaxed constraint and adaptive hypotheses will require additional work, from a combination of epidemiological, population genetic, and functional studies. Particular examination of the inferred nine amino acid insertion in African apes appears promising given its unique phylogenetic pattern and properties, including an inferred alternate start codon. Interestingly, a recent case report found evidence for an altered initiation site in *SLC4A1* in a severely anemic individual, a mutation that leads to *P. falciparum* resistance *in vitro* (Perrotta et al., 2005). It is worth mentioning that the case for a relationship between this putatively adaptive evolution and *Plasmodium* will require additional functional work as a rigorous test, as it is possible that the evolution of this gene may be related to a wholly different selective pressure than *Plasmodium*.

Because it is restricted to humans and African apes, an adaptive interpretation supports the hypothesis that the SAO mutation of *SLC4A1* is targeting *P. falciparum* in humans. Furthermore, it suggests that African apes and humans may share some common mechanisms of adaption to *P. falciparum*-related parasites. This is especially interesting because examination of other human malarial adaptations in non-human primates has found mixed support for common adaptations between humans and other primates. At the *G6PD* and *β-globin* loci of chimpanzees there is no evidence for selection (MacFie et al., 2009; Verrelli et al., 2006) and in orangutans the evidence for selection at the *α-globin* locus is limited (Steiper et al., 2005, 2006), though there is evidence for selection at these loci in humans (Allison, 1954; Flint et al., 1986). Tung et al. (2009) recently examined the relationship between variants of the baboon *FY* locus (known to be important in human malarial resistance) and infection with *Hepatocystis* (a relative of *Plasmodium*). Using multiple lines of evidence, Tung et al. clearly showed that particular *FY* variants conferred resistance to *Hepatocystis*, revealing that some mechanisms of parasitic adaptation are broadly similar between humans and non-human primates (2009). Further studies examining the population genetics of *SLC4A1* in African apes and humans, including functional assessments of segregating variants and their relationships with particular *Plasmodium* parasites, will enable direct tests of the hypotheses forwarded here. Finally, the nine amino acid insertion inferred to have occurred in African apes, yet lost in humans, is worthy of special attention in future functional analyses of the primate *SLC4A1* gene.

The results presented here also bear on our understanding of the evolutionary history of the human Wright blood group antigen. In humans, the Wr antigen is formed when the 658th amino acid at SLC4A1 links to the 61st amino acid of the GYPA protein (arginine) (Bruce et al., 1995). The human Wr^a allele is exceptionally rare (e.g. Arriaga et al., 2005). Interestingly, the 658th amino acid was identified in multiple tests as evolving positively in *SLC4A1* across mammals. Furthermore, *GYPA* is one of the most rapidly evolving loci in the primate genome (Baum et al., 2005; Wang et al., 2003). The positive selection at this site linking *SLC4A1* to *GYPA* suggests that these genes may be evolving adaptively and in a correlated manner. Previous evolutionary work exploring the relationship between *GYPA* and *SLC4A1* in primates suggested that the specific amino acids bonding these proteins may differ across species (Huang et al., 1996). Because SLC4A1 amino acid 658 is evolving

positively across all mammals studied, it is likely that the correlated evolution is not related to *Plasmodium* but is a more ancient or generalized selective force. Although a recent study of the population genetics of *SLC4A1* in humans suggests the selective history of this gene may relate to anion transport (Wilder et al., 2009), the Wr antigens do not differ in this feature (Bruce et al., 1995). Future study will help to elucidate the relationship between the *SLC4A1* and *GYPA* genes across primate evolution and also determine the selection pressure acting on these loci.

5. Conclusions

These results show that there has been a differential selective regime operating on *SLC4A1* gene in humans and African apes relative to other primates and mammals. The effect is mainly found in the cytosolic domain of this protein. This finding has two interpretations. One interpretation is that the cytosolic domain of *SLC4A1* is experiencing relaxed selective constraint in humans and African apes. A second interpretation is that there have been a small number of adaptive evolutionary changes within the human and African ape lineages, potentially in response to *P. falciparum* and its relatives. The adaptive hypothesis suggests that these African hominoid primates have been evolving to combat *Plasmodium* since their last common ancestor. Future studies investigating the relationship between *SLC4A1* variants and *Plasmodium* in humans and wild African apes can further evaluate this hypothesis. Also of note is the finding that the amino acid in *SLC4A1* that encodes the human Wright blood group antigen is under positive selection across mammals. The Wright blood group antigen is little studied, and additional research is required to determine the potential evolutionary, functional, and clinical significance of this finding.

6. Database ID

HM065568, HM065581, HM065569, HM065570, HM065571, HM065572, HM065580, HM065573, HM065573, HM065575, HM065576, HM065577, HM065578, HM065579 (all at Genbank, the Genetic sequence database at the National Center for Biotechnical Information (NCBI)).

Acknowledgments

The authors thank A. Lobell, W. Qiu, and the reviewers for comments on the manuscript and Z. Yang for statistical advice. The infrastructure of the Anthropological Genetics Lab at Hunter College was supported by Grant Number RR03037 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH.

