

THAIS CRIPPA DE OLIVEIRA

Genômica populacional de *Plasmodium vivax*: níveis e mecanismos de diversidade genética na América

Dissertação apresentada ao Programa de Pós-Graduação em Biologia da Relação Patógeno-Hospedeiro do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção de Título de Mestre em Ciências.

São Paulo
2016

THAIS CRIPPA DE OLIVEIRA

Genômica populacional de *Plasmodium vivax*: níveis e mecanismos de diversidade genética na América

Dissertação apresentada ao Programa de Pós-Graduação em Biologia da Relação Patógeno-Hospedeiro do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção de Título de Mestre em Ciências.

Área de concentração: Biologia da Relação Patógeno-Hospedeiro

Orientador: Prof. Dr. Marcelo Urbano Ferreira.

Co-Orientador: Prof. Dr. João Marcelo Pereira Alves

Versão Original

São Paulo
2016

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e informação Biomédica
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Crippa de Oliveira, Thais
Genômica populacional de Plasmodium vivax: níveis
e mecanismos de diversidade genética na América /
Thais Crippa de Oliveira; orientador Marcelo
Urbano Ferreira; coorientador João Marcelo
Pereira Alves. -- São Paulo, 2016.
94 p.

Dissertação (Mestrado)) -- Universidade de São
Paulo, Instituto de Ciências Biomédicas.

1. Plasmodium vivax. 2. Malária. 3. SNP. 4.
Diversidade genética. I. Urbano Ferreira, Marcelo ,
orientador. II. Pereira Alves, João Marcelo ,
coorientador. III. Título.

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a): Thais Crippa de Oliveira

Titulo da Dissertação/Tese: Genômica populacional de Plasmodium vivax: níveis e mecanismos de diversidade genética na América

Orientador: Marcelo Urbano Ferreira

A Comissão Julgadora dos trabalhos de Defesa da Dissertação de Mestrado/Tese de Doutorado, em sessão pública realizada a/...../....., considerou o(a) candidato(a):

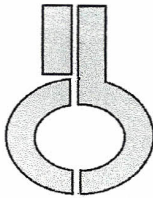
() **Aprovado(a)** () **Reprovado(a)**

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Presidente: Assinatura:
Nome:
Instituição:



**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS**

Cidade Universitária "Armando de Salles Oliveira"
Av. Prof. Lineu Prestes, 2415 - cep. 05508-000 São Paulo, SP - Brasil
Telefone : (55) (11) 3091.7733 telefax : (55) (11) 3091-8405
e-mail: cep@icb.usp.br

São Paulo, 03 de junho de 2014.

PARECER 1183/CEPSH

A Comissão de *Ética em Pesquisas em Seres Humanos* do ICB, na sessão realizada no dia 21.05.2014, **APROVOU** o projeto intitulado: "*Genômica populacional de Plasmodium vivax na Amazônia rural brasileira*" do pesquisador **MARCELO URBANO FERREIRA** e aluna **THAIS CRIPPA DE OLIVEIRA**.

Cabe aos pesquisadores elaborar e apresentar a este Comitê, relatórios anuais (parciais e final), de acordo com a Resolução nº 466/12, item II, II.19 e II.20, do Conselho Nacional de Saúde, conforme modelo constante no site: icb.usp.br.

Aos pesquisadores cabe também finalizar o processo junto à Plataforma Brasil quando do encerramento deste.

O primeiro relatório deverá ser encaminhado à Secretaria deste CEP em **21.05.2015**.

Atenciosamente,

Profª. Dra. **PAOLO M.A. ZANOTTO**
Coordenador da Comissão de Ética em
Pesquisas com Seres Humanos - ICB/USP

AGRADECIMENTOS

Nas seguintes linhas irei mostrar minha gratidão a todos aqueles que participaram do meu projeto, direta ou indiretamente, tendo certeza de não ter esquecido de ninguém. :)

Eu agradeço ao apoio da USP e da CAPES que custearam meu projeto além da Fundação Faculdade de Medicina e das agências de fomento CNPq e FAPESP.

Agradeço aos membros da banca que aceitaram meu convite e que espero que possam apreciar a leitura dessa dissertação.

Agradeço ao meu orientador, Marcelo, primeiramente por ter me aceito no seu grupo de trabalho e por me dar a oportunidade de trabalhar em um campo tão lindo como a evolução de populações de *P. vivax*. Obrigada por toda a compressão com meus horários malucos de trabalho, com todo o entusiasmo com meus resultados e por ser um orientador presente e atencioso.

Agradeço ao meu co-orientador, J, que calmamente me ensinou o mundo dos servidores e as técnicas clássicas de bioinformática. Obrigada por sempre estar disponível e por nunca achar que minhas indagações eram estranhas.

Agradeço a todos os colaboradores do projeto, incluindo: Laboratório Nacional de Computação Científica (Ana Tereza, Alexandra, Guilherme Loss), Instituto Pasteur de Montevideu (Luiza Berná e Carlos Robello) e Jody Phelan. As colaboradoras do laboratório, Susana Barbosa e Priscila Thihara, obrigada por sempre estarem de prontidão para as minhas dúvidas no mundo do R e na revisão do trabalho.

Agradeço a Zezé e a Dani, técnicas do laboratório, que por muitas vezes me salvaram com meus documentos e por sempre me deixarem menos estressada com as dúvidas em relação a pós-graduação, aos intercâmbios e as bolsas.

Agradeço a todas as meninas do meu laboratório: Zezé, Dani, Van, Nath, Raquel, Ana, Lais, Pri, Su e Si (e aos meninos também, Pablo e Jaques) por criarem uma ambiente tão gostoso de trabalho, repleto de conversas, risadas e, é claro, por me darem ânimo e comida.

Aos meus amigos, Tha, Leo, Su e Dan que sempre que possível me traziam à normalidade. Ao Thales, com carinho, agradeço pelas horas intermináveis que passamos juntos em outros trabalhos e pela dedicação para me ajudar a encontrar meu atual orientador.

A Aninha, Clarissa, ConCon e Zezinho que por vezes me viram chorar e me viram sorrir. Obrigada por fazerem parte da minha vida desde sempre e pelas diversas conversas que muitas vezes me fizeram entender o rumo da vida.

Agradeço aos meus pais, Sonia e Isano, ao meu irmão Vinicius, a minha avó Daisy e a

minha tia Deise, que por vezes me ajudaram simplesmente falando: "Você vai conseguir! Tenha fé." Vocês são e sempre vão ser meu pilar de sustentação e se hoje estou aonde estou, devo isso a vocês. Amo vocês.

Por fim, agradeço a Deus, aos amigos deste e de outros mundos que me fizeram ser quem sou.

"Os limites da minha linguagem são os limites do meu mundo."
(Ludwig Wittgenstein)

"Vai. E se der medo, vai com medo mesmo."
(Autor desconhecido)

RESUMO

CRIPPA, T. O. Genômica populacional de *Plasmodium vivax*: níveis e mecanismos de diversidade genética na América. 2016. 95 f. Dissertação (Mestrado em Parasitologia) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2016.

Malária é um problema de saúde pública global. No Brasil, *Plasmodium vivax* é o causador de 85% dos 142 mil casos de malária relatados em 2013. Estudos genômicos tem o potencial de complementar os estudos *in vitro*, provendo novas oportunidades para o descobrimento de alvos para vacinas e drogas, além de auxiliar em testes de expressão de genes. Deste modo, este projeto teve como objetivos, a partir do sequenciamento de 9 genomas nucleares de isolados simpátricos brasileiros de *P. vivax* coletados entre 2012 e 2013, avaliar: (a) os níveis e os potenciais mecanismos de geração de diversidade genética utilizando polimorfismos de base única (SNPs), (b) os níveis de estrutura populacional na população brasileira simpátrica em comparação com outras regiões da América, (c) os *loci* sob pressão seletiva. Utilizamos técnicas de sequenciamento de nova geração associadas à identificação de marcadores moleculares do tipo SNPs e análises subsequentes de diversidade, estrutura populacional e de evidências de seleção natural. Nossos resultados mostraram a população do Brasil compartilhando mais ancestrais com a do Peru, bem como a da Colômbia compartilhando mais ancestrais com a do México. Genes de famílias previamente descritas como hipervariáveis foram observados com os maiores valores de diversidade; alguns genes envolvidos com a interação parasito-hospedeiro apresentaram evidência de seleção balanceada, a partir do valor D de Tajima. A alta frequência de recombinação meiótica encontrada em amostras das populações de parasitos, apesar da baixa transmissão local de malária, resultou no declínio do desequilíbrio de ligação entre pares de SNPs situados em uma distância de 50 pares de bases ao longo do mesmo cromossomo. Os dados apresentados corroboram com dados prévios de análises com uso de SNPs, complementando as informações sobre diversidade em escala genômica de populações de *P. vivax* nas Américas.

Palavras-chaves: *Plasmodium vivax*. Malária. SNP. Diversidade genética.

ABSTRACT

CRIPPA, T. O. Population Genomics of *Plasmodium vivax*: levels and mechanisms of genetic diversity in America. 2016. 95 p. Master thesis (Parasitology) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2016.

Malaria is a public health problem worldwide. In Brazil, *Plasmodium vivax* is responsible for 85% of the 142 thousand malaria cases reported in 2013. Genomic studies complement *in vitro* experiments, providing new opportunities for the discovery of targets for vaccines and drugs, as well as enabling gene expression tests. Our objectives for this project, using genome sequences from 9 *P. vivax* Brazilian sympatric isolates collected between 2012 and 2013, were to evaluate the: a) genetic diversity levels and potential generating mechanisms using single nucleotide polymorphisms (SNPs); b) population structure levels in the Brazilian population in comparison with other regions in America; c) *loci* under selective pressure. We used new generation sequencing techniques in association with the identification of SNPs as molecular markers, and subsequent analysis of diversity, population structure and natural selection. Our results show the Brazilian population sharing more ancestrals with Peru, and in addition Colombia sharing more ancestrals with Mexico. Families of genes previously described as hyper variable were observed to have the same diversity values and some genes involved in the host-parasite interaction presented evidences of positive selection, considering Tagima's D. The high meiotic recombination frequency found in the parasite population samples, despite local low transmission, produced a decrease in the linkage disequilibrium between SNPs pairs located in the same chromosome at a 50 base pairs distance. The results presented here corroborate previous results using SNPs analysis, complementing knowledge on diversity at a genomic scale of *P. vivax* in the Americas.

Keywords: *Plasmodium vivax*. Malaria. SNP. Genetic diversity.

LISTA DE FIGURAS

Figura 1 - Mapa da distribuição espacial da endemicidade de <i>Plasmodium vivax</i> no mundo em 2010	15
Figura 2 - Ciclo de vida do gênero <i>Plasmodium</i> que infectam humanos	18
Figura 3 - Mapa da distribuição do local de coleta das amostras da América utilizadas neste estudo	25
Figura 4 - Alelos compartilhados entre populações de <i>P. vivax</i> nas Américas	35
Figura 5 - Estrutura populacional baseada em SNPs de 84 amostras de <i>P. vivax</i> da América.	37
Figura 6 - Análise de Componentes Principais com três <i>eigenvectors</i> para o <i>barcode</i> de 37 SNPs	39
Figura 7 - Diversidade de nucleotídeos encontrada nos cromossomos 4, 7, 10 e 12, nas 84 amostras de <i>P. vivax</i> nas 4 populações das Américas.	41
Figura 8 - Valores de r^2 calculados entre pares de SNPs entre distâncias definidas a partir de dados de SNPs por região.	56
Figura S1 - Estrutura populacional baseada em SNPs de 86 amostras de amostras de <i>P.vivax</i> nas Américas	75
Figura S2 - Valores de <i>clusters</i> populacionais do ADMIXTURE obtidos segundo o menor valor de erro encontrado para K durante corridas de entre 2 e 10, com 84 amostras das quatro regiões da América (Brasil, Colômbia, México e Peru)	76
Figura S3 - Análises de ADMIXTURE com o banco de dados com K=2 e K=3 com 84 amostras de <i>P. vivax</i> nas Américas.	77

LISTA DE TABELAS

Tabela 1 – Genomas de <i>P. vivax</i> disponíveis nos bancos de dados para países do continente americano	23
Tabela 2 – Amostras brasileiras de <i>P. vivax</i> para sequenciamento genômico nuclear	24
Tabela 3 – Genomas nucleares de <i>P. vivax</i> das regiões do Brasil, Colômbia, México e Peru que passaram pelos testes de qualidade desse estudo, com seus respectivos: local de coleta da amostra, alinhamento com Sal-I (referência) e quantidade de SNPs encontrados para cada amostra	31
Tabela 4 – Valores de <i>FST</i> para cada par de regiões da América	38
Tabela 5 – 50 maiores valores de diversidade nucleotídica (π) encontrada nas quatro regiões da América (Brasil, Colômbia, México e Peru) divididos pelos cromossomos. Genes associados as regiões, quando não em regiões intergênicas, foram identificados	43
Tabela 6 – Funções do genes associados aos maiores e menores valores de D de Tajima para cada uma das regiões da América (Brasil, Colômbia, México e Peru). As funções foram separadas em três grandes grupos: interação parasita-hospedeiro, regulação do ciclo de vida e fertilidade. Foram apenas estabelecidas as funções de genes que estivessem descritas em alguma espécie de <i>Plasmodium</i>	52
Tabela S1 – Amostras americanas com seus números de acesso e filtro para as análises.	78
Tabela S2 – 50 maiores valores de D de Tajima encontrados nas quatro regiões da América (Brasil, Colômbia, México e Peru) divididos pelos cromossomos	81
Tabela S3 – 50 menores valores de D de Tajima encontrados nas quatro regiões da América (Brasil, Colômbia, México e Peru) divididos pelos cromossomos	88

LISTA DE ABREVIATURAS E SIGLAS

DNA	<i>Deoxyribonucleic acid</i>
DL	<i>Desequilíbrio de ligação</i>
IPA	<i>Índice parasitário Anual</i>
kb	<i>Kilo pares de bases</i>
MAF	<i>Minore allele frequency</i>
Mpb	<i>Mega pares de bases</i>
nsSNP	<i>SNPs não-sinônimos</i>
pb	<i>Pares de bases</i>
RNA	<i>Ribonucleic acid</i>
Sal-I	<i>Salvador I</i>
SNP	<i>Single nucleotide polymorphism</i>
sSNP	<i>SNPs sinônimos</i>
WHO	<i>World Health Organization</i>

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Malária	14
1.2 Malária nas Américas	14
1.3 Ciclo de vida e tratamentos	16
1.4 Diversidade genética em <i>P. vivax</i>	18
1.5 Genômica de <i>P. vivax</i>	21
2 MATERIAL E MÉTODOS	23
2.1 Genomas de <i>P. vivax</i> disponíveis nos bancos de dados	23
2.2 Amostras de parasitos brasileiros para sequenciamento genômico nuclear.	23
2.3 Preparo das amostras dos genomas	25
2.4 Sequenciamento por <i>Ion Torrent</i> e <i>Ion Proton</i>	26
2.5 Análises de polimorfismos (SNPs)	27
2.6 Análise de dados	28
2.6.1 Estrutura populacional	28
2.6.2 Diversidade e seleção	28
2.7 Análise de recombinação	29
3 RESULTADOS	30
3.1 Análise de 9 genomas completos de <i>P. vivax</i> do Brasil	30
3.2 Análise comparativa de amostras das Américas	30
3.2.1 Estrutura populacional em diferentes regiões da América	36
3.2.2 Diversidade nucleotídica e seleção em regiões dos genomas da América	39
3.3 Declínio do desequilíbrio de ligação em diferentes populações	55
4 DISCUSSÃO	57
4.1 Estrutura Populacional de <i>P. vivax</i> na América	59
4.2 Diversidade e seleção em populações de <i>P. vivax</i> americanas	60
4.3 Desequilíbrio de ligação	61
5 CONCLUSÃO	62
REFERÊNCIAS	63
ANEXOS	74

1 INTRODUÇÃO

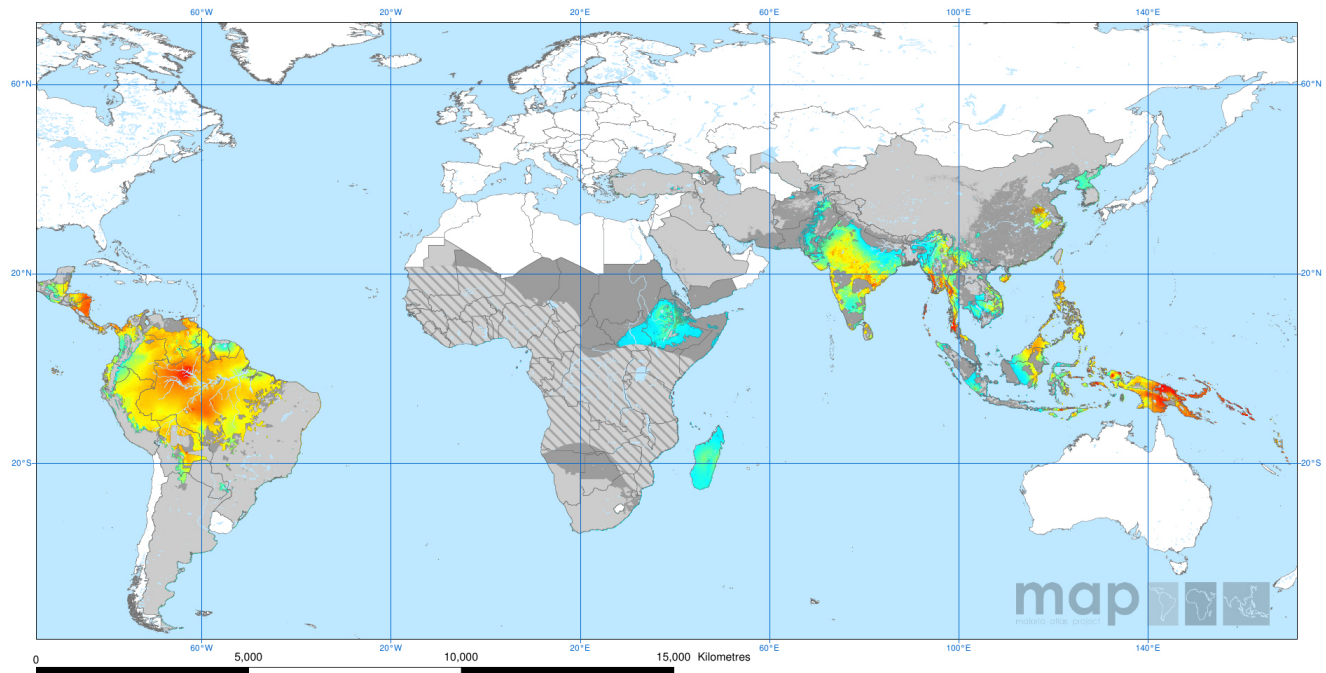
1.1 Malária

Malária é um problema de saúde pública global e a terceira principal causa de morte por doenças infecciosas em crianças menores que cinco anos em todo o mundo (BLACK et al., 2010) (BLACK et al., 2010). A malária é uma doença causada por parasitos apicomplexos do gênero *Plasmodium*, transmitidos pela picada da fêmea dos mosquitos do gênero *Anopheles*. Há cinco agentes etiológicos causadores de infecção em seres humanos: *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* e *P. knowlesi* (World Health Organization - WHO -, 2014). Embora as metas de eliminação abordem todas as malárias humanas, a alocação de fundos e recursos está em dois parasitos de maior importância, *P. falciparum* e *P. vivax*. Essas duas espécies são epidemiológica e biologicamente diferentes, portanto, não é possível supor que os métodos de controle desenvolvidos para a malária falciparum sejam transferíveis para malária vivax (BAIRD, 2007; BOCKARIE; DAGORO, 2006; BOUSEMA; DRAKELEY, 2011; LUXEMBURGER et al., 1994). Evidências de infecções por *P. vivax* em populações cuidadosamente monitoradas mostram que a malária vivax não deve mais ser pensada como uma doença benigna e raramente fatal, mas que pode ser grave e letal (ANSTEY et al., 2009; BAIRD, 2007; PRICE et al., 2007)

1.2 Malária nas Américas

P. vivax é o principal causador de malária fora da África, principalmente atingindo a Ásia e as Américas, sendo ausente em regiões da África Central e Ocidental. Nas Américas os casos ocorrem normalmente em áreas de densidade populacional muito baixa, enquanto que o fenômeno oposto é encontrado na Ásia, com alta densidade populacional (BATTLE et al., 2012) (Figura 1). Na última década, a incidência de malária vem diminuindo na América do Norte (México), Central e do Sul. No final do ano de 2011, 12 dos 21 países endêmicos para malária na América chegaram ao objetivo traçado para 2015 (redução de 75% da incidência de malária quando comparado ao ano de 2000), enquanto outros 5 países estão próximos dessa meta (WHO, 2012). As conquistas recentes são consideradas um reflexo da vontade política e empenho aos níveis nacional e global acompanhada do aumento do financiamento para a malária e do controle de doenças transmitidas por vetores.

Figura 1 – Mapa de distribuição espacial da endemidade de *Plasmodium vivax* no mundo em 2010



Cores tendendo para vermelho representam uma maior endemidade. Fonte: Adaptado de Malaria Atlas Project (MAP, 2016)

México, El Salvador e Costa Rica, três países em fase de pré-eliminação de malária, vem mantendo reduções da incidência de malária por mais de duas décadas (CARTER et al., 2015). Responsável por 99% dos casos de malária no México, *P. vivax*, em 1980, atingia mais de 100.000 hospedeiros humanos. Devido aos esforços no controle intensivo de malária realizadas pelo Programa Nacional de Controle da Malária na década de 1990, o número de casos tem diminuído desde então (GONZALEZ-CERON et al., 2013). A transmissão da malária nesse país é geralmente baixa e sazonal, e está concentrada em focos residuais ao longo da costa do Pacífico. Nos últimos 10 anos, aproximadamente 80% dos casos de malária relatados no México provêm das províncias de Oaxaca e Chiapas (BETANZOS, 2011). Estes focos de malária são vulneráveis a eventos climáticos severos, como evidenciado por surtos de malária ocorridos coincidente com *el Niño* e o furacão Paulina em 1998, em Oaxaca e Chiapas, e do furacão Stan, no final de 2005 no sul do Chiapas.

Na América Central e do Sul, República Dominicana, Haiti and Venezuela mostraram aumento da incidência de malária na última década e países que compartilham a Bacia Amazônica apresentaram declínio (CARTER et al., 2015), tais como a Colômbia, Peru e Brasil. Apesar do declínio, o Peru é um dos países da América Latina com maior valor de endemidade

em malária, com impacto negativo na saúde e no desenvolvimento econômico (WHO, 2014). A maioria dos casos de malária (76% de 64,673) foram reportados na área da bacia Amazônia (região de Loreto) em 2014, sendo 83% de *P. vivax* (CHAPILLIQUÉN, 2015). A maior parte das infecções por *P. vivax* eram assintomáticas e não detectáveis por microscópica, provendo uma potencial importância na transmissão local (BRANCH et al., 2005; DELGADO-RATTO et al., 2014; GRIFFING; GAMBOA; UDHAYAKUMAR, 2013). Além disso, infecções múltiplas ocorreram mesmo depois da administração de tratamentos indicados (primaquina e cloroquina) pela *World Health Organization* (WHO) (DELGADO-RATTO et al., 2014; GRAF et al., 2012).

A malária na Colômbia, vem experimentando declínio entre os anos de 2006 a 2011, que pode ser explicado por esforços para reduzir a carga da doença, gravidade e mortalidade por uso extensivo de mosquiteiros com inseticidas, diagnóstico precoce e tratamento imediato (WHO, 2014). Entre 1990 e 2011, houve predominância de *P. vivax* sobre *P. falciparum* nos 2.964.818 casos, com média anual de 134.764; picos de índice parasitário anual (IPA) em 1994, nas cidade de Guainía e Guaviare, e em 1998, nas cidades de Guaviare e Chocó (RAMIREZ et al., 2014).

