

SIMONE SILVA DOS SANTOS

# ANÁLISE DA VARIABILIDADE GENÉTICA DE POPULAÇÕES DOS ANDES.

Dissertação apresentada ao Curso de Mestrado do Departamento de Biologia Geral do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Genética.

Orientador: Fabrício Rodrigues dos Santos

Co-orientador: Sérgio D.J. Pena

Belo Horizonte-MG

Departamento de Biologia Geral

Instituto de Ciências Biológicas

2004

## Dedicatória

Dedico este trabalho aos meus pais a quem devo todo meu respeito e consideração, e por serem os patrocinadores dos meus sonhos. E ao meu Avô, José Vicente, de quem herdei a coragem e determinação (*In Memoriam*).

# Agradecimentos

Em primeiro lugar gostaria de agradecer ao meu orientador Fabrício Santos por permitir a realização deste trabalho, por entender minhas dificuldades e principalmente pela paciência dos últimos meses, que apesar de cansativos foram engrandecedores. Mesmo sendo um pouco fechado e sério (um típico mineiro) é uma pessoa maravilhosa com quem aprendi muito durante esses dois anos. E a sua esposa Adriana, pelo brilho no olhar do nosso “chefito”.

Ao Dr. Sérgio Pena, pela colaboração e por todas as discussões das segundas-feiras que foram essenciais para a realização deste trabalho. Por quem tenho grande respeito e admiração.

Aos colaboradores deste projeto pelas amostras de DNA cedidas. E ao Eduardo Tarazona, por toda ajuda.

Aos ameríndios que doaram as amostras e aos colaboradores que nos cederam as amostras de DNA.

Á toda minha família por todo incentivo e torcida, que mesmo estando distantes não deixaram me faltar nada, principalmente a minha mãe Marquina, meu padrasto Domingos, meu pai e meu tio Paulino pela confiança e por todo apoio emocional e financeiro.

A minha família mineira, Raquel, Gabriela, Bia, Beth, Leandro, pela ajuda computacional, Juliana, Gisele, Rodrigo, Luciana, Eloisa, Tia Dani, Renata, Paula, Savio e Ivan, enfim a todos os membros atuais e que passaram pelo LBEM que durante esses dois anos foram tudo para mim e em especial ao Rodrigo, por ser meu confidente e co-orientador, a Gisele pela amizade, Beth e Luciana Werneck pela ajuda nos momentos mais difíceis. Aos técnicos do LBEM, Wlademir e Cid Antonio.

Aos Membros do LGB, Pacu, Cadu, Carlos Gustavo, Helder, Léo, Marina, Paula, Cacá, Carol Peçanha, Jorge, Luis, Marcela, Dani, Débora Aline, Debrinha, Alice, Alessandra, Luciana Bastos, Vanessinha, Roseane, Simone, Chico Lobo, Flavia Parra, em especial Charles e Claudia, pela força e ajuda. Aos professores Nem, Andreia e Gloria. As técnicas do LGB, Miroca, Neuzinha e Katita, em especial a Katita a quem devo muitas caixas de Skol.

Aos funcionários do departamento de biologia geral, em especial a Marina Miranda pela disposição e carinho.

Aos professores da pós-graduação em genética por todo aprendizado, em especial Bernadete, Marisa, Dolores e Monica por todo carinho e atenção.

A todos colegas da pós-graduação, em especial: Chico Tosco, Fred, Lu Lara, Camilli, Cris, Dani Pontes, Carol, Ferhdy, Fernanda Dorela, e Juliana Roriz.

Aos amigos de todas as horas, Miguelito, Renatinha e Ricardo Camelo por serem meus verdadeiros amigos, por estarem presentes nos momentos bons e ruins, amo todos vocês.

Ao pessoal do grupo de discussão, Claudia, Charles, Carlos Gustavo, Léo, Dani, Rodrigo, Luis, Andreia, por todas as grandes discussões e lanches das nossas quintas. Em especial, ao Rodrigo, Claudia e Dani, pelas discussões dentro e fora do grupo, pelos conselhos e orientações.

Aos amigos de todas as festas, Lu Lara, Carol, Luis, Ferhdy, Filipe, Renata, Andrea, Fernanda Dorela, Dani Pontes, Gustavo, Sávio, Miguel, Chico Tosco, Bruno, Eder, Flavia Redondo e Anna Luiza.

A Anna Luiza, ao Mãozinha (Tiago) e ao pessoal do Vira e Mexe em especial Dudu e Marcley por animar e divertir as minhas quintas-feiras.

Aos meus amigos de Maceió, pela amizade sincera, principalmente a Juli, Jeane, Natan e Rala.

Ao meu primeiro e grande orientador Luiz Antônio Ferreira da Silva, pelos primeiros ensinamentos.

Aos meus amigos, que estão na luta (Mestrado e Doutorado) no Rio, São Paulo e Recife, que as nossas realizações sejam completas.

Aos meninos que dividiram seu espaço comigo, Leandro, Tiago, Helder e Dado, pelo cuidado e carinho.

Ao Alexandre, por existir.

Enfim, a todos muito obrigada.

# Sumário

<b>AGRADECIMENTOS</b>	<b>3</b>
<b>LISTA DE FIGURAS</b>	<b>7</b>
<b>LISTA DE TABELAS</b>	<b>8</b>
<b>LISTA DE ABREVIATURAS</b>	<b>9</b>
<b>RESUMO</b>	<b>10</b>
<b>ABSTRACT</b>	<b>12</b>
<b>1 INTRODUÇÃO</b>	<b>14</b>
1.1 ASPECTOS EVOLUTIVOS HUMANOS	14
1.2 O POVOAMENTO DAS AMÉRICAS	16
1.2.1 ESTUDOS EVOLUTIVOS COM OS NATIVOS AMERICANOS E ASIÁTICOS	17
1.3 POPULAÇÕES ANDINAS	20
1.3.1 OS QUÉCHUAS	20
1.3.2 OS CAYAPAS	21
1.3.3 OS COYAIMAS	21
1.4 CARACTERÍSTICAS DO CROMOSSOMO Y	23
1.4.1 ORIGEM E EVOLUÇÃO	23
1.4.2 USO DO CROMOSSOMO Y EM ESTUDOS EVOLUTIVOS	24
1.4.3 MARCADORES POLIMÓRFICOS DE EVOLUÇÃO LENTA (UEPs)	24
1.4.4 MARCADORES POLIMÓRFICOS DE EVOLUÇÃO RÁPIDA	25
1.4.5 TERMINOLOGIA E NOMENCLATURA DE HAPLOGRUPOS	25
1.5 CARACTERÍSTICAS DO DNA MITOCONDRIAL	28
<b>2 OBJETIVOS</b>	<b>30</b>
2.1 OBJETIVO GERAL	30
2.2 OBJETIVOS ESPECÍFICOS	30
<b>3 METODOLOGIA</b>	<b>31</b>
3.1 OS INDIVÍDUOS	31
3.1.1 AMOSTRAS DE POPULAÇÕES NATIVAS DA AMÉRICA DO SUL	31
3.1.2 AMOSTRAS DE POPULAÇÕES NATIVAS DA ÁSIA CENTRAL	31
3.2 EXTRAÇÃO DO DNA	32
3.3 TIPAGEM DOS EVENTOS POLIMÓRFICOS ÚNICOS(UEPs)	32
3.4 TIPAGEM POR ENZIMA DE RESTRIÇÃO	34
3.5 TIPAGEM POR MINI-SEQÜENCIAMENTO	34
3.6 ELETROFORESE EM GEL DE POLIACRILAMIDA CORADO POR SAIS DE PRATA	35
3.7 TIPAGEM DOS MICROSSATÉLITES	35
3.8 ANÁLISE DOS DADOS POPULACIONAIS COM O Y	38

<b>3.9</b>	<b>TIPAGEM E ANÁLISE DO DNA MITOCONDRIAL</b>	<b>38</b>
3.9.1	SEQÜENCIAMENTO DA REGIÃO CONTROLE	38
3.9.2	ANÁLISE DAS SEQÜÊNCIAS DO DNA MITOCONDRIAL	39
3.9.3	MINI-SEQÜENCIAMENTO DO DNA MITOCONDRIAL	39
<b>3.10</b>	<b>DESENVOLVIMENTO DO MULTIPLEX DE SNPs DO CROMOSSOMO Y</b>	<b>43</b>
<b>4</b>	<b>RESULTADOS</b>	<b>47</b>
<b>4.1</b>	<b>ANÁLISE DOS UEPs NAS POPULAÇÕES NATIVAS AMERICANAS</b>	<b>47</b>
<b>4.2</b>	<b>POPULAÇÕES DA ÁSIA CENTRAL</b>	<b>51</b>
<b>4.3</b>	<b>VARIABILIDADE DO HAPLOGRUPO Q NAS POPULAÇÕES ANDINAS</b>	<b>51</b>
<b>4.4</b>	<b>VARIABILIDADE INTERPOPULACIONAL</b>	<b>54</b>
<b>4.5</b>	<b>SEQÜENCIAMENTO DA REGIÃO CONTROLE DO DNA MITOCONDRIAL DA POPULAÇÃO COYAIMA DA COLÔMBIA.</b>	<b>57</b>
<b>4.6</b>	<b>DESENVOLVIMENTO DE NOVOS MARCADORES DO Y</b>	<b>59</b>
<b>4.7</b>	<b>PADRONIZAÇÃO DO MULTIPLEX DE SNPs DO CROMOSSOMO Y</b>	<b>63</b>
<b>5</b>	<b>DISCUSSÃO</b>	<b>65</b>
<b>5.1</b>	<b>VARIABILIDADE GENÉTICA DAS POPULAÇÕES ANDINAS</b>	<b>66</b>
<b>5.2</b>	<b>ORIGEM DA POPULAÇÃO COYAIMA ATUAL</b>	<b>68</b>
<b>5.3</b>	<b>QUANTAS ONDAS MIGRATÓRIAS?</b>	<b>69</b>
<b>6</b>	<b>CONCLUSÕES</b>	<b>71</b>
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>73</b>
<b>8</b>	<b>APÊNDICES</b>	<b>81</b>

# Lista de Figuras

<b>Figura.1.</b> Mapa do Estreito de Bering.	<b>16</b>
<b>Figura 2.</b> Mapa com a distribuição das freqüências dos haplogrupos Q3, Q e C e suas relações filogenéticas.	<b>19</b>
<b>Figura 3.</b> Mapa da distribuição lingüística dos Quéchuas	<b>20</b>
<b>Figura 4.</b> Índio cayapa utilizando canoa para locomoção.	<b>21</b>
<b>Figura 5.</b> Mapa da região de Tolima, Colombia.	<b>22</b>
<b>Figura.6.</b> Região específica do cromossomo Y.	<b>23</b>
<b>Figura 7.</b> Árvore de haplogrupos do cromossomo Y .	<b>27</b>
<b>Figura 8.</b> Mapa estrutural do DNA mitocondrial humano	<b>28</b>
<b>Figura 9.</b> Mapa com a localização geográfica das amostras analisadas.	<b>32</b>
<b>Figura-10.</b> Protocolo da reação de mini-seqüenciamento.	<b>36</b>
<b>Figura 11.</b> Árvore filogenética do cromossomo Y com os UEPs utilizados no multiplex de SNPs.	<b>46</b>
<b>Figura 12.</b> Gel de poliacrilamida 6% com tipagem da mutação <i>DYS199</i> .	<b>48</b>
<b>Figura 13.</b> Gel de poliacrilamida 6% com tipagem da mutação <i>RPS4Y</i> .	<b>48</b>
<b>Figura 14.</b> Mini-seqüenciamento do <i>DYS199</i>	<b>49</b>
<b>Figura 15.</b> Mini-seqüenciamento do marcador <i>Mid227</i>	<b>61</b>
<b>Figura 16.</b> Gel de poliacrilamida 6%, com tipagem <i>Mid967</i> .	<b>62</b>
<b>Figura 17.</b> Multiplex de SNPs por mini-seqüenciamento, iniciadores não purificados.	<b>63</b>
<b>Figura 18.</b> Multiplex de SNPs por mini-seqüenciamento, iniciadores purificados.	<b>64</b>

## Lista de tabelas

<b>Tabela 1</b> –Principais polimorfismos que definem os haplogrupos A, B, C,D e X.	<b>29</b>
<b>Tabela 2</b> - Seqüência dos iniciadores dos UEPs utilizados.	<b>33</b>
<b>Tabela 3</b> - Enzimas de restrição utilizadas nas análises dos UEPs	<b>34</b>
<b>Tabela 4</b> -Seqüência dos iniciadores dos microssatélites analisados	<b>37</b>
<b>Tabela 5</b> -Seqüências dos iniciadores externos para amplificação das regiões de interesse a serem analisadas por mini-seqüenciamento	<b>41</b>
<b>Tabela 6.</b> – Polimorfismos, iniciadores e condições ideais para PCR para o estudo do DNAmT de nativos americanos através de mini-seqüenciamento	<b>42</b>
<b>Tabela 7.</b> Seqüência dos iniciadores do multiplex sistema 1 (utilizando ddGTP-F*)	<b>44</b>
<b>Tabela 8.</b> Seqüência dos iniciadores do multiplex sistema 2 (utilizando ddATP-F*)	<b>45</b>
<b>Tabela 9.</b> Haplogrupos determinados pelos UEPs analisados	<b>50</b>
<b>Tabela 10.</b> Freqüências dos diferentes haplogrupos do cromossomo Y nas populações andinas analisadas.	<b>50</b>
<b>Tabela 11.</b> Haplótipos de microssatélites presentes nos indivíduos do haplogrupo Q, das populações analisadas e suas respectivas freqüências	<b>53</b>
<b>Tabela 12.</b> Análise da Variância Molecular das Populações Andinas.	<b>55</b>
<b>Tabela 13.</b> Matriz com resultado do teste exato de diferenciação (haplótipos com base em todos UEPs).	<b>56</b>
<b>Tabela 14.</b> Matriz com resultado do teste exato de diferenciação (haplótipos com base em microssatélites do haplogrupo Q).	<b>57</b>
<b>Tabela 15.</b> Linhagens do DNA mitocondrial identificadas na população Coyaima.	<b>57</b>
<b>Tabela 16.</b> Freqüência dos haplogrupos de DNAmT na população Coyaima.	<b>59</b>
<b>Tabela 17.</b> Freqüências dos alelos do marcador Mid967 observadas em nosso estudo.	<b>62</b>

## Lista de Abreviaturas

ACIT (Asociación de Cabildos Indígenas del Tolima) – Associação de grupos indígenas da região de Tolima.

ALF-Automated Labeled Fluorescent

AMOVA (Analysis of Molecular Variance) - Análise de variância molecular

ANDICOL - Comissão de direitos humanos Interamericana.

DHPLC (denaturation high-performance liquid chromatography) - Cromatografia líquida denaturante de alta performance.

DNA- (Deoxyribonucleic acid) - Ácido deoxiribonucléico

DNAmt- DNA mitocondrial

dNTP desoxi-nucleotídeo tri-fosfato

ddNTP-F didesoxinucleotídeo marcado com fluorescência

ExoI Exonuclease I

HPLC (High Performance Liquid Chromatography) - Cromatografia líquida de alto rendimento

HVSI- Segmento Hipervariável da Região controle do DNA mitocondrial

Mb- Megabase

MSY(male-specific region of the Y) – Região específica de homens do cromossomo Y

NCBI- National Center for Biotechnology Information

NRY (non-recombining portion of the Y) - Porção não recombinante do cromossomo Y

nt- nucleotídeo

OEA- Organização dos Estados Americanos

PAR (pseudo-autosomal regions) - Região pseudo-autossômica

pb- pares de base

PCR (Polimerase Chain Reaction) - Reação em cadeia da polimerase

PPE- Polimorfismo específicos de populações

RFLP- (Restriction Fragment Length Polymorphism) - Polimorfismo de tamanho de fragmento de restrição

RNA Ribonucleic acid- Ácido ribonucléico

RPS4Y- Ribosomal protein subunit 4 Y-gene- Gene da proteína ribossomal S4

SAP Shrimp Alkaline Phosphatase-

SNP (Single nucleotide polymorphism) - Polimorfismo de ponto

UEPs (Unique Events Polymorphisms) – Polimorfismos de eventos únicos.

YAP – (Y- Alu insertion polymorphism) (Polimorfismo de inserção Alu no cromossomo Y)

YCC– Consórcio do cromossomo

## Resumo

A variabilidade genética de populações andinas e da Ásia Central foi analisada utilizando marcadores moleculares do cromossomo Y e do DNA mitocondrial. Com o objetivo de investigar a origem e história evolutiva de populações andinas, estudamos um total de 240 indivíduos de cinco populações ameríndias: Tayacaja, Arequipa e San Martín, do Peru, Coyaima da Colômbia e Cayapa do Equador; e 61 indivíduos da Ásia Central, populações de Kirghiz e Kazakos. Neste estudo buscamos determinar o grau de diversidade das populações nativas dos Andes bem como avaliar o nível de miscigenação dos andinos com populações européias e africanas. Utilizamos nove polimorfismos de evento único (UEPs) que definem os seguintes haplogrupos continente-específicos: RPS4Y (haplogrupo C), encontrado em asiáticos e alguns raros nativos americanos; YAP (haplogrupo D-E\*); SRY-8299 (haplogrupo E\*), presente na Eurásia e África; sY81 (haplogrupo E3a), de populações africanas; 92R7, (haplogrupo P\*), presente em populações eurasiáticas; P36, (haplogrupo Q\*), ocorre em asiáticos e em alguns nativos americanos; DYS199 (haplogrupo Q3), exclusivo dos nativos americanos; M207, (haplogrupo R\*) e M173, (haplogrupo R1), presentes em populações européias e asiáticas. Cinco microssatélites foram analisados entre indivíduos do haplogrupo Q: DYS19, DYS389a, DYS389b, DYS391 e DYS393. Os UEPs analisados permitiram estimar o nível de miscigenação de cada população andina estudada, sendo a população de Coyaima a que apresentou maior frequência de haplogrupos não nativos, enquanto que a população de Cayapa possui apenas um haplogrupo ameríndio (Q3). Os haplogrupos C, D e R1 não foram observados nas populações andinas analisadas. Não foi observada a presença do haplogrupo Q\* nas populações asiáticas, um cromossomo encontrado em populações asiáticas com ancestrais em comum com nativos americanos. A análise do DNAmT da população Coyaima indicou que todos indivíduos possuem haplogrupos ameríndios, confirmando o padrão de colonização que ocorreu nas Américas, onde as linhagens maternas são nativas e as paternas possuem contribuições de imigrantes. A análise de variância molecular com base nos UEPs e nos microssatélites do haplogrupo Q (Q\*+ Q3) e Q3, confirmou o alto nível de variabilidade dentro das populações, entretanto, a variabilidade entre os grupos é maior do que entre as populações na análise com UEPs. Na análise com microssatélites do haplogrupo Q, a variação entre grupos e populações é significativa ( $P < 0,05$ ), apesar de ser bem menor que na análise com UEPs. O valor de  $F_{st}$  (baseado em UEPs) total entre tribos ameríndias utilizando todos haplogrupos foi de 0,18, enquanto que após retirada dos cromossomos não nativos o  $F_{st}$  subiu para 0,35. Isto sugere que a miscigenação esteja homogeneizando essas populações andinas. A diferenciação significativa observada

entre as populações andinas, principalmente entre Quéchuas e não-Quéchuas, pode ser atribuída à deriva genética e à restrição do fluxo gênico entre elas. No entanto, as populações do grupo dos Quéchuas ainda apresentam um considerável fluxo gênico entre elas, tornando-as mais homogêneas. Este trabalho contribuiu para o entendimento da dinâmica populacional dos povos andinos, que corroboram com os dados antropológicos destas populações.

## Abstract

The genetic variability of Andean and Central Asian populations has been analyzed by the use of Y chromosome and mitochondrial DNA molecular markers. In order to investigate the origin and evolutionary history of Andean populations, we have studied 240 Individuals from five amerindian populations: Tayacaja, Arequipa and San Martin from Peru, Coyaima from Colombia, Cayapa from Ecuador; and 61 individuals from Central Asian populations: Kirghiz and Kazak. We have analyzed the genetic diversity of Andean populations as well as their admixture level with Africans and Europeans. We have used nine unique event polymorphisms (UEPs) which define the following continent-specific haplogroups: RPS4Y, that defines haplogroup C, present in Asians and few Native American tribes; YAP (haplogroup D-E\*), which occurs in Asia (specially Japan); SRY-8299 (haplogroup E\*) present in Eurasians and Africans; sY81 (haplogroup E3a), found almost exclusively in Africans; 92R7 (haplogroup P\*), present in Eurasian populations; P36 (haplogroup Q\*), found in Asians and some American populations; DYS199 (haplogroup Q3), found exclusively in Native American populations; M207 (haplogroup R\*) and M173 (haplogroup R1), both found in Europeans and Asians. Five Y-linked microsatellites were studied in haplogroup Q individuals: DYS19, DYS389a, DYS389b, DYS391 and DYS393. The UEP haplogroups allowed estimating the admixture level of each Andean population. The Coyaima showed higher frequencies of non-Native American haplogroups, besides the Cayapa showed only the Native American haplogroup Q3. The haplogroups C, D and R1 were not found in the analyzed Andean populations. It was not detected the occurrence of haplogroup Q\* in Asian populations, a chromosome type present in Asian populations with recent common ancestors with Native Americans. The mtDNA analysis of Coyaima population showed that all individuals possess Amerindian haplogroups, corroborating with the American colonization pattern whose maternal lineages show native haplogroups and paternal lineages show also immigrant derived ones. The analysis of molecular variance using UEPs and microsatellites within haplogroup Q (Q\*+ Q3) confirm the higher variability observed within populations. However the among groups variability is a little higher (although  $P=0.10$ ) than among populations variability in the UEP analysis. The microsatellite analysis within haplogroup Q also revealed a very low percent of among groups and among populations variability, although statistically significant ( $P<0,05$ ). Total  $F_{st}$  (based in UEPs) values between Amerindian tribes using all haplogroups was 0.18 and this value raised to 0.35 when admixed haplogroups were left out of the analysis. Thus, admixture may be turning homogeneous the Andean populations. The significant differentiation observed among Andean populations, mainly between Quechuas and non-Quechuas, can be attributed to genetic drift and restriction of gene flow between them.

However, Quechua populations still show a considerable genetic flow among them, causing them homogeneous. The present study contributes to the understanding of the populational dynamics of Andean populations and confirms anthropological data on these populations.

# 1 Introdução

## 1.1 Aspectos evolutivos humanos

Tradicionalmente, os eventos evolutivos relacionados à nossa espécie eram investigados utilizando como evidências vestígios de materiais deixados pelas culturas antigas ou restos fósseis. Enfoques diferentes também foram utilizados tais como o estudo das línguas nativas, os mitos e as lendas, os nomes de localidades geográficas e outros relatos da nossa evolução cultural (Santos e Tarazona-Santos, 2002).

Entretanto, nos últimos anos foi acrescentada a esses estudos a análise de dados que utiliza como evidência o DNA. De fato, as técnicas da genética acrescentaram à arqueologia, etnologia e lingüística, uma poderosa fonte de informação para o estudo das populações atuais, o que torna possível fazer inferências de eventos ocorridos no passado como migrações, gargalos populacionais e expansões demográficas. Estes fenômenos demográficos podem de alguma forma serem *lidos* na estrutura genética das populações atuais (Cavalli-Sforza *et al.*, 1994). A genética de populações também contribui para a resolução de problemas históricos, tais como a antigüidade e origem geográfica dos povos das Américas, as migrações ligadas à difusão da agricultura nos diversos continentes, etc. Para este tipo de estudo histórico-molecular a genética nos fornece vários modelos que permitem predizer a distribuição de genes dentro e entre populações, em presença de um ou mais fatores evolutivos (deriva genética, migração, seleção natural e frequências de mutação) (Cavalli-Sforza, 1998). Assim, com o uso de modelos de genética de populações, pode-se formular hipóteses sobre os possíveis cenários evolutivos que geraram esta distribuição da variabilidade humana (Cavalli-Sforza, 1998).

Embora o comportamento das populações humanas seja de natureza complexa, modelos simplificados, mesmo considerando o predomínio de um ou dois fatores evolutivos (ex. deriva genética e mutação), podem ser úteis, já que os efeitos associados às ações de fatores adicionais são menores do que erros inerentes da estocasticidade do processo evolutivo e à própria amostragem experimental (Hammer *et al.*, 1998).

