Faculdade de Medicina de São José do Rio Preto
Programa de Pós-graduação em Ciências da Saúde

Natália Guelfi Furlani

Pesquisa da Ancestralidade Genômica em uma População da Amazônia Ocidental Brasileira

São José do Rio Preto
2011

Natália Guelfi Furlani

Pesquisa da Ancestralidade Genômica em uma População da Amazônia Ocidental Brasileira

Dissertação apresentada à Faculdade de Medicina de São José do Rio Preto para obtenção do Título de Mestre no Curso de Pós-graduação em Ciências da Saúde, Eixo Temático: Medicina e Ciências Correlatas.

Orientador: Profa. Dra. Andréa Regina de Souza Baptista

Co-orientador: Profa. Dra. Flávia Cristina Lisoni

São José do Rio Preto
2011

Furlani, Natália Guelfi
Pesquisa da Ancestralidade Genômica em População da Amazônia Ocidental Brasileira
São José do Rio Preto, 2011
p.

Dissertação (Mestrado) – Faculdade de Medicina de São José do Rio Preto – FAMERP
Eixo Temático: Medicina e Ciências Correlatas

Orientadora: Profa. Dra. Andréa Regina de Souza Baptista
1. Marcadores moleculares; 2. Ancestralidade; 3. Susceptibilidade Genética; 4. Malária; 5. População Brasileira; 6. Amazônia.

NATÁLIA GUELFI FURLANI

Pesquisa da Ancestralidade Genômica em uma População da Amazônia Ocidental Brasileira

BANCA EXAMINADORA

DISSERTAÇÃO PARA OBTENÇÃO DO TÍTULO DE MESTRE

Presidente e Orientador: Profa. Dra. Andréa Regina de Souza Baptista
Co-orientador: Profa. Dra. Flávia Cristina Lisoni

1º Examinador: Profa. Dra. Lílian Madi-Ravazzi
2º Examinador: Profa. Dra. Érika Pavarino
1º Suplente: Profa. Dra. Cláudia Regina Bonini Rodrigues
2º Suplente: Profa. Dra. Eloiza Helena Tajara da Silva

São José do Rio Preto, 06/06/2011.

SUMÁRIO

Dedicatória i

Agradecimentos ii

Epígrafe v

Lista de Abreviaturas e Símbolos.. vi

Resumo.. vii

Abstract.. ix

1. Introdução 1

1.1. Objetivos 5

2. Artigos Científicos.. 6

2.1. Artigos Científico 1 8

2.2. Artigos Científico 2 23

4. Conclusões 33

5. Referências Bibliográficas.. 35

Aos meus pais Odair e Teresinha

Pelo amor incondicional, pela educação, renúncia e esforço desempenhado para a minha formação profissional. Essa conquista é de vocês!

Agradecimentos

A Deus, pela vida e por me oferecer mais essa oportunidade de ter outras perspectivas no campo profissional, e conhecer um mundo investigativo, cansativo, competitivo, mas extremamente contagiante e apaixonante que é o mundo da pesquisa e do conhecimento.

A minha orientadora Profa. Dra. Andréa Regina Baptista Rossit, pela confiança e aos ensinamentos acumulados no decorrer desse período.

A minha co-orientadora Profa. Dra. Flávia Cristina Rodrigues Lisoni, pesquisadora e orientadora exemplar, pela paciência, atenção, amizade,pela oportunidade e confiança depositada em mim desde o estágio básico.

Com carinho especial, a minha mãe Teresinha, que sempre esteve ao meu lado incentivando-me e apoiando-me, e sempre acreditando nos meus sonhos. Amo você!

Aos meus avós, pelo carinho, compreensão da minha ausência e orgulho que sentem de mim. Amo vocês!

Ao meu irmão Caio que mesmo de longe sempre se mostrou interessado e presente no meu trabalho, me dando força

Ao querido Danilo, que com seu amor e carinho soube me mostrar o significado da palavra companheirismo e sempre soube me ouvir, mesmo às vezes não entendendo muito bem aonde eu queria chegar.

Aos amigos Gustavo, Ana Luiza, Caíque, Jackeline e Evelise pela companhia, torcida e auxílio nas etapas do meu trabalho.

As minhas amigas queridas: Adriana, Camila, Tatienne, Marcela e Liane, pelo apoio e torcida, sempre me incentivando na busca de novos conhecimentos.

A todos os estágio e pós-graduandos do LMMBM (Laboratório de Marcadores Moleculares e Bioinformática Médica) e CIM (Centro de Investigação de Microrganismos) pelo apoio e colaboração de todos.

Ao Prof. Dr. Eduardo Tarazona-Santos pela colaboração e por ceder um espaço em seu laboratório, Laboratório de Diversidade Genética Humana (LDGH), na UFMG (Universidade Federal de Minas Gerais), para realização de uma parte dos experimentos, como também as suas alunas, e em especial a doutoranda Luciana Werneck, que me receberam com atenção, ajudando sempre que necessário.

As companheiras de trabalho, de quarto, Daphene e Daiane, que dividiram um poucos das suas vidas comigo durante a estadia de três meses em Belo Horizonte para execução desse projeto.

A Profa. Dra. Isabel Cabral, que me acolheu e me recebeu muito bem em sua casa durante minha estadia em Belém.

Ao Prof. Dr. Sidney Santos por me auxiliar nas analises e por ceder uma parte de seu laboratório, Laboratório de Genética Humana e Medica (LGHM), na UFPA (Universidade Federal do Pará), para realização de uma parte do meu trabalho.

Aos membros da banca examinadora, pela disponibilidade e atenção.

Ao Programa de Pós-Graduação em Ciências da Saúde da FAMERP pela constante dedicação na manutenção e fortalecimento do curso de pós-graduação da Instituição.

A toda minha família que sempre me apoio e me ajudou.

Ao órgão de incentivo a pesquisa, FAPESP, pela bolsa de mestrado, e pela oportunidade de conhecer novos laboratórios, ampliando meus conhecimentos.

Aos voluntários que cederam a matéria prima para que esse estudo acontecesse.

Agradeço a todos que de alguma forma me auxiliaram em mais um degrau em minha vida. Muito obrigado!

"- O senhor poderia me dizer qual caminho tomar para sair daqui? Interroga Alice.

- Isso depende muito de para onde você quer ir. Respondeu o Gato.

- Não me importo muito para onde… Retrucou Alice.

- Então não importa o caminho que você escolha. Disse o gato."

(Lewis Carrol – Alice no País das Maravilhas)

Lista de Abreviaturas e Símbolos

%	-	Porcentagem
MIAs	-	Marcadores Informativos de Ancestralidade
AIMs	-	Ancestry Informative Markers
mtDNA	-	DNA mitocondrial
DNA	-	Ácido desoxiribonucléico
SNPs	-	Polimorfismos de base única
FAMERP	-	Faculdade de Medicina de São José do Rio Preto
INDELs	-	Polimorfismos de inserção/deleção
STR	-	Repetições curtas em tandem
IPA	-	Índice Parasitário Anual
RFLP	-	Polimorfismos de comprimento de fragmentos de restrição
VNTRs	-	Número variável de repetições em série
δ	-	Delta

Resumo

Introdução: A população brasileira é formada majoritariamente por três populações parentais: Nativos Americanos, Europeus e Africanos. Níveis significativos de miscigenação tri-híbrida já foram detectados em todas as regiões e níveis sócio-econômicos do País. Métodos estatísticos recentes aliados à possibilidade de genotipagem de um grande número de marcadores permitem estimar a miscigenação em nível individual. Em estudos epidemiológicos com desenho de caso-controle, a heterogeneidade étnica entre estas categorias pode produzir resultados falso-positivos. Por este motivo, para este tipo de estudo, é importante estudar a miscigenação individual, assim como avaliar como a miscigenação neste nível influencia a estrutura genética dos diferentes subgrupos das populações brasileiras. O controle do efeito da miscigenação, para evitar as associações espúrias em estudos do tipo caso-controle pode ser feito estudando Marcadores Informativos de Ancestralidade (MIAs), os quais apresentam grandes diferenças de freqüência entre as populações parentais. Este conhecimento poderá contribuir para a futura definição dos níveis de estruturação populacional em grupos e subgrupos que constituirão alvo de estudos epidemiológicos, permitindo a otimização dos mesmos, a fim de evitar resultados falso-positivos. Dentre estes grupos, são de particular relevância estudos que abordem os fatores genéticos que modulam a suscetibilidade à malária em regiões onde a população se encontra exposta em nível endêmico como, por exemplo, a que reside no município de Porto Velho (RO) e região. **Objetivos:** a) descrever a ancestralidade genômica individual Ameríndia,

Européia e Africana em indivíduos sadios e portadores de malária por *Plasmodium falciparum* da cidade de Porto Velho, RO (Amazônia Ocidental), Brasil, e avaliar seu impacto no desenho de estudos epidemiológicos e b) determinar, a partir dos marcadores genotipados, a estrutura genética das populações de portadores de “malária falciparum” e indivíduos sadios da mesma região, em função dos níveis de miscigenação.

Palavras-chave: 1 Marcadores moleculares; 2. Ancestralidade; 3. Susceptibilidade Genética; 4. Malária; 5. População Brasileira; 6. Amazônia.

Abstract

Introduction: The Brazilian population is mainly composed of three parental populations: Native Americans, Europeans and Africans. Significant levels of tri-hybrid admixture have been detected in all regions and socioeconomic levels within the country. Recent statistical methods allied to the ability to genotype a large number of markers permit the estimate of the admixture at the individual level. In epidemiological studies with case-control design, ethnic heterogeneity between subgroups can produce false positive results. Therefore, for this type of study, it is important to evaluate individual admixture, as well as to measureits influence on the genetic structure of diverse subgroups within the Brazilian populations. Ancestry Informative Markers (AIMs), which frequencies show large differences between parental populations are suitable for tracking the effect of mixing, in order to avoid spurious associations in case control studies. This knowledge could help to further define the levels of population structure in groups and subgroups that constitute the subject of epidemiological studies, optimizing their designto avoid false positive results. Among these groups are particularly relevant studies addressing genetic factors that modulate susceptibility to malaria in regions where the population is exposed to endemic levels, for example, those residing in the municipality of Porto Velho (RO), Brazilian amazon region and surroundings. **Objectives:** a) to describe the Amerindian, European and African individual genomic ancestry in healthy individuals and patients with “falciparum malaria” in Porto Velho, RO (Western Amazonia) and assess its impact on the design of epidemiological studies and b) to determine, from the genotyped markers, the genetic structure of

populations of *falciparum* malaria patients and healthy individuals from the same region, depending on the admixture levels .

Keywords: 1. Molecular Markers, 2. Ancestry; 3. Genetic Susceptibility; 4. Malaria; 5. Brazilian Population; 6. Amazon.

1. INTRODUÇÃO

1. INTRODUÇÃO

A população brasileira é uma das mais heterogêneas do mundo, resultante de uma mistura entre nativos americanos, europeus (principalmente portugueses), e africanos. O processo de miscigenação aconteceu diversamente em diferentes regiões geográficas, com uma contribuição mais acentuada de Nativos Americanos no Norte, uma alta contribuição de Africanos no Nordeste, e relativamente baixas contribuições dos Americanos e Africanos no Sul. [(1)]

Do total de imigrantes que chegaram ao Brasil entre 1500 e 1978, estima-se que 58% foram europeus, 40% africanos e 2% asiáticos,[(2)] e as contribuições de regiões tais como Itália, Espanha, Alemanha, Síria e Líbano foram intensificadas e contínuas até o inicio do século XX. Contribuições de outras regiões do mundo também aconteceram, mas foi numericamente menor. [(3)]

A miscigenação no Brasil tem sido avaliada utilizando diferentes tipos de indicadores, mostrando um processo claramente tendencioso envolvendo predominantemente descendentes homens, de europeus, e mulheres, de ambos, nativos Americanos e Africanos. [(1)] Portanto o povoamento brasileiro foi conduzido a partir da diversidade genética de, principalmente, três grupos étnicos – ameríndios, europeus e africanos – em um processo de miscigenação que já se estende por meio milênio. Devido a sua grande extensão territorial, a distribuição destes grupos no território brasileiro não foi homogênea, a qual é refletida na composição da população atual. [(4)] Porém, houve também um processo de incorporação de várias outras culturas à

brasileira, e, consequentemente suas características genotípicas e fenotípicas também foram adicionadas à nossa população. [5]

Estimativas precisas das proporções ancestrais em populações miscigenadas têm sido utilizadas para responder a questões relacionadas com a evolução antropológica, e são fundamentais para a avaliação dos efeitos da subestrutura da população em estudos de associação. [1]

Combinar marcadores genéticos do cromossomo X é um interessante recurso para ambos os marcadores, autossômicos e uniparental. Isto ocorre, pois, a recombinação do cromossomo X apresenta diversas regiões, cada um proveniente de histórias diferentes, ao contrário do mtDNA e cromossomo Y. Haploidia no sexo masculino reduz o tamanho efetivo da população, o que aumenta o acumulo de suscetibilidade à deriva genética e uma população mais propícia a diferenciação. Portanto, marcadores ligados ao cromossomo X, podem ser de grande utilidade para abordar a história de populações humanas e as questões de antropologia evolutiva. [6]

Marcadores informativos de ancestralidade (MIAs), anteriormente denominados de alelos específicos da população, [7] são definido como marcadores genéticos que apresentam freqüência alélica diferente entre as populações parentais de menor ou igual a 30%. [8] Eles são considerados ferramentas eficientes para a obtenção precisa de estimativas de miscigenação e detecção da estrutura da população. [1,8,9]

Além dos SNPs, marcadores do tipo INDEL (inserção e deleção) são muito freqüentes no genoma humano e apresentam distintas vantagens em estudos populacionais e forenses, tais como a baixa taxa de mutação, menor

risco que os STRs (repetições curtas em tandem) de interpretações incorretas de alelos, amplicons menores (50 pares de base ou menos) que permitem trabalhar com DNA degradado. [10]

Em 2002 Weber et al. identificaram e caracterizaram 2000 polimorfismos bialélicos de inserção-deleção (INDELS) no genoma humano. A grande adaptabilidade dos INDELS para a amplificação de DNA degradado combinada com a facilidade da tipagem e com este altíssimo poder de descriminação tornando os MULTINDELS uma poderosa plataforma para a determinação de identidade genética pelo DNA. [11]

Variações nas seqüências de DNA podem afetar a forma na qual os seres humanos desenvolvem doenças e respondem aos agentes patogénicos, produtos químicos, medicamentos, vacinas e outros agentes, sendo assim, sua maior importância na pesquisa biomédica é para na comparação de regiões do genoma entre gerações. [12]

Estima-se que mais de 40% da população mundial está exposta ao risco de adquirir malária. No ano de 2009, o Brasil registrou 306.000 casos de malária, sendo a espécie *Plasmodium vivax* de maior incidência (83%)[13] Embora esta seja a espécie mais prevalente, o *P. falciparum* é responsável pela maior gravidade da doença. A Região da Amazônia Legal concentra 99,7% dos casos de malária, essa região é composta pelos estados do Acre, Amazonas, Amapá, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins, e totaliza 807 municípios, e tendo sido identificados cerca de 90 municípios como sendo de alto risco para a malária, ou seja, com um Índice Parasitário Anual (IPA) igual ou maior a 50 casos por 1.000 habitantes. [14]

A transmissão nessa área está relacionada a fatores biológicos (presença de alta densidade de mosquitos vetores, agente etiológico e população suscetível); geográficos (altos índices de pluviosidade, amplitude da malha hídrica e a cobertura vegetal); ecológicos (desmatamentos, construção de hidroelétricas, estradas e de sistemas de irrigação, açudes); e sociais (presença de numerosos grupos populacionais, morando em habitações com ausência completa ou parcial de paredes laterais e trabalhando próximo ou dentro das matas). [(14)] Estes dados retratam bem a região Norte do país, que por sua vez, trata-se de uma área endêmica.

