

Molecular evolution of a malaria resistance gene (*DARC*) in primates

Thiago Yukio Kikuchi Oliveira · Eugene E. Harris ·
Diogo Meyer · Chong K. Jue · Wilson Araújo Silva Jr

Received: 17 October 2011 / Accepted: 8 February 2012
© Springer-Verlag 2012

Abstract Genes involved in host-pathogen interactions are often strongly affected by positive natural selection. The Duffy antigen, coded by the Duffy antigen receptor for chemokines (*DARC*) gene, serves as a receptor for *Plasmodium vivax* in humans and for *Plasmodium knowlesi* in some nonhuman primates. In the majority of sub-Saharan Africans, a nucleic acid variant in GATA-1 of the gene promoter is responsible for the nonexpression of the Duffy antigen on red blood cells and consequently resistance to invasion by *P. vivax*. The Duffy antigen also acts as a receptor for chemokines and is expressed in red blood cells and many other tissues of the body. Because of this dual role, we sequenced a ~3,000-bp region encompassing the entire *DARC* gene as well as part of its 5' and 3' flanking regions in a phylogenetic sample of primates and used statistical methods to evaluate the

nature of selection pressures acting on the gene during its evolution. We analyzed both coding and regulatory regions of the *DARC* gene. The regulatory analysis showed accelerated rates of substitution at several sites near known motifs. Our tests of positive selection in the coding region using maximum likelihood by branch sites and maximum likelihood by codon sites did not yield statistically significant evidence for the action of positive selection. However, the maximum likelihood test in which the gene was subdivided into different structural regions showed that the known binding region for *P. vivax/P. knowlesi* is under very different selective pressures than the remainder of the gene. In fact, most of the gene appears to be under strong purifying selection, but this is not evident in the binding region. We suggest that the binding region is under the influence of two opposing selective pressures, positive selection possibly exerted by the parasite and purifying selection exerted by chemokines.

Electronic supplementary material The online version of this article (doi:10.1007/s00251-012-0608-2) contains supplementary material, which is available to authorized users.

T. Y. K. Oliveira
Laboratory of Molecular Immunology, The Rockefeller University,
New York, NY 10065, USA

T. Y. K. Oliveira (✉) · W. A. Silva Jr
Departamento de Genética-Bloco H,
Faculdade de Medicina de Ribeirão Preto, University of São Paulo,
Av. Bandeirantes, 3900-Monte Alegre,
Ribeirão Preto, Brazil
e-mail: stratus@lgmb.fmrp.usp.br

E. E. Harris · C. K. Jue
Department of Biological Sciences and Geology, Queensborough
Community College, City University of New York,
New York, NY11364, USA

D. Meyer
Departamento de Genética e Biologia Evolutiva,
Universidade de São Paulo,
São Paulo, Brazil

Keywords Positive selection · Duffy · Coevolution · Antigen · Erythrocyte · *Plasmodium vivax*

Introduction

The Duffy antigen receptor for chemokines (*DARC*) gene encodes the seven-transmembrane glycoprotein Duffy antigen (*Fy*) expressed on cell surfaces in diverse tissues of the body including endothelial cells, brain, and erythrocytes (Hadley and Peiper 1997). The Duffy protein was first discovered to be responsible for the polymorphic human antigens *Fy*^a and *Fy*^b. Furthermore, it was discovered that individuals lacking the Duffy antigen (*Fy*⁻/*Fy*⁻) are resistant to invasion by the human malarial parasite *Plasmodium vivax* (Miller et al. 1976) and the simian malarial parasite *Plasmodium knowlesi* (Miller et al. 1975). *P. knowlesi* is geographically distributed in Southeast Asia and is known

to naturally infect some monkey species within the genus *Macaca* (i.e., the macaques). *P. knowlesi* as well as other species of *Plasmodium* are believed to have been persistent parasites of macaque monkeys during their evolution (Escalante et al. 2005).

In humans, the absence of the Duffy antigen is nearly fixed in sub-Saharan African populations, yet it is present almost universally in all other world populations (Cavalli-Sforza et al. 1994). The absence of the antigen is due to the transition from T to C within the GATA promoter specific for the expression of the antigen on erythrocytes. Recent analyses of single nucleotide polymorphism within and between human populations strongly suggest that the null genotype (in which the antigen is not expressed) experienced a recent selective sweep in sub-Saharan African populations (Hamblin and Di Rienzo 2000; Hamblin et al. 2002) and is one of the strongest cases of human population-level adaptation (Harris and Meyer 2006).

In monkeys, results from early experimental studies (Butcher et al. 2010; Schmidt et al. 1977) have strongly suggested that some macaque species, for example, *Macaca fascicularis* (the crab-eating macaque) and possibly *Macaca nemestrina* (the pig-tailed macaque) and *Macaca nigra* (Celebes crested macaque), who are natural hosts for *P. knowlesi*, are able to control infection by *P. knowlesi* and turn it into a chronic infection. These macaque species are either wholly or partially distributed across the Southeast Asian islands where *P. knowlesi* is also distributed. In contrast, other macaque species, with geographic distributions outside the range of *P. knowlesi*, for example, *Macaca mulatta* in mainland India and China, show virtually no ability to control infection to *P. knowlesi* and usually display a fulminant infection that quickly leads to death (Butcher 1996; Butcher et al 2010). Interestingly, in contrast to island population of *M. fascicularis*, mainland populations of *M. fascicularis* apparently have little to no ability to control infection by *P. knowlesi* and are similar to *M. mulatta* in this respect (Butcher et al. 2010; Fooden 1994; Schmidt et al. 1977). The possible molecular basis accounting for the differential susceptibility to *P. knowlesi* in different macaque species is unknown.