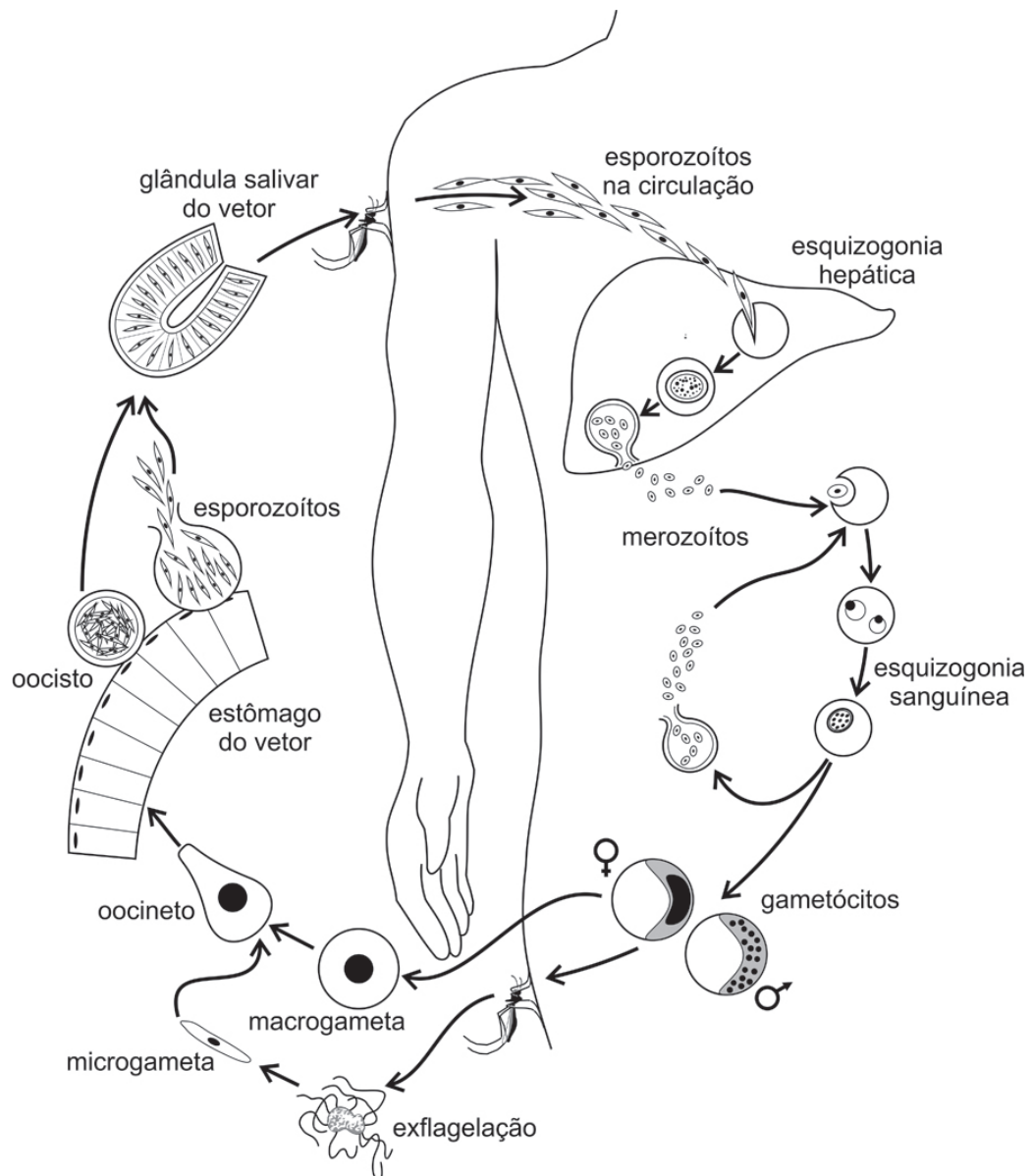
O Brasil também reportou decréscimo na incidência de malária (76.8%) nos anos de 2000 a 2014, ainda assim atualmente contribui com 42% de todos os casos reportados de malária na América (FERREIRA; CASTRO, 2016). A transmissão se mantém basicamente associada a bacia Amazônica, que contribui com 99.5% dos casos de malária do país (FERREIRA; CASTRO, 2016). No Brasil, ambos *P. falciparum* e *P. vivax* são transmitidos através da Bacia Amazônica, com infecções raras por *P. malariae*. A incidência relativa das duas espécies principais de *Plasmodium* transmitidos no Brasil era de cerca de 50% cada, em 1988. Isso mudou a partir de 1990 (quando 44,3% dos casos foram devidos ao *P. falciparum*) atingindo a situação em 2009, quando *P. vivax* tornou-se a espécie predominante (apenas 16,3% dos casos devido à *P. falciparum*) (BARBOSA et al., 2014). *P. vivax* atualmente é o causador de 85% dos 142 mil casos de malária relatados neste país em 2014 (WHO, 2014). Essa recente predominância de *P. vivax* no Brasil pode ser parcialmente explicado por características biológicas que tornam essa espécie menos responsiva que *P. falciparum* para estratégias de controle baseadas somente no diagnóstico precoce e tratamento imediato de infecções de estágio-sanguíneo. *P. vivax*, apresenta a forma de hipnozoito que pode persistir no hospedeiro humano durante vários meses, estágio do parasito no fígado em que ele fica dormente e eventualmente pode causar recaídas (WHITE; IMWONG, 2012).

1.3 Ciclo de vida e tratamentos

A transmissão do parasita ocorre através de um ciclo de vida complexo (Figura 2) que se inicia com o repasto sanguíneo da fêmea do vetor *Anopheles*, a qual inocula no hospedeiro, humano ou símio, centenas de formas esporozoítas. Essas formas atravessam a pele, alcançando

a corrente sanguínea, o que possibilita a sua migração ao fígado. Após um processo de replicação mitótica assexuada denominado esquizogônia, os hepatócitos se rompem, liberando milhares de merozoítos novamente para a corrente sanguínea. No momento em que essas formas invadem os eritrócitos, inicia-se o ciclo intra-eritrocítico que se mantém por uma constante reprodução assexuada, acompanhada da diferenciação para formas intra-eritrocíticas: anel, trofozoito maduro e esquizontes respectivamente. Neste momento, os eritrócitos são rompidos e os merozoítos jovens infectam novas hemácias. Durante esses ciclos de infecção, uma parcela dos merozoítos originam as formas sexuadas: gametócitos feminino e masculino, que podem ser aspiradas pelo vetor, quando este pica um hospedeiro infectado. No intestino do mosquito os gametas se diferenciam em o microgameta (masculino) e macrogameta (feminino) que se fundem gerando a forma móvel chamada de oocineto, permitindo a recombinação entre cromossomos homólogos. O oocineto após sua maturação na membrana basal origina o oocisto, que se rompe liberando milhares de esporozoítos, que então migram para o epitélio da glândula salivar maturando para a forma infectiva e possibilitando assim o recomeço do ciclo.

Figura 2 – Ciclo de vida do gênero *Plasmodium* que infectam humanos



Fonte: Adaptado de *National Institute of Allergy and Infectious Diseases - NIAID*

1.4 Diversidade genética em *P. vivax*

Conhecer os mecanismos que geram variação genética em populações de parasitos é essencial para compreender a origem e dispersão de fenótipos de interesse. A recente emergência de resistência à cloroquina em *P. vivax* e a ocorrência de casos graves, eventualmente fatais, de malária vivax sugerem a aquisição de novas características fenotípicas por parasitos de diferentes regiões geográficas (PRICE; DOUGLAS; ANSTEY, 2009). No Brasil, tanto a ocorrência de resistência à cloroquina (SANTANA-FILHO et al., 2007; MARQUES et al., 2014) como a de

doença grave (ANDRADE et al., 2010) estão bem caracterizadas. Os conhecimentos atuais sobre a biologia populacional do parasito não permitem prever padrões de emergência e dispersão desses fenótipos. De fato, nem sequer se conhece a contribuição relativa de mecanismos dependentes de seleção natural e de processos evolutivamente neutros (como a deriva gênica ou a migração) para a dispersão espacial e a dinâmica temporal de polimorfismos de relevância clínica ou em saúde pública.

Um dos problemas chave que medeiam o controle efetivo da malária brasileira é a falta de compreensão na estrutura populacional de *P. vivax* e sua dinâmica de transmissão. Desta forma, mais estudos de diversidade genética são necessários para se definir a diversidade, distribuição e dinâmica das populações de *P. vivax* e como as populações de parasitos se diferem entre as localidades (HUPALO et al., 2016; JOSHI et al., 2008; JOY et al., 2008; PEARSON et al., 2016; REID et al., 2010; SCHULTZ et al., 2010). O mapeamento global e local das populações de *P. vivax* é essencial para o estabelecimento de metas para eliminação e intervenção. Portanto, estudar a história demográfica dos parasitos de malária é útil por dois motivos principais. Primeiro, inferências sobre tempo e magnitude nas mudanças do tamanho populacional podem prover pistas das causas dessas mudanças. Em segundo, a influência da história demográfica ou da subestrutura populacional na variação do genoma devem ser consideradas na identificação de regiões sob pressão de seleção e de alelos de resistência a drogas (CHANG et al., 2012).

Estudar uma população local em profundidade possui duas vantagens sobre um estudo análogo de amostras de todo o mundo: (a) os polimorfismos em uma população local são melhores para revelar assinaturas de seleção recente e (b) a existência da estrutura populacional em amostras globais pode resultar em padrões de desequilíbrio de ligação e de espectro de frequência de alelos que artificialmente se assemelham aos esperados de seleção. Estudos de subestrutura em diferentes populações mundiais acometidas por malária vivax mostraram uma mesma face: a existência de pelo menos duas subpopulações dentro de uma mesma população (ABDULLAH et al., 2013; CHENET et al., 2012; DELGADO-RATTO et al., 2014; GUNAWARDENA et al., 2014; JOY et al., 2008; LARRAÑAGA et al., 2013; LIU et al., 2014a). Assim, a comparação de genomas obtidos de amostras simpátricas de parasitos é uma estratégia extremamente poderosa para avaliar a importância relativa de diferentes mecanismos (mutações pontuais, inserções e deleções, proliferação de motivos repetitivos e recombinação meiótica) na geração de diversidade genética em *P. vivax*. Além disso, essa análise permite inferir a ocorrência de pressões seletivas (diversificadora e purificadora) em diferentes regiões genômicas, que afetam os padrões de variação observados em nível populacional.

O maior obstáculo, então, para esses estudos de genética de populações de *P. vivax* foi, até recentemente, a falta de marcadores moleculares adequados para uso em larga escala. Microsatélites, por exemplo, são marcadores úteis sob certas circunstâncias, pois eles permitem a detecção de estruturas genéticas em divergências recentes (PAYSEUR; JING; HAASL, 2011)

e eles são abundantes tanto em genomas de *P. falciparum* quanto em *P. vivax* (RUSSELL; SUWANARUSK; LEK-UTHAI, 2006). Níveis altos de diversidade genética em *P. vivax*, com uso de microssatélites, vem sido demonstrado em países como Papua Nova Guiné, Tailândia, Laos, Índia, Myanmar, Sri Lanka, Honduras, Brasil, Colômbia, Peru e México (FERREIRA et al., 2007; GUNAWARDENA et al., 2014; KARUNAWEERA et al., 2008; MENEGON et al., 2014; MOBEGI et al., 2012; REZENDE et al., 2010; SUTTON, 2013). No caso do Brasil, Sri Lanka e Myanmar, foi observada a alta diversidade genética mesmo com baixos níveis de transmissão (FERREIRA et al., 2007; GUNAWARDENA et al., 2014). Gunawardena et al. (2010) observou, usando 12 microssatélites (marcadores tri e tetranucleotídeos) em 425 isolados de Sri Lanka, Myanmar e Etiópia, que há a clusterização de acordo com a origem geográfica. Esse resultado é congruente com os de Karunaweera et al. (2008) em que isolados da América Central, África, Sudeste Asiático e Oceania apresentaram uma fração de semelhança entre os *clusters* para os 14 marcadores de microssatélites utilizados. Neste mesmo estudo, em isolados de mesma área no Brasil (Acre), no entanto, foi possível observar subestrutura populacional. Assim, a maior diferenciação genética foi encontrada no Brasil, ao longo de um período de 5 a 6 anos. Na Amazônia peruana, as poucas observações disponíveis em genética de populações *P. vivax* foram coletadas em pequenas áreas (dispersas aldeias e comunidades) e relatado que as populações de parasitas eram heterogêneas e clonais. O uso em larga escala de microssatélites, porém, tem limitações. Os métodos de tipagem usados atualmente estão sujeitos a variação de resultados entre laboratórios mesmo com um conjunto de marcadores iguais, reduzindo as possibilidades de comparação entre resultados obtidos.

Em contrapartida, SNPs (*single nucleotide polymorphism*) vem tendo uma considerável atenção por causa do seu potencial como marcadores para mapeamento genético e para estudo de evolução molecular e dinâmica populacional em *P. vivax* e em espécies aparentadas (BRIGHT et al., 2012; CHAN et al., 2012; DHARIA et al., 2010; FENG et al., 2003; NEAFSEY et al., 2008; NEAFSEY et al., 2012; ORJUELA-SÁNCHEZ et al., 2010; SACHIDANANDAM et al., 2001) pois são relativamente fáceis de se manejar e se apresentam com grande frequência. Estudo realizado por Orjuela-Sanchez et al. (2010), baseado em um conjunto de 85 SNPs ao longo do cromossomo 8 de *P. vivax* provenientes de isolados de campo do Brasil, Camboja, Vietnã e Sri Lanka, mostrou uma clusterização dessas amostras em três populações distintas: uma formada por isolados do Camboja, Vietnã e Sri Lanka e outras duas de isolados do Brasil. Alguns trabalhos vem propondo SNPs significativos, como Baniecki et al. (2015) que propuseram 42 SNPs, testados em 87 amostras de países da América, África e Ásia, que poderiam ser utilizados como *barcode* para *P. vivax*, utilizando técnica similar de geração de *barcode* de SNP usada por este grupo em *P. falciparum* (DANIELS et al., 2008).

1.5 Genômica de *P. vivax*

A conclusão do sequenciamento de um genoma de referência de *P. vivax* (CARLTON et al., 2008), gerado a partir do isolado Salvador I (paciente de El Salvador (Sal-I)), representou um avanço crucial para os estudos de diversidade genética dessa espécie. Em muitos aspectos, os genomas das espécies de *Plasmodium* (*P. falciparum*, *P. knowlesi*, *P. vivax* e *P. y. yoelii*) são uniformes, tendo cerca de 23 a 27 Mpb entre os 14 cromossomos e contendo ~5500 genes, com a maioria (~51%) contendo pelo menos 1 intron. O genoma de ~26,8 Mpb de *P. vivax*, consiste em 14 cromossomos em tamanhos que variam de 0,76 a 3,12 Mpb e 2733 pequenos *contigs* de tamanhos 200-101.928 bases, com conteúdo G+C de ~42,3%; 4,3 Mb de pequenos *contigs* nas regiões subteloméricas ainda permanecem não especificados devido a sua natureza repetitiva (CARLTON et al., 2008).

Atualmente, há cinco genomas completamente montados e anotados (NEAFSEY et al., 2012) e outros com disponibilização apenas dos *reads* (BRIGHT et al., 2012; CHAN et al., 2012; CHAN et al., 2015; DHARIA et al., 2010; FENG et al., 2003; HUPALO et al., 2016; NEAFSEY et al., 2008; NEAFSEY et al., 2012; ORJUELA-SÁNCHEZ et al., 2010; PEARSON et al., 2016; SACHIDANANDAM et al., 2001), que fornecem uma base de dados para estudos de variação genética de *P. vivax* em escala genômica. A primeira análise comparativa de genomas completos de *P. vivax* (NEAFSEY et al., 2012) com uso de SNPs revelou uma característica extremamente interessante dessa espécie: para todos os tipos de polimorfismos estudados (regiões codificadoras e não-codificadoras, polimorfismos sinônimos e não-sinônimos, regiões repetitivas e não-repetitivas), os isolados analisados de *P. vivax* apresentaram muito mais diversidade genética do que um conjunto de isolados de *P. falciparum* provenientes das mesmas regiões geográficas. Uma análise comparativa de diversidade de microssatélites em amostras de ambas as espécies provenientes do Camboja havia chegado à mesma conclusão (ORJUELA-SÁNCHEZ et al., 2013). Não se conhecem, entretanto, os mecanismos responsáveis por essa grande diversidade genética em *P. vivax*.

Este ano, dois artigos publicaram 228 (PEARSON et al., 2016) e 117 (HUPALO et al., 2016) genomas completos de isolados de regiões do Sudeste Asiático, Índia, Oceania e América, sendo 303,616 SNPs e 134,800 SNPs de alta qualidade, respectivamente. Desses genomas, 25 são provenientes de amostras da região noroeste do Brasil, contudo nesses trabalhos os genomas apresentaram baixa cobertura. Ambos os trabalhos observaram estruturas populacionais primeiramente determinadas pela localização das amostras, sendo esses resultados condizentes com resultados prévios de microssatélites e estudos de SNPs (CARLTON et al., 2008; KOEPFLI et al., 2015; NEAFSEY et al., 2012; SCHOUSBOE et al., 2014), porém promoveram uma resolução maior dado a quantidade de informações agregadas. Amostras de Velho Mundo separaram-se de amostras de Novo Mundo (HUPALO et al., 2016), sendo que a Ásia apresentou de modo geral *clusters* populacionais por país (HUPALO et al., 2016; PEARSON et al., 2016), excluindo-se a Índia, que apresentou uma população composta por ancestrais

distintos (HUPALO et al., 2016). No Novo Mundo, amostras americanas apresentaram presença de 3 ancestrais distintos, claramente diferentes dos encontrados no Velho Mundo, sendo que o Brasil mostrou ser mais miscigenado, mesmo com amostras provindas de mesmo localidade (HUPALO et al., 2016).

Além disso, evidenciaram fortes sinais de seleção ocasionada por evolução recente tanto em genes de resistência a drogas, como DHPS (*dihydropteroate synthase*) e DHFR-TS (*dihydrofolate reductase-thymidylate synthase*) (HUPALO et al., 2016; PEARSON et al., 2016), quanto em novos *loci*, como MAEBL (*merozoite adhesive erythrocytic binding protein*) (Hupalo, et al, 2016). Huppalo et al. (2016) observaram que esses genes sob seleção estavam adaptados as diferenças regionais nos hospedeiros humano, mosquito e a outras variáveis ambientais. De modo geral, esses dados reforçam a epidemiologia variável desses parasitas e destacam a amplitude das abordagens que poderão ser necessárias para a eliminação global de *P. vivax*.

Assim, estudos genômicos tem o potencial para complementar os estudos *in vitro*, provendo novas oportunidades para o descobrimento de alvos para vacinas e drogas e, auxílio em testes de expressão de genes (BOZDECH et al., 2008; WESTENBERGER et al., 2010), preenchendo assim, algumas das lacunas em nossa compreensão da biologia *P. vivax*. Os novos genomas brasileiros adicionam informações importantes ao banco de dados atual melhorando o panorama sobre a variação genética em diferentes regiões genômicas e em diferentes tipos de sequências (codificadoras e não-codificadoras). A disponibilidade de genomas adicionais em regiões brasileiras permite ampliar a base de comparação com dados de outras origens geográficas, elucidando padrões de divergência genética em cada população e entre as diversas populações já analisadas na América.

Neste projeto foram analisados 9 genomas completos de isolados simpátricos de *P. vivax* da Amazônia brasileira e genomas completos das Américas já disponíveis no bando de dados, avaliando: (a) os níveis e os potenciais mecanismos de geração de diversidade genética, como mutações ou recombinações; (b) os níveis de diferenciação entre a população simpátrica brasileira e outras regiões das Américas e (c) os *loci* sob pressão seletiva.

2 MATERIAL E MÉTODOS

2.1 Genomas de *P. vivax* disponíveis nos bancos de dados

Todas as seguintes amostras foram sequenciadas via tecnologia *Illumina paired end* (*reads* com tamanhos ~ 75 pb) (Tabela 1) (Figura 3). Belem e Brasill, foram adquiridas através do banco de dados SRA (SRA, 2016), sendo amostras mantidas em macacos dos gêneros *Aotus* e *Saimiri*. Amostra PV4 foi adquirida através do banco de dados MalariaGen (MalariaGen, 2016) (PEARSON et al., 2016). Outras amostras do Brasil e de outras localidades da América (Colômbia, México e Peru), foram adquiridas do banco de dados SRA (HUPALO et al., 2016). Nessas amostras, a seleção híbrida foi usada para capturar o DNA de *P. vivax* de amostras de sangue humano sem utilização da depleção de leucócitos. O DNA genômico da referência Sal-I foi usado como modelo para a criação de pequenos fragmentos, utilizando a transcrição *in vitro*, que foram usados para hibridização com o DNA de *P. vivax* das amostras. Dessa forma apenas os fragmentos de *P. vivax* foram extraídos e submetidos ao sequenciamento.

Tabela 1 – Genomas de *P. vivax* disponíveis nos bancos de dados para países do continente americano

Região	Número de amostras	Ano	Referência
Brasil	1 (Belem)	1980	(Chan, 2012)
	1 (Brazil I)	1994	(Nayar et al 1997)
	1 (PV4)	2009	(Pearson, et al, 2016)
	20	2011	(Hupalo, et al, 2016)
Colômbia	31	2012-2013	(Hupalo, et al, 2016)
México	19	2000-2008	(Hupalo, et al, 2016)
Peru	34	2010-2013	(Hupalo, et al, 2016)

2.2 Amostras de parasitos brasileiros para sequenciamento genômico nuclear

As amostras de *P. vivax* no Brasil foram colhidas entre novembro de 2012 e junho de 2013, em zonas rurais com baixa transmissão de malária: Acrelândia, com 8 amostras e Remansinho, com 1 amostra (Tabela 2). Todas as amostras foram obtidas com consentimento do paciente e guardiões em estudo prévio.

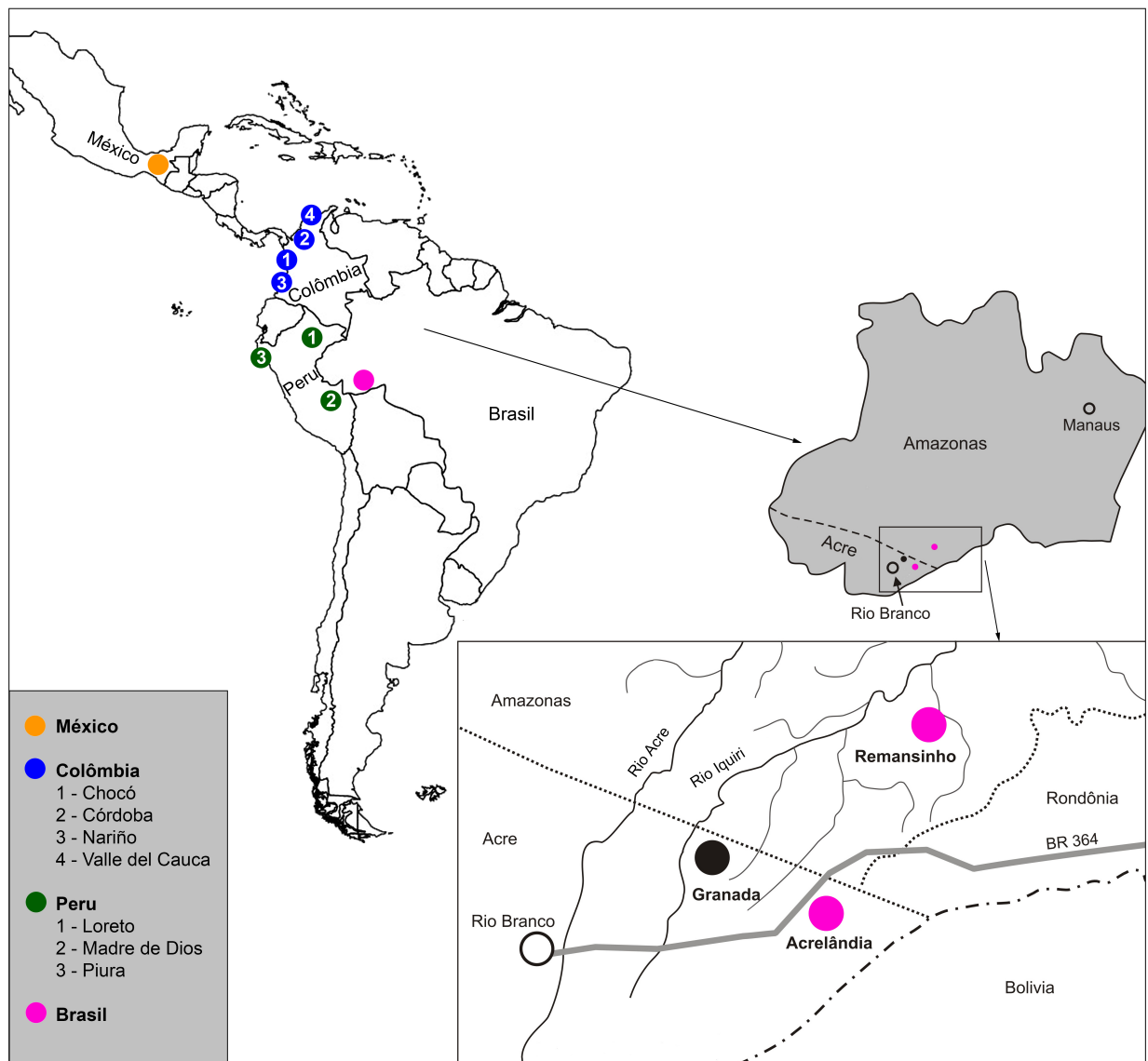
Acrelândia está situada entre os rios Abunã e Iquiri, no Vale do Rio Acre. Situa-se a 112 km a leste de Rio Branco (capital do estado do Acre) (3), apresentando IPA < 50/mil

Tabela 2 – Amostras brasileiras de *P. vivax* para sequenciamento genômico nuclear

Amostra	Data da coleta	Local
PV17	jun/13	Acrelândia
PV18	dez/12	Acrelândia
PV19	nov/12	Acrelândia
PV20	abr/13	Acrelândia
PV32	jun/13	Acrelândia
PV51	jun/13	Acrelândia
PV52	jun/13	Acrelândia
PV207	abr/13	Remansinho
PVACR	nov/12	Acrelândia

habitantes, em 2010 (Ministério da Saúde, dados não publicados, 2011). Remansinho situa-se no extremo oeste do estado de Rondônia e no sul do estado do Amazonas, junto à fronteira com o Acre, a cerca de 120 km de Acrelândia (sede de município mais próxima) (Figura 3). De 2010 a 2013 houve uma queda na prevalência de infecções por *P. vivax* de 23.8% para 3.0% (BARBOSA et al., 2014)

Figura 3 – Mapa da distribuição do local de coleta das amostras da América utilizadas neste estudo



Foco no Noroeste brasileiro com a localização do município de Acrelândia (Acre) e do assentamento conhecido como Remansinho (Amazonas) em rosa

2.3 Preparo das amostras dos genomas

Foram colhidos tubos de ACD, na proporção 5:1, de sangue de pacientes infectados por *P. vivax*, preferencialmente com parasitemia igual ou superior a 2-20 parasitos por campo microscópico de grande aumento. O diagnóstico foi confirmado pelo exame microscópico de gotas espessas. O sangue foi passado dos tubos de coleta (ACD) para um tubo estéril e foi cortada a saída da mangueira onde flui o sangue. Cortou-se a entrada da mangueira (que foi

fixada a parede do fluxo). Nesta posição, foi injetado o sangue a ser leucorreduzido, utilizando uma seringa com agulha. Foi repetido este procedimento até que todo o volume de sangue coletado tenha sido filtrado. Após isso, cetrifugaram-se os tubos, desprezando o sobrenadante. Mensurou-se o volume do precipitado restante. Ao *pellet* foi adicionado solução de glicerol na proporção adequada (volume do precipitado restante x 1,66). Os criotubos com as amostras foram mantidos a 80 °C por 24h, contendo álcool isopropílico.

Após o procedimento de leucodepleção da amostra, as hemácias infectadas foram lavadas com 5 mL de meio e centrifugadas por 5 min. Foi desprezado o sobrenadante e estimado o volume do *pellet* com micropipeta. Em seguida, foi adicionado glicerol. As amostras foram congeladas a -80 °C *overnight* e depois transferidas para nitrogênio.

A seguir foi feita a maturação dos esquizontes *in vitro*, seguindo o protocolo adaptado do laboratório do Professor Manoj Duraisingh (*Department of Immunology and Infectious Diseases - Harvard School of Public Health*). Os criotubos com amostras descongeladas foram transportados para banho de gelo, durante 1 minuto a temperatura ambiente e, logo em seguida, durante 1 min a 37 °C. Os tubos foram transferidos para tubo de 50 mL e foi adicionado 1/5 do volume da amostra de NaCl 12%, e deixado em repouso durante um minuto. Em seguida foram adicionados 10 x o volume da amostra inicial de NaCl 1,6% e centrifugados os tubos por 8 min. Foi aspirado o sobrenadante e ressuspendido o *pellet* utilizando meio Mc Coy's incompleto (sem soro). Depois, o *pellet* foi ressuspendido em meio McCoy's completo (25% soro AB inativado) para hematócrito de 2%. Foi aplicada a mistura de gases (5% CO₂, 5% O₂ e 90% N₂) e os parasitos mantidos a 37 °C por 48 h.

