Species-Specific Features of DARC, the Primate Receptor for *Plasmodium vivax* and *Plasmodium knowlesi*

Ann Demogines,[1] Kimberly A. Truong,[1] and Sara L. Sawyer*,[1]

[1]Section of Molecular Genetics and Microbiology, Institute for Cellular and Molecular Biology, The University of Texas at Austin

*Corresponding author: E-mail: saras@mail.utexas.edu.

Associate editor: Anne Stone

Abstract

The *DARC* (*Duffy antigen/receptor for chemokines*) gene, also called *Duffy* or *FY*, encodes a membrane-bound chemokine receptor. Two malaria parasites, *Plasmodium vivax* and *Plasmodium knowlesi*, use DARC to trigger internalization into red blood cells. Although much has been reported on the evolution of *DARC* null alleles, little is known about the evolution of the coding portion of this gene or the role that protein sequence divergence in this receptor may play in disease susceptibility or zoonosis. Here, we show that the *Plasmodium* interaction domain of DARC is nearly invariant in the human population, suggesting that coding polymorphism there is unlikely to play a role in differential susceptibility to infection. However, an analysis of *DARC* orthologs from 35 simian primate species reveals high levels of sequence divergence in the *Plasmodium* interaction domain. Signatures of positive selection in this domain indicate that species-specific mutations in the protein sequence of DARC could serve as barriers to the transmission of *Plasmodium* between primate species.

Key words: malaria, positive selection, arms race, species tropism, zoonosis.

The *DARC* gene (*Duffy antigen/receptor for chemokines*), also called *Duffy* or *FY*, encodes a membrane-bound chemokine receptor. Different human alleles of this gene underlie the designation of different Duffy blood groups (Meny 2010), but *DARC* also plays a role in the biology of malaria. In humans, the most severe form of malaria is caused by *Plasmodium falciparum*, with less severe forms of the disease caused by the related organisms *Plasmodium knowlesi*, *Plasmodium vivax*, *Plasmodium ovale*, and *Plasmodium malariae*. Two of these, *P. vivax* and *P. knowlesi*, exploit DARC for internalization into red blood cells. Interaction with DARC is mediated by a *Plasmodium* surface ligand called *Duffy-binding protein* (DBP) (Haynes et al. 1988; Wertheimer and Barnwell 1989). A *cis*-regulatory polymorphism that silences *DARC* expression in erythrocytes has arisen independently in human populations from different parts of the world and is highly protective against *P. vivax* and *P. knowlesi* infection (Miller et al. 1975, 1976; Tournamille et al. 1995; Zimmerman et al. 1999). In a fascinating example of convergent evolution, polymorphisms in the *cis*-regulatory region of *DARC* in African baboons are also associated with resistance to a malaria-like parasite common in baboon populations, *Hepatocystis kochi* (Tung et al. 2009). Although much has been reported on the evolution of *DARC* regulatory elements, little is known about the evolution of the coding portion of this gene or the role that protein sequence divergence in this receptor may play in disease susceptibility or zoonosis.

Lack of Human Diversity in the *Plasmodium* Interaction Domain

The N-terminal extracellular tail of DARC interacts with the DBP of *P. vivax* and *P. knowlesi* (fig. 1) and also contains some of the determinants for chemokine binding (Chitnis et al. 1996; Tournamille et al. 2003, 2005). In mutagenesis studies, several mutations in this domain have been shown to destroy the interaction with *P. vivax* while not interfering with chemokine binding (Tournamille et al. 2003, 2005). Given that *P. vivax* is the most widely distributed malaria parasite in the world (Mendis et al. 2001), it is intriguing that polymorphisms at these sites have not been reported in the human population. In fact, we find that no nonsynonymous single-nucleotide polymorphisms have ever been reported in the portion of *DARC* encoding the *Plasmodium* interaction domain (fig. 1). The only exception to this is the common polymorphism at position 42, which defines the *FY*A* and *FY*B* Duffy blood group alleles (Meny 2010). Thus, polymorphism in the coding sequence of *DARC* is unlikely to play a role in differential susceptibility of humans to *P. knowlesi* and *P. vivax*, since most human genomes will encode an identical DARC N-terminal tail. Destabilizing mutations elsewhere in the protein which convey reduced levels of DARC expression at the cell surface will still be important, and one such human polymorphism (R89C) is known (Meny 2010).

