Neglected Burden of Human *vivax* Malaria:

Comparative Analysis of *Plasmodium vivax* and Key Related Species

Plasmodium Comparative Genomics Writing Group:

John Barnwell:	Malaria Research and Development Laboratories Unit, CDC
Jane Carlton:	Dept. Medical Parasitology, NYU School of Medicine
William Collins:	Division of Parasitic Diseases, CDC
Ananias Escalante:	School of Life Sciences, Arizona State University
Jim Mullikin:	Comparative Genomics Unit, NHGRI
Allan Saul:	Laboratory of Malaria and Vector Research, NIAID

August 13, 2007

This proposal represents the first of a trio of white papers which are being proposed by the malaria and anopheline research communities to target specific areas of malaria research that will be significantly enhanced through the availability of genome sequence data. This first paper proposes sequencing isolates of Plasmodium vivax*, the major cause of malaria outside Africa, and several related monkey species which are used as robust and faithful model systems for the study of human malaria. The second white paper is concerned with the generation of additional sequence data from* Plasmodium falciparum*, the major cause of malaria in Africa, which will add to the list of strains that have already been sequenced during the generation of a genetic diversity map. The third white paper addresses the need for additional sequence data of the mosquito vector* Anopheles*, species of which transmit the malaria parasite. All three papers are complimentary and represent a focused attempt to integrate of the needs of the various scientific communities The* Anolpheles *white paper is nearing completion and development of the* P. falciparum *white paper is well underway; both are anticipated to be submitted this fall.*

Table of Contents

I. Executive Summary

The magnitude of the malaria global health problem is unquestionable: 1-3 million deaths per year, ~515 million total clinical cases per year, 2.6 billion people at risk from infection. Malaria is endemic in more than half of the countries of the world, most notably in South America, Southeast Asia, Africa and Oceania, although the majority of malaria fatalities occur in sub-Saharan Africa. Malaria is a disease associated with poverty and is also a major cause of poverty. There is no commercially available vaccine and the parasite has developed resistance to almost all antimalarial drugs commonly used in endemic countries. Although the majority of malaria-related deaths are caused by *Plasmodium falciparum*, *Plasmodium vivax* is the most widely distributed species and the major cause of the disease outside Africa, especially in countries of Asia and the Americas. Since *P. vivax* kills less frequently and is not amenable to some experimental methods, including routine continuous *in vitro* cell culture, this malaria parasite remains a relatively neglected disease in the shadow of *P. falciparum*. Currently the ~27 Mb sequence of a single reference strain of *P. vivax* is available for the malaria research community. The lack of sequence data from additional *P. vivax* genomes undoubtedly hampers whole genome analyses and elucidation of the unique biology associated with the species.

In consultation with members of the greater malaria and vivax-specific research communities, we have developed this document describing a consensus viewpoint of the best way to approach sequencing of additional *Plasmodium* species. The species and isolates proposed for sequencing have been chosen according to three main criteria: (1) their human health and biomedical importance; (2) their ability to contribute data that has the potential to drive compelling scientific questions of direct relevance in the fight against human malaria; and (3) the potential of their genome sequence data to have an immediate and significant impact upon the malaria research community. Accordingly, we propose here a tiered strategy for sequencing the genomes of several isolates of *P. vivax* which differ in certain important phenotypes such as antimalarial drug susceptibility, and the genomes of three key simian species used as models for human malaria from the related 'monkey malaria clade'. We anticipate that the availability of extensive sequence data from *P. vivax* isolates and related monkey malaria species will stimulate further research and overcome some of the limitations imposed by limited access to laboratory-derived parasite material. Specifically, we are requesting the sequencing of:

First priority: Six *P. vivax* genomes to 8x coverage and 40,000 ESTs; three model monkey malaria genomes to 8x coverage and 120,000 ESTs. Sequence requested: 243 Mb and 160,000 ESTs.
Second priority: Seven additional *P. vivax* genomes to 2-3x coverage. Sequence requested: 189 Mb.

Total sequence requested: 432 Mb and 160,000 ESTs. (*N.B.* Average *Plasmodium* genome ~27 Mb)

All genome sequence data will be promptly deposited in GenBank following the NHGRI and NIAID policies on data release. Two NIAID-funded resources will also be used to disseminate the genome sequence data and biological resources to the malaria research community: the malaria genome database *PlasmoDB*, and the malaria repository, *MR4*.

II. Introduction and Rationale

Fundamental questions remain concerning many aspects of malaria: What parasite and human genetic determinants are correlated with disease and with disease severity? What parasite factors influence transmission? When and how did the pathogen originate and how is it evolving? How effective and valuable can we expect new malaria intervention strategies to be? How do new parasite variants arise? How is the parasite's genotype coupled to its phenotype? Genomic technology is now being used to approach these questions. A key contribution has been the generation of genome sequence data from ~15 world-wide geographical isolates of *Plasmodium falciparum* (the most deadly species) and ~9 other species of *Plasmodium*, and their analysis using comparative genomics tools. Many of these *P. falciparum* whole genome sequences have been used in pilot projects to generate diversity maps for prospective association studies between genotype and phenotype, as well as to identify potential novel antigens, study *Plasmodium* genome evolution, and infer phylogenies [1-3].

However, a major gap in *Plasmodium* genomics is the relative lack of sequence data for the second most important human malaria species, *Plasmodium vivax*. *P. vivax* accounts for the majority of malaria outside Africa with approximately 80-100 million cases annually, and more people are at risk from infection with *P. vivax* than with *P. falciparum.* A disease of poor people living on the margins of developing economies, vivax malaria traps many societies in a relentless cycle of poverty [4]. Protective immunity against *P. vivax* is infrequent due to intermittent transmission, and the disease occurs at all ages, though especially in young adult men. Morbidity results from repeated acute febrile episodes of a debilitating intensity that can persist for months. Drug resistance in *P. vivax* is becoming more widespread, hindering management of clinical cases [5]. Since *P. vivax* kills much less frequently and is not amenable as yet to continuous *in vitro* culture, the species remains comparatively neglected in the shadow of *P. falciparum.* Thus, while the genomes of ~15 primarily *in vitro* cultured strains of *P. falciparum* have been wholly or partially completed, data exist for only one *P. vivax* genome (monkey-adapted laboratory strain Salvador I; [6]). We anticipate that as more extensive sequence data from *P. vivax* isolates and related monkey malaria species become available, it will stimulate further research and overcome some of the restrictions imposed by current limited access to laboratory-derived parasite material.

A. Unique aspects of *P. vivax* biology

Phylogenetically and biologically, *P. vivax* and *P. falciparum* are very distant and very different from each other. *P. vivax* parasites predominantly infect reticulocytes, which results in long-term chronic infections and anemia, but less mortality. In contrast, *P. falciparum* parasites invade cells of various ages and cause acute anemia and frequently death. Moreover, *P. vivax* cannot infect **Duffy negative reticulocytes**, and so is exceedingly rare in West Africa where Duffy negativity predominates in the human population. Depending on the geographic isolate, *P. vivax* sporozoites differ in their potential for development to maturity in the liver. Some may initiate a primary blood stage infection, while others may remain in the liver as a resting stage (the **hypnozoite**) until conditions are more favorable for mosquito infection. Thus, in *P. vivax*, there are **relapses** of malaria due to reactivation of the hypnozoite stage in the liver, a specific adaptation to diminished mosquito populations during the winter months that increases the likelihood of transmission when environmental conditions are more favorable. This contrasts with **recrudescences**, the patent reappearance of antigenically distinct populations of parasites in the blood that are produced during the course of an existing infection, and which occur with both *P. falciparum* and *P. vivax*. *P. vivax* populations in different geographical regions and environmental conditions have different time courses of relapse, ranging from very delayed in the northern latitudes (*e.g*, Korea), to very frequent in the tropics of the Pacific and Indonesia. The biological phenomenon of

relapse presents serious challenges to, and is responsible for, vivax malaria being significantly resistant to elimination and control efforts.

B. Drug resistance

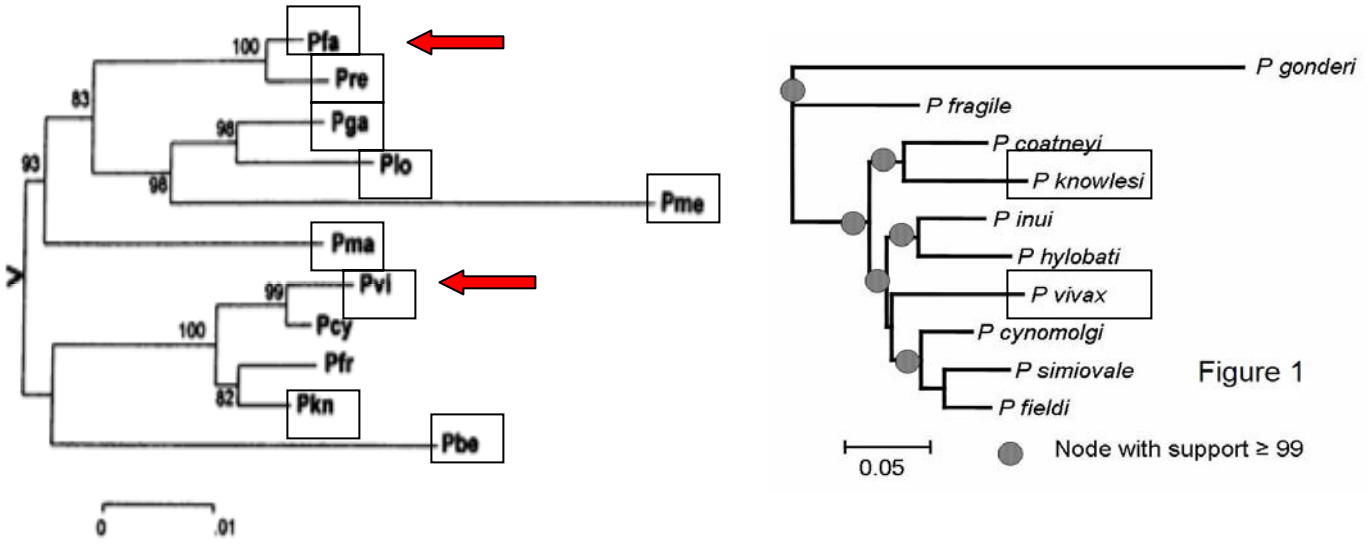
Current treatment for vivax malaria primarily relies upon two antimalarial drugs, chloroquine and primaquine, with the latter reducing the number of subsequent relapses through killing of the hypnozoite stage. Reports of chloroquine resistant *P. vivax* strains being found frequently in several regions of Asia and also South America have been published [7], and likewise, recent reports of primaquine resistance are worrisome [8], although the epidemiology and molecular mechanisms of resistance remain to be determined. Relapse of vivax malaria often confounds drug efficacy tests, since the relapse interval can coincide with the time when recrudescence of a drug resistant parasite would occur following drug treatment. The emergence of drug resistance in *P. vivax* -- particularly to the only class of compounds available for killing the dormant liver stage -- is alarming and of high priority for research.