2.4 Sequenciamento por *Ion Torrent* e *Ion Proton*

Foram construídas 9 bibliotecas de fragmentos de DNA de acordo com o protocolo *IonXpressTM Plus gDNA Fragment Library Preparation*. Cada biblioteca foi preparada a partir de 1 µg de DNA. Para a fragmentação do DNA utilizou-se o sistema de sonicação Bioruptor® UCD-200 TS. Para bibliotecas com tamanho de leitura de 200 bases, o protocolo recomenda 3 ciclos de 15 minutos de sonicação, entretanto após os 3 ciclos de sonicação verificou-se que o DNA ainda não estava fragmentado na faixa de tamanho adequado, e sendo assim foram feitos mais 2 ciclos de 15 min e para algumas das amostras ainda mais dois ciclos, totalizando 5 ou 7 ciclos de sonicação. Todas as demais etapas de construção das bibliotecas foram feitas de acordo com o protocolo sem modificações. A PCR em emulsão foi feita no *IonOneTouchTM 2 System*, de acordo com o protocolo *Ion PGM Template OT2 200 Kit*. Cada biblioteca amplificada foi sequenciada em um *Ion318TM Chip v2* de acordo com o descrito no *IonPGMTM Sequencing200 Kit v2*. Todos os procedimentos descritos foram realizados pela equipe chefiada pela Dra. Ana Tereza Ribeiro de Vasconcelos do Laboratório Nacional de Computação (LNCC, Petrópolis/RJ).

De modo a assegurar um bom suporte à identificação de polimorfismos sequenciamos todos os genomas até obter um mínimo de 40X de cobertura média.

2.5 Análises de polimorfismos (SNPs)

116 arquivos fastqs das amostras obtidas do sequenciamento supracitado e dos dados do banco de dados foram submetidos a análises de qualidade e de presença de adaptadores. 6 amostras foram excluídas por não apresentarem alta qualidade (*Phred score* > 30). Posteriormente, foram alinhados com a referência Sal-I (CARLTON et al., 2008) usando Bowtie2 versão 2.2.6 (LANGMEAD; SALZBERG, 2012) com parâmetros restritivos. Amostras com mapeamento abaixo de 80% em relação a referência foram excluídas (11 amostras foram excluídas do Brasil e Peru). As duplicatas foram identificadas e marcadas usando a ferramenta Picard versão 2.0.1 MarkDuplicates (PICARD TOOLS, 2016). Os *reads* mapeados foram então processados com Samtools (LI et al., 2009), a fim de apresentar as conversões, classificações e a indexação. O GATK versão 2.0 (MCKENNA et al., 2010) foi usado para o SNPcall a partir do manual de melhores práticas em detecção de variantes. SNPs foram identificados em cada amostra usando individualmente GATK *UnifiedGenotyper* com a opção *Base Alignment Quality* (-baq CALCULATE_AS_NECESSARY) e filtros rigorosos foram aplicados para atingir SNPs de alta confiança com GATK *VariantFiltration*. Os filtros utilizados incluíam mínimo de 20X de cobertura, *mapping quality* de 30, *base quality* de 30, *ReadDepthAndAllelicFractionBySample* mínimo de 1.0 e *HaplotypeScore* máximo de 3,0 (~50%). Os SNPs de elevada confiança foram então genotipados na amostra usando GATK *UnifiedGenotyper*, com parâmetros *default*, excluindo-se: -stand_call_conf 50.0 -stand_emit_conf 10.0 -dcov 200, que foram modificados. SNPs "heterozigotos" (mais de um alelo por locus, indicando a coexistência de clones distintos na mesma amostra) foram convertidos para o genótipo majoritário se a proporção fosse 75:25 ou mais. Utilizamos o SnpEff (CINGOLANI et al., 2012) para anotar cada uma das ocorrências dos polimorfismos em regiões intergênicas e codantes (sinônimas e não-sinônimas) com as anotações de *P. vivax* Sal-I do banco de dados PlasmoDB (PLASMODB, 2016)

Outros filtros foram aplicados para garantir a qualidade das amostras e SNPs. Foi estabelecido limiar de 3000 SNPs para a continuidade do uso da amostra (13 amostras excluídas do Brasil e Peru). Variantes que possuíam frequência de alelos minoritários (MAF) maior que 0.01 e/ou com múltiplos alelos alternativos foram excluídas da análise (total de 5441 alelos em amostras brasileiras), porque em uma única infecção as hipóteses seriam que estes indicam: erro de sequenciamento ou multiplicidade de infecção. Ao total, obtivemos 84 amostras com 94122 SNPs confiáveis não repetitivos.

2.6 Análise de dados

2.6.1 Estrutura populacional

Para a identificação de estruturas genéticas com um grande banco de dados sem a presença de um modelo de genética populacional delimitado utilizamos a análise de componentes principais (PCA) feita com o *software* PLINK (PURCELL et al., 2007). A inferência de uma árvore de *neighbor joining* foi feita utilizando os pacotes *adegenet* e *ape* do R (JOMBART, 2012). Calculamos o índice de fixação (F_{ST}) através VCFtools (DANECEK et al., 2011), a fim de verificar a divergência das populações americanas. Este índice mede a quantidade de variação genética que pode ser explicada pela estrutura populacional, sendo a influência na estrutura populacional provinda da relação entre deriva e fluxo gênico.

Análise de ancestralidade compartilhada foi feita entre todas as amostras disponíveis das Américas que passaram em nossos critérios de qualidade, através do *software* ADMIXTURE (ALEXANDER; LANGE, 2011). ADMIXTURE é um *software* para a estimativa da probabilidade máxima (*maximum likelihood*) de linhagens individuais a partir de conjuntos de dados de SNPs. Ele usa o mesmo modelo estatístico do STRUTURE (PRITCHARD; STEPHENS; DONNELLY, 2000), mas calcula as estimativas muito mais rapidamente usando um algoritmo de otimização numérica rápida. Como o modelo utilizado assume desequilíbrio de ligação (DL) que os pares de SNP sejam independentes entre si, excluímos SNPs em pares que estivessem ligados. Como seria impossível remover todos os SNPs em DL, especialmente em populações recentemente misturadas, usamos uma estratégia para mitigar os efeitos de DL que incluem marcadores separados entre eles por uma certa distância física no cromossomo. Com esse algoritmo, dos 94122 SNPs iniciais, apenas 12762 SNPs não ligados foram usados na análise com os mesmos critérios estabelecidos por Pearson et al. 2016. O número ideal de *clusters* (K) foi determinado com múltiplas corridas com diferentes valores de K (2 - 10), estimando-se o valor mais plausível de K usando a métrica de *cross-validation*; um bom valor de K irá exibir um valor mínimo de erro de *cross-validation* quando comparado com os outros valores de K.

2.6.2 Diversidade e seleção

As análises subsequentes foram feitas excluindo-se as amostras Belem e Brasil. A diversidade de nucleotídeos (π) e o valor de D de Tajima, para identificação de evidência de seleção natural, em médias por janelas de 1 kb, foram calculadas ao longo dos cromossomos, usando VCFtools. Os resultados foram então plotados de forma semelhante, com uso do *software* R, utilizando uma média móvel ao longo de cada cromossomo. π global foi calculado utilizando a média total dos valores de diversidade encontrados para cada par de nucleotídeos, utilizando VCFtools. Para encontrar regiões com maior diversidade foram estabelecidos que os 50 maiores valores encontrados ao longo do genoma para π (*cut off*: Brasil 0.0034; Colômbia

0.0034; Peru 0.0040; México 0.0053) e, maiores (*cut off*: Brasil 2.024; Colômbia 2.078; Peru 1.890; México 2.407) e menores (*cut off*: Brasil -1.096; Colômbia -1.724; Peru -1.622; México -1.376) para D de Tajima .

Classificamos os genes encontrados nas regiões de interesse usando anotações do PlasmoDB para *P. vivax* e genes anotados como hipotéticos tiveram suas funções inferidas através do PFAM (FINN et al., 2016). As funções dos genes de interesse foram averiguadas em artigos associados ao gênero *Plasmodium*.

2.7 Análise de recombinação

Para avaliar a frequência de recombinação meiótica nas populações americanas, utilizamos o decaimento do desequilíbrio de ligação com distâncias genéticas para cada conjunto de amostras separadamente utilizando VCFtools. DL foi medido computando uma medida comum, coeficiente de determinação r^2 , para pares de SNPs com distância variável. A mediana de r^2 foi calculada em janelas de 50 pb ao longo de cada cromossomo. Para estimar os valores de r^2 esperados entre *loci* independentes em cada população (linha de base), utilizamos a mediana de todos os valores de r^2 para todos os pares de SNPs situados em cromossomos diferentes.

3 RESULTADOS

3.1 Análise de 9 genomas completos de *P. vivax* do Brasil

Analizamos 9 amostras (PV17, PV18, PV19, PV20, PV207, PV32, PV51, PV52, PVACR) de sangue contendo *P. vivax* providas de uma região simpátrica da Amazônia rural brasileira. Após a preparação das bibliotecas de forma a evitar os contaminantes providos de DNA humano, sequenciamos cada amostra através de corridas por *Ion Torrent* e *Ion Proton*, sendo no total 4 corridas e seus resultados concatenados para a geração dos dados que serão mostrados. Juntamente com nossas amostras brasileiras, adicionamos genomas já sequenciados e disponíveis no banco de dados de outras amostras americanas.

Mapeamos todos os *reads* no genoma de referência do isolado Sal-I. Filtramos os *reads* de baixa qualidade ($MQ < 30$ - *mapping quality score*) e com cobertura menor que a colocada como limite com o genoma de referência ($DP < 20$ - *depht*). Mais de 91% dos *reads* foram mapeados no genoma de Sal-I (Tabela 3). Os genomas sequenciados por outros autores apresentaram de 18 a 94% de mapeamento contra a referência Sal-I (apenas amostras com mapeamentos acima de 80% com Sal I foram consideradas neste estudo). Em primeira instância, mantivemos especificamente a amostra Brazil I, mesmo com 25% de mapeamento, provavelmente pela alta taxa de contaminação provinda de DNA de símios. Esse parasito foi mantido em algumas análises por ser uma das duas amostras adicionais de regiões distintas brasileiras disponíveis para comparação com aquelas sequenciadas neste trabalho.

3.2 Análise comparativa de amostras das Américas

A princípio foram estabelecidos três banco de dados: (A) Dentre as 116 amostras das Américas (Brasil, Peru, Colômbia e México) (Tabela S1) (Figura 2), restringimos nossas análises a 11 amostras do Brasil, 31 da Colômbia, 23 do Peru e 19 do México, em um total de 84 amostras com 94122 SNPs filtrados e de alta qualidade, utilizado para todas as análises (Tabela 3);(B) com as 84 amostras mais as amostras Brazil I e Belem e (C) com 37 dos 42 SNPs indicados pelo *barcode* de SNPs para *P. vivax* (BANIECKI et al., 2015), utilizados apenas nas análises de estrutura populacional.

Ao todo foram obtidos 94122 SNPs de alta qualidade, para o primeiro banco de dados, sendo destes 55.0% (50690) em regiões intergênicas, 8.8% (8288) em íntrons e 36.2% (33122) em regiões codantes, dos quais 38.79% (12849) são SNPs que geraram mutações sinônimas e 61.20% (20273) não sinônimas (Tabela 3).

Tabela 3 – Genomas nucleares de *P. vivax* das regiões do Brasil, Colômbia, México e Peru que passaram pelos testes de qualidade desse estudo, com seus respectivos: local de coleta da amostra, alinhamento com Sal I (referência) e quantidade de SNPs encontrados para cada amostra

País	Região	Amostra	Alinhamento com Sal-I (%)	Número de SNPs
Brasil	Remansinho (Acre)	17	94.70	6445
		18	93.47	6232
		19	94.74	5524
		20	93.38	6188
		32	91.83	5528
		51	94.01	5819
		52	94.03	6158
		207	94.57	6041
		ACR	93.45	5414
		PV4	95.00	8036
	Plácido de Castro (Acre)	Brazil32	86.21	3685
	Belem (Pará)	Belem	88.68	4173
		Brasill	25.08	4943
Colômbia	Quibdó (Chocó)	Col015	87.08	6240
		Col020	86.13	5812
		Col026	88.84	5147
		Col036	89.71	6340
	Buenaventura, (Valle del Cauca)	Col038	88.69	6195
		Col040	87.98	6818
	Tumaco (Narino)	Col280	87.27	6832
		Col295	90.57	5875
		Col437	88.43	6320
		Col438-A	91.69	5378
Col438-B		92.91	5687	
	Col439	91.51	5258	

Tabela 3- Continuação da página anterior

País	Região	Amostra	Alinhamento com Sal I1 (%)	Número de SNPs
		Col440	87.31	6878
		Col441-A	85.11	5928
		Col441-B	92.05	4473
		Col443-A	87.08	7186
		Col443-B	87.76	6195
		Col445	87.93	6687
		Col446	86.94	7130
		Col448	89.60	5392
	Tierralta (Córdoba)	Col449	89.31	6407
		Col485	88.79	5733
		Col486	88.55	5605
		Col487	87.13	5543
		Col488	89.90	5092
		Col489	88.44	5224
		Col490	88.88	5745
		Col491	89.33	5951
		Col492	73.22	4297
		Col493	90.35	5852
		Col504	88.62	6391
Peru		Peru06	86.67	4829
	Santo Tomás (Loreto - Norte da Amazônia)	Peru09	77.52	5228
		Peru10	78.35	5394
		Peru257	86.81	5198
		Peru258	92.35	5972
	Delta 1 (Madre de Dios - Sul da Amazônia)	Peru259	86.67	4831
		Peru260	89.31	5627
		Peru262	88.18	5464
		Peru852	75.67	4732
	Mazán (Loreto - Norte da Amazônia)	Peru858	75.57	5468
		Peru869	82.61	5081

Tabela 3- Continuação da página anterior

País	Região	Amostra	Alinhamento com Sal I1 (%)	Número de SNPs
		Peru1008	87.36	5537
		Peru1021	91.95	4128
	Puerto America (Loreto - Norte da Amazonia)	Peru1022	93.38	4697
		Peru2025	85.42	7145
		Peru3133	90.73	4295
		Peru3136	81.12	6613
	Padrecocha (Loreto - Norte da Amazonia)	Peru3232	85.14	5066
		Peru3280	86.31	6895
	Iquitos (Loreto - Norte da Amazonia)	Peru4017	80.14	6589
		Peru4023	92.14	4007
	Sulina (Piura - Costa Norte)	Peru00622	83.91	5944
		Peru00699	84.07	4566
México	Frontera Hidalgo (Chiapas)	Mexico938-A	87.97	6112
	Carrillo (Chiapas)	Mexico32-E-03	84.38	6832
	Tuxtla Chico (Chiapas)	Mexico566-A	88.25	6120
	Huehuetán (Chiapas)	Mexico533-07	87.75	6061
		Mexico330-A	87.70	6342
		Mexico267-A	94.22	5592
		Mexico980-A	87.96	6498
		Mexico1086-A	85.32	7411
		Mexico55-03	88.45	6457
		Mexico203-04	88.91	6478
		Mexico118-A	88.34	6491
	Tapachula (Chiapas)	Mexico165-A	88.88	5571
		Mexico161-04	88.52	6014
		Mexico760-A	92.36	4904
		Mexico63-08	90.65	6036
		Mexico21-A	94.54	5233
		Mexico10-E-05	87.01	6977
		Mexico31A-03	88.02	7081

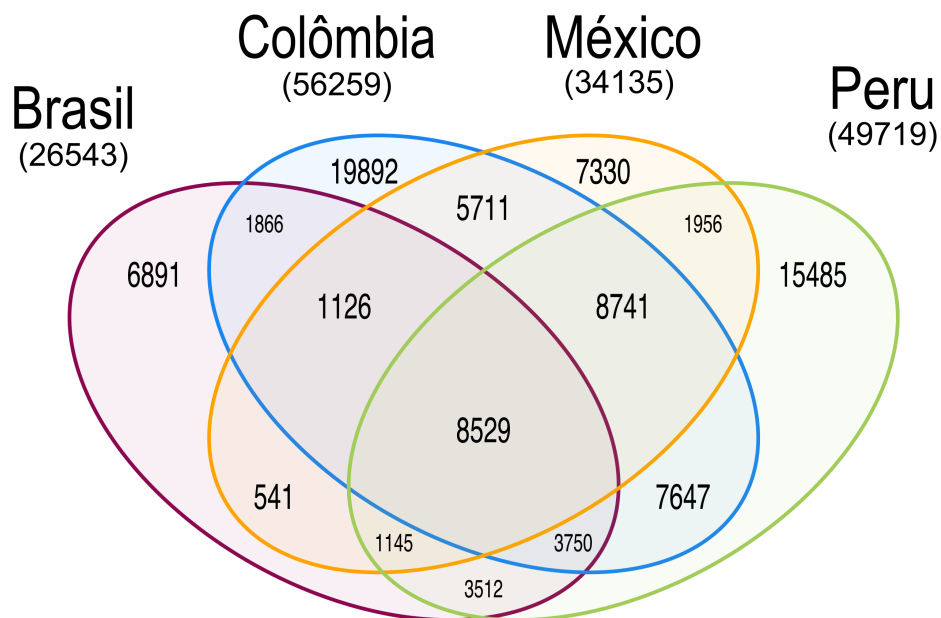
Tabela 3- Continuação da página anterior

País	Região	Amostra	Alinhamento com Sal I1 (%)	Número de SNPs
		Mexico1067-A	89.13	6831

O diagrama de Venn dos alelos compartilhados (Figura 4) entre os quatro países mostra uma maior quantidade de SNPs compartilhados entre as amostras do México e Colômbia (5711 SNPs, aproximadamente três vezes maior do que o encontrado entre México e Peru) e em seguida entre as amostras do Brasil e Peru (3512 SNPs, cerca de duas vezes maior o encontrado entre Brasil e Colômbia). Os 8529 SNPs únicos compartilhados entre todas as regiões, foram divididos em 36,6% (3133) em SNPs de regiões intergênicas e 50,9% (4339) em genes de regiões codantes, destes 57,8% (3123) foram considerados não-sinônimos.

6891 SNPs foram encontrados exclusivamente nos 11 isolados amazônicos. Dos SNPs em regiões codantes, 40,0% (2759), temos que 65,2% (1798) são devido a mutações não-sinônimas.

Figura 4 – Alelos compartilhados entre populações de *P. vivax* nas Américas.



Abaixo dos nomes dos países há o número total de SNPs encontrados para aquela população

Identificamos alguns *loci* associados a residência a drogas com mutações não-sinônimas com relação a Sal-I, incluindo: *pvDHFR*, *pvDHPS* e *pvMDR1*. Os três genes apresentaram uma única mutação cada, sendo: no gene *pvDHPS* (A383G), associado à resistência à sulfadoxina (PORNTHANAKASEM et al., 2016; RUNGSIHIRUNRAT et al., 2008); *pvDHFR* (S117N), associado à resistência à pirimetamina (IMWONG et al., 2003) e o gene *pvMDR1* (T958M), associado à resistência à cloroquina (GOMES et al., 2016). Os dois últimos apresentaram mutação em três das quatro regiões e, *pvDHPS* em apenas duas regiões.

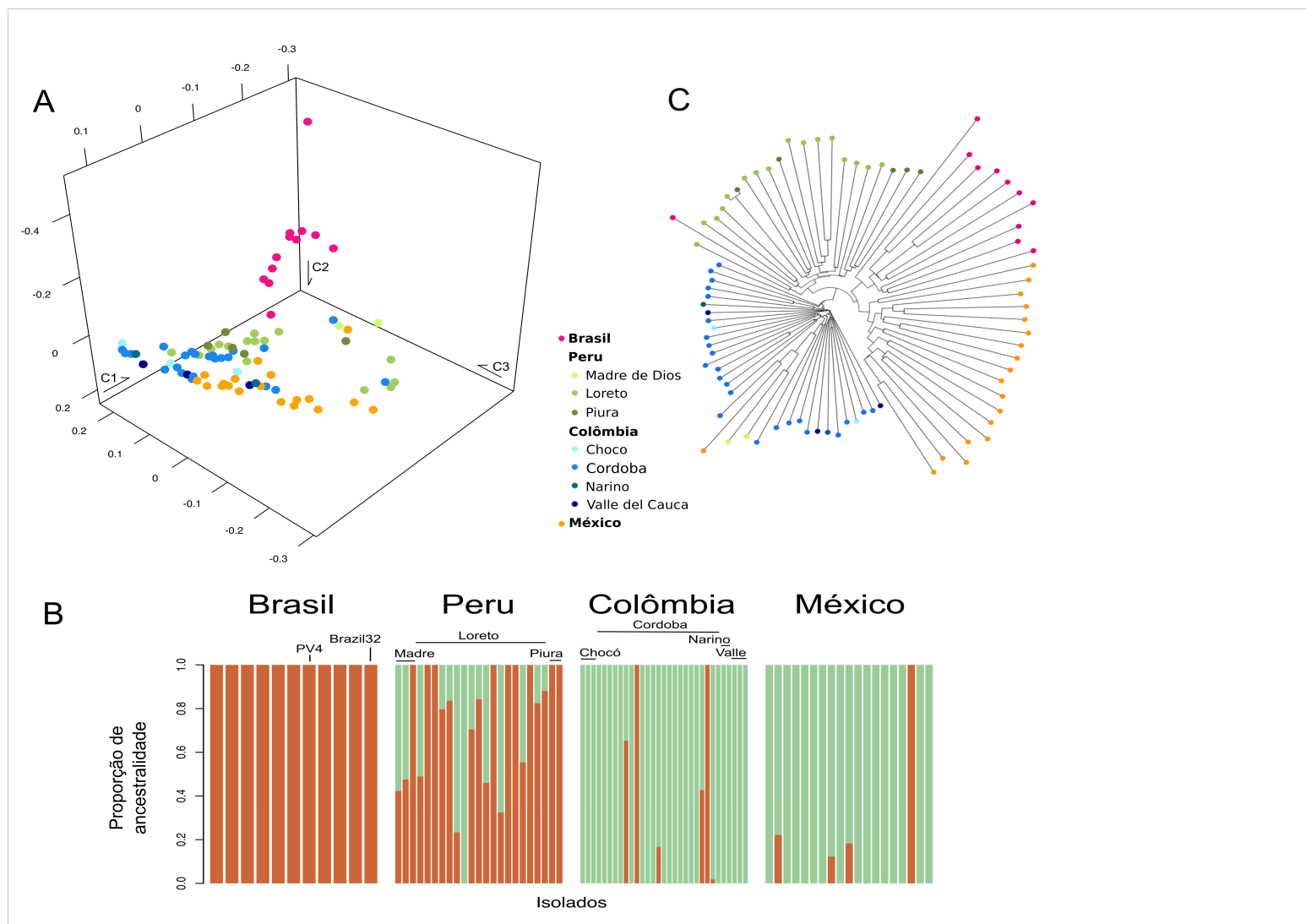
3.2.1 Estrutura populacional em diferentes regiões da América

Uma árvore construída com o método de *neighbour joining* (Figura 5c), utilizada para representar visualmente a matrix de distância genética, mostrou uma razoável separação das amostras segundo o país de origem. As análises subsequentes da divergência genética entre populações de parasitas utilizaram um método de análise multivariado (análise de componentes principais), uma análise pareada de índices de fixação (F_{ST}) e um método bayesiano de identificação de *clusters* implementado no programa ADMIXTURE.

Realizamos a PCA com os marcadores do tipo SNPs para explorar a estrutura populacional de *P. vivax* nas Américas, com três bancos de dados distintos. O primeiro possuía 84 amostras da América, os segundo essas 84 mais as amostras Brasil I e Belem e o terceiro foi formado por SNPs propostos como *barcode* para *P. vivax* (BANIECKI et al., 2015). Os resultados mostraram que essas amostras brasileiras, Brasil e Belem, destuavam as análises tanto de componentes principais (Figura S1a), que forçou a aglomeração das outras amostras nos componentes C2 e C3, quanto de ancestralidade em que a amostra Brasil apresentou ser um único ancestral (Figura S1b).

A PCA das 84 amostras sem as amostras Brasil e Belem (Figura 5a) revelou estrutura primeiramente determinada pela localização geográfica. Com a adição de isolados brasileiros simpátricos ao estudo que originou os genomas do Peru, Colômbia e México, temos uma diferenciação maior entre as amostras dos países, principalmente das do Brasil, do que a encontrada nos resultados de Hupaló et al. (2016). Em concordância com os dados da PCA, os dados de clusterização obtidos por ADMIXTURE, com 2 *clusters* (K), mostram as amostras brasileiras formadas por um único ancestral e as amostras mexicanas formadas por majoritariamente um outro ancestral (Figura 5b). Esse valor de *clusters* populacionais foi obtido segundo o menor valor de erro encontrado para K (Figura S2). As amostras peruanas e colombianas, no entanto, mostraram influência de 2 ancestrais distintos, sendo a colombiana mais próxima da mexicana enquanto o contrário acontece com a peruana; valor de K=3 também mostra essa tendência, porém mais bem resolvida (Figura S3).

Figura 5 – Estrutura populacional baseada em SNPs de 84 amostras de *P. vivax* nas Américas.



(A) Análise de Componentes Principais com três *eigen*vectors. Cada isolado foi colorido de acordo com sua origem, e percentuais dos *eigen*vectors são: C1 = 22,27%, C2 = 14,15% e C3 = 8,15%; (B) Análise de ADMIXTURE com a base de dados total de SNPs usando valor de $K = 2$; (C) Árvore baseada em *neighbor-joining* iniciando a partir dos isolados com menor variação ancestral no painel A.

Em um panorama geral, podemos observar a atribuição de ancestrais devido a geografia dos países, isto é, países mais próximos compartilham mais ancestrais que países mais distantes. Valores de F_{ST} (Tabela 4) foram condizentes com as estruturas apresentadas nos resultados do ADMIXTURE, com Colômbia e México, apresentando o menor valor, $F_{ST} = 0,025$, e Brasil e México com o maior valor, $F_{ST} = 0,092$. Valores de F_{ST} tendendo a 1 implicam que quase toda a variação genética é explicada pela divergência entre populações, enquanto valores próximos de 0 indicam predomínio do componente de variação intra-populações. Valores obtidos com os pares de populações Brasil vs. Peru e Brasil vs. Colômbia apresentaram valores intermediários (0,037 e 0.055, respectivamente), condizentes com a proximidade geográfica do local de coleta das amostras.