Divergence of DARC Protein Sequence during Primate Evolution

Plasmodium vivax and *P. knowlesi*, as well as *P. falciparum*, are all known to be zoonotic transmissions from nonhuman primates (Escalante et al. 2005; Mu et al. 2005; Liu et al. 2010; Lee et al. 2011). Furthermore, species-specific susceptibility of primate cells to *P. vivax* and *P. knowlesi* is well documented (Wertheimer and Barnwell 1989). There is evidence to suggest that this is at least partly
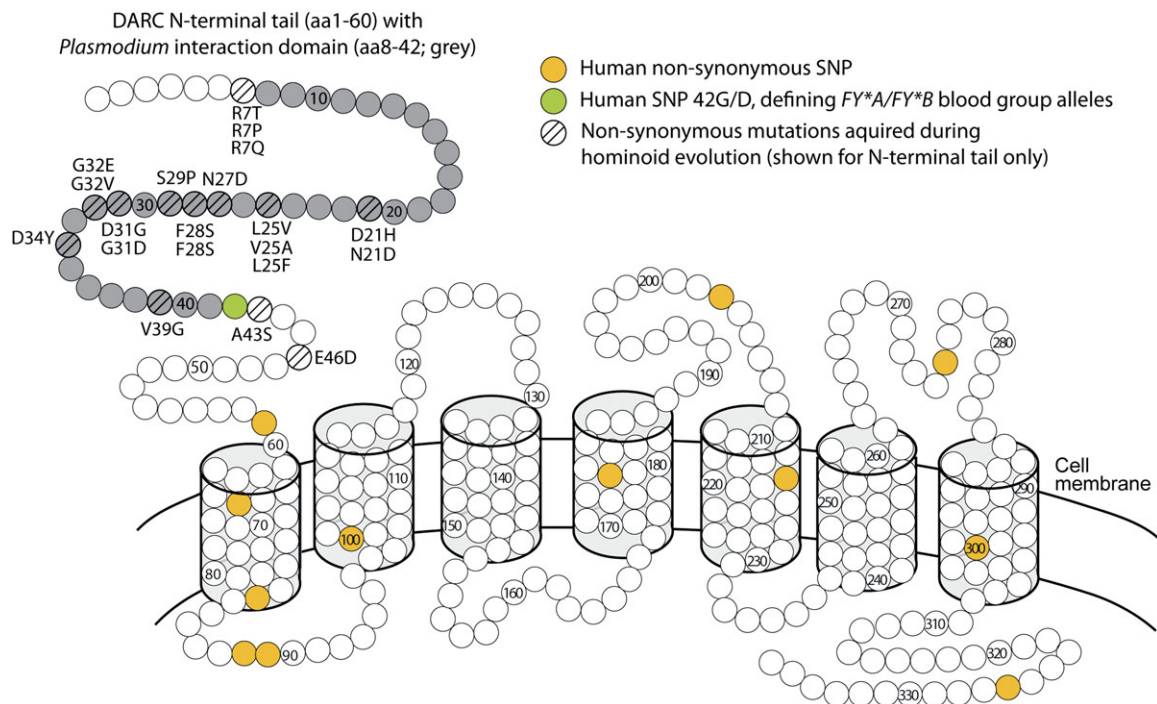
Letter

FIG. 1. Divergence but not diversity shapes the *Plasmodium* interaction domain of DARC. A model of DARC embedded in the cell membrane is shown (Tournamille et al. 2003). In yellow and green are residue positions where a human nonsynonymous single-nucleotide polymorphism (SNP) has been reported (dbSNP and 1000 Genomes databases). The N-terminal tail contains the interaction domain with *Plasmodium vivax* and *Plasmodium knowlesi* (Chitnis et al. 1996). Labeled next to residues in the N-terminal tail are the nonsynonymous mutations predicted to have occurred during hominoid evolution by free-ratio analysis. F28S occurred twice, on separate branches.

