

Dissertação de Mestrado
Programa de Pós Graduação em Genética
Universidade Federal de Minas Gerais



**Estrutura genética do DNA mitocondrial em
populações nativas da América do Sul:
testes de modelos ancestrais de fluxo gênico**

Luciana Werneck Zuccherato
Orientador: Eduardo Martin Tarazona-Santos
Outubro de 2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Luciana Werneck Zuccherato

***Estrutura genética do DNA mitocondrial em
populações nativas da América do Sul: testes
de modelos ancestrais de fluxo gênico***

Dissertação apresentada ao Programa de Pós-graduação em Genética do Departamento de Biologia Geral do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito básico para obtenção do grau de Mestre em Genética

Orientador: Prof. Dr. Eduardo Martín Tarazona Santos

**Belo Horizonte
Departamento de Biologia Geral
Instituto de Ciências Biológicas
Universidade Federal de Minas Gerais
2008**

“Tudo no mundo começou com um sim. Uma molécula disse sim a outra molécula e nasceu a vida. Mas antes da pré-história havia a pré-história da pré-história e havia o nunca e havia o sim. Sempre houve. Não sei o quê, mas sei que o universo jamais começou.”

Clarice Linspector – A Hora da Estrela

Agradecimentos

Gostaria prestar meus sinceros agradecimentos:

Ao professor Eduardo Tarazona pela orientação, carinho e atenção durante essa conquista da minha vida.

Aos professores da Pós-Graduação em Genética pelos importantes ensinamentos que contribuíram para a minha formação e pela amizade e carinho recebidos.

A todos os colegas do Departamento de Biologia Geral, que sempre me trataram com carinho e simpatia. Em especial a Marina, por todas as ajudas burocráticas e o sorriso diário, e à Dani, pelas correções essenciais na minha dissertação.

Aos colegas do LDGH, Maria Clara, Wagner, Moara, Juliana, Fernanda, Marília, Giordano, Roxana e Laélia pelo carinho e companheirismo nos momentos de difíceis.

A Cláudia Márcia Benedetto, por ter me transformado de técnica em cientista, pelo exemplo de dedicação e carinho pela Ciência e pelo mundo.

Aos professores Fabrício Rodrigues dos Santos, Sérgio Danilo Pena e Santuza Teixeira, que serão sempre meus orientadores do coração, e pelos quais nutro muito carinho e respeito pelo profissionalismo e exemplo de vida.

Aos colegas do LBEM, LGB e Laboratório de Genética Molecular de Parasitas, que, desde o início da minha carreira, me acolheram com simpatia, carinho e colaboração e me ensinaram muito do que eu sei hoje.

Em especial ao Carlos, Ferdi, Dani, Jorge, Rê, Rodrigo, Pati, Raquel, Sara e Simone, que se transformaram em amigos que eu levarei pra sempre no meu coração, e que tento sempre seguir os exemplos de vida.

Aos meus pais, minha irmã Cristiane e meu sobrinho Felipe, por sempre me mostrarem o quanto a vida é linda! Por sempre me apoiarem nas minhas escolhas e sempre me lembrarem do quão especial eu sou. Aos meus irmãos Aninha e Pepe pelo orgulho que sentem por mim.

A todos os meus colegas de coração do ICB: Carol, Catita, Charles, Elimar, Gabriela, Helder, Jarina, Josimar, Leandro, Marilza, Neuzinha, Renan, Sibelle, Thiago, Vanessa e Vinício, por todas as alegrias que me trouxeram no dia-a-dia e por nutrirem por mim o mesmo carinho que eu sinto por eles.

As minhas amigas Elisa, Débora, Sandra, Danusa e Mirinha, por sempre terem ficado ao meu lado nos momentos mais difíceis.

Enfim, a todos que plantaram uma semente de carinho em mim e me transformaram sempre numa pessoa melhor! Obrigada!

	<u>Índice</u>
Índice de Figuras	vii
Índice de Tabelas	viii
Índice de Anexos	ix
Lista de Abreviaturas e Símbolos	x
Resumo	xi
Abstract	xiii
I. Introdução	1
1. Evidências sobre o povoamento das Américas	2
2. Modelo de evolução das populações da América do Sul	6
3. Shima e Monte Carmelo	3
II. Objetivos	13
III. Metodologia	15
1. Extração, quantificação e armazenagem do DNA	16
2. Amplificação e Sequenciamento	16
3. Determinação dos haplogrupos de DNAMt	16
4. Coleta de dados da literatura	17
5. Atualização do Banco de Dados	17
6. Análises populacionais	18
7. Teste de Modelo de Isolamento por Distância	18
IV. Resultados	22
1. Análise da sequência HVI do DNAMt	23
2. Atualização do Banco de Dados	24
3. Diversidade das populações nativas da América do Sul	24
4. Teste da hipótese nula de Isolamento por Distância	26
V. Discussão	28
VI. Conclusão e Perspectivas	32
VII. Referências Bibliográficas	34

FIGURA	PÁGINA
1. Haplogrupos de DNAMt e sua distribuição no continente americano.	4
2. Árvore de máxima- verossimilhança (A) e <i>Bayesian Skyline plot</i> (B) de divergência dos haplogrupos de DNAMt.	6
3. Modelo evolutivo da América do Sul proposto por Tarazona-Santos <i>et al.</i> (2001).	7
4. Modelo de <i>Isolamento por Distância</i> (IBD, Wrigth, 1943).	8
5. Regressão do eixo maior reduzido das distâncias genéticas <i>versus</i> o logarítmico das distâncias geográficas das populações do Oeste, Leste e toda a América do Sul.	9
6. Mapa do Departamento de Cusco/Peru e Detalhe da localização da Reserva Machiguenga.	10
7. Mapa da localização da Reserva Machiguenga.	11
8. Distribuição das populações nativas da América do Sul utilizadas no presente estudo.	20
9. Haplótipos e haplogrupos observados nas populações de Monte Carmelo e Shimaa.	22
10. Regressão do eixo maior reduzido das distâncias genéticas <i>versus</i> o logarítmico da distância geográfica de todas as populações da América do Sul.	25
11. Regressão do eixo maior reduzido das distâncias genéticas <i>versus</i> o logarítmico da distância geográfica das populações do Leste e Oeste da América do Sul.	26
12. Gráfico dos valores decrescentes de θ_K observados em nossas análises.	29

Índice de Tabelas

TABELA	PÁGINA
1. Descrição das Populações Nativas da América do Sul utilizadas no presente estudo.	19
2. Valores dos três estimadores de θ , seus intervalos de 95% de confiança (95% IC) ou Desvios Padrões (DP), e valores de testes de Neutralidade de D de Tajima e Fs de Fu.	24

Índice de Anexos

ANEXO	PÁGINA
1. Caixa 1- Protocolo Adaptado de Extração de DNA.	40
2. Caixa 2- Amplificação da região HVI do DNAMt.	41
3. Caixa 3- <i>Pipeline</i> de Análise de sequências.	42
4. Caixa 4- Algoritmo para atualização do database de seqüências.	43

Lista de Abreviaturas e Símbolos

Encontram-se abaixo as definições das abreviaturas e símbolos que foram citados no texto:

DNA: ácido desoxirribonucléico

DNAm: DNA mitocondrial

HVI: região hipervariável I

PCR: reação em cadeia da polimerase

IBD: *Isolation by distance* (Isolamento por Distância)

FWD: *forward*

REV: reverso

TS2001: Modelo Tarazona-Santos *et al.*, 2001

NRY: *Nonrecombining portion of the Y chromosome* (Região não recombinante do cromossomo Y)

RFLP: polimorfismo de tamanho de fragmento de restrição

Resumo

Entre 2001 e 2003, nosso grupo propôs um modelo de povoamento peculiar da América do Sul (*TS2001*), em que o padrão de diversidade genética das populações sul-ameríndias pode ser explicado por uma ação diferencial da deriva genética e do fluxo gênico nas populações do Oeste e do Leste do continente. Na parte Oeste do continente sul-americano, que compreende a região dos Andes, foi evidenciada uma menor divergência genética interpopulacional e uma maior diversidade intrapopulacional. Isso sugere que as populações exibiram ao longo do tempo um grande tamanho efetivo e altas taxas de fluxo gênico, determinando uma tendência de homogeneização do *pool* gênico. De forma contrastante, populações do Leste (Amazônicas) foram sujeitas a forte ação da deriva genética e mantiveram baixas taxas de fluxo gênico entre si, com resultante tendência à diferenciação populacional. Fuselli *et al.* (2003), utilizando o modelo genético populacional de *Isolamento por Distância* (IBD), evidenciaram que não existe correlação entre diversidade genética e distância geográfica quando se analisa a América do Sul como um todo, mas sim quando as populações do Leste e do Oeste são analisadas separadamente, sugerindo a existência de dois complexos de populações evoluindo sob a ação de fatores evolutivos diferenciados. Para testar a vigência do modelo *TS2001*, geramos dados da região hipervariável I do DNA mitocondrial de duas populações nativas peruanas de etnia Machiguenga (Shimaa, n=37 e Monte Carmelo, n=28), localizadas numa região de transição entre os Andes e o Leste do continente. Além disso, foi atualizada uma base de dados genéticos de 1698 indivíduos de 45 populações nativas da América do Sul, permitindo uma ampla cobertura geográfica do continente. As análises da diversidade intrapopulacional e do espectro de freqüências alélicas sugerem que as populações Andinas têm mantido um tamanho efetivo maior que as populações do Leste da América do Sul, o que é compatível com maiores níveis de fluxo gênico entre elas e tamanho censitário mantido ao longo do tempo. As populações localizadas na região de transição entre as duas partes do continente se adaptam ao padrão observado na região Oeste. Enquanto no Oeste da América do Sul o padrão de variabilidade se adapta ao modelo de IBD, a re-análise dos dados nas populações do Leste sugere padrões evolutivos mais complexos. Análises estatísticas mais robustas poderão contribuir

para elucidar o resultado observado, mas os resultados obtidos são compatíveis com o modelo proposto anteriormente pelo nosso grupo.

Palavras chave: DNAm, populações nativas, estrutura genética, inferências evolutivas.

.

Abstract

Between 2001 and 2003, our group proposed the following model for the settlement of South America (TS2001): the pattern of genetic diversity of south Amerindian populations could be explained by the differential action of genetic drift and gene flow over the populations located West or East of the continent. On the Western of the continent, the Andean region, we observed low interpopulacional divergence and high intrapopulacional diversity, suggesting a larger long term effective size with high gene flow among populations, therefore determining a trend towards homogenization of the gene pool. On the other hand, populations on the East (Amazonia) underwent stronger genetic drift and decreased levels of gene flow, leading to the differentiation of the populations. Fuselli *et al.* (2003), using the population genetics model of Isolation by Distance (IBD), found no correlation between geographic distance and genetic variability when populations from all South America were analyzed together, but found such correlation when Eastern and the Western populations were analyzed separately. This result suggested the existence of two complexes of populations evolving under the action of different evolutionary factors. To test the validity of the model, we present above data for the hipervariable region I of the mitochondrial DNA for two Peruvian native populations from the Machiguenga ethnic group (Shimaa, n=37 and Monte Carmelo, n=28), located in a region of transition between the Andes and the Amazon. Moreover, we prepared an updated database of genetic data of 1698 individuals of 45 native populations of the South America, which allowed us to have a wide geographic coverage of the continent. The analysis of intrapopulacional diversity and the spectrum of allelic frequencies suggest that the Andean populations have maintained a larger effective size than populations from the East of South America, what is coherent with higher levels of gene flow among them, and large census size over the time. The populations located in the transition area between the two parts of the continent fit the pattern observed in the West region. While in the West of the South America the pattern of variability fit IBD model, re-analysis of the data of the East suggests more complex evolutionary pattern. More robust statistical analyses will contribute to elucidate the observed pattern.

Key words: mtDNA, native populations, genetic structure, evolution inferences.

I- Introdução

1. Evidências sobre o povoamento das Américas

A América foi o último continente a ser colonizado pelos humanos modernos durante o Pleistoceno (há mais de 11.000 anos atrás). No final do século XV, período do expansionismo europeu, as populações existentes no continente americano apresentavam diferentes culturas e línguas, desenvolvidas durante o processo de crescimento e adaptação aos diversos ambientes (Mulligan *et al.*, 2004).

Apesar dos esforços de caracterizar geneticamente a população nativa sul-americana, o conhecimento dos processos evolutivos pré-colombianos ainda é incerto. Nos últimos anos os números de estudos genéticos sobre os nativos americanos têm aumentado consideravelmente, sendo a maioria deles direcionada aos vários aspectos do povoamento das Américas: data de colonização, número e localização das rotas migratórias, *pool* genético dos primeiros colonizadores e outros (Corella *et al.*, 2007).