The *DARC* gene also serves as a chemokine receptor binding C-C and C-X-C families; however, its physiological significance is unclear (Hadley and Peiper 1997). Tournamille et al. (2004) found that *DARC* in all the nonhuman primates he analyzed bound the chemokines CXCL-8 and CCL-5. Nonetheless, because of the Duffy antigen's dual functions, associated with invasion of erythrocytes by *P. vivax/P. knowlesi* and in binding chemokines, Tournamille et al. (2004) suggested that *DARC* may be under dual selective pressures: an internal pressure acting to preserve chemokine binding function (i.e., purifying selection) and an external selective pressure due to *Plasmodium* infection that could drive adaptive changes that provide resistance (i.e., positive selection). Studies of the

Duffy binding protein in the *P. vivax* parasite have demonstrated repeated positive selection in the past strongly suggesting adaptive coevolutionary interactions with the *DARC* receptor on erythrocytes (Baum et al 2003).

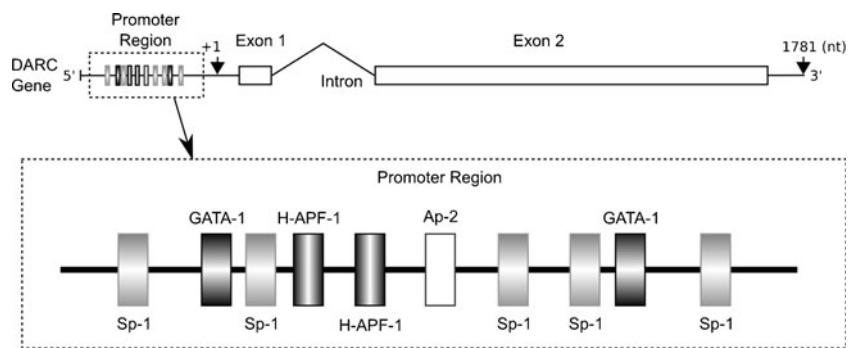
To investigate the nature of the selection pressures acting on the Duffy antigen, we sequenced and analyzed a ~3,000-bp region including the entire *DARC* gene as well as portions of its 5' and 3' flanking regions in a phylogenetic sample of nonhuman primates. We applied phylogenetic-based likelihood approaches to the full coding region as well as to the proximal 5' cis-regulatory region. Figure 1 shows the structure of the *DARC* gene in humans and shows the location of the erythroid GATA site as well as several other known regulatory binding sites (Iwamoto et al. 1996). The *DARC* gene is comprised of a very short exon (21 bp), a single intron (480 bp), and a second exon (990 bp). In this study, we characterize the nature of molecular evolution occurring in different subregions of the gene and identify possible amino acid sites as well as nucleotide sites within the regulatory region that show evidence of positive selection.

Material and methods

Primate samples

We sequenced the *DARC* gene and the regulatory region in the New World monkeys (*Saimiri sciureus*, *Saimiri ustus*, and *Cebus apella*), the Old World monkeys (*M. mulatta*, *M. fascicularis*, *M. nemestrina*, *M. nigra*, *Macaca thibetana*, *Mandrillus sphinx*, *Mandrillus leucophaeus*, *Lophocebus aterrimus*, *Cercocebus torquatus lunatus*, *Cercocebus galeritus*, *Cercopithecus mitis*), and the ape (*Gorilla gorilla*). The *M. mulatta* and *M. fascicularis* samples used in this study have geographic origins from China and Philippines, respectively (information supplied by Covance, Inc.). The precise geographic origins of the remaining samples used in this study are unknown. However, all Old World monkey species of the genus *Macaca* sampled in this study have geographic distributions in continental Southeast Asia and/or in the islands of Southeast Asia, while all remaining Old World monkey species as well as the *Gorilla* sampled in this study have distributions in regions of sub-Saharan Africa. The New World monkeys are all from regions within Brazil. See supplementary material for information regarding the sources of the DNA samples and details about the PCR and sequencing primers, PCR and sequencing protocols, and methods of sequence assembly. In addition, the human and chimpanzee (*Pan troglodytes*) *DARC* sequences were downloaded from GenBank. All nonhuman sequences generated for this study were deposited in GenBank under the accession numbers HQ285843-HQ285857.