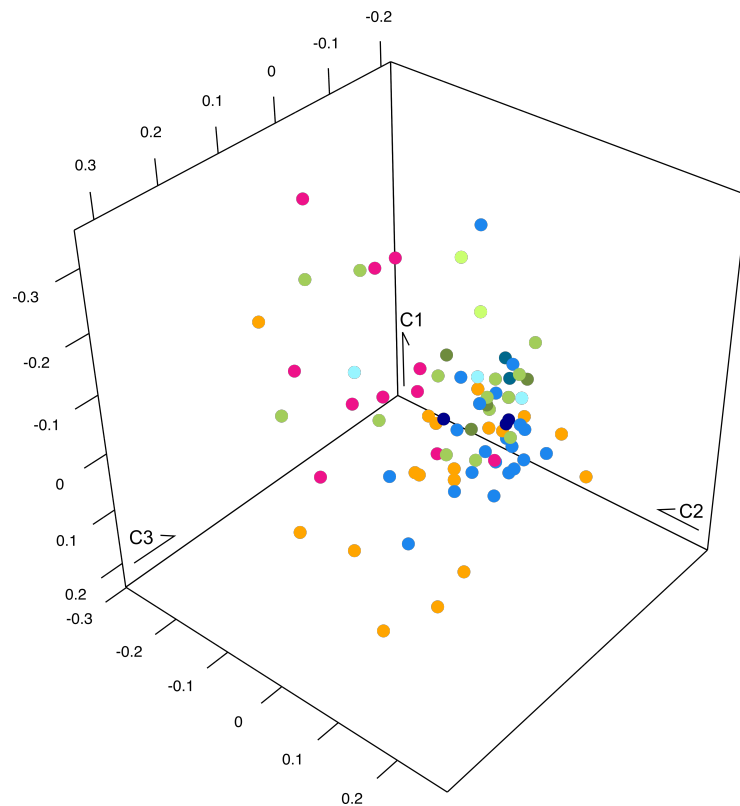
Tabela 4 – Valores de F_{ST} para cada par de regiões da América

	Brasil	Colômbia	Peru	México
Brasil	-	0.055	0.037	0.092
Colômbia	-	-	0.03	0.025
Peru	-	-	-	0.047
México	-	-	-	-

Os autores do ADMIXTURE acreditam ser uma boa ideia diminuir a quantidade total de marcadores para aqueles que não possuam DL, já que o modelo que eles criaram não considera explicitamente os SNPs com DL. O resultado foi similar ao banco de dados com SNPs totais, porém menos informativo; as amostras brasileiras apresentaram ancestralidade com mais de um *cluster*, independente do $K=2$ ou $K=3$, melhores valores de *clusters* encontrados (Figura S3).

A terceira PCA foi feita utilizando apenas 37 SNPs dos 42 SNPs do *barcode* que passaram em nossos critérios de seleção; os 5 outros SNPs propostos não se apresentaram como mutação em nenhuma de nossas amostras, assim, foi tido como igual a referência Sal-I. Esse *barcode* foi proposto em sítios que possuíam seleção neutra, isto é, regiões intergênicas, introns e regiões codantes com a *4-fold-degenerated*, com base em 87 amostras da África (Etiópia), Ásia (Sri Lanka) e América do sul (Guiana Francesa e Brasil). O resultado obtido por Bacaniecki et al. (2015) mostra a distinção entre os três continentes com utilização, por exemplo, da PCA, porém dentro da América do Sul não houve clara distinção. Ao refazermos a PCA com os 37 SNPs em nossas 84 amostras do continente americano, observamos que o *barcode* não é informativo para a distinção de populações intracontinentais (Figura 6).

Figura 6 – Análise de componentes Principais com três *eigenvectors* para o *barcode* de 37 SNPs.



Percentuais dos *eigenvectors* são: C1 = 19,93%, C2 = 12,03% e C3 = 11,42%. Cores correspondem a: Brasil em rosa, Colômbia em azul, México em laranja e Peru em verde

3.2.2 Diversidade nucleotídica e seleção em regiões dos genomas da América

Análises de diversidade nucleotídica a partir dos SNPs mostraram que as 4 regiões apresentaram valores de π médio similares (0.00037 - Brasil; 0.00040 - Colômbia; 0.00039 - Peru; Mexico 0.00034). Os maiores valores de π por pares de SNPs em janelas de 1 Kb foram identificados em regiões previamente descritas como hipervariáveis, principalmente em genes que codificam antígenos de estágios sanguíneos do parasito (Figura 7), como:

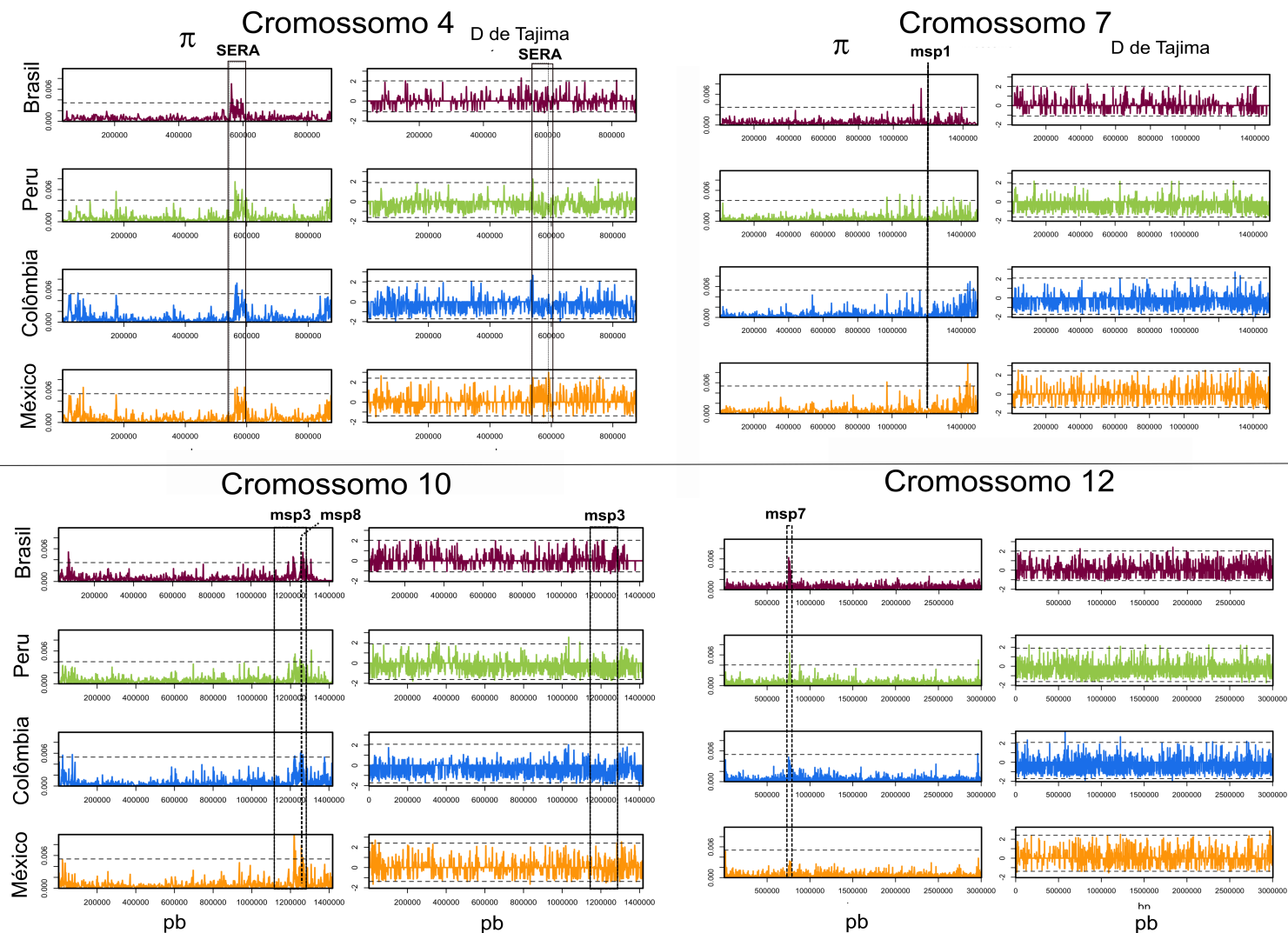
- genes *vir*: a maior família multigênica de *P. vivax*, com cerca de 346 genes que estão implicados na variação antigênica, correspondendo a 10% dos genes de *P. vivax* (CARLTON et al., 2008). Localizados nas regiões subteloméricas, podem exercer diferentes funções durante a evasão do sistema imune (SINGH; GUPTA; PANDE, 2014)

- família SERA (*Serine Repeat Antigen* - 8 genes de uma famílias de 12 membros) que são altamente expressos entre os estágios de trofozoito maduro a esquizontes do desenvolvimento do parasito, servindo de alvo para intervenções (HORII et al., 2010).

- famílias de proteínas associadas a superfície do merozoito (MSP - *Merozoite Surface Protein*): se apresentam em 10 famílias no genoma de *P. vivax* (CARLTON et al., 2008).

- - MSP-7 (8 genes de uma família de 11 membros): proteína que em *P. falciparum* liga-se a MSP1 e possui uma expansão do número de cópias em *P. vivax*, que, nesta espécie, ainda não se sabe se há ligação com MSP-1 nesta última espécie (GARZÓN-OSPINA et al., 2012; GARZÓN-OSPINA; FORERO-RODRÍ GUEZ; PATARROYO, 2014).
- MSP-3 (11 genes de uma família de 11 membros): os onze membros da família que ocorrem *in tandem* no cromossomo 10, foram identificados como antigênicos e alguns de seus membros estão sujeitos a pressão do sistema imune, fixando ou acumulando diversas mutações não-sinônimas como mecanismo de evasão (GARZÓN-OSPINA; FORERO-RODRÍ GUEZ; PATARROYO, 2015).
- MSP-8: esse gene, utilizando métodos filogenéticos, foi identificado sob evidência de seleção purificadora em linhagens de *P. vivax* (PACHECO M ANDREÍ et al., 2012).
- MSP-1: a proteína codificada por esse gene em *P. vivax* pode ter um papel importante no processo de fixação e invasão nos reticulócitos (ROY et al., 2009).

Figura 7 – Diversidade encontrada nos cromossomos 4, 7, 10 e 12, nas 84 amostras de *P. vivax* nas 4 populações das Américas.



Diversidade de nucleotídeos (π) e D de Tajima foram calculadas em janelas de 1kb para infecções simples, com o banco de dados de SNPs de alta qualidade. As linhas horizontais em cada gráfico representam o *cut-off* dos 50 maiores valores de π e, maiores valores para D de Tajima encontrados por região. Os box representam regiões de alta diversidade. Cores representam: Rosa - Brasil Azul - Colômbia Laranja - México e Verde - Peru

Nos 50 maiores picos de diversidade dentro de todos os cromossomos (Tabela 5), que aparecem somente nas amostras brasileiras, temos que metade das regiões são compostas por genes *vir*, proteínas hipotéticas e regiões intergênicas. A outra metade é composta por:

- três genes com domínios que envolvem tRNA;
- um gene com domínio do tipo *Phrist*, que possui um domínio identificado primeiramente em *P. falciparum* em proteínas do tipo RESA (*Ring-infected erythrocyte surface antigen*).
- um gene que codifica proteína TATA-*binding* (TBP), que é caracterizado como um fator de transcrição que se liga especificamente a regiões do DNA chamadas de TATA *box*.
- um gene com domínio de actina, que é uma família de proteínas globulares que formam filamentos de actina.

Tabela 5 – 50 maiores valores de diversidade nucleotídica (π) encontrada nas quatro regiões da América (Brasil, Colômbia, México e Peru) divididos pelos cromossomos. Genes associados as regiões, quando não em regiões intergênicas, foram identificados

Cromossomo	Região		π				Genes
	Início (pb)	Fim (pb)	Brasil	Peru	Colômbia	México	
1	773001	774000	X				PVX_093680, (Phist protein Pf-fam-b))
	800001	801000	X				PVX_093710(VIR)
	805001	806000			X	X	PVX_093715(VIR)
	829001	830000				X	PVX_093735(VIR)
2	1000	3000			X	X	PVX_096910(VIR)
	66001	93000	X	X	X	X	PVX_096965(VIR), PVX_096970(VIR), PVX_096975(VIR), PVX_096980(VIR), PVX_096985(VIR), PVX_096990(VIR)
3	3001	4000			X		NC
	9001	27000	X	X	X	X	PVX_001100 (RAD protein (Pv-fam-e)), PVX_001097 (tRNA Asparagine), PVX_001095(PH), PVX_001090(PH), PVX_001085(PH), PVX_001080(PH)
	472001	473000	X				NC

Tabela 5- Continuação da página anterior

Cromossomo	Região		π				Genes
	Início (pb)	Fim (pb)	Brasil	Peru	Colômbia	México	
	901001	972000			X	X	PVX_096001(VIR), PVX_096001a(VIR), PVX_096000(VIR), PVX_095997(VIR), PVX_095995(VIR), PVX_095990(VIR), PVX_103145(VIR), PVX_103150(VIR), PVX_103155(VIR), PVX_103160(VIR), PVX_000000 (Plasmodium exported protein, unknow function), PVX_000005 (Plasmodium exported protein, unknow function), PVX_000010 (Plasmodium exported protein, unknow function), PVX_000015(Pv- fam-d protein),
4	51001	69000			X	X	PVX_002507 (Pv-fam-b protein), PVX_002510 (Nucleosomal binding protein 1), PVX_002512(PH),
	175001	176000		X	-	-	NC

Tabela 5- Continuação da página anterior

Cromossomo	Região		π				Genes
	Início (pb)	Fim (pb)	Brasil	Peru	Colômbia	México	
	562001	600000	X	X	X	X	PVX_003840(SERA), PVX_003835(SERA), PVX_003830(SERA), PVX_003825(SERA), PVX_003820(SERA), PVX_003815(SERA), PVX_003810(SERA), PVX_003805(SERA)
	861001	873000		X			PVX_003485(VIR), PVX_003475(PH)
5	17001	18000				X	PVX_088795(VIR)
	37001	39000		X		X	PVX_088805(VIR)
	1256001	1257000	X			X	PVX_090240(CyRPA)
	1291001	1298000		X	X		PVX_090280(PST-A protein)
	1313001	1314000			X	X	PVX_110935(PH)
6	18001	31000		X	X	X	PVX_001635(VIR), PVX_001640(VIR), PVX_001645(VIR)
	841001	842000		X			PVX_110935(PH)
7	969001	971000		X		X	PVX_099690(TLP-2)
	1044001	1045000		X			NC
	1115001	1117000	X	X			PVX_099915 (RNA-binding protein)

Tabela 5- Continuação da página anterior

Cromossomo	Região		π				Genes
	Início (pb)	Fim (pb)	Brasil	Peru	Colômbia	México	
	1159001	1161000	X	X			PVX_099980(MSP1)
	1356001	1357000		X			PVX_086940(PH)
	1392001	1393000	X				PVX_086905(PH)
	1401001	1472000		X	X	X	PVX_086900(PH), PVX_086895(VIR), PVX_086893(VIR), PVX_086890(VIR), PVX_086880(VIR), PVX_086875(VIR), PVX_086870(VIR), PVX_086865(VIR), PVX_086863(VIR), PVX_086860(VIR), PVX_086855(VIR), PVX_086850(VIR), PVX_086845(VIR), PVX_005040(PH), PVX_005045(VIR),
8	1001	12000	X	X	X	X	PVX_094230(PH)
	24001	25000			X	X	PVX_094240(VIR)
	1639001	1640000	X				NC
	1672001	1673000				X	PVX_119210(VIR)
9	3001	10000		X	X		PVX_090835(PH)

Tabela 5- Continuação da página anterior

Cromossomo	Região		π				Genes
	Início (pb)	Fim (pb)	Brasil	Peru	Colômbia	México	
	21001	29000	X				PVX_090840(PH), PVX_090845(60S acidic ribosomal protein P1), PVX_090847 (tRNA Glycine), PVX_090848 (tRNA Leucine),
	36001	37000	X	X			PVX_090860 (CPW-WPC family protein)
	40001	41000	X				PVX_090865(actin)
	337001	338000				X	PVX_091200(PH), PVX_091205 (replication factor C subunit 5)
10	21001	22000			X		PVX_079695(PH)
	67001	69000	X				PVX_079750(PH)
	71001	72000	X		X		PVX_079755 (phosphoenolpyruvate/ phosphate translocator (PPT))

Tabela 5- Continuação da página anterior

Cromossomo	Região		π				Genes
	Início (pb)	Fim (pb)	Brasil	Peru	Colômbia	México	
	1191001	1308000	X	X	X	X	PVX_097755(MyoB), PVX_097750(protein kinase), PVX_097745(ADF1), PVX_097740(SRSF12), PVX_097735(PH), PVX_097730(PH), PVX_097725(MSP3), PVX_097720(MSP3), PVX_097715(PH), PVX_097705(MSP3), PVX_097700(MSP3), PVX_097695(MSP3), PVX_097690(MSP3), PVX_097685(MSP3), PVX_097680(MSP3), PVX_097675(MSP3), PVX_097670(MSP3), PVX_097665(MSP3), PVX_0977660(1spE), PVX_097655(50S ribosomal protein L28, apicoplast), PVX_097650(PH), PVX_097645(PH), PVX_097640(PH), PVX_097635(PH), PVX_097630(PH), PVX_097625(MSP8), PVX_097620(PH), PVX_097615(PH), PVX_097610(PH), PVX_097605(VPS11), PVX_097600(PH)
	1377001	1378000			X		NC
11	2027001	2029000				X	PVX_124725(VIR)
	2062001	2063000			X	X	NC

Tabela 5- Continuação da página anterior

Cromossomo	Região		π				Genes
	Início (pb)	Fim (pb)	Brasil	Peru	Colômbia	México	
12	751001	776000	X	X			PVX_082710(PH), PVX_082700(MSP7), PVX_082695(MSP7), PVX_082685(MSP7), PVX_082680(MSP7), PVX_082675(MSP7), PVX_082670(MSP7), PVX_082665(MSP7), PVX_082660(PH), PVX_082655(MSP7)
	881001	882000		X			PVX_082480(PH)
	2960001	2961000			X		NC
	2967001	2968000		X			PVX_118675(PH)
13	1001	2000			X		NC
	5001	7000				X	PVX_084090 (liver stage associated protein 1(LSAP1))
	130001	131000	X				PVX_084260(PH)
	909001	910000	X				PVX_085105 (transcription initiation TFIID-like)
	1992001	1994000	X	X	X	X	PVX_086305 (ubiquitin-40S ribosomal protein S27a)
	2025001	2026000			X	X	NC

Tabela 5- Continuação da página anterior

Cromossomo	Região		π			Genes	
	Início (pb)	Fim (pb)	Brasil	Peru	Colômbia		México
14	43001	44000			X	X	PVX_1211885 (cytoadherence linked asexual protein (CLAG))
	996001	997000		X			PVX_122920(PH)
	2019001	2020000		X			PVX_124065(PH)
	2032001	2033000	X	X			PVX_124975(PH)

Símbolos indicam: 'X' - SNPs incluídos entre os mais 50 polimórficos em cada país, 'PH' - Proteína Hipotética e 'NC' - Nada Consta

Regiões de família SERA e MSP3 ainda apresentaram valor de D de Tajima dentre os 50 maiores e menores valores encontrados entre todos os SNPs. Além das famílias já destacadas anteriormente para o teste D de Tajima temos entre os maiores valores, 44 genes identificados distribuídos entre as quatro regiões, 76 genes caracterizados como proteínas hipotéticas e 8 relacionados a processos ou subunidades de RNA (Tabela S2). Para o teste D de Tajima com os menores valores, temos 121 genes distribuídos pelas 4 regiões, sendo que 21 deles codificantes de proteínas vir, 55 caracterizados como proteínas hipotéticas e 10 relacionados a processos ou subunidades de RNA (Tabela S3).

Três grandes grupos de genes dividiram os valores de D de Tajima, em primeiro os associados a interação parasito-hospedeiro, em seguida os associados ao ciclo de vida do parasito e em terceiro um grupo associado a fertilidade, com apenas 5 dos 51 genes discriminados (Tabela 6). Podemos destacar, para os associados a interação parasito-hospedeiro, em amostras do Peru, genes que codificam proteínas TLP2 e CSP - valor D de Tajima positivo - e em amostras da Colômbia, a gene codificante da proteína MSP5 - valor D de Tajima negativo. Para os relacionados ao ciclo de vida temos o ApiAP2, que apresentou D de Tajima positivo em três das quatro regiões e negativo no Peru; único gene com valores de D de Tajima estudado encontrado nas quatro regiões. O Brasil e o Peru compartilharam 5 genes, sendo um deles o ApiAP2, mencionado acima, e: RAD, PSOP24, SRPK1 e CuTP.

Tabela 6 – Funções do genes associados aos maiores e menores valores de D de Tajima para cada uma das regiões da América (Brasil, Colômbia, México e Peru). As funções foram separadas em três grandes grupos: interação parasita-hospedeiro, regulação do ciclo de vida e fertilidade. Foram apenas estabelecidas as funções de genes que estivessem descritas em alguma espécie de *Plasmodium*.

Região	Função	Genes	D de Tajima	Citação
INTERAÇÃO PARASITO- HOSPEDEIRO				
Brasil	Evasão da reposta do sistema imune	RAD	+	Cornejo, et al, 2014
	Superfície do merozoito	6cysteine	+	Hostetler, et al, 2015
	Superfícies de estágios sanguíneos e esporozoitos	CRMP1	+	Thompson, et al, 2007
	Superfícies de estágios sanguíneos e esporozoitos	CRMP3	+	Douradinha,et al, 2011
	Superfícies de estágios sanguíneos e esporozoitos	Pvstp1		Winter, et al, 2005
	Associado a infecção do intestino pelos oocinetos	PPLP5	+	Ecker, et al, 2007
	Associado ao estágio de oocineto	PSOP24		Baum, et al, 2013
	Sistema de remoção de antioxidante	TrxR	+	Tiwari, et al, 2016
	Proteção contra estresse nitrosativo	Trx1	+	Wang, et al, 2014
	Resposta do parasita a drogas	ABCG2	+	Eastman, et al, 2013
Peru	Proteção contra extresse oxidativo	TLP2	+	Afanador, et al, 2014
	Invasão de esporozoitos em hepatócitos	CSP	+	Ancsin, et al, 2004
	Resistência a drogas	MDR1	+	Sanchez, et al, 2010
	Motilidade do esporozoito e ligação no hepatócito	TREP		Combe, et al, 2009
Colômbia	Superfície de estágios sanguíneos	MSP5		Perraut, et al, 2014
	Associado a invasão de eritrócitos ou interações com o sistema imune do hospedeiro	P41	+	Taechalertpaisarn, et al, 2012
	Associado a regulação da secreção do micronema durante a invasão dos eritrócitos	EPAC	+	Dawn, et al,
	Associado a mediação do sequestro de hemácias e ao tráfico de proteínas específicas para dentro da célula do hospedeiro	PTEX88	+	Chisholm, et al, 2016
México	Proteção do parasito de efeitos danosos durante episódios de febre no hospedeiro	HSP110c	+	Muralidharan, et al, 2013

Tabela 6- Continuação da página anterior

Região	Função	Genes	D de Tajima	Citação
	Ligação com eritrócitos ou associado a membrana durante a invasão	RhopH2	+	SamYellowe, et al, 1992
	Ligação com eritrócitos ou associado a membrana durante a invasão	RON3	+	Zhao, et al, 2014
	Ligação com eritrócitos ou associado a membrana durante a invasão	RBP1a		Gallinski, et al, 1992
	Associado ao desenvolvimento da rede tubovesicular (TVN) no eritrócito	EVP1		Tamez, et al, 2008
ASSOCIADO AO CICLO DE VIDA				
Brasil	Regulação da transcrição	ApiAP2	+	Lindner, et al, 2010
	Regulação de genes estágio específicos	AP2O	+	Kaneko, et al, 2015
	Regulação de expressão de genes associados a virulência	HDA2		Coleman, et al, 2014
	Expressão de genes de estágio intraeritrocítico	PREBP	+	KomakiYasuda, et al, 2013
	Expressão estágio específico	Pvfama		Zeeshan, et al, 2015
	Síntese de repetições teloméricas	TERT	+	Bottius, et al, 1998
	Metilação de histonas e alvos não histônicos	MYND		Browm, et al, 2006
	Associado a proliferação do parasito	SRPK1	+	
	Componente facilitador de produtos integrados a membrana ou transportados ao lumen do RE	Sec61gamma		Erdmann, et al, 2011
	Formação da vesícula COPII	Sec24a		Lee, et al, 2008
	Peru	Associado a desenvolvimento e diferenciação	methyltransferase	+
Associada ao centróssomo		CEP76	+	Kumar, et al, 2013
Associado a retenção de proteínas no RE		ERD2	+	Elmendorf, et al, 1993
Síntese de lipoamida		histone acetyltransferase GCN5		Siwo, et al, 2015
Associado a manutenção dos estágios eritrocíticos		SufD		Gisselberg, et al, 2013
Associada a divisão de organelas, endocitose e criação de vesículas		dynamiclike protein		Li, et al, 2004
Associado a sumoilação		SUMO		Reiter, et al, 2013