Uma questão pertinente e ainda controversa é o número de migrações que ocorreram durante o povoamento das Américas. Em 1986, Greenberg, Turner e Zegura propuseram um dos primeiros modelos interdisciplinares de ocupação baseados em dados lingüísticos, dentários e genéticos. O modelo foi chamado de *Teoria de Três Ondas Migratórias* e sugere que os três diferentes grupos lingüísticos existentes na América (Ameríndios, Na-Denes e Aleutas-Esquimós) se originaram a partir de três ondas migratórias cronologicamente independentes (Greenberg, Turner e Zegura, 1986). Embora o modelo tenha sido aceito posteriormente por Cavalli-Sforza *et al.* (1994), utilizando dados de polimorfismos protéicos, e por Torroni *et al.* (1992), utilizando dados de DNA mitocondrial (DNAmt), estudos genéticos posteriores mostraram que a variação genética reduzida e a distribuição única dos polimorfismos de DNAmt e cromossomo Y sugerem que apenas uma única onda migratória é condizente com o perfil de diversidade encontrado nas Américas (Pena *et al.*, 1995; Santos *et al.*, 1995, 1996, 1999; Tarazona-Santos e Santos, 2002; Zegura *et al.*, 2004; Fagundes *et al.*, 2008).

Estudos paleoantropológicos recentes também são controversos: variações craniofaciais parecem suportar um cenário em que a América teria sido

sucessivamente ocupada por duas populações morfologicamente diferenciadas, chamado de *Modelo de Dois Componentes* (Neves *et al.*, 2003). Segundo os autores, os indivíduos do final do Pleistoceno não possuem a morfologia cranial típica dos Mongolóides, que é característica comum dos atuais nativos americanos. Portanto os não-Mongolóides teriam sido extintos ou substituídos pelos Mongolóides que vieram em migrações subseqüentes. Outros estudos paleoantropológicos sugerem que os pré-Mongolóides precederam os Mongolóides, pois há uma heterogeneidade morfológica nos crânios que persiste até tempos atuais (González-José *et al.*, 2005). A proposta de uma migração adicional foi então refutada, pois a variabilidade morfológica encontrada sugere que as duas formas, Pré-Mongolóides e Mongolóides, são encontradas nas populações de Paleoíndios e de Índios Holocênicos. Considerando os dados genéticos que indicam que os ancestrais dos nativos americanos atuais vieram para a América no Pleistoceno, quando não existiam ainda Mongolóides, nem na Ásia, uma possível interpretação consensual é de que há uma continuidade, uma linhagem ancestral-descendente entre os primeiros migrantes com morfologia não-Mongolóide e os povos indígenas de hoje (González-José *et al.*, 2008).

Embora diferentes disciplinas científicas possam averiguar a história do povoamento inicial da América, suas limitações não podem ser ignoradas. Dados arqueológicos e paleoantropológicos nos fornecem informações sobre a existência e passagem de povos do passado, que podem não ter deixado descendentes, e portanto não estar relacionados ao passado direto dos povos atuais. Os dados genéticos, por sua vez, obtidos de povos atuais, dizem respeito exclusivamente ao seu passado, sobre a história dos seus ancestrais, mas não dizem nada sobre povos do passado que não deixaram descendentes.

Apesar das controvérsias existentes a respeito do povoamento das Américas, um cenário comum parece estar surgindo a partir de novas evidências (González-José *et al.*, 2008). Muitos dos avanços na reconstrução histórica do povoamento da América são devidos ao uso de marcadores polimórficos de herança uniparental tais como o DNAmT e o cromossomo Y humano. O DNA mitocondrial é caracterizado por ser uma única região genômica não-recombinante e herdado matematicamente,

enquanto o cromossomo Y apresenta uma região não-recombinante (NRY) e é herdado paternalmente.

Nas últimas duas décadas o DNA mitocondrial tem sido extensivamente usado para caracterizar a estrutura das populações nativas americanas e sua história (Torrioni *et al.*, 1992; Schmitt *et al.*, 2004; Dornelles *et al.*, 2004; Bert *et al.*, 2004; Vona *et al.*, 2005; Garcia *et al.*, 2006; Cabana *et al.*, 2006; Tamm *et al.*, 2007; Melton *et al.*, 2007; Fagundes *et al.*, 2008). O DNA mitocondrial autóctone das populações nativo-americanas é caracterizado por cinco haplogrupos designados de A, B, C, D e X, definidos por várias e específicas mutações de ponto localizadas na região controle do genoma mitocondrial. O haplogrupo A ocorre em altas freqüências nas regiões do norte da América do Sul continente, os haplogrupos C e D são freqüentes em várias partes da América do Sul, o haplogrupo B é abundante no sul do Peru, região dos Andes Bolivianos, norte do Chile e Argentina, e o haplogrupo X é restrito a América do Norte (Schurr e Sherry, 2004) (Figura 1).

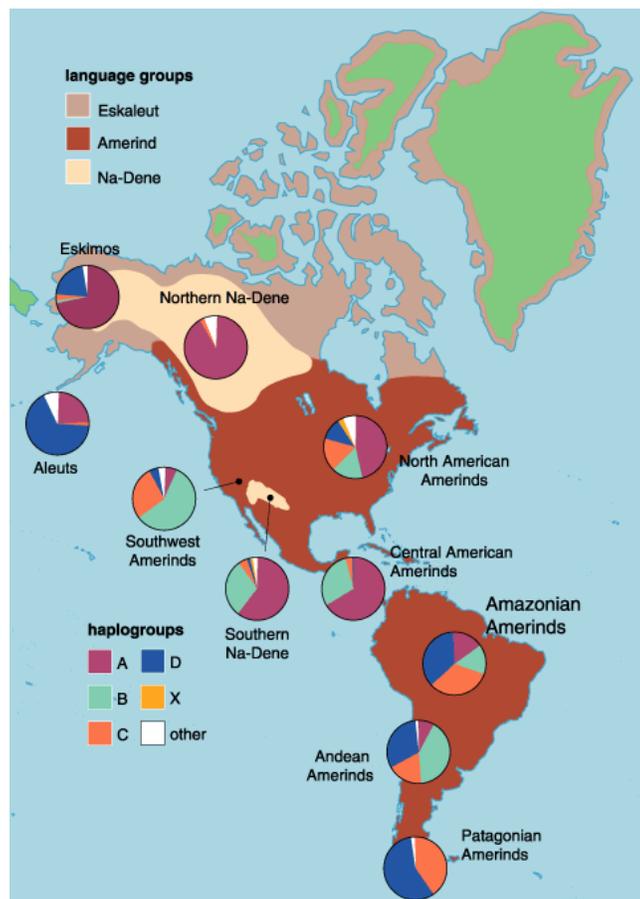


Figura 1- Haplogrupos de DNAm^t e sua distribuição no continente Americano (American Science, v.88, p.246, Maio-Junho 2000)

Estas linhagens do DNAMt também são encontradas na Ásia, corroborando a hipótese de origem a partir do continente asiático (Tamm *et al.*, 2004). Recentes estudos analisando marcadores autossômicos (Battilana *et al.*, 2007) e genomas completos de DNAMt confirmam a presença de sub-linhagens que refletem o acúmulo de mutações específicas nas populações nativas, sugerindo que os migrantes asiáticos foram isolados por um período de aproximadamente 5.000 anos antes da dispersão para as Américas (Tamm *et al.*, 2007; Fagundes *et al.*, 2008). Evidências arqueológicas e paleoclimáticas apontam a Beríngia como um lugar de refúgio climático e ecológico na formação preliminar do *pool* gênico americano (Fagundes *et al.*, 2008).

As questões relativas à data de ocupação ancestral do continente também estão sendo elucidadas. A datação de aproximadamente 18.000 anos foi demonstrada por trabalhos recentes (Fagundes *et al.*, 2008; Figura 2). Através do seqüenciamento de uma região de 9 mil pares de bases do DNAMt de alguns Ameríndios (Silva-Jr *et al.*, 2002), foi sugerida uma data de cerca de 16.000 anos atrás para o início da expansão da população que deu origem à população atual. Este ponto de vista consensual, também revelado pelos estudos com cromossomo Y (Tarazona-Santos e Santos, 2002; Bortolini *et al.*, 2003), afirma que populações ancestrais chegaram a Beríngia (e América) durante o Pleistoceno, >15.000 anos atrás, vindos do nordeste da Ásia (Santos *et al.*, 1999; Schurr e Sherry, 2004; Wang *et al.* 2007). Esses dados, associados a descobertas arqueológicas recentes no Sítio de Monte Verde, no Chile, e Pedra Pintada, no Brasil, sugerem que há aproximadamente 12.000 anos caçadores-coletores já estavam presentes na América do Sul, fazendo com que a idéia de ocupação pré-Clovis do continente pareça ser a mais verossímil. Consistentemente, estudos do DNAMt também sugerem um povoamento Pleistocênico da América do Sul (Fuselli *et al.*, 2003).

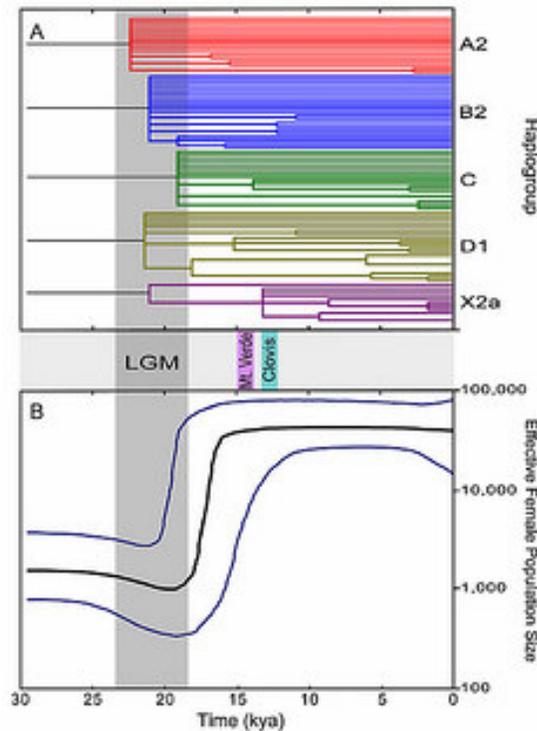


Figura 2- Árvore de máxima- verossimilhança (A) e *Bayesian Skyline plot* (B) indicam a divergência dos haplogrupos de DNAm_t a aproximadamente 20.000 anos e uma expansão populacional condizente com o povoamento das Américas, o que corrobora estimativas de recentes achados arqueológicos (Fagundes *et al.*, 2008).

A quantidade de dados genéticos gerados na última década, independente dos marcadores utilizados, está convergindo para um cenário histórico comum, embora complexo, para a origem dos nativos americanos. Informações mais precisas sobre a data exata de colonização, número e localização das rotas migratórias e *pool* genético dos colonizadores ainda estão por ser determinadas (Corella *et al.*, 2007).

2. Modelo de evolução das populações de América do Sul

As dúvidas a respeito da ocupação da América do Sul também são pertinentes entre os pesquisadores. Padrões demográficos (Schmitt *et al.*, 2004; Dornelles *et al.*, 2004; Bert *et al.*, 2004; Vona *et al.*, 2005; Garcia *et al.*, 2006; Cabana *et al.*, 2006; Melton *et al.*, 2007), geográficos e lingüísticos (Lewis *et al.*, 2005; Hunley *et al.*, 2007; Marrero *et al.*, 2007) e paleoantropológicos (Moraga *et al.*, 2005; Shinoda *et al.*, 2006) são frequentemente utilizados para explicar a diversidade genética observada nas populações da América do Sul.

Em 2001, Tarazona-Santos e colaboradores, a partir de dados do cromossomo Y de populações nativas dos Andes, Planalto Central do Brasil, Amazônia e Chaco, propuseram um modelo evolucionário peculiar da América do Sul que distingue as populações andinas e populações não-andinas. Este modelo evidencia que populações Andinas possuem alta diversidade genética intrapopulacional, porém baixa diversidade interpopulacional, enquanto as populações ameríndias localizadas na Amazônia brasileira apresentam características opostas, ou seja, baixa diversidade intrapopulacional e grande diversidade interpopulacional (Tarazona-Santos *et al.*, 2001).

A partir destes dados, os autores sugeriram que o padrão de diversidade das populações sul-ameríndias poderia ser explicado por uma ação diferencial da deriva genética e do fluxo gênico nas populações do Oeste e do Leste da América do Sul. Na parte Oeste do continente, que compreende a região dos Andes, as populações exibiram ao longo do tempo um grande tamanho efetivo e grande fluxo gênico entre elas, determinando a tendência de homogeneização do *pool* gênico, demonstrado por uma menor divergência genética interpopulacional e uma maior diversidade intrapopulacional. De forma contrastante, populações do Leste (Amazônicas) estiveram sujeitas a altas taxas de deriva e baixas taxas de fluxo gênico, com resultante tendência à diferenciação populacional (Figura 3).



Figura 3- Modelo evolutivo da América do Sul proposto por Tarazona-Santos *et al.* (2001). Segundo o modelo, as populações do Oeste do continente apresentaram maior tamanho efetivo populacional e altas taxas de fluxo gênico entre elas. Já as populações do Leste do continente apresentam um padrão inverso.