References

Allen SJ, O'Donnell A, Alexander ND, Mgone CS, Peto TE, Clegg JB, Alpers MP, Weatherall DJ. Prevention of cerebral malaria in children in Papua New Guinea by southeast Asian ovalocytosis band 3. Am J Trop Med Hyg. 1999; 60:1056–1060. [PubMed: 10403343]

Allison AC. Protection afforded by sickle-cell trait against subtertian malarial infection. Br Med J. 1954; 1:290–294. [PubMed: 13115700]

Anisimova M, Yang Z. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol Biol Evol. 2007; 24:1219–1228. [PubMed: 17339634]

Arriaga F, Llopis F, de la Rubia J, Carpio N, Moscardo J, Marty ML. Incidence of Wra antigen and anti-Wra in a Spanish population. Transfusion. 2005; 45:1324–1326. [PubMed: 16078920]

Baer A, Lie-Injo LE, Welch QB, Lewis AN. Genetic factors and malaria in the Temuan. Am J Hum Genet. 1976; 28:179–188. [PubMed: 817597]

Baum J, Maier AG, Good RT, Simpson KM, Cowman AF. Invasion by *P. falciparum* merozoites suggests a hierarchy of molecular interactions. PLoS Pathog. 2005; 1:e37. [PubMed: 16362075]

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2009; 37:D26–31. [PubMed: 18940867]

Bininda-Emonds OR. TransAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. BMC Bioinf. 2005; 6:156.

Bruce L. Mutations in band 3 and cation leaky red cells. Blood Cells Mol Dis. 2006; 36:331–336. [PubMed: 16531080]

Bruce LJ, Ring SM, Anstee DJ, Reid ME, Wilkinson S, Tanner MJ. Changes in the blood group Wright antigens are associated with a mutation at amino acid 658 in human erythrocyte band 3: a site of interaction between band 3 and glycophorin A under certain conditions. Blood. 1995; 85:541–547. [PubMed: 7812009]

Bruce LJ, Ring SM, Ridgwell K, Reardon DM, Seymour CA, Van Dort HM, Low PS, Tanner MJ. South-East Asian ovalocytic (SAO) erythrocytes have a cold sensitive cation leak: implications for in vitro studies on stored SAO red cells. Biochim Biophys Acta. 1999; 1416:258–270. [PubMed: 9889381]

Cattani JA, Gibson FD, Alpers MP, Crane GG. Hereditary ovalocytosis and reduced susceptibility to malaria in Papua New Guinea. Trans R Soc Trop Med Hyg. 1987; 81:705–709. [PubMed: 3329776]

Coatney, GR.; Collines, WE.; Warren, M.; Contacos, PG. The Primate Malarias. US Department of Health, Education, and Welfare; Bethesda: 1971.

Cortés A, Benet A, Cooke BM, Barnwell JW, Reeder JC. Ability of *Plasmodium falciparum* to invade Southeast Asian ovalocytes varies between parasite lines. Blood. 2004; 104:2961–2966. [PubMed: 15265796]

Cortés A, Mellombo M, Mgone CS, Beck HP, Reeder JC, Cooke BM. Adhesion of *Plasmodium falciparum*-infected red blood cells to CD36 under flow is enhanced by the cerebral malaria-protective trait South-East Asian ovalocytosis. Mol Biochem Parasitol. 2005; 142:252–257. [PubMed: 15978955]

Dluzewski AR, Nash GB, Wilson RJ, Reardon DM, Gratzer WB. Invasion of hereditary ovalocytes by *Plasmodium falciparum* in vitro and its relation to intracellular ATP concentration. Mol Biochem Parasitol. 1992; 55:1–7. [PubMed: 1435863]

Duval L, Fourment M, Nerrienet E, Rousset D, Sadeuh SA, Goodman SM, Andriaholinirina NV, Randrianarivelojosia M, Paul RE, Robert V, Ayala FJ, Ariey F. African apes as reservoirs of *Plasmodium falciparum* and the origin and diversification of the Laverania subgenus. Proc Natl Acad Sci USA. 2010; 107:10561–10566. [PubMed: 20498054]

Duval L, Nerrienet E, Rousset D, Sadeuh MbaSA, Houze S, Fourment M, Le Bras J, Robert V, Ariey F. Chimpanzee malaria parasites related to *Plasmodium ovale* in Africa. PLoS One. 2009; 4:e5520. [PubMed: 19436742]

Ellegren H. Comparative genomics and the study of evolution by natural selection. Mol Ecol. 2008; 17:4586–4596. [PubMed: 19140982]

Escalante AA, Barrio E, Ayala FJ. Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. Mol Biol Evol. 1995; 12:616–626. [PubMed: 7659017]

Flint J, Hill AVS, Bowden DK, Oppenheimer SJ, Sill PR, Serjeantson SW, Bana-Koiri J, Bhatia K, Alpers MP, Boyce AJ, et al. High frequencies of α-thalassaemia are the result of natural selection by malaria. Nature. 1986; 321:744– 750. [PubMed: 3713863]

Foo LC, Rekhraj V, Chiang GL, Mak JW. Ovalocytosis protects against severe malaria parasitemia in the Malayan aborigines. Am J Trop Med Hyg. 1992; 47:271–275. [PubMed: 1524139]

Fooden J. Malaria in macaques. Intl J Primatol. 1984; 15:573–596.