Tabela 6- Continuação da página anterior

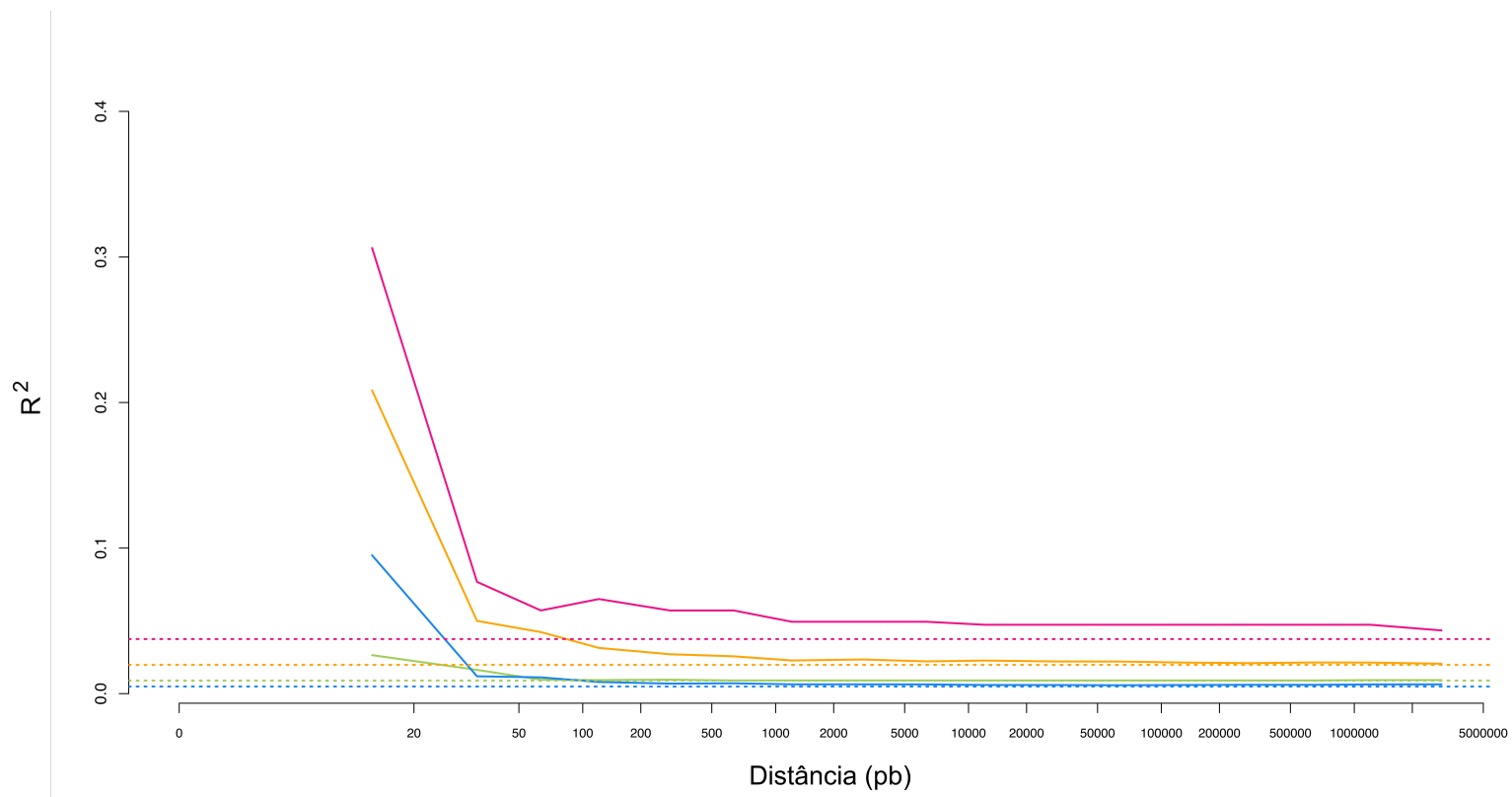
Região	Função	Genes	D de Tajima	Citação
Colômbia	Associado a montagem da estrutura do corpo basal	delta tubulin	+	Francia, et al, 2016
	Sobrevivência do parasito e degradação de proteínas ubiquitinadas	RPN6	+	Muralidharan, et al, 2011
	Associado a recombinação homóloga durante a mitose	RAD54		Kumar, et al, 2013
	Regulação do desenvolvimento dos estágios iniciais do gametócito	GDV1		Mobegi, et al, 2014
	Associado a atividade do hipnozoito	PI4K		Zeeman, et al, 2016
México	Envolvido na iniciação e alongamento da transcrição	CCr4NOT transcription complex subunit 1		Bischoff, et al, 2010
FERTILIDADE				
Brasil	Associado a fertilidade	CuTP	+	Kenthirapalan, et al, 2013
México	Diferenciação sexual	Guanylyl cyclase	+	Baker, et al, 2004
	Controle de translação em gametócitos	DOZI		Bunnik, et al, 2016
	Regulação da frequência da conversão gametocítica	HDA2		Coleman, et al, 2014
	Associado a motilidade em gametas masculinos	dynein heavy chain	+	Khan, et al, 2005

A ausência de símbolo na coluna 'D de Tajima' representa que o valor de D é negativo

3.3 Declínio do desequilíbrio de ligação em diferentes populações

Observou-se um rápido declínio de DL em função da distância nucleotídica entre pares de SNPs em todas as populações estudadas (Figura 8); em geral, os valores de r^2 entre pares de SNPs situados a mais de 50 pb no mesmo cromossomo foram semelhantes àqueles encontrados entre SNPs situados em cromossomos distintos (isto é, sem ligação física entre eles). O decaimento é ligeiramente mais rápido do que aquele encontrado por Pearson et al. (2016) em amostras do Sudeste Asiático e da Oceania (valores de r^2 equivalentes aos de *loci* não-ligados são observados a partir de 100 e 200 pb de distância entre pares de SNPs), regiões essas com alta transmissão de malária. Em amostras peruanas, observou-se mínimo DL mesmo entre *loci* situados a menos de 50 pb de distância.

Figura 8 – Valores de r^2 calculados entre pares de SNPs entre distâncias definidas a partir de dados de SNPs por região.



Linha de base definida a partir de valores de r^2 de SNPs em cromossomos diferentes.

4 DISCUSSÃO

Para expandir nosso entendimento da diversidade genética em *P. vivax*, sequenciamos 9 genomas nucleares completos de parasitos da Amazônia brasileira de mesma origem geográfica e os comparamos ao nível continental, utilizando genomas disponíveis de países americanos. A alta cobertura gerada pela tecnologia de sequenciamento por *Ion Proton* e *Ion Torrent*, em nossas amostras, e da tecnologia *Illumina*, nas amostras disponíveis nos bancos de dados, associada à triagem subsequente de SNPs com o uso dos mesmos métodos e *softwares* específicos, assegurou uma maior comparabilidade entre os dados gerados neste trabalho e aqueles derivados da literatura.

Estudos com *P. vivax* obtidos de amostras de sangue de pacientes infectados são dificultados pela presença de DNA humano. Devido à diferença entre os tamanhos dos genomas, se um único leucócito estiver presente em uma amostra que contém 10 parasitos, mais de 95% do DNA extraído será de origem humana. Este problema agrava-se no estudo de *P. vivax*, em comparação com *P. falciparum*, já que a parasitemia em infecções naturais é tipicamente inferior a 10.000 parasitos por μl de sangue (CHAN et al., 2015). Nas amostras do Brasil, Colômbia, México e Peru sequenciadas depois de um processo de enriquecimento de DNA do parasito por seleção híbrida (HUPALO et al., 2016), entre 18 e 91% dos reads foram mapeados contra o genoma de referência; em comparação, em nossas amostras, que passaram pelo processo de depleção de leucócitos, a percentagem de mapeamento foi superior a 91%. Esses dados mostram o sucesso obtido pela técnica de leucodepleção adaptada em nosso laboratório.

No primeiro estágio das análises, com 94122 SNPs de alta confiança e com 60% do genoma de referência codante (cromossomos variando entre 54 a 64%) observamos que mais metade dos SNPs estavam presentes em regiões intergênicas em amostras americanas, mesmo que em amostras majoritariamente do Sudoeste asiático e da Oceania haja similaridade na quantidade de SNPs entre as regiões intergênicas e codantes (PEARSON et al., 2016). Um fato recorrente em genomas de *P. vivax*, com relação às regiões codantes, é a verificação de uma quantidade maior de SNPs não-sinônimos (nsSNPs) em relação aos sinônimos (sSNPs). A diferença nesses genomas sugere a existência de seleção diversificadora atuando em escala genômica (PEARSON et al., 2016). Em Chan et al. (2012), a partir de dados de SNPs de genomas nucleares completos de três amostras do Camboja e duas de Madagascar, foi verificada a existência de 1,5 vezes mais nsSNPs em relação a sSNPs. Neafsey et al. (2012), em dados suplementares mostraram que existe também essa relação de maior quantidade de nsSNPs do que sSNPs, tanto em *P. vivax* quanto em *P. falciparum*, em 5 amostras de regiões diferentes para cada espécie. Com um número mais expressivo de genomas sequenciados, como no caso de Pearson et al. (2016), cerca de 58% foram considerados nsSNPs. Os dados obtidos neste estudo corroboram esses resultados, apresentando amostras americanas com mais de 60% de

nsSNPs, mesmo quando restringimos a análise para SNPs únicos compartilhados por amostras brasileiras.

Devido à natureza conservada das enzimas da via metabólica do folato, sítios polimórficos semelhantes nos genes DHFR e DHPS de *P. vivax* e *P. falciparum* podem conferir resistência às drogas pirimetamina e sulfadoxina (HYDE, 1990; SAMAD et al., 2015; WANG et al., 1997). Embora a combinação dessas drogas não seja normalmente usada como um tratamento para a infecção por *P. vivax*, essas duas drogas tem sido utilizadas para tratar pacientes co-infectados com *P. falciparum* e *P. vivax*; uma prática que teria exposto *P. vivax* a pressão do fármaco e que pode ter resultado em mutações-chave para os fenótipos resistentes relatados em infecção por *P. falciparum* (SNOUNOU; WHITE, 2004). Polimorfismos em *pvDHFR* nos sítios S58R e S117N são altamente relacionados à resistência de pirimetamina, sendo que mutações adicionais podem aumentar o grau da resistência (IMWONG et al., 2003). A Colômbia, único país que apresentou as duas mutações em *pvDHFR* na maioria das amostras, passou por um processo de uso dessas drogas para tratamento de *P. falciparum* durante mais de duas décadas e hoje pode-se observar um alto nível de resistência a pirimetamina principalmente na Bacia Amazônica colombiana (OSORIO; PÉREZ; GONZÁLEZ, 2007). Apesar de algumas mutações nos genes DHFR e DHPS em conjunto estarem associadas a resistência (GREGSON; PLOWE, 2005; HERNÁNDEZ et al., 2013; HYDE, 1990; SIBLEY et al., 2001), não encontramos evidência dessas em nossas amostras. Em duas únicas amostras, uma do Peru e outra da Colômbia, a mutação A383G em *pvDHPS*, foi encontrada. A mutação A383G identificada sozinha, na dupla com S382A ou na tripla com S382A e A553G, foi encontrada em 90% das amostras resistentes onde malária é endêmica na Ásia (PORNTHANAKASEM et al., 2016; RUNGSIHIRUNRAT et al., 2008).

Por sua vez, a resistência à cloroquina é o principal desafio para os programas nacionais de controle da malária vivax. Os primeiros casos de resistência a essa droga foram descritos em países Asiáticos (BAIRD, 2004; LEE et al., 2009) e da Oceania; países da América do Sul, incluindo o Brasil (ALECRIM; ALECRIM; MACÊDO, 1999; GONÇALVES LÍ; CRAVO; FERREIRA, 2014; MELO et al., 2014) também identificaram casos de resistência. Os polimorfismos Y976F e F1076L no gene *pvmdr1* foram descritos como marcadores moleculares associados a resistência à cloroquina (BREGA et al., 2005). Nossas análises indicaram a existência da mutação T958M em três das quatro regiões estudadas, excluindo-se o Brasil. Estudos prévios com amostras da região amazônica brasileira mostraram que a mutação T958M no gene *pvmdr1* foi encontrada com frequência no Brasil e em amostras extra amazônicas (CHEHUAN et al., 2013; GOMES et al., 2016; MARTHA et al., 2005).

4.1 Estrutura Populacional de *P. vivax* na América

Dada a recente evidência da origem de *P. vivax* na África (LIU et al., 2014b), tem sido proposto que populações de *P. vivax* do Novo Mundo foram fundadas na era colonial por uma população de parasitos europeus que agora estão erradicados (CULLETON et al., 2011). Assim, discutem-se os motivos que levaram a uma grande divergência entre amostras de parasitos de Novo Mundo e do Velho Mundo, observada com o uso de marcadores como microssatélites ou SNPs. Análises de F_{ST} , clusterização e PCA mostraram resultados similares em microssatélites, com as amostras da América do Sul parcialmente ou totalmente separadas das demais (Ásia e África) (KOEPLI et al., 2015; SCHOUSBOE et al., 2014); em resultados de clusterização com uso de SNPs, foi sugerido que os isolados da África e Índia exibiam grande heterogeneidade relativa aos ancestrais de isolados de outras populações, sendo inferido que isto derivava tanto do Novo quanto do Velho Mundo (HUPALO et al., 2016; PEARSON et al., 2016). Clusterização por continente também foi observada em amostras genômicas de *P. falciparum* (OCHOLLA et al., 2014; SAMAD et al., 2015).

Em amostras do Novo Mundo, podemos observar a atribuição de ancestrais com base na geografia dos países; países mais distantes compartilham menos ancestrais, como por exemplo Brasil vs. México, do que países vizinhos, como por exemplo Brasil vs. Peru (ver também (HUPALO et al., 2016)). No entanto, com o uso de marcadores genéticos mais polimórficos, como os microssatélites, observa-se a formação de grupos separados contendo amostras do Brasil e Peru, com pouca miscigenação entre eles (KOEPLI et al., 2015; SCHOUSBOE et al., 2014). Com o uso de microssatélites, pode-se observar divergência acentuada entre algumas subpopulações provenientes de uma mesma província do Peru (VANDENEDE et al., 2010), embora nem sempre esse padrão se mantenha (DELGADO-RATTO et al., 2016). Porém, em nossos dados de três regiões peruanas, obtivemos que 10 das 23 amostras apresentaram um único ancestral, sendo ele o mesmo encontrado no Brasil, mas não pudemos observar nessas regiões clara estrutura populacional. Em estudos com microssatélites unindo amostras da Ásia e da Colômbia, com $K=5$, temos a predominância de 2 ancestrais na Colômbia, enquanto os outros 3 predominam nas outras amostras (IMWONG et al., 2007), sendo que em dados com SNPs foi possível enxergar um ancestral na Colômbia que raramente ocorreu em amostras da Índia, de Madagascar e da Mauritania (HUPALO et al., 2016). Em concordância os dados de SNPs de outros autores, as nossas amostras colombianas apresentaram majoritariamente um ancestral. Dados de microssatélites da região de Chiapas, no México, mesma região de nossas amostras genômicas, mostraram a presença forte de 3 ancestrais distintos (JOY et al., 2008). Em contraste, os genomas do México que analisamos têm um único ancestral predominante, mesmo sendo provenientes de regiões distintas dentro do estado de Chiapas.

Dados de microssatélites em amostras brasileiras revelaram divergência entre 4 diferentes localidades (REZENDE et al., 2010). Esse padrão de diversidade pode ter sido resultado de poucos parasitos com sua origem claramente atribuída a uma única subpopulação e em outros

tantos com mistura desse ancestral. Essa inferência se torna mais óbvia quando olhamos atentamente ao resultado da nossa análise de clusterização com o banco de dados de amostras brasileiras, provindas de mesmo local de coleta, ou seja, simpátricas, e tendo um único ancestral presente.

4.2 Diversidade e seleção em populações de *P. vivax* americanas

Estudos prévios de diversos autores mostraram que *P. vivax* é mais diverso comparado a *P. falciparum*, sugerindo um maior e/ou mais estável tamanho efetivo populacional (HUPALO et al., 2016; NEAFSEY et al., 2012; PEARSON et al., 2016), mesmo quando o banco de dados de *P. vivax* é menor que o de *P. falciparum* (HUPALO et al., 2016). Para analisar essa diversidade, obtivemos dados de diversidade nucleotídica em cada uma das regiões de interesse (Brasil, Colômbia, Peru e México). Os resultados mostraram que a maior parte da diversidade ocorre em famílias de genes previamente descritos como hipervariáveis, como *vir*, MSP e SERA (CARLTON et al., 2008).

Durante a última fase de invasão das hemácias, diversas proteínas de superfície de merozoítos (MSP-1, MSP-6, MSP-7 e MSP-9) formam um complexo que parece ser essencial na ligação a receptores (por exemplo, banda 3) da superfície das células hospedeiras (ROY et al., 2009). Os antígenos envolvidos nesse complexo são muito polimórficos, aparentemente sujeitos à seleção balanceada (WEEDALL; CONWAY, 2010), constituindo-se em alvos importantes imunidade naturalmente adquirida. Além disso, temos que a grande quantidade relativa de polimorfismos, é explicada pelo acúmulo de substituições de aminoácidos que permitem que o parasita consiga evitar a resposta imune do hospedeiro (ESCALANTE et al., 2004; WEEDALL; CONWAY, 2010).

Em um olhar mais atento as outras regiões de picos de diversidade podemos observar, no caso do Brasil, genes envolvidos com formação de filamentos de actina, codificante de *TATA-binding* e com domínio *Phrist*. A produção de actina é essencial para o desenvolvimento do ciclo do parasito; a forma do parasita transmitida pelo mosquito, chamada esporozoito, precisa de motilidade para entrar nas glândulas salivares, para migrar pela pele e para entrar nos capilares sanguíneos e eventualmente hepatócitos. A motilidade dos esporozoitos baseia-se exatamente no motor de actina-miosina (BANE et al., 2016). No caso de gene codificante de *TATA-binding*, a simples existência de um gene homólogo ao gene TBP de *P. falciparum*, em *P. vivax*, sugere a existência de *TATA box* ou um equivalente ao *motif* de *PfTBP-binding* (RUVALCABA-SALAZAR et al., 2005). Quanto ao gene com domínio *Phrist*, existem algumas evidências de que as proteínas codificadas por esse gene se localizam na membrana do parasito (OAKLEY et al., 2007).

Para inferir seleção natural em escala genômica, utilizamos o teste D de Tajima em janelas de 1 kb. O propósito desse teste é identificar sequências ou regiões que não se encaixam

no modelo teórico de neutralidade em um contexto de equilíbrio mutação-deriva. Valores positivos do teste D de Tajima sugerem a manutenção de múltiplos alelos na população - alguns em baixa e outros em alta frequência - em consequência de seleção balanceada. Entre os SNPs associados com valores positivos de teste D de Tajima em nossa população, encontramos mais exemplos daqueles situados em genes envolvidos com a interação parasito-hospedeiro do que associados ao ciclo de vida e a fertilidade do parasito. Desse primeiro grande grupo, os genes com maior destaque são: TLP2 (proteção contra estresse oxidativo), CSP (invasão do esporozoítos nos hepatócitos), RAD (evasão da resposta do sistema imune), SRPK1 (associado à proliferação do parasito) e CuTP (associado à fertilidade). O gene ApiAP2, que foi encontrado associado a altos valores de D de Tajima nas 4 populações analisadas, está associado à regulação da transcrição. Já entre os SNPs com valores de teste D de Tajima menores que 0, que podem representar a perda de alelos raros na população como resultado de seleção negativa, temos como representantes os genes MSP5 (superfície de estágios sanguíneos), RAD (evasão da resposta do sistema imune) e PSOP24 (associado ao estágio de oocineto) (Tabela 6).

A maioria dos genes identificados neste estudo apresentaram índices positivos de seleção balanceada em estudos prévios, ou possuem ortólogos em *P. falciparum* (OCHOLLA et al., 2014). Lembramos que a evidência de seleção natural, positiva ou negativa, em genes que codificam antígenos podem indicar variação ou restrição funcional que são imunologicamente relevantes.

4.3 Desequilíbrio de ligação

Em todas as populações estudadas americanas, o declínio do desequilíbrio de ligação em função da distância nucleotídica foi abrupto e rápido, atingindo menos da metade do valor de r^2 em uma distância de 50 pb. Esse rápido declínio foi previamente observado tanto em *P. vivax* (em amostras do Sudeste Asiático e da Oceania (PEARSON et al., 2016), quanto em *P. falciparum* (em amostras de genes de populações africanas (CONWAY et al., 1999); ainda assim, o declínio descrito nesses exemplos anteriores é menos abrupto do que aquele observado em nossas amostras, com a possível exceção de algumas amostras africanas de *P. falciparum* (NEAFSEY et al., 2008). Esses resultados são compatíveis com elevadas taxas de recombinação meiótica nas populações de *P. vivax* das Américas, a despeito dos baixos níveis de transmissão de malária observados na maioria das áreas endêmicas de malária desse continente. Um fator importante para assegurar uma taxa de endocruzamento relativamente baixa em *P. vivax* é a elevada proporção de hospedeiros que albergam simultaneamente múltiplos clones geneticamente distintos de *P. vivax*, mesmo em regiões com baixos níveis de transmissão de malária (KOEPLI et al., 2015). A presença de infecções multiclonais torna mais provável a fecundação cruzada entre gametas geneticamente distintos do parasito no interior do inseto vetor, elevando a taxa de recombinação efetiva (CONWAY et al., 1999) mesmo em regiões com baixa transmissão, como o caso do Brasil.

5 CONCLUSÃO

Encontramos divergência entre as populações americanas estudadas, com uma tendência a ancestrais distintos em função ao aumento da distância geográfica. Deste modo, as populações de *P. vivax* do Brasil compartilham mais ancestrais com o Peru e as da Colômbia com as populações do México. A divergência entre populações do mesmo continente é evidente quando se analisa o conjunto de SNPs observados nos genomas nucleares completos, porém não é claramente observada quando se analisa um conjunto restrito de 42 marcadores propostos como um *barcode* para essa espécie.

Os níveis de diversidade encontrados em nossas análises mostraram similaridade com o proposto na literatura, em que genomas de *P. vivax* possuem mais diversidade do que genomas de *P. falciparum*, mesmo em uma população simpátrica como a brasileira. A alta taxa de recombinação meiótica encontrada em amostras da população brasileira sugere que a recombinação seja um importante mecanismo de geração de diversidade genética nessa espécie a partir de polimorfismos presentes nos ancestrais e de novos polimorfismos gerados por mutações.

Genes de famílias previamente descritas como hipervariáveis foram observados com os maiores valores de diversidade, seguidos, no caso do Brasil, de genes codificantes de fatores de transcrição e de actina. Identificamos, sob seleção balanceada, mais genes envolvidos com a interação parasito-hospedeiro do que associados ao ciclo de vida e a fertilidade, apesar do único gene com alto valor de D de Tajima encontrado nas 4 regiões, *ApiAP2*, estar associado a regulação da transcrição.

REFERÊNCIAS

- ABDULLAH, N. R.; BARBER, B. E.; WILLIAM, T.; NORAHMAD, N. A.; SATSU, U. R.; MUNIANDY, P. K.; ISMAIL, Z.; GRIGG, M. J.; JELIP, J.; PIERA, K.; Von Seidlein, L.; YEO, T. W.; ANSTEY, N. M.; PRICE, R. N.; AUBURN, S. Plasmodium vivax population structure and transmission dynamics in Sabah Malaysia. **PLoS ONE**, v. 8, n. 12, 2013.
- ALECRIM, G. C.; ALECRIM, W.; MACÊDO, V. Plasmodium vivax resistance to chloroquine (R2) and mefloquine (R3) in Brazilian Amazon region. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 32, n. 1, p. 67–68, 1999.
- ALEXANDER, D.; LANGE, K. A quasi-Newton acceleration for high-dimensional optimization algorithms. **Statistics and Computing**, v. 21, n. 2, p. 261–273, 2011.
- ANDRADE, B. B.; REIS-FILHO, A.; SOUZA-NETO, S. M.; CLARÊNCIO, J.; CAMARGO, L. M. A.; BARRAL, A.; BARRAL-NETTO, M. Severe Plasmodium vivax malaria exhibits marked inflammatory imbalance. **Malaria Journal**, v. 9, p. 13, 2010.
- ANSTEY, N. M.; RUSSELL, B.; YEO, T. W.; PRICE, R. N. The pathophysiology of vivax malaria. **Trends in Parasitology**, v. 25, n. 5, p. 220–227, 2009.
- BAIRD, J. K. Chloroquine Resistance in Plasmodium vivax. **Antimicrobial Agents and Chemotherapy**, v. 48, n. 11, p. 4075–4083, 2004.
- BAIRD, J. K. Neglect of Plasmodium vivax malaria. **Trends in Parasitology**, v. 23, n. 11, p. 533–539, 2007.
- BANE, K. S.; LEPPER, S.; KEHRER, J.; SATTTLER, J. M.; SINGER, M.; REINIG, M.; KLUG, D.; HEISS, K.; BAUM, J.; MUELLER, A. K.; FRISCHKNECHT, F. The Actin Filament-Binding Protein Coronin Regulates Motility in Plasmodium Sporozoites. **PLoS Pathogens**, v. 12, n. 7, 2016.
- BANIECKI, M. L.; FAUST, A. L.; SCHAFFNER, S. F.; PARK, D. J.; GALINSKY, K.; DANIELS, R. F.; HAMILTON, E.; FERREIRA, M. U.; KARUNAWEEERA, N. D.; SERRE, D.; ZIMMERMAN, P. A.; S??, J. M.; WELLEMS, T. E.; MUSSET, L.; LEGRAND, E.; MELNIKOV, A.; NEAFSEY, D. E.; VOLKMAN, S. K.; WIRTH, D. F.; SABETI, P. C. Development of a Single Nucleotide Polymorphism Barcode to Genotype Plasmodium vivax Infections. **PLoS Neglected Tropical Diseases**, v. 9, n. 3, 2015.
- BARBOSA, S.; GOZZE, A. B.; LIMA, N. F.; BATISTA, C. L.; BASTOS, M. d. S.; NICOLETE, V. C.; FONTOURA, P. S.; GONCALVES, R. M.; VIANA, S. A. S.; MENEZES, M. J.; SCOPEL, K. K. G.; CAVASINI, C. E.; MALAFRONTI, R. d. S.; SILVA-NUNE, M. d.; VINETZ, J. M.; CASTRO, M. C.; FERREIRA, M. U. Epidemiology of Disappearing Plasmodium vivax Malaria : A Case Study in Rural Amazonia. **PLoS Neglected Tropical Diseases**, v. 8, n. 8, 2014.
- BATTLE, K. E.; GETHING, P. W.; ELYAZAR, I. R. F.; MOYES, C. L.; SINKA, M. E.; HOWES, R. E.; GUERRA, C. A.; PRICE, R. N.; BAIRD, K. J.; HAY, S. I. **The Global Public Health Significance of Plasmodium vivax**. [S.l.]: Elsevier, 2012. 1–111 p. ISBN 9780123979001.