Fuselli *et al.* (2003) utilizou o modelo genético populacional de *Isolamento por Distância* (IBD, Wrigth, 1943) para explorar aspectos adicionais do modelo evolutivo de Tarazona-Santos *et al.* (2001). O modelo de IBD descreve uma dinâmica do fluxo gênico caracterizada por uma probabilidade de migrar que diminui com o incremento da distância geográfica entre populações. Esta dinâmica determina uma relação linear entre o logaritmo da distância geográfica e a variabilidade genética entre populações (Figura 4). As análises evidenciaram que a hipótese nula de *isolamento por distância* não se adapta à variabilidade genética encontrada em

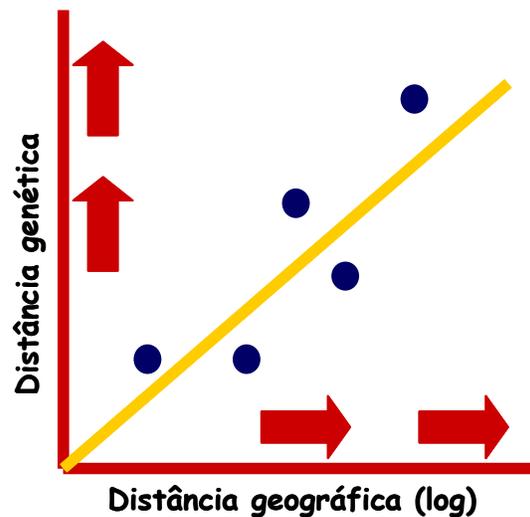


Figura 4- Modelo de *Isolamento por Distância* (IBD, Wrigth, 1943). Com o aumento da distância geográfica entre as populações se observa também um aumento da distância genética.

toda a América do Sul, mas sim quando as populações do Leste e do Oeste são analisadas separadamente, sugerindo a existência de dois complexos de populações evoluindo sob a ação de fatores evolutivos diferenciados (Figura 5). Fatores históricos se correlacionam com o modelo proposto: toda a área geográfica andina foi envolvida em um único processo cultural por pelo menos 4.000 anos, culminando com o Império Inca, e levando a uma alta homogeneidade cultural e linguística. Por outro lado, o padrão observado na parte Leste da América do Sul, correspondente à bacia amazônica, está associado à uma maior variabilidade linguística e cultural entre populações. Porém, o modelo de Tarazona-Santos *et al.* (2001) ainda deixa abertas algumas questões pertinentes, como por exemplo, se houve a existência de fluxo gênico entre as duas partes do continente e se o padrão de diversidade encontrado pode ser explicado por uma ou duas ondas migratórias (Fuselli *et al.*, 2003).

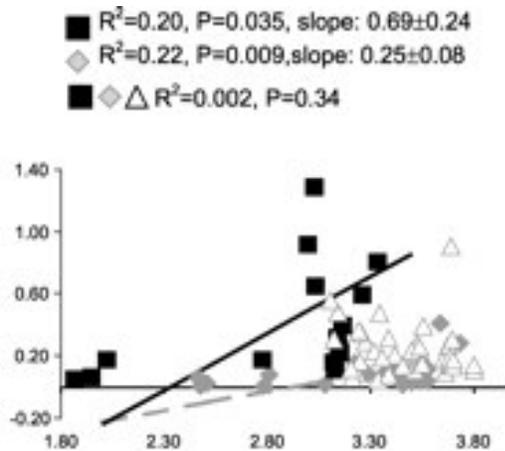


Figura 5- Regressão do eixo maior reduzido das distâncias genéticas *versus* o logarítmico da distância geográfica das populações do Oeste (quadrado) e Leste (diamantes). Os triângulos representam as distâncias entre as populações do Leste e do Oeste. O modelo de *IBD* explica 20 e 22% da diversidade genética encontrada nas populações do Oeste e Leste da América do Sul, respectivamente.

A partir de informações sobre populações estrategicamente localizadas entre as regiões Andinas e Amazônicas, espera-se elucidar algumas das questões propostas. Para isso, o nosso grupo fez novas coletas em duas populações cujos dados são apresentadas no presente trabalho: Shimaa e Monte Carmelo.

3. Shimaa e Monte Carmelo

As populações de Shimaa e Monte Carmelo pertencem à etnia Machiguenga, e ao grupo linguístico Arawak Pré-Andino (<http://www.ethnologue.com>). Encontram-se situadas às margens altas do Rio Urubamba, Província de *La Convencion*, Departamento de Cusco no Peru (Misioneros Dominicos, 2006) (Figura 6 e 7). Em 2004, a população de Shimaa era estimada em cerca de 80 famílias e 400 indivíduos, enquanto a de Monte Carmelo possuía 110 famílias e um tamanho censitário de 650 indivíduos (<http://www.pmacperu.org/resultados.htm>). Pouco se conhece dessas populações e não existem trabalhos sobre sua estrutura genética.

Os Machiguengas constituem cerca de 10.000 indivíduos, representando 3,62% da população indígena. Trata-se de um grupo originário do Rio Urubamba e seus afluentes, onde a maioria se encontra localizada há milhares de anos (Misioneros Dominicos, 2006). Possuem uma história marcada por esforços

constantes de proteção de sua região com os limites do Império Inca, que dominavam as regiões de bosque ao norte de Cusco.. Não existem relatos de subordinação dos Machiguengas ao Império Inca, embora acredita-se que houve intensas relações comerciais (Misioneros Dominicanos, 2006).

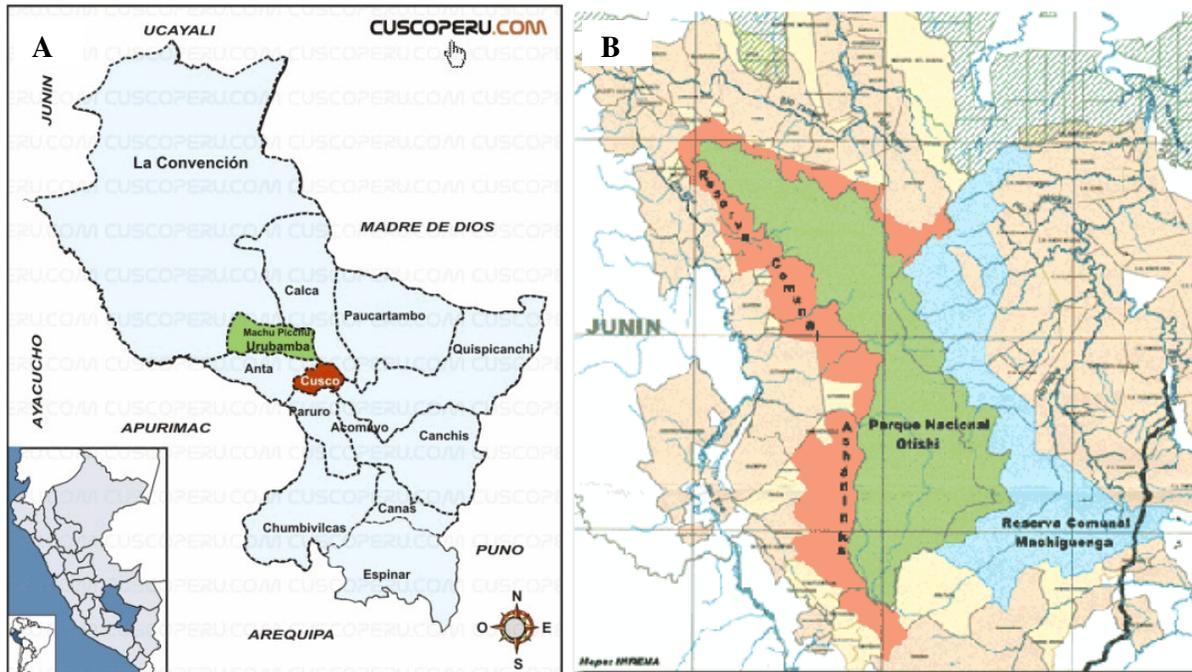


Figura 6- A) Mapa do Departamento de Cusco/Peru (<http://www.cuscoperu.com>); B) Detalhe da localização da Reserva Machiguenga (www.parkswatch.org).

A organização social dos Machiguengas é constituída por parentesco cognático, cujos membros se encontram dispersos em diferentes assentamentos. Um indivíduo deve casar-se, segundo uma clara regra de exogamia, fora do âmbito social dos consangüíneos. A regra de residência pós-matrimonial masculina provoca uma grande dispersão dos membros e uma máxima extensão do espaço social. A subsistência depende da horticultura, caça e pesca. Os principais cultivos são mandioca, batata-doce, amendoim, milho e banana. A criação de aves e animais menores como o *Cavia porcellus* se estendeu entre as famílias (Misioneros Dominicanos, 2006).

Embora a partir de 1650 os Jesuítas fundassem escolas nas regiões ao norte de Cusco, não houve esforços para evangelizar essas populações. Ao final do século XIX, com o interesse do mercado na exploração de quinino e borracha, o interesse pelas populações próximas a Cusco aumentou consideravelmente, tendo como resultado uma alta taxa de mortalidade entre os Machiguengas devido às epidemias de malária e varíola que ocorreram na época, levando-os a se refugiarem em zonas isoladas (Misioneros Dominicicos, 2006).

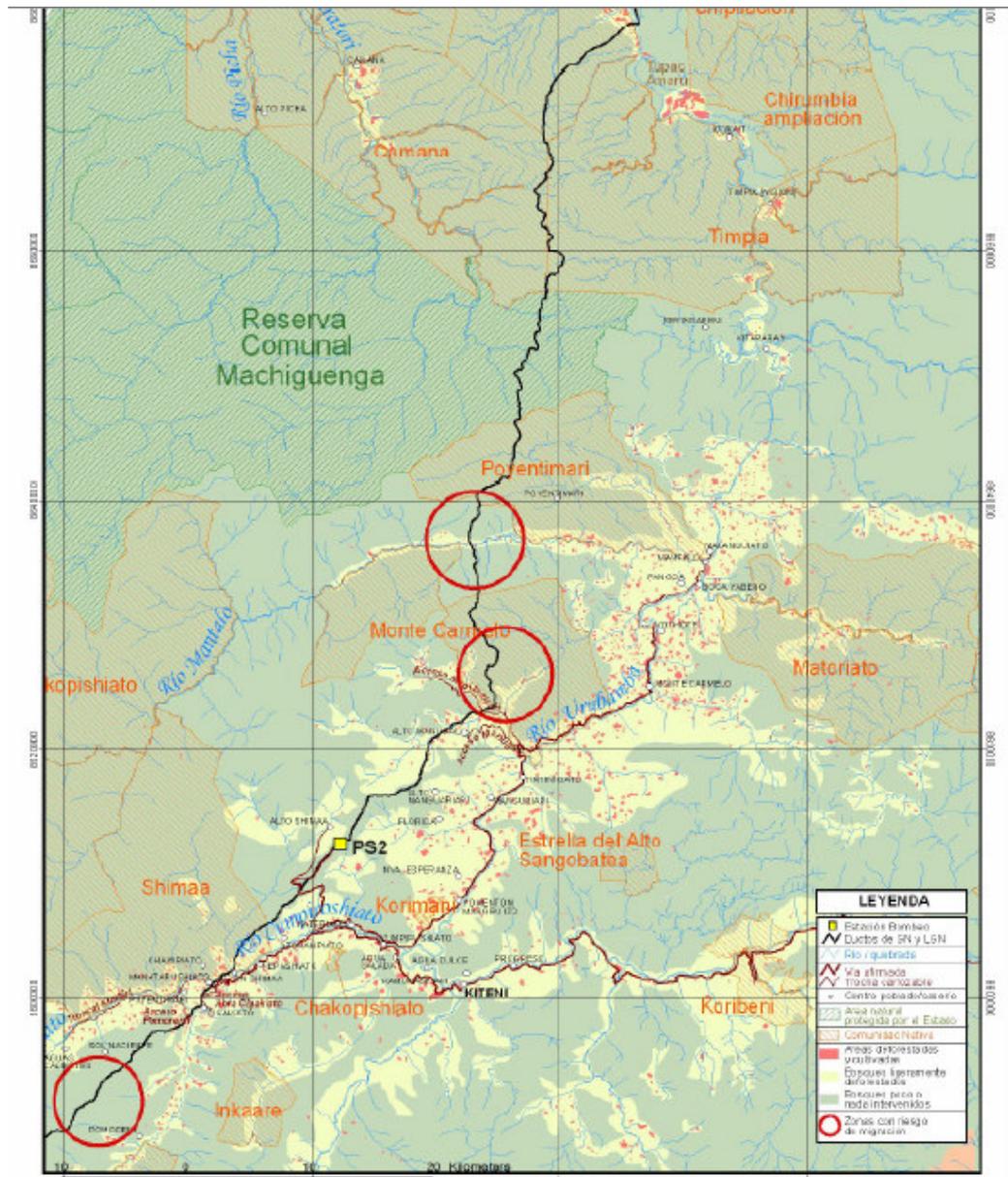


Figura 7 – Mapa da localização das Comunidades de Shimaa e Monte Carmelo. Fonte: <http://www.camisea.com.pe>.