Genton B, al-Yaman F, Mgone CS, Alexander N, Paniu MM, Alpers MP, Mokela D. Ovalocytosis and cerebral malaria. Nature. 1995; 378:564–565. [PubMed: 8524388]

Goel VK, Li X, Chen H, Liu SC, Chishti AH, Oh SS. Band 3 is a host receptor binding merozoite surface protein 1 during the *Plasmodium falciparum* invasion of erythrocytes. Proc Natl Acad Sci USA. 2003; 100:5164–5169. [PubMed: 12692305]

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol. 1998; 9:585–598. [PubMed: 9668008]

Hadley T, Saul A, Lamont G, Hudson DE, Miller LH, Kidson C. Resistance of Melanesian elliptocytes (ovalocytes) to invasion by *Plasmodium knowlesi* and *Plasmodium falciparum* malaria parasites in vitro. J Clin Invest. 1983; 71:780–782. [PubMed: 6338046]

Hayakawa T, Arisue N, Udono T, Hirai H, Sattabongkot J, Toyama T, Tsuboi T, Horii T, Tanabe K. Identification of *Plasmodium malariae*, a human malaria parasite, in imported chimpanzees. PLoS One. 2009; 4:e7412. [PubMed: 19823579]

Hayakawa T, Culleton R, Otani H, Horii T, Tanabe K. Big bang in the evolution of extant malaria parasites. Mol Biol Evol. 2008; 25:2233–2239. [PubMed: 18687771]

Huang CH, Reid ME, Xie SS, Blumenfeld OO. Human red blood cell Wright antigens: a genetic and evolutionary perspective on glycophorin A-band 3 interaction. Blood. 1996; 87:3942–3947. [PubMed: 8611724]

Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L, Coates G, Fairley S, Fitzgerald S, Fernandez-Banet J, Gordon L, Graf S, Haider S, Hammond M, Holland R, Howe K, Jenkinson A, Johnson N, Kahari A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E, Lawson D, Longden I, Megy K, Meidl P, Overduin B, Parker A, Pritchard B, Rios D, Schuster M, Slater G, Smedley D, Spooner W, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wilder S, Zadissa A, Birney E, Cunningham F, Curwen V, Durbin R, Fernandez-Suarez XM, Herrero J, Kasprzyk A, Proctor G, Smith J, Searle S, Flicek P. Ensembl 2009. Nucleic Acids Res. 2009; 37:D690–697. [PubMed: 19033362]

Jarolim P, Palek J, Amato D, Hassan K, Sapak P, Nurse GT, Rubin HL, Zhai S, Sahr KE, Liu SC. Deletion in erythrocyte band 3 gene in malaria-resistant Southeast Asian ovalocytosis. Proc Natl Acad Sci USA. 1991; 88:11022– 11026. [PubMed: 1722314]

Jay DG. Role of band 3 in homeostasis and cell shape. Cell. 1996; 86:853–854. [PubMed: 8808620]

Kariuki MM, Li X, Yamodo I, Chishti AH, Oh SS. Two *Plasmodium falciparum* merozoite proteins binding to erythrocyte band 3 form a direct complex. Biochem Biophys Res Commun. 2005; 338:1690–1695. [PubMed: 16289042]

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002; 12:996–1006. [PubMed: 12045153]

Kidson C, Lamont G, Saul A, Nurse GT. Ovalocytic erythrocytes from Melanesians are resistant to invasion by malaria parasites in culture. Proc Natl Acad Sci USA. 1981; 78:5829–5832. [PubMed: 7029547]

Kochar DK, Saxena V, Singh N, Kochar SK, Kumar SV, Das A. *Plasmodium vivax* malaria. Emerg Infect Dis. 2005; 11:132–134. [PubMed: 15705338]

Krief S, Escalante AA, Pacheco MA, Mugisha L, Andre C, Halbwax M, Fischer A, Krief JM, Kasenene JM, Crandfield M, Cornejo OE, Chavatte JM, Lin C, Letourneur F, Gruner AC, McCutchan TF, Renia L, Snounou G. On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from Bonobos. PLoS Pathog. 2010; 6:e1000765. [PubMed: 20169187]

Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. Am J Hum Genet. 2005; 77:171–192. [PubMed: 16001361]

Li X, Chen H, Bahamontes-Rosa N, Kun JF, Traore B, Crompton PD, Chishti AH. *Plasmodium falciparum* signal peptide peptidase is a promising drug target against blood stage malaria. Biochem Biophys Res Commun. 2009; 380:454– 459. [PubMed: 19174148]

Li X, Chen H, Oh SS, Chishti AH. A Presenilin-like protease associated with *Plasmodium falciparum* micronemes is involved in erythrocyte invasion. Mol Biochem Parasitol. 2008; 158:22–31. [PubMed: 18160114]

Li X, Chen H, Oo TH, Daly TM, Bergman LW, Liu SC, Chishti AH, Oh SS. A co-ligand complex anchors *Plasmodium falciparum* merozoites to the erythrocyte invasion receptor band 3. J Biol Chem. 2004; 279:5765–5771. [PubMed: 14630931]