- BETANZOS, A. Challenges and progress in the elimination of malaria in Mexico. **Boletín Médico del Hospital Infantil de México**, v. 68, n. 2, p. 159–168, 2011.
- BLACK, R. E.; COUSENS, S.; JOHNSON, H. L.; LAWN, J. E.; RUDAN, I.; BASSANI, D. G.; JHA, P.; CAMPBELL, H.; WALKER, C. F.; CIBULSKIS, R.; EISELE, T.; LIU, L.; MATHERS, C. Global, regional, and national causes of child mortality in 2008: a systematic analysis. **The Lancet**, Elsevier Ltd, v. 375, n. 9730, p. 1969–1987, 2010.
- BOCKARIE, M. J.; DAGORO, H. Plasmodium falciparum than Plasmodium vivax- infected mosquitoes ? **Malaria Journal**, v. 5, p. 1–5, 2006.
- BOUSEMA, T.; DRAKELEY, C. Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination. **Clinical Microbiology Reviews**, v. 24, n. 2, p. 377–410, 2011.
- BOZDECH, Z.; MOK, S.; HU, G.; IMWONG, M.; JAIDEE, A.; RUSSELL, B.; GINSBURG, H.; NOSTEN, F.; DAY, N. P. J.; WHITE, N. J.; CARLTON, J. M.; PREISER, P. R. The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 42, p. 16290–16295, 2008.
- BRANCH, O.; CASAPIA, W. M.; GAMBOA, D. V.; HERNANDEZ, J. N.; ALAVA, F. F.; RONCAL, N.; ALVAREZ, E.; PEREZ, E. J.; GOTUZZO, E. Clustered local transmission and asymptomatic Plasmodium falciparum and Plasmodium vivax malaria infections in a recently emerged, hypoendemic Peruvian Amazon community. **Malaria Journal**, v. 4, p. 27, 2005.
- BREGA, S.; MESLIN, B.; MONBRISON, D.; SEVERINI, C.; GRADONI, L.; UDOMSANG-PETCH, R.; SUTANTO, I. Identification of the Plasmodium vivax mdr- Like Gene (pvmdr1) and Analysis of Single-Nucleotide Polymorphisms among Isolates from Different Areas of Endemicity. **Journal of Infectious Diseases**, v. 191, 2005.
- BRIGHT, A. T.; TEWHEY, R.; ABELES, S.; CHUQUIYAURI, R.; LLANOS-CUENTAS, A.; FERREIRA, M. U.; SCHORK, N. J.; VINETZ, J. M.; WINZELER, E. A. Whole genome sequencing analysis of Plasmodium vivax using whole genome capture. **BMC Genomics**, v. 13, n. 1, p. 262, 2012.
- CARLTON, J. M.; ADAMS, J. H.; SILVA, J. C.; BIDWELL, S. L.; LORENZI, H.; CALER, E.; CRABTREE, J.; ANGIUOLI, S. V.; MERINO, E. F.; AMEDEO, P.; CHENG, Q.; COULSON, R. M. R.; CRABB, B. S.; Del Portillo, H. A.; ESSIEN, K.; FELDBLYUM, T. V.; FERNANDEZ-BECERRA, C.; GILSON, P. R.; GUEYE, A. H.; GUO, X.; KANG'A, S.; KOUIJ, T. W. A.; KORSINCZKY, M.; MEYER, E. V.-S.; NENE, V.; PAULSEN, I.; WHITE, O.; RALPH, S. A.; REN, Q.; SARGEANT, T. J.; SALZBERG, S. L.; STOECKERT, C. J.; SULLIVAN, S. A.; YAMAMOTO, M. M.; HOFFMAN, S. L.; WORTMAN, J. R.; GARDNER, M. J.; GALINSKI, M. R.; BARNWELL, J. W.; FRASER-LIGGETT, C. M. Comparative genomics of the neglected human malaria parasite Plasmodium vivax. **Nature**, v. 455, n. 7214, p. 757–763, 2008.
- CARTER, K. H.; SINGH, P.; MUJICA, O. J.; ESCALADA, R. P.; ADE, M. P.; CASTELLANOS, L. G.; ESPINAL, M. A. Malaria in the Americas: Trends from 1959 to 2011. **American Journal of Tropical Medicine and Hygiene**, v. 92, n. 2, p. 302–316, 2015.

CHAN, E. R.; BARNWELL, J. W.; ZIMMERMAN, P. A.; SERRE, D. Comparative Analysis of Field-Isolate and Monkey-Adapted *Plasmodium vivax* Genomes. **PLoS Neglected Tropical Diseases**, v. 9, n. 3, 2015.

CHAN, E. R.; MENARD, D.; DAVID, P. H.; RATSIMBASOA ARSÈ, n.; KIM, S.; CHIM, P.; DO, C.; WITKOWSKI, B.; MERCEREAU-PUIJALON, O.; ZIMMERMAN, P. A.; SERRE, D. Whole Genome Sequencing of Field Isolates Provides Robust Characterization of Genetic Diversity in *Plasmodium vivax*. **PLoS Neglected Tropical Diseases**, v. 6, n. 9, 2012.

CHANG, H.-H.; PARK, D. J.; GALINSKY, K. J.; SCHAFFNER, S. F.; NDIAYE, D.; NDIR, O.; MBOUP, S.; WIEGAND, R. C.; VOLKMAN, S. K.; SABETI, P. C.; WIRTH, D. F.; NEAFSEY, D. E.; HARTL, D. L. Genomic sequencing of *Plasmodium falciparum* malaria parasites from Senegal reveals the demographic history of the population. **Molecular biology and Evolution**, v. 29, n. 11, p. 3427–3439, 2012.

CHAPILLIQUÉN, F. Situación epidemiológica de la malaria (a la SE 53). **Ministry Health of Peru**, 2015.

CHEHUAN, Y. F.; COSTA, M. R. F.; COSTA, J. S.; ALECRIM, M. G. C.; NOGUEIRA, F.; SILVEIRA, H.; BRASIL, L. W.; MELO, G. C.; MONTEIRO, W. M.; LACERDA, M. V. G. In vitro chloroquine resistance for *Plasmodium vivax* isolates from the Western Brazilian Amazon. **Malaria Journal**, p. 1–5, 2013.

CHENET, S. M.; SCHNEIDER, K. a.; VILLEGAS, L.; ESCALANTE, A. a. Local population structure of *Plasmodium*: impact on malaria control and elimination. **Malaria Journal**, *Malaria Journal*, v. 11, n. 1, p. 412, 2012.

CINGOLANI, P.; PLATTS, A.; WANG, L. L.; COON, M.; NGUYEN, T.; WANG, L.; LAND, S. J.; LU, X.; RUDEN, D. M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w 1118; iso-2; iso-3. **Fly**, v. 6, n. 2, p. 80–92, 2012.

CONWAY, D. J.; ROPER, C.; ODUOLA, A. M. J.; ARNOT, D. E.; KREMSNER, P. G.; GROBUSCH, M. P.; CURTIS, C. F.; GREENWOOD, B. M. High recombination rate in natural populations of *Plasmodium falciparum*. **Proceedings of the National Academy of Sciences**, v. 96, n. 8, p. 4506–4511, 1999.

CULLETON, R.; COBAN, C.; ZEYREK, F. Y.; CRAVO, P.; KANEKO, A.; RANDRIANARIVE-LOJOSIA, M.; ANDRIANARANJAKA, V.; KANO, S.; FARNERT, A.; AREZ, A. P.; SHARP, P. M.; CARTER, R.; TANABE, K. The origins of african *plasmodium vivax*; insights from mitochondrial genome sequencing. **PLoS ONE**, v. 6, n. 12, 2011.

DANECEK, P.; AUTON, A.; ABECASIS, G.; ALBERS, C. A.; BANKS, E.; DEPRISTO, M. A.; HANDSAKER, R. E.; LUNTER, G.; MARTH, G. T.; SHERRY, S. T.; MCVEAN, G.; DURBIN, R. The variant call format and VCFtools. **Bioinformatics**, v. 27, n. 15, p. 2156–2158, 2011.

DANIELS, R.; VOLKMAN, S. K.; MILNER, D. A.; MAHESH, N.; NEAFSEY, D. E.; PARK, D. J.; ROSEN, D.; ANGELINO, E.; SABETI, P. C.; WIRTH, D. F.; WIEGAND, R. C. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. **Malaria Journal**, v. 7, p. 223, 2008.

- DELGADO-RATTO, C.; GAMBOA, D.; SOTO-CALLE, V. E.; DEN, P. V.; TORRES, E.; SÁNCHEZ-MARTÍ NEZ, L.; CONTRERAS-MANCILLA, J.; ROSANAS, A.; VAN, P.; ALESSANDRO, U. D. Population Genetics of *Plasmodium vivax* in the Peruvian Amazon. **PLoS Neglected Tropical Diseases**, 2016.
- DELGADO-RATTO, C.; SOTO-CALLE, V. E.; Van den Eede, P.; GAMBOA, D.; ROSAS, A.; ABATIH, E. N.; Rodriguez Ferrucci, H.; LLANOS-CUENTAS, A.; Van Geertruyden, J.-P.; ERHART, A.; D'ALESSANDRO, U. Population structure and spatio-temporal transmission dynamics of *Plasmodium vivax* after radical cure treatment in a rural village of the Peruvian Amazon. **Malaria Journal**, v. 13, p. 8, 2014.
- DHARIA, N. V.; BRIGHT, a. T.; WESTENBERGER, S. J.; BARNES, S. W.; BATALOV, S.; KUHEN, K.; BORBOA, R.; FEDERE, G. C.; MCCLEAN, C. M.; VINETZ, J. M.; NEYRA, V.; LLANOS-CUENTAS, A.; BARNWELL, J. W.; WALKER, J. R.; WINZELER, E. a. Whole-genome sequencing and microarray analysis of ex vivo *Plasmodium vivax* reveal selective pressure on putative drug resistance genes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 46, p. 20045–20050, 2010.
- ESCALANTE, A. A.; CORNEJO, O. E.; ROJAS, A.; UDHAYAKUMAR, V.; LAL, A. A. Assessing the effect of natural selection in malaria parasites. **Trends in Parasitology**, v. 20, n. 8, p. 388–395, 2004.
- FENG, X.; CARLTON, J. M.; JOY, D. A.; MU, J.; FURUYA, T.; SUH, B. B.; WANG, Y.; BARNWELL, J. W.; SU, X.-Z. Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 14, p. 8502–8507, 2003.
- FERREIRA, M. U.; CASTRO, M. C. Challenges for malaria elimination in Brazil. **Malaria Journal**, BioMed Central, v. 15, n. 1, p. 284, 2016.
- FERREIRA, M. U.; KARUNAWEEERA, N. D.; SILVA-NUNES, M. da; SILVA, N. S. da; WIRTH, D. F.; HARTL, D. L. Population structure and transmission dynamics of *Plasmodium vivax* in rural Amazonia. **Journal of infectious diseases**, v. 195, p. 1218–1226, 2007.
- FINN, R. D.; COGGILL, P.; EBERHARDT, R. Y.; EDDY, S. R.; MISTRY, J.; MITCHELL, A. L.; POTTER, S. C.; PUNTA, M.; QURESHI, M.; SANGRADOR-VEGAS, A.; SALAZAR, G. A.; TATE, J.; BATEMAN, A. The Pfam protein families database: Towards a more sustainable future. **Nucleic Acids Research**, v. 44, n. D1, p. 279–285, 2016.
- GARZÓN-OSPINA, D.; FORERO-RODRÍ GUEZ, J.; PATARROYO, M. A. Heterogeneous genetic diversity pattern in *Plasmodium vivax* genes encoding merozoite surface proteins (MSP) -7E, -7F and -7L. **Malaria Journal**, v. 13, n. 1, p. 495, 2014.
- GARZÓN-OSPINA, D.; FORERO-RODRÍ GUEZ, J.; PATARROYO, M. A. Inferring natural selection signals in *Plasmodium vivax*-encoded proteins having a potential role in merozoite invasion. **Infection, Genetics and Evolution**, v. 33, p. 182–188, 2015.
- GARZÓN-OSPINA, D.; LÓPEZ, C.; FORERO-RODRÍ GUEZ, J.; PATARROYO, M. A. Genetic Diversity and Selection in Three *Plasmodium vivax* Merozoite Surface Protein 7 (Pvmsp-7) Genes in a Colombian Population. **PLoS ONE**, v. 7, n. 9, 2012.

GOMES, L. R.; OLIVEIRA, N. K. Almeida-de; LAVIGNE, A. R. de; LIMA, S. R. F. de; PINA-COSTA, A. de; BRASIL PATRÍ, c.; DANIEL-RIBEIRO, C. T.; MÉNARD, D.; CRUZ, M. D. F. Ferreira-da. Plasmodium vivax mdr1 genotypes in isolates from successfully cured patients living in endemic and non-endemic Brazilian areas. **Malaria Journal**, BioMed Central, v. 15, n. 1, p. 96, 2016.

GONÇALVES LÍ, g. A.; CRAVO, P.; FERREIRA, M. U. Emerging Plasmodium vivax resistance to chloroquine in South America : an overview. **Memórias do Instituto Oswaldo Cruz**, v. 109, n. August, p. 534–539, 2014.

GONZALEZ-CERON, L.; MU, J.; SANTILLÁN, F.; JOY, D.; SANDOVAL, M. a.; CAMAS, G.; SU, X.; CHOY, E. V.; TORREBLANCA, R. Molecular and epidemiological characterization of Plasmodium vivax recurrent infections in southern Mexico. **Parasites & vectors**, v. 6, p. 109, 2013.

GRAF, P. C.; DURAND, S.; Alvarez Antonio, C.; MONTALVAN, C.; Galves Montoya, M.; GREEN, M. D.; SANTOLALLA, M. L.; SALAS, C.; LUCAS, C.; BACON, D. J.; FRYAUFF, D. J. Failure of Supervised Chloroquine and Primaquine Regimen for the Treatment of Plasmodium vivax in the Peruvian Amazon. **Malaria Research an Treatment**, v. 2012, p. 936067, 2012.

GREGSON, A.; PLOWE, C. V. Mechanisms of Resistance of Malaria Parasites to Antifolates. **Pharmacological Reviews**, v. 57, n. 1, p. 117–145, 2005.

GRIFFING, S. M.; GAMBOA, D.; UDHAYAKUMAR, V. The history of 20th century malaria control in Peru. **Malaria Journal**, v. 12, n. 1, p. 303, 2013.

GUNAWARDENA, S.; FERREIRA, M. U.; KAPILANANDA, G. M. G.; WIRTH, D. F.; KARUNAWEEERA, N. D. The Sri Lankan paradox: high genetic diversity in Plasmodium vivax populations despite decreasing levels of malaria transmission. **Parasitology**, v. 141, p. 880–890, 2014.

HERNÁNDEZ, D. C.; GUERRA, A. P.; MILENA, C. Z.; NICHOLLS, R. S.; BARRERA, S. M. Haplotypes associated with resistance to sulfadoxine-pyrimethamine in Plasmodium falciparum in two malaria endemic locations in Colombia. **Infection, Genetics and Evolution**, v. 18, p. 183–190, 2013.

HORII, T.; SHIRAI, H.; JIE, L.; ISHII, K. J.; PALACPAC, N. Q.; TOUGAN, T.; HATO, M.; OHTA, N.; BOBOGARE, A.; ARAKAKI, N.; MATSUMOTO, Y.; NAMAZUE, J.; ISHIKAWA, T.; UEDA, S.; TAKAHASHI, M. Evidences of protection against blood-stage infection of Plasmodium falciparum by the novel protein vaccine SE36. **Parasitology International**, Elsevier Ireland Ltd, v. 59, n. 3, p. 380–386, 2010.

HUPALO, D. N.; LUO, Z.; MELNIKOV, A.; SUTTON, P. L.; ROGOV, P.; ESCALANTE, A.; VALLEJO, A. F.; HERRERA, S.; ARÉVALO-HERRERA, M.; FAN, Q.; WANG, Y.; CUI, L.; LUCAS, C. M.; DURAND, S.; SANCHEZ, J. F.; BALDEVIANO, G. C.; LESCOANO, A. G.; LAMAN, M.; GAMBOA, D.; KOSEK, M.; VINETZ, J. M.; GONZÁLEZ-CERÓN, L.; BIRREN, B. W.; NEAFSEY, D. E.; CARLTON, J. M. Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. **Nature Genetics**, v. 48, n. May, p. 953–959, 2016.

HYDE, J. E. The dihydrofolate reductase-thymidylate synthetase gene in the drug resistance of malaria parasites. **Pharmacology & Therapeutics**, v. 48, n. 1, p. 45–59, 1990. ISSN 01637258.

IMWONG, M.; NAIR, S.; PUKRITTAYAKAMEE, S.; SUDIMACK, D.; WILLIAMS, J. T.; MAYXAY, M.; NEWTON, P. N.; KIM, J. R.; NANDY, A.; OSORIO, L.; CARLTON, J. M.; WHITE, N. J.; DAY, N. P. J.; ANDERSON, T. J. C. Contrasting genetic structure in *Plasmodium vivax* populations from Asia and South America. **International Journal for Parasitology**, v. 37, n. 8-9, p. 1013–1022, 2007.

IMWONG, M.; PUKRITTAYAKAMEE, S.; RÉNIA, L.; LETOURNEUR, F.; CHARLIEU, J.-p.; LOOAREESUWAN, S.; NICHOLAS, J.; SNOUNOU, G.; RE, L.; LEARTSAKULPANICH, U. Novel Point Mutations in the Dihydrofolate Reductase Gene of *Plasmodium vivax* : Evidence for Sequential Selection by Drug Pressure Novel Point Mutations in the Dihydrofolate Reductase Gene of *Plasmodium vivax* : Evidence for Sequential Selection by Drug Pr. **Antimicrobial Agents and Chemotherapy**, v. 47, n. 5, p. 1514–1521, 2003.

JOMBART, T. Analysing genome-wide SNP data using adegenet. **Software User Manual**, p. 1–37, 2012.

JOSHI, H.; PRAJAPATI, S. K.; VERMA, A.; KANG'A, S.; CARLTON, J. M. *Plasmodium vivax* in India. **Trends in Parasitology**, v. 24, n. 5, p. 228–235, 2008.

JOY, D. A.; GONZALEZ-CERON, L.; CARLTON, J. M.; GUEYE, A.; FAY, M.; MCCUTCHAN, T. F.; SU, X. Z. Local adaptation and vector-mediated population structure in *Plasmodium vivax* malaria. **Molecular Biology and Evolution**, v. 25, n. 6, p. 1245–1252, 2008.

KARUNAWEEERA, N. D.; FERREIRA, M. U.; MUNASINGHE, A.; BARNWELL, J. W.; COLLINS, W. E.; KING, C. L.; KAWAMOTO, F.; HARTL, D. L.; WIRTH, D. F. Extensive microsatellite diversity in the human malaria parasite *Plasmodium vivax*. **Gene**, v. 410, n. 1, p. 105–112, 2008.

KOEPFLI, C.; RODRIGUES, P. T.; ANTAO, T.; ORJUELA-SÁNCHEZ, P.; EEDE, P. V. D.; GAMBOA, D.; HONG, N. V.; BENDEZU, J. *Plasmodium vivax* Diversity and Population Structure across Four Continents. **PLOS Neglected Tropical Diseases**, 2015.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–359, 2012. ISSN 1548-7105.

LARRAÑAGA, N.; MEJÍA, R. E.; HORMAZA, J. I.; MONTOYA, A.; SOTO, A.; FONTECHA, G. A. Genetic structure of *Plasmodium falciparum* populations across the Honduras-Nicaragua border. **Malaria Journal**, v. 12, p. 354, 2013.

LEE, K. S.; KIM, T. H.; KIM, E. S.; LIM, H.-s.; YEOM, J.-s.; JUN, G.; PARK, J.-w. Short Report : Chloroquine-resistant *Plasmodium vivax* in the Republic of Korea. **The American Journal of Tropical Medicine and Health**, v. 80, n. 2, p. 215–217, 2009.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNEL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 2009.

LIU, W.; LI, Y.; SHAW, K. S.; LEARN, G. H.; PLENDERLEITH, L. J.; MALENKE, J. a.; SUNDARARAMAN, S. a.; RAMIREZ, M. a.; CRYSTAL, P. a.; SMITH, A. G.; BIBOLLET-RUCHE, F.; AYOUBA, A.; LOCATELLI, S.; ESTEBAN, A.; MOUACHA, F.; GUICHET, E.; BUTEL, C.; AHUKA-MUNDEKE, S.; INOGWABINI, B.-I.; NDJANGO, J.-B. N.; SPEEDE, S.; SANZ, C. M.; MORGAN, D. B.; GONDER, M. K.; KRANZUSCH, P. J.; WALSH, P. D.; GEORGIEV, A. V.; MULLER, M. N.; PIEL, A. K.; STEWART, F. a.; WILSON, M. L.; PUSEY, A. E.; CUI, L.; WANG, Z.; FÄ RNERT, A.; SUTHERLAND, C. J.; NOLDER, D.; HART, J. a.; HART, T. B.; BERTOLANI, P.; GILLIS, A.; LEBRETON, M.; TAFON, B.; KIYANG, J.; DJOKO, C. F.; SCHNEIDER, B. S.; WOLFE, N. D.; MPOUDI-NGOLE, E.; DELAPORTE, E.; CARTER, R.; CULLETON, R. L.; SHAW, G. M.; RAYNER, J. C.; PEETERS, M.; HAHN, B. H.; SHARP, P. M. African origin of the malaria parasite *Plasmodium vivax*. **Nature Communications**, v. 5, p. 3346, 2014.

LIU, Y.; AUBURN, S.; CAO, J.; TRIMARSANTO, H.; ZHOU, H.; GRAY, K.-A.; CLARK, T. G.; PRICE, R. N.; CHENG, Q.; HUANG, R.; GAO, Q. Genetic diversity and population structure of *Plasmodium vivax* in Central China. **Malaria Journal**, v. 13, n. 1, p. 262, 2014.

LUXEMBURGER, C.; PEREA, W. A.; DELMAS, G.; PRUJA, C.; PECOU1, B.; MOREN, A. Permethrin-impregnated bed nets for the prevention of malaria in schoolchildren on the thai-burmese border. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 88, n. 2, p. 155–159, 1994.

MalariaGen. Malaria Genomic Epidemiology Network. Banco de Dados. Disponível em: <<https://www.malariagen.net/projects>>. acesso em: 05 jan. 2016.

MAP. Malaria Atlas Project. Banco de Dados. Disponível em: <<http://www.map.ox.ac.uk/>>. acesso em: 12 jul. 2016.

MARQUES, M. M.; COSTA, M. R. F.; Santana Filho, F. S.; VIEIRA, J. L. F.; NASCIMENTO, M. T. S.; BRASIL, L. W.; NOGUEIRA, F.; SILVEIRA, H.; REYES-LECCA, R. C.; MONTEIRO, W. M.; LACERDA, M. V. G.; ALECRIM, M. G. C. *Plasmodium vivax* chloroquine resistance and anemia in the western brazilian amazon. **Antimicrobial Agents and Chemotherapy**, v. 58, n. 1, p. 342–347, 2014.

MARTHA, S. J.; NOMURA, T.; NEVES, J. D. A.; BAIRD, J. K.; WELLEMS, T. E.; Del Portillo, H. A. *Plasmodium vivax*: Allele variants of the *mdr1* gene do not associate with chloroquine resistance among isolates from Brazil, Papua, and monkey-adapted strains. **Experimental Parasitology**, v. 109, n. 4, p. 256–259, 2005.

MCKENNA, A.; HANNA, M.; BANKS, E.; SIVACHENKO, A.; CIBULSKIS, K.; KERNYTSKY, A.; GARIMELLA, K.; ALTSHULER, D.; GABRIEL, S.; DALY, M.; DEPRISTO, M. A. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. **Genome Research**, v. 20, n. 9, p. 1297–1303, 2010.

MELO, G. C.; MONTEIRO, W. M.; SIQUEIRA, A. M.; SILVA, S. R.; MAGALHÃES, B. M. L.; ALENCAR, A. C. C.; KUEHN, A.; Del Portillo, H. A.; FERNANDEZ-BECERRA, C.; LACERDA, M. V. G. Expression levels of *pvcrt-o* and *pvm-dr-1* are associated with chloroquine resistance and severe *Plasmodium vivax* malaria in patients of the Brazilian Amazon. **PLoS One**, v. 9, n. 8, 2014.

MENEGON, M.; DURAND, P.; MENARD, D.; LEGRAND, E.; PICOT, S.; NOUR, B.; DAVIDYANTS, V.; SANTI, F.; SEVERINI, C. Genetic diversity and population structure of

Plasmodium vivax isolates from Sudan, Madagascar, French Guiana and Armenia. **Infection, Genetics and Evolution**, Elsevier B.V., v. 27, p. 244–249, 2014.

MOBEGI, V. A.; LOUA, K. M.; AHOUIDI, A. D.; SATOQUINA, J.; NWAKANMA, D. C.; AMAMBUA-NGWA, A.; CONWAY, D. J. Population genetic structure of *Plasmodium falciparum* across a region of diverse endemicity in West Africa. **Malaria Journal**, v. 11, p. 223, 2012.

NEAFSEY, D. E.; GALINSKY, K.; JIANG, R. H. Y.; YOUNG, L.; SYKES, S. M.; SAIF, S.; GUJJA, S.; GOLDBERG, J. M.; YOUNG, S.; ZENG, Q.; CHAPMAN, S. B.; DASH, A. P.; ANVIKAR, A. R.; SUTTON, P. L.; BIRREN, B. W.; ESCALANTE, A. a.; BARNWELL, J. W.; CARLTON, J. M. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. **Nature Genetics**, Nature Publishing Group, v. 44, n. 9, p. 1046–50, 2012.

NEAFSEY, D. E.; SCHAFFNER, S. F.; VOLKMAN, S. K.; PARK, D.; MONTGOMERY, P.; MILNER, D. A.; LUKENS, A.; ROSEN, D.; DANIELS, R.; HOUDE, N.; CORTESE, J. F.; TYNDALL, E.; GATES, C.; STANGE-THOMANN, N.; SARR, O.; NDIAYE, D.; NDIR, O.; MBOUP, S.; FERREIRA, M. U.; MORAES, S. d. L.; DASH, A. P.; CHITNIS, C. E.; WIEGAND, R. C.; HARTL, D. L.; BIRREN, B. W.; LANDER, E. S.; SABETI, P. C.; WIRTH, D. F. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. **Genome Biology**, v. 9, n. 12, p. 171, 2008.

OAKLEY, M. S. M.; KUMAR, S.; ANANTHARAMAN, V.; ZHENG, H.; MAHAJAN, B.; HAYNES, J. D.; MOCH, J. K.; FAIRHURST, R.; MCCUTCHAN, T. F.; ARAVIND, L. Molecular factors and biochemical pathways induced by febrile temperature in intraerythrocytic *Plasmodium falciparum* parasites. **Infection and Immunity**, v. 75, n. 4, p. 2012–2025, 2007.