Há aproximadamente 120 anos se iniciou o extrativismo de madeira na região, e, desde 1980, o território Machiguenga onde Shimaa e Monte Carmelo se encontram, faz parte de uma área de intenso extrativismo de gás natural, denominado Projeto *Camisea*, sob responsabilidade do governo Peruano (http://www.camisea.com.pe/fra_eia.htm), e motivo de conflitos constantes nesta região (<http://www.washingtonpost.com>). Diversas ONGs internacionais trabalham para a preservação ambiental da região que, segundo os moradores locais, tem sofrido impactos com as instalações do Projeto *Camisea* e alterado a subsistência das famílias.

II. Objetivos

1. Objetivo Geral

Testar a partir de novos dados do DNA mitocondrial a validade do modelo de estruturação genética das populações nativas da América do Sul proposto por Tarazona-Santos *et al.* (2001) (TS2001):

2. Objetivos Específicos

1.1 Gerar dados de sequenciamento da região hipervariável I do mtDNA de duas populações estrategicamente localizadas entres os Andes a Amazônia: Monte Carmelo e Shimaá.

1.2 Atualizar o banco de dados disponível no LDGH com sequências de DNAm_t de populações nativas peruanas e publicadas recentemente na literatura.

1.3 Estimar os parâmetros de diversidade genética das populações estudadas.

1.4 Avaliar a correlação de diversidade genética entre populações e distância geográfica testando a hipótese nula de Isolamento por Distância (*IBD*) nas regiões ocidentais e orientais da América do Sul.

1.5 Testar as seguintes hipóteses consistentes com o modelo *TS2001*:

- O modelo de *IBD* não se correlaciona com a variabilidade genética encontrada em toda a América do Sul, mas apenas quando as populações do Leste e do Oeste são analisadas separadamente.

- Caso o modelo *IBD* se ajustar no Leste e Oeste do continente, evidenciar que no Leste se observa um maior aumento da distância genética entre populações com o incremento da distância geográfica, quando comparado ao Oeste.

- Os menores índices de diversidade intrapopulacional e maiores de diversidade interpopulacional são encontrados no Leste, enquanto as populações localizadas no Oeste apresentam um padrão contrário.

III. Metodologia

1. Extração, quantificação e armazenagem do DNA

O DNA dos indivíduos pertencentes às populações de Monte Carmelo (n=28) e Shimaá (n=37) foi extraído de acordo com o protocolo padronizado e adaptado para o nosso laboratório do kit Gentra® Puregene® Blood Kit (Qiagen) após comparação com os métodos de extração de fenol/clorofórmio (<http://www.icb.ufmg.br/lbem>) e do Wizard® Genomic DNA Purification Kit da Promega® (dados não mostrados). As amostras foram quantificadas utilizando o espectrofotômetro Nanodrop® (Thermo Scientific) e armazenadas a -70°C na concentração de ~50 ng/μl. As amostras de sangue foram gentilmente doadas pelo Dr. Robert Gilman, da Universidade Cayetano Heredia situada em Lima, Peru (Anexo1).

2. Amplificação e Sequenciamento (Objetivo 1.1)

As amostras de DNA foram amplificadas com a utilização de iniciadores para a região hipervariável I (HVI) do DNAm (L15989 – H16410), adaptados de acordo com o protocolo desenvolvido por Alves-Silva *et al.* (2000) (Anexo2). Os produtos de PCR foram precipitados de acordo com o protocolo de polietilenoglicol (PEG) 8000 (<http://www.icb.ufmg.br/lbem>). As reações de sequenciamento foram realizadas utilizando os *primers* M13FWD (5'-TGTAACGACGGCCAGT-3') e M13REV (5'-CAGGAAACAGCTATGACC-3') e precipitadas conforme o protocolo descrito pelos kits *DYEnamic Et Dye Terminator MegaBACE®* (GE HealthCare) e *BigDye® Terminator v3.1 Cycle Sequencing* (Applied Biosystems). As sequências foram obtidas usando os sequenciadores automáticos *MegaBACE®* (GE Amersham) e *Applied Biosystems 3130*. Para cada indivíduo foram efetuados pelo menos dois sequenciamentos diretos e dois reversos. A análise das sequências foi feita utilizando o *pipeline* desenvolvido no LDGH (Anexo 3).

3. Determinação dos haplogrupos de DNAm

As sequências geradas foram alinhadas e comparadas com a Sequência Referência de Cambridge (Anderson *et al.*, 1981) utilizando o programa MEGA versão 4 (<http://www.megasoftware.net>). Variações entre posições 16182-16186 e o número de repetições da citosina homopolimérica nestas posições não foram consideradas. As amostras foram classificadas conforme seu padrão de polimorfismos em sítios que caracterizam os haplogrupos de DNAm (Merriwether *et al.*, 1995). As análises de RFLP para a confirmação dos haplogrupos não foi efetuada neste trabalho.

4. Coleta de dados da literatura

Foram selecionadas, junto com os dados de populações de Fuselli *et al.* (2003), os trabalhos publicados após o ano de 2003 sobre 29 populações nativas da América do Sul que utilizaram a região HVI do DNAm_t como marcador informativo. Descrições sobre as características de cada população utilizada neste estudo encontram-se detalhadas em suas respectivas publicações. As sequências de cada população foram gentilmente cedidas pelos respectivos autores, e as coordenadas geográficas, quando não disponibilizadas, foram inferidas a partir dos dados presentes na publicação. Embora erros de aproximação possam ser introduzidos com esse procedimento, estes não devem alterar o padrão observado no nível continental. As informações sobre as populações utilizadas neste trabalho encontram-se na Tabela 1 e a cobertura geográfica encontra-se na Figura 8. As populações foram classificadas segundo o critério de localização: pertencentes à região Oeste do continente (localizadas na região andina), pertencentes à região Leste do continente (localizadas na região Amazônica, Platô Central Brasileiro e Chacos) e pertencentes à região de transição aquelas que se encontram entre essas duas regiões.

5. Atualização do Banco de Dados (Objetivo 1.2)

As seqüências geradas, assim como as disponíveis na literatura, foram formatadas e incluídas no banco de dados desenvolvido por Sílvia Fuselli antes de 2003 e que continha as informações sobre as populações utilizadas em seu trabalho. Nesta atualização, foi considerada a região de 360 nucleotídeos localizada entre as posições 16024 e 16383 da região HVI do DNAm_t. Variações entre posições 16182-16186 e o número de repetições da citosina homopolimérica também foram desconsiderados em nossas análises.

Utilizando um *script* desenvolvido em linguagem *Perl* pelo Prof. Eduardo Tarazona, novos haplótipos foram automaticamente acrescentados ao banco de dados e haplótipos já existentes foram identificados. As frequências haplotípicas de cada população também foram computadas (Anexo 4).

6. Análises populacionais (Objetivo 1.3)

Calculamos diferentes estimadores do parâmetro $\theta = 2N_e\mu$ para comparação dos índices de diversidade intrapopulacional:

- θ_K (Ewens, 1972), baseado no número de haplótipos observados (κ);
- θ_S (Waterson, 1975), baseado no número de sítios segregantes observados (S);
- θ_π (Tajima, 1983), baseado no número de diferenças par-a-par entre seqüências (π).

Devido ao nosso interesse em detectar eventos demográficos acontecidos logo após o povoamento do continente (500 gerações anteriores), baseamos nossas comparações nos parâmetros de θ_K e θ_S , estimadores mais sensíveis a mudanças de tamanhos populacionais (Tajima, 1989a).

Para o teste de hipótese nula de tamanho constante contra a hipótese alternativa de gargalo/expansões populacionais acontecidas no passado, utilizamos as estatísticas D de Tajima (1989b) e F_s de Fu (1997), investigando quais populações apresentam uma frequência de alelos raros acima do esperado pelo equilíbrio entre mutação-deriva, o que seria sugestivo de expansão demográfica acontecida no passado.

As distâncias genéticas entre as populações foram calculadas para investigar a estrutura genética da diversidade do DNAm_t utilizando Φ_{st} , sob o modelo mutacional de Tamura-Nei (1993), com uma correção *gamma* para heterogeneidade de taxas mutacionais de $\alpha=0.26$ (Meyer *et al.*, 1999). Todas as análises descritas acima foram realizadas utilizando o programa Arlequin (Excoffier e Schneider, 2005).

7. Teste de modelo de Isolamento por Distância (Objetivo 1.4)

A análise de regressão foi utilizada para testar se o traço de dispersão de $\Phi_{st}/(1-\Phi_{st})$ *versus* o logarítmico da distância geográfica corresponde a uma função linear com inclinação positiva, como predito pelo modelo de *Isolamento por Distância* (IBD, Wright, 1943). Foi utilizado o Modelo II de regressão (Regressão da Redução do Eixo Maior; Sokal e Rohlf, 1995) com 5000 simulações, utilizando o programa ***IBD-Isolation by distance*** (Bohonak, 2002). As distâncias geográficas entre populações foram calculadas em quilômetros, a partir das coordenadas geográficas, utilizando o programa desenvolvido por Lucia Simoni (doutoranda em Antropologia Biológica pela Universidade de Bologna/Itália).

Tabela 1- Descrição das Populações Nativas da América do Sul utilizadas no presente estudo.

População	Número	Tamanho Amostral	País	Coordenadas Geográficas	Referência
Tayacaja	1	61	Peru	12°20'S, 75°50'W	 Fuselli et al. 2003
Arequipa	2	22		16°23'S, 71°31'W	
San Martin	3	22		07°00'S, 77°00' W	
Cayapa	4	120	Equador	00°00'S, 79°00'W	
Mapuche (1)	5	39	Argentina	40°00'S, 69°00' W	
Mapuche (2)	6	34	Chile	40°15'S, 72°25' W	
Pehuenche	7	24		37°43'S, 71°16' W	
Wai-Wai	8	26	Brasil	-01°35'S, 58°39'W	
Zoro	9	30		10°20'S, 60°20' W	
Gavião	10	27		10°10' S, 61°08' W	
Xavante	11	25		13°20'S, 51°40' W	
Suruí	12	24		10°50'S, 61°10' W	
Yanomama	13	129		-02°00'S, 64°00'W	
Yaghan	14	14		Antártica	
Aché	15	63	Paraguay	24°10'S, 56°30'W	
Ayoreo	16	90	Paraguay/ Bolívia	18°41'S, 60°19'W	 Dornelles et al. 2004
Ancash	17	33	Peru	09°01'S, 77°39'W	 Lewis et al. 2004
Puno	18	44		15°12'S, 69°58'W	
Tupe	19	17		10°50'S, 76°40'W	
Yungay	20	37		09°08'S, 77°44'W	
Ignaciano	21	15	Bolívia	13°33'S, 65°45'W	 Bert et al. 2004
Trinidadio	22	12		14°50'S, 65°54'W	
Movima	23	12		13°45'S, 65°23'W	
Yuracare	24	15		17°53'S, 64°22'W	
Guahibo	25	62	Venezuela	-05° 15'S, 67°20'W	 Vona et al. 2005
Mata Chaco	26	32	Argentina	24°60'S, 61°50'W	 Cabana et al. 2006
Mata Formosa	27	67		24°00' S, 62°20'W	
Toba Chaco	28	43		26°00'S, 60°60'W	
Toba Formosa	29	24		25°20'S, 58°30'W	
Pilagá	30	38		24°40' S, 59°60'W	
Guarani Kaiowá	31	120	Brasil	23°06'S, 55°12'W	 Marrero et al. 2007
Guarani	32	56		23°48'S, 54°30'W	
Ñandeva	33	24		25°18'S, 52°32'W	
Guarani M'byá	34	21		24°30'S, 54°10'W	
Kaygang PR	35	57		28°00'S, 53°00'W	
Kaygang RS	36	21		-11°00'S, 73°30'W	
Kogi	37	31	Colômbia	-10°26'S, 73°30'W	 Melton et al. 2007
Ijca	38	28		-10°40'S, 73°00'W	
Arsario	39	30		-12°00' S, 71°40'W	
Wayuú	40	10	Bolívia	14°04'S, 66°08'W	 Corella et al. 2007
Chimane	41	10		14°39'S, 66°58'W	
Moseten	42	10		14°33'S, 68°19'W	
Aymara	43	16		16°37'S, 68°01'W	
Quechua	44	37	Peru	12°33'S, 73°05'W	 Presente Trabalho
Shimaa	45	28		12°13'S, 72°58'W	



Figura 8- Distribuição das populações nativas da América do Sul utilizadas no presente estudo. A descrição das populações encontram-se resumidas na Tabela 1.

IV. Resultados

1. Análise da sequência HVI do DNAMt

A variação da sequência do DNAMt observada e os haplogrupos inferidos das populações de Shimaa e Monte Carmelo encontram-se resumidos na Figura 9, assim como frequências haplotípicas. Todos os indivíduos analisados pertencem aos haplogrupos autóctones A, B e C de populações nativas, e o haplogrupo D não foi encontrado em nossas análises. O haplogrupo B foi o mais frequente (89,2% nos Shimaa e 60,7% em Monte Carmelo), compatível com as frequências encontradas nas populações andinas, e as frequências do haplogrupo A e C variaram entre as populações (A: 2,7% e 32,1%; C: 8,1% e 7,2%, Shimaa e Monte Carmelo, respectivamente), sendo que apenas o sub-haplogrupo B2 foi encontrado. Dos dez haplótipos encontrados, seis (Apl0 915-920) estão representados unicamente nessas duas amostras da América do Sul, quando comparados aos haplótipos encontrados em outras populações do presente estudo.