Liu SC, Jarolim P, Rubin HL, Palek J, Amato D, Hassan K, Zaik M, Sapak P. The homozygous state for the band 3 protein mutation in Southeast Asian Ovalocytosis may be lethal. Blood. 1994; 84:3590–3591. [PubMed: 7949112]

Liu SC, Zhai S, Palek J, Golan DE, Amato D, Hassan K, Nurse GT, Babona D, Coetzer T, Jarolim P, et al. Molecular defect of the band 3 protein in southeast Asian ovalocytosis. N Engl J Med. 1990; 323:1530–1538. [PubMed: 2146504]

Liu W, Li Y, Learn GH, Rudicell RS, Robertson JD, Keele BF, Ndjango JB, Sanz CM, Morgan DB, Locatelli S, Gonder MK, Kranzusch PJ, Walsh PD, Delaporte E, Mpoudi-Ngole E, Georgiev AV, Muller MN, Shaw GM, Peeters M, Sharp PM, Rayner JC, Hahn BH. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature. 2010; 467:420–425. [PubMed: 20864995]

MacFie TS, Nerrienet E, Bontrop RE, Mundy NI. The action of *falciparum* malaria on the human and chimpanzee genomes compared: absence of evidence for a genomic signature of malaria at *HBB* and *G6PD* in three subspecies of chimpanzee. Infect Genet Evol. 2009; 9:1248–1252. [PubMed: 19631293]

Mendis K, Sina BJ, Marchesini P, Carter R. The neglected burden of *Plasmodium vivax* malaria. Am J Trop Med Hyg. 2001; 64:97–106. [PubMed: 11425182]

Mgone CS, Koki G, Paniu MM, Kono J, Bhatia KK, Genton B, Alexander ND, Alpers MP. Occurrence of the erythrocyte band 3 (AE1) gene deletion in relation to malaria endemicity in Papua New Guinea. Trans R Soc Trop Med Hyg. 1996; 90:228–231. [PubMed: 8758056]

Mohandas N, Winardi R, Knowles D, Leung A, Parra M, George E, Conboy J, Chasis J. Molecular basis for membrane rigidity of hereditary ovalocytosis. A novel mechanism involving the cytoplasmic domain of band 3. J Clin Invest. 1992; 89:686–692. [PubMed: 1737855]

Murphy WJ, Pevzner PA, O'Brien SJ. Mammalian phylogenomics comes of age. Trends Genet. 2004; 20:631–639. [PubMed: 15522459]

Okoye VC, Bennett V. *Plasmodium falciparum* malaria: band 3 as a possible receptor during invasion of human erythrocytes. Science. 1985; 227:169–171. [PubMed: 3880920]

Okoyeh JN, Pillai CR, Chitnis CE. *Plasmodium falciparum* field isolates commonly use erythrocyte invasion pathways that are independent of sialic acid residues of glycophorin A. Infect. Immun. 1999; 67:5784–5791.

Ollomo B, Durand P, Prugnolle F, Douzery E, Arnathau C, Nkoghe D, Leroy E, Renaud F. A new malaria agent in African hominids. PLoS Pathog. 2009; 5:e1000446. [PubMed: 19478877]

Pasvol G, Wainscoat JS, Weatherall DJ. Erythrocytes deficiency in glycophorin resist invasion by the malarial parasite *Plasmodium falciparum*. Nature. 1982; 297:64–66. [PubMed: 7040988]

Perkins SL, Schall JJ. A molecular phylogeny of malarial parasites recovered from cytochrome *b* gene sequences. J Parasitol. 2002; 88:972–978. [PubMed: 12435139]

Perrotta S, Borriello A, Scaloni A, De Franceschi L, Brunati AM, Turrini F, Nigro V, del Giudice EM, Nobili B, Conte ML, Rossi F, Iolascon A, Donella-Deana A, Zappia V, Poggi V, Anong W, Low P, Mohandas N, Della Ragione F. The N-terminal 11 amino acids of human erythrocyte band 3 are critical for aldolase binding and protein phosphorylation: implications for band 3 function. Blood. 2005; 106:4359–4366. [PubMed: 16118313]

Prugnolle F, Durand P, Neel C, Ollomo B, Ayala FJ, Arnathau C, Etienne L, Mpoudi-Ngole E, Nkoghe D, Leroy E, Delaporte E, Peeters M, Renaud F. African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. Proc Natl Acad Sci USA. 2010; 107:1458–1463. [PubMed: 20133889]

Rich SM, Leendertz FH, Xu G, Lebreton M, Djoko CF, Aminake MN, Takang EE, Diffo JL, Pike BL, Rosenthal BM, Formenty P, Boesch C, Ayala FJ, Wolfe ND. The origin of malignant malaria. Proc Natl Acad Sci USA. 2009

Rogerson SJ, Carter R. Severe vivax malaria: newly recognised or rediscovered. PLoS Med. 2008; 5:e136. [PubMed: 18563965]

Ruvolo M. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. Mol Biol Evol. 1997; 14:248–265. [PubMed: 9066793]

Serjeantson S, Bryson K, Amato D, Babona D. Malaria and hereditary ovalocytosis. Hum Genet. 1977; 37:161–167. [PubMed: 328370]

Sim BK, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH. Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. Science. 1994; 264:1941–1944. [PubMed: 8009226]

