Infection, Genetics and Evolution 40 (2016) 243–252

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Characterizing the genetic diversity of the monkey malaria parasite *Plasmodium cynomolgi*

Patrick L. Sutton [a,1], Zunping Luo [a], Paul C.S. Divis [b,c], Volney K. Friedrich [d,2], David J. Conway [b,c], Balbir Singh [c], John W. Barnwell [e], Jane M. Carlton [a], Steven A. Sullivan [a,*]

[a] Center for Genomics and Systems Biology, Department of Biology, New York University, 12 Waverly Place, New York, NY 10003, United States
[b] Pathogen Molecular Biology Department, London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, United Kingdom
[c] Malaria Research Centre, Faculty of Medicine and Health Sciences, University Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
[d] Department of Anthropology, New York University, 38 Waverly Place, New York, NY 10003, United States
[e] Laboratory Research and Development Unit, Malaria Branch, Centers for Disease Control and Prevention, Atlanta, GA, United States

ARTICLE INFO

Article history:
Received 4 January 2016
Received in revised form 1 March 2016
Accepted 2 March 2016
Available online 14 March 2016

Keywords:
Plasmodium cynomolgi
Plasmodium vivax
Genetic variation
Microsatellite repeats
Evolution

ABSTRACT

Plasmodium cynomolgi is a malaria parasite that typically infects Asian macaque monkeys, and humans on rare occasions. *P. cynomolgi* serves as a model system for the human malaria parasite *Plasmodium vivax*, with which it shares such important biological characteristics as formation of a dormant liver stage and a preference to invade reticulocytes. While genomes of three *P. cynomolgi* strains have been sequenced, genetic diversity of *P. cynomolgi* has not been widely investigated. To address this we developed the first panel of *P. cynomolgi* microsatellite markers to genotype eleven *P. cynomolgi* laboratory strains and 18 field isolates from Sarawak, Malaysian Borneo. We found diverse genotypes among most of the laboratory strains, though two nominally different strains were found to be genetically identical. We also investigated sequence polymorphism in two erythrocyte invasion gene families, the reticulocyte binding protein and Duffy binding protein genes, in these strains. We also observed copy number variation in *rbp* genes.

1. Introduction

The apicomplexan parasite *Plasmodium cynomolgi* causes malaria in Asian monkeys as well as experimental and rare zoonotic infections in humans (Coatney et al., 1971; Eyles et al., 1960; Garnham, 1966; Ta et al., 2014). It shares important biologic features with its sister taxon, the human malaria parasite *Plasmodium vivax*, including a dormant liver stage (Krotoski et al., 1982a; Krotoski et al., 1982b), a preference for invading immature red blood cells (Warren et al., 1966), early formation of infectious sexual stages (Dissanaike et al., 1965), modifications of the infected erythrocyte membrane known as Schuffner's stippling (Aikawa et al., 1975), and tertian periodicity (Mulligan, 1935). *P. vivax* is a serious global health problem that threatens more than 50% of the world's population (Guerra et al., 2010). While *P. vivax* cannot be cultured *in vitro*, severely hampering laboratory studies, *P. cynomolgi* has been successfully adapted to short-term *in vitro* culture (Nguyen-Dinh et al., 1981). Thus *P. cynomolgi* represents an ideal model system for investigating *P. vivax* biology, evolution, and pathology, and for identifying novel drugs against the dormant liver stage.

P. cynomolgi was first described by Mayer in 1907, though this 'type strain' was not preserved (Mayer, 1907). In 1935 H.W. Mulligan comprehensively re-described the species and maintained his isolate as the first laboratory strain (Mulligan, 1935). A second laboratory strain of *P. cynomolgi* was described and established by Garnham in 1959, to which he gave subspecies status as *P. cynomolgi bastianellii* (Garnham, 1959). In the decades following its isolation, stocks of Mulligan's strain were distributed among labs worldwide, and a proliferation of alternate strain names appeared in the literature. Initially it was referred to as the 'Rockefeller' strain (Garnham, 1959), or the 'TC' ('typical cynomolgi') strain (Eyles, 1960). In 1961 it as designated the 'M' strain in honor of Mulligan (Coatney et al., 1961; Schmidt et al., 1961), though it has also been referred to simply as 'Mulligan' (Cochrane et al., 1985). Adding further confusion, it was also designated as the 'neotype' strain of *P. cynomolgi* and given subspecies status as *P. cynomolgi cynomolgi* (Eyles et al., 1963). Meanwhile *P. cynomolgi bastianellii* came to be called the B strain (Contacos et al., 1962), because malariologists at the National Institutes of Health did not think that the differences between it and the neotype strain were significant enough to warrant subspecies status. Later, a line of the B strain was termed the 'NIH' strain and also

* Corresponding author.
E-mail address: sullis02@nyu.edu (S.A. Sullivan).
[1] Present address: Acsel Health, 500 5th Ave, Suite 2760, New York, NY 10110, United States.
[2] Present address: Department of Anthropology, University of Pennsylvania Museum of Archaeology and Anthropology, Room 325, 3260 South Street, Philadelphia, PA 19104, United States.

http://dx.doi.org/10.1016/j.meegid.2016.03.009

referred to as the 'London' strain; however, these two strains were found to be different from each other, and the NIH strain was found to be identical to the M strain (Cochrane et al., 1985; Enea et al., 1986; Galinski et al., 1987). Identity of ostensible B and M stocks was more recently noted during genome sequencing of several *P. cynomolgi* strains (Tachibana et al., 2012). Between 1960 and 1967 eight more strains of *P. cynomolgi* were isolated from monkeys captured in Southeast Asia and Sri Lanka and some of these, too, have borne multiple strain names in the literature and sample repositories (*e.g.*, strain Berok has the alias PT-II in the ATCC catalog, apparently after its original pig-tailed macaque host). Geographic origins and aliases of known strains are shown in Fig. 1.

In addition to exhibiting similar biologic features, there is a strong body of evidence that demonstrates a close taxonomic relationship between *P. cynomolgi* and *P. vivax*. In 1991, there were hints of homology between these parasites with respect to the 135-kDa Duffy binding protein (DBP) (Fang et al., 1991). Between 1992 and 1998 this homology would be explicitly demonstrated across several proteins, including the small subunit (SSU) ribosomal RNA (Water et al., 1993), reticulocyte binding protein (RBP) (Galinski et al., 1992), circumsporozoite protein (Escalante et al., 1995), and cytochrome *b* from the linear mitochondrial gene (Escalante et al., 1998). These early investigations set the stage for extensive research that would occur over the next two decades, including the recent publication of the genome of the *P. cynomolgi* B strain from Peninsular Malaysia, which serves as the reference genome for the species, as well as draft genomes of strains Berok and Cambodian from Peninsular Malaysia and western Cambodia, respectively (Tachibana et al., 2012). The genome size of *P. cynomolgi* is ~26.3 Mb, with an estimated GC content of 40.4%, and ~5700 genes predicted on 14 chromosomes. Approximately 96% of genes are orthologous between *P. cynomolgi* and *P. vivax*, with ~3800 genes inferred to be under purifying selection and only 83 genes inferred to be under positive selection, indicating that *P. cynomolgi* will be a useful model for studying *P. vivax* (Tachibana et al., 2012).

An understanding of genetic diversity in *Plasmodium* parasites is crucial for vaccine development, malaria control, and elimination. The population genetics *of P. vivax* has been widely studied in endemic regions of the world (Arnott et al., 2012; Carlton et al., 2013), while the genetic diversity and population structure of *P. cynomolgi* has received much less attention. Diversity in microsatellite (MS) markers is generated by replication slippage/slipped-strand mismatch repair, which causes expansions or contractions of repeat. Approximately 180 polymorphic intergenic MS were identified in *P. cynomolgi* B and Berok strains by comparative genomics (Tachibana et al., 2012). Along with their utility for characterizing and comparing the genetic diversity and structure of populations, MS markers can be particularly useful for

