Universidade Federal do Rio Grande do Sul

Instituto de Biociências

Programa de Pós-graduação em Genética e Biologia Molecular

VARIABILIDADE GENÉTICA EM POPULAÇÕES AMERÍNDIAS E ASIÁTICAS: REGIÃO 16p13.3 E INSERÇÕES *ALU*

Jaqueline Battilana

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências

Orientador: Dr. Francisco Mauro Salzano

Co-orientador: Dr. Sandro Luis Bonatto

Porto Alegre

Março de 2005

Este trabalho foi realizado nas instalações do Laboratório de Evolução Molecular do Departamento de Genética do Instituto de Biociências da Universidade Federal do Rio Grande do Sul (UFRGS) e no Centro de Biologia Genômica e Molecular da Faculdade de Biociências da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), subvencionado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela Financiadora de Estudos e Projetos (FINEP), pela Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

Aos povos dos Continentes Americano e Asiático, dedico este trabalho como uma modesta contribuição ao conhecimento de suas origens.

Quem, de três milênios,
Não é capaz de se dar conta
Vive na ignorância, na sombra,
À mercê dos dias, do tempo.

Johann Wolfgang von Goethe

AGRADECIMENTOS

Aos meus orientadores,

...do que unicamente precisamos para nos tornarmos bons filósofos é a capacidade de nos admirarmos com as coisas...

Ao Dr. Francisco M. Salzano, a quem respeito, admiro e por quem tenho grande carinho, agradeço pela confiança depositada.
Ao Dr. Sandro L. Bonatto, por me proporcionar um ambiente que enriqueceu meus conhecimentos e no qual cresci cientificamente.

À Dra Loreta B. de Freitas, por ter disponibilizado os recursos humanos e materiais necessários para elaboração de parte deste trabalho.

Aos professores do Programa de Pós-Graduação em Genética e Biologia Molecular e do Departamento de Genética do Instituto de Biociências da UFRGS, por contribuírem para minha formação.

À Universidade Federal do Rio Grande do Sul, por me acolher e proporcionar os meios pelos quais aqui cheguei.

Aos colegas e amigos do PPGBM e Departamento de Genética.

À Laci Krupahtz pela amizade, ajuda incansável e exemplo de garra.

Ao Clênio D. Machado, pelo auxílio laboratorial e pelo agradável convívio.

Ao Elmo J. A. Cardoso e a Ellen Mezzeck, pelo imprescindível trabalho e companheirismo.

Aos Coordenadores do Centro de Biologia Genômica e Molecular da PUCRS, Sandro Luis Bonatto, Jomar Laurino e Mauricio Bogo, por me receberem e disponibilizarem todos os recursos humanos e materiais necessários para a elaboração de parte deste trabalho.

Ao Dr. Jomar Laurino pela assistência no processo de clonagem.

À Luana Cardoso pela ajuda significativa da parte experimental.

À Cladinara Sarturi Roberts e ao André Schnorr pelo auxílio laboratorial, amizade e alegria.

Ao Nelson J. R. Fagundes, colega científico e amigo, por sua inestimável ajuda na finalização deste trabalho.

Aos colegas e amigos do Laboratório de Evolução Molecular do Departamento de Genética, por compartilharmos dificuldades e alegrias e ainda estarmos juntos.
Aline, Carlos André, Cláudia E. Thompson, Cláudia Dornelles, Dânae, Franceli, Geraldo, Joana, Nicole, Pakisa, Patricia, Valéria e Verônica.

Aos colegas e amigos do Centro de Biologia Genômica e Biologia Molecular, por me receberem e proporcionar um amigável ambiente de trabalho.
Ana Maria, André, Arethuza, Cladinara, Cláudia, Eduardo, Fabiana, Felipe, Fernanda Britto, Fernanda Munari, Fernanda Vianna, Fernando, Gabriela, Gabrielle, Helena, Leonara, Lisiane, Luana, Luiza, Márcio, Mariana, Manuel, Nadiesca, Nelson, Paulo, Peter, Raquel A., Raquel L, Ricardo, Roberta e Rodrigo.

Vocês todos moram no meu coração!

Aos amigos Carlos André, Lutécia Pereira e Renata Medina, pela força e apoio que sempre me deram, gosto muito de vocês.

À Cláudia Schmitz, a great woman, a great teacher.

À Maria Regina Sana e Mônica Giaretton Bicca, pelo incentivo e tudo mais.

E porque todo cientista também é humano...

Às minhas famílias Battilana e Lehmann, por tudo.

Aos meus amigos, que enriquecem meu mundo.

Ao Mauricio, por sempre ter me apoiado, compreendido e me aceitado como sou. Amo-te.

SUMÁRIO

RESUMO

Um estudo envolvendo populações asiáticas e ameríndias foi realizado para avaliar os relacionamentos históricos e genéticos entre eles através de polimorfismos moleculares autossômicos. Um deles é uma seqüência polimórfica localizada na região 16p13.3. Um total de 1558 pares de base foram investigados em 98 indivíduos da Mongólia, Beringia e das Américas. Estes resultados foram comparados com aqueles obtidos em uma prévia investigação por outros autores. Cinqüenta e cinco sítios polimórficos foram classificados em trinta e cinco haplótipos. Uma árvore de haplótipos (median joining network) baseada neles, revelou dois grupos distintos, um mais compacto com haplótipos derivados das cinco categorias etno-geográficas estabelecidas; enquanto o outro, com haplótipos mais divergentes, era composto principalmente por africanos e ameríndios. Quase todos os parâmetro de neutralidade apresentaram valores negativos. Simulações realizadas para interpretar estes resultados foram executadas com dois conjuntos de dados: um composto exclusivamente de ameríndios e outro com populações mundiais. O primeiro, sugerindo crescimento e declínio populacional, rejeitou os cenários com crescimento abaixo de cinco vezes e anteriores a ~18 mil anos atrás; enquanto que o segundo, sugerindo crescimento populacional, não rejeitou cenários nos quais a magnitude de crescimento foi maior que dez vezes e anteriores a ~21 mil anos atrás, provavelmente refletindo um crescimento antigo fora da África.

Doze polimorfismos de inserções *Alu* foram também estudados em 170 indivíduos pertencentes a 7 grupos nativos sul-americanos, 60 a dois siberianos, e 91 a dois mongóis. Estes dados foram integrados com aqueles de 488 indivíduos associados a outros 13 grupos, para determinar as relações entre asiáticos, Beringianos e ameríndios. Nestes três grupos, foi observado um decréscimo da heterozigosidade e da quantidade de fluxo gênico, na mesma ordem indicada acima. A solidez destas subdivisões foi demonstrada nas distâncias genéticas, nas análises de componentes principais, de variância molecular, no teste de Mantel, e numa abordagem Bayesiana de atribuição genética. Entretanto, não pode ser observada uma clara estrutura entre os nativos Sul-americanos, indicando a importância dos fatores dispersivos (deriva genética, efeito fundador) na sua diferenciação.

A congruência dos resultados obtidos com os dois marcadores são: (a) não foi confirmada uma história de declínio populacional forte associado com a chegada do

homem pré-histórico nas Américas; (b) os ameríndios não apresentaram estruturação clara dentro do continente e não se diferenciaram dos asiáticos do nordeste da Ásia (beringianos); e (c) o povo Aché foi o grupo mais divergente.

ABSTRACT

A study involving Asian and Amerindian populations was performed to evaluate the historical and genetic relationships among them using autosomal molecular polymorphisms. One of them is a polymorphic sequence located in the 16p13.3 region. A total of 1558 base pairs was investigated in 98 individuals of Mongolian, Beringian, and Amerindian affiliation, and the results compared with those obtained in a previous research, by other authors. Fifty-five polymorphic sites could be classified in thirty-five haplotypes. A median joining network based on them revealed two distinct clusters, one more compact with haplotypes derived from the five geographic-ethnic categories established; while the other, with the most divergent haplotypes, was composed mainly by Africans and Amerindians. Almost all neutrality parameters yielded negative values. Simulations to interpret these results were performed with two datasets: one exclusively of Amerindians and another with worldwide populations. The first, suggesting population growth and decline, rejected scenarios with a growth below 5-fold and before ~18 thousand years BP; while the second also rejected population growth starting ~21 thousand years BP and below 10-fold of magnitude of growth, the signal of population increase probably being a reflection of an ancient growth out of Africa.

Twelve *Alu* insertion polymorphisms, were also studied in 179 individuals belonging to 7 South American Native, 60 to two Siberian, and 91 to two Mongolian populations. These data were integrated with those from 488 persons affiliated with 13 other groups, to ascertain the relationships between Asian, Beringian, and Amerindian populations. A decreasing trend concerning heterozygosities and amount of gene flow was observed in the three sets, in the order indicated above. Genetic distances, principal components analysis, analysis of molecular variance, Mantel test, and a Bayesian approach to genetic assignment, all indicated the validity of these subdivisions. However, no clear structure could be observed within South American Natives, indicating the importance of dispersive (genetic drift, founder effects) factors in their differentiation.

The congruence of results obtained with the two markers are: a history of strong population decline associated with the prehistoric human arrival in the Americas was not supported; the Amerindians did not presented intracontinental structure and did not differentiated from Northeast Asians (beringians); the Aché people was the most divergent group.

CAPÍTULO I

INTRODUÇÃO GERAL

1. INTRODUÇÃO

1.1. O microcosmo americano

Os povos Ameríndios são uma rica fonte de diversidade cultural e genética. Muitos deles vivem em comunidades nas quais suas atividades de subsistência (caça, coleta de alimentos e agricultura incipiente) assemelham-se àquelas prevalentes entre grupos nas primeiras fases da história humana, o que faz deles um ótimo modelo para estudos evolutivos. Várias características apresentadas por estes grupos justificam sua importância. Um aspecto importante é o demográfico, já que os grupos geralmente são pequenos, com isolamento relativo e sua ecologia é simples, o que os torna adequados para investigações sobre a relação entre estrutura populacional e variabilidade genética. Muitos deles possuem uma dinâmica populacional conhecida como processo de fissão-fusão (Salzano e Callegari-Jacques, 1988). Esse processo, com raízes sócio-culturais, leva os grupos a flutuações no tamanho populacional, tornando-os sujeitos a agentes aleatórios no processo de transmissão genética. Estes fatores interagem com aqueles determinísticos (mutação e seleção natural) resultando na variabilidade que vemos hoje (Salzano, 1978, 1982). Outro aspecto importante, é que a data de sua entrada no Continente Americano está estabelecida dentro de razoáveis limites. Além disso, muitos estudos têm sido realizados entre os Ameríndios envolvendo diversas áreas, as quais são essenciais às interpretações evolutivas, tais como demografia, epidemiologia, antropologia social, linguagem e arqueologia.

Embora a informação a respeito dos povos Ameríndios seja abundante, ela é significativamente heterogênea em relação às populações e aos tipos de sistemas genéticos investigados (revisão em Salzano, 2002). De um modo geral, o nível de variabilidade genética dos Ameríndios é considerável quando comparado a outros grupos, mas não pode ser facilmente explicada de acordo com a estrutura geográfica, linguística e histórica. Existe restrição de variabilidade para alguns dos marcadores de mtDNA e HLA, mas não necessariamente para outras características genéticas. Por outro lado, a variação interpopulacional parece ser mais marcada em Ameríndios que em qualquer outro lugar, provavelmente devido à sua estrutura populacional. As diferenças mais marcantes entre Ameríndios e não-Ameríndios são aquelas relatadas para o sistema HLA, provavelmente devido a processos históricos e por exposição diversificada a agentes infecciosos. Certamente, muitas diferenças genéticas puderam ser detectadas ao longo do continente

algumas delas sendo graduais, enquanto outras são mais abruptas. De certo modo isto seria esperado devido à quantidade variada de movimentos populacionais e de ambientes distintos com os quais eles se depararam.

Muitos estudos relacionando as populações nativas americanas e seus possíveis ancestrais asiáticos (beringianos, siberianos, mongóis, etc) vêm sendo realizados, os quais resultam em questionamentos a respeito do povoamento do Novo Mundo. As principais questões levantadas são: (a) quando os nativos americanos ancestrais chegaram pela primeira vez nas Américas ? (b) quantas expansões ou migrações populacionais estiveram envolvidas neste processo de colonização; e (c) de onde na Ásia/Eurásia estes grupos ancestrais vieram? A afinidade entre siberianos e ameríndios é apoiada por evidências antropológicas (Kozintsev *et al.*, 1999), dentais (Turner, 1984) e genéticas, incluindo os polimorfismos de DNA mitocondrial (mtDNA) (Starikovskaya *et al.*, 2005), do cromossomo Y (Lell *et al.*, 2002), e da variação autossômica nos genes do HLA (Human Leukocite Antigens) de classe II (Uinuk-ool *et al.*, 2002).

Os dois sistemas genéticos mais comumente usados em estudos de populações de Nativos Americanos e grupos afins, têm sido o mtDNA e a porção não recombinante do cromossomo Y (NRY). Estes dois marcadores (mtDNA e NRY) fornecem informação sobre uma entrada inicial no continente pelos Nativos Americanos ancestrais entre 20.000 e 15.000 anos atrás, o que favorece uma expansão tardia dos primeiros americanos. Como estas datas referem-se ao último período glacial médio, que seria antes de haver um corredor livre de gelo disponível para a passagem das populações humanas, os grupos colonizadores podem ter usado a rota costeira durante seus movimentos iniciais em direção à América do Norte. Os primeiros imigrantes aparentemente trouxeram com eles para a América os haplogrupos A-D (talvez X) do DNA mitocondrial, o haplogrupo P-M45a e haplótipo Q-242/M3 do cromossomo Y, os quais se dispersaram através das áreas continentais no Novo Mundo. Uma expansão posterior teria trazido o haplogrupo X do mtDNA (e talvez mais haplótipos A-D) e contribuído com os haplogrupos P-M45b, C-M130 e R1a1-M17 do NRY para as populações nativas Americanas. Estes teriam se disseminado apenas nas Américas do Norte e Central. Esta expansão poderia ter coincidido com a abertura do corredor livre de gelo por volta de 12.550 anos atrás. Uma outra expansão posterior à última mencionada provavelmente envolveu o surgimento das populações do Círculo Ártico, tais como Eskimós, Aleutas e Na-Dené, quando estes

migraram da Beringia para o norte da América do Norte depois do último período glacial máximo (revisão em Schurr, 2004 e Schurr e Sherry, 2004).

Entre os polimorfismos genéticos autossômicos estão os genes do HLA de classe II, polimorfismos de repetições curtas em tandem (STRs) ou microssatélites, polimorfismos de inserção/deleção (*Alu* e LINE ou long interspersed elements), polimorfismos de comprimento de fragmento de restrição (RFLPs), polimorfismos de número variável de repetições em tandem (VNTRs) ou minissatélites, e polimorfismos de nucleotídeos únicos (SNP) (revisão em Salzano, 2002). Estes, estudados em número crescente nas populações Ameríndias, até o momento não foram suficientes para estabelecer um padrão geral de variabilidade que explique a relação inter- e intra-continental ameríndia, devido à heterogeneidade dos dados (conforme mencionado anteriormente).

Estudos que envolvem sociedades humanas são complexos, porque as relações entre os indivíduos determinam sua estrutura populacional. Sendo assim, nenhuma área do conhecimento que se dedica a estes estudos é independente. Os estudos sobre linguagem, antropologia e arqueologia ajudam na interpretação dos dados genéticos e vice-versa. As populações nativas americanas continuam sendo consideradas importantes do ponto de vista da genética de populações e evolução porque elas unificam os conceitos de isolamento geográfico das populações fundadoras (com ou sem gargalos-de-garrafa), e diversificação lingüística. As populações ameríndias sofreram grandes baixas desde a chegada dos colonizadores europeus (~500 anos), estima-se uma redução populacional de 95% (Cavalli-Sforza *et al.*, 1994), extinguindo-se portanto grande parte desta diversidade. Mesmo assim, vários grupos sobrevivem até o presente em relativo isolamento. Devido ao rápido avanço no processo de aculturação, a perda de seus valores culturais e genéticos vem se tornando inevitável. Por esse motivo, os estudos referentes a eles merecem grande atenção e urgência. O presente estudo se propõe a fazer uso da ferramenta genética para analizar a dinâmica histórica dos povos ameríndios, mas sem deixar de lado o ponto de vista lingüístico, antropológico e arqueológico.

1.2. O segmento hipervariável subterminal 16p13.3 humano

A seqüência hipervariável do cromossomo 16p foi descrita por Alonso e Armour (2001) como um segmento altamente variável, localizado na região subterminal do braço

curto do cromossomo 16 humano, de aproximadamente 50 quilobases (kb) de extensão. Esta região é vizinha à extremidade 5' final do minissatélite MS205 e presume-se que seja neutra, pois está dentro de um grande íntron do gene do canal de Ca^{2+} tipo T ativado por baixa voltagem (CACNA1H) (Badge *et al.*, 2000). Além disto, esta região é rica em guanina e citosina (65% G+C), a qual, quando sujeita à desaminação mediada por metilação, pode alcançar taxas de transição 5 vezes maiores do que a taxa de mutação basal. Esta região contém CpGs metilados no DNA de células somáticas e espermáticas. Adicionalmente, ela localiza-se em uma região sujeita a uma alta taxa de recombinação, que pode ajudar a protegê-la de futuros efeitos de distorção causados por "hitchhiking" genético (carona genética) ou seleção basal.

No trabalho de Alonso e Armour (2001), foi estudado o DNA genômico de 50 indivíduos [20 africanos (10 pigmeus e 10 quenianos), 10 japoneses, 10 britânicos e 10 bascos]. Foram detectadas 42 substituições e um evento de deleção (envolvendo 5 pares de bases a partir da posição 219) nesses 100 cromossomos (todas as 100 seqüências haplotípicas foram submetidas ao GenBank – acessos AJ391838 a AJ391937). A diversidade nucleotídica π estimada variou entre 0,3% para os pigmeus e 0,04% para os britânicos. A divergência (K) foi estimada comparando-se uma seqüência escolhida ao acaso de um indivíduo pigmeu com a seqüência de um chimpanzé (GenBank, acessos AJ252012, AJ252013 e AJ252014). A estimativa da taxa média de mutação por sítio por ano foi de $2{,}19 \times 10^{-9}$. Ela é maior que aquelas encontradas em várias outras seqüências descritas na literatura (Harding *et al.*, 1997; Harris e Hey, 1999; Jaruzelska *et al.*, 1999; Kaessmann *et al.*, 1999) e também mais elevada que a taxa autossômica média estimada em $1{,}28 \times 10^{-9}$ (Nachman e Crowell, 2000). A taxa de mutação por seqüência (1.742 sítios) por geração (20 anos) foi estimada em $7{,}63 \times 10^{-5}$. A abundância de pares CpG poderia ser uma explicação desta taxa bastante elevada, porque mais de 40% das mutações detectadas estavam em dinucleotídios CpG.

Os resultados encontrados apontam para uma maior diversidade nas populações africanas, enquanto que nos euro-asiáticos a variabilidade parece ter derivado recentemente de um pequeno subconjunto de linhagens africanas. O ancestral comum mais recente foi estimado ter vivido a 1,04 milhões de anos atrás. Além disso, os autores encontraram evidência de forte crescimento populacional para algumas das populações, conciliando

inferências de marcadores nucleares e mitocondriais. Os testes de neutralidade também mostraram evidências de crescimento populacional para as populações euro-asiáticas.

Em relação à recombinação, mesmo que ela possa diminuir o poder dos testes de neutralidade, acredita-se que ela tenha uma freqüência baixa o suficiente para não distorcer a reconstrução genealógica da região analisada.

Alonso e Armour (2001) acreditam que, embora deva ser analisado um número maior de populações, o crescimento populacional encontrado, geograficamente associado com populações não-africanas, estaria mais provavelmente ligado a um processo de expansão, de acordo com o modelo de origem africana recente (out-of-Africa).

Como indicado, portanto, essa seqüência pode constituir-se em uma fonte rica de variação polimórfica, útil para estudos de evolução humana. O número de estudos envolvendo seqüências nucleares em povos ameríndios ainda é reduzido se comparado aos dos marcadores clássicos (grupos sangüíneos e proteínas) e uniparentais (mtDNA e cromossomo Y), e até o momento não houve investigação deste segmento em povos ameríndios. Estes são os principais argumentos que justificam a escolha deste marcador para os estudos aqui propostos.