C. Geographical variation

Despite the essential role that assessing the genetic variation of malaria parasites plays in developing, testing and deploying control interventions, investigation of genetic variation in human *Plasmodium* species to date has been limited and biased toward *P. falciparum.* There have been a few population-based studies of *P. vivax* (*e.g.,* [9-11]), mostly of genes encoding parasite surface proteins or neutral markers such as microsatellites and mitochondrial loci. Significantly, even these limited studies indicate high levels of genetic polymorphism, large numbers of gene duplication events in the species' evolutionary past, and rapid evolution of particularly repetitive tandem repeats in *P. vivax* proteins. (Two studies reporting low levels of genetic variation in *P. vivax* [12, 13] are at odds with these, most likely due to their use of short repeat microsatellites which are less diverse than long repeats because slippage mutations become exponentially more common with an increase in array length [14].) They also indicate a distinct geographic structure with at least two major *P. vivax* "clusters", one in Southeast Asia and another in Melanesia [15]. Thus, biologically and geographically, *P. vivax* is a more ancient and much more diverse parasite than *P. falciparum*, requiring extensive sampling of widely different geographical isolates in order to capture the full range of geographic diversity exhibited by the species.

D. The monkey malaria clade

P. vivax is intimately related to a large clade of malaria parasite species that infect the cercopithicoid (Old World) monkeys and the lesser apes (gibbons) of South Asia, from Taiwan to Sri Lanka and India. The Southeast Asian non-human primate malaria parasites form a monophyletic group containing *P. fragile, P. knowlesi, P. coatneyi, P. inui, P. hylobati, P. simiovale, P. fieldi,* and *P. cynomolgi*, together with *P. vivax.* These species in turn form a divergent sister group with *P. gonderi* (and other species) from African monkeys (**Figure 1**). An important feature of this group of parasites is that several biological traits, such as periodicity, morphology and virulence, have little value for inferring their phylogenetic relationships [16, 17]. For instance, the length of periodicity is a convergent characteristic: the 'quartan' (72-hour erythrocytic cycle) parasites *P. inui* and *P. malariae* do not form a monophyletic group, and *P. knowlesi*, the only 'quotidian' (24-hour erythrocytic cycle) primate parasite shares this characteristic with several species parasitic to rodents. Not all species of the monkey malaria clade are capable of relapse; *P. vivax, P. cynomolgi, P. simiovale* and *P. fieldi* are among those that exhibit this trait. Based on the phylogeny depicted in **Figure 1**, it appears that the capacity to relapse appeared *de novo* in *P. vivax* and the clade including *P. cynomolgi*.



A. **B.**

Figure 1. Two phylogenies showing relationships between *Plasmodium* species, especially the divergence between the *P. falciparum* and *P. vivax* clades. **A.** Phylogenetic relationships among eleven *Plasmodium* species inferred from rRNA gene sequences. The five main clades shown include *P. falciparum* with its sister taxon *P. reichenowi*, a clade containing malaria parasites of birds and lizards, the monkey malaria clade containing *P. vivax*, and the rodent malaria clade showing *P. berghei*. Arrows indicate the divergent positions of *P. falciparum* and *P. vivax*. Abbreviations: Pfa: *P. falciparum*; Pre, *P. reichenowi*; Pga, *P. gallinaceum*; Plo, *P. lophurae*; Pme; *P. mexicanum*; Pma, P. *malariae*; Pvi, *P. vivax*; Pcy, *P. cynomolgi*; Pfr, *P. fragile*; Pkn, *P. knowlesi*; Pbe, *P. berghei*. Taken from Ref. [18]. **B.** Detailed phylogeny of the monkey malaria clade based upon complete mitochondrial genomes and nuclear genes. The tree was inferred with Mr. Bayes using Bayesian methods under an HKY substitution model. The parasites found in Southeast Asian non-human primates form a monophyletic group with *P. vivax*. *P. vivax* appears as a sister taxon of two non-human primate malaria lineages, one that includes *P. cynomolgi*, *P. simiovale*, and *P. fieldi*, and another that includes *P. inui* and the gibbon parasite *P. hylobati*. *P. coatneyi* appears as sister taxon to *P. knowlesi*. Species which have ongoing or completed sequencing projects are shown boxed.

Two of the non-human primate malaria species, *P. cynomolgi* and *P. coatneyi*, are excellent and faithful models for the biology, immunology and pathology of human species with the same phenotypic characteristics, *i.e.*, *P. vivax* and *P. falciparum*, respectively. A third, *P. inui*, is the only other known quartan parasite aside from its human equivalent *P. malariae* (the latter is currently being sequenced at the Wellcome Trust Sanger Institute).

E. Zoonotic *Plasmodium* infections

Of the seven malaria parasite species currently known to naturally infect Asian macaque monkeys, three (*P. cynomolgi*, *P. inui*, and *P. knowlesi*) are capable of infecting humans [19]. The ability of macaques to thrive in human-altered environments has made them the most common non-human primate species in Asia [20]. In addition, several Asian cultures honor centuries-old traditions of human - non-human - primate commensalism (interactions associated with habitually sharing a space); such close interaction may facilitate the emergence of zoonotic malaria parasites. In fact, there have been increasing reports of naturally-acquired human infections with *P. knowlesi* [21, 22]. In line with this apparent frequent host-switching undertaken by members of the monkey malaria clade, *P. vivax* is thought to have arisen in humans through several host switches from Asian macaque monkeys [16, 23].

F. Refractoriness of *P. vivax* to experimental manipulation

There are several reasons why genome sequencing presents one of the easiest and most cost effective methods of obtaining genetic data about *P. vivax*. The parasite cannot be grown in long term *in vitro* culture, although short term culture has been reported. While several groups within the research community have ongoing projects to develop a culture system -- for example, through using stem cell technology to establish and maintain human erythrocytic cell lines for the generation of reticulocytes which can support *P. vivax* growth -- genetic crosses, which have identified candidate loci for several phenotypes in *P. falciparum*, currently are not feasible in *P. vivax* since the cloning and phenotyping of progeny clones would have to be carried out *in vivo*. Long-oligo arrays have been developed through TIGR's Pathogen Genomics Resource Center for the community, but it is unclear at this stage whether gene expression studies or CGH studies will be possible using these arrays due to inherent problems of obtaining enough parasite DNA and RNA that is free from contaminating host material; *P. vivax* infects reticulocytes that contain host DNA and RNA, and parasitemias rarely exceed ~1% due to restriction to this class of erythrocyte. However, association studies that identify links between genetic markers and a particular phenotype in populations of parasites are one of the few methods available to connect phenotype to genotype in *P. vivax*. Using microsatellites developed from the genome project, the first association mapping studies are underway in *P. vivax*. However, such studies would benefit tremendously from the development of a haplotype map, which requires sequencing of more *P. vivax* isolates. Once candidate loci have been identified, recent advances in transient *P. vivax* transfection [24] and heterologous transfection (see **Section VII**) ensure that the next steps to determine candidate gene function and molecular mechanisms are possible.

In summary, *P. vivax* and related monkey malaria species exhibit tremendous phenotypic and genotypic diversity within a monophyletic taxonomic group of relatively recent origin. Issues of spreading drug resistance and zoonotic infections are cause for concern, in addition to the current incidence and prevalence of malaria cases caused by the parasite. Refractoriness of *P. vivax* to standard laboratory manipulation makes genome sequencing a particularly compelling method for investigating its biology and diversity. Comparative genomics will provide invaluable information on, for example, the rate and mode of evolution of proteins involved in biological diversity, with the ultimate goal of generating better methods of disease surveillance and control.

III. Secrets Revealed by the First *P. vivax* Genome Sequence

The *P. vivax* genome sequencing project began at The Institute for Genomic Research (TIGR) in 2002 with the goal of producing a finished sequence at least as good as the sequence of *P. falciparum*. Surplus funds from the US Department of Defense and NIAID, which supported part of the *P. falciparum* genome sequencing project, were used to finance the project [6]. After a nine month halt in 2004 due to diminution of funds, the project was rescued by additional funding from the Burroughs Wellcome Fund and NIAID. The paper describing the genome is under pre-submission consideration at *Nature* (Carlton *et al*. 2007) in conjunction with the paper describing the genome of the monkey malaria parasite and close relative, *P. knowlesi*, undertaken at the Sanger Institute.

Throughout the project, *P. vivax* genome sequence data have been made available to the community and have been used in multiple ways to understand the basic biology of this pathogen. Some of the major findings are:

- Screening of ~330 microsatellites in the *P. vivax* genome has identified ~150 that are polymorphic among eight strains. Many researchers have used these for world-wide population

studies of *P. vivax* that have confirmed the tremendous genetic variation exhibited by the parasite (see for example [10, 25, 26]). Another study has used the microsatellites to determine that *P. vivax* populations emerging from hypnozoites routinely differ from the populations that caused the acute episode, and that activation of heterologous hypnozoite populations is the most common cause of first relapse in patients with vivax malaria [26]. The first association studies using these microsatellites are now underway.

- *P. vivax*-specific genes involved in biologically interesting phenotypes such as red blood cell invasion, antigenic variation, and host-cell interactions have been identified (*e.g.* [27]). In particular, the complete repertoire of the *vir* gene family found in the subtelomeric regions of *P. vivax* chromosomes and implicated in antigenic variation and immune evasion has been characterized (Carlton *et al*. 2007). The *rbp* gene family involved in invasion of reticulocytes was also found to contain many more members than previously thought, dispelling a view that *P. vivax* has a relatively uncomplicated erythrocyte invasion mechanism and likely has alternate invasion pathways similar to *P. falciparum* (Carlton *et al*. 2007).
- Crystal structures and homology models of proteins implicated in drug interactions and drug resistance have been compared between *P. vivax* and *P. falciparum*, in order to predict the *P. vivax* parasite's active sites involved in drug interaction (Carlton *et al*. 2007).
- Candidate loci for the genetic switch to dormancy in *P. vivax* have been identified (Carlton *et al*. 2007).
- A network of *P. vivax* protein interactions from yeast-two hybrid experiments, Rosetta stone analysis and phylogenetic profiling, has been generated and compared to the interactome of *P. falciparum* (Carlton *et al*. 2007).