Nos últimos anos com a possibilidade de se estudar o DNA antigo conservado em ossadas e dentes foi acrescentada uma dimensão temporal aos estudos da variabilidade genética, tornando possível o estudo de populações nativas extintas ou cujos

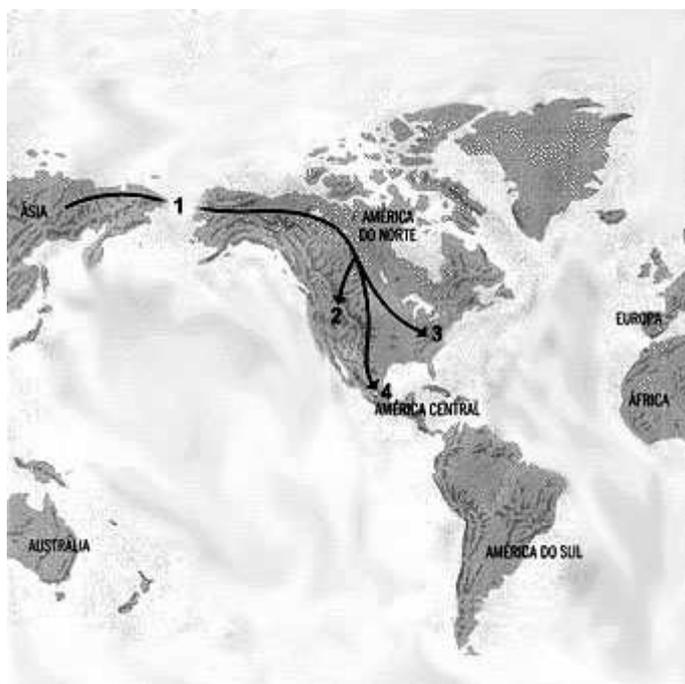
descendentes têm experimentado um alto grau de miscigenação (Stone e Stoneking 1998, Alves-Silva *et al.*, 2000).

Os primeiros trabalhos utilizando o DNA como evidência foi na questão da origem do homem moderno. O trabalho foi conduzido pelo grupo de Douglas Wallace da Universidade de Emory, seguido pelo grupo de Alan Wilson, na Universidade da Califórnia, em Berkeley, com o DNA mitocondrial. Em seu trabalho inicial, o grupo de Wilson estudou RFLP no DNA mitocondrial de 147 indivíduos de várias origens geográficas e elaborou uma árvore filogenética que apontava um ancestral comum, propondo que o DNA mitocondrial de todos os seres humanos atuais tinha sua origem na África há aproximadamente 200.000 anos (Cann *et al.*, 1987; Vigilant *et al.*, 1991). Este estudo sofreu algumas críticas, principalmente na sua metodologia estatística, sendo que a idade estimada foi reduzida para ~150 mil anos, o que foi corroborado por diversos outros estudos (Stoneking e Soodyall, 1996; Templeton, 1997). Este trabalho deu origem à teoria da Eva mitocondrial, que é consistente com o modelo de origem única e recente (Fora da África Recente) dos humanos atuais e não sustenta o modelo da origem Multirregional do homem moderno. A teoria Fora da África Recente sugere que todo o conjunto genético dos humanos atuais tem uma origem africana recente há 100-200 mil anos, e durante a expansão geográfica pelos diversos continentes, teria substituído por completo as populações de humanos arcaicos (derivadas dos *H. erectus* que migraram da África há 1,5 milhão de anos) que povoaram anteriormente Europa, Ásia e África. Enquanto isto, a teoria Multirregional, propõe um modelo baseado em um grau elevado de fluxo gênico entre as populações arcaicas dos diversos continentes, durante o período de evolução de 1 a 2 milhões até a formação do *Homo sapiens* moderno (Lewin, 1999).

Nos últimos anos, o DNA mitocondrial passou a ser intensamente analisado em várias populações nativas do continente americano e asiático, validando seu uso como uma ferramenta molecular bastante interessante, capaz de fornecer informações sobre várias questões polêmicas como é o caso do povoamento das Américas (Bonatto e Salzano, 1997, Fagundes *et al.*, 2002, Silva *et al.*, 2002). Seguindo a tendência da utilização do DNA mitocondrial como ferramenta para estudos evolutivos, o cromossomo Y, que possui algumas características semelhantes ao DNAMt, começou a ser utilizado em diversos estudos. Um dos primeiros foi o Adão do cromossomo Y, ou seja, o cromossomo Y mais antigo e ancestral que deu origem aos cromossomos atuais realizado por Michael Hammer da Universidade do Arizona em 1995, no qual estimou a idade do cromossomo Y ancestral humano em 188.000 anos, sugerindo que ambas regiões não-recombinantes (o DNAMt e o cromossomo Y humano) concordam quanto ao local e data estipulados para a origem do homem moderno (Hammer, 1995).

## 1.2 O Povoamento das Américas

O processo de colonização pré-histórica do continente americano vem sendo desvendado, principalmente, através dos achados da arqueologia, antropologia física e da lingüística, além da genética. Alguns estudos nestas áreas foram agregadas e são os alicerces do “Modelo das três ondas migratórias”, de Greenberg *et al.*, (1986). Estes autores defendiam que o povoamento do continente americano teria ocorrido em três levadas migratórias temporariamente isoladas. Segundo o modelo, a primeira onda ocorreu há 11 mil anos, composta por ameríndios e teria originado a maioria dos nativos americanos atuais, tanto da América do Norte quanto das Américas Central e Sul. A segunda onda migratória, ocorrida há nove mil anos, é representada pelos atuais Nadenes, um grupo de populações nativas restritas à costa noroeste dos Estados Unidos e à costa ocidental do Canadá. Por fim, a terceira onda migratória, há quatro mil anos, teria originado os Aleutas-Esquimós do extremo norte do continente. Estas migrações teriam se dado através do Estreito de Bering (Figura.1), que durante a última glaciação (pico ao redor de 18 mil anos atrás) quando o nível do mar esta até 130 metros mais baixo, seria uma ponte de terra entre a Ásia e América, a Beríngia.



**Figura.1.** Mapa do Estreito de Bering, a rota de migração mais aceita para as Américas.

Embora existam outros modelos a respeito do povoamento do continente americano, o modelo das três ondas migratórias foi bastante aceito na tentativa de

explicar o povoamento e geração de toda a diversidade observada nos nativos do continente americano, até surgirem novas propostas. No entanto, existem muitas controvérsias sobre o povoamento do continente americano, principalmente no que diz respeito ao número de ondas migratórias, à data da primeira migração e à localização geográfica precisa das populações que deram origem aos nativos americanos. Na literatura encontramos autores que defendem a ocorrência de quatro (Neves *et al.*, 2003), três (Greenberg *et al.*, 1986, Wallace 1995), duas (Lell *et al.*, 2002) e uma (Santos *et al.*, 1999; Silva *et al.*, 2002, Zegura *et al.*, 2004) ondas migratórias principais distintas.

Considerando a quantidade de dados obtidos nas diversas áreas, como arqueologia, antropologia, lingüística e genética, o único consenso que existe sobre o povoamento das Américas é que os grupos que deram origem aos nativos americanos eram caçadores-coletores asiáticos que atravessaram o estreito de Bering (Figura.1) no final do Pleistoceno e se espalharam rapidamente por todo continente (Santos e Tarazona-Santos, 2002).

### 1.2.1 Estudos evolutivos com os Nativos Americanos e Asiáticos

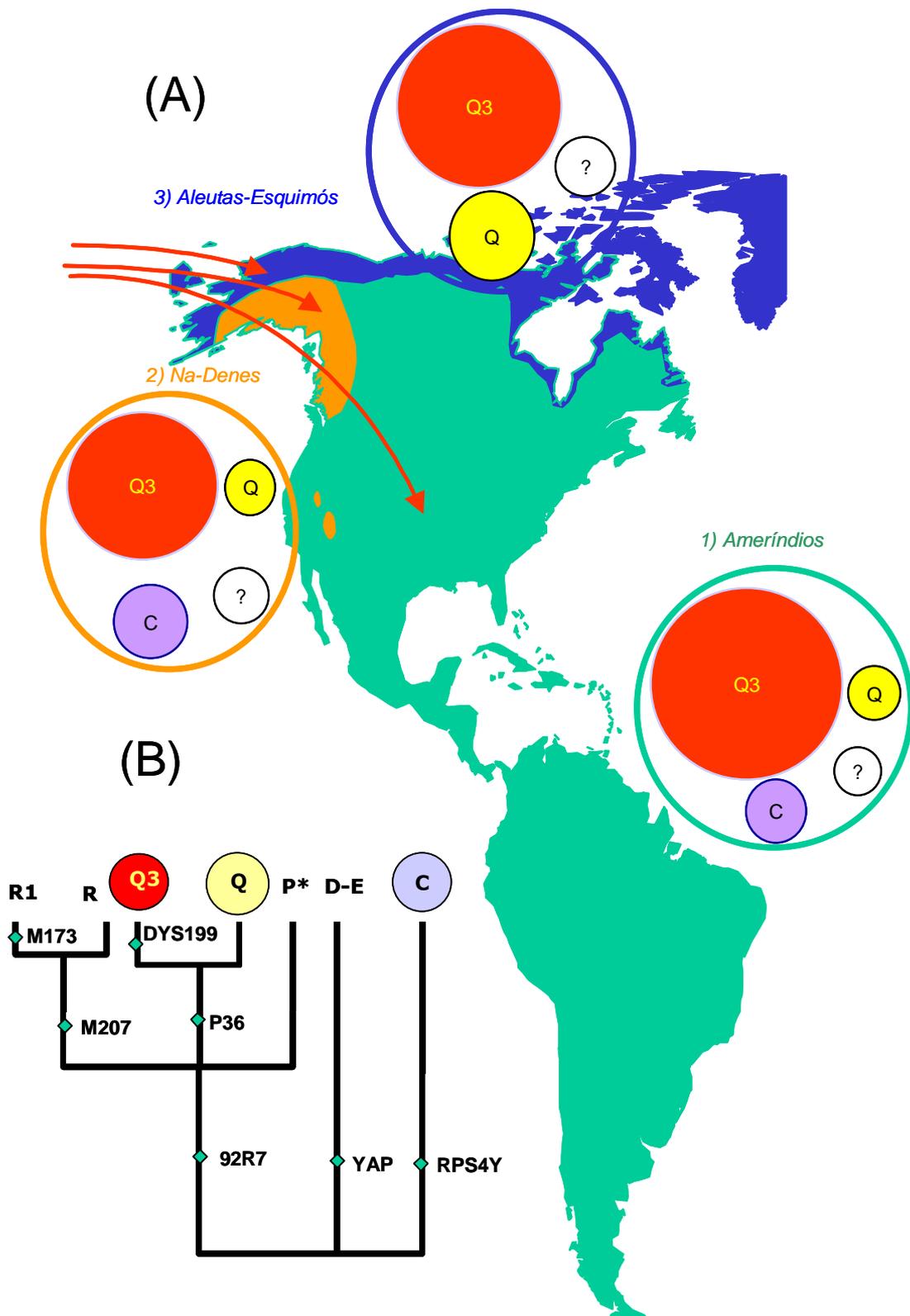
Os primeiros dados relativos à variabilidade do cromossomo Y nos Nativos Americanos evidenciaram um forte efeito fundador com dois marcadores moleculares multilocais (Pena *et al.*, 1995; Santos *et al.*, 1995), o que foi posteriormente confirmado por estudos nos EUA (Underhill *et al.*, 1996) utilizando a transição de C para T no loco *DYS199*. Esses estudos evidenciaram que o tipo de cromossomo Y que apresenta o alelo derivado desta mutação é o mais comum entre os nativos americanos, estando presente em mais de 70% dos indivíduos. Também encontrada em algumas populações na Ásia, provavelmente originou-se na América do Norte ou na Beríngia, depois que os imigrantes asiáticos deixaram a Sibéria (Santos *et al.*, 1999, Santos e Tarazona-Santos, 2002). A mutação denominada *DYS199-T* (Underhill *et al.*, 1996) define o haplogrupo chamado Q3 (Figura 2) conforme YCC (2002). Além das Américas, Q3 também foi encontrado em poucos indivíduos na Sibéria (próximo ao Estreito de Bering), muito provavelmente devido à migração reversa da América para Sibéria, o que é um fato comum hoje entre os grupos esquimós (existem também esquimós siberianos) (Santos *et al.*, 1999; Karafet *et al.*, 1999). O haplogrupo Q3 originou-se a partir do haplogrupo Q, definido pela mutação P36 que também está presente em populações da Sibéria (Hammer e Zegura, 2002).

A mutação P36 permite diferenciar os cromossomos nativos *DYS199-C* daqueles que chegaram depois da conquista nos últimos 500 anos fornecendo uma classificação mais precisa dos cromossomos nativos americanos. Estas mutações (P36 e *DYS199*)

constituem alelos população-específicos (PPE) e são eventos genéticos que podem ser usados para datar eventos demográficos. Neste trabalho, a datação de P36 (Hammer e Zegura, 2002) permite estimar uma data máxima, enquanto *DYS199* estima a data mínima de ocorrência da primeira migração para as Américas, justamente porque a primeira mutação ocorreu na Ásia antes da migração para as Américas, enquanto *DYS199* ocorreu após esta primeira migração.

Diversos estudos de genética molecular, relativos aos nativos americanos, têm se direcionado a responder questões sobre o primeiro povoamento do continente. No entanto, em níveis geográficos mais específicos dentro da América do Norte e do Sul, as relações históricas entre as diferentes populações não foram determinadas. Em relação à América do Sul, o trabalho de Tarazona-Santos e Santos (2002) verificou a existência de padrões diferenciais de deriva genética e fluxo gênico no oeste e leste do continente. As populações dos Andes (oeste), considerando aquelas do agrupamento linguístico Quéchua, apresentam maior tamanho efetivo e fluxo gênico do que as populações do leste do continente. Este padrão tem resultado em uma tendência à homogeneização do conjunto gênico nas populações do oeste e uma diferenciação entre as populações no leste (Tarazona-Santos e Santos, 2001).

Várias populações ainda não foram analisadas e algumas apresentam dúvidas quanto à sua origem étnica devido à aculturação com perda da identidade étnica e provável fluxo gênico com outras populações. Um exemplo é a população de índios da tribo Coyaima da região de Tolima na Colômbia. A tribo remanescente ocupa hoje um pequeno vilarejo nos Andes colombianos e teve sua língua original extinta (Coyaima) que foi substituída pelo espanhol. A relação da língua original com outras línguas nativas americanas não é bem conhecida. Isto torna a análise de dados genéticos ainda mais importante para realizar inferências históricas sobre o passado dos nativos americanos cuja cultura esteja se exaurindo.



**Figura 2.** (a) Mapa com a distribuição das freqüências inferidas para os haplogrupos Q3, Q e C nos três grupos lingüísticos de nativos americanos (Santos e Tarazona-Santos, 2002), o tamanho dos círculos são referentes a freqüências dos haplogrupos, a interrogação no círculo branco é referente aos outros haplogrupos estudados. (b) Relações filogenéticas entre os haplogrupos e UEPs que os definem de acordo com a nova nomenclatura (YCC 2002).

### 1.3 Populações Andinas

Neste projeto foram estudadas três populações Andinas do Peru pertencentes ao grupo linguístico Quéchua, predominante na parte ocidental da América do Sul, tendo sua distribuição geográfica (norte, centro e sul dos Andes peruanos). Outra população estudada constitui-se de índios equatorianos, os Cayapas, que pertencem ao grupo linguístico Chibcha-Paezan (Greenberg *et al.*, 1987) ou Chachi (Ethnologue Data Base), mas geograficamente se localizam na área de influência das culturas andinas. A população de índios colombianos Coyaima pertence ao grupo linguístico Coyaima hoje extinto, e se localiza na região entre duas cadeias de montanhas nos Andes da Colômbia.

#### 1.3.1 Os Quéchuas

Os Quéchuas são descendentes dos Incas pré-colombianos conquistados por Francisco Pizarro. Os povos quéchuas na sua origem habitavam as montanhas meridionais do Peru, hoje suas populações se estendem além das fronteiras do Peru. Sua língua é hoje uma das mais difundidas da América ocorrendo na Colômbia, Equador, Peru, Bolívia, Argentina e Chile (Figura 3), sendo reconhecida como língua oficial no Peru. ([www.mapahumano.fietras.com](http://www.mapahumano.fietras.com)).

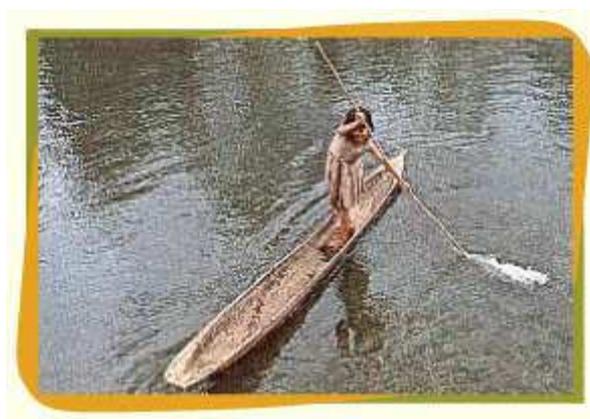


**Figura 3.** Mapa da distribuição lingüística dos Quechuas (porção vermelha do mapa). [www.prodiversitas.bioetica.org/quechuas.htm](http://www.prodiversitas.bioetica.org/quechuas.htm)

Os quéchuas elaboraram uma cultura bastante rica antes da chegada dos conquistadores espanhóis e têm permanecido fiéis às suas tradições, conservando grande parte de sua religião, cantos, orações, deuses e seres mágicos. Às vezes, nos Andes, ocorrem disputas entre os grupos indígenas pelas fronteiras dos campos cultiváveis produzindo mortos e feridos ([www.mapahumano.fietras.com](http://www.mapahumano.fietras.com)).

### 1.3.2 Os Cayapas

Os cayapas chamam a si mesmos de Chachi, vivem ao redor do Rio Cayapas e seus afluentes no Equador. Vivem da terra e seu principal meio de transporte é a canoa. Provavelmente tiveram origem nas montanhas andinas, mas a região que habitam hoje possuem uma vegetação bastante densa e se encontra pouco habitada, com exceção das vias aquáticas. Cada grupo é governado por um chefe (uñi) que se encarrega de encaminhar os matrimônios e resolver as disputas. O adultério, poligamia e incesto não são permitidos. Estes povos são resistentes a mudanças e continuam sendo tribos campestres com pequeno mercado baseado em suas canoas (Figura 4) ([www.mapahumano.fietras.com](http://www.mapahumano.fietras.com)).

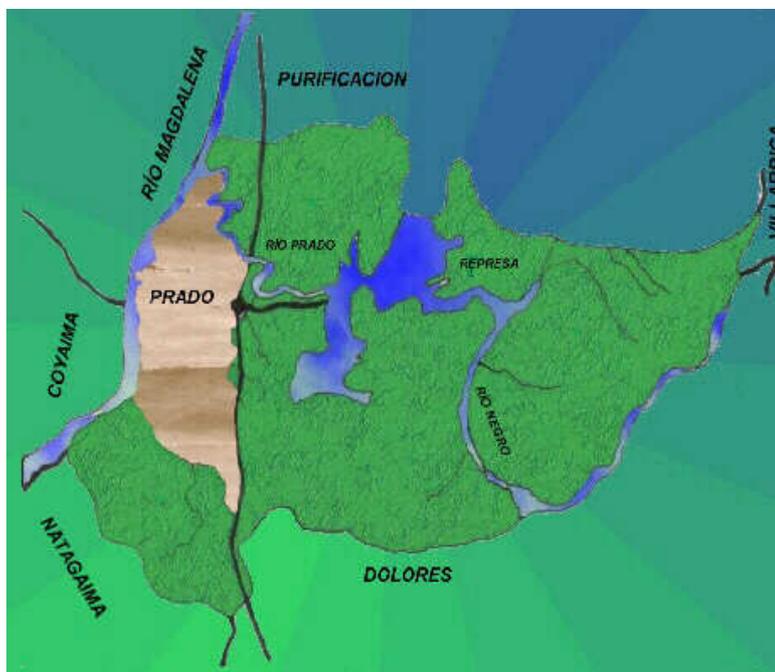


**Figura 4.** Índio cayapa utilizando canoa para locomoção.

### 1.3.3 Os Coyaimas

A população de Coyaima é descendente dos Pijaos que constituíam o grupo indígena colombiano que ofereceu maior resistência à conquista espanhola. Viviam ao lado do Rio Magdalena, a partir do baixo Saldanha, ao longo do Vale de Neiva, na região

de Tolima (Figura 5) (Gómez,1967). Os povos indígenas da etnia Pijao estão distribuídos nos municípios de Natagaima, Ortega, Coyaima, Chaparral, Ataco, Dolores, Saldaña, Purificación, Coello e Prado. Esses indígenas estão organizados em uma associação ACIT (Asociación de Cabildos Indígenas del Tolima) que os agrupam em 60 comunidades (Antorveza,1992).



**Figura 5.** Mapa da região de Tolima, Colômbia.

Este grupo (Coyaima) viviam em constante conflito com outras tribos da região o que facilitou a conquista pelos espanhóis. A tribo ainda existe como entidade, mas não fala mais a língua nativa, o coyaima, que é considerada extinta, atualmente eles falam apenas o espanhol. A população dos índios Coyaima permanece até hoje sendo perseguida, agora por guerrilheiros da Colômbia com interesse na região que é caracterizada por ser uma zona agro-industrial no centro-sul do país e tendo como maior riqueza as minas de ouro ([www.indigenascolombia.org](http://www.indigenascolombia.org)).

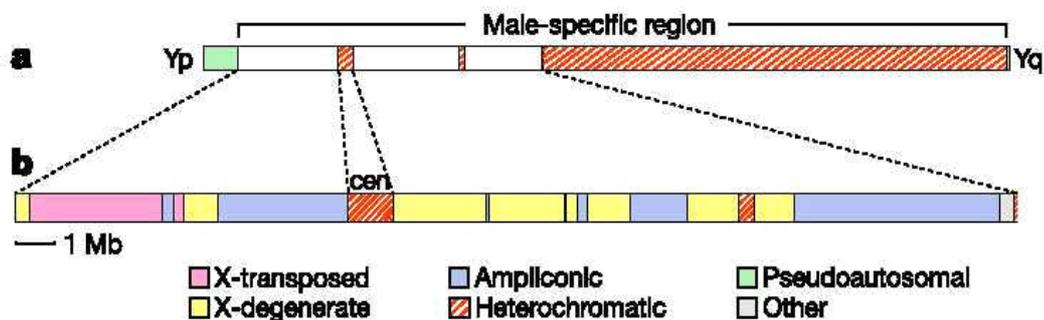
Historicamente, desde os anos 50 há presença de guerrilhas na região que entraram em conflito com os conservadores da época. Nos anos 90 iniciou a presença de grupos paramilitares como “*rojo ata*”, grupo que age no sul do departamento de Tolima. O conflito armado na região intensificou-se em 2001 quando o assassinato de alguns indígenas e líderes políticos tornou-se freqüente, despertando o interesse de líderes sociais e da comissão de direitos humanos Interamericana (ANDICOL,2003). (Em anexo carta da comissão Interamericana de direitos humanos da OEA para o Defensor nacional da Colômbia com data de 30 de Janeiro de 2003).

## 1.4 Características do cromossomo Y

### 1.4.1 Origem e evolução

Os atuais cromossomos sexuais de mamíferos (X e Y) evoluíram a partir de um par ancestral de autossomos. Quando um componente desse par ancestral assumiu o gene que determinava o sexo masculino, a recombinação entre os pares homólogos foi suprimida, dando origem aos proto-X e proto-Y (figura 6) (Lahn e Page 1999). A supressão da recombinação e perda de segmentos homólogos com o X foi se estendendo ao longo do Y até que atingiu 95% do cromossomo no homem, com exceção das extremidades do braço curto (Yp) e longo (Yq), as únicas regiões que sofrem pareamento e recombinação durante a meiose. Estas regiões são denominadas PAR1 e 2 (pseudo-autosomal regions) e, no homem possuem 2,6 e 0,34 Mb, respectivamente (Vogt *et al.*, 1997). O restante do Y humano, uma região de ~60 Mb, é denominada NRY (non-recombining portion of the Y) (Lahn e Page 1999) ou MSY (male-specific region of the Y) (Lahn *et al.*, 2003).

Após o seqüenciamento completo do cromossomo Y, foi possível construir um mapa físico de alta resolução da MSY, sendo esta região um mosaico de seqüências de heterocromatina e três classes seqüências de eucromatina no total de 23 Mb (Lahn *et al.*, 2003) e divididas em: i) seqüências transpostas do X, ii) seqüências degeneradas do X e iii) seqüências repetitivas (Figura 6).



**Figura.6.** Região específica do cromossomo Y. **a**,esquema completo do cromossomo Y, incluindo regiões pseudo-autossômica e heterocromatina. **b**,região específica do cromossomo Y. Figura transcrita de Lahn, 2003.

### 1.4.2 Uso do cromossomo Y em estudos evolutivos

A natureza não recombinante do cromossomo Y, presente em única cópia nos homens, e transmitido clonalmente de pai a filho, oferece importantes possibilidades de aplicações em estudos de evolução humana. Como todos os locos estão em completa ligação, as mutações se acumulam seqüencialmente nas linhagens paternas através das gerações, dando informações a respeito da história biológica das linhagens paternas humanas (Santos e Tyler-Smith, 1996).