Steiper ME, Wolfe ND, Karesh WB, Kilbourn AM, Bosi EJ, Ruvolo M. The population genetics of the alpha-2 globin locus of orangutans (*Pongo pygmaeus*). J Mol Evol. 2005; 60:400–408. [PubMed: 15871050]

Steiper ME, Wolfe ND, Karesh WB, Kilbourn AM, Bosi EJ, Ruvolo M. The phylogenetic and evolutionary history of a novel alpha-globin-type gene in orangutans (*Pongo pygmaeus*). Infect Genet Evol. 2006; 6:277–286. [PubMed: 16172024]

Tanner MJ. Molecular and cellular biology of the erythrocyte anion exchanger (AE1). Semin Hematol. 1993; 30:34–57. [PubMed: 8434259]

Tung J, Primus A, Bouley AJ, Severson TF, Alberts SC, Wray GA. Evolution of a malaria resistance gene in wild primates. Nature. 2009; 460:388–391. [PubMed: 19553936]

Verrelli BC, Tishkoff SA, Stone AC, Touchman JW. Contrasting histories of *G6PD* molecular evolution and malarial resistance in humans and chimpanzees. Mol Biol Evol. 2006; 23:1592–1601. [PubMed: 16751255]

Wang HY, Tang H, Shen CK, Wu CI. Rapidly evolving genes in human. I The glycophorins and their possible role in evading malaria parasites. Mol Biol Evol. 2003; 20:1795–1804. [PubMed: 12949139]

WHO. World Malaria Report. World Health Organization; Geneva: 2008.

Wilder JA, Stone JA, Preston EG, Finn LE, Ratcliffe HL, Sudoyo H. Molecular population genetics of *SLC4A1* and Southeast Asian Ovalocytosis. J Hum Genet. 2009; 54:182–187. [PubMed: 19229254]

Wildman DE, Jameson NM, Opazo JC, Yi SV. A fully resolved genus level phylogeny of neotropical primates (Platyrrhini). Mol Phylogenet Evol. 2009; 53:694– 702. [PubMed: 19632342]

Williamson RC, Toye AM. Glycophorin A: band 3 aid. Blood Cells Mol Dis. 2008; 41:35–43. [PubMed: 18304844]

Wolfe, ND. PhD thesis. Harvard School of Public Health; Cambridge, MA: 1999. Pathogen evolution and exchange in Bornean orangutans.

Yang, Z. Computational Molecular Evolution. Oxford University Press; Oxford: 2006.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007; 24:1586–1591. [PubMed: 17483113]

Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005; 22:1107–1118. [PubMed: 15689528]

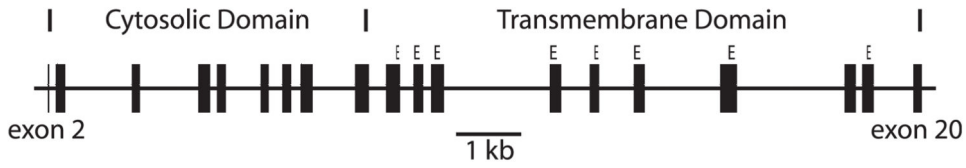

Fig. 1.

Schematic organization of the human *SLC4A1* gene, showing all translated exons (2–20). The cytosolic and transmembrane domains are shown. The external residues of the transmembrane are indicated with the letter “E.” To enhance legibility, these “E” notations approximate the codon lengths of the external residues but are not exact.