1.3. Polimorfismo de inserção-deleção: inserções *Alu*

Uma porcentagem significante do genoma dos mamíferos consiste de seqüências espalhadas de DNA repetitivo. Estes geralmente são classificados como SINEs (short interspersed elements) ou LINEs (long interspersed elements) (Okada *et al.*, 1997). A família *Alu* de SINEs constitui-se em um dos elementos genéticos móveis mais bem sucedidos e extensivamente estudados do genoma humano. Estima-se que os elementos *Alu* estejam espalhados pelo genoma com uma freqüência média de um elemento a cada 5kb (quilobases) de DNA (Deininger *et al.*, 1981; Rinehart *et al.*, 1981; Jelnick e Schmid, 1982; Batzer *et al.*, 1990; Bailey e Shen, 1993; Novick *et al.*, 1996), representando aproximadamente 10% do genoma nuclear em humanos (Smit, 1996). Assim como outros SINEs, as seqüências *Alu* ocorrem em alta freqüência dentro de domínios não codificantes (por exemplo, regiões inter-gênicas, introns, UTRs 5' e 3', etc.) e nas bandas reversas (R) de cromossomos metafásicos (Daniels e Deininger, 1985; Korenberg e Rykowski, 1988; Bailey e Shen, 1993).

Novas cópias destes elementos deslocam-se dentro do genoma por retrotransposição (revisão em Weiner *et al.* 1986). Nos SINEs, o RNA envolvido neste processo é transcrito pela RNA Polimerase III (Weiner *et al.*, 1986). As seqüências *Alu* possuem em geral aproximadamente 300 pares de bases (pb) (Schmid e Jelinek, 1982), e consistem de dois monômeros distintos ligados por uma região de oligo-d(A). Sua estrutura é semelhante à do RNA 7SL humano (Labuda e Zietkiewicz, 1994). Sugere-se que um "*Alu* fóssil" teria surgido pela deleção de um "domínio-S" central do RNA 7SL e pela adição de uma cauda de d(A) na porção 3', que poderia ter facilitado a transcrição reversa (Quentin, 1992) do cDNA da seqüência *Alu*, que é a cópia que se integra ao genoma do hospedeiro (Shaikh *et al.*, 1997).

Comparações entre membros de diferentes subfamílias de *Alu* mostraram que, em média, 70% das seqüências são idênticas, mostrando alto grau de homogeneidade. Isto sugere que apenas um ou um pequeno número de membros desta família estiveram envolvidos nas amplificações das mesmas em um dado momento (Britten *et al.*, 1988; Deininger e Slagel, 1988; Matera *et al.*, 1990; Batzer e Deininger, 1991). Estas seqüências com amplificação ativa no passado têm sido referidas como genes *Alu master* (Shen *et al.*, 1991). Todas as evidências atuais indicam que as subfamílias surgiram preferivelmente via amplificação genômica, e não através do processo de conversão gênica (Deininger *et al.*, 1992).

Mudanças nos genes *Alu master*, devido a substituições nucleotídicas em suas seqüências, possibilitaram a organização da família *Alu* em 12 subfamílias distintas. Estas subfamílias são classificadas de acordo com suas idades em "velhas, intermediárias e jovens", sendo que as duas subfamílias mais velhas, Jo e Jb, surgiram de eventos de retrotransposição independentes envolvendo um único *Alu* ancestral, há aproximadamente 81 milhões de anos atrás (Kapitonov e Jurka, 1996).

A amplificação de subfamílias *Alu* não tem ocorrido a uma taxa constante durante a evolução, sendo que as principais retrotransposições de *Alu* completaram-se há 30 milhões de anos atrás (Mighell *et al.*, 1997). Mesmo assim, em subfamílias jovens ainda pode ocorrer este processo como um evento raro (Shaikh e Deininger, 1996). A prova disso é que muitas seqüências *Alu* ainda não se encontram fixadas no genoma humano (Batzer *et al.*, 1990, 1994, 1995; Batzer e Deininger, 1991; Novick *et al.*, 1995, 1998; Arcot *et al.*,

1997; Stoneking *et al.*, 1997). Elas, portanto, são polimórficas quanto à ausência ou presença da inserção em um loco específico.

Os elementos *Alu* polimórficos podem servir como marcadores para elucidar vários aspectos da evolução humana (Salem *et al.*, 2003), além de revelar padrões de diversidade (Batzer e Deininger, 1991; Stoneking *et al.*, 1997). As vantagens do seu uso como marcadores genéticos são substanciais. Os dados bi-alélicos dos elementos *Alu* são relativamente livres do estado de homoplasia. É altamente improvável que uma seqüência tenha ocupado um sítio cromossômico específico mais de uma vez na evolução humana, ou seja, todos os *loci* que possuem a mesma inserção *Alu* são derivados de um único evento de transposição. Além disso, inserções recentes representam polimorfismo estável e elas são raramente perdidas sem deixar vestígio (Novick *et al.*, 1996; Stoneking *et al.*, 1997). Parece não haver um mecanismo preciso de excisão dos elementos *Alu* do genoma humano, e conseqüentemente, vestígios reconhecíveis das seqüencias *Alu* em algum episódio passado de inserção e remoção parcial (Novick *et al.*, 1996). Então, devido à identidade por descendência e estabilidade decorrente de deleção incompleta, as inserções *Alu* são literalmente fósseis moleculares. Outro aspecto interessante dos estudos populacionais é que a condição ancestral é definida pela ausência do elemento *Alu*. Mesmo sabendo-se que os elementos *Alu* afetam o genoma de diferentes maneiras causando mutações, recombinação entre elementos, conversão gênica e alterações na expressão gênica (Batzer e Deininger, 2002; Jurka *et al.*, 2004), o polimorfismo de suas inserções continua rendendo bons resultados nos estudos de genética de populações humanas.

As doze inserções que serão analisadas no presente trabalho incluem membros das subfamílias AluYa5, AluYa8 e AluYb8. Em estudos anteriores, as seqüências *Alu* mostraram ser uma ótima ferramenta, gerando resultados interessantes (Batzer *et al.*, 1994, 1996; Arcot *et al.*, 1995a,b, 1996; Stoneking *et al.*, 1997), inclusive em populações indígenas sul-americanas (Novick *et al.*, 1998; Battilana *et al.*, 2002; Oliveira, 1999). Estes marcadores têm se mostrado eficientes no agrupamento das populações do continente sul-americano, separando-as das européias e africanas, mostrando, como grupo intermediário, os asiáticos.

1.4. Objetivos

Os objetivos do presente estudo são os seguintes:

- Descrever a variabilidade nucleotídica na seqüência 16p13.3 em populações do continente americano e asiático.

- Descrever a variabilidade genotípica de 12 inserções *Alu* em 14 populações ameríndias, 2 do Círculo Ártico e 8 asiáticas, e estabelecer o seu padrão intra e intercontinental.

- Comparar os resultados obtidos com as inserções *Alu* e a seqüência 16p13.3 para as populações do continente americano, avaliando a sua variabilidade genética intra e inter populacional e outros parâmetros genéticos.

- Inferir sobre o povoamento das Américas, comparando os resultados obtidos com os de outros marcadores e evidências derivadas de outras disciplinas.

CAPÍTULO II

Molecular Variability of the 16p13.3 Region in Amerindians and its Anthropological Significance

---

Manuscrito a ser submetido ao Annals of Human Genetics

Molecular variability of the 16p13.3 region in Amerindians and its anthropological significance

J. Battilana[1], L. Cardoso-Silva[1], R. Barrantes[2], M. H. Hutz[1], F. M. Salzano[1] and S. L. Bonatto[3]

[1]Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil

[2]Sección de Genética Humana, Instituto Nacional de Investigaciones en Salud (INISA), Universidad de Costa Rica, San José, Costa Rica

[3]Centro de Biologia Genômica e Molecular, Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, 90619-900 Porto Alegre, RS, Brazil

Running title: 16p13.3 variability in Amerindians

Key words: 16p13.3, Amerindians, molecular variability

Address for correspondence and reprints: Francisco M. Salzano, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. Tel.: 555133166747; Fax: 555133167311, E-mail: francisco.salzano@ufrgs.br.

Summary

A total of 1558 base pairs in the 16p13.3 region was investigated in 98 individuals of Mongolian, Northern Arctic, and Amerindian affiliation, and the results compared with those obtained in a previous research of the same sequence. Fifty-five polymorphic sites could be classified in thirty-five haplotypes. A median joining network based on them revealed two distinct clusters, one more compact with haplotypes derived from the five geographic-ethnic categories established; while the other, with the most divergent haplotypes, was composed mainly by Africans and Amerindians. Almost all neutrality parameters yielded negative values. Simulations performed to interpret these results considered two dataset: one exclusively of Amerindians and other with worldwide populations. The first, suggesting population growth and decline, rejected scenarios with growth below 5-fold and before ~18 thousand years BP; the second also rejected population growth starting ~21 thousand years BP and below 10-fold of magnitude of growth, the signal of population increase probably being a reflection of an ancient growth out of Africa.

Introduction

Studies of human genetic variation provide a powerful means for elucidating the genetic, evolutionary, and demographic factors shaping the human genome. Considerable work has accumulated over the last decades, documenting DNA sequence variation in humans. Early studies focused primarily on mitochondrial DNA (mtDNA) (Vigilant *et al.* 1991) and Y chromosome (Hammer, 1995; Whitfield *et al.* 1995; Underhill *et al.* 2000). More recent single-locus studies have focused on the X chromosome (Nachman *et al.* 1998; Harris & Hey, 1999; Kaessmann *et al.* 1999; Nachman & Crowell, 2000; Gilad *et al.* 2002; Saunders *et al.* 2002; Verrelli *et al.* 2002; Yu *et al.* 2002) and autosomes [reviewed in Przeworski *et al.* (2000); Excoffier (2002)]. Particularly on autosome noncoding regions, the number of surveys is still low but are beginning to reveal possible demographic events and mechanisms of evolutionary change at the molecular level. In relation to Amerindians, the number of surveys and diversity of markers have been growing [review in Salzano (2002)]. Recent studies on noncoding regions include Fagundes *et al.* (2005) and Heller *et al.* (2005).

Alonso & Armour (2001) studied a region localized immediately flanking the 5' end of the minisatellite MS205 at 16p13.3 in 100 chromosomes sampled from different African and Euroasiatic populations. It maps within a large intron approximately 50 kb long of a low voltage-activated T-type Ca^{2+} channel gene (CACNA1H) (Badge *et al.* 2000) and is G+C rich (65%). G+C-rich regions can contain frequent CpG dinucleotides, which, if subjected to methylation-mediated deamination, may reach transition rates five times the background mutation rate. They argued that this region does contain CpGs methylated in both somatic and sperm DNA, and that in addition, it maps to a region of high recombination. This fact may help to shield it from the distorting effects of genetic hitchhiking or background selection. Actually, the estimate of the average mutation rate per site per year found by these authors (2.19×10^{-9}), was higher than several others estimates in the literature (Harding *et al.* 1997; Harris & Hey, 1999; Jaruzelska *et al.* 1999; Kaessmann *et al.* 1999; Nachman & Crowell, 2000). Consequently, the region could constitute a source of sequence polymorphism useful for human evolution studies. We have sequenced 1558 base pairs of this region in a set of different ethnic Amerindian and Asian groups and integrated these date with those of Alonso & Armour (2001) with the

following questions in mind: (a) In what ways do the investigated region differ, in terms of molecular variability, from most recent studies, described after Alonso & Armour's (2001)? (b) Can the extended set of data shade new light on the process of continental human diversification? and (c) What inferences can be made about early peopling of the Americas and the subsequent process of Amerindian molecular-genetic diversification?

Subjects and Methods

Samples

Information about the studied populations is presented in Table 1. A total of 98 individuals had been tested, 4 Mongolians (2 Khalkh, 2 Khoton), 4 Northern Arctic (2 Chukchi, 2 Eskimo), 12 Central Americans (from six tribes) and 78 South Americans (affiliated with 17 tribes or ethnic groups). As is indicated by the geographical coordinates, they are spread all over Central and South America, while non-Americans are located in places from which the prehistoric colonization of the New World presumably has taken place. Wide linguistic representation also occurs among Central and South Americans.

DNA Extraction, PCR Amplification, and DNA Sequencing

The region analyzed includes the first 1558 base pairs (bp) of the 1742bp sequence studied by Alonso & Armour (2001). Part of the genomic DNA samples were extracted from plasmas and glycerolized red blood cells stored in Porto Alegre, using the QIAamp DNA Mini Kit (Qiagen). Six primer pairs were designed with overlapping regions to amplify the 1558 base pairs (one of them designed by Alonso & Armour, 2001. Primer sequences are available on request). Different methods of amplification were performed to obtain the target fragments (Touchdown, Hot Start and Nested PCR). Amplification was performed using 10-50 ng of genomic DNA, 1.5 mM of $MgCl_2$, 0.2 mM of each dNTP, 0.2 μM of each primer, and 0.5 U of *Taq* DNA polymerase (AmpliTaq Gold [Applied Biosystems] or Platinum [Invitrogen Life Technologies]). Cycle conditions were 94ºC for 1 min, 62ºC for 2 min, and 72ºC for 2 min, with an initial denaturing step of 94ºC for 1 min and a final extension step of 72ºC for 10 min. The PCR products were purified with exonuclease I and alkaline phosphatase (1U/μl; Amersham Biosciences) and sequenced on both strands using the amplification primers. Sequencing was performed with the DYEnamic ET Dye Terminator Kit (MegaBACE, Amersham Biosciences) as instructed and read in a MegaBace1000 (Amersham Biosciences) automated system. Samples with uncertain phase results were sequenced again in an ABI Prism 3100 (Applied Biosystems) machine following the manufacturer's protocols. All determinations which indicated possible variants were confirmed by exhaustive re-sequencing.

Haplotype Assignment

Samples with confirmed multiple variants had their haplotypes determined experimentally, by cloning the PCR products with the Topo TA Cloning Kit for Sequencing (Invitrogen Life Technologies). Plasmidial DNA extraction was done according to Sambrook & Russell (2001). Sequencing of the cloned fragments was performed as indicated above, in the MegaBace1000 machine.

Data Analysis

All chromatograms were checked using the CHROMAS 1.45 (www.technelysium.com.au/index.html) program, and all sequences were manually aligned using BIOEDIT 6.0.7 (www.mbio.ncsu.edu/BioEdit/bioedit.html). To test departures from the neutral model, such as selection or population change, Tajima's D (Tajima, 1989), Fu & Li's D* and F* (Fu & Li, 1993) and Fu's Fs (Fu, 1997) statistics were calculated; additionally, diversity parameters such as haplotype (Hd) and nucleotide (π) diversity (Nei, 1987), as well as theta (θ) (Watterson, 1975) were obtained employing the DnaSP 4.00 (http://www.ub.es/DNASP) software. The relationships among haplotypes were obtained by the median joining method using NETWORK 4.1.0.8 (www.fluxus-engineering.com). ARLEQUIN V. 2000 (Schneider *et al.* 2000) was used to perform the Analysis of Molecular Variance-AMOVA.

Mean divergence (K) between humans (all diferent haplotypes) and one chimpanzee (GenBank accession no. AJ252012) was calculated using PAUP*, with a Tamura-Nei with invariants distance selected by MODELTEST (Posada & Crandall, 1998). The mutation rate was inferred from the divergence value by using the formula $\mu = K/(2t)$, considering a divergence time (t) of 6 million years.

To test if there are signals of past demographic fluctuations in the samples of all individuals (worldwide sample), and Amerindians alone, simulations were performed using Rogers' algorithm (Rogers, 1995), as implemented by Wooding *et al*. (2004), in the DFSC 1.0 program (http://www.xmission.com/~wooding/DFSC/). These simulations were also performed with the original data of Alonso & Armour (2001) alone. This algorithm assumes that *t* generations ago, a population increased suddenly from an ancient population size (N_0) to a larger population size (N_1), with an infinite-sites mutation rate (μ). The simulations considered a combination of demographic scenarios, from growth factors of 1-

fold to 500-fold, beginning 3,000 to 150,000 years ago for all individuals and for the Amerindians dataset. In Amerindians, assumptions about population decline were performed considering reduction factors of 1-fold to 50-fold, beginning 3 to 30,000 years ago.

Results

Polymorphic Sites and Haplotype Determination

A total of 1558 base pairs of the 196 chromosomes were sequenced here. Of the 98 individuals tested, 91 had their haplotypes directly determined, since they were either homozygous or heterozygous for a single site. The remaining seven had their phase assigned by cloning (seven clones of each individual were sequenced in the MegaBACE1000 machine) as well as by sequencing them again in other instrument (ABI Prism 3100). They were: one Aché, one Kaingang, one Pacaás Novos, three Parakanã, and one Quechua. Four of them (the Pacaás Novos, the Quechua and two Parakanã) did not confirm the initial indication of heterozygosis, and were classified as homozygotes while the Aché, Kaingang and the other Parakanã are heterozigotes. In the Aché and the Kaingang, we found some recombinant clones between the two alleles, that may be explained as an artifact of the PCR process, perhaps originated as described by Ennis *et al.* (1990).

Combining our data with those found by Alonso & Armour (2001), 55 polymorphic sites (including the 5 bp indel in two haplotypes) were observed in this segment. Of these, 12 were observed in this study for the first time. These new substitutions can be characterized as follows: nine transitions on sites 143, 735, 1535 (C/T); 264, 468, 491, 535, 722 (G/A); 1510 (T/C); and three transversions on sites 1132 (A/C); and 1141, 1251 (A/T).

The haplotypes observed by Alonso & Armour (2001) and those inferred in this study are assembled in Table 2. As is indicated there, a total of 35 haplotypes could be identified. By far the most common was HP1, occurring in 62% of the 296 chromosomes studied. HP5 has a frequency of 11%, while all the others had prevalence of 5% or less. Twelve occur in Amerindians and Northern Arctic only, and were newly identified in the present investigation. Nineteen others have been found by Alonso & Armour (2001) only,

in samples from Asian, European, and African subjects, while four were observed both by them and ourselves. HP1 occurs in the five ethnic-geographic categories established; HP13 in four, HP5 in three, and HP10 in two of them. In Amerindians, among the 180 chromosomes tested, the three most common haplotypes were (in percentages) HP1 (68), HP5 (14), and HP6 (8).

Of the 27 ethnic groups listed in Table 1, seven present all individuals homozygous for the most common allele (HP1): South Americans, Arara, Lengua, Pacaás Novos, Wai Wai; Central Americans, Huetar, Teribe; and the Khalkh Mongolians. The following haplotypes occurred in only one ethnic group: HP3 (Quechua), HP4 (Waiãpi), HP11 (Gavião), HP12 (Foz do Içana), HP14 (Cinta Larga), HP15 (Chukchi), and HP16 (Aché). Haplotype six occurs in 26 individuals belonging to the Aché, Parakanã, Suruí, and Xavante tribes. The Parakanã are the most diversified tribe, showing, besides HP6, the HP1, HP2, HP5, HP7, HP8, and HP9 haplotypes.

Median Joining Network

A Median Joining Network was constructed with the above-mentioned haplotypes, and is pictured in Figure 1. Two clusters are apparent, one with the most divergent haplotypes and the other with a starlike structure, where haplotype 1 is the central and most numerous one. This second cluster is more compact, with most haplotypes differing from HP1 by just one mutation. Included there are haplotypes present in the five geographic-ethnic categories established in the present work. Of the haplotypes found exclusively in Amerindians, the most frequent is HP6, which differs from HP1 by a mutation in site 491. Four other haplotypes, HP7, HP8, HP9, HP11, also differ from HP1 by just one mutation. The most derived Amerindian haplotype in this cluster is HP2, found in the Parakanã, which differs from HP1 in three, and from HP8 in two sites.

The second cluster originated from HP27, that was identified by Alonso & Armour (2001) as the root of their tree. Represented there are mostly exclusively African haplotypes, but HP28 was found in only one Basque individual. HP10, found by Alonso & Armour (2001) in Kenya, was also observed among the Kaingang. Four other haplotypes in this clade were found exclusively in Amerindians: HP3 in the Quechua, HP16 in one Aché and HP15 in the Chukchi. The reticulations observed derive from haplotypes confirmed by cloning.

Mutation and Evolutionary Rates

Mean divergence (K) between humans and one chimpanzee was 0.026 ± 0,00082. The average mutation rate, estimated from this divergence, was 2.16×10^{-9} per site per year. The mutation rate per sequence (1558 sites) per generation (20 years) was estimated as 6.75×10^{-5}.