Desmentindo as primeiras triagens que sugeriam pouquíssima variabilidade genética no cromossomo Y (Malaspina *et al.*, 1990), nesta última década foi demonstrada a existência de distintos tipos de polimorfismos que estão permitindo uma análise detalhada da sua variabilidade e esclarecer aspectos da história da humanidade. No cromossomo Y foram identificados inúmeros polimorfismos de eventos únicos (UEPs) que são normalmente bialélicos (YCC 2002) e vários outros marcadores polimórficos de evolução rápida, os microsatélites (Bosch *et al.*, 1999).

### 1.4.3 Marcadores polimórficos de evolução lenta (UEPs)

UEP (*unique event polymorphism*) são mutações geralmente bialélicas de ocorrência única (apomorfias) e podem ser SNPs, indels ou retrotransposons:

- SNP (*single nucleotide polymorphism*) - mutações pontuais que envolvem a mudança de uma única base (Marshall 1997);
- Indels são pequenas inserções e deleções nucleotídicas,
- Inserções de retrotransposons (Alus e L1s humanas) (Santos *et al.*, 2000).

Estes polimorfismos possuem uma baixa taxa de mutação ( $10^{-7}$  a  $10^{-9}$  mutações por sítio por geração) e a maior parte deles são eventos únicos de mutação, nos quais os estados ancestrais e derivados podem ser facilmente reconhecidos a partir de estudos com primatas. Aquele estado alélico compartilhado com os demais primatas é normalmente o estado ancestral, assim é possível usar um enfoque cladístico para reconstruir a filogênese dos cromossomos Y humanos que existem hoje (Santos *et al.*, 1999, 2000, Carvalho-Silva *et al.*, 2001, Underhill *et al.*, 2000, 2001). Esta análise busca caracterizar a relação genealógica entre haplogrupos através do compartilhamento de um determinado alelo derivado do UEP que os define.

#### 1.4.4 Marcadores polimórficos de evolução rápida

Os marcadores de evolução rápida presentes no cromossomo Y podem ser mini- e microssatélites. Os microssatélites são regiões de DNA repetitivo em tandem, com unidades de 1 a 6 pares de bases, que são abundantes, ubíquos e altamente polimórficos nos genomas de eucariotos (Carvalho-Silva *et al.*, 1999). A variabilidade inter-alélica é geralmente determinada pela diferença no número de repetições (Tautz, 1989). A alta variabilidade dos locos de microssatélites é produto da alta taxa de mutação deste tipo de *loco* ( $10^{-2}$  -  $10^{-5}$  mutações por loco por geração) (Brinkmann *et al.*, 1998) em comparação a outras regiões do genoma ( $10^{-7}$ - $10^{-9}$  mutações por loco por geração) (Nei *et al.*, 1987). O tipo mais comum de mutações em microssatélites geralmente leva a um aumento ou diminuição no número de repetições de uma unidade. Acredita-se que este tipo de mutação ocorre no nível molecular através de um *deslize* da DNA polimerase durante a replicação (Schlotterer e Tautz, 1992). Precisamente, pelo alto nível de variabilidade que apresentam, os microssatélites são usados para estudos populacionais em diversas espécies, incluída a humana, e na resolução de problemas de identificação individual (Pena *et al.*, 1995) com aplicações forenses (Pena e Chakraborty, 1994) e na etologia animal (Prodöhl *et al.*, 1996). Atualmente, já foram descritos aproximadamente 300 microssatélites no cromossomo Y (Jobling e Tyler-Smith, 2003).

#### 1.4.5 Terminologia e nomenclatura de haplogrupos

De posse dos conceitos de alelos (derivado/ancestral) e de UEPs, devemos esclarecer o conceito de haplogrupo do cromossomo Y. O haplogrupo pode ser definido como a combinação de alelos derivados para uma série de UEPs do cromossomo Y, que apresenta uma distribuição geográfica conhecida. Por exemplo, o haplogrupo Q3 pode ser caracterizado pela presença dos alelos derivados em 3 UEPs (92R7, P36 e *DYS199*) e é observado quase exclusivamente no continente americano (Underhill *et al.*, 1996; Santos e Tyler-Smith, 1996).

Por outro lado o haplogrupo P\*, que difere do Q3 pela presença do alelo ancestral nos locos P36 e *DYS199* (Santos e Tarazona-Santos, 2002), apresenta uma distribuição mais ampla na Europa, em algumas partes da Ásia e provavelmente também nas Américas (Santos *et al.*, 1999). No entanto, os haplótipos no cromossomo Y são definidos com base nos microssatélites. Quando uma mutação de ponto acontece em um cromossomo Y originando um novo haplogrupo, este apresenta um determinado

haplótipo (combinação de alelos para distintos locos de microssatélites). Se essa mutação sobrevive por um tempo considerável e a frequência de indivíduos portadores é aumentada, os microssatélites associados a esta mutação começarão a acumular variabilidade com o número de gerações passadas e com o incremento populacional, porque os dois fatores aumentam a incidência de eventos de mutação. Analisando os padrões de acúmulo da variabilidade nos microssatélites em qualquer das linhagens (haplogrupos) definidas por UEPs é também possível fazer inferências históricas, em escalas temporais recentes (em relação aos UEPs), o que inclui a resolução de problemas evolutivos em nível microgeográfico (Zerjal *et al.*, 1997, Hurles *et al.*, 1999).

A distribuição geográfica de um estado alélico derivado no UEP é geralmente proporcional ao tempo quando esta mutação ocorreu. Quanto mais restrita geograficamente é a mutação, mais recente ela é, e quanto mais difundida pelo mundo mais antiga ela é (Hammer, 1998).

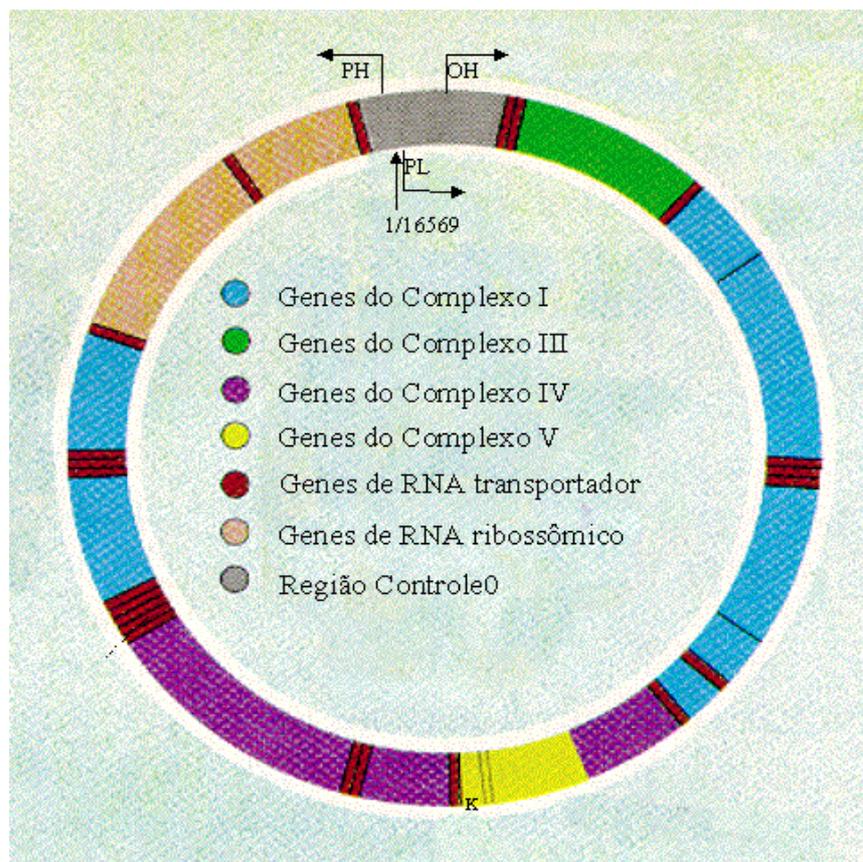
A possibilidade de formar haplótipos compostos a partir de marcadores com velocidade de mutação diferenciada está fornecendo interessantes possibilidades em toda a genética de populações, pois suas potencialidades vão além de estudos evolutivos com o cromossomo Y. Estes foram aplicados também em estudos com interesse histórico (Zerjal *et al.*, 1997, Forster *et al.*, 1998, Hurles *et al.*, 1998), para determinação de casos de paternidade em dúvida de várias gerações passadas como o caso dos descendentes de Thomas Jefferson com a sua escrava Sally Hemmings (Foster *et al.*, 1998) e a maternidade de Cristovão Colombo (Lorente 2003, comunicação pessoal).

Até pouco tempo eram utilizados três conjuntos distintos de UEPs para definir os haplogrupos do cromossomo Y (Underhill *et al.*, 2001, Hammer *et al.*, 2001, Jobling e Tyler-Smith, 2000). Nos últimos anos houve uma explosão de dados a partir da porção não recombinante do cromossomo Y: em 1997 eram 19 UEPs tipados por PCR e com o desenvolvimento de novas técnicas como DHPLC (denaturation high-performance liquid chromatography) foram publicados mais de 200 UEPs (Underhill, 2001). Com isso aumentou o número de UEPs conhecidos, de publicações e de diferentes sistemas de nomenclatura, o que dificultava a comparação entre diferentes autores (YCC, 2002). A filogênese humana do cromossomo Y está atualmente padronizada em um único sistema flexível de nomenclatura (Figura 7) (YCC, 2002).



## 1.5 Características do DNA mitocondrial

O DNA mitocondrial (DNAm<sub>t</sub>) é o material genético encontrado no interior das mitocôndrias, organelas produtoras de energia presentes no citoplasma das células. O DNAm<sub>t</sub> humano, constituído de 16,569 pares de bases, pode ser dividido em duas regiões principais: a região codificadora que constitui cerca de 94% da molécula e contém os genes para os diferentes tipos de RNA e proteínas necessários para síntese de componentes catalíticos do sistema de fosforilação oxidativa. A taxa de mutação na região codificadora é aproximadamente 5 vezes maior do que a do DNA nuclear. A região controle com cerca de 1,122 pb contém alguns sítios de replicação e de transcrição da molécula e evolui cinco vezes mais rápido que o resto da molécula, portanto 25 vezes mais rápido que o DNA nuclear (Figura 8) (Anderson,1981;Wallace,1995).



**Figura 8.** Mapa estrutural do DNA mitocondrial humano. As posições 1/16569 definem o primeiro e o último nucleotídeo da molécula circular, sendo que os demais nucleotídeos são numerados no sentido horário, de acordo com Anderson e colaboradores (1981). Cerca de 90% da seqüência do DNAm<sub>t</sub> humano é transcrita em algum produto gênico, estando, em sua maioria, relacionados com a síntese de componentes catalíticos da fosforilação oxidativa. Figura adaptada de Wallace,1997 por Santos (1998).

O DNAMt foi o primeiro genoma a ser seqüenciado na sua totalidade e interpretado em relação ao conteúdo gênico. São 37 genes que codificam dois RNA ribossômicos, 22 RNA transportadores e 13 polipeptídeos envolvidos no processo de fosforilação oxidativa (Anderson *et al.*, 1981). As propriedades do DNA mitocondrial incluem: o modo de herança quase exclusivamente materno, alto número de cópias e taxa de evolução rápida fornecendo uma ferramenta formidável para estudos evolutivos. Uma propriedade comum ao DNAMt em mamíferos é o baixo nível de transferência do DNAMt paterno, da ordem de 0,001% a cada geração. Essa pequena contribuição paterna é consistente com a quantidade relativa do DNAMt materno e paterno no zigoto, imediatamente após a fertilização (Stoneking e Soodyall, 1996).

As linhagens de DNAMt de todo o mundo dividem-se em três grandes conjuntos: os super-haplogrupos L1, L2 e L3. Os dois primeiros são específicos africanos, e o último ocorre em todos os continentes, mas podem ser divididos em haplogrupos típicos de populações nativas africanas, européias, asiáticas e americanas (Parra, 2003).

Estudos realizados mostraram que a maioria das linhagens mitocondriais encontradas nas populações nativas americanas atuais podem ser agrupadas em cinco haplogrupos denominados A, B, C, D e X (Tabela 1), cada um com mutações específicas; As freqüências desses haplogrupos diferem entre as diversas populações nativas americanas distribuídas pelo continente (Torroni *et al.*, 1992).

**Tabela 1** –Principais polimorfismos que definem os haplogrupos A, B, C, D e X.

HAPLOGRUPO	POLIMORFISMO DA REGIÃO CONTROLE*
A	<b>16111(C-T)</b> ,16223(C-T),16290(C-T),16319(G-A),16362(T-C).
B	16189(T-C),16217(T-C),
C	16223(T-C),16298(T-C), <b>16325(T-C)</b> ,16327(C-T)
D	16223(C-T), <b>16325(T-C)</b> ,16362(T-C)
X	16213(G-A),16200(A-G)

\* Os números representam a posição do nucleotídeo da extremidade 5' do sítio de reconhecimento da endonuclease de restrição onde ocorrem as mutações (entre parêntesis), de acordo com a seqüência referência (Anderson *et al.*, 1981). Os polimorfismos encontrados principalmente em nativos americanos estão em negrito.

Após o seqüenciamento total da molécula de DNAMt foi possível identificar e diferenciar os haplogrupos nativos americanos dos seus ancestrais asiáticos com mais precisão, utilizando polimorfismos específicos de cada população (Bandelt *et al.*, 2003).

# 2 Objetivos

## 2.1 *Objetivo geral*

Caracterizar distintos tipos de marcadores polimórficos do cromossomo Y, incluindo novos UEPs, nas amostras procedentes de populações da América do Sul e Ásia Central para avaliação da estruturação da variabilidade genética do cromossomo Y nessas populações.

## 2.2 *Objetivos específicos*

1. Analisar os UEPs (YAP, RPS4Y, SRY-8299, sY81, 92R7, P36, *DYS199*, M207 e M173) nas populações nativas da América e Ásia Central.
2. Obter dados de variabilidade genética dentro do haplogrupo Q e Q3 através da tipagem dos microssatélites *DYS19*, *DYS389a*, *DYS389b*, *DYS391* e *DYS393* nas populações nativas das Américas;
3. Avaliar o nível de diferenciação genética e a estruturação populacional dos povos andinos analisados.
4. Determinar os níveis de mistura entre as populações andinas estudadas com as populações de origem europeia e africana que chegaram na América do Sul nos últimos 500 anos, principalmente a população dos Coyaimas da Colômbia, um povo isolado com uma história de resistência à colonização mas com perda de sua identidade linguística.
5. Análisar o DNA mitocondrial da população de Coyaima da Colômbia.
6. Desenvolver novos marcadores polimórficos do cromossomo Y, úteis para estudos evolutivos e populacionais.
7. Padronização de um multiplex de SNPs do cromossomo Y.

# 3 Metodologia

## 3.1 Os indivíduos

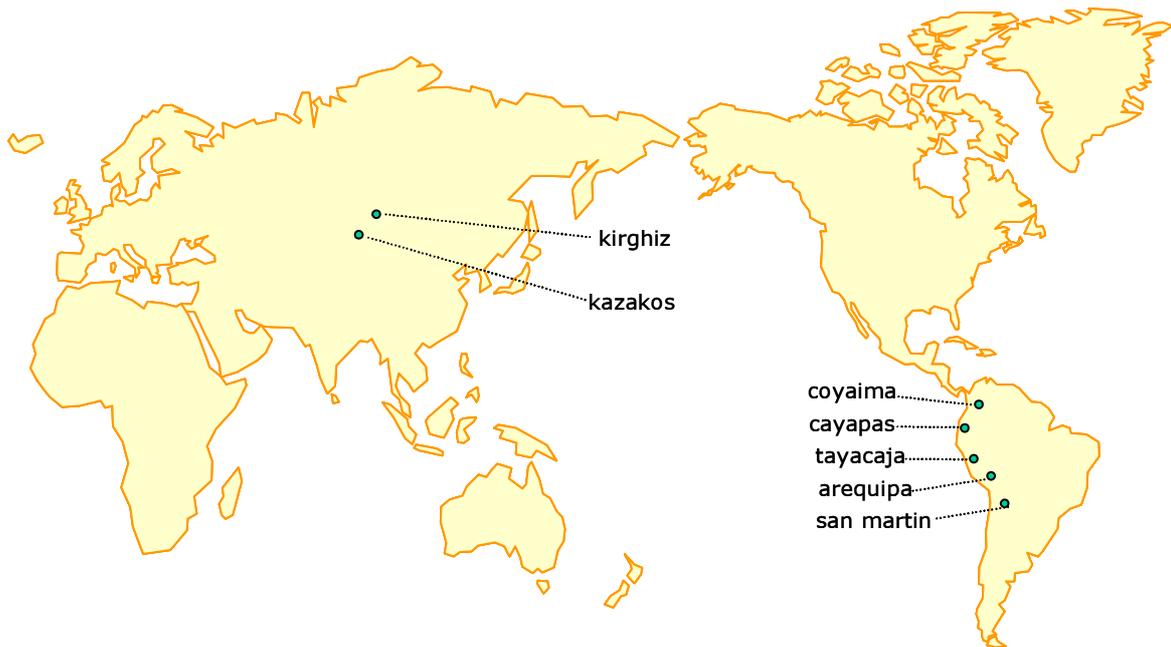
### 3.1.1 Amostras de populações nativas da América do Sul

- 97 amostras de DNA da localidade de Tayacaja, Andes Centrais do Peru e 23 amostras das localidades de Arequipa, nos Andes Meridionais do Peru. Estas amostras de Quéchuas foram parcialmente analisadas durante o mestrado do aluno Eduardo Tarazona-Santos e obtidas com a colaboração de pesquisadores da Universidade Peruana Cayetano Herédia de Lima, Peru.
- 59 amostras de DNA da localidade de San Martin, Andes Setentrionais do Peru, obtidas com a colaboração da Dra. Leli Solari, do Projeto Malária da Marinha de Guerra dos Estados Unidos da América.
- 40 amostras de DNA de índios Cayapas do Equador, cedidos pelo Prof. Gianfranco De Stefano e pela Dra. Olga Rickards do Departamento de Biologia da Universidade Tor Vergata-Roma II (Itália).
- 21 amostras de sangue de índios Coyaima, região de Tolima da Colômbia, cedidas pelo Professor Gustavo Adolfo Vallejo da Universidad del Tolima, Ibagué, Colômbia.

### 3.1.2 Amostras de populações nativas da Ásia Central

- 36 amostras de Kirghiz das Montanhas do Pamir (localidade de Sar Tash, Kirgizistão, fronteira com a China).
- 31 amostras de Kazakos das Montanhas do Tien Shian (Kazakhistão).

As amostras da Ásia foram cedidas pelo Professor Davide Pettener do Departamento de Biologia Evolucionista Experimental da Universidade de Bolonha, Itália. Todas as amostras haviam sido previamente coletadas com o consentimento informado dos doadores por colaboradores do Peru, Colômbia e Itália. Na Figura 9 é apresentada a localização geográfica das populações analisadas.



**Figura 9.** Mapa com a localização geográfica das amostras analisadas.

### **3.2 Extração do DNA**

A maior parte das amostras foi cedida como alíquotas de DNA extraídas segundo protocolo padrão (Sambrook *et al.*, 1989). No caso da população de índios colombianos Coyaima amostras de sangue foram enviadas e extraídas pelo método de guanidina com algumas modificações, utilizando o Kit de extração DNeasy<sup>®</sup> - Quiagen.

### **3.3 Tipagem dos Eventos Polimórficos Únicos (UEPs)**

Para a maioria dos UEPs foram utilizados protocolos baseados em PCR previamente estabelecidos em nosso laboratório (Santos *et al.*, 1999, Tarazona-Santos *et al.*, 2001). Para detecção da variação, ou seja, identificação dos alelos (ancestral ou derivado), utilizamos diferentes métodos como enzima de restrição e/ou mini-seqüenciamento (ver Tabela 2).

**TABELA 2-** Seqüência dos iniciadores dos UEPs utilizados.

UEP	Iniciadores direto e reverso (5' -- 3')	Tamanho do produto	Modo de detecção	Tamanho dos fragmentos da digestão		Haplogrupo
				Alelo ancestral	Alelo derivado	
RPS4Y <sup>1</sup>	CCACAGAGATGGTGTGGGTA GAGTGGGAGGGACTGTGAGA	528 pb	<i>BsII</i> * Mini-seqüenciamento	234 +154 + 140 pb	388 +140 pb	C
YAP <sup>2</sup>	CAG GGG AAG ATA AAG AAA TA ACT GCT AAA AGG GGA TGG AT	150 ou 450 pb	Em gel **	150 pb	450 pb	D-E*
SRY- 8299 <sup>3</sup>	ACAGCACATTAGCTGGTATGAC TCTCTTTATGGCAAGACTTACG	509 pb	Mini-seqüenciamento	-	-	E*
sY81 <sup>4</sup>	AGGCACTGGTCAGAATGAAG AATGGAAAATACAGCTCCCC	209 pb	<i>Hsp92II</i> * Mini-seqüenciamento	102 + 65 + 42 pb	144 + 65 pb	E3a
92R7 <sup>5</sup>	TGCATGAACACAAAAGACGTA GCATTGTTAAATATGACCAGC	55 pb	<i>Hind III</i> *	30 + 25 pb	55 pb	P*
P36 <sup>6</sup>	TGA AGG ACA GTA AGT ACA CA TAA GTC CAT TGA TCT ACA GA	627pb	Mini-seqüenciamento	-	-	Q*
DYS199 <sup>6</sup>	TAATCAGTCTCCTCCCAGCA AGGTACCAGCTCTTCCCAATT	202 pb	<i>MfeI</i> * Mini-seqüenciamento	182 + 20 pb	202 pb	Q3
M207 <sup>6</sup>	AGGAAAAATCAGAAGTATCCC TTAATCCTATGCCTATGGAA	338 pb	Mini-seqüenciamento	-	-	R*
M173 <sup>6</sup>	TTACAATTCAAGGGCATTAG AAGGTGTATCTGGCATCCGTT	252 pb	Mini-seqüenciamento	-	-	R1

\*enzima de restrição, \*\*gel de poliacrilamida 6% ou agarose 1%, detecção presença ou ausência do retroelemento. 1-Bergen *et al.*, (1999); 2-Hammer e Horai (1995); 3-Whitfield *et al.*, (1995); 4-Seielstad *et al.*, (1994); 5-Santos *et al.*, (1999); 6-Underhill *et al.*, (1996) e Santos *et al.*, (1999).

Esses polimorfismos foram analisados em reações separadas de PCR, em um volume final de 12,5 µl, contendo de 20-50 ng de DNA genômico, 0,5-1,0 µM de cada iniciador, 200 µM de dNTPs , 0,5-1 unidade de *Taq* DNA Polimerase (Phonetrria) em tampão de reação contendo Tris-HCl 10 mM, pH 8,5, 1,5 mM de MgCl<sub>2</sub>. As reações de amplificação foram realizadas nos termocicladores PT-100 MJResearch e/ou Eppendorf programados para uma desnaturação inicial a 94°C por 3 minutos, seguida de 29-34 ciclos com temperaturas de 94°C por 20 segundos, 51 -55°C por 30-60 segundos, 72°C por 30 segundos, e uma extensão final a 72°C por 5 minutos.

### 3.4 Tipagem por enzima de restrição

No caso dos UEPs detectados por enzima de restrição, 5 µl do produto de PCR foram digeridos com uma enzima de restrição específica que permitia diferenciar os alelos alternativos. O produto da digestão (10 µl) era submetido à eletroforese em gel de poliacrilamida. Na Tabela 3 estão relacionadas às condições de digestão de cada enzima.