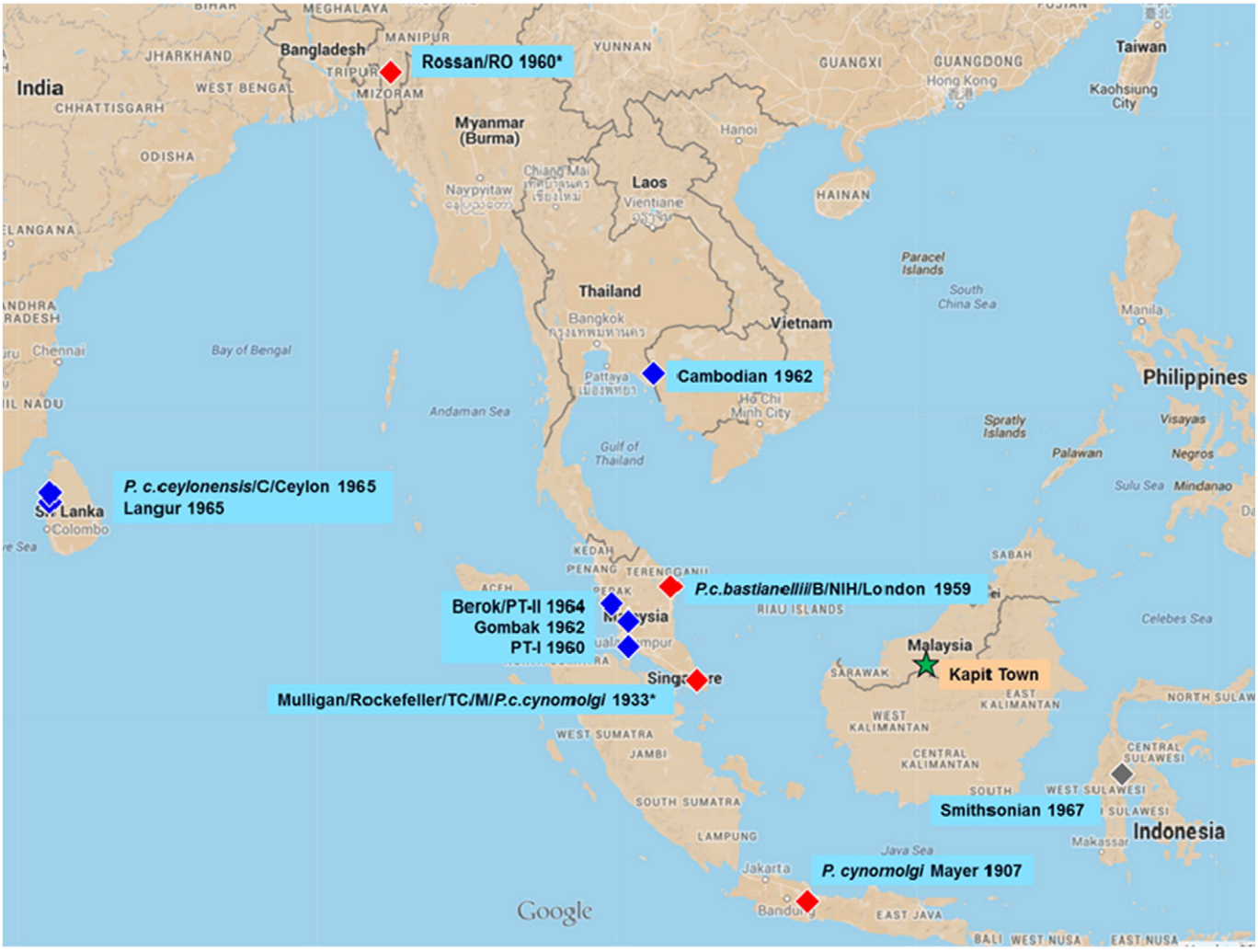

Fig. 1. Geographical origins of *P. cynomolgi* strains. Strain name, aliases, and year of isolation are shown. Blue, red, and gray markers indicate specific, approximate, and dubious origins respectively, gleaned from the literature (see Supplementary materials for references). The reported Sulawesi (Celebes) origin for the Smithsonian strain is dubious because the island lies outside of the range of the host monkey species *Macaca speciosa*. Asterisk (*) indicates that the site was the shipping point of the host monkey, but the location of capture of the monkey is unknown. A star shows the location of Kapit Town in Malaysian Borneo, where *P. cynomolgi* field isolates genotyped in this paper were collected.

describing and resolving phylogeographic data, population bottlenecks, gene flow, and evolutionary history (Anderson et al., 1999, 2000; Branch et al., 2011; Oliveira et al., 1998; Sutton et al., 2011).

The observation that *P. cynomolgi* and *P. vivax* target immature erythrocytes (reticulocytes) for invasion has focused research on receptor–ligand interactions underlying this preference (Warren et al., 1966). The RBL (Reticulocyte Binding-Like) gene superfamily encodes large ligand proteins (230–350 kDa) that are localized to the apical membrane of the invasive stage of the parasite (Galinski et al., 1992; Rayner et al., 2004). Though all RBL genes share a conserved two-exon structure, subfamilies have different target cell preferences in different *Plasmodium* species. The reticulocyte binding proteins (RBPs), an RBL subfamily found in some *Plasmodium* species, are thought to specifically recognize the reticulocyte subpopulation of erythrocytes (Galinski et al., 1992). RBPs in *P. vivax* are promising targets for vaccines that block receptor–ligand interactions; however, high sequence polymorphism and copy number variations of *rbp* genes could hinder vaccine effectiveness (Rayner et al., 2005), making it important to consider diversity in these genes in *P. vivax* and *P. cynomolgi* strains. The *rbp* genes can be classified into *rbp1*, *rbp2*, and *rbp3* subgroups based on sequence similarity. In both *P. vivax* and *P. cynomolgi*, the gene pairs *rbp1a*/*rbp1b* and *rbp2c*/*rbp2d* are adjacent to each other on chromosome 7 and chromosome 5, respectively, whereas other *rbp* genes are located on different chromosomes. All three RBP groups have putatively functional representatives in the *P. cynomolgi* genome, whereas *P. vivax* and the zoonotic monkey malaria species *Plasmodium knowlesi* lack predicted functional RBP3 and RBP1 genes, respectively (Carlton et al., 2008; Pain et al., 2008; Tachibana et al., 2012). Therefore, *rbp* gene family diversification seems to have arisen before speciation, but losses have occurred since, suggesting a possible evolution of multiple species-specific erythrocyte invasion mechanisms. Variation in *rbp* gene repertoire extends beyond interspecies difference, as our genomic analysis of three *P. cynomolgi* strains revealed *rbp1b* to be present in the Berok strain but absent in the B and Cambodian strains (Tachibana et al., 2012), the first reported case of strain-specific copy number variation (CNV) for an *rbp* gene in *Plasmodium*. The *rbp2e* gene was subsequently found to be missing in the Salvador I strain of *P. vivax* but present in other strains (Hester et al., 2013). The receptors on reticulocytes are still largely unknown for the various expressed RBP proteins.

Erythrocyte Binding-Like (EBL) proteins comprise a second superfamily of erythrocyte-binding invasion ligands (Chitnis and Miller, 1994; Siddiqui et al., 2012; Sim et al., 1994) in *Plasmodium*. In *P. cynomolgi* and *P. vivax*, an EBL subfamily of Duffy-binding proteins (DBPs) are released during parasite invasion, where they bind to the Duffy Antigen Receptor for Chemokines (DARC) (Adams et al., 1990; Adams et al., 1992; Batchelor et al., 2011) on reticulocyte surfaces. Because DARC-negative individuals typically are resistant to *P. vivax*, the DBP/DARC interaction was thought for many years to be an absolute requirement for *P. vivax* malaria (Miller et al., 1976; Miller et al., 1975). However, the recent discovery of *P. vixax* malaria in DARC-negative individuals indicates that the parasite can exploit heretofore unknown pathways for erythrocyte invasion (Ntumngia et al., 2012). As with RBPs, DBPs show species- and strain-specific CNV. The reference strain of *P. vivax* (Salvador I) has one *dbp* gene while *P. cynomolgi* strains B, Berok, and Cambodian have two very similar DBP paralogs (Carlton et al., 2008; Neafsey et al., 2012; Tachibana et al., 2012). A recent study revealed tandem duplication of *dbp* in several *P. vivax* field isolates from Madagascar (Menard et al., 2013). The duplication was prevalent in the regions where the highest frequencies of *P. vivax* infected Duffy-negative people were observed, suggesting that *P. vivax* might overcome the barrier to infection presented by Duffy negativity through the duplication of the *dbp* gene. Moreover, the two copies of *dbp* in *P. vivax* are identical except for one non-synonymous difference in the signal peptide, which is even greater sequence similarity than *dbp1 vs. dbp2* of *P. cynomolgi*. The highly conserved nature of the *P. vivax* DBP duplication in these patients from Madagascar indicated that it was a recent event. Another *dbp*-like gene distantly related to DBP1/DBP2 also exists in three *P. cynomolgi* strains and in strains of *P. vivax*, except for the Salvador I strain (Chan et al., 2015).

Here we describe development and validation of a panel of 14 *P. cynomolgi* microsatellite (MS) markers using the three *P. cynomolgi* reference genome strains and eleven nominally different laboratory reference strains collected between 1933 and 1967. We subsequently used the panel to explore the diversity of 18 field samples taken from wild monkeys in Malaysian Borneo to determine their effectiveness at strain differentiation. We show that the 14 MS are capable of characterizing the genetic diversity of laboratory strains and field isolates. We also surveyed the diversity of RBPs and DBPs in all samples, comparing the non-neutral diversity in these functionally important genes to the purported neutral diversity observed in the MS. Also, we discovered CNV of *rbp* genes among the nine *P. cynomolgi* strains. Finally, we were able to identify potential strain archiving errors and address some nomenclature discrepancies that have arisen during the long-term maintenance of *P. cynomolgi* laboratory stocks.

2. Materials and methods

2.1. P. cynomolgi reference strains

DNA samples from 20 cryostabilate specimens representing eleven laboratory-maintained reference strains of *P. cynomolgi* were obtained from the Centers for Disease Control (Dr. John Barnwell), the American Type Cell Culture (ATCC), and the Malaria Reagent Reference Resource Center (MR4; Table 1). Samples were supplied as DNA, blood spots, or whole blood. DNA was extracted from blood spots or cryopreserved whole blood using a Qiagen DNeasy Blood and Tissue Kit (Qiagen®, Valencia, CA). All DNA samples were whole genome amplified (WGA) using the Qiagen REPLI-G Mini Kit (Qiagen®, Valencia, CA). WGA DNA was used to amplify microsatellite loci, while non-WGA DNA was used for multigene family amplification and sequencing.