In addition, comparative analysis with completed *Plasmodium* genomes *P. falciparum*, *P. yoelii* (and other rodent malaria species) and *P. knowlesi*, presented in Carlton *et al*. 2007, provided:

- the fundamental characteristics of mamalian *Plasmodium* genomes (genome size range, number of genes, gene length, GC content, codon bias etc).
- the major *Plasmodium* gene families shared across species, which could be targeted to generate a genus-specific vaccine.
- four-way synteny maps, used to infer the chromosomal rearrangements that have occurred since divergence of the species from a common ancestor and the evolution of the *Plasmodium* genus (**Figure 2**).
- the number of synonymous substitutions per synonymous site (dS) and non-synonymous substitutions per non-synonymous site (dN) between orthologous pairs of *P. vivax* and *P. knowlesi* genes, and used these values to visualize how the mutation rate varies over the length of chromosomes, and identify genes under possible positive selection pressure that may represent infected erythrocyte surface proteins that could be targeted for vaccine development.

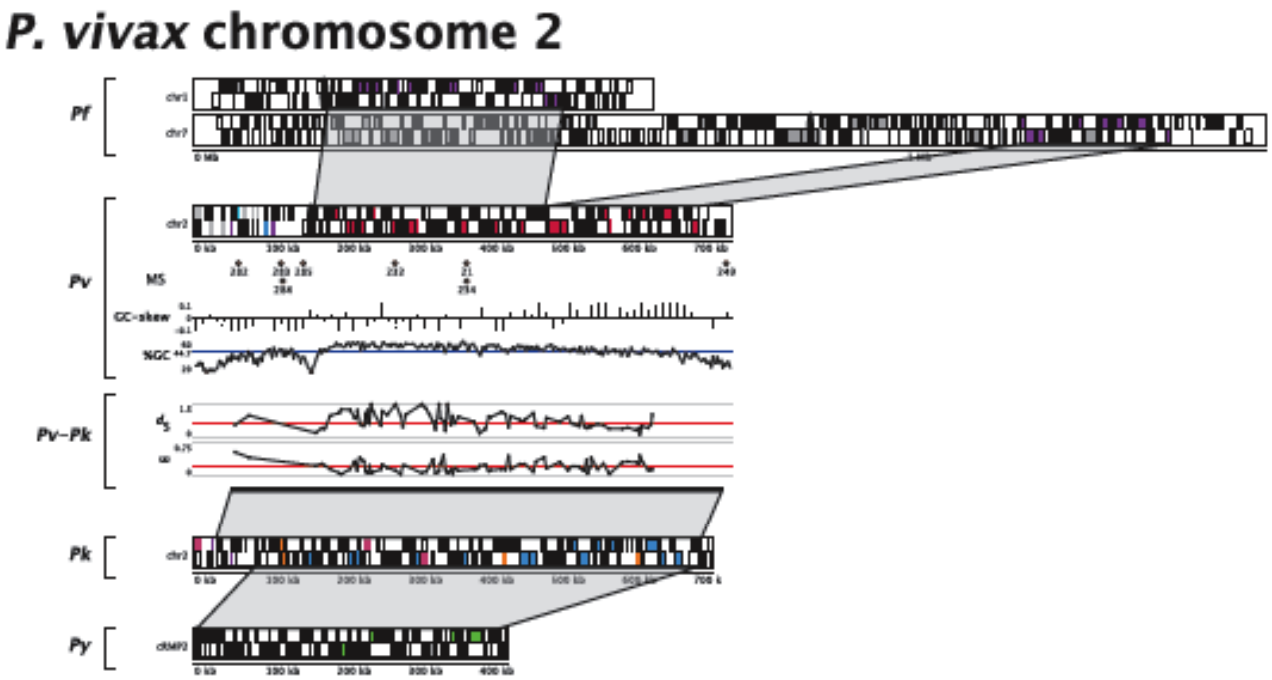


Figure 2. Map of *P. vivax* chromosome 2 showing syntenic regions and orthologous genes between four species of malaria parasite (cRMP refers to a composite rodent malaria chromosome generated from partial sequence data of three rodent malaria species). The plots between *P. vivax* and *P. knowlesi* show: (i) the position of polymorphic microsatellite markers (MS) in the genome; (ii) GC skew along the *P. vivax* chromosome; and (iii) plots of molecular evolution statistics such as dS (a rough approximation to the mutation rate) and dN/dS (areas of the genome under possible positive selection) determined from analysis of orthologs between *P. vivax* and *P. knowlesi.* Taken from the *P. vivax* genome paper (Carlton *et al*. 2007).

A wealth of information has been generated from the first genome sequence of *P. vivax*. However, further analyses, in particular studying the natural variation of *P. vivax* genes such as those involved in invasion have been limited by the lack of sequence data from additional isolates. We hope to ameliorate this situation via the plans outlined in this proposal.

IV. Essential Research into the *Plasmodium* Monkey Malaria Clade

We propose a two-tiered strategy for sequencing several species and isolates of the monkey malaria clade (**Table 1**): **The isolates and species in Tier 1 have priority over the isolates in Tier 2.**

Tier 1A: Due to the biomedical importance of the human malaria parasite *P. vivax*, and the tremendous benefits that will accrue to the malaria research community through access to additional sequence data, the first tier contains isolates of *P. vivax* that differ in several important phenotypes. Generating sequence data from these isolates will have a significant impact on at least three key research areas of *P. vivax* biology: drug resistance, the phenomenon of relapse, and genetic diversity.
Tier 1B: This tier contains three simian malaria species, *P. cynomolgi*, *P. coatneyi* and *P. inui*, which are excellent and robust models for the study of human malaria. Sequencing their genomes will provide genetic data that can be added to the growing list of resources that these model organisms provide, such as *in vitro* culture, transfection and transmission.

Tier 2: This tier comprises additional *P. vivax* isolates to be considered for sequencing, if the results of the studies from the Tier 1A strains provide compelling scientific reasons to generate a more detailed genetic diversity map.

A. The neglected human malaria parasite *P. vivax* (Tiers 1A & 2)

Sequence data from Tier 1A/Tier 2 targets will significantly advance studies of *P. vivax* in four key areas:

1. Study of the relapse stage

The range of *P. vivax* extends from the tropics to the colder northern temperate regions and, as described above, the organism possesses relapse mechanisms that enable it to survive in the primate host through periods of drought and long winters when mosquito vectors for transmission are unavailable. In order to capture the diverse relapse patterns that have evolved as a result of geographic isolation, we need to sequence the genomes of *P. vivax* isolates with different relapse phenotypes:

North Korean strain survives the long winters as a quiescent hypnozoite stage in the liver of the primate host [28]. Once the resting hypnozoite is reactivated, with an ensuing erythrocytic stage infection in late spring, mosquitoes can be infected and successful transmission to a new host achieved. Thus this strain is considered of the long relapse type.

Indonesia XIX strain is an example of parasites found on the islands of the Southwestern Pacific Ocean, which are continually transmitted; the liver stages relapse frequently and almost continuously [29]. Thus this strain is considered of the short relapse type. Additionally, this strain is highly resistant to chloroquine and complements studies on AMRU I described below.

Salvador I, the isolate that has already been sequenced and is mentioned here solely for comparison, is a New World form that relapses infrequently in an erratic pattern. It has evolved in an area of periodic dry and wet seasons during which vectors are not always available.

How will obtaining sequence data from long and short relapse isolates move studies on the nature of *P. vivax* relapse forward? Briefly, screening the Salvador I proteome and association studies using microsatellites are beginning to identify candidate loci that may be involved in the relapse phenotype. The North Korean, AMRU I (see below) and Indonesian XIX strains will provide reference sequences that can be investigated for polymorphisms (SNPs, gene amplifications etc) at these candidate loci.

2. Drug resistance

There are few studies on the epidemiology of drug resistance in *P. vivax*, primarily due to the lack of genetic markers of resistance and the lack of well-defined and validated *in vitro* drug tests for use in the field. Sequencing two laboratory-adapted isolates of *P. vivax* that are highly resistant to chloroquine or primaquine, and which are from two separate continents, will provide reference genome data for downstream studies on the molecular mechanism of drug resistance:

AMRU I is an isolate of *P. vivax* from an Australian soldier infected in Papua New Guinea where chloroquine resistant vivax malaria is highly prevalent. The isolate is extremely resistant to chloroquine, and like Indonesia XIX is a frequent relapsing parasite.

Brazil I was recently isolated from a patient who had traveled to Brazil and whose subsequent malaria, diagnosed in Florida and verified as *P. vivax*, failed to be radically cured by primaquine at the standard dose of 15 mg base daily for 14 days; it is thus considered to be highly resistant [30]. It is, however, chloroquine sensitive.

How will obtaining sequence data from two drug resistant *P. vivax* isolates advance studies concerning the nature of *P. vivax* drug resistance? Briefly, association studies using microsatellites are beginning to identify candidate loci that may be involved in drug resistance. The AMRU I and Brazil I, in comparison

with sequence data from drug sensitive strains will provide reference sequences that can be screened for polymorphisms in or around these candidate loci. These may be, for example, point mutations within drug binding sites, or gene amplification(s) that could result in over-expression of a drug binding target.

3. Genetic diversity studies

To extend our current knowledge of genetic diversity and population structure of *P. vivax*, we propose sequencing two additional isolates which, in combination with the selection of isolates mentioned above from Indonesia, Brazil, El Salvador and North Korea, will drive further genome diversity studies:
India VII was isolated from cryopreserved blood and adapted to a variety of primates [31]. India has the most cases of *P. vivax* annually of any country, and obtaining a reference sequence from this country is of high importance.
Mauritania I was acquired by a traveler in this West African country [32]. Since *P. vivax* occurs very infrequently due to the lack of the Duffy antigen in most West African populations, the isolate is a rare example of an African *P. vivax* strain that occurs among West Africans with Berber and Arab genetic backgrounds. Thus generating its sequence is of high importance.