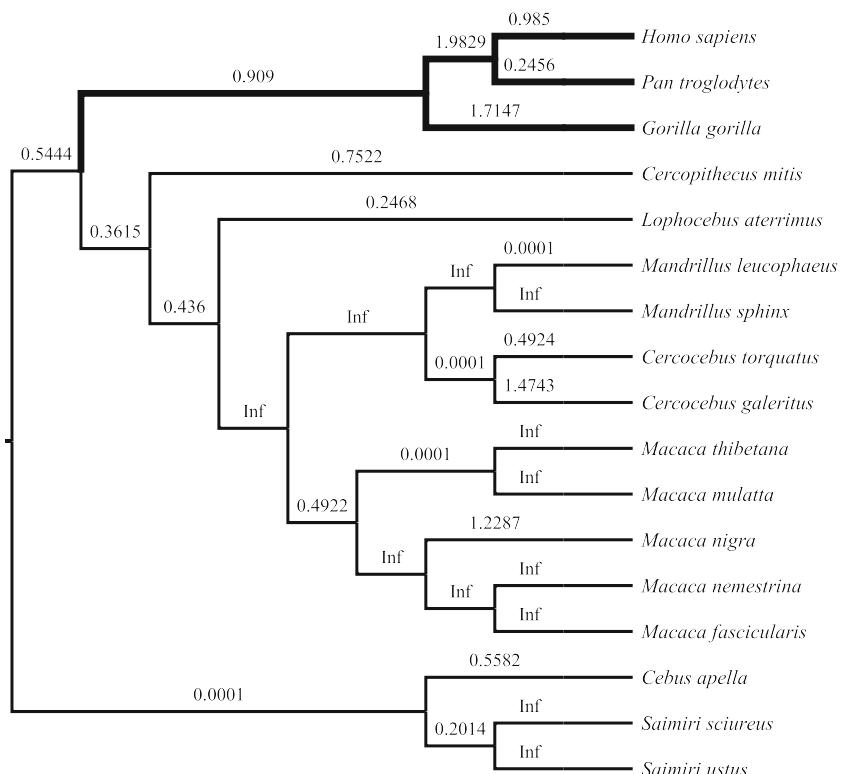
Fig. 1 Structure of the *DARC* gene in humans. The figure indicates the location of the erythroid GATA site as well as several other known regulatory binding sites (see text for more details)



Phylogenetic analysis

The maximum likelihood phylogenetic analyses used the PhyML software (Guindon and Gascuel 2003). For maximum parsimony analysis, the MEGA 4.1 program was used (Tamura et al. 2007). The neighbor-joining phylogeny was done using NEIGHBOR from the PHYLIP package (Felsenstein 2005). All methods used the coding and non-coding sequence (~3,000 bp) and a bootstrap of 1,000 replicates. Nucleotide substitution models (HKY85 for ML and F84 for NJ) were chosen by ModelGenerator v0.84 (Keane et al. 2006) from all *DARC* sequences. All trees were rooted using the *DARC* gene from the mouse genome from GenBank. The outgroup is not shown in the tree in Fig. 2 because phylogenetic maximum likelihood (PAML) only accepts unrooted trees for analyses.

Fig. 2 The gene tree for *DARC* inferred using maximum likelihood and used in positive selection analyses using the software PAML



Tests of positive selection

We tested for evidence of positive selection in the coding region of *DARC* using the codeml software in the PAML v.4.0 package (Yang 2007). All gaps and deletions were treated as ambiguity codes (parameter clean data=0). We applied three positive selection tests (maximum likelihood branch-site test, maximum likelihood site test, and partitioned site test).

To test for positive selection in the regulatory region, we used the maximum likelihood method implemented in the program EvoNC (Wong et al. 2004) to analyze the 5' region of *DARC*. The method implemented in EvoNC calculates a ratio (ζ) between the rate of nucleotide substitution in the regulatory region and the rate of synonymous substitutions, using rates of substitution at synonymous sites as a neutral

standard. As in the d_N/d_S approach, $\zeta=1$ indicates neutral evolution, $\zeta<1$ indicates purifying selection, and $\zeta>1$ indicates positive selection (Wong et al. 2004).

Protein structure visualization

The three-dimensional structure of the *DARC* protein showing amino acids that are hypothesized to be under positive selection (Fig. 3) was generated by PyMOL 1.3 using the RasMol.txt output file from PAML (site method M3), which contained all *a posteriori* probabilities (ω) for each codon. We used an in silico prediction of *DARC* generated by de Brevern et al. (2005).

Results

GATA-1 region

The resistance to invasion by *P. vivax* in sub-Saharan Africans is caused by replacement of a nucleotide in the binding site of the transcription factor GATA-1 inhibiting the *DARC* gene expression on red blood cells. The GATA-1 motif did not show any differences among the nonhuman primate species sequenced, indicating that this nucleotide change occurs only in humans, consistent with the results obtained by Tournamille et al. (2004).

Comparison of sequence alignments

According to Tournamille et al. (1997), the conservation of extracellular cysteines is important because it is believed that at least four of them play important structural roles in the Duffy protein's function as a chemokine receptor. *DARC* has 12 cysteines. Residues C4, C51, C54, C129, C195, and C276 are predicted to be extracellular whereas, C93, C134, C149, C214, C297, and C308 are predicted to be in the transmembrane or intracellular region. We found that all cysteine residues are conserved among our primate samples, with the exception of residue C308 where the gorilla sequence has tyrosine. Thus, the distribution of conserved codons confirms the findings of Tournamille et al. (2004) indicating conservation of structure for Duffy's chemokine function.

The expression of *Fy*^a and *Fy*^b (human antigens) in nonhuman primates was analyzed by Palatnik and Rowe (1984). Using flow cytometry, Tournamille et al. (2004) confirmed that *Fy*^a expression is restricted to humans with the difference being due to a change from aspartic acid (*Fy*^b) to glycine (*Fy*^a) at residue 42. This result reinforces the hypothesis of Chaudhuri et al. (1995) that the gene for *Fy*^b is the ancestral allele of *DARC*. Our results also confirm these findings; all of the nonhuman primates we sequenced have an aspartic acid at position 42.

