

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA



Tese de Doutorado

**História Genética
das
Populações Peruanas**

Orientado: MSc. José Raúl Sandoval Sandoval

Orientador: Dr. Fabrício Rodrigues dos Santos

Fevereiro - 2013

José Raúl Sandoval Sandoval

**História Genética
das
Populações Peruanas**

Tese apresentada ao Departamento de Biologia Geral do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção de título de Doutor em Genética. Área de concentração: Genética Evolutiva e de Populações

Orientador:

Dr. Fabrício Rodrigues dos Santos

Universidade Federal de Minas Gerais

Instituto de Ciências Biológicas

Departamento de Biologia Geral

Belo Horizonte-MG-Brasil

2013

Agradecimentos

Agradeço a meu orientador Fabrício Rodrigues dos Santos por todo o apoio no Laboratório de Biodiversidade e Evolução Molecular (LBEM), pela oportunidade oferecida, a qual me possibilita apresentar esta tese.

A Sergio Pena pela parceria e por compartilhar suas experiências.

A Daniela Lacerda e Marilza Jota pelo apoio na parte experimental assim como nos procedimentos e análises.

A Heloisa Pena (GENE) e Gisele Dantas pelas sugestões em algumas análises.

A Ricardo Fujita, Alberto Salazar e Oscar Acosta (Centro de Genética y Biología Molecular, Facultad de Medicina Humana, Universidad de San Martín de Porres, Lima), Cesar Ñique (USAT, Lambayeque), Jaime Descailleaux e Margarita Velásquez (UNMSM, Lima), Douglas C. Wallace e Olga Derbeneva (Center for Molecular and Mitochondrial Medicine and Genetics (MAMMAG), University of California, Irvine, CA, USA), pela parceria associada a este projeto.

Aos amigos(as) do Projeto Genográfico Sul-Americano por compartilhar experiências nas expedições de campo.

Aos amigos(as) do LBEM e de outros labs por compartilhar a amizade e convivência neste período de tempo.

A todos os voluntários-participantes da amostragem que nos permitiram montar este projeto.

A Romeu Cardoso Guimarães, pelas correções do português.

À CAPES e ao CNPq pela bolsa PEC-PG e à National Geographic Society e à FAPEMIG pelos fundos necessários à realização deste projeto. Muito obrigado!

Agradeço a minha esposa Roxana e meu filho Lev por ser o núcleo da minha esperança e horizonte na vida, assim como pelo apoio, e a meus irmãos(ãs) Salvador, Susana, Justa, Benito, Juan, Julia, e familiares.

Sumário

I - LISTA DE FIGURAS	I
II - LISTA DE TABELAS	V
III - LISTA DE ABREVIATURAS E SÍMBOLOS	VIII
1 – RESUMO	1
2 – ABSTRACT	2
3 – INTRODUÇÃO	3
3.1- POVOAMENTO PRÉ-COLOMBIANO DO PERU	3
3.2- HISTÓRIA DOS HABITANTES DO LAGO TITICACA	4
3.3- HISTÓRIA DA COLONIZAÇÃO PÓS-COLOMBIANA DO PERU	5
3.3.1- Os fluxos migratórios e a miscigenação no Peru	5
3.4- DIVERSIDADE GENÉTICA E EVOLUÇÃO HUMANA	7
3.4.1- Estudos históricos com o mtDNA humano	7
3.4.2- Estudos históricos com o cromossomo Y humano	9
3.4.3- Estudos com marcadores autossômicos do tipo INDELS	12
3.5- HIPÓTESES E JUSTIFICATIVAS DO ESTUDO COM POPULAÇÕES PERUANAS	13
4 – OBJETIVOS	14
4.1- OBJETIVO GERAL	14
4.2- OBJETIVOS ESPECÍFICOS	14

5 - MATERIAL E MÉTODOS	15
5.1- AMOSTRAGEM POPULACIONAL	15
5.2- ANÁLISES DOS MARCADORES DE DNA	16
5.2.1- Região controle do mtDNA	16
5.2.2- Genotipagem do cromossomo Y	19
5.2.3- Painel de 40 marcadores autossômicos de tipo INDEL	22
6 - RESULTADOS	24
6.1- SOBRE A REGIÃO CONTROLE DO MTDNA	24
6.1.1- Polimorfismos da Região Controle do mtDNA nas populações peruanas	28
6.1.2- Análises de regressão linear dos sub-haplogrupos A2, B2, C1 e D1 segundo a latitude	32
6.1.3- Determinação dos agrupamentos A2, B2, C1 e D1 de mtDNA nas populações peruanas por PhyML	33
6.1.4- Índices de diversidade dos sub-haplogrupos A2, B2, C1 e D1 e teste de neutralidade	35
6.1.6- Comparação das linhagens A2, B2, C1 e D1 com outras linhagens maternas das populações continentais	45
6.1.7- Determinação filogenética de outras linhagens de mtDNA não autóctones reveladas nas populações peruanas	49
6.2- SOBRE OS SNPs E STRs DO CROMOSSOMO Y	52
6.2.1- Análises das linhagens paternas através do programa Network	54
6.2.2- Variabilidade genética das populações peruanas segundo os 17 Y-STRs	59
6.3- A HISTÓRIA GENÉTICA UNIPARENTAL DOS AIMARÁS E QUÉCHUAS DO LAGO TITICACA	60
6.4- ANÁLISES DE GENOMAS COMPLETOS DE MTDNA DOS HAPLOGRUPOS A2, B2, C1 E D1 ENCONTRADOS NO PERU	69
6.4.1- Determinação da divergência genética das linhagens A2, B2, C1 e D1 por mtPHYL	74

6.5- SOBRE OS INDELS AUTOSSÔMICOS	77
6.5.1- Determinação das proporções de ancestralidade das populações peruanas através de STRUCTURE	77
6.5.2- Teste de ancestralidade das populações peruanas através do método Mínimos Quadrados	83
6.5.3- Agrupamento das populações peruanas segundo BAPS	84
6.5.4- Agrupamento das populações peruanas por Componentes Principais	86
6.5.5- Análises da variabilidade genética nas populações peruanas	87
7 – DISCUSSÃO	92
7.1- SOBRE O mTDNA	92
7.2- SOBRE O CROMOSSOMO Y	96
7.3- SOBRE OS INDELS	97
8 – CONSIDERAÇÕES FINAIS	99
9 – REFERÊNCIAS BIBLIOGRÁFICAS	100
10 – ANEXOS	109

I - Lista de Figuras

Figura 1: Árvore das principais linhagens maternas segundo o PhyloTree (http://www.phylotree.org) (van Oven & Kayser, 2009), baseada nas sequências completas de mtDNA.	9
Figura 2: Árvore das principais linhagens paternas baseado nos SNPs do cromossomo Y, segundo Chiaroni <i>et al.</i> (2009).	12
Figura 3: Localização aproximada das cidades ou comunidades, de onde procedem as amostras, no mapa do Peru.	16
Figura 4: Análises do sequenciamento do DNA mitocondrial.	19
Figura 5: Genotipagem dos 17 Y-STRs (amostra de Ayacucho) e visualização através de GeneMapper v.3.2. Os alelos para cada marcador são mostrados em forma de picos que correspondem ao número de repetições, que está anotado abaixo de cada pico.	20
Figura 6: Genotipagem dos 40 <i>loci</i> com INDELS de uma amostra testada (referência) e visualizada através de GeneMapper ID v3.2.	23
Figura 7: Distribuição das frequências absolutas dos sub-haplogrupos A2, B2, C1 e D1 do mtDNA nas 29 locais peruanas (n=700).	26
Figura 8: Gráfico bidimensional PCA com análise das frequências dos sub-haplogrupos A2, B2, C1 e D1 do mtDNA das subpopulações peruanas.	27
Figura 9: Análise de regressão linear entre a latitude e o sub-haplogrupo A2 (A).	32
Figura 10: Análise de regressão linear entre a latitude e o sub-haplogrupo B2 (B).	33
Figura 11: Árvore gerada pelo PhyML sob modelo TN93 para os sub-haplogrupos A2, B2, C1 e D1.	34
Figura 12: Mapa fatorial das análises de PCA de haplótipos dos sub-haplogrupos A2, B2, C1 e D1 nas populações peruanas baseado na região controle de mtDNA.	34
Figura 13: Plotagem MDS baseada na distância genética (Fst de Reynolds) das 29 subpopulações segundo todos os sub-haplogrupos mtDNA (A2, B2, C1 e D1).	43
Figura 14: Plotagem MDS baseada na distância genética (Fst de Reynolds) das populações segundo o sub-haplogrupo A2.	43
Figura 15: Plotagem MDS baseado na distância genética (Fst de Reynolds) das populações segundo o sub-haplogrupo B2.	44
Figura 16: Plotagem MDS baseada na distância genética (Fst de Reynolds) das populações segundo o sub-haplogrupo C1.	44
Figura 17: Plotagem MDS baseada na distância genética (Fst de Reynolds) das populações segundo o sub-haplogrupo D1.	45

Figura 18: Rede MJ (median joining) dos haplogrupos ou sub-haplogrupos de mtDNA comparados no nível continental.....	48
Figura 19: Rede MJ (median joining) dos haplogrupos ou sub-haplogrupos de mtDNA comparados no nível continental segundo por regiões geográficas.	48
Figura 20: Plotagem MDS das distâncias genéticas F_{st} de Reynolds entre as populações selecionadas do painel de HGDP-CEPH e do Peru, segundo as sequências de mtDNA da Região Controle.....	49
Figura 21: Rede MJ feita pelo programa Network. Os sub-haplogrupos A2, B2, C1 e D1 correspondem à população de Cabanaconde (Cb). nd=não determinado por predição.	51
Figura 22: Plotagem da distribuição das frequências dos Y-SNPs (Tabela 10) através de componentes principais (PC).	54
Figura 23: Rede MJ sobre as linhagens paternas e baseado nos SNPs e 17 Y-STRs (n=371).	55
Figura 24: Rede MJ de 178 haplótipos dos 17 Y-STRs do Hg Q-M3 do cromossomo Y nas 23 populações peruanas (n=229). As cores representam as regiões: verde oliva (Amazônia); prata (norte dos Andes), marron (centro dos Andes), branco (sul dos Andes), vermelho (Taquile), celeste (Amantani), verde (Uros), amarelo (Anapia) e azul (Lima). Os círculos representam haplótipos com área proporcional à frequência de indivíduos.	56
Figura 25: Rede MJ da distribuição dos haplótipos dos 17 Y-STR do subHg Q-M3 em algumas populações do Peru.....	57
Figura 26: Gráfico bidimensional PCA sobre a distribuição dos 178 haplótipos dos 17 Y-STR da linhagem Q-M3 nas 23 populações peruanas (n=229) agrupadas por regiões geográficas. Cada cor representa indivíduos-haplótipos de sua correspondente população. AndesN (Andes do Norte); AndesC (Andes do Centro); AndesS (Andes do Sul). Titicaca (Taquile, Amantani, Uros, Anapia, Puno); Coast (Lima, Lambayeque); Amazon (Andoas, Pucallpa, Chachapoyas).....	58
Figura 27: Plotagem MDS através da distância genética (R_{st} de Reynolds) entre as populações peruanas usando 17 Y-STRs.....	60
Figura 28: Rede MJ da distribuição de 42 haplótipos (65 sítios polimórficos) da Região Controle do mtDNA (1.116 pb) dos indivíduos de Taquile, Amantani, Uros, Anapia e Puno (n=139).	63
Figura 29: Matriz de 45 haplótipos mtDNA com 75 sítios polimórficos em relação à rCRS (Hap_1).....	65
Figura 30: Rede de haplótipos gerada pelo programa TCS com haplótipos de indivíduos do Lago Titicaca. Hd (diversidade haplotípica)=0,95; haplótipos=45; S=75; tamanho das sequências de mtDNA=1.114 pb.	66