How will obtaining sequence data from Tier 1 *P. vivax* isolates advance studies on population structure and genetic diversity? Briefly: (i) we will be able to make preliminary statements concerning the gene repertoires (for example the range of antigen genes such as *vir, msp, rbp, trag*) in parasites from seven different countries; (ii) we will obtain a first glimpse of what a typical *P. vivax* parasite from India, West Africa and Brazil looks like in comparison to parasites from other countries, allowing us to answer such questions as: Does the African parasite have additional copies of invasion genes, enabling it to use alternative pathways of invasion that do not use the Duffy antigen receptor, as has been postulated recently in a study of Duffy negative patients from western Kenya infected with *P. vivax* [33]?; (iii) we will be able to determine the rate and mode of evolution of *P. vivax*-specific proteins, allowing identification of groups of genes that may be evolving fast within *P. vivax* compared to other species; and (iv) we will be able to compare genome architectures, including the isochore structure in *P. vivax*; is the isochore structure the same or does it vary geographically? What implications does this have for the evolution of genes found within these regions?

If genetic diversity studies with the limited number of isolates proposed in Tier 1 provides compelling scientific reasons to generate a more detailed genetic diversity map, a highly likely scenario, we suggest that the an additional seven isolate genomes be sequenced at a later stage (**Table 1, Tier 2**): one isolate from Sri Lanka (**Sri Lanka**, to compare with India VII), an isolate from the west coast of South America (**Ecuador**, to compare with Brazil I), an isolate from Vietnam (**Vietnam Palo Alto**, highly virulent), an isolate from Nicaragua (**NICA**, to compare with Salvador I), an isolate from Thailand (**Thai III**, to compare with Vietnam Palo Alto), and a further isolate from Melanesia (**Chesson**, also used as a standard strain for drug studies in humans, to compare with Indonesia XIX and AMRU I). Obtaining low-coverage sequencing of strains located in close geographical proximity to the Tier 1 isolates will also provide preliminary data as to whether any of the Tier 1 isolates are atypical parasites from that particular region. With the genome sequence of a total of 14 *P. vivax* isolates, the vivax research community will start to be able to (i) identify SNPs, indels and gross chromosomal rearrangements and construct a map of genetic diversity/a haplotype map of the species; (ii) calculate allele frequencies and identify patterns of linkage disequilibrium; (iii) estimate recombination rates and patterns; (iv) identify correlates of protein evolution; and (v) identify potential vaccine candidates through identification of genes under selective pressure.

4. Studies on the potential for zoonotic infections

The potential for zoonotic infections transmitted from the monkey malaria clade to humans requires further exploration. Obtaining the genome sequences of several zoonotic (see **Section 4**) parasites has the potential to provide important biological data concerning, for example, the mechanism(s) of invasion utilized by zoonotic parasites, and identifying molecular markers which can be used to monitor zoonotic epidemics.

Brazil VII (**Table 1, Tier 2**) is a recent *P. vivax* isolate that appears to be nearly identical genetically to *P. simium*, a vivax-like parasite isolated from a South American howler monkey several years ago; it is, however, genetically distinct from other human isolates of *P. vivax* from Central and South America. It infects New World monkeys easily. Through parasitological and genetic studies, *P. simium* appears to be *P. vivax* that has adapted to New World primates sometime during the last five hundred years, and thus might also be transmitted to humans as well as between monkeys.

How will obtaining sequence data from Brazil VII advance studies on the nature of *Plasmodium* zoonoses? Since it is the mechanism of invasion, either of liver cells or erythrocytes, which is the likely 'switch' that enables a successful infection of a new host, analysis of genome sequence data from Brazil VII and comparison to Brazil I and other *P. vivax* isolates may reveal novel alleles or identify additional copies of known invasion genes, as well as identify novel gene families which might be involved in zoonotic invasion mechanisms. In addition, gross chromosomal rearrangements or gene duplications that might have occurred during host switching may be identified. Sequencing the Brazil I genome, a strain that is highly similar to a species that infects monkeys, will enable some truly interesting comparisons between *P. vivax* isolates.

B. Monkey models of human *Plasmodium* species (Tier 1B)

We propose sequencing the genomes of three simian malaria species (**Table 1**, **Tier 1B**) which are used as robust and faithful models for the study of *P. falciparum*, *P. vivax* and *P. malariae*. These species already serve as experimental resources for *in vitro* culture, transfection and mosquito transmission of human malaria; obtaining their genome sequences will accelerate use of these model systems within the malaria research community, and facilitate biological experiments that are not readily and easily performed with human malaria species.

1. *Plasmodium cynomolgi*: a robust monkey model for *P. vivax*

Plasmodium cynomolgi possesses many if not most of the phenotypic, biologic and genetic characteristics of *P. vivax.* In fact, it is the closest relative of *P. vivax* and resides in the same subclade in the ancestral tree of the simian malaria parasites (**Figure 1**). It infects both Old World and New World monkeys as well as humans. There is a great deal of experimental information about this species, and it has the same dormant liver stage (the hypnozoite) as *P. vivax*. The range of this parasite found to infect a variety on macaque species stretches from the Philippines to Sri Lanka, and it infects a variety of anopheline mosquitoes. *P. cynomolgi* has been used extensively as a model for *P. vivax* in drug efficacy and biological studies. There are over a dozen different strains of this parasite available to choose for use in studies, although the Berok strain of the parasite is recommended for sequencing because it can also be cultured *in vitro* [34]. It is the only other monkey malaria species known to harbor an isochore genome structure (see **Section V**).

Obtaining the genome sequence of *P. cynomolgi* would enhance the experimental capability of this species manyfold. In the absence of continuous culture, and until it has been developed, there is an urgent need to have a good model system for *P. vivax.* With *P. cynomolgi*, infections of both New World and Old World monkeys (which are more closely related to humans) can be used to study the hypnozoite

stage, to transfect *P. vivax* vaccine candidates for challenge experiments, for homologous gene knock-out studies, and for mosquito transmission studies. In addition, *P. vivax* parasite material such as DNA, RNA and protein can be generated from long term *in vitro* cultures of transgenic *P. cynomolgi* and used in molecular and biochemical assays. However, in order to carry out these sorts of studies, the genome sequence of *P. cynomolgi* is required to determine the genetic relatedness of the two species, and to compare the gene complement. For example, construction of a transgenic line of *P. cynomolgi* carrying a copy of a *P. vivax* gene would be senseless unless the copy number of the *P. cynomolgi* homolog(s) was known. In addition, between-species comparisons of *P. cynomolgi* and *P. vivax* genomes will shed light on (i) genome architecture, such as the degree of synteny and number of orthologous genes, and whether the isochore structure is shared in location and contains orthologous genes in the two species; (ii) gene families that are shared or have undergone expansion in one lineage; and (iii) identify genes which are under positive selection pressure, which can be compared to the genes identified as under positive selection pressure between *P. vivax* and *P. knowlesi* (**Figure 2**).

Of the three models of human malaria presented here, determining the genome sequence of *P. cynomolgi* is the most important because of how it will accelerate use of the species to study the experimentally less tractable *P. vivax*.

2. *Plasmodium coatneyi*: a monkey model for cerebral malaria

Plasmodium coatneyi grows in Old World monkeys and is characterized by marked sequestration of the mature parasite forms in the deep vascular tissues of various organs and tissues -- sites which are very similar to the location of sequestration for *P. falciparum*. *P. coatneyi* sequestration affects the brain in particular, and symptoms of cerebral malaria occur in *Macaca mulatta* and *M. fuscata* monkeys. *P. coatneyi* characteristically produces severe acute and long-term chronic infections, with waves of parasitemia that represent the appearance of new variant gene antigens on the surface of the infected erythrocytes (antigenic variation), highly similar to *P. falciparum* infections. The parasite has variant antigen genes similar to the *SICAvars* found in *P. knowlesi*, which are also functional paralogs of the *var* gene family of *P. falciparum*. *P. coatneyi* is also morphologically similar to *P. falciparum* (excluding the gametocyte stages). *P. coatneyi* is genetically close to and falls in the same subclade as *P. knowlesi*, a parasite with quite different phenotypic and biologic characteristics, such as variable sequestration and a 24 hour asexual growth cycle (*P. knowlesi* has been sequenced at the Wellcome Trust Sanger Institute).

Thus, *P. coatneyi* represents a robust and faithful model species for *P. falciparum* in terms of sequestration, cerebral malaria, antigenic variation, severe pathology (such as acute nephritis, major organ dysfunction and anemia), immunobiology and determination of the genetic characteristics responsible for wide phenotypic differences amongst closely-related sibling species. Arguments for obtaining the genome sequence follow those presented above for *P. cynomolgi*, but in addition identifying the complete complement of genes in the species will enable insight into the evolution of the phenotypes specific to *P. falciparum* and *P. coatneyi* such as sequestration and severe pathology, *i.e* whether they are the result of convergent evolution or loss of the phenotypes in other *Plasmodium* species.

3. *Plasmodium inui*: a monkey model for *P. malariae*

Plasmodium inui infects Old World monkeys and has been experimentally transmitted to humans and *Aotus* monkeys. *P. inui* infections, like those of *P. malariae* in humans, are characterized by being extremely long-term, often lasting for the lifetime of the host, despite lacking a relapse mechanism. Infections frequently last up to 13 or 14 years [35], whereas human malaria parasites (with the exception of *P. malariae*) last one to three years on average. The asexual stages of both *P. malariae* and *P. inui* require approximately 72 hours for development. In addition, chronic infections of *P. inui* cause renal disease in rhesus monkeys, much as *P. malariae* is known to cause chronic nephrosis in human and

nonhuman primates [36]. The parasite is readily transmitted by a variety of anopheline mosquitoes from Taiwan throughout Southeast Asia and on to India. It is quite possible that humans in Southeast Asia could be naturally infected with *P. inui*, but be misdiagnosed as having *P. malariae* because of their morphological similarity. Many different isolates of *P. inui* have been adapted to *M. mulatta* monkeys with over a dozen different strains available. In addition, the OS strain has been experimentally transmitted to humans and to *Aotus* monkeys, and adapted to *in vitro* culture [37]. Sequencing and subsequent investigation of the *P. inui* genome have the potential to explain how a malaria parasite such as *P. malariae* establishes a lifelong relationship with the host, whereas *P. falciparum* and *P. vivax* infections terminate within several years.

Generating the genome sequence of *P. inui* will provide the community with a robust model system with which to study *P. malariae* and the long-term effects of chronic malaria infection. Briefly: (i) novel gene families or regulatory systems may be identified which ensure long-term survival of the parasite; (ii) since *P. inui* may be frequently under-diagnosed in humans, identification of genetic markers which can be used to develop diagnostic kits will enable better tracking and epidemiology of the parasite; and (iii) comparison with genome sequence data from other zoonotic species such as *P. knowlesi* and *P. cynomolgi* may lead to important discoveries concerning, for example, the methods of erythrocyte invasion utilized by zoonotic species.