attributable to species-specific differences in the DARC protein sequence. For example, the protein sequence of rhesus macaque DARC conveys a special glycosylation pattern that blocks binding of the *P. vivax* DBP (Chitnis et al. 1996). To better understand how the DARC protein varies between primates, we sequenced *DARC* from 17 simian primate species, including nine hominoid species (great apes and gibbons) closely related to humans. Primate cell lines were obtained and grown under standard conditions (supplementary table S1, Supplementary Material online), and polymerase chain reaction was performed on genomic DNA harvested from these cells (supplementary table S2, Supplementary Material online). The coding sequence of each ortholog was assembled from genomic sequence based on the structure of the human isoform "b" transcript, most of which (990 bases) is encoded by a single exon, with 21 bases coming from an upstream exon (Iwamoto et al. 1996). These novel sequences (starred in fig. 2; Genbank accessions JN544123–JN544139) were combined with 18 *DARC* sequences available in Genbank to create a final data set of 35 full-length simian primate *DARC* orthologs. The alignment of these sequences was straightforward with a single codon deletion being the only indel present, and their phylogenetic relationship was in agreement with accepted primate phylogeny (Perelman et al. 2011).

We used the free-ratio model, implemented in PAML (Yang 1997), to calculate the dN/dS value on each branch of the *DARC* phylogeny. This ratio summarizes the accumulation of nonsynonymous and synonymous mutations, with dN representing the number of nonsynonymous

mutations per nonsynonymous site and dS representing the number of synonymous mutations per synonymous site (Hurst 2002). It is well appreciated that, over evolutionary time, most protein coding genes accumulate nonsynonymous mutations at a rate far slower than synonymous mutations (dN/dS ≪ 1) (Meyerson and Sawyer 2011). However, on several branches *DARC* has accumulated nonsynonymous mutations at a rate as fast or faster than synonymous changes (dN/dS ≥ 1) (bold type; fig. 2A).