**TABELA 3** - Enzimas de restrição utilizadas nas análises dos UEPs.

ENZIMA	FABRICANTE/TAMPÃO	TEMPERATURA
<i>BsII</i>	New England Biolabs NEBuffer 3	55°C
<i>Hsp92II</i>	Promega Buffer K	37°C
<i>HindIII</i>	Promega Buffer E	37°C
<i>MfeI</i>	New England Biolabs NEBuffer 4	37°C

### 3.5 Tipagem por mini-seqüenciamento

Alguns SNPs foram analisados pela técnica de mini-seqüenciamento (Syvanen *et al.*, 1999) com algumas modificações. Nesta técnica uma DNA polimerase (*Thermo sequenase*) é usada especificamente para estender um iniciador que anela imediatamente adjacente à posição do nucleotídeo polimórfico a ser analisado. Utilizando-se um didesoxinucleotídeo (ddNTP-F) marcado com fluorescência e complementar ao nucleotídeo variante. A extensão do iniciador irá finalizar assim que o ddNTP adicionado encontrar seu nucleotídeo complementar, adiciona-se dNTPs para continuar a extensão até a base complementar do ddNTP-F utilizado, assim permitindo a detecção do alelo através do aparecimento de um pico de fluorescência de determinado tamanho, de acordo com a presença do estado alélico alternativo. Os alelos variantes correspondem a distintos tamanhos dos produtos de mini-seqüenciamento, cuja leitura é feita em seqüenciador automático (Figura 10). Mais de um UEP pode ser analisado desta maneira em uma mesma canaleta de gel de seqüenciamento, desde que os fragmentos alélicos não se sobreponham. A reação de mini-seqüenciamento foi feita em 12,5 µl de reação, utilizando 2 µl de produto de PCR multiplex ou individual purificado e 0,5 µM de

iniciadores. O mix contém 0,5 mM de ddATP (NEL402) ou ddGTP (NEL403) marcados com fluorescência (Perkin Elmer Life Sciences, USA), 0,5 mM de dCTP, dTTP, dGTP ou dATP não marcados, tampão 1X da enzima Thermo Sequenase, 1 U de Thermo Sequenase DNA Polimerase (Amersham Biosciences, Sweden). A reação foi feita como “hot start”, ou seja, a enzima só foi acrescentada após 2 minutos de desnaturação inicial. A reação de amplificação foi: 80°C por 5 min, 30 ciclos a 95°C por 30 seg, 55°C por 30 seg, 72°C por 20 seg.

### **3.6 Eletroforese em gel de poliacrilamida corado por sais de prata**

A eletroforese em gel de poliacrilamida era realizada para tipagem dos polimorfismos utilizando enzimas de restrição, para observação da qualidade dos produtos de PCR e análise de alguns microssatélites. Os géis de poliacrilamida não desnaturante eram feitos na concentração de 4-12%, variando de acordo com a resolução necessária para visualizar os polimorfismos e o tamanho do fragmento analisado, em tampão TBE 1X (90mM Tris-Borato, 2mM EDTA, pH 8,0). Alíquotas de 5 µl dos produtos de PCR ou da digestão foram misturadas a um volume igual de tampão de amostra (0,25 % de azul de bromofenol, 0,25 % de xilenocianol e 50 % de glicerol), aplicadas no gel e submetidas à eletroforese a 100V (25 a 30 mA) por aproximadamente duas horas. Os géis foram posteriormente corados pelo método de nitrato de prata (Santos *et al.*, 1995): foram incubadas por 10 minutos à temperatura ambiente em solução fixadora (etanol 10 % e ácido acético 0,5 % v/v), seguindo-se de uma incubação por 10 minutos em solução de nitrato de prata 0,2 % (p/v) e a revelação em uma solução de NaOH 3 % (p/v) e Formaldeído 1,1% (v/v a partir de solução 37 % p/v) durante o tempo necessário para o aparecimento das bandas. A revelação era interrompida com a solução fixadora. Posteriormente os géis eram secos e armazenados envoltos por papel celofane permeável.

### **3.7 Tipagem dos Microssatélites**

Foram analisados cinco microssatélites ligados ao cromossomo Y: *DYS19*, *DYS389a*, *DYS389b*, *DYS391* e *DYS393*. Os locos *DYS389a*, *DYS389b* e *DYS393* foram detectados por PCR em uma única reação de amplificação (triplex).



Os microssatélites *DYS389a* e *DYS389b* são amplificados com um único par de iniciadores, por serem produtos de uma duplicação parcial, fornecendo produtos de amplificação com tamanhos diferentes (Cooper *et al.*, 1996). A outra reação foi feita em duplex com *DYS19* e *DYS391*, analisados em gel de poliacrilamida 8% não desnaturante. Os iniciadores utilizados estão especificados na Tabela 4 e as condições de PCR foram: volume final de 12,5 µl contendo 20-50 ng de DNA genômico, 0,5 µM de cada iniciador, 200 µM de dNTPs e 0,625 U de *Taq* DNA polimerase (Phoentria).

A tipagem dos microssatélites *DYS389a*, *DYS389b* e *DYS393* foi realizada utilizando o seqüenciador automático ALF (Automated Labeled Fluorescent DNA sequencer - Pharmacia Biotech, Uppsala, Sweden). Depois da amplificação, cerca de 1 µl do produto foi misturado a 2 µl de tampão de formamida. A mistura foi fervida por 5 minutos e submetida à eletroforese em gel de poliacrilamida desnaturante 6% com 6 M de uréia, a 1500 V, 45 mA, 34 W e 45°C durante 420 minutos. Os alelos foram analisados usando o programa AlleLinks ver 1.0 (Pharmacia Biotech, Uppsala, Suécia). Os alelos de cada loco foram identificados usando uma escada alélica de 50-500 pares de bases variando 50 pb (Amershan Biosciences) cada banda, e marcadores internos com tamanhos diferentes dos microssatélites tipados.

**Tabela 4**-Seqüência dos iniciadores dos microssatélites analisados.

<i>DYS19</i>	5' CTA CTG AGT TTC TGT TAT AGT 3' 5' ATG GCA TGT AGT GAG GAC A 3'
<i>DYS391</i>	5' CTA TTC ATT CAA TCA TAC ACC CA 3' 5' GAT TCT TTG TGG TGG GTC TG 3'
<i>DYS389*</i>	5' FTC CAA CTC TCA TCT GTA TTA TCT ATG 3' 5' TCT TAT CTC CAC CCA CCA GA 3'
<i>DYS393*</i>	5' FGT GGT CTT CTA CTT GTG TCA ATA C 3' 5' AAC TCA AGT CCA AAA AAT GAG G 3'

**Condições de amplificação:** 30 ciclos a 51°C por 20 segundos, extensão a 72°C por 30 seg. e desnaturação a 94°C por 20 seg.

\*Iniciadores com fluorescência (*DYS389* e *DYS393*).

### **3.8 Análise dos Dados Populacionais com o Y**

Todas as análises populacionais foram feitas através do programa Arlequin (Schneider *et al.*, 2000). A variabilidade intra-populacional foi estimada utilizando cálculos de diversidade de Nei (Nei, 1987). Para análise de divergência inter-populacional foi feito um teste exato de diferenciação baseado na frequência dos haplótipos utilizando 10.000 passos de cadeias de Markov (Raymond e Rousset, 1996). Para cálculos dos componentes da variância genética total foi utilizada a análise de variância molecular (AMOVA). A AMOVA foi feita com agrupamentos das populações em um (considerando as cinco populações) ou três grupos (separadas de acordo com as línguas: quéchua, cayapa e coyaima).

### **3.9 Tipagem e análise do DNA Mitocondrial**

#### **3.9.1 Seqüenciamento da região controle**

A seqüência do segmento hipervariável I da Região controle do DNA mitocondrial (HVSI), foi determinada para todos os indivíduos da população Coyaima da Colômbia (21 indivíduos). A amplificação inicial por PCR foi realizada com os iniciadores L15926 5'-TCAAAGCTTACACCAGTCTTGTAACC-3' e H16498 5'-CCTGAAGTAGGAACCAGATG-3'. As amplificações foram realizadas em volume final de 12,5 µl contendo 0,5 µM dos iniciadores, 200 µM de cada dNTP e 0,5 unidade de *Taq* DNA polimerase (Phonectria) em tampão de reação contendo 10mM Tris-HCl pH 8,5, 1,5 mM MgCl<sub>2</sub>. Em cada tubo de PCR foi adicionado 20ng de amostra de DNA genômico. O programa de amplificação executava uma etapa inicial de desnaturação de 3 min a 94°C, seguida de 30 ciclos de desnaturação por 1 minuto, anelamento a 55°C por 30 segundos e extensão a 72°C por 1 minuto. Em cada grupo de reação de PCR foram feitos controles negativos (sem adição de DNA) para verificar contaminação. Os produtos de PCR de 612 pb foram visualizados em géis de agarose 1% corados com brometo de etídio (0,10 µg/ml). Confirmada amplificação, os produtos de PCR eram purificados utilizando as enzimas exonuclease I (Exol) e SAP (Shrimp Alkaline Phosphatase). Uma mistura de 1 U de SAP e 10 U de Exol é colocada na proporção de 1:1 com o produto de PCR e a reação é incubada a 37°C durante 45 minutos e 80°C por 15 minutos. Foi realizado o seqüenciamento direto a partir de produtos de PCR das fitas direta e reversa de cada amostra. As reações de seqüenciamento foram realizadas utilizando Kits DYEnamic ET<sup>®</sup> Terminator (Amershan Biosciences) com protocolo sugerido pelo fabricante e

posteriormente corridas no seqüenciador automático capilar de alto desempenho MegaBACE1000® (Amershan Biosciences).

### *3.9.2 Análise das seqüências do DNA mitocondrial*

A qualidade das seqüências foi verificada inicialmente no programa Sequence Analyzer 2.4 que acompanha o sequenciador de DNA e posteriormente utilizando o programa Phred v.0.20425 (Ewing *et al.*, 1998). Os fragmentos sequenciados foram montados e alinhados no programa Phrap v.0.990319 (Green, 1994), e posteriormente visualizados e editados para geração de seqüências consenso no programa Consed 12.0 (Gordon *et al.*, 1998).

As seqüências consenso de cada indivíduo foram posteriormente alinhadas entre si com o programa ClustalX 1.81 (Thompson *et al.*, 1997). As análises filogenéticas das seqüências eram feitas no programa MEGA 3.0 (Kumar, *et al.*, 2001) através do algoritmo neighbor-joining de construção de árvores.

### *3.9.3 Mini-seqüenciamento do DNA mitocondrial*

Vários polimorfismos do DNAm<sub>t</sub> abrangem sítios de restrição dispersos por todo genoma mitocondrial. Para uma análise mais específica utilizamos iniciadores desenhados para amplificar regiões restritas de polimorfismos de nativos americanos que eram identificados com o método de mini-seqüenciamento. Primeiramente era feita amplificação das regiões de interesse para o mini-seqüenciamento segundo Parra (2003) sendo que o tamanho médio dos produtos de amplificação era em torno de 800 pb. As amplificações foram realizadas em volume final de 12,5 µl contendo 0,25 µM dos iniciadores, 200 µM de cada dNTP e 1 unidade de *Taq* DNA polimerase (Phoentria) em tampão de reação contendo Tris-HCl 10 mM pH 8,5, 1,5mM de MgCl<sub>2</sub>. O programa de amplificação executava uma etapa inicial de desnaturação de 5 min a 94°C, seguida de 45 ciclos de desnaturação a 94°C por 1 min, anelamento a 55°C por 1 min e extensão a 72°C por 2 min, com uma extensão final de 10 min. Após verificação da amplificação em gel de agarose 1% ou de poliacrilamida 6% era feita a purificação dos produtos de PCR.

A técnica mini-seqüenciamento foi feita utilizando iniciadores desenhados imediatamente adjacentes ao nucleotídeo polimórfico, utilizando um ddNTP na reação, sendo um ddGTP-F ou um ddATP-F (F=fluorescentes). Na reação era promovida a extensão do iniciador até a primeira base completar ao ddNTP marcado e a presença do produto (pico) fluorescente era dependente do alelo. Os iniciadores foram desenhados

com cauda de tamanhos variados, equivalentes à seqüência da região M13 do pUC18 (ver tabelas 5 e 6), permitindo a produção de um sistema multiplex. Foram analisados 13 polimorfismos, cada um com um par de iniciadores e para diferenciar os alelos A e G do mesmo polimorfismo foram acrescentados 2 pb de diferença entre as caudas. Para polimorfismos de A para T, foi utilizada a fita complementar para o desenho do iniciador.

**Tabela 5-**Seqüências dos iniciadores externos para amplificação das regiões de interesse no DNAm<sup>t</sup> a serem analisadas por mini-seqüenciamento.

<i>H a p l o g r u p o</i>	<i>P o l i m o r f i s m o</i>	<i>I n i c i a d o r</i>	<i>S e q u ê n c i a</i>
<b>A - A 2</b>	G 8 0 2 7 A	1 1 F 7 1 4 8	G T A A A A C G A C G G C C A G T A C G C C A A A A T C C A T T T C A C T
		1 1 R 8 0 9 5	C A G G A A A C A G C T A T G A C C C G G G A A T T G C A T C T G T T T T T
	G 1 2 0 0 7 A	1 7 F 1 1 3 1 4	G T A A A A C G A C G G C C A G T T C A C T C T C A C T G C C C A A G A A
		1 7 R 1 2 0 7 6	C A G G A A A C A G C T A T G A C C G G A G A A T G G G G G A T A G G T G T
C 8 7 9 4 T	1 3 F 8 6 2 1	G T A A A A C G A C G G C C A G T T T T C C C C C T C T A T T G A T C C C	
	1 3 R 9 3 9 7	C A G G A A A C A G C T A T G A C C G T G G C C T T G G T A T G T G C T T T	
<b>B - B 4 b 1</b>	C 1 5 5 3 5 T	2 2 F 1 4 8 5 6	G T A A A A C G A C G G C C A G T T G A A A C T T C G G C T C A C T C C T
		2 2 R 1 5 9 7 8	C A G G A A A C A G C T A T G A C C A G C T T T G G G T G C T A A T G G T G
	C 1 1 1 7 7 T	1 6 F 1 0 6 7 2	G T A A A A C G A C G G C C A G T G C C A T A C T A G T C T T T G C C G C
		1 6 R 1 1 4 7 2	C A G G A A A C A G C T A T G A C C T T G A G A A T G A G T G T G A G G C G
	A 3 5 4 7 G	5 F 3 1 6 9	G T A A A A C G A C G G C C A G T T A C T T C A C A A A G C G C C T T C C
		5 R 3 9 6 1	C A G G A A A C A G C T A T G A C C A T G A A G A A T A G G G C G A A G G G
	T 4 9 7 7 C	7 F 4 4 8 5	G T A A A A C G A C G G C C A G T A C T A A T T A A T C C C C T G G C C C
		7 R 5 4 2 0	C A G G A A A C A G C T A T G A C C C C T G G G G T G G G T T T T G T A T G
	C 6 4 7 3 T	9 F 5 8 5 5	G T A A A A C G A C G G C C A G T G A G G C C T A A C C C C T G T C T T T
		9 R 6 6 4 2	C A G G A A A C A G C T A T G A C C A T T C C G A A G C C T G G T A G G A T
T 9 9 5 0 C	1 4 F 9 2 3 0	G T A A A A C G A C G G C C A G T C C C A C C A A T C A C A T G C C T A T	
	1 4 R 1 0 1 3 0	C A G G A A A C A G C T A T G A C C T G T A G C C G T T G A G T T G T G G T	
<b>C - C 1</b>	A 1 5 4 8 7 T *	2 2 F 1 4 8 5 6	G T A A A A C G A C G G C C A G T T G A A A C T T C G G C T C A C T C C T
		2 2 R 1 5 9 7 8	C A G G A A A C A G C T A T G A C C A G C T T T G G G T G C T A A T G G T G
	T 1 4 3 1 8 C	2 1 F 1 4 0 0 0	G T A A A A C G A C G G C C A G T G C A T A A T T A A A C T T T A C T T C
		2 1 R 1 4 9 9 8	C A G G A A A C A G C T A T G A C C A G A A T A T T G A G G C G C C A T T G
<b>D - D 1</b>	C 2 0 9 2 T	3 F 1 8 5 4	G T A A A A C G A C G G C C A G T G G A C T A A C C C C T A T A C C T T C T G C
		3 R 2 6 6 9	C A G G A A A C A G C T A T G A C C G G C A G G T C A A T T T C A C T G G T
	C 8 4 1 4 T	2 1 F 1 4 0 0 0	G T A A A A C G A C G G C C A G T G C A T A A T T A A A C T T T A C T T C
		2 1 R 1 4 9 9 8	C A G G A A A C A G C T A T G A C C A G A A T A T T G A G G C G C C A T T G
<b>D - D 2</b>	C 1 1 2 1 5 T	1 6 F 1 0 6 7 2	G T A A A A C G A C G G C C A G T G C C A T A C T A G T C T T T G C C G C
		1 6 R 1 1 4 7 2	C A G G A A A C A G C T A T G A C C T T G A G A A T G A G T G T G A G G C G

\*Para detectar este polimorfismo utilizamos a fita complementar para desenhar o iniciador

Tabela transcrita de Parra (2003).

**Tabela 6.**– Polimorfismos, iniciadores e condições ideais para PCR para o estudo do DNAmT de nativos americanos através de mini-seqüenciamento.

<b>Haplog. Polimorf.</b>	<b>Cauda - pUC 18 - início M13univ</b>		<b>Iniciador de MS</b>	<b>ddNTP</b>
<b>A - A2</b>	G8027A	1g	GTAA AAC gagtagtactccccgattgaa	ddGTP
		1a	GTAAA gagtagtactccccgattgaa	ddATP
	G12007A	2g	GTAA AACGACGG atatttaccacaacacaatg	ddGTP
		2a	GTAA AACGAC atatttaccacaacacaatg	ddATP
	C8794T	3g	GTAA AACGACGGCCAGT GGTGGTTGGTGTAAATGAGT	ddGTP
		3a	GTAA AACGACGGCCA GGTGGTTGGTGTAAATGAGT	ddATP
<b>B - B4b1</b>	C15535T	4g	GTAA AACGACGGCCAGTGCCAA GGCTTGATGTGGGGAGGGGT	ddGTP
		4a	GTAA AACGACGGCCAGTGCC GGCTTGATGTGGGGAGGGGT	ddATP
	C11177T	5g	GTAA AACGACGGCCAGTGCCAAGCTTG GCCTGCGTTTCAGGCGTTCTG	ddGTP
		5a	GTAA AACGACGGCCAGTGCCAAGCT GCCTGCGTTTCAGGCGTTCTG	ddATP
	A3547G	6a	GTAA AACGACGGCCAGTGCCAAGCTTGCCAT ccccgaccttagctctcacc	ddATP
		6g	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGC ccccgaccttagctctcacc	ddGTP
	T4977C	7a	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTG CCGTAGCTGGGTTTGGTTTA	ddATP
		7g	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCA CCGTAGCTGGGTTTGGTTTA	ddGTP
	C6473T	8g	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGC AGGAGAAGTAGGACTGCTGT	ddGTP
		8a	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGT AGGAGAAGTAGGACTGCTGT	ddATP
	T9950C	9a	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGT GAGACATACAGAAATAGTCA	ddATP
		9g	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCT GAGACATACAGAAATAGTCA	ddGTP
<b>C - C1</b>	A15487T *	10a	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCTAGA acattaacactattctcacc	ddATP
		11a	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCTAGAGG TCTGGGTCGCCTAGGAGGTC	ddATP
	T14318C	12a	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCTAGAGGATC GGTGGTTGTGGTAAACTTTA	ddATP
		12g	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCTAGAGGATCCC GGTGGTTGTGGTAAACTTTA	ddGTP
<b>D - D1</b>	C2092T	13g	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCTAGAGGATCCCCGGT TGTTCCTCTTTGGACTAACA	ddGTP
		13a	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCTAGAGGATCCCCGG TGTTCCTCTTTGGACTAACA	ddATP
	C8414T	14g	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCTAGAGGATCCCCGGGTACCGA GATGAGGAATAGTGTAAAGGA	ddGTP
		14a	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCTAGAGGATCCCCGGGTACCG GATGAGGAATAGTGTAAAGGA	ddATP
<b>D - D2</b>	C11215T	15g	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCTAGAGGATCCCCGGGTACCGAGCTCG GGAAGGGAGCCTACTAGGGT	ddGTP
		15a	GTAA AACGACGGCCAGTGCCAAGCTTGCCATGCCTGCAGGTGCAGTCTAGAGGATCCCCGGGTACCGAGCT GGAAGGGAGCCTACTAGGGT	ddATP

Tabela transcrita de Parra (2003).

### **3.10 Desenvolvimento do Multiplex de SNPs do cromossomo Y**

Para produção de um multiplex de SNPs do cromossomo Y foi utilizada a técnica de mini-seqüenciamento descrita na seção 3.5 da metodologia. Um PCR de multiplex inicial é feito com todos os marcadores, visualizamos em gel de poliacrilamida 6 %, purificamos o produto de PCR com *exo/sap* e fazemos uma segunda reação de multiplex com todos iniciadores do mini-seqüenciamento que possuem cauda de M13 que permite diferenciação dos marcadores e evita a sobreposição na utilização do sistema de multiplex. Os SNPs foram selecionados de modo que abrangessem os principais haplogrupos do cromossomo Y de acordo com YCC (2002), mas com ênfase especial naqueles informativos para o povoamento das Américas (Figura 11).

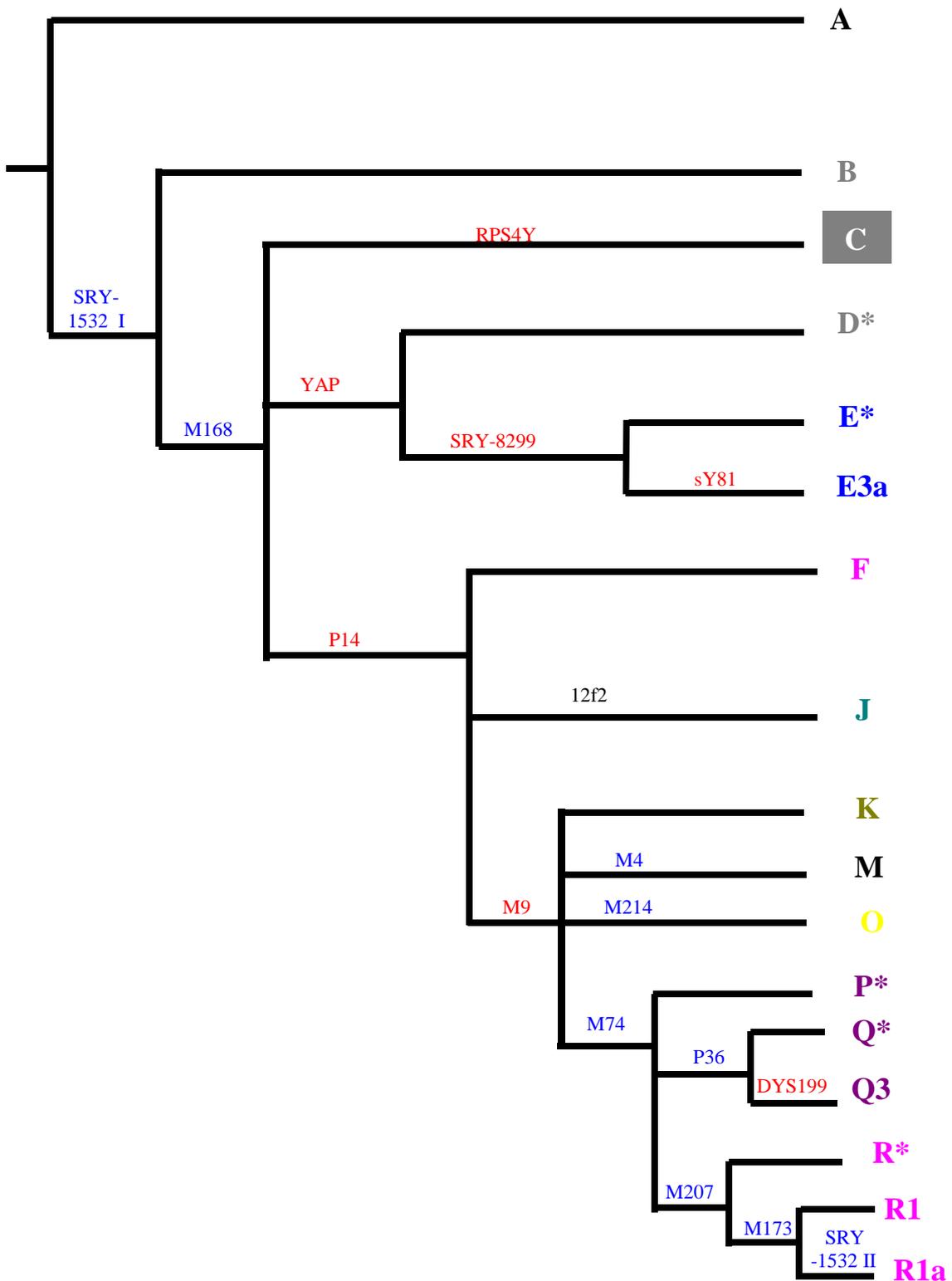
Foram padronizados dois sistemas de multiplexes: sistema 1- utilizando ddGTP-F com seis marcadores, e sistema 2- utilizando ddATP-F com oito marcadores. Em cada sistema os iniciadores foram desenhados de acordo com o nucleotídeo polimórfico e caudas de M13 foram acrescentadas nos iniciadores do mini-seqüenciamento para permitir a montagem do multiplex (ver tabelas 7 e 8).