1.1- OBJETIVOS

Considerando o acima exposto e tendo como sujeitos indivíduos sadios e portadores de malária por *Plasmodium falciparum* que residem na cidade de Porto Velho, RO (Amazônia Ocidental), Brasil, o presente estudo tem por objetivos:

1) Descrever a ancestralidade genômica individual (Ameríndia, Européia e Africana).

2) Determinar a estrutura genética das populações de portadores de malária falciparum” e indivíduos sadios da mesma região, em função dos níveis de miscigenação.

3) Avaliar o possível impacto da diversidade étnica individual/populacional sobre o desenho de estudos epidemiológicos de caso-controle.

2. ARTIGOS CIENTÍFICOS

2. ARTIGOS CIENTÍFICOS

Os resultados referentes aos objetivos dessa dissertação estão apresentados na forma de artigo.

Artigo 1

Tílulo: Estratégias moleculares para investigação da composição étnica da população brasileira e seu impacto na saúde
Autores: Natália Guelfi Furlani, Flávia Lisoni, Ricardo Luiz Dantas Machado e Andréa Regina Baptista Rossit
Periódico: Arquivos em Ciências da Saúde. (A ser submetido).

Artigo 2

Título: "Population genetics of *GYPB* and association study between *GYPB**S/s* polymorphism and susceptibility to *P.falciparum* infection in the Brazilian Amazon".
Autores: Eduardo Tarazona-Santos; Lilian Castilho; Daphne Renata Tavares Amaral; Daiane Cobianchi Costa; **Natália Guelfi Furlani**; Luciana Werneck Zuccherato; Moara Machado; Marion E Reid; Mariano Gustavo Zalis; Andréa Regina Baptista Rossit; Sidney EB Santos; Ricardo Luiz Dantas Machado; Sara Lustigman.

Periódico: PLoS One. 2011 Jan 24;6(1):e16123.

Estratégias moleculares para investigação da composição étnica da população brasileira e seu impacto na saúde.

Natália Guelfi Furlani[1], Flávia Lisoni[2], Ricardo Luiz Dantas Machado[1] e Andréa Regina Baptista Rossit[3]

[1] Departamento de Doenças Dermatológicas, Infecciosas e Parasitárias, Faculdade de Medicina de São José do Rio Preto, São José do Rio Preto, SP.
[2] Departamento de Biologia e Zootecnia, Univerdade Estadual Paulista, Ilha Solteira, SP.
[3]Departamento de Microbiologia e Parasitologia, Instituto Biomédico, Universidade Federal Fluminense, Niterói, RJ.

Resumo

A população brasileira é formada majoritariamente por três populações parentais: Nativos Americanos, Europeus e Africanos. Níveis significativos de miscigenação tri-híbrida já foram detectados em todas as regiões e níveis sócio-econômicos do país. Métodos estatísticos recentes e a possibilidade de genotipar um grande número de marcadores permitem estimar a miscigenação em nível individual. Em estudos epidemiológicos com desenho de caso controle, a heterogeneidade gênica entre estas categorias pode produzir resultados falso-positivos. Por este motivo, para tais estudos, é importante avaliar a miscigenação individual e sua influência sobre a estrutura genética de subgrupos da população brasileira. O controle do efeito da miscigenação, para evitar associações espúrias em estudos de caso controle pode ser feito estudando Marcadores Informativos de Ancestralidade (MIAs) e inserções/deleções (INDELS), que são polimorfismos apresentam grandes diferenças de frequência entre as populações parentais. Este conhecimento poderá contribuir para a futura definição dos níveis de estruturação populacional em grupos e subgrupos que constituirão alvo de estudos epidemiológicos. Nesse contexto, são de particular relevância os estudos que abordem os fatores genéticos que modulam a susceptibilidade às doenças mais prevalentes no cenário Nacional, incluindo as infecto-parasitárias, que acometem as populações expostas em nível endêmico.

ABSTRACT

The Brazilian population is composed mainly of three parental populations: Native Americans, Europeans and Africans. Significant levels of tri-hybrid admixture have been detected in all regions and socioeconomic levelsin this country. Recent statistical methods and the ability to genotype a large number of markers allow the estimate of the admixture at the individual level. In epidemiological studies with case-control design, the genetic heterogeneity between these categories can produce false positives. Therefore, for this type of study, it is important to determine individual admixture, as well as evaluating its influence on the genetic structure of Brazilian populations. Ancestry Informative Markers (AIMs) and insertions/deletions (INDELs) are polymorphisms which differ significantly in frequency between the parental population and, therefore, are suitable for tracking the effect of admixture to avoid spurious associations in case control studies, This knowledge could help to further define the levels of population structure in groups and subgroups that constitute the subject of epidemiological studies in this scenario, particularly relevant studies will be those addressing genetic factors that modulate susceptibility to the nationwide most prevalent diseases , including infectious and parasitical ones, affecting endemically exposed populations.

INTRODUÇÃO

A população brasileira é uma das mais variáveis de todo o mundo, formada por Nativos Americanos, Europeus (principalmente portugueses), e Africanos - em um processo de miscigenação que já se estende por meio milênio, com contribuição mais acentuada de Nativos Americanos no Norte, alta contribuição de Africanos no Nordeste, e relativamente baixas contribuições dos Americanos e Africanos no Sul (Choudhry et al., 2006).

O processo de miscigenação que deu origem a atual população brasileira ocorreu de forma muito diversificada, nas diferentes regiões geográficas do país. Salzano e Bortilini, em 2002, mostraram que as principais contribuições de populações ancestrais do povo brasileiro ocorreram por meio de complexas interações étnicas, sendo que a proporção de participação de cada grupo parental difere dependendo da região brasileira analisada.

Os índios brasileiros, os primeiros habitantes, são descendentes do primeiro grupo de caçadores asiáticos que chegaram às Américas, provavelmente através do estreito de Bering (Figura 01), durante o período glacial, há 25.000 anos ou mais, até chegarem ao Alasca. De lá, se espalharam pelas Américas desde o extremo norte até o extremo sul (Cunha, 1992). Esta população que originalmente saiu da Ásia, sofreu uma redução drástica no seu tamanho populacional diminuindo a variabilidade genética existente nestes primeiros povos que migraram para o continente americano (Wallace *et al.*, 1985; Schurr *et al.*, 1990; Horai *et al.*, 1993; Santos *et al* 1999). Porém, culturalmente falando, estes povos estavam inseridos numa diversidade de nações com línguas e costumes distintos.

Estima-se que a chegada dos ameríndios ao território conhecido hoje como Brasil tenha ocorrido por volta de 12.000 anos atrás. Não é possível uma determinação precisa do tamanho da população que habitava este território antes da

chegada dos portugueses, devido a precariedade de dados históricos nesse período (Salzano e Callegari-Jaques, 1988). No entanto, muitos afirmam que entre um e 10 milhões de indivíduos habitavam a região, sobretudo pertencentes aos grupos Tupi e Jê ou Tapuia (Fundação Nacional do Índio - FUNAI, 2006).

Figura 01 – Mapa esquemático do Estreito de Bering (1), mostrando rota de entrada dos povos asiáticos no continente americano, segundo Cunha (1992).

A partir da colonização muitas sociedades indígenas que viviam no território dominado foram sendo exterminadas seja pela ação de armas, em decorrência do contágio por doenças trazidas dos países distantes ou, ainda, pela aplicação de políticas visando a "assimilação" dos índios à nova sociedade implantada, com forte influência européia. O fato é que restam, hoje, apenas cerca de 350 mil índios (aproximadamente 0,2% da população do país) que habitam, principalmente, a região amazônica (FUNAI, 2006).

Os portugueses chegaram ao litoral brasileiro oficialmente em 22 de abril de 1500 e já trouxeram séculos de interação genética e cultural com outros povos de origens geográfica distintas.Como os portugueses já utilizavam mão de obra escrava

africana, supõe-se que os primeiros escravos africanos brasileiros tenham vindo junto com a esquadra de Cabral (Curtin, 1969). Os primeiros navios negreiros chegaram ao Brasil em meados do século XVI trazendo os escravos para trabalhar na agricultura e na mineração. Registros históricos sugerem que entre 1551 e 1850 cerca de 3,5 milhões de africanos foram introduzidos compulsoriamente no Brasil (Salzano e Freire-Maia, 1967; Curtin, 1969; Ribeiro, 1995). Mais tarde, o Brasil também recebeu outros imigrantes europeus além dos portugueses. A estimativa é que entre 1800 e a metade do século XX, em torno de 4 milhões de indivíduos entraram no pais, principalmente portugueses, italianos, espanhóis e alemãs (Santos et al., 2009).

A distribuição de grupos étnicos baseada no autorelato da cor da pele é heterogênea entre as regiões geopolíticas brasileiras. A classificação do auto-relato como "Pardos" o grupo que incluía aqueles que declararam não possuir nenhuma miscigenação ou origem multietnica, representando mais de 42% do total da população brasileira, variando entre 15% na região Sudeste a 71,5% na região Nordeste (Lins et al., 2010). Ademais, estudo baseados em marcadores genéticos mostram que a autodeclaração ou mesmo a classificação por critérios de observação visual da etnia não representa a real participação de cada grupamento étnico na composição individual do sujeito pesquisado (Pena, 2005).

Marcadores Genéticos

A miscigenação no Brasil tem sido avaliada utilizando diferentes tipos de indicadores, mostrando um processo claramente tendencioso envolvendo predominantemente descendentes europeus, homens e, para as mulheres tanto a partir de nativos Americanos como Africanos (Choudhry et al., 2006). Devido a sua grande extensão territorial, a distribuição destes grupos no território brasileiro não foi homogênea, a qual é refletida na composição da população atual (IBGE, 2000). A obtenção de estimativas precisas das proporções de contribuição de subpopulações

formadoras (parentais) em populações miscigenadas, feitas a partir do estudo com marcadores genéticos, tem sido usadas para responder questões relacionadas à evolução e antropologia (Bedoya *et al.*, 2006; Laan *et al.*, 2005; Pereira e Pena, 2006) e são fundamentais como controle dos efeitos de uma possível subestruturação populacional em estudos de associação (Parra *et al.*, 1998; Pfaff *et al.*, 2001; Reiner *et al.*, 2005; Bonilla *et al.*, 2004; Choudhry *et al.*, 2006). Além de auxiliar na reconstrução das rotas e origens dos povoamentos (Salzano e Callegari-Jacques, 1988; Saitou, 1995).

Dados de Carvalho-Silva et al.(2001) demonstram que a grande maioria dos cromossomos Y em brasileiros brancos do sexo masculino, independentemente da sua origem regional, é de origem Européia, com uma freqüência muito baixa de cromossomos de populações do sub-Saara Africano e uma completa ausência de contribuições Ameríndias. Esses resultados configuraram uma imagem forte de acasalamento direcional no Brasil envolvendo Europeus e Africanos do sexo masculino e Ameríndios do sexo feminino. Isto está em consonância com a conhecida história da população do Brasil desde 1500.

Os estudos dos polimorfismos de DNA (regiões do genoma nas quais existem variações não associadas a doenças) permitem a construção de um perfil genético absolutamente indivíduo-específico (Pena, 2005). Para ser considerado um polimorfismo, o alelo raro de um determinado loco deve estar presente em mais de 1% dos indivíduos da população. Assim, dado o grande o número e tipo de variações disponíveis para avaliação, fica possível a identificação de uma pessoa com base no seu padrão único de polimorfismos (Brown, 2001). Entre os diferentes polimorfismos genéticos utilizados, estão os “RFLP” (*Restriction fragment lenght polymorphism* – polimorfismos de comprimento de fragmentos de restrição), os Minissatélites ou VNTRs (*Variable number of tandem repeats* - número variável de repetições em série), os Microssatélites ou STRs (*Short tandem repeats* - repetições curtas em série) e,

mais recentemente, os abundantes polimorfismos de base única ou SNPs (Hinds *et al.*, 2005) e os polimorfismos de inserção/ deleção (INDELs; Weber *et al.*, 2002). Esses últimos polimorfismos apresentam como uma grande vantagem a possibilidade de pesquisa de produtos de amplificação muito curtos (50 pares de bases ou menores) e, assim, apresentam distintas vantagens sobre os microssatélites no estudo de DNA extremamente degradado (Pena, 2005). No entanto, o melhor conhecimento da história evolutiva e da diversidade genética de populações humanas surge dos diferentes resultados obtidos pela análise de marcadores baseados em estratégias distintas e de natureza diversa (protéicos, uniparentais e moleculares) (Lum *et al.,* 1998).

Marcadores Informativos de Ancestralidade e INDELS

Um número crescente de publicações tem relatado o uso de marcadores informativo de ancestralidade (MIAs), marcadores cuja freqüência alélica varia significativamente entre populações de origem geográfica distinta a fim de estimar a mistura individual e identificar a composição étnica populacional (Freitas et al,2010). Os MIA se caracterizam, por apresentarem uma grande diferença de frequências alélicas entre as populações formadoras, que pode ser estimada através de uma medida simples denominada delta (δ), podendo chegar a ausência ou presença exclusiva de um alelo qualquer em uma determinada população que se deseja estudar. Qualquer locus polimórfico é considerado um MIA com um δ acima de 30%. Porém quanto maior o valor de delta mais informativo é o marcador o que, por sua vez, diminui o número de marcadores necessários para medir com precisão a proporção de mistura de cada indivíduo (Parra *et al.*, 2003).

Na maioria dos estudos, estes MIAs consistem de polimorfismos de nucleotídeo simples (SNPs) (Benn-Torres et al. 2008; Choudhry et al., 2006; Kosoy e et al., 2009; Parra et al., 1998, 2001; Shriver et al., 2003), todavia a inserção-deleção

de polimorfismos (INDELs) de pequenos fragmentos de DNA (Bedoya et al., 2006) e pequenas repetições em Tandem (Pimenta et al. 2006) também são utilizados.