This pattern was also observed when the portion encoding the N-terminal tail of DARC was analyzed alone (fig. 2B), a region where high sequence divergence between primate orthologs has previously been noted (Tournamille et al. 2004). In these 60 codons, four nonsynonymous mutations have accumulated along the branch leading to gorilla from its last common ancestor with humans, whereas zero synonymous mutations became fixed during this time. An alignment of the human and gorilla N-terminal tail is shown along with the predicted sequence of DARC that existed at the time of their last common ancestor, illustrating these four nonsynonymous mutations (R7T, V25A, G31D, D34Y; fig. 2C). To illustrate the extent of mutation that has occurred in this N-terminal tail, all nonsynonymous mutations predicted to have occurred during the evolution of our closest ancestors, the hominoids, are diagrammed in figure 1. This diagram illustrates that there has been a constellation of divergent N-terminal tails encoded by primate genomes during their evolution. Divergence in this tail probably does not affect chemokine binding, as different primate orthologs of DARC are all able to bind
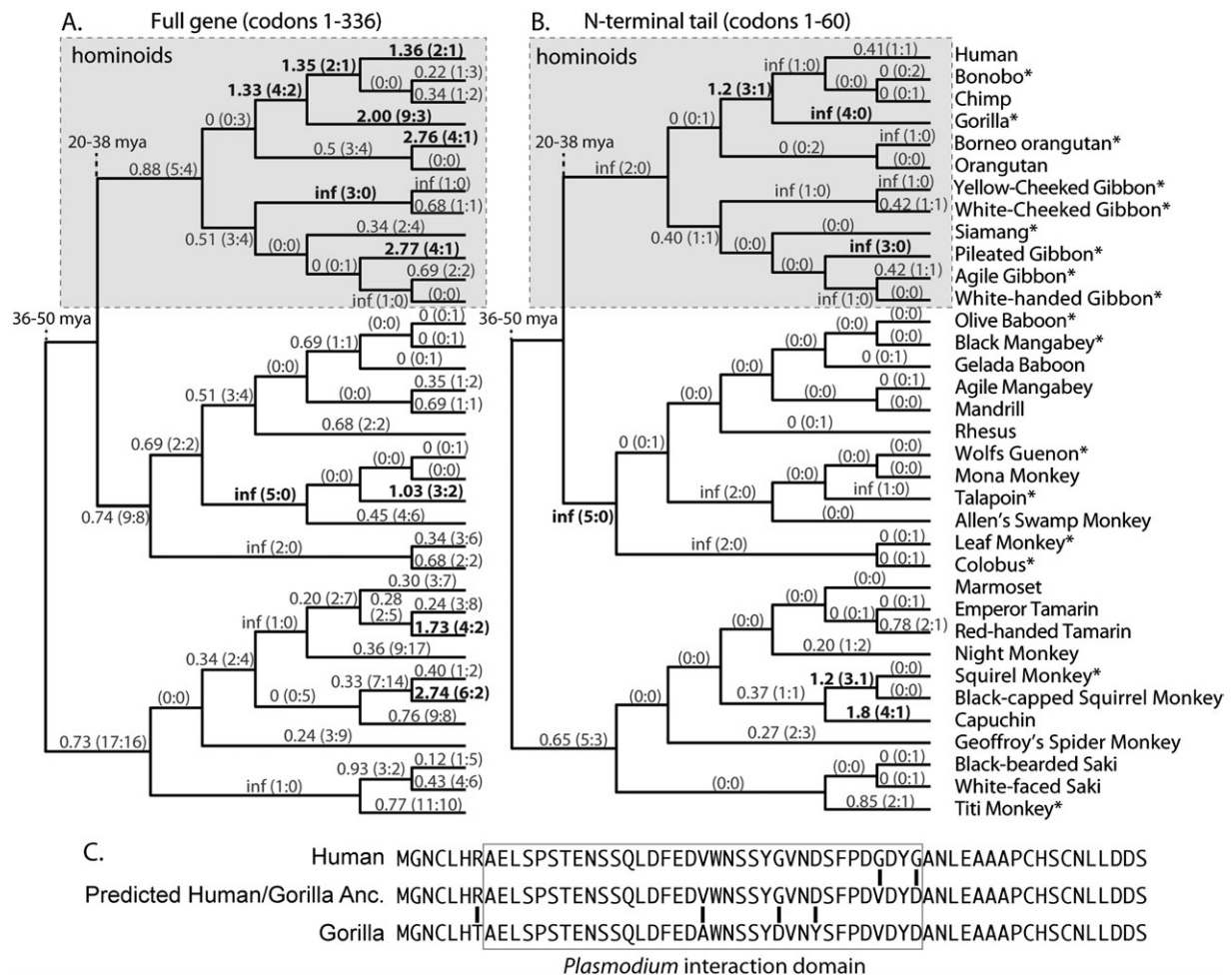
FIG. 2. Divergence of *DARC* during simian primate speciation. dN/dS was calculated along each branch over either (*A*) the entire gene length or (*B*) the first 60 codons, which encode the N-terminal tail. Actual numbers of nonsynonymous and synonymous mutations predicted to have occurred along each branch are indicated in parentheses (*N:S*). dN/dS is indicated as "inf" (infinity) where dS = 0 and in bold where dN/dS > 1 (or, arbitrarily, where *N:S* ≥ 3:0 in cases where dS = 0). Asterisks indicate sequences acquired as part of this study. (*C*) An alignment of the N-terminal tail sequence of human and gorilla DARC, along with their predicted ancestral sequence. Vertical lines represent nonsynonymous mutations acquired since that ancestor.

a relevant human chemokine (Tournamille et al. 2004), but could potentially have affected *Plasmodium* susceptibility by modulating the interaction with DBP. If so, species-specific features of the DARC protein sequence could impede the movement of *Plasmodium* between primate species.

Residues under Positive Selection in the *Plasmodium* Interaction Domain

We tested the hypothesis that nonsynonymous mutations in *DARC* can confer an adaptive fitness advantage, regardless of what that might be. We did this by assessing whether *DARC* bears the signatures of positive natural selection for such mutations. Codon-based models in PAML were used to analyze the 35-species data set. Positive selection of *DARC* was supported ($P \leq 0.03$) under all models of codon evolution (supplementary table S3, Supplementary Material online), which was verified with the program Hy-Phy (supplementary table S4, Supplementary Material online). We noticed that *DARC* evolution appeared to

be least conservative in the hominoids (boxed in fig. 2). This could either be due to relaxed selection or to more acute positive selection in these species. However, positive selection of *DARC* was also supported in a codon-based analysis of only the hominoid species ($P \leq 0.003$; supplementary table S5, Supplementary Material online). This result is more significant than it was with the full data set, even though fewer species and reduced tree length should reduce statistical power (Anisimova et al. 2001). A branch-sites analysis also supports hominoid-specific positive selection ($P < 0.001$; supplementary table S6, Supplementary Material online).

