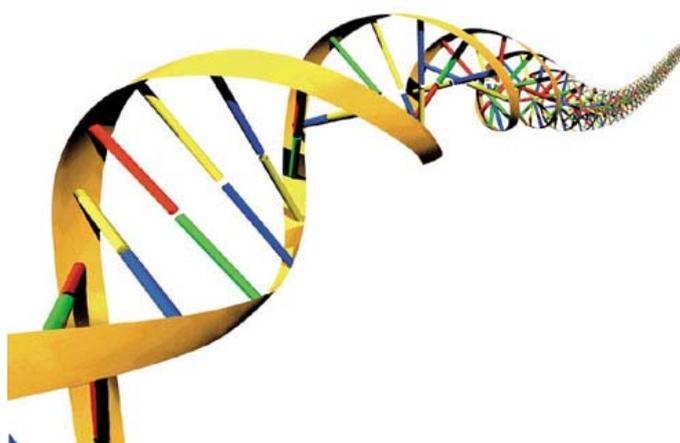


**Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde**

**Estudo da variabilidade de isolados de *Plasmodium vivax* de áreas endêmicas na
Amazônia Brasileira**



Antonio Mauro Rezende

**Belo Horizonte
Fevereiro/2009**

**Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde**

**Estudo da variabilidade de isolados de *Plasmodium vivax* de áreas endêmicas na
Amazônia Brasileira**

por

Antonio Mauro Rezende

**Dissertação apresentada com vistas à
obtenção do Título de Mestre em
Ciências da Saúde na área de
concentração de Biologia Celular e
Molecular.**

**Orientação: Dra. Cristiana F. A. de
Brito**

**Belo Horizonte
Fevereiro/2009**

Catálogo-na-fonte
Rede de Bibliotecas da FIOCRUZ
Biblioteca do CPqRR
Segemar Oliveira Magalhães CRB/6 1975

R467e
2009

Rezende, Antonio Mauro.

Estudo da variabilidade de *Plasmodium vivax* de áreas endêmicas na Amazônia Brasileira, ou, Variability study of *Plasmodium vivax* from endemic areas in Brazilian Amazon / Antonio Mauro Rezende. – Belo Horizonte, 2009.

xi, 79 f.: il.; 210 x 297mm.

Bibliografia: f.: 88 - 90

Dissertação (Mestrado) – Dissertação para obtenção do título de Mestre em Ciências pelo Programa de Pós - Graduação em Ciências da Saúde do Centro de Pesquisas René Rachou. Área de concentração: Biologia Celular e Molecular.

1. Malária vivax/genética 2. *Plasmodium vivax*/genética 3. Repetições de microssatélites I. Título. II. Brito, Cristiana Ferreira Alves de. (Orientação)

CDD – 22. ed. – 616.936 2

**Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde**

**“Estudo da variabilidade de isolados de *Plasmodium vivax* de áreas endêmicas na
Amazônia Brasileira”**

por

Antonio Mauro Rezende

Foi avaliada pela banca examinadora composta pelos seguintes membros:

Prof. Dra. Cristiana Ferreira Alves de Brito (Presidente)

Prof. Dr. Guilherme Corrêa de Oliveira

Prof. Dr. Gerhard Wunderlich

Suplentes: Dra. Roberta Lima Caldeira

Dissertação ou tese defendida e aprovada em: 09/02/2009

Agradecimentos

Ao meu Deus por sua presença inestimável nos momentos de dúvida, medo e escuridão.

Aos meus pais, que são exemplos de honestidade, caráter e que sempre me incentivaram para não desistir dos meus sonhos e desejos em uma vida repleta de conhecimentos.

À Tatiana (Taty ou Minha Linda), que já me acompanha a quatro anos e com certeza estará comigo até o fim da vida. Agradeço pela paciência, pelos incentivos e principalmente por todo amor e carinho.

Aos meus irmãos que de alguma maneira sempre me fizeram sentir orgulho de mim mesmo.

À minha Tia Rozana, que também foi sempre um grande exemplo em minha vida e que sempre me ofereceu confiança, carinho e amor.

Aos meus amigos Silvio e Leopoldo. Pois mesmo de longe estavam sempre próximos com e-mails e telefonemas.

À Cris, minha orientadora, pela paciência e grandes ensinamentos ao longo destes quatro anos e dois meses.

Aos amigos do LAMAL, principalmente, Sabrina, Armando e Bruno (Moço) pela ajuda, incentivo, companheirismo e por fazer com que eu tenha muito orgulho de ter feito parte da equipe deste laboratório.

Aos colegas do Centro de Pesquisas René Rachou por estes anos de ótimo convívio e por oferecer a estrutura necessária para realização deste projeto.

Ao Nilton e a Elisângela pela imensa ajuda com os seqüenciadores automáticos A.L.F. e MegaBACE na realização das genotipagens.

Ao programa de pós-graduação em Ciências da Saúde do Centro de Pesquisas René Rachou pelo oferecimento deste curso de mestrado com ênfase em Biologia Molecular e Celular.

À todas agências de fomento pelo apoio financeiro: CAPES, FAPEMIG, CNPq e FIOCRUZ.

Sumário

Lista de Figuras _____	VIII
Lista de Tabelas _____	VIII
Lista de Abreviaturas e Símbolos _____	IX
Resumo _____	X
Abstract _____	XI
1 Introdução _____	12
1.1 Malária _____	12
1.2 Contexto histórico e panorama atual da Malária _____	12
1.2.1 Malária no Mundo _____	12
1.2.2 Malária no Brasil _____	13
1.3 Ciclo de vida do <i>Plasmodium sp</i> _____	15
1.4 A variabilidade genética do <i>P. vivax</i> _____	19
1.4.1 TRs, Microssatélites e genética de populações em <i>P. vivax</i> _____	21
2 Justificativa _____	24
3 Objetivos _____	25
3.1 Objetivos Geral _____	25
3.2 Objetivos Específicos _____	25
4 Materiais e Métodos _____	26
4.1 Áreas de estudo e isolados de <i>P. vivax</i> _____	26
4.2 Extração de DNA _____	27
4.3 Busca de Novos Microssatélites _____	28
4.4 Amplificação dos Novos Microssatélites _____	29
4.5 Visualização da Amplificação em Gel de Agarose _____	30
4.6 Genotipagem no Seqüenciador Automático de DNA MegaBACE _____	31

4.7 Amplificação dos tandem repeats (TRs)	33
4.8 Genotipagem no Seqüenciador Automático de DNA A.L.F. (Automatic Laser Fluorescence)	34
4.9 Análises Populacionais	34
4.9.1 Número de alelos por locus e por população	34
4.9.2 Análise molecular de Variância	35
4.9.3 Índice de Fixação (F_{ST})	35
4.9.4 Heterozigosidade Esperada (H_E)	36
4.9.5 Teste de Mantel	37
4.9.6 Desequilíbrio de Ligação	37
4.9.7 Análise de estrutura de população utilizando inferência Bayesiana	38
4.9.8 “Bottleneck” ou Gargalo de Garrafa populacional	40
5 Artigos	42
5.1 Primeiro artigo: Analysis of genetic variability of <i>Plasmodium vivax</i> isolates from different Brazilian Amazon areas using tandem repeats	42
5.2 Segundo artigo: Determining the genetic variability of <i>Plasmodium vivax</i>	60
6 Considerações Finais	85
7 Referências Bibliográficas	88

Lista de Figuras

Figura 1 - Distribuição global do risco de transmissão da malária _____	12
Figura 2 - Mapa de risco de infecção na Amazônia Legal _____	14
Figura 3 - Ciclo de vida do <i>Plasmodium sp.</i> _____	17
Figura 4 – Gel de agarose para o microsatélite 6605 _____	30
Figura 5 - Eletroforese capilar no seqüenciador automático MegaBACE _	31
Figura 6 - Eletroforese realizada no seqüenciador automático A.L.F _____	33

Lista de Tabelas

Tabela 1 - Características Demográficas e Epidemiológicas dos isolados em estudo _____	26
Tabela 2 - Características dos Microsatélites utilizados _____	28
Tabela 3 - Iniciadores e condições de amplificação dos microsatélites _	29
Tabela 4 - Características, iniciadores e condições de amplificação dos TRs _____	32

Lista de Abreviaturas e Símbolos

AM	Estado do Amazonas
AMOVA	Análise Molecular de Variância
AP	Estado do Amapá
CEM	Campanha de Erradicação da Malária
CSP	Proteína de Superfície de circunesporozoíto
DDT	Dicloro Difenil Tricloreto
dNTP	Dideoxiribonucleotídeos
EDTA	Ácido Etilenodiamino Tetra-acético
H_E	Heterozigidade Esperada
ICBs	Blocos Conservados Inter-espécies
IPA	Índice Parasitário Anual
LD	Desequilíbrio de ligação
MSP	Proteína de Superfície de Merozoíto
MT	Estado do Mato-Grosso
PA	Estado do Pará
PCR	Reação em Cadeia da Polimerase
PIACM	Plano de Intensificações das Ações de Controle da Malária
PNCM	Plano Nacional de Controle da Malária
RO	Estado de Rondônia
SNP	Polimorfismo de Base Única
TIGR	Instituto de Pesquisas Genômicas
TR	Repetições em Tandem
WHO	Organização Mundial da Saúde

Resumo

A malária humana é causada por protozoários do gênero *Plasmodium*. Atualmente, mais de 2 bilhões de pessoas estão sob risco de contrair malária sobretudo em áreas tropicais e subtropicais. No Brasil, a malária está concentrada na região da Amazônia Legal, onde se registram mais de 99% dos casos, sendo cerca de 80% causados pelo *Plasmodium vivax*. Estratégias efetivas de combate ao parasito são essenciais para o controle da doença, e estas dependem do entendimento de diversos fatores como, por exemplo, a variabilidade genética e a estrutura populacional do *P. vivax*. Desta maneira, os marcadores moleculares são importantes ferramentas que podem auxiliar na elucidação destes aspectos biológicos do parasito. Contudo, poucos marcadores moleculares estão disponíveis para estudos populacionais de *P. vivax*, comparado ao *P. falciparum*, principalmente marcadores neutros ou quase neutros, ideais para esse tipo de estudo. Portanto, o objetivo deste trabalho foi identificar novos microssatélites em *P. vivax* e realizar análises de genética de população em isolados de cinco áreas da Amazônia Brasileira utilizando esses microssatélites e os cinco marcadores do tipo TR mais polimórficos descritos na literatura. Deste modo, uma busca *in silico* foi realizada utilizando o rascunho do genoma do parasito, e foram selecionados microssatélites com unidades repetitivas variando de 2 a 4 pares de base. Cada locus foi amplificado utilizando iniciadores específicos, incluindo os TRs cujos iniciadores foram baseados na literatura. O DNA molde para a realização da amplificação foi extraído de pacientes com infecção aguda de *P. vivax*. Os produtos amplificados foram analisados em seqüenciadores automáticos de DNA utilizando programas de análise de tamanhos de fragmentos. As análises de genética de população foram realizadas utilizando vários pacotes computacionais. Ambos marcadores detectaram uma alta diversidade nas populações estudadas. Porém, algumas diferenças foram identificadas nas análises com cada tipo de marcador molecular, dentre elas: as populações geneticamente próximas foram diferentes juntamente com o padrão de desequilíbrio de ligação em cada população e a taxa de multiplicidade de infecção também variou entre os tipos de marcadores. Além disso, foi possível, utilizando os microssatélites, estruturar as populações, o que não ocorreu utilizando os TRs. Por fim, foi possível perceber que os microssatélites se mostraram mais polimórficos, mais bem distribuídos pelo genoma do parasito e ainda assim, o seu mecanismo de manutenção de variabilidade é conhecido, logo, eles aparentam ser mais aptos na utilização de estudos de genética de população de *P. vivax*.

Abstract

Human malaria is caused by a protozoan from genus *Plasmodium*. Currently, more than 2 billion people are under risk to contract malaria mainly in tropical and subtropical areas. In Brazil, malaria is concentrated in the Legal Amazon, where more than 99% of the cases are registered, being around 80% caused by *Plasmodium vivax*. Effective strategies of combating are essential to control the disease, and these ones depend on understanding of many factors such as the genetic variability and population structure of *P. vivax*. Hence, molecular markers are important tools that can help to elucidate these biological aspects of the parasite. However, few molecular markers are available for population studies of *P. vivax*, compared with *P. falciparum*, mainly neutral or near neutral markers, which are indicated for this kind of study. Thus, the aim of this work was to identify new polymorphic microsatellite loci in *P. vivax* and to perform population genetic analyses with isolates from five areas of Brazilian Amazon using these microsatellites and also the five most polymorphic tandem repeats markers described in literature. For this, an *in silico* search was performed using the genome draft of the parasite, and microsatellite with repeat units ranging from 2 to 4 base pairs were selected. Each locus was amplified using specific primers, including the TRs whose primers were previously described. The template DNA for amplification was obtained from patient's blood with acute *P. vivax* infection. The amplified products were analyzed in automatic DNA sequencer using software to identify fragments size. The genetic population analyses were performed with several computational packages. Although, some differences were identified among the measured parameters with each kind of molecular marker: the populations genetically close, linkage disequilibrium pattern in each population and level of multiple infection were different. Besides, it was possible, using the microsatellites, to detect structuring of the populations, what did not happen using the TRs. In conclusion, it was possible to notice that the microsatellite are more polymorphic, more spread across parasite genome, and its variability maintenance mechanism is known; thus, they seem to be more useful in population genetic studies of *P. vivax* field isolates.

1 INTRODUÇÃO

1.1 Malária

A malária é causada por protozoários do filo *Apicomplexa*, gênero *Plasmodium*. As principais espécies causadoras da doença no homem são: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malarie* e *Plasmodium ovale*. Estes parasitos são organismos heteroexenos, tendo como hospedeiro vertebrado o homem e como hospedeiro invertebrado e definitivo mosquitos do gênero *Anopheles*, que são os vetores da doença.

A espécie humana e os parasitos causadores da malária possuem um longo tempo de interação evolutiva. Avanços nas metodologias para análise de dados apóiam hipóteses de que a cerca de 10.000 anos atrás, a população de *P. falciparum* rapidamente expandiu-se na África e disseminou-se mundialmente, coincidindo com crescimento da população humana e subseqüentes diásporas facilitadas pelo desenvolvimento da agricultura (1). Porém mesmo sendo uma doença tão antiga, a malária ainda é um dos principais problemas de saúde pública no mundo.

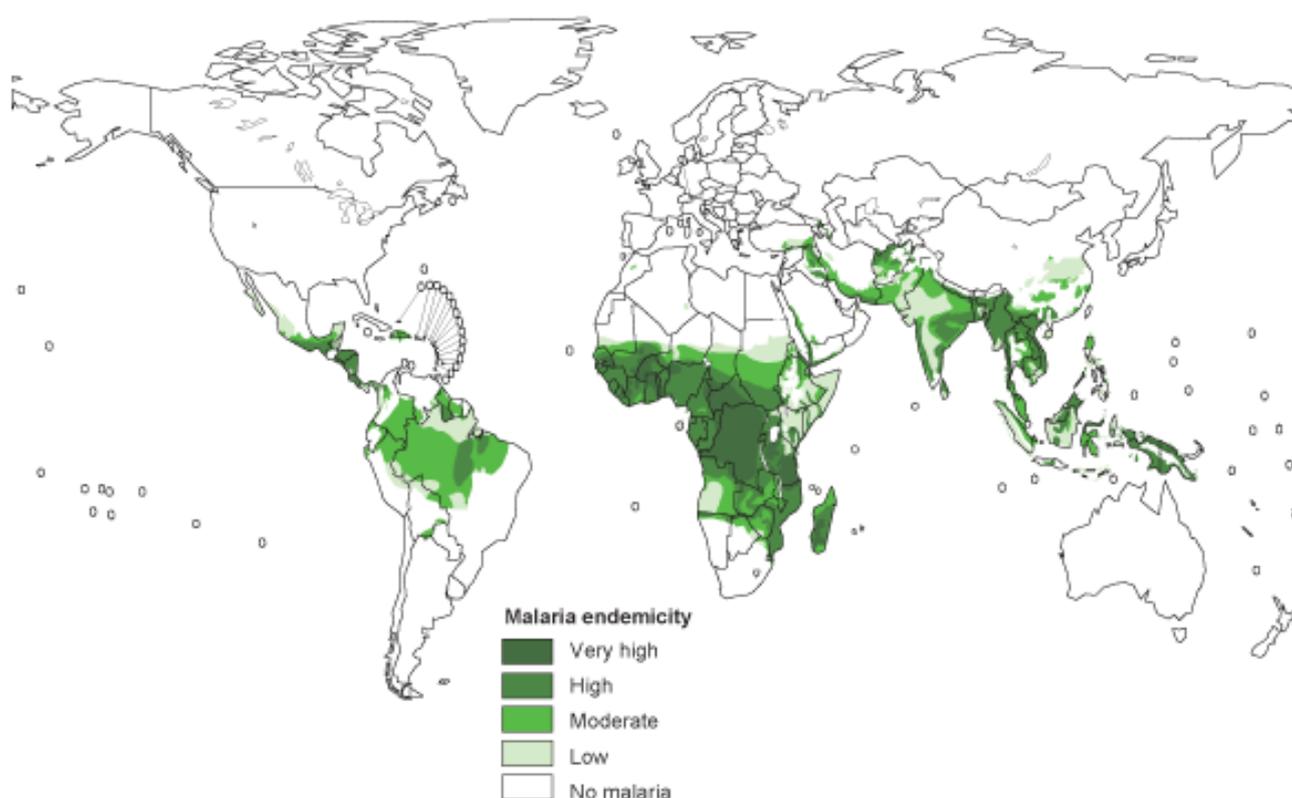
1.2 Contexto histórico e panorama atual da Malária

1.2.1 Malária no Mundo

Mais de 2 bilhões de pessoas estão sob risco de contrair a malária e as principais populações afetadas são majoritariamente pobres, vivendo em áreas tropicais e subtropicais (2) (Figura 1). Anualmente estima-se que ocorram de 300-500 milhões de casos globalmente, e 1 milhão de mortes, principalmente entre crianças com menos de cinco anos de idade na África Sub-Sahariana (3). Além disso, as perdas econômicas são graves, particularmente nesta região africana, uma vez que a malária pode custar mais de um per cento do crescimento econômico anual (3).

Porém, apesar da malária ser um problema de saúde pública atual, parte do seu cenário contemporâneo se remete ao século XIX. Neste período muitas

doenças, inclusive a malária, tiveram um melhoramento no controle devido às melhoras nas condições sociais, particularmente nas moradias e mudanças nas práticas agrícolas. A melhora no controle da malária frequentemente era coincidente com o desenvolvimento econômico e social da região. Contudo, no século XX, as diferenças na taxa de desenvolvimento das diferentes regiões do globo não diminuíram. Assim a malária não é necessariamente apenas uma doença tropical, mas mais precisamente é uma doença restrita aos trópicos através das diferenças no desenvolvimento socioeconômico e nas medidas de controle adotadas (1).



Fonte: <http://www.rbm.who.int/wmr2005/html/map1.htm>

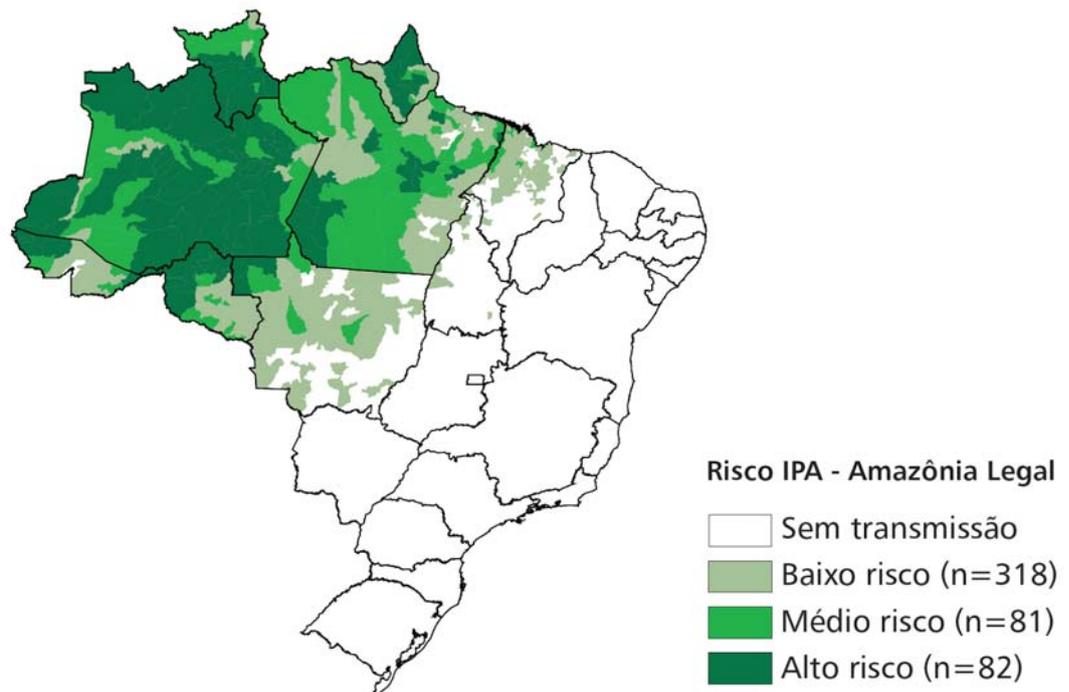
Figura 1 – Distribuição global do risco de transmissão da malária segundo o Relatório de Malária de 2005 da Organização Mundial de Saúde (WHO).