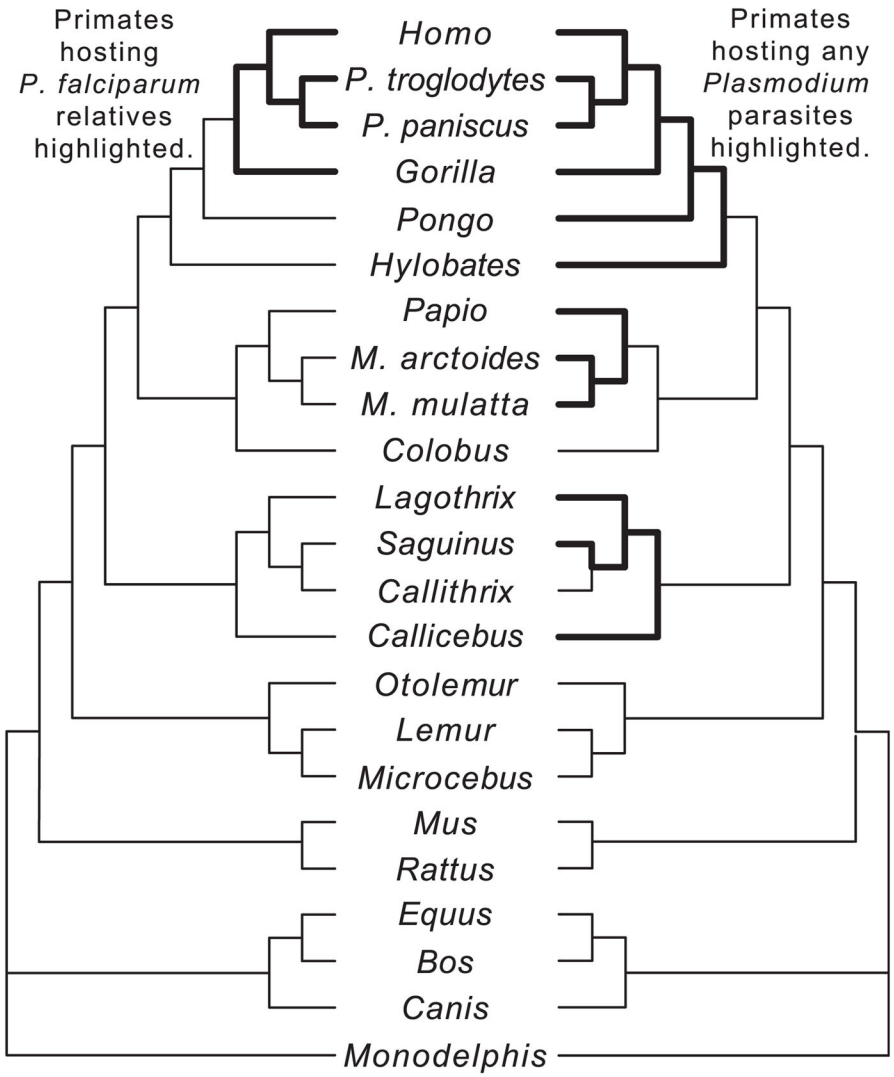

Fig. 2.

Phylogeny of the species examined in the present study. To the left of the species names, the phylogenetic tree highlights the lineages inferred to have harbored *P. falciparum* and related parasites (see Table 1). To the right of the species names, the phylogenetic tree highlights the lineages inferred to have harbored any species of *Plasmodium*, except for rodents (see Section 2 and Table 1). These bold branches are those used by *PAML* to enable the branch tests and branch site tests as described in the Methods.

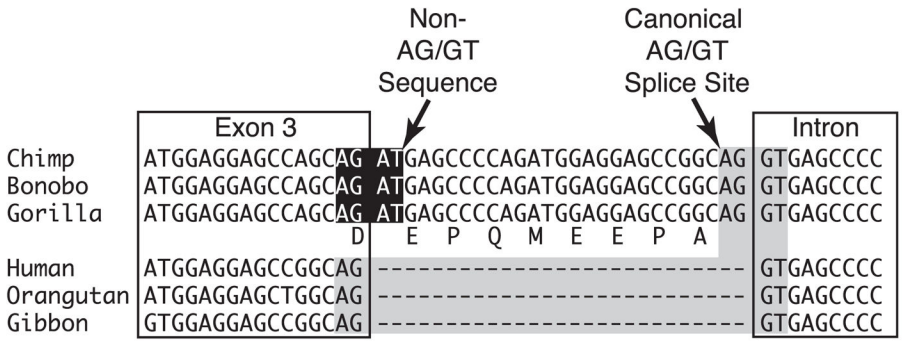

Fig. 3.

Detail of the insertion at the 3′ end of exon 3 that is reconstructed to have occurred in African Apes. There is clear homology among the sites in exon 3, the insertion, and following intron. The canonical splice sites are highlighted in gray; the obliterated splice site is highlighted in black. The in frame inferred amino acid sequence of the insertion is also given.

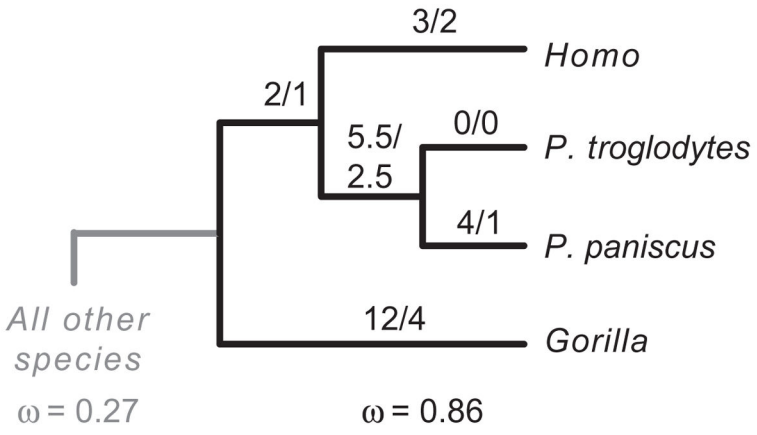

Fig. 4.

Summary of the significant branch test for the cytosolic domain. This simplified phylogeny shows the two branch classes (species harboring relatives of *P. falciparum* vs. all other species) and their respective estimated ω values for the cytosolic domain. The *codeml* program was used (with Rateancestor = 1) to infer the number of nonsynonymous (numerator) and synonymous (denominator) along each branch. These estimates are shown above each branch. Gapped regions were not included in these estimates.

Infect Genet Evol. Author manuscript; available in PMC 2014 June 30.

Table 1

Species examined, sample information, accession numbers, and *Plasmodium* infection status. (See below-mentioned references for further information.)