Five codons, R7, V25, G31, T68, and L197 (human coordinates), were identified in these analyses as belonging to the class of codons with dN/dS >1 (supplementary tables S3–S6, Supplementary Material online). Residues V25 and G31 fall within the *Plasmodium* interaction domain of DARC (comprised of residues 8–42), whereas residue R7 sits directly adjacent (fig. 3). Even though no other pathogen is currently known to interact with the N-terminal tail
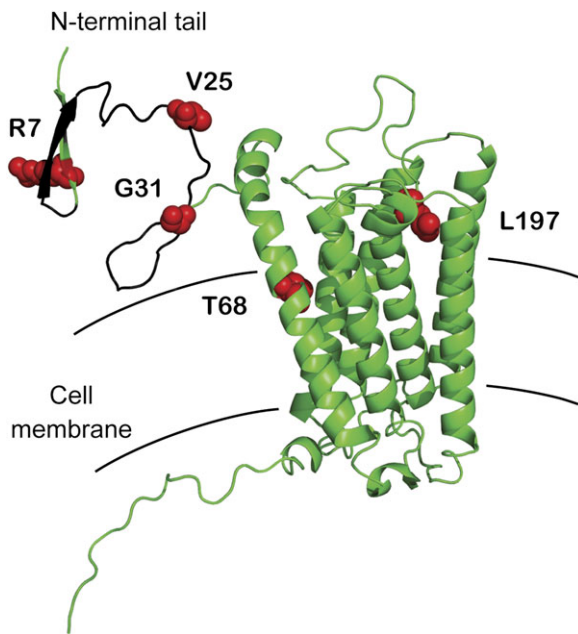
FIG. 3. Three residues under positive selection map to the N-terminal extracellular tail. A structural model of human DARC (de Brevern et al. 2005) is shown, with sites under positive selection indicated in red (human coordinates). The domain that interacts with the *Plasmodium vivax* and *Plasmodium knowlesi* ligand (residues 8–42) is colored black (Chitnis et al. 1996).

of DARC, it is possible that the rapid evolution of these codons may have been driven by a selective pressure other than *Plasmodium*. However, it has been demonstrated that mutating V25 to alanine disrupts the interaction between DARC and the *P. vivax* DBP without affecting chemokine binding (Tournamille et al. 2003, 2005), making *Plasmodium* a plausible selective force to explain the rapid evolution of at least some of these residues in DARC. Interestingly, nature has also made this very mutation (fig. 2C). Gorillas encode an alanine at this residue position, suggesting that *P. vivax* would not be capable of invading gorilla cells. Such a mutation could potentially serve as a species-specific barrier to infection.

Although the zoonotic transmissions of *P. vivax* and *P. knowlesi* from Asian macaques are estimated to have occurred less than 250,000 years ago (Escalante et al. 2005; Mu et al. 2005; Lee et al. 2011), the signatures of positive selection that we observe are consistent with a far more ancient evolutionary "arms race" between *Plasmodium* and *DARC*, one that has lasted on the order of tens of millions of years (Meyerson and Sawyer 2011). In support of an arms race model, the DBP of *P. vivax* also shows evidence for positive selection (Baum et al. 2003). The stronger signal of positive selection of *DARC* observed in the hominoid clade may be evidence for a more intense evolutionary battle in these species, our closest relatives. Indeed, evidence exists that African apes are currently under intense pressure, as sampling projects in wild chimpanzees, bonobos, and gorillas have uncovered *P. vivax* variants and closely related *Plasmodium* species beyond those known to infect humans (Kaiser et al. 2010; Krief et al. 2010; Liu et al. 2010).