A maior parte dos estudos de miscigenação em populações brasileiras tem estimado a miscigenação ao nível populacional. Métodos de análise recentes (Pritchard et al. 2000, revisão de Tarazona-Santos et al. 2007) e a possibilidade de genotipar um grande número de marcadores permitem hoje estimar componentes de miscigenação no nível individual. Em particular, marcadores com grandes diferenças de freqüência entre as populações parentais são especialmente informativos para estimar a miscigenação em níveis populacional e individual. O uso de marcadores altamente informativos é desejável para a redução custo e tempo de genotipagem, uma vez que permite a redução do número de marcadores necessários para estimar a miscigenação. Uma possível aplicação destas estimativas, por exemplo, é detectar estratificação populacional devido à miscigenação entre casos e controles em estudos epidemiológicos com este tipo de desenho experimental (Rosenberg et al., 2003, Pffaf et al., 2004). Em geral, a estratificação populacional existe quando uma população é constituída pela mistura de diferentes grupos étnicos, denominados de subpopulações, e quando a proporção de mistura (proporção do genoma que teve origem em cada subpopulação) varia entre indivíduos. Como no conceito de "associação decorrente de subestruturação" (Pritchard *et al.*, 2000; Pritchard e Donnelly, 2001) que diz que em uma população na qual proporções de mistura variam entre os indivíduos que a compõem; e na qual as frequências alélicas de um determinado marcador variam entre as diferentes subpopulações ancestrais, poderá haver associações entre alelos de loci não ligados.

Recentemente, os INDELs têm sido o foco de múltiplas investigações (Bastos-Rodrigues et al., 2006, Mills et al., 2006; Ribeiro Rodrigues et al. 2009, Weber et al., 2002). Os polimorfismos do tipo INDEL se caracterizam pela inserção ou deleção de um ou alguns nucleotídeos, portanto são polimorfismos bialélicos. Apesar

de parecerem abundantes no genoma humano e de estarem presentes em regiões promotoras, introns e éxons, esse tipo de polimorfismo é relativamente menos estudado quando comparado aos outros tipos já citados (Mills *et al.*, 2006). Este tipo de polimorfismo apresenta características interessantes como marcadores genéticos: (1) INDELs estão espalhados por todo o genoma humano (2); INDELs derivam de um único evento (que não apresentam homoplasia), (3) uma vez que as freqüências alélicas dos INDELs são significativamente diferentes em populações separadas, elas podem ser usadas assim como os AIMs, (4) pequenos INDELs podem ser analisados usando *amplicons* curtos, o que melhora a amplificação de DNA degradado e facilita o *multiplex* e (5) INDELs pode ser facilmente genotipados com uma simples eletroforese (Santos et al, 2009).

Todos os INDEL identificados até o presente podem ser acessados através dos sites: Marshfield Clinic (http://research.marshfieldclinic.org/genetics/INDEL/default.asp) e NCBI (http://www.ncbi.nlm.nih.gov/SNP/). Através destes sites é possível identificar cada um dos polimorfismos, sua localização específica dentro do cromossomo em questão e, em muitos casos, a distribuição alélica em diferentes populações: europeus, africanos e nativos ameríndios (Weber *et al.*, 2002).

Combinar marcadores genéticos do cromossomo X é um interessante recurso para ambos os marcadores, autossômicos e uniparental. Isto ocorre, pois, a recombinação do cromossomo X apresenta diversas regiões, cada um proveniente de histórias diferentes, ao contrário do mtDNA e cromossomo Y. A haploidia no sexo masculino reduz o tamanho efetivo da população, o que aumenta o acumulo de suscetibilidade à deriva genética e uma população mais propicia a diferenciação (Schaffner, 2004). Portanto, marcadores ligados ao cromossomo X, podem ser de grande utilidade para abordar a história de populações humanas e as questões de antropologia evolutiva.

Aplicações dos marcadores genéticos em saúde humana

O uso de marcadores genéticos informativos de ancestralidade para ajudar a inferir os processos demográficos históricos e antropológicos está no centro das atenções em estudos de miscigenação da população humana atual (Benn-Torres et al., 2008; Martinez-Marignacet al., 2007; Seldin et al., 2007), aplicações na genômica médica para mapear os genes que influenciam a susceptibilidade à doenças complexas e fatores subjacentes de etnia (Bonilla et al., 2004b; Shaffer et al., 2007), bem como a sua utilidade para controlar os efeitos de confusão das associações genéticas em populações estratificadas (Gentil et al. 2007, 2009; Matsuzaki et al., 2004; Moreno Lima et al., 2007; Parra et al., 2004; Suarez-Kurtz et al., 2007).

Se uma doença apresenta uma maior incidência em um dos três grupos étnicos principais (europeus, ameríndios ou africanos) que contribuíram na formação da população brasileira, este grupo étnico pode estar sobre-representado na amostra de casos em estudos casos-controle. Nesta situação, qualquer variante gênica associada estatisticamente com o grupo étnico de maior incidência, vai apresentar associação com a doença, independentemente do seu envolvimento na patogênese, podendo levar a resultados falsos positivos **(Figura 2)**. Para evitar associações espúrias, em estudos casos-controle em populações miscigenadas é necessário controlar o efeito da ancestralidade genômica, o que pode ser feito estudando MIAs (Tarazona-Santos et al. 2007).

O risco de obter falsos positivos em estudos de associação genótipo-fenótipo depende do nível de estruturação ou estratificação da população. Numa população na qual todos os indivíduos têm o mesmo nível de miscigenação, o problema não existe. Contrariamente, o problema se incrementa se o nível de miscigenação é variável entre indivíduos, de modo que casos e controles podem ter níveis de miscigenação diferentes. Por exemplo, no Rio Grande do Sul, Zembrzuski et al. (2006) evidenciaram uma ancestralidade genômica homogênea em um estudo epidemiológico de caso-

controle de doenças cardiovasculares. Porém, algumas regiões do Brasil como, por exemplo, Rondônia, podem apresentar uma maior variância na miscigenação inter-individual.

Considerações Finais

Por estas razões, para planejar futuros estudos de genética epidemiológica e de susceptibilidade genética às doenças humanas, é importante identificar a variação dos níveis de miscigenação individual. Este conhecimento poderá contribuir para a futura definição dos níveis de estruturação populacional em grupos e sub-grupos que constituirão alvo de estudos epidemiológicos. Assim, permitirão o desenho otimizado dos mesmos, a fim de evitar resultados falsos positivos. Dentre estes grupos, são de particular relevância estudos que abordem os fatores genéticos que modulam a susceptibilidade às doenças mais prevalentes no cenário Nacional ou mesmo aquelas que ocorrem em regiões onde a população se encontra exposta em nível endêmico como, por exemplo, a malária na Amazônia brasileira (Cavasini et al., 2006; Tarazona-Santos et al., 2011).

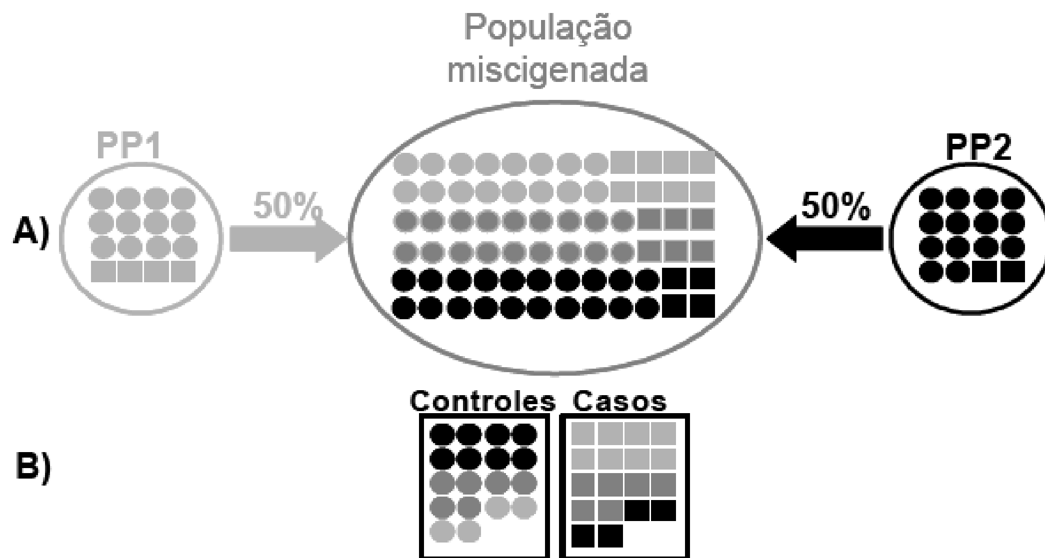

Figura 2. Representação esquemática da miscigenação e do efeito da estratificação de populações em estudos genético-epidemiológicos. **A)** Cada uma das populações parentais PP1 (cinza claro) e PP2 (preto) contribuíram com 50% do pool gênico da população miscigenada (cinza escuro). Indivíduos da população miscigenada possuem diferentes níveis de miscigenação. Na figura, simplificamos esta situação, apresentando apenas três níveis de miscigenação individual: indivíduos cujos genomas derivam completamente de PP1 (cinza claro), de PP2 (preto) e indivíduos miscigenados (cinza escuro). Nesta figura, indivíduos normais e doentes são representados por círculos e quadrados, respectivamente. Note que a incidência da doença é maior em PP1 do que em PP2. Na população miscigenada, a incidência da doença é maior entre os indivíduos cinza claros, intermediária entre os cinza escuros e menor entre os pretos. **B)** Um estudo caso-controle: casos são representados por quadrados e controles por círculos. Como a incidência da doença é maior entre indivíduos com maior ancestralidade cinza claro (PP1), estes estarão super-representados entre os casos. Como conseqüência, qualquer alelo que tenha freqüência mais alta em PP1 do que em PP2 pode apresentar associação com a doença, independentemente se ele é a o alelo causal.

REFERÊNCIAS BIBLIOGRÁFICAS:

Bastos-Rodrigues L, Pimenta JR e Pena SDJ. The genetic structure of human populations studied through short insertion-deletion polymorphisms. Am J Hum Genet 70: 1-8, 2006.

Bedoya G, Montoya P, Garcya J, Soto I, Bourgeois S, Carvajal L, Labuda D, Alvarez V, Ospina J, Hedrick PW, Ruiz-Linares A. Admixture dynamics in Hispanics: A shift in the nuclear genetic ancestry of a South American population isolate. PNAS 103:7234-7239, 2006.

Benn-Torres J, Bonilla C, Robbins CM, Waterman L, Moses TY, Hernandez W, Santos ER, Bennett F, Aiken W, Tullock T, Coard K, Hennis A, Wu S, Nemesure B, Leske MC, Freeman V, Carpten J, Kittles RA. Admixture and population stratification in African Caribbean populations. Ann Hum Genet 72:90–98, 2008.

Bonilla C, Parra EJ, Pfaff CL, Marshall DSJA, Hamman RF, Ferrell RE, Hoggart CL, McKeigue PM, Shriver MD. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. Am Hum Genet 68:139–153, 2004.

Brown TA. Clonagem Gênica e Análise de DNA: Uma introdução. 4.ed. Porto Alegre: Artmed. 376pp. 2001.

Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ. The phylogeography of Brazilian Y-Chromosome lineages. Am J Hum. Gen. 68:281-286, 2001.

Cavasini CE, De Mattos LC, Alves RT, et al. Frequencies of ABO, MNSs, and Duffy phenotypes among blood donors and malaria patients from four Brazilian Amazon areas. Hum Biol. 78(2):215-9, 2006.

Choudhry S, Coyle NE, Tang H, Salari HTK, Lind D, Clark SL, Tsai SLCH, Naqvi M, Ung APN, Matallana H, Avila PC, Casal J, Torres A, Nazario S, Castro R, Perez-Stable EJ, Kwok P-Y, Sheppard D, Shriver MD, Rodriguez-Cintron W, Risch N, Burchard EG. Population stratification confounds genetic association studies among Latinos. Hum Genet. 118:652–664, 2006.

Cunha MC, Salzano FM. História dos Índios no Brasil, Fundação de Amparo a Pesquisa do estado de São Paulo.1992.

Curtin PD. The Atlantic Slave Trade: a Census. 1ª edição. The University of Wisconsin, Milwaukee, 457 pp, 1969.

Freitas NS, Resque RL, Ribeiro-Rodrigues EM, Guerreiro JF, Santos NP, Ribeiro-dos-Santos A, Santos S. X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel. Int J Legal Med. 124(6):589-93, 2010.

Fundação Nacional do Índio – FUNAI, disponível em, www.funai.gov.br. Acessado em 21/Jun. 2010.

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. Whole-genome patterns of common DNA variation in three human populations. Science, n. 307, p. 1072-1079. 2005.

Gentil P, Lima RM, Lins TC, Abreu BS, Pereira RW, Oliveira RJ. Physical activity, Cdx-2 genotype, and BMD. Int J Sports Med 28:1065–1069. 2007.

Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S e Tajima K. Peopling of the Americas, founded by four major lineages of mithocondrial DNA. Mol Biol Evol. 10:23-47, 1993.

IBGE – HTTP: //www.ibge.gov.br/home/estatística/população. Censo 1991, 2000 e Brasil 500 anos de povoamento. Acesso em 14/dez. 2009.

Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, Kittles R, Alarcon-Riquelme ME, Gregersen PK, Belmont JW, De La Vega FM, Seldin MF. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat 30:69–78, 2009.

Laan M, Wiebe V, Khusnutdinova E, Remm M, Paabo S. X-chromosome as a marker for population history: linkage disequilibrium and haplotype study in Eurasian populations. J Hum Genet 13:452-462, 2005.

Lins TC, Vieira RG, Abreu BS, Grattapaglia D, Pereira RW. Genetic Composition of Brazilian Population Samples Based on a Set of Twenty Eight Ancestry Informative SNPs. Am J Hum Biol. 2010.

Lum JK, Cann RL, Martinson JJ e Jorde LB. Mitochondrial and nuclear genetic relationship among Pacific Island and Asian populations. Am J Hum Genet 63:613-624,1998.

Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, Di X, Liu WM, Yang G, Liu G, Huang J, Kennedy GC, Ryder TB, Marcus GA, Walsh PS, Shriver MD, Puck JM, Jones KW, Mei R. Parallel genotyping of populations. Hum Genet 118:424–433. 2004

Martinez-Marignac VL, Valladares A, Cameron E, Chan A, Perera A, Globus-Goldberg R, Wacher N, Kumate J, McKeigue P, O'Donnell D, et al. Admixture in Mexico City: implications for admixture mapping of type 2 diabetes genetic risk factors. Hum Genet 120:807–819, 2007.

Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS e Devine SE. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16:1182-1190, 2006.

Moreno Lima R, Abreu BS, Gentil P, Lins TC, Grattapaglia D, Pereira RW, Oliveira RJ. Lack of association between vitamin D receptor genotypes and haplotypes with fat-free mass in postmenopausal Brazilian women. J Gerontol A Biol Sci Med Sci 62:966–972. 2007.

Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SDJ. Color and genomic ancestry in Brazilians. Nat Genet 36(11 Suppl):S54–S60, 2004.

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD. Estimating African American admixture proportions by use of population- specific alleles. Am J Hum Genet 63:1839–1851, 1998.
Parra EJ, Kittles RA, Shriver MD. Implications of correlations between skin color and genetic ancestry for biomedical research. Nat Genet 36(11 Suppl):S54–S60, 2004.
Pena SDJ. Segurança Pública: determinação de identidade genética pelo DNA. In: Seminários Temáticos para a 3ª Conferência Nacional de C, T and I. Parcerias Estratégicas, v. 20, p. 447 – 460, 2005.
Pereira RW, Pena SD. Phylogeography of haplotypes of five microsatellites located in a low-recombination region of the X chromosome: studies worldwide and in Brazilian populations. Genetica 126:243-250, 2006.
Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am J Hum Genetic 68:198-207, 2001.
Pfaff CL, Barnholz-Sloan J, Wagner JK, Long JK. Information on ancestry from genetic markers. Gen Epidem (26): 305-15, 2004.
Pimenta JR, Zuccherato LW, Debes AA, Maselli L, Soares RP, Moura-Neto RS, Rocha J, Bydlowski SP, Pena SD. Color and genomic ancestry in Brazilians: a study with forensic microsatellites. Hum Hered. 62(4):190-5, 2006.
Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data.Genetics 155(2):945-59, 2000.
Reiner AP, Ziv E, Lind DL, Nievergelt CM, Schork NJ, Cummings SR, Phong A, Burchard EG, Harris TB, Psaty BM, Kwok PY. Population Structure, Admixture, and Aging-Related Phenotypes in African American Adults: The Cardiovascular Health Study. Am J Hum Genet 76:463–477, 2005.
Ribeiro D. O Povo Brasileiro: A Formação e o Sentido do Brasil. Companhia das Letras, São Paulo, 242 pp, 1995.
Ribeiro-Rodrigues EM, Santos NPC, Ribeiro dos Santos, AKC, Pereira R, Zago MA, Amorin A, Gusmão L, Santos SEB. Assessing interethnic admixture using an X-linked insertion-deletion multiplex. Am J Hum Biol. 21:707–709, 2009.
Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers of inference of ancestry. Am J Hum Genet. 73(6): 1402-22, 2003
Saitou NA. A genetic affinity analysis of human populations. Hum Evol 10:17-33, 1995.
Salzano FM, Bortolini MC. Evolution and Genetics of Latin American Populations. Cambridge University Press, Cambridge, 512 pp, 2002.
Salzano, FM, Freire-Maia, N: Populações brasileiros: aspectos demográficos, genéticos e antropológicos. São Paulo, EDUSP, 1967.
Salzano FM e Callegari-Jacques SM. South American Indians: a Case Study in Evolution. Research Monographs on Human Population Biology, number 06. Clarendon Press, Oxford, 259 pp, 1988.
Santos NPC, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos AKC, Pereira R, Gusmão L, Amorim A, Guerreiro JF, Zago MA, Matte C, Hutz MH, Santos SEB. Assessing Individual Interethnic Admixture and Population Substructure Using a 48–Insertion-Deletion (INSEL) Ancestry-Informative Marker (AIM) Panel. Hum Mut. 2009.
Santos SEB, Ribeiro-dos-Santos AKC, Santos EJM e Guerreiro JF. The Amazonian microcosm. Ciência e Cultura. 51:181-190, 1999.
Schaffner SF. The X chromosome in population genetics. Nat Rev Genet 5:43–51, 2004.
Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM e Wallace DC. Amerindian mitochondrial DNAs have rare asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet. 46:613-623,1990.
Seldin MF, Tian C, Shigeta R, Scherbarth HR, Silva G, Belmont JW, Kittles R, Gamron S, Allevi A, Palatnik SA, et al. Argentine population genetic structure: large variance in Amerindian contribution. Am J Phys Anthropol 132:455–462, 2007.
Shaffer JR, Kammerer CM, Reich D, McDonald G, Patterson N, Goodpaster B, Bauer DC, Li J, Newman AB, Cauley JA, et al. Genetic markers for ancestry are correlated with body composition traits in older African Americans. Osteoporos Int 18:733–741, 2007.
Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM, Kittles RA. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 112:387–399, 2003.
Suarez-Kurtz G, Vargens DD, Struchiner CJ, Bastos-Rodrigues L, Pena SD. Self-reported skin color, genomic ancestry and the distribution of GST polymorphisms. Pharmacogenet Genomics 17:765–771. 2007.
Tarazona-Santos E, Castilho L, Amaral DR, Costa DC, Furlani NG, Zuccherato LW, Machado M, Reid M, Zalis MG, Rossit AR, Santos SEB, Machado ML, Lustigman S. Population Genetics of *GYPB* and Association Study between *GYPB*S/s* Polymorphism and Susceptibility to *P. falciparum* Infection in the Brazilian Amazon. Plos One. 6(1): e16123, 2011.
Tarazona-Santos E, Raimondi S, Fuselli S. Controlling the effects of population stratification by

admixture in pharmacogenetics. In: Guilherme Suarez-Kurtz. Pharmacogenomics in Admixed populations. Austin, TX: Landes Bioscience, P.1-16, 2007.

Wallace DC, Garrison K e Knowler WC. Dramatic founder effects in Amerindian mitochondrial DNAs. Am J Phys Anthropol. 68:149-155,1985.

Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G. Human diallelic insertion/deletion polymorphisms. Am J Hum Genet 71:854–862, 2002.

Zembrzuski VM, Callegari-Jacques SM, Hutz MH. Application of an African Ancestry Index as a genomic control approach in a Brazilian population. Ann Hum Genet 70::822-8, 2006.

OPEN ACCESS Freely available online PLoS one

Population Genetics of *GYPB* and Association Study between *GYPB*S/s* Polymorphism and Susceptibility to *P. falciparum* Infection in the Brazilian Amazon

Eduardo Tarazona-Santos[1,9], Lilian Castilho[2,9], Daphne R. T. Amaral[2], Daiane C. Costa[2], Natália G. Furlani[3], Luciana W. Zuccherato[1], Moara Machado[1], Marion E. Reid[4], Mariano G. Zalis[5], Andréa R. Rossit[3¤], Sidney E. B. Santos[6], Ricardo L. Machado[3], Sara Lustigman[7]*

1 Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, **2** Laboratório de Pesquisa em Hemoterapia- Hemocentro Campinas, São Paulo, Brazil, **3** Centro de Investigação de Microrganismos, Faculdade de Medicina de São José do Rio Preto, São José do Rio Preto, Brazil, **4** Laboratory of Immunochemistry, Lindsley F. Kimball Research Institute, New York Blood Center, New York, New York, United States of America, **5** Laboratório de Infectologia e Parasitologia Molecular, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, **6** Laboratório de Genética Humana e Médica, Universidade Federal do Pará, Belem, Brazil, **7** Laboratory of Molecular Parasitology, Lindsley F. Kimball Research Institute, New York Blood Center, New York, New York, United States of America

Abstract

Background: Merozoites of *Plasmodium falciparum* invade through several pathways using different RBC receptors. Field isolates appear to use a greater variability of these receptors than laboratory isolates. Brazilian field isolates were shown to mostly utilize glycophorin A-independent invasion pathways via glycophorin B (GPB) and/or other receptors. The Brazilian population exhibits extensive polymorphism in blood group antigens, however, no studies have been done to relate the prevalence of the antigens that function as receptors for *P. falciparum* and the ability of the parasite to invade. Our study aimed to establish whether variation in the *GYPB*S/s* alleles influences susceptibility to infection with *P. falciparum* in the admixed population of Brazil.

Methods: Two groups of Brazilian Amazonians from Porto Velho were studied: *P. falciparum* infected individuals (cases); and uninfected individuals who were born and/or have lived in the same endemic region for over ten years, were exposed to infection but have not had malaria over the study period (controls). The GPB Ss phenotype and *GYPB*S/s* alleles were determined by standard methods. Sixty two Ancestry Informative Markers were genotyped on each individual to estimate admixture and control its potential effect on the association between frequency of *GYPB*S* and malaria infection.

Results: *GYPB*S* is associated with host susceptibility to infection with *P. falciparum*; *GYPB*S/GYPB*S* and *GYPB*S/GYPB*s* were significantly more prevalent in the in the *P. falciparum* infected individuals than in the controls (69.87% vs. 49.75%; $P<0.02$). Moreover, population genetics tests applied on the *GYPB* exon sequencing data suggest that natural selection shaped the observed pattern of nucleotide diversity.

Conclusion: Epidemiological and evolutionary approaches suggest an important role for the GPB receptor in RBC invasion by *P. falciparum* in Brazilian Amazons. Moreover, an increased susceptibility to infection by this parasite is associated with the GPB S+ variant in this population.

Citation: Tarazona-Santos E, Castilho L, Amaral DRT, Costa DC, Furlani NG, et al. (2011) Population Genetics of *GYPB* and Association Study between *GYPB*S/s* Polymorphism and Susceptibility to *P. falciparum* Infection in the Brazilian Amazon. PLoS ONE 6(1): e16123. doi:10.1371/journal.pone.0016123

Editor: Georges Snounou, Université Pierre et Marie Curie, France

Received July 29, 2010; **Accepted** December 14, 2010; **Published** January 24, 2011

Copyright: © 2011 Tarazona-Santos et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This project has been funded in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health number 5R03TW007349-3 and 1R56AI069547-01A2 to SL. This research was also supported in part by the intramural funding of the New York Blood Center to SL and by funding from Brazilian National Research Council (CNPq) to LC, ETS and MM; São Paulo State Research Agency (FAPESP) to LC, DRTA, DCC, NGF, ARBR, RLDM; Minas Gerais State Research Agency (FAPEMIG) to ETS, Brazilian Ministry of Education (CAPES) to LWZ. The Hematology and Hemotherapy Center - Hemocentro UNICAMP, forms part of the National Institute of Science and Technology of Blood, Brazil (INCT do Sangue -CNPq / MCT / FAPESP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: slustigman@nybloodcenter.org

9 These authors contributed equally to this work

¤ Current address: Departamento de Microbiologia e Parasitologia, Universidade Federal Fluminense, Niterói, Brazil

Introduction

Specificity of invasion *of Plasmodium falciparum* merozoites into human red blood cells (RBCs) was the first indication that malaria parasites possess ligands that recognize and interact exclusively with receptors on the surface of the host RBCs [1]. The specificity of malaria parasites for RBC depends on a number of ligand-receptor interactions that are dynamic in *P. falciparum* and provide a greater flexibility to the parasite to overcome the variability in host RBCs and to evade immune responses. The *P. falciparum*

merozoites invade through several pathways using different RBC receptors. Receptors include the glycophorin A, B and C (GPA, GPB and GPC), Band 3 and others (Receptors Y, E, Z and X), whose molecular identity has not yet been determined [2]. Field isolates have shown even greater variability than the laboratory strains in which the known invasion pathways have been defined [3–9]. Our initial studies of field isolates from Mato Grosso, Brazil have revealed that these parasites differ from common laboratory strains, and mostly utilized GPA-independent invasion pathways – the so-called alternative invasion pathways via receptors such as GPB, Receptor Y, Receptor Z and/or others [2].

The genetic polymorphisms of glycophorins or other known receptors play an important role in the resistance to the invasion of erythrocytes by *P. falciparum* [10–19]. In hyperendemic malaria regions of Papua New Guinea deletions in Band 3 (*SLC4A1Δ27*) or in GPC (Ge phenotype, *GYPCΔ*ex3) are highly prevalent and these RBC phenotypes confer selective advantage on morbidity [13–19]. *In vitro* invasion of RBCs that have GPC deficiency is significantly reduced but not eliminated [10,11,20–23].

The polymorphisms identified so far in the two other known *P. falciparum* glycophorin receptors, GPA (GenBank M60707) and GPB (GenBank M60708), have been shown to be only mildly associated with the efficiency of *P. falciparum* invasion and they conferred only partial protection against invasion of RBCs [24,25]. For example, the invasion of *P. falciparum* malaria of En(a−) (lack of GPA) and S−s−U− (lack of GPB) RBCs is significantly reduced but not eliminated [10,11,20–23], presumably because the parasites can use one or more of the other RBC receptors for invasion.

The *GYPA* and *GYPB* genes code for Type I membrane RBC proteins that carry antigens of the MNS blood group system. *GYPA* has two codominant allelic forms, which determine the M or N antigens at the N-terminus of GPA (^{1}SSTTG^{5} for M and ^{1}LSTTE^{5} for N). GPB is identical to GPA^{N} for the first 26 amino acids and, thus, also encodes the N antigen at the N-terminus. *GYPB* also has two codominant alleles: *GYPB**S and *GYPB**s corresponding to S and s antigens, respectively, on the RBC surface. The Ss antigens are defined by an amino acid change at position 29 [Met(S)/Thr(s)] of GPB and are displayed as the S+s−; S−s+ or S+s+ phenotypes [26]. In addition to their sequence homology, *GYPA* and *GYPB* recombination and gene conversion hotspots have been identified, generating on the RBC surface many different hybrid *GYP* gene products that encode GPB bearing low-prevalence antigens like He and Dantu [27]. The GP.Dantu glycophorin is specified by a hybrid gene whose N-terminal sequence is encoded by the *GYPB* gene and C-terminal sequence by the *GYPA* gene. In this hybrid, the genomic breakpoint is located in its intron 4, a composite of *GYPB* and *GYPA* [28]. The hallmark of GPB.He antigen is the change of the N-terminal sequence (from ^{1}LSTTE^{5} for N to ^{1}WSTSG^{5} for He) that abolishes the expression of the common N antigen. GP.He isoform of GPB may or may not carry the S or the s antigen depending on whether it associates with mutations that affect the splicing of exon 4 [29]. Dantu invariably expresses s, albeit weakly. The Dantu variant is prevalent in Africans; 4% vs. 0% in Europeans [30,31]. The *in vitro* invasion of GP.Dantu RBCs by *P. falciparum* is severely compromised [30,32].