2.2. P. cynomolgi field isolates

Sampling of wild long-tailed (*Macaca fascicularis*) and pig-tailed (*Macaca nemestrina*) macaques, two of the natural host species for *P. cynomolgi*, was conducted as previously described (Lee et al., 2011). Permission was granted and the study protocol for the capture, blood collection, and forest release of macaques was approved by the Sarawak Forestry Department (Permit Numbers: NPW.907.4.2-32, NPW.907.4.2-97, NPW.907.4.2-98, 57/2006 and 70/2007) and Sarawak Biodiversity Centre (Permit Number: SBC-RP-0081-BS). A total of 37 long-tailed macaques and seven pig-tailed macaques were sampled within a 30 km radius of Kapit town, Sarawak, Malaysian Borneo from 2004 to 2008. The macaques were anesthetized by intramuscular injection of tiletamine and zolazepam and venous blood was drawn by a veterinarian. Macaques were then tagged with a microchip to prevent re-sampling. All efforts to collect blood samples from macaques were made at the trap sites and the macaques were released immediately into the forest after the blood samples had been obtained. Whole blood samples were frozen in a liquid nitrogen dry shipper and subsequently transferred to the Malaria Research Centre, University Malaysia Sarawak (UNIMAS) for further experiments.

DNA was extracted from whole blood samples using the QIAamp DNA Blood Mini Kit (Qiagen®, Germany). Species detection of *P. cynomolgi*, *P. knowlesi*, *Plasmodium fieldi*, *Plasmodium coatneyi*, and *Plasmodium inui* was performed by nested PCR assays using species-specific primers as previously described (Lee et al., 2011). All procedures (DNA extraction and detection of species DNA by nested PCR assays) were conducted at UNIMAS.

2.3. Microsatellite genotyping

Microsatellite loci and primers were selected from a set of 175 (83 intergenic, 92 intragenic) published by Tachibana et al. (Tachibana et al., 2012), and NCBI Primer Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) used to confirm their uniqueness. Seventy MS were selected for optimization and tested at DNA concentrations of 0.5 ng/μL, 0.05 ng/μL, 0.005 ng/μL, and 0.0005 ng/μL, with ≤0.5 ng/μL chosen as the optimal DNA concentration for our non-nested PCR protocol. Primer pairs were tested for specificity on nine simian malaria parasite species: *Plasmodium simium* (MRA-353), *Plasmodium simiovale* (MRA-488), *P. fieldi* (MRA-553), *P. coatneyi* (MRA-445), *Plasmodium fragile* (MRA-352), *P. inui* Mulligan (MRA-508), *P. inui* Taiwan I (MRA-551), *P. inui shortii* (MRA-551), and *Plasmodium gonderi* (MRA-447); three region-specific *P. vivax* reference strains Indonesia XIX (MRA-378), ONG (MRA-341), Pakchong (MRA-342G); and ten nominal *P. cynomolgi* strains: Berok (PT-II), Gombak, B (*P.c. bastianellii*), M (Mulligan), Smithsonian, PT-I, Rossan (RO), Cambodian, Ceylon (*Plasmodium cynomolgi ceylonensis*), and Langur (Table 1). All 14 validated MS markers had a 20 nucleotide tag sequence (5′-CCACGACG TTGTAAAACGAC-3′) added to the 5′ end of the forward primer. A universal primer was synthesized with the same nucleotide tag sequence and 5′-labeled with the phosphoramidite conjugate 6-FAM (Sigma-Aldrich Co. LLC, St. Louis, MO). This fluorescently-labeled tag sequence anneals to each forward primer during amplification, reducing the cost of synthesizing a probe specific for each primer-pair combination.

PCR amplifications (25 μL volume total) contained 2.5 μM forward primer, 5 μM reverse primer, 2.5 μM of the primer 6-FAM-5′-CCACGA CGTTGTAAAACGAC-3′, 2× Green GoTaq® Master Mix Reaction Buffer (pH 8.5) (Promega Corporation, Madison, WI), molecular grade water, and DNA (adjusted to 10–20 ng/μL) or whole-genome amplified DNA adjusted similarly. PCR cycling in an Eppendorf Mastercycler (Westbury, NY) was as follows: denaturation at 95 °C for 5 min; followed by 40 cycles of denaturation at 95 °C for 20 s, annealing at 58 °C for 45 s, extension at 60 °C for 45 s; and a final extension at 72 °C for 5 min. PCR products were electrophoresed on 1.5% Agarose HS (Denville Scientific Inc., Metuchen, NJ) and visualized under ultraviolet.

Primer optimization and MS genotyping of laboratory strains was performed at New York University (NYU) (New York, NY, USA), while the MS genotyping of the field isolates was performed at the London School of Tropical Medicine and Hygiene (London, UK). To ensure synchronization between sites, identical positive controls served as a means to standardized allele scores by grouping each allele into locus- and size-specific bins. Any inconsistency in allele size between the identical positive controls resulted in a score adjustment which was applied to each of the field isolates tested. To minimize data analysis error, data generated at each site was analyzed by a single individual.

Amplicon sizes were determined using an ABI 3730 genetic analyzer (Applied Biosystems, Foster, CA).The allele length of each MS marker

Table 1
Plasmodium cynomolgi laboratory strains. A unique identifier was assigned to each sample for this study. Alternating gray and white row shading indicates stocks nominally from the same strain. The sample name, sample code, and archive date (date on which the sample was isolated and frozen) were data supplied by the source. Original monkey host information was supplied by source or mined from literature, as were the historical aliases (see Supplementary materials for references). The revised names for extant US stocks reflect strain identities indicated by our data in the case of M and B strains, and otherwise our preference where multiple aliases exist. Abbreviations: CDC: Centers for Disease Control; MR4: Malaria Research and Reference Reagent Resource Centre; ATCC: American Type Culture Collection. * Isolated from mosquito vector. Note: ATCC and MR4 strain isolates originated from archived cryopreserved strains maintained by CDC.

No.	Unique sample identifier for this study	Sample source	Name on sample	Sample code or monkey number	Sample archive date	Original monkey host	Aliases in literature and repositories	Revised name
1	Bastianellii_T644	CDC	CynoB	Monkey #T644	21-Mar-69	*Macaca fascicularis*	*P. c. bastianellii*; B strain; NIH; London	M/B
2	Bastianellii_9903	CDC	B-strain	Monkey #9903	12-Apr-01			
3	Bastianellii_MRA-350G	MR4	Bastianellii	MRA-350G (DNA)	Unknown			
4	Berok_446	CDC	Berok	Monkey # 84-46	9-Jan-01	*Macaca nemestrina*	Berok; PT-II	Berok
5	PT-II_Monkey-8-42	CDC	PT-II (Berok)	Monkey # 8-42	1-Dec-83			
6	Cambodian_9903	CDC	Cambodian	Monkey #9903	31-Jul-01	*Macaca fascicularis*	Cambodian	Cambodian
7	Ceylonensis_ATCC-30144	ATCC	Ceylonensis	ATCC 30144, lot 67687	17-Mar-71	*Macaca sinica*	*P.c. ceylonensis*; C strain	Ceylon
8	Ceylonensis_Monkey-8-11	CDC	Ceylonensis	Monkey 8-11	30-Nov-82			
9	Gombak_MRA-550	MR4	Gombak	MRA-550	Unknown	None*	Gombak	Gombak
10	Gombak_RH0002	CDC	Gombak	Monkey #RH0002	12-Nov-02			
11	Langur_RH0001	CDC	Langur	Monkey #RH0001	12-Nov-02	*Presbytis entellus thersites*	Langur	Langur
12	Mulligan_ATCC-30155	ATCC	M strain	ATCC 30155, lot 67713	5-May-71	*Macaca fascicularis*	Mulligan; Rockefeller; TC; M strain; *P. c. cynomolgi*	M/B
13	Mulligan_Monkey-18IX97	CDC	Mulligan	Monkey #89-04	18-Sep-97			
14	Mulligan_30037	ATCC	M-strain	ATCC 30037, lot 67427	Mar-70			
15	Mulligan_T426	CDC	CynoM	Monkey #T426	4-Aug-67			
16	NIH-U-53_Monkey-U-53	CDC	NIH	Monkey #U-53	13-May-87	Unknown	NIH	M/B
17	PT-I_RH0001	CDC	PT-I	Monkey #RH0001	27-Apr-04	*Macaca nemestrina*	PT-I	PT-I
18	RO_ATCC-30146	ATCC	RO strain	ATCC 30146, lot 67691	12-Mar-71	*Macaca mulatta*	RO; Rossan	Rossan
19	RO_RNQ9	CDC	Rossan	Monkey #RNQ9	17-Apr-12			
20	Smithsonian_MRA-351G	MR4	Smithsonian	MRA-351G (DNA)	Unknown	*Macaca speciosa*	Smithsonian	Smithsonian

was determined using internal size standards (GeneScan 500 LIZ Size Standard, Applied Biosystems) with GeneMapper v4.0 software (ABI, Foster City, CA). Only alleles detected with a peak height ≥200 fluorescent units were considered. Mixed genotype infections were defined as two or more MS loci containing multiple alleles, and minor peaks were defined as peaks equal to one-third the height of the major peak. Negative controls included reactions containing water instead of DNA to monitor for contamination, and the use of *P. vivax* and *Plasmodium falciparum* DNA to rule out spurious bands that may be generated by non-specific hybridization of human DNA.