**Tabela 7.** Seqüência dos iniciadores do multiplex, sistema 1 (utilizando ddGTP-F\*).

UEP	Tamanho produto de PCR (pb)	Cauda M13	Iniciador Mini-sequenciamento	Tamanho total do iniciador	Tamanho do produto final	
					Alelo ancestral	Alelo derivado
RPS4Y	528	-	GCC CTT TCC CCT GGG CAG 18 nt	18 nt	19 nt	20 nt
SRY-8299	509	-	AAA GTC TGA CAA CCT TCG AGA GGT CAA GGC 30 nt	30 nt	31 nt	33 nt
P14	414	GTA AAA CGA CGG 12 nt	CAT CTA TCT GCA TAT GGA AAA TG 23 nt (reverso)	35 nt	36 nt	45 nt
M9	340	GTA AAA CGA CGG CC 14 nt	TGT CTA AAT TAA AGA AAA ATA AAG AG 26 nt (reverso)	40 nt	41 nt	56 nt
sY81	209	-	TTA TAT TTC ATT GTT AAC AAA AGT CC 26 nt	26 nt	29 nt	27 nt
DYS199	202	-	AGG TAC CAG CTC TTC CTA ATT (reverso) 21 nt	21 nt	22 nt	26 nt

**Tabela 8.** Seqüência dos iniciadores do multiplex, sistema 2 (utilizando ddATP-F\*).

UEPs	Tamanho produto de PCR (pb)	Cauda M13	Iniciador mini-sequenciamento	Tamanho total do iniciador	Tamanho do produto final	
					Alelo ancestral	Alelo Derivado
P36	635	GTA A 4 nt	TCA TCT ATC TAT CCA TTA TTC TCT CT (reverso) 26 nt	30 nt	33 nt	31 nt
M214	460	GTA AAA CGA CGG CCA GTG CCA AGC TTG CAT GC 32 nt	GTG AGA CAC TGT CTG AAA ACA AC (reverso) 23 nt	55 nt	56 nt	58 nt
M74	385	GTA AAA CGA CGG C 13 nt	AAT AGA TTA GGT ATT TTA AAA ACT GGT 27 nt	40 nt	41 nt	43 nt
M207	338	-	AAT GTA AGT CAA GCA AGA AAT TTA G 25 nt	25 nt	26 nt	27 nt
M4	273	GTA AAA CGA CGG CCA GTG CCA A 22 nt	CAA ACT AAT GAT TAA TGA TAT TCA TAT 27 nt	49 nt	50 nt	52 nt
M168	261	GTA AAA CGA C 10 nt	CTA TTG TTT TAA TTC TTC AGC TAG C (reverso) 25 nt	35 nt	37 nt	36 nt
M173	252	GTA AAA CGA CGG CCA GTG CCA AGC TTG CAT GCC TG 35 nt	CTT ACA ATT CAA GGG CAT TTA GAA C 25 nt	60 nt	61 nt	63 nt
SRY-1532	167	GTA AAA CGA CGG CCA GTG CC 20 nt	TTG TAT CTG ACT TTT TCA CAC AGT 24 nt	44 nt	45 nt	47 nt



**Figura 11** - Árvore filogenética do cromossomo Y com os UEPs utilizados no multiplex de SNPs. Em vermelho sistema multiplex 1 utilizando ddGTP-F, em azul sistema de multiplex 2 utilizando ddATP-F e em preto tipagem pelo método convencional.

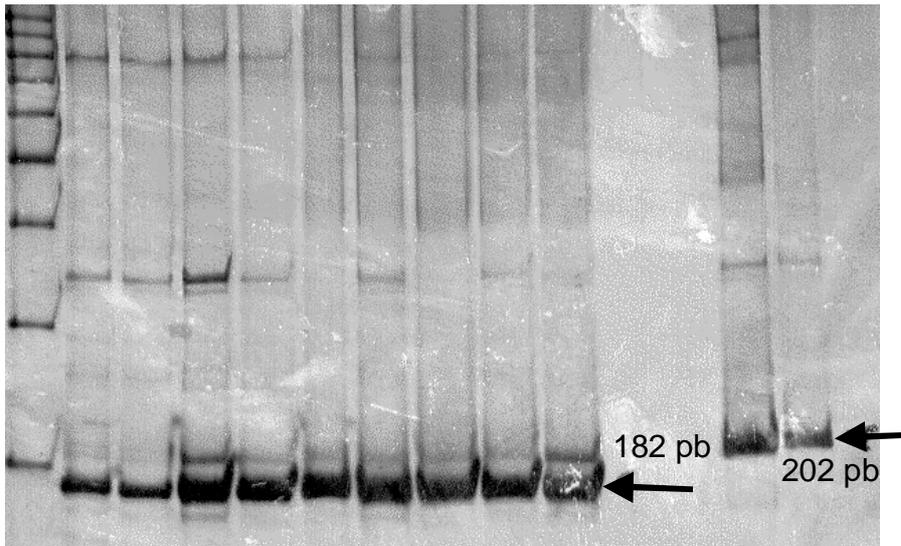
# 4 Resultados

## 4.1 Análise dos UEPs nas populações nativas americanas

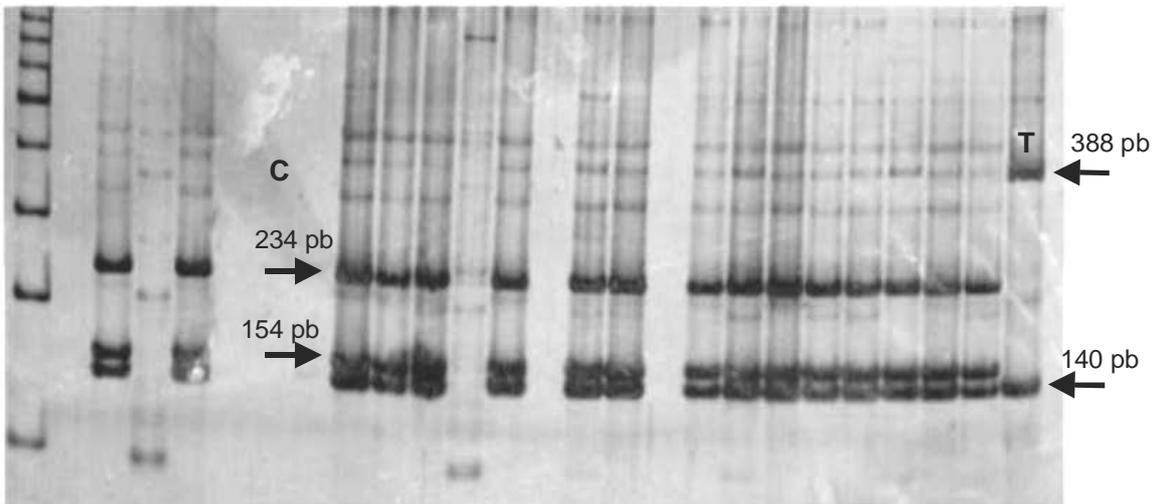
Foram analisados os seguintes UEPs: YAP, RPS4Y, SRY-8299, sY81, 92R7, P36, *DYS199*, M207 e M173; em 240 indivíduos das cinco populações andinas estudadas.

Os UEPs analisados definem os seguintes haplogrupos continente específicos: RPS4Y, define haplogrupo C, encontrado em asiáticos e alguns raros nativos americanos, principalmente na América do Norte; YAP determina o haplogrupo D-E\*; que ocorre na Ásia (principalmente Japão); SRY-8299, haplogrupo E\*, que ocorre na Eurásia e África; sY81, haplogrupo E3a, quase exclusivo de populações africanas; 92R7, define haplogrupo P\*, presente em populações eurasiáticas; P36, haplogrupo Q\*, ocorre em asiáticos e também em alguns nativos americanos; *DYS199* define o haplogrupo Q3, praticamente exclusivo dos nativos americanos; M207, o haplogrupo R\*, e M173, haplogrupo R1, presentes em populações européias e asiáticas; consideramos como haplogrupo Y\*, os cromossomos que não foram inclusos em nenhum haplogrupo definidos pelos UEPs analisados. Utilizando esses UEPs foi possível medir a contribuição de diferentes populações na variabilidade genética do cromossomo Y das populações andinas estudadas. A tabela 9 mostra os distintos haplogrupos definidos pelos UEPs estudados e a tabela 10, a frequência de cada haplogrupo nas populações andinas. Nas populações ameríndias estudadas não foi observada a presença dos haplogrupos C, D e R1.

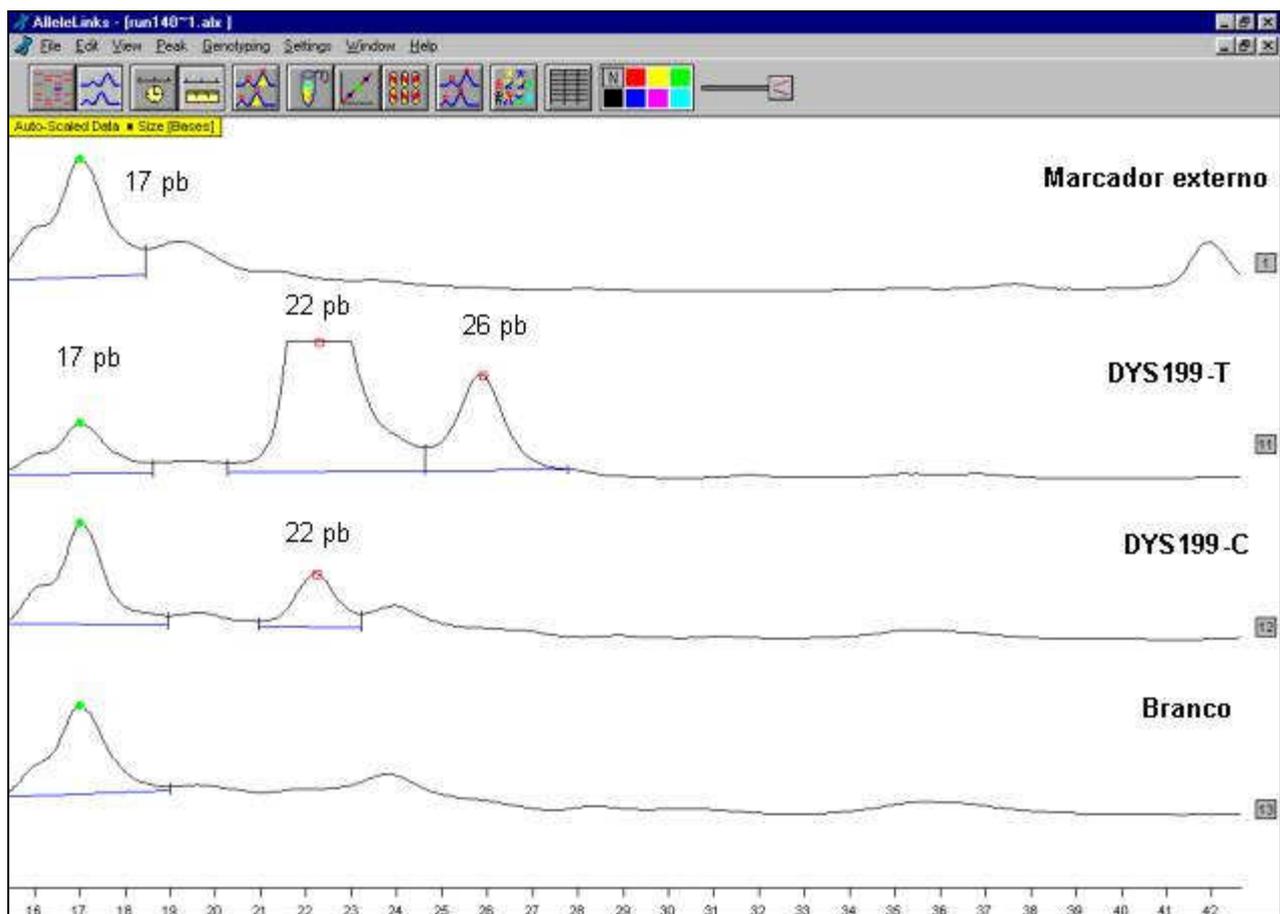
As figuras 12 e 13 apresentam as tipagens das mutações *DYS199* e RPS4Y, por digestão com enzima de restrição. Para a análise da mutação *DYS199* com enzima de restrição foi utilizado um iniciador que produz um sítio artificial para a enzima *MfeI* (Santos *et al.*, 1999) quando aparece o estado ancestral (*DYS199-C*), que por ser uma enzima rara e com custo elevado padronizamos também o mini-seqüenciamento deste marcador. Usamos os mesmos iniciadores utilizados para tipagem com enzima de restrição para a reação de PCR. Com o iniciador reverso foi feita também a reação de mini-seqüenciamento usando didesoxinucleotídeos (ddGTP-F) marcados com fluorescência (Figura 14).



**Figura 12.** Gel de poliacrilamida 6% com visualização da tipagem por enzima de digestão da mutação *DYS199*. Alelo ancestral *DYS199-C* (182 pb) e alelo derivado *DYS199-T* (202 pb).



**Figura 13.** Gel de poliacrilamida 6% com tipagem da mutação *RPS4Y* utilizando a enzima de restrição *BsaI*. *RPS4Y-C*, alelo ancestral digerido gera fragmentos de 140, 154 e 234 pb; *RPS4Y-T*, alelo derivado gera fragmentos de 140 e 388 pb.



**Figura 14.** Mini-sequenciamento do *DYS199*. Canaleta 1 - marcador externo 50-500 pb, com marcador interno de 17 pb; Canaleta 11 - indivíduo *DYS199-T* com alelo derivado apresenta um pico de 22 pb e um de 26 pb; Canaleta 12 - indivíduo *DYS199-C*, alelo ancestral apresenta apenas um pico de 22 pb. A presença do pico do alelo ancestral em indivíduos com alelo derivado é provavelmente devido a um artefato da técnica de mini-sequenciamento, mas a comparação com os dados de enzima de restrição confirma o estado alélico destes indivíduos.

**Tabela 9.** Haplogrupos determinados pelos UEPs analisados.

RPS4Y	YAP	SRY- 8299	sY81	92R7	P36	DYS199	M207	M173	N <sup>^</sup>	Haplogrupo
0	0	0	0	0	0	0	0	0	6	Y*
1	0	0	0	0	0	0	0	0	0	C
0	1	0	0	0	0	0	0	0	0	D-E*
0	1	1	0	0	0	0	0	0	6	E*
0	1	1	1	0	0	0	0	0	2	E3a
0	0	0	0	1	0	0	0	0	1	P*
0	0	0	0	1	1	0	0	0	12	Q*
0	0	0	0	1	1	1	0	0	183	Q3
0	0	0	0	1	0	0	1	0	31	R*
0	0	0	0	1	0	0	1	1	0	R1a

<sup>^</sup>N=Número de indivíduos. 0-representa alelo ancestral e 1-alelo derivado de cada UEP.

**Tabela 10.** Frequências dos diferentes haplogrupos do cromossomo Y nas populações andinas analisadas.

Populações	HAPLOGRUPOS							Número de cromossomos
	R*	E*	E3a	Y*	P*	Q*	Q3	
Tayacaja	21,6%	1,03%	1,03%	3,09%	1,03%	1,03%	71,2%	97
Arequipa	0	4,3%	0	0	0	0	95,7%	23
San Martin	10,3%	3,3%	0	5,1%	0	3,3%	78%	59
Coyaima	19%	9,5%	4,76%	0	0	42,8%	23,8%	21
Cayapa	0	0	0	0	0	0	100%	40

Nota: Em cinza aparecem os haplogrupos “autênticos” de Nativos Americanos.

Estudos prévios evidenciaram que o tipo de cromossomo Y mais comum, presente em no mínimo 70% dos nativos americanos é o haplogrupo Q3 definido pelo alelo *DYS199-T* (Underhill et al., 1996, Santos et al., 1999, Karafet et al., 1999). Nossos resultados confirmam esses estudos prévios da variabilidade do cromossomo Y em nativos americanos, sugerindo uma frequência média de 75,8% do haplogrupo Q3 nas populações analisadas. No entanto, a população de Coyaima da Colômbia apresenta uma frequência significativamente diminuída deste haplogrupo (23,8 %), que poderia ser uma consequência da miscigenação recente com outras populações, já que as frequências dos haplogrupos de origem européia (~24%; 4,5 - E\* + 19 - R) e africana (~9%; 4,5 - R + 4,76 - E3a) são altas em comparação às outras populações andinas (haplogrupo E\* ocorre tanto na África quanto Europa e foi atribuído a estas duas origens no cálculo aproximado de miscigenação), além disso, dados antropológicos desta

população corroboram nossos resultados. A princípio não seria possível distinguir os cromossomos DYS199-C nativos dos que chegaram após a conquista (europeus P\*), porém a utilização da mutação P36, que define o haplogrupo Q (Q\*+Q3), permite discriminar os cromossomos nativos dos que chegaram nos últimos 500 anos, introduzidos por outras populações (P\*). Na população Coyaima, observamos uma frequência alta de 42% do haplogrupo Q\*, o que demonstra que apesar da miscigenação confirmada pela herança paterna, essa população ainda apresenta alta frequência de cromossomos nativos americanos (mínimo de 66% = Q\* + Q3). O estado derivado das mutações RPS4Y e M173 que definem os haplogrupos C e R1, respectivamente, não foram observados nas populações andinas analisadas.

## **4.2 Populações da Ásia Central**

A Ásia Central tem sido vista como uma provável fonte de populações que vieram para as Américas (Santos *et al.*, 1999, Karafet *et al.*, 1999) e neste projeto procuramos por evidências destas hipóteses nestas populações. Indivíduos das populações de Kirghiz e Kazakos da Ásia central foram analisados com a mutação P36, que define o haplogrupo Q\*, para verificar a presença desse haplogrupo nessas populações. Em 2002, Karafet *et al.* verificaram a frequência de 18,8 % do haplogrupo Q\* em populações da Sibéria, os Kets e Selkups. Do total de 67 indivíduos da Ásia Central, 45 foram previamente classificados como do haplogrupo P (Tarazona-Santos, 2002), foram analisados dos quais apenas 24 foram tipados com P36, devido à qualidade do DNA, e nenhum indivíduo apresentou o alelo derivado desta mutação. A ausência do alelo derivado P36-T nas populações da Ásia Central sugere que essas populações não são ou fazem parte do grupo das populações derivadas das que fizeram a viagem inicial ao continente americano dando origem aos povos nativos americanos.

## **4.3 Variabilidade do haplogrupo Q nas populações Andinas**

Foram analisados 159 indivíduos do haplogrupo Q3 (cromossomos com o alelo DYS199-T) e 12 indivíduos do haplogrupo Q\* (cromossomos com o alelo P36-T e DYS199-C) das cinco populações estudadas, com os seguintes microssatélites do cromossomo Y: DYS389a, DYS389b, DYS19, DYS391 e DYS393. A Tabela 11 apresenta os resultados dos indivíduos do haplogrupo Q tipados com estes microssatélites. Foram observados 74 haplótipos de microssatélites presentes em cromossomos do haplogrupo

Q, sendo 72 haplótipos no haplogrupo Q3 e 6 haplótipos no haplogrupo Q\* (Q\* significa haplogrupo Q sem a presença do Q3), sendo 2 diferentes dos encontrados no Q3. A diversidade haplotípica total observada nas populações estudadas foi de  $0,978 \pm 0,0038$ . Tarazona-Santos (1999), utilizando duas populações dos Andes Centrais (Tayacaja e Arequipa) e seis microssatélites, observou uma diversidade total de  $0,964 \pm 0,013$ . Os índices de diversidade de Nei encontrados dentro do haplogrupo Q3 nas populações do Peru foram:  $0,9631 \pm 0,0103$  em Tayacaja,  $0,9105 \pm 0,053$  em Arequipa e  $0,9922 \pm 0,0071$  na população de San Martin. Na população Coyaima da Colômbia:  $0,800 \pm 0,1640$  e na população de Cayapa do Equador:  $0,9242 \pm 0,0319$ . Enquanto que diversidade encontrada dentro do haplogrupo Q, foi um pouco maior apenas na população de Coyaima ( $0,8791 \pm 0,0576$ ), devido à inclusão dos indivíduos do haplogrupo Q\*. As populações San Martin (Peru) e Coyaima (Colômbia) apresentaram os maiores e menores níveis de diversidade haplotípica respectivamente.

Dentro do Haplogrupo Q, observamos 46 haplótipos que apareceram apenas uma vez em determinada população, sendo exclusivos de uma única população.

**Tabela 11.** Haplótipos de microssatélites presentes nos indivíduos do haplogrupo Q, das populações analisadas e suas respectivas frequências.

Haplótipo	<i>DYS19</i>	<i>DYS389a</i>	<i>DYS389b</i>	<i>DYS391</i>	<i>DYS393</i>	Frequência
Y1	14	15	10	10	13	3
Y2	14	17	10	11	13	3
Y3	13	16	10	10	13	8
Y4	13	18	10	10	13	5
Y5	13	17	11	10	14	11
Y6	13	18	10	10	14	3
Y7	13	16	11	10	14	8
Y8	13	16	10	9	13	1
Y9	13	19	10	10	13	1
Y10	13	17	10	10	13	8
Y11	13	16	11	10	13	3
Y12	14	17	10	10	13	2
Y13	13	18	11	10	13	4
Y14	13	17	10	10	14	3
Y15	13	15	10	10	14	1
Y16	13	17	11	10	13	4
Y17	13	16	9	10	14	1
Y18	13	15	11	10	13	1
Y19	13	18	9	10	13	5
Y20	14	18	11	10	13	1
Y21	13	18	10	11	13	2
Y22	13	17	9	11	13	3
Y23	14	16	11	10	13	3
Y24	13	17	11	11	13	1
Y25	13	20	10	10	13	1
Y26	14	17	10	10	14	1
Y27	13	16	11	11	14	2
Y28	13	17	11	11	14	1
Y29	13	16	10	11	13	5
Y30	13	15	10	10	13	1
Y31	13	16	10	9	12	1
Y32	12	17	10	10	13	2
Y33	14	16	10	10	14	1
Y34	13	15	9	10	13	1
Y35	13	15	10	11	13	1
Y36	14	15	11	11	13	1
Y37	13	16	11	11	13	12
Y38	13	16	9	11	13	2
Y39	13	18	10	11	12	1
Y40	14	16	10	11	13	1
Y41	13	17	10	11	12	2
Y42	13	17	10	10	15	2
Y43	14	16	10	11	12	1
Y44	13	17	10	11	13	4
Y45	13	16	10	11	15	1
Y46	13	18	11	12	13	1
Y47	14	17	9	12	13	1
Y48	14	16	11	11	13	1
Y49	14	16	11	12	13	1

**Tabela 11.** continuação

Y50	14	17	10	12	13	1
Y51	14	16	9	11	13	1
Y52	13	16	10	10	14	1
Y53	13	17	9	12	13	1
Y54	13	16	10	12	13	1
Y55	13	16	11	11	15	1
Y56	13	16	11	10	15	2
Y57	13	16	11	12	15	1
Y58	13	15	11	10	15	1
Y59	13	16	10	11	14	1
Y60	14	16	11	11	14	1
Y61	14	15	10	11	13	1
Y62	13	15	11	11	13	2
Y63	12	17	10	10	12	4
Y64	13	19	11	10	13	1
Y65	13	18	9	11	13	1
Y66	13	18	10	11	14	1
Y67	13	18	11	11	13	1
Y68	13	17	10	11	14	1
Y69	12	17	12	10	12	1
Y70	13	16	12	11	13	1
Y71	13	18	11	11	14	1
Y72	11	17	11	10	14	1
Y73	12	16	14	13	11	1
Y74	12	17	13	13	10	1
Total						171

#### ***4.4 Variabilidade interpopulacional***

Para avaliar a partição da variabilidade genética entre as populações andinas, fizemos a análise da variância molecular (AMOVA) com base nos haplogrupos formados pelos UEPs analisados ou com base nos microssatélites analisados em indivíduos do haplogrupo Q (Q\*+ Q3) e Q3, cromossomos autenticamente nativos americanos. Haplogrupos P\* podem também ser cromossomos nativos, mas com os marcadores utilizados não podem ser discriminados de P\* europeus. Para a análise da AMOVA, utilizando UEPs ou microssatélites-Q ou Q3, agrupamos as cinco populações andinas em três grupos de acordo com a língua e localização geográfica: grupo 1- populações do Peru (Tayacaja, Arequipa e San Martin), grupo 2-população da Colômbia (Coyaima) e grupo 3-população do Equador (Cayapa). A tabela 12 apresenta o resultado da AMOVA, entre os três agrupamentos das populações analisadas para o cromossomo Y.

**Tabela 12.** Análise da Variância Molecular das Populações Andinas a partir de UEPs e de microssatélites dos haplogrupos Q, Q3 do cromossomo Y.