Genetic Diversity

Results of haplotype and nucleotide diversity, as well as the neutrality tests, are presented in Table 3. Haplotype and nucleotide diversity indices are very similar among the major ethnic groups, with the exception of Africans, that present significantly higher values. Northern Arctic populations also present higher values, but these are probably due to one Chukchi that is homozygote for a very divergent haplotype (HP15). Our results for the world sample are similar to those found by Alonso & Armour (2001), although diversity values from our data are usually smaller, as the low diversity Amerindians comprise a high proportion of the sample. Almost all neutrality statistics (Tajima's D, Fu and Li's D* and F*, Fu's Fs) presented negative values, except those for Northern Arctic populations, that were positive, but not significant; and Fu and Li's D* for Amerindians, that was also non-significantly positive. Europeans and the world sample present significant values for all neutrality tests and Amerindians for Tajima's D and Fu's Fs. Asian estimates were significant only for Fu's Fs.

These negative numbers suggest population growth (assuming selective neutrality, see discussion). To better investigate this implication for the worldwide data, a wide variety of scenarios (see Subjects and Methods) were considered using coalescent simulation (Figure 2B). The simulations generated for Tajima's D statistic (at a two-tailed P value cutoff of 0.05) showed that we could reject for our data all scenarios in which population expansion occurred before ~21,000 years ago and also, reject all scenarios of a stationary population or of growth magnitude below 10–fold. To test if these results could have been influenced by the large sample size of the Amerindians, simulations were also run with statistics estimated exclusively with Alonso & Armour (2001) data. The results were very similar to those obtained with the whole dataset.

Two historical events were tested with an exclusively Amerindian dataset: population growth or decline (bottleneck) (Figure 2A). Simulations considering population increase rejected scenarios for all magnitude of growth inferred below ~18,000 years ago (calculated from an effective population size of 10,000, 20 years per generation) as well as any scenario of growth below 5-fold. All scenarios of population decline (>1-fold to 50-fold) were rejected.

To investigate the partition of variance within and among populations, an AMOVA analysis was performed comparing five geographic arrangements. The results are shown in Table 4. As is usually found in human populations, most of the variation was found within populations. In the first comparison, involving all the geographic-ethnic categories considered in Table 3, ~87% of the variance occurs at the intrapopulation level, ~10% among populations within these categories and ~3% among these categories. The next three other comparisons relate to the Amerindians and their putative ancestors (Asians and Northern Arctic populations). There are no clear differentiation between Central and South Amerindians (negative value), but there are variation among the different tribes or populations studied in Central and South America (~15%). The numbers in the Amerindians vs Asians comparison are similar, while the large among-groups variation in the Amerindians vs Northern Arctic contrast may be related to the small sample size of the latter, and the already mentioned Chukchi who presented an unusually high number of mutations. The last comparison (Africans and non-Africans) shows more than 10% of difference among these two groups, what was expected since Africans showed the most distinctive haplotypes.

Discussion

Both similarities and differences can be found when the present study is compared to that of Alonso & Armour (2001). In relation to haplotype frequencies, in both studies HP1 (or Hpa in the former nomenclature) was the most frequent (respectively 62% and 52%). But the second most frequent haplotype was different in the two investigations, namely, HP5 (or HPb) here and HP13 (Hpe) in the study of these authors, both occurring at a prevalence of 11%. All the others had frequencies of 7% or less in both researches. The five geographic-ethnic categories established were differentially sampled. By far the most

studied is the Amerindian, with 90 individuals. Therefore, it is not surprising that they show the largest number (11) of specific haplotypes.

Among Amerindian populations, a Brazilian tribe called Parakanã, showing four specific haplotypes and sharing three of the most numerous with the other populations studied, are quite peculiar in relation to this DNA region. They also showed unusual features in other morphological (high proportion of light skin) and genetic traits (Black *et al.* 1980, 1988). HP16 is notable because it appears in the Aché, a group that was also found to be different from other Amerindians in morphological and genetic traits (Hill & Hurtado, 1996; Battilana *et al.* 2002; Schmitt *et al.* 2004).

Similarly to Alonso & Armour (2001), the median-joining network has two clusters, one with a starlike shape that suggests a strong population expansion from a restricted source, and another with very divergent haplotypes, mainly found in Africans. However, we also found some very divergent haplotypes in this cluster that are presented in Amerindians and Northern Arctic populations. While Alonso & Armour (2001) findings suggest that the out-of-Africa exodus was followed by a very small set of closely related haplotypes, our expanded sampling support a more diverse founding population. As these divergent haplotypes are found in very low frequency, they would certainly be missing if our sample size were small and/or ethnic diversity low. These results recommend a cautionary note on conclusions about genetic diversity in humans supported by samples from very limited sources.

Considering the mutation and evolutionary rates estimated for all populations together, it differed little from those calculated by Alonso & Armour (2001). Divergence (K) between humans and chimpanzee as calculated here and by the indicated authors were very similar (respectively 0.029 and 0.026), the same being true for the mutation rates (respectively 2.18×10^{-9} and 2.19×10^{-9}). Nucleotide diversities were recalculated to include the same numbers of base pairs. As expected, Africans showed the highest value (0.243%), Amerindians presenting a number that was of the same order of magnitude (0.064%) than Europeans (0.050%) and Asians (0.048%).

When the data were hierarchically analyzed, the Central and South Amerindians showed no significant difference. In Central America only the two most common haplotypes were found (HP5 and HP1), and this low genetic variability may have occurred by isolation followed by a population reduction (Barrantes *et al.* 1990; Azofeifa *et al.*

2001). The difference observed among populations within groups (~15%) is probably due to a genetic structure in South America, althought it can not be explained by geography, language or history.

The statistical neutrality tests for ours and Alonso & Armour (2001) worldwide datasets resulted in negative values, corroborating the signal of a past population growth. Our simulations with the worldwide sample support a minimum 10-fold population growth at least 21,000 years ago for humans, in special in non-African samples (Alonso & Armour 2001), although the ancestral population may not have been as small as thought before. The simulations with the exclusively Amerindian dataset suggested similar demographic scenarios, although usually more moderate than those from the worldwide sample (population growth at least 18,000 years ago, Fig. 2A). Interestingly, the simulations also rejected all scenarios of a population decline (bottleneck) in the Amerindian sample. Therefore, the hypothesis of a strong bottleneck on the entrance of prehistoric humans in the Americas is not supported by the locus 16p13.3, which on the contrary indicates that the genetic variability of the source population was maintained in Amerindians. The signal of population growth observed in the Amerindians is probably the signal of the out-of-Africa exodus that was not erased during and after the demographic events that led to the colonization of the New World.

The history of population decline associated with the modern human arrival in Americas is controversial. A recent study in a noncoding region with Amerindian and Asian populations (Heller *et al.,* 2005) did not find bottleneck evidence in Amerindian populations in the entrance of Americas, the same being true for the investigation of Hutz *et al*. (2002) in a study about five Brazilian Indian populations and fifteen short tandem repeat polymorphisms (STRPs). On the other hand, Fagundes *et al.* (2005) found a signal of population bottleneck in Amerindians studying the low density lipoprotein receptor gene (*LDLR*) of 222 chromosomes of individuals from African, Asian, Caucasian, and Amerind ancestry. Their results agreed with polymorphisms in the Y chromosome (Bortolini *et al.* 2003) and mtDNA (Bonatto & Salzano, 1997). Nevertheless, mtDNA and NRY (non-recombined region of the Y chromosome) are uniparentally inherited single loci that are subject to the stochastic processes of genetic drift, as well as to natural selection (including genetic hitchhiking), which can alter the shape of the haplotype phylogenies, patterns of variation, and estimates of TMRCA (the time of the most recent common ancestor)

(Mishmar *et al.* 2003; Excoffier, 1990). More importantly for this question, both mtDNA and NRY have effective population sizes one fourth of the autosomal locus considered here, and are therefore more strongly affected by population size reduction than the later. Therefore, scenarios of bottlenecks of moderate intensity may be compatible with these apparently conflicting results.

Acknowledgments

This work was supported by grants from Programa de Apoio a Núcleos de Excelência (PRONEX), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS). We thank Kim Hill, and Madaglena Hurtado for assistance in the Amerindian field collections. The latter have been approved by the Brazilian Ethics National Committee (CONEP, resolution 123/98). We would like to thank Prof. Dr. Jomar Laurino for his assistance and support on cloning process; Nelson J. R. Fagundes and Felipe Gobbi Grazziotin for their contribution to some of the results, and M. L. Petzl-Erler, L. T. Tsuneto, Eduardo Tarazona-Santos, Mikhail Krylov, L. Benevolenskaya, B. Munkhbat for providing part of DNA samples analyzed in the present study.

References

Alonso, S. & Armour, J. A. (2001) A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outstide Africa. *Proc Natl Acad Sci U S A* **98,** 864-869.

Andrade, F. M., Coimbra, C. E. A., Jr., Santos, R. V., Goicoechea, A., Carnese, F. R. & Salzano, F. M. *et al.* (2000) High heterogeneity of apolipoprotein E gene frequencies in South American Indians. *Ann Hum Biol* **27,** 29-34.

Azofeifa, J., Ruiz, E. & Barrantes, R. (2001) Blood group, red cell, and serum protein variation in the Cabecar and Huetar, two Chibchan Amerindian tribes of Costa Rica. *Am J Hum Biol* **13,** 57-64.

Badge, R. M., Yardley, J., Jeffreys, A. J. & Armour, J. A. (2000) Crossover breakpoint mapping identifies a subtelomeric hotspot for male meiotic recombination. *Hum Mol Genet* **9,** 1239-1244.

Barrantes, R. (1993) *Evolución en el trópico: los Amerindios de Costa Rica e Panamá.* San José: Editorial de la Universidad de Costa Rica.

Barrantes, R., Smouse, P. E., Mohrenweiser, H. W., Gershowitz, H., Azofeifa, J., Arias, T. D., *et al.* (1990) Microevolution in Lower Central America - genetic-characterization of the Chibcha-speaking groups of Costa-Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *Am J Hum Genet* **46,** 63-84.

Battilana, J., Bonatto, S. L., Freitas, L. B., Hutz, M. H., Weimer, T. A. & Callegari-Jacques, S. M. *et al.* (2002) *Alu* insertions versus blood group plus protein genetic variability in four Amerindian populations. *Ann Hum Biol* **29,** 334-347.

Black, F. L., Salzano, F. M., Layrisse, Z., Franco, M. H., Harris, N. S. & Weimer, T. A. (1980) Restriction and persistence of polymorphisms of HLA and other blood genetic traits in the Parakanã Indians of Brazil. *Am J Phys Anthropol* **52,** 119-132.

Black, F. L., Santos, S. E., Salzano, F. M., Callegari-Jacques, S. M., Weimer, T. A. & Franco, M. H. *et al.* (1988) Genetic variation within the Tupi linguistic group: new data on three Amazonian tribes. *Ann Hum Biol* **15,** 337-351.

Bonatto, S. L. & Salzano, F. M. (1997) A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci U S A* **94,** 1866-1871.

Bortolini, M. C., Salzano, F. M., Thomas, M. G., Stuart, S., Nasanen, S. P. & Bau, C. H. *et al.* (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* **73,** 524-539.

Brown, S. M., Gajdusek, D. C., Leyshon, W. C., Steinberg, A. G., Brown, K. S. & Curtain, C. C. (1974) Genetic studies in Paraguay: blood group, red cell, and serum genetic patterns of the Guayaki and Ayore Indians, Mennonite settlers, and seven other Indian tribes of the Paraguayan Chaco. *Am J Phys Anthropol* **41,** 317-343.

Callegari-Jacques, S. M., Salzano, F. M., Weimer, T. A., Franco, M. H. L. P., Mestriner, M. A. & Hutz, M. H. *et al.* (1996) The Wai Wai Indians of South America: History and genetics. *Ann Hum Biol* **23,** 189-201.

Callegari-Jacques, S. M., Salzano, F. M., Weimer, T. A., Hutz, M. H., Black, F. L. & Santos, S. E. *et al.* (1994) Further blood genetic studies on Amazonian diversity-data from four Indian groups. *Ann Hum Biol* **21,** 465-481.

Campbell, L. (2000) *American Indian Languages*. New York: Oxford University Press.

Coimbra, C. E. A. Jr, Flowers, N. M., Salzano, F. M. & Santos, R. V. (2002) *The Xavánte in transition : health, ecology, and bioanthropology in central Brazil.* Ann Arbor: University of Michigan Press.

Dornelles, C. L., Battilana, J., Fagundes, N. J., Freitas, L. B., Bonatto, S. L. & Salzano, F. M. (2004) Mitochondrial DNA and *Alu* insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. *Am J Hum Biol* **16,** 479-488.

Ennis, P. D., Zemmour, J., Salter, R. D. & Parham, P. (1990) Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. *Proc Natl Acad Sci USA* **87,** 2833-2837.

Erdesz, S., Shubin, S. V., Shoch, B. P., Krylov, M., Mylov, N. M. & Chekalina, N. A. *et al.* (1994) Spondyloarthropathies in circumpolar populations of Chukotka (Eskimos and Chukchi): epidemiology and clinical characteristics. *J Rheumatol* **21,** 1101-1104.

Excoffier, L. (1990) Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. *J Mol Evol* **30**, 125-139.

Excoffier, L. (2002) Human demographic history: refining the recent African origin model. *Curr Opin Genet Dev* **12,** 675-682.

Fagan, B. M. (2004) *The great journey : the peopling of ancient America.* Florida: University Press of Florida.

Fagundes, N. J., Salzano, F. M., Batzer, M. A., Deininger, P. L. & Bonatto, S. L. (2005) Worldwide genetic variation at the 3'-UTR region of the *LDLR* gene: possible influence of natural selection. *Ann Hum Genet* **(in press)**.

Fu, Y. X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147,** 915-925.

Fu, Y-X. & Li, W-H. (1993) Statistical tests of neutrality of mutations. *Genetics* **133,** 693-709.

Gilad, Y., Rosenberg, S., Przeworski, M., Lancet, D. & Skorecki, K. (2002) Evidence for positive selection and population structure at the human MAO-A gene. *Proc Nat Acad Sci USA* **99,** 862-867.

Goicoechea, A. S., Carnese, F. R., Dejean, C., Avena, S. A., Weimer, T. A. & Estalote, A. C. *et al.* (2001) New genetic data on Amerindians from the Paraguayan Chaco. *Am J Hum Biol* **13,** 660-667.

Hammer, M. F. (1995) A recent common ancestry for human Y-chromosomes. *Nature* **378,** 376-378.

Harding, R. M., Fullerton, S. M., Griffiths, R. C., Bond, J., Cox, M. J. & Schneider, J. A, *et al.* (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet* **60,** 772-789.

Harris, E. E. & Hey, J. (1999) X chromosome evidence for ancient human histories. *Proc Natl Acad Sci USA* **96,** 3320-3324.

Heidrich, E. M., Hutz, M. H., Salzano, F. M., Coimbra, C. E. A. Jr & Santos, R. V. (1995) D1S80 locus variability in three Brazilian ethnic-groups. *Hum Biol* **67,** 311-319.

Heller, A. H., Salzano, F. M., Barrantes, R., Krylov, M., Benevolenskaia, L. & Arnett, F. C. *et al.* (2005) Intra and intercontinental molecular variability in the 3'-UTR region of the *LDLR* gene. *Hum Biol* **(in press)**.

Hill, K. & Hurtado, A. M. (1996) *Ache life history : the ecology and demography of a foraging people*. New York: Aldine de Gruyter.

Hutz, M.H., Callegari-Jacques, S.M., Almeida, S.E., Armborst, T. & Salzano FM (2002) Low levels of STRP variability are not universal in American Indians. *Hum Biol* **74**, 791-806.

Hutz, M. H., Mattevi, V. S., Callegari-Jacques, S. M., Salzano, F. M., Coimbra, C. E. A. Jr & Santos, R. V. *et al.* (1997) D1S80 locus variability in South American Indians. *Ann Hum Biol* **24,** 249-255.

Jaruzelska, J., Zietkiewicz, E., Batzer, M., Cole, D. E., Moisan, J. P. & Scozzari, R. *et al.* (1999) Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. *Genetics* **152,** 1091-1101.

Kaessmann, H., Heissig, F., von Haeseler, A. & Paabo, S. (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat Genet* **22,** 78-81.

Kolman, C. J., Sambuughin, N. & Bermingham, E. (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* **142,** 1321-1334.

Krylov, M., Erdesz, S., Alexeeva, L., Benevolenskaya, L., Arnett, F. C. & Reveille, J. D. (1995) HLA class II and HLA-B27 oligotyping in two Siberian native population groups. *Tiss Antig* **46,** 382-386.

Mestriner, M. A. & Salzano, F. M. (1998) Monomorphic and polymorphic enzyme genetic markers of the Waiãpi Indians of Amapá and of inhabitants of Manaus, Amazonas. *Gen Mol Biol* **21,** 311-314.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G. & Hosseini, S., *et al.* (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100, 171-176.

Munkhbat, B., Sato, T., Hagihara, M., Sato, K., Kimura, A. & Munkhtuvshin, N. *et al.* (1997) Molecular analysis of HLA polymorphism in Khoton-Mongolians. *Tiss Antig* **50,** 124-134.

Nachman, M. W., Bauer, V. L., Crowell, S. L. & Aquadro, C. F. (1998) DNA variability and recombination rates at X-linked loci in humans. *Genetics* **150,** 1133-1141.

Nachman, M. W. & Crowell, S. L. (2000) Contrasting evolutionary histories of two introns of the Duchenne muscular dystrophy gene, Dmd, in humans. *Genetics* **155,** 1855-1864.

Nei, M. (1987) *Molecular evolutionary genetics.* New York: Columbia University Press.

Posada, D. & Crandall, K. A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14,** 817-818.

Przeworski, M., Hudson, R. R. & Di Rienzo, A. (2000) Adjusting the focus on human variation. *Trends Genet* **16,** 296-302.

Rogers, A. R. (1995) Genetic evidence for a Pleistocene population explosion. *Evolution* **49,** 608-615.

Salzano, F. M. (2002) Molecular variability in Amerindians: widespread but uneven information. *An Acad Bras Cienc* **74,** 223-263.

Salzano, F. M., Black, F. L., Callegari-Jacques, S. M., Santos, S. E., Weimer, T. A. & Mestriner, M. A, *et al.* (1991) Blood genetic systems in four Amazonian tribes. *Am J Phys Anthropol* **85,** 51-60.

Salzano, F. M., Callegari-Jacques, S. M., Franco, M. H., Hutz, M. H., Weimer, T. A. & Silva, R. S. *et al.* (1980) The Caingang revisited: blood genetics and anthropometry. *Am J Phys Anthropol* **53,** 513-524.

Salzano, F. M., Callegari-Jacques, S. M., Weimer, T. A., Franco, M. H. L. P., Hutz, M. H. & PetzlErler, M. L. (1997) Electrophoretic protein polymorphisms in Kaingang and Guarani Indians of southern Brazil. *Am J Hum Biol* **9,** 505-512.

Salzano, F. M., Gershowitz, H., Mohrenweiser, H., Neel, J. V., Smouse, P. E. & Mestriner, M. A., *et al.* (1986) Gene flow across tribal barriers and its effect among the Amazonian Içana river Indians. *Am J Phys Anthropol* **69,** 3-14.

Salzano, F. M., Weimer, T. A., Franco, M. H. L. P., Mestriner, M. A., Simões, A. L. & Constans, J., *et al.* (1985) Population-structure and blood genetics of the Pacaás Novos Indians of Brazil. *Ann Hum Biol* **12,** 241-249.

Sambrook, J. & Russell, D. W. (2001) *Molecular cloning : a laboratory manual.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Saunders, M. A., Hammer, M. F. & Nachman, M. W. (2002) Nucleotide variability at G6pd and the signature of malarial selection in humans. *Genetics* **162,** 1849-1861.

Schmitt, R., Bonatto, S. L., Freitas, L. B., Muschner, V. C., Hill, K. & Hurtado, A. M. *et al.* (2004) Extremely limited mitochondrial DNA variability among the Aché natives of Paraguay. *Ann Hum Biol* **31,** 87-94.

Schneider, S., Roessli, D. & Excoffier, L. (2000) *Arlequin ver. 2.000: a software for population genetics.* Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585-595.