2.4. Population genetic and phylogenetic analyses

H_E (expected heterozygosity) was calculated using the formula $H_E = [n/(n-1)][1 - \Sigma^n{}_{i=1}\, p^2]$ where p is the frequency of the ith allele and n is the number of alleles sampled, as implemented in Arlequin 3.11 (Excoffier et al., 2005). Distance matrices were developed using Nei's DA calculation for genetic distance *via* POPTREE2 (Takezaki et al., 2010).

2.5. P. cynomolgi RBP and DBP amplification and sequencing

Primers for *P. cynomolgi rbp1a*, *rbp2*, *rbp3*, *dbp1*, and *dbp2* genes were designed manually by inspecting conserved regions of each sequence from *P. cynomolgi* strains B, Berok, and Cambodian. All sequences were obtained from GenBank (PCYB_033090, PCYB_063270, PCYB_071060, PCYB_147650, JQ422036, JQ422035, JQ422037, JQ422043, AB617789, AB617788, JQ422044, JQ422050) and aligned in Sequencher (version 5.3, Gene Codes Corporation, Ann Arbor, MI USA) or Geneious Software v6.1.6 (Kearse et al., 2012). Primers and PCR conditions are shown in Supplementary Table 3 and Supplementary Table 4, and schematics showing the combinations and locations of these primers are provided in Supplementary Fig. 1 and Supplementary Fig. 2. GenBank accession numbers correlated to individual isolates and genes can be found in Supplementary Table 5.

PCR amplifications were performed in a total volume of 20 μL in the presence of 10× LA PCR Buffer II (Mg^{2+} free), 2.5 mM $MgCl_2$, 400 μM dNTPs, 0.2 μM primers and 2.5 U/50 μL of TaKaRa LA Taq Polymerase (TaKaRa Bio Inc., Mountain View, CA). The PCR products were analyzed on a 1% agarose gel and purified (QIAquick Gel Extraction Kit, Valencia, CA). Purified products were sequenced using BigDye v3.1 Terminator (Applied Biosystems, Grand Island, NY) and then analyzed on an ABI machine (Applied Biosystems, Grand Island, NY). Sequences were assembled using Sequencher 5 (Gene Codes, Ann Arbor, MI).

2.6. RBP and DBP phylogenetic analyses

Sequence alignments were constructed using Sequencher and then adjusted manually. Regions that could not be unambiguously aligned were omitted from subsequent phylogenetic analyses. Trees were constructed from nucleotide sequences of *rbp1a* (1965 bp), *rbp3* (4200 bp), *dbp1* (1866 bp), *dbp2* (1026 bp), and concatenated nucleotide sequence of *rbp1a*, *rbp3*, *dbp1* and *dbp2* (9057 bp) using Neighbor Joining in the Geneious software package with bootstrap support of 1000 replicates using the HKY substitution model.

3. Results

3.1. Validation of a panel of P. cynomolgi microsatellite markers in laboratory strains

We developed a panel of 14 microsatellite markers from a set of 175 originally mined from the *P. cynomolgi* strain B and strain Berok genomes (Tachibana et al., 2012). After excluding loci on the basis of primer sequence similarity to *P. vivax* and missing sequence data using NCBI Primer BLAST, a total of 70 were selected for optimization that: (1) conformed to motif size >1–≤4 bp, (2) were expected to have a fragment length between 100–400 bp, (3) included both purines and pyrimidines, and (4) were distributed across the 14 *P. cynomolgi* chromosomes. None of the 70 MS markers were amplified from three *P. vivax* DNA strains tested. However, amplification of bands from simian malaria species *P. simium*, *P. simiovale*, *P. fieldi*, *P. coatneyi*, *P. fragile*, *P. inui*, and *P. gonderi* resulted in the exclusion of 36 MS loci. Of the remaining 34 MS markers, 20 were discarded due to non-specific amplification of nine *P. cynomolgi* reference strains, leaving 14 MS markers. The characteristics of the 14 MS markers including the primers used to amplify them are shown in Supplementary Table 1.

We tested the final panel of 14 MS markers on a total of 20 specimens nominally representing 11 *P. cynomolgi* reference strains. These strains have been used for as long as 80 years in research laboratories worldwide. Fig. 1 illustrates their geographical origins and years of isolation.

Six strains were sampled more than once, including three strain B samples, four M samples, two Berok samples, two Gombak samples, two Ceylon samples, and two Rossan samples (Table 1). Some of these strains are archived at the ATCC and/or the MR4, while others were contributions from the Primate Malaria Archive at the Centers for Disease Control and Prevention (CDC), where all *P. cynomolgi* strains and other primate malaria species are maintained.

Our genotyping showed the specimens nominally representing B, M, and NIH strains to be genetically identical (Table 2), and hereafter we

Table 2
Allele scoring for *P. cynomolgi* laboratory strains. The 'allele score' is the size, in nucleotides, of the amplicon for each of the 14 MS markers (columns) from each sample (rows).

	1.307	2.36	3.574	4.41	4.462	5.956	6.455	7.1006	8.1086	10.179	10.621	11.1096	12.286	14.429
Bastianellii_9903	315	215	274	287	264	212	220	292	208	135	289	167	206	341
Bastianellii_MRA-350G	315	215	274	287	264	212	220	292	208	135	289	167	206	341
NIH-U-53_Monkey-U-53	315	215	274	287	264	212	220	292	208	135	289	167	206	341
Mulligan_Monkey-18IX97	315	215	274	287	264	212	220	292	208	135	289	167	206	341
Mulligan_ATCC-30155	315	215	274	287	264	212	220	292	208	135	289	167	206	341
Mulligan_30037	315	215	274	287	264	212	220	292	208	135	289	167	206	341
Berok_446	300	192	218	290	252	164	214	308	193	121	241	140	155	331
PT-II_Monkey-8-42	300	192	218	290	252	164	214	308	193	121	241	140	155	331
Cambodian_9903	327	190	282	293	255	218	223	244	208	125	281	155	209	333
Ceylonensis_Monkey-8-11	327	225	302	287	258	233	223	292	214	125	250	150	206	337
Ceylonensis_ATCC-30144	327	225	302	287	258	233	223	292	214	125	250	150	206	337
Langur_RH0001	312	225	242	287	249	218	223	344	208	125	250	155	206	337
Smithsonian_MRA-351G	294	190	266	290	264	212	211	284	214	133	285	159	206	339
PT-I_RH0001	294	190	302	296	264	209	211	296	223	133	289	163	206	335
Gombak_RH0002	318	190	210	281	270	170	232	236	247	119	253	140	158	321
RO_ATCC-30146	312	232	302	293	267	218	217	292	214	121	285	155	206	333

designate these stocks collectively as the 'M/B strain'. Our subsequent data analysis thus encompassed nine unique laboratory strains. The average number of alleles per locus across the 14 MS loci for the nine laboratory strains was 5.85 ± 0.95, with a range of from four to seven alleles per locus (Table 3). The H_E was 0.87 ± 0.09, with a minimum of 0.58 and a maximum of 0.94. Our panel of 14 *P. cynomologi* MS was thus highly polymorphic and potentially useful as markers for genetic diversity studies.

MS allele frequency data was used to calculate a distance matrix for the nine genetically distinct laboratory strains (Supplementary Table 2). With the caveat that the sample size is small, Berok and Gombak appear more distant from other Peninsular Malaysian isolates (M/B, PT-I, and Smithsonian), and the four non-Malaysian isolates (Cambodian, Ceylon, Langur, and Rossan) appear more closely related to one another than to the isolates from Peninsular Malaysia.

3.2. Allelic diversity of microsatellite markers in P. cynomolgi field isolates

We next used the 14 MS loci to genotype extant *P. cynomolgi* isolates collected from long-tailed and pig-tailed macaques in the Malaysian state of Sarawak on the island of Borneo during 2004–2008. A total of 27 isolates were collected from eight sites within a 30 km radius of Kapit Town, Sarawak (Fig. 1). MS loci 2.36, 4.462, 11.1096, and 12.286 consistently failed to amplify the field isolates and were excluded from subsequent analyses (Table 3). Largely due to host DNA contamination, low parasitemia and sample volume are characteristic of clinically-isolated *Plasmodium* species infections. In addition to high host contamination, numerically high mixed-species and mixed-genotype infection rates are expected to further lower the DNA yield for individual species and clones within the infection. The amplification failures observed in this study were likely due to the low DNA quality and quantity attributable to these factors. However, pathological parasite suppression by competing species and/or clones could also play a role.

Six (33.3%) of the 18 field isolates were identified as mixed genotype infections. These isolates were not excluded in subsequent analyses because the major clone could be clearly identified. No shared haplotypes were identified in the 18 field isolates and the average number of alleles per MS locus was 7.20 ± 2.49, with a range of four to ten alleles per locus. The average H_E across the 10 MS loci was 0.80 ± 0.15, with a minimum of 0.47 and a maximum of 0.94, very similar to that observed in the laboratory strains (average H_E 0.87 ± 0.09, with a minimum of 0.58 and a maximum of 0.94). Haplotype information for the field isolates are shown in Table 4.