Origem da variação	gl	Soma de quadrados	Componentes da variância	Porcentagem de variabilidade	P>obs.
<b>Haplogrupos de UEPs</b>					
Entre grupos	2	6,193	0,05424 Va	23,18	0,09
Entre populações do Peru	2	1,120	0,00745 Vb	3,19	<0,05
Dentro das populações	235	40,479	0,17225 Vc	73,63	<0,05
<b>Haplótipos de microssatélites-Q</b>					
Entre grupos	2	2,650	0,01735 Va	3,47	<0,05
Entre populações do Peru	2	1,565	0,00805 Vb	1,61	<0,05
Dentro das populações	166	78,873	0,47514 Vc	94,93	<0,05
<b>Haplótipos de microssatélites-Q3</b>					
Entre grupos	2	2,321	0,01751 Va	3,49	<0,05
Entre populações do Peru	2	1,519	0,00760 Vb	1,52	<0,05
Dentro das populações	154	73,323	0,47613 Vc	94,99	<0,05
<b>Haplotipos de UEPs com dois grupos</b>					
Entre populações do Peru	4	6,138	0,00172 Vb	0,179	<0,05
Dentro das populações	235	54,589	0,49873 Vc	99,82	<0,05
<b>Haplotipos de UEPs ameríndios com dois grupos</b>					
Entre populações do Peru	4	7,998	0,00172 Vb	0,35	<0,05
Dentro das populações	235	47,632	0,45982 Vc	99,65	<0,05

A AMOVA confirma o alto nível de variabilidade dentro das populações, independente da origem dos haplótipos, entretanto, a variabilidade entre os grupos é maior do que entre as populações na análise com UEPs (apesar de não significativo  $P=0,09$ ), enquanto na análise com microssatélites do haplogrupo Q e Q3 a variação entre grupos é menor mas significativa ( $P<0,05$ ). A variabilidade entre e dentro das populações

utilizando todos os haplótipos de UEPs ( $F_{st}$  0,179) foi menor do que considerando apenas os haplogrupos ameríndios ( $F_{st}$  0,35), esse aumento pode ser devido a homogeneização das populações pelos haplogrupos miscigenados por fluxo gênico ou mesmo pelo fato dos haplogrupos miscigenados terem a mesma origem (Africana e Européia). A diferença entre os haplótipos de UEPs e microssatélites pode ser explicada pelas diferenças de taxas de mutação entre os marcadores utilizados e a escala evolutiva comparada, nos níveis intra e inter-populacional e inter-grupos. Devido à alta homoplasia, microssatélites perdem o sinal filogenético, ou seja, o poder de diferenciação entre os agrupamentos, e esta perda é maior nas comparações entre grupos de populações, que envolve teoricamente um tempo de separação evolutiva maior. Nas análises com UEPs apenas a variação intra-populacional é significativa e pode ser explicado pelo fato de que os haplogrupos são bem anteriores ao processo de diferenciação entre estas populações que possuem entre si composições similares de haplogrupos, mas que no nível inter-grupos está ligeiramente aumentada (~23%,  $P=0,09$ ) provavelmente porque os Coyaima e os Cayapas formam grupos separados e são um pouco diferenciados dos demais. O teste exato de diferenciação populacional foi realizado para verificar a divergência entre as populações comparadas aos pares com base na frequência haplotípica de UEPs observada, considerando  $P < 0,05$ , as populações de Coyaima e Cayapas são as que mais diferem entre si e com as outras populações (Tabelas 13). Enquanto que, analisando os haplótipos de microssatélites do haplogrupo Q, observamos um resultado similar aos UEPs mas com uma divergência ligeiramente maior entre a população de San Martin com as demais, inclusive do grupo quéchua (tabela 14).

**Tabela 13.** Matriz com resultado do teste exato de diferenciação (valores P) entre as populações, baseado nas frequências haplotípicas (haplótipos com base em todos UEPs). Considerando  $P < 0,05$ .

	Tayacaja	Arequipa	San Martin	Coyaima
Arequipa	0,05845 ± 0,0093			
San Martin	0,19500 ± 0,0215	0,30935 ± 0,0146		
Coyaima	0,00000 ± 0,0000	0,00000 ± 0,0000	0,00000 ± 0,0000	
Cayapas	0,00170 ± 0,0006	0,36575 ± 0,0047	0,01950 ± 0,0033	0,00000 ± 0,0000

**Tabela 14.** Matriz com resultado do teste exato de diferenciação (valores P) entre as populações, baseado nas frequências haplotípicas (haplótipos com base em microssatélites do haplogrupo Q). Considerando  $P < 0,05$ .

	Tayacaja	Arequipa	San Martin	Coyaima
Arequipa	0,66690 ± 0,0376			
San Martin	0,00000 ± 0,0000	0,05195 ± 0,0198		
Coyaima	0,00170 ± 0,0013	0,00000 ± 0,0000	0,42110 ± 0,0401	
Cayapas	0,00000 ± 0,0000	0,00000 ± 0,0000	0,03825 ± 0,0116	0,00000 ± 0,0000

#### 4.5 Seqüenciamento da região controle do DNA mitocondrial da população Coyaima da Colômbia.

A classificação das linhagens do DNA mitocondrial em haplogrupos continente-específicos foi feita através do seqüenciamento da região hipervariável I do DNA mitocôndria (302 pb), para os 21 indivíduos da população Coyaima (tabela 15). As seqüências obtidas foram alinhadas e comparadas com a seqüência de referência revisada de Anderson (Anderson *et al.*, 1991, Andrews *et al.*, 1999). Os quatro principais haplogrupos nativos americanos (A, B, C e D) foram representados na população de Coyaima. As frequências dos haplogrupos estão representadas na tabela 16. Para uma análise mais específica fizemos o RFLP de polimorfismos restritos de nativos americanos com o método de mini-seqüenciamento (tabela 16), o que permite identificar e diferenciar os haplogrupos nativos americanos dos de outros continentes e principalmente dos seus ancestrais asiáticos com mais precisão, utilizando polimorfismos específicos de cada população.

**Tabela 15.** Linhagens do DNA mitocondrial identificadas na população Coyaima.

	1	1	1	1	1	1	1	1	1	1	
	6	6	6	6	6	6	6	6	6	6	
	1	1	2	2	2	3	3	3	3	3	Haplogrupos
	1	8	1	2	9	0	1	2	2	6	
	1	9	7	3	0	1	9	5	7	2	
<b>CRS</b>	<b>C</b>	<b>T</b>	<b>T</b>	<b>C</b>	<b>C</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>C</b>	<b>T</b>	
C13	T	.	.	T	T	.	A	.	.	C	A
C46	T	.	.	T	T	.	A	.	.	C	A
C104	.	.	.	T	.	T	A	.	.	C	A
C258	T	.	.	T	T	.	A	.	.	C	A

C17	.	<b>C</b>	<b>C</b>	.	.	.	.	.	.	.	.	B
C60	?	<b>C</b>	<b>C</b>	.	.	.	.	.	.	.	.	B
C151	.	<b>C</b>	<b>C</b>	.	.	.	.	.	.	.	.	B
C323	.	<b>C</b>	<b>C</b>	.	.	.	.	.	.	.	<b>C</b>	B
C14	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C22	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C23	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C39	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C45	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C62	.	<b>C</b>	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C72	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C73	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C74	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C125	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C133	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C135	.	.	.	<b>T</b>	.	.	.	<b>C</b>	<b>T</b>	.	.	C
C93	.	.	.	<b>T</b>	.	.	.	<b>C</b>	.	<b>C</b>	.	D

As posições dos nucleotídeos na molécula de DNA mitocondrial estão representadas na primeira linha. Em **negrito**, a seqüência de referência revisada de Anderson (Anderson, *et al.*, 1991, Andrews, *et al.*, 1999). O ponto significa as bases que são idênticas às das seqüências de referência e o sinal de interrogação marca uma base não identificada. Os indivíduos estão representados à esquerda e os haplogrupos à direita da tabela. Estão sendo representadas apenas as principais mutações que definem os haplogrupos nativos americanos.

**Tabela 16.** Frequência dos haplogrupos de DNAmT na população Coyaima.

Haplogrupos (seqüenciamento da HVI-I)	Haplogrupos (RFLP por mini-seqüenciamento)	Freqüências
A	A2	19%
B	B4b1	19%
C	C1	57%
D	D1	5%

#### 4.6 Desenvolvimento de novos marcadores do Y

Em 2002, Weber *et al.* publicaram um estudo de identificação e caracterização de 2.000 polimorfismos inserção/deleção (indels), distribuídos por todo o genoma humano. Treze marcadores indels estavam localizados no cromossomo Y e dois destes foram selecionados para nosso estudo.

##### Nome do Marcador Mid227

Diferença entre os alelos: 2 pb

Iniciador direto: 5' TATCAAACATTACTCCCCCA 3'

Iniciador reverso: 5' AATCTGGCACTCTGAAAATTT 3'

Seqüência consenso:

TATCAAACATTACTCCCCCA [ -- / AA ] TATAGTAAAGTTTCTCTTTGTCATCTGAGAAAATTCTCTAGTGATA  
ATATGATCTTATATTCATAATAGCAATAACATGGAGAAGAAAGCAAACACACCTAAATGTAAAATTTTCAGAG  
TGCCAGATT

Frequências dos alelos do marcador Mid227:

Populações	Cromossomos amostrados	alelo maior 150pb	alelo menor 148pb
Africanos	42	0,38	0,61
Europeus	200	0,41	0,58
Nativos americanos	50	0,36	0,63
Japoneses	50	0,33	0,66

Fonte: [www.marshfieldclinic.org/research/genetics/indels/default.asp](http://www.marshfieldclinic.org/research/genetics/indels/default.asp)

##### Nome do Marcador: Mid967

Diferença entre os alelos: 11 pb

Iniciador direto: 5' GGAGACAGTGAGGTCAAGAA 3'

Iniciador reverso: 5' ATGGGAATGGAATCTAGTC 3'

### Seqüência consenso:

GGAGACAGTGAGGTCAAGAATCCGGCCATCTTTCATTGATAGCTGCCTCTGGGGTTTAAGTATGATTCTATCA  
CCCAAACAACCCTCA[-----/ACAACACACTG]ACAACACACT GGACTAGATTCCATTCCCCAT

### Freqüências dos alelos do marcador Mid967:

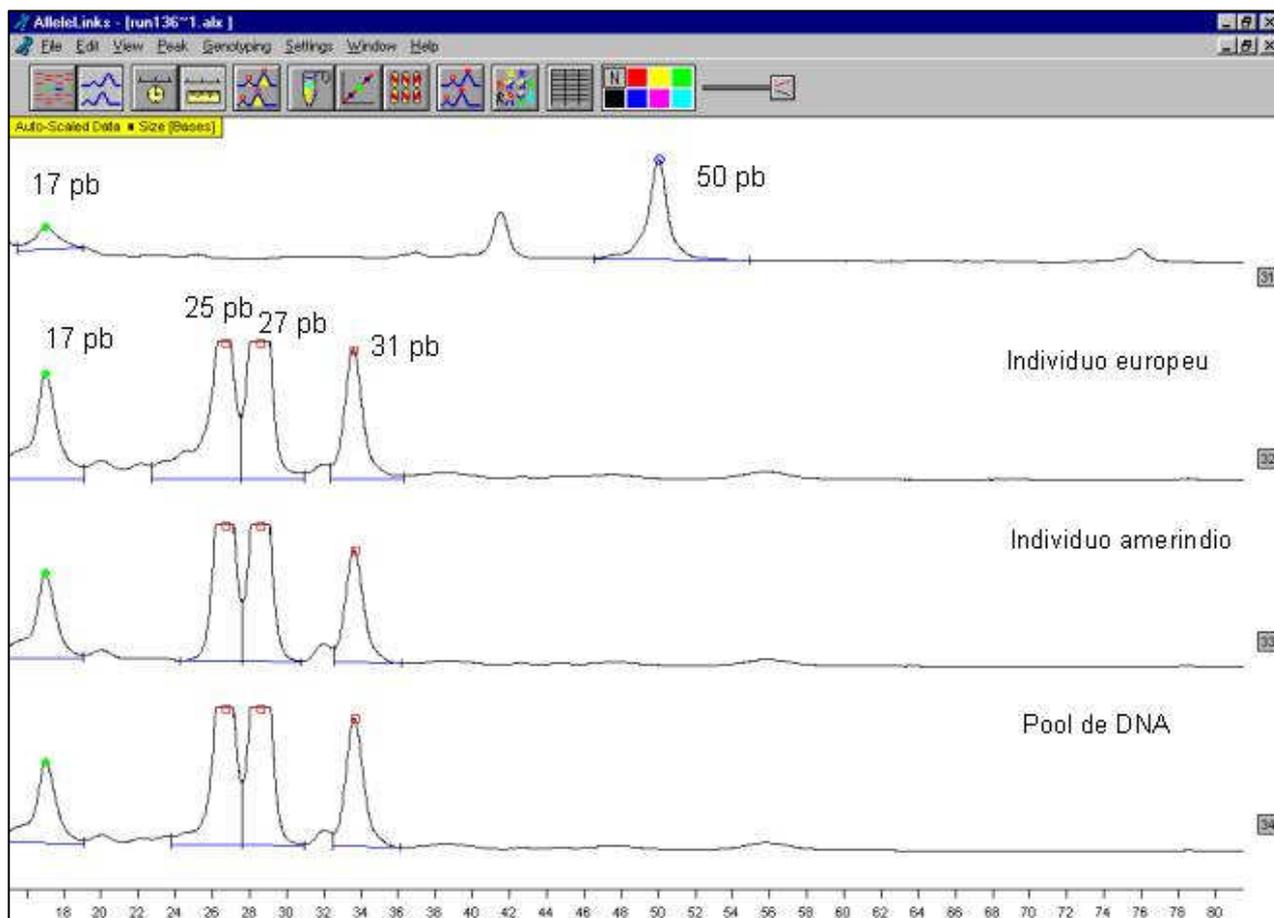
Populações	Cromossomos amostrados	alelo maior 130 pb	alelo menor 119
Africanos	42	0,001	0,99
Europeus	200	0,23	0,76
Nativos americanos	50	0,001	0,99
Japoneses	50	0,01	0,98
Coriell M44PDR	88	0,12	0,87

Fonte: [www.marshfieldclinic.org/research/genetics/indels/default.asp](http://www.marshfieldclinic.org/research/genetics/indels/default.asp)

Para estudo desses marcadores do cromossomo Y: Mid227 e Mid967, utilizamos os mesmos iniciadores descritos pelo autor (Weber *et al.*, 2002). Para visualização dos polimorfismos, utilizamos um pool de DNA de 203 indivíduos. Observamos a diferença do polimorfismo de 11 pb do Mid967 em gel de acrilamida 6%, no entanto, o Mid227, não pode ser visualizado devido ao tamanho da inserção ser muito pequena, apenas 2 pb. Para visualização da inserção utilizamos a técnica de mini-seqüenciamento. O iniciador direto, descrito pelo autor, foi usado na reação de mini-seqüenciamento com o didesoxinucleotídeo (ddGTP-F) marcado com fluorescência. Os resultados esperados eram que indivíduos sem a inserção de 2 pb, apresentariam um pico com tamanho de 25 nt e os indivíduos com a inserção um pico de 27 nt. Porém, o resultado obtido demonstrou que o marcador não permitia a diferenciação entre os alelos, todos os indivíduos apresentaram picos de 25 nt, 27 nt e 31 nt, este deve ser devido a um artefato da técnica de mini-seqüenciamento (Figura 15).

Através de pesquisa no Blast do NCBI, observamos que os dois marcadores possuem seqüências repetidas tanto no Y como em outros cromossomos.

O polimorfismo Mid227, anela em 4 regiões diferentes no cromossomo Y, o que não nos permite a discriminação dos alelos na presença ou ausência da inserção de 2 pb. Este resultado demonstrou que a seqüência do Mid227 é repetitiva no cromossomo Y, e está presente também em outros cromossomos como X e 16. Utilizamos uma amostra de 10 europeus e 4 ameríndios para tipar o polimorfismo Mid227 e não observamos diferenças entre indivíduos das populações analisadas como também no pool de DNA (figura 15).



**Figura 15.** Mini-sequenciamento do marcador Mid227. Canaleta 31-marcador externo 50-500pb, com marcador interno de 17 pb; Canaleta 32- individuo europeu; Canaleta 33- ameríndio e canaleta 34- pool de DNA com 203 indivíduos brasileiros (Gene/MG). Todos apresentam o mesmo perfil. Não há diferenciação entre a presença e ausência da inserção de 2 pb.

O polimorfismo Mid967, também anela em outras 3 regiões no cromossomo Y, porém é possível discriminar os alelos, alelo menor (119 pb), sem a inserção de 11 pb, está presente na maioria dos indivíduos enquanto o alelo maior (130 pb), com a inserção de 11 pb, foi observado no pool de DNA e em indivíduos europeus.

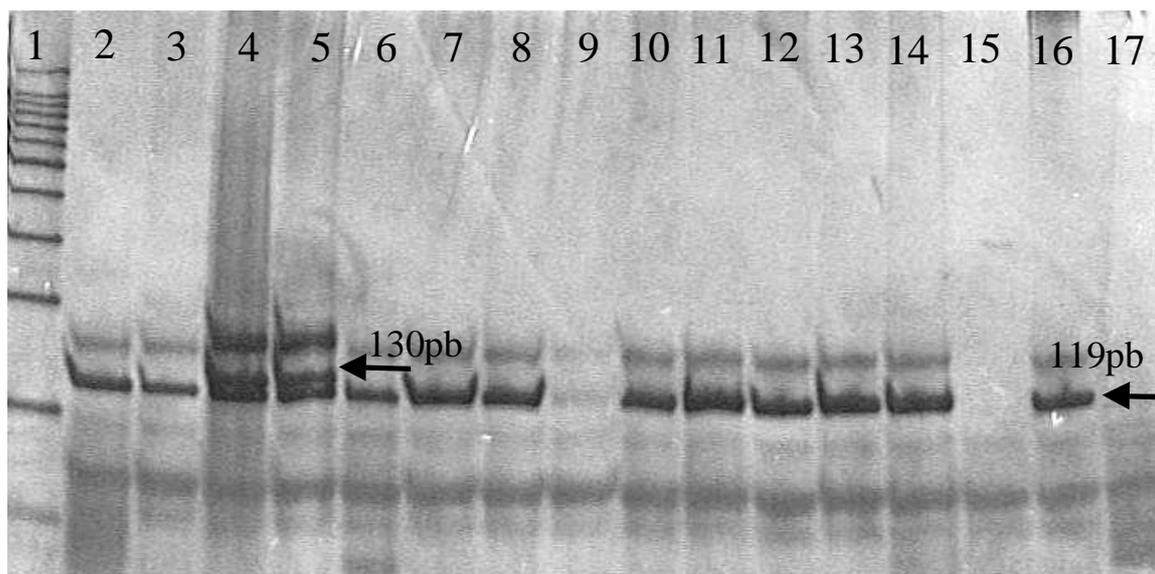
Para tipar o polimorfismo Mid967, utilizamos amostras de 23 indivíduos de diferentes populações (2 asiáticos, 12 europeus, 1 brasileiro do haplogrupo J, 2 japoneses e 6 ameríndios) (Figura 16) e apenas os europeus apresentaram o alelo maior de 130 pb. Após esse teste trabalhamos com duas das populações andinas Coyaima da Colômbia e San Martin do Peru. Dos 21 indivíduos colombianos analisados apenas dois apresentaram o alelo maior (130 pb), estes indivíduos já haviam sido classificados como pertencentes ao haplogrupo europeu R\*, com o marcador M207. Na população de San Martin, 54 indivíduos foram analisados e apenas quatro apresentaram o alelo maior,

sendo estes indivíduos também classificados previamente como pertencentes a haplogrupos europeus. Os indivíduos classificados como ameríndios estavam inclusos nos haplogrupos Q\* e Q3 (nativos americanos). A frequência observada do alelo maior (130 pb) para os europeus foi maior que descrita por Weber e colaboradores que é de 0,23 (ver tabela de frequência do Mid967), enquanto, a nossa foi de 0,888 (Tabela 17). Estudos complementares serão posteriormente realizados para inclusão do marcador Mid967 na árvore filogenética do cromossomo Y (YCC, 2002).

**Tabela 17.** Frequências dos alelos do marcador Mid967 observadas em nosso estudo.

Populações	Cromossomos amostrados	alelo maior 130 pb	alelo menor 119
Europeus	18 <sup>1</sup>	0,888	0,112
Ameríndios	69	0,001	0,999

<sup>1</sup> 12 indivíduos europeus mais 6 indivíduos classificados posteriormente como europeus, antes classificados como ameríndios.

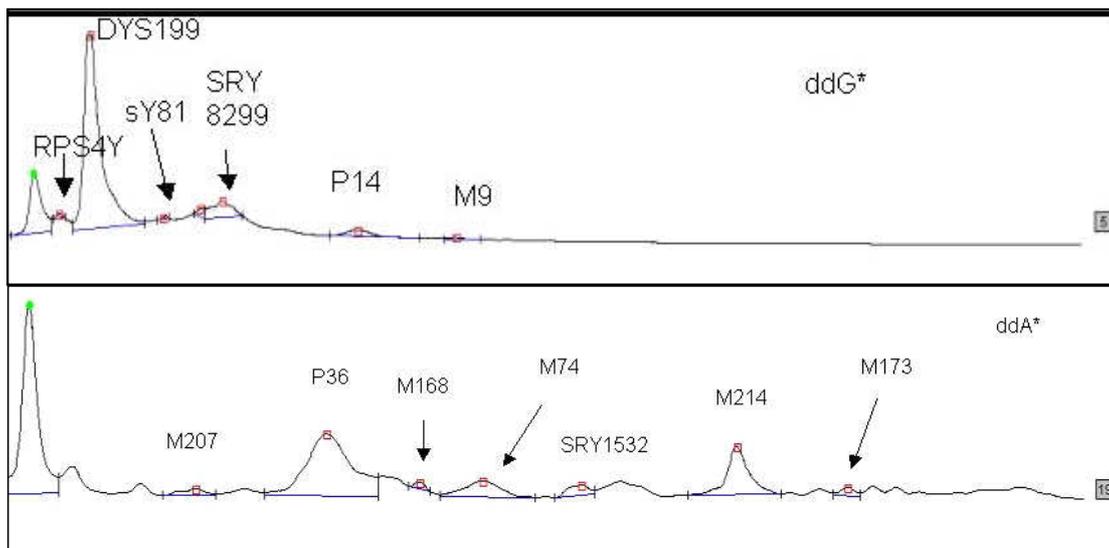


**Figura 16.** Gel de poliacrilamida 6%, das amostras de indivíduos de diferentes populações para o marcador Mid967. 1-marcador de peso 1Kb, 2 e 3 indivíduos da Ásia central; 4, 5, 6 e 8 indivíduos europeus, 7-indivíduo do haplogrupo J, 9 e 10-japoneses, 11-16 - ameríndios e 17-negativo. Setas ilustrando alelo maior (119 pb) e menor (130 pb).