Tarazona-Santos, E., Carvalho-Silva, D. R., Pettener, D., Luiselli, D., De Stefano, G. F. & Labarga, C. M. *et al.* (2001) Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* **68,** 1485-1496.

Underhill, P. A., Shen, P. D., Lin, A. A., Jin, L., Passarino, G. & Yang, W. H, *et al.* (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* **26,** 358-361.

Verrelli, B. C., McDonald, J. H., Argyropoulos, G., Destro-Bisol, G., Froment, A. & Drousiotou, A. *et al.* (2002) Evidence for balancing selection from nucleotide sequence analyses of human G6PD. *Am J Hum Genet* **71,** 1112-1128.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. (1991) African populations and the evolution of human mitochondrial DNA. *Science* **253,** 1503-1507.

Watterson, G. A. (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7,** 256-276.

Whitfield, L. S., Sulston, J. E. & Goodfellow, P. N. (1995) Sequence variation of the human Y-chromosome. *Nature* **378,** 379-380.

Wooding, S., Kim, U. K., Bamshad, M. J., Larsen, J., Jorde, L. B. & Drayna, D. (2004) Natural selection and molecular evolution in *PTC*, a bitter-taste receptor gene. *Am J Hum Genet* **74,** 637-646.

Yu, N., Chen, F. C., Ota, S., Jorde, L. B., Pamilo, P. & Patthy, L. *et al.* (2002) Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161,** 269-74.

Table 1 Sample size, geographic location, and linguistic information about the studied populations

Population	No. of individuals	Geographic location	Linguistic family[a]	References for further information
South Americans				
Aché	5	23°30'-24°10'S; 55°50'-56°30'W	Guayaki	Brown *et al.* (1974); Hill & Hurtado (1996)
Ayoreo	5	16-22°S; 58-63°W	Ayore	Dornelles *et al.* (2004)
Arara	3	3°30’S; 53°W	Arara	Salzano *et al.* (1991)
Cinta Larga	4	9°50’-12°30’S; 59°10’-60°50’W	Tupi Mondé	Callegari-Jacques *et al.* (1994)
Foz do Içana	3	1°N; 67°30’W	Cubeo, Tucano, Baniwa, Tariana	Salzano *et al.* (1986)
Gavião	4	10°10’S; 61°08’W	Tupi Mondé	Hutz *et al.* (1997); Andrade *et al.* (2000)
Guarani	3	23°6'S, 55°12'W; 23°12'S, 55°6'W 23°48'S, 54°30'W	Guarani	Salzano *et al.* (1997)
Kaingang	4	27º20’S; 52º45’W	Ge-Kaingan	Salzano *et al.* (1980)
Lengua	3	23ºS; 56ºW	Macro-Panoan	Brown *et al.* (1974); Goicoechea *et al.* (2001)
Pacaás Novos	7	11°8’S; 65°5’ W	Chapacura	Salzano *et al.* (1985)
Parakanã	14	5°S, 50°10’W; 4°30’S, 50°W 5°55’S; 52°42’W	Tupi	Black *et al.* (1988)
Quechua	5	12°33’S, 75°83’W; 16°38’S, 71°52’W	Quechua	Tarazona-Santos *et al.* (2001)
Suruí	4	10°50’S; 61°10’W	Tupi Mondé	Hutz *et al.* (1997); Andrade *et al.* (2000)
Waiãpi	3	1°N; 53°W	Oyampi	Mestriner & Salzano (1998)
Wai Wai	3	0°40’S; 58°W	Carib	Callegari-Jacques *et al.* (1996)
Xavante	4	13°20'S; 51°40'W	Gê	Coimbra *et al.* (2002)
Zoró	4	10°20’S; 60°20’W	Tupi Mondé	Heidrich *et al.* (1995); Andrade *et al.* (2000)

Central Americans				
Bribri	2	9º38’N; 82º50’W	Chibchan	Barrantes (1993)
Cabecar	2	9º26’N; 83º09’W	Chibchan	Barrantes (1993)
Guatuso	2	10º40’N; 84º49’W	Chibchan	Barrantes (1993)
Guaymi	2	8º13’N; 82º57’W	Chibchan	Barrantes (1993)
Huetar	2	9º53’N; 84º14’W	Chibchan	Barrantes (1993)
Teribe	2	9º20’N; 82º35’W	Chibchan	Barrantes (1993)
Northern Arctic				
Chukchi	2	64ºN; 175ºW	Chukot	Erdesz *et al.* (1994); Krylov *et al.* (1995)
Eskimo	2	64ºN; 175ºW	Yupic	Erdesz *et al.* (1994); Krylov *et al.* (1995)
Mongolians				
Khalkh	2	46ºN; 106ºE	Altaic	Kolman *et al.* (1996)
Khoton	2	45ºN; 94ºE	Qotong	Munkhbat *et al.* (1997)

[a] according to Campbell (2000)

Table 2 16p haplotypes observed in the present investigation and by Alonso & Armour (2001) showing their geographical range[a]

Haplo-types[b]	Sites 671112222233333444455567777778881111111111111111111 942472367723567366938991233558015001111111223455555 731194491463033815180425667890578344447856720012344212359912232601 5	N° found	Geographical range[c]
HP1/a	GGGCA+CGCCGGGCCGGGGGACCCGCCCGACGGGGAATAAGCATAAGGTGC	184	1(122)2(4)3(17)4(28)5(13)
HP2	...T.....A.....T.....	2	1(2)
HP3	.....CA.T.....A....A.....AT...G.....	1	1(1)
HP4	..C.....A.....	1	1(1)
HP5/b	..C.....	33	1(25)2(1)3(4)4(3)
HP6	.....A.....	14	1(14)
HP7	.....T.....	2	1(2)
HP8	.....A.....	2	1(2)
HP9	.....T	5	1(5)
HP10/n	....G.....G.....A.....A....G.....	3	1(1)5(2)
HP11	.....C.....	2	1(2)
HP12	.....A.....C..	2	1(2)
HP13/e	.....T.....	12	2(1)3(2)4(1)5(8)
HP14	....G.....G....T.....A.....A....G.....	2	1(2)
HP15	.....CA.T.....AG.....A....A..T...AT...G.....	2	2(2)
HP16	.....C..T.....G.....A....A..T...AT...G.....	1	1(1)
HP17/c	.....C.....	3	4(3)
HP18/j	.....T.....	1	5(1)
HP19/h	.....T.....	1	3(1)
HP20/g	.....C.....	1	3(1)
HP21/p	.....A.	1	4(1)
HP22/d	.....C.....	2	4(1)5(1)
HP23/l	.....T...A.....	2	5(2)
HP24/r	.....T.....C.....	1	4(1)
HP25/o	.A.....A.....	1	4(1)
HP26/i	..C.....A.....	1	3(1)
HP27/k	.....G.....G.....	2	5(2)
HP28/f	.....T.....G.....G.....	1	4(1)
HP29/w	.....T...A.....A...	1	5(1)
HP30/u	.....A.....T...A.....	1	5(1)
HP31/s	A.....T.....GT.....A.....GA....	3	5(3)
HP32/x	.....T.....G.T.....A...ACG....G.....	1	5(1)
HP33/t	.....CA.T.....G.....A....A.....AT...G.....	1	5(1)
HP34/y	.....-...T.....G.....A.....G.....	3	5(3)
HP35/z	.....-..GT.....G.....A.....G.....	1	5(1)

[a] Dots indicate agreement and dashes deletions in relation to the most common haplotype (HP1).

[b] Haplotypes inferred in the present study and their equivalents in Alonso & Armour (2001). Those in boldface were found in this study for the first time.

[c] 1. Amerindians; 2. Northern Arctic; 3. Asians (Mongolia and Japan); 4. Europeans; 5. Africans; number of haplotypes found in these groups indicated in parentheses.

Table 3 16p haplotype and nucleotide diversity, as well as neutrality test results.

Characteristics	Amerindians		Northern Arctic	Asian	Europeans	Africans	World	World A & A[b]
	Central	South						
No. of chromosomes	24	156	8	28	40	40	296	100
No. of haplotypes	14		4	6	9	14	35	28
No. of polymorphic sites	23		13	5	11	28	49	42
Haplotype diversity (Hd)	0.519		0.750	0.598	0.508	0.853	0.598	0.765
Nucleotide diversity (π) (SE)[a] %	0.064 (0.011)		0.335 (0.116)	0.048 (0.010)	0.050 (0.013)	0.243 (0.037)	0.098 (0.012)	0.147 (0.020)
Tajima's D	-2.08*		0.20	-1.18	-2.12*	-1.46	-2.32**	-2.14*
Fu and Li's F*	-0.66		0.90	-1.58	-3.55*	-2.07	-2.88*	-3.76*
Fu and Li's D*	0.46		1.00	-1.44	-3.45*	-1.90	-2.46*	-3.80*
Fu's Fs	-7.60**		2.30	-2.68*	-5.97**	-2.72	-32.70**	-18.42**

[a]SE, standard error, *: $p<0.05$; **: $p<0.01$ [b]Alonso & Armour (2001)

Table 4 AMOVA results: intra and interpopulation variability considering different hierarchical arrangements

Population Structure	Among Groups (%)	Among Populations Within Groups (%)	Within Populations (%)
Among Geographic-Ethnic Categories[a]	3.32[ns]	9.70	86.98
Central vs South Amerindians	-0.34[ns]	15.12	85.22
Amerindians vs Asian	-3.33[ns]	14.00	89.33
Amerindians vs Northern Arctic	31.52	6.45	62.03
Africans vs non-Africans[b]	11.44	8.05	80.51

[a]Groups: Africans, Amerindians (South and Central), Asians, Europeans and Northern Arctic.
[b]non-Africans: Amerindians, Asians, Europeans and Northern Arctic.
[ns]: $p>0.1$; all the other values are significant at the 0.1% or lower level.

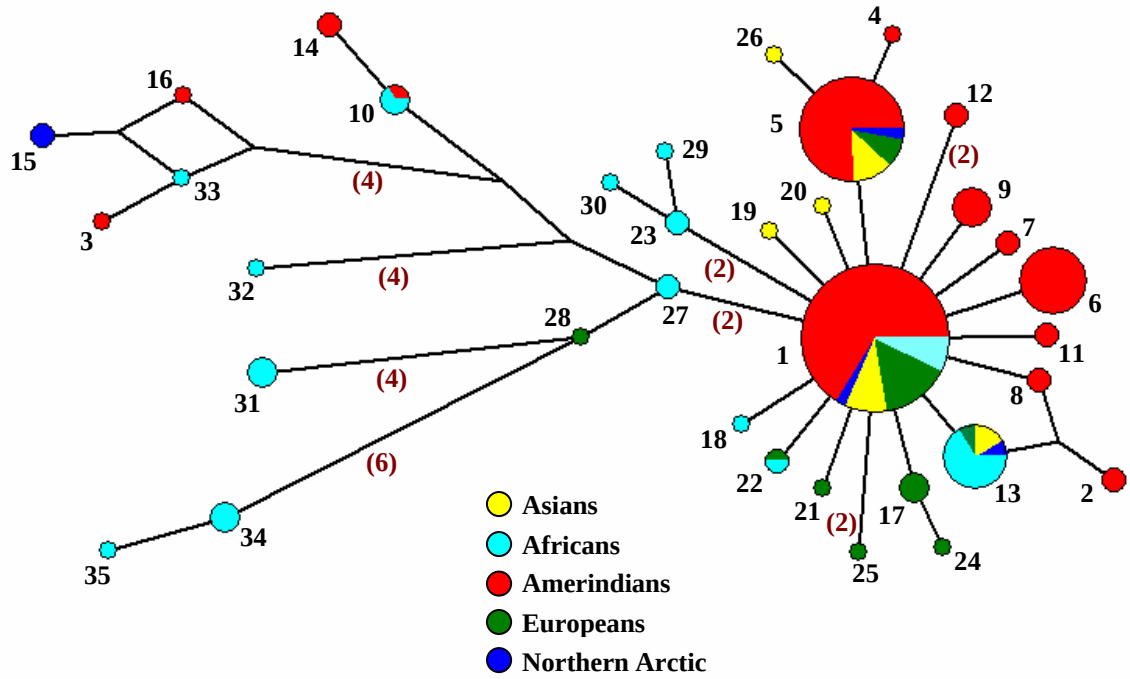

Figure 1 Median-joining network showing the relationships among haplotypes. Each circle represents a different haplotype, and its size is proportional to their relative frequencies. For each haplotype, the different colors indicate the fraction of observations in Africans, Amerindians, Asians, Europeans and Northern Arctic. The branches correspond to nucleotides substitutions and the number beside each circle is the haplotype assignment. Reticulations indicate ambiguous relationships. HP27 is the root of the tree.

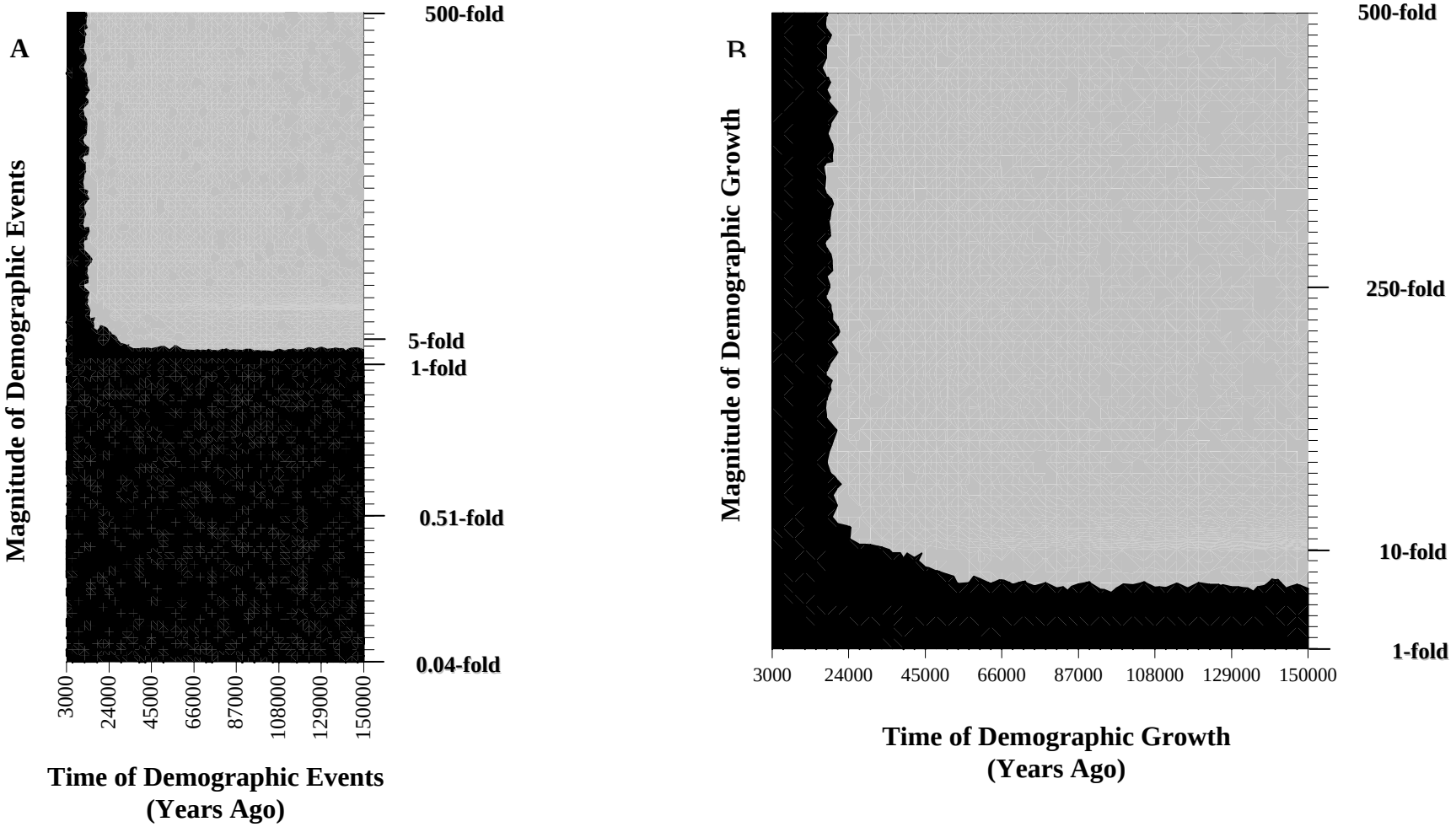

Figure 2 Simulation tests considering growth and bottleneck in Amerindians (A) and growth in world populations (B). Black shading indicate rejected scenarios (P<0.025). Light shading indicates scenarios that could not be rejected.

CAPÍTULO III

Alu Insertion Polymorphisms in Native Americans and Related Asian Populations

Manuscrito a ser submetido ao Annals of Human Biology

Alu Insertion Polymorphisms in Native Americans and Related Asian Populations

JAQUELINE BATTILANA,[1,2] NELSON J. R. FAGUNDES,[1,2] ANA H. HELLER,[1] ANGELA GOLDANI,[2] LORETA B. FREITAS,[1] EDUARDO TARAZONA-SANTOS,[3] BATMUNKH MUNKHBAT,[4,5] NAMID MUNKHTUVSHIN,[4] KIMIYOSHI TSUJI,[5] MLU KRYLOV,[6] LIDIA BENEVOLENSKAIA,[6] FRANK ARNETT,[7] MARK A. BATZER,[8] PRESCOTT L. DEININGER,[9] FRANCISCO M. SALZANO,[1] AND SANDRO L. BONATTO[2]

Abbreviated title: *ALU POLYMORPHISMS IN AMERINDIANS*

[1]*Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil*
[2]*Centro de Biologia Genômica e Molecular, Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, 90619-900 Porto Alegre, RS, Brazil*
[3]*Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, MG, Brazil*
[4] *National Institute of Health, National Medical University of Mongolia, Ulaanbaatar, Mongolia*
[5]*School of Medicine, Tokay University, Japan*
[6]*Department of Epidemiology and Genetics of Rheumatic Disease, Russian Academy of Science, Moscow*
[7]*Health Science Center, University of Texas, Houston, Texa, USA*
[8]*Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA*
[9]*Tulane Cancer Center, Tulane University Health Center, New Orleans, LA 70112, USA*

Contract grant sponsors: Programa de Apoio a Núcleos de Excelência (PRONEX); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS); Pró-Reitoria de Pesquisa da Universidade Federal do Rio Grande do Sul (PROPESQ-UFRGS).

Correspondence to: Sandro L. Bonatto, Centro de Biologia Genômica e Molecular, Faculdade de Biociências, PUCRS, 90619-900 Porto Alegre, RS, Brazil. E-mail: slbonatto@pucrs.br

ABSTRACT: Twelve *Alu* insertion polymorphisms were studied in 179 individuals belonging to 7 South American Native, 60 to two Siberian, and 91 to two Mongolian populations. These data were integrated with those from 488 persons affiliated with 13 other groups, to ascertain the relationships between Asian, Northern Arctic, and Amerindian populations. A decreasing trend concerning heterozygosities and amount of gene flow was observed in the three sets, in the order indicated above. Genetic distances, principal components analysis, analysis of molecular variance, Mantel test, and a Bayesian approach to genetic assignment, all indicated the validity of these subdivisions. However, no clear structure could be observed within South American Natives, indicating the importance of dispersive (genetic drift, founder effects) factors in their differentiation.

Key Words: *Alu* polymorphisms, Amerindians, Asians, Siberians.

The relationships among Native Americans, Northern Arctic and Northeast Asian populations have been extensively studied by different authors considering diverse genetic markers. Data gathered from the Y chromosome (Santos et al., 1999) and mtDNA (Bonatto and Salzano, 1997) have furnished indications that Native American haplotypes derive from a subsample of the genetic diversity present in Siberia, and suggested that the Northern Arctic populations underwent a second bottleneck by the Last Glacial Maximum (LGM) (review in Schurr, 2004). However, data from autosomal markers did not found evidences for a bottleneck in Amerindian populations (Hutz et al., 2002). It is not established, also, the amount of genetic diversity present in Native American ancestral population (s).