Host Group	Host Species	Common Name	Coriell Sample ID	Accession Number	Genome Project	Genome Sequence Coordinates	Infections from *P. falciparum* and similar spp.	Other infecting *Plasmodium* spp.
Humans	*Homo sapiens*	Human			NG_007498.1 †		*P. falciparum*	*P. vivax*, *P. malariae*, *P. ovale*
African Apes								
	Pan paniscus	Bonobo	NG05253	**HM065568**			*P. falciparum* (Krief *et al*. 2010)	Spp. related to *P. malariae* (Krief *et al*. 2010)
	Pan troglodytes	Chimpanzee	NG06939	**HM065581**	panTro2	Chr 17:13296154-13309268	*P. reichenowi*, *P. gaboni*, unnamed parasite spp. (Krief *et al*. 2010; Ollomo *et al*. 2009; Rich *et al*. 2009; Prugnolle *et al*. 2010)	*P. rodhaini*, *P. schwetzi*, unnamed parasite spp. (similar to *P. malariae*, *P. vivax* and/or *P. ovale*, Coatney 1971, Duval *et al*. 2009, Krief *et al*. 2010, Hayakawa et al 2009)
	Gorilla gorilla	Gorilla	NG05251	**HM065569**			*P. falciparum*, unnamed parasite spp. (Prugnolle *et al*. 2010)	*P. rodhaini*, *P. schwetzi* (similar to *P. malariae*, *P. vivax* and/or *P. ovale*, Coatney 1971)
Asian Apes								
	Pongo pygmaeus	Orangutan	NA04272	**HM065570**			none	*P. silvaticum*, *P. pitheci* (similar to *P. vivax*, Wolfe 1999)
	Hylobates gabriellae	Yellow Cheeked Gibbon	PR00381	**HM065571**			none	*P. hylobati* (similar to *P. vivax*, Perkins and

Host Group	Host Species	Common Name	Coriell Sample ID	Accession Number	Genome Project	Genome Sequence Coordinates	Infections from *P. falciparum* and similar spp.	Other infecting *Plasmodium* spp.
								Schall 2002, Hayakawa *et al* 2008)
Old World Monkeys								
	Macaca arctoides	Stump Tailed Macaque	GM03443	**HM065572**			none	Multiple spp. (similar to *P. vivax*, Perkins and Schall 2002, Hayakawa *et al*. 2008)
	Macaca mulatta	Rhesus Macaque	NG07109	**HM065580**	rheMac2	Chr 16:54355363-54341551	none	Multiple spp. (similar to *P. vivax*, Perkins and Schall 2002, Hayakawa *et al*. 2008)
	Colobus guereza	Black and White Colobus	PR00980	**HM065573**			none	none
	Papio anubis	Baboon	PR00036	**HM065574**			none	*Hepatocystis* (silimar to *P. ovale*, Perkins and Schall 2002)
New World Monkeys	*Callithrix jacchus*	Marmoset			calJac1	Contig103:955511- 939806	none	none
	Callicebus moloch	Dusky Titi Monkey	NG06115	**HM065575**			none	*P. brasilanum* (similar to *P. malariae*, Escalante *et al*. 1995)
	Lagothrix lagothricha	Wooly Monkey	NG05356	**HM065576**			none	”
	Saguinus labiatus	White Lipped Tamarin	NG05308	**HM065577**			none	”
Strepsirhines	*Otolemur garnettii*	Small Eared Galago	PR00049	**HM065578**			none	none
	Lemur catta	Ring Tailed Lemur	ID#6351*	**HM065579**			none	none
	Microcebus murinus	Gray Mouse Lemur			ensembl	Genescaffold1055:63426-52166	none	none
Rodents	*Mus musculus*	Mouse			mm9	Chr 11:102222709 102211601	none	Murinae infected with *Plasdmodium*

Host Group	Host Species	Common Name	Coriell Sample ID	Accession Number	Genome Project	Genome Sequence Coordinates	Infections from *P. falciparum* and similar spp.	Other infecting *Plasmodium* spp.
								that is outgroup to a that is outgroup to a a that is outgroup to a a la *P. vivax, P. malariae, I* *P. vivax, P. malariae, F* *P. vivax, P. malariae, F* *P. vivax, P. malariae,* group, Perkins and Schall 2002, Hayakawa et al. 2008)
	Rattus norvegicus	Rat			rn4	Chr 10:91467178-91456310	none	”
Other Mammals	*Bos taurus*	Cow			bosTau4	Chr 19:45496100-45485259	none	none
	Canis familiaris	Dog			canFam2	Chr 9:22463842-22475102	none	none
	Equus caballus	Horse			equCab2	Chr 11:19271365-19283645	none	none
	Monodelphis domestica	Opossum			monDom5	Chr 2:204285378-204264654	none	none

Table 2

Branch tests of adaptive evolution.