Figura 31: Rede MJ baseada na distribuição de 47 haplótipos dos 17 Y-STRs que correspondem à linhagem Q-M3 (Q1a3a) distribuída nas ilhas Taquile, Amantani, Uros e Anapia (n=76). Os círculos correspondem às proporções das frequências; a cor vermelha à Taquile; celeste à Amantani; verde à Uros, amarelo à Anapia.	67
Figura 32: Rede MJ sobre a localização dos indivíduos do Lago Titicaca em relação aos demais indivíduos do Peru.	68
Figura 33: Gráfico bidimensional PCA dos haplótipos dos 17 Y-STRs do Haplogrupo Q-M3 das populações do Lago Titicaca (n=79, incluindo 3 de Puno).	68
Figura 34: Árvore filogenética dos 21 mtDNA completos segundo EBS e o modelo HKY85 que está de acordo com o modelo padrão sobre a árvore mtDNA humano sugerido pelo PhyloTree (van Oven & Kayser, 2009).	72
Figura 35: Resultados da coalescência das 21 sequências mtDNA completas. Valores segundo a mediana (entre parêntese o desvio padrão) em anos.	72
Figura 36: Resultados da coalescência das 21 sequências mtDNA da região codificadora. Valores segundo a mediana (entre parêntese o desvio padrão) em anos.	73
Figura 37: Reconstrução filogenética das 21 sequências de genomas de mtDNA completos usando o programa PhyML segundo o modelo TN93, aLRT, NNI, BioNJ, bootstrap=100 Bootstraps: HgA2=100%; HgB2=98%; HgC1=100%; HgD1=79%.	74
Figura 38: Filogenia dos genomas completos de mtDNA dos sub-haplogrupos A2, B2, C1 e D1 dos Aimarás do Lago Titicaca, através do programa mtPhyl.	76
Figura 39: Filogenia das 21 sequências de genomas completos de mtDNA por Máxima Parcimônia através de mtPHYL.	77
Figura 40: Resultado de $\Delta K=2$ (valor da moda=190.05) segundo o método de Evanno que foi obtida através do programa STRUCTURE HARVESTER. Os 551 indivíduos do Peru são agrupados em dois conglomerados. $\ln P(K)=-24225.1$; $SD=2.32$; $\ln'(K)=213.33$; $ \ln'(K) =440.45$; 10 réplicas.	79
Figura 41: Plotagem por DISTRUCT da partição K=2 das populações peruanas (códigos em Material e Métodos) usando como referência as populações da Europa (n=161), Leste Asiático (n=251) e da América (n=108) do painel do HGDP-CEPH. As populações estão representadas por segmentos delimitados e os indivíduos pelas linhas verticais coradas segundo sua proporção de ancestralidade.	80
Figura 42: Plotagem por DISTRUCT das partições desde K=3 até K=6 obtido por STRUCTURE. As populações de referência do painel HGDP-CEPH foram divididas <i>a priori</i> em quatro macro-regiões (África; Eurásia (Europa, Oriente Médio, Ásia Central, Leste Asiático); Oceania; América). As populações estão representadas por segmentos delimitados e os indivíduos pelas linhas verticais coradas segundo sua proporção de ancestralidade. Amer=América; O=Oceania.	80

Figura 43: Plotagem por DISTRUCT da partição K=5 das populações peruanas (códigos em Material e Métodos) usando como referência as 52 populações do painel HGDP-CEPH. As populações estão representadas por segmentos delimitados e os indivíduos pelas linhas verticais coradas segundo sua proporção de ancestralidade.	81
Figura 44: Plotagem 3D das populações peruanas segundo os valores Q (<i>ancestry proportion</i>) na partição K=3 (Eurásia_Oceania; África, e América). Referência: HGDP-CEPH: n=157 da África; n=760 da Eurásia; n=108 da América; n=39 da Oceania.	81
Figura 45: Plotagem por STRUCTURE da partição K=3 (África, Europa e América).	82
Figura 46: “Trained” admixture clustering por BAPS assumindo <i>a priori</i> K=4. AFR=África, O=Oceania; A=América; Eurásia (Europa, Oriente Médio, Ásia Central, Leste Asiático). Veja os códigos das localidades peruanas em Material e Métodos. As análises estão baseadas em 10.000 iterações das probabilidades <i>a posteriori</i> das frequências alélicas e 10 réplicas.	85
Figura 47: Plotagem dos indivíduos segundo as populações em agrupamento K=5 por STRUCTURE, segundo o modelo “não miscigenação” e “frequências alélicas não correlacionadas”. MCMC=2x10 ⁶ , Burn-in=100.000.	85
Figura 48: Plotagem das populações pelo método de PCA. Os quadrados representam as localidades mais miscigenadas que estão do lado contrário daquelas representadas por triângulos (menos miscigenadas).	86
Figura 49: Plotagem pelo método FCA (Factorial Correspondence Analysis) dos indivíduos que são agrupados segundo seus genótipos nos dois principais eixos (Axis1 e Axis2). Os quadrados representam aos indivíduos (de CA, CAsm, Char, Ch, LA, LI) formando um gradiente próximo à Eurásia que corresponde ao grau de miscigenação genética. Os triângulos representam indivíduos com menor grau de miscigenação (Taq, Amt, Ap, Ur e Yke).	87
Figura 50: Plotagem MDS das distâncias genéticas (Fst de Reynolds) entre as 25 populações peruanas (códigos em Material e Métodos) e as 18 populações selecionadas do painel HGDP-CEPH.	91
Figura 51: Plotagem dos índices de heterozigosidade esperada (He) e dos coeficientes da proporção de ancestralidade (<i>Q-membership ancestry proportions</i>) de Not-America quando as populações peruanas são agrupadas em K=2 (da Tabela 15) usando como referência o painel HGDP-CEPH.	91
Figura A1: Painel de 10 marcadores SNPs usados (círculos cheios) no estudo da determinação das linhagens do cromossomo Y nas populações peruanas. (Figura modificada de Karafet <i>et al.</i> , 2008 e de YCC, 2002)	131

II - Lista de Tabelas

Tabela 1: Distribuição de frequências (absolutas) dos sub-haplogrupos A2, B2, C1 e D1(+D4) de mtDNA nas subpopulações peruanas segundo as regiões (Amazônia, Andes e Costa) e locais.....	25
Tabela 2: Índices de diversidade baseados nas sequências da Região Controle do mtDNA. As amostras (n=700) correspondem às 29 localidades das três regiões do Peru (Amazônia, Andes e Costa).....	35
Tabela 3: Teste de neutralidade de Tajima baseado nas sequências mtDNA da Região Controle. As amostras (n=700) correspondem às 29 localidades das três regiões do Peru (Amazônia, Andes e Costa).....	36
Tabela 4: Resultado de AMOVA. Análises do mtDNA nas 29 subpopulações peruanas segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100	37
Tabela 5: Resultado de AMOVA. Análises do sub-haplogrupo A2 do mtDNA em 24 subpopulações peruanas (cinco não apresentam A2) segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100	38
Tabela 6: Resultado de AMOVA. Análises do sub-haplogrupo B2 do mtDNA em 27 subpopulações peruanas (duas não apresentam B2) segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100	39
Tabela 7: Resultado de AMOVA. Análises do sub-haplogrupo C1 do mtDNA em 24 subpopulações peruanas (cinco não apresentam C1) segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100	40
Tabela 8: Resultado de AMOVA. Análises do sub-haplogrupo D1 do mtDNA em 22 subpopulações peruanas (sete não apresentam D1) segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100	41
Tabela 9: Resultado de AMOVA. Análises da Região Controle mtDNA nas 47 subpopulações continentais (n=1033) segundo o agrupamento hierárquico. Iterações=10.100.....	47
Tabela 10: Frequências (absolutas) dos haplogrupos do cromossomo Y nas populações peruanas.....	53
Tabela 11: Resultado de AMOVA. Análises de 17 Y-STRs correspondentes ao haplogrupo Q-M3 nas 23 subpopulações peruanas (n=229, 3 regiões) segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100.....	59
Tabela 12: Testes de neutralidade entre as populações do Lago Titicaca.....	62
Tabela 13: Resultado de BEAST usando o EBSP (Extended Bayesian Skyline Plot) segundo o modelo de substituição nucleotídica HKY85. MCMC=100.000.000 de iterações para a	

corrida separadamente por haplogrupos, e MCMC=80.000.000 de iterações para a corrida de 21 sequências de mtDNA completas. Foram feitas mínimo 5 corridas para cada teste. HPD=Highest Posterior Density; T=Tempo; s.e.=standard error 71

Tabela 14: Determinação do tempo de coalescência dos sub-haplogrupos mtDNA através do parâmetro estatístico *Rho* (ρ) no programa mtPHYL. σ =desvio padrão 75

Tabela 15: Coeficientes (média) de proporção de ancestralidade (Q) segundo a partição K=5 obtida por STRUCTURE em relação às 52 populações de referência do painel HGDP-CEPH. Também são apresentados os resultados da partição K=2 para a comparação das proporções de ancestralidade entre dois grupos: not-America e America. ME=Middle East; CA=Central Asia. 82

Tabela 16: Comparação dos coeficientes ou proporções de ancestralidade em agrupamento K=2 obtido pelos programas STRUCTURE e ADMIX. 83

Tabela 17: Resultado de AMOVA em níveis hierárquicos sobre 25 populações peruanas através dos 40 marcadores INDELS. 90

Tabela 18: Comparação dos tempos de coalescência determinados através do parâmetro estatístico *Rho* (ρ) usando diferentes taxas de mutação para a região codificadora do mtDNA. 95

Tabela A1: Haplótipos (Hp) com seus respectivos SNPs / variantes na Região Controle mtDNA do sub-haplogrupo A2 nas populações peruanas (n=90). Os sítios ou posições nucleotídicas são anotadas segundo o rCRS (revised Cambridge Reference Sequence). +=inserção; d=deleção; as transversões são anotadas depois do sítio com seu variante, com letra. 109

Tabela A2: Haplótipos (Hp) com seus respectivos SNPs / variantes na Região Controle mtDNA do sub-haplogrupo B2 nas populações peruanas (n=234). Os sítios ou posições nucleotídicas são anotadas segundo o rCRS (revised Cambridge Reference Sequence). +=inserção; d=deleção; as transversões são anotadas depois do sítio com seu variante, com letra. 113

Tabela A3: Haplótipos (Hp) com seus respectivos SNPs / variantes na Região Controle mtDNA do sub-haplogrupo C1 nas populações peruanas (n=110). Os sítios ou posições nucleotídicas são anotadas segundo o rCRS (revised Cambridge Reference Sequence). +=inserção; d=deleção; as transversões são anotadas depois do sítio com seu variante, com letra. 120

Tabela A4: Haplótipos (Hp) com seus respectivos SNPs / variantes na Região Controle mtDNA do sub-haplogrupo D1 nas populações peruanas (n=65). Os sítios ou posições nucleotídicas são anotadas segundo o rCRS (revised Cambridge Reference Sequence). +=inserção; d=deleção; as transversões são anotadas depois do sítio com seu variante, com letra. Entre parêntese a linhagem diferente de D1. 124

Tabela A5: Predição de haplogrupos para 34 amostras para as quais foram determinados seus correspondentes <i>clusters</i> na árvore filogenética mitocondrial. HPT=Haplogroup Prediction Tool do projeto Genográfico; n.d.=não determinado.....	126
Tabela A6: Lista dos haplótipos únicos de STR do haplogrupo Q-M3 do cromossomo Y. A ordem dos 17 Y-STRs corresponde da esquerda à direita: DYS393-DYS390-DYS19-DYS391-DYS385a-DYS385b-DYS439-DYS389a-DYS392-DYS389b-DYS458-DYS437-DYS448-GATAH4-DYS456-DYS438-DYS635.....	127

III - Lista de Abreviaturas e símbolos

HGDP-CEPH- Human Genome Diversity Panel-CEPH in France

NRY- non-recombinant region of chromosome Y

YCC- Y Chromosome Consortium

INDEL- insertion/deletion

SNP- single nucleotide polymorphism

Q- coefficient membership of ancestry proportion

AIMs- ancestry informative markers

BAPS- Bayesian Analyses of Population Structure

AMOVA- analysis of molecular variance

MDS- multidimensional scaling

MCMC- Markov Chain Monte Carlo

PCA- principal components analysis

HVS- hypervariable segment of mtDNA

rCRS- revised Cambridge Reference Sequence

RT-PCR- real time-polymerase chain reaction

TCS- Templeton, Crandall and Sing method

MJ-MP- median joining and maximum parsimony

PhyML- phylogeny by maximum likelihood

TN93- Tamura-Nei 1993 model

HKY85- Hasegawa Kishino Yano 1985 model

BEAST- Bayesian Evolutionary Analysis by Sampling Trees

BEAUti- Bayesian Evolutionary Analysis Utility

EBSP- Extended Bayesian Skyline Plot

MRCA- most recent common ancestor

1 – RESUMO

As populações peruanas atuais devem sua formação a dois movimentos migratórios principais, um antigo ou pré-colombiano e outro mais recente ou pós-colombiano. Neste estudo, tentamos reconstruir o impacto destes fluxos migratórios através do estudo genômico dos peruanos atuais usando a comparação de genótipos (marcadores autossômicos INDELS) e haplótipos (mtDNA e Y-DNA) dos mesmos com populações nativas continentais da Europa, Ásia, África, Oceania e América.

Nossos resultados com dados autossômicos mostram que há uma elevada proporção de ancestralidade autóctone americana (entre 53,8% e 96,5%) e um gradiente de miscigenação pós-colombiana nas populações peruanas que é consistente com os registros que deixaram os cronistas e historiadores. Em conjunto, a população atual do Peru apresenta uma proporção de ancestralidade autóctone de aproximadamente 80%.