Alu insertions, the most abundant and extensively studied mammalian class of SINEs (short interspersed elements) of repetitive sequences represent about 11% of human genome. With ~1 million copies, *Alu* insertions are found on average once every 4 kb interval of the human genome (International Human Genome Sequencing Consortium, 2001). New copies of these elements can be inserted in the genome by retrotransposition (review in Weiner et al., 1986). Due to changes during the evolution of the source genes, there are at least 12 major *Alu* subfamilies, which may be classified as Old, Intermediate or Young. The two oldest subfamilies, Jo and Jb, arose from independent retroposition events involving a single, ancestral source gene that existed approximately 81 million years ago (Kapitonov and Jurka, 1996). The Young family contains approximately 100,000 members, and can be further subdivided into several subfamilies. Many of these inserts reflect recent retroposition events that have not yet been fixed within the human species (Batzer and Deininger, 1991; Batzer et al., 1994, 1995; Novick et al., 1995, 1998; Arcot et al., 1997; Stoneking et al., 1997; Xing et al., 2003; Carter et al., 2004; Otieno et al., 2004). These polymorphic *Alu* elements may serve as markers to investigate various aspects of human evolution. There are many advantages in their use for this purpose. It is highly improbable that a specific *Alu* sequence has retroposed into a particular site more than once in human evolution; and recent insertions represent stable polymorphisms that are rarely lost without leaving a trace. Due to these properties, identity by descent and evidential stability due to incomplete deletion, *Alu* insertions are literally molecular fossils. Another attractive feature of *Alu*-based population studies is that the ancestral condition is defined by the lack of an *Alu* element (Batzer and Deininger, 2002; Jurka *et al.*, 2004). Knowledge

of the original character state is useful in rooting phylogenies. Additionally, they are easily determined by a rapid PCR-based assay.

Previous studies using *Alu* insertions in Native American populations reinforced its affinities to Asian populations (Mateus-Pereira et al., 2005), and suggested the occurrence of multiple migrations from a single, Asian, source population during the peopling of the Americas (Novick et al., 1998). However, these studies used limited (seven and five insertion loci, respectively), and therefore have to be checked with a larger members of markers.

The present study furnishes new information on twelve *Alu* insertions for seven South Amerindian, two Siberian, and two Mongolian groups. They have been integrated with 13 other American and Asian samples, trying to answer the following questions: (a) do the population relationships found agree with the ethnic, historical, and geographical data? and (b) what can heterozygote levels and associated results inform us about the events that led to the colonization of the New World?

SUBJECTS AND METHODS

Subjects

Genotypes for a total of 818 individuals from 24 populations were analyzed. The total sample is formed by: twelve South American (Aché, Cinta Larga, Gavião, Guarani, Kaingang, South-Andean Quechua [Quechua A], Central-Andean Quechua [Quechua QT], Suruí, Wai Wai, Xavante, Yanomami, Zoró), one Central American (Maya), one North American (Mvskoke), three Northeast Asians (China, Khalkh, Khoton), and four Northern Arctic (Chukchi, Siberian Eskimo, Alaskan Native, Geenlander) populations (Fig. 1). Two East Indonesian (Molucca and Nusa-Tengarra) and one population from Oceania (Papua New Guinea Coastal) were used as an outgroup. Data for the Cinta Larga, Gavião, Quechua A, Quechua QT, Suruí, Wai Wai, Zoró, Khalkh, Khoton, Chukchi and Siberian Eskimo populations are being reported here for the first time. The Aché, Guarani, Kaingang and Xavante populations had been previously studied for these markers by Battilana et al. (2002) and data for the remainder populations have been obtained by M.A.B.'s research group. All data used in this paper is available with the authors upon

request.

DNA amplification and Alu typing

DNA was extracted following standard procedures (Lahiri and Nurnberger, 1991; Miller et al., 1988). Polymerase Chain Reaction (PCR) amplification was carried out following Batzer and Deininger (1991) using a final volume of 25μl. Each sample was subjected to the following amplification procedures: one minute at 94°C, two minutes at the appropriate annealing temperature, two minutes at 72°C, plus 5 minutes at 72°C for 40 cycles. Fifteen microliters of the PCR products were run in 2% 1x TEB agarose gels containing ethidium bromide, and the reaction products were directly visualized using ultra violet fluorescence. The loci studied were *APO* (Karathanasis, 1985), *ACE* (Tiret et al., 1992), TPA25 (Batzer and Deininger, 1991), PV92 (Batzer et al., 1994), *FXIII* and *MABDI* (Batzer et al., 1996), *A25* (Arcot et al., 1995), and *HS3.23, HS4.32, HS4.59, HS4.65, HS4.75* (Arcot et al., 1996). *Alu* insertion frequencies for each population were determined by gene counting.

Data analysis

Since several of the methods we used require individual genotypic data, only those populations for which this kind of information was available were considered. Individuals for whom data were missing for five or more loci were also excluded from the analysis. To verify if the observed genotype frequencies agreed with those expected under Hardy-Weinberg equilibrium a chi-square test for goodness of fit using Levene's correction for small sample sizes and Bonferroni's correction for multiple comparisons (Brown and Russell, 1997) was performed in the BIOSYS-2 program (Black, 1996). To infer the historical relationships among populations, an unrooted neighbor-joining (NJ) tree (Saitou and Nei, 1987) was constructed from the D_A distance matrix (Nei et al., 1983). Branch support for the population tree was assessed by 1,000 bootstrap replications using the DISPAN program (Ota, 1993). Principal components analysis (PCA, Sneath and Sokal, 1973) based on the gene frequency data was done employing the NTSYSpc program (Rohlf, 1998). The AMOVA analysis was computed in the ARLEQUIN program (Excoffier et al., 1992; Schneider et al., 2000), the statistical significance of the values being computed with 2,000 replications. Additionally, we used the Bayesian framework

implemented in the STRUCTURE program (Pritchard et al., 2000). This approach probabilistically assigns each individual into one of the *K* user-defined populations. The "non-admixture" model, in which each individual owes its whole genome from a single *K* population, was used. Although this is a very simplistic model, it has a better performance when the genetic structure is not marked (Pritchard et al., 2000). The posterior probabilities were calculated considering 1,000,000 steps, after a burn in of 100,000 steps. For each *K*, five independent runs were performed, to check the consistency of the results. Given that the structure shown by these markers is not high, the populations were clustered using the average probabilities of its individuals belonging to each *K* group using UPGMA, in the DISPAN program (Ota, 1993).

For every population, the linear relationship between heterozygosity and the distance from the centroid (defined as the arithmetic mean of the allele frequencies) was plotted as described by Harpending and Ward (1982). The theoretical expectation is that populations that have experienced more gene flow than average will fall above the theoretical prediction (regression line), while populations that have had less gene flow will fall below that prediction. Correlation between the genetic and geographical distance matrices and their associated significance levels were obtained through the use of Mantel's test (1,000 permutations) (Mantel, 1967; Smouse et al., 1986). The geographical distance matrix was computed using the geographical coordinates of each sample site (or average values when multiple sample sites were available) and the great-circle distance. Additionally, the distance between populations falling on opposite sides of the Pacific Ocean was calculated in two-steps, from one population to the Bering Strait and summed with the distance from this point to the second population. The D_A genetic distance matrix of Nei (1987) was calculated using the DISPAN package (Ota, 1993).

RESULTS

Genotype distributions and probability values concerning Hardy-Weinberg equilibrium expectations, as well as allele frequencies and heterozygosity levels for the twelve loci in the twenty-four populations are reported in Table 1. Of 245 Hardy-Weinberg equilibrium tests, 7 (2.86%) showed a significant departure at the 0.05 level, considering the Bonferroni correction. Rejection of the Hardy-Weinberg equilibrium was evenly distributed among populations and loci ($P>0.05$, Likelihood chi-squared test). Therefore

these departures were considered as chance deviations. Most of the *Alu* insertions were polymorphic in all populations, the exceptions being *4.75*, *APO*, *FXIII* and *ACE*, in which the Alu^{+} allele was fixed in eighteen, eight, six and one populations respectively; and the loci *4.65* and *A25* that were fixed for the Alu^{-} allele on three and five populations respectively. Considering the mean heterozygosity for each population, there was a general trend for higher diversity in Asians (0.34 against 0.30 for Northern Arctic and 0.25 for Amerindians). Indeed, the highest value for an Amerindian tribe was found in the Mvskoke, who are known to have a high degree of admixture (Kasprisin et al., 1987). The lowest diversity was encountered in the Aché of Paraguay. This was expected, given that they have been living in isolation for some hundred years and have a low genetic diversity in other genetic systems (Battilana et al., 2002; Schmitt et al., 2004).

Several clusters of high historical consistency were formed in the neighbor-joining population tree (Fig. 2), despite the generally low bootstrap values. Northern Arctic and Northeast Asian populations formed a first group, except for the Alaskan Natives, which occupied a position closer to the outgroup, while the remaining populations formed the second group, which contained all Amerind populations, except Mvskoke, which was included in the first group. Populations such as Aché and China are highly divergent in the tree. The last cluster was formed by populations considered as outgroup.

The plotting of the two principal components (Fig. 3) was also informative about the relationships among populations, and the general picture was highly concordant with the distance tree. The two first principal components accounted for 34% and 17% of the variance, respectively. The first PC separated the populations according to their geographical origin, Native Americans at the left, Northern Arctic and Asians at the right, with an extreme displacement of the PNG Coastal population, one of the outgroup population. Northern Arctic tended to be in a somewhat intermediate position between Amerinds and Asians, the exceptions being the Alaskan Natives, closer to the Asians, and the Mvskoke with the Northern Arctic. The second component distinguished between Northeast and Southeast Asians and strengthened the distinctiveness of the PNG Coastal population. The model-based clustering method implemented in the STRUCTURE program (Pritchard et al., 2000) was used to assign individuals to genetically homogeneous clusters and thus infer the underlying population structure. Four different numbers of clusters were tested: K=2, K=3, K=4 and K=5. The highest likelihood was found for K=5.

However, for *K*=3 the structure recovered was very similar to that of *K*=4 and *K*=5. The results for *K*=3 and *K*=5 are shown in Fig. 4a. It is clear the high similarity of the Amerind populations, with exception of the Mvskoke. The clustering of the average probabilities of assignment for each population suggests three to four subclusters (Fig. 4b,c). For *K=3* the first cluster was composed by the PNG Coastal population only, the second subclusters consisted of the remaining outgroup populations plus the Alaskan Natives, and third was formed by Northeast Asians, the remaining Northern Arctic, and Mvskoke. The fourth subcluster was formed by all South Amerinds and Maya. The inferred clustering for *K*=5 was very similar to that one, except that the subcluster formed by Alaskan Natives and East Indonesia populations (outgroup) received now the Mvskoke, which showed a greater affinity with Alaskan Natives, the two East Indonesia forming a second pair of populations. These results are in accordance with both the allele frequency NJ tree and the PCA.

A hierarchical analysis of genetic variation (AMOVA) was performed for all *Alu* loci (Table 2). As is true for all human groups, most of the genetic variability occurs within the population level (~80%). Several alternative groupings were tried based on a geographical criterion. The two larger values found among groups are 9.4 and 11%, and refers to comparisons of Northern Arctic with the other groups. When the Northeast Asians are taken in count in this comparison, the partition of variance referred among groups are the lesser, except in one case. It shows that Amerindian were more similar to Northeast Asians than to Northern Arctic, and that these last had higher variability. A similar pattern was observed on the NJ tree, where Northeast Asians were close to Amerindians in the intermediate cluster. An uncommon difference between Northern Arctic and Amerindians was observed among groups (11%). The variation among populations within groups were reasonable and represents the loci variability in the studied populations.

To estimate the relative amount of gene flow experienced by each population studied, we plotted the heterozygosity of each population against its distance from the centroid (Fig. 5). Asians present a greater than predicted heterozygosity, indicating that they have had more gene flow than the average, whereas Amerindians had a lower than predicted heterozygosity. Northern Arctic tended to be closer to the Amerindians, below the expectation line.

This difference between Asians and Amerindians is also reflected by the Mantel test, since these groups of populations are the most distantly located and genetically distinct. Considering all populations, there was a significant correlation between genetic and geographical distances (r: 0.523; *P*=0.001). When only Northern Arctic and Amerindian populations were considered such correlation persisted (r: 0.430; *P*=0.001), but when only the Amerindian populations (r: 0.268; *P*=0.123), or when only South Amerindians were taken into account (r: 0.105; *P*=0.322) it did not. These results are also corroborated by the AMOVA analysis, in which further subdividing South Amerinds in geographical or linguistic domains did not increase the among-group component of the test (data not shown).

DISCUSSION

A trend towards a reduction in the average heterozygosity from Southeast Asians to South Amerinds was found in the present study. This pattern is compatible with a model in which the population sizes in Asia are higher than those of Native Americans, thus allowing genetic drift to have a stronger impact over the genetic variability of the latter, in accordance with other genetic studies considering independent markers (O'Rourke et al., 1992; Tarazona-Santos et al., 2001; Bortolini et al. 2002; Fagundes et al., 2002). Alternatively, this pattern could have been caused by a population bottleneck during the peopling of the American continent, which would have caused a general reduction in genetic diversity in these populations. The existence and the intensity of such a bottleneck are controversial, being supported mainly by analysis of the Y-chromosome and of some mtDNA data (Bonatto and Salzano, 1997; Tarazona-Santos and Santos, 2002), but not by some studies on independent nuclear markers (Mattevi et al., 2000; Hutz et al., 2002; Heller et al., 2005) and on a smaller number of *Alu* insertions (Novick et al., 1998; Mateus-Pereira et al., 2005). However, it should be stressed that uniparental loci (mtDNA and Y-chromosome) have effective population size one fourth of the autosomal loci considered here, and are therefore much more affected by population size reduction than the later. Therefore, there are colonization scenarios involving moderate intensity bottlenecks that are compatible with both kinds of loci. Additionally, if our data reflect only this ancient bottleneck, we would expect that the Northern Arctic populations would also present a

similar reduction, and this was not the case (although it should be noted that some of our samples from this region have records of recent admixture). We believe that the current pattern of diversity is probably due: an initial moderate bottleneck intensified by more recent historical events, such as the depopulations that happened after the European conquest. Genetic drift caused by the isolation and inbreeding typical of Amerind populations could than occur (Salzano and Bortolini, 2002). When the heterozygosity was plotted against the distance from the centroid, a similar pattern of differentiation emerges, in which Amerindians presented lower, Asians higher, and Northern Arctic intermediate values in relation to those expected.

The different approaches employed to analyze the genetic structure of our *Alu* polymorphism data yielded concordant results. Considering the NJ population tree and PCA, three major clusters emerge, representing Southeast Asians (outgroup), Northeast Asians plus Northern Arctic, and Native Americans. The distinctiveness showed by one of the populations considered as outgroup (PNG) was expected, as they are known to be very divergent in several genetic markers, including the *Alu* polymorphism (Stoneking et al., 1997). This differentiation of the PNG population is also visible in the Bayesian assignment approach.

Two earlier studies using *Alu* markers found no major differences between Asians and Amerindians. Novick et al. (1998) used five *Alu* loci, and finding a close resemblance between Asian and Native American populations, suggested a model of multiple migrations to the Americas from a single “source” population. Mateus-Pereira et al. (2005) used four *Alu* loci and three *L1* insertion polymorphisms to corroborate the close genetic affinities between East Asians and Native American populations, but did not find support for claims of multiple migrations for the peopling of the Americas. Our findings corroborate the (Northeast) Asian genetic affinities of Native Americans, but we do find a gap between Amerindians and Asians, and a slight structure between Northern Arctic and Asians, possibly because we have studied more loci. Similarity between Northeast Asians and Amerindians was observed by Foster et al. (2001) in a study with mtDNA and Asian populations. They believe that repopulation of northern Asian latitudes occurred after the Last Glacial Maximum, obscured the ancestral Asian gene pool of Amerindians.

Concerning the question of a single or multiple migrations for the origin of Native Americans, we think that our data do not have sufficient information to answer it

with confidence. For example, the grouping of Northern Arctic and Northeast Asian populations in one cluster and the Amerindians in other in the NJ tree do not provide support for a dual origin, as this kind of method is not intended to depict historical branching, but is a hierarchical representation of the distance matrix. And, as explained above, genetic drift should have played a very important role in Native American populations (see Fig. 5). Therefore, while the differences between Amerindians and both Northern Arctic and Northeast Asian populations is reinforced by the lesser diversity of the former, the similarity between the latter two groups of populations maybe simply because they maintain comparable levels of diversity since they shared a recent common ancestor, although their differences could be noted when compared hierarchically in AMOVA, reporting 10% of variability between groups (data not shown).

Two populations consistently stood away from its geographically neighboring populations, the Mvskoke and Alaskan Natives. The Mvskoke were always closer to the Northern Arctic and Northeast Asians and the Alaskan Natives clustered with outgroup populations in the NJ tree, PCA, and in the Structure analysis. The position of these two populations did not find any historical support. However, as said above, the Mvskoke, are known to have a high degree of admixture (Kasprisin et al., 1987) and the Alaskan Native data studied here is a mixed sample of Eskimos, Native Amerindians, and Aleuts (Batzer et al. 1994). Therefore, the likely explanation for the results concerning these samples is that in both cases, genetic admixture with non-Indian populations has moved these populations away from their historical counterparts.

LITERATURE CITED

Arcot SS, Fontius JJ, Deininger PL, Batzer MA. 1995. Identification and analysis of a 'young' polymorphic *Alu* element. Biochim Biophys Acta 1263:99-102.

Arcot SS, Adamson AW, Lamerdin JE, Kanagy B, Deininger PL, Carrano AV, Batzer MA. 1996. *Alu* fossil relics - distribution and insertion polymorphism. Genome Res *6:*1084-1092.

Arcot SS, DeAngelis MM, Sherry ST, Adamson AW, Lamerdin JE, Deininger PL, Carrano AV, Batzer MA. 1997. Identification and characterization of two polymorphic Ya5 *Alu* repeats. Mutat Res 382:5-11.

Battilana J, Bonatto SL, Freitas LB, Hutz MH, Weimer TA, Callegari-Jacques SM, Batzer MA, Hill K, Hurtado AM, Tsuneto LT, Petzl-Erler ML, Salzano FM. 2002. *Alu* insertions versus blood group plus protein genetic variability in four Amerindian populations. Ann Hum Biol 29:334-347.

Batzer MA, Deininger PL. 1991. A human-specific subfamily of *Alu* sequences. Genomics 9:481-487.

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nature Reviews Genetics 3:370-379.

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL. 1994. African origin of human-specific polymorphic *Alu* insertions. Proc Natl Acad Sci USA 91:12288-12292.

Batzer MA, Rubin CM, Hellmann-Blumberg U, Alegria-Hartman M, Leeflang EP, Stern JD, Bazan HA, Shaikh TH, Deininger PL, Schmid CW. 1995. Dispersion and insertion polymorphism in two small subfamilies of recently amplified human *Alu* repeats. J Mol Biol 247:418-427.

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M. 1996. Genetic variation of recent *Alu* insertions in human populations. J Mol Evol 42:22-29.

Black W. 1996. BIOSYS-2: A computer program for the analysis of allelic variation in genetics. Fort Collins, CO: Department of Microbiology, Colorado State University. Available at ftp://lamar.colostate.edu/pub/wcb4.

Bonatto SL, Salzano FM. 1997. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. Am J Hum Genet 61:1413-1423.

Bortolini MC, Salzano FM, Bau CHD, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Bedoya G e Rutz-Linares A. 2002. Y-chromosome biallelic polymorphisms and Native American population structure. Annals of Human Genetics 66:255-259.

Brown BW, Russell K. 1997. Methods correcting for multiple testing: operating characteristics. Stat Med 16:2511-2528.

Carter AB, Salem AH, Hedges DJ, Keegan CN, Kimball B, Walker JA, Watkins WS, Jorde LB, Batzer MA. 2004. Genome-wide analysis of the human *Alu* Yb-lineage. Hum Genomics 1:167-178.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Fagundes NJ, Bonatto SL, Callegari-Jacques SM, Salzano FM. 2002. Genetic, geographic, and linguistic variation among South American Indians: possible sex influence. Am J Phys Anthropol 117:68-78.

Forster P, Torroni A, Renfrew C e Rohl A. 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. Mol Biol Evol 18:1864-1881.

Harpending HC, Ward RH. 1982. Chemical systematics and human population. In: Nitecki M, editor. Biochemical aspects of evolutionary biology. Chicago, IL: University of Chicago Press. p 213-252.