1.2.2 Malária no Brasil

No Brasil, a malária está concentrada na região da Amazônia Legal (Figura 2), onde se registram mais de 99% de todos os casos de malária do

país (4). Alguns planos de controles da doença já foram implantados desde década de 70, quando o número de casos começou a aumentar. O primeiro foi a Campanha de Erradicação da Malária (CEM) baseada na aplicação intradomiciliar do inseticida DDT e no uso de drogas antimaláricas (5). Porém em 1999, o número de casos chegou ao número máximo de 635.646 casos. Assim em 2000, outro plano de controle, Plano de Intensificação das Ações de Controle da Malária na Amazônia Legal (PIACM), foi implementado pelo Ministério da Saúde. Em 2003, visando melhorar e ampliar o atendimento aos portadores de malária, o Ministério implementou o Programa Nacional de Controle da Malária (PNCM), que procurou manter os ganhos obtidos pelo PIACM (4).

Apesar dos esforços dos programas de controle, a malária continua sendo uma endemia grave no Brasil. Cerca de 466.000 casos de malária foram relatados no último ano, sendo cerca de 80% causados pelo protozoário *P. vivax*. Além disso, a distribuição de casos e a incidência por estado em 2006 demonstram que a ocorrência da malária não é homogênea na Amazônia Legal, os estados do Amazonas, Rondônia, Pará e Acre foram responsáveis por 87,9% dos casos em 2006. Em 2007, os estados do Acre, Rondônia e Amazonas foram classificados como áreas de alto risco de transmissão, sendo a incidência, medida pelo Índice Parasitário Anual (IPA), maior de 50 casos por mil habitantes. Roraima, Amapá e Pará foram classificados de médio risco (IPA entre 10 – 49/1000 hab.) e os estados do Mato Grosso, Maranhão e Tocantins de baixo risco (IPA < 10/1000 hab.) (4). O IPA é usado para determinar o risco de transmissão, pois corresponde ao número de lâminas positivas para o *Plasmodium* sp. em cada 1000 habitantes.



Fonte: portal.saude.gov.br/portal/arquivos/pdf/folder_malaria_2008_final.pdf

Figura 2 – Mapa de risco da malária por município de infecção na Amazônia Legal, área de transmissão brasileira, em 2007, segundo Secretaria de Vigilância em Saúde (SVS, 2008). Municípios classificados segundo o IPA (Índice parasitário anual): baixo risco ($IPA < 10$), médio risco ($10 > IPA < 50$) e alto risco ($IPA > 50$).

1.3 Ciclo de vida do *Plasmodium sp.*

A infecção do hospedeiro humano com o parasito *Plasmodium sp.* inicia com a picada de mosquitos do gênero *Anopheles* infectado, que inocula os parasitos na forma de esporozoítos (2) (Figura 3). Neste estágio o parasito é alongado e ativamente móvel. Após serem inoculados, cerca de 30% dos esporozoítos invadem e movimentam-se ao longo dos vasos linfáticos. Estes parasitos que migram através da linfa, provavelmente não estarão hábeis para posterior etapa no ciclo. Assim, aparentemente os esporozoítos cujos destinos foram o sistema linfático cessam sua jornada no nódulo linfático mais próximo ao sítio da inoculação (6). Contudo, cerca de 70% dos esporozoítos invadem os vasos sanguíneos e através da corrente sanguínea, estes parasitos alcançam os capilares sinusóides no fígado ligando-se aos receptores de heparan-

sulfato, atravessam o endotélio e posteriormente as células de Kupffer e invadem ativamente vários hepatócitos antes de se instalarem em um hepatócito final (7). Os co-receptores no esporozoítos que medeiam estes processos invasivos são em parte os domínios trombospondina da proteína de circunsporozoítos (CS) e a proteína adesiva relacionada à trombospondina (TRAP) (8).

O esporozoítos dentro do hepatócito diferencia-se em um esquizonte multinucleado e este então gera vários merozoítos que são uma nova forma morfológica invasiva aos eritrócitos. Porém, para ter acesso aos eritrócitos, o merozoítos hepático deve alcançar o lúmen dos capilares sinusóides no fígado, atravessando a própria membrana do hepatócito, o espaço de Disse, posteriormente a camada de matriz extracelular e o endotélio do sinusóide. Além disso, ele deve evitar a fagocitose pelos inúmeros macrófagos, células de Kupffer, que patrulham os sinusóides hepáticos (9). Assim, para conseguir ultrapassar estes obstáculos biológicos, o parasito apresenta mecanismos para “manipular” os hepatócitos e garantir com segurança a chegada dos merozoítos na corrente sanguínea. Assim, o parasito faz com que os hepatócitos formem grandes vesículas, chamadas de merosomas, que possuem de poucos a milhares de merozoítos. Além disso, sabe-se que as células infectadas acabam morrendo. Contudo o parasito consegue fazer com que estas células, e também os merosomas, expressem marcadores de células viáveis, como por exemplo, distribuição assimétrica de resíduos de fosfatidilserina na membrana celular, para adiar a morte celular. Deste modo as células infectadas e os merosomas não são reconhecidos pelo sistema imune, assim os merosomas vão para os sinusóides, agindo como entregadores que asseguram a liberação dos merozoítos na corrente sanguínea (9).

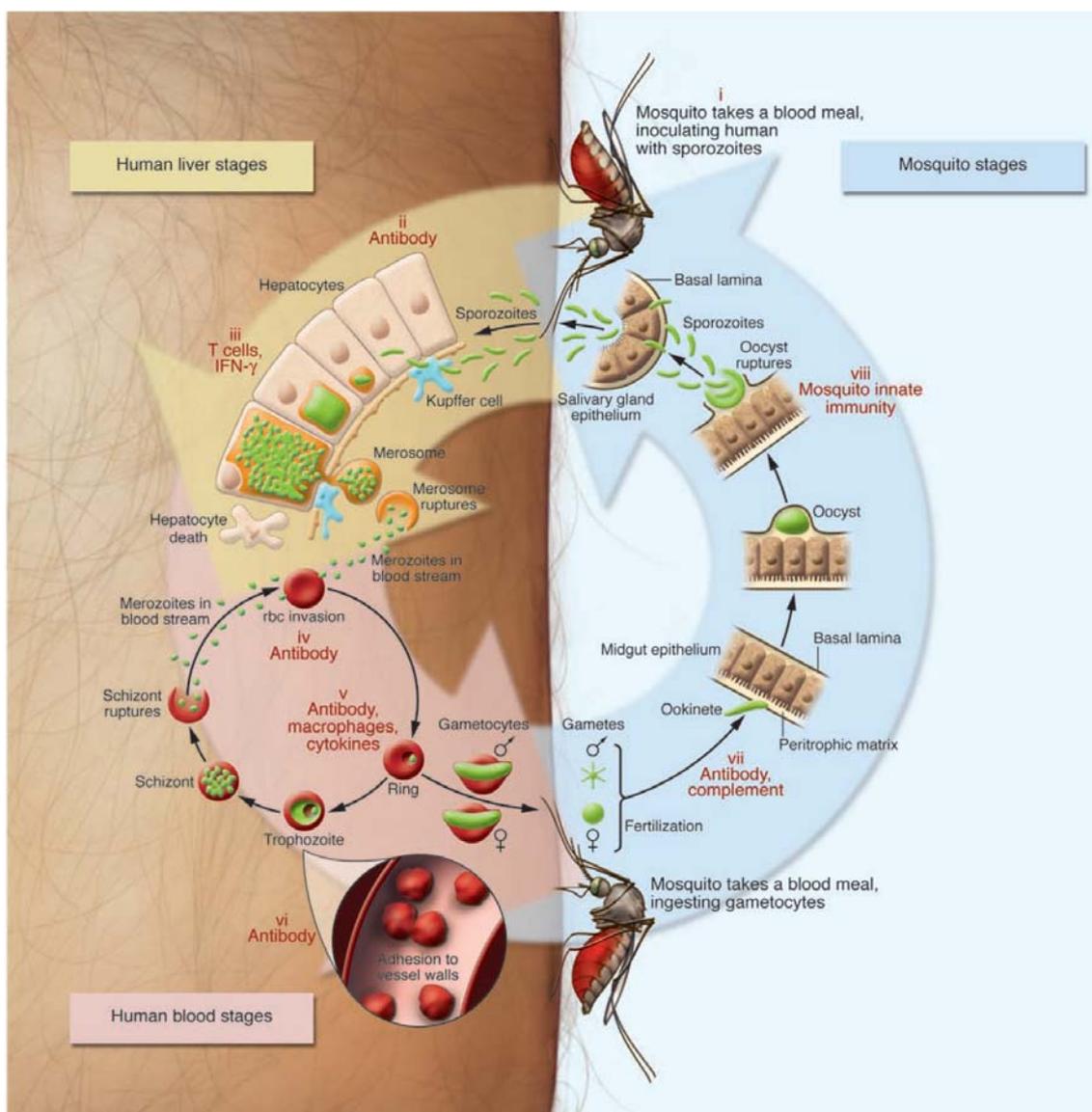
Após os merozoítos serem liberados na corrente sanguínea, eles reconhecem receptores na membrana dos eritrócitos, se ligando a eles, depois sofrem reorientação apical e ocorre a formação da junção irreversível. Os parasitos então induzem a formação de um vacúolo derivado da membrana celular do eritrócito e entram neste movendo a junção. O merozoítos dentro do eritrócito inicia a fase assexuada do ciclo transformando-se em trofozoítos jovens e posteriormente trofozoítos maduros e após a sua multiplicação através de esquizogonia dentro da célula, produz cerca de 20 merozoítos, que estão aptos

a invadirem novas hemácias (10). *P. falciparum* e *P. vivax* levam cerca 48 horas para completarem o desenvolvimento nas hemácias em cada rodada do ciclo assexuado (8).

É importante ressaltar que grande parte dos sintomas e da patogenia da doença é decorrente deste momento do ciclo de vida do parasito. Ambos, *P. falciparum* e *P. vivax* podem causar anemia grave, porém somente *P. falciparum* causa complicações como malaria cerebral, hipoglicemia e acidose metabólica. Certamente diferenças na biologia dos dois parasitos parcialmente explicam as diferenças nos padrões da doença. Primeiro, *P. falciparum* pode invadir um grande percentagem dos eritrócitos, enquanto *P. vivax* está limitado aos reticulócitos (hemácias jovens). Segundo, existe um número maior de vias de invasão do eritrócito utilizadas pelo *P. falciparum* quando comparado ao *P. vivax*. Este último invade principalmente as hemácias positivas para o grupo sanguíneo Duffy. O primeiro pode usar várias vias redundantes para invadir as hemácias que não possuem um receptor em particular como por exemplo os receptores que possuem ácido siálico (8).

Dando continuidade ao ciclo, uma pequena proporção dos parasitos assexuados converte-se em gametócito. Estes podem então ser ingeridos pelo mosquito vetor após seu repasto sanguíneo. Estas mudanças abruptas de ambiente e fatores específicos do vetor estimulam o parasito a sair da hemácia. Dentre estes estímulos estão: queda de temperatura em torno de 5°C, um aumento do pH e exposição ao ácido xanturênico. Desta maneira, uma cascata de transdução de sinal resulta na liberação de cálcio no citoplasma do gametócito causando então o início do seu desenvolvimento sexual (11). O gametócito macho se transforma então em oito microgametócitos móveis após três rodadas de rápida replicação genômica seguida por divisão nuclear e uma quinase cálcio-dependente, CDPK4 é ativada regulando a exflagelação. Já o gametócito-fêmea emerge do eritrócito com uma morfologia arredondada. Então, o zigoto, resultante da fertilização entre micro e macrogametócito, será o único parasito sobrevivente no agressivo ambiente do intestino do mosquito (11,12). Ele se transforma em oocineto, que atravessa a parede do intestino. Uma minoria dos oocinetos sobreviventes ao processo de invasão transforma-se em oocistos no espaço basal subepitelial. Deste modo, após um passo de reprodução, em cada oocisto são produzidos milhares de esporozoítos que

migram para a glândula salivar do vetor. Estes então infectarão um novo hospedeiro vertebrado durante o próximo repasto sanguíneo do vetor, fechando assim o ciclo de vida do parasito (13). Esta fase sexuada do ciclo possui uma grande importância evolutiva para o parasito, pois é neste momento em que ocorrem recombinações meióticas, podendo haver geração de isolados com diferentes perfis haplotípicos, o que pode ter reflexo no fenótipo do parasito.



Fonte: Greenwood *et al.* 2008

Figura 3 – Desenho esquemático representando o ciclo de vida do *Plasmodium* sp. Ilustrando suas diversas etapas, incluindo as recentes alterações.

1.4 A variabilidade genética do *P. vivax*

Alguns parasitos do gênero *Plasmodium* já possuem seus genomas sequenciados como, por exemplo, *P. falciparum*, *P. yoelli* e *P. vivax*. Em muitos aspectos, os genomas destes parasitos de mamíferos são uniformes, variando de 23 a 27 Mega bases ao longo de 14 cromossomos. Além disso, possuem cerca de 5500 genes sendo que a maioria deles (51%) contém pelo menos um íntron. O genoma de *P. vivax* possui 26,8 Mega bases sendo que o conteúdo de G + C chega a 42,3%, consideravelmente maior que o do *P. falciparum*. Além disso, os cromossomos do genoma de *P. vivax* são os únicos, das espécies de *Plasmodium* que infectam humanos, que exibem uma estrutura onde as regiões subteloméricas possuem baixo conteúdo de G+C e as regiões internas com um conteúdo de G+C significativamente mais alto (16). Sobre a diversidade deste genoma, estudos de polimorfismos de base única (single nucleotide polymorphisms – SNPs) mostraram que o genoma de *P. vivax* possui uma grande variabilidade, em muitos casos a variabilidade em alguns loci é maior em *P. vivax* do que em *P. falciparum* (15).

Mais detalhadamente, duas abordagens básicas estão sendo empregadas para estudos da variabilidade genética de parasitos causadores da malária. A primeira tem como objetivo entender a estrutura genética e a dinâmica populacional do parasito. Já na segunda, o objetivo é avaliar a diversidade de genes específicos que codificam antígenos alvos para vacina e genes associados com resistência às drogas antimaláricas (14).

Os estudos sobre estrutura populacional avaliam como os alelos estão espacialmente distribuídos e como a diversidade genética está organizada. Estes estudos fornecem indicações importantes sobre a origem, dispersão e estabilidade de genótipos multilocus. Informação esta essencial para predizer e monitorar a efetividade das estratégias de intervenção tais como: o uso de combinações específicas de drogas ou regimes terapêuticos; a dispersão de resistência a drogas e aparecimento de parasitos resistente a multidroga; o impacto a longo prazo de intervenções físicas como mosquiteiros; e a origem de genótipos que podem enfraquecer a resposta imune gerada por vacinas multivalentes (14).

Porém, poucos estudos sobre a diversidade genômica de *P. vivax* foram realizados devido à escassez de marcadores moleculares, provavelmente em decorrência da baixa taxa de mortalidade associada à infecção pelo *P. vivax* e a indisponibilidade de um sistema de cultivo *in vitro* deste parasito (15). Contudo, recentemente o genoma de *P. vivax* foi publicado e junto a este trabalho uma lista de marcadores moleculares do tipo microssatélite foi divulgada (16). Assim, apesar do aumento significativo no número de marcadores disponíveis para estudos de diversidade em isolados de *P. vivax*, comparado ao *P. falciparum* ainda são bastante reduzidos. Além disso, os estudos até então realizados focaram moléculas que estão sob seleção pelo sistema imune, particularmente antígenos ortólogos de *P. falciparum*, como MSP (merozoite surface protein) e CSP (circumsporozoite surface protein).

Um número de genes codificando proteínas MSPs em *P. vivax* já foi identificado, incluindo *pvmsp-1*, *pvmsp-3 α* , *pvmsp-3 β* e *pvmsp-3 γ* , *pvmsp-4*, *pvmsp-5* e *pvmsp-9* (17). Os genes *pvmsp-1* e *pvmsp-3 α* são altamente polimórficos e são atualmente usados para genotipagem de isolados de campo de *P. vivax*. A estrutura primária do gene *pvmsp-1*, caracterizado originalmente em duas linhagens de *P. vivax* adaptadas a macacos, exibe regiões conservadas, semiconservadas e regiões polimórficas. Este gene tem a organização genética similar ao *pfmsp-1* e no ortólogo de *Plasmodium yoelli* (*pymmsp-1*), com regiões altamente conservadas entre as espécies, denominadas bloco conservados interespecies (interspecies conserved blocks – ICBs). Análise de seqüência da região entre ICB5 e ICB6 de parasitos isolados do Brasil, Sri Lanka, Colômbia, Papua Nova Guiné e Tailândia demonstra a natureza dimórfica desta região e sugere a recombinação intragênica como mecanismo de manutenção de diversidade antigênica. Contudo muitos estudos com *pvmsp-1* mostram evidências de recombinação genética de diferentes alelos e alguns ainda detectam agrupamentos geográficos de variações específicas (14). Amplificação e análise de uma série de seqüências do gene *pvmsp-3 α* de isolados de *P. vivax* de pacientes de diversas regiões geográficas (América do Sul e Central, Índia e sul da Ásia) e linhagens do parasito adaptadas a macacos mantidos em laboratório indicam um alto grau de polimorfismo. A diversidade presente na seqüência tende a estar largamente restrita ao domínio central a molécula. A natureza polimórfica

desta molécula sugere que ela pode estar sobre pressão do sistema imune (18).

Genes codificadores das CSPs de várias espécies de *Plasmodium* já foram clonados, sendo este gene um dos primeiros alvos para estudos de vacina. Em todas as espécies de *Plasmodium*, ele é um gene de cópia única, com um domínio central de seqüências repetidas em série sendo flanqueado predominantemente por seqüências conservadas não repetitivas. O domínio repetitivo central varia na seqüência e no tamanho entre as espécies de *Plasmodium*. Análises da seqüência do gene *pvcsp* revelaram que o parasito possui dois tipos de unidades repetitivas, denominadas cada uma como VK210 e VK247 (19, 20). Ambas variantes têm uma distribuição global, e análise de seqüências do *pvcsp* em amostras do sul da Ásia indica muito pouco polimorfismo fora da região repetitiva. Há pouco tempo atrás *pvcsp* foi empregado com sucesso em estudos epidemiológicos de *P. vivax* usando métodos bem estabelecidos: análise de hibridização molecular de DNA do parasito isolado de amostras de sangue e ELISA para detectar proteínas de esporozoítos em mosquitos artificialmente infectados ou coletados do campo (14).

Porém, apesar de todos estes estudos utilizando como marcadores os genes comentados acima, alguns parâmetros populacionais são melhores investigados usando marcadores que são neutros ou não estão sob forte pressão seletiva, como os microssatélites e as repetições em série (Tandem Repeats – TRs) (15, 21)

1.4.1 Tandem Repeats (TRs), Microssatélites e genética de populações em *P. vivax*

Os TRs, até o presente momento, somente foram utilizados para estudos populacionais em *P. vivax*. Estes marcadores moleculares aparentemente são típicos deste parasito. Eles poderiam ser classificados como uma espécie de minisatélite, devido ao tamanho do arranjo e disposição das unidades repetitivas. Foram descritos primeiramente por Feng e colaboradores em 2003 (15). Em 2004 Leclerc e colaboradores (21) utilizaram dos oito TRs dos 33 descritos no trabalho anterior, para análise populacional de

92 amostras de *P. vivax* provenientes do Azerbaijão, Tailândia, Turquia, Venezuela e Etiópia. A variabilidade destes marcadores nestas amostras variou de 2 a 10 alelos. Infelizmente, até o presente momento nenhum outro estudo foi realizado utilizando estes marcadores e o trabalho original possuiu apenas uma descrição superficial sobre a origem e manutenção da diversidade destas seqüências repetitivas.