Some other *GYPB* nucleotide changes are known to influence the expression of S or s antigens on the RBCs [27,33]. The GPB U antigen is defined by aa 33–39 [34]. Thus, a deletion of *GYPB* exons 2–5 results in S−s−U− phenotype; absence of GPB on the RBCs. Although, the GPB U− RBCs are also S−s−, approximately 16% of S−s− RBCs are U+ (S−s−U+^{var} phenotype) and encoded by a hybrid glycophorin gene [29]. Of these, ~23% are associated with a variant GPB that usually expresses the He antigen, albeit variably [35–37]. Nucleotide changes in or around *GYPB* exon 5 were suggested to be its molecular origin [29]. Notably, a higher prevalence of the S−s−U− phenotype is found in Africa (2–8%); among the pygmies (20%) [38–40], up to 37% of West Africans and ~1% of African Americans [41]. Africans also have a higher prevalence of the Henshaw phenotype (S−s−U+^{var} or GP.He phenotype) [27]. The prevalence of GP.He in African Americans is 3% and up to 7% in people of African origin in Natal, Brazil [27]. The existence of these GPB variants in people of African origin has led to the speculation that these variants may have been selected as a result of the relative resistance that they confer against *P. falciparum* malaria. Recently, a new *GYPB-A-B* recombinant allele (Morobe allele) was found in a highly endemic area in Papua New Guinea [42], but its precise protective effect has not been yet characterized. In South and SE Asia significantly higher prevalence of antigens carried by GPA/GPB hybrids, such as the MUT, MINY, HIL, Hop, St^{a}, Mur and Mi^{a} antigens, are present (0.68–15% vs. 0% in Caucasian), some of which also affect the expression of the GPB S antigen on RBC surface [27]. All these recombinant variants, largely described by blood group scholars, are consistent with the results of a genome wide survey for Copy Number Polymorphisms (CNPs) in the human genome developed using Comparative Genome Hybridization [43], which identified the *GYPB* locus as a CNP in African populations. A recent study have identified the *P. falciparum* ligand for GPB, which raised the possibility that mutations in the gene encoding Glycophorin B in malaria endemic areas could affect susceptibility to malaria through the inability of the ligand to bind to the varied receptor [44].

The Brazilian population exhibits extensive polymorphism in blood group antigens. Although many of these are well documented [45–47], no studies have been done to specifically relate the frequencies of defined polymorphic blood group antigens that function as receptors for *P. falciparum* and the ability of the parasite to invade them. An uncontrolled GPB phenotype-based study of four different ethnic groups in Colombia, suggested an association between the GPB S−s+ variant and a greater resistance to malaria (*P. vivax* and/or *P. falciparum*) in people of African origin [48]. A study by Beiguelman et al. [49] in a rural area of Rondônia was not able to substantiate the studies in Colombia [48] as they did not find any significant associations between GPB SS, Ss or ss phenotypes and *Plasmodium* infection status. However, preliminary studies in four other endemic regions of the Brazilian Amazon including Porto Velho of Rondônia also observed higher frequencies of the GPB S+s+ phenotype among *P. falciparum* malaria patients from Belém and Rio Branco, while higher frequencies of the GPB S−s+ phenotype were found in uninfected blood donors from Belém, Porto Velho, and Rio Branco [50]. The discrepancy between the two studies in Rondônia was attributed to potential differences of the populations studied.

These preliminary studies pointed to a potential association between GPB S+ carriers and their *P. falciparum* infection status, although the significance of these results was impacted by the lack of control for ethnicity, a potential confounding factor. Therefore, our present study in Porto Velho, Rondônia was aimed to further establish whether molecular variation in the *GYPB* gene, particularly the one that generates the *GYPB*S/s* alleles, influences host susceptibility to infection with *P. falciparum*, taking into account the possible confounding factor of ethnicity. In addition, we interpreted our results in the context of the coding nucleotide diversity and haplotype structure of *GYPB*, which might also influence the prevalence of *P. falciparum* infections in this population.

Materials and Methods

Ethics Statement

The research reported here was approved by the IRB of each of the collaborating institutions Hemocentro, UNICAMP, State of Sao Paulo; the Faculdade de Medicina, São José do Rio Preto, State of Sao Paulo; and the Universidade Federal do Rio de Janeiro. The collaborating institutions are registered with the OHRP (FWA00007713; FWA00000377; and FWA00003452, respectively) and their IRB are also registered. The overall protocol was also approved by the IRB from the New York Blood Center (Protocol 415-05). A written informed consent was obtained from all adults, as well as from the parents or legal guardians of minors who participated in the present study.

Study population

Two groups of individuals have been recruited for this study over the period of 2006–2007 in Porto Velho, Rondônia: 1) *P. falciparum* infected individuals; and 2) uninfected individuals who were born and/or have lived in the same endemic region for over ten years, were exposed to infection but have not had malaria in the past or over the 2–3 year study period. All consenting individuals have been interviewed and information regarding their gender, age, date of birth, place of birth, mother's name and maiden surname, ethnic origin of their parent and their grandparents, length of residence in their present locality, history of exposure and/or number of malaria episodes in the last 10 years, and past treatment for malaria were recorded.

Recruitment of *P. falciparum* infected individuals (N = 83) was done in the local healthcare clinic, Centro de Pesquisa em Medicina Tropical (CEPEM), Porto Velho – Rondônia. The consented individuals were of ranging age (18–62) and a female/male ratio of 27/56 (Table 1). The blood samples of the consented infected individuals were analyzed for *P. falciparum* parasitemia using Giemsa stained thick blood smears. The density of parasitemia in the infected individuals was recorded and expressed as the number of asexual *P. falciparum* per microliter of blood assuming a leukocyte count of 8000/µl. All patients with any symptoms of malaria and/or microscopically confirmed infections were given standard and appropriate treatment. The *P. falciparum* treatments involved artemether-lumefantrine (Coartem®) or quinine-doxycycline therapy, the first-line anti-malarial therapies recommended by the Brazilian Ministry of Health, which are superior to other available and affordable treatments. DNA prepared from whole blood samples was later analyzed by PCR for parasite species specific genotyping using established protocols [51] to confirm their infection status with the malaria parasites. It appeared that only 32.3% of all the *P. falciparum* infected individuals had mono-infection with *P. falciparum*; the majority of the individuals had mixed infection with *P. vivax* (66.3%) or had a triple infection with *P. vivax* and *P. malariae* (1.4%).

Active recruitment of the uninfected individuals who were born and/or have lived in the same endemic region for over ten years and have reported that they never had an episode of malaria although their family members or neighbors had it (markers for exposure) was done in the villages around Porto Velho (Ouro Preto do Oeste, Guajará-Mirim, Ji-Paraná and Candeias do Jamari). Those who met the inclusion criteria and consented to participate were tested on site with the OptiMal® (DiaMed AG, Switzerland), a rapid malaria diagnosis test, to validate their non-infection status. Only those that were negative by the OptiMal® kit were bled. To further verify the non-infection status of the individuals in this study group, the DNA extracted from their whole blood sample was later analyzed by PCR for parasite species specific genotyping to identify those that might have asymptomatic malaria infections or to confirm their malaria infection free status using established protocols [51]. It appeared that 30 of the individuals were PCR positive for malaria DNA; 27 of them had *P. vivax* specific DNA and 7 had a mixed *P. vivax* and *P. falciparum* DNA. These individuals were therefore excluded from the uninfected control group but also were not included in the infected group. The final control uninfected group constituted of 199 individuals; 18–56 years of age and a female/male ratio of 97/102. A follow up of these individuals 2 years after the start of the study confirmed that they still had no episodes of malaria in the past or over the study period.

Phenotyping for GPB Ss blood group antigens

The presence of GPB Ss antigens on the surface of the RBCs was detected by the hemagglutination test using specific gel cards (Diamed AG, Morat, Switzerland) and appropriate commercial antibodies. The testing was done using fresh blood from all individuals from both study groups at the Blood Bank of Porto Velho (Centro de Hemoterapia e Hematologia de Rondônia-Fundação Hemeron/Hemotereapy and Hematology Center of the Rondônia State – HEMERON Foundation) on the same day of the blood collection.

Genotyping for *GYPB*S/s*

DNA samples of each individual were prepared from frozen blood samples. Genomic DNA was isolated by a whole blood DNA extraction kit (QIAmp, Qiagen, Valencia, CA) according to the manufacturer's instructions. The DNA solutions were analyzed for quality by agarose gel electrophoresis. The *GYPB* S/s* genotyping was performed using one or more of the following assays:

Allele-specific PCR. Allele-specific PCR (AS-PCR) for the *GYPB*S/s* alleles were performed in all samples. The sequences of primer combinations and control primers that amplified an unrelated gene (human growth hormone gene) were previously published [29]. AS-PCR was carried out under the following conditions: 1× PCR buffer, 1.5 mmol/L $MgCl_2$, 0.2 mmol/L dNTP mix, 100 ng of sense and antisense primers, 100 ng of control primers, 2.5 U *Taq* DNA polymerase and ~50 ng of genomic DNA per 50 µl of total volume. The amplification was performed using a 35-cycles protocol, with an annealing temperature of 62°C.

***GYPB* Exon 5 combination AS/PCR-RFLP assay.** To determine if the S allele was silenced, genomic DNA samples from S−s+ samples genotyped as *GYPB*S/s* were amplified with the GPB4/5, GPBIVS5 and GPB5T primers [29], using a combination AS/PCR-RFLP assay to determine whether *GYPB* is present or absent and to distinguish the variant *GYPB* genes products in S−s+ (*GYPB*S* silent gene) individuals. The PCR products were digested with *Eco*RI overnight at 37°C. The uncut and digested products were analyzed on a 10% polyacrylamide gel.

BeadChip DNA analysis. For some of the *P. falciparum* infected and controls samples for which there was discordance between the two genotyping methods described above and the phenotyping by hemagglutination were re-tested for the *GYPB*S/s* genotypes using a DNA array, BeadChip™ Human Erythrocyte Antigen ("HEA"), containing specific probes directed to polymorphic sites in *RHCE*, *FY* (including *FY-GATA* and *FY265*), *DO* (including *HY and JO*), *CO*, *DI*, *SC*, *GYPA*, *GYPB* (including markers permitting the identification of U-negative and U-variant types), *LU*, *KEL*, *JK*, *LW* and one mutation associated with hemoglobinopathies (HgbS) (BioArray Solutions, Warren, NJ,

Table 1. Demographic characteristics, ancestry estimations and *GYPB**S/s* genotype frequencies in cases and controls, tests for Hardy-Weinberg equilibrium and association between *GYPB**S/s* genotype frequency and infection with malaria.

	Mean Age (SD*)	Females/Males	African ancestry (SD)	European ancestry (SD)	Native American ancestry (SD)	SS (%)	Ss (%)	ss (%)	Significance of test for Hardy-Weinberg equilibrium
Controls (n = 199)	28.29 (9.16)	97/102	0.18 (0.14)	0.54 (0.19)	0.28 (0.17)	15 (7.54)	84 (42.21)	100 (50.25)	$P = 0.65$
Cases (n = 83)	31.76 (11.99)	27/56	0.18 (0.14)	0.54 (0.18)	0.28 (0.18)	3 (3.61)	55 (66.27)	25 (30.12)	$\mathbf{P<0.01}$
Total (n = 282)	29.30 (10.16)	124/158	0.18 (0.14)	0.54 (0.19)	0.28 (0.17)	Association test assuming dominance of S: $\mathbf{P<0.02}$**			

*SD, standard deviation.
**Association persists ($P<0.02$) if age, gender and European, African or Native American ancestry are included as covariates.
doi:10.1371/journal.pone.0016123.t001

USA). The HEA BeadChip assay was performed in accordance with a previously described protocol [52,53].

***GYPB* exon sequencing of *GYPB*.** Specific PCR amplification of *GYPB* is difficult because of its high homology with the two distinct glycophorin genes, *GYPA* and *GYPE*. We therefore designed the *GYPB*-specific primers using the following procedure: we aligned human *GYPA* (GenBank m60707), *GYPB* (GenBank m60708), and *GYPE* (GenBank m29609) sequences and identified sites that are variable and specific for *GYPB* gene segments encompassing exons 2, 4, 5 and 6. The *GYPB*-specific primers had to be at their 3′ terminals sites 100% specific for their own sequences. After PCR and sequencing, we verified the identity of the PCR amplicon by verifying the presence of *GYPB* specific sites. Primers used for PCR amplification contained a M13F or M13R tails (Table 2) and their PCR products included the respective exon and part of its flanking intron regions. PCR amplification was done using 10 ng of genomic DNA, 200 nM of each primer and Platinum PCR SuperMix (Invitrogen, Carlsbad, CA, US) in final volume of 50 µL. The cycling profile used was 2 minutes at 94°C, 45 cycles of 30 seconds at 94°C, 45 seconds at annealing temperature, 2 minutes at 72°C and a final extension of 5 minutes at 72°C. PCR products were purified and bi-directionally sequenced using M13 forward or M13 reverse primers and Applied Biosystem technology (Genewiz, NJ, US). Sequences were analyzed using phred-phrap-consed software [54] and the pipeline described by Machado et al. [55], using the Genbank sequence NC_000004.11 as reference.

Genotyping a *GYPB* tagSNP using Taqman real time PCR assay. To complement the sequencing data and obtain a better coverage of the *GYPB* haplotype structure, we identified SNPs from the HapMap (June 2009) database that were not included in the re-sequenced regions (for example, intronic SNPs), were common (MAF>0.10) and were polymorphic in at least one of the HapMap population. We analyzed the pattern of linkage disequilibrium across *GYPB* from HapMap data and selected two tag-SNPs (that can be used as surrogate for untested SNPs, due to the pattern of linkage disequilibrium): rs4835511 and rs9685167. We genotyped them by TaqMan assays (Applied Biosystems, Palo Alto, CA, US); rs9685167 genotyping did not work properly, while rs4835511 did using 10 ng of genomic DNA, TaqMan® Genotyping Assay 20× and TaqMan® Genotyping Master Mix (Applied Biosystems®, Foster City, CA, US) in final volume of 10 µL.

Ancestry Informative Markers genotyping

We genotyped 62 Ancestry Informative Markers (AIMs) in the DNA samples of all cases and controls. The first set of AIMs consisted of 14 SNPs reported and genotyped in two multiplex reactions as in Da Silva et al. [56] The second set of AIMs included 48 INDELs reported and genotyped in three multiplex reactions as in Santos et al. [57].

Statistical and population genetics analyses

We used the Fisher exact test to assess the Hardy-Weinberg equilibrium. To measure association between S/s genotypes and malaria infection we used the haplotype score test by Schaid et al. [58] and Lake et al. [59] implemented in the software *Haplostats* v.1.4, assuming dominance for the S allele and when necessary, including age, gender, and African, European or Native American

Table 2. Primers designed to amplify *GYPB* exons 2, 4, 5 and 6 by PCR.

Specificity	Primer name	Sequences (in bold are the M13F or M13R tails of primers)	Annealing temperature	Size of PCR product
Exon 2	**M13F**-Exon2-for	**TGTAAAACGACGGCCAGT**GGACTGGAGGGATGTGAGA	55°C	402 bp
	M13R-Exon2-rev	**CAGGAAACAGCTATGACC**TAGAATTCCTCTGTAGTAA		
Exon 4	**M13F**-Exon4-for	**TGTAAAACGACGGCCAGT**GCATGGGACTGGCATCTC	60°C	722 bp
	M13R-Exon4-rev	**CAGGAAACAGCTATGACC**GCTTGGCCTCCCAAAATTATA		
Exon 5	**M13F**-Exon5-for	**TGTAAAACGACGGCCAGT**ATAGTATGTTAACTGTACTTTG	48°C	393 bp
	M13R-Exon5-rev	**CAGGAAACAGCTATGACC**TCTATGTGTCCAGTTGAAAA		
Exon 6	**M13F**-Exon6-for	**TGTAAAACGACGGCCAGT**CAGAGGCTGAAGTGGAGTCT	55°C	276 bp
	M13R-Exon6-rev	**CAGGAAACAGCTATGACC**TAGAGAATACAGTAATAGTG		

doi:10.1371/journal.pone.0016123.t002

ancestry (see below) as covariates. Although *Haplostats* has been developed keeping in mind haplotype analyses, the association test (analogous to Fisher exact tests) is also suitable for single SNPs.