	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Shimaa (n=37)	Monte Carmelo (n=28)
	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6		
	0	0	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3		
	5	9	1	2	8	1	2	3	6	9	9	0	1	2	2	6			
	1	3	1	9	9	7	3	9	6	0	8	3	9	5	7	2			
Anderson	A	T	C	G	T	T	C	C	C	C	T	G	G	T	C	T			
Apl0915-B2	C	C	A	7 (0.19)	4 (0.14)	
Apl0916-B2	G	.	.	A	C	C	18 (0.49)	12 (0.43)	
Apl0917-B2	G	.	.	A	C	C	.	.	.	G	1 (0.03)	1 (0.04)	
Apl0918-B2	G	C	.	A	C	C	1 (0.03)	-	
Apl0919-B2	G	.	.	A	C	C	A	.	.	.	1 (0.03)	-	
Apl0498-B2	C	C	5 (0.14)	-	
Apl0427-A2	.	.	T	.	.	.	T	A	T	T	.	.	A	.	.	C	1 (0.03)	6 (0.21)	
Apl0521-A	C	.	T	.	.	T	.	.	A	.	.	C	-	3 (0.11)	
Apl0531-C1	G	T	.	.	.	C	.	.	C	T	.	3 (0.08)	1 (0.04)	
Apl0920-C1	.	C	.	A	.	.	T	.	.	.	C	.	.	C	T	.	-	1 (0.04)	

Figura 9 – Haplótipos e haplogrupos observados nas populações de Monte Carmelo e Shimaa.

2. Atualização do Banco de Dados

A atualização da base de dados nos permitiu um acréscimo de 1103 indivíduos de 31 populações nativas da América do Sul (comparado aos 597 indivíduos de 14 populações disponíveis anteriormente no trabalho de Fuselli *et al.*, 2003), o que nos permitiu obter uma ampla cobertura geográfica do continente (Tabela 1, Figura 8).

3. Diversidade das populações Nativas da América do Sul

As populações de Shimaa e Monte Carmelo apresentaram valores de θ_K , θ_S e θ_π semelhantes aos encontrados nas populações localizadas no Leste do continente (Tabela 2). Como previsto pelo modelo TS2001, as populações localizadas no Oeste da América do Sul apresentaram os maiores valores de θ encontrados no continente, quando comparado com as populações do Leste. Várias das populações localizadas na região de transição apresentaram valores de θ que se assemelham mais aos encontrados nas populações do Oeste andino, ao contrário do observado nas populações Shimaa e Monte Carmelo. Um excesso de alelos raros determinando valores negativos significativos de D de Tajima e/ou F_s de Fu foram encontrados apenas nas populações do Oeste do continente e em algumas populações consideradas de transição.

Tabela 2- Valores dos três estimadores de θ , seus intervalos de confiança (95% IC) ou desvios padrões (DP), e valores de testes de Neutralidade D de Tajima e F_s de Fu. As populações estão organizadas pelos valores decrescentes de θ_K . As populações em branco são as consideradas da região Oeste do continente, em cinza as do Leste, e em azul as de transição.

População	Localização	N	θ_K (95% IC)	θ_s (DP)	θ_π (DP)	D Tajima	F_s Fu
Tayacaja	1	61	69.4 (40.8-120.3)	11.11 (3.28)	6,20 (3,31)	-1,44*	-25,22**
Ancash	17	33	67.0 (31.0-153.8)	9.61 (3.22)	5,71 (3,12)	-1,47*	-20,11**
Trinitário	22	12	58.4 (14.4-252.4)	6.95 (3.00)	6,39 (3,67)	-0,35	-4,46**
Puno	18	44	45.3 (24.7-85.2)	10.34 (3.26)	5,17 (2,83)	-1,75*	-22,92**
Arequipa	2	22	43.9 (17.6-117.4)	7.13 (2.68)	5,12 (2,88)	-1,07	-10,60**
Moseten	41	10	38.8 (9.4-169.1)	6.36 (2.90)	6,36 (3,72)	-0,01	-2,83*
Ignaciano	21	15	25.7 (9.1-78.0)	6.77 (2.78)	6,23 (3,51)	-0,33	-3,78*
Pilagá	30	38	23.7 (12.7-44.8)	7.85 (2.63)	6,40 (3,44)	-0,65	-8,51**
San Martin	3	22	19.5 (8.6-45.4)	6.31 (2.41)	5,23 (2,93)	-0,64	-5,30**
Yurucare	24	15	17.1 (6.4-47.8)	5.84 (2.45)	5,94 (3,37)	0,07	-2,65
Yungay	20	37	17.0 (9.0-32.1)	7.19 (2.44)	5,25 (2,89)	-0,94	-6,87**
Chimane	40	10	16.4 (4.9-59.0)	5.66 (2.61)	5,13 (3,07)	-0,43	-1,92
Yanomama	13	129	13.3 (8.7- 20.1)	5.71(1.64)	4,79(2,61)	-0,47	-10,64**
MataFormosa	27	67	12.0 (7.1-19.9)	6.70 (2.08)	4,85 (2,66)	-0,89	-6,24*
Pehuenche	7	24	10.8 (5.0-23.4)	5.89 (2.23)	5,91 (3,26)	0,01	-1,94
Movima	23	12	9.3 (3.3-27.3)	3.97 (1.85)	2,94 (1,86)	-1,09	-2,63*
Mapuche Arg	5	39	8.4 (4.4-16.0)	4.73 (1.70)	5,39 (2,95)	0,46	-1,76
Tupe	19	17	7.0 (2.9-17.1)	5.32 (2.20)	5,24 (2,98)	-0,07	-0,48
TobaFormosa	29	24	5.9 (2.7-12.9)	5.09 (1.97)	4,80 (2,71)	-0,2	-0,38
Quechua	43	16	5.7 (2.3-14.2)	5.42 (2.26)	4,08 (2,40)	-0,99	-0,56
Ayoreo	16	90	4.9 (2.7-8.5)	3.55 (1.19)	2,44 (1,48)	-0,91	-3,57
Guahibo	25	62	4.7 (2.5-8.7)	2.77 (1.04)	4,92 (2,69)	2,25	0,42
Wai-Wai	8	26	4.4 (2.0-9.7)	6.29 (2.33)	6,35 (3,47)	0,04	1,66
TobaChaco	28	43	4.4 (2.2-8.7)	5.55 (1.91)	4,94 (2,72)	-0,37	0,7
Zoró	9	30	4.0 (1.8-8.5)	4.29 (1.64)	3,98 (2,28)	-0,25	0,31
Mapuche Chile	6	34	3.7 (1.7-7.7)	4.16 (1.57)	5,53 (3,03)	1,11	1,97
Cayapa	4	120	3.5 (1.9-6.2)	3.73 (1.19)	5,01 (2,71)	0,97	2,12
Shimaa	44	37	3.5 (1.6-7.2)	3.59 (1.37)	2,36 (1,46)	-1,11	-1,11
Yaghan	14	14	3.4 (1.2-9.1)	4.09 (1.83)	4,73 (2,76)	0,63	1,36
Aymara	42	10	3.3 (1.1-10.0)	3.18 (1.61)	1,80 (1,28)	-1,90**	-0,84
Kaingang RS	35	57	3.3 (1.6-6.4)	4.55 (1.55)	6,12 (3,28)	1,08	3,4
Gavião	10	27	2.7 (1.1-6.2)	4.15 (1.63)	4,28 (2,44)	0,11	1,94
Monte Carmelo	45	28	2.7 (1.1-6.1)	4.11 (1.61)	5,60 (3,09)	1,25	3,26
MataChaco	26	32	2.5 (1.0-5.5)	4.22 (1.60)	5,59 (3,07)	1,1	3,74
Wayuú	39	30	2.0 (0.8-4.6)	4.04 (1.56)	5,26 (2,91)	1,03	4,37
Nandeva	32	56	1.9 (0.8-4.1)	2.39 (0.95)	3,47 (1,99)	1,28	2,78
Kaingang PR	34	21	1.2 (0.4-3.3)	3.34 (1.43)	5,09 (2,87)	1,86	5,9
Kaiowá	31	120	1.2 (0.5-2.6)	1.87 (0.72)	1,57 (1,05)	-0,39	1,33
Xavante	11	25	1.1 (0.4-3.0)	2.65 (1.16)	3,00 (1,81)	0,44	3,72
Surui	12	24	1.1 (0.4-3.0)	1.87 (0.90)	1,82 (1,21)	-0,1	1,78
Arsário	38	28	1.0 (0.3-2.8)	2.57 (1.11)	4,15 (2,37)	1,98	5,74
Kogi	36	21	0.7 (0.2-2.2)	2.78 (1.24)	3,25 (1,95)	0,58	5,4
M'byá	33	24	0.7 (0.2-2.1)	1.34 (0.71)	2,48 (1,55)	2,47	4,43
Ijka	37	31	0.6 (0.2-1.8)	2.75 (1.15)	1,47 (1,02)	-1,49	2,81
Aché	15	63	0.5 (0.1-1.4)	1.49 (0.67)	1,08 (0,80)	-0,69	2,57

*P<0.05; ** P<0.01.

4. Teste da Hipótese Nula de Modelo de Isolamento por Distância

Quando todas as populações da América do sul são analisadas conjuntamente, não se observa correlação entre distância genética e distância geográfica (Figura 10). Esses resultados sugerem que um modelo simples como o IBD não explica uma parte significativa do padrão espacial de diversidade genética de DNAmT na América do Sul. Entretanto, quando as populações são divididas em dois grupos, Leste e Oeste, uma estrutura geográfica aparece nas populações do Oeste (Figura 11A): o logarítmico da distância geográfica explica 16% da diversidade genética encontrada nessa região, se adaptando ao modelo de *IBD*. Ao contrário do observado anteriormente por Fuselli *et al.* (2003), a distância geográfica não é capaz de explicar a diversidade genética observada entre as populações do Leste do continente com o acréscimo de novas populações, sugerindo padrões evolutivos mais complexos nessa região (Figura 11B).

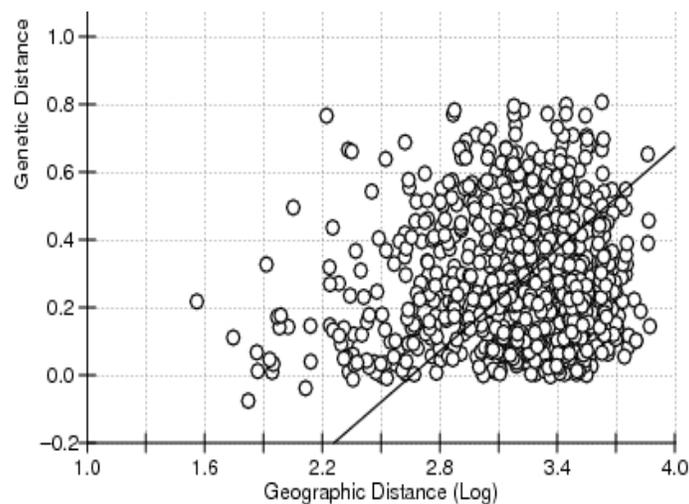


Figura 10- Regressão do eixo maior reduzido das distâncias genéticas versus o logarítmico da distância geográfica de todas as populações da América do Sul. As populações, quando analisadas conjuntamente, não se adaptam ao Modelo de IBD.

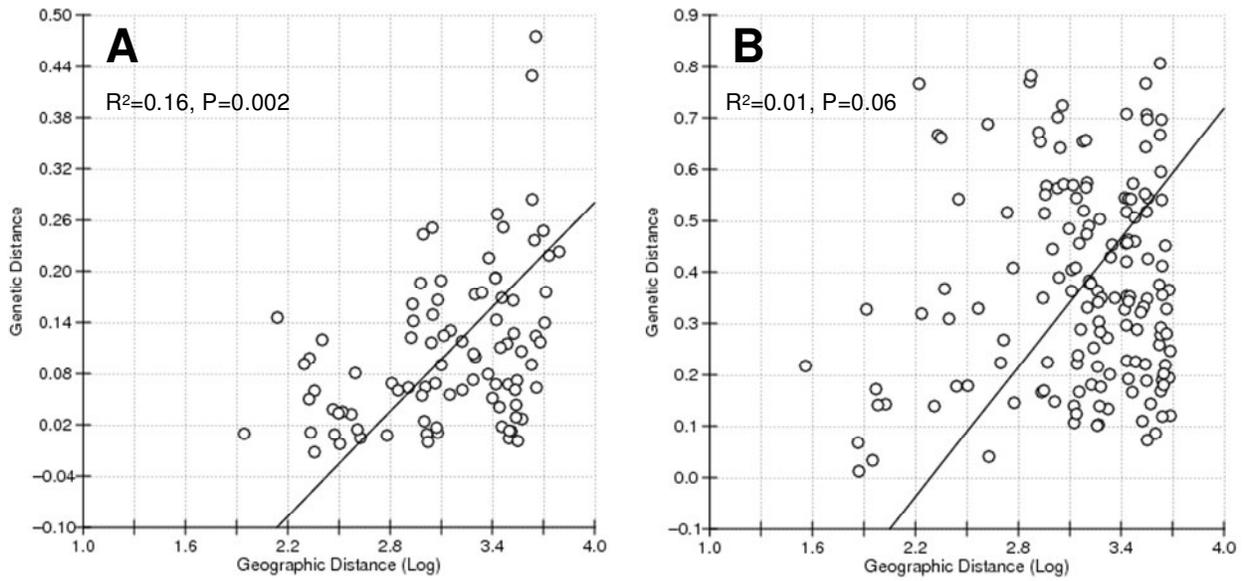


Figura 11- Regressão do eixo maior reduzido das distâncias genéticas *versus* o logarítmico da distância geográfica das populações do oeste (A) e leste (B) da América do Sul. A distância geográfica explica 16% da diversidade genética das populações do Oeste do continente, aceitando a hipótese nula de IBD. Com o acréscimo do número de populações do leste, a hipótese nula de IBD é rejeitada.