Maximum likelihood branch-site test for positive selection

The branch model implemented in PAML allows detection of positive selection acting on particular lineages permitting different d_N/d_S ratios (ω) among codon sites. A phylogenetic tree was inferred by maximum likelihood (Fig. 2). The topology of this tree is concordant with the known phylogeny of these species. This tree and the entire coding region of the *DARC* gene were used for all PAML analyses.

The d_N/d_S ratios estimated by the free-ratio model in PAML are shown on the tree. Some of these values are considerably greater than 1.0, like on the branch leading to humans and chimpanzees as well as the branch leading to gorillas. For this reason, we carried out the branch-site analysis using the entire hominoid clade as a foreground clade testing for positive selection against the remainder of the tree. We compared model A (which assumes positive selection) versus the corresponding null model (i.e., the neutral model) where $\omega_2=1$. The results of this test are given in the supplementary file to this article as Table 1 but indicate that the high d_N/d_S ratios do not reach statistical significance.

Test for positive selection at individual amino acid sites

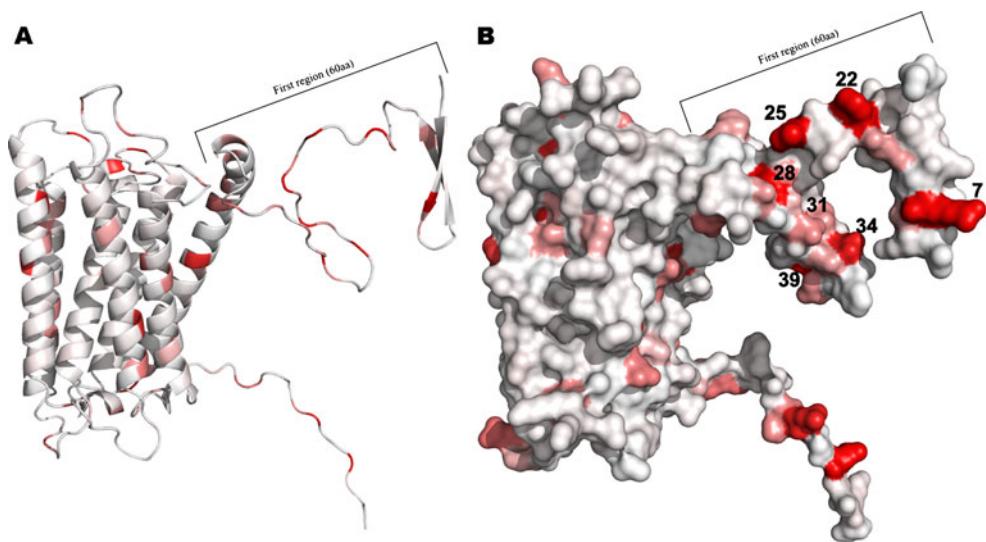
Four pairs of models were compared in our site model analysis. The first pair of models was M1a (nearly neutral) versus M2a (positive selection), the second pair was M0 (one ratio) versus M3 (discrete), the third pair was M7 (beta) versus M8 (beta and ω), and the fourth pair was M8a (beta and $\omega=1$) versus M8 (beta and ω). All results from tests are given in the supplementary material in Table 2.

Model M2a suggests that in *DARC*, nearly 15% of the amino acid sites are under positive selection with $\omega_2=2.4734$. Comparing model M2a versus M1a using a likelihood ratio test ($2\Delta\ln L$, likelihood ratio test (LRT)) yields the result of 3.7739, with $P=0.1$ and 2 degrees of freedom (df). Therefore, model M2a is not significantly better than model M1a in explaining the data.

The discrete model (M3) that allows three classes of ω suggests that 14% of the amino acid sites are under positive selection with $\omega_2=2.4734$, but only six of these amino acid sites were identified as under positive selection with the cutoff posterior probability, $P=95\%$. Model M3 was significantly better than the one-ratio model (M0) ($2\Delta\ln L=22.4915$ with $P=0.0001$ and $df=4$).

Sixteen amino acids were identified to be under positive selection in model M3 with a probability greater than 80%. These amino acid sites are shown in red on the predicted three-dimensional structure of the *DARC* gene (Fig. 3). It is notable that seven of these sites are located in the initial 60 amino acids (extracellular region) of the protein. According to Chitnis et al. (1996), the binding region of *P. vivax*

Fig. 3 Predicted three-dimensional structure of the human DARC protein. **a** The cartoon model. **b** The surface model. It shows (in red) the seven amino acid sites (7, 22, 25, 28, 31, 34, 39) within the binding region hypothesized to be under positive selection (as predicted by model M3) with more than 80% significance. White regions indicate strongly conserved amino acid regions



extends between amino acids 8 and 42, indicating that this region may actually be under strong positive selection to prevent the binding of *P. vivax*/*P. knowlesi*.

When we compared model 8 versus model 7, it was hypothesized that about 15% of the amino acid sites are under positive selection with $\omega_2=2.4908$ and identified six sites under positive selection with posterior probability values greater than 80%. However, the difference between the likelihoods of M7 and M8 is not statistically significant ($2\Delta\ln L=3.8817$ with $P=0.1$ and $df=4$).