Os microssatélites são os marcadores moleculares mais utilizados para estudos populacionais em vários organismos, isso se deve principalmente por suas características como: alta reprodutibilidade, natureza multi-alélica, padrão de herança codominante, abundância e ampla cobertura genômica (22). Eles consistem de pequenas repetições em série, cada repetição possui entre um a seis nucleotídeos. A diversidade dos microssatélites é mantida através da perda e ganho repetições, e estas variações genéticas ocorrem primariamente através de um processo onde as fitas complementares de DNA se pareiam de modo errado ocorrendo um subsequente erro durante a replicação. Este processo é denominado deslizamento das fitas ou “slippage” da DNA polimerase, e além deste, outros processos biológicos como o reparo de DNA e a recombinação estão envolvidos na manutenção da diversidade dos microssatélites (23). Além disso, os microssatélites foram classificados por Weber (24) como; perfeitos, quando existe um motivo único de repetição, sem interrupções; imperfeitos, quando existem bases diferentes intercaladas na seqüência que se repete; compostos, quando uma repetição perfeita ou imperfeita está associada à outra repetição com motivo diferente.

Análises de variação alélica em loci independentes múltiplos, como microssatélites, fornecem o meio mais efetivo para determinar estrutura e diversidade populacional de diferentes organismos incluindo os plasmódios (25). Em razão disto, alguns autores têm discutido o quão polimórfico são os microssatélites de *P. vivax* (21),(26), (27). Leclerc e colaboradores em 2004 encontraram uma baixa variabilidade nos microssatélites estudados. Neste estudo, dos 13 microssatélites apenas um mostrou alta variabilidade, sendo 9 completamente monomórficos entre oito populações analisadas. Já em 2006 usando o rascunho do genoma de *P. vivax* não publicado, Imwong e colaboradores identificaram 11 microssatélites polimórficos com repetições de dinucleotídeo. Este estudo apresentou dados sugerindo que *P. vivax* possui

uma alta diversidade alélica. Nos dados de Leclerc e colaboradores de 2004, os microssatélites isolados possuem arranjos muito pequenos (média de 5,5 repetições) e não seria esperado mostrar altos níveis de variação. Já Imwong e colaboradores em 2006 analisaram microssatélites com 12-18 repetições (média de 16 repetições). Por esta razão, a explicação dada por Imwong para essa clara diferença entre os resultados dela e Leclerc foi que a variação nos microssatélites é dependente do número de repetições, isto é, do tamanho do arranjo. Arranjos longos são mais variáveis do que curtos, pois, mutações se tornam exponencialmente mais comuns com o aumento do arranjo.

Recentemente, outros estudos populacionais para *P. vivax* foram publicados utilizando como ferramenta molecular os microssatélites. Imwong e colaboradores em 2007 (25) utilizando 9 loci de microssatélites e 345 amostras de populações da Ásia e Colômbia, conseguiram relatar a existência de uma alta diversidade genética nas populações, além disso foi possível analisar a estruturação populacional entre as populações estudadas. Ferreira e colaboradores em 2007 (28) através da genotipagem de 74 amostras de uma região endêmica para malária no Brasil (Acre) e utilizando 14 loci de microssatélites, relataram uma baixa taxa efetiva de recombinação devida provavelmente à alta taxa de endogamia entre as amostras. Ainda neste estudo foi possível observar um alto nível de substituição dos haplótipos em um período de tempo de 15 meses. Karunaweera e colaboradores em 2008 (29) usando os 14 loci de microssatélites descritos anteriormente, mostraram a existência de uma grande diversidade genética dentro do táxon *P. vivax* e na análise de estrutura de populações foram identificadas diferenças genéticas maiores em amostras do mesmo continente. Joy e colaboradores em 2008 (30) também utilizando marcadores moleculares do tipo microssatélite, 24 loci, conseguiram visualizar a estrutura de população existente entre uma população de parasito de uma região do interior do México e uma população da região litorânea. Esta estruturação refletiu a distribuição de duas diferentes espécies de mosquitos vetores da região (*Anopheles pseudopunctipennis* e *Anopheles albimanus*). Desta maneira, o estudo sugere que uma seleção recíproca entre os parasitas da malária e os mosquitos vetores tem levado a adaptação local do parasito e esta adaptação pode ter um importante papel na geração da estrutura de população de *Plasmodium* (30).

2 JUSTIFICATIVA

A epidemiologia da malária, como o resultado das complexas interações entre humanos, vetores e parasita, varia consideravelmente em diferentes regiões geográficas. Resultados de pesquisas extensivas sobre a estrutura de população de *Plasmodium falciparum*, na última década, demonstraram que a estrutura das populações deste parasito é fortemente associada aos padrões de transmissão (25, 31). Este conhecimento pode contribuir enormemente para o entendimento da dinâmica da transmissão e auxiliar na determinação das estratégias de controle da doença. Contudo, em comparação com o *P. falciparum* ainda existem poucos estudos populacionais sobre o *P. vivax*, utilizando marcadores neutros ou quase neutros, que são os ideais para estimar determinados parâmetros populacionais. Além disso, poucos trabalhos têm analisado simultaneamente isolados de diferentes regiões geográficas de um mesmo país utilizando análise multilocus. Assim, este trabalho propõe utilizar marcadores moleculares já disponíveis e aumentar o número de novos marcadores moleculares para estudos populacionais de isolados de campo de *P. vivax* de áreas endêmicas do Brasil com diferentes características epidemiológicas. Os marcadores já disponíveis e que foram utilizados primeiramente neste trabalho foram os “tandem repeats” – (TRs) devido a falta de microssatélites polimórficos no princípio do projeto. E devido a esta mesma razão, o trabalho propõe em um segundo momento buscar novos microssatélites polimórficos.

3 Objetivos

3.1 Objetivo Geral

Identificar novos microssatélites polimórficos e utilizá-los junto a cinco marcadores do tipo “tandem repeats” (TRs), previamente descritos, em estudos populacionais de isolados de *Plasmodium vivax* de cinco diferentes regiões da Amazônia brasileira.

3.2 Objetivos Específicos

- a) Determinar a variabilidade dos TRs, previamente descritos, em isolados de *P. vivax* das regiões da Amazônia Brasileira;
- b) Identificar novos microssatélites em seqüências genômicas de *P. vivax* através de ferramentas computacionais;
- c) Determinar a variabilidade dos novos microssatélites nos isolados brasileiros de *P. vivax*;
- d) Elucidar a estrutura populacional dos isolados de *P. vivax* de cinco diferentes regiões da Amazônia brasileira.

4 Materiais e Métodos

4.1 Áreas de estudo e isolados de campo de *P. vivax*

Os 64 isolados de *P. vivax* foram obtidos a partir de sangue de pacientes com infecção ativa coletados em cinco diferentes regiões da Amazônia Brasileira: 11 amostras de Cuiabá (15°35'45"S, 56°05'49"W) no estado de Mato Grosso (MT), 11 amostras de Macapá (00°02'20"S, 51°03'59"W) no estado do Amapá (AP), 16 amostras de Porto Velho (8°45'30", 63°24'28"W) no estado de Rondônia (RO), 15 amostras de Augusto Corrêa (01°01'18"S, 46°38'06"W) no estado do Pará (PA) e 11 amostras de Manaus (03°06'07"S, 60°01'30"W) no estado do Amazonas (AM). As amostras foram coletadas nas diferentes regiões em: AM, maio de 2003; AP, Novembro de 2004; PA, Outubro de 2005; RO, outubro a dezembro de 2004; e MT, de Julho de 2003 a julho de 2004. As amostras provenientes de Cuiabá, Macapá, Augusto Corrêa e Manaus foram utilizadas para o estudo populacional de *P. vivax* usando como marcador molecular os TRs. Para o estudo realizado com os microssatélites como marcadores moleculares, foram utilizadas as amostras de Macapá, Augusto Corrêa, Manaus e Rondônia.

Os dados epidemiológicos disponíveis indicam que as regiões estudadas são consideradas de hipo a mesoendêmicas para malária (32). A taxa de transmissão, medida pelo Índice Parasitário Anual (IPA), reflete o número de esfregaços sanguíneos positivos por 1000 habitantes. De acordo com a Secretária de Vigilância em Saúde do Ministério da Saúde, este índice é usado para classificar as áreas baseado no risco de infecção por malária: alto risco de infecção (IPA>50), médio risco de infecção (IPA entre 10 e 50) e baixo risco de infecção (IPA<10) (4). No momento da coleta de sangue das amostras estudadas, as áreas deste estudo foram classificadas: Augusto Correa/PA e Porto Velho/RO como áreas de alto risco, Manaus/AM como área de médio risco e Macapá/AP como área de baixo risco de infecção. Pacientes que vieram do centro de saúde em Cuiabá/MT relataram diferentes municípios de contágio, com diferentes valores de IPA, assim, esta área foi classificada como tendo risco variado. As infecções foram confirmadas por microscopia óptica de

esfregaços sanguíneos corados com Giemsa. Todos os pacientes tinham entre 16 a 90 anos de idades (média de idade de 32 anos) (Tabela 1).

Tabela 1 - Características Demográficas e Epidemiológicas dos isolados em estudo

Pacientes infectados com <i>P. vivax</i>			Localidade da coleta de sangue		
N	Homens (%)	Média de idade (anos)	Cidade/Estado	IPA ^a	Risco de Infecção ^b
11	62	30 (8-58)	Macapá / AP	9	Baixo risco
11	55	35 (17-58)	Manaus / AM	50	Médio risco
15	40	29 (6-90)	Augusto Correa/PA	119	Alto risco
16	80	32 (16-84)	Porto Velho / RO	139 ^c	Alto risco
11	76	30 (8-52)	Cuiabá / MT	ND ^c	Risco variado

^a Índice parasitário anual (número de esfregaços positivos por 1000 habitantes) das localidades de contágio na época da coleta das amostras.

^b Estratificação baseada no IPA: alto risco (IPA>50); médio risco (IPA 10 to 50); baixo risco (IPA<10).

^c ND – não determinado

4.2 Extração de DNA

O DNA do parasito foi extraído do sangue total das amostras coletadas usando o kit para extração Purgene DNA (Gentra Systems, Minneapolis, MN) de acordo com o protocolo do fabricante: Para cada 1 mL de sangue total 3 mL de solução de lise para eritrócitos são utilizados por 10 minutos em temperatura ambiente. Posteriormente a mistura é centrifugada a 2000 x g por 10 minutos; o sobrenadante então é removido, ficando no tubo somente o sedimento. Após este passo, o sedimento é ressuspensionado em 1 ml de solução de lise celular. Adiante, 333 µL de solução de precipitação de proteína são adicionados ao lisado celular e esta mistura é homogeneizada. Posteriormente ocorre uma centrifugação a 2000 x g por 10 minutos e, depois o sobrenadante é colocado em um tubo contendo 1 mL de isopropanol. Esta solução é homogeneizada e posteriormente centrifugada a 2000 x g por 3 minutos. O sobrenadante então é jogado fora e é adicionado ao sedimento 1 mL de etanol

70%. Outra centrifugação é realizada a 2000 x g por 1 minuto e o sobrenadante é novamente descartado. Finalmente são adicionados 100 µL de solução de hidratação de DNA ao sedimento. O DNA deve ser hidratado incubando-o a 65°C por 1 hora ou 16 horas em temperatura ambiente. O DNA é então armazenado a -20°C.

A extração de DNA foi realizada pelos membros do laboratório que haviam participado da coleta das amostras nas áreas endêmicas.

4.3 Busca de Novos Microssatélites

Através do endereço eletrônico do TIGR (The Institute for Genomic Research - www.tigr.org), foi possível ter acesso às seqüências genômicas de *P. vivax*. Estas seqüências foram obtidas em fevereiro de 2006. O arquivo com as seqüências genômicas continha muitos contigs, que serviram como arquivo de entrada para o programa RepeatMasker (<http://www.repeatmasker.org/>).

O Repeat Masker gerou uma lista de seqüências repetitivas, incluindo microssatélites perfeitos e regiões de baixa complexidade. A partir desta lista nós selecionamos algumas seqüências utilizando os seguintes critérios: microssatélites perfeitos, tamanho da unidade repetitiva, número de repetições e presença de regiões flanqueadoras para desenho dos iniciadores. Foram escolhidos 23 loci, dos quais 11 foram polimórficos e os outros não amplificaram bem ou foram monomórficos. Dos 11 loci polimórficos, 4 loci continham repetições de dinucleotídeos, 3 loci com repetições de trinucleotídeos e 4 loci com repetições de tetranucleotídeos (Tabela 2). Os iniciadores foram desenhados utilizando o programa Oligo versão 4.0. (Tabela 3). Os iniciadores senso de cada locus polimórfico foram adquiridos com fluoresceína para realização da genotipagem dos produtos amplificados no seqüenciador automático MegaBACE (Amershan Bioscience).

Os 11 loci que foram utilizados no estudo foram localizados no genoma de *P. vivax* através do programa Blastn. O programa foi executado comparando as sequências dos loci com as sequências genômicas do parasito, que foram atualizadas em setembro de 2008. Os parâmetros para execução do programa foram: wordsize com valor 7 e filtro para repetições desligado.

Tabela 2 - Características dos Microssatélites utilizados

Microssatélite	Cromossomo ^a	Unidade Repetitiva ^b	Varição no tamanho do Arranjo (pb) ^c	Número de Alelos ^d
PvMS1	12	GT ₉	224-248	4
PvMS2	3	CA ₁₂	292-312	8
PvMS3	8	TA ₁₈	145-197	10
PvMS4	6	TA ₂₁	130-174	15
PvMS5	3	CAT ₁₀	172-218	12
PvMS6	14	TGA ₁₉	185-301	19
PvMS7	2	TAA ₂₂	349-439	10
PvMS8	13	TGTA ₇	225-321	9
PvMS9	6	CATA ₁₁	193-217	7
PvMS10	5	TATG ₁₄	265-376	10
PvMS11	8	CATA ₄₉	292-596	31

^a Determinado da identidade dos iniciadores com sequências dos bancos de dados utilizando a ferramenta Primer-Blast (www.ncbi.nlm.nih.gov)

^b Número de repetições para Sal-1 de acordo com a sequência obtida no TIGR

^c Variação no tamanho do arranjo encontrada neste trabalho

^d Número de alelos detectados neste estudo

4.4 Amplificação dos novos microssatélites

A reação em cadeia da polimerase ou PCR foi padronizada usando o termociclador Eppendorf Gradient. As temperaturas de anelamento dos iniciadores e a concentração de cloreto de magnésio variaram de 50°C a 60°C e de 0,75mM a 1,50mM, respectivamente (Tabela 3).

Tabela 3 – Iniciadores e condições de amplificação dos microssatélites

Microssatélite	Senso (5' →. 3')	Anti-senso (5' →. 3')	Tm	[Mg ⁺²]
PvMS1	CTATCTGAGGAATGGGGA	ATTTACTATGACGAAGGTGA	53,4°C	1,50mM
PvMS2	CATCATTGGGTAAGTCGGG	GCAGCCACAAAATCAACACC	60,0°C	1,50mM
PvMS3	GGGAAGCACAAAATCGTA	TCAGCAGGGACAAAACG	60,0°C	1,50mM
PvMS4	TTATTTCCCCCTTTGCC	AAATGGATGTTCTTGTCAA	55,7°C	1,00mM
PvMS5	TGCTATTTGCTCGGTCTGT	GAGCGTTATCATCATTAG	56,0°C	1,50mM
PvMS6	ACACATTTGACACAGTTCC	ATGCCCTGGTCCCTACAA	58,6°C	1,50mM
PvMS7	GTATTCCCCGTCTTGTCC	CTTTCTCCGTTCTTATTTCT	56,0°C	1,50mM
PvMS8	TCCGTTGTTTTGTTGCC	CACTTGTTCCGTTCCGCTC	60,0°C	1,50mM
PvMS9	TGTGGATAAGGGGAAAAA	TTTTTTTCCTTGAGTTTACG	50,0°C	0,75mM
PvMS10	AAGTGTATTTTCCCGACG	CTTTTGCTTGCTCCGTTT	54,7°C	1,50mM
PvMS11	CGATGCGTTCACTTGGAT	TATTCTTCTCCCCTCGTG	54,0°C	0,75mM

Foi utilizado um volume de reação de 20 µL contendo 20 pmoles de cada iniciador (senso e antisenso), 0,125 mM de dNTP (desoxiribonucleotídeos), 2 µL tampão da Taq DNA polimerase 10X, 1 u de Taq DNA polimerase (Invitrogen).

Os ciclos para as reações de amplificação foram padronizados do seguinte modo: 94°C por 2 minutos, 35 ou 40 ciclos com 94°C por 30 segundos, 50°C a 60°C por 20 segundos (Tabela 3) e 72°C por 30 segundos. Após estes ciclos um passo de extensão final foi realizado a 72°C por 2 minutos.

4.5 Visualização da Amplificação em Gel de Agarose

Com o objetivo de verificar se o locus amplificado era polimórfico, após a reação de amplificação de algumas amostras escolhidas aleatoriamente das diferentes regiões estudadas, o produto amplificado era visualizado em gel de agarose 3% corado com brometo de etídio com concentração final de 0,5 µg/mL. Deste modo era possível identificar polimorfismos mesmo em microssatélites com repetições de dinucleotídeos (Figura 4).

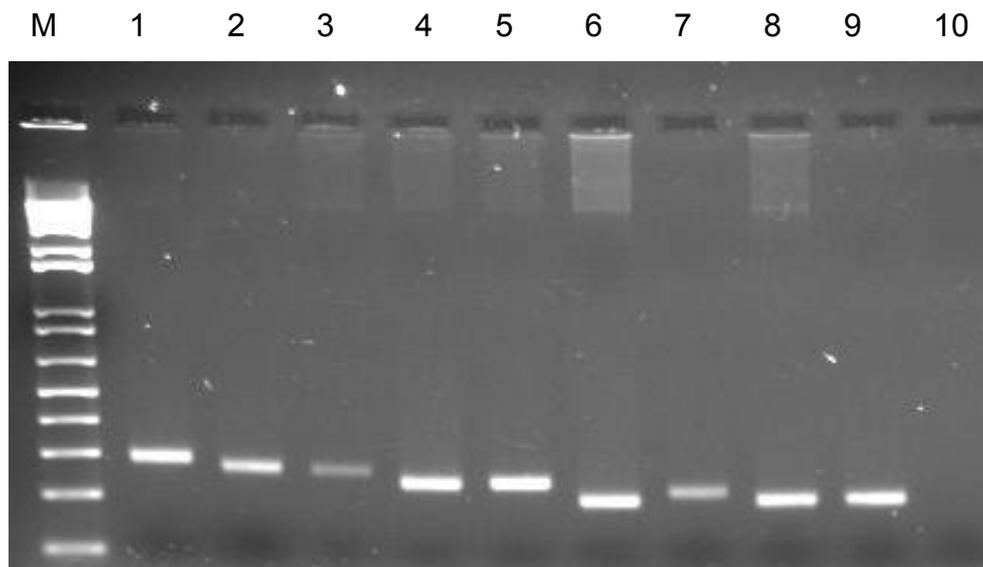


Figura 4 – Gel de agarose 3% corado pelo brometo de etídeo representando os resultados da amplificação do microssatélite PVMS6. Canaletas: M – marcador de peso molecular 100bp, 1 a 3 – isolados do *P. vivax* do Amapá, 4 a 6 – isolados do *P. vivax* do Pará, 7 a 9 – isolados do *P. vivax* de Rondônia e 10 – controle negativo da PCR (sem DNA).

Posteriormente à identificação dos loci polimórficos, os iniciadores senso eram adquiridos com fluoresceína na extremidade 5' para realização da genotipagem dos diversos isolados no seqüenciador automático.

4.6 Genotipagem no Seqüenciador Automático MegaBACE

Para realização da genotipagem dos microssatélites no seqüenciador automático de DNA foram utilizados, além dos produtos de PCR com iniciadores senso marcados com fluoresceína, placa de 96 poços do modelo “Ultraamp™ Skirted 96 PCR Plate” (marca Sorenson BioScience TNC), Tween 20 e um padrão de peso molecular, MegaBACE™ ET550-R Size Standard (GE Healthcare), com todas suas bandas marcadas com ET-ROX. Em cada poço da placa foram colocados 7,75 μ L de Tween 20 a 0,1%, 0,25 μ L do padrão de peso molecular mencionado acima e 2 μ L do produto amplificado. As amostras eram homogêneas na placa, logo em seguida a placa era centrifugada por um tempo suficiente para que as gotas na parede de cada poço baixassem. A

placa era então armazenada a 4°C até que fosse levada para o seqüenciador automático. Antes das amostras serem lidas no aparelho, elas eram desnaturadas a 94°C por 1 minuto.

Após a leitura das amostras, os resultados da eletroforese capilar foram analisados utilizando o programa MegaBACE Fragment Profiler versão 1.2, o qual possibilita a verificação dos picos na corrida e posterior geração dos tamanhos dos fragmentos previamente amplificados (Figura 5). Pelo fato do parasito ser haplóide na fase sanguínea do ciclo, a presença de dois picos numa mesma amostra, era uma evidência de infecção múltipla. Entretanto, o segundo pico só era considerado se este possuísse no mínimo um terço da altura do pico maior, o qual foi denominado predominante. Para as análises populacionais, somente o pico predominante foi utilizado na montagem do haplótipos, o segundo pico foi somente utilizado para cálculo do número total de alelos e da taxa de infecção múltipla.