V. Discussão

Os estimadores de θ baseados no número de diferentes haplótipos (θ_K) e sítios segregantes (θ_S) mostram que as populações andinas e a maioria das populações da região de transição apresentam maiores valores de diversidade intra-populacional entre as populações da América do Sul (Tabela 2). Além disso, estas populações tendem a apresentar um excesso de alelos raros (com valores significativos) no que diz respeito à expectativa de equilíbrio entre mutação e deriva, enquanto as populações localizadas no Leste do continente (com exceção dos Yanomama) não apresentam. Dois fatores podem contribuir para este padrão: expansão populacional (Slatkin e Hudson, 1991) e seleção natural.

Bonato e Salzano (1997) demonstraram que o povoamento das Américas foi seguido por uma expansão demográfica, cujos sinais devem estar presentes nos nativos americanos modernos caso essas populações não tenham sofrido eventos recentes de *bottlenecks* típicos de caçadores-coletores (Excoffier e Scheneider, 1999). De fato, evidências arqueológicas e históricas demonstram que a região andina abrigou as maiores populações existentes na América do Sul, pelo menos desde o desenvolvimento de sociedades complexas, enquanto as populações do Leste do continente foram mais fragmentadas, tiveram menor tamanho censitário e adquiriram menor desenvolvimento socioeconômico (Salzano e Callegari-Jacques, 1988).

Embora existam evidências de que a seleção natural atue na variação encontrada no DNAmT em humanos (Mishmar *et al.*, 2003), dados de variabilidade do cromossomo Y apontaram para um padrão semelhante de diversidade. Dados recentes de microssatélites também indicam altas taxas de variabilidade nas populações do Oeste do continente (Wang *et al.*, 2007), excluindo-se a hipótese de que os valores negativos significativos de D de Tajima e F_s de Fu sejam devidos unicamente à ação da seleção natural, e sugerem que, apesar do stress causado por ambientes de alta altitude, a população andina manteve grande tamanho efetivo populacional e altas taxas de fluxo gênico (Tarazona-Santos *et al.*, 2001).

Em geral, populações consideradas de transição entre o Leste e o Oeste da América do Sul apresentaram um padrão de diversidade similar ao encontrado nas populações da região Oeste do continente (Figura 12). Trabalhos recentes apontam

diversos fatores como responsáveis para a variação encontrada nas populações estudadas. Em particular, nas populações da região do Grande Chaco na Argentina (Cabana *et al.*, 2006), os autores sugerem que o padrão de alta diversidade encontrado pode ser atribuído ao fato da região ter historicamente abrigado rotas inter-regionais de fluxo gênico, associados a eventos evolutivos como o povoamento inicial da região. Nas populações de Llanos de Moxos na Bolívia (Bert *et al.*, 2004), a alta diversidade genética encontrada foi atribuída à ausência de *bottlenecks* e processos contínuos de fluxo gênico, sendo que a influencia genética das populações andinas foi refutada como fator relevante.

Por outro lado, As populações de Shima e Monte Carmelo, apresentaram padrões de diversidade semelhantes aos das populações do Leste do continente (θ_K : 3.5 [1.6-7.2] e 2.7 [1.1-61] respectivamente). Evidências históricas dos Machiguengas sugerem que o contato com os Incas se limitou a relações comerciais (Misioneros Dominicanos, 2006) e o fato dessas populações estarem geograficamente isoladas na selva peruana, dificultando o contato com outras populações, podem ter contribuído para o padrão de diversidade genética observado.

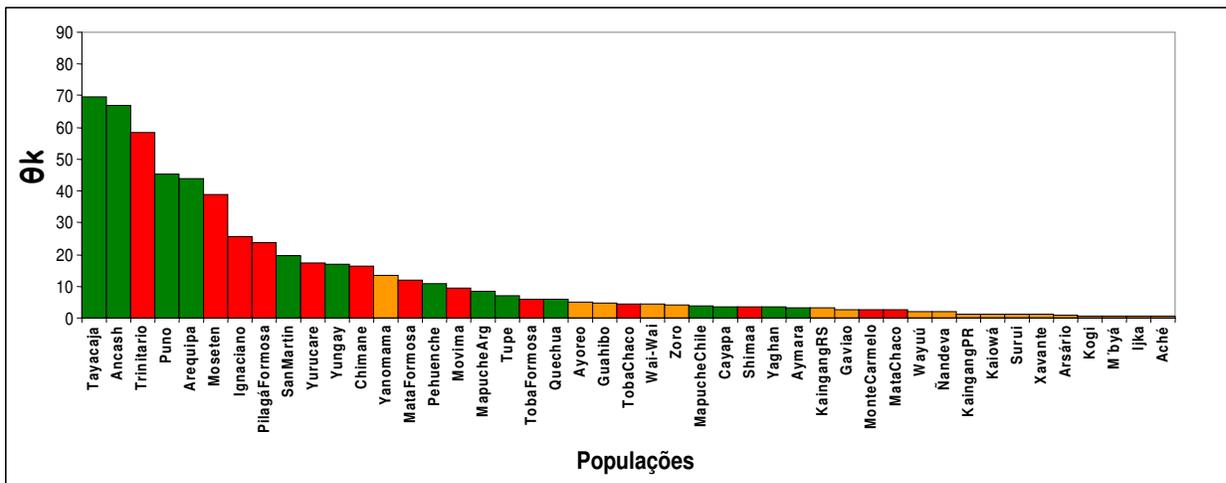


Figura 12– Gráfico dos valores decrescentes de θ_K observados em nossas análises. As populações em verde representam as do Oeste do continente, as amarelas do Leste e vermelhas as consideradas nas zonas de transição.

Análises de regressão da diversidade do DNAm sugerem que apenas as populações do Oeste do continente apresentam um padrão de isolamento por distância (Figuras 10 e 11). Embora Fuselli *et al.* (2003) tenha observado este padrão também na região Leste do continente, o incremento do tamanho amostral nesta região mostrou que apenas a distância geográfica não é capaz de explicar o padrão encontrado.

Uma possibilidade para ausência de correlação entre diversidade genética e distância geográfica pode ser atribuída a um padrão de maior diversidade encontrado entre as populações da parte Leste do continente sul-americano. De fato, Lewis e Long (2008) mostraram que as populações do Leste apresentavam um nível de variação maior do que se esperava, não permitindo que essas populações fossem agrupadas de forma coesa em suas análises. Porém, trabalhos recentes de populações nativas Americanas utilizando grandes quantidades de marcadores autossômicos (Wang *et al.*, 2007) e variáveis craniométricas (Pucciarelli *et al.*, 2006) demonstraram que as populações do Leste apresentam os menores valores de diversidade encontrados na América do Sul, corroborando o que foi proposto anteriormente pelo nosso grupo (Tarazona-Santos *et al.*, 2001; Fuselli *et al.*, 2003). De forma similar, nossos resultados também demonstram esse padrão de diversidade na parte Leste do continente.

A explicação mais plausível para a ausência de correlação entre diversidade genética e distância geográfica nas populações do Leste do continente sul-americano seria a de que diferentes fatores evolutivos, como *bottlenecks*, deriva e fluxo gênico; históricos, como o isolamento das populações na selva amazônica; ou comportamentais e linguísticos possam ter contribuído para uma alta divergência genética encontrada entre as populações nesta parte do continente. De fato, Marrero *et al.* (2007) demonstraram que, embora as populações Guarani e Kaingang estejam localizadas próximas uma a outra, elas exibem alta divergência genética e, eventos como *bottlenecks*, o padrão comportamental nômade dos Guaranis, e fatores linguísticos, seriam responsáveis por este fenômeno.

Nossos dados sugerem que um modelo simplista como o de IBD na parte Leste da América do Sul não é capaz de explicar a diversidade encontrada nessas

populações e que padrões mais complexos de fatores evolutivos, comportamentais, lingüísticos e históricos atuaram nesta parte do continente. Novas evidências estão surgindo para tentar elucidar o padrão de diversidade encontrado: a descoberta de sítios arqueológicos recentes na região Amazônica sugere que populações mais complexas se desenvolveram nessa parte do continente, propondo que maiores tamanhos populacionais e o desenvolvimento de tecnologias possam ter contribuído para a história evolutiva dessas populações (Heckenberger *et al.*, 2008), e trazendo novas perspectivas para a compreensão do povoamento da América do Sul.

VI. Conclusão e Perspectivas

Nosso trabalho utilizou dados de DNAm_t e um tamanho amostral significativo para testar o Modelo de ocupação da América do Sul proposto por Tarazona-Santos *et al.* (2001). Nossos dados corroboram o padrão encontrado no continente sul-americano: as populações do Oeste apresentam altas taxas de diversidade comparadas às populações situadas no Leste. Valores negativos significativos de D de Tajima e F_s de Fu encontrados nas populações do Oeste do continente também sugerem expansão populacional, corroborando os dados históricos dessas populações.

Em geral, populações localizadas na região de transição entre as duas áreas apresentaram um padrão semelhante ao encontrado nas populações Andinas. Entretanto, nosso trabalho também revelou o padrão de diversidade de duas populações nativas de etnia Machiguenga (Shimaa e Monte Carmelo), localizadas na selva peruana, e demonstrou que, apesar de suas localizações geográficas próximas as populações andinas, essas apresentaram uma baixa diversidade genética, o que as torna mais similar ao que é observado nas populações do Leste de continente.

Embora a distância geográfica esteja correlacionada à diversidade genética nas populações do Oeste, esse padrão não foi observado nas populações do Leste do continente, sugerindo que padrões evolutivos mais complexos atuaram nessas últimas, e que a distância geográfica não é capaz de explicar o padrão observado.

Não é surpreendente que existam incertezas a respeito das origens e estruturação das populações nativas da América do Sul, considerando que padrões complexos de variação genética dificultam a elaboração de modelos com significância estatística (Lewis e Long, 2008). Futuras análises estatísticas mais robustas e modelos mais flexíveis que considerem diferentes fatores evolutivos e correlacionem diferentes dados, como linguísticos e craniométricos, podem contribuir para elucidar os padrões observados.