Analyzing partitions within the coding region

The maximum likelihood method can be performed by partitioning the protein into regions hypothesized to be subjected to different selective pressures and that therefore might be expected to have different ω ratios. We divided the DARC gene into four regions based on its protein structure: region 1, the extracellular binding region of *P. vivax*/*P. knowlesi* (the first 60 amino acids); region 2, the seven transmembrane regions; region 3, the four cytoplasmic regions; and region 4, the extracellular regions other than region 1 (Fig. 4). When applying maximum likelihood to partitioned gene regions, different ω ratios are allowed for each region. Similarly, other parameters may be allowed to differ between the regions.

Four models were used to accommodate various levels of heterogeneity among the regions. The simplest model assumes the four regions have the same pattern of substitution and therefore share identical parameters (model A in Table 3 in the supplementary file). The more complex model (model D in Table 3 in the supplementary file) assumes the branch lengths (b) are similar among partitions but with different codon frequencies (rs), different transition/transversion ratio (κ), different nonsynonymous/synonymous

substitution ratios (ω), and different equilibrium nucleotide frequency (πS). Models B and C are between these two extremes.

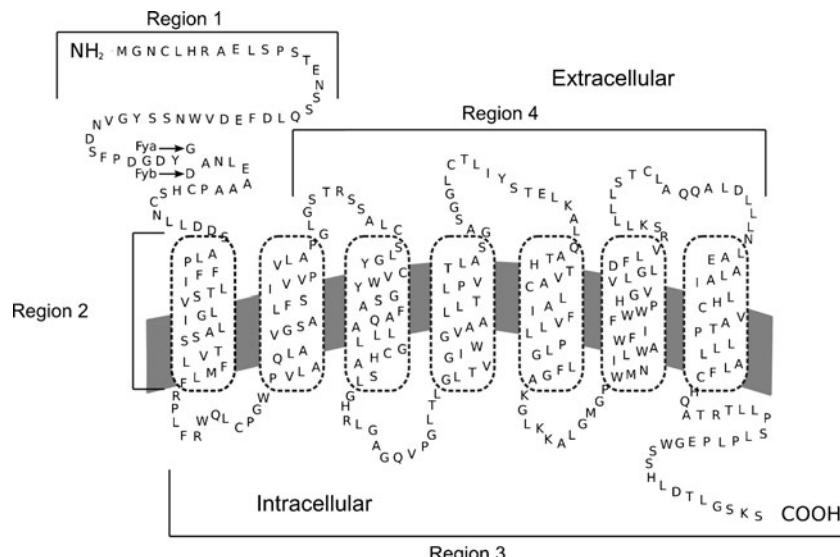
The results from each protein partition are shown in Table 3 (see supplementary file). Model A showed $\ell_A=-2396.7621$. Model B represents the data better than model A (the LRT result is 46.6336 with $df=9$). This result suggests that the pattern in the use of codons is different among regions. Model C also represents the data better than model A (the LRT result is 16.6662 with $df=2$). Model D fits the data even better compared with models A, B, and C. This indicates similar results to those of model C.

To test whether the ratio ω of region 1 is significantly greater than 1, we calculated the log likelihood for models C and D in which we set $\omega_1=1$ (called C_{null} and D_{null}). The LRT statistics for C_{null} versus C and D_{null} versus D are 2.9799 ($P=0.084$) and 3.7032 ($P \sim 0.05$) (see Table 3 in supplementary file), respectively, with $df=1$. Therefore, the d_N/d_S ratio for region 1 from models C and D approaches significance at the 5% level.

Analysis of positive selection within the 5' regulatory region

Within the 5' regulatory region, we found evidence of positive selection at 28 nucleotide sites having a posterior probability of $P \geq 90\%$. For a list of these sites and their posterior probability values, see the [supplementary material](#). When we compared model M1 versus model M2 as well as model M1 versus model M3, they yield almost identical results with both having an LRT ~ 7.77 , with 2 degrees of freedom. This is significant with $P < 0.05$, meaning we can reject the neutral model (M1). Most of these sites do not fall within known regulatory motifs (Iwamoto et al. 1996); however, site 704 falls within an Ap-2 motif, and site 843

Fig. 4 The *DARC* gene was subdivided into four subregions based on its protein structure in order to test whether natural selection has acted differently within these different subregions. Region 1, the extracellular binding region of *P. vivax/P. knowlesi* (the first 60 amino acids); region 2, the seven transmembrane regions; region 3, the four cytoplasmic regions; and region 4, the extracellular regions other than region 1 (Modified from Hadley and Peiper 1997.)



is located very near to the GATA-1 motif that regulates expression of Duffy on erythrocytes.

Discussion

The *DARC* is widely expressed and acts as both a promiscuous chemokine receptor and erythrocyte receptor for *P. vivax/P. knowlesi* (Hadley and Peiper 1997; Langhi and Bordin 2006; Miller et al. 1976; Miller et al. 1975; Tournamille et al. 2004). Because of its dual biological properties, *DARC* is potentially subjected to two selective pressures: an internal purifying selection for maintenance of the binding properties of chemokines (with the endogenous chemokines as a selective factor) and an external positive selection (where parasites are a potential selective factor) (Tournamille et al. 2004).