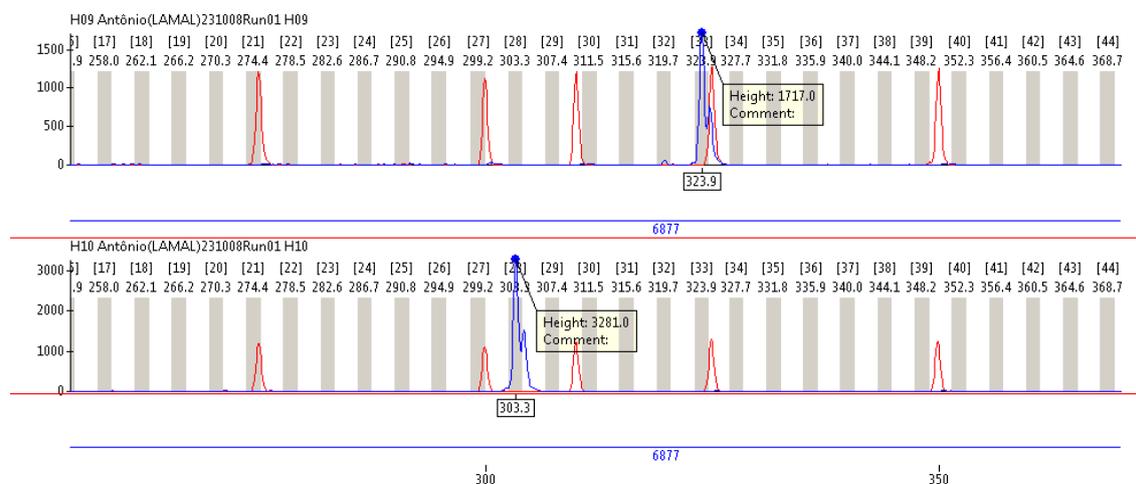


Figura 5 – Representação dos resultados da eletroforese capilar no seqüenciador automático MegaBACE. Os picos em vermelho representam o padrão de peso molecular e os picos em azul representam as amostras genotipadas. As barras em cinza representam os possíveis alelos para o marcador genotipado e os números abaixo dos picos azuis são os tamanhos dos alelos para as amostras acima genotipadas.

4.7 Amplificação dos tandem repeats (TRs)

A reação em cadeia da polimerase ou PCR foi padronizada usando o termociclador Eppendorf Gradient. As temperaturas de anelamento dos iniciadores e a concentração de cloreto de magnésio variaram de 50°C a 60°C e de 1,00mM a 3,00mM, respectivamente (Tabela 4).

Foi utilizado um volume de reação de 20 µL contendo 20 pmoles de cada iniciador (senso e antisenso), 0,125 mM de dNTP (deoxiribonucleotídeos), 2 µL tampão da Taq DNA polimerase 10X, 1 u de Taq DNA polimerase (Invitrogen).

Os ciclos para as reações de amplificação foram padronizados do seguinte modo: 94°C por 2 minutos, 35 ou 40 ciclos com 94°C por 30 segundos, 52.5°C a 60°C por 20 segundos (Tabela 4) e 72°C por 30 segundos. Após estes ciclos um passo de extensão final foi realizado a 72°C por 2 minutos.

Todos os loci de TRs foram localizados no cromossomo 8 do genoma de *P. vivax* utilizando a mesma metodologia para a localização dos microsatélites.

Tabela 4 – Características, iniciadores e condições de amplificação dos TRs

TR	Unidade de Repetição ^a	Senso (5' → 3')	Anti-senso (5' → 3')	Ta(°C) ^b	[Mg ⁺²] (mM)
MN2	(CCGGTGATGAGAA TCATT) ₂	AATGGTGACTACCAAAACGA	ATACACTTCGCTTTCTTCAC	60,0	3,00
MN7	(CATTTTG) ₅	CAGCGCTTTGCAATTTAAA	GAGTAGTCTTCGAAAATTG	52,5	1,50
MN21	(CCACT) ₅	TGCCACGATGAAAGGCCCTT	TTTGAGGTGGCACTGCTATA	60,0	1,00
MN23	(CACC) ₆	AGCACAAGGTGACTCAAAAA	AGCTGCGTTTTGATGGAGAA	58,0	1,50
MN25	(AAGTCAACATGGCA AAAAA) ₃	GGGGAAAAACAATGGCAA	AGGCAGTCGATTTGAAACT	55,0	1,50

^a Número de repetições de acordo com o trabalho de Feng et al. 2003 (15).

^b Ta: temperatura de anelamento.

4.8 Genotipagem no Seqüenciador Automático de DNA A.L.F. (Automatic Laser Fluorescence)

A genotipagem dos TRs foi realizada utilizando o seqüenciador A.L.F. (Amersham Pharmacia Biotech) e também produtos de PCR com iniciadores senso marcados com fluoresceína. Além disso, foram utilizados géis de poliacrilamida 6% desnaturante com 7 M de uréia, utilizando-se tampão de corrida TBE 0,6 x (1M de Tris, 0,83M de ácido bórico e 10 mM de EDTA). Foram aplicados no gel 2 μ L de cada amostra e 10 fmoles de um padrão de peso molecular fluorescente de 50pb, ALFexpress sizer 50-500 bp – concentração de 5 fmol/ μ l (Amersham Pharmacia) utilizado como referência de mobilidade dos fragmentos. Os resultados obtidos foram analisados com o programa Allele Link (versão 1.0, Amersham) (Figura 6).

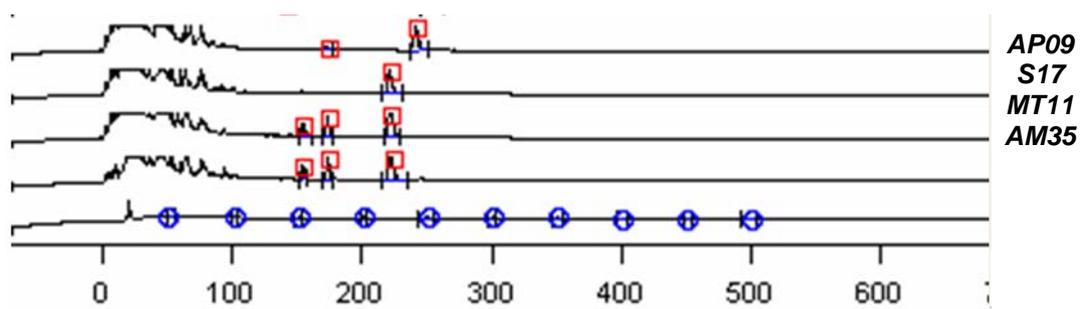


Figura 6 – Representação dos resultados da eletroforese realizada no seqüenciador automático A.L.F. A linha com círculos em azul representa a canaleta com o padrão de peso molecular. As linhas com picos e quadrados em vermelho representam as canaletas com as amostras genotipadas.

4.9 Análises Populacionais

4.9.1 Número de alelos por locus e por população

Fragmentos amplificados possuindo tamanhos diferentes encontrados na realização da genotipagem para um determinado locus foram considerados como sendo diferentes alelos. Para as amostras que apresentavam dois alelos em um mesmo locus, e o pico representando o segundo alelo satisfazia a condição de ter um terço da altura do pico maior, estes eram considerados na

contagem para o número total de alelos de um determinado locus. Foi também contabilizado o número de alelos encontrados para cada locus nas populações estudadas, sendo esta contagem realizada da mesma maneira como foi descrita logo acima.

4.9.2 Análise Molecular de Variância (AMOVA)

Foram utilizados diversos programas computacionais para realização das análises populacionais. O primeiro, que nós utilizamos, foi o programa Arlequin versão 3.1 (<http://lgb.unige.ch/arlequin/>), que é um conjunto amplo de métodos básicos e testes estatísticos com o objetivo de extrair informações sobre características genéticas e demográficas de amostras populacionais (33). O primeiro método utilizado foi o AMOVA, que é uma abordagem baseada na análise de variância das frequências gênicas que leva em consideração o número de mutações entre os haplótipos moleculares (manual do arlequin). Este método permitiu testar a estrutura de populações, por nós definida em 4 populações, de acordo com sua origem geográfica. O AMOVA também utiliza uma análise hierárquica de variância, a qual divide a variância total em componentes devido às diferenças intra-individuais, diferenças interindividuais e/ou diferenças inter-populacionais. Desta maneira, a variância molecular é a soma do componente de covariância devido às diferenças entre os haplótipos dentro da população, do componente de covariância devido às diferenças entre os haplótipos em diferentes populações dentro de um grupo, e do componente de covariância devido às diferenças entre os grupos de populações. Em nosso caso, como nós trabalhamos apenas com um grupo de populações, o AMOVA teve apenas dois componentes de covariância: um devido às diferenças haplotípicas entre as populações e o outro componente devido às diferenças haplotípicas dentro de cada população.

4.9.3 Índice de Fixação (F_{ST})

Dentre diversas abordagens para estimar a quantidade de diferenciação genética entre subpopulações, Wrigth (34, 35) desenvolveu uma abordagem, a qual particiona a variação genética em uma população subdivida. Este método

consiste de três diferentes coeficientes F , usados para alocar a variabilidade genética ao nível de população total (T), subpopulações (S), e indivíduos (I).

Assim, o F_{ST} é uma medida de diferenciação genética entre subpopulações e é sempre positivo. Os outros coeficientes, F_{IS} e F_{IT} , são medidas de desvio das proporções genotípicas em Equilíbrio de Hardy-Weinberg dentro de subpopulações e no total da população, respectivamente. Derivando a fórmula do F_{ST} a partir dos conceitos de F_{IS} e F_{IT} nós temos que:

$$F_{ST} = \frac{V(q)}{\bar{q}(1-\bar{q})}$$

Onde $V(q)$ é a variância na frequência do alelo q e \bar{q} é a frequência do alelo q na população total. Conseqüentemente se a variação das frequências alélicas entre duas subpopulações é zero, logo, o F_{ST} será zero, pois não existe diferença de frequência alélica entre as mesmas. Do mesmo modo se o F_{ST} tiver valor acima de zero significa que existe diferença nas frequências alélicas. Os valores deste índice podem variar de 0 a 1. Estas medidas de diferença nas frequências alélicas pode ser interpretada como uma medida da distância genética entre as subpopulações comparadas.

Em nosso estudo, nós utilizamos o F_{ST} para medir a distância genética entre as populações estudadas. Essas populações foram consideradas pelo Arlequin como sendo subpopulações de um mesmo grupo. Desta maneira, nós tínhamos somente uma população subdividida em quatro.

4.9.4 Heterozigosidade Esperada (H_E)

A mais ampla medida de variação genética em uma população é a quantidade de heterozigosidade. Devido indivíduos em espécies diplóides serem heterozigotos ou homozigotos para um dado locus, esta medida representa uma quantificação da variação genética biologicamente útil. Porém, o *P. vivax* na fase eritrocítica é um organismo haplóide, assim, não é possível observar a frequência de indivíduos heterozigotos, pois estes não existem neste momento do ciclo de vida do parasito. Desta maneira, nós somente calculamos a heterozigosidade esperada baseada na frequência dos alelos encontrados para cada loci nas populações estudadas e assumindo uma

condição de Equilíbrio de Hardy-Weinberg nas populações. Portanto, podemos interpretar o H_E como sendo uma medida referente à chance de em um conjunto de amostras da população, nós retirarmos dois indivíduos e eles juntos formarem um genótipo heterozigoto para o locus em estudo. Foi calculado o H_E para cada locus e para cada subpopulação.

4.9.5 Teste de Mantel

O teste de Mantel consiste em testar a significância da correlação entre duas ou mais matrizes através de um procedimento de permutação, permitindo gerar a distribuição nula empírica do coeficiente de correlação. Primeiramente o produto Z entre duas matrizes é calculado e posteriormente o produto Z^* é calculado a partir da permutação de colunas e linhas de uma das matrizes. Este novo produto é comparado ao original e destas comparações é testada a significância da correlação entre as matrizes.

A correlação entre o índice F_{ST} e a distância geográfica entre as populações estudadas foi testada através do método acima descrito. As matrizes usadas no teste foram: a matriz par-a-par do índice F_{ST} , calculada pelo programa Arlequin e uma matriz, em que cada ponto representava a distância geográfica em quilômetros entre duas populações. Desta maneira foi possível verificar se existe alguma correlação entre um índice de distância genética e as distâncias geográficas entre as populações estudadas. O teste de Mantel foi realizado também no programa Arlequin.

4.9.6 Desequilíbrio de Ligação

O equilíbrio de ligação é caracterizado pela independência estatística dos alelos em todos os loci estudados, isto é, pela associação aleatória dos alelos dos loci nas populações. Quando esta independência não é satisfeita nós dizemos que os loci estudados estão em desequilíbrio de ligação.

Para testar a hipótese nula de equilíbrio de ligação entre os loci por nós estudados, foi utilizado o programa LIAN versão 3.5 (LInkage ANalysis - <http://guanine.evolbio.mpg.de/cgi-bin/lian/lian.cgi.pl>), que testa esta hipótese nula utilizando dados multilocus. De maneira mais detalhada, o programa testa

a associação independente dos alelos primeiramente calculando o número de loci que possuem alelos diferentes para cada par de indivíduos analisados. Da distribuição dos valores dos loci com alelos diferentes uma variância (V_D) é calculada. Esta é comparada com a variância esperada para equilíbrio de ligação (V_e). Assim, a hipótese nula $H_0 : V_D = V_e$ é testada através de simulações computacionais. Utilizando os valores das variâncias é calculado um Índice Padronizado de Associação (I_A^s), que é uma medida de desequilíbrio geral para os loci estudados em uma determinada população. O índice acima é calculado da seguinte maneira:

$$I_A^s = \frac{1}{l-1} \left(\frac{V_D}{V_e} - 1 \right)$$

Onde l é o número de loci investigados.

Além do desequilíbrio de ligação global, foi também testado o desequilíbrio de ligação entre cada par de loci, para isto o programa Arlequin foi utilizado. Deste modo, os dados observados foram organizados em tabelas de contingência. As entradas $K_1 \times K_2$ da tabela são as freqüências haplotípicas observadas (valores absolutos), com K_1 e K_2 sendo o número de alelos no locus 1 e 2, respectivamente. O teste consiste em obter a probabilidade de encontrar uma tabela com os mesmo valores em conjunto de dados simulados e que tenha uma probabilidade igual ou menor do que a tabela observada. O P -valor do teste é calculado em razão da proporção de tabelas visitadas tendo uma probabilidade menor ou igual à tabela de contingência observada. Ao invés de visitar todas as possibilidades de tabela, cadeias Markov são utilizadas para explorar eficientemente todo espaço de tabelas possíveis. Neste estudo foram utilizados 1.000.000 passos da cadeia de Markov (33).

4.9.7 Análise de estrutura de população utilizando inferência Bayesiana

A inferência Bayesiana é baseada no teorema de Bayes e tem como objetivo maximizar a informação sobre determinada variável para reduzir a incerteza sobre a mesma. Desta maneira, toda informação é útil e deve ser aproveitada. Dado que o teorema de Bayes é apresentado na forma:

$$p(q|Y) \propto p(q) p(Y|q)$$

Em que $p(q)$ é a densidade de probabilidade *a priori* de q , a qual reflete o grau de conhecimento acumulado sobre os possíveis valores de q , antes da obtenção de informações contidas em Y ; $p(Y|q)$ é a função de verossimilhança de Y , que representa a contribuição de Y ao conhecimento de q ; e $p(q|Y)$ é a densidade de probabilidade *posterior* de q , a qual inclui o grau de conhecimento prévio sobre q “atualizado por informações adicionais” contidas em Y (36).

Neste estudo o $p(q)$ representa o número de K de populações utilizado como informação *a priori* para análise de estrutura de população. $p(Y|q)$ representa a probabilidade dos modelos evolutivos e o conjunto de dados utilizados *a priori* condicionada ao valor de K , isto é, a probabilidade de que os modelos evolutivos e o conjunto de dados sejam explicados pelo valor de K . E finalmente, $p(q|Y)$ representa a probabilidade do valor de K condicionada aos modelos evolutivos e o conjunto de dados, isto é, probabilidade de que o valor de K seja explicado pelos modelos evolutivos utilizados e pelo conjunto de dados.

Para utilizar a inferência Bayesiana nas análises de estrutura de populações foi utilizado o programa chamado Structure versão 2.1 (<http://pritch.bsd.uchicago.edu/structure.html>), o qual assume um número K de populações, sendo que cada uma é caracterizada por um conjunto de frequências alélicas em cada locus. O método utilizado no programa tenta prever a população de cada indivíduo com base no seu haplótipo, enquanto simultaneamente estima a frequência alélica populacional. A abordagem utilizada pelo programa também permite identificar a presença de indivíduos miscigenados nas amostras, os quais possuem uma constituição genética representada por mais de uma população do número total de K populações (37).

Para utilizar o programa alguns modelos foram utilizados como informações *a priori*. Foi utilizado como modelo de ancestralidade o modelo de miscigenação, que permite que indivíduos possam ter constituição genética mista. Assim é possível saber a proporção de miscigenação de cada um dos indivíduo. Como modelo de frequência alélica, foi utilizado o modelo de frequência alélica correlacionada, pois este modelo assume que populações próximas possuem frequências alélicas similares, desta maneira o programa se

torna mais sensível, melhorando as inferências em casos em que as diferentes populações divergem pouco (38).

Os valores utilizados de K variaram de 2 a 8 e para cada valor o programa foi rodado três vezes. Cada rodada foi realizada com 1.000.000 de passos na cadeia de Markov e 100.000 passos de burnin, isto é, passos da cadeia que são descartados. Para cada valor de K um logaritmo da probabilidade de K condicionada a X (modelos e dados utilizados) foi calculado. Este logaritmo foi transformado em uma probabilidade posterior, em que os valores variam de 0 a 1, e quanto maior o valor, maior a probabilidade do K correlacionado com o maior valor de probabilidade posterior calculado.

Para representação dos resultados desta análise, um gráfico de pizza foi construído. Cada padrão de cores no gráfico representa a constituição genética de cada população estudada. Além disso, as cores estão representando as populações ancestrais e elas são distribuídas de acordo com constituição genética média da população, que foi calculada com base na constituição genética dos indivíduos da população. Estes valores são inferidos e apresentados pelo programa Structure.

4.9.8 “Bottleneck” ou Gargalo de Garrafa populacional

Para verificar se as populações analisadas haviam passado por algum tipo de redução no número de indivíduos, como gargalo de garrafa ou efeito fundador, nós utilizamos o programa Bottleneck versão 1.2.02 (<http://www1.montpellier.inra.fr/URLB/bottleneck/bottleneck.html>). Este programa é baseado na suposição de que populações que passaram por uma recente redução do seu tamanho populacional efetivo exibem uma redução do número de alelos e da heterozigidade em loci polimórficos. Porém, a diversidade alélica é reduzida de maneira mais rápida do que a heterozigidade. Assim, a heterozigidade observada é maior do que a esperada dado um número de alelos e um número de loci sob equilíbrio entre mutação e deriva.

Estritamente falando, o excesso de diversidade genética (heterozigidade observada maior que a esperada) tem sido demonstrado somente para loci evoluindo sobre o modelo de mutação IAM (Infinite Allele

Model) por Maruyama e Fuerst (1985). Se o locus evolui estritamente sobre o modelo SMM (Stepwise Mutation Model), podem existir situações em que este excesso de diversidade não seja observado. Contudo, somente alguns loci seguem estritamente este modelo, e se eles desviam uma pequena porcentagem do modelo em direção ao modelo IAM, eles exibirão um excesso de diversidade genética como consequência de um evento de gargalo de garrafa. Como somente alguns poucos microssatélites seguem estritamente o modelo SMM, nós utilizamos também na análise o modelo TPM (Two-phased model of mutation). Este modelo é intermediário aos modelos SMM e IAM.

O programa também utiliza três testes estatísticos para determinar se a população exibe um significativo número de loci com excesso de diversidade genética. O primeiro denominado teste sinal, o segundo teste de diferença padronizado e o último teste de Wilcoxon. O primeiro possui baixo poder estatístico. O segundo teste não é muito útil, uma vez que ele demanda no mínimo 20 loci polimórficos. O teste de Wilcoxon fornece relativamente um alto poder estatístico e ele pode ser usado com até quatro loci polimórficos e qualquer número de indivíduos (de 15 a 40 indivíduos e de 10 a 15 loci polimórficos é recomendado para alcançar o um alto poder estatístico).