Individual European, African and Native American ancestry were estimated using the Bayesian clustering algorithms developed by Pritchard and implemented in the program STRUCTURE v2.3.2 [60,61]. We assumed that three parental populations (K = 3 clusters) contributed to the genome of the admixed individuals. STRUCTURE estimates individual admixture conditioning on Hardy-Weinberg and linkage equilibrium on each of the K = 3 clusters, that represent the parental populations. We run the program using a length of burn-in Period of 100 000 and 10 0000 repetitions of MCMC after burning. We used prior population information for individuals from the parental populations to assist clustering (USEPOPINFO = 1) and assumed the admixture model for individuals from the admixed populations, inferring the alpha parameter for each population. We also used the parameters GENSBACK = 2 and MIGRPRIOR = 0.05. Moreover, we assumed that allele frequencies were correlated (i.e. similar across parental populations) and that the parental populations show different levels of differentiation (F_{ST}, with prior mean of 0.01 and standard deviation of 0.05). The admixture in each group (cases and controls) is calculated by Structure as the average of the admixture for each individual within the group. We performed these analyses with the two sets of data we produced for this study: (1) the 48 INDELS previously used by Santos et al. [57], using the ancestral populations reported in that publication, and (2) using 62 AIMs (those in Santos et al. [57] and the 14 SNPs reported by da Silva et al. [56], using the parental populations from the latter publication. Both measurements of ancestry were highly correlated (Spearman correlation coefficients: 0.82, 0.89 and 0.92 for African, European and Native American admixture, respectively, with P always$<$0.01).

We inferred haplotypes considering SNPs with a Minor Allele Frequency (MAF) $\geq$0.05, using the method by Stephens and Scheet [62]. The recombination parameter ρ was also calculated for each population by using the method of Li and Stephens [63]. These inferences were performed by the software Phase v.2.1.1., using 10.000 iterations, thinning intervals of 100 and burn in of 1000. Linkage disequilibrium (LD) was estimated by r^2 for SNPs with MAF$\geq$0.05 in at least one population [64] and its significance assessed by LOD scores, using software Haploview v.3.2 [65].

For the analyses of the sequencing data, we assessed intra-population variability using the following estimators of the θ parameter based on the infinite-site-model of mutations: π, the per-site mean number of pair-wise differences between sequences [66], and by θ_w, based on the number of segregating sites [67]. To investigate if the observed patterns of variability in the studied Brazilian population is consistent with the neutral model of evolution, we used the statistical tests of Tajima's D [68], Fu and Li's D* and Fu and Li 's F* [69] on the re-sequencing data, testing these statistics against the standard null hypothesis of neutrality (no natural selection) under constant population size.

Results and Discussion

*GYPB*S/s* frequencies in the control and *P. falciparum* infected populations

Two groups of individuals were studied; the uninfected group (N = 199) and the *P. falciparum* infected group (N = 83). Host DNA from these individuals was used for *GYPB**S/s genotyping by AS-PCR, PCR-RFLP and/or DNA array analysis. Specific PCR amplification of *GYPB*S/GYPB*s* is difficult because of the high homology between *GYPA*, *GYPB* and *GYPE*. To have accurate results we have chosen two different methods for *GYPB*S/GYPB**s genotyping: the allele-specific AS-PCR, an "in house" method, and the HEA (i.e. Human Erythrocyte Antigen) BeadChip, a microarray method commercially available. All samples were analyzed by both methods. When the genotype results in an individual were inconsistent using both methods, we excluded these individuals from further analyses, even thought the HEA BeadChip is known to be more accurate than the AS-PCR [52,53]. Only 3 individuals had mismatching genotypes and were not included in the association studies. Genomic DNA samples from individuals phenotyped S−s+ but genotyped as *GYPB*S/s* were analyzed using a combination AS/PCR-RFLP assay [29,70] in order to determine if the S allele was silenced (*GYPB*S* silent gene).

When the genotype distributions of *GYPB**S/s alleles in the two study groups were compared, the differences in the frequencies of the *GYPB*S/GYPB*S* and *GYPB*S/GYPB*s* vs. *GYPB*s/GYPB*s* genotypes between the *P. falciparum* infected individuals (cases) and the uninfected individuals (control) were significant (69.87% vs. 49.75% of *GYPB*S/GYPB*S* and *GYPB*S/GYPB*s* and 30.1% vs. 50.25% of *GYPB*s/GYPB*s*, respectively; $P<0.02$, Odds Ratio = 1.55 with 95%CI = [1.02, 2.38]) (Table 1). In these analyses, we assumed that the presence of the *GYPB**S allele is a dominant risk factor for susceptibility to infection (i.e., regardless if it is a homozygote or heterozygote), as it results in the phenotypic expression of GPB S+ on the surface of the RBCs. Intriguingly, only the infected individuals do not fit the Hardy-Weinberg expectation; with an excess of the heterozygous *GYPB*S/GYPB*s* genotype (66.27%, Table 1) as compared to what was expected (46.66%).

We observed a discordance of 5% between Ss phenotyping (performed in the field) and genotyping, which is compatible with previous studies [70]. Noteworthy, the results of phenotypes match those of genotypes, both for the excess of S+s+ or *GYPB*S/GYPB*s* heterozygous individuals among cases and for the association of the presence of the GPB S+ or *GYPB*S* putatively dominant allele with their infection status. Although *GYPB*, *GYPA* and *GYPE* genotyping is associated with technical difficulties due to their extensive sequence homology, the high concordance of our phenotype and genotype results and the significance of the association even when phenotypes are analyzed, suggest that our genotyping results are robust.

Interestingly, we have found a higher frequency of heterozygous *GYPB*S/s* genotypes among the *P. falciparum* infected individuals (66.27% in cases versus 42.21% in controls). Considering that heterozygous are more frequent that the homozygous S+s−, it seems that the amount of GPB S receptor molecules doesn't influence the susceptibility to *P. falciparum* infection. If we make an analogy to the Duffy blood group (FY), the receptor for *P. vivax*, and susceptibility of *P. vivax* malaria, different studies [20,71,72], including one in the Amazonian region of Brazil [73], have demonstrated that individuals with the *FYA/FYB* genotype have higher susceptibility to malaria infection. The *Fy* gene has two antigens (Fy^a and Fy^b) that are encoded by the co-dominant alleles *FYA* and *FYB*, located on chromosome 1. The corresponding anti-Fy^a and anti-Fy^b antibodies define four different phenotypes; Fy(a+b+), Fy(a+b−), Fy(a−b+) and Fy(a−b−). The *FYA* and *FYB* alleles differ by one nucleotide change in exon 2 encoding glycine in Fy^a or aspartic acid in Fy^b at residue 42 [73]. In the Brazilian study [73], the authors reported a larger number of malaria episodes among patients with the heterozygote (*FY*A/FY*B*) genotype than the homozygote (*FY*A/FY*A* or *FY*B/FY*B*) genotypes. Individuals homozygous for *FYA* or *FYB* alleles

expressed a lower quantity of the Duffy antigen, which is required for the *P. vivax* invasion, than those who were heterozygotes. Apart from the different levels of the expression, the specific conformation of the Fy^a and Fy^b antigens may also determine differences in the susceptibility to infection. Nevertheless, it was concluded that one of the possible consequences of differential susceptibility to *P. vivax* malaria could be modifications in allelic frequencies of *FY*A* and *FY*B* in populations exposed to *P. vivax*, the most prevalent malaria species in the Brazilian Amazon region.

Our study has verified for the first time that molecular variation in the *GYPB* gene, particularly in the *GYPB**S/s alleles, influenced host susceptibility to infection with *P. falciparum* in Porto Velho, Rondônia. In this study we took into consideration the possible confounding factor of ethnicity by performing association analyses adjusted for admixture, as well as for age and gender (Table 1). An association between other human receptor polymorphisms and variations in the parasite ligands of *P. falciparum* that modulate susceptibility to malaria, was also demonstrated [25,74–76].

*GYPB*S/s* allele frequencies vary across human populations; the *GYPB*S* allele (supposedly associated with infection) is less common in East Asians (~10%) than in Sub-Saharan Africans (>25%) and Europeans (~30–40%). Therefore, if the cases have more European ancestry than controls; a false positive result may emerge due to the association of any variant more common in Europeans, and thus the association results we got might not be at all related with susceptibility to infection conferred by the GPB Ss blood group antigens. To avoid a false positive result, we also measured the association among GPB Ss variants and infection by controlling the effect of admixture. Because we ascertained that African, European and Native American admixture do not differ among cases and controls (Figure 1 and Table 1), we exclude the possibility that our result is a spurious association. Controlling for admixture is essential in genetic epidemiology studies performed in Latin American populations, where large inter-individual differences in admixture are the rule [27].

We hypothesize that this Met29Thr polymorphism might be associated with changes in the structure of the GPB molecule that is used by *P. falciparum* to enter the RBCs. By having a Thr residue (GPB s+) on the RBCs instead of a Met residue (GPB S+), GPB s+ gains potentially a new site for O-glycosylation, which can likely alter its conformation and thus influence the efficiency of invasion by the parasites and ultimately, susceptibility to infection. Alternatively, this polymorphism may also affect dimerization of GPB molecules with other GPB or with GPA molecules [77].

Population genetics of *GYPB* and inferences about natural selection

In principle, the *GYPB**S/s alleles might not be the only functional GPB variants that modify *P. falciparum* invasion efficiencies in the Brazilian population studied. To understand the relationships between the observed association and the haplotype structure of *GYPB* in this population, we sequenced exons 2, 4, 5 and 6 of *GYPB* and their flanking regions, for a total of 1492 bp in sub-samples of cases and controls, matching the proportion of SS, Ss and ss genotypes observed in the total number of cases and controls studied. These sequences are publicly available under the GenBank accession numbers HQ639948–HQ640229. This sub-sample includes 41 cases (2 SS, 26 Ss, 13 ss) and 100 controls (8 SS, 42 Ss and 50 ss). We also used HapMap data available in June 2009 to explore the pattern of linkage disequilibrium across *GYPB* and selected the tag-SNP rs4835511 to be genotyped in the *GYPB* sequenced individuals using *TaqMan* (Applied Biosystem) assay.

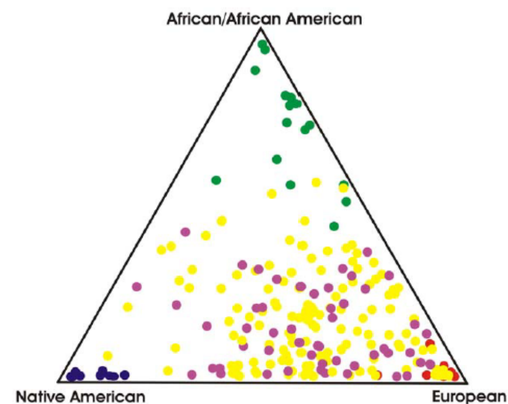

Figure 1. Estimation of admixture using Ancestry Informative Markers genotyping. Individual European, African and Native American ancestry were inferred from 60 ancestry informative markers in cases (magenta) and controls (yellow). Admixture was inferred by comparison with individuals from the putative parental populations: Europeans (red), African/African American (green) and Native Americans (blue). Admixture was estimated using the software Structure and average admixture over cases and controls is shown in Table 1.
doi:10.1371/journal.pone.0016123.g001

Table 3 shows the common *GYPB* haplotypes (i.e. a combination of alleles on the same chromosome) based on eight common SNPs across the sequenced region and the tag-SNP rs4835511. Haplotypes are sorted on the basis of their S/s (rs7683365) allele, and Figure 2 shows that the S/s SNP is in linkage disequilibrium with some common silent polymorphisms in exon 4 and its adjacent introns, but not with all the common SNPs reported in the *GYPB* sequenced regions. In particular, there is no linkage disequilibrium between the S/s alleles (rs7683365) and the non-synonymous common SNP rs1132783 (Ser/Thr) in exon 5. Since this polymorphism is predicted by PolyPhen [78] as benign, we assumed that its role in determining susceptibility to malaria infection is minor in respect to the S/s allele; SNP rs1132783 is located in the transmembrane domain of the protein, thus unlikely to have an effect on the host-parasite interaction. The site-specific PolyPhen algorithm (http://genetics.bwh.harvard.edu/pph/), which uses protein structure and/or sequence conservation information from each gene to predict whether a nonsynonymous mutation is "benign," "possibly damaging," or "probably damaging", was shown to be the best predictor of the fitness effects of nonsynonymous mutations in a study analyzing a large polymorphism data set from 301 human genes [79].

The pattern of nucleotide diversity of *GYPB* revealed by the sequencing data is also informative about the role of the S/s alleles in susceptibility to malaria infection, because it allows inferences about the action of natural selection driven by malaria during the human evolutionary history [80]. Inferences about the action of natural selection have two implications. First, variants on genes inferred to be under selection have contributed to determine phenotype variability and perhaps, differential susceptibility to diseases such as malaria. Second, by definition of natural selection, these variants have been associated with relatively different reproductive efficiencies (i.e. *fitness*) of their carriers, and therefore, they have biomedical relevance. In particular for malaria, Ayodo et al. [81] have evidenced that combining information from

Table 3. *GYPB* haplotype frequencies determined on the re-sequencing panel on the basis of common SNPs (MAF>0.05).

	rs4835511	rs12499907	rs12499906	rs41338748	rs7662277	rs7683365[a]	rs7661933	rs1132783[b]	Cases	Controls	Total
Ancestral allele	C	T	T	T	T	C	T	G			
GYPB-s1	T	.	.	.	.	.	.	.	5	13	18
GYPB-s2	.	.	.	.	.	.	.	.	37[c]	109	146
GYPB-s3	.	.	.	.	.	.	.	C	4	0	4
GYPB-s4	.	.	.	A	.	.	.	C	6	19	25
GYPB-s5	.	.	.	A	.	.	.	.	0	1	1
GYPB-S6	.	G	.	.	A	T	A	.	1	1	2
GYPB-S7	.	.	G	.	A	T	A	.	0	1	1
GYPB-S8	.	G	G	.	A	T	A	.	26	49	75
GYPB-S9	.	G	G	A	A	T	A	.	3	7	10
Number of chromosomes									**82**	**200**	**282**

[a]SNP accounting for S (Thr) and s (Met) phenotypes.
[b]Ser(G)/Thr(C).
[c]The modal haplotype in each group is underlined.
Non-synonymous substitutions are underlined.
doi:10.1371/journal.pone.0016123.t003

association studies and evolutionary inferences about natural selection increases the probability of identifying susceptibility genes of malaria infection.