The free-ratio model in PAML indicated that along some branches on the tree, especially within the hominoid clade, there were more nonsynonymous changes than synonymous changes (Fig. 2). This could indicate that positive selection has occurred along hominoid branches. We also know from previous studies that *DARC* has experienced positive selection in sub-Saharan African populations for resistance to *P. vivax* infection. Indeed, new studies have shown that wild populations of African great apes are naturally infected with *Plasmodium* species that are very similar to human *P. vivax* (Rayner et al. 2011; Prugnolle et al. 2010). This suggests the possibility that African apes have been the source of *P. vivax* infection in sub-Saharan African populations through zoonotic transmission in the past which then drove positive selection in *DARC* for resistance to *P. vivax*. However, even when we restricted our tests for positive selection specifically to the hominoid clade, the branch-site model was not able to

detect statistically significant evidence for positive selection. It is possible that if additional hominoid species are included in the analysis, statistically significant support for positive selection could be identified. Indeed, Demogines et al. (2012), who applied similar methods as we used, found statistical evidence supporting positive selection in the hominoid group when additional species were included.

It should be noted that additional nonhominoid branches of the tree also showed d_N/d_S values considerably higher than 1 which could also indicate positive selection. This is seen, for example, on the branch to *M. nigra* as well as to *C. galeritus*. It is possible that the high d_N/d_S value in *M. nigra* is due to *P. knowlesi* which is known to infect macaque monkeys throughout the Indonesian islands and the Philippines in Southeast Asia (Fooden 1994). Although *P. knowlesi* is not known from Africa, it is possible that *P. vivax*, which infects African apes (Rayner et al. 2011; Prugnolle et al. 2010), also infects *C. galeritus* though we have seen no study that has investigated this. However, some *Plasmodium* species are known to naturally infect some African monkeys, particularly close relatives of *Cercopithecus* (Prugnolle et al. 2011; Poiriez et al. 1994).

Given the fact that we see high d_N/d_S values on different branches of the tree, we can use a less stringent approach to help identify positive selection in *DARC* within the primates. Specifically, we used PAML site models over the entire *DARC* coding region without targeting any particular branches. This approach allows for the detection of positive selection at specific codons within the protein (Yang et al. 2000). In the three models that allow sites to evolve non-neutrally (M2a, M3, and M8), the d_N/d_S ratio (ω_2) was greater than 1, but only M3 versus M0 presents LRT results that are statistically significant (see Table 2 in supplementary material). According to Yang and Nielsen (2002), the LRT

between M0 and M3 models is more of a variability test of the ω ratio between sites than a test of positive selection. Nevertheless, we localized the sites that were hypothesized to be under positive selection within the three-dimensional structure of *DARC* protein (Fig. 3 in supplementary file). Interestingly, the distribution of these sites is not random, and many sites are concentrated in or are very close to the *P. vivax/P. knowlesi* binding region.

Applying the selection tests to the separate subregions showed that the amino-terminal binding region of *P. vivax/P. knowlesi* (defined as region 1) has an ω greater than 1, though not significantly different from 1. The transmembrane, intracellular, and the remaining extracellular regions, however, showed very low values of ω indicating that they are under purifying selection with the transmembrane region (region 2) being most conserved ($\omega < 0.35$). Tournamille et al. (2004) conducted a similar study with the *DARC* gene in primates using slightly different methods. They applied their tests in two ways: in the first case, the test was applied over the entire gene, and in the other case, the test was applied only to the amino-terminal region (our region 1). Their results were similar to our results and did not yield strong evidence supporting positive selection, either over the whole gene or in the amino-terminal region. However, Tournamille et al. (2004) did not perform tests of positive selection, like we did, on different structural subregions of the gene. Our analysis was able to contrast the d_N/d_S ratio (ω) among separate regions and therefore was able to reveal that the d_N/d_S ratio (ω) in the binding region (i.e., the first 60 amino acids) is strikingly different compared to the other regions of the gene. This discrepancy itself strongly suggests that there are different selective forces acting on these different regions. Moreover, the results in Tournamille et al. (2004), which show a correlation between hydrophilicity and variability along the protein, also lead to the conclusion that the evolutionary constraints differ along the protein.

We analyzed the binding region for *P. vivax/P. knowlesi* (region 1) separately from the other extracellular regions (region 4) for study purposes, but theoretically, the entire extracellular region might be expected to be under the same selective pressure because it is exposed at the cell surface. Thus, it might be expected that the amino-terminal end (binding region) would be under levels of purifying selection similar to that seen in the other extracellular regions. The difference between the two regions however could indicate the existence of a second type of selection (e.g., positive selection) acting simultaneously with purifying selection on the *P. vivax/P. knowlesi* binding region. Even though the relatively high d_N/d_S ratio in the binding region could indicate relaxed selection, for several reasons, we believe that positive selection is acting in this domain. It is known a priori that *P. vivax/P. knowlesi* binds to the *DARC* antigen (Miller et al. 1975). Also, as we said previously,

DARC is under positive selection in humans. Furthermore, recent data indicate that positive selection has acted on the ligands of different species of *Plasmodium* by which they attach to erythrocytes (Sawai et al. 2010).