Nos estudos com linhagens maternas (mtDNA), a distribuição de suas frequências indica que a contribuição matrilinear é em grande parte de proporção autóctone americana (97,1% de A2, B2, C1 e D1), representando principalmente o legado pré-colombiano do Peru. Nos estudos com linhagens paternas (cromossomo Y), há predominância do haplogrupo Q-M3 (60,1%); seguida pelos haplogrupos R-M207 (18,8%); F-M89* (12,5%); E-M96 (4,3%), e o paragrupo Q-M346* (4,1%). A alta prevalência da linhagem Q (Q-M3 e Q-M346*) sugere que a contribuição paterna autóctone americana é três vezes maior do que a contribuição do haplogrupo R-M207, a principal linhagem de ascendência européia. Por outro lado, observa-se uma contribuição menor de linhagens da Eurásia, do Oriente Médio e da África (F-M89*, e E-M96).

Os eventos de “miscigenação” pós-colombiana no Peru, embora menores do que observado em outros países latino-americanos, também ocorreram, em sua maioria, entre homens de linhagens euro-asiáticas com mulheres nativas peruanas. O gradiente de miscigenação observado com as linhagens paternas ao longo das diferentes regiões do Peru é, em geral, muito similar ao padrão obtido pelos marcadores autossômicos do tipo INDEL. Desta forma, os três sistemas genéticos utilizados se mostraram complementares na reconstrução histórica do povoamento do Peru, pois descreveram o mesmo panorama a partir de diferentes perspectivas, revelando importantes inferências demográficas e históricas do ser humano na América.

2 – ABSTRACT

The current Peruvian populations were mostly derived from two major migratory events, one ancient or pre-Columbian and another more recent or post-Columbian. In this study, we attempt to reconstruct the impact of these migration flows through a genomic study of the Peruvians using the comparisons of genotypes (autosomal markers-INDELS) and haplotypes (mtDNA and Y-DNA) of them with the autochthonous populations from Europe, Asia, Africa, Oceania and America regions.

Our results with autosomal data show that there is a high proportion of autochthonous American ancestry (between 53.8% and 96.5%) and a gradient of post-Columbian miscegenation, which is consistent with the historical records left by historians and chroniclers. Overall, the current population of Peru shows a genomic proportion corresponding to approximately 80% of native American ancestry.

The distribution and frequencies of maternal lineages (mtDNA) indicates a major contribution of autochthonous Americans (97.1% of A2, B2, C1 and D1), representing mainly the pre-Columbian genetic legacy of Peruvians. In the paternal lineages (Y chromosome) there is a predominance of Q-M3 (60.1%), followed by R-M207 (18.8%), F-M89* (12.5%), E-M96 (4.3%) and Q-M346* paragroup (4.1%). The high prevalence of the Q lineage (Q-M3 and Q-M346*) shows that the autochothonous paternal contribution is at least three times larger than the contribution R-M207 lineage, the main haplogroup from European descent. Besides, there is also a minor contribution of lineages from Eurasia, Middle East and Africa regions (F-M89* and E-M96).

The post-Columbian “miscegenation” events in Peru, although lower than those observed in other Latin American countries, also occurred mostly among men from Eurasian lineages and autochthonous Peruvian women. The gradient of miscegenation of paternal lineages observed throughout the different regions of Peru is very similar to the genetic pattern obtained by autosomal INDELS. Thus, the three genetic systems were shown to complement each other on the reconstruction of history of peopling of Peru, describing the same scenario from different perspectives, revealing important demographic and historical inferences of the human being in America.

3 – INTRODUÇÃO

Estudos prévios sobre as populações peruanas foram iniciados no ano 2000 por nosso grupo na UFMG em Belo Horizonte, Brasil, e pelos colaboradores deste projeto no Centro de Genética e Biologia Molecular da Faculdade de Medicina Humana da Universidade de San Martín de Porres (USMP) em Lima, Peru. Estes estudos tinham por objetivo em seu início, desvendar as relações intra e interpopulacionais, bem como identificar os componentes genéticos que revelam parte da história peruana, suas associações às doenças, assim como outros processos evolutivos do *Homo sapiens*. Neste período, esses estudos também permitiram a formação de um banco de amostras de diversas populações do Peru que foi utilizado nesta tese sobre ancestralidade e miscigenação dos peruanos. Neste estudo tentamos desvendar a estrutura populacional atual do Peru, em localidades da Costa, Andes e Amazônia, considerando processos estocásticos similares ao modelo de “ilhas em equilíbrio”, no qual o fluxo gênico e a deriva são fatores que influenciam marcadamente a história demográfica das populações. Foi levada em consideração a história conhecida de povoamento do país, tanto sobre a ancestralidade autóctone americana, da época pré-colombiana, quanto os influxos migratórios pós-colombianos. Além disto, um estudo detalhado foi elaborado com as populações da região do Lago Titicaca.

3.1- Povoamento pré-colombiano do Peru

Ao longo da Costa (litoral do Oceano Pacífico), no sul do Peru, foram encontrados vestígios dos primeiros grupos de caçadores e coletores litorâneos datados ao redor de 13.000 anos atrás, em Quebrada Jaguay (Sandweiss *et al.*, 1998). Um sítio arqueológico ao norte de Lima, também registra a cidade de Caral, uma das civilizações mais antigas da América do Sul, com data aproximada de 4.500 anos atrás (León Canales, 2007). Ao redor de 3.000 anos atrás (Cáceres, 1989; Lumbreras, 1981), outras civilizações apareceram, tal como Chavin no norte; Paracas no sul; e logo após 2.100 anos atrás apareceram outras como os Moche (litoral norte); Nazca (costa central); Wari (centro-sul dos Andes), Chimu (litoral norte); Chachapoyas (atual Amazonas), e, finalmente, a civilização do Império Tawantin Suyu (1432-1532) que foi dominado pelos Incas (Lexus, 1998).

Alguns restos arqueológicos e ósseos indicam que os primeiros grupos de caçadores e coletores chegaram à região andina do Peru ao redor de 12.000 anos atrás (transição Pleistoceno/Holoceno), habitando áreas próximas a Guitarrero Cave em Ancash (Lynch & Kennedy, 1970) e Ayacucho (MacNeish, Berger & Protscha, 1970). Foram encontrados esqueletos humanos de 10.000 anos atrás nos Andes, em Lauricocha e Huánuco, além de outros lugares do Altiplano andino (Lexus, 1998). Algumas áreas próximas ao Lago Titicaca

tiveram primeira ocupação humana com uma datação de aproximadamente 4.000 anos atrás (Aldenderfer, Craig, Speakman & Popelka-Filcoff, 2008).

O império incaico (Tawantin Suyu), em seu apogeu, abrangia os atuais territórios andinos do Peru, Equador e Bolívia, parte da Colômbia, Argentina e Chile, além de algumas áreas da Amazônia peruana e boliviana. Durante o seu apogeu, os Incas deslocavam algumas famílias (“recompensados” ou “punidos”) de uma região para outra, denominadas “mitimaes” ou “mitma” (por exemplo, alguns grupos de Cañaris do Equador foram deslocados para a região de Cusco e do Lago Titicaca, os “pastos” da Colômbia foram também levados a Cusco, alguns quéchuas foram transferidos para o sul da Bolívia e norte da Argentina, etc.). Assim, os Incas asseguravam uma administração política sem muitas oposições, facilitavam o intercâmbio de labores, além de impor a língua Quéchua como a dominante na maior parte do Império. No entanto, devido às lutas de poder dentro do Tawantin Suyu, os Incas estavam divididos quando apareceu um grupo de exploradores espanhóis, um sinal de presságio da decadência do império incaico (Lexus, 1998).

3.2- História dos habitantes do Lago Titicaca

Desde os tempos pré-colombianos até hoje, os habitantes do lago Titicaca trabalham na agricultura, pecuária e pesca. No século XV, a maior parte desta região era habitada pelos Collas (de língua aimará), que foram subjugados pelo Inca Pachacutec.

Atualmente, os moradores das ilhas do Lago Titicaca e de maior parte da região de Collao (região do Altiplano ao redor do lago) falam Quéchua e Aimará, línguas irmãs do tronco andino. No entanto, as crônicas datadas a partir da conquista espanhola e expansão colonial europeia, também registraram as línguas Uruquilla e Puquina, faladas originalmente por outros grupos da região. Historicamente é descrito um grupo de pessoas chamadas em aimará de “Uru” ou “*uma jaqe*” (pessoas da água) que falavam Uruquilla e habitavam as margens do Lago Titicaca, exercendo atividades de pesca, caça e coleta.

Até o século XVI, os Uros (ou Urus, como são chamados na Bolívia) estavam distribuídos por um eixo aquático compreendendo o Rio Azangaro, Lago Titicaca, Rio Desaguadero no Peru/Bolívia; Lago Poopó, Rio Lacajahuira e o Lago Coipasa na Bolívia, onde ainda existem falantes do idioma Uru-Chipaya, mas também há alguns registros de que habitaram o litoral de Arica e Iquique (Vellard, 1959; Wachtel, 1986). Desde a conquista espanhola até o presente, houve uma diminuição da população dos Uros (Manelis de Klein, 1973). Outros autores sugerem que, de forma gradual, os Uros foram transculturados pelo sistema incaico, colonial e republicano, muitos deles “aimarizando-se” (Wachtel, 1986).

Sobre os Puquinas, as crônicas indicam que eles ocuparam previamente a área do vasto império pré-inca do Tiwanaku, mas esta língua foi extinta, em contraste ao Uruquilla que ainda é falada pelos Uru-Chipaya (Bouysse-Cassagne, 1991).

Nas ilhas Taquile e Amantani do Lago Titicaca foram achadas evidências arqueológicas e históricas de assentamentos sucessivos desde as épocas pré-incaicas, incaica e colonial. Foram também descobertos restos da cultura Tiwanaku (relacionado ao sítio arqueológico do mesmo nome localizado na Bolívia), e possivelmente também foram habitadas por mitimaes durante o Tawantin Suyu (apogeu do Império Inca, entre 1432 a 1532).

Segundo as crônicas dos espanhóis, a península de Capachica e as ilhas Amantani, Titicaca (Ilha do Sol) e Coati (Ilha da Lua) foram ocupadas pelos Collas, considerados “os melhores tecelões do Império Inca” (Bouysse-Cassagne, 1991). Atualmente, o artesanato têxtil é ainda muito significativo nas ilhas de Taquile e Amantani, as quais, depois de 1650, foram adquiridas pelo comendador espanhol Pedro Gonzáles Taquile, sendo repovoadas com pessoas de lugar indeterminado (Espinoza, 1989). Em algumas crônicas se relatou que Amantani foi evacuada por seu proprietário antes de ser repovoada, mas não há qualquer referência sobre Taquile, embora se presuma que tenha sofrido o mesmo destino. Taquile e Amantani permaneceram relativamente isoladas até 1960, quando se tornaram destinos turísticos temporários, de modo que podemos assumir que estas populações são majoritariamente descendentes do repovoamento ocorrido no século XVII. Por outro lado, na ilha Anápiá-Suana (localizada no Lago Menor ou Wiñaimarca) e nas ilhas vizinhas Suriki, Pariti, Anayutani, bem como Coati e Titicaca, também se encontraram vestígios das culturas Tiwanaku e Inca (Bouysse-Cassagne, 1991). Posteriormente, na era do colonialismo europeu, as ilhas do Lago Wiñaimarca foram utilizadas como fazendas pelos espanhóis e, posteriormente, retornaram às comunidades nativas.

3.3- História da colonização pós-colombiana do Peru

3.3.1- Os fluxos migratórios e a miscigenação no Peru

Após várias explorações no Caribe e na América Central, um grupo de aproximadamente 250 espanhóis sob o comando de Francisco Pizarro chegou à costa peruana em 1531, via Panamá, com a ideia de conquistar novas terras. Estes espanhóis chegaram a Piura, no litoral peruano, de onde partiram para Cajamarca, e ali executaram o Inca Atawalpa. Em 1533, os espanhóis chegaram à cidade de Cusco, capital do império incaico, e, gradualmente, as imigrações europeias aumentaram, levando à ocupação do Peru pelos primeiros colonizadores europeus (Lexus, 1998).

Nota-se que os primeiros espanhóis que chegaram à América durante o século XVI vieram principalmente de Andaluzia, Extremadura e Castilla (Lexus, 1998). Os costumes e tradições desses lugares influenciaram fortemente a sociedade colonial no Peru, com a formação de “categorias” como “criollos” (filhos de espanhóis, nascidos na América).

Os primeiros africanos foram trazidos como escravos pelos espanhóis para a exploração da agricultura na costa do país (por exemplo, Chíncha e Cañete) e como servidão dos latifundiários espanhóis. Os escravos vieram de diferentes etnias africanas como a mandinga, carabali, congos, bantu, etc.