Heller AH, Salzano FM, Barrantes R, Krylov M, Benevolenskaia L, Arnett FC, Munkhbat B, Munkhtuvshin N, Tsuji K, Monsalve MV, Devine DV, Hutz MH, Carnese FR, Goicoechea AS, Freitas LB, Bonatto SL. 2005. Intra and intercontinental molecular variability in the 3'-UTR region of the LDLR gene. Hum Biol (in press).

Hutz MH, Callegari-Jacques SM, Almeida SE, Armborst T, Salzano FM. 2002. Low levels of STRP variability are not universal in American Indians. Hum Biol 74:791-806.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.

Jurka J, Kohany O, Pavlicek A, Kapitonov VV, Jurka MV. 2004. Duplication, coclustering, and selection of human *Alu* retrotransposons. Proc Natl Acad Sci USA 101:1268-1272.

Kapitonov V, Jurka J. 1996. The age of Alu subfamilies. J Mol Evol 42:59-65.

Karathanasis SK. 1985. Apolipoprotein multigene family: tandem organization of human apolipoprotein AI, CIII, and AIV genes. Proc Natl Acad Sci USA 82:6374-6378.

Kasprisin DO, Crow M, McClintock C, Lawson J. 1987. Blood types of the native Americans of Oklahoma. Am J Phys Anthropol 73:1-7.

Lahiri DK, Nurnberger JI, Jr. 1991. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. Nucleic Acids Res 19:5444.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res 27:209-220.

Mateus-Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Batzer MA, Bonatto SL, Salzano FM, Herrera RJ. 2005. Levels of LINE and *Alu* variability, and derived population affinities among Native Americans and Asians. Am J Phys Anthropol (in press).

Mattevi VS, Fiegenbaum M, Salzano FM, Weiss KM, Moore J, Monsalve MV, Devine DV, Hutz MH. 2000. Beta-globin gene cluster haplotypes in two North American indigenous populations. Am J Phys Anthropol 112:311-317.

Miller SA, Dykes DD, Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucl Ac Res 16:1215.

Nei M. 1987. Molecular evolutionary genetics. New York, NY: Columbia University Press.

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol 19:153-170.

Novick GE, Novick CC, Yunis J, Yunis E, Martinez K, Duncan GG, Troup GM, Deininger PL, Stoneking M, Batzer MA, Herrera RJ. 1995. Polymorphic human specific *Alu* insertions as markers for human identification. Electrophoresis 16:1596-1601.

Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ. 1998. Polymorphic *Alu* insertions and the Asian origin of Native American populations. Hum Biol 70:23-39.

O'Rourke DH, Mobarry A, Suarez BK. 1992. Patterns of genetic variation in Native America. Hum Biol 64:417-434.

Ota T. 1993. DISPAN: genetic distance and phylogenetic analysis. University Park, PA: Pennsylvania State University.

Otieno AC, Carter AB, Hedges DJ, Walker JA, Ray DA, Garber RK, Anders BA, Stoilova N, Laborde ME, Fowlkes JD, Huang CH, Perodeau B, Batzer MA. 2004. Analysis of the human *Alu* Ya-lineage. Journal of Molecular Biology 342:109-118.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Rohlf FJ. 1998 NTSYSpc Numerical taxonomy and multivariate analysis system. Version 2.02i. New York, NY: Applied Biostatistics Inc., Department of Ecology and Evolution State University of New York.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

Salzano FM, Bortolini MC. 2002. The evolution and genetics of Latin American populations. Cambridge, Cambridge University Press.

Santos FR, Pandya A, Tyler-Smith C, Pena SD, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ. 1999. The central Siberian origin for native American Y chromosomes. Am J Hum Genet 64:619-628.

Schmitt R, Bonatto SL, Freitas LB, Muschner VC, Hill K, Hurtado AM, Salzano FM. 2004. Extremely limited mitochondrial DNA variability among the Aché Natives of Paraguay. Ann Hum Biol 31:87-94.

Schneider S, Roessli D, Excoffier L. 2000. Arlequin ver. 2.000: a software for population genetics. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva.

Schurr TG. 2004. The peopling of the New World: perspective from molecular anthropology. Annu Rev Anthropol 33:551-583

Smouse PE, Long JC, Sokal RR. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. System Zool 35:627-632.

Sneath PHA, Sokal RR 1973. Numerical taxonomy: the principles and practice of numerical classification. San Francisco, CA: W. H. Freeman.

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. 1997. *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061-1071.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. Am J Hum Genet 68:1485-1496.

Tarazona-Santos E, Santos FR. 2002. The peopling of the Americas: a second major migration? Am J Hum Genet 70:1377-1380.

Tiret L, Rigat B, Visvikis S, Breda C, Corvol P, Cambien F, Soubrier F. 1992. Evidence, from combined segregation and linkage analysis, that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. Am J Hum Genet 51:197-205.

Weiner AM, Deininger PL, Efstratiadis A. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631-661.

Xing J, Salem AH, Hedges DJ, Kilroy GE, Watkins WS, Schienman JE, Stewart CB, Jurka J, Jorde LB, Batzer MA. 2003. Comprehensive analysis of two *Alu* Yd subfamilies. J Mol Evol 57:S76-89.

TABLE 1. Genotype distributions, Alu insertion frequencies(+), heterozygosity (h), Hardy Weinberg probability fitness values, and average heterozygosities

Population	*HS3.23*					*HS4.32*					*HS4.59*				
	Genotypes **(**++/+-/--**)**[2]	**(+)** freq	N[1]	h	*P*-Value[3]	Genotypes **(**++/+-/--**)**[2]	**(+)** freq	N[1]	h	*P*-Value[3]	Genotypes **(**++/+-/--**)**[2]	**(+)** freq	N[1]	h	*P*-Value[3]
South America															
Ache	28/06/02	0.861	36	.243	.110	03/08/25	0.194	36	.318	.097	26/01/00	0.981	27	.037	1.000
Cinta Larga	08/11/07	0.519	26	.509	.449	01/04/19	0.125	24	.223	.298	23/03/00	0.942	26	.111	1.000
Gavião	03/10/11	0.333	24	.454	1.000	02/04/20	0.154	26	.265	.077	05/12/08	0.440	25	.503	1.000
Guarani	18/07/01	0.827	26	.292	1.000	01/05/16	0.159	22	.274	.430	12/11/06	0.603	29	.487	.265
Kaingang	05/21/12	0.408	38	.489	.509	02/15/21	0.250	38	.380	1.000	29/08/02	0.846	39	.264	.196
Quechua A	08/05/08	0.500	21	.512	.026	03/09/09	0.357	21	.470	1.000	15/04/03	0.773	22	.359	.040
Quechua QT	04/13/06	0.457	23	.507	.684	01/10/09	0.300	20	.431	.619	13/07/03	0.717	23	.414	.299
Suruí	07/06/09	0.455	22	.507	.039	00/09/12	0.214	21	.345	.533	08/09/05	0.568	22	.502	.419
Waiwai	10/09/04	0.630	23	.476	.411	00/04/16	0.100	20	.185	1.000	13/05/03	0.738	21	.396	.093
Xavante	16/13/03	0.703	32	.424	1.000	02/11/18	0.242	31	.373	1.000	14/10/02	0.731	26	.401	1.000
Yanomami	10/13/03	0.635	26	.473	1.000	01/11/15	0.241	27	.372	1.000	06/03/14	0.326	23	.449	.001
Zoró	06/14/09	0.448	29	.503	1.000	01/11/18	0.217	30	.345	1.000	07/08/06	0.524	21	.511	.380
Central America															
Maya	11/12/05	0.607	28	.486	.693	03/09/16	0.268	28	.399	.349	15/06/06	0.667	27	.453	.011
North America															
Mvskoke	24/07/00	0.887	31	.204	1.000	09/00/21	0.300	30	.427	**<.001**	15/13/03	0.694	31	.432	1.000
Northern Arctic															
Alaska Natives	28/05/06	0.782	39	.345	**<.001**	04/10/30	0.205	44	.329	.056	12/18/10	0.525	40	.505	.537
Chukchi	09/01/00	0.950	10	.100	1.000	00/05/02	0.357	07	.495	.441	04/03/00	0.786	07	.363	1.000
Greenlanders	47/03/00	0.970	50	.059	1.000	13/18/17	0.458	48	.502	.089	08/23/19	0.390	50	.481	.774
Siberian Eskimo	34/01/00	0.986	35	.029	1.000	01/05/13	0.184	19	.309	.489	06/06/03	0.600	15	.497	.600
Southeast Asia															
Molucca	16/17/09	0.583	42	.492	.340	00/26/20	0.289	46	.410	.009	17/23/06	0.622	46	.477	.764
Nusa-Tengarra	23/25/02	0.710	50	.416	.183	07/20/23	0.340	50	.453	.528	20/18/12	0.580	50	.492	.081
PNG Coastal	06/19/20	0.344	45	.457	.742	04/22/23	0.306	49	.429	1.000	09/15/24	0.344	48	.456	.051
Northeast Asia															
China	47/02/01	0.960	49	.078	.060	09/16/22	0.362	36	.467	.110	15/26/09	0.560	46	.498	.781
Khalkh	38/03/00	0.963	41	.071	1.000	07/18/16	0.390	41	.482	.742	17/15/09	0.598	41	.487	.190
Khoton	47/03/00	0.970	50	.059	1.000	11/25/14	0.470	50	.503	1.000	20/22/08	0.620	50	.476	.763

TABLE 1. (Cont.)

Population	*HS4.65*					*HS4.75*					*TPA25*				
	Genotypes (++/+-/--)[2]	(+) freq	N[1]	h	*P*-Value[3]	Genotypes (++/+-/--)[2]	(+) freq	N[1]	h	*P*-Value[3]	Genotypes (++/+-/--)[2]	(+) freq	N[1]	h	*P*-Value[3]
South America															
Ache	00/00/28	0.000	28	.000	1.000	37/00/00	1.000	37	.000	1.000	30/02/03	0.886	35	.205	.001
Cinta Larga	00/02/22	0.042	24	.082	1.000	23/00/00	1.000	23	.000	1.000	02/17/05	0.438	24	.503	.091
Gavião	04/03/21	0.196	28	.321	.002	29/00/00	1.000	29	.000	1.000	18/10/01	0.793	29	.334	1.000
Guarani	00/01/16	0.029	17	.059	1.000	25/00/00	1.000	25	.000	1.000	17/09/04	0.717	30	.413	.179
Kaingang	02/08/28	0.158	38	.269	.206	37/00/00	1.000	37	.000	1.000	16/20/03	0.667	39	.450	.480
Quechua A	00/00/23	0.000	23	.000	1.000	23/00/00	1.000	23	.000	1.000	10/07/04	0.643	21	.470	.335
Quechua QT	00/04/18	0.091	22	.169	1.000	22/00/00	1.000	22	.000	1.000	13/04/04	0.714	21	.418	.021
Suruí	02/00/21	0.087	23	.162	.002	24/00/00	1.000	24	.000	1.000	02/14/06	0.409	22	.495	.214
Waiwai	00/01/17	0.028	18	.056	1.000	23/00/00	1.000	23	.000	1.000	10/08/00	0.778	18	.356	.529
Xavante	00/12/20	0.188	32	.310	.557	28/00/00	1.000	28	.000	1.000	06/13/10	0.431	29	.499	.708
Yanomami	00/00/26	0.000	26	.000	1.000	26/00/00	1.000	26	.000	1.000	14/09/04	0.685	27	.440	.366
Zoró	01/00/25	0.038	26	.075	.020	30/00/00	1.000	30	.000	1.000	13/10/03	0.692	26	.434	.653
Central America															
Maya	00/01/27	0.018	28	.036	1.000	27/01/00	0.982	28	.036	1.000	11/14/03	0.643	28	.468	1.000
North America															
Mvskoke	00/01/31	0.016	32	.031	1.000	32/00/00	1.000	32	.000	1.000	08/18/06	0.531	32	.506	.722
Northern Arctic															
Alaska Natives	01/04/39	0.068	44	.129	.166	43/00/00	1.000	43	.000	1.000	00/23/20	0.267	43	.396	.021
Chukchi	00/01/03	0.125	04	.250	1.000	11/00/00	1.000	11	.000	1.000	02/08/02	0.500	12	.522	.563
Greenlanders	00/03/46	0.031	49	.060	1.000	49/00/00	1.000	49	.000	1.000	08/30/12	0.460	50	.502	.252
Siberian Eskimo	00/01/16	0.029	17	.059	1.000	21/00/00	1.000	21	.000	1.000	10/21/11	0.488	42	.506	1.000
Southeast Asia															
Molucca	01/16/33	0.191	50	.298	1.000	28/08/01	0.871	37	.237	.503	18/19/13	0.564	50	.500	.097
Nusa-Tengarra	00/18/32	0.180	50	.298	.327	42/08/00	0.920	50	.149	1.000	08/21/21	0.370	50	.471	.545
PNG Coastal	01/07/39	0.096	47	.175	.344	48/01/00	0.990	49	.020	1.000	00/17/32	0.173	49	.290	.320
Northeast Asia															
China	01/10/39	0.120	49	.213	.527	03/47/00	0.530	50	.503	**<.001**	14/22/14	0.500	50	.505	.405
Khalkh	02/07/32	0.134	41	.235	.128	40/01/00	0.988	41	.024	1.000	10/21/10	0.500	41	.506	1.000
Khoton	00/02/47	0.020	49	.040	1.000	50/00/00	1.000	50	.000	1.000	12/19/19	0.430	50	.495	.147

TABLE 1. (Cont.)

Population	*ACE*					*APO*					*FXIIIB*				
	Genotypes (++/+-/--)[2]	(+) freq	N[1]	h	*P*-Value[3]	Genotypes (++/+-/--)[2]	(+) freq	N[1]	h	*P*-Value[3]	Genotypes (++/+-/--)[2]	(+) freq	N[1]	h	*P*-Value[3]
South America															
Ache	39/00/00	1.000	39	.000	1.000	33/00/00	1.000	33	.000	1.000	20/14/04	0.711	38	.417	.691
Cinta Larga	17/07/01	0.820	25	.301	1.000	23/02/00	0.960	25	.078	1.000	22/01/01	0.938	24	.120	.064
Gavião	24/02/01	0.926	27	.140	.111	29/00/00	1.000	29	.000	1.000	29/00/00	1.000	29	.000	1.000
Guarani	23/07/01	0.855	31	.252	.491	27/00/02	0.931	29	.131	.001	20/04/00	0.917	24	.156	1.000
Kaingang	15/09/02	0.750	26	.382	.626	34/02/00	0.972	36	.055	1.000	29/09/00	0.882	38	.212	1.000
Quechua A	17/04/02	0.826	23	.294	.098	23/00/00	1.000	23	.000	1.000	21/00/00	1.000	21	.000	1.000
Quechua QT	11/07/05	0.630	23	.476	.171	22/01/00	0.978	23	.043	1.000	20/01/02	0.891	23	.198	.008
Suruí	18/04/01	0.870	23	.232	.310	24/00/00	1.000	24	.000	1.000	24/00/00	1.000	24	.000	1.000
Waiwai	20/01/00	0.976	21	.048	1.000	23/00/00	1.000	23	.000	1.000	19/02/02	0.870	23	.232	.023
Xavante	15/11/03	0.707	29	.422	.662	32/00/00	1.000	32	.000	1.000	29/00/00	1.000	29	.000	1.000
Yanomami	10/04/02	0.750	16	.387	.200	26/00/00	1.000	26	.000	1.000	06/00/00	1.000	06	.000	1.000
Zoró	24/02/00	0.962	26	.075	1.000	29/01/00	0.983	30	.033	1.000	27/00/00	1.000	27	.000	1.000
Central America															
Maya	16/03/07	0.673	26	.449	**<.001**	26/02/00	0.964	28	.070	1.000	19/04/01	0.875	24	.223	.298
North America															
Mvskoke	09/15/05	0.569	29	.499	1.000	31/00/00	1.000	31	.000	1.000	10/07/07	0.563	24	.503	.049
Northern Arctic															
Alaska Natives	06/28/09	0.465	43	.503	.068	38/01/00	0.987	39	.026	1.000	13/11/17	0.845	41	.501	-
Chukchi	07/05/00	0.792	12	.344	1.000	08/04/00	0.833	12	.290	1.000	09/03/00	0.875	12	.228	1.000
Greenlanders	17/23/08	0.594	48	.488	1.000	49/01/00	0.990	50	.020	1.000	21/14/13	0.845	48	.491	-
Siberian Eskimo	09/12/01	0.682	22	.444	.361	44/04/00	0.958	48	.081	1.000	15/04/00	0.895	19	.193	1.000
Southeast Asia															
Molucca	10/37/04	0.553	51	.498	.001	29/17/04	0.734	50	.379	.468	27/18/01	0.783	46	.344	.661
Nusa-Tengarra	14/33/03	0.610	50	.481	.009	31/17/02	0.790	50	.335	1.000	34/15/01	0.830	50	.285	1.000
PNG Coastal	19/19/05	0.663	43	.452	1.000	21/17/07	0.656	45	.457	.322	04/09/34	0.181	47	.299	.027
Northeast Asia															
China	20/20/10	0.600	50	.485	.247	28/07/00	0.900	35	.183	1.000	33/12/04	0.796	49	.328	.089
Khalkh	14/22/05	0.610	41	.482	.524	33/07/01	0.890	41	.198	.388	31/06/04	0.829	41	.287	.007
Khoton	12/28/10	0.520	50	.504	.570	42/04/04	0.880	50	.213	.001	32/11/07	0.750	50	.379	.005

TABLE 1. (Cont.)

Population	PV92					*MABD1*					*A25*				
	Genotypes (++/+-/--)[2]	(+) freq	N[1]	h	*P*-Value[3]	Genotypes (++/+-/--)[2]	(+) freq	N[1]	h	*P*-Value[3]	Genotypes (++/+-/--)[2]	(+) freq	N[1]	h	*P*-Value[3]
South America															
Aché	27/11/01	0.833	39	.281	1.000	10/16/04	0.600	30	.488	.711	00/02/37	0.026	39	.051	1.000
Cinta Larga	09/10/07	0.538	26	.507	.256	02/09/12	0.283	23	.414	1.000	00/00/27	0.000	27	.000	1.000
Gavião	24/04/01	0.897	29	.189	.249	15/03/10	0.589	28	.493	**<.001**	00/00/28	0.000	28	.000	1.000
Guarani	19/07/03	0.776	29	.354	.111	07/07/15	0.362	29	.470	.014	00/06/23	0.103	29	.189	1.000
Kaingang	27/09/04	0.788	40	.339	.049	24/00/09	0.727	33	.403	**<.001**	00/02/22	0.042	24	.082	1.000
Quechua A	13/06/04	0.696	23	.433	.127	03/07/08	0.361	18	.475	.612	01/04/17	0.136	22	.241	.324
Quechua QT	13/08/02	0.739	23	.394	.609	13/01/06	0.675	20	.450	**<.001**	00/02/21	0.043	23	.085	1.000
Suruí	21/03/00	0.938	24	.120	1.000	02/08/11	0.286	21	.418	1.000	00/04/20	0.083	24	.156	1.000
Waiwai	17/06/00	0.870	23	.232	1.000	08/09/06	0.543	23	.507	.401	00/02/21	0.043	23	.085	1.000
Xavante	23/06/03	0.813	32	.310	.050	10/13/08	0.532	31	.506	.470	01/13/18	0.234	32	.365	.652
Yanomami	24/02/00	0.962	26	.075	1.000	04/06/11	0.333	21	.455	.137	00/00/24	0.000	24	.000	1.000
Zoró	21/08/01	0.833	30	.282	1.000	09/17/04	0.583	30	.494	.472	00/01/27	0.018	28	.036	1.000
Central America															
Maya	15/08/04	0.704	27	.425	.165	05/08/13	0.346	26	.462	.104	00/00/28	0.000	28	.000	1.000
North America															
Mvskoke	09/16/07	0.531	32	.506	1.000	08/07/10	0.460	25	.507	.042	01/07/24	0.141	32	.246	.479
Northern Arctic															
Alaska Natives	07/10/22	0.308	39	.432	.020	05/15/17	0.338	37	.454	.712	00/04/37	0.049	41	.094	1.000
Chukchi	02/07/02	0.500	11	.524	.581	03/05/01	0.611	09	.503	1.000	00/01/11	0.042	12	.083	1.000
Greenlanders	29/13/08	0.710	50	.416	.013	13/22/15	0.480	50	.504	.405	02/14/34	0.180	50	.298	.642
Siberian Eskimo	07/11/03	0.595	21	.494	1.000	07/07/06	0.525	20	.512	.198	03/03/12	0.250	18	.386	.032
Southeast Asia															
Molucca	26/17/07	0.691	50	.432	.183	02/11/24	0.203	37	.328	.616	00/00/47	0.000	47	.000	1.000
Nusa-Tengarra	15/23/12	0.530	50	.503	.579	02/17/31	0.210	50	.335	1.000	00/07/43	0.070	50	.132	1.000
PNG Coastal	05/18/25	0.292	48	.418	.498	00/07/35	0.083	42	.155	1.000	00/02/47	0.020	49	.040	1.000
Northeast Asia															
China	36/12/02	0.840	50	.272	.589	13/14/17	0.455	44	.502	.018	20/17/11	0.080	48	.488	-
Khalkh	20/15/06	0.671	41	.447	.294	03/06/06	0.400	15	.497	.600	01/10/30	0.146	41	.253	1.000
Khoton	24/22/04	0.700	50	.424	1.000	04/07/10	0.357	21	.470	.335	01/19/30	0.210	50	.335	.666

TABLE 1. (Cont.)