The model of two selective pressures (purifying selection and positive selection) acting on the binding region while the other regions are under influence of only one of them (i.e., purifying selection) is a reasonable explanation for the results. Thus, we cannot rule out the existence of positive selection acting on the *DARC* gene during the evolutionary history of primates. Although in our analysis we employed likelihood tests that have proven valuable in revealing positive selection in studies of other genes (e.g., Yang and Swanson 2002), Yang and Nielsen (2002) argued that LRT tests may fail to provide significant statistical support for positive selection especially when analyzing short sequences. This may explain the lack of statistical significance for positive selection in the *P. vivax/P. knowlesi* binding site because it has only 60 amino acids.

Moreover, it is known that *P. vivax* is a human parasite, while in nonhuman primates, specifically species within the genus *Macaca*, infection is by the closely related parasite *P. knowlesi*. According to Iyer et al. (2007), *P. vivax* is highly dependent on the Duffy antigen to invade red blood cells in humans. However, it is known that *P. knowlesi* also has alternative ways to invade red blood cells in macaque species (Allison 2009; Iyer et al. 2007). This could indicate that *DARC* does not suffer a strong selective pressure in monkeys because it is not the only determining factor for infection, which may also explain the lack of significance in positive selection tests.

Our results of the regulatory region yielded several sites hypothesized to be under positive selection. Yet, only two of these sites were near-known regulatory motifs for *DARC* (see Iwamoto et al. 1996). It is interesting that one of the sites is very close to the GATA motif that harbors the nucleotide variant responsible for the absence of the Duffy protein on the erythrocytes of sub-Saharan Africans and provides resistance to invasion by *P. vivax*. Interestingly, another possible selective agent on *DARC* is the malaria-like pathogen *Hepatocystis kochi* that is known to infect baboons. Indeed, Tung et al (2009) found evidence that some sites within the regulatory region of *DARC* show evidence of having been positively selected possibly due to this pathogen. However, without experimental assays, we cannot know the functional consequence of the substitutions we identified in regulating *DARC*.

While we have not been able to obtain strong statistical support for positive selection on *DARC* using a phylogenetic approach, future analyses would benefit by using population-level methods for detecting selection. The genus *Macaca* should be an important focus because certain species appear to have evolved immunological defenses against

this parasite (Butcher et al. 2010; Butcher 1996; Fooden 1994). For example, *M. fascicularis* usually shows low parasitemias of *P. knowlesi* and is able to live quite comfortably even at peak infections of 1–3% (Butcher 1996). In contrast, experimental infection of *M. mulatta* by *P. knowlesi* results in very high levels of parasitemias and almost always leads rapidly to death (Butcher 1996). Moreover, additional plasmodia are known to infect macaques including *Plasmodium coatneyi*, *Plasmodium fragile*, *P. knowlesi*, *Plasmodium inui*, and *Plasmodium cynomolgi* (Fooden 1994; Sasseville and Ditors 2008); therefore, these plasmodia might also exert a selective pressure.

In conclusion, positive selection tests do not provide statistical support in favor of positive selection acting on the *DARC* gene in primates. However, our results do indicate comparatively high rates of nonsynonymous change along some branches, particularly in the hominoid clade. Additionally, our analyses convincingly demonstrate that the region of the *DARC* gene containing the binding domain for *P. vivax*/*P. knowlesi* is under very different selective pressures than the remainder of the gene. For this region, our results indicate a scenario in which two selective forces (i.e., purifying selection and positive selection) counterbalance each other within the binding region and for which current statistical tests are limited to detect the signature of positive selection. Therefore, it is challenging to prove that positive selection played a role in the evolution of the *DARC* gene. The *DARC* protein has a complex relationship between its role in malaria and its function as a promiscuous receptor for chemokines. This, coupled with the expression of *DARC* in different body tissues, makes it difficult to detect the contribution of the different forces that have influenced its evolution.

Acknowledgments We thank the two anonymous reviewers for their helpful comments that helped us improve the article. This research was supported by a PSC CUNY 38 grant to EEH and a C3IRG award from City University of New York to EEH and CKJ as well as grants from INCTC/CNPq, CTC/CEPID/FAPESP, Regional Blood Center of Ribeirão Preto, and Millennium Institute/CNPq to WA Silva-Jr. The authors would like to thank Adriana Marques, Cristiane A. Pereira, and Anemare R. Dinarte for their technical assistance.