Por outro lado, durante a época colonial se formaram novas “categorias” sociais: os “mestiços”, (filhos de pai espanhol e mãe indígena), que não estavam isentos de impostos, mas conseguiram trabalhar em alguns ofícios como artesões, agricultores ou cronistas (exemplo, Garcilazo de La Vega, escritor de *Los Comentarios Reales*); os “mulatos” (filhos de pai espanhol e mãe escrava africana) que também eram considerados escravos e não tiveram tanta sorte quanto os mestiços; os “zambos” (filhos de pai africano e mãe nativa) foram mais afortunados, pois sendo filhos de mães livres, também eram considerados livres (embora o pai fosse um escravo), e como não eram registrados nas listas de indígenas, não pagavam qualquer tipo de imposto.

De acordo com os registros históricos, na época da colonização espanhola de 1532 até 1821, as miscigenações entre conquistadores e conquistados (ou escravizados) eram comuns. A miscigenação (e em alguns casos o isolamento) entre os diversos grupos foi aumentando gradualmente até formar as classes sociais denominadas como “criollos”, “mestiços” e “indígenas” (tanto andinos quanto amazônicos).

Em 1782, a migração dos espanhóis para o Peru cresceu bastante e, por exemplo, em Cajamarca houve um aumento considerável da miscigenação com aporte de pessoas de procedência espanhola. A província de Hualgayoc (centro mineiro) registrava naquela época uma população de 7.835 espanhóis, 22.299 mestiços e 29.692 indígenas (Lexus, 1998).

As populações miscigenadas peruanas, até o início de século XIX, foram ainda influenciadas por novas imigrações de espanhóis e africanos, além de migrações autóctones (internas). Por exemplo, em cidades coloniais como Lima, se refletiu a estrutura social que caracterizou toda a época da colônia, com poucos europeus acompanhados de muitos criollos, mestiços e indígenas. Além disso, apareceu um fenômeno que mudou a configuração das principais cidades, a imigração de pessoas do interior do país produzindo o crescimento exponencial da população, sobretudo, na área da capital, resultando nas grandes aglomerações urbanas.

Após a independência do Peru em 1821, aumentou-se a imigração, a miscigenação e o intercâmbio cultural, assim como o povoamento de diversas zonas rurais e urbanas da Costa, dos Andes e da Amazônia. As atuais populações peruanas são compostas por uma

mistura étnica de diversos povos, culturas originadas da Europa, África, Ásia e, principalmente, da América (nativa americana).

Em 1849, houve a imigração chinesa (“coolie”) que se deu através de contratos de trabalho (Herrera, 1986). Em 1899, iniciou-se também um processo de imigração japonesa, primeiro através de contratos de trabalho como os chineses, mas depois em pequenas ondas imigratórias de forma espontânea (Lexus, 1998).

Em 1853, chegaram os primeiros grupos de famílias imigrantes alemãs, a fim de colonizar a floresta. Eles foram estabelecidos em Tingo Maria, Tarapoto, Moyobamba, Amazonas, e em seguida, numa segunda onda em 1868, se estabeleceram em Pozuzo (Cerro de Pasco), onde desenvolveram suas atividades ao estilo europeu, mantendo alguns costumes e língua até hoje (Lexus, 1998).

Durante a Primeira Guerra Mundial e início da Segunda, devido à crise econômica na Europa, ocorreu também uma imigração italiana (1918-1938), assim como de outros estrangeiros de diversos países.

No entanto, a partir de 1940, aconteceu uma grande onda migratória interna com a chegada de muitos migrantes andinos à região de Lima, formando-se bairros e, em seguida, assentamentos humanos em torno da capital com grandes povoados novos em Villa El Salvador, Comas, etc., que atualmente são grandes distritos (Lexus, 1998).

Até o ano 1981, segundo o censo feito pelo INEI (Instituto Nacional de Estatística e Informática: <http://www.inei.gob.pe>), os percentuais da distribuição de migrantes andinos para a região de Lima mostra a região de Ancash com a maior contribuição de migrantes, seguida de Junin, Ayacucho, La Libertad, Ica e uma menor proporção de migrantes vindos de Cajamarca, Lambayeque, Piura e de outras regiões do país.

3.4- Diversidade genética e evolução humana

3.4.1- Estudos históricos com o mtDNA humano

Cada célula contém muitas mitocôndrias e centenas de milhares de cópias do genoma mitocondrial. Essas moléculas de mtDNA só são transmitidas através da linhagem materna e suas variações são frequentemente utilizadas em estudos de genética populacional (por exemplo, traçar rotas migratórias), identidade genética, medicina forense, assim como doenças humanas. Como o mtDNA não sofre recombinação e possui uma elevada taxa de mutação quando comparado com o genoma nuclear, essa propriedade é uma vantagem no mapeamento das variações moleculares ou perfis genéticos nas populações.

Devido a sua herança uniparental (materna), o mtDNA acumula variações sequencialmente ao longo do tempo, irradiando linhagens femininas. Portanto, o número de diferenças de

nucleotídeos entre as sequências de mtDNA de dois indivíduos é proporcional ao tempo que compartilharam um ancestral comum do sexo feminino. Assim, esses estudos têm revelado que todas as sequências de mtDNA coalescem em uma única sequência “fundadora” (haplótipo ancestral ou MRCA-most recent common ancestor) que existiu na África entre 290.000 e 140.000 anos atrás (Cann *et al.*, 1987; Vigilant *et al.*, 1991). A partir desta origem, quatro linhagens específicas da África subsaariana (L0, L1, L2, L3) se irradiaram para formar grupos de haplótipos relacionados filogeneticamente, chamado de haplogrupos. Cerca de 70.000 anos atrás, duas linhagens de mtDNA, M e N, surgiram a partir da linhagem L3 no nordeste ou sul da África (Mishmar *et al.*, 2003; Henn *et al.*, 2011) se diversificando em todas as linhagens da Eurásia, Oceania e América.

Nas Américas, os primeiros estudos com RFLPs (Restriction Fragment Length Polymorphism) e sequenciamento de mtDNA identificaram inicialmente quatro matrilineagens fundadoras dos nativos americanos, denominadas A, B, C e D (Torroni *et al.*, 1993). Atualmente, considera-se a existência de cinco haplogrupos (ou linhagens) do mtDNA entre nativos americanos, incluindo o haplogrupo X (Sukernik *et al.*, 1996; Merriwether *et al.*, 1996; Easton *et al.*, 1996; Stone & Stoneking, 1998). As outras populações nativas do mundo foram catalogadas segundo as mutações específicas nos seguintes haplogrupos: L0, L1, L2, L3, L4, L5, L6 africanas; H, V, U, J, T, R, K, W, X, I européias; Q, G, E, D, C, Z; A, B, S, O, Y, F, P asiáticas, segundo o PhyloTree (<http://www.phylotree.org>) (van Oven & Kayser, 2009) (Figura 1). Os haplogrupos mtDNA têm sido classificados em relação à rCRS.

Estudos mais recentes com genomas mitocondriais completos nas populações nativas americanas indicam a existência de sub-linhagens (A2, B2, C1, D1, X2a) derivadas dos haplogrupos principais, que estão distribuídas heterogeneamente entre os diferentes povos nativos de diferentes grupos lingüísticos (Soares *et al.*, 2009; Fagundes *et al.*, 2008; Tamm *et al.*, 2007).

Nas populações peruanas, nossos resultados preliminares (Sandoval *et al.*, 2004) e de outros grupos de pesquisa (Barbieri *et al.*, 2011; Fehren-Schmitz *et al.*, 2010; Kemp *et al.*, 2009; Shinoda *et al.*, 2006; Shimada *et al.*, 2004; Lewis *et al.*, 2004; Fuselli *et al.*, 2003; Rodriguez-Delfin *et al.*, 2001) concordam com a predominância do sub-haplogrupo B2 (de acordo com nomenclatura PhyloTree) entre os povos andinos do Peru, assim como em grande parte dos outros povos andinos sul-americanos (Merriwether *et al.* 1996, Bisso-Machado *et al.*, 2012). Por exemplo, nas populações do Lago Titicaca se observa a mais alta frequência dos haplótipos B2 (B1 na antiga nomenclatura RFLP) do que em qualquer estudo prévio de populações nativas da América do Sul: Taquile (100%), Amantani (88,6%), Anapia (87,5%) e nos Uros (75%).

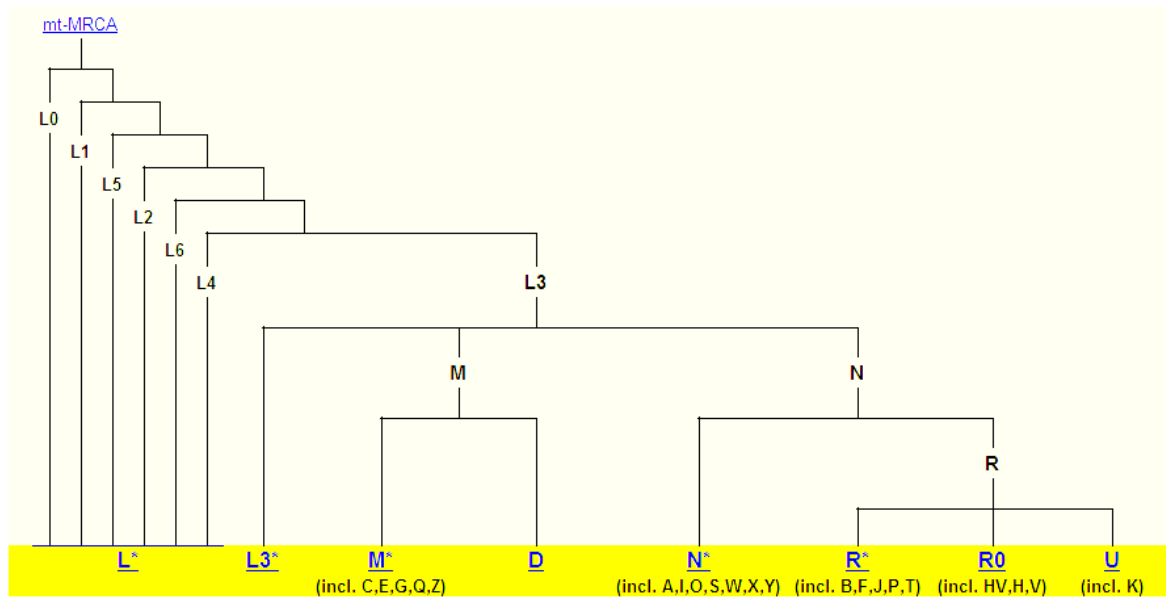


Figura 1: Árvore das principais linhagens maternas segundo o PhyloTree (<http://www.phylotree.org>) (van Oven & Kayser, 2009), baseada nas sequências completas de mtDNA.

3.4.2- Estudos históricos com o cromossomo Y humano

O cromossomo Y é uma parte do genoma nuclear masculino, virtualmente haplóide e transmitido pela linhagem paterna (do pai ao filho), em contraste ao mtDNA (da mãe aos filhos e filhas). As variações nucleotídicas na região haplóide e não recombinante (NRY) do cromossomo Y têm sido utilizadas em estudos populacionais, evolutivos, testes de identidade genética e paternidade, da mesma forma do que o mtDNA (Jobling & Tyler-Smith, 2003).

Os estudos das linhagens paternas têm mostrado também que todos haplótipos das diferentes populações do mundo coalescem em um ancestral comum na África, ao redor de 120.000 anos atrás, complementando assim a história materna, mostrando o mesmo cenário sobre a origem inicial e diversificação das populações humanas atuais pelos diversos continentes (Shi *et al.*, 2010; Karafet *et al.*, 2008; YCC, 2002; Macpherson *et al.*, 2004; Jobling & Tyler-Smith, 2003; Walsh, 2001; Underhill *et al.*, 2000).

O Consórcio Internacional do cromossomo Y (YCC, 2002) estabeleceu a nomenclatura dos haplótipos e classificou inicialmente as linhagens masculinas em 20 haplogrupos, de A até T, além de inúmeros sub-haplogrupos e haplótipos definidos por mutações do tipo SNPs (Figura 2). As linhagens A e B são exclusivas da África; a linhagem E é muito frequente na África, principalmente em populações Bantu (sub-haplogrupo E1b1a), mas ocorre também em partes da Europa e Ásia; R é predominante na Europa; F, G, I, J, M, N são encontradas

tipicamente na Eurásia; as linhagens C, D, H, K, O, P e Q ocorrem com maior frequência na Ásia; S em Nova Guiné, L e T na África e sudeste Asiático (Chiaroni *et al.*, 2009). As linhagens nativas americanas são derivadas dos haplogrupos asiáticos Q e C. No entanto, o principal sub-haplogrupo autóctone nativo americano Q-M3 (ou Q1a3a) (Underhill *et al.*, 1996; YCC, 2002) é encontrado em altíssima frequência em todos os povos nativos americanos analisados, de todos os grupos linguísticos, fenômeno provavelmente resultado de um efeito fundador paterno durante o povoamento pré-colombiano do continente (Pena *et al.*, 1995; Santos *et al.*, 1995a).