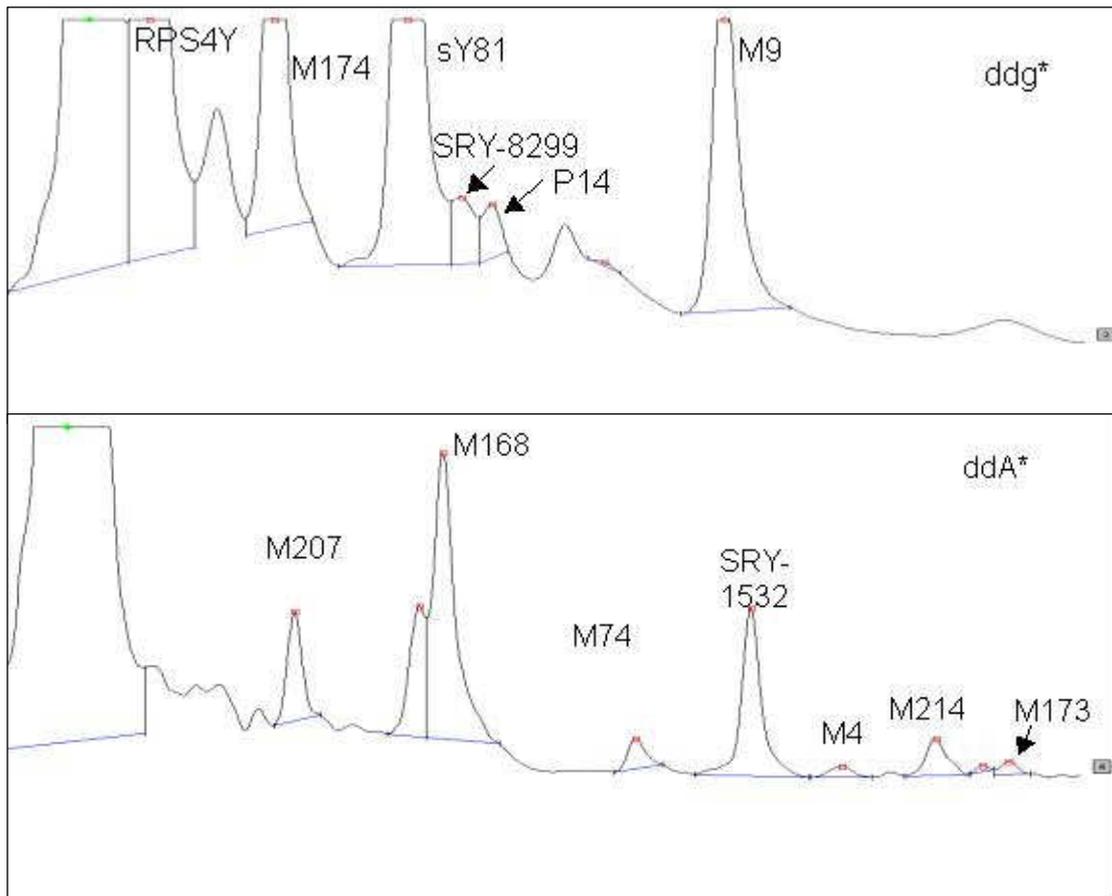
#### 4.7 Padronização do Multiplex de SNPs do cromossomo Y

Utilizamos a técnica de mini-seqüenciamento para produzir um multiplex de SNPs do cromossomo Y. Os SNPs foram selecionados de modo que abrangesse os principais haplogrupos do cromossomo Y de acordo com YCC (2002), mas com ênfase especial naqueles informativos para o povoamento das Américas. Para melhor resolução dos 16 SNPs estudados (figura 11), dividimos o multiplex em dois sistemas, com oito SNPs. A técnica de mini-seqüenciamento descrita na seção 3.5 da metodologia, utiliza ddNTPs (didesoxinucleotídeos) fluorescentes (ddATP-F ou ddGTP-F) para discriminação dos dois alelos de cada SNP. Foi feita apenas a padronização inicial dos multiplexes, utilizando indivíduos já tipados em nosso laboratório com os SNPs pelos outros métodos descritos. A padronização ainda está sendo finalizada. No entanto, no momento já conseguimos caracterizar por mini-seqüenciamento todos os SNPs individuais. No primeiro teste do multiplex de SNPs, apenas seis marcadores do sistema 1, utilizando ddGTP-F foram visualizados e sete do sistema 2 com ddATP-F (Figura 17). Os iniciadores do mini-seqüenciamento foram purificados com HPLC para melhorar a definição e visualização dos sistemas estudados. Na figura 18 podemos verificar uma melhora na definição dos picos.

Os multiplexes ainda estão sendo padronizados, e diferentes concentrações de iniciadores estão sendo testadas para obtenção dos sistemas completos com todos os marcadores visualizados.



**Figura 17.** Multiplex de SNPs por mini-seqüenciamento. Sistema 1-utilizando ddGTP-F, apenas seis marcadores; sistema 2-utilizando ddATP-F com sete marcadores. Iniciadores não purificados.



**Figura 18.** Multiplex de SNPs por mini-seqüenciamento. Sistema 1-utilizando ddGTP-F, apenas seis marcadores; sistema 2-utilizando ddATP-F, com sete marcadores. Iniciadores purificados.

## 5 Discussão

Nosso grupo vem trabalhando com populações nativas da América no intuito de clarear várias questões sobre o povoamento do continente americano. Há ainda muita controvérsia a respeito do número de ondas migratórias principais ao continente, da data da primeira migração, sobre a origem asiática destas populações que deram origem aos nativos americanos bem como quais as relações entre distintos povos nativos da América do Sul e como foi a ocupação deste subcontinente. Algumas dessas questões vêm sendo detalhadas por vários estudos e foram também abordadas neste projeto. As primeiras inferências: as primeiras evidências com marcadores do Y em nativos americanos foram obtidas em 1995, no qual foi evidenciado um forte efeito fundador nas Américas, a presença do de um haplótipo ( $\alpha$ -II e *DYS19-A*) em altas frequências nas populações de nativos americanos (Pena *et al.*, 1995; Santos *et al.*, 1995, Santos *et al.*, 1996). Estas conclusões foram posteriormente confirmadas por estudos nos EUA (Underhill *et al.*, 1996) utilizando a transição de C para T no *locus* *DYS199* que define o haplogrupo Q3. Este tipo de cromossomo Y (Q3) é o mais comum, presente em mais de 70% dos nativos americanos (Santos e Tarazona-Santos, 2002), e provavelmente originou-se na América do Norte ou Beríngia, depois que os imigrantes asiáticos deixaram a Sibéria. Estes resultados foram confirmados em nosso trabalho onde a frequência desse haplogrupo foi de 75,5% nas populações andinas estudadas.

Em 1999, foi apontada como provável origem dos fundadores da América a população ancestral dos Kets e dos Altai na Sibéria, povos atuais que compartilham com nativos americanos vários marcadores do cromossomo Y, hoje reconhecido como haplogrupo Q\* (Santos *et al.*, 1999, Santos e Tarazona-Santos, 2002). Neste estudo avaliamos a ocorrência do haplogrupo Q\* em populações da Ásia Central, em localidades próximas às das populações de Kets e Altai, mas este haplogrupo não foi encontrado.

A datação da primeira migração com dados do Y (Bortolini *et al.*, 2003) tem sugerido o final Pleistoceno, ao redor de 14 a 15 mil anos atrás, como o período mais provável para a chegada dos primeiros povos ancestrais dos atuais. No entanto, existem outras datas alternativas com dados do Y e do DNAm (Hammer *et al.*, 2001, Bonatto e Salzano, 1997, Silva *et al.*, 2002) que são geralmente mais antigas. Devido à controvérsia no uso de distintas taxas de mutação e métodos de datação (Carvalho-Silva *et al.*, 1999, Santos e Tarazona-Santos, 2002), associada à nossa amostra limitada a povos andinos, preferimos não apresentar cálculos de datações neste projeto, mas utilizando metodologias e taxas de mutações disponíveis na literatura os resultados são similares aos previamente reportados (Bortolini *et al.*, 2003).

Em 2001, Tarazona-Santos e colaboradores elaboraram um modelo de evolução diferenciada entre populações do Leste e Oeste da América do Sul. Neste modelo as populações dos Andes, principalmente da família lingüística Quéchuas, se diferenciavam de populações do Leste (Amazônia, Planalto Central do Brasil, etc) por apresentarem maior diversidade intrapopulacional e menor interpopulacional. Neste trabalho, estudamos em detalhe cinco populações andinas, inclusive duas populações não previamente caracterizadas, com objetivo determinar os níveis de diversidade intra e interpopulacional com novos marcadores do tipo UEPs (YCC, 2002) e microssatélites, e avaliar a miscigenação européia e africana nas populações andinas, com enfoque na população Coyaima da Colômbia.

### **5.1 Variabilidade Genética das Populações Andinas**

O índice de diversidade de Nei intrapopulacional observado nas populações estudadas foi: 0,9631 na população de Tayacaja, 0,9105 na população de Arequipa, 0,9922 na população de San Martin, 0,800 na população Coyaima e 0,9242 na população de Cayapa. Os valores encontrados vão de acordo com o descrito por (Tarazona-Santos,1999) para populações dos Andes (0,964). A população Coyaima apresenta o menor índice de diversidade intrapopulacional comparando com as demais populações andinas analisadas, o que pode ser atribuído à redução populacional que esta população sofreu durante a colonização espanhola.

A partição da variabilidade genética, medida através da análise de variância molecular (AMOVA) confirma o nível de variabilidade dentro das populações maior que entre as populações, independente da origem dos haplótipos (UEPs ou microssatélites). Entretanto, com os dados de UEPs a variabilidade entre os grupos é maior que nos microssatélites porém não significativa. A diferenciação entre os grupos pode ser devido a processos de diferenciação promovidos por deriva genética, e homogeneização das populações dentro do grupo lingüístico Quéchuas pelo fluxo migratório entre populações andinas do Peru (Tarazona-Santos *et al.*, 2001). Na análise de haplótipos gerados por microssatélites nos haplogrupos Q e Q3, foi demonstrado uma menor variabilidade entre grupos comparado com UEPs, porém todas significativas ( $P < 0,05$ ). O valor de  $F_{st}$  encontrado com todos os haplogrupos foi de 0,18 (calculado na AMOVA com haplótipos de UEPs), enquanto que o valor encontrado retirando os haplogrupos considerados miscigenados ( $E^*$ ,  $E_{3a}$  e  $R^*$ ) foi de 0,35. O aumento no valor de  $F_{st}$  com a retirada dos haplogrupos miscigenados sugere que esses cromossomos não nativos promovem uma homogeneização das populações, devido ao fluxo gênico ou apenas pela fato das

população não nativas terem a mesma origem (Europeias e Africanas). A diferenciação observada com o teste exato entre as populações, principalmente entre Quéchuas e não-Quéchuas pode ser atribuída à deriva genética e à restrição do fluxo gênico entre elas. Nossos dados corroboram com a homogeneização das populações do grupo dos Quéchuas, apresentando um maior fluxo gênico entre elas (Tarazona-Santos *et al.*, 2001). Este padrão difere do encontrado por Zegura *et al.*, (2004), em 18 populações nativas americanas, retirando haplogrupos miscigenados (E, F, G, I, J, R\* e R1b) o valor de  $F_{st}$  diminui de 0,21 para 0,19. A variabilidade genética encontrada entre populações nativas americanas é considerada baixa em comparação com populações asiáticas que em 18 populações da Sibéria o valor de  $F_{st}$  foi de 0,41 (Karafet *et al.*, 2002).

O uso de marcadores genéticos para o estudo do povoamento inicial do continente americano tem demonstrado que parte da diversidade observada nas populações nativas atuais é devida à miscigenação recente com povos imigrantes ou colonizadores. Este problema afeta os estudos com cromossomo Y de uma maneira mais significativa do que os estudos com DNA mitocondrial, pelo fato de que a miscigenação ocorre envolvendo preferencialmente o homem imigrante e a mulher nativa (Mesa *et al.*, 2000). Para estudos mais detalhados de miscigenação aumenta-se cada vez mais a necessidade de marcadores do cromossomo Y que sejam específicos de cada população de distintas localidades, e que possam realmente distinguir os cromossomos nativos dos de origem recente. Os marcadores disponíveis atualmente (YCC, 2002) podem ser utilizados para este propósito, mas é necessário conhecer sua ocorrência entre as populações aborígenes dos distintos continentes.

Dentre os haplogrupos analisados, Q\* e Q3 podem ser considerados nativos americanos autênticos com uma frequência média de 83,2% (Q\*= 9,43% + Q3= 73,7%). Alguns dos raros indivíduos com haplogrupos P\* (0,21%) e Y\* (1,6%) podem ser possivelmente nativos americanos, mas estes haplogrupos ocorrem em alta frequência nas populações da Eurásia e são considerados aqui como produtos de miscigenação (apesar de não terem sido excluídos da AMOVA com UEPs). Seriam necessários mais marcadores para um estudo detalhado de discriminação destes haplogrupos. Os demais haplogrupos encontrados são considerados produtos reais de miscigenação. O haplogrupo miscigenado mais encontrado foi o R\* (5,1%), muito frequente na Europa, seguido dos haplogrupos E\* (3,6%), comum na Eurásia e África e E3a (1,16%), frequente na África (tabela 10). Os haplogrupos C, D-E\* e R1, não foram observados nas populações analisadas, sendo que apenas C, se encontrado, poderia ser considerado como um cromossomo autêntico de nativos americanos.

O haplogrupo Q3 originou-se a partir do haplogrupo Q, este último definido pela mutação P36 (YCC, 2002), que nos permite distinguir os cromossomos nativos P que não

possuem a mutação *DYS199*, daqueles que são provavelmente produto da miscigenação após a colonização das Américas por outras populações. O haplogrupo Q\* é freqüente dentre grupos de nativos americanos incluindo Ameríndios e Na-denes (Karafet *et al.*, 2002; Bortolini *et al.*, 2003) com freqüência de 23,8% (Zegura *et al.*, 2004) e está presente também na Sibéria, com freqüência de 18,8% nas populações Kets e Selkup (Karafet, *et al.*, 2002). A presença dessa mutação sugere que uma migração inicial (pelo menos de cromossomos Y) se direcionou às terras virgens americanas, a partir de um único grupo, ou grupos relacionados da Ásia, com cromossomos Y do haplogrupo Q (Santos *et al.*, 1999), após entrada nas Américas surgiu o alelo derivado *DYS199-T*. Os cromossomos com esta mutação se fixaram nas populações nativas americanas. No entanto, é possível encontrar indivíduos nativos que não possuam o alelo *DYS199-T*, mas que pertençam ao haplogrupo ancestral Q. Diversos fatores podem ter contribuído para a manutenção de indivíduos com haplogrupo Q\*, que possuem o alelo ancestral *DYS199-C* nessas populações. Analisamos a mutação P36 (haplogrupo Q\* - nativo) e a M207 (haplogrupo R\* - europeu) nas cinco populações andinas estudadas neste projeto e a freqüência observada do haplogrupo Q\* foi baixa em comparação ao haplogrupo R\*, que é freqüente na Europa, o que significa que a grande parte dos *DYS199-C* (anteriormente classificados com P\*), encontrados nas populações analisadas são provavelmente produto de miscigenação recente com outras populações, principalmente européias, resultado da colonização espanhola, na região dos Andes. No entanto, na população Coyaima da Colômbia, mesmo apresentando o maior nível de miscigenação (tabela 10) dentre as populações andinas, apresentou a maior freqüência do haplogrupo Q\* (42,8%). A população Cayapa do Equador, ao contrário das outras populações não apresentou miscigenação com o Y, pois todos os indivíduos pertencem ao haplogrupo Q3, que pode ser explicado pelos hábitos dessa população: vivem isolados ao redor do Rio Cayapa, são resistentes a mudanças culturais e contatos com outros povos e continuam sendo tribos que vivem na floresta.

## **5.2 Origem da População Coyaima atual**

A maioria dos estudos relativos aos nativos americanos tem se direcionado a responder questões sobre o primeiro povoamento do continente. No entanto, em níveis geográficos mais específicos dentro da América do Norte e do Sul, as relações históricas entre as diferentes populações e grupos lingüísticos não foram determinadas. Várias tribos ainda não foram analisadas e algumas destas apresentam dúvidas quanto sua origem étnica devido à aculturação com perda da identidade étnica. Um exemplo é a

população de índios da tribo Coyaima da região de Tolima na Colômbia. A tribo remanescente ocupa hoje um pequeno vilarejo nos Andes colombianos, teve sua língua original extinta (Coyaima ou Pijao) sendo substituída pelo espanhol. O estudo da variabilidade genética desta população confirma os dados históricos, apesar da resistência contra a colonização espanhola, a população fala hoje apenas o espanhol e apresenta a linhagem paterna bastante miscigenada, com altas frequências de haplogrupos europeus (~24%) e africanos (~9%), porém, a frequência do haplogrupo Q\* encontrada foi de 42,8%, sugerindo que a maioria dos cromossomos *DYS199-C* são realmente nativos (tabela 10). Esta alta frequência do haplogrupo Q\*, que estava provavelmente em alta frequência entre os primeiros migrantes nas Américas (Santos e Tarazona-Santos, 2002), pode indicar que os Coyaima de hoje sejam remanescentes de povos com descendência direta dos primeiros migrantes. Frequência tão alta deste haplogrupo não é observada em nenhuma outra tribo ameríndia, apenas comparável a algumas tribos de Na-Dene e Aleuta-Eskimos da América do Norte (Zegura *et al.*, 2004, F.R. Santos, comunicação pessoal). No entanto, esta alta frequência de Q\* poderia ser um resultado de deriva genética recente, o que não é corroborado pelos microssatélites pois quatro haplótipos distintos aparecem entre os Coyaima Q\*.

Os Coyaima vivem atualmente isolados, sofreram uma grande diminuição populacional durante a colonização espanhola por apresentarem maior resistência, perderam sua língua e atualmente vivem restritos a um pequeno vilarejo na Colômbia sofrendo pressão da guerrilha colombiana. Os resultados do estudo do DNAMt nos indivíduos da população Coyaima demonstraram que as linhagens maternas nessa população são ameríndias, estando todos inclusos nos haplogrupos: A, B, C e D. As análises do DNAMt baseadas em sítios específicos da região codificadora por mini-seqüenciamento e RFLP, permitiram a classificação mais precisa das linhagens mitocondriais da população Coyaima, nos haplogrupos A2, B4b1, C1 e D1 (Tabela 13), específicos de nativos americanos. Este resultado reafirma o padrão de colonização que ocorreu nas Américas, no qual entre os colonizados as linhagens maternas são nativas e as paternas com contribuições da América, Europa e África.

### **5.3 Quantas ondas migratórias?**

Lell *et al.*, (2002), ao analisarem a distribuição de haplogrupos do cromossomo Y em indivíduos da Sibéria e Nativos Americanos, propuseram um modelo com duas ondas migratórias principais provenientes de regiões da Sibéria que teriam dado origem aos povos da América. Com base na divisão do haplogrupo P em duas sublinhagens através da associação com quatro microssatélites: M45a, primeira onda migratória, que inclui

haplogrupo Q, e M45b, segunda onda migratória, que inclui o haplogrupo R. Esta segunda onda migratória teria trazido também o haplogrupo C, proveniente da Região do Mar de Okhotsk. A primeira coincide com a proposta pelo nosso grupo (Santos *et al.*, 1999), porém a segunda onda migratória trazendo os haplogrupos R e C, pode ser apenas produto da miscigenação recente (no caso de cromossomos R) com populações européias. O que foi sugerido por Tarazona-Santos e Santos (2002) é que a estrutura genética das populações de nativos americanos é mais compatível com a hipótese de uma única onda migratória principal e que a presença do haplogrupo R seria devido à miscigenação. A presença de indivíduos com haplogrupo C e P\* (e talvez Y\*), raros entre nativos americanos, mas comuns na Sibéria, poderia ser explicada por uma migração principal exclusiva com maior diversidade de haplogrupos ou por subseqüentes migrações pequenas e minoritárias (não principais) que trouxeram outros cromossomos para as Américas (Tarazona-Santos e Santos, 2002). Outros trabalhos recentes confirmam o modelo de uma única onda migratória principal (Zegura *et al.*, 2004) trabalhando com 63 UEPs e 10 microssatélites em 18 populações nativas americanas, 28 asiáticas e 5 européias. Nossos resultados reafirmam a hipótese de que populações andinas com a presença do haplogrupo R (R\*+ R1) refletem um quadro que pode ser mais facilmente explicado por miscigenação recente, corroborada por fatos históricos.

## 6 Conclusões

Com a análise da variabilidade genética de populações Andinas e da Ásia Central através do uso de marcadores moleculares do cromossomo Y e do DNA mitocondrial, podemos concluir:

- ◆ A análise de variância molecular (AMOVA) confirma o alto nível de variabilidade dentro das populações, bem maior do que entre as populações, tanto para haplótipos gerados por UEPs ou por microssatélites do cromossomo Y.
- ◆ A AMOVA com UEPs demonstrou que a variabilidade entre os grupos é maior do que entre as populações (apesar da baixa significância), o que é esperado pois o agrupamento separou populações de países e grupos lingüísticos diferentes.
- ◆ A AMOVA com haplótipos de microssatélites em indivíduos do haplogrupo Q (Q\* + Q3) e Q3 revelou menor variabilidade entre grupos e entre populações (embora altamente significativa) comparando com haplótipos de UEPs, sugerindo que populações andinas de distintos grupos lingüísticos possuem uma heterogeneidade significativa entre si. Isto demonstra que no oeste do continente, a cultura dos povos está correlacionada com o fluxo gênico, sendo que populações de um mesmo grupo lingüístico apresentam pouca diferenciação entre si, mas há uma significativa diferenciação entre populações de línguas distintas.
- ◆ Os testes exatos com UEPs e microssatélites confirmam que a diferenciação observada entre populações de distintos grupos lingüísticos andinos é significativa.
- ◆ Os marcadores genéticos do cromossomo Y utilizados neste projeto, específicos de cada população, permitem a distinção da maior parte dos cromossomos nativos dos de origem miscigenada. Esta análise demonstrou que parte significativa da diversidade observada nas populações nativas atuais nos Andes é devida à miscigenação com povos de origem europeia e africana. A alta frequência do haplogrupo R\*, freqüente em europeus, encontrada nas populações analisadas pode ser atribuída ao processo de colonização espanhola que ocorreu na região dos Andes.

- ◆ Os valores de  $F_{st}$  encontrados com todos haplogrupos de UEPs (0,18) ou retirando os haplogrupos não nativos (0,35) sugerem que o fluxo gênico está homogeneizando as populações. A diferenciação observada entre as populações, principalmente entre Quéchuas e não-Quéchuas pode ser atribuída à deriva genética e à restrição do fluxo gênico entre elas. Enquanto, as populações do grupo dos Quéchuas apresentam um grande fluxo gênico entre elas, tornando-as homogêneas.
- ◆ A população Coyaima apresenta grande miscigenação por análise das linhagens paternas (~33%), mas a frequência do haplogrupo Q\* foi alta (42,8%), indicando que a maioria dos cromossomos *DYS199-C* (anteriormente P\*), antes vistos como prováveis produtos de miscigenação, são realmente nativos. A alta frequência do Q\*, não observada em nenhum outro estudo feito atualmente com ameríndios, sugere que esta população talvez seja um remanescente direto dos primeiros migrantes para as Américas que chegaram da Ásia portando a mutação P-36.
- ◆ Os estudos do DNAm demonstraram que as linhagens maternas dos Coyama são tipicamente ameríndias, estando todos indivíduos inclusos nos haplogrupos A2, B4b1, C1 e D1. Os resultados de análise de miscigenação confirmam o padrão de colonização que ocorreu nas Américas, onde os colonizados receberam grande contribuição de linhagens paternas dos colonizadores enquanto suas linhagens maternas permaneceram nativas.
- ◆ Dos dois marcadores do cromossomo Y avaliados (Mid227 e Mid967), apenas o Mid967 pode ser utilizado em estudos populacionais e evolutivos, após a conclusão de estudos complementares para inclusão dele na árvore filogenética do cromossomo Y (YCC, 2002).
- ◆ O desenvolvimento dos sistemas multiplexes de SNPs do cromossomo Y, que se mostrou exequível embora ainda incipiente, será de grande utilidade, aumentando a eficiência, diminuindo os custos e tempo gasto na tipagem de indivíduos em estudos populacionais e evolutivos.

# 7 Referências Bibliográficas

- Alves-Silva, J., Santos, M.S., Guimarães, P.E.M., Ferreira, A.C.S., Bandelt, H.J., Pena, S.D.J., Prado, V. F. (2000) The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet* 67:444-461.
- Anderson, S., Bankier, A.T., Barrell, B.G., Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Rose, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., Young, I.G. (1981) Sequence and organization of the human mitochondrial genome. *Nature*. 290:457-465.
- Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowers, R.N., Turnbull, D.M., Howell, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
- Antorveza, A.T. (1992) *La colonización española en el Tolima siglos XVI y XVII*. 1st edn. Funcol, Colombia.
- Bandelt, H.T., Herrnstadt, C., Yao, Y.G., Kong, Q.P., Kivisild, T., Rengo, C., Scozzari, R., Richards, M., Villems, R., Macaulay, V., Howell, N., Torroni, A., Zhang, Y.P. (2003) Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Am. J. Hum. Genet* 67:512-524.
- Bergen, A.W., Wang, C.Y., Tsai, J., Jefferson, K., Dey, C., Smith, K.D., Park, S.C., Tsai, S.J., Goldman, D. (1999) An Asian- Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann. Hum. Genet* 63:63-80.
- 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 USA*. 94(5):1866-1871.
- Bortolini, M.C., Salzano, F.M., Bau, C.H.D., Layrisse, Z., Petzl-Erler, M.L., Tsuneto, L.T., Hill, K., Hurtado, A.M., Castro-de-Guerra, Bedoya, , Ruiz-Linares, A. (2002) Y-chromosome biallelic polymorphisms and Native American population structure. *Am. J. Hum. Genet* 66:255-259.
- Bortolini, M.C., Salzano, F.M., Thomas, M.G., Stuart, S., Nasanen, S.P.K., Bau, C.H.D., Hutz, M.H., Layrisse, Z., Petzl-Erler, M.L., Tsuneto, L.T., Hill, K., Hurtado, A.M., Castro-de-Guerra, D., Torres, M.M., Groot, H., Michalski, R., Nymadawa, P., Bedoya, G., Bradman, N., Labuda, D., Ruiz-Linares, A. (2003) Y-chromosome Evidence for differing ancient demographic histories in the Americas. *Am. J. Hum. Genet* 73:524-539.
- Bosch, E., Calafell, F., Santos, F.R., Pèrez-Lezaun, A., Comas, D., Benchemsi, N., Tyler-Smith, C., Bertranpetit, J. (1999) Variation in short tandem repeats is deeply structure by genetic background on the the human Y chromosome. *Am. J. Hum. Genet* 65:1623-1638.
- Brinkman, B., Klitschar, M., Neuhuber, F., Huhne, J., Rolf B. (1998) mutation rate in human microsatellites: influence of the structure and length of tandem repeats. *Am. J. Hum Genet* 62:1408-1415.