Analysis of *GYPB* nucleotide diversity in the Brazilian population studied (Table 4) reveals that: (a) *GYPB* shows a level of diversity that is similar to the most variable loci (*ABO*, *SERPINA5*) observed in African populations [82]; the most diverse continental human population. This is due in part to the genomic structure of the glycophorins' region, that encodes the *GYPA*, *GYPB* and *GYPE* highly homologous genes, and which might be affected by high rates of gene conversion among these loci. The observed high diversity may also be due the admixed nature of the Brazilian population, whose diversity reflects the combining effect of the parental population's diversity. (b) When the diversity of *GYPB* is measured separately for haplotypes carrying S or s alleles, diversity is consistently lower across cases and controls for the s haplotypes (associated with resistance to infection) than for the S haplotypes (associated with infection), notwithstanding the higher frequency of the s allele, that being the most common in humans, is expected to be the ancestral one, a condition typically associated with higher nucleotide diversity.

The pattern of genetic diversity on a specific genomic region depends both on the demographic history of populations, as well as on locus specific evolutionary factors such as mutation, recombination and natural selection. Almost 60 years ago, Haldane [83] proposed the so-called "Malaria Hypothesis" - that malaria might act as a selective force on human populations. Since then, several studies have tested and verified this hypothesis in the human host [81] and the malaria parasites [84,85]. To infer if malaria-driven natural selection has shaped the diversity of *GYPB*, we used statistical tests of the null hypothesis of neutrality: that the *GYPB* pattern of diversity may be explained considering only the demographic history of the studied population and mutation and recombination patterns of *GYPB*, without the need to invoke the action of other factors such as natural selection. These statistical tests [68,69] (Tajima's D and Fu-Li's D* and F*) are based on the proportion of rare and common polymorphisms expected in a population under neutrality, and this proportion is informative about natural selection. Table 4 shows that Tajima's and Fu-Li's statistical tests are negative and significantly different from 0, which is indicative of an excess of rare alleles in respect to neutral expectations. This result is consistent with the following scenario involving the action of natural selection (i.e. a *selective sweep*): a beneficial substitution (putatively the s allele) rapidly increases in

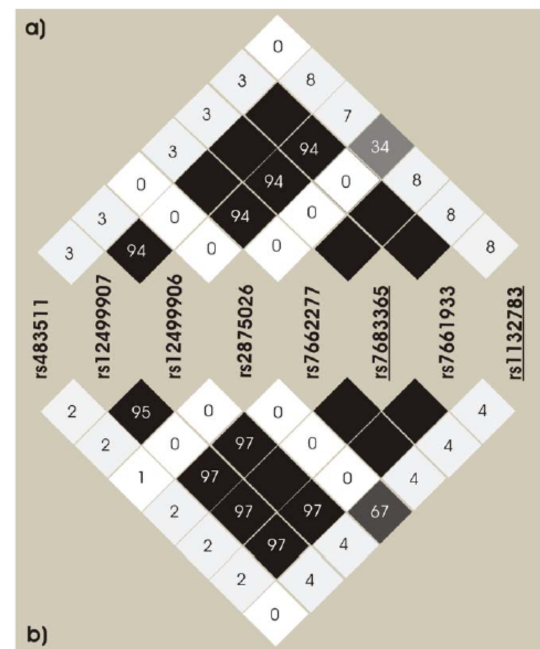

Figure 2. Linkage disequilibrium among common SNPs in *GYPB*. Linkage disequilibrium among common SNPs in *GYPB* was estimated in both study groups: the controls (a) and in the cases, malaria infected individuals from Brazilian Amazon (b). Underlined SNPs are non-synonymous substitutions: rs7683365 is the SNP determining S/s antigens; rs1132783 is a Ser/Thr polymorphism (see Table 3).
doi:10.1371/journal.pone.0016123.g002

Table 4. Summary of *GYPB* diversity indexes and tests of neutrality based on re-sequencing data of a subset of cases and controls and their partitions in S and s alleles (rs7683365) of the Ss blood group antigens.

Populations	Controls*	Controls-S	Controls-s	Cases*	Cases-S	Cases-s
Number of chromosomes	200	58	142	82	30	52
Segregating sites	26	17	8	21	15	4
Singletons	17	16	5	13	14	2
Non-synonymous (total/singletons)	5/3	3/3	1/0	5/3	3/3	1/0
ρ (per adjacent sites $\times 10^3$)	0.32	-	-	0.02	-	-
θ estimators						
$\pi \pm$ SD ($\times 10^3$)	1.80±0.13	0.51±0.29	0.39±0.06	2.09±0.24	0.75±0.53	0.40±0.09
θW ± SD ($\times 10^3$) (per site)	2.97±0.85	2.46±0.87	0.97±0.40	2.83±0.92	2.54±0.99	0.59±0.33
Neutrality tests						
Tajima's D	−1.092	**−2.413**[b]	−1.376	−0.777	**−2.372**[b]	−0.720
Fu and Li's D*	**−5.400**[a]	**−5.355**[a]	**−3.002**[c]	**−3.578**[a]	**−3.972**[a]	−1.217
Fu and Li's F*	**−4.385**[a]	**−5.138**[a]	**−2.891**[c]	**−3.035**[a]	**−4.071**[a]	−1.243

*The samples of cases and controls were selected so the proportion of SS, Ss and ss genotypes observed in the total set of cases and controls was matching.
[a]$P<0.02$,
[b]$P<0.01$,
[c]$P<0.001$.
doi:10.1371/journal.pone.0016123.t004

frequency (i.e. incomplete sweep) carrying its associated haplotype through a hitchhiking effect. This process, driven by natural selection, is too rapid for recombination to shuffle the surrounding haplotype. As a consequence, it is expected that the nucleotide diversity (π) of the haplotypes carrying the beneficial allele (s in this case) is lower than the alternative allele (S), as observed in our data (Table 4). During a selective sweep, rare substitutions become common both because: the rise in frequency of the s haplotype; other substitutions associated with S become rare; and new (and therefore rare) substitutions appear in the expanding positively selected haplotype. The negative values observed for the neutrality tests for *GYPB* is consistent with this scenario (Table 4).

In the case of the Brazilian admixed population, the observed excess of rare alleles in *GYPB* can not be a consequence of admixture, which generally results in a reduced proportion of rare alleles in respect to neutral expectations (and therefore, to positive values of statistical tests such as D, D* and F*) [69]. Instead, we are likely observing the signature of a selective sweep that occurred during the last thousands of years of the human evolution in malaria affected regions mainly in Southern Europe and more likely Africa [86], where the ancestors of our studied Brazilian population settled (see Table 1 for the predominant European admixture proportion). Although our interpretation may be true and consistent with the results of the present association study, an unambiguous inference about the action of natural selection would require a comparison of *GYPB* diversity with a set of other loci not affected by natural selection, that would allow to obtain a more realistic null neutral distribution of neutrality statistics (D, D*, F*) that incorporate the specific demographic history of the studied population. On the other hand, the lack of evidence of the action of natural selection on African or European populations on genomic screenings of signatures of natural selection [87–89] may be due to the difficulties in genotyping/sequencing of the glycophorins' genomic regions.

Altogether, our results suggest that the S domain on GPB is important for its binding to the specific ligand of the *P. falciparum* parasite, EBL-1, which was recently identified and characterized [44]. RBCs carrying glycophorin B but not RBCs lacking glycophorin B (S−s−U−) were shown to adsorb the native EBL-1 from *P. falciparum* culture supernatants. Future studies are needed to demonstrate whether EBL-1 binds differentially to GPB S−s+ vs. S+s+ RBCs, and thus indirectly substantiate at a molecular level our observation of association studies at the population level and our population genetics inferences about the action of natural selection. Performing similar studies in other regions of the Amazons and other endemic regions of the world is also needed to further substantiate our observations; including both association studies with larger sample sizes and with a population genetics approach that include sequencing and genotyping at a higher resolution. Because *P. falciparum* shows also high population structure, in particular in the Amazon Region [90], it is also important to understand the extent of variability in host-parasite interaction and their co-evolution. The statistical association between the presence of the S allele and infection supports also the hypothesis that *P. falciparum* parasites in the Brazilian Amazon regions utilize GPB as a key receptor for invasion, and consequently individuals who carry distinct *GYPB* gene variants, which might facilitate erythrocyte invasion, will be more susceptible to *P. falciparum* infection. Thus, our results reinforce the need of studies focusing on *in vitro* invasion assays using erythrocytes with diverse *GYPB* genotypes and *P. falciparum* strains from different origin to establish the role of the GPB receptor for *P. falciparum* parasites of this region.

Acknowledgments

We thank Dr. Cheng-Han Huang and Yogi Vedvyas for designing and optimizing the GYPB primers sets, Professor Luiz Hildebrando Pereira da Silva for allowing us to use the facilities at Centro de Pesquisa em Medicina Tropical (CEPEM) and Elieth Mesquita for helping in the recruitment of *P. falciparum* infected individuals and blood collections at health clinic at CEPEM. We also thank the Managing Director of the Blood Bank of Porto Velho (Centro de Hemoterapia e Hematologia de Rondônia- Fundação Hemeron// Hemotereapy and Hematology Center of the Rondônia State – HEMERON Foundation) for helping us process the blood samples collected from uninfected individuals. We also thank Maria Clara da Silva and Marilia Scliar for sharing information about genotyping and sequencing protocols.

Author Contributions

Conceived and designed the experiments: SL LC ETS RLM ARR MGZ SEBS. Performed the experiments: LC DRTA DCC NGF SEBS RLM. Analyzed the data: SL LC ETS RLM ARR DRTA DCC NGF LWZ MM SEBS. Contributed reagents/materials/analysis tools: ETS SEBS. Wrote the paper: SL ETS LC RLM MR.