Os primeiros estudos detalhados sobre o povoamento inicial da América com dados de linhagens paternas foram feitos por nosso grupo (Pena *et al.*, 1995; Santos *et al.*, 1995a), que indicavam que mais de 70% dos cromossomos Y se encaixavam em um haplótipo II-A ($\alpha_{HII} + DYS19-A$), hoje pertencente ao haplogrupo Q. Isto indicava um forte efeito fundador, relativo às linhagens paternas, no povoamento pré-colombiano das Américas. Posteriormente, o grupo de Cavalli-Sforza (Underhill *et al.*, 1996) identificou o mesmo efeito fundador em nativos americanos, utilizando um SNP no *locus* DYS199 (M3), hoje um marcador característico do sub-haplogrupo Q-M3 (YCC, 2002).

Um estudo detalhado de nosso grupo (Santos *et al.*, 1999a) com populações nativas americanas e asiáticas verificou a existência de linhagens relacionadas ao haplogrupo Q-M3 entre os povos Ketí e Altaianos, na região siberiana entre o Rio Ienissei e montes Altai. Estas linhagens siberianas são consideradas hoje como pertencentes ao haplogrupo Q (YCC, 2002), estreitamente relacionadas às linhagens nativas americanas.

A existência em alta frequência dos haplogrupos Q-M3 e Q* (ou Q-M346*) entre todos os povos nativos americanos indica uma origem majoritariamente derivada de uma fonte populacional principal, provavelmente beringiana. Esta visão foi, posteriormente, reafirmada em outros estudos e, atualmente, pode ser explicada por uma origem a partir de uma população beringiana ao redor de 18 mil anos atrás, que, durante o final de Pleistoceno e o Holoceno manteve algum fluxo gênico com a Ásia, resultando, por exemplo, na ocorrência de outras linhagens paternas nativas americanas (do haplogrupo C e outras Q-M346*) entre os nativos americanos atuais, além de línguas relacionadas entre América e Ásia como as dos Inuits e Na-Dene (González-José *et al.*, 2008).

Atualmente, a inferência dos haplogrupos principais pode também ser feita a partir da genotipagem de STRs (ou microssatélites) do cromossomo Y com o uso de algoritmos de classificação (Schlecht *et al.*, 2008). Esta inferência pode ser feita a partir de STRs amplificados com o kit Amp/STR Yfiler de Applied Biosystems (Mulero *et al.*, 2006). O kit consiste do sistema “*European minimal haplotype*” (DYS19, DYS385a/b, DYS389a/b, DYS390, DYS391, DYS392 e DYS393), do Scientific Working Group-DNA Analysis Methods (SWGAM) – “*European minimal haplotype*” mais DYS438 e DYS439, assim como *loci*

adicionais (DYS437, DYS448, DYS456, DYS458, DYS635 e GATAH4). Estes 17 STRs estão distribuídos ao longo da região não recombinante do cromossomo Y, sendo que apenas três não são tetranucleotídeos: o DYS392 que é de unidade trinucleotídica, DYS438 –pentanucleotídica, e DYS448 –hexanucleotídica. Existe uma grande quantidade de dados do Y *filer* em inúmeras populações de todo o mundo, incluindo vários bancos de dados forenses, sendo que muitas das amostras também foram genotipadas com SNPs do cromossomo Y. Além de permitir a inferência do haplogrupo com certo grau de acurácia (Schlecht *et al.*, 2008), os STRs do cromossomo Y podem ser associados a SNPs para estudos filogeográficos, facilitando a análise de estrutura populacional e a identificação de relações filogenéticas e rotas migratórias (Karafet *et al.*, 2008; Bortolini *et al.*, 2003; Tarazona-Santos *et al.*, 2001; Zerjal *et al.*, 1997).

Nos estudos genéticos de populações do Peru, o uso dos marcadores Y-STRs e STRs autossômicos foram dedicados a fins forenses (Builes *et al.*, 2006; Iannacone *et al.*, 2005; Tito *et al.*, 2004). Entretanto, alguns outros estudos enfocaram na história das populações, incluindo as pré-colombianas (Baca *et al.*, 2012; Fehren-Schmitz *et al.*, 2010). Um estudo prévio com SNPs do Y em populações peruanas do noroeste do país (Carbajal-Caballero *et al.*, 2005) mostrou que o variante DYS199-T (Q-M3) é frequente na localidade Yamayakat (96,7%), do grupo nativo Awajun, localizado na Amazônia, o mesmo observado em Santiago de Chuco (53,3%), área andina do departamento de La Libertad, e Lambayeque (73,3%), localizada na Costa. Em contraste, na população de Trujillo (cidade na Costa) foi detectada a menor porcentagem (33,3%) de Q-M3, indicando uma elevada miscigenação com linhagens não autóctones, principalmente espanhóis, em concordância com os registros históricos (Lexus, 1998; Hunefeldt, 2004). Em outro estudo anterior numa localidade da região andina (n=25, de Cerro de Pasco) também foram encontrados 48% de indivíduos Q-M3 (Rodriguez-Delfin *et al.*, 2001).

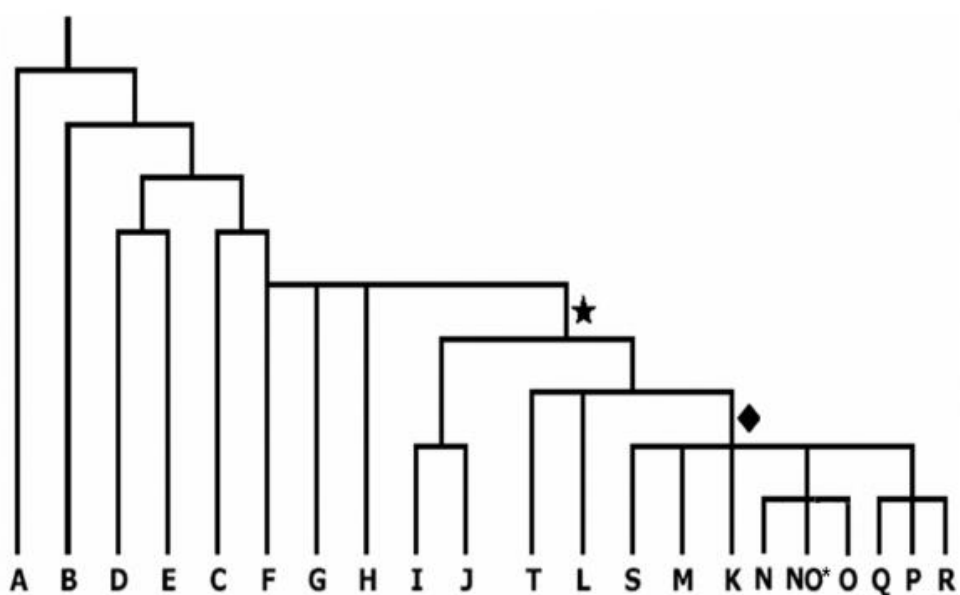


Figura 2: Árvore das principais linhagens paternas baseado nos SNPs do cromossomo Y, segundo Chiaroni *et al.* (2009).

3.4.3- Estudos com marcadores autossômicos do tipo INDELS

Diferentemente dos estudos com marcadores uniparentais, como mtDNA e cromossomo Y, os marcadores autossômicos têm natureza diploide e se distribuem por vários cromossomos, além estarem sujeitos ao processo de recombinação. Portanto, comportam-se como marcadores mendelianos típicos, de herança biparental. Com o avanço do Projeto Genoma Humano, vários marcadores autossômicos se tornaram disponíveis, inicialmente com os microssatélites ou STRs, atualmente muito utilizados em genética forense, e mais recentemente marcadores bialélicos como os SNPs e INDELS (inserção-deleção) que são muito úteis nos estudos evolutivos, demográficos e análises de estruturação populacional, fluxo-gênico e miscigenação.

Os INDELS são blocos de sequências de DNA que podem envolver um ou alguns milhares de nucleotídeos e se distribuem ao longo do todo o genoma humano. Os INDELS envolvendo regiões pequenas do DNA são menos susceptíveis à influência da recombinação do que os INDELS maiores, por isto são mais apropriados para estudos de estruturação populacional, testes de paternidade ou ancestralidade, por exemplo, como foi feito com a população do Brasil (Pimenta & Pena, 2010; Pena *et al.*, 2011; Santos *et al.*, 2010). Os INDELS funcionam muito bem como marcadores informativos de ancestralidade ou AIMs (Ancestry Informative Markers), que são definidos como marcadores que

apresentam elevada heterogeneidade na distribuição alélica entre as populações, permitindo fazer associação com a origem geográfica das mesmas (Bastos-Rodrigues *et al.*, 2006). Por esta razão, o uso dos INDELS como AIMs permite determinar a proporção de genoma herdado de seus ancestrais continentais (de povos autóctones de cada continente) ou o nível de contribuição genética de cada um deles nas miscigenações (Pena *et al.*, 2011; Santos *et al.*, 2010).

Com o uso de 40 INDELS bialélicos encontrou-se que as populações “originárias” da América são muito mais próximas geneticamente às do leste asiático do que as da Europa, e mais distantes das populações da África (Bastos-Rodrigues *et al.* 2006). Neste estudo, também observaram agrupamentos das populações humanas autóctones do painel de células imortalizadas de diferentes etnias, o HGDP-CEPH, em sete regiões do mundo, o mesmo resultado observado por Li *et al.* (2008) ao analisar 642.690 SNPs autossômicos em 938 indivíduos do mesmo painel.

3.5- Hipóteses e justificativas do estudo com populações peruanas

A população do Peru é composta atualmente por um dos maiores percentuais de indivíduos “autóctones” do continente americano (censo 2007, <http://www.inei.gob.pe>). Além disto, como revelado por sua história escrita (dos últimos 500 anos), também deve apresentar uma grande contribuição genética nativa americana na população urbana e rural, não “autóctone” ou “mestiça”. De acordo com os registros históricos do Peru, desde 1532, aconteceu uma crescente “miscigenação” com os imigrantes pós-colombianos da Eurásia e África, cuja ascendência genômica deve também estar presente nos atuais habitantes peruanos.

Através de diferentes marcadores de DNA vamos tentar recontar essa história ainda pouco conhecida sobre a herança genética dos peruanos. Da mesma forma, pretendemos caracterizar as relações interpopulacionais dos povos peruanos, com foco nos movimentos migratórios pré e pós-colombianos, comparando com os poucos registros históricos e dados interdisciplinares disponíveis.

Nosso interesse neste contexto é traçar os padrões de migração e os níveis de miscigenação dos grupos humanos que formam as populações rurais e urbanas do Peru. Por exemplo, nos perguntamos: *i)* qual será a contribuição das linhagens maternas e paternas nos atuais povoadores de Cajamarca, que tiveram uma massiva imigração de espanhóis?; *ii)* quais os perfis de ancestralidade e miscigenação dos atuais habitantes urbanos e camponeses dos Andes e da Amazônia peruana e sua relação com os povos nativos de cada continente?; *iii)* Como estão relacionadas geneticamente as populações insulares do Titicaca com as demais populações autóctones da região?; *iv)* quais as

relações interpopulacionais entre habitantes da região do Lago Titicaca, reconhecidamente uma região com marcada herança nativa americana?

Para tal estudo, foram selecionados marcadores uniparentais (mtDNA e Y) assim como 40 INDELS autossômicos distribuídos ao longo de 20 cromossomos (dos 22 pares), para a genotipagem em indivíduos amostrados de populações peruanas. Essa abordagem permitiu uma estimativa dos níveis de miscigenação genômica dos grupos populacionais peruanos e a identificação da herança de linhagens autóctones americanas, euro-asiáticas (Europa, Oriente Médio, Ásia Central, Leste Asiático, Oceania) e africanas na formação das atuais populações da Costa, dos Andes e da Amazônia do Peru.

4 – OBJETIVOS

4.1- Objetivo geral

Estudar a ancestralidade, miscigenação e estrutura genética das populações peruanas atuais através de marcadores genéticos autossômicos e uniparentais, confrontando com a história do povoamento pré e pós-colombiano do país.