References

- Allison AC (2009) Genetic control of resistance to human malaria. *Curr Opin Immunol* 21:499–505
- Baum J, Thomas AW, Conway DJ (2003) Evidence for diversifying selection on erythrocyte-binding antigens of *Plasmodium falciparum* and *P. vivax*. *Genetics* 163(4):1327–1336
- Butcher GA (1996) Models for malaria: nature knows best. *Parasitol Today* 12(10):378–382
- Butcher GA, Mitchell GH, Cohen S (2010) *Plasmodium knowlesi* infections in a small number of non-immune natural hosts (*Macaca fascicularis*) and in rhesus monkeys (*M. mulatta*). *Trans R Soc Trop Med Hyg* 104:75–77
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton
- Chaudhuri A, Polyakova J, Zbrzezna V, Pogo AO (1995) The coding sequence of Duffy blood group gene in humans and simians: restriction fragment length polymorphism, antibody and malarial parasite specificities, and expression in nonerythroid tissues in Duffy-negative individuals. *Blood* 85:615–621
- Chitnis CE, Chaudhuri A, Horuk R, Pogo OA, Miller LH (1996) The domain on the Duffy blood group antigen for binding *Plasmodium vivax* and *P. knowlesi* malarial parasites to erythrocytes. *J Exp Med* 184:1531–1536
- de Brevern AG, Wong H, Tournamille C, Colin Y, Le Van KC, Etchebest C (2005) A structural model of a seven-transmembrane helix receptor: the Duffy antigen/receptor for chemokine (*DARC*). *Biochim Biophys Acta* 1724(3):288–306
- Demogines A, Truong KA, Sawyer SL (2012) Species-specific features of *DARC*, the primate receptor for *Plasmodium vivax* and *Plasmodium knowlesi*. *Mol Biol Evol* 29(2):445–449
- Escalante AA, Cornejo OE, Freeland DE, Poe AC, Durrego E, Collins WE, Lal AA (2005) A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. *Proc Natl Acad Sci U S A* 102:1980–1985
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genetics, University of Washington, Seattle. Distributed by the author
- Fooden J (1994) Malaria in macaques. *Int J Primatol* 15(4):573–596
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704
- Hadley TJ, Peiper SC (1997) From malaria to chemokine receptor: the emerging physiologic role of the Duffy blood group antigen. *Blood* 89:3077–3091
- Hamblin MT, Di Rienzo A (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am J Hum Genet* 66:1669–1679
- Hamblin MT, Thompson EE, Di Rienzo A (2002) Complex signatures of natural selection at the Duffy blood group locus. *Am J Hum Genet* 70:369–383
- Harris EE, Meyer D (2006) The molecular signature of selection underlying human adaptations. *Am J Phys Anthropol Suppl* 43:89–130
- Iwamoto S, Li J, Omi T, Ikemoto S, Kajii E (1996) Identification of a novel exon and spliced form of Duffy mRNA that is the predominant transcript in both erythroid and postcapillary venule endothelium. *Blood* 87:378–385
- Iyer J, Gruner AC, Renia L, Snounou G, Preiser PR (2007) Invasion of host cells by malaria parasites: a tale of two protein families. *Mol Microbiol* 65:231–249
- Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6:29
- Langhi DM Jr, Bordin JO (2006) Duffy blood group and malaria. *Hematology* 11:389–398
- Miller LH, Mason SJ, Dvorak JA, McGinniss MH, Rothman IK (1975) Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. *Science* 189:561–563
- Miller LH, Mason SJ, Clyde DF, McGinniss MH (1976) The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, *FyFy*. *N Engl J Med* 295:302–304
- Palatnik M, Rowe AW (1984) Duffy and Duffy-related human antigens in primates. *J Hum Evol* 13:173–179
- Poiriez J, Dei-Cas E, Landau I (1994) Further description of blood stages of *Plasmodium petersi* from *Cercopithecus albigena* monkey. *Folia Parasitol* 41(3):168–172
- Prugnolle F, Ollomo B, Durand P, Yalcindag E, Arnathau C, Elguero E, Berry A, Pourrut X, Gonzalez JP, Nkoghe D, Akiana J, Verrier D,

- Leroy E, Ayala FJ, Renaud F (2011) African monkeys are infected by *Plasmodium falciparum* nonhuman primate-specific strains. Proc Natl Acad Sci U S A 108(29):11948–11953
- 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 (2010) African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. Proc Natl Acad Sci U S A 107(4):1458–1463
- Rayner JC, Liu W, Peeters M, Sharp PM, Hahn BH (2011) A plethora of *Plasmodium* species in wild apes: a source of human infection? Trends Parasitol 27(5):222–229
- Sasseville VG, Ditors RW (2008) Impact of infections and normal flora in nonhuman primates on drug development. ILAR J 49(2):179–190
- Sawai H, Otani H, Arisue N, Palacpac N, de Oliveira ML, Pathirana S, Handunnetti S, Kawai S, Kishino H, Horii T, Tanabe K (2010) Lineage-specific positive selection at the merozoite surface protein 1 (msp1) locus of *Plasmodium vivax* and related simian malaria parasites. BMC Evol Biol 10:52
- Schmidt LH, Fradkin R, Harrison J, Rossan RN (1977) Differences in the virulence of *Plasmodium knowlesi* for *Macaca irus (fascicularis)* of Philippine and Malayan origins. Am J Trop Med Hyg 26:612–622
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599
- Tournamille C, Blancher A, Le Van KC, Gane P, Apoil PA, Nakamoto W, Cartron JP, Colin Y (2004) Sequence, evolution and ligand binding properties of mammalian Duffy antigen/receptor for chemokines. Immunogenetics 55:682–694
- Tournamille C, Le Van KC, Gane P, Blanchard D, Proudfoot AE, Cartron JP, Colin Y (1997) Close association of the first and fourth extracellular domains of the Duffy antigen/receptor for chemokines by a disulfide bond is required for ligand binding. J Biol Chem 272:16274–16280
- Tung J, Primus A, Bouley AJ, Severson TF, Alberts SC, Wray GA (2009) Evolution of a malaria resistance gene in wild primates. Nature 460(7253):388–391
- Wong WS, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168:1041–1051
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449
- Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917
- Yang Z, Swanson WJ (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol Biol Evol 19(1):49–57