3.3. Diversity of RBP and DBP genes in P. cynomolgi laboratory strains

We next surveyed the diversity of *P. cynomolgi* reticulocyte binding proteins (RBPs) and Duffy binding proteins (DBPs) in *P. cynomolgi* laboratory strains. The status of *rbp1*, *rbp2*, *rbp3*, *dbp1*, and *dbp2* genes was determined by designing primers (Supplementary Table 3, Supplementary Fig. 1) for the flanking regions of each gene, such that its presence or absence could be determined from the size of the PCR product. Similar to our previous results that found *rbp1b* in the Berok strain but not in the M/B and Cambodian strains (Tachibana et al., 2012), we found *rbp1b* in the Gombak strain but not in the Rossan, Smithsonian, Ceylon, or Langur strains (Table 5). No consistent amplification could be obtained from the PT-I strain. Furthermore, *P. cynomolgi* strains that lack the *rbp2a* gene (Berok and Gombak) contained an *rbp1b* gene, and while those that lacked *rbp1b* contained *rbp2a*. No copy number variation was observed for any of the other *rbp* genes, or for *dbp1* and *dbp2* genes.

We sampled the genetic diversity of RBP and DBP genes by partially sequencing amplification products of *rbp1a*, *rbp3*, *dbp1*, and *dbp2* genes in the *P. cynomolgi* laboratory strains. For *rbp1a* and *rbp3*, five and six different combinations of primers, respectively, were used to amplify the whole gene in fragments. For *dbp1* and *dbp2*, we designed primers targeting the N-terminal cysteine-rich region, where the binding domain is located. We generated a phylogenetic tree using concatenated nucleotide sequences of these genes from eight *P. cynomolgi* samples and five *P. vivax* reference strains (Fig. 2). Two main clades of *P. cynomolgi* were apparent, one containing the Berok and Gombak strains, and a second containing six strains in two subclades: the two M/B sequences plus Smithsonian and Ceylon in one branch, and the Cambodian and Rossan strains in a second branch. Sequencing and trees furthermore confirmed 100% identity of B strain and M strain samples. We also generated dendrograms for each gene (Supplementary Fig. 3). Berok and Gombak strains consistently branch together in these trees as well.

4. Discussion

The publication of multiple reference genomes for *P. cynomolgi* and *P. vivax* presented an opportunity to develop molecular tools for understanding the global diversity within each species. We used data from three *P. cynomolgi* reference genomes (Tachibana et al., 2012) to develop genetic tools to assess genetic variation of this species. First, we developed a panel of 14 *P. cynomolgi*-specific microsatellite (MS) markers. Our results indicate that the 14 MS loci optimized in this study are robust genetic markers for measuring diversity. Distributed across 12 of 14 *P. cynomolgi* chromosomes, the MS panel was shown

Table 3
Characteristics of 14 MS loci amplified from laboratory strains and field isolates. The number of alleles, allele size range, and expected heterozygosity for the 14 MS loci is shown for nine lab strains, 18 field isolates, and then all strains and isolates combined. MS loci 2.36, 4.462, 11.1096, and 12.286 consistently failed to amplify the field isolates and were excluded from subsequent analyses. ND: not determined.

ID	Chr. no.	Motif	No. of alleles			Allele size range			Expected heterozygosity		
			Lab strains	Field isolates	All strains	Lab strains	Field isolates	All strains	Lab strains	Field isolates	All strains
1.307	1	TCT	6	6	11	294–327	285–351	285–351	0.92	0.76	0.84
2.36	2	CA	5	ND	ND	190–232	ND	ND	0.81	ND	ND
3.574	3	GTGA	7	10	14	210–302	206–282	206–302	0.92	0.93	0.95
4.41	4	TCT	5	10	13	281–296	273–315	273–315	0.86	0.94	0.93
4.462	4	TCC	7	ND	ND	249–270	ND	ND	0.92	ND	ND
5.956	5	AAG	6	4	10	164–233	167–206	164–233	0.89	0.65	0.87
6.455	6	AGG	6	4	7	211–232	211–229	211–232	0.89	0.47	0.74
7.1006	7	ATAC	7	6	13	236–344	153–260	153–344	0.92	0.82	0.90
8.1086	8	GAG	5	9	11	193–247	184–220	184–247	0.83	0.91	0.90
10.179	10	TA	5	5	8	119–135	117–127	117–135	0.86	0.71	0.85
10.621	10	TGTA	6	10	14	241–289	233–305	233–305	0.92	0.92	0.95
11.1096	11	TGTA	6	ND	ND	140–167	ND	ND	0.89	ND	ND
12.286	12	AAT	4	ND	ND	155–209	ND	ND	0.58	ND	ND
14.429	14	AT	7	8	15	321–341	311–329	311–341	0.94	0.90	0.95

Table 4
Allele scoring for *P. cynomolgi* field isolates. Alleles scored in each of the field isolates genotyped at 10 microsatellite loci. Italicized *isolate names* represent field-collected isolates. Highlighted (light gray) allele scores represent the predominant allele within a mixed genotype infections (n = 6 mixed-genotype infections), while single unmixed alleles are not highlighted. Missing genotype data is reflected by gaps.

Strain/isolate	1.307	3.574	4.41	5.956	6.455	7.1006	8.1086	10.179	10.621	14.429
LT040		258	315	173	229	240	211	123	257	
LT042	315	246		203	229		205	117	265	323
LT043	285	282		167	229		214	127	265	325
LT058	285	282	293		217	248	205		265	311
LT072	294	246		167	229	244	202	123	281	
LT051	294	262	300		229	240	184	117	241	325
LT073	288	234	306	206	229	252	214		233	325
LT082	294	258	290	203	217	240	202	123	265	
PTK012	351	206	275		229	256	193	117	237	311
LT020	285	266	278	203	211	244	214	123	257	315
LT049	297	254	300	203	217	240	208	123	257	319
LT055	294	254	303	203	220	244	202	123	281	319
LT065	297	246	296	203	229	248	199	119	277	311
LT070	294	242	300	167	229	260	202	125	273	329
LT084	294	246	309	203	229	244	211	123	305	315
LT089	294	270	290	206	229	244	220	123	233	327
LT091	294	266	296	203	229	240	211	125	297	317
PTK014	306	206	273	170	229	153	193	121	237	311

to be sensitive, specific, and capable of differentiating laboratory strains isolated from various sources. An earlier publication developed a panel of 13 *P. vivax* microsatellite loci and tested them in 108 *P. vivax* samples from eight localities in Asia, Africa, South America, and New Guinea, as well as in nine other *Plasmodium* species (Leclerc et al., 2004). These authors demonstrated significant overlap of microsatellite loci between *P. vivax* and several other species, including *P. cynomolgi*, concluding that the flanking regions of their microsatellite loci appear to be genetically similar, and thus evolutionarily conserved. To avoid such overlap, species-specificity of each marker, which can be a concern when deploying this methodology on field isolates that can contain more than one *Plasmodium* species, was tested and confirmed using nine different simian malaria species and three *P. vivax* laboratory strains. To ensure the utility of these MS loci as a genotyping tool for field collected isolates, 18 non-human primate samples (collected from Kapit Town, Sarawak, Malaysian Borneo from 2004 to 2008) were tested. High genetic diversity among *P. cynomolgi* isolates combined with the high mixed species prevalence (specifically with *P. knowlesi*) reflects the high parasite transmission within the macaque population. However, limitations in the study design prevented the consideration of high genetic diversity attributable to chronic infections and hypnozoite relapse. Whether or not *P. cynomolgi* in the macaque population is associated with different macaque hosts, as described in *P. knowlesi* populations (Divis et al., 2015), needs to be determined by more sampling.

Observation of strain-specific CNV of *rbp* and *dbp* genes, noted first during genome sequencing of *P. cynomolgi* (Tachibana et al., 2012) is described more comprehensively here, We note the most variation in the *rbp1b* and *rbp2a*. We observed that all *P. cynomolgi* strains which contain one of this pair lack the other, *e.g.*, the Berok and Gombak strains lack *rbp2a* but contain *rbp1b*. During the invasion of reticulocytes by *P. cynomolgi*, perhaps the presence of *rbp1b* could compensate for the absence of *rbp2a* and *vice versa*. The parasite might first lose *rbp1b* and then gain *rbp2a* or the other way around. Such changes could underlie important phenotypic differences. Polymorphism of a gene encoding an erythrocyte binding protein in rodent malaria *Plasmodium yoelii yoelii* was associated with disease virulence (Otsuki et al., 2009; Pattaradilokrat et al., 2009). Particular alleles of two erythrocyte binding protein genes in *P. knowlesi* were found to be associated with specific disease progression profiles, such as differences in parasitemia, and markers of disease severity such as hemoglobin levels, platelet levels, renal dysfunction, *etc.* in a population of knowlesi malaria patients (Ahmed et al., 2014). These two genes are also known to show differential binding between macaque and human erythrocytes (Semenya et al., 2012). Strain-specific CNV has also been found in the antigen gene merozoite surface protein 3 (MSP3) of *P. cynomolgi* strains B and Berok (Rice et al., 2014).