To summarize, having the genome sequence of these three monkey malaria models at hand will provide the means and incentives for scientists to participate in genetic and biological research on human species of malaria parasite. Monkeys and non-human primates are more appropriate animal models than rodents because of their close phylogenetic relationship to humans. Establishment of such animal models is an essential prerequisite of the research on human diseases such as malaria. Such disease animal models can be used effectively as common research resources. Over 100,000 primates are used in US-based biomedical research each year, because of their utility in medical science.

Thus, generating genome sequence data from 14 *P. vivax* isolates and three monkey malaria models will enable comparative analyses between very many members of the *Plasmodium* genues. For example, there are an increasing number of studies addressing the evolution of introns, codon bias, and amino acid composition of proteins in *Plasmodium* species [38, 39]. This interest is motivated, in part, by the strong codon bias and low GC content observed in many malaria parasites, including *P. falciparum*. However, since *P. vivax* and related species have 40-50% GC content in gene encoding proteins, these species offer a unique opportunity to explore how such changes in base composition affect the evolution of gene encoding proteins in a group of eukaryotic species that are relatively closely related. Understanding how the genome base composition and mutational bias affects the evolution of proteins has been, historically, a controversial issue in molecular evolution [40]. In addition, by comparing two groups of malaria parasites, *P. vivax* and related primate species, and three species of rodent malaria parasite that have been sequenced, we can explore how amino acid substitution rates are affected by differences in genome AT content. We can also explore how differences in AT content affect the evolution of regions of low complexity and the dynamic of microsatellite loci.

Table 1. Characteristics of *Plasmodium* species and isolates proposed in this study.
* Generated from a Thai patient isolate: ~48% of *P. vivax* genes were verified during annotation using these ESTs. ** ~40,000 *P. vivax* ESTs from one of these six strains for gene verification. *** *Anopheles* vector likely to be proposed for sequencing by the *Anopheles* Writing Group.
# Genome project complete. Abbreviations: TBD, to be determined

No.	Species	Isolate	Hosts	Origin	In vitro culture	Phenotype or geography	Laboratory transmission	Strategy	Principal *Anopheles* vector
Ref.	*P. vivax*	Salvador I **(sequence complete)**	Human, *Aotus, Pan, Saimiri*	El Salvador	No	New World species; erratic relapse pattern	Very good	10x WGS; 21,500 ESTs* **(done)**	*An. albimanus****
	TIER 1A: sampling of *P. vivax* isolates for genetic diversity, relapse, and drug resistance studies								
1.	*P. vivax*	North Korean	Human, *Aotus, Pan*	North Korea	No	Long relapse interval	Good	8x WGS**	*An. sinensis*
2.	*P. vivax*	Indonesia XIX	Human, *Aotus Saimiri*	New Guinea	No	Short relapse interval	Fair	8x WGS**	*An. farauti**** *An. epiroticus****
3.	*P. vivax*	AMRU I	Human, *Aotus, Saimiri*	Papua New Guinea	No	Highly chloroquine resistant	Good	8x WGS**	*An. farauti****
4.	*P. vivax*	Brazil I	Human, *Aotus, Saimiri*	Brazil	No	Primaquine resistant	Good	8X WGS**	*An. darlingi*
5.	*P. vivax*	Mauritania I	Human, *Aotus*	Mauritania	No	West African isolate (rare)	Fair	8x WGS**	*An. gambiae#* *An. arabiensis**** *An. fenestus****
6.	*P. vivax*	India VII	Human, *Aotus, Pan, Saimiri*	India	No	Indian isolate	Very good	8x WGS**	*An. stephensi**** *An. culicifacies**** *An. minimus****
	TIER 1B: species used as informative models to study human malaria, and possible zoonoses								
1.	*P. cynomolgi*	Berok	Old World monkeys Humans, *Aotus*	Malaysia	Yes	Model for *P. vivax*; infects both New and Old World monkeys	Very good; infects many lab species	8x WGS; closure; 40,000 ESTs; auto.	*An. dirus**** *An. maculatus**** *An. stephensi**** *An. epiroticus ****

								annotation	
2.	*P. coatneyi*	Type species	Old World monkeys	Malaysia	Not deter-mined	Severe pathology model for *P. falciparum*: sequesters; undergoes antigenic variation; causes cerebral malaria	Very good; infects lab strains of *An. dirus*	8x WGS; closure; 40,000 ESTs; auto. annotation	*An. dirus**** *An. hackeri*
3.	*P. inui*	OS	Old World monkeys Humans, *Aotus*	India	Yes	Model for *P. malariae;* infectious to humans, probably zoonotic; quartan periodicity	Very good, infects many lab species	8x WGS; closure; 40,000 ESTs; auto. annotation	*An. dirus**** *An. maculatus**** *An. stephensi****
	TIER 2: continued sampling of *P. vivax* for genetic diversity studies								
1.	*P. vivax*	Sri Lanka	Human, *Saimiri*	Sri Lanka	No	Sri Lankan isolate	Fair	2-3x WGS	*An. culicifacies**** *An. annularis**** *An. subpictus****
2.	*P. vivax*	Ecuador I	Human *Aotus, Saimiri*	Ecuador	No	S. American west coast isolate	Good	2-3x WGS	*An. darlingi*
3.	*P. vivax*	Vietnam Palo Alto	Human *Aotus*	Vietnam	No	S.E. Asian isolate	No	2-3x WGS	*An. dirus**** *An. minimus***** *An. epiroticus ****
4.	*P. vivax*	Thai III	Human *Aotus, Pan*	Thailand	No	S.E. Asian isolate	Good	2-3x WGS	*An. dirus**** *An. minimus***** *An. maculatus****
5.	*P. vivax*	Chesson	Human *Aotus Pan Saimiri*	New Guinea	No	Melanesian isolate	Good	2-3x WGS	*An. farauti****
6.	*P. vivax*	Brazil VII	Human, *Aotus, Saimiri*	Brazil	No	Anthroponotic; identical to *P. simium*	Very good	2-3x WGS	*An. darlingi*
7.	*P. vivax*	NICA	Human *Aotus*	Nicaragua	No	Transmitted to humans	Good	2-3x WGS	*An. albimanus****

V. Proposal Logistics

A. Parasite material

Unlike *P. falciparum*, *P. vivax* and *P. coatneyi* have not been maintained in continuous culture, and all parasite material including DNA for sequencing is most readily obtained from infection of non-human primates (*P. inui* and *P. cynomolgi* can be grown *in vitro*.) All of the *P. vivax* isolates described above have been adapted to growth in New World monkeys, and *P. coatneyi* is maintained as frozen parasite stocks from rhesus monkeys. These frozen parasite stocks, plus detailed genealogies of both primate passage and mosquito transmission, are available as a unique resource at the Division of Parasitic Diseases, Centers for Disease Control. Most stocks have also been provided to the malaria repository, *MR4*, under a contract with the CDC Foundation. Similarly, the CDC Foundation would be able to provide parasite material for the genome sequencing proposed here via a contract funding mechanism.

Of paramount importance with many parasite genome sequencing projects is the need to limit the amount of contaminating host material in preparations of parasite material for genomic and cDNA library construction. During the *P. vivax* Salvador I sequencing project [6], protocols were perfected which minimized the amount of contaminating monkey DNA, including methods utilizing acid-washed glass beads to remove activated platelets, and cellulose fiber columns, filters and affinity columns to remove monkey white blood cells. These procedures successfully removed greater than 99.7% of host DNA from the Salvador I parasite material, and can be used during preparation of parasite material from the isolates discussed here. In addition, genotyping of monkey-adapted parasite lines has shown that the lines of parasites generated during the adaptation process are genetically homogenous, *i.e.*, they consist of a single clone. The completion of the Salvador I genome sequence has also confirmed these results. Thus, heterozygosity or mixed infections of the isolates discussed here is unlikely to be a problem.

B. Sequencing strategies

The *P. vivax* genome is ~27 Mb, haploid and distributed among 14 chromosomes (all *Plasmodium* species studied so far share genomic characteristics similar to these). The haploidy of *Plasmodium* genomes makes them uniquely amenable to large-scale sequencing. Unlike the *P. falciparum* and rodent malaria parasite genomes, which are extremely biased in their genome content, members of the monkey malaria parasite clade have a more balanced GC content (~35-45%) (**Table 2**). *P. vivax* and *P. cynomolgi* are unique in containing an 'isochore' chromosome structure, *i.e.*, GC-rich regions interspersed with GC-poor regions; in *P. vivax* the GC-poor regions predominate in subtelomeric regions of the chromosomes, which made it more difficult to assemble the ends of *P. vivax* chromosomes during the sequencing project. Due to their phylogenetic relatedness, it is likely that all the species belonging to the monkey malaria clade have similar genome size and GC content, although only *P. vivax* and *P. cynomolgi* have an isochore structure. For the purpose of this proposal we have assumed that the isolates proposed here have an average genome size of ~27 Mb.

1. Genomic libraries

Since the GC content of the *P. coatneyi* and *P. inui* species is sufficiently high that vector/insert instability issues are minimal, we propose that large insert BAC and/or fosmid libraries be prepared for these species. End sequencing of large clones from these libraries will provide scaffolding information for the genome assemblies and, since the availability of DNA from monkey malaria species is limited, a minimal tiling path of clones across each genome in addition to the BAC/fosmid library itself can be provided as a resource for the malaria community (a similar tiling path of 10-12 kb plasmid clones for *P. vivax* Salvador I is available through *MR4*, reagent #MRA 840). The isochore structure of *P. cynomolgi* chromosomes reduces the usefulness of large insert vectors for cloning, since low GC regions in the genomes will be less stable resulting in possible biased libraries. However, low GC isochore regions constitute less than

18% of the *P. vivax* genome, and if this holds true for *P. cynomolgi* then the majority of the species' genome is likely to be stably cloned in large insert vectors and could still be made available to the community as a source for DNA.

Table 2. Genome characteristics of several *Plasmodium* species. *Approximate GC content derived from a limited number of protein coding genes. ** Tandem repeats and low complexity sequences. NK, not known.