4.2- Objetivos específicos

- Avaliar ancestralidade das linhagens paternas (Y) e maternas (mtDNA) nas amostras das populações peruanas através de genotipagem e sequenciamento;
- Estimar as proporções de ancestralidade genômica autóctone americana e não autóctone nas populações peruanas através dos 40 INDELS em comparação aos genótipos das populações do painel HGDP-CEPH;
- Utilizar análises refinadas de genética populacional, atribuição demográfica (miscigenação) e filogeografia para teste de hipóteses históricas (pré e pós-colombiana) sobre a variabilidade, diferenciação e dinâmica populacional dos povos do Peru, comparando com dados publicados para diferentes continentes e o painel HGDP-CEPH.
- Desvendar os componentes de ancestralidade autóctone de dinâmica populacional na região do Lago Titicaca, confrontando diferentes hipóteses sobre origens das comunidades locais.

5 - MATERIAL E MÉTODOS

5.1- Amostragem populacional

De acordo com protocolos aprovados pelo Comitê de Ética da Universidad San Martin de Porres (USMP, Lima, Peru), após a obtenção do consentimento informado foram obtidas amostras de sangue ou raspado bucal dos indivíduos de diferentes localidades urbanas e rurais da Costa, Andes e Amazônia do Peru. O critério de inclusão era ser nascido na localidade amostrada, de preferência, sem parentesco de primeiro grau com outros indivíduos amostrados no mesmo local.

Foram amostrados 721 indivíduos voluntários, tanto homens como mulheres, das diferentes regiões do Peru (Localidade_DEPARTAMENTO=N): 137 peruanos da Amazônia (Andoas_LO=75; Iquitos_LO=7; Pucallpa_UC=10; Chachapoyas_AM=15, Tarapoto_SM=7; Lamas_SM=23); 460 peruanos dos Andes (Cajamarca_CA=40; San Marcos_CA=24; Huaraz_AN=9; Chogo_AN=17; Ocopon_AN=14; Huarochiri_LI=23; Huancayo_JU=33; Huancavelica_HV=5; Ayacucho_AY=32; Andahuaylas_AP=43; Kaquiabamba_AP=9; Arequipa_AR=19 (Characato e Mollebaya); Chivay_AR=20; Yanque_AR=11; Cabacaconde_AR=18; Taquile_PU=51; Amantani_PU=34; Uros_PU=28; Anapia_PU=24; Puno_PU=6); e 124 peruanos da Costa (Piura_PI=5; Lima_LI=75; Lambayeque_LA=44). Incluímos também um indivíduo da Alemanha como amostra de referência europeia, que corresponde ao haplogrupo H do mtDNA. Os códigos dos Departamentos do Peru aos quais correspondem as localidades são: LO=Loreto, UC=Ucayali, AM=Amazonas, SM=San Martin, CA=Cajamarca, AN=Ancash, JU=Junín, AY=Ayacucho, AP=Apurimac, AR=Arequipa, PU=Puno, PI=Piura, LA=Lambayeque, LI=Lima.

Em alguns gráficos e análises, as localidades foram etiquetadas com as seguintes abreviações: AY (Ayacucho), Hyo (Huancayo), HV (Huancavelica), Cb (Cabanaconde), Cy (Chivay), Yke (Yanque), Char (Characato), Mll (Mollebaya), AR (Arequipa), ANh (Huaraz), Oco (Ocopon), Ch (Chogo), CA (Cajamarca), CAsm (San Marcos), Ahy (Andahuaylas), Kaq (Kaquiabamba), Llhr (Huarochiri), Ur (Uros), Ap (Anapia), Amt (Amantani), Taq (Taquile), PU (Puno), And (Andoas), Iq (Iquitos), Puc (Pucallpa), Chp (Chachapoyas), SMLa (Lamas), Tar (Tarapoto), PI (Piura), LA (Lambayeque) e LI (Lima) (Figura 3).

No procedimento geral de coleta realizado por uma pessoa treinada, 3 ml de sangue foram obtidos por punção venosa. O sangue foi preservado em solução com EDTA (anticoagulante) e logo as amostras foram armazenadas em refrigeração até serem levadas ao Laboratório do Centro de Genética e Biologia Molecular (CGBM), USMP, Lima. Em algumas coletas, amostras foram obtidas de células de raspado bucal através de swabs (escovinhas bucais). Estas amostras representam a população peruana urbana e rural dos Andes, Costa e Amazônia, coletadas ao longo de 10 anos.

A extração do DNA foi por executada por lavagem com Tris-EDTA, lise com N-lauryl-sarcosine, digestão com proteinase K, precipitação e extração com álcool e sal, de acordo com a rotina laboratorial do IGBM-USMP (Peru). O DNA de algumas amostras coletadas entre 2008 e 2010 foi extraído segundo os protocolos padronizados (www.icb.ufmg.br/labs/lbem/protocolos) no LBEM/ICB-UFMG (Brasil), e preservadas em um freezer. A quantificação do DNA foi realizada através do espectrofotômetro NanoDrop-1000.

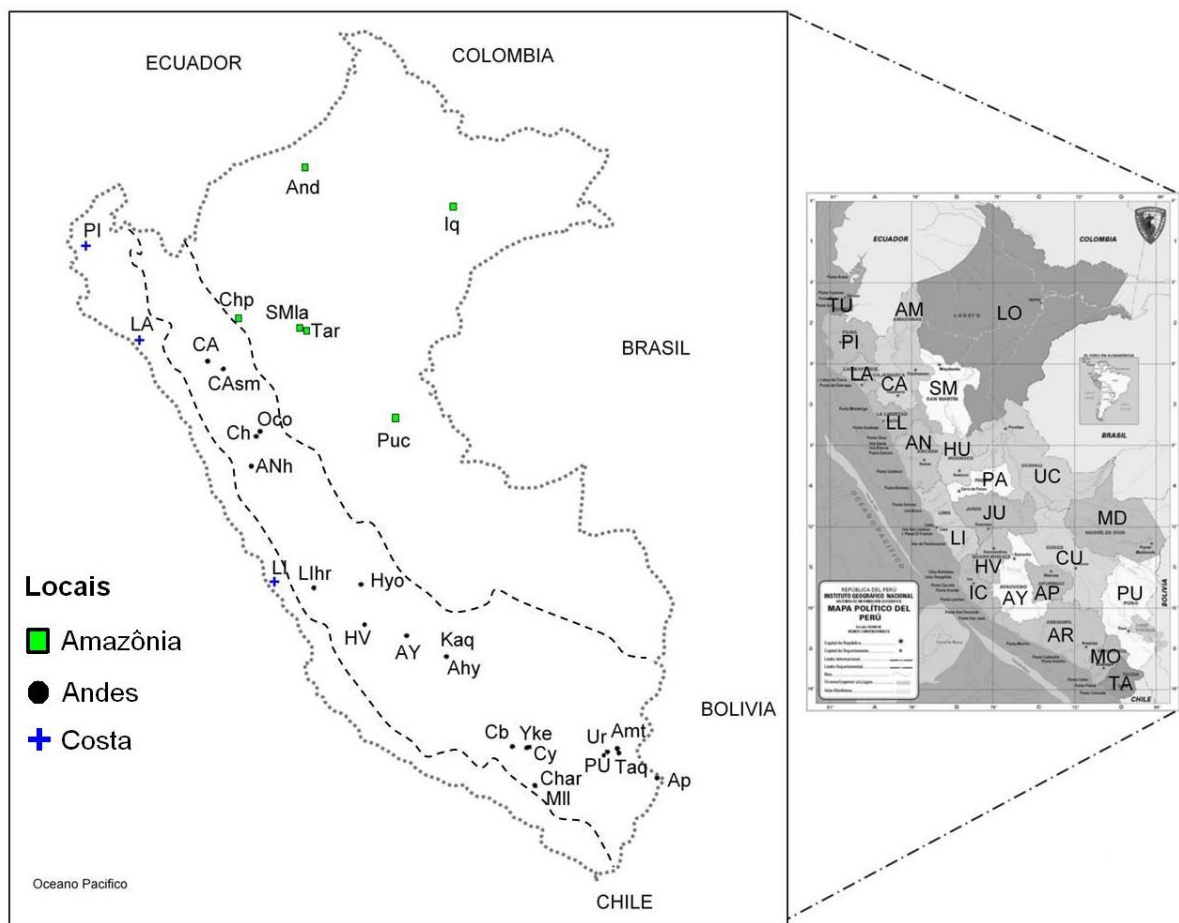


Figura 3: Localização aproximada das cidades ou comunidades, de onde procedem as amostras, no mapa do Peru.

5.2- Análises dos marcadores de DNA

5.2.1- Região controle do mtDNA

Para a amplificação da Região Controle (inclui as posições 15907-580 do *Cambridge Reference Sequence- CRS*), que inclui HVS-I (342 pb, posição 16024-16365), domínio central conservado (276 pb, a posição 16366-72), HVS-II (268 pb, a posição 73-340) e HVS-

III (137 pb, a posição 438-574); se procedeu segundo os protocolos pré-estabelecidos no LBEM para o Projeto Genográfico, utilizando os seguintes reagentes com concentração final: 1X buffer, 3 mM MgCl₂, 200 uM dNTPs, 0,0375 uM de cada dos primers 15876F/639R e 0,2 u/tubo da Taq DNA Pol Platinum, 1-5 ng DNA genómico, em volume final de 20 ul. A PCR foi efetuada através do termociclador GeneAmp PCR System 9700 AB. Após realizar a corrida eletroforética em gel de agarose com Brometo de Etídio e verificar o segmento DNA amplificado, foi feita a “purificação” dos fragmentos seguindo o protocolo PEG (polietilenoglicol)-NaCl.

Foi utilizado o kit BigDye Terminator v3.1 Cycle Sequencing (Applied Biosystems) para a reação de sequenciamento bidirecional, seguindo as instruções e condições do fabricante (Applied Biosystems) com os primers padrão 15946-Forward/132-Reverse e 16436-Forward/637-Reverse que abarca a Região Controle do mtDNA. Assim, a reação foi realizada no termociclador GeneAmp PCR System 9700 AB, com os seguintes parâmetros de temperatura: 96°C-10 seg. de desnaturação; 50°C-5 seg. de anelamento; 60°C-4 min. de extensão e por 26 ciclos. O passo seguinte foi purificar e precipitar os fragmentos DNA, com acetato de amônio e vários lavados com álcool absoluto gelado (99% e 70%) e centrifugações a 4000 rpm, segundo o protocolo padronizado do LBEM, até ser secado e ressuspendido em 10 ul de Hi-Di Formamida (Applied Biosystems). Finalmente, o sequenciamento dos fragmentos mitocondriais foi executado no sequenciador automatizado ABI 3130xl/Genetic Analyzer (Applied Biosystems) de 16 capilares.

Além das sequências de região controle, a região codificadora do genoma mtDNA de algumas amostras peruanas (réplicas) foi sequenciada independentemente pelo grupo colaborador (Douglas Wallace, MAMMAG; University of California, Irvine-USA). Esses dados serviram por um lado para controle e comparações dos dados de sequenciamento e por outro para as simulações bayesianas através de BEAST.

Em alguns estudos preliminares com aproximadamente 150 amostras desta coleção de peruanos, os haplogrupos de mtDNA foram determinados utilizando-se o sistema PCR-RFLP com enzimas de restrição: sítio de corte na posição 663 por *HaeIII* assinala o haplogrupo A; a deleção de 9 pb na posição 8272-8280, entre COII e tRNA^{Lys}, indica o haplogrupo B; a ausência do sítio de corte na posição 13259 para *HincII*, o haplogrupo C; e sítio de corte na posição 5176 para *AluI*, corresponde ao haplogrupo D (Sandoval *et al.*, 2004).

Sequências dos iniciadores de mtDNA:

16436F_HVRhum: 5'-gtg cta ctc tcc tcg ctc-3'

637R_HVRhum: 5'-gtg atg tga gcc cgt cta-3'

15946F_HVRhum: 5'-caa gga caa atc aga gaa aa-3'

132R_HVRhum: 5'-gac aga tac tgc gac ata gg-3'

5.2.1.1- Haplotipagem de mtDNA

5.2.1.2- Predição de linhagens de mtDNA

O programa *Haplogroup Prediction Tool* (HPT) foi desenvolvido pelo Projeto Genográfico e está baseado no algoritmo denominado *Nearest Neighboring-End* (<http://nnhgtool.nationalgeographic.com/classify>). Os SNPs ou mutações específicas das sequências mtDNA do segmento HVS-I (desde a posição 16023 até 16569) que mostram diferenças do *Cambridge Reference Sequence* (CRS) são comparadas à base de dados do Projeto Genográfico, que classifica sugestivamente essas sequências em seus haplogrupos correspondentes segundo a nomenclatura padrão (Wolfgang *et al.*, 2010). No entanto, usamos também os sistemas de predição MitoTool (Fan & Yao, 2011) e mtPhyl (<http://eltsov.org>) para fazer as respectivas comparações e conferir a acurácia da predição dos haplogrupos. Se os três sistemas concordaram na predição do mesmo haplogrupo, este foi considerado como “aceito”. Posteriormente, os agrupamentos dos haplogrupos foram conferidos através do algoritmo filogenético de máxima verossimilhança no programa PhyML 3.0 (<http://www.atgc-montpellier.fr/phyml/>) ou outros programas de reconstrução de árvores filogenéticas.