Population	Avg Het	St Error
South America		
Aché	.170	.051
Cinta Larga	.236	.058
Gavião	.224	.058
Guarani	.255	.045
Kaingang	.277	.046
Quechua A	.271	.062
Quechua QT	.299	.054
Suruí	.244	.058
Waiwai	.213	.053
Xavante	.301	.055
Yanomami	.221	.064
Zoró	.235	.062
Central America		
Maya	.295	.058
North America		
Mvskoke	.322	.061
Northern Arctic		
Alaska Natives	.291	.053
Chukchi	.308	.053
Greenlanders	.300	.061
Siberian Eskimo	.292	.060
Southeast Asia		
Molucca	.366	.041
Nusa-Tengarra	.362	.037
PNG Coastal	.303	.048
Northeast Asia		
China	.354	.047
Khalkh	.331	.051
Khoton	.325	.056

[1]Number of individuals; [2]The presence or absence of the *Alu* insertion is denoted by + and -, respectively; [3]Values in boldface, $P<0.05$ considering Bonferroni's correction.

TABLE 2. Hierarchical AMOVA analysis for Alu allelic frequency results, considering different levels of population hierarchy and different sources of variation

	Hierarchical classification	Comparisons[1]		
		Among groups	Among populations within groups	Within populations
I.	Northeast Asians and Northern Arctic *vs.* North, Central and South Amerindians	5.8	14.0	80.2
II.	Northeast Asians *vs.* Northern Arctic, North, Central and South Amerindians	6.0	14.6	79.4
III.	Northeast Asians *vs.* Northern Arctic *vs.* North, Central and South Amerindians	9.4	11.2	79.4
IV.	Northeast Asians *vs.* North, Central and South Amerindians	7.6	10.6	81.7
V.	Northern Arctic *vs.* North, Central and South Amerindians	11.0	10.0	79.0

[1]The values for the three levels of analysis refer to percentage of variation; all P values < 0.05.

Fig. 1. Geographic distribution of the populations considered. 1. Molucca, 2. Nusa-Tengarra, 3. PNG Coastal, 4. China, 5. Khalkh, 6. Khoton, 7. Chukchi, 8. Siberian Eskimo, 9. Alaskan Natives, 10. Greenlanders, 11. Mvskoke, 12. Maya, 13. Yanomami, 14. Wai Wai, 15. Quechua QT, 16. Quechua A, 17. Cinta Larga, 18. Surui, 19. Gavião, 20. Zoró, 21. Xavante, 22. Kaingang, 23. Guarani, 24. Aché.

Fig. 2. Neighbor-joining tree depicting the population relationships. Numbers on the branches are bootstrap values based on 1000 replications. Bootstrap values lower than 20% are not indicated.

Fig. 3. Principal Component (PC) of the allele frequencies for the twelve *Alu* insertion polymorphic loci. Asians: circles; Northern Arctic: triangles; Amerindians: squares.

Fig. 4. a) Estimated population structure. Each individual is represented by a thin vertical line, which is partitioned into *K* colored segments that represent the individual's estimated membership fractions in *K* clusters. Black lines separate individuals of different populations; b) UPGMA tree relative to the average population membership probability based on *K*=3; c) UPGMA tree relative to the average population membership probability based on *K*=5.

Fig. 5. Plot of heterozygosity vs distance from the centroid for Amerinds, Asians and Northern Arctic. The numbers refer to populations, as indicated in Fig. 1. Circles: Asians; triangles: Northern Arctic; squares: Amerindians.

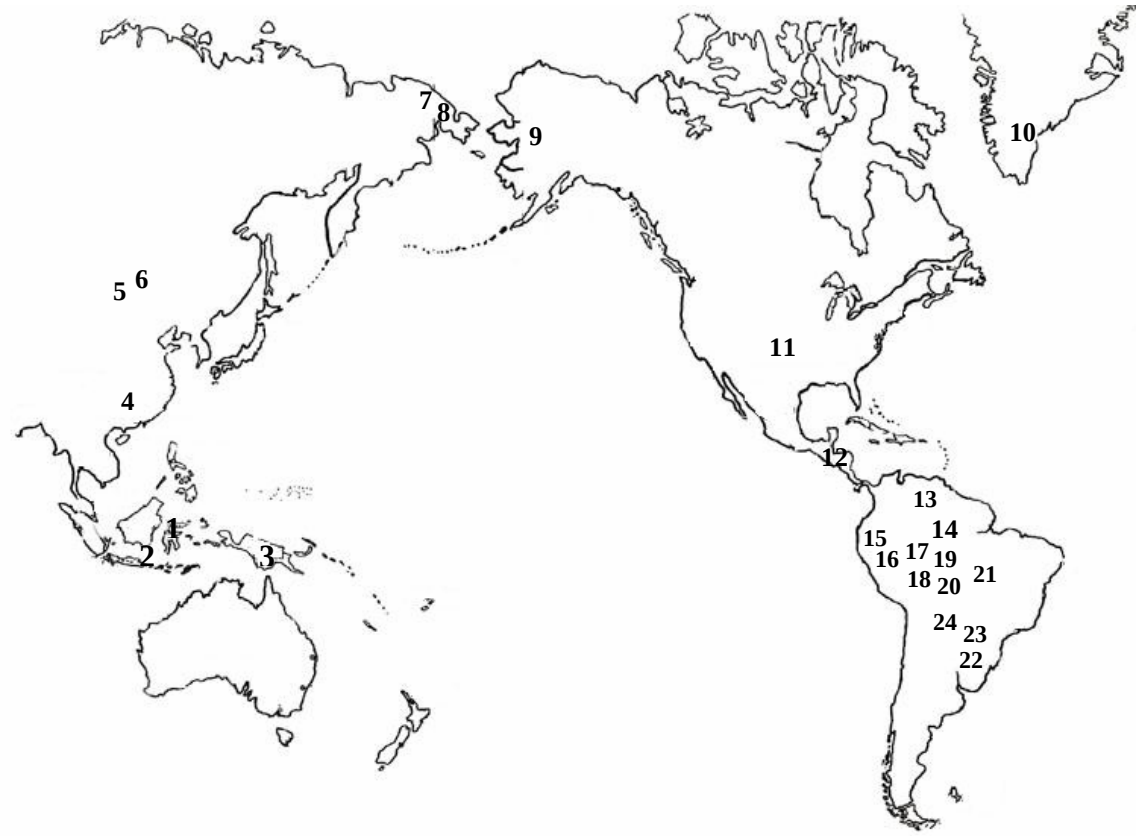

Figure 1

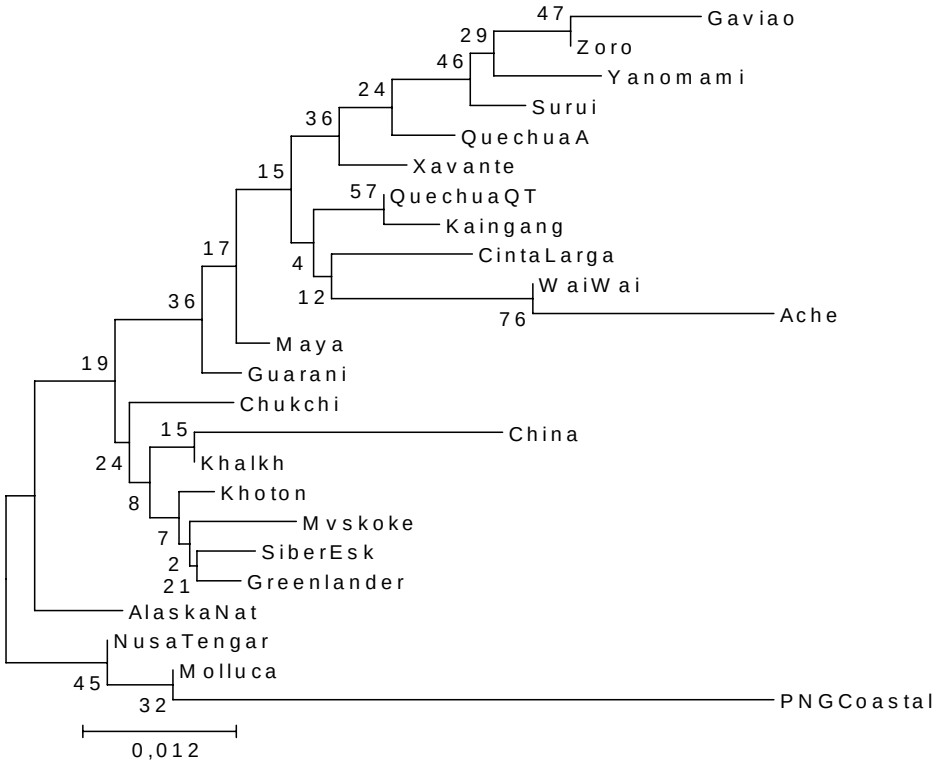
47
Gaviao
29
Zoro
46
Yanomami
24
Surui
36
QuechuaA
Xavante
15
57
QuechuaQT
Kaingang
17
4
CintaLarga
WaiWai
36
12
76
Ache
Maya
19
Guarani
Chukchi
15
China
24
Khalkh
8
Khoton
Mvskoke
7
SiberEsk
2
21
Greenlander
AlaskaNat
NusaTengar
Molluca
45
32
PNGCoastal
0,012

Figure 2

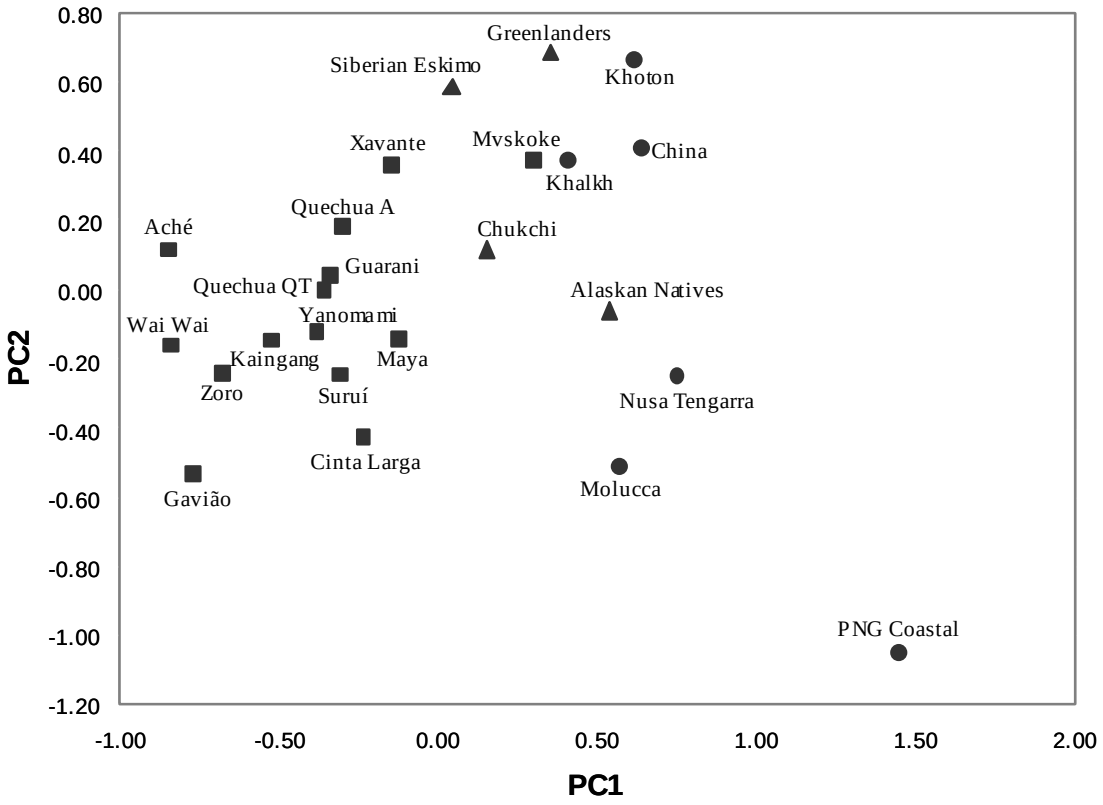
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
-0.80
-1.00
-1.20
PC2
-1.00
-0.50
0.00
0.50
1.00
1.50
2.00
PC1
Greenlanders
Siberian Eskimo
Khoton
Xavante
Mvskoke
China
Khalkh
Quechua A
Aché
Chukchi
Guarani
Quechua QT
Alaskan Natives
Yanomami
Wai Wai
Kaingang
Maya
Zoro
Suruí
Nusa Tengarra
Cinta Larga
Molucca
Gavião
PNG Coastal

Figure 3

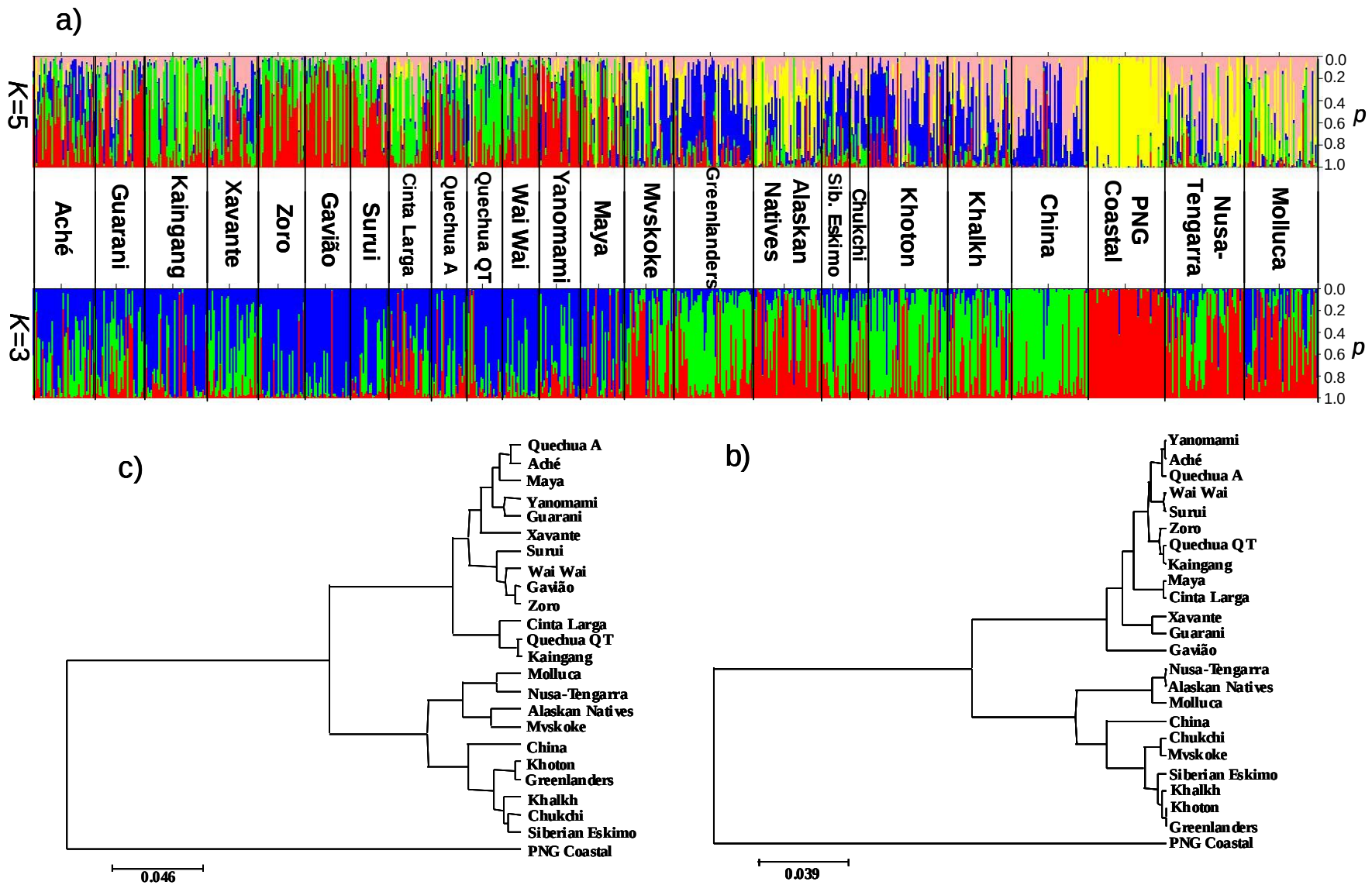
a)
K=5
K=3
p
0.0
0.2
0.4
0.6
0.8
1.0
Aché
Guarani
Kaingang
Xavante
Zoro
Gavião
Surui
Cinta Larga
Quechua A
Quechua QT
Wai Wai
Yanomami
Maya
Mvskoke
Greenlanders
Alaskan Natives
Sib. Eskimo
Chukchi
Khoton
Khalkh
China
PNG Coastal
Nusa-Tengarra
Molluca
c)
Quechua A
Aché
Maya
Yanomami
Guarani
Xavante
Surui
Wai Wai
Gavião
Zoro
Cinta Larga
Quechua QT
Kaingang
Molluca
Nusa-Tengarra
Alaskan Natives
Mvskoke
China
Khoton
Greenlanders
Khalkh
Chukchi
Siberian Eskimo
PNG Coastal
0.046
b)
Yanomami
Aché
Quechua A
Wai Wai
Surui
Zoro
Quechua QT
Kaingang
Maya
Cinta Larga
Xavante
Guarani
Gavião
Nusa-Tengarra
Alaskan Natives
Molluca
China
Chukchi
Mvskoke
Siberian Eskimo
Khalkh
Khoton
Greenlanders
PNG Coastal
0.039

Figure 4

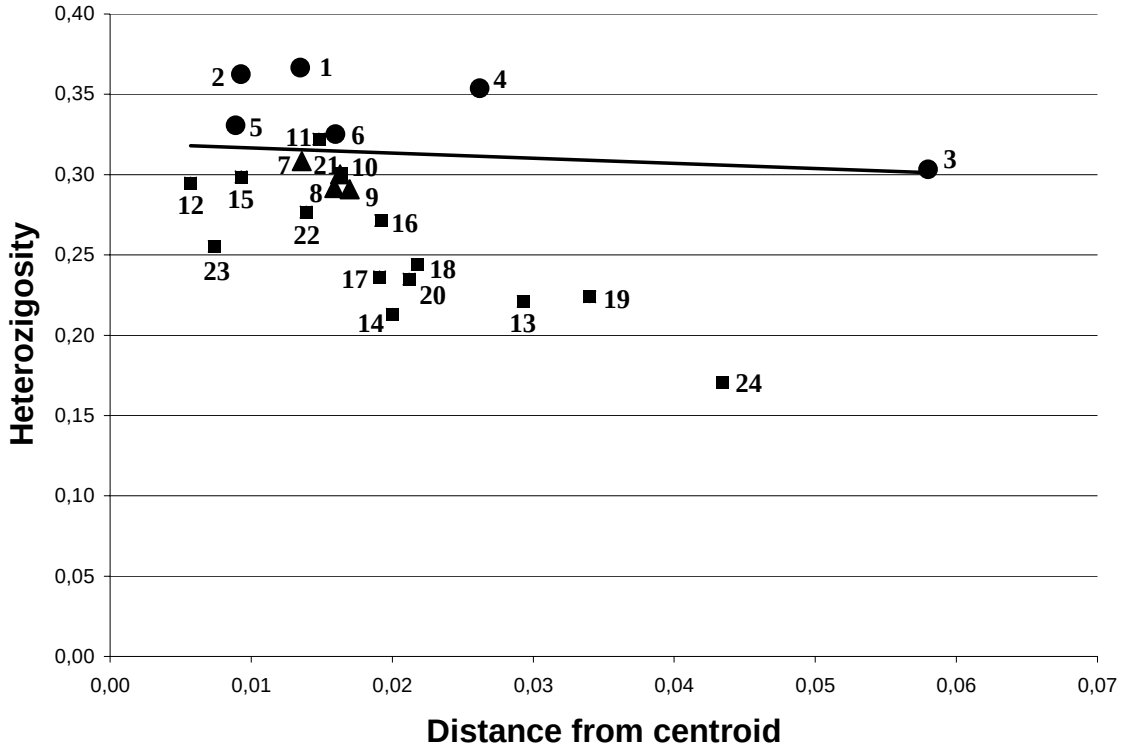
Heterozigosity
Distance from centroid
0,40
0,35
0,30
0,25
0,20
0,15
0,10
0,05
0,00
0,00
0,01
0,02
0,03
0,04
0,05
0,06
0,07
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Figure 5

CAPÍTULO IV

DISCUSSÃO GERAL

4. Discussão Geral

Os dois artigos científicos apresentados nessa tese tratam, de um modo geral, dos povos ameríndios e sua variabilidade genética, seus relacionamentos intracontinentais, e intercontinentais com povos que supostamente seriam seus ancestrais; além disso, tentam esclarecer algumas questões demográficas e evolutivas envolvendo estes povos. Tais aspectos foram avaliados pelo estudo de marcadores genéticos autossômicos (seqüência 16p13.3 e inserções *Alu*) levando em conta, além dos aspectos genéticos, os culturais, lingüisticos, antropológicos e arqueológicos. Ambos marcadores revelaram-se informativos em relação às questões levantadas. Nos dois trabalhos os ameríndios apresentam variabilidade genética semelhante àquela apresentada pelos Beringianos, reforçando a hipótese de que entre os povos asiáticos estariam (ou estiveram) os povos ancestrais dos ameríndios. Essa relação tem sido registrada há algumas décadas por diferentes pesquisadores de diversas áreas (Turner, 1984; Kozintsev *et al.*, 1999; Lell *et al.*, 2002; Uinuk-ool *et al.*, 2002; Starikovskaya *et al.*, 2005).