Clade	Species	Genome size (Mb)	No. genes	% GC	Repeat content**	Structure
Monkey	*P. vivax*	26.8	~5,400	42.3	Medium/high	Isochore
Monkey	*P. cynomolgi*	NK	NK	34.5*	NK	Isochore
Monkey	*P. coatneyi*	NK	NK	34.5*	NK	Uniform
Monkey	*P. inui*	NK	NK	40.2*	NK	Uniform
Monkey	*P. knowlesi*	23.5	~5,200	37.5	Medium	Uniform
Human	*P. falciparum*	23.3	~5,400	19.4	High	Uniform
Rodent	*P. yoelii*	23.1	~5,800	22.6	High	Uniform

2. Sequence coverage

We propose sequencing the species and isolates in Tier 1 to 8x (full) coverage, and sequencing of the additional isolates in Tier 2 to 2-3x (sample) coverage. Previous low-coverage genome projects (*e.g.* the rodent malaria parasites [41, 42]) have highlighted the inherent problems of analysis of partial sequence data, and thus it is important that at the very least the three simian model parasite genomes be sequenced to full coverage since these will be the reference sequences for each of these species. In addition, we request that the six *P. vivax* isolates in Tier 1 be sequenced to full coverage, since the genetic diversity within *P. vivax* is much greater than seen for example within *P. falciparum*, and genome changes such as gene duplications and gross rearrangements within *P. vivax* isolates are possible and would not be visible through low-coverage sequencing. Moreover, one of the important analyses that can be addressed through the availability of high-quality sequence data from a range of *P. vivax* isolates will be analysis of genes and intergenic regions of the subtelomeres, which are of unique interest since these are where *P. vivax*-specific genes involved in interaction with the host immune system are located (such as the *rbp* and *vir* genes mentioned previously). We recommend sample sequencing for the Tier 2 *P. vivax* isolates since high-coverage is not required for the generation of a diversity map. Obtaining deep coverage of the Tier 1 isolates will thus provide the community with a set of gold-standard genome sequences against which low-coverage genome sequence, such as that obtained from Tier 2 isolates, can be compared.

The decision as to which sequencing platform to use (Sanger versus the newer sequencing technologies, *e.g.* 454 and Solexa) awaits result of analyses recently carried out by the Wellcome Trust Sanger Institute and the Broad, which sequenced the highly repetitive *P. falciparum* genome using new sequencing technologies. The genomes of two members of the monkey malaria clade analyzed so far, *P. vivax* and *P. knowlesi*, show fewer repeats than the genome of *P. falciparum* (**Table 2**), which might favor the use of one of the new sequencing platforms, but a more likely scenario may be a combination of the latter with Sanger sequencing.

3. ESTs/flcDNAs

There have been significant developments in *Plasmodium* full-length (fl) cDNA cloning technology [43], and we would like to propose the construction of flcDNA libraries for paired-end sequencing for several of the species mentioned here. We propose that ~40,000 ESTs/flcDNA sequences be generated from a blood stage cDNA library of *P. vivax*, and ~40,000 ESTs/flcDNA sequences be generated from blood

stage cDNA libraries of *P. coatneyi*, *P. cynomolgi* and *P. inui*. During the Salvador I genome project, ~2,600 gene predictions were verified using 21,500 ESTs generated from a cDNA library from a Thai patient, so transcript verification data for the remaining 2,800 genes is desirable. The 40,000 ESTs/flcDNA sequences for each of the monkey models will also be used for gene model verification during automated annotation.

4. Closure

The *P. vivax* Salvador I draft genome sequence contains gaps. Four of the 28 chromosome ends have been closed, but the majority of the subtelomeric/telomeric regions are fragmented, contained within 2,745 contigs representing 4.2 Mb of sequence. The subtelomeric regions of *P. falciparum* chromosomes harbor species specific genes such as those involved in antigenic variation and immune evasion, and these regions are considered hot-spots of recombination that give rise to novel alleles during mitosis. *P. vivax* subtelomeric regions likely serve a similar function, but verification of this requires that all 14 chromosome ends be complete. We request closure of the 24 remaining Salvador I chromosome ends through HAPPY mapping, or some other targeted approach, that stitches the telomeric/subtlomeric regions together with the main body of each chromosome. Alternatively, should the closure process be more easily performed as part of the sequencing of one of the six Tier 1 isolates, one of these genomes could be closed.

The availability of the *P. vivax* reference sequence, and the high degree of synteny between mammalian *Plasmodium* species [44], will greatly facilitate genome assembly and automated closure of all of the genomes proposed here. However, in the case of the three monkey malaria parasite genomes, additional manual closure may be necessary, including contig assignment to chromosomes through hybridization of contig probes to chromosomes separated by pulsed-field gel electrophoresis, and closing of physical gaps through targeted PCR. The Carlton lab at NYU has considerable expertise in the generation of such physical maps and closure of physical gaps.

5. Annotation

The *P. vivax* Salvador I sequence underwent extensive manual annotation by Bioinformatics Analysts at TIGR, including curation for gene model structure and gene function. Each gene has been manually curated at least three times. Members of the community have commented that the annotation of the *P. vivax* genome appears to be better quality than that of the *P. falciparum* annotation. Thus we propose automated annotation of the *P. coatneyi*, *P. cynomolgi* and *P. inui* genomes using the *P. vivax* and *P. knowlesi* manual annotation as a reference.

C. Resources for the malaria research community

Two NIAID-funded resources will be used to disseminate the genome sequence data and resources to the malaria research community and thus ensure a fast and immediate impact of the sequencing projects proposed here. First, all genome sequence data will be promptly deposited in GenBank following the NHGRI policies on data release (http://www.genome.gov/10506537). Data will also be released to the malaria genome sequence database and data mining resource *PlasmoDB* (http://www.plasmodb.org [45]), part of the NIAID-funded *ApiDB* Bioinformatics Resource Center [46]. This database displays *Plasmodium* sequence data in a format which is easily searched and manipulated by malaria researchers. Second, the NIAID-funded malaria repository *MR4* (http://www.mr4.org/ [47]) provides a convenient and efficient mechanism for archiving and distributing physical material generated by this project. As indicated above, the availability of *P. vivax* material through this portal will be of prime importance since *P. vivax* biological resources, either patient or monkey derived, are not easily obtained.

VI. Community Input and Support

The malaria community is small and the vivax community an even smaller subset. Thus members of the Comparative Genomics Writing Group sought comments and feedback from a wide circle of researchers taken from the extended malaria community concerning the next set of priority *Plasmodium* species to sequence. Comments were provided during the initial drafting stages from:

Professor Richard Carter, University of Edinburgh, UK
Professor Robert Sinden, Imperial College, London UK
Dr Thomas McCutchan, Laboratory of Malaria and Vector Research, NIH
Dr Xin-zhuan Su, Laboratory of Malaria and Vector Research, NIH

All of them favored sequencing further *Plasmodium* genomes, and all supported the proposal for concentrating on species that would add to the value of comparative studies. In addition to sequencing more *P. vivax* and monkey isolates, both Professors Carter and Sinden suggested sequencing further rodent isolates to aid in the interpretation of genetic crossing experiments aimed at mapping growth and immunological phenotypes of rodent malaria to their genotypes. The writing group considered this suggestion but decided that this had a lower priority than the work proposed in this white paper at this time, since the monkey-adapted lines of *P. vivax* and the simian model species such as *P. coatneyi* and *P. cynomolgi* provide excellent systems for the study of human malaria, and monkeys are more appropriate animal models than rodents because of their close phylogenetic relationship to humans.

A draft of the proposal was distributed by e-mail to more than ten researchers in the vivax community, and the following scientists provided very useful written comments:

- Prof Richard Carter, University of Edinburgh. Considerable background in *P. vivax* biology and control.
- Dr Qin Cheng, Australian Army Malaria Unit. Drug resistance in *P. vivax* and interests in the biology of *P. vivax* in the South Pacific.
- Dr Chetan Chitnis, Institute for Genetic Engineering and Biotechnology, New Delhi, India. Actively working on *P. vivax* vaccines and biology.
- Dr Socrates Herrera, Immunology Institute, Universidad del Valle, Cali, Colombia. Vivax biology and immunology in South America.
- Professor David Conway, MRC Unit, The Gambia. Expert on *Plasmodium* parasite evolution, selection and biology.

There was universal support amongst this group for further sequencing of multiple *P. vivax* isolates and the proposed monkey malaria species. None of these experts proposed major changes to the proposal. Several made suggestions regarding priorities of the work and specific isolates that should be considered. These suggestions were considered in developing the list in **Table 1**. In addition, members of the *P. falciparum* writing group who are drafting a white paper identifying areas of *P. falciparum* research that that will be significantly enhanced through the availability of genome sequence data, are highly supportive of this proposal (**Appendix 1**). A writing group is currently working on an *Anopheles* white paper (some of the species that are being proposed for sequencing are indicated in **Table 1**), and both *Plasmodium* groups and the malaria community will be coordinating with the *Anopheles* community. Finally, colleagues at the Pathogen Sequencing Unit of the Wellcome Trust Sanger Institute, in particular Drs. Matt Berriman and Arnab Pain, provided verbal support for the sequencing of members of the monkey malaria clade. They are currently working with members of the *Plasmodium* Comparative Genomics Writing Group to publish side-by-side papers describing the first genomes of *P. vivax* Salvador

I and *P. knowlesi*. The PSU have no plans to sequence further members of the monkey malaria clade, thus there is no duplication of effort or overlap with this proposal (**Appendix 2**).

The writing group also considered sequencing clinical isolates before and after adaptation to a New World monkey host, out of concern voiced by some in the Pathogens and Vectors group that adaptation of the proposed isolates to growth in monkeys may have been accompanied by significant genome changes. All members of the writing group agreed that the greatly increased cost, time expended and difficulty of such experiments would be significant and the premise of substantial genome changes during adaptation is not likely to be be valid for species of the monkey malaria clade. Specifically:

- The monkey malaria clade is highly zoonotic, with multiple leaps between human and non-human primate hosts proposed to have occurred during its evolutionary history [16, 23]. There are several examples of monkey malaria parasites infecting humans, and of *P. vivax*-like parasites infecting monkeys, thus it would seem that the parasites require little adaptation in order to switch from one host to another, *i.e.*, there is little genetic information lost or gained in the process, simply selection of exisiting sub-population(s) of parasites.
- Both patient samples and their adapted lines would need to be sequenced to detect changes resulting from adaptation, and this would take several attempts since adaptations take 9-12 months and are not always successful (~30-40% success rate). Lack of success is also not a measure of the lack of adaptability of the patient sample (which might suggest that significant changes in the genome *had* to occur in order for adaptation to occur), but more often due to the quality of the clinical sample.
- Several of the *P. vivax* isolates mentioned in this proposal have been used to infect chimpanzees and humans (either deliberately or by laboratory accident) via mosquitoes, making it unlikely that any vital function has been lost as a result of monkey adaptation.
- Finally, there is a no evidence that significant alterations occur in the genome of *P. vivax* when passaged in non-human primate hosts, other than expected sequence selections that can occur even when *P. vivax* is passaged in different human or vector populations. In contrast, there is considerable evidence for large genomic changes occurring during *in vitro* culture, one of the reasons why sequencing of a clinical isolate of *P. falciparum* has been attempted since all other genome sequences have been from patient isolates adapted to and maintained for generations in *in vitro* culture.