OCHOLLA, H.; PRESTON, M. D.; MIPANDO, M.; JENSEN, A. T. R.; CAMPINO, S.; MACINNIS, B.; ALCOCK, D.; TERLOUW, A.; ZONGO, I.; OUDRAOGO, J. B.; DJIMDE, A. A.; ASSEFA, S.; DOUMBO, O. K.; BORRMANN, S.; NZILA, A.; MARSH, K.; FAIRHURST, R. M.; NOSTEN, F.; ANDERSON, T. J. C.; KWIATKOWSKI, D. P.; CRAIG, A.; CLARK, T. G.; MONTGOMERY, J. Whole-genome scans provide evidence of adaptive evolution in malawian *Plasmodium falciparum* isolates. **Journal of Infectious Diseases**, v. 210, n. 12, p. 1991–2000, 2014.

ORJUELA-SÁNCHEZ, P.; KARUNAWEEERA, N. D.; SILVA-NUNES MÔ, n. da; SILVA, N. S. da; SCOPEL, K. K. G.; GONÇALVES, R. M.; AMARATUNGA, C.; M, S. J.; SOCHEAT, D.; FAIRHURST, R. M.; GUNAWARDENA, S.; THAVAKODIRASAH, T.; GALAPATHTHY, G. L. N.; ABEYSINGHE, R.; KAWAMOTO, F.; WIRTH, D. F.; FERREIRA, M. U. Single-nucleotide polymorphism, linkage disequilibrium and geographic structure in the malaria parasite *Plasmodium vivax*: prospects for genome-wide association studies. **BMC Genetics**, v. 11, p. 65, 2010.

ORJUELA-SÁNCHEZ, P.; M., S. J.; BRANDI, M. C. C.; RODRIGUES, P. T.; BASTOS, M. S.; AMARATUNGA, C.; DUONG, S.; FAIRHURST, R. M.; FERREIRA, M. U. Higher microsatellite diversity in *Plasmodium vivax* than in sympatric *Plasmodium falciparum* populations in Pursat, Western Cambodia. **Experimental Parasitology**, v. 134, n. 3, p. 318–326, 2013.

OSORIO, L.; PÉREZ, P.; GONZÁLEZ, I. J. Evaluación de la eficacia de los medicamentos antimaláricos en Tarapacá, Amazonas colombiano. **Biomédica**, v. 27, p. 133–140, 2007.

PACHECO M ANDREÍ, n.; ELANGO, A. P.; RAHMAN, A. A.; FISHER, D.; COLLINS, W. E.; BARNWELL, J. W.; ESCALANTE, A. A. Infection , Genetics and Evolution Evidence of purifying selection on merozoite surface protein 8 (MSP8) and 10 (MSP10) in Plasmodium spp . **Infection, Genetics and Evolution**, v. 12, p. 978–986, 2012.

PAYSEUR, B. A.; JING, P.; HAASL, R. J. A genomic portrait of human microsatellite variation. **Molecular Biology and Evolution**, v. 28, n. 1, p. 303–312, 2011.

PEARSON, R. D.; AMATO, R.; AUBURN, S.; MIOTTO, O.; ALMAGRO-GARCIA, J.; AMARATUNGA, C.; SUON, S.; MAO, S.; NOVIYANTI, R.; TRIMARSANTO, H.; MARFURT, J.; ANSTEY, N. M.; WILLIAM, T.; BONI, M. F.; DOLECEK, C.; TRAN, H. T.; WHITE, N. J.; MICHON, P.; SIBA, P.; TAVUL, L.; HARRISON, G.; BARRY, A.; MUELLER, I.; FERREIRA, M. U.; KARUNAWEEERA, N.; RANDRIANARIVELOJOSIA, M.; GAO, Q.; HUBBART, C.; HART, L.; JEFFERY, B.; DRURY, E.; MEAD, D.; KEKRE, M.; CAMPINO, S.; MANSKE, M.; CORNELIUS, V. J.; MACINNIS, B.; ROCKETT, K. A.; MILES, A.; RAYNER, J. C.; FAIRHURST, R. M.; NOSTEN, F.; PRICE, R. N.; KWIATKOWSKI, D. P. Genomic analysis of local variation and recent evolution in Plasmodium vivax. **Nature Genetics**, Advanced o, n. 8, p. 959–964, 2016.

PICARD TOOLS. Version 2.0.1. Broad Institute. Software. Disponível em: <<http://broadinstitute.github.io/picard/>>. Acesso em: 12 jul . 2016.

PLASMODB. Plasmodium Database. Banco de Dados. Disponível em: <<http://broadinstitute.github.io/picard/>>. acesso em: 06 jan. 2016.

PORNTHANAKASEM, W.; RIANGRUNGROJ, P.; CHITNUMSUB, P.; ITTARAT, W.; KONGKASURIYACHAI, D.; UTHAIPIBULL, C. Role of Plasmodium vivax Dihydropteroate Synthase Polymorphisms in Sulfa Drug Resistance. **Antimicrobial Agents and Chemotherapy**, v. 60, n. 8, p. 4453–4463, 2016.

PRICE, R. N.; DOUGLAS, N. M.; ANSTEY, N. M. New developments in Plasmodium vivax malaria: severe disease and the rise of chloroquine resistance. **Current Opinion in Infectious Diseases**, v. 22, p. 430–435, 2009.

PRICE, R. N.; TJITRA, E.; GUERRA, C. A.; YEUNG, S.; WHITE, N. J.; ANSTEY, N. M. Vivax malaria: Neglected and not benign. **American Journal of Tropical Medicine and Hygiene**, v. 77, n. 79–87, p. 220–227, 2007.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v. 155, n. 2, p. 945–959, 2000.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; BAKKER, P. I. W. de; DALY, M. J.; SHAM, P. C. PLINK: A tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

RAMIREZ, A. P.; BUITRAGO, J. I. G.; Pimentel González, J. P.; Herrera Moráles, A.; CARRASQUILLA, G. Frequency and tendency of malaria in Colombia, 1990 to 2011: a descriptive study. **Malaria Journal**, v. 13, n. 1, p. 202, 2014.

REID, H.; VALLELY, A.; TALEO, G.; TATEM, A. J.; KELLY, G.; RILEY, I.; HARRIS, I.; HENRI, I.; IAMAHER, S.; CLEMENTS, A. C. A. Baseline spatial distribution of malaria prior to an elimination programme in Vanuatu. **Malaria Journal**, v. 9, n. 1, p. 150, 2010.

REZENDE, A. M.; TARAZONA-SANTOS, E.; FONTES, C. J. F.; SOUZA, J. M.; COUTO, A. D. A.; CARVALHO, L. H.; BRITO, C. F. A. Microsatellite loci: Determining the genetic variability of *Plasmodium vivax*. **Tropical Medicine and International Health**, v. 15, n. 6, p. 718–726, 2010.

ROY, S. W.; WEEDALL, G. D.; SILVA, R. L. da; POLLEY, S. D.; FERREIRA, M. U. Sequence diversity and evolutionary dynamics of the dimorphic antigen merozoite surface protein-6 and other *Msp* genes of *Plasmodium falciparum*. **Gene**, v. 443, n. 1-2, p. 12–21, 2009.

RUNGSIHIRUNRAT, K.; SIBLEY, C. H.; MUNGTHIN, M.; NA-BANGCHANG, K. Geographical distribution of amino acid mutations in *Plasmodium vivax* DHFR and DHPS from malaria endemic areas of Thailand. **American Journal of Tropical Medicine and Hygiene**, v. 78, n. 3, p. 462–467, 2008.

RUSSELL, B.; SUWANARUSK, R.; LEK-UTHAI, U. *Plasmodium vivax* genetic diversity: microsatellite length matters. **Trends in Parasitology**, v. 22, n. 9, p. 399–401, 2006.

RUVALCABA-SALAZAR, O. K.; RAM, C.; VARGAS, M.; HERN, R. Recombinant and native *Plasmodium falciparum* TATA-binding-protein binds to a specific TATA box element in promoter regions. **Molecular & Biochemical Parasitology**, v. 140, p. 183–196, 2005.

SACHIDANANDAM, R.; WEISSMAN, D.; SCHMIDT, S. C.; KAKOL, J. M.; STEIN, L. D.; MARTH, G.; SHERRY, S.; MULLIKIN, J. C.; MORTIMORE, B. J.; WILLEY, D. L.; HUNT, S. E.; COLE, C. G.; COGGILL, P. C.; RICE, C. M.; NING, Z.; ROGERS, J.; BENTLEY, D. R.; KWOK, P. Y.; MARDIS, E. R.; YEH, R. T.; SCHULTZ, B.; COOK, L.; DAVENPORT, R.; DANTE, M.; FULTON, L.; HILLIER, L.; WATERSTON, R. H.; MCPHERSON, J. D.; GILMAN, B.; SCHAFFNER, S.; Van Etten, W. J.; REICH, D.; HIGGINS, J.; DALY, M. J.; BLUMENSTIEL, B.; BALDWIN, J.; STANGE-THOMANN, N.; ZODY, M. C.; LINTON, L.; LANDER, E. S.; ALTSHULER, D. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. **Nature**, v. 409, n. 6822, p. 928–933, 2001.

SAMAD, H.; COLL, F.; PRESTON, M. D.; OCHOLLA, H.; FAIRHURST, R. M.; CLARK, T. G. Imputation-Based Population Genetics Analysis of *Plasmodium falciparum* Malaria Parasites. **PLoS Genetics**, v. 11, n. 4, 2015.

SANTANA-FILHO, F. S. de; ARCANJO, A. R.; CHEHUAN, Y. M.; COSTA, M. R.; MARTINEZ-ESPINOSA, F. E.; VIEIRA, J. L.; BARBOSA, M.; ALECRIM, W. D.; ALECRIM, M. Chloroquine-resistant *Plasmodium vivax*, Brazilian Amazon. **Emerging Infectious Diseases**, v. 13, n. 7, p. 1125–1126, 2007.

SCHOUSBOE, M. L.; RANJITKAR, S.; RAJAKARUNA, R. S.; AMERASINGHE, P. H.; KONRADSEN, F.; MORALES, F.; ORD, R.; PEARCE, R.; LESLIE, T.; ROWLAND, M.; GADALLA, N.; BYGBJERG, I. C.; ALIFRANGIS, M.; ROPER, C. Global and local genetic diversity at two microsatellite loci in *Plasmodium vivax* parasites from Asia, Africa and South America. **Malaria Journal**, v. 13, p. 392, 2014.

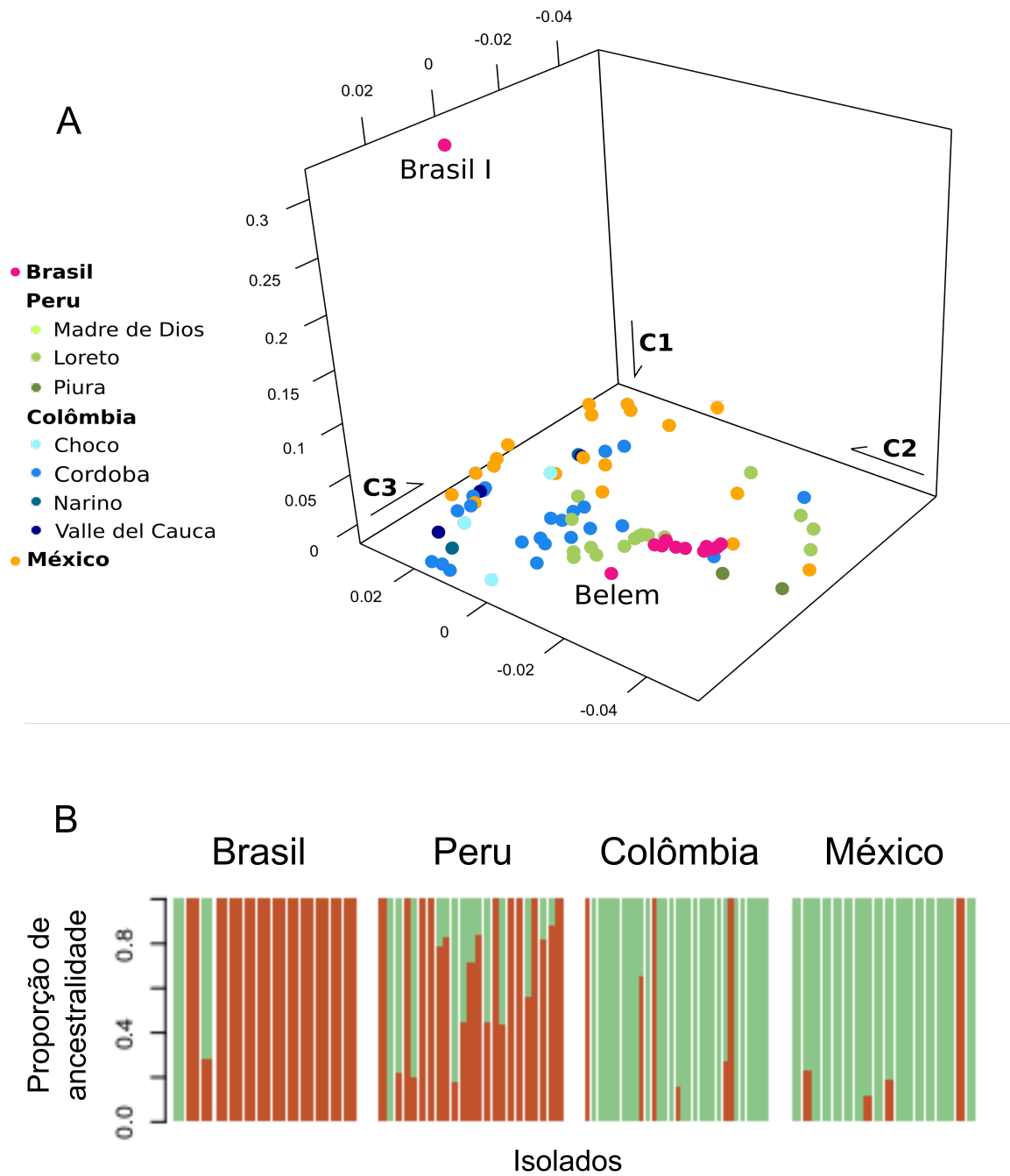
SCHULTZ, L.; WAPLING, J.; MUELLER, I.; NTSUKE, P. O.; SENN, N.; NALE, J.; KINIBORO, B.; BUCKEE, C. O.; TAVUL, L.; SIBA, P. M.; REEDER, J. C.; BARRY, A. E. Multilocus haplotypes reveal variable levels of diversity and population structure of *Plasmodium falciparum* in Papua New Guinea, a region of intense perennial transmission. **Malaria Journal**, v. 9, n. 1, p. 336, 2010.

- SIBLEY, C. H.; HYDE, J. E.; SIMS, P. F.; PLOWE, C. V.; KUBLIN, J. G.; MBERU, E. K.; COWMAN, a. F.; WINSTANLEY, P. a.; WATKINS, W. M.; NZILA, a. M. Pyrimethamine-sulfadoxine resistance in *Plasmodium falciparum*: what next? **Trends in Parasitology**, v. 17, n. 12, p. 582–588, 2001.
- SINGH, V.; GUPTA, P.; PANDE, V. Revisiting the multigene families: *Plasmodium var* and *vir* genes. **Journal of Vector Borne Diseases**, v. 51, n. 2, p. 75–81, 2014.
- SNOUNOU, G.; WHITE, N. J. The co-existence of *Plasmodium*: Sidelights from *falciparum* and *vivax* malaria in Thailand. **Trends in Parasitology**, v. 20, n. 7, p. 333–339, 2004.
- SRA. Sequence Read Archive. Banco de Dados do National Center for Biotechnology Information. Disponível em: <<http://www.ncbi.nlm.nih.gov/sra/>>. acesso em: 03 jan. 2016.
- SUTTON, P. L. A call to arms: on refining *Plasmodium vivax* microsatellite marker panels for comparing global diversity. **Malaria Journal**, v. 12, p. 447, 2013. ISSN 1475-2875.
- VANDENEDEE, P.; VANDERAUWERA, G.; DELGADO, C.; HUYSE, T.; SOTO-CALLE, V. E.; GAMBOA, D.; GRANDE, T.; RODRIGUEZ, H.; LLANOS-CUENTAS, A.; JOZEF, A.; ERHART, A.; D’ALESSANDRO, U. Multilocus genotyping reveals high heterogeneity and strong local population structure of the *Plasmodium vivax* population in the Peruvian Amazon. **Malaria Journal**, v. 9, p. 151, 2010.
- WANG, P.; READ, M.; SIMS, P. F.; HYDE, J. E. Sulfadoxine resistance in the human malaria parasite *Plasmodium falciparum* is determined by mutations in dihydropteroate synthetase and an additional factor associated with folate utilization. **Molecular microbiology**, v. 23, n. 5, p. 979–986, 1997.
- WEEDALL, G. D.; CONWAY, D. J. Detecting signatures of balancing selection to identify targets of anti-parasite immunity. **Trends in Parasitology**, v. 26, n. 7, p. 363–369, 2010.
- WESTENBERGER, S. J.; MCCLEAN, C. M.; CHATTOPADHYAY, R.; DHARIA, N. V.; CARLTON, J. M.; BARNWELL, J. W.; COLLINS, W. E.; HOFFMAN, S. L.; ZHOU, Y.; VINETZ, J. M.; WINZELER, E. A. A systems-based analysis of *Plasmodium vivax* lifecycle transcription from human to mosquito. **PLoS Neglected Tropical Diseases**, v. 4, n. 4, 2010. ISSN 1935-2735.
- WHITE, N.; IMWONG, M. . Relapse. **Advances in Parasitology**, v. 80, p. 113–150, 2012.
- WHO. World Malaria report. 2012.
- WHO. World Malaria Report. 2014.

¹ De acordo com: ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 6023: informação e documentação: referências: elaboração. Rio de Janeiro, 2002.

ANEXOS

Figura S1 – Estrutura populacional baseada em SNPs de 86 amostras de amostras de *P. vivax* nas Américas



(A) Análise de Componentes Principais com três *eigenvectors*. Cada isolado foi colorido de acordo com sua origem, e percentuais dos *eigenvectors* são: C1 = 35,86%, C2 = 14,36% e C3 = 9,32%; (B) Análises de ADMIXTURE com o banco de dados total de SNPs usando valor de K=2

Figura S2 – valores de *clusters* populacionais do ADMIXTURE obtidos segundo o menor valor de erro encontrado para K durante corridas de entre 2 e 10, com 84 amostras das quatro regiões da América (Brasil, Colômbia, México e Peru)

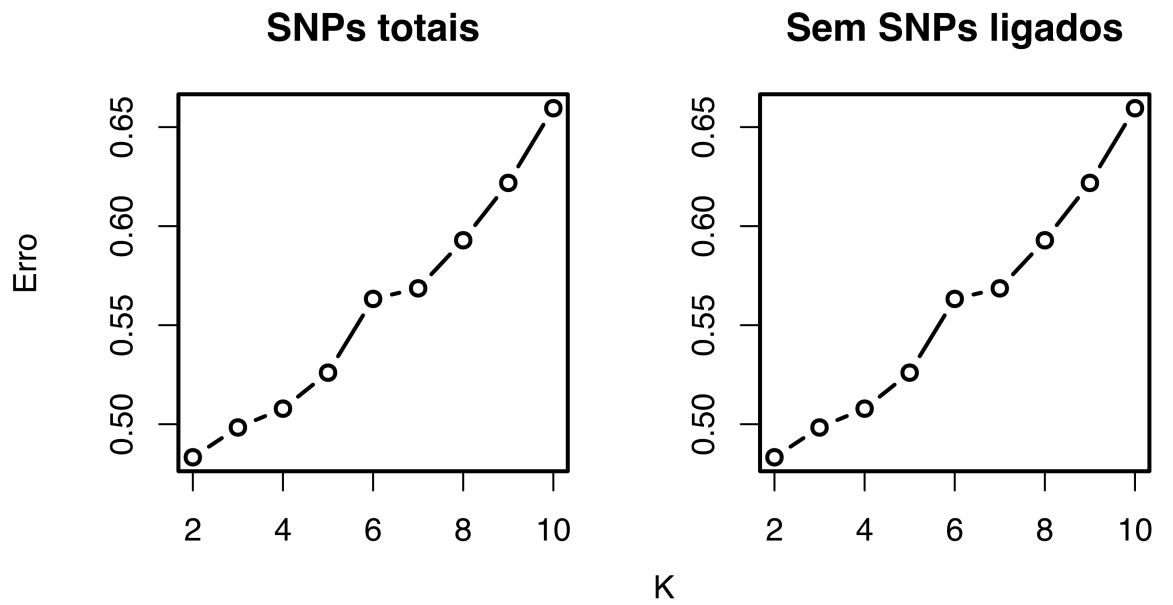
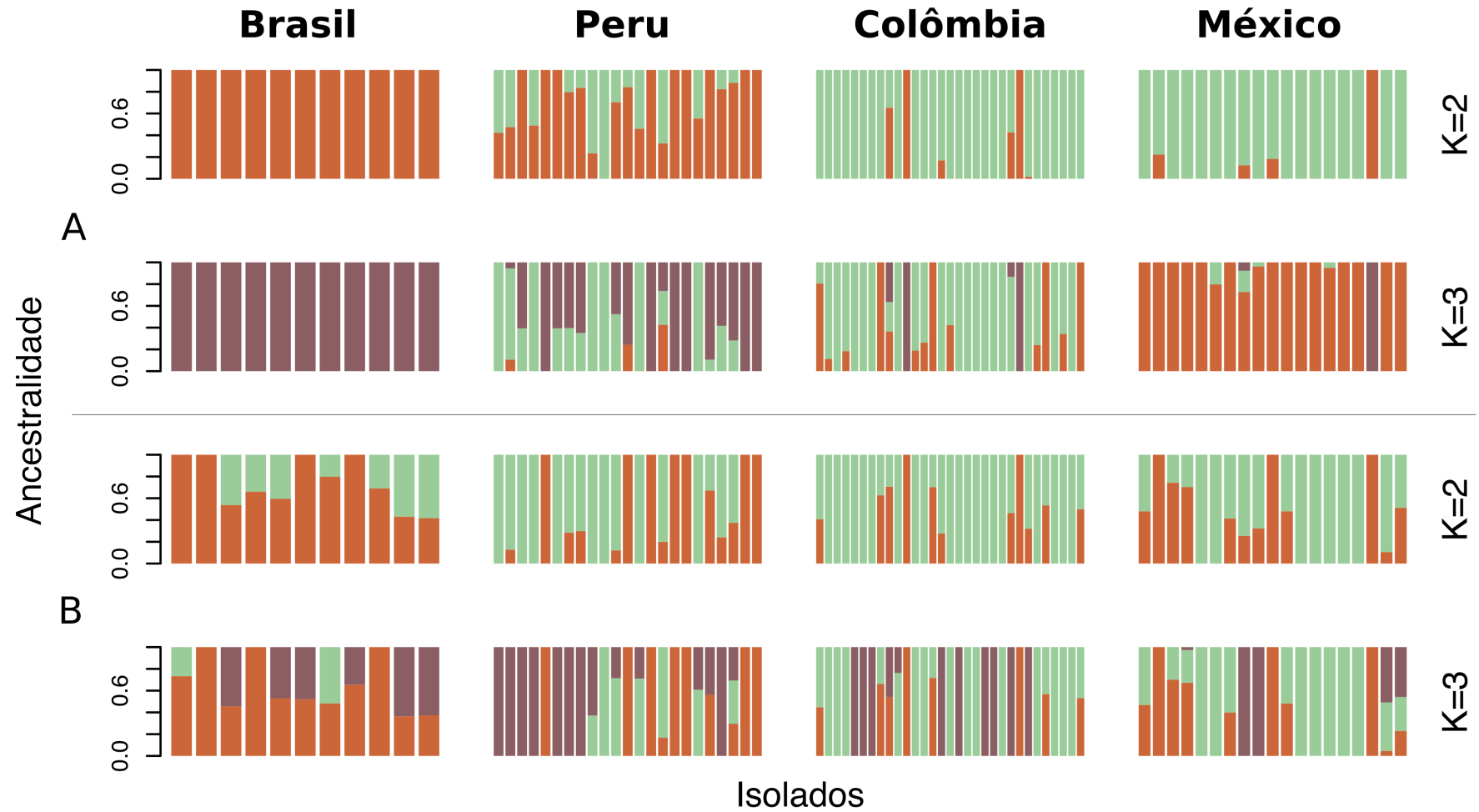


Figura S3 – Análises de ADMIXTURE com o banco de dados com K=2 e K=3 com 84 amostras de *P. vivax* nas Américas.