5 Artigos

5.1 Primeiro Artigo

Carta de aceite do artigo "Analysis of genetic variability of Plasmodium vivax isolates from different Brazilian Amazon areas using tandem repeats" na revista American Journal of Tropical Medicine and Hygiene.

----- Mensagem Original -----

 Assunto: American Journal of Tropical Medicine & Hygiene - AJTMH-08-0362.R1
 De: ajtmh@hotmail.com
 Data: Ter, Dezembro 2, 2008 1:27 pm
 Para: cristiana@cpqrr.fiocruz.br

02-Dec-2008
 To: Mrs. Cristiana Brito
 RE: AJTMH-08-0362.R1, Analysis of genetic variability of Plasmodium vivax isolates from different Brazilian Amazon areas using tandem repeats by Rezende, Antonio; Tarazona-Santos, Eduardo; Couto, Alvaro; Fontes, Cór Jesús; Souza, José Maria; Carvalho, Luzia; Brito, Cristiana

Dear Dr. Brito:

On behalf of Dr. Joseph Vinetz I would like to thank you for submitting your manuscript to the American Journal of Tropical Medicine & Hygiene. Your manuscript has been accepted for publication.

Please be sure that you have sent in your signed copyright form, which can be located on the journal's submission site (<http://mc.manuscriptcentral.com/ajtmh>) under the link to "Instructions and Forms" and that all authors have sent in a signed statement indicating participation in the study and acceptance of the original submission and subsequent revisions submitted by you. We will NOT send your manuscript to press until these items have been received and will let you know if any of these items are missing from your file.

We will contact you if questions arise during the copyediting process. Otherwise, you can expect to receive your galley proofs by email within the next six to eight weeks. Please be sure to return your corrected galley proofs, page charge form, and reprint order form (if you wish to purchase reprints) within 48 hours of receiving your galley proofs.

Sincerely,
 Laura Buckley
 Editorial Assistant
 American Journal of Tropical Medicine and Hygiene
ajtmh@hotmail.com

--
 Centro de Pesquisas Rene Rachou/CPqRR - A FIOCRUZ em Minas Gerais.
 Rene Rachou Research Center/CPqRR - The Oswaldo Cruz Foundation in the State of Minas Gerais-Brazil.

Analysis of genetic variability of *Plasmodium vivax* isolates from different Brazilian Amazon areas using tandem repeats

ANTÔNIO M. REZENDE, EDUARDO TARAZONA-SANTOS, ÁLVARO D'ALMEIDA COUTO, CÓR JÉSUS F. FONTES, JOSÉ MARIA DE SOUZA, LUZIA H. CARVALHO AND CRISTIANA F. A. BRITO

Laboratory of Malaria, Centro de Pesquisas René Rachow/Fiocruz, Belo Horizonte, Brazil

Abstract. Few genetic markers have been described to analyze populations of *Plasmodium vivax*. The genetic variability of *P. vivax* has been analyzed mainly among isolates taken from areas ranging from hyper- to holoendemic areas. These studies of genetic variability have neglected many areas with different epidemiological profiles. The purpose of this study was to analyze the genetic variability of *P. vivax* isolates from four different Brazilian Amazon areas. We chose to study the five most polymorphic tandem repeats (TRs) identified so far. All TRs studied were polymorphic in at least one studied population, with a modal allele at nearly all loci. Expected heterozygosity ranged from 0.462 - 0.666 and did not correlate with the repeat array length. The genetic distances among the populations varied from 0.027 to 0.241, and did not correlate with their geographic separation. TRs identified in *P. vivax* isolates failed to allow geographic clustering.

INTRODUCTION

Malaria is a public health problem in over 100 countries; it causes at least 300 million cases of acute illness, and it kills more than one million people every year. *Plasmodium vivax* is the most widespread malaria species, and up to 40% of the world population is at risk of infection from this species.¹ Currently, malaria caused by *P. vivax* is endemic in many countries of Asia, the South Pacific, North Africa, the Middle East, and South and Central America. In Brazil, 540,000 new cases of malaria were

reported in the year 2006. Nearly all of them (99.5%) occurred in the Amazon states, where *P. vivax* was responsible for more than 70% of cases.² Despite a lower mortality than *Plasmodium falciparum* infection, *P. vivax* is responsible for intense morbidity and significant economic losses.³ Parasite control strategies depend on an understanding of the genetic variability and population structure of the parasite and the dynamics of disease transmission.⁴ Moreover, studies of genetic variability, particularly of field isolates collected directly from patients, can provide markers to predict drug resistance, virulence, and periodic relapse.⁵ Several studies have assessed the genetic variability, population structure, and evolution of *P. falciparum*. In contrast, the lower virulence of *P. vivax* and the difficulty of propagating it continuously in culture means that less is known about its genetic variability.⁶ Furthermore, few studies have analyzed the genetic diversity among *P. vivax* populations, which could lead to inferences about its population dynamics and patterns of gene flow.

Molecular markers for population studies of *P. vivax* have mostly been based on orthologs of *P. falciparum* antigen genes, such as merozoite surface antigen, erythrocyte-binding proteins, and circumsporozoite protein, which are under selection by host immunity.^{7,8} However, some population parameters are better inferred from neutral or nearly-neutral markers, including microsatellite loci.⁹ In contrast with the approximately 1,000 polymorphic microsatellites described for *P. falciparum*, only 40 or so have been described for *P. vivax*.¹⁰⁻¹⁵ Microsatellite variability in *P. vivax* is strongly dependent on the repeat array length: short arrays exhibit little variation and long arrays are highly polymorphic.^{11,12} When repeat array length is taken into account, *P. vivax* microsatellites show levels of variation comparable to those in *P. falciparum*.¹² However, the geographical clustering of *P. vivax* isolates was not detected using

microsatellite loci.¹⁶ This may be due to high within-population heterogeneity or to the complex phylogeny of *P. vivax*.¹⁶

On the other hand, Feng and colleagues¹⁵ described 33 polymorphic tandem repeats (TRs) or minisatellites containing repeating units of five nucleotides or more. Leclerc and colleagues¹¹ showed a limited variability among isolates from different areas using microsatellites and TR loci. TR loci tend to be less variable than microsatellite loci and, for this reason, they can allow for geographic clustering of isolates. In addition, PCR-amplified fragments of TRs may be separated using simple agarose gel electrophoresis and easily analyzed because of their significant variation in the number of repeats.

Thus, the aim of this study was to analyze the variability of *P. vivax* isolates from different endemic areas in the Amazon using the five most polymorphic TRs employed by Leclerc.¹¹ Our specific objectives were to assess the genetic diversity, structure, and differentiation of *P. vivax* populations in relation to geographic distances between these populations and the level of malaria transmission in different areas of Brazilian Amazon. We report the results of the genetic analysis and discuss the evolutionary forces that may explain the observed patterns of diversity.

MATERIALS AND METHODS

***Plasmodium vivax* isolates.** Blood samples were collected from 44 *P. vivax*-infected patients (11 individuals from each area) in different areas of the Brazilian Amazon (Figure 1): Macapá in Amapá State (AP; 00°02'20"S, 51°03'59"W), Cuiabá in Mato Grosso State (MT; 15°35'45"S, 56°05'49"W), Augusto Correa in Pará State (PA; 01°01'18"S, 46°38'06"W), and Manaus in Amazonas State (AM; 03°06'07"S, 60°01'30"W). The samples were collected from the different regions as follows: AM, May 2003; AP, November 2004; PA, October 2005; and MT, from July 2003 to July

2004. The available epidemiological data indicated that the studied regions were areas of hypo- to mesoendemic malaria with intermittent transmission.¹⁷ The rate of transmission for each area was measured by the Annual Parasite Index (API), which reflects the number of positive blood smears per 1,000 inhabitants. By following the guidelines of the Health Surveillance Secretariat of the Ministry of Health¹⁸, the API index was used to stratify the risk of malaria infection in the different areas at the time of blood collection: high risk (API >50), Augusto Correa (PA); medium risk (API 10–50), Manaus (AM); and low risk (API < 10), Macapá (AP). Patients who came to the health center in Cuiabá (MT) reported many locations of infection with different API, so the area was classified as variable risk. Patient infection was confirmed by microscopic analysis on Giemsa-stained blood smears. All patients were between 16 and 58 years old (mean, 32 years).

Genomic DNA isolation and PCR amplification. DNA was extracted from whole blood samples using the PUREGENE DNA isolation Kit (Gentra Systems, Minneapolis, MN) according to the manufacturer's protocol. The five loci (MN2, MN7, MN21, MN23, and MN25¹⁵) were amplified using the primers and conditions described by Leclerc,¹¹ but forward primers were labeled with fluorescent dye (FAM). The repeat unit ranged from 4 (MN23) to 20 nucleotides (MN25) (Table 1). These markers were assigned to chromosome 8 of the *Plasmodium vivax* SaI-1 (unfinished sequence, whole genome shotgun sequencing project, GenBank accession number AAKM00000000). The markers were separated from each other at a distance of 1,727 bp (Mn21/Mn23) to 32,934 bp (Mn2/Mn7) (Table 1). PCR products were separated using agarose gel electrophoresis and the exact size of PCR products was identified using the ALF automatic DNA sequencer (Amersham Pharmacia Biotech), comparing to the

ALFexpress sizer 50-500 bp (5 fmol/ μ l; Amersham), and analyzed using Allele Link software (version 1.0, Amersham).

Variability analysis. We measured allele frequencies using the single or predominant allele at each locus from each isolate; non-predominant alleles were recorded and used to estimate multiplicity of infection, since all markers are single-copy loci and blood stage malaria parasites are haploid. The highest peak in the electropherogram traces was defined as the predominant allele and electropherogram traces with at least one-third the height of the predominant allele were defined as additional alleles in a multiple-clone infection. We measured intra-population diversity using the statistic H_E , which is the average probability that two alleles randomly obtained for each locus are different. AMOVA was performed to measure whether the resulting variability could be explained by variations among and within populations. We performed a Kruskal-Wallis test to verify whether the differences between the H_E of each locus for individual populations were significant.

We assessed the linkage disequilibrium for each locus pair and each population using the Fisher Exact test of linkage disequilibrium available in Arlequin 3.1¹⁹ software with 100,000 steps in the Markov Chain and 1000 dememorization steps. We also used LIAN 3.5 software²⁰ to test for evidence of multilocus linkage disequilibrium in each population. This test compares the variance (V_D) of the number D of alleles shared between all pairs of haplotypes observed in the population (D) with the variance expected when alleles are assigned randomly (V_E) as follows: $I_A^S = (V_D / V_E - 1)(r - 1)$, where r is the number of loci analyzed. V_E is derived from 100,000 simulated data sets in which alleles were randomly reshuffled among haplotypes. Significant linkage disequilibrium is detected if V_D is 95% of the values derived from the re-shuffled data sets. Two analyses were done: one considering all haplotypes and another with unique

haplotypes in order to remove the effect of the epidemic expansion of a particular haplotype on linkage disequilibrium in randomly mating populations.²¹

Between-population genetic diversity was quantified using pairwise F_{ST} , as implemented in Arlequin 3.1.¹⁹ We also estimated the correlation among the matrices of pairwise genetic and geographic distances using the Mantel test.²²

We attempted to assess the population structure of samples using Structure 2.1 software²³. This uses Bayesian approaches to infer the most likely number of populations or clusters (K) represented in the total sample, and then measure the probability that individual parasites come from each of these K populations (1). The parameters set to run the software were an admixture model and linkage model that assumed dependent allele frequencies. For each K value, we ran the software three times, each one with 100,000 burnins and 1,000,000 iterations. K values ranged from 2 to 8.

RESULTS

Five TR loci were selected to analyze 44 isolates of *P. vivax* taken from four different endemic areas in the Brazilian Amazon; their characteristics are shown in Table 1. The TR markers used in this study map to chromosome 8, and they include non-coding loci (MN7, MN21, and MN25), a hypothetical protein locus (MN23), and a protein containing a kinase domain locus (MN2). The number of alleles per locus ranged from 2 to 9, with a modal allele in most cases (frequency ≥ 0.5) (Table 2). For all the TR loci studied, isolates were polymorphic in at least one population, and only MN2 was monomorphic in isolates from the PA population. We used a Kruskal-Wallis test to verify whether the difference between the heterozygosity expected (H_E) for each locus was significant (Table 3). We noted no difference between the loci, indicating that

the length of the repeating unit of each locus did not correlate with its variability. Multiplicity of infection was estimated at 27% and 66% in the MT and AM populations, respectively. However, we did not find any individual infected with more than one haplotype in the PA and AP populations.

The comparative analysis of the four Amazon populations studied showed that all of them were highly variable, with H_E ranging from 0.46182 to 0.66600. In addition, when AMOVA was performed, we found that approximately 84% of the diversity detected was due to variation within each population. Only 16% of the diversity could be explained by variation among populations.

The F_{ST} values between populations (Table 3) were between 0.02 and 0.24, and they did not correlate with geographical distance between populations (Mantel Test, $Z = 0.031$, $P = 0.295$). The AM and AP populations could not be differentiated, and all of the populations diverged significantly ($P \leq 0.05$) from the PA population. This F_{ST} distribution and the derived estimations for the number of effective migrants ($M = Nm$) suggest that gene flow is higher among the AM and AP populations.

Linkage disequilibrium (LD) was estimated using the Fisher test. The number of steps in the Markov Chain used in the analysis allowed that the results converged. We found LD to vary as a function of population and loci pairs, but these results should be interpreted with caution given our small samples sizes. Analysis of LD for the loci pairs in all isolates showed that six of 10 possible pair combinations showed high LD, and did not correlate with the physical distance between the marker loci (Figure 2). Presence of marker loci in LD varied among populations. In the AP population, the loci pairs MN7/MN21, MN2/MN23, and MN21/MN25 showed high LD; in the MT population, the loci pairs MN7/MN23 and MN2/MN7 showed high LD; in the AM population, only the loci pair MN7/MN25 showed LD. LD was not found in the PA population. When

LD analysis was performed with AP and AM as one population, because they both had low F_{ST} values, we noted only the locus pair MN7/MN21 in LD. The AP and AM populations showed significantly higher multilocus LD. However, when the analysis was carried out only on the unique haplotype infections, only the AP population showed significant LD (Table 4).

It was not possible to cluster the samples using Structure 2.1²³ (data not shown). This was probably due to the limited number of markers used and the linkage disequilibrium between the markers. Consequently, little independent information was available for the software to use for clustering.

DISCUSSION

All molecular markers that were previously reported to be polymorphic in populations from different countries (Azerbaijan, Ethiopia, Turkey, Venezuela, and Thailand)¹¹ have also turned out to be polymorphic in Brazilian Amazonic populations. As reported by Leclerc¹¹, we observed a predominant allele at almost all loci, yet at only one locus was the same allele prevalent in all geographic areas studied. We detected in our isolates 68% of the alleles previously described¹¹.

Our data showed a high variability within populations, with high values of H_E . Similar values of H_E were also reported using microsatellite analysis of *P. falciparum* isolates from different regions of Brazil.²⁴ However, a study by Ferreira and colleagues analyzing a multilocus microsatellite in *P. vivax* isolates from one Brazilian population in Acre State reported slightly higher H_E values than ours.²¹ This population showed high percentage of multiple-clone infections, whereas only one population in our study showed a high multiplicity of infection: AM population, 66%. All of our populations showed a similar level of within-population variability. The populations studies here

were found to be structured with minimal gene flow, based on F_{ST} values. Our results corroborate the findings of multilocus microsatellite analysis in *P. falciparum* populations in Brazil.²⁴ Moreover, our study and that of Machado and colleagues²⁴ did not show a correlation between genetic and geographic distances for populations of *P. vivax* or *P. falciparum*, respectively.

Despite the fact that all the molecular markers map to the same chromosome, we did not observe a correlation between physical distance and linkage disequilibrium. For instance, the loci MN21 and MN23 lie less than 2,000 bp away from each other, but they did not show significant linkage disequilibrium. When we performed multilocus linkage disequilibrium analysis, we observed a correlation between the I_A^S value and number of loci pairs in linkage disequilibrium in each population. If we excluded the MT population, the correlation was $r = 0.9425$. The MT population was excluded because their isolates were obtained from individuals infected in many localities. Moreover, the blood samples in this group were collected over the longest period (one year). These two factors may increase the variability in these populations. Interestingly, we did not find a negative correlation between multiplicity of infection and linkage disequilibrium in populations, contrary to what has been observed with *P. falciparum*. This result is similar to that of Ferreira *et al* (2006), and it may indicate that these markers are under some selection pressure, which would alter the pattern of linkage disequilibrium.²¹ Nevertheless, according to the Ewens-Watterson neutrality test, the loci used in this work do not suffer from selection pressure. Indeed, none of the TR markers used here map to known surface antigens likely to be under immune-mediated diversifying natural selection. A higher variability was observed at loci MN7, MN21, and MN25, which lie in non-coding regions (Table 1). However, these molecular markers may not be considered to be strictly neutral, because of their physical proximity

and linkage disequilibrium with genes encoding proteins under natural selective pressure.

In *P. falciparum*, the genetic structure of the population is strongly linked to patterns of transmission.²⁵ A widely accepted model is that high levels of single infection, inbreeding, and rare recombination characterize low transmission regions with low genetic variability. In contrast, frequent mosquito inoculation, multiple infections, frequent outbreeding, and extensive recombination characterize high transmission regions with high genetic variability.¹³ A similar result was obtained in *P. falciparum* from the Amazon region, which showed a negative correlation between linkage disequilibrium and multiplicity of infection.²⁴ This correlation between genetic variability and patterns of transmission was less clear in *P. vivax* infections. For South America, Laos, and Thailand, the isolates fit the model well, as demonstrated by Anderson et al.^{13, 25} However, Indian isolates were found to have a low linkage disequilibrium and low frequency of multiple-clone infections.¹³ Conversely, Ferreira and colleagues showed a high LD and high frequency of multiple-clone infection in their study of Brazilian isolates from Acre State.²¹ Here, we observed high LD in the absence of multiple infections only in one of our populations.

According to Inwong and colleagues¹² and Russell and colleagues,⁵ the repeat array length of microsatellites is directly associated with genetic variability. The fact that we did not observe this correlation for TRs suggests that the relationship is too weak to be detected by our small number of TRs, or it may not exist for these particular markers. These molecular markers were shown to be highly polymorphic, though slightly less so than the microsatellites that have already been described.^{13, 26} In addition, this approach allows easy analysis based only on agarose gel electrophoresis, which suggest that high-throughput processing is possible using an automatic DNA sequencer. However, there is

no evidence for a putative mutation mechanism to explain the variability at these markers and this makes it difficult to understand how TR diversity is maintained. Therefore, further research on this issue is necessary.

Acknowledgments

This work was supported by the Fundação de Amparo à Pesquisa de Minas Gerais (Fapemig), the Conselho Nacional de Pesquisa (CNPq), and PAPES/FIOCRUZ.

Authors' addresses: Antônio M. Rezende, Luzia H. Carvalho, and Cristiana F. A. Brito, Laboratory of Malaria, Centro de Pesquisa Rene Rachou /Fiocruz, Barro Preto, Belo Horizonte, MG, Brazil, 30190 002, Telephone: +55 31 33497772, fax: +55 31 32953115. Eduardo Tarazona-Santos, Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. Álvaro D'Almeida Couto, Faculdade SEAMA, Macapá, Brazil. Cór Jesús F. Fontes, Departamento de Clínica Médica, Universidade Federal do Mato Grosso, Cuiabá, Brazil. José Maria de Souza Instituto Evandro Chagas, Belém, Brazil.