VII. Summary and Final Justification for Proposed Sequencing

The publication of the *P. falciparum* genome in October 2002 [48], made a significant impact on the malaria field. Of the 5,680 scientific papers published since 2003 that report work with *P. falciparum*, more than 1,000 cite the genome paper. In contrast, only ~970 papers have been published dealing with any aspect of *P. vivax* during this period. The problems associated with working on *P. vivax* are partly to blame: the parasite cannot be routinely cultured in the laboratory, and the primate models, while important, are costly and more difficult to work with than other lab animals such as rodents. Our prime motivation for this proposal is the conviction that the ready availability of genomic sequences and resources from multiple *P. vivax* isolates and from other members of the monkey malaria clade will constitute an enormous improvement to this situation. As an example, recent advances in *Plasmodium* heterogeneous transfection technology mean that *P. vivax* gene products can now be investigated as drug or vaccine targets without needing continuous access to a source of *P. vivax* parasites (**Box 1**). The availability of genome sequence data from *P. vivax* and simian malaria parasite species proposed here, will enable extension of such transgenic vaccine studies to simian malaria transgenic model systems in New World, and more importantly, Old World primate hosts, which are the most relevant physiologically,

biologically and immunologically for studying human malaria parasites and human host responses [24, 49].

Box 1. *The utility of transgenic models in malaria vaccine research.*

Since the GC content of P. vivax *is more balanced than for* P. falciparum *and therefore the amino acid abundance more closely reflects that of other organisms, it is often easier to express proteins of* P. vivax *than of* P. falciparum*. Frequently the native sequences can be used directly without the need for extensive re-engineering of codon usage and the generation of synthetic genes that is almost always required for* P. falciparum*. This was exemplified recently during experiments with the leading mosquito-stage vaccine candidate Pvs25 from* P. vivax*. A prime difficulty was in testing the vaccine, which required either blood from P. vivax infected chimpanzees or from naturally infected human patients, neither of which are readily available. In order to circumvent this problem, a rodent malaria parasite was transfected with the Pvs25 gene from* P. vivax *(Ramjanee et al., Vaccine 25: 886-94, 2007), and anti–P. vivax Pvs25 antisera effectively blocked transmission of this recombinant parasite to mosquitoes. Thus, using transgenic rodent parasites expressing* P. vivax *transmission-blocking immunogens has become a new laboratory tool for investigating transmission-blocking activity. This type of transgenic research can now be extended to* P. knowlesi *or* P. cynomolgi*, providing more relevant host-parasite models for studying* P. vivax *and other human malaria species.*

While the availability of the Salvador 1 sequence, released to the community through *PlasmoDB* prior to final publication, has jump-started *P. vivax* research, the availability of sequences from several *P. vivax* isolates with different phenotypes and from several monkey malaria models will enable an altogether higher and more sophisticated level of analysis, since it will provide the context for understanding where sequence conservation or divergence is critical (**Figure 3**), and provide initial insights into sequence/function relationships. Moreover, the genomes proposed for sequencing here have a markedly different base composition from the already-sequenced *P. falciparum* and rodent malaria parasite genomes (which comprise the majority of the genomes sequenced so far), making comparisons within *P. vivax* and the monkey malaria parasites, and between these species and *P. falciparum* and the rodent malarias, particularly informative. For example, *Plasmodium* intergenic regions show low sequence conservation when aligned, yet intron/exon boundaries are highly conserved between species. Recent studies using alignment of ESTs to predicted *P. falciparum* gene models show that there is approximately ~25% error in the assignment of intron/exon boundaries in the *P. falciparum* genome annotation. Many of these can be corrected by comparison between orthologous gene structures across species (**Figure 3**). This is most powerful when there are genomes from multiple species with different GC content. Thus we predict that the availability of the sequence data generated by this proposal will lead to a substantial interest in both the level and success of research into *P. vivax* and comparative studies with other human malaria species.

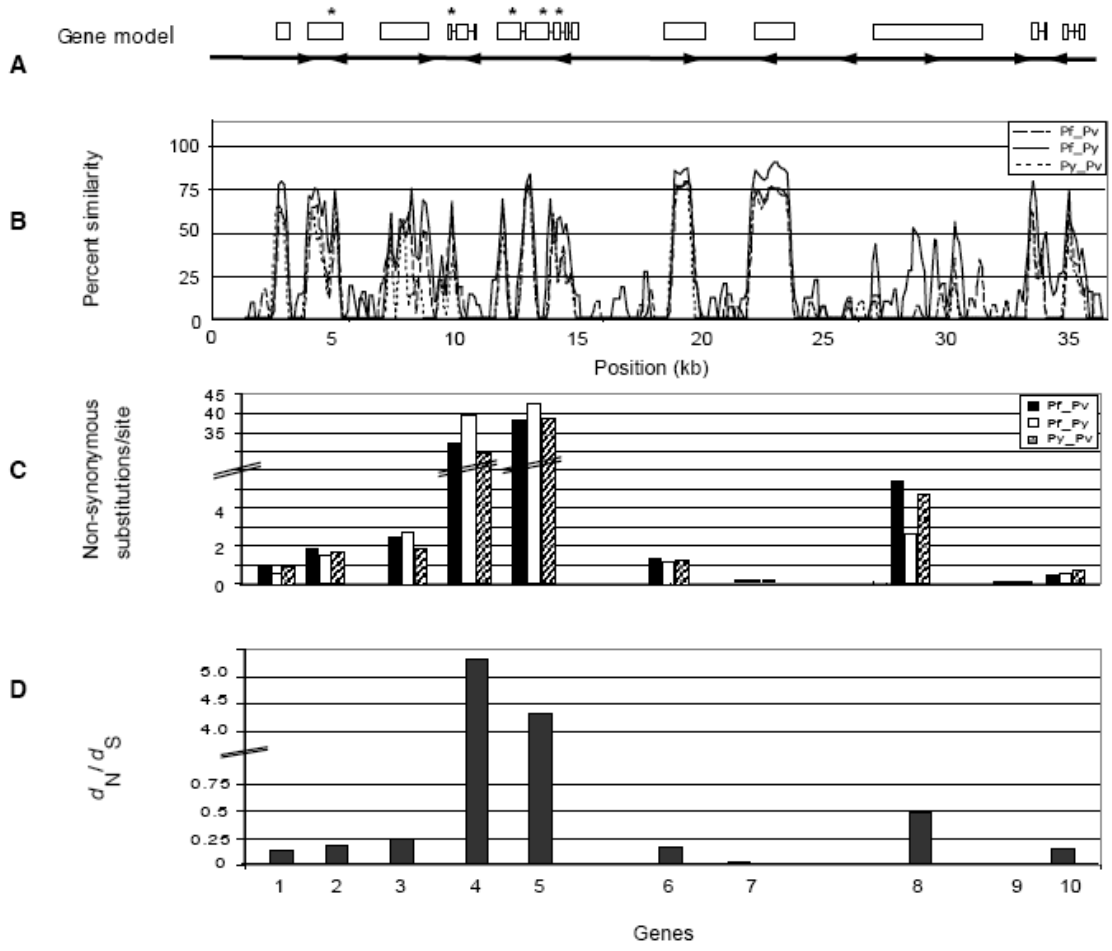


Figure 3. Comparison of 35 kb syntenic region between three *Plasmodium* species, *P. vivax, P. falciparum* and *P. yoelii*. The predicted gene models are shown in panel A, percent similarity at the DNA level in panel C, and non-synonymous substitutions per site and the ratio of dN to dS for ten genes in panels C and D, respectively. As can be seen, intergenic regions between *Plasmodium* species are less conserved than coding regions. Five of the genes indicated with an asterix were refined based upon manual inspection of the orthologous regions in *P. vivax* and *P. yoelii*. Pair-wise calculation of the number of non-synonymous substitutions per site in the ten genes reveal different patterns of sequence conservation or divergence. Note that the mutation saturation observed in synonymous sites between the species means that accurate determination of dN/dS between such divergent species is not possible.

Appendices

Appendix 1.

Letter of support from the *P. falciparum* writing group.

Appendix 2.

Letter of support from the Pathogen Sequencing Unit at the Wellcome Trust Sanger Institute.