5.2.1.3- Análises de sequências da Região Controle do mtDNA

O sequenciamento dos fragmentos de mtDNA foram padronizados no ABI3130XL (Applied Biosystems), cujos resultados são analisados no programa Sequencing Analysis v5.3.1 (Applied Biosystems) que tem algoritmos para converter os picos eletroforéticos fluorescentes DNA em leitura de nucleotídeos. Com ajuda do programa SeqScape v2.6 (Applied Biosystems) mostrado na Figura 4 foi realizado o alinhamento múltiplo das sequências de mtDNA, tendo como sequência de referência a rCRS (Anderson *et al.*, 1981). Após conferir os SNPs de cada sequência de nucleotídeos, assim como os cromatogramas pelo programa Chromas Lite, as sequências foram salvas no formato FASTA para gerar outros alinhamentos com sequências de bancos de dados ou publicações, através dos algoritmos Muscle ou ClustalW no programa MEGA5 (Tamura *et al.*, 2011).

As análises de variabilidade genética entre e dentro das populações, testes de neutralidade, entre outras, foram feitas com os programas Arlequin 3.5 (Excoffier & Lischer, 2010) e DnaSP (Librado & Rozas, 2009). As relações filogenéticas foram reconstruídas através dos programas Network 4.6 (www.fluxus-engineering.com), PhyML 3.0 (<http://www.atgc-montpellier.fr/phyml/>), mtPHYL (<http://eltsov.org>), BEAST (Drummond & Rambaut, 2007), TCS (Clement *et al.*, 2000), entre outros. As análises exploratórias sobre as relações entre grupos de populações foram feitas usando diferentes métodos de análise multivariada tal

como PCA (componentes principais) com os programas ade4 e FactoMineR (<http://www.r-project.org>), baseando-se nas frequências dos haplogrupos/haplótipos ou SNPs.

Para a visualização das relações interpopulacionais no escalonamento multidimensional (MDS) foram usadas as distâncias genéticas linearizadas de Reynolds, geradas no programa Arlequin. Para caracterizar relações de similaridade e/ou parentesco entre as populações e avaliar correlações dos haplogrupos com a latitude foi usado o programa GenGIS (Parks *et al.*, 2009). O isolamento por distância foi testado através do software GenAIEx (Peakall & Smouse, 2006).



Figura 4: Análises do sequenciamento do DNA mitocondrial.

5.2.2- Genotipagem do cromossomo Y

5.2.2.1- STRs com o sistema AmpFISTR Y Filer

A genotipagem de 17 microssatélites (STRs) da região não recombinante do cromossomo Y foi executada de acordo com o kit AmpFISTR-YFiler-PCR Amplification Kit (Mulero *et al.*, 2006) em um sistema PCR multiplex no volume total de 6,25 ul por amostra, que consiste de 2,3 ul de *reaction mix*, 1,25 ul de *primer set*, 0,2 ul de AmpliTaq Gold e completado com 1,5 ul de H₂O milliQ. Na genotipagem incluem-se dois controles (um positivo e outro negativo). A reação de PCR foi feita nas condições de 1 min. para a desnaturação (94°C), anelamento (61°C) e extensão (72°C) por 30 ciclos. Afinal com 80 min. adicionais a 60°C.

Na corrida eletroforética feita no capilar do sequenciador é utilizado o GeneScan-500 LIZ Standard AB como padrão de medida dos tamanhos de fragmentos de DNA, neste caso 16 fragmentos de fita simples de DNA, entre 35 e 500 nucleotídeos.

O preparo da reação para uma placa (96 poços) consistiu de $0,25 \times 100 = 25$ ul de GeneScan e de $9,25 \times 100 = 925$ ul de Hi-Di formamide, sendo que em cada reação foi utilizado 9,5 ul desta mistura com 1 ul de DNA amplificado para a corrida. Foi utilizado também o AmpFISTR Allelic Ladder como padrão molecular de tamanho alélico dos STRs, e um controle positivo com alelos referenciais conhecidos para cada um dos seguintes marcadores STRs: DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, GATA_H4, DYS437, DYS438, DYS448.

As corridas eletroforéticas foram executadas no sequenciador ABI 3130xl/Genetic Analyzer (Applied Biosystems), determinando-se os tamanhos dos alelos segundo os seus espectros de fluorescência (forma de picos em grupos de quatro cores: azul, verde, amarelo e vermelho) segundo os parâmetros pré-estabelecidos, utilizando o programa GeneMapper ID v3.2 para as análises correspondentes (Figura 5).

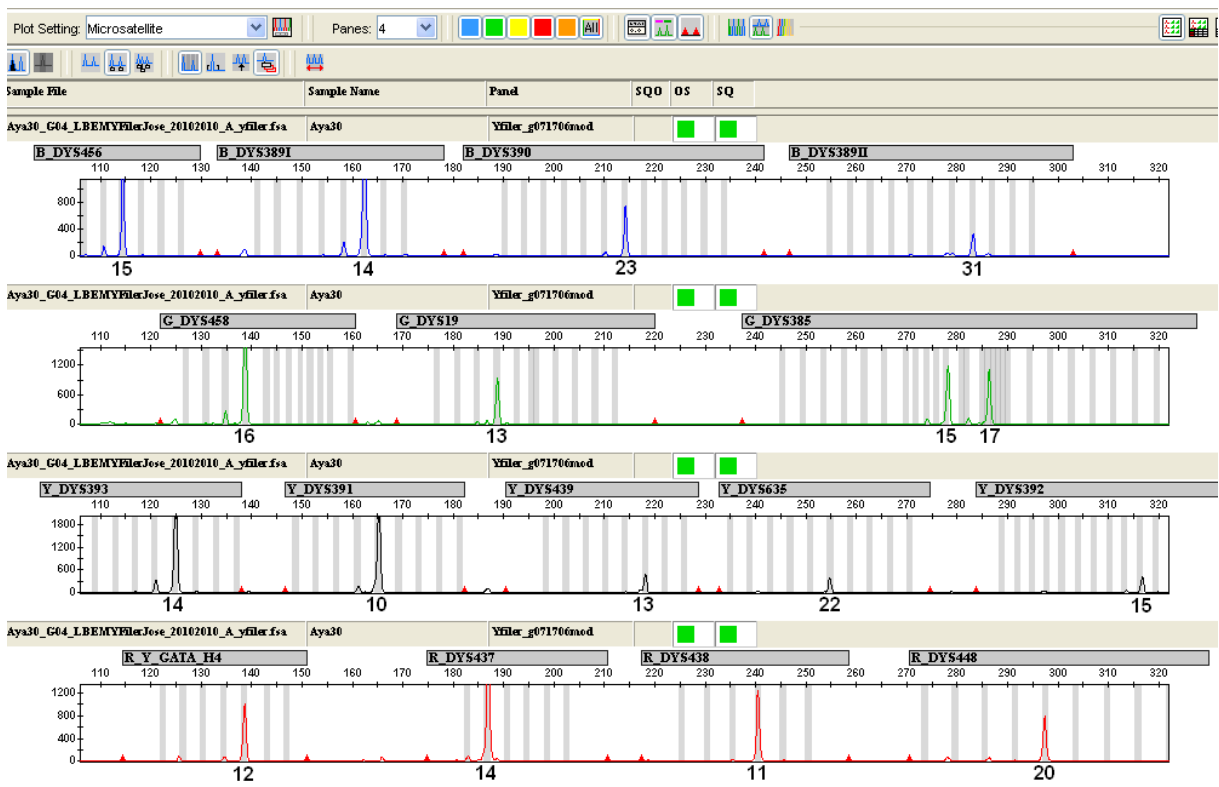


Figura 5: Genotipagem dos 17 Y-STRs (amostra de Ayacucho) e visualização através de GeneMapper v.3.2. Os alelos para cada marcador são mostrados em forma de picos que correspondem ao número de repetições, que está anotado abaixo de cada pico.

5.2.2.2- Detecção de SNPs através de RT-PCR

Para determinar as linhagens paternas das populações peruanas foi realizada inicialmente a detecção dos SNPs usando os marcadores correspondentes aos haplogrupos e alguns sub-haplogrupos (Figura A1, Anexo) segundo a classificação de *Y Chromosome Phylogenetic Tree* (YCC, 2002):

Hg Q: M242, M346, M3

Hg P: M45

Hg R: M207, M124 (sublinhagem R-2)

Hg C: M130 (ou RPS4Y711)

Hg D: M174

Hg E: M96

Hg F*: M89

Além do marcador M3, específico de nativos americanos, outro marcador, o SA01 (Jota *et al.*, 2011) que ocorre em alguns indivíduos Q-M3 dos Andes, foi também utilizado. Os iniciadores para genotipagem destes SNPs pelo sistema TaqMan foram desenvolvidos por Janet Ziegler (Applied Biosystems) através do software *Primer Express* (Applied Biosystems) assim como os iniciadores (*Forward* e *Reverse*).

Para realizar o RT-PCR (Real Time PCR) o protocolo TaqMan consiste em usar Genotyping Universal PCR Master Mix (2x), TaqMan Genotyping Assay Mix (20x) e a amostra DNA a ser testada com os respectivos iniciadores/sondas. Para um volume de 2,75 ul, a reação consiste de 2,5 ul de TaqMan Genotyping Master Mix (2x) e 0,25 ul de TaqMan Genotyping Assay Mix (20x), sendo pipetadas 2,75 ul em cada poço. Um volume de 2,25 ul da amostra DNA (1-10 ng) é adicionado à reação. A reação de PCR foi feita nas condições de 15 segundos para a desnaturação (95°C), 1,5 minutos para o anelamento e extensão (60°C) por 50 ciclos no equipamento RT-PCR 7900HT. As duas sondas marcadas (com fluorescências FAM e VIC) em seus extremos 5' são utilizadas para o anelamento ao sítio no DNA alvo a partir da amplificação do segmento que inclui o SNP de interesse. A leitura da genotipagem foi feita pelo software AutoCaller de gerenciamento do RT-PCR 7900HT (Applied Biosystems).

5.2.2.3- Análises dos dados de SNPs e STRs do cromossomo Y

Para a discriminação alélica dos SNPs foram utilizados os softwares SDS e AutoCaller v1.1. O processo da detecção específica do SNP é a seguinte: a) Se a sonda pareia com o DNA alvo, então significa complementaridade ao SNP, b) Se durante a RT-PCR no estágio da polimerização do DNA (pela AmpliTaq Gold Pol) há sinal da fluorescência reportada (ruptura da FAM ou VIC), então amplifica o segmento DNA com o SNP. O produto se acumula

exponencialmente ao mesmo tempo em que é monitorada por uma câmera CCD e um programa de análise de discriminação bialélica. Para a análise dos haplótipos de STR em relação ao seu haplogrupo correspondente, é feita uma rede filogeográfica usando os métodos *Median Joining* e Máxima Parcimônia no programa Network 4.6.

Para as análises da variabilidade e diferenciação entre populações foi utilizado o programa Arlequin 3.5., e para testar o isolamento por distância através do programa GenAlEx. Para as análises espaciais das frequências das linhagens masculinas (Y-SNPs) ou Y-STRs, através dos programas GenGIS, ade4, FactoMineR.

5.2.3- Painel de 40 marcadores autossômicos de tipo INDEL

Os 40 marcadores INDELS denominados MID# são os seguintes: MID-1 (rs3917), MID-15 (rs4181), MID-17 (rs4183), MID-51 (rs16343), MID-89 (rs16381), MID-107 (rs16394), MID-131 (rs16415), MID-132 (rs16416), MID-150 (rs16430), MID-159 (rs16438), MID-170 (rs16448), MID-258 (rs16695), MID-278 (rs16715), MID-420 (rs140709), MID-444 (rs140733), MID-468 (rs140757), MID-470 (rs140759), MID-663 (rs1305047), MID-788 (rs1610874), MID-857 (rs1610942), MID-914 (rs1610997), MID-918 (rs1611001), MID-1002 (rs1611084), MID-1092 (rs2067180), MID-1100 (rs2067188), MID-1129 (rs2067217), MID-1291 (rs2067373), MID-1352 (rs2307548), MID-1428 (rs2307624), MID-1537 (rs2307733), MID-1549 (rs2307745), MID-1586 (rs2307782), MID-1642 (rs2307838), MID-1654 (rs2307850), MID-1759 (rs2307955), MID-1763 (rs2307959), MID-1847 (rs2308043), MID-1861 (rs2308057), MID-1943 (rs2308135), MID-1952 (rs2308144). As sequências nucleotídicas para cada polimorfismo estão detalhadas no sítio de Marschfield Clinic, USA (<http://research.marshfieldclinic.org/genetics/home/index.asp>). Nas análises como marco de referência se usou o painel do HGDP-CEPH (<http://www.ceph.fr/HGDP-CEPH-Panel>) que consiste de amostras de 52 populações (n=1066) de 7 regiões geográficas: África, Oriente Médio, Europa, Ásia Central, Leste Asiático, Oceania e América (Bastos-Rodrigues *et al.*, 2006).