All *P. cynomolgi* strains that we tested have two very similar *dbp* genes (~92% DNA identity) and one other closely related gene. A duplication of *dbp* in *P. vivax*, which typically contains only one *dbp* gene, was prevalent in Madagascar where the highest frequencies of *P. vivax*-infected, Duffy-negative people were reported, raising the possibility that *P. vivax* may be able to invade Duffy-negative erythrocytes through

Table 5
Presence of *rbp* and *dbp* genes in nine *P. cynomolgi* laboratory reference strains. Strain names are the revised names from Table 1. Results for *rbp1b* in bold were previously published by us. *: represents multiple samples analyzed; +: gene present; −: gene absent; ND: not determined due to insufficient DNA; NP: no product;

	Berok*	Gombak*	M/B*	Smithsonian	Rossan*	Ceylon*	Langur	Cambodian	PT-I
rbp1a	+	+	+	+	+	+	+	+	ND
rbp1b	+	+	−	−	−	−	−	−	ND
rbp2a	−	−	+	+	+	+	+	+	+
rbp2b	+	+	+	+	+	+	+	+	ND
rbp2c	+	+	+	+	+	+	+	+	ND
rbp2d	+	+	+	+	+	+	+	+	+
rbp2e	+	+	+	+	+	NP	+	+	ND
rbp2f	+	+	+	+	NP	NP	+	+	+
rbp3	+	+	+	+	+	+	+	+	ND
dbp1	+	+	+	+	+	+	+	+	+
dbp2	+	+	+	+	+	+	+	+	+

the duplication of *dbp* (Menard et al., 2013). Similarly, *P. cynomolgi* might enable the parasite to infect Duffy-negative red blood cells of macaques, although the apparent rarity of Duffy-negative macaques in the wild (Palatnik and Rowe, 1984) argues against this.

We observed two subclades of *P. cynomolgi*, one of which contains Berok and Gombak, in phylogenetic trees of *rbp* and *dbp* genes. This pattern in *P. cynomolgi* has been reported for several antigen genes, including circumsporozoite protein gene (CSP) (Pacheco et al., 2012b), merozoite surface protein 1 (MSP1) (Tanabe et al., 2007), MSP3 (Rice et al., 2014), MSP8, and MSP10 (Pacheco et al., 2012a), suggesting that the fact that the genetic differences shown by Berok and Gombak might be due to the geographical location or host variation. Further sampling of this widespread parasite of macaques is clearly needed to investigate in detail whether there is geographical population genetic structure in *P. cynomolgi*, and whether particular polymorphic genes are under selection. We suggest that the loci that we present here will be useful for this purpose, and will allow interesting comparisons between patterns seen in this and other parasite species.

Finally, during the MS optimization process, we observed that stocks of some commonly available strains of *P. cynomolgi* are genetically identical. As early as the mid-1980s, published immunological, electrophoretic, and genetic evidence (Cochrane et al., 1985) suggested that stocks maintained at NIH of *P. c. bastianellii* (also known as the B or NIH strain) from stock originally sent to the USA from London by Professor P.C.C. Garnham in 1959 (Bennett et al., 1966), and the Mulligan strain (also known as the M, Rockefeller, or TC strain, and at one point given the subspecific name *P. c. cynomolgi*) (see Supplemental materials), originally from stock imported from India to the USA during World War II) (Eyles et al., 1963; Schmidt et al., 1949) were similar if not identical. This confusion of the B and M strains had been noted briefly by one of us (JWB) before (Tachibana et al., 2012), and probably resulted from an error committed over the course of the complex monkey passage and stock exchange history of these two strains at NIH and between laboratories since 1959. The oldest samples of B and M strains archived at the CDC, which originated from the old NIH/NIAID Laboratory of Parasite Chemotherapy, date from 1969 and 1967, respectively. As these two specimens are genetically identical in our study, we conclude that this archiving/strain mix-up must have occurred no later than 1967. We here refer to these M and B samples as the 'M/B strain' to acknowledge the ambiguity. However, review of the M and B strain literature, and of US specimen archives, and public database sequences annotated as M or B strain (and aliases thereof), and comparison with our sequences, suggests that all B and M stocks currently available from US sources, as well as public database 'B strain' sequences derived from US stock (including the reference *P. cynomolgi* strain B genome), are all probably M strain. Public database analyses, however suggest that uncontaminated B strain stock may still exist in India (SAS, unpublished). Definitive sorting of M and B stocks and sequences will require sequencing of such uncontaminated B strain stock, if it still exists.

5. Conclusions

The 14 *P. cynomologi* MS described here were validated against nine unique laboratory strains as being highly polymorphic, and thus useful as markers for genetic diversity studies of the species, as indicated by successful genotyping of field isolates. Diversity in this species is also manifested structurally by strain-specific presence/absence of some *rbp* genes, as well as by the sequences of *rbp* and *dbp* genes. Our study further reveals that extant US stocks of strains M and B, the first two

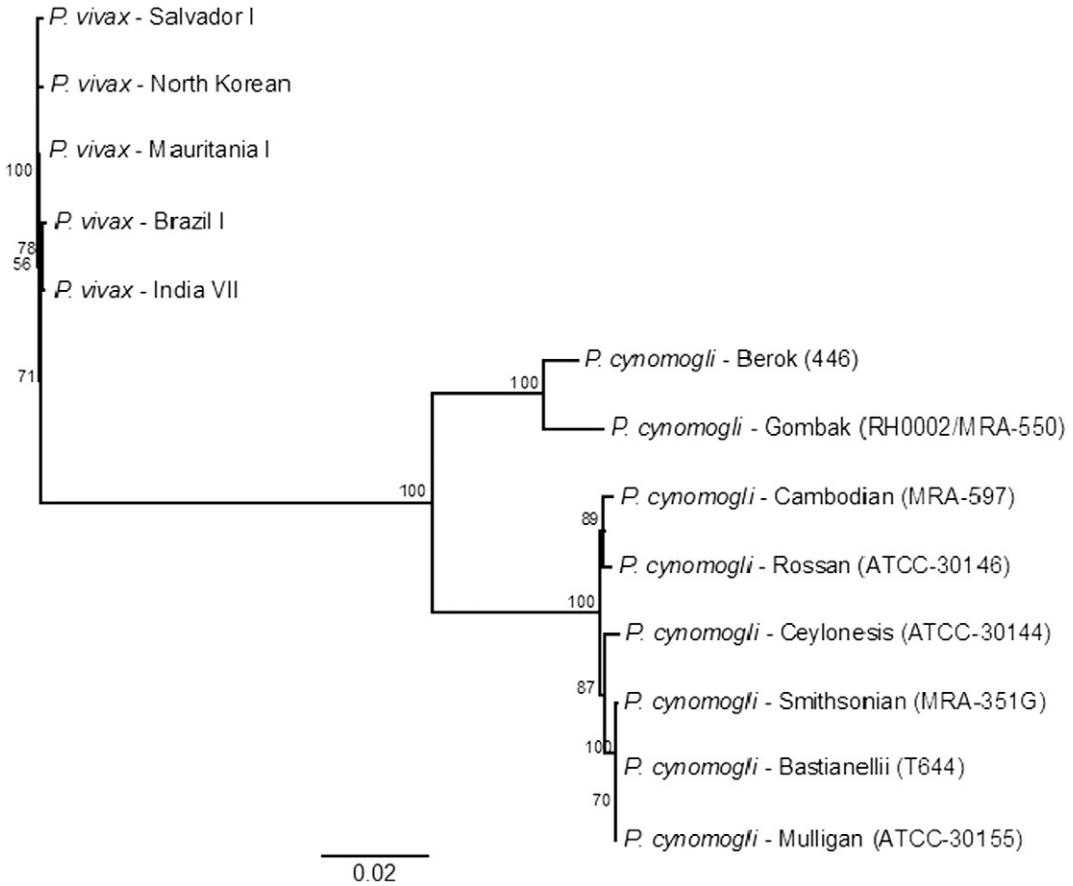

Fig. 2. Dendrogram showing inter-isolate differentiation based on concatenated *P. cynomolgi* nucleotide sequences from four genes. Dendrogram based on distance matrices from 8031 bp of concatenated nucleotide sequence of *rbp1a*, *rbp3*, and *dbp1* from eight *P. cynomolgi* strains and five *P. vivax* reference strains built by Neighbor Joining with 1000 bootstraps using the HKY substitution model. The numbers shown along nodes represent bootstrap values. Identical topology was obtained using Maximum Likelihood (not shown).

P. cynomolgi strains isolated, are genetically identical, likely reflecting a long-standing archiving error.

Acknowledgments

ZL is supported by the MacCracken Program in the Graduate School of Arts and Science at New York University (New York, NY). This work was supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH) International Centers of Excellence in Malaria Research Grant U19AI089676 to JMC. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.meegid.2016.03.009.

References

Adams, J.H., Hudson, D.E., Torii, M., Ward, G.E., Wellems, T.E., Aikawa, M., Miller, L.H., 1990. The Duffy receptor family of *Plasmodium knowelsi* is located within the micronemes of invasive malaria merozoites. Cell 63, 141–153.

Adams, J.H., Sim, B.K., Dolan, S.A., Fang, X., Kaslow, D.C., Miller, L.H., 1992. A family of erythrocyte binding proteins of malaria parasites. Proc. Natl. Acad. Sci. U. S. A. 89, 7085–7089.