References

1. Jeffares, D.C., A. Pain, A. Berry, A.V. Cox, J. Stalker, C.E. Ingle, A. Thomas, et al., *Genome variation and evolution of the malaria parasite Plasmodium falciparum.* Nat Genet, 2007. **39**(1): p. 120-5.
2. Mu, J., P. Awadalla, J. Duan, K.M. McGee, J. Keebler, K. Seydel, G.A. McVean, et al., *Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome.* Nat Genet, 2007. **39**(1): p. 126-30.
3. Volkman, S.K., P.C. Sabeti, D. DeCaprio, D.E. Neafsey, S.F. Schaffner, D.A. Milner, Jr., J.P. Daily, et al., *A genome-wide map of diversity in Plasmodium falciparum.* Nat Genet, 2007. **39**(1): p. 113-9.
4. Mendis, K., B.J. Sina, P. Marchesini, and R. Carter, *The neglected burden of Plasmodium vivax malaria.* Am J Trop Med Hyg, 2001. **64**(1-2 Suppl): p. 97-106.
5. Baird, J.K., E. Schwartz, and S.L. Hoffman, *Prevention and treatment of vivax malaria.* Curr Infect Dis Rep, 2007. **9**(1): p. 39-46.
6. Carlton, J., *The Plasmodium vivax genome sequencing project.* Trends Parasitol, 2003. **19**(5): p. 227-31.
7. Baird, J.K., *Chloroquine resistance in Plasmodium vivax.* Antimicrob Agents Chemother, 2004. **48**(11): p. 4075-83.
8. Collins, W.E. and G.M. Jeffery, *Primaquine resistance in Plasmodium vivax.* Am J Trop Med Hyg, 1996. **55**(3): p. 243-9.
9. Figtree, M., C.J. Pasay, R. Slade, Q. Cheng, N. Cloonan, J. Walker, and A. Saul, *Plasmodium vivax synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA 1 and MSP 1 genes.* Mol Biochem Parasitol, 2000. **108**(1): p. 53-66.
10. Imwong, M., D. Sudimack, S. Pukrittayakamee, L. Osorio, J.M. Carlton, N.P. Day, N.J. White, et al., *Microsatellite variation, repeat array length, and population history of Plasmodium vivax.* Mol Biol Evol, 2006. **23**(5): p. 1016-8.
11. Jongwutiwes, S., C. Putaporntip, T. Iwasaki, M.U. Ferreira, H. Kanbara, and A.L. Hughes, *Mitochondrial Genome Sequences Support Ancient Population Expansion in Plasmodium vivax.* Mol Biol Evol, 2005. **22**(8): p. 1733-1739.
12. Leclerc, M.C., P. Durand, C. Gauthier, S. Patot, N. Billotte, M. Menegon, C. Severini, et al., *Meager genetic variability of the human malaria agent Plasmodium vivax.* Proc Natl Acad Sci U S A, 2004. **101**(40): p. 14455-60.
13. Lim, C.S., L. Tazi, and F.J. Ayala, *Plasmodium vivax: recent world expansion and genetic identity to Plasmodium simium.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15523-8.
14. Russell, B., R. Suwanarusk, and U. Lek-Uthai, *Plasmodium vivax genetic diversity: microsatellite length matters.* Trends Parasitol, 2006. **22**(9): p. 399-401.
15. Cornejo, O.E. and A.A. Escalante, *The origin and age of Plasmodium vivax.* Trends Parasitol, 2006. **22**(12): p. 558-63.
16. Escalante, A.A., O.E. Cornejo, D.E. Freeland, A.C. Poe, E. Durrego, W.E. Collins, and A.A. Lal, *A monkey's tale: the origin of Plasmodium vivax as a human malaria parasite.* Proc Natl Acad Sci U S A, 2005. **102**(6): p. 1980-5.
17. Escalante, A.A., A.A. Lal, and F.J. Ayala, *Genetic polymorphism and natural selection in the malaria parasite Plasmodium falciparum.* Genetics, 1998. **149**(1): p. 189-202.
18. Escalante, A.A. and F.J. Ayala, *Phylogeny of the malarial genus Plasmodium, derived from rRNA gene sequences.* Proc Natl Acad Sci U S A, 1994. **91**(24): p. 11373-7.
19. Coatney, G.R., W.E. Collins, M. Warren, and C. P.G., *The Primate Malarias*. 1971, Washington, D.C.: U.S. Department of Health, Education and Welfare.

20. *Conservation Assessment and Management Plan for the Primates of Indonesia*. Final Report 2001, Conservation International Indonesia, Taman Safari Indonesia.
21. Jongwutiwes, S., C. Putaporntip, T. Iwasaki, T. Sata, and H. Kanbara, *Naturally acquired Plasmodium knowlesi malaria in human, Thailand.* Emerg Infect Dis, 2004. **10**(12): p. 2211-3.
22. Singh, B., L. Kim Sung, A. Matusop, A. Radhakrishnan, S.S. Shamsul, J. Cox-Singh, A. Thomas, et al., *A large focus of naturally acquired Plasmodium knowlesi infections in human beings.* Lancet, 2004. **363**(9414): p. 1017-24.
23. Mu, J., D.A. Joy, J. Duan, Y. Huang, J. Carlton, J. Walker, J. Barnwell, et al., *Host Switch Leads to Emergence of Plasmodium vivax Malaria in Humans.* Mol Biol Evol, 2005. **22**(8): p. 1686-1693.
24. Pfahler, J.M., M.R. Galinski, J.W. Barnwell, and M. Lanzer, *Transient transfection of Plasmodium vivax blood stage parasites.* Mol Biochem Parasitol, 2006. **149**(1): p. 99-101.
25. Ferreira, M.U., N.D. Karunaweera, M. da Silva-Nunes, N.S. da Silva, D.F. Wirth, and D.L. Hartl, *Population structure and transmission dynamics of Plasmodium vivax in rural Amazonia.* J Infect Dis, 2007. **195**(8): p. 1218-26.
26. Imwong, M., G. Snounou, S. Pukrittayakamee, N. Tanomsing, J.R. Kim, A. Nandy, J.P. Guthmann, et al., *Relapses of Plasmodium vivax Infection Usually Result from Activation of Heterologous Hypnozoites.* J Infect Dis, 2007. **195**(7): p. 927-33.
27. Mongui, A., O. Perez-Leal, J. Rojas-Caraballo, D.I. Angel, J. Cortes, and M.A. Patarroyo, *Identifying and characterising the Plasmodium falciparum RhopH3 Plasmodium vivax homologue.* Biochem Biophys Res Commun, 2007. **358**(3): p. 861-6.
28. Collins, W.E., J.C. Skinner, W.A. Krotoski, F.B. Cogswell, R.W. Gwadz, J.R. Broderson, N.S. Ma, et al., *Studies on the North Korean strain of Plasmodium vivax in Aotus monkeys and different anophelines.* J Parasitol, 1985. **71**(1): p. 20-7.
29. Collins, W.E., J.S. Sullivan, D.J. Fryauff, J. Kendall, V. Jennings, G.G. Galland, and C.L. Morris, *Adaptation of a chloroquine-resistant strain of Plasmodium vivax from Indonesia to New World monkeys.* Am J Trop Med Hyg, 2000. **62**(4): p. 491-5.
30. Nayar, J.K., R.H. Baker, J.W. Knight, J.S. Sullivan, C.L. Morris, B.B. Richardson, G.G. Galland, et al., *Studies on a primaquine-tolerant strain of Plasmodium vivax from Brazil in Aotus and Saimiri monkeys.* J Parasitol, 1997. **83**(4): p. 739-45.
31. Sullivan, J.S., E. Strobert, C. Yang, C.L. Morris, G.G. Galland, B.B. Richardson, A. Bounngaseng, et al., *Adaptation of a strain of Plasmodium vivax from India to New World monkeys, chimpanzees, and anopheline mosquitoes.* J Parasitol, 2001. **87**(6): p. 1398-403.
32. Collins, W.E., P. Nguyen-Dinh, J.S. Sullivan, C.L. Morris, G.G. Galland, B.B. Richardson, and S. Nesby, *Adaptation of a strain of Plasmodium vivax from Mauritania to New World monkeys and anopheline mosquitoes.* J Parasitol, 1998. **84**(3): p. 619-21.
33. Ryan, J.R., J.A. Stoute, J. Amon, R.F. Dunton, R. Mtalib, J. Koros, B. Owour, et al., *Evidence for transmission of Plasmodium vivax among a duffy antigen negative population in Western Kenya.* Am J Trop Med Hyg, 2006. **75**(4): p. 575-81.
34. Nguyen-Dinh, P., A.L. Gardner, C.C. Campbell, J.C. Skinner, and W.E. Collins, *Cultivation in vitro of the vivax-type malaria parasite Plasmodium cynomolgi.* Science, 1981. **212**(4499): p. 1146-8.
35. Schmidt, L.H., R. Fradkin, J. Harrison, R.N. Rossan, and W. Squires, *The course of untreated Plasmodium inui infections in the rhesus monkey (Macaca mulatta).* Am J Trop Med Hyg, 1980. **29**(2): p. 158-69.
36. Nimri, L.F. and N.H. Lanners, *Immune complexes and nephropathies associated with Plasmodium inui infection in the rhesus monkey.* Am J Trop Med Hyg, 1994. **51**(2): p. 183-9.
37. Nguyen-Dinh, P., C.C. Campbell, and W.E. Collins, *Cultivation in vitro of the quartan malaria parasite Plasmodium inui.* Science, 1980. **209**(4462): p. 1249-51.

38. DePristo, M.A., M.M. Zilversmit, and D.L. Hartl, *On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins.* Gene, 2006. **378**: p. 19-30.
39. Roy, S.W. and D.L. Hartl, *Very little intron loss/gain in Plasmodium: intron loss/gain mutation rates and intron number.* Genome Res, 2006. **16**(6): p. 750-6.
40. Sueoka, N., *On the genetic basis of variation and heterogeneity of DNA base composition.* Proc Natl Acad Sci U S A, 1962. **48**: p. 582-92.
41. Carlton, J.M., S.V. Angiuoli, B.B. Suh, T.W. Kooij, M. Pertea, J.C. Silva, M.D. Ermolaeva, et al., *Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii.* Nature, 2002. **419**(6906): p. 512-9.
42. Hall, N., M. Karras, J.D. Raine, J.M. Carlton, T.W. Kooij, M. Berriman, L. Florens, et al., *A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses.* Science, 2005. **307**(5706): p. 82-6.
43. Watanabe, J., H. Wakaguri, M. Sasaki, Y. Suzuki, and S. Sugano, *Comparasite: a database for comparative study of transcriptomes of parasites defined by full-length cDNAs.* Nucleic Acids Res, 2007. **35**(Database issue): p. D431-8.
44. Carlton, J.M., M.R. Galinski, J.W. Barnwell, and J.B. Dame, *Karyotype and synteny among the chromosomes of all four species of human malaria parasite.* Mol Biochem Parasitol, 1999. **101**(1-2): p. 23-32.
45. Stoeckert, C.J., Jr., S. Fischer, J.C. Kissinger, M. Heiges, C. Aurrecoechea, B. Gajria, and D.S. Roos, *PlasmoDB v5: new looks, new genomes.* Trends Parasitol, 2006. **22**(12): p. 543-546.
46. Aurrecoechea, C., M. Heiges, H. Wang, Z. Wang, S. Fischer, P. Rhodes, J. Miller, et al., *ApiDB: integrated resources for the apicomplexan bioinformatics resource center.* Nucleic Acids Res, 2007. **35**(Database issue): p. D427-30.
47. Adams, J.H., Y. Wu, and A. Fairfield, *Malaria Research and Reference Reagent Resource Center.* Parasitol Today, 2000. **16**(3): p. 89.
48. Gardner, M.J., N. Hall, E. Fung, O. White, M. Berriman, R.W. Hyman, J.M. Carlton, et al., *Genome sequence of the human malaria parasite Plasmodium falciparum.* Nature, 2002. **419**(6906): p. 498-511.
49. Kocken, C.H., H. Ozwara, A. van der Wel, A.L. Beetsma, J.M. Mwenda, and A.W. Thomas, *Plasmodium knowlesi provides a rapid in vitro and in vivo transfection system that enables double-crossover gene knockout studies.* Infect Immun, 2002. **70**(2): p. 655-60.