Referências Bibliográficas

- Alves-Silva, J., M. D. Santos, P. E. M. Guimaraes, A. C. S. Ferreira, H. J. Bandelt, S. D. J. Pena e V. F. Prado. The ancestry of Brazilian mtDNA lineages. American Journal of Human Genetics, v.67, n.2, Aug, p.444-461. 2000.
- Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. Debruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden e I. G. Young. Sequence and Organization of the Human Mitochondrial Genome. Nature, v.290, n.5806, p.457-465. 1981.
- Battilana, J., L. Cardoso-Silva, R. Barrantes, K. Hill, A. M. Hurtado, F. M. Salzano e S. L. Bonatto. Molecular variability of the 16p13.3 region in Amerindians and its anthropological significance. Annals of Human Genetics, v.71, Jan, p.64-76. 2007.
- Bert, F., A. Corella, M. Gene, A. Perez-Perez e D. Turbon. Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. Annals of Human Biology, v.31, n.1, Jan-Feb, p.9-28. 2004.
- Bohonak, A. J. IBD (isolation by distance): A program for analyses of isolation by distance. Journal of Heredity, v.93, n.2, Mar-Apr, p.153-154. 2002.
- Bonatto, S. L. e F. M. Salzano. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the new world. American Journal of Human Genetics, v.61, n.6, Dec, p.1413-1423. 1997.
- Bortolini, M. C., F. M. Salzano, M. G. Thomas, S. Stuart, S. P. K. Nasanen, C. H. D. Bau, M. H. Hutz, Z. Layrisse, M. L. Petzl-Erler, L. T. Tsuneto, K. Hill, A. M. Hurtado, D. Castro-De-Guerra, M. M. Torres, H. Groot, R. Michalski, P. Nymadawa, G. Bedoya, N. Bradman, D. Labuda e A. Ruiz-Linares. Y-chromosome evidence for differing ancient demographic histories in the Americas. American Journal of Human Genetics, v.73, n.3, Sep, p.524-539. 2003.
- Cabana, G. S., D. A. Merriwether, K. Hunley e D. A. Demarchi. Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. American Journal of Physical Anthropology, v.131, n.1, Sep, p.108-119. 2006.
- Cavalli-Sforza, L. L., Menozzi, P., Piazza, A. The history and geography of human genes. Princeton, New Jersey: Princeton University Press, 1994. 1088p.
- Corella, A., F. Bert, A. Perez-Perez, M. Gene e D. Turbon. Mitochondrial DNA diversity of the Amerindian populations living in the Andean Piedmont of Bolivia: Chimane, Mosen, Aymara and Quechua. Annals of Human Biology, v.34, n.1, Jan-Feb, p.34-55. 2007.
- Dornelles, C. L., J. Battilana, N. J. R. Fagundes, L. B. Freitas, S. L. Bonatto e F. M. Salzano. Mitochondrial DNA and Alu insertions in a genetically peculiar population: The ayoreo Indians of Bolivia and Paraguay. American Journal of Human Biology, v.16, n.4, Jul-Aug, p.479-488. 2004.
- Ewens, W. J. Sampling Theory of Selectively Neutral Alleles. Theoretical Population Biology, v.3, n.1, p.87-&. 1972.
- Excoffier, L. e S. Schneider. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. Proceedings of the National Academy of Sciences of the United States of America, v.96, n.19, Sep 14, p.10597-10602. 1999.
- Excoffier, L. e Schneider, S. Arlequin v.3.0: An integrated software package for population genetics analyses. Evolutionary Bioinformatics online, v.1, p.47-50. 2005.
- Fagundes, N. J. R., R. Kanitz, R. Eckert, A. C. S. Valls, M. R. Bogo, F. M. Salzano, D. G. Smith, W. A. Silva, M. A. Zago, A. K. Ribeiro-Dos-Santos, S. E. B. Santos, M. L. Petzl-Erler e S. L. Bonatto. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. American Journal of Human Genetics, v.82, n.3, Mar, p.583-592. 2008.

- Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics, v.147, n.2, Oct, p.915-925. 1997.
- Fuselli, S., E. Tarazona-Santos, I. Dupanloup, A. Soto, D. Luiselli e D. Pettener. Mitochondrial DNA diversity in south America and the genetic history of Andean highlanders. Molecular Biology and Evolution, v.20, n.10, Oct, p.1682-1691. 2003.
- Garcia, F., M. Moraga, S. Vera, H. Henriquez, E. Llop, E. Aspillaga e F. Rothhammer. mtDNA microevolution in Southern Chile's archipelagos. American Journal of Physical Anthropology, v.129, n.3, Mar, p.473-481. 2006.
- Gonzalez-Jose, R., W. Neves, M. M. Lahr, S. Gonzalez, H. Pucciarelli, M. H. Martinez e G. Correa. Late Pleistocene/Holocene craniofacial morphology in mesoamerican paleoindians: Implications for the peopling of the new world. American Journal of Physical Anthropology, v.128, n.4, Dec, p.772-780. 2005.
- Gonzalez-Jose, R., Bortolini, M.C., Santos, F.R., Bonatto, S.L. The peopling of America: Craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. American Journal of Physical Anthropology, v. 137, n.2, May, p.175-187. 2008.
- Greenberg, J. H., C. G. Turner e S. L. Zegura. The Settlement of the America - a Comparison of the Linguistic, Dental, and Genetic-Evidence. Current Anthropology, v.27, n.5, Dec, p.477-497. 1986.
- Heckenberger, M. J., J. C. Russell, C. Fausto, J. R. Toney, M. J. Schmidt, E. Pereira, B. Franchetto e A. Kuikuro. Pre-columbian urbanism, anthropogenic landscapes, and the future of the Amazon. Science, v.321, n.5893, Aug 29, p.1214-1217. 2008.
- Hunley, K. L., G. S. Cabana, D. A. Merriwether e J. C. Long. A formal test of linguistic and genetic coevolution in Native Central and South America. American Journal of Physical Anthropology, v.132, n.4, Apr, p.622-631. 2007.
- Lewis, C. M. e J. C. Long. Native south american genetic structure and prehistory inferred from hierarchical modeling of mtDNA. Molecular Biology and Evolution, v.25, n.3, Mar, p.478-486. 2008.
- Lewis, C. M., R. Y. Tito, B. Lizarraga e A. C. Stone. Land, language, and loci: mtDNA in native Americans and the genetic history of Peru. American Journal of Physical Anthropology, v.127, n.3, Jul, p.351-360. 2005.
- Marrero, A. R., W. A. Silva-Junior, C. M. Bravi, M. H. Hutz, M. L. Petzl-Erler, A. Ruiz-Linares, F. M. Salzano e M. C. Bortolini. Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. American Journal of Physical Anthropology, v.132, n.2, Feb, p.301-310. 2007.
- Melton, P. E., I. Briceno, A. Gomez, E. J. Devor, J. E. Bernal e M. H. Crawford. Biological relationship between central and south American Chibchan speaking populations: Evidence from mtDNA. American Journal of Physical Anthropology, v.133, n.1, May, p.753-770. 2007.
- Merriwether, D. A., F. Rothhammer e R. E. Ferrell. Distribution of the 4 Founding Lineage Haplotypes in Native-Americans Suggests a Single Wave of Migration for the New-World. American Journal of Physical Anthropology, v.98, n.4, Dec, p.411-430. 1995.
- Meyer, S., G. Weiss e A. Von Haeseler. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics, v.152, n.3, Jul, p.1103-1110. 1999.
- Mishmar, D., E. Ruiz-Pesini, P. Golik, V. Macaulay, A. G. Clark, S. Hosseini, M. Brandon, K. Easley, E. Chen, M. D. Brown, R. I. Sukernik, A. Olckers e D. C. Wallace. Natural selection shaped regional mtDNA variation in humans. Proceedings of the National Academy of Sciences of the United States of America, v.100, n.1, Jan 7, p.171-176. 2003.
- Misioneros Dominicanos. La vida del pueblo Matsiguenga. Lima: Centro Cultural José Pío Aza. 2006. 880p.

- Moraga, M., C. M. Santoro, V. G. Standen, P. Carvallo e F. Rothhammer. Microevolution in prehistoric Andean populations: Chronologic mtDNA variation in the desert valleys of northern Chile. American Journal of Physical Anthropology, v.127, n.2, Jun, p.170-181. 2005.
- Mulligan, C. J., K. Hunley, S. Cole e J. C. Long. Population genetics, history, and health patterns in native Americans. Annual Review of Genomics and Human Genetics, v.5, p.295-315. 2004.
- Neves, W. A., A. Prous, R. Gonzalez-Jose, R. Kipnis e J. Powell. Early Holocene human skeletal remains from Santana do Riacho, Brazil: implications for the settlement of the New World. Journal of Human Evolution, v.45, n.1, Jul, p.19-42. 2003.
- Pena, S. D. J., F. R. Santos, N. O. Bianchi, C. M. Bravi, F. R. Carnese, F. Rothhammer, T. Gerelsaikhan, B. Munkhtuja e T. Oyunsuren. A Major Founder Y-Chromosome Haplotype in Amerindians. Nature Genetics, v.11, n.1, Sep, p.15-16. 1995.
- Pucciarelli, H. M., W. A. Neves, R. Gonzalez-Jose, M. L. Sardi, F. R. Rozzi, A. Struck e M. Y. Bonilla. East-West cranial differentiation in pre-Columbian human populations of South America. Homo-Journal of Comparative Human Biology, v.57, n.2, p.133-150. 2006.
- Santos, F. R., M. H. Hutz, C. E. A. Coimbra, R. V. Santos, F. M. Salzano e S. D. J. Pena. Further evidence for the existence of a major founder Y-chromosome haplotype in Amerindians. Brazilian Journal of Genetics, v.18, n.4, Dec, p.669-672. 1995.
- Santos, F. R., A. Pandya, C. Tyler-Smith, S. D. J. Pena, M. Schanfield, W. R. Leonard, L. Osipova, M. H. Crawford e R. J. Mitchell. The central Siberian origin for Native American Y chromosomes. American Journal of Human Genetics, v.64, n.2, Feb, p.619-628. 1999.
- Santos, F. R., L. Rodriguezdelfin, S. D. J. Pena, J. Moore e K. M. Weiss. North and South Amerindians may have the same major founder Y chromosome haplotype. American Journal of Human Genetics, v.58, n.6, Jun, p.1369-1370. 1996.
- Salzano, F. M., and S. Callegari-Jacques. South American Indians: a case study in evolution. Oxford: Clarendon Press. 1988. p.272.
- Schmitt, R., S. L. Bonatto, L. B. Freitas, V. C. Muschner, K. Hill, A. M. Hurtado e F. M. Salzano. Extremely limited mitochondrial DNA variability among the Ache Natives of Paraguay. Annals of Human Biology, v.31, n.1, Jan-Feb, p.87-94. 2004.
- Schurr, T. G. e S. T. Sherry. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: Evolutionary and demographic evidence. American Journal of Human Biology, v.16, n.4, Jul-Aug, p.420-439. 2004.
- Shinoda, K., N. Adachi, S. Guillen e I. Shimada. Mitochondrial DNA analysis of ancient Peruvian highlanders. American Journal of Physical Anthropology, v.131, n.1, Sep, p.98-107. 2006.
- Silva, W. A., S. L. Bonatto, A. J. Holanda, A. K. Ribeiro-Dos-Santos, B. M. Paixao, G. H. Goldman, K. Abe-Sandes, L. Rodriguez-Delfin, M. Barbosa, M. L. Paco-Larson, M. L. Petzl-Erler, V. Valente, S. E. B. Santos e M. A. Zago. Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. American Journal of Human Genetics, v.71, n.1, Jul, p.187-192. 2002.
- Slatkin, M. e R. R. Hudson. Pairwise Comparisons of Mitochondrial-DNA Sequences in Stable and Exponentially Growing Populations. Genetics, v.129, n.2, Oct, p.555-562. 1991.
- Tajima, F. Evolutionary Relationship of DNA-Sequences in Finite Populations. Genetics, v.105, n.2, p.437-460. 1983.
- Tajima, F. The Effect of Change in Population-Size on DNA Polymorphism. Genetics, v.123, n.3, Nov, p.597-601. 1989.

- Tajima, F. Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics, v.123, n.3, Nov, p.585-595. 1989.
- Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D.G., Mulligan, C.J., Bravi, C.M., Rickards, O., Martinez-Labarga, C., Khusnutdinova, E.K., Fedorova, S.A., Golubenko, M.V., Stepanov, V.A., Gubina, M.A., Zhadanov, S.I., Ossipova, L.P., Damba, L., Voevoda, M.I., Dipierri, J.E., Villems, R., Malhi, R.S. Beringian standstill and spread of Native American founders. PLoS ONE, v.2, n.9, Sep, p.829. 2007.
- Tamura, K. e M. Nei. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-DNA in Humans and Chimpanzees. Molecular Biology and Evolution, v.10, n.3, May, p.512-526. 1993.
- Tarazona-Santos, E., D. R. Carvalho-Silva, D. Pettener, D. Luiselli, G. F. De Stefano, C. M. Labarga, O. Rickards, C. Tyler-Smith, S. D. J. Pena e F. R. Santos. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: Evidence from the Y chromosome. American Journal of Human Genetics, v.68, n.6, Jun, p.1485-1496. 2001.
- Tarazona-Santos, E. e F. R. Santos. The peopling of the Americas: A second major migration? American Journal of Human Genetics, v.70, n.5, May, p.1377-1380. 2002.
- Torrioni, A., T. G. Schurr, C. C. Yang, E. J. E. Szathmary, R. C. Williams, M. S. Schanfield, G. A. Troup, W. C. Knowler, D. N. Lawrence, K. M. Weiss e D. C. Wallace. Native-American Mitochondrial-DNA Analysis Indicates That the Amerind and the Nadene Populations Were Founded by 2 Independent Migrations. Genetics, v.130, n.1, Jan, p.153-162. 1992.
- Vona, G., A. Falchi, P. Moral, C. M. Calo e L. Varesi. Mitochondrial sequence variation in the Guahibo Amerindian population from Venezuela. American Journal of Physical Anthropology, v.127, n.3, Jul, p.361-369. 2005.
- Wang, S., C. M. Lewis, M. Jakobsson, S. Ramachandran, N. Ray, G. Bedoya, W. Rojas, M. V. Parra, J. A. Molina, C. Gallo, G. Mazzotti, G. Poletti, K. Hill, A. M. Hurtado, D. Labuda, W. Klitz, R. Barrantes, M. C. Bortolini, F. M. Salzano, M. L. Petzl-Erler, L. T. Tsuneto, E. Llop, F. Rothhammer, L. Excoffier, M. W. Feldman, N. A. Rosenberg e A. Ruiz-Linares. Genetic variation and population structure in Native Americans. Plos Genetics, v.3, n.11, Nov, p.2049-2067. 2007.
- Watterson, G. A. Number of Segregating Sites in Genetic Models without Recombination. Theoretical Population Biology, v.7, n.2, p.256-276. 1975.
- Wright, S. Isolation by Distance. Genetics, v.28, n.2, Mar, p.114-138. 1943.
- Zegura, S. L., T. M. Karafet, L. A. Zhivotovsky e M. F. Hammer. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Molecular Biology and Evolution, v.21, n.1, Jan, p.164-175. 2004.