- Cann, R.L., Stoneking, M., Wilson, A.C.** (1987) Mitochondrial DNA and human evolution. *Nature*.325:31-36.
- Capelli, C., Wilson, J.F., Richards, M., Stumpf, M.P., Gratrix, F., Oppenheimer, S., Underhill, P., Pascali, V.L., Ko, T.M., Goldstein, D.B.** (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am. J. Hum. Genet* 68:432-443.
- Carvalho-Silva D.R., Santos F.R., Hurtz M.H., Salzano F.M., Pena S.D.** (1999) Divergent human Y-chromosome microsatellite evolution rates. *Journal Molecular Evolution*. 49:204-214.
- Carvalho-Silva D. R.**(2000).*Estudos moleculares e filogeográficos de cromossomos Y em populações humanas do Brasil*.Tese apresentada para obtenção do título de Doutor ao Departamento de Bioquímica e Imunologia, UFMG, Belo Horizonte.
- Carvalho-Silva, D.R., Santos, F.R., Rocha, J., Pena, S.D.**(2001)The philogeography of Brazilian Y-chromosome lineages. *Am. J. hum. Genet* 68:281-286.
- Cavalli-Sforza, L., Menozzi, P., Piazza, A.**(1994) *The History and Geography of Human Genes*. Princeton University Press, Princenton, pp 541.
- Cavalli-Sforza, L.** (1998) The DNA revolution in population genetics. *Trends in genetics*.14(2):60-65.
- Cooper G, Amos W, Hoffman D, Rubinsztein DC** (1996) Network analysis of human Y microsatellite haplotypes. *Hum Mol Genet* 5: 1759-1766.
- Dib, C., Faure, S., Fizames, C., Samson D., Drouot, N., Vignal, A., Milasseau, P., Marc, S., Hazan J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J., Weissenbach, J.** (1996). A comprehensive genetic map of human genome based on 5,256 microsatellites. *Nature*. 380:152-154.
- Ewing, B.; Hillier,L.; Wendl, M. , Green, P.** (1998). Basecalling of automated sequencer traces using Phred I: Accuracy assesment. *Genome Research* 8: 175-185.
- Excoffier, L., Smouse, P.E., Quattro, J.M.** (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. 131:479-491.
- Fagundes, J.R., Bonatto, S.L., Callegari-Jacques, S.M., Salzano, F.M.** (2002) Genetic, Geographic, and linguistic variation among South American Indians: Possible Sex influence. *Am. J. Phys. Anthropol*.117:68-78.
- Foster, E.A., Jobling, M.A., Taylor,P.G.,Donnelly, P., Knijff, P., Mieremet, R., Zerjal, T., Tyler-Smith, C.** (1998) Jefferson fathered slave's last child.*Nature*.396:27-28.
- Gómez,L. D.**(1967). La cordillera central. In:*Historia extensa de Colombia:Tribus indígenas y sitios arqueologicos*. Vol.1(Ediciones Lerner).Academia colombiana de historia,pp. 225-247.
- Gordon, D.; Abajian, C., Green, P.** (1998).Consed: A graphical tool for sequence finishing. *Genome Research* 8:195-202.
- Green, P.** (1994). *Phrap*. [www.genome.washington.edu/UWGC/analysistools/phrap.htm](http://www.genome.washington.edu/UWGC/analysistools/phrap.htm)

- Greenberg, J.H., Turner, C.G., Zegura, S.L.** (1986) The settlement of the Americas : a comparison of the linguistic, dental and genetic evidence. *Curr. Anthropol.*27:477-498.
- Hammer, M.F.**(1995) A recent common ancestry for human Y chromosomes. *Nature.* 378:376-378.
- Hammer, M.F., Horai, S.** (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet.*56(4):951-962.
- Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.R., Zegura, S.L.** (1998) Out of Africa and back again:nested cladistic analysis of human Y chromosome variation. *Molecular Biology Evolution.* 15:427-41.
- Hammer, M.F., Karafet, T., Redd ,J.A., Jarjanazi, H., Santachiara-Benecetti, S., Soodyall, H. , Zegura, L. S.** (2001) Hierarchical Patterns of Global Human Y-Chromosome Diversity. *Molecular Biology Evolution* 18:1189-1203.
- Hammer M.F. , Zegura S.L.** (2002). The Human Y Chromosome Haplogroup Tree: Nomenclature and Phylogeography of Its Major Divisions. *Annu.Rev.Anthropol.*31:303-21.
- Hurles, M.E., Irven, C., Nicholson, C., Taylor, P.G., Santos, F.R., Loughin, J., Jobling, M.A., Sybes, B.C.** (1998) European Y-chromosomal lineages in Polynesia: a contrast to the population structure revealed by mitochondrial DNA. *Am. J. Hum. Genet* 63:1793-1806.
- Hurles, M.E., Veitia, R., Arroyo, E., Armenteros, M., Bertranpetit, J., Perez-Lezaun, A., Bosch, E., Schlumukova, M., Cambon-Thomsen, A., McElreavey, K., Lopez, A.M., Rohl, A., Wilson, I.J., Singh, L., Pandya, A., Santos, F.R., Tyler-Smith, C., Jobling, M.A.** (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am. J. Hum. Genet* 65:1437-1448.
- Jobling, M. A., Pandya, A., Tyler-Smith C.**(1997) The Y chromosome in forensic analysis and paternity testing. *Int. J. Legal. Med.* 110:118-124.
- Jobling, M.A., Tyler-Smith, C.** (2000) New uses for new haplotypes the human Y chromosome, disease and selection. *Trends in Genet* 16:356-362.
- Jobling, M.A., Tyler-Smith, C.** (2003) The human chromosome comes of age. *Nature Reviews Genetics.*4:598-612.
- Karafet, M.T., Zegura, S. L., Posukh, O., Osipova,L., Bergen, A., Long, J.,Goldman, D., Klitz, W., Harihara, S., Knijff, P., Wiebe,V., Griffiths,R.C., Templeton,A.R. , Hammer,F.M.**(1999).Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am. J. Hum. Genet* 64:817-831.
- Karafet, T., Xu, L. Du, R., Wang, W., Feng, S., Wells, R.S., Redd, A.J., Zegura, S.L., Hammer, M.F.** (2001) Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet* 69:615-628.
- Karafet, M.T., Zegura, Osipova, L.P., Gubina, M.A., Posukh, O.L. , Hammer, F.M.** (2002). High levels of Y-chromosome differentiation among Native Siberian

- populations and the genetic signature of a boreal hunter-gatherer way of life. *Human Biology*. 74(6):761-789.
- Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M.** (2001) MEGA2: Molecular Evolutionary Genetics Analysis software, *Bioinformatics* 17:1244-1245.
- Lahn, B.T., Page, D.C.** (1999) Four evolutionary strata on the human X chromosome. *Science*. 289:964-967.
- Lahn, B.T., Pearson, N.M., Jegalian, K.** (2001) The human Y chromosome, in the light of evolution. *Nature Reviews Genetics*. 2:207-216.
- Lahn, B.T., et al.** (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. 423:825-837.
- Lell, J.L., Sukernik, R.L., Starikovskaya, Y.B., Su, B., Li, J., Schurr, T.G., Underhill, P. A., Wallace, D. C.** (2002). The dual origin and Siberian affinities of native American Y chromosomes. *Am. J. Hum. Genet* 70:192-206.
- Lewin, R.** (1999) *Evolução Humana*. 1ed. Atheneu Ed. São Paulo. pp526.
- Lorente, A. J.** (2003) Universidade de Granada, Lab. de Identificação Genética, Espanha.
- Malaspina, P., persichetti, F., Novelletto, A., Jodice, C., Terrenato, L., Wolfe, J., Ferraro, M., Prantera, G.** (1990) The human Y chromosome shows a low level of DNA polymorphism. *Ann. Hum. Genet* 3:297-305.
- Marshall, E.** (1997) Snipping Away at genome patenting. *Science*. 277:1752-1753.
- Mesa, N.R., Mondragon, M.C., Soto, I.D., Parra, M.V., Duque, C., Ortiz-Barrientos, D., Garcia, L.F., Velez, I.D., Bravo, M.L., Munera, J.G., Bedoya, G., Bortolini, M.C., Ruiz-Linares, A.** (2000). Autosomal, mtDNA and Y-chromosome diversity in Amerinds: pre- and post-colombian patterns of gene flow in south America. *Am. J. Hum. Genet* 67:1277-1286.
- Nei, M.** (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, pp512.
- Neves, W., Hubbe, M., Mazzuia, E.T.A., Atui, J.P.V.** (2003) *A primeira descoberta da América*. 1 ed. Sociedade Brasileira de genética. Ribeirão Preto. pp70.
- Parra, F.C.** (2003). *Ancestralidade Brasileira sob a ótica da genética*. Tese apresentada para obtenção do título de Doutor ao Departamento de Bioquímica e Imunologia, UFMG, Belo Horizonte.
- Pena, S.D.J., Chakraborty, R.** (1994) Paternity testing in the DNA era. *Trends in Genet* 10:204-209.
- Pena, S.D., Santos, F.R., Biachi, N.O., Bravi, C.M., Carnese, F.R., Rothlammer, F., Gerelsaikhhan T., Munkhtuja B., Oyunsuren T.** (1995) A major founder Y-chromosome haplotype in Amerindians. *Nature Genetics* .11:15-16.
- Prodöhl, P.A., Loughry, W.J., McDonough, C.M., Nelson, W.S., Avise, J.C.** (1996) Molecular documentation of polyembryony and the micro-spatial dispersion of clonal

- sibships in the nine-banded armadillo, *Dasypus novemcinctus*. *Proc R Soc Lond B Biol Sci.* 263(1377):1643-16499.
- Raymond, M. Rousset, F., Goudet, J.** (1996) Testing differentiation in diploid populations. *Genetics.*144(4):1933-1940.
- Sambrook, J., Fritsch, E.F., Maniats, T.** (1989) *Molecular Clonig: A laboratory Manual.* 2 ed. New York: Cold Spring Harbor Press.
- Santos, F.R., Hurtz M.H., Coimbra C.E.A., Santos R.V., Salzano F.M., Pena S.D.** (1995) Further evidence for existence of major founder Y chromosome haplotype in Amerindians. *Brazilian Journal of Genetics.*18:669-672.
- Santos, F.R., Tyler-Smith, C.** (1996) Reading the human Y chromosome: the emerging DNA markers and human genetic history. *Braz. J. Genet.* 19:665-670.
- Santos, F.R., Bianchi, N.O., Pena, S.D.J.** (1996) Worldwide distribution of human Y chromosome haplotypes. *Genome Res.*6:601-611.
- Santos, F.R., Pandya A., Tyler-Smith C., Pena S.D., Schanfield M., Leonard W.R., Osipova L., Crawford M.H., Mitchell R.J.**(1999) The central Siberian origin for native American Y chromosomes. *Am. J. H. Genet* 64:619-628.
- Santos ,F.R. e Tarazona-Santos,E.** (2002) A história do povoamento pré-colombiano das Américas e o cromossomo Y humano. In: "Homo brasilis : Aspectos Genéticos, Linguísticos, Históricos e Culturais da Formação do Povo Brasileiro" (Editor: Sérgio D.J. Pena, Editora FUNPEC).
- Santos, F. R., Carvalho-Silva, D. R. and Pena, S. D. J.** (1999). PCR-based DNA profiling of human Y chromosomes. In: DNA profiling and DNA fingerprinting. (Epplen, J. T. and T. Lubjuhn, ed.) edn. Birkhauser, basel, pp. 133-152.
- Santos, F.R., Pandya, A., Kayser, M., Mitchell, J., Liu, A., Singh, L., Destro-Bisol, G., Novelletto, A., Qamar, A., Mehdi, Q., Adhikari, R., Clegg, J., Kniff, P., Tyler-Smith, C.** (2000) A polymorphic L1 retroposon insertion in the centromere of human Y chromosome. *Hum. Mol. Genet* 9:421-430.
- Santos, M. S.**(1998) *Estudo da variabilidade do mtDNA na População brasileira.* Tese apresentada para btenção do título de Mestre ao Departamento de Bioquímica e Imunologia, UFMG, Belo Horizonte.
- Schlötterer, C., Tautz, D.** (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20:211-215.
- Schneider S, Roessli D, Excoffier L.** (2000) Arlequin, v 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory. University of Geneva, Geneva.
- Seielstad, M., Hebert, J., Lin, A., Underhill, P., Ibrahim, M., Vollrath, D., Cavalli-Sforza, L.L.** (1994) Constrution of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum. Mol. Genet* 3:2159-2161.
- Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S.** (2000) The genetic

legacy of paleolithic homo sapiens in extant Europeans: A Y chromosome perspective. *Science*.290:1155-1159.

- Silva, W.A.Jr., Bonatto, S.L., Holanda, A.J., Ribeiro-dos-Santos, A.K., Paixão, B.M., Goldman, G.H., Abe-Sandes, K., Rodriguez-Delfin, L., Barbosa, M., Paçó-Larson, M.L., Petzi-Erlar, M.L., Valente, V., Santos, S.E.B., Zago, M.A.** (2002) Mitochondrial genome diversity of native American supports a single early entry of founder populations into America. *Am. J. Hum. Genet* 71:187-192.
- Stone, A.C., Stoneking M.** (1998) mtDNA analysis of a prehistoric Oneota population: implications for peopling of the New World. *Am. J. Hum. Genet* 62:1153-1170.
- Stoneking, M., Soodyall.** (1996) Human evolution and the mitochondrial genome. *Current Opinion in Genetics & Development*. 6:731-736.
- Su, B., Xiao, J., Underhill, P., Deka, R., Zhang, W., Akey, J., Huang, W., Shen, D., Lu, D., Luo, J.** (1999) Y-chromosome evidence for northward migrations of modern humans into Eastern Asia during the last Ice age. *Am. J. Hum. Genet* 65:1718-1724.
- Syvänen, A.C.**(1999) From gels to chip: "Minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Human Mutation*.13:1-10.
- Tarazona-Santos, E.** (1999) *Variabilidade molecular do cromossomo Y em populações andinas*. Tese apresentada para obtenção do título de Mestre ao Departamento de Bioquímica e Imunologia, UFMG, Belo Horizonte.
- Tarazona-Santos, E.** (2002) *Variabilidade molecular do cromossomo Y em populações nativas americanas*. Tese apresentada para obtenção do título de Doutor ao Departamento de Bioquímica e Imunologia, UFMG, Belo Horizonte.
- Tarazona-Santos, E., Carvalho-Silva, D.R., Pettener, D., Luiselli, D., Stefano, G.F., Labarga, C.M., Rickards, O., Tyler-Smith, C., Pena, S.D.J., Santos, F.R.** (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, F.R.** (2002) The peopling of the Americas: a second major migration? *Am. J. Hum. Genet* 70:1377-1380.
- Tautz, D.** (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17:6463-6471.
- Templeton, A.** (1997) Out of Africa? What do genes tell us? *Curr. Opin. Genet Dev.* 7:841-847.
- Thompson, J.D.; Gibson, T.J.; Plewniak, F., Jeanmougin, F., Higgins, D.G.**(1997 ). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24:4876-4882.
- Torroni, A., Schurr, T.G., Young, C.C., Szathmari, E.J.E., Williams, R.C., Schanfield, M.S., Troup, G.A., Knowler, W.C., Lawrence, D.N., Weiss, K.M., Wallace, D.C.** (1992) Native American mitochondrial DNA analyses indicates that Amerind and Nadene populations were founded by two independent migrations. *Genetics*. 130:153-162.
- Underhill, P.A., Jin, L., Zemans, R., Oefner, P.J., Cavalli-Sforza, L.L.** (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci U S A* 93:196-200.

- Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L., Oefner, P.J.** (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996-1005.
- Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J.R., Mehdi, S.Q., Seielstad, M.T., Wells, R.S., Piazza, A., Davis, R.W., Feldman, M.W., Cavalli-Sforza, L.L., Oefner, P.J.** (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26(3):358-361.
- Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazón-Lahr, M., Foley, R.A., Oefner, P.J., Cavalli-Sforza, L.L.** (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Am. J. Hum. Genet* 65:43-62.
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., Wilson, A.C.** (1991) African populations and evolution of human mitochondrial DNA. *Science*.253:1503-1507.
- Vogt, P.H., Affara, N., Davey, P., Hammer, M., Jobling, M.A., Lau, Y.F., Mitchell, M., Schempp, W., Tyer-Smith, C., Williams, G., Yen, P., Rappold, G.A.**(1997) Report of the third international workshop on Y chromosome mapping. *Cytogenet Cell Genet* 79:1-20.
- Wallace, D.C.** (1995) Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am. J. Hum. Genet* 57:201-223.
- Weber, J.L., David, D., Heil, J., Fan, Y., Zhao, C., Marth, G.** (2002) Human diallelic insertion/deletion polymorphisms. *Am. J. Hum. Genet* 71:854-862.
- Whitfield, L.S., Sulston, J.E., Goodfellow, P.N.**(1995) Sequence variation of the human Y chromosome. *Nature*.378:379-380.
- Wright, S.** (1943) Isolation by distance. *Genetics*. 28:114-138.
- YCC** (2002). A nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Research*.12:339-448.
- Zegura, L.S., Karafet, T.M., Zhivoloviky, A. Hammer, M.F.** (2004) High-resolution SNPs and microsatellite haplotypes point to a single recent entry of Native American Y chromosomes into the Americas. *Mol. Biol. Evol.* 21(1):164-175.
- Zerjal, T., Dashnyam, B., Pandya, A., Roewer, L., Santos, F.R., Schiefenhovel, W., Fretwell, N., Jobling, M.A., Hariahara, S., Shimizu, K., Semjiddmaa, D., Sajantila, A., Salo, P., Crawford, M.H., Ginter, E.K., Evgrafov, O.V., Tyler-Smith, C.** (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet* 60:1174-1183.

#### Internet

- **Populações andinas:** [www.mapahumano.fietras.com](http://www.mapahumano.fietras.com)
- **Índios da Colômbia:** [www.indigenascolombia.org](http://www.indigenascolombia.org)
- **Mapa Quéchua:** [www.prodiversitas.bioetica.org/quechuas.html](http://www.prodiversitas.bioetica.org/quechuas.html)

- **Novos Indels:** [www.marshfieldclinic.org/research/genetics/indels/default.asp](http://www.marshfieldclinic.org/research/genetics/indels/default.asp)

# 8 Apêndices

Apêndice 1:

Carta da comissão Interamericana de direitos humanos da OEA para o Defensor nacional da Colômbia com data de 30 de Janeiro de 2003.

Bogotá, 30 de enero del 2003.

Doctor:

**EDUARDO CIFUENTES MUÑOZ**

Defensor Nacional del Pueblo.

Ciudad.

**REF: ALERTA TEMPRANA POR MASACRE ANUNCIADA CONTRA LA POBLACIÓN INDÍGENA Y CAMPESINA DE COYAIMA (TOLIMA).**

Estimado Doctor:

Con el presente escrito informo a su despacho sobre la grave situación que vive la población indígena de Coyaima, municipio del departamento del Tolima, con la finalidad de que se ponga en marcha el Sistema de Alertas Tempranas para garantizar la vida e integridad física de todas estas personas.

También queremos solicitar que a través suyo sean alertadas de estos hechos las siguientes autoridades civiles y militares:

- Fiscalía General de la Nación
- Procuraduría Nacional
- Gobernador del departamento del Tolima
- Comandante del Ejército en el Tolima
- Comandante de la Policía en el Tolima
- Alcalde de Coyaima

**ANTECEDENTES:**

En Coyaima (Tolima) actúa el Bloque Tolima de los paramilitares de las Autodefensas Unidas de Colombia (AUC), quienes realizan retenes en el kilómetro 14 de la vía Castilla-Coyaima y en la vereda Santa Rita la Mina de la vía Coyaima-Ataco. Los habitantes de la zona aseguran que en los citados controles ilegales, “bajan a la gente de los carros, comparan con la lista que ellos cargan y si está en la lista, los dejan y los desaparecen”. Este grupo está compuesto por unos diez paramilitares de alrededor de 30 años de edad, que se presentan como AUC-Bloque Tolima y se transportan en una camioneta marca Mazda con cabina, de 4 puertas, color blanco y vidrios oscuros.

Los mismos vecinos de la región afirman que para instalar los retenes los paramilitares cruzan el casco urbano de Coyaima sin que la fuerza pública ni el alcalde, David Loaiza Culma, hagan nada para evitarlo; de hecho, la Administración Municipal no se ha pronunciado frente a la situación de alto riesgo en que se encuentran los indígenas.

Dieciocho líderes indígenas, la mayoría de ellos pertenecientes a la Asociación de Cabildos Indígenas del Tolima (ACIT), fueron asesinados o desaparecidos a lo largo del pasado año a pesar de las denuncias que se interpusieron tras cada crimen ante las instituciones oficiales. Entre las víctimas figuran Egidio Matoma, Aquilino Oyola y Roque Oyola, del Cabildo Indígena Guayaquil y miembros del Partido Comunista Colombiano (PCC) y de la Unión Patriótica (UP); Marceliano Liz, en el corregimiento de Castilla; Efrén Pamo, en el casco urbano de Coyaima, y Agustín Poloche Matoma, líder indígena de 78 años y fundador del PCC de Coyaima. Además, sus familiares y otras personas se vieron obligados a abandonar la región por las constantes amenazas que recibieron de los paramilitares.

Muchos de estos casos ya fueron presentados por Reiniciar ante la Comisión Interamericana de Derechos Humanos en las audiencias que este alto tribunal ofreció en octubre del pasado año.

#### HECHOS RECIENTES:

A las 9 de la noche del 15 de enero de 2003, integrantes del Bloque Tolima incursionaron en la vereda Dolares Porvenir y tirotearon las puertas de la casa de Rosa Tique. Este grupo armado ilegal estaba buscando a Hernán Loaiza Tique, gobernador indígena de esa localidad durante el año pasado; al no encontrarlo, se llevaron a su hermano, Fredy Loaiza Tique, a quien golpearon y torturaron.

Ocho días después, el 23 de enero, los paramilitares regresaron y volvieron a atacar la misma vivienda provocando el terror entre los habitantes de la población. En esta ocasión lanzaron granadas contra la casa, que quedó en ruinas; afortunadamente, las personas que la ocupaban la abandonaron momentos antes de la agresión.

A todo esto se añade el anuncio efectuado por los propios paramilitares de que disponen de una lista de personas indígenas y campesinas que van a ser asesinadas o desaparecidas. En el listado aparecen, entre otros, los siguientes nombres: Crispín Otavo, dirigente indígena y secretario político del PCC en Coyaima; Joselino Alape Ortiz, asesor de la ACIT; Porfidio Esquivel, presidente de la ACIT; David Quintero; Martín Loaiza Rodríguez, funcionario del municipio de Coyaima; María Nubia Loaiza Oviedo, funcionaria de Coyaima; Hernán Loaiza Tique, ex gobernador del cabildo indígena de Dolares Porvenir; Henry Loaiza Tique, concejal de la UP en Coyaima; Arquímedes Onatra, dirigente indígena y concejal de la UP en Coyaima; José Ángel Zapata Ducuara, ex concejal de la UP en Coyaima y dirigente indígena en la Comisión de Conciliación de la ACIT; Diubino Santa, ex concejal de la UP en Coyaima y líder indígena, y Rosalía Poloche, tesorera de la ACIT.

De los hechos descritos anteriormente se deduce que se está preparando una nueva acción criminal del Bloque Tolima de las AUC de mayor proporción que las cometidas en el pasado cuyo objetivo sería aterrorizar a toda la población indígena y campesina de Coyaima atacando a los líderes de sus organizaciones sociales y políticas, (ACIT, PCC y UP).

Por todo lo expuesto, solicitamos de su oficina de forma respetuosa y urgente que:

-Una vez activada esta Alerta Temprana, se dé seguimiento a la reacción de las distintas autoridades civiles y militares implicadas para lograr el fin de preservar la integridad de la población amenazada.

-Se inste a la Fuerza Pública de la zona a cumplir con su obligación de proteger a los civiles de Coyaima de la actuación de los grupos armados ilegales que operan en la zona y, por tanto, se impida su accionar y se persiga contundentemente a sus integrantes.

-Se garantice la vida de todos los pobladores, especialmente, de las personas que mayor riesgo corren por figurar sus nombres en la lista que manejan los paramilitares.

#### CORPORACIÓN REINICIAR

C.C.: Comisión Interamericana de Derechos Humanos de la OEA.