Understanding the host genetics that define species tropism of these pathogens is extremely important, both for the development of primate model systems for studying malaria and for understanding how zoonotic processes give rise to new human diseases.

Supplementary Material

Supplementary tables S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

References

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol.* 18:1585–1592.

Baum J, Thomas AW, Conway DJ. 2003. Evidence for diversifying selection on erythrocyte-binding antigens of Plasmodium falciparum and P. vivax. *Genetics* 163:1327–1336.

Chitnis CE, Chaudhuri A, Horuk R, Pogo AO, 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 Kim C, Etchebest C. 2005. A structural model of a seven-transmembrane helix receptor: the Duffy antigen/receptor for chemokine (DARC). *Biochim Biophys Acta.* 1724:288–306.

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.

Haynes JD, Dalton JP, Klotz FW, McGinniss MH, Hadley TJ, Hudson DE, Miller LH. 1988. Receptor-like specificity of a Plasmodium knowlesi malarial protein that binds to Duffy antigen ligands on erythrocytes. *J Exp Med.* 167:1873–1881.

Hurst LD. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* 18:486.

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.

Kaiser M, Löwa A, Ulrich M, et al. (15 co-authors). 2010. Wild chimpanzees infected with 5 Plasmodium species. *Emerg Infect Dis.* 16:1956–1959.

Krief S, Escalante AA, Pacheco MA, et al. (18 co-authors). 2010. On the diversity of malaria parasites in African apes and the origin of Plasmodium falciparum from Bonobos. *PLoS Pathog.* 6:e1000765.

Lee K-S, Divis PCS, Zakaria SK, Matusop A, Julin RA, Conway DJ, Cox-Singh J, Singh B. 2011. Plasmodium knowlesi: reservoir hosts

and tracking the emergence in humans and macaques. *PLoS Pathog*. 7:e1002015.

Liu W, Li Y, Learn GH, et al. (22 co-authors). 2010. Origin of the human malaria parasite Plasmodium falciparum in gorillas. *Nature* 467:420–425.

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

Meny GM. 2010. The Duffy blood group system: a review. *Immunohematology* 26:51–56.

Meyerson NR, Sawyer SL. 2011. Two-stepping through time: mammals and viruses. *Trends Microbiol*. 19:286–294.

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.

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.

Mu J, Joy DA, Duan J, et al. (11 co-authors). 2005. Host switch leads to emergence of Plasmodium vivax malaria in humans. *Mol Biol Evol*. 22:1686–1693.

Perelman P, Johnson WE, Roos C, et al. (14 co-authors). 2011. A molecular phylogeny of living primates. *PLoS Genet*. 7:e1001342.

Tournamille C, Blancher A, Le Van Kim C, Gane P, Apoil PA, Nakamoto W, Cartron J-P, Colin Y. 2004. Sequence, evolution and ligand binding properties of mammalian Duffy antigen/receptor for chemokines. *Immunogenetics* 55:682–694.

Tournamille C, Colin Y, Cartron J-P, Le Van Kim C. 1995. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet*. 10:224–228.

Tournamille C, Filipe A, Badaut C, Riottot M-M, Longacre S, Cartron J-P, Le Van Kim C, Colin Y. 2005. Fine mapping of the Duffy antigen binding site for the Plasmodium vivax Duffy-binding protein. *Mol Biochem Parasitol*. 144:100–103.

Tournamille C, Filipe A, Wasniowska K, Gane P, Lisowska E, Cartron J-P, Colin Y, Le Van Kim C. 2003. Structure-function analysis of the extracellular domains of the Duffy antigen/receptor for chemokines: characterization of antibody and chemokine binding sites. *Br J Haematol*. 122:1014–1023.

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

Wertheimer SP, Barnwell JW. 1989. Plasmodium vivax interaction with the human Duffy blood group glycoprotein: identification of a parasite receptor-like protein. *Exp Parasitol*. 69:340–350.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13:555–556.

Zimmerman PA, Woolley I, Masinde GL, et al. (11 co-authors). 1999. Emergence of FY*A(null) in a Plasmodium vivax-endemic region of Papua New Guinea. *Proc Natl Acad Sci U S A*. 96:13973–13977.