Reprint requests: Cristiana F. A. Brito, Laboratory of Malaria, Centro de Pesquisa Rene Rachou /Fiocruz, Av. Augusto de Lima 1715, Barro Preto, Belo Horizonte, MG, Brazil, 30190 002, Telephone: +55 31 33497772, fax: +55 31 32953115, E-mail: cristiana@cpqrr.fiocruz.br

REFERENCES

1. Price R, Tjitra E, Guerra C, Yeung S, White N, Anstey N, 2007. Vivax malaria: neglected and not benign. *Am J Trop Med Hyg* 77: 79-87.
2. Health HMoBSovi, 2007. Malaria Epidemiologic Situation in Brazil.
3. Mendis K, Sina B, Marchesini P, Carter R, 2001. The neglected burden of *Plasmodium vivax* malaria. *Am J Trop Med Hyg* 64: 97-106.
4. Carlton J, 2003. The *Plasmodium vivax* genome sequencing project. *Trends Parasitol* 19: 227-31.
5. Russell B, Suwanarusk R, Lek-Uthai U, 2006. *Plasmodium vivax* genetic diversity: microsatellite length matters. *Trends Parasitol* 22: 399-401.
6. Golenda C, Li J, Rosenberg R, 1997. Continuous in vitro propagation of the malaria parasite *Plasmodium vivax*. *Proc Natl Acad Sci U S A* 94: 6786-91.
7. del Portillo H, Longacre S, Khouri E, David P, 1991. Primary structure of the merozoite surface antigen 1 of *Plasmodium vivax* reveals sequences conserved between different *Plasmodium* species. *Proc Natl Acad Sci U S A* 88: 4030-4.
8. Bruce M, Galinski M, Barnwell J, Snounou G, Day K, 1999. Polymorphism at the merozoite surface protein-3 α locus of *Plasmodium vivax*: global and local diversity. *Am J Trop Med Hyg* 61: 518-25.
9. Jarne P, David P, Viard F, 1998. Microsatellites, transposable elements and the X chromosome. *Mol Biol Evol* 15: 28-34.
10. Gomez J, McNamara D, Bockarie M, Baird J, Carlton J, Zimmerman P, 2003. Identification of a polymorphic *Plasmodium vivax* microsatellite marker. *Am J Trop Med Hyg* 69: 377-9.
11. Leclerc M, Durand P, Gauthier C, Patot S, Billotte N, Menegon M, Severini C, Ayala F, Renaud F, 2004. Meager genetic variability of the human malaria agent *Plasmodium vivax*. *Proc Natl Acad Sci U S A* 101: 14455-60.
12. Imwong M, Sudimack D, Pukrittayakamee S, Osorio L, Carlton J, Day N, White N, Anderson T, 2006. Microsatellite variation, repeat array length, and population history of *Plasmodium vivax*. *Mol Biol Evol* 23: 1016-8.
13. Imwong M, Nair S, Pukrittayakamee S, Sudimack D, Williams J, Mayxay M, Newton P, Kim J, Nandy A, Osorio L, Carlton J, White N, Day N, Anderson T, 2007. Contrasting genetic structure in *Plasmodium vivax* populations from Asia and South America. *Int J Parasitol* 37: 1013-22.
14. Karunaweera ND, Ferreira MU, Hartl DL, Wirth DF, 2007. Fourteen polymorphic microsatellite DNA markers for the human malaria parasite *Plasmodium vivax*. *Mol. Ecol. Notes* 7: 172-175.
15. Feng X, Carlton J, Joy D, Mu J, Furuya T, Suh B, Wang Y, Barnwell J, Su X, 2003. Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. *Proc Natl Acad Sci U S A* 100: 8502-7.
16. Karunaweera N, Ferreira M, Munasinghe A, Barnwell J, Collins W, King C, Kawamoto F, Hartl D, Wirth D, 2008. Extensive microsatellite diversity in the human malaria parasite *Plasmodium vivax*. *Gene* 410: 105-12.
17. Camargo L, Ferreira M, Krieger H, De Camargo E, Da Silva L, 1994. Unstable hypoendemic malaria in Rondonia (western Amazon region, Brazil): epidemic

- outbreaks and work-associated incidence in an agro-industrial rural settlement. *Am J Trop Med Hyg* 51: 16-25.
18. Health HMoBSovi, 2005. Malaria Epidemiologic Situation in Brazil.
 19. Excoffier L, Laval G, Schneider S, 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolut. Bioinf. Online*, 47-50.
 20. Haubold B, Hudson R, 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Linkage Analysis. Bioinformatics* 16: 847-8.
 21. Ferreira M, Karunaweera N, da Silva-Nunes M, da Silva N, Wirth D, Hartl D, 2007. Population structure and transmission dynamics of *Plasmodium vivax* in rural Amazonia. *J Infect Dis* 195: 1218-26.
 22. Mantel N, 1967. Detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209-&.
 23. Pritchard J, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-59.
 24. Machado R, Povoá M, Calvosa V, Ferreira M, Rossit A, dos Santos E, Conway D, 2004. Genetic structure of *Plasmodium falciparum* populations in the Brazilian Amazon region. *J Infect Dis* 190: 1547-55.
 25. Anderson T, Haubold B, Williams J, Estrada-Franco J, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, Whitworth J, Velez I, Brockman A, Nosten F, Ferreira M, Day K, 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol* 17: 1467-82.
 26. Leclerc M, Gauthier C, Villegas L, Urdaneta L, 2005. Genetic diversity of merozoite surface protein-1 gene of *Plasmodium vivax* isolates in mining villages of Venezuela (Bolívar State). *Acta Trop* 95: 26-32.

Legends

Figure 1. Map of Brazil showing the four locations where blood was collected: Macapá (Amapá State, AP), Manaus (Amazonas State, AM), Cuiabá (Mato Grosso State, MT), and Augusto Correa (Pará state, PA).

Figure 2. Patterns of linkage disequilibrium between pairs of TR markers in four populations of *Plasmodium vivax* from the Brazilian Amazon. Shading denotes significant LD at the 5% level.

TABLE 1 Description of tandem repeats used to analyze variability in *P. vivax* isolates

Name	Repeat unit	First position*	Gene ID [†]	Description of coding gene or flanking gene (genbank accession number)
MN2	(CCGGTGATG AGAATCATT)	1625109	Coding	A protein containing a kinase domain (5472301)
MN7	(CATTTTG)	1591966	Non-coding	1683 bp at 5' side: hypothetical protein (5472308) 1224 bp at 3' side: hypothetical protein (5472307)
MN21	(CCACT)	1567132	Non-coding	358 bp at 5' side: hypothetical protein (5472314) 739 bp at 3' side: dihydrolipoamide acyltransferase, putative (5472313)
MN23	(CACC)	1565219	Coding	hypothetical protein (5472314)
MN25	(AAGTCAACA TGGCAAAAA AA)	1556283	Non-coding	1021 bp at 5' side: hypothetical protein (5472317) 2408 bp at 3' side: membrane skeletal protein, putative (5472316)

*First position at chromosome 8 (GenBank accession number NC 009913.1).

[†]Top hit obtained in a BlastN search of the *P. vivax* Sal-1 database at NCBI genomes (<http://blast.ncbi.nlm.nih.gov/>)

TABLE 2
Alleles detected at five tandem repeats (TRs) among *Plasmodium vivax* isolates from different endemic areas in Brazil

Locus*	Allele [†]	Allele number (frequency)			
		PA [‡]	MT	AP	AM
MN2	133		2 (22.2)	1 (9.1)	2 (25)
	152	11 (100)	6 (77.8)	10 (90.9)	3 (37.5)
	171				1 (12.5)
	189				2 (25)
MN7	209	6 (60)	5 (45.5)	9 (81.8)	7 (70)
	220	4 (40)	6 (54.7)	2 (18.2)	3 (30)
MN21	254	2 (18.2)	7 (58.3)	6 (54.5)	5 (35.7)
	258	8 (72.7)	3 (25.0)	2 (18.2)	1 (7.1)
	266		1 (8.4)		2 (14.3)
	270	1 (8.4)			
MN23	278				3 (21.4)
	290		1 (8.4)	3 (27.2)	3 (21.4)
	219		1 (8.4)	5 (45.5)	8 (53.3)
	223	2 (18.2)	4 (33.3)	3 (27.2)	
	237			3 (27.2)	1 (6.7)
	240		1 (8.4)		6 (40.0)
	247		5 (41.6)		
	251		1 (8.4)		
	278	5 (45.5)			
	282	3 (27.2)			
302	1 (9.1)				
MN25	169	8 (72.7)		1 (9.1)	1 (11.1)
	191			1 (9.1)	
	212				1 (11.1)
	220	1 (9.1)			
	234		3 (27.3)	2 (18.2)	4 (44.4)
	241	1 (9.1)	2 (18.2)		
	256	1 (9.1)	6 (54.5)	4 (36.4)	1 (11.1)
	261				2 (22.2)
	276			3 (27.2)	

* Locus named according to Leclerc¹¹

[†] Mean size of amplicons determined using Allele Link software (version 1.0, Amersham).

[‡] Geographic regions: PA, Para State; MT, Mato Grosso State; AP, Amapá State; AM, Amazonas State.

TABLE 3
Expected heterozygosity (H_E) for the TR loci and F_{ST} among *P. vivax* isolates from different endemic areas in Brazil

	H_E^\dagger					Mean in population	$F_{ST} (X100)^\dagger$		
	MN2	MN7	MN21	MN23	MN25		PA	MT	AP
PA	0	0.51	0.56	0.74	0.49	0.46 (0.248)			
MT	0.47	0.54	0.56	0.71	0.69	0.59 (0.092)	20.355*		
AP	0.18	0.33	0.65	0.71	0.82	0.54 (0.242)	19.697*	9.373	
AM	0.82	0.47	0.84	0.51	0.69	0.67 (0.153)	24.160*	16.175	2.692
Mean at locus	0.37 (0.310)	0.46 (0.083)	0.65 (0.111)	0.67 (0.092)	0.67 (0.117)				

* Statistically significant ($P < 0.005$)

$^\dagger H_E$ – Expected heterozygosity and F_{ST} were calculated using Arlequin 3.1 software.

TABLE 4
Linkage disequilibrium (standardized index of association, I^S_A) in all TR loci in different populations

	I^S_A				
	Amapa	Amazonas	Mato-Grosso	Para	Total
All infections	0.2360**	0.2038*	0.0626	0.0358	0.0278
Infection with unique haplotype	0.1870** (N = 9)	0.1122 (N = 6)	-0.0007 (N = 9)	-0.0159 (N = 9)	0.0088 (N = 33)

* $P < 0.05$

** $P < 0.01$



Figure 1

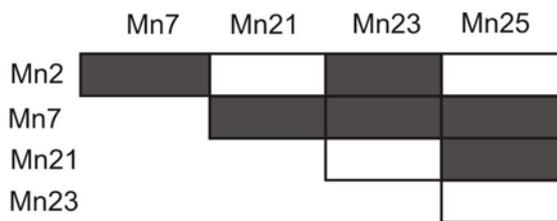


Figure 2

5.2 Segundo Artigo

Carta de submissão do artigo "Microsatellite loci: Determining the genetic variability of *Plasmodium vivax*" na revista Molecular & Biochemical Parasitology.

-----Mensagem original-----

De: Molecular & Biochemical Parasitology
Data: 01/20/09 17:43:00
Para: cristiana@cpqrr.fiocruz.br
Assunto: Submission Confirmation

Dear Dr Cristiana Ferreira Alves de Brito,

Your submission entitled "Microsatellite loci: Determining the genetic variability of *Plasmodium vivax*" has been received by Molecular & Biochemical Parasitology

You may check on the progress of your paper by logging on to the Elsevier Editorial System as an author. The URL is <http://ees.elsevier.com/molbio/>.

Your manuscript will be given a reference number once an Editor has been assigned.

Thank you for submitting your work to this journal.

Kind regards,

Elsevier Editorial System
Molecular & Biochemical Parasitology

Microsatellite loci: Determining the genetic variability of *Plasmodium vivax*

Antônio M. Rezende^a, Eduardo Tarazona-Santos^b, Cór Jesús F. Fontes^c, José Maria de Souza^d, Luzia H. Carvalho^a and Cristiana F. A. Brito^{a,*}

^a Laboratory of malaria, Centro de Pesquisas René Rachou/Fiocruz, Belo Horizonte, MG, Brazil

^b Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

^c Universidade Federal do Mato Grosso, Cuiabá, MT, Brazil

^d Instituto Evandro Chagas, Belém, PA, Brazil

*Corresponding author: Laboratory of Malaria, Rene Rachou Research Institute, Fundação Oswaldo Cruz, Avenida Augusto de Lima, 1715, Belo Horizonte, Brazil, 30190-002. Phone: (55) 31 33497772. Fax: (55) 31 32953115. E-mail: cristiana@cpqrr.fiocruz.br.

Running title: Microsatellite analysis of *Plasmodium vivax* isolates from Brazil

Abbreviations: LD – linkage disequilibrium ; He – gene diversity; API - Annual Parasite Index.

Abstract

Plasmodium vivax is currently the most globally widespread species causing human malaria, and control strategies depend on an understanding of its genetic variability and population structure. However, only few neutral or near neutral markers have been described for *P. vivax* analysis. Therefore, the aim of this work was to identify new polymorphic microsatellites for use in variability analysis of *P. vivax* isolates. An *in silico* search of the *P. vivax* genome was conducted, and microsatellites with repeat units comprising from two to four base pairs were selected. Following DNA extraction from the whole blood of patients with *P. vivax* acute infections from four different geographic regions of the Brazilian Amazon, each locus was amplified using specific primers. The amplified products were analyzed with the automatic sequencer MegaBACE using MegaBACE™ Genetic Profiler version 2.2 software. Substantial variability was observed among *P. vivax* isolates within populations, concurrent with high levels of multiple-clone infections and high linkage disequilibrium. Overall, moderate to high genetic differentiation was detected between these populations and, according to Structure software, structured populations were also observed. The results revealed differences among populations, suggesting that these and other microsatellites can be used to elucidate the population structure of *P. vivax*.

Key words: Malaria, *Plasmodium vivax*, microsatellite genetic variability, linkage disequilibrium, population structure

1. Introduction

More than two billion people are at risk of malaria infection, and approximately 300–500 million clinical cases are reported annually [1]. Malaria primarily affects impoverished populations in tropical and subtropical areas, where the temperature and rainfall are suitable for the development of malaria-causing *Plasmodium* parasites in *Anopheles* mosquitoes [2]. *Plasmodium vivax*, which causes human malaria, is the most widespread species globally. Because development of *P. vivax* within the mosquito can occur at a lower environmental temperature than that required by *Plasmodium falciparum*, infection with *P. vivax* occur in tropical as well as in subtropical regions.

Within an endemic area, parasite genetic diversity is frequently observed, and many field parasites circulate within the human host, as well as within mosquito vectors. The ability to distinguish between different isolates and populations is a pre-requisite for understanding the local and global epidemiology of *P. vivax*. In addition, it is also clear that studies concerning malaria parasite genetic diversity are not only of academic interest, but also have practical significance to implement novel control strategies.

Compared with the more virulent parasite *P. falciparum*, our knowledge about the genetic variability of *P. Vivax* is limited, likely because of difficulties associated with its *in vitro* maintenance [3]. Although *P. vivax* surface antigens show high genetic diversity [4]; this is likely due to the combined effects of population history of the parasite and natural selection acting on these loci [5]. Thus, these surface antigens are not the best loci to make inferences about population structures [6]. On the other hand, analyses of allelic variation at multiple independent loci such as microsatellite, that more likely are neutral or

nearly neutral, provide a most effective way to measure the population structure of *P. vivax*.

Several authors reported polymorphism levels for *P. vivax* microsatellites [7-10]. Leclerc et al. (2004) reported a low variability in the microsatellites loci studied [9]. In this study, only one of thirteen microsatellite loci identified showed extensive variability, and nine were completely monomorphic among the eight *P. vivax* populations analyzed. Using a draft of the unpublished *P. vivax* genome, Imwong et al. (2006) designed primer pairs for eleven dinucleotide microsatellites [8]. Their data illustrate that *P. vivax* possesses high allelic diversity. Imwong et al. (2006) explained the clear difference in their results compared to those of Leclerc et al. (2004) as due to the dependence of microsatellite variation on the repeat array length [8]. Long arrays would be more diverse than short arrays because slippage mutations become exponentially more common with an increase in array length. However, the same isolates were not analyzed using both types of microsatellites (with long and short repeat arrays).

In this context, the aim of the present work was to identify new polymorphic microsatellite loci to better elucidate the pattern of genetic diversity and population structure of *P. vivax* isolates from four malaria endemic areas of the Brazilian Amazon.

2. Materials and methods

2.1. Field isolates and DNA extraction

Fifty-three *P. vivax* isolates were obtained from the blood of infected patients from four geographical areas in the Brazilian Amazon, including 11 isolates from Macapá (00°02'20"S, 51°03'59"W) in Amapá State (AP), 16 isolates from Porto Velho (8°45'30", 63°24'28"W) in Rondônia State (RO), 15 isolates from Augusto Correa (01°01'18"S, 46°38'06"W) in Pará State (PA), and 11 isolates from Manaus (03°06'07"S, 60°01'30"W) in Amazonas State (AM) (Fig. 1). The samples were collected in May 2003 for AM, in November 2004 for AP, in October 2005 for PA, and from July 2003 to July 2004 for RO. The available epidemiological data indicate that endemic regions in Brazil exhibit hypo to mesoendemic malaria [11]. The rate of transmission is measured by the Annual Parasite Index (API), which reflects the number of positive blood smears per 1,000 inhabitants. According to the guidelines of the Health Surveillance Secretariat of the Ministry of Health, this index was used to stratify the areas based on the risk of malaria infection. High risk was designated as (API >50), medium risk as (API > 10 < 50), and low risk as (API <10) [12]. At time of blood collection, the studied areas were classified based on their respective API levels. Augusto Correa/PA and Porto Velho/RO were classified as high risk, Manaus/AM as medium risk, and Macapá/AP as low risk. Patient infection was confirmed by microscopic analysis of Giemsa-stained blood smears. Patients ranged from 16 – 58 years of age, with a mean age of 32 years.

DNA was extracted from whole blood samples using the PUREGENE DNA isolation Kit (Gentra Systems, Minneapolis, MN), according to the manufacturer's protocol.

2.2. Search for and amplification of new microsatellite loci

A *P. vivax* Salvador-I strain genome draft was downloaded from the Institute of Genomic Research (TIGR) website in February 2006 [13]. The draft genome contained many contigs, and these were scanned for repetitive sequences using RepeatMasker software (<http://www.repeatmasker.org>). This software released a list of repetitive sequences found in the genome draft, and all sequences that were not microsatellites were manually deleted. Microsatellite sequences were then selected using the following criteria: perfect tandem repeats, size of the repeats, number of repeats, and the availability of flanking sequences for primer design. Primers were designed using Oligo software version 4.0 [14, 15] (Table 1). The forward primers for each polymorphic locus were labeled with fluorescein to genotype the amplified products in an automatic DNA sequencer.

Polymerase Chain Reactions (PCRs) were standardized using an Eppendorf Gradient thermocycler. The melting temperatures and magnesium concentration ranged from 50°C to 60°C and from 0.75 mM to 1.50 mM, respectively (Table 1). A reaction volume of 20 µL containing 20 pmol of each primer (forward and reverse), 0.125 mM of dNTP, 1X buffer, and 1 U of Taq DNA polymerase (Invitrogen) was used. To check the amplification, products were visualized in an agarose gel stained with ethidium bromide. Amplification with fluorescent primers allows to visualize and differentiate alleles after capillary electrophoresis in an automatic DNA sequencer (MegaBACE, Amersham Biosciences, USA). The lengths of PCR products were determined by reference to internal size standards (MegaBACE™ ET550-R) using MegaBACE™ Fragment Profiler version 1.2 software.

2.3. Data analysis

The population genetics analysis was performed using the predominant infection haplotype, which was defined using the predominant allele observed for each locus, which is associated to the highest signal. When an isolate had more than one allele for a locus, the second allele was considered as signal of multiple infections if its height was at least one third of the predominant allele [16].

We calculated the gene diversity (expected heterozygosity, H_E), which is a measure of intrapopulation variability. The H_E may be defined as the probability that a pair of alleles randomly obtained from the population differs [6]. The Pearson correlation were performed between the number of repetitive units in each locus and the number of alleles, as well as among the size of repetitive units and the number of alleles using the BioEstat version 4.0 program.

To verify whether the population had suffered a population reduction event, such as a bottleneck or a founder effect, we used the statistical test proposed by Luikart & Cornuet and implemented in the software Bottleneck (version 1.2.02 software, <http://www1.montpellier.inra.fr/URLB/bottleneck/bottleneck.html>) [17]. We assumed a two-phase mutational model, which means that besides a step-wise mutation model, we assumed that microsatellites have a component of the infinite allele model in its process to generate new alleles. The test is based on the assumption that populations having recently experienced a reduction in their

effective population size exhibit a reduction of their number of alleles as well as the gene diversity at polymorphic loci. However, the number of alleles is reduced faster than the gene diversity (H_E). Thus, in a recently bottlenecked population, the observed gene diversity is higher than the expected equilibrium gene diversity, as computed from the observed number of alleles assuming a constant-size (equilibrium) population [17].

We calculated the fixation index F_{ST} to measure the differentiation between populations. The diversity measure H_E and index F_{ST} , were calculated using the Arlequin 3.1 software package (<http://cmpg.unibe.ch/software/arlequin3/>).

We also assessed population structure using the Structure version 2.1 software [18], which employs Bayesian approaches to infer the most likely number of populations (K) represented in the total sample, and then measures the probability that individual parasites derive from each of these K populations [19]. To run the software we assumed the admixture model, which consider that the genome of each individual parasite may have ancestries in more than one of the K parental populations. We also assumed a model of correlated allele frequencies, and we did not use prior information about population origin for each individual (PopFlag=0). For each K value considered, we ran the program three times, each one with 100,000 burnings and 1,000,000 iterations. K values ranged from 2-8 and K with the highest probability posteriori value was considered.