As reações multiplex-PCR e os iniciadores (*primers*) usados foram processados seguindo os protocolos pré-estabelecidos (Pimenta & Pena, 2010; Bastos-Rodrigues *et al.*, 2006), mas a determinação dos tamanhos dos alelos foi executada através do ABI 3130xl/Genetic Analyzer (Applied Biosystems) acrescentando soluções de Hi-Di Formamide/GeneScan-500-LIZ size standard aos amplicones e utilizando o programa GeneMapper ID v3.2. (Applied Biosystems) para a visualização dos eletroforegramas (Figura 6).

As localidades peruanas foram etiquetadas com os seguintes códigos:

1=AY (Ayacucho), 2=Hyo (Huancayo), 3=Cb (Cabanaconde), 4=Cy (Chivay), 5=Yke (Yanque), 6=Char (Characato), 7=Mll (Mollebaya), 8=Oco (Ocopon), 9=Ch (Chogo), 10=CA

(Cajamarca), 11=CAsm (San Marcos), 12=Ahy (Andahuaylas), 13=Kaq (Kaquiabamba), 14=Llhr (Huarochiri), 15=Ur (Uros), 16=Ap (Anapia), 17=Amt (Amantani), 18=Taq (Taquile), 19=And (Andoas), 20=Iq (Iquitos), 21=Puc (Pucallpa), 22=Chp (Chachapoyas), 23=SMla (Lamas), 24=LA (Lambayeque), 25=LI (Lima).

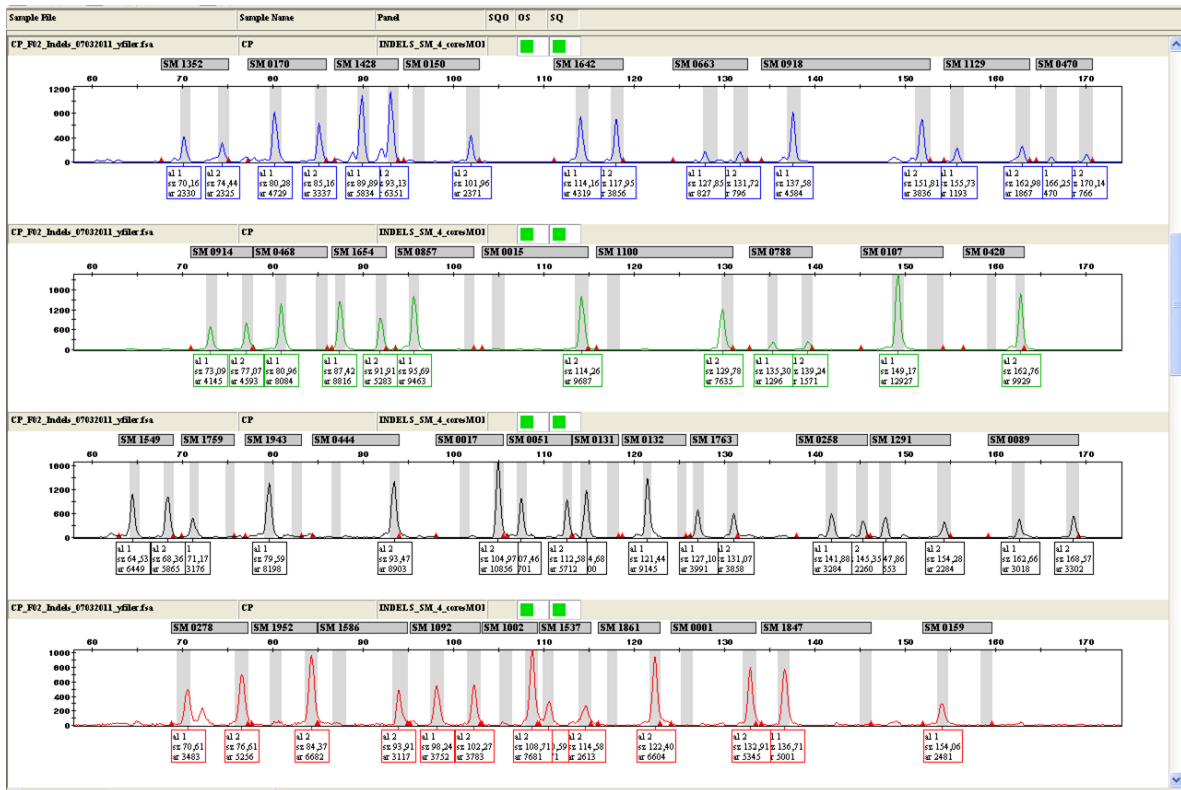


Figura 6: Genotipagem dos 40 loci com INDELS de uma amostra testada (referência) e visualizada através de GeneMapper ID v3.2.

5.2.3.1- Análises de fragmentos dos 40 INDELS autossômicos

Para as análises dos agrupamentos foram usados os programas STRUCTURE v2.3 (Falush, Stephens & Pritchard, 2003; Pritchard, Stephens & Donnelly, 2000), STRUCTURE HARVESTER (Earl & vonHoldt, 2011), BAPS (Corander *et al.*, 2008), ADMIX (www.genetica.fmed.edu.uy), PCAGEN (<http://www2.unil.ch/popgen/softwares/pcagen.htm>), CLUMPP (Jakobsson & Rosenberg, 2007), Distruct (Rosenberg, 2003), simco, ade4, FactoMineR (<http://www.r-project.org>) e GENETIX (<http://www.univ-montp2.fr/~genetix/>); e para as análises da variabilidade genética entre e dentro das populações os programas ARLEQUIN 3.5 e GENEPOP 4.0 (Rousset, 2008).

6 - RESULTADOS

6.1- Sobre a Região Controle do mtDNA

A distribuição das frequências absolutas dos sub-haplogrupos A2, B2, C1 e D1 (inclui o sub-haplogrupo D4) nas populações peruanas (n=700), das três diferentes regiões (Amazônia, Andes, Costa) são mostradas na Tabela 1, e visualizadas na Figura 7 segundo suas coordenadas geográficas. Observa-se que a frequência do subhaplogrupo B2 é predominante em grande parte das populações analisadas.

Para visualizar as conglomerações populacionais segundo a distribuição das frequências dos sub-haplogrupos A2, B2, C1 e D1 (Tabela 1), foi usado o método de PCA, cujo gráfico é apresentado na Figura 8. Segundo a configuração bidimensional incluindo a análise de frequências das quatro linhagens, as populações de Taq, Amt, Ap e Ur (do Lago Titicaca) mostram proximidade genética entre si, as quais estão mais distantes das populações de Iq=Iquitos e Tar=Tarapoto (ambas na Amazônia), assim como de PI=Piura (Costa norte), em relação às demais populações.

A análise de isolamento por distância (IBD) foi realizada entre as 29 populações peruanas para testar correlações entre a distância geográfica (km) e a distância genética (Fst de Reynolds) par a par dessas populações utilizando 999 permutações com o teste de Mantel através do software GenAlEx. A distância geográfica entre os locais foi medida através do programa `Geographic-Distance-Matrix-Generator` v1.2.3 (http://biodiversityinformatics.amnh.or/open_source/gdmg) segundo a latitude e longitude de cada local. O teste Mantel mostrou uma pequena correlação entre os dois fatores ($R^2=0,075$).

Tabela 1: Distribuição de frequências (absolutas) dos sub-haplogrupos A2, B2, C1 e D1(+D4) de mtDNA nas subpopulações peruanas segundo as regiões (Amazônia, Andes e Costa) e locais.

Região /Local	n	A2	B2	C1	D1 (+D4)
Amazônia					
Andoas	74	22	30	12	10 (+1)
Iquitos	6	4	-	1	1 (+1)
Pucallpa	10	3	4	2	1
Chachapoyas	14	4	6	2	2
Tarapoto	6	2	2	-	2
Lamas	23	7	13	3	-
Andes					
Cajamarca	36	5	16	9	6 (+1)
San Marcos	24	6	5	7	6 (+1)
Huaraz	8	3	1	3	1
Chogo	17	4	12	1	-
Ocopon	14	-	11	-	3
Huancayo	32	11	9	10	2
Huancavelica	5	2	3	-	-
Ayacucho	32	5	15	11	1
Andahuaylas	42	2	29	10	1
Kaquiabamba	8	1	1	4	2
Huarochiri	23	5	8	7	3
Taquile	51	-	51	-	-
Amantani	34	-	30	4	-
Uros	28	5	21	-	2
Anapia	24	-	21	2	1
Puno	6	-	3	3	-
Cabanaconde	18	2	5	5	6 (+1)
Chivay	20	2	11	2	5
Yanque	11	2	5	3	1
Arequipa	18	1	11	4	2
Costa					
Lima	69	18	27	13	11
Piura	5	2	-	3	-
Lambayeque	42	9	19	6	8 (+1)
Total	700	127	369	127	77

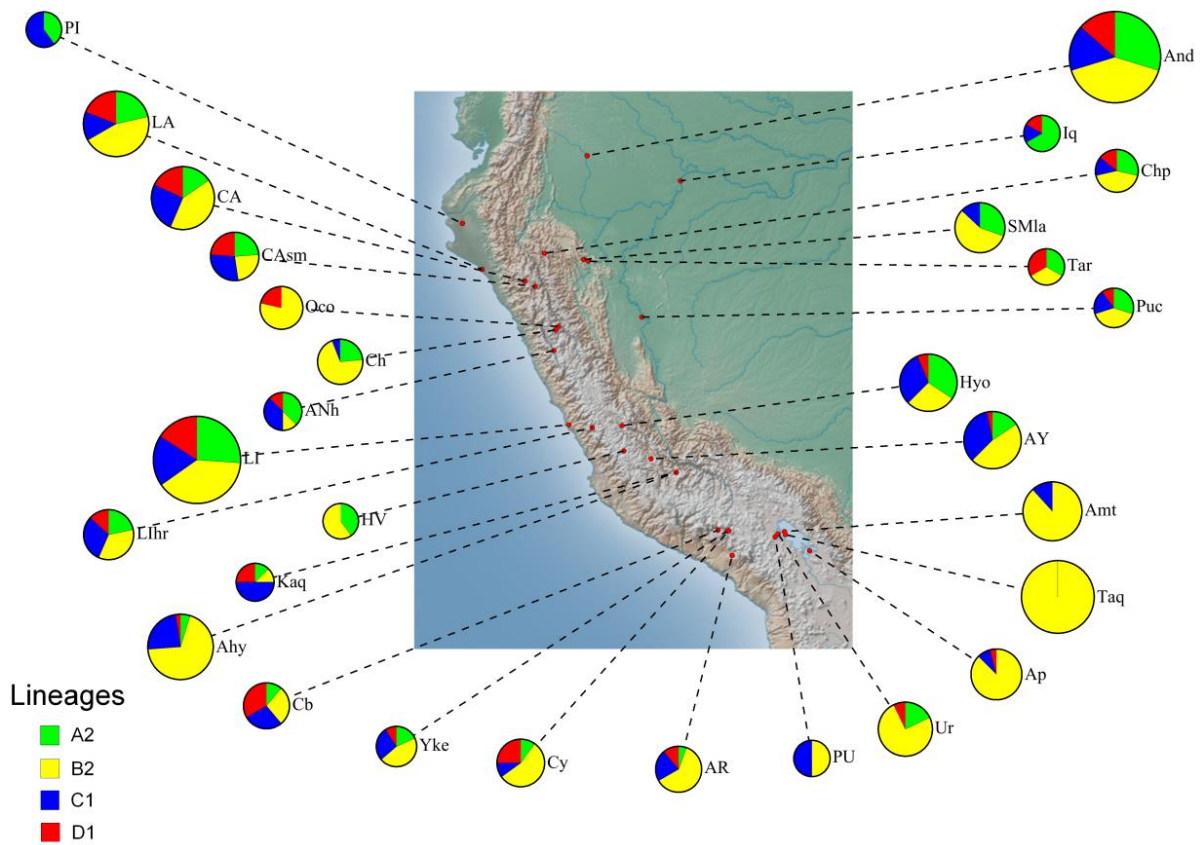


Figura 7: Distribuição das frequências absolutas dos sub-haplogrupos A2, B2, C1 e D1 do mtDNA nas 29 locais peruanas (n=700).