Ahmed, A.M., Pinheiro, M.M., Divis, P.C., Siner, A., Zainudin, R., Wong, I.T., Lu, C.W., Singh-Khaira, S.K., Millar, S.B., Lynch, S., Willmann, M., Singh, B., Krishna, S., Cox-Singh, J., 2014. Disease progression in *Plasmodium knowlesi* malaria is linked to variation in invasion gene family members. PLoS Negl. Trop. Dis. 8, e3086.

Aikawa, M., Miller, L.H., Rabbege, J., 1975. Caveola–vesicle complexes in the plasmalemma of erythrocytes infected by *Plasmodium vivax* and *P cynomolgi*. Unique structures related to Schuffner's dots. Am. J. Pathol. 79, 285–300.

Anderson, T.J., Su, X.Z., Bockarie, M., Lagog, M., Day, K.P., 1999. Twelve microsatellite markers for characterization of *Plasmodium falciparum* from finger-prick blood samples. Parasitology 119, 113–125.

Anderson, T.J., Haubold, B., Williams, J.T., Estrada-Franco, J.G., Richardson, L., Mollinedo, R., Bockarie, M., Mokili, J., Mharakurwa, S., French, N., Whitworth, J., Velez, I.D., Brockman, A.H., Nosten, F., Ferreira, M.U., Day, K.P., 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol. Biol. Evol. 17, 1467–1482.

Arnott, A., Barry, A.E., Reeder, J.C., 2012. Understanding the population genetics of *Plasmodium vivax* is essential for malaria control and elimination. Malar. J. 11, 14.

Batchelor, J.D., Zahm, J.A., Tolia, N.H., 2011. Dimerization of *Plasmodium vivax* DBP is induced upon receptor binding and drives recognition of DARC. Nat. Struct. Mol. Biol. 18, 908–914.

Bennett, G.F., Warren, M., Cheong, W.H., 1966. Biology of the simian malarias of Southeast Asia. II. The susceptibility of some Malaysian mosquitoes to infection with five strains of *Plasmodium cynomolgi*. J. Parasitol. 52, 625.

Branch, O.H., Sutton, P.L., Barnes, C., Castro, J.C., Hussin, J., Awadalla, P., Hijar, G., 2011. *Plasmodium falciparum* genetic diversity maintained and amplified over 5 years of a low transmission endemic in the Peruvian Amazon. Mol. Biol. Evol. 28, 1973–1986.

Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S.V., Merino, E.F., Amedeo, P., Cheng, Q., Coulson, R.M., Crabb, B.S., Del Portillo, H.A., Essien, K., Feldblyum, T.V., Fernandez-Becerra, C., Gilson, P.R., Gueye, A.H., Guo, X., Kang'a, S., Kooij, T.W., Korsinczky, M., Meyer, E.V., Nene, V., Paulsen, I., White, O., Ralph, S.A., Ren, Q., Sargeant, T.J., Salzberg, S.L., Stoeckert, C.J., Sullivan, S.A., Yamamoto, M.M., Hoffman, S.L., Wortman, J.R., Gardner, M.J., Galinski, M.R., Barnwell, J.W., Fraser-Liggett, C.M., 2008. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature 455, 757–763.

Carlton, J.M., Das, A., Escalante, A.A., 2013. Genomics, population genetics and evolutionary history of *Plasmodium vivax*. Adv. Parasitol. 81, 203–222.

Chan, E.R., Barnwell, J.W., Zimmerman, P.A., Serre, D., 2015. Comparative analysis of field-isolate and monkey-adapted *Plasmodium vivax* genomes. PLoS Negl. Trop. Dis. 9, e0003566.

Chitnis, C.E., Miller, L.H., 1994. Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in erythrocyte invasion. J. Exp. Med. 180, 497–506.

Coatney, G.R., Elder, H.A., Contacos, P.G., Getz, M.E., Greenland, R., Rossan, R.N., Schmidt, L.H., 1961. Transmission of the M strain of *Plasmodium cynomolgi* to man. Am.J.Trop. Med. Hyg. 10, 673–678.

Coatney, G.R., Collins, W.E., Warren, M., Contacos, P.G., 1971. The Primate Malarias. U. S. Government Printing Office, Washington, D.C.

Cochrane, A.H., Gwadz, R.W., Ojo-Amaize, E., Hii, J., Nussenzweig, V., Nussenzweig, R.S., 1985. Antigenic diversity of the circumsporozoite proteins in the *Plasmodium cynomolgi* complex. Mol. Biochem. Parasitol. 14, 111–124.

Contacos, P.G., Elder, H.A., Coatney, G.R., Genther, C., 1962. Man to man transfer of two strains of *Plasmodium cynomolgi* by mosquito bite. Am.J.Trop. Med. Hyg. 11, 186–193.

Dissanaike, A.S., Nelson, P., Garnham, P.C.C., 1965. Two new malaria parasites, *Plasmodium cynomolgi ceylonensis* subsp. nov. and *Plasmodium fragile* sp. nov., from monkeys in Ceylon. Ceylon J. Med. Sci. 14, 1–9.

Divis, P.C., Singh, B., Anderios, F., Hisam, S., Matusop, A., Kocken, C.H., Assefa, S.A., Duffy, C.W., Conway, D.J., 2015. Admixture in humans of two divergent *Plasmodium knowlesi* populations associated with different macaque host species. PLoS Pathog. 11, e1004888.

Enea, V., Galinski, M., Schmidt, E., Gwadz, R., Nussenzweig, R.S., 1986. Evolutionary profile of the circumsporozoite gene of the *Plasmodium cynomolgi* complex. J. Mol. Biol. 188, 721–726.

Escalante, A.A., Barrio, E., Ayala, F.J., 1995. Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. Mol. Biol. Evol. 12, 616–626.

Escalante, A.A., Freeland, D.E., Collins, W.E., Lal, A.A., 1998. The evolution of primate malaria parasites based on the gene encoding cytochrome b from the linear mitochondrial genome. Proc. Natl. Acad. Sci. U. S. A. 95, 8124–8129.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinformatics Online 1, 47–50.

Eyles, D.E., 1960. The exoerythrocytic cycle of *Plasmodium cynomolgi* and *P. cynomolgi bastianellii* in the rhesus monkey. Am.J.Trop. Med. Hyg. 9, 543–555.

Eyles, D.E., Coatney, G.R., Getz, M.E., 1960. Vivax-type malaria parasite of macaques transmissible to man. Science 131, 1812–1813.

Eyles, D.E., Dunn, F.L., Warren, M., Guinn, E., 1963. *Plasmodium coatneyi* from the Philippines. J. Parasitol. 49, 1038.

Fang, X.D., Kaslow, D.C., Adams, J.H., Miller, L.H., 1991. Cloning of the *Plasmodium vivax* Duffy receptor. Mol. Biochem. Parasitol. 44, 125–132.

Galinski, M.R., Arnot, D.E., Cochrane, A.H., Barnwell, J.W., Nussenzweig, R.S., Enea, V., 1987. The circumsporozoite gene of the *Plasmodium cynomolgi* complex. Cell 48, 311–319.

Galinski, M.R., Medina, C.C., Ingravallo, P., Barnwell, J.W., 1992. A reticulocyte-binding protein complex of *Plasmodium vivax* merozoites. Cell 69, 1213–1226.

Garnham, P.C.C., 1959. A new sub-species of *Plasmodium cynomolgi*. Riv. Parassitol. 20, 273–278.

Garnham, P.C.C., 1966. Malaria Parasites and Other Haemosporidia. Blackwell Scientific Publications, Oxford, U.K.

Guerra, C.A., Howes, R.E., Patil, A.P., Gething, P.W., Van Boeckel, T.P., Temperley, W.H., Kabaria, C.W., Tatem, A.J., Manh, B.H., Elyazar, I.R., Baird, J.K., Snow, R.W., Hay, S.I., 2010. The international limits and population at risk of *Plasmodium vivax* transmission in 2009. PLoS Negl. Trop. Dis. 4, e774.

Hester, J., Chan, E.R., Menard, D., Mercereau-Puijalon, O., Barnwell, J., Zimmerman, P.A., Serre, D., 2013. De novo assembly of a field isolate genome reveals novel *Plasmodium vivax* erythrocyte invasion genes. PLoS Negl. Trop. Dis. 7, e2569.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Krotoski, W.A., Bray, R.S., Garnham, P.C., Gwadz, R.W., Killick-Kendrick, R., Draper, C.C., Targett, G.A., Krotoski, D.M., Guy, M.W., Koontz, L.C., Cogswell, F.B., 1982a. Observations on early and late post-sporozoite tissue stages in primate malaria. II. The hypnozoite of *Plasmodium cynomolgi bastianellii* from 3 to 105 days after infection, and detection of 36- to 40-hour pre-erythrocytic forms. Am.J.Trop. Med. Hyg. 31, 211–225.

Krotoski, W.A., Garnham, P.C., Bray, R.S., Krotoski, D.M., Killick-Kendrick, R., Draper, C.C., Targett, G.A., Guy, M.W., 1982b. Observations on early and late post-sporozoite tissue stages in primate malaria. I. Discovery of a new latent form of *Plasmodium cynomolgi* (the hypnozoite), and failure to detect hepatic forms within the first 24 hours after infection. Am.J.Trop. Med. Hyg. 31, 24–35.

Leclerc, M.C., Durand, P., Gauthier, C., Patot, S., Billotte, N., Menegon, M., Severini, C., Ayala, F.J., Renaud, F., 2004. Meager genetic variability of the human malaria agent *Plasmodium vivax*. Proc. Natl. Acad. Sci. U. S. A. 101, 14455–14460.