(A) 94122 SNPs (B) 12762 SNPs não ligados

Tabela S1 – Amostras americanas com seus números de acesso e filtro para as análises

Região	Amostra	Número de Acesso	Filtro
Brasil	17	-	Sim
	18	-	Sim
	19	-	Sim
	20	-	Sim
	32	-	Sim
	51	-	Sim
	52	-	Sim
	207	-	Sim
	ACR	-	Sim
	PV4	ERP000194	Sim
	Brazil01	SRR1568121	Não
	Brazil02	SRR1568116	Não
	Brazil03	SRR1568197	Não
	Brazil08	SRR1568176	Não
	Brazil12	SRR1568108	Não
	Brazil13	SRR1562973	Não
	Brazil15	SRR1568233	Não
	Brazil17	SRR1568173	Não
	Brazil22	SRR1562939	Não
	Brazil24	SRR1568111	Não
	Brazil25	SRR1568188	Não
	Brazil27	SRR1568215	Não
	Brazil29	SRR1568193	Não
	Brazil30	SRR1564653	Não
	Brazil31	SRR1568784	Não
	Brazil32	SRR1568192	Sim
	Brazil33	SRR1568212	Não
	Brazil34	SRR1562969	Não
	Brazil35	SRR1568222	Não
	Brazil87	SRR1562536	Não
	Belem	SRR575087	Sim
	Brasill	AFMK01000000	Sim
	Colômbia	Col015	SRR1564664
Col020		SRR1568235	Sim
Col026		SRR1568128	Sim
Col036		SRR1573226	Sim
Col038		SRR1562524	Sim
Col040		SRR1568221	Sim
Col280		SRR1568236	Sim
Col295		SRR1562967	Sim
Col437		SRR1568171	Sim
Col438-A		SRR1568155	Sim
Col438-B		SRR1568227	Sim
Col439		SRR1562975	Sim
Col440		SRR1568230	Sim
Col441-A		SRR1568159	Sim

Tabela S1- Continuação da página anterior

Região	Amostra	Número de Acesso	Filtro
	Col441-B	SRR1567977	Sim
	Col443-A	SRR1562971	Sim
	Col443-B	SRR1562555	Sim
	Col445	SRR1568207	Sim
	Col446	SRR1568118	Sim
	Col448	SRR1564670	Sim
	Col449	SRR1562870	Sim
	Col485	SRR1568213	Sim
	Col486	SRR1568160	Sim
	Col487	SRR1562518	Sim
	Col488	SRR1564665	Sim
	Col489	SRR1564660	Sim
	Col490	SRR1568169	Sim
	Col491	SRR1564650	Sim
	Col492	SRR1562965	Sim
	Col493	SRR1562818	Sim
	Col504	SRR1568112	Sim
Peru	Peru03	SRR1759047	Não
	Peru04	SRR1759122	Não
	Peru05	SRR1759139	Não
	Peru07	SRR1568211	Não
	Peru06	SRR1568172	Sim
	Peru08	SRR1562512	Não
	Peru09	SRR1562513	Sim
	Peru10	SRR1562521	Sim
	Peru257	SRR1568149	Sim
	Peru258	SRR1562871	Sim
	Peru259	SRR1562958	Sim
	Peru260	SRR1568196	Sim
	Peru262	SRR1568174	Sim
	Peru612	SRR1759307	Não
	Peru00622	SRR1568166	Sim
	Peru00692	SRR1568191	Não
	Peru697	SRR1759336	Não
	Peru00699	SRR1568216	Sim
	Peru852	SRR1568182	Sim
	Peru858	SRR1568195	Sim
	Peru869	SRR1562534	Sim
	Peru872	SRR1568122	Não
	Peru1008	SRR1568206	Sim
	Peru1021	SRR1562567	Sim
	Peru1022	SRR1562851	Sim
	Peru2025	SRR1568162	Sim
	Peru3133	SRR1564630	Sim
	Peru3136	SRR1568157	Sim
	Peru3232	SRR1568232	Sim
	Peru3280	SRR1562615	Sim

Tabela S1- Continuação da página anterior

Região	Amostra	Número de Acesso	Filtro
	Peru3323	SRR1568184	Não
	Peru4017	SRR1562538	Sim
	Peru4023	SRR1568117	Sim
	Peru99622	SRR1568178	Não
México	Mexico330-A	SRR1568231	Sim
	Mexico267-A	SRR1568225	Sim
	Mexico980-A	SRR1568223	Sim
	Mexico1086-A	SRR1568219	Sim
	Mexico55-03	SRR1568218	Sim
	Mexico938-A	SRR1568201	Sim
	Mexico32-E-03	SRR1568190	Sim
	Mexico203-04	SRR1568181	Sim
	Mexico118-A	SRR1568158	Sim
	Mexico165-A	SRR1568153	Sim
	Mexico161-04	SRR1568150	Sim
	Mexico760-A	SRR1568127	Sim
	Mexico566-A	SRR1568126	Sim
	Mexico63-08	SRR1568110	Sim
	Mexico21-A	SRR1568077	Sim
	Mexico10-E-05	SRR1562840	Sim
	Mexico533-07	SRR1562526	Sim
	Mexico31A-03	SRR1562839	Sim
	Mexico1067-A	SRR1562968	Sim

Símbolo '-' representa que essas amostras ainda não foram disponibilizadas no banco de dados;

Sim na coluna Filtro representa que a amostra foi mantida nas análises de diversidade e estrutura populacional

Tabela S2 – 50 maiores valores de D de Tajima encontrados nas quatro regiões da América (Brasil, Colômbia, México e Peru) divididos pelos cromossomos

Cromossomo	Região	Início	SNPs	Genes
1	Brasil	379000	5	-
	México	235000	6	PVX_087865(PH)
	México	324000	5	PVX_087970(heat shock protein 110 (HSP110c))
	Peru	45000	7	PVX_087700(PH)
	Peru	546000	2	-
	Peru	778000	4	-
2	Brasil	642000	2	PVX_081770(PH)
	Colômbia	297000	7	PVX_081390(PH)
	México	662000	8	PVX_081792(PH)
	Peru	680000	3	-
3	Brasil	607000	4	PVX_096310(PH)
	Brasil	656000	2	PVX_096265(40S ribosomal protein S5)
	Colômbia	23000	7	PVX_001080(PH)
	Colômbia	72000	3	PVX_001040(transcription factor with AP2 domain(s)(ApiAP2))
	Colômbia	128000	8	PVX_000995(6-cysteine protein (P41))
	Colômbia	622000	3	PVX_096295(PH)
	México	9000	8	PVX_001100(RAD protein (Pv-fam-e))
	México	11000	6	PVX_001095(PH)
	México	20000	7	PVX_001080(PH)
	México	26000	19	PVX_001080(PH)
	México	393000	5	PVX_000660(PH)
	México	615000	6	PVX_096300(PH)
	México	960000	9	-
	México	962000	5	PVX_000010(PH)
	Peru	114000	3	PVX_001010(PH)
	Peru	160000	2	PVX_000970(pre-mRNA-processing-splicing factor 8)

Tabela S2- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Colômbia/México	961000	41223	-
4	Brasil	157000	2	PVX_002645(PH)
	Brasil	518000	3	PVX_003900(6-cysteine)
	Brasil	815000	4	-
	Colômbia	533000	2	PVX_003870(PH)
	México	46000	7	PVX_002505(PH)
	México	583000	11	PVX_003820(SERA)
	México	591000	9	PVX_003810(SERA)
	México	757000	5	-
	Peru	753000	8	PVX_003590(serine/threonine-specific protein kinase)
	Colômbia/México/Peru	540000	4 /4/5	PVX_003865(PH)
5	Colômbia	471000	3	PVX_089325(2-oxoglutarate dehydrogenase E1 component, mitochondrial precursor)
	Colômbia	742000	4	-
	Colômbia	1036000	2	-
	México	111000	7	-
	México	1236000	6	-
	Peru	496000	2	PVX_089365(helicase)
	Peru	847000	2	PVX_089810(RAD protein (Pv-fam-e))
	Peru	952000	4	PVX_089935(PH)
6	Colômbia	155000	2	-
	Colômbia/México	907000	40971	PVX_110855(PH)
7	Brasil	36000	2	-
	Brasil	129000	2	PVX_098690(copper-transporting ATPase (CuTP))
	Brasil	348000	2	PVX_098910(exoribonuclease)
	Brasil	438000	6	PVX_099005(cysteine repeat modular protein 1, putative (CRMP1))
	Brasil	657000	2	-
	Brasil	969000	3	PVX_099690(thioredoxin-like protein 2 (TLP2))

Tabela S2- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Brasil	1406000	2	-
	Colômbia	1037000	2	PVX_099770(PH)
	Colômbia	1295000	7	PVX_087025(delta tubulin), PVX_087020(PH)
	Colômbia	1325000	2	PVX_086960(vacuolar protein sorting-associated protein 33 (VPS33))
	México	40000	3	PVX_098610(PH)
	México	1322000	6	PVX_086970(PH)
	México	1123000	4	PVX_099930(high molecular weight rhoptry protein 2 (RhopH2))
	Peru	131000	3	PVX_098690(copper-transporting ATPase (CuTP))
	Peru	628000	2	PVX_099255(methyltransferase), PVX_099257(RNA-binding protein musash (HoMu))
	Peru	928000	2	-
	Peru	970000	12	PVX_099690(thioredoxin-like protein 2 (TLP2))
	Peru	1286000	3	PVX_087040(gamma-tubulin complex component)
8	Brasil	538000	2	PVX_094810(PRE-binding protein(PREBP))
	Brasil	683000	2	-
	Brasil	808000	2	-
	Brasil	1068000	2	PVX_094450(PH), PVX_094455(PH)
	Brasil	1362000	2	PVX_119585(asparagine synthetase [glutamine-hydrolyzing])
	Brasil	1565000	3	PVX_119315(PH)
	Brasil	1625000	5	PVX_119250(serine/threonine protein kinase (SRPK1))
	Colômbia	214000	4	PVX_094400(60S ribosomal protein L13a)
	Colômbia	604000	10	PVX_094885(PH)
	Colômbia	618000	2	PVX_094895(phospholipase C-like)
	Colômbia	794000	5	PVX_095095(PH)
	Colômbia	1642000	9	-
	México	276000	11	PVX_094470(PH)
	México	965000	4	PVX_095290(PH)
	México	1119000	5	PVX_095452(PH)

Tabela S2- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	México	1643000	8	-
	Peru	123000	3	-
	Peru	214000	4	PVX_094400(60S ribosomal protein L13a)
	Peru	232000	2	PVX_094410(E3 ubiquitin-protein ligase)
	Peru	301000	2	PVX_094510(PH)
	Peru	1427000	3	PVX_119505(PH)
	Peru	1538000	5	PVX_119355(circumsporozoite (CS) protein (CSP))
9	Brasil	26000	6	-
	Brasil	1125000	3	-
	Brasil	1196000	2	PVX_092230(PH)
	Brasil	1705000	5	PVX_0192760(transcription factor with AP2 domain(s) (AP2-O))
	Colômbia	131000	2	PVX_090960(translocon component PTEX88)
	Colômbia	836000	3	PVX_091775(leucine-rich repeat protein (LRR11))
	Colômbia	973000	2	-
	Colômbia	1101000	2	-
	Colômbia	1213000	2	PVX_092260(PH)
	México	1524000	5	PVX_092570(transcription factor with AP2 domain(s)(ApiAP2))
	México	559000	5	PVX_091480(PH)
	México	223000	3	PVX_019065(transcription factor with AP2 domain(s)(ApiAP2))
	México	1907000	3	PVX_092985(serine/threonine protein kinase)
	México	310000	4	-
	Peru	187000	2	-
	Peru	791000	4	PVX_091755(calcium-dependent protein kinase 6 (CDPK6))
	Peru	968000	5	-
	Colômbia/Peru	46000	41127	-
	Colômbia/Peru	588000	40940	PVX_091530(PH)
10	Brasil	227000	3	PVX_079935(ADP-ribosylation factor GTPase-activating protein, putative (ARF-GAP))

Tabela S2- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Brasil	370000	3	PVX_080105(PH)
	Brasil	1063000	6	PVX_097895(TBC domain)
	Brasil	1217000	10	PVX_097725(MSP3)
	México	17000	4	PVX_079695(PH)
	México	33000	5	PVX_079710(PH)
	México	53000	6	PVX_079730(PH)
	México	1283000	8	PVX_097635(PH)
	México	1307000	10	PVX_097600(PH)
	Peru	352000	2	-
	Peru	363000	2	PVX_080100(multidrug resistance protein 1, putative (MDR1))
	Peru	1090000	2	PVX_097852(conserved Plasmodium protein, unknown function)
	Peru	1033000	7	PVX_097940(cell differentiation protein rcd1)
11	Brasil	434000	2	PVX_1149990(guanylyl cyclase)
	Brasil	1489000	2	-
	Brasil	1920000	2	PVX_113315(PH)
	Colômbia	31000	10	-
	Colômbia	906000	5	PVX_114510(PH)
	Colômbia	985000	2	PVX_114410(oxidoreductase, short-chain dehydrogenase family)
	Colômbia	1151000	2	PVX_114240(PH)
	Colômbia	1513000	4	-
	Colômbia	1750000	2	PVX_113505(coatomer alpha subunit)
	México	396000	4	PVX_115015(protein transport protein Sec24A (SEC24A))
	México	1543000	4	PVX_113731(PH)
	México	1765000	3	PVX_113495(polypyrimidine tract binding protein)
	México	2062000	15	-
	Peru	48000	2	-
	Peru	724000	3	PVX_114670(ER lumen protein retaining receptor (ERD2))
	Peru	1205000	2	PVX_114167(tRNA Leucine), PVX_114165(PH)

Tabela S2- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Peru	1486000	3	PVX_113815(PH)
	Peru	1887000	3	PVX_113355(centrosomal protein CEP76 (CEP76))
	Brasil/Colômbia	1261000	40940	PVX_114090(PH)
12	Brasil	751000	14	PVX_082710(PH)
	Brasil	1769000	2	PVX_117320(PH)
	Brasil	1831000	4	PVX_117365(PH)
	Brasil	1990000	2	PVX_117595(PH), PVX_117600(PH), PVX_117605(thioredoxin 1(TRX1))
	Brasil	2633000	2	PVX_118355(PH)
	Brasil	2908000	4	PVX_118625(trailer hitch homolog (CITH))
	Colômbia	136000	2	-
	Colômbia	577000	6	PVX_082935(nucleolar complex protein 2)
	Colômbia	2720000	2	PVX_118450(transporter)
	México	2967000	8	PVX_118675(PH)
	Peru	156000	2	PVX_083385(PH)
	Peru	279000	3	-
	Peru	1092000	2	PVX_116560(RNA-binding protein)
	Peru	1223000	3	-
	Peru	1513000	4	PVX_117060(PH)
	Peru	1686000	3	PVX_117225(ATP synthase (C/AC39) subunit)
	Peru	2255000	3	-
	Peru	2626000	4	PVX_118345(protein transport protein SEC7(SEC7))
	Colômbia/Peru	2698000	40940	PVX_118425(serine/threonine protein kinase)
	Colômbia/México	1223000	41001	-
	Colômbia/Peru	2300000	40940	PVX_117970(flavodoxin domain containing protein)
13	Brasil	43000	5	PVX_084160(dynein heavy chain)
	Brasil	330000	2	PVX_084460(PH)
	Brasil	1013000	2	PVX_085205(ABC transporter G family member 2 (ABCG2))
	Brasil	1638000	2	-

Tabela S2- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Colômbia	731000	4	PVX_084860(PH)
	Colômbia	1774000	3	PVX_086035(transcription factor with AP2 domain(s) (AP2-G2))
	Colômbia	1320000	2	PVX_085580(rap guanine nucleotide exchange factor (EPAC))
	Colômbia	1999000	12	PVX_086315(26S proteasome regulatory subunit RPN6)
	México	40000	8	PVX_084160(dynein heavy chain)
	México	1988000	5	PVX_086295(U2 snRNP-associated SURP motif-containing protein (SR140))
	México/Peru	1905000	41032	-
14	Brasil	265000	3	PVX_122077(NLI interacting factor-like phosphatase(NIF1))
	Brasil	388000	2	PVX_122222(PH)
	Brasil	669000	3	PVX_122530(telomerase reverse transcriptase (TERT))
	Brasil	1738000	4	-
	Brasil	2786000	4	-
	Brasil	2974000	4	PVX_101505
	Colômbia	1304000	2	-
	Colômbia	1767000	2	-
	Colômbia	1642000	2	PVX_123670(PH)
	Colômbia	2249000	6	PVX_100635(PH)
	México	1329000	4	PVX_12330(PH)
	México	1645000	4	PVX_123675(PH)
	México	2808000	4	PVX_101335(26S protease regulatory subunit 8 (RPT6))
	México	2928000	5	PVX_101485(rhoptry neck protein 3 (RON3))
	Peru	725000	6	PVX_122620(exportin-1)
	Peru	1371000	2	PVX_123340(eukaryotic translation initiation factor 3 subunit 10)
	Peru	1514000	3	-
	Brasil/Colômbia	1465000	40941	PVX_123480(serine-tRNA ligase)

Símbolo 'PH' indica Proteína Hipotética

Tabela S3 – 50 menores valores de D de Tajima encontrados nas quatro regiões da América (Brasil, Colômbia, México e Peru) divididos pelos cromossomos

Cromossomo	Região	Início	SNPs	Genes
1	Brasil	802000	12	PVX_093710(VIR)
	Colômbia	653000	10	PVX_093575(DNA repair protein RAD54)
	México	45000	6	PVX_087700(PH)
	México	163000	4	PVX_087785(PH)
	México	445000	5	-
	Peru	372000	10	PVX_088025 (PH)
	Peru	376000	8	PVX_088025(PH)
	Brasil/Colômbia	544000	7/9	-
2	Brasil	35000	6	-
	Brasil	78000	6	-
	Brasil	266000	6	PVX_081345(secreted ookinete protein, putative (PSOP24))
	Brasil	501000	5	PVX_081625(PH)
	Colômbia	88000	9	PVX_096985(VIR)
	Colômbia	267000	8	PVX_081350(PH)
	México	34000	4	PVX_096938(VIR)
	México	53000	11	-
	México	133000	4	-
	México	405000	4	-
	México	601000	4	PVX_081710(actin-like protein(ALP3))
	México	623000	4	PVX_081745(RNA-binding protein)
	Peru	264000	13	PVX_081345(secreted ookinete prote]na (PSOP24))
3	Brasil	3000	6	-
	Brasil	768000	7	PVX_096110(PH)
	Brasil	896000	6	PVX_096003(VIR)
	Colômbia	12000	9	PVX_001095(PH)
	Colômbia	27000	16	PVX_001080(PH)

Tabela S3- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Colômbia	909000	18	PVX_096000(VIR)
	Colômbia	908000	23	PVX_096000(VIR)
	Colômbia	936000	14	PVX_103155(VIR)
	Colômbia	1009000	14	-
	México	17000	9	PVX_001085(PH)
	México	337000	4	PVX_000745(erythrocyte vesicle protein 1(EVP1))
	México	342000	4	PVX_000737(tRNA Isoleucine)
	México	449000	4	PVX_000590(eukaryotic translation initiation factor 3 subunit 2)
	México	864000	4	-
	México	883000	4	-
	México	910000	7	PVX_096000(VIR)
	Peru	393000	6	PVX_000660(PH)
4	Brasil	68000	7	-
	Brasil	590000	7	-
	Brasil	681000	6	PVX_003695(PH)
	Brasil	872000	7	PVX_003475(PH)
	Colômbia	1000	11	PVX_106210(VIR)
	Colômbia	93000	13	PVX_002550(PH)
	Colômbia	629000	10	PVX_003770(MSP5)
	Colômbia	867000	29	-
	México	29000	12	-
	México	321000	4	PVX_002845(PH)
	México	575000	5	PVX_003830(SERA)
	México	603000	6	PVX_003800(SERA)
	México	875000	5	PVX_003475(PH)
	Peru	571000	10	PVX_003835(SERA)
	Peru	838000	6	PVX_003510(VIR)
5	Brasil	471000	5	PVX_003950(transport protein SEC61 gamma subunit)

Tabela S3- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Brasil	1284000	5	PVX_090270(tryptophan-rich antigen (Pv-fam-a))
	Brasil	1296000	11	-
	Brasil	1299000	5	PVX_090285(Pvstp1)
	Colômbia	1238000	11	PVX_090220(PH)
	Colômbia	1322000	25	-
	Colômbia	1341000	18	-
	México	20000	5	-
	México	1066000	5	PVX_090045(PH)
	México	1368000	14	PVX_090335(VIR)
	Peru	313000	10	PVX_089110(PH)
	Peru	375000	6	PVX_089200(histone acetyltransferase GCN5)
	Peru	437000	7	-
	Peru	1302000	10	-
	Peru	1303000	8	-
	Peru	1304000	7	-
	Peru	1318000	11	-
6	Brasil	18000	7	PVX_001635(VIR)
	Brasil	20000	5	PVX_001635(VIR)
	Colômbia	87000	10	PVX_001720(DNA polymerase delta catalytic subunit)
	México	351000	5	-
	México	1013000	5	PVX_004525(VIR)
	Peru	3000	7	-
	Peru	32000	15	-
	Peru	255000	9	PVX_001925(PH)
	Peru	896000	7	-
	Peru	532000	17	PVX_111365(U3 small nucleolar ribonucleoprotein protein MPP10)
7	Brasil	537000	5	-
	Brasil	1265000	6	PVX_087065(PH)

Tabela S3- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Brasil	1378000	8	-
	Brasil	1424000	6	-
	Colômbia	410000	9	PVX_098965(DNA repair protein REV1)
	Colômbia	505000	18	PV_099100(PH)
	Colômbia	623000	7	PVX_099250(ubiquitin fusion degradation protein 1 (UFD1))
	Colômbia	111000	7	PVX_098670(prefoldin subunit)
	Colômbia	1336000	10	PVX_086950(gametocyte development protein 1(GDV1))
	Colômbia	1397000	8	PVX_086903(PH)
	Colômbia	1413000	18	PVX_086895(VIR)
	Colômbia	1461000	16	PVX_086845(VIR)
	México	20000	5	PVX_098585(RBP1a)
	México	1475000	9	PVX_005050(VIR)
	México	1402000	6	PVX_086900(PH)
	México	1476000	6	PVX_005050(VIR)
	México	1490000	6	-
	Peru	1468000	6	-
	Peru	1471000	9	PVX_005045(VIR)
8	Brasil	366000	5	PVX_094595(histone deacetylase 2(HDA2))
	Brasil	452000	7	PVX_094690(PH)
	Brasil	625000	5	PVX_094900(PH)
	Brasil	881000	5	PVX_095195(ATP-dependent RNA helicase DDX6 (DOZI))
	Brasil	1233000	6	PVX_119735(transporter)
	Colômbia	199000	9	PVX_094385(PH)
	México	19000	5	PVX_094235(PH)
	México	345000	5	PVX_094580(transcription factor with AP2 domain(s) (ApiAP2))
	México	1434000	7	PVX_119500(PH)
	México	1460000	10	PVX_119450(PH), PVX_119445(FAD-dependent glycerol-3-phosphate dehydrogenase)

Tabela S3- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Peru	1651000	7	PVX_119225(PH)
	Peru	1654000	9	-
9	Brasil	6000	5	PVX_090835(PH)
	Brasil	61000	7	PVX_090878(PH), PVX_090880(phenylalanine-tRNA ligase beta subunit)
	Brasil	452000	5	PVX_091355(PH)
	Colômbia	4000	22	-
	Colômbia	691000	12	PVX_091660(PH)
	Colômbia	1773000	8	PVX_092865(protein kinase domain containing protein)
	México	58000	6	PVX_090876(CCR4-NOT transcription complex subunit 1)
	México	101000	5	-
	México	979000	5	PVX_091950(autophagy-related protein 23 (ATG23))
	Peru	11000	16	-
	Peru	32000	14	PVX_090850(FeS cluster assembly protein SufD (SufD))
	Peru	260000	9	-
	Peru	1563000	7	PVX_092605(PH)
	Peru	1788000	6	PVX_092875(dynamin-like protein)
10	Brasil	832000	5	PVX_080660(RNA pseudouridylate synthase)
	Brasil	854000	6	PVX_080660(RNA pseudouridylate synthase)
	Brasil	1033000	5	PVX_097940(cell differentiation protein rcd1)
	Brasil	1054000	7	PVX_097910(PH)
	Brasil	1220000	5	PVX_097720(MPS3)
	Brasil	1251000	10	PVX_097685(MSP3)
	Brasil	1275000	5	PVX_097655(50S ribosomal protein L28, apicoplast)
	Colômbia	938000	10	PVX_098050(phosphatidylinositol 4-kinase (PI4K))
	Colômbia	1074000	10	PVX_097885(PH)
	Colômbia	1212000	27	PVX_097730(PH)
	Colômbia	1227000	18	PVX_097710(MSP3)
	Colômbia	1235000	28	PVX_097700(MSP3)

Tabela S3- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Colômbia	1249000	12	PVX_097685(MSP3)
	Colômbia	1258000	22	PVX_097675(MSP3)
	Colômbia	1260000	9	PVX_097675(MSP3)
	Colômbia	1275000	9	PVX_097655(50S ribosomal protein L28, apicoplast)
	Colômbia	1399000	13	-
	Colômbia	1411000	16	PVX_097530(VIR)
	México	22000	8	PVX_079700(PH)
	México	834000	14	PVX_080660(RNA pseudouridylate synthase)
	México	1227000	8	PVX_097710(MSP3)
	Peru	227000	9	PVX_079935(ADP-ribosylation factor GTPase-activating protein (ARF-GAP))
	Peru	650000	9	PVX_080410(transcription factor with AP2 domain(s)(ApiAP2))
	Peru	991000	6	PVX_097995(ATP-dependent RNA helicase)
	Peru	1092000	7	PVX_097850(small ubiquitin-related modifier (SUMO))
	Peru	1236000	6	PVX_097700(MSP3)
	Peru	1256000	35	PVX_097680(MSP3)
	Peru	1251000	7	PVX_097685(MSP3)
	Peru	1378000	6	PVX_097555(VIR)
11	Brasil	396000	5	PVX_115015(protein transport protein Sec24A (SEC24A))
	Brasil	975000	6	-
	Brasil	1120000	7	-
	Brasil	1736000	6	PVX_1153520(MYND finger protein)
	Colômbia	10000	11	-
	Colômbia	174000	8	PVX_115290(signal recognition particle receptor)
	Colômbia	2017000	7	-
	México	11000	9	-
	México	1726000	9	PVX_113535(PH)
	México	1681000	5	PVX_113590(PH)
12	Brasil	1038000	6	PVX_116515(PH)

Tabela S3- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Brasil	2960000	5	-
	Colômbia	811000	9	-
	Colômbia	1830000	10	PVX_117365(PH)
	México	15000	5	-
	México	774000	10	-
	México	1086000	6	PVX_116555(PH)
	México	2817000	7	PVX_118540(copper transporter)
	Peru	974000	6	PVX_082375(PST-A protein)
	Peru	2642000	7	PVX_118360(TRAP-like protein (TREP))
	Peru	2727000	7	PVX_118455(clathrin coat assembly protein AP50)
	Peru	2927000	8	PVX_118648(PH)
	Brasil/Colômbia	2444000	6/9	-
13	Brasil	19000	6	PVX_084110(PH)
	Brasil	675000	7	PVX_084815(PH)
	Colômbia	745000	8	PVX_084870(chromatin assembly factor 1 P55 subunit)
	Peru	790000	12	PVX_084935(PH), PVX_084940(voltage-dependent anion-selective channel protein), PVX_084945(syntaxin)
	Peru	799000	9	PVX_084945(syntaxin)
	Peru	827000	12	PVX_084985(PH)
	Peru	897000	7	PVX_085045(ADP-dependent DNA helicase RecQ (WRN))
	Peru	1101000	7	PVX_085320(cyclophilin)
	Peru	2024000	9	PVX_086345(PH)
14	Brasil	16000	6	-
	Brasil	996000	7	PVX_122920(PH)
	Brasil	1565000	6	PVX_123600(PH)
	Colômbia	2064000	9	PVX_124105(PH)
	México	631000	7	PVX_122500(PH)
	México	1919000	5	PVX_123945(haloacid dehalogenase-like hydrolase (HAD2))

Tabela S3- Continuação da página anterior

Cromossomo	Região	Início	SNPs	Genes
	Peru	43000	14	PVX_121885(cytoadherence linked asexual protein (CLAG))
	Peru	2062000	8	PVX_124100(T-complex protein 1 subunit gamma (CCT3)), PVX_124105(PH)
	Peru	2396000	8	PVX_100810(PH)
	Peru	2803000	11	PVX_101345(PH)
	Colômbia/Peru	389000	15/12	PVX_122222(PH)

Símbolo 'PH' indica Proteína Hipotética