In addition, a linkage disequilibrium (LD) analysis was performed using LIAN (Linkage Analysis) software version 3.5 [20]. This software tests the null hypothesis of linkage equilibrium for multilocus data. Linkage equilibrium is

characterized by statistical independence of the alleles at all loci. LIAN tests for this independent assortment by first computing the number of loci at which each pair of individuals differs. From the distribution of the mismatch values (the number of loci with distinct allele for each pair of patient) a variance, V_D , is calculated and compared to the variance expected for linkage equilibrium, V_e . The null hypothesis $H_0: V_D = V_e$ is tested by computer simulation and by a parametric test [21]. The software calculates an association index (I_A^S) that allows for comparison between different datasets. Linkage disequilibrium analysis was performed for every pair of loci using Arlequin software. The observed data were arranged into contingency tables, and Markov chains were used to explore all contingency tables obtained with 1,000,000 simulated datasets in which alleles were randomly reshuffled among haplotypes. The significance level was defined as 0.05.

3. Results

3.1. Identification of new polymorphic microsatellite loci

The *P. vivax* genome draft was initially used to discover new microsatellites for our analysis. After the genome was scanned, we generated a list of the repetitive sequences contained in the *P. vivax* genome. A large number of repetitive sequences were observed in this genome, and many of them can be classified as microsatellites. Of the repetitive elements sequenced, 13,000 were classified as single repeats. Twenty-three microsatellites were

selected to design primers, and twelve of them did not work well or were not polymorphic in a preliminary PCR test. The other eleven microsatellites amplified properly and were polymorphic (Table 1 and 2).

3.2. Population Diversity

Fifty-three isolates from four different geographic areas located in the Brazilian Amazon were genotyped using the 11 selected microsatellite loci (Table 2). Table 2 shows the distribution the 11 selected microsatellites across the *P. vivax* genome. A great variation in the number of alleles for the different loci, and consequently a large variation in fragment sizes were observed (Table 2). The number of alleles varied from 4 (locus PVMS1) to 31 (locus PVMS11), with an average of 12 alleles. Concerning the multiple infections, Para and Amapa populations demonstrated a lower percentage of isolates comprising multiple-clone infections (0.40 and 0.54, respectively), while Amazonas and Rondonia populations presented higher percentages (0.69 and 0.72, respectively).

All of the studied populations demonstrated a high level of diversity. The mean H_E ranged from 0.57 to 0.77, with the Amapá and the Amazonas populations exhibiting the lowest and highest values, respectively (Table 3). Of interest, 51 haplotypes were found, and no genotypes in common between populations and within populations were detected. Only Amapa possessed two haplotypes that each appeared twice.

Using the two-phase model as our mutational model (Bottleneck software), for no population, we reject the null hypothesis of constant population

size (data not shown). However, these results should be interpreted with caution since 11 loci may not have enough power to detect a recent reduction in population size.

The F_{ST} values ranged from 0.06 to 0.23, with the highest values obtained when comparing Amapá with the other populations (Table 3). The Rondonia population appeared to be the most genetically distant from the Amapá population. Nevertheless, the Amazonas and Rondonia populations were genetically closest. Furthermore, to better summarize the matrix of F_{ST} values, we constructed a Neighbour-Joining tree (Figure 2).

A positive correlation was found between the number of repeat units and the number of alleles (Figure 3). This correlation was independent of the repeat unit size, i.e. di, tri, or tetranucleotide repeats. However, the reasons are unclear, but only the locus PVMS7 behaved differently. In addition, an ANOVA analysis revealed no statistical difference in the number of alleles between loci containing repetitive units of different sizes.

3.3. Population Structure

The population structure was evaluated using Structure software (Fig. 1). The most likely K is four, and this result was in agreement with the F_{ST} values. In the Figure 1, for each geographic sample, we represent in the circles the ancestry proportion for each of the $K=4$ populations as inferred by Structure. Each population showed a different pattern of ancestry, thus the parasites from these regions seem to possess admixed genetic backgrounds and one distinct predominant ancestral population (Figure 1). One of these ancestral

populations (yellow) was predominant in Amazonas population and was the second predominant ancestry population in all the other three populations.

3.4. Linkage disequilibrium analysis

Linkage disequilibrium (LD) analysis was performed using LIAN software (Table 3). Strong evidence for LD was found in the Amapa and Para populations. The Rondonia population showed evidence of LD, but it was not as strong as that observed for the two above mentioned populations. The Amazonas population showed no evidence of LD (Table 3). The loci pair analysis of linkage disequilibrium performed using Arlequin software showed that the Amapa and Para populations contain more loci pairs in linkage disequilibrium than Rondonia and Amazonas, which demonstrated the smallest number (Figure 4).

4. Discussion

In the present work, we wished to elucidate the population structure and analyze the genetic diversity of *P. vivax* isolates from different geographical regions of the Brazilian Amazon. The Brazilian Amazon is a very large region, and it demonstrates epidemiological characteristics that are geographically variable. For these reasons, the Brazilian Amazon is a particularly interesting region for population structure analysis. To achieve our aims, we identified eleven new polymorphic microsatellite loci. In addition, we analyzed the influence of the repeat unit and array length in genetic variability analysis of *P.*

vivax field isolates. Recently, the *P. vivax* genome was published by Carlton et al. (2008) [22], who searched for microsatellites using a different tool from those used in the present analysis. They found 333 loci, of which approximately 160 were polymorphic. By using Repeat Masker, we found a large number of microsatellites across *P. vivax* genome, relevant information because these molecular markers can be used for different purposes. This complete list of loci could be available upon request from the authors. Dense microsatellite panels may be used for instance, for association mapping of important phenotypic traits [23, 24]. This approach is particularly attractive for *P. vivax*, because long-term culture is difficult [19]. Interestingly, among the eleven loci used in our study, five were also found by Carlton et al. (2008) [22] (Table 1).

Large variations in the number of alleles found within each locus were demonstrated. A positive correlation was found between the number of alleles and the number of repeat units. This finding is consistent with the fact that microsatellite variation is strongly dependent on the length of repeat arrays due to an exponential increase in the amount of DNA polymerase slippage that is observed with increasing repeat numbers [25, 26]. Our results corroborate the studies performed by Imwong et al. (2006) demonstrating the importance of microsatellite array length [8]. However, because of unknown reasons, the locus PVMS7 behaved differently compared to all other microsatellite loci analyzed. Might be because of this locus could be under constraints of selective pressures acting in genes flanking this locus, the small rRNA subunit upstream this locus (apart 4,022 bp) and ABC transporter downstream this locus (apart 1,028 bp).

Genetic distances were not associated with geographic distances, as previously reported for *P. falciparum* isolates from Brazil [27]. When we used

the Structure software to assess population structure, the results were similar to F_{ST} analysis. Based on this result, many isolates from Amazonas and Rondonia were found to be genetically similar, and this can be explained by high gene flow between these geographic regions.

In addition, we observed significant multilocus linkage disequilibrium in three out of the four studied populations. The Amapa population showed the strongest evidence of LD, while the Amazonas population revealed no evidence of LD. High LD may also result from recent bottlenecks; however, evidence for such an event was not found. LD has also been reported for *P. falciparum* and *P. vivax* natural populations from low transmission areas, such as the Brazilian Amazon [27, 28]. However, we showed that isolates from various geographic areas display extensive genetic diversity, and we also illustrated that frequent multiple-clone infections concur with a high level of LD, as previously demonstrated for the Acre population, another Brazilian State [28]. Since we obtained populations from areas exhibiting a low, medium, and high risk of malaria, a negative correlation between the transmission level and the LD and a positive one between transmission level and multiplicity of infections were not observed. Nevertheless, the genetic population structure of *P. falciparum* is strongly linked to the patterns of transmission [29]. However, *P. vivax* appears to be more diverse and more frequently associated with high multiple-clone infections than sympatric *P. falciparum* isolates [28]. These results suggest that the genetic structure of *P. vivax* is unlikely to be determined by recombination.

One of our populations, the Amapa, was the most homogenous and differentiated population among those studied in the present report. This population displayed high LD, genetic diversity, and elevated multiple-clone

infections. We have some hypotheses for these findings that are not mutually exclusive. First, this population may have been established more recently than other populations, and therefore it did not have time to accumulate new mutational events. However, no evidence for a bottleneck or founder effect was revealed by our analysis conducted using the Bottleneck software. The second hypothesis postulates the presence of a different vector in this area. The major vector in Brazil is *Anopheles darlingi*, although there are several reports demonstrating that *Anopheles marajoara* is the major vector in the Amapa state [30, 31]. If this is the case, this vector could be applying a selective pressure that differs from other regions and decreases the overall genetic diversity in this particular region. Joy *et al.* (2008) [32] demonstrated the importance of the vector for the population structure of parasites. The third hypothesis involves recombination events, which occur inside the mosquito, where the process of meiosis occurs during the sexual phase of the parasite life cycle. Therefore, when numerous infected individuals are present in a location with multiple parasites, mosquitoes can bite one or more persons and gain access to more than one parasite strain. Thus, a recombination event can occur between these parasites, consequently decreasing the linkage disequilibrium among loci. The Amapa state has a very low API value of 8.9. This value directly affects recombination levels because the transmission rate is low, and consequently the chances of recombination between different parasites are also low.

In conclusion, we demonstrated a high diversity in Amazon parasite populations. These populations, however, are structured, and thus they can be differentiated genetically. In addition, we demonstrated that the Amapa population is the most divergent population, likely due to many peculiarities

associated with this state, such as a low API and multiple malaria vectors. Finally, the present work presents epidemiological data associated with the variability data that could increase our overall understanding of the local malaria situation, and this is very important in defining regional strategies for disease control.

Acknowledgements

We thank the Brazilian patients for their substantial cooperation. We also thank Elisângela Monteiro Coser, technician from sequencing platform (PDTIS) and Nilton Barnabé for helping in DNA genotyping and Daniel Coscarelli for providing a Brazilian geographic map. In addition, we are thankful for the financial support of FAPEMIG, CNPq, CAPES, PDTIS program (Fiocruz) and CPqRR/Fiocruz.

References

- [1] Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present, and future. *Lancet Infectious Diseases* 2004; 4:327-336.
- [2] Greenwood B, Fidock D, Kyle D, Kappe S, Alonso P, Collins F, Duffy P. Malaria: progress, perils, and prospects for eradication. *J Clin Invest* 2008;118:1266-1276.
- [3] Feng XR, Carlton JM, Joy DA, Mu JB, Furuya T, Suh BB, Wang YF, Barnwell JW, Su XZ. Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. *Proceedings of the National Academy of Sciences of the United States of America* 2003;100:8502-8507.
- [4] Cui LW, Escalante AA, Imwong M, Snounou G. The genetic diversity of *Plasmodium vivax* populations. *Trends in Parasitology* 2003;19:220-226.
- [5] Escalante AA, Cornejo OE, Rojas A, Udhayakumar V, Lal AA. Assessing the effect of natural selection in malaria parasites. *Trends in Parasitology* 2004;20:388-395.
- [6] Karunaweera ND, Ferreira MU, Munasinghe A, Barnwell JW, Collins WE, King CL, Kawamoto F, Hartl DL, Wirth DF. Extensive microsatellite diversity in the human malaria parasite *Plasmodium vivax*. *Gene* 2008;410:105-112.
- [7] Russell B, Suwanarusk R, Lek-Uthai U. *Plasmodium vivax* genetic diversity: microsatellite length matters. *Trends Parasitol* 2006;22:399-401.
- [8] Imwong M, Sudimack D, Pukrittayakamee S, Osorio L, Carlton J, Day N, White N, Anderson T. Microsatellite variation, repeat array length, and population history of *Plasmodium vivax*. *Mol Biol Evol* 2006;23:1016-1018.
- [9] Leclerc M, Durand P, Gauthier C, Patot S, Billotte N, Menegon M, Severini C, Ayala F, Renaud F. Meager genetic variability of the human malaria agent *Plasmodium vivax*. *Proc Natl Acad Sci U S A* 2004;101:14455-14460.
- [10] Gomez J, McNamara D, Bockarie M, Baird J, Carlton J, Zimmerman P. Identification of a polymorphic *Plasmodium vivax* microsatellite marker. *Am J Trop Med Hyg* 2003;69:377-379.
- [11] Marcelo L, Camargo A, Ferreira MU, Krieger H, Decamargo EP, Dasilva LP. Unstable Hypoendemic Malaria In Rondonia (Western Amazon Region, Brazil) - Epidemic Outbreaks And Work-Associated Incidence In An Agroindustrial Rural Settlement. *American Journal of Tropical Medicine and Hygiene* 1994;51:16-25.
- [12] Health Ministry of Brazil, Malaria Epidemiologic Situation in Brazil, Secretary of Vigilance in Health, 2007.
- [13] Carlton J. The *Plasmodium vivax* genome sequencing project. *Trends in Parasitology* 2003;19:227-231.
- [14] Wojciech R, Rhoads R. A Computer Program for Choosing Optimal Oligonucleotides for Filter Hybridization, Sequencing and in vitro Amplification of DNA. *Nucleic Acids Res* 1989;17:8543-8551.
- [15] Wojciech R, Spencer W, Rhoads R. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res* 1990;18:6409-6412.
- [16] Anderson TJC, Su XZ, Bockarie M, Lagog M, Day KP. Twelve microsatellite markers for characterization of *Plasmodium falciparum* from finger-prick blood samples. *Parasitology* 1999;119:113-125.
- [17] Luikart G, Cornuet JM. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* 1998;12:228-237.

- [18] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945-959.
- [19] Imwong M, Nair S, Pukrittayakamee S, Sudimack D, Williams JT, Mayxay M, Newton PN, Kim JR, Nandy A, Osorio L, Carlton JM, White NJ, Day NPJ, Anderson TJC. Contrasting genetic structure in *Plasmodium vivax* populations from Asia and South America. *International Journal for Parasitology* 2007;37:1013-1022.
- [20] Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Bioinformatics* 2000; 16:847-848.
- [21] Haubold B, Travisano M, Rainey PB, Hudson RR. Detecting linkage disequilibrium in bacterial populations. *Genetics* 1998;150:1341-1348.
- [22] Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, Cheng Q, Coulson RMR, Crabb BS, del Portillo HA, Essien K, Feldblyum TV, Fernandez-Becerra C, Gilson PR, Gueye AH, Guo X, Kang'a S, Kooij TWA, Korsinczky M, Meyer EVS, Nene V, Paulsen I, White O, Ralph SA, Ren QH, Sargeant TJ, Salzberg SL, Stoeckert CJ, Sullivan SA, Yamamoto MM, Hoffman SL, Wortman JR, Gardner MJ, Galinski MR, Barnwell JW, Fraser-Liggett CM. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 2008;455:757-763.
- [23] Su XZ, Wootton JC. Genetic mapping in the human malaria parasite *Plasmodium falciparum*. *Molecular Microbiology* 2004;53:1573-1582.
- [24] Anderson TJC. Mapping drug resistance genes in *Plasmodium falciparum* by genome-wide association. *Curr. Drug Targets Infect. Dis.* 2004;4:65-78.
- [25] Imwong M, Sudimack D, Pukrittayakamee S, Osorio L, Carlton JM, Day NPJ, White NJ, Anderson TJC. Microsatellite variation, repeat array length, and population history of *Plasmodium vivax*. *Molecular Biology and Evolution* 2006;23:1116-1018.
- [26] Russell B, Suwanarusk R, Lek-Uthai U. *Plasmodium vivax* genetic diversity: microsatellite length matters. *Trends in Parasitology* 2006;22:399-401.
- [27] Machado RLD, Pova MM, Calvosa VSP, Ferreira MU, Rossit ARB, dos Santos EJM, Conway DJ. Genetic structure of *Plasmodium falciparum* populations in the Brazilian Amazon region. *Journal of Infectious Diseases* 2004;190:1547-1555.
- [28] Ferreira MU, Karunaweera ND, da Silva-Nunes M, da Silva NS, Wirth DF, Hartl DL. Population structure and transmission dynamics of *Plasmodium vivax* in rural amazonia. *Journal of Infectious Diseases* 2007;195:1218-1226.
- [29] Anderson TJC, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, Whitworth J, Velez ID, Brockman AH, Nosten F, Ferreira MU, Day K.P. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Molecular Biology and Evolution* 2000;17:1467-1482.
- [30] Conn JE, Wilkerson RC, Segura MNO, De Souza RTL, Schlichting CD, Wirtz RA, Pova MM. Emergence of a new neotropical malaria vector facilitated by human migration and changes in land use. *American Journal of Tropical Medicine and Hygiene* 2002;66:18-22.
- [31] Li C, Wilkerson RC. Identification of *Anopheles (Nyssorhynchus) albitarsis* complex species (Diptera : Culicidae) using rDNA internal transcribed spacer 2-based polymerase chain reaction primers. *Memorias Do Instituto Oswaldo Cruz* 2005;100:495-500.
- [32] Joy DA, Gonzalez-Ceron L, Carlton JM, Gueye A, Fay M, McCutchan TF, Su XZ. Local adaptation and vector-mediated population structure in *Plasmodium vivax* malaria. *Molecular Biology and Evolution* 2008;25:1245-1252.

Table1

Description of the microsatellite-specific primers and amplification conditions.

Microsatellite	Forward (5' →. 3')	Reverse (5' →. 3')	Ta(°C) ^a	[Mg+2] (mM)
PvMS1	CTATCTGAGGAATGGGGA	ATTTACTATGACGAAGGTGA	53.4	1.50
PvMS2	CATCATTGGGTAAGTCGGG	GCAGCCACAAAATCAACACC	60.0	1.50
PvMS3*	GGGAAGCACAAAATCGTA	TCAGCAGGGACAAAACG	60.0	1.50
PvMS4	TTATTTCCCCCTTTGCC	AAATGGATGTTCTTGTCAA	55.7	1.00
PvMS5*	TGCTATTTGCTCGGTCTGT	GAGCGTTATCATCATTAG	56.0	1.50
PvMS6	ACACATTTGACACAGTTCC	ATGCCCTGGTCCCTACAA	58.6	1.50
PvMS7*	GTATTTCCCGTCTTGTCC	CTTTCTCCGTTCTTATTTCT	56.0	1.50
PvMS8	TCCGTTGTTTTGTTGCC	CACTTGTTTCGTTCCGCTC	60.0	1.50
PvMS9	TGTGGATAAGGGGAAAAA	TTTTTTTCCTTGAGTTTACG	50.0	0.75
PvMS10*	AAGTGTATTTTCCCGACG	CTTTTGCTTGCTCCGTTT	54.7	1.50
PvMS11*	CGATGCGTTCACCTGGAT	TATCTTCTCCCCTCGTG	54.0	0.75

^a Temperature of primers annealing

* Loci also described by Carlton et al. 2008 [22]

Table2

Microsatellite characteristics: Repeat unit, array length, and allele numbers

Microsatellite	Chromosome ^a	Repeat Unit ^b	Range of array length (bp) ^c	Numbers of alleles ^d
PvMS1	12	GT ₉	224-248	4
PvMS2	3	CA ₁₂	292-312	8
PvMS3	8	TA ₁₈	145-197	10
PvMS4	6	TA ₂₁	130-174	15
PvMS5	3	CAT ₁₀	172-218	12
PvMS6	14	TGA ₁₉	185-301	19
PvMS7	2	TAA ₂₂	349-439	10
PvMS8	13	TGTA ₇	225-321	9
PvMS9	6	CATA ₁₁	193-217	7
PvMS10	5	TATG ₁₄	265-376	10
PvMS11	8	CATA ₄₉	292-596	31

^a Chromosome location determined using Primer-blast tool at www.ncbi.nlm.nih.gov.

^b Repeat number in Sal-1 sequence according to the GenBank sequence.

^c Range of array length detected herein..

^d Allele numbers detected herein.

Table 3

Microsatellite-based analysis of the gene diversity (expected heterogeneity, H_E), linkage disequilibrium (I_A^S), and genetic distances (F_{ST}) between populations of *P. vivax* from different Brazilian geographical regions.