Lee, K.S., Divis, P.C., Zakaria, S.K., Matusop, A., Julin, R.A., Conway, D.J., Cox-Singh, J., Singh, B., 2011. *Plasmodium knowlesi*: reservoir hosts and tracking the emergence in humans and macaques. PLoS Pathog. 7, e1002015.

Mayer, M., 1907. Ueber malaria beim affen. Med. Klin. Berl. 3, 579–580.

Menard, D., Chan, E.R., Benedet, C., Ratsimbasoa, A., Kim, S., Chim, P., Do, C., Witkowski, B., Durand, R., Thellier, M., Severini, C., Legrand, E., Musset, L., Nour, B.Y., Mercereau-Puijalon, O., Serre, D., Zimmerman, P.A., 2013. Whole genome sequencing of field isolates reveals a common duplication of the Duffy binding protein gene in Malagasy *Plasmodium vivax* strains. PLoS Negl. Trop. Dis. 7, e2489.

Miller, L.H., Mason, S.J., Dvorak, J.A., McGinniss, M.H., Rothman, I.K., 1975. Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. Science 189, 561–563.

Miller, L.H., Mason, S.J., Clyde, D.F., McGinniss, M.H., 1976. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. N. Engl. J. Med. 295, 302–304.

Mulligan, H.W., 1935. Descriptions of two species of monkey *Plasmodium* isolated from *Silenus irus*. Arch. Protist. 84, 285–314.

Neafsey, D.E., Galinsky, K., Jiang, R.H., Young, L., Sykes, S.M., Saif, S., Gujja, S., Goldberg, J.M., Young, S., Zeng, Q., Chapman, S.B., Dash, A.P., Anvikar, A.R., Sutton, P.L., Birren, B.W., Escalante, A.A., Barnwell, J.W., Carlton, J.M., 2012. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. Nat. Genet. 44, 1046–1050.

Nguyen-Dinh, P., Gardner, A.L., Campbell, C.C., Skinner, J.C., Collins, W.E., 1981. Cultivation in vitro of the vivax-type malaria parasite *Plasmodium cynomolgi*. Science 212, 1146–1148.

Ntumngia, F.B., King, C.L., Adams, J.H., 2012. Finding the sweet spots of inhibition: understanding the targets of a functional antibody against *Plasmodium vivax* Duffy binding protein. Int. J. Parasitol. 42, 1055–1062.

Oliveira, R.P., Broude, N.E., Macedo, A.M., Cantor, C.R., Smith, C.L., Pena, S.D., 1998. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. Proc. Natl. Acad. Sci. U. S. A. 95, 3776–3780.

Otsuki, H., Kaneko, O., Thongkukiatkul, A., Tachibana, M., Iriko, H., Takeo, S., Tsuboi, T., Torii, M., 2009. Single amino acid substitution in *Plasmodium yoelii* erythrocyte ligand determines its localization and controls parasite virulence. Proc. Natl. Acad. Sci. U. S. A. 106, 7167–7172.

Pacheco, M.A., Elango, A.P., Rahman, A.A., Fisher, D., Collins, W.E., Barnwell, J.W., Escalante, A.A., 2012a. Evidence of purifying selection on merozoite surface protein 8 (MSP8) and 10 (MSP10) in *Plasmodium* spp. Infect. Genet. Evol. 12, 978–986.

Pacheco, M.A., Reid, M.J., Schillaci, M.A., Lowenberger, C.A., Galdikas, B.M., Jones-Engel, L., Escalante, A.A., 2012b. The origin of malarial parasites in orangutans. PLoS One 7, e34990.

Pain, A., Bohme, U., Berry, A.E., Mungall, K., Finn, R.D., Jackson, A.P., Mourier, T., Mistry, J., Pasini, E.M., Aslett, M.A., Balasubrammaniam, S., Borgwardt, K., Brooks, K., Carret, C., Carver, T.J., Cherevach, I., Chillingworth, T., Clark, T.G., Galinski, M.R., Hall, N., Harper, D., Harris, D., Hauser, H., Ivens, A., Janssen, C.S., Keane, T., Larke, N., Lapp, S., Marti, M., Moule, S., Meyer, I.M., Ormond, D., Peters, N., Sanders, M., Sanders, S., Sargeant, T.J., Simmonds, M., Smith, F., Squares, R., Thurston, S., Tivey, A.R., Walker, D., White, B., Zuiderwijk, E., Churcher, C., Quail, M.A., Cowman, A.F., Turner, C.M., Rajandream, M.A., Kocken, C.H., Thomas, A.W., Newbold, C.I., Barrell, B.G., Berriman, M., 2008. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. Nature 455, 799–803.

Palatnik, M., Rowe, A.W., 1984. Duffy and Duffy-related human-antigens in primates. J. Hum. Evol. 13, 173–179.

Pattaradilokrat, S., Culleton, R.L., Cheesman, S.J., Carter, R., 2009. Gene encoding erythrocyte binding ligand linked to blood stage multiplication rate phenotype in *Plasmodium yoelii yoelii*. Proc. Natl. Acad. Sci. U. S. A. 106, 7161–7166.

Rayner, J.C., Huber, C.S., Galinski, M.R., Barnwell, J.W., 2004. Rapid evolution of an erythrocyte invasion gene family: the *Plasmodium reichenowi* Reticulocyte Binding Like (RBL) genes. Mol. Biochem. Parasitol. 133, 287–296.

Rayner, J.C., Tran, T.M., Corredor, V., Huber, C.S., Barnwell, J.W., Galinski, M.R., 2005. Dramatic difference in diversity between *Plasmodium falciparum* and *Plasmodium vivax* reticulocyte binding-like genes. Am.J.Trop. Med. Hyg. 72, 666–674.

Rice, B.L., Acosta, M.M., Pacheco, M.A., Carlton, J.M., Barnwell, J.W., Escalante, A.A., 2014. The origin and diversification of the merozoite surface protein 3 (msp3) multi-gene family in *Plasmodium vivax* and related parasites. Mol. Phylogenet. Evol. 78, 172–184.

Schmidt, L.H., Genther, C.S., et al., 1949. Development of resistance to chlorguanide (paludrine) during treatment of infections with *Plasmodium cynomolgi*. J. Pharmacol. Exp. Ther. 95, 382–398.

Schmidt, L.H., Greenland, R., Genther, C.S., 1961. The transmission of *Plasmodium cynomolgi* to man. Am.J.Trop. Med. Hyg. 10, 679–688.

Semenya, A.A., Tran, T.M., Meyer, E.V., Barnwell, J.W., Galinski, M.R., 2012. Two functional reticulocyte binding-like (RBL) invasion ligands of zoonotic *Plasmodium knowlesi* exhibit differential adhesion to monkey and human erythrocytes. Malar. J. 11, 228.

Siddiqui, A.A., Xainli, J., Schloegel, J., Carias, L., Ntumngia, F., Shoham, M., Casey, J.L., Foley, M., Adams, J.H., King, C.L., 2012. Fine specificity of *Plasmodium vivax* Duffy binding protein binding engagement of the Duffy antigen on human erythrocytes. Infect. Immun. 80, 2920–2928.

Sim, B.K., Chitnis, C.E., Wasniowska, K., Hadley, T.J., Miller, L.H., 1994. Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. Science 264, 1941–1944.

Sutton, P.L., Torres, L.P., Branch, O.H., 2011. Sexual recombination is a signature of a persisting malaria epidemic in Peru. Malar. J. 10, 329.

Ta, T.H., Hisam, S., Lanza, M., Jiram, A.I., Ismail, N., Rubio, J.M., 2014. First case of a naturally acquired human infection with *Plasmodium cynomolgi*. Malar. J. 13, 68.

Tachibana, S., Sullivan, S.A., Kawai, S., Nakamura, S., Kim, H.R., Goto, N., Arisue, N., Palacpac, N.M., Honma, H., Yagi, M., Tougan, T., Katakai, Y., Kaneko, O., Mita, T., Kita, K., Yasutomi, Y., Sutton, P.L., Shakhbatyan, R., Horii, T., Yasunaga, T., Barnwell, J.W., Escalante, A.A., Carlton, J.M., Tanabe, K., 2012. *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. Nat. Genet. 44, 1051–1055.

Takezaki, N., Nei, M., Tamura, K., 2010. POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol. Biol. Evol. 27, 747–752.

Tanabe, K., Escalante, A., Sakihama, N., Honda, M., Arisue, N., Horii, T., Culleton, R., Hayakawa, T., Hashimoto, T., Longacre, S., Pathirana, S., Handunnetti, S., Kishino, H., 2007. Recent independent evolution of msp1 polymorphism in Plasmodium vivax and related simian malaria parasites. Mol Biochem Parasitol. 156 (1), 74–79 PMID: 17706800.

Warren, M., Skinner, J.C., Guinn, E., 1966. Biology of the simian malarias of Southeast Asia. I. Host cell preferences of young trophozoites of four species of *Plasmodium*. J. Parasitol. 52, 14–16.

Waters, A.P., Higgins, D.G., McCutchan, T.F., 1993. Evolutionary relatedness of some primate models of *Plasmodium*. Mol. Biol. Evol. 10, 914–923.