Model	*N* parameters	All sites		Cytosolic sites		Transmembrane sites		External sites	
		ln*L*	*ω*	ln*L*	*ω*	ln*L*	*ω*	ln*L*	*ω*
A									
(A) $\omega = 1$	44	−18873.419	1.00	−9687.625	1.00	−9010.701	1.00	−2581.484	1.00
(B) Estimate one ω	45	−18039.469	0.19	−9412.785	0.28	−8413.696	0.12	−2462.885	0.20
(C) *P. falciparum* branches vs. others	46	−18036.433	0.37, 0.19	−9407.341	0.86, 0.27	−8413.530	0.09, 0.12	−2462.808	0.14, 0.209
(D) *Plasmodium* branches vs. others	46	−18039.340	0.21, 0.19	−9410.411	0.38, 0.26	−8412.150	0.09, 0.13	−2462.599	0.15, 0.20

Likelihood ratio tests	All sites		Cytosolic sites		Transmembrane sites		External sites	
	2Δln*L*	*P* value	2Δln*L*	*P* value	2Δln*L*	*P* value	2Δln*L*	*P* value
B								
Neutral evolution (A vs B (df = 1))	1667.9	**0.000**	549.7	**0.000**	1194.0	**0.000**	237.2	**0.000**
P. falciparum branches different (C vs B (df = 1))[a]	6.07	**0.014**[b]	10.9	**0.001**[b]	0.3	0.565	0.2	0.694
Plasmodium branches different (D vs B (df = 1))[a]	0.3	0.611	4.7	0.029[c]	3.1	0.079	0.6	0.449

A: parameters and likelihoods estimated for the different models.

B: likelihood ratio tests described in Methods and Results. Bold indicates $P < 0.05$.

[a] A correction for multiple tests is used for C vs. B and D vs. B tests.

[b] Significant after Bonferroni correction for multiple tests.

[c] Not significant after correction for multiple tests.

Table 3

Site tests of adaptive evolution, fixed class model.

Model	*N* parameters	All sites		Cytosolic sites		Transmembrane sites		External sites	
		ln *L*	ω Values	ln*L*	ω Values	ln*L*	ω Values	ln*L*	ω Values
A									
(A) 'M0', one ω class	45	−18039.469	0.19	−9412.785	0.28	−8413.696	0.12	−2462.885	0.20
(B) 'M1a', two ω classes (<1,1)	46	−17475.971	0.07, 1.00	−9178.578	0.10, 1.00	−8172.866	0.05, 1.00	−2367.630	0.07, 1.00
(C) 'M2a', three ω classes (<1, 1,>1)	48	−17469.195	0.07, 1.00, 2.86	−9178.578	0.10, 1.00, 1.00	−8166.515	0.05, 1.00, 3.12	−2364.868	0.07, 1.00, 2.88

Likelihood ratio tests	All sites			Cytosolic sites			Transmembrane sites			External sites		
	2Δln*L*	*P* value	Selected sites	2Δln*L*	*P* value	Selected sites	2Δln*L*	*P* value	Selected sites	2Δln*L*	*P* value	Selected sites
B												
Two ω classes vs. One ω class (A vs. B (df = 1))	1127	**0**	n/a	468.4	**0**	n/a	481.7	**0**	n/a	190.5	**0**	n/a
Two ω classes vs. Three ω classes (B vs. C (df = 2))	13.6	**0.001**	367-1[a], 658	0	1	n/a	12.7	**0.002**	658	5.5	0.063	n/a

A: parameters and likelihoods estimated for the different models.

B: likelihood ratio tests described in Sections 2 and 3. Bold indicates $P < 0.05$.

[a]The position under selection is a gap in the human alignment corresponding to the codon before the 367th codon in human.

Table 4

Site tests of adaptive evolution, beta distribution model.

Model	*N* parameters	All sites		Cytosolic sites		Transmembrane sites		External sites	
		ln*L*	ω of selected class	ln*L*	ω of selected class	ln*L*	ω of selected class	ln*L*	ω of selected class
A									
(A) 'M7', ω beta distribution	46	−17452.735		−9173.066		−8159.716		−2358.899	
(B) 'M8', ω beta distribution with ω class < 1	48	−17424.343	1.59	−9164.167	1.25	−8134.548	2.04	−2352.06	1.79

Likelihood ratio tests	All sites			Cytosolic sites			Transmembrane sites			External sites		
	2Δln*L*	*P* value	Selected sites	2Δln*L*	*P* value	Selected sites	2Δln*L*	*P* value	Selected sites	2Δln*L*	*P* value	Selected sites
B												
Beta dist. For ω vs. Beta with a site class with ω < 1 (A vs. B (df = 2))	56.8	**0**	658	17.8	**0**	None	50.3	**0**	558, 560, 656, 658	13.7	**0**	658

These site tests are based on a beta distribution for site classes for ω values for each site. A: parameters and likelihoods estimated for the different models. B: likelihood ratio tests described in Sections 2 and 3. Bold indicates $P < 0.05$.

Table 5

Branch-site tests of adaptive evolution.

	Null model		Selection model		Likelihood ratio tests
	N **parameters**	**ln*L***	***N*** **parameters**	**ln*L***	***P*** **value**
A					
All	47	−17473.9418	48	−17472.7279	0.119
Cytosolic sites	47	−9174.2083	48	−9172.7565	0.088[a]
Transmembrane sites	47	−8172.8663	48	−8172.8663	1.000
External sites	47	−2367.4868	48	−2367.4868	1.000
B					
All	47	−17475.2612	48	−17473.9605	0.107
Cytosolic sites	47	−9176.0386	48	−9175.5231	0.310
Transmembrane sites	47	−8172.8663	48	−8172.4822	0.381
External sites	47	−2367.6301	48	−2367.6301	1.000

A: tests on lineages harboring *P. falciparum* and related parasites.

B: tests on lineages harboring any *plasmodium* parasites.

[a]See text for additional information regarding this comparison.