Populations	$(H_E) \pm SD$	I_A^S	F_{ST}		
			AP	RO	PA
AP	0.57 ± 0.27	0.4228*	-		
RO	0.75 ± 0.13	0.0770*	0.2370*	-	
PA	0.74 ± 0.13	0.1706*	0.1732*	0.1115*	-
AM	0.77 ± 0.16	0.0344	0.2352*	0.0681*	0.1355*

* $P < 0.05$

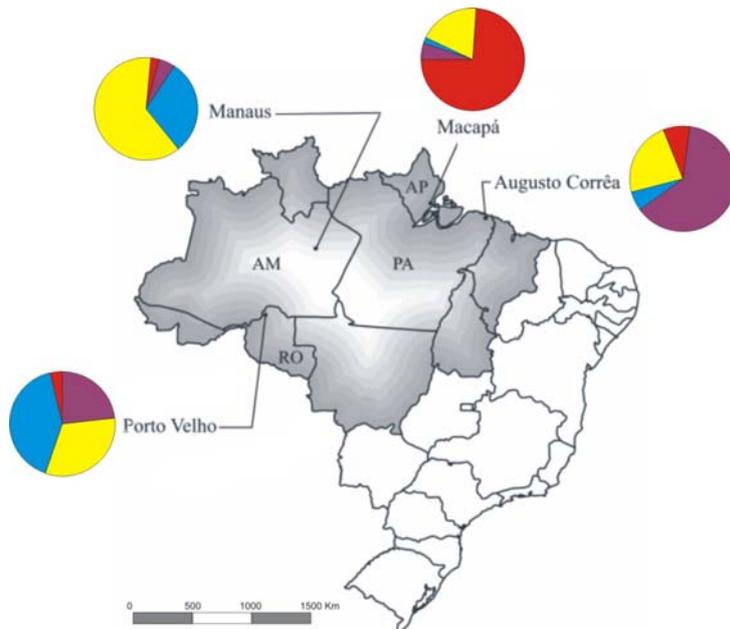


Figure 1 – Map of Brazil, with the grey color representing regions endemic for malaria. AM, AP, PA, and RO indicate the Amazonas, Amapá, Pará, and Rondonia states, respectively. The localities of Manaus, Macapá, Augusto Corrêa, and Porto Velho indicate the localities where the samples were collected. Results from the Structure analyses are summarized for the populations in the colored circles. Each color represents the contribution of each inferred $K=4$ ancestral population to the sampled population.

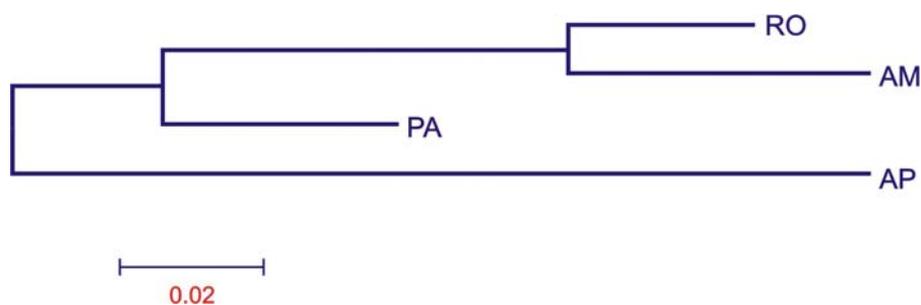
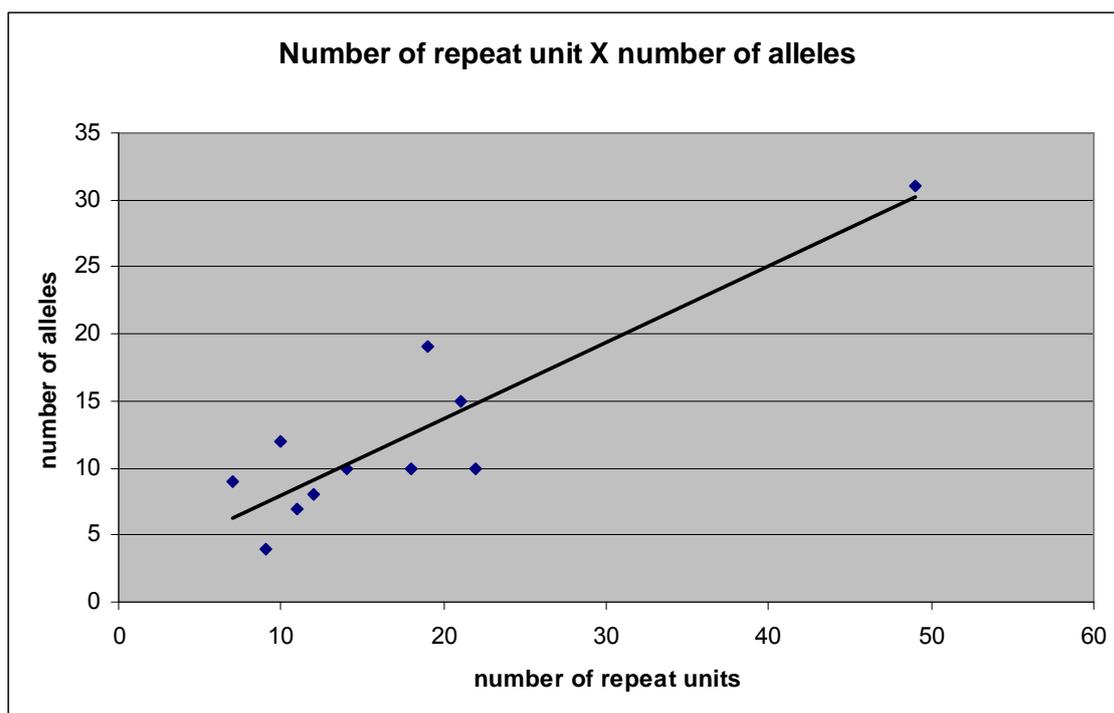


Figure 2 – Neighbor-Joining tree based on F_{ST} values calculated using Arlequin version 3.1 software. The tree was constructed using Mega 4.0 software.



* Correlation coefficient of Pearson ($r=0.8989$ – $p<0.05$)

Figure 3 - Correlation analysis between the number of repeat units of microsatellite loci and the number of alleles detected among *P. vivax* Brazilian isolates.

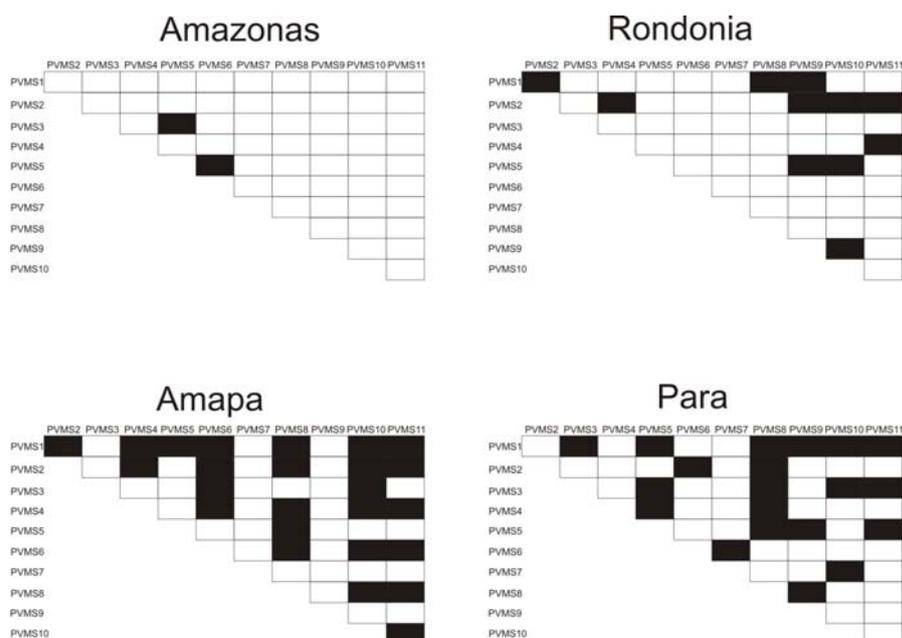


Figure 4 – Linkage Disequilibrium analyses for pair of loci in each studied population. Each box represents the level of LD for a pair of loci. Black box represents a pair of loci in linkage disequilibrium. The level of significance was defined as 0.05. These analyses were performed using Arlequin version 3.1 software.

6 Considerações Finais

O trabalho aqui descrito foi composto basicamente por duas partes publicadas em dois artigos, um aceito e outro submetido. As duas partes tinham como principal objetivo a elucidação da estrutura populacional de isolados de *P. vivax* oriundos de diferentes regiões da Amazônia Brasileira. Contudo, a diferença crucial entre as duas partes deste trabalho foi a metodologia empregada na análise dos isolados. A primeira parte foi concretizada utilizando como ferramenta molecular os marcadores do tipo TR, já descritos na literatura, enquanto a segunda foi realizada através da utilização de novos marcadores moleculares do tipo microssatélite. Além disso, para os estudos realizados com os microssatélites, a população MT não foi utilizada, pois as amostras desta população eram provenientes de diferentes lugares o que poderia acrescentar mais variáveis genéticas para as análises de estrutura de população.

Quando os TRs foram utilizados foi possível notar que as populações de *P. vivax* oriundas das áreas em estudos possuíam uma alta diversidade genética, com altos valores de H_E . Através da análise do tipo AMOVA, grande parte da diversidade encontrada foi endereçada como sendo variações dentro da própria população e não entre elas. Ainda relacionado à diversidade, não foi possível perceber correlação entre a diversidade medida pelo H_E e o tamanho do arranjo, o que pode sugerir que o mecanismo de manutenção de diversidade envolvido nos loci de TR não seja o mesmo para os microssatélites. Além disso, como os marcadores estavam todos ligados fisicamente no mesmo cromossomo, cromossomo 8, alto desequilíbrio de ligação (LD) foi detectado entre eles, entretanto, não foi possível visualizar nenhuma correlação entre a distância física entre eles e os pares de marcadores em desequilíbrio de ligação. Foi identificada uma estruturação das populações com base nos valores de F_{ST} , entretanto, não foi possível visualizar essa estruturação utilizando o programa Structure. Provavelmente isso ocorreu pelo pequeno número de marcadores e pelo fato de alguns deles estarem em LD; assim, a quantidade de informações independentes sobre cada população teve seu número reduzido e conseqüentemente, diminuindo a possibilidade de

inferir os estados estruturais destas populações. Ainda sobre os padrões de LD encontrados, era esperado, como ocorre em *P. falciparum*, que quanto maior a taxa de infecções múltiplas, mais fraco seria o LD. Porém, este padrão não ficou claro para *P. vivax* quando os TRs foram utilizados. A população AM, que apresentou uma taxa de multiplicidade de infecção de 66%, apresentou um LD mais forte do que as populações de MT e PA, que possuíam baixas taxas de multiplicidade de infecção.

Do mesmo modo que os TRs, as análises realizadas utilizando marcadores moleculares do tipo microssatélite detectaram também uma alta diversidade genética para cada uma das populações em estudo, fato este evidenciado também pelos altos valores de H_E , que foram em geral superiores aos valores encontrados para os TRs, comparando as mesmas populações. Além disso, foi possível observar uma correlação positiva entre a diversidade dos loci (H_E) e o número de alelos encontrados, em relação ao número de repetições dos mesmos. Portanto, devido ao mecanismo de manutenção de variabilidade dos microssatélites, quanto maior o número de repetições do locus mais polimórfico este é. Fato este corroborado por trabalhos prévios na literatura (26, 27). Além disso, estes marcadores possibilitaram inferir a estrutura populacional das populações estudadas, ao contrário dos TRs, provavelmente devido ao maior número de marcadores e ao fraco LD encontrado, pois os loci estavam em sua maioria em cromossomos diferentes e mesmo os pares de loci localizados no mesmo cromossomo estavam também, em sua grande maioria, em equilíbrio. Assim, utilizando os microssatélites foi possível obter um maior número de informações independentes sobre cada população em estudo. Este aumento na quantidade de dados populacionais foi importante para balancear uma característica inerente aos loci polimórficos de microssatélite; devido a uma maior taxa de mutação destes loci, poderiam surgir alelos iguais (mesmo número de repetições) no mesmo locus em diferentes populações independentemente, sem origem comum, o que afetaria as inferências populacionais. Finalmente, vale ressaltar que o LD encontrado utilizando os microssatélites, chamado de “admixture LD” (38) foi levado em conta pelo modelo adotado na metodologia utilizada para estruturação das populações (structure). Diferente do TRs, os quais o LD encontrado, chamado

de “background LD” (38), que não é possível ser tratado nos modelos utilizados na metodologia de estruturação.

Além dos resultados comentados acima, outro fato interessante foi a identificação das populações geneticamente mais semelhantes, baseado no índice F_{ST} , nas duas diferentes etapas do projeto. Quando os TRs foram utilizados, as populações AM e AP foram as mais próximas; porém, quando os microssatélites foram utilizados como marcadores, as populações AM e RO foram as mais próximas. Como a população RO não está incluída na primeira etapa, não é possível saber se utilizando os TRs esta população seria mais próxima da população AM. Porém, a população AP nos estudos realizados utilizando microssatélites se apresentou como sendo a população mais distante geneticamente quando comparada com as outras três populações, sendo PA a população mais próxima da mesma. Além disso, os dados de microssatélites possibilitaram a estruturação das populações através de uma outra metodologia, diferente do F_{ST} , e estes resultados corroboram a evidência de que AM e RO são mais próximas geneticamente e AP a mais distante geneticamente. Estes resultados também foram corroborados por outros dados do nosso grupo através da análise de SNPs no gene codificador da Duffy binding protein (39).

Concluindo, os microssatélites se apresentaram como sendo marcadores moleculares mais úteis para as diversas análises populacionais. Algumas razões para isto podem ser pontuadas. Primeiro, os microssatélites estão mais dispersos pelo genoma do parasito, isto diminui a possibilidade de LD e um possível viés de seleção natural em determinadas regiões do genoma. Por outro lado, os TRs foram localizados somente em um cromossomo. Segundo, os microssatélites possuem seu mecanismo de manutenção da variabilidade conhecido, assim, se torna mais fácil escolher ferramentas e elaborar modelos populacionais para estes marcadores. Em contraste, os TRs não possuem um modelo de mutação definido.

7 Referências Bibliográficas

1. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present, and future. *Lancet Infectious Diseases*. 2004;4(6):327-36.
2. Greenwood BM, Fidock DA, Kyle DE, Kappe SHI, Alonso PL, Collins FH, et al. Malaria: progress, perils, and prospects for eradication. *Journal of Clinical Investigation*. 2008;118(4):1266-76.
3. Keiser J, Utzinger J, De Castro MC, Smith TA, Tanner M, Singer BH. Urbanization in sub-Saharan Africa and implication for malaria control. *American Journal of Tropical Medicine and Hygiene*. 2004;71(2):118-27.
4. SVS. Malaria Epidemiologic Situation in Brazil. Health Ministry of Brazil; 2007.
5. Loiola CCP, da Silva CJM, Tauil PL. Controle da malária no Brasil: 1965 a 2001. *Rev Panam Salud Publica*. 2002;11(4):10.
6. Amino R, Thiberge S, Shorte S, Frischknecht F, Menard R. Quantitative imaging of Plasmodium sporozoites in the mammalian host. *Comptes Rendus Biologies*. 2006;329(11):858-62.
7. Chen Q, Wang H. Implications of imaging malaria sporozoites. *Trends in Parasitology*. 2008;24(3):106-9.
8. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. *Nature*. 2002;415(6872):673-9.
9. Sturm A, Amino R, van de Sand C, Regen T, Retzlaff S, Rennenberg A, et al. Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science*. 2006;313(5791):1287-90.
10. Weatherall DJ, Miller LH, Baruch DI, Marsh K, Doumbo OK, Casals-Pascual C, et al. Malaria and the Red Cell. *Hematology*. 2002.
11. Vlachou D, Schlegelmilch T, Runn E, Mendes A, Kafatos FC. The developmental migration of Plasmodium in mosquitoes. *Current Opinion in Genetics & Development*. 2006;16(4):384-91.
12. Alano P. Plasmodium falciparum gametocytes: still many secrets of a hidden life. *Molecular Microbiology*. 2007 Oct;66(2):291-302.
13. Vlachou D, Zimmermann T, Cantera R, Janse CJ, Waters AP, Kafatos FC. Real-time, in vivo analysis of malaria ookinete locomotion and mosquito midgut invasion. *Cellular Microbiology*. 2004;6(7):671-85.
14. Cui LW, Escalante AA, Imwong M, Snounou G. The genetic diversity of Plasmodium vivax populations. *Trends in Parasitology*. 2003;19(5):220-6.

15. Feng XR, Carlton JM, Joy DA, Mu JB, Furuya T, Suh BB, et al. Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. *Proceedings of the National Academy of Sciences of the United States of America*. 2003 Jul;100(14):8502-7.
16. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, et al. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*. 2008;455(7214):757-63.
17. Zakeri S, Barjesteh H, Djadid ND. Merozoite surface protein-3 alpha is a reliable marker for population genetic analysis of *Plasmodium vivax*. *Malaria Journal*. 2006;5.
18. Bruce MC, Galinski MR, Barnwell JW, Snounou G, Day KP. Polymorphism at the merozoite surface protein-3 alpha locus of *Plasmodium vivax*: Global and local diversity. *American Journal of Tropical Medicine and Hygiene*. 1999;61(4):518-25.
19. Arnot DE, Barnwell JW, Tam JP, Nussenzweig V, Nussenzweig RS, Enea V. Circumsporozoite Protein Of *Plasmodium-Vivax* - Gene Cloning And Characterization Of The Immunodominant Epitope. *Science*. 1985;230(4727):815-8.
20. Rosenberg R, Wirtz RA, Lanar DE, Sattabongkot J, Hall T, Waters AP, et al. Circumsporozoite Protein Heterogeneity In The Human Malaria Parasite *Plasmodium-Vivax*. *Science*. 1989;245(4921):973-6.
21. Leclerc MC, Durand P, Gauthier C, Patot S, Billotte N, Menegon M, et al. Meager genetic variability of the human malaria agent *Plasmodium vivax*. *Proceedings of the National Academy of Sciences of the United States of America*. 2004 Oct;101(40):14455-60.
22. Sharma PC, Grover A, Kahl G. Mining microsatellites in eukaryotic genomes. *Trends in Biotechnology*. 2007;25:490-8.
23. Levinson G, Gutman GA. Slipped-Strand Mismatching - A Major Mechanism For Dna-Sequence Evolution. *Molecular Biology and Evolution*. 1987;4(3):203-21.
24. Weber JL. Informativeness Of Human (Dc-Da)N.(Dg-Dt)N Polymorphisms. *Genomics*. 1990;7(4):524-30.
25. Imwong M, Nair S, Pukrittayakamee S, Sudimack D, Williams JT, Mayxay M, et al. Contrasting genetic structure in *Plasmodium vivax* populations from Asia and south America. *International Journal for Parasitology*. 2007 Jul;37(8-9):1013-22.
26. Imwong M, Sudimack D, Pukrittayakamee S, Osorio L, Carlton JM, Day NPJ, et al. Microsatellite variation, repeat array length, and population history of *plasmodium vivax*. *Molecular Biology and Evolution*. 2006 May;23(5):1116-018.
27. Russell B, Suwanarusk R, Lek-Uthai U. *Plasmodium vivax* genetic diversity: microsatellite length matters. *Trends in Parasitology*. 2006 Sep;22(9):399-401.

28. Ferreira MU, Karunaweera ND, da Silva-Nunes M, da Silva NS, Wirth DF, Hartl DL. Population structure and transmission dynamics of plasmodium vivax in rural amazonia. *Journal of Infectious Diseases*. 2007 Apr;195(8):1218-26.
29. Karunaweera ND, Ferreira MU, Munasinghe A, Barnwell JW, Collins WE, King CL, et al. Extensive microsatellite diversity in the human malaria parasite Plasmodium vivax. *Gene*. 2008 Feb;410(1):105-12.
30. Joy DA, Gonzalez-Ceron L, Carlton JM, Gueye A, Fay M, McCutchan TF, et al. Local adaptation and vector-mediated population structure in Plasmodium vivax malaria. *Molecular Biology and Evolution*. 2008;25(6):1245-52.
31. Anderson TJC, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. *Molecular Biology and Evolution*. 2000 Oct;17(10):1467-82.
32. Marcelo L, Camargo A, Ferreira MU, Krieger H, Decamargo EP, Dasilva LP. Unstable Hypoendemic Malaria In Rondonia (Western Amazon Region, Brazil) - Epidemic Outbreaks And Work-Associated Incidence In An Agroindustrial Rural Settlement. *American Journal of Tropical Medicine and Hygiene*. 1994;51(1):16-25.
33. Excoffier L, Laval G, Schneider S. Arlequin ver 3.1 - An integrated software package for population genetics data analysis. 2006. p. 145.
34. Wrigth S. The genetical structure of populations. *Ann Eugen*. 1951;15:32.
35. Wrigth S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evol*. 1965;19:26.
36. Faria, U C, Magnabosco CU, Reyes AL, Lôbo RB, Bezerra LAF. Inferência Bayesiana e sua aplicação na avaliação genética de bovinos da raça nelore: revisão bibliográfica. *Ciência Animal Brasileira*. 2007;8(1):12.
37. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155(2):945-59.
38. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*. 2003;164(4):1567-87.
39. Sousa T, Cerávolo I, Fontes C, Couto A, Carvalho L, Brito C. The pattern of major polymorphism in the Duffy binding protein ligand domain among Plasmodium vivax isolates from the Brazilian Amazon area. *Molecular & Biochemical Parasitology*. 2006;146:4.