Quanto à questão demográfica, não houve sinal de “bottleneck” nas populações ameríndias nem nas beringianas em nossas investigações. Essa questão é ainda controversa, porque alguns pesquisadores defendem que houve diminuição significativa no tamanho da população fundadora seguida de crescimento populacional na entrada do homem pré-histórico nas Américas (Bonatto e Salzano, 1997; Tarazona-Santos e Santos, 2002; Fagundes *et al.*, 2005). Por outro lado, muitos autores não têm encontrado um sinal significativo de que isso tenha de fato acontecido ou que, pelo menos tenha sido forte o suficiente para ter restringido o reservatório genético das populações atuais (Novick *et al.*, 1998; Mattevi *et al.*, 2000; Hutz *et al.*, 2002; Heller *et al.*, 2005; Mateus-Pereira *et al.*, 2005). Encontramos uma redução na heterozigosidade das populações estudadas com as inserções *Alu* desde o sudoeste da Ásia até o sul da América do Sul. Isso pode ser explicado por eventos históricos recentes de isolamento populacional e endocruzamento e também por efeitos de deriva genética. Estes eventos, podem ser resumidos em dois modelos. De acordo com o primeiro, os grupos ancestrais estabeleceram-se em locais específicos muito cedo no processo de colonização e permaneceram naquelas áreas desde então. Este padrão levaria a significante continuidade biológica e cultural entre os grupos ancestrais e seus modernos antecedentes. De fato, este modelo explicaria porque alguns

haplótipos da sequência 16p13.3 encontrados em quatro ameríndios (Quechua, Kaingang, Cinta Larga e Aché) e em um beringiano (Chukchi), aparecem em uma posição tão divergente na árvore de haplótipos (median joining network) localizando-se entre os africanos. Estes indivíduos teriam conservado os haplótipos de seus ancestrais, os quais os teriam trazido em suas migrações para as Américas. Alternativamente (segundo modelo), a composição genética das populações, poderia não ser a mesma da ancestral devido a fatores como reassentamentos, fusões entre tribos adjacentes, deriva genética, ou outros processos estocásticos que poderiam ter alterado os padrões de diversidade biológica através do tempo, condicionando, no entanto, uma certa similaridade entre os grupos modernos e os ancestrais.

Quando houve tentativa de estabelecer relações entre os diversos grupos ameríndios estudados, sem considerar asiáticos e beringianos, não foi observada qualquer estruturação significativa. Quando todas elas eram comparadas com asiáticos (não beringianos), elas permaneceram unidas e distintas, principalmente em relação às inserções *Alu*. Alguns autores sugeriram que o processo de colonização ocorrido na América do Sul provavelmente envolveu um padrão de colonização bidirecional, que usou rotas andinas e amazônicas (Salzano e Callegari-Jacques, 1988; Tarazona-Santos *et al.*, 2001; Keyeux *et al.*, 2002). A partir deste modelo, os grupos colonizadores teriam inicialmente experimentado efeitos significantes de deriva genética devido ao relativo isolamento entre eles, mas teria sido conservado parte do mesmo reservatório genético devido ao fluxo gênico entre as populações. Logo depois da ocupação destas regiões, as populações nativas teriam sofrido um processo de tribalização, marcado por uma significante redução no fluxo gênico entre elas (Salzano e Callegari-Jacques, 1988; Torroni *et al.*, 1993; Malhi *et al.*, 2002). Esta transição pôde ser vista em sítios arqueológicos do Período Arcaico, onde apareceram evidências de especialização e intensificação do uso dos recursos locais (Roosevelt *et al.*, 1996; Dillehay, 1999; Fiedel, 1999; Fagan, 2000). O aumento no crescimento populacional e no sedentarismo destes grupos acompanharam estas mudanças e, conseqüentemente, teriam reduzido os efeitos da deriva genética nestas populações e aumentado o fluxo gênico entre grupos locais, contribuindo assim para a formação do reservatório genético regional (Malhi *et al.*, 2002). Estas flutuações impossibilitaram, em certo grau, que houvesse estruturação significativa dentro da América do Sul.

Em relação aos resultados encontrados na sequência 16p13.3 em ameríndios, eles refletem o padrão encontrado para outras populações mundiais com essa mesma seqüência (Alonso e Armour, 2001), sugerindo um crescimento populacional antigo ocorrido fora da África que, de alguma forma se manteve nas populações ameríndias atuais.

O grupo Aché é uma população ameríndia situada no Paraguai, que possui características ecológicas e morfológicas de grupos caçadores-coletores, mas sua linguagem pertence ao tronco Tupi. Alguns trabalhos envolvendo polimorfismos genéticos tentaram estabelecer a origem deste grupo, se seria Tupi ou Gê (Battilana *et al.*, 2002; Gaspar *et al.*, 2002; Tsuneto *et al.*, 2003; Schmitt *et al.*, 2004), mas os resultados não foram congruentes, e até o momento esta questão não pôde ser resolvida. Na análise com as inserções *Alu*, os Aché apresentaram o menor valor de heterozigosidade entre todas as populações estudadas, sugerindo alto grau de endocruzamento. Nas demais análises, eles se posicionaram ora próximos ao grupo Yanomami, ora aos Wai Wai, os quais pertencem a famílias lingüísticas distintas daquelas propostas para o povo Aché (Chibcha-Paeza e Caribe, respectivamente). Já quanto à árvore de haplótipos da análise da seqüência 16p13.3, além do haplótipo divergente já mencionado, os outros não apresentam padrões de relação que possam ser úteis à elucidação da questão proposta.

CAPÍTULO V

BIBLIOGRAFIA GERAL

5. Bibliografia Geral

Alonso S e Armour JA (2001) A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outstide Africa. Proc Natl Acad Sci U S A 98:864-869.

Arcot SS, Fontius JJ, Deininger PL e Batzer MA (1995a) Identification and analysis of a 'young' polymorphic *Alu* element. Biochim Biophys Acta 1263:99-102.

Arcot SS, Wang Z, Weber JL, Deininger PL e Batzer MA (1995b) *Alu* repeats: a source for the genesis of primate microsatellites. Genomics 29:136-144.

Arcot SS, Adamson AW, Lamerdin JE, Kanagy B, Deininger PL, Carrano AV e Batzer MA (1996) *Alu* fossil relics--distribution and insertion polymorphism. Genome Res 6:1084-1092.

Arcot SS, DeAngelis MM, Sherry ST, Adamson AW, Lamerdin JE, Deininger PL, Carrano AV e Batzer MA (1997) Identification and characterization of two polymorphic Ya5 *Alu* repeats. Mutat Res 382:5-11.

Badge RM, Yardley J, Jeffreys AJ e Armour JA (2000) Crossover breakpoint mapping identifies a subtelomeric hotspot for male meiotic recombination. Hum Mol Genet 9:1239-1244.

Bailey AD e Shen CK (1993) Sequential insertion of *Alu* family repeats into specific genomic sites of higher primates. Proc Natl Acad Sci U S A 90:7205-7209.

Battilana J, Bonatto SL, Freitas LB, Hutz MH, Weimer TA, Callegari-Jacques SM, Batzer MA, Hill K, Hurtado AM, Tsuneto LT, Petzl-Erler ML e Salzano FM (2002) *Alu* insertions versus blood group plus protein genetic variability in four Amerindian populations. Ann Hum Biol 29:334-347.

Batzer MA e Deininger PL (1991) A human-specific subfamily of *Alu* sequences. Genomics 9:481-487.

Batzer MA e Deininger PL (2002) *Alu* repeats and human genomic diversity. Nature Reviews Genetics 3:370-379.

Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL e Deininger PL (1990) Structure and variability of recently inserted *Alu* family members. Nucleic Acids Res 18:6793-6798.

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ e et al. (1994) African origin of human-specific polymorphic *Alu* insertions. Proc Natl Acad Sci U S A 91:12288-12292.

Batzer MA, Rubin CM, Hellmann-Blumberg U, Alegria-Hartman M, Leeflang EP, Stern JD, Bazan HA, Shaikh TH, Deininger PL e Schmid CW (1995) Dispersion and insertion polymorphism in two small subfamilies of recently amplified human *Alu* repeats. J Mol Biol 247:418-427.

Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E e Zuckerkandl E (1996) Standardized nomenclature for *Alu* repeats. J Mol Evol 42:3-6.

Bonatto SL e Salzano FM (1997) Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. Am J Hum Genet 61:1413-1423.

Britten RJ, Baron WF, Stout DB e Davidson EH (1988) Sources and evolution of human *Alu* repeated sequences. Proc Natl Acad Sci U S A 85:4770-4774.

Cavalli-Sforza LL, Menozzi P e Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, 518 p.

Daniels GR e Deininger PL (1985) Integration site preferences of the *Alu* family and similar repetitive DNA sequences. Nucleic Acids Res 13:8939-8954.

Deininger PL, Jolly DJ, Rubin CM, Friedmann T e Schmid CW (1981) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J Mol Biol 151:17-33.

Deininger PL e Slagel VK (1988) Recently Amplified *Alu* Family Members Share a Common Parental *Alu* Sequence. Mol Cel Biol 8:4566-4569.

Deininger PL, Batzer MA, Hutchison CA, 3rd e Edgell MH (1992) Master genes in mammalian repetitive DNA amplification. Trends Genet 8:307-311.

Dillehay TD (1999) The late Pleistocene cultures of South America. Evol Anthropol 7:206-216.

Fagan BM (2000) Ancient North America: The Archaeology of a Continent. Thames & Hudson, New York, pp.

Fagundes NJ, Salzano FM, Batzer MA, Deininger PL e Bonatto SL (2005) Worldwide genetic variation at the 3'-UTR region of the *LDLR* gene: possible influence of natural selection. Ann Hum Genet (in press).

Fiedel SJ (1999) Older than we thought: Implications of corrected dates for Paleoindians. American Antiquity 64:95-115.

Gaspar PA, Hutz MH, Salzano FM, Hill K, Hurtado AM, Petzl-Erler ML, Tsuneto LT e Weimer TA (2002) Polymorphisms of *CYP1a1*, *CYP2e1*, *GSTM1*, *GSTT1*, and *TP53* genes in Amerindians. Am J Phys Anthropol 119:249-256.

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS e Clegg JB (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:772-789.

Harris EE e Hey J (1999) X chromosome evidence for ancient human histories. Proc Natl Acad Sci U S A 96:3320-3324.

Heller AH, Salzano FM, Barrantes R, Krylov M, Benevolenskaia L, Arnett FC, Munkhbat B, Munkhtuvshin N, Tsuji K, Monsalve MV, Devine DV, Hutz MH, Carnese FR, Goicoechea AS, Freitas LB e Bonatto SL (2005) Intra and intercontinental molecular variability in the 3'-UTR region of the *LDLR* gene. Hum Biol (in press).

Hutz MH, Callegari-Jacques SM, Almeida SE, Armborst T e Salzano FM (2002) Low levels of STRP variability are not universal in American Indians. Hum Biol 74:791-806.

Jaruzelska J, Zietkiewicz E, Batzer M, Cole DE, Moisan JP, Scozzari R, Tavare S e Labuda D (1999) Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. Genetics 152:1091-1101.

Jelnick WR e Schmid CW (1982) Repetitive sequences in eukariotic DNA and their expression. Annu Rev Biochem 51:813-844.

Jurka J, Kohany O, Pavlicek A, Kapitonov VV e Jurka MV (2004) Duplication, coclustering, and selection of human *Alu* retrotransposons. Proc Natl Acad Sci U S A 101:1268-1272.

Kaessmann H, Heissig F, von Haeseler A e Paabo S (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat Genet 22:78-81.

Kapitonov V e Jurka J (1996) The age of *Alu* subfamilies. J Mol Evol 42:59-65.

Keyeux G, Rodas C, Gelvez N e Carter D (2002) Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations. Hum Biol 74:211-233.

Korenberg JR e Rykowski MC (1988) Human genome organization: *Alu*, lines, and the molecular structure of metaphase chromosome bands. Cell 53:391-400.

Kozintsev AG, Gromov AV e Moiseyev VG (1999) Collateral relatives of American Indians among the Bronze age populations of Siberia? American Journal of Physical Anthropology 108:193-204.

Labuda D e Zietkiewicz E (1994) Evolution of secondary structure in the family of 7SL-like RNAs. J Mol Evol 39:506-518.

Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA e Wallace DC (2002) The dual origin and Siberian affinities of Native American Y chromosomes. Am J Hum Genet 70:192-206.

Malhi RS, Eshleman JA, Greenberg JA, Weiss DA, Schultz Shook BA, Kaestle FA, Lorenz JG, Kemp BM, Johnson JR e Smith DG (2002) The structure of diversity within New World mitochondrial DNA haplogroups: implications for the prehistory of North America. Am J Hum Genet 70:905-919.

Matera AG, Hellmann U e Schmid CW (1990) A transpositionally and transcriptionally competent *Alu* subfamily. Mol Cell Biol 10:5424-5432.

Mateus-Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Batzer MA, Bonatto SL, Salzano FM e Herrera RJ (2005) Levels of LINE and *Alu* variability, and derived population affinities among Native Americans and Asians. Am J Phys Anthropol (in press).

Mattevi VS, Fiegenbaum M, Salzano FM, Weiss KM, Moore J, Monsalve MV, Devine DV e Hutz MH (2000) Beta-globin gene cluster haplotypes in two North American indigenous populations. Am J Phys Anthropol 112:311-317.

Mighell AJ, Markham AF e Robinson PA (1997) *Alu* sequences. FEBS Lett 417:1-5.

Nachman MW e Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. Genetics 156:297-304.

Novick GE, Novick CC, Yunis J, Yunis E, Martinez K, Duncan GG, Troup GM, Deininger PL, Stoneking M, Batzer MA e et al. (1995) Polymorphic human specific *Alu* insertions as markers for human identification. Electrophoresis 16:1596-1601.

Novick GE, Batzer MA, Deininger PL e Herrera RJ (1996) The mobile genetic element *Alu* in the human genome. Bioscience 46:32-41.

Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA e Herrera RJ (1998) Polymorphic *Alu* insertions and the Asian origin of Native American populations. Hum Biol 70:23-39.

Oliveira SF (1999) Inserções *Alu* em populações indígenas da Amazônia brasileira. Tese de Doutorado, Universidade de São Paulo, São Paulo.

Okada N, Hamada M, Ogiwara I e Ohshima K (1997) SINEs and LINEs share common 3' sequences: a review. Gene 205:229-243.

Quentin Y (1992) Origin of the *Alu* family: a family of *Alu*-like monomers gave birth to the left and the right arms of the *Alu* elements. Nucleic Acids Res 20:3397-3401.

Rinehart FP, Ritch TG, Deininger PL e Schmid CW (1981) Renaturation rate studies of a single family of interspersed repeated sequences in human deoxyribonucleic acid. Biochemistry 20:3003-3010.

Roosevelt AC, daCosta ML, Machado CL, Michab M, Mercier N, Valladas H, Feathers J, Barnett W, daSilveira MI, Henderson A, Sliva J, Chernoff B, Reese DS, Holman JA, Toth N e Schick K (1996) Paleoindian cave dwellers in the Amazon: The peopling of the Americas. Science 272:373-384.

Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB e Batzer MA (2003) *Alu* elements and hominid phylogenetics. Proc Natl Acad Sci U S A 100:12787-12791.

Salzano FM (1978) Multidisciplinary studies in tribal societies and human evolution. In: Meier RJ, Otten CM e Abdel-Hameed F (eds) Evolutionary models and studies in human diversity. Mouton, The Hague, pp 181-199.

Salzano FM (1982) Fatores determinísticos e estocásticos no processo microevolucionário humano. Actas V Congr Latinoam Genet: 81-89.

Salzano FM (2002) Molecular variability in Amerindians: widespread but uneven information. An Acad Bras Cienc 74:223-263.

Salzano FM e Callegari-Jacques SM (1988) South American Indians: a case study in evolution. Clarendon Press, Oxford, 259 pp.

Schmid CW e Jelinek WR (1982) The *Alu* family of dispersed repetitive sequences. Science 216:1065-1070.

Schmitt R, Bonatto SL, Freitas LB, Muschner VC, Hill K, Hurtado AM e Salzano FM (2004) Extremely limited mitochondrial DNA variability among the Ache Natives of Paraguay. Ann Hum Biol 31:87-94.

Schurr TG (2004) The peopling of the New World: perspectives from molecular anthropology. Annu Rev Anthropol 33:551-583.

Schurr TG e Sherry ST (2004) Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. Am J Hum Biol 16:420-439.

Shaikh TH e Deininger PL (1996) The role and amplification of the HS *Alu* subfamily founder gene. J Mol Evol 42:15-21.

Shaikh TH, Roy AM, Kim J, Batzer MA e Deininger PL (1997) cDNAs derived from primary and small cytoplasmic *Alu* (*scAlu*) transcripts. J Mol Biol 271:222-234.

Shen MR, Batzer MA e Deininger PL (1991) Evolution of the master *Alu* gene(s). J Mol Evol 33:311-320.

Smit AF (1996) The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6:743-748.

Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K e Wallace DC (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. Ann Hum Genet 69:67-89.

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL e Batzer MA (1997) *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061-1071.

Tarazona-Santos E e Santos FR (2002) The peopling of the Americas: a second major migration? Am J Hum Genet 70:1377-1380.

Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD e Santos FR (2001) Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. Am J Hum Genet 68:1485-1496.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM e Wallace DC (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563-590.

Tsuneto LT, Probst CM, Hutz MH, Salzano FM, Rodriguez-Delfin LA, Zago MA, Hill K, Hurtado AM, Ribeiro-dos-Santos AK e Petzl-Erler ML (2003) HLA class II diversity in seven Amerindian populations. Clues about the origins of the Ache. Tissue Antigens 62:512-526.

Turner CG, 2nd (1984) Advances in the dental search for Native American origins. Acta Anthropogenet 8:23-78.

Uinuk-ool TS, Takezaki N, Sukernik RI, Nagl S e Klein J (2002) Origin and affinities of indigenous Siberian populations as revealed by HLA class II gene frequencies. Human Genetics 110:209-226.

Weiner AM, Deininger PL e Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631-661.