References

1. McGhee RB (1953) The infection by Plasmodium lophurae of duck erythrocytes in the chicken embryo. J Exp Med 97(6): 773–82.
2. Cowman AF, Crabb BS (2006) Invasion of red blood cells by malaria parasites. Cell 124(4): 755–66.
3. Okoyeh JN, Pillai CR, Chitnis CE (1999) Plasmodium falciparum field isolates commonly use erythrocyte invasion pathways that are independent of sialic acid residues of glycophorin A. Infect Immun 67(11): 5784–91.
4. Baum J, Pinder M, Conway DJ (2003) Erythrocyte invasion phenotypes of Plasmodium falciparum in The Gambia. Infect Immun 71(4): 1856–63.
5. Baum J, Thomas AW, Conway DJ (2003) Evidence for diversifying selection on erythrocyte-binding antigens of Plasmodium falciparum and P. vivax. Genetics 163(4): 1327–36.
6. Lobo CA, de Frazao K, Rodriguez M, Reid M, Zalis M, et al. (2004) Invasion profiles of Brazilian field isolates of Plasmodium falciparum: phenotypic and genotypic analyses. Infect Immun 72(10): 5886–91.
7. Bei AK, Membi CD, Rayner JC, Mubi M, Ngasala B, et al. (2007) Variant merozoite protein expression is associated with erythrocyte invasion phenotypes in Plasmodium falciparum isolates from Tanzania. Mol Biochem Parasitol 153(1): 66–71.
8. Deans AM, Nery S, Conway DJ, Kai O, Marsh K, et al. (2007) Invasion Pathways and Malaria Severity in Kenyan Plasmodium falciparum Clinical Isolates. Infect Immun 75(6): 3014–20.
9. Jennings CV, Ahouidi AD, Zilversmit M, Bei AK, Rayner J, et al. (2007) Molecular analysis of erythrocyte invasion in Plasmodium falciparum isolates from Senegal. Infect Immun 75(7): 3531–8.
10. Pasvol G, Wainscoat JS, Weatherall DJ (1982) Erythrocytes deficiency in glycophorin resist invasion by the malarial parasite Plasmodium falciparum. Nature 297(5861): 64–6.
11. Pasvol G, Jungery M, Weatherall DJ, Parsons SF, Anstee DJ, et al. (1982) Glycophorin as a possible receptor for Plasmodium falciparum. Lancet 2(8305): 947–50.
12. Hadley TJ (1986) Invasion of erythrocytes by malaria parasites: a cellular and molecular overview. Annu Rev Microbiol 40: 451–77.
13. Booth PB, Albrey JA, Whittaker J, Sanger R (1970) Gerbich blood group system: a useful genetic marker in certain Melanesians of Papua and New Guinea. Nature 228(270): 462.
14. Booth PB, McLoughlin K, Hornabrook RW, Macgregor A, Malcolm LA (1972) The Gerbich blood group system in New Guinea. II. The Morobe District and North Papuan Coast. Hum Biol Oceania 1(4): 259–66.
15. Macgregor A, Booth PB (1973) A second example of anti-Ge1, and some observations on Gerbich subgroups. Vox Sang 25(5): 474–8.
16. Booth PB, Tills D, Warlow A, Kopec AC, Mourant AE, et al. (1982) Red cell antigen, serum protein and red cell enzyme polymorphisms in Karkar Islanders and inhabitants of the adjacent North Coast of New Guinea. Hum Hered 32(6): 385–403.
17. Serjeantson SW, White BS, Bhatia K, Trent RJ (1994) A 3.5 kb deletion in the glycophorin C gene accounts for the Gerbich-negative blood group in Melanesians. Immunol Cell Biol 72(1): 23–7.
18. Patel SS, King CL, Mgone CS, Kazura JW, Zimmerman PA (2004) Glycophorin C (Gerbich antigen blood group) and band 3 polymorphisms in two malaria holoendemic regions of Papua New Guinea. Am J Hematol 75(1): 1–5.
19. Allen SJ, O'Donnell A, Alexander ND, Mgone CS, Peto TE, et al. (1999) Prevention of cerebral malaria in children in Papua New Guinea by southeast Asian ovalocytosis band 3. Am J Trop Med Hyg 60(6): 1056–60.
20. Miller LH, Haynes JD, McAuliffe FM, Shiroishi T, Durocher JR, et al. (1977) Evidence for differences in erythrocyte surface receptors for the malarial parasites, Plasmodium falciparum and Plasmodium knowlesi. J Exp Med 146(1): 277–81.
21. Facer CA (1983) Merozoites of P. falciparum require glycophorin for invasion into red cells. Bull Soc Pathol Exot Filiales 76(5): 463–9.
22. Facer CA (1983) Erythrocyte sialoglycoproteins and Plasmodium falciparum invasion. Trans R Soc Trop Med Hyg 77(4): 524–30.
23. Pasvol G (1984) Receptors on red cells for Plasmodium falciparum and their interaction with merozoites. Philos Trans R Soc Lond B Biol Sci 307(1131): 189–200.
24. Weatherall DJ (1996) Host genetics and infectious disease. Parasitology 112 Suppl: S23–9.
25. Miller LH (1994) Impact of malaria on genetic polymorphism and genetic diseases in Africans and African Americans. Proc Natl Acad Sci U S A 91(7): 2415–9.
26. Wang HY, Tang H, Shen CK, Wu CI (2003) Rapidly evolving genes in human. I. The glycophorins and their possible role in evading malaria parasites. Mol Biol Evol 20(11): 1795–804.
27. Reid ME, Lomas-Francis C (2004) *The Blood Group Antigen.* 2nd ed. FactBook Seriers Elsevier, Academic Press.
28. Huang CH, Blumenfeld OO (1988) Characterization of a genomic hybrid specifying the human erythrocyte antigen Dantu: Dantu gene is duplicated and linked to a delta glycophorin gene deletion. Proc Natl Acad Sci U S A 85(24): 9640–4.
29. Storry JR, Reid ME, Fetics S, Huang CH (2003) Mutations in GYPB exon 5 drive the S−s−U+(var) phenotype in persons of African descent: implications for transfusion. Transfusion 43(12): 1738–47.
30. Unger P, Procter JL, Moulds JJ, Moulds M, Blanchard D, et al. (1987) The Dantu erythrocyte phenotype of the NE variety. II. Serology, immunochemistry, genetics, and frequency. Blut 55(1): 33–43.
31. Contreras M, Green C, Humphreys J, Tippett P, Daniels G, et al. (1984) Serology and genetics of an MNSs-associated antigen Dantu. Vox Sang 46(6): 377–86.
32. Field SP, Hempelmann E, Mendelow BV, Fleming AF (1994) Glycophorin variants and Plasmodium falciparum: protective effect of the Dantu phenotype in vitro. Hum Genet 93(2): 148–50.
33. Storry JR, Reid ME, MacLennan S, Lubenko A, Nortman P (2001) The low-incidence MNS antigens M(v), s(D), and Mit arise from single amino acid substitutions on GPB. Transfusion 41(2): 269–75.
34. Dahr W, Moulds JJ (1987) High-frequency antigens of human erythrocyte membrane sialoglycoproteins, IV. Molecular properties of the U antigen. Biol Chem Hoppe Seyler 368(6): 659–67.
35. Blumenfeld OO, Huang CH (1997) Molecular genetics of glycophorin MNS variants. Transfus Clin Biol 4(4): 357–65.
36. Reid ME, Storry JR, Ralph H, Blumenfeld OO, Huang CH (1996) Expression and quantitative variation of the low-incidence blood group antigen He on some S−s−red cells. Transfusion 36(8): 719–24.
37. Huang CH, Blumenfeld OO, Reid ME, Chen Y, Daniels GL, et al. (1997) Alternative splicing of a novel glycophorin allele GPHe(GL) generates two protein isoforms in the human erythrocyte membrane. Blood 90(1): 391–7.
38. Fraser GR, Giblett ER, Motulsky AG (1966) Population genetic studies in the Congo. 3. Blood groups (ABO, MNSs, Rh, Jsa). Am J Hum Genet 18(6): 546–52.
39. Vos GH, Moores P, Lowe RF (1971) A comparative study between the S−s−U−phenotype found in Central African Negroes and the S−s+U−phenotype in Rhnull individuals. S Afr J Med Sci 36(1): 1–6.
40. Lowe RF, Moores PP (1972) S−s−U−red cell factor in Africans of Rhodesia, Malawi, Mozambique and Natal. Hum Hered 22(4): 344–50.
41. Issitt PD, Anstee DJ (1998) Applied blood group serology. 4th ed, ed. N.C. Durham Montgomery Scientific Publications.
42. Patel SS, Katsnelson S, McNamara D, Reeder JC, Kazura JW, et al. (2004) Multiple Erythrocyte Polymorphisms Including a Glycophorin B Variant in a malaria endemic area of Papua New Guinea. ASTMH Abstract #913 2004.
43. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, et al. Origins and functional impact of copy number variation in the human genome. Nature 464(7289): 704–12.
44. Mayer DC, Cofie J, Jiang L, Hartl DL, Tracy E, et al. (2009) Glycophorin B is the erythrocyte receptor of Plasmodium falciparum erythrocyte-binding ligand, EBL-1. Proc Natl Acad Sci U S A 106(13): 5348–52.
45. Baleotti W, Jr., Rios M, Reid ME, Fabron A, Jr., Pellegrino J, Jr., et al. (2003) A novel DI*A allele without the Band 3-Memphis mutation in Amazonian Indians. Vox Sang 84(4): 326–30.
46. Castilho L, Rios M, Pellegrino J, Jr., Saad ST, Costa FF, et al. (2004) A Novel FY Allele in Brazilians. Vox Sang 87(3): 190–5.
47. Baleotti W, Jr., Rios M, Reid ME, Hashmi G, Fabron A, Jr., et al. (2006) Dombrock gene analysis in Brazilian people reveals novel alleles. Vox Sang 91(1): 81–7.
48. Montoya F, Restrepo M, Montoya AE, Rojas W (1994) Blood groups and malaria. Rev Inst Med Trop Sao Paulo 36(1): 33–8.
49. Beiguelman B, Alves FP, Moura MM, Engracia V, Nunes AC, et al. (2003) The association of genetic markers and malaria infection in the Brazilian Western Amazonian region. Mem Inst Oswaldo Cruz 98(4): 455–60.
50. Cavasini CE, De Mattos LC, Alves RT, Couto AA, Calvosa VS, et al. (2006) Frequencies of ABO, MNSs, and Duffy phenotypes among blood donors and malaria patients from four Brazilian Amazon areas. Hum Biol 78(2): 215–9.
51. Kimura M, Kaneko O, Liu Q, Zhou M, Kawamoto F, et al. (1997) Identification of the four species of human malaria parasites by nested PCR that targets variant sequences in the small subunit rRNA gene. Parasitology International 46: 91–95.
52. Hashmi G, Shariff T, Seul M, Vissavajjhala P, Hue-Roye K, et al. (2005) A flexible array format for large-scale, rapid blood group DNA typing. Transfusion 45(5): 680–8.
53. Hashmi G, Shariff T, Zhang Y, Cristobal J, Chau C, et al. (2007) Determination of 24 minor red blood cell antigens for more than 2000 blood donors by high-throughput DNA analysis. Transfusion 47(4): 736–47.

54. Montgomery KT, Iartchouck O, Li L, Loomis S, Obourn V, et al. (2008) PolyPhred analysis software for mutation detection from fluorescence-based sequence data. Curr Protoc Hum Genet Chapter 7: Unit 7 16.
55. Machado M, Magalhães WCS, Sene A, Araújo B, Faria-Campos A, et al. (2010) Phred-Phrap-Consed-Polyphred to analyses tools: a pipeline to facilitate population genetics re-sequencing studies. Investigative Genetics In Press.
56. da Silva MC, Zuccherato LW, Soares-Souza GB, Vieira ZM, Cabrera L, et al. (2010) Development of two flexible multiplex mini-sequencing panels of ancestry informative SNPs for admixture studies in Latin Americans and their application to populations of the state of Minas Gerais in Brazil. Genetics and Molecular Research 9(4): 2069–85.
57. Santos NP, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK, Pereira R, Gusmao L, et al. (2010) Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. Hum Mutat 31(2): 184–90.
58. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 70(2): 425–34.
59. Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, et al. (2003) Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. Hum Hered 55(1): 56–65.
60. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67(1): 170–81.
61. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2): 945–59.
62. Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet 76(3): 449–62.
63. Li N, Stephens M (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165(4): 2213–33.
64. Hill WG, Robertson AR (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38: 226–231.
65. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21(2): 263–5.
66. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105(2): 437–60.
67. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7(2): 256–76.
68. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3): 585–95.
69. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133(3): 693–709.
70. Omoto R, Reid ME, Castilho L (2008) Molecular analyses of GYPB in African Brazilians. Immunohematology 24(4): 148–53.
71. Miller LH, Mason SJ, Dvorak JA, McGinniss MH, Rothman IK (1975) Erythrocyte receptors for (Plasmodium knowlesi) malaria: Duffy blood group determinants. Science 189(4202): 561–3.
72. Michon P, Woolley I, Wood EM, Kastens W, Zimmerman PA, et al. (2001) Duffy-null promoter heterozygosity reduces DARC expression and abrogates adhesion of the P. vivax ligand required for blood-stage infection. FEBS Lett 495(1–2): 111–4.
73. Cavasini CE, de Mattos LC, Couto AA, Couto VS, Gollino Y, et al. (2007) Duffy blood group gene polymorphisms among malaria vivax patients in four areas of the Brazilian Amazon region. Malar J 6: 167.
74. Lell B, May J, Schmidt-Ott RJ, Lehman LG, Luckner D, et al. (1999) The role of red blood cell polymorphisms in resistance and susceptibility to malaria. Clin Infect Dis 28(4): 794–9.
75. Carter R, Mendis KN (2002) Evolutionary and historical aspects of the burden of malaria. Clin Microbiol Rev 15(4): 564–94.
76. Zimmerman PA, Patel SS, Maier AG, Bockarie MJ, Kazura JW (2003) Erythrocyte polymorphisms and malaria parasite invasion in Papua New Guinea. Trends Parasitol 19(6): 250–2.
77. Dahr W (1986) Recent Advances in Blood Group Biochemistry VVengelen-Tyler, WJJudd, eds. ArlingtonVA: American Association of Blood Banks. pp 23–65.
78. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30(17): 3894–900.
79. Williamson SH, Hernandez R, Fledel-Alon A, Zhu L, Nielsen R, et al. (2005) Simultaneous inference of selection and population growth from patterns of variation in the human genome. Proc Natl Acad Sci U S A 102(22): 7882–7.
80. Sabeti PC (2008) Natural selection: Uncovering mechanisms of evolutionary adaptation to infectious disease. Nature Education 1(1).
81. Ayodo G, Price AL, Keinan A, Ajwang A, Otieno MF, et al. (2007) Combining evidence of natural selection with association analysis increases power to detect malaria-resistance variants. Am J Hum Genet 81(2): 234–42.
82. Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, et al. (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol 2(10): e286.
83. Haldane JBS (1949) Disease and evolution. Ricerca Science Supplement 19: 3–10.
84. Escalante AA, Barrio E, Ayala FJ (1995) Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. Mol Biol Evol 12(4): 616–26.
85. Neafsey DE, Schaffner SF, Volkman SK, Park D, Montgomery P, et al. (2008) Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. Genome Biol 9(12): R171.
86. Campbell MC, Tishkoff SA (2008) African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet 9: 403–33.
87. Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, et al. (2010) A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327(5967): 883–6.
88. International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, et al. (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467(7311): 52–8.
89. 1000 Genomes Project Consortium, Durbin RM, Abecasis GR, Altshuler DL, Auton A, et al. (2010) A map of human genome variation from population-scale sequencing. Nature 467(7319): 1061–73.
90. Machado RLD, Povoa MM, Calvosa VSP, Ferreira MU, Rossit ARB, et al. (2004) Genetic Structure of *Plasmodium falciparum* populations in the Brazilian Amazon Region. The Journal of Infectious Disease 190(9): 1547–1555.

3 .CONCLUSÕES

3. CONCLUSÕES

A partir do uso de marcadores genômicos de ancestralidade em populações de indivíduos sadios (controles) e portadores de malária por *Plasmodium falciparum* (casos), que residem na cidade de Porto Velho, RO (Amazônia Ocidental), Brasil, foi possível concluir que:

A pesquisa da ancestralidade genômica individual por meio dos AIMs é útil na determinação da proporção de contribuição das populações parentais e, portanto, da ancestralidade de grupos e subgrupos a serem pareados em estudo de caso controle. As populações portadora de malária por *Plasmodium. falciparum* e respectivos controles saudáveis são igualmente miscigenadas.

- Predomina neste grupamento da população Brasileira a ancestralidade Européia.

- O grau de miscigenação individual/populacional não é fator preocupante e não gerará resultados espúrios em futuros estudos de caso-controle e/ou epidemiológicos que tenham como sujeitos indivíduos que residem na região aqui considerada.

4. REFERÊNCIAS BIBLIOGRÁFICAS

4. REFERÊNCIAS BIBLIOGRÁFICAS

1. Choudhry S, Coyle NE, Tang H, Salari HTK, Lind D, Clark SL, Tsai SLCH, Naqvi M, Ung APN, Matallana H, Avila PC, Casal J, Torres A, Nazario S, Castro R, Perez-Stable EJ, Kwok P-Y, Sheppard D, Shriver MD, Rodriguez-Cintron W, Risch N, Burchard EG. Population stratification confounds genetic association studies among Latinos. Hum Genet. 2006;118:652–664.
2. Callegari-Jacques SM., Salzano FM. Brasilian Indian/non-Indian interections and their effects. Ciênc Cult. 1999; 51:166-174.
3. Carvalho-Silva, DR, Santos, FR, Rocha, J, Pena, SD: The phylogeography of Brazilian Y-chromosome lineages. AM J Hum Genet 2001; 68:281-286.
4. IBGE – HTTP: //www.ibge.gov.br/home/estatística/população. Censo 1991, 2000 e Brasil 500 anos de povoamento. Acesso em 14/dez. 2010.
5. Salzano FM, Bortolini MC. Evolution and Genetics of Latin American Populations. Cambridge University Press, Cambridge, 2002; 512 pp.
6. Schaffner SF. The X chromosome in population genetics. Nat Rev Genet 2004; 5:43–51.
7. Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrell RE Ethnic-affiliation estimation by use of population-specific DNA markers. Am J Hum Genet 1997; 60:957–964.
8. Bonilla C, Parra EJ, Pfaff CL, Marshall DSJA, Hamman RF, Ferrell RE, Hoggart CL, McKeigue PM, Shriver MD. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. Am Hum Genet 2004; 68:139–153.

9. Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am J Hum Genetic 2001; 68:198-207.
10. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G. Human diallelic insertion/deletion polymorphisms. Am J Hum Genet. 2002; 71:854–862.
11. Pena SDJ. Segurança Pública: determinação de identidade genética pelo DNA. In: Seminários Temáticos para a 3ª Conferência Nacional de C, T and I. Parcerias Estratégicas, 2005; v. 20, p. 447 – 460.
12. Carlson B. SNPs - A Shortcut to Personalized Medicine. Genetic Engineering & Biotechnology News (Mary Ann Liebert, Inc.): p. 12. Disponível em http://www.genengnews.com/articles/chitem.aspx?aid=2507. Acesso em 15/Dez 2010.
13. Oliveira-Ferreira J, Lacerda MV, Brasil P, Ladislau JL, Tauil PL, Daniel-Ribeiro CT. Malaria in Brazil: an overview.Malar J. 2010; Apr 30;9:115.
14. MINISTÉRIO DA SAÚDE. Disponível em http://portal.saude.gov.br/portal/saude/profissional/area.cfm?id_area=1526. Acesso em 10/nov. 2009.