Anexos

Caixa 1- Protocolo Adaptado de Extração de DNA a partir do kit Gentra® Puregene Blood Kit® (Quiagen)

O protocolo foi padronizado e adaptado para o nosso laboratório com o objetivo de diminuir custos e aumentar a eficiência, praticidade e pureza na extração de DNA. Para a diminuição de custos, apenas a *Protein Precipitation Solution* precisa ser adquirida do fabricante (<http://www.qiagen.com>). O tempo de extração é reduzido comparado aos outros protocolos (<http://www.icb.ufmg.br/lbem>, <http://www.promega.com>), além de não haver manipulação de substâncias tóxicas (fenol-clorofórmio). Todas as amostras extraídas corretamente com o protocolo adaptado apresentaram alta pureza, baseados na relação 260/280, quando dosadas do Nanodrop® (Thermo Scientific) e valores não detectáveis de proteínas quando dosadas com o Qubit® (Invitrogen).

Solução 1 :RCLB (*Red Cell Lysis Buffer*)

1mM NH₄HCO₃
115 mM NH₄Cl

Solução 2 :WCLB (*White Cell Lysis Buffer*)

100 mM Tris-HCL (pH 7.6)
40mM EDTA (pH 8.0)
50 mM NaCl
0.2% SDS
0.05% (8 mM) Sodium Azide

1. Coloque 300 µl de sangue total em 900 µl da **Solução 1** e misture invertendo o tubo.
2. Incube por 1 min, a temperatura ambiente (15–25 °C).
3. Centrifugue por 1 min a 13.000rpm para precipitar as células brancas.
4. Descarte cuidadosamente o sobrenadante e repita o procedimento.
5. Adicione 300µl da **Solução 2** e vortexe vigorosamente. Incube a 37 °C até que a solução fique homogênea.
6. Adicione 5-10µl de **Proteinase K** (20mg/ul) e vortexe. Incube a 57 °C por no mínimo 2h ou *overnight*.
7. Adicione 1,5 µl de **RNaseA (10mg/ml)**, e misture invertendo o tubo. Incube por 15 min a 37 °C.
8. Adicione 100 µl **Protein Precipitation Solution®**, e vortexe vigorosamente.
9. Centrifugue por 3 min a 13.000rpm. Se o pellet de proteína não estiver firme, incube por 5 min no gelo e repita a centrifugação.
10. Pipete 300µl de **isopropanol 100%** em um tubo e adicione o sobrenadante do passo anterior, despejando cuidadosamente. Certifique-se de que o pellet de proteína não seja pipetado.
11. Inverta gentilmente o tubo até que o DNA seja visível.
12. Centrifugue por 3 min a 13.000rpm.
13. Descarte cuidadosamente o sobrenadante.
14. Adicione 300 µl de **ethanol 70%** e centrifugue por 1 min a 13.000 rpm.
15. Descarte cuidadosamente o sobrenadante e vire o tubo sobre um papel absorvente limpo. Aguarde até a evaporação do etanol.
16. Adicione 50 µl de **Low TE** (a quantidade de Low TE pode variar dependendo do *pellet* de DNA obtido). Incube a 65 °C por 5 min (ou a 37 °C por 1 hora), para dissolver o DNA.
17. Quantifique o DNA da amostra usando o Qubit ou o Nanodrop. Transfira o DNA para um tubo devidamente identificado e congele a amostra.

Caixa 2- Amplificação da região hipervariável I do DNAm

A sequência de nucleotídeos da região HVI do DNAm entre as posições 16024 e 16363 de todos os indivíduos foi determinada utilizando o protocolo adaptado de Alves-Siva *et al.* (2001). A amplificação por PCR foi feita em um volume final de 20 µl de reação contendo 0,3µM de *primers* com cauda M13 (L15989_M13F 5'-TGTAACGACGGCCAGTCCCAAAGCTAAGATTCTAAT-3' e H16410_M13R 5'- CAGGAAACAGCTATGACCGAGGATGGTGGTCAAGGGAC-3'), 200µM dNTPs, 0,1U de Platinum® Taq DNA Polimerase (Invitrogen), Tampão 1X (60 mM Tris-SO₄ [pH 8.9], 18mM [NH₄]₂ SO₄), 15mM MgSO₄ e 50ng de DNA. O programa de ciclagem consistia em uma desnaturação inicial de 4 min a 94°C, 35 ciclos de desnaturação a 94°C por 1 min, anelamento a 55°C por 30s, e extensão a 72°C por 1 min. Controles negativos foram feitos simultaneamente para detectar contaminação nas reações. Os produtos foram visualizados em gel de agarose 1% e corados com brometo de etídeo.

Referências:

Alves-Silva, J., M. D. Santos, P. E. M. Guimaraes, A. C. S. Ferreira, H. J. Bandelt, S. D. J. Pena e V. F. Prado. The ancestry of Brazilian mtDNA lineages. American Journal of Human Genetics, v.67, n.2, Aug, p.444-461. 2000.

Caixa 3: Pipeline de Análise de Seqüências (modificado a partir do protocolo elaborado pela aluna de iniciação científica Moara Machado)

A análise das seqüências foi realizada utilizando os softwares Phred-Phrap-Consed e Polyphred distribuídos gratuitamente para fins acadêmicos e de acordo com os seguintes passos:

1. Criação de pastas

Para o funcionamento dos programas que viabilizam a análise de seqüências é necessário criar 4 pastas (*edit_dir*; *phd_dir*; *poly_dir* e *chromat_dir*) no Linux.

2. PHRED-PHRAP (comando Unix: */genome/bin/phredPhrap*)

O programa Phred é rodado no Linux a partir do comando acima e ele lê as fluorescências dos arquivos gerados pelo seqüenciador, chama as bases e atribui um valor de qualidade a cada uma delas. O programa Phrap roda junto com o Phred e é responsável pelo alinhamento das seqüências e formação de contigs.

3. POLYPHRED (comando Unix: */genome/bin/polyphred -t p -t g -o out.txt*)

A função do PolyPhred está integrada ao uso dos programas Phred, Phrap e Consed. Ele compara as fluorescências obtidas de cada indivíduo para identificar locais heterozigotos para substituições únicas de nucleotídeos (SNPs). Os SNPs potenciais são marcados com uma cor azul para a inspeção rápida e um arquivo de texto contendo todos os possíveis polimorfismos identificados é produzido na pasta *edit_dir*. O PolyPhred não é perfeito e ocasionalmente podem ocorrer erros. No caso de seqüências haplóides, como o DNAm, todos os indivíduos são considerados homozigotos para os loci polimórficos.

4. CONSED (comando Unix: */genome/bin/consed_linux2.4*)

O Consed é uma ferramenta para a visualização e edição das seqüências produzidas, permitindo uma avaliação manual dos contigs.

5. Análise dos dados

Cada um dos picos dos fluorogramas é avaliado e comparado com uma seqüência referência, sendo que, se houver alterações a serem feitas, estas serão realizadas no arquivo texto gerado pelo PolyPhred. Posteriormente, os dados de todos os SNPs encontrados são agrupados em um único arquivo de texto para serem submetidos a um algoritmo criado pelo aluno de Doutorado em Bioinformática da UFMG, Wagner Carlos Santos Magalhães, para a produção de uma tabela (*algoritmo Wagner 1*). Essa tabela é utilizada para inferência das fases cromossômicas utilizando o software PHASE e, posteriormente as seqüências são reconstruídas utilizando uma algoritmo desenvolvido pelo Prof. Eduardo Tarazona (*Phase to DNAsp*).

Referências:

- Ewing, B., Hillier, L., Wendl, M., Green, P. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research, v.8, p.175-185. 1998.
- Ewing, B., Green, P. Basecalling of automated sequencer traces using phred. II. Error probabilities. Genome Research, v.8, p.186-194. 1998.
- Gordon, D., Abajian, C., Green, P. Consed: a graphical tool for sequence finishing. Genome Research, v.8, p.195-202. 1998

Caixa 4- Algoritmo desenvolvido em linguagem *Perl* pelo Prof. Eduardo Tarazona para atualização do *database* de seqüências da região HVI do DNAm

```

# For the mt project,
# What does this script do?
# The following script uses two files: (1) a database of sequences (database.txt) and (2) a set of new sequences
(sequencias.txt), and for each sequence of this file, check if a matching sequence exists in the database. If it exists, the
script change the name of the analyzed sequence by the name of the matching sequence of the database. If a
sequence of the sequencias.txt file does not match any of the sequences of the database, the script generate a new
name for this sequence, and append it to the database. The sequences with the changed names are in the file
'sequenciasmod.txt'. The script also calculate the haplotype frequencies in the sequencias sample and creates a file
with these frequencies. To test this script, we use two testing files:
#database.txt (file with the known sequences, to take from the Silvia's Arlequin file)
#hap1 CTTTCT
#hap2 CTTTCC
#hap3 CTTTTC
#sequencias.txt (file with the sequences from a population)
#seq1 CTTCCCT
#seq2 CTTTCC
#seq3 CTTCCCT
#seq1 and seq3 are equal and do not match any sequence in the database. Must be added as hap4
#seq2 matches hap2 in the database
#BOTH SEQUENCES MUST BE ALIGNED.
#/usr/bin/perl -w
use strict;

#variables definition
my $flag;
my @row; #data from population sequences ('sequencias.txt')
my $rowseq; #data from population sequences ('sequencias.txt')
my @hap; #data from database ('database.txt')
my $rowhap; #data from database ('database.txt')
my $l; #length of the aligned sequences
$l = 360;
my $j; #counter of the l characters of the sequence
my $baserow; # storage each base of the sequence file
my $basehap; # storage each base of the database file
my $seqdif; # 0 if sequences are equal, 1 if sequences are different

#IMPORTANT
my $nhap = 947; #number of haplotypes in the database. THIS VALUE MUST BE UPDATED AFTER EVERY RUN OF
THE SCRIPT
my @count; #store the sequences of the sequencias file for counting
my @sortedcount; #sorted @count
my $i;

#open sequencias.txt file and create sequenciasmod.txt file
open(FILE1,'sequencias.txt') || die "can't open file: $!";
open(FILE3,'>>sequenciasmod.txt') || die "can't open file: $!";

#reading each sequence of the sequencias population and taking sequence by sequence for comparisons with
#the database. The name of the sequences are in $row[0] and the sequence in $row[1]
$i = 0;
while ( defined ($rowseq = <FILE1>)){
    $flag = 0;
    chomp ($rowseq);
    @row = split(' ', $rowseq);

    #opening and parsing the 'database' to check for a matching sequence
    open(FILE2,'+<database.txt') || die "can't open file: $!";
    while (defined ($rowhap = <FILE2>)){
        chomp ($rowhap);
        @hap = split(' ', $rowhap);

```

Caixa 4- Continuação

```

#controlling if a pop sequence matches a database sequence
$seqdif = 0;
for ($j=0;$j<$l;$j++){
$baserow = substr($row[1],$j,1);
$basehap = substr($hap[1],$j,1);
print "$j\t";
print "$baserow\t";
print "$basehap\n";

if ($baserow ne $basehap && $baserow ne '?' && $basehap ne '?'){
$seqdif = 1;
$j = $l;
}
}

# what happen if a pop sequence matches a seq in the database
if ($seqdif == 0) {
# the name is changed using the database name and the sequence is stored in
#sequenciasmod.txt
print FILE3 "$hap[0]\t";
print FILE3 "$hap[1]\n";
$count[$i] = $hap[0];
$flag = 1;
}

# what happen if non of the pop sequence matches a seq. in the database
if ($flag == 0){
# a new sequence name is generated
$nhap = $nhap+1;
$row[0] = 'aplo'. $nhap;
$count[$i] = $row[0];
#the new sequence is stored in sequenciasmod.txt
print FILE3 "$row[0]\t";
print FILE3 "$row[1]\n";
#the new sequence is also stored in database.txt
print FILE2 "$row[0]\t";
print FILE2 "$row[1]\n";
}

close FILE2;
$i = $i + 1;
}
close FILE1;
close FILE3;

#counting the haplotypes in sequencias.txt and genetating a frequencies.txt file
@sortedcount = sort @count;
open(FILE4,'>>frequencies.txt') || die "can't open file: $!";
my $k; #index for haplotype names (0,1,2,3...)
$k = 0;
my $flag1;
my $count;
$count = 1;
for ($k = 0; $k < $i; $k = $k + 1){
if ($sortedcount[$k] ne $sortedcount[$k+1]){
print FILE4 "$sortedcount[$k]\t";
print FILE4 "$count\n";
$count = 1;
}
else {
$count = $count + 1;
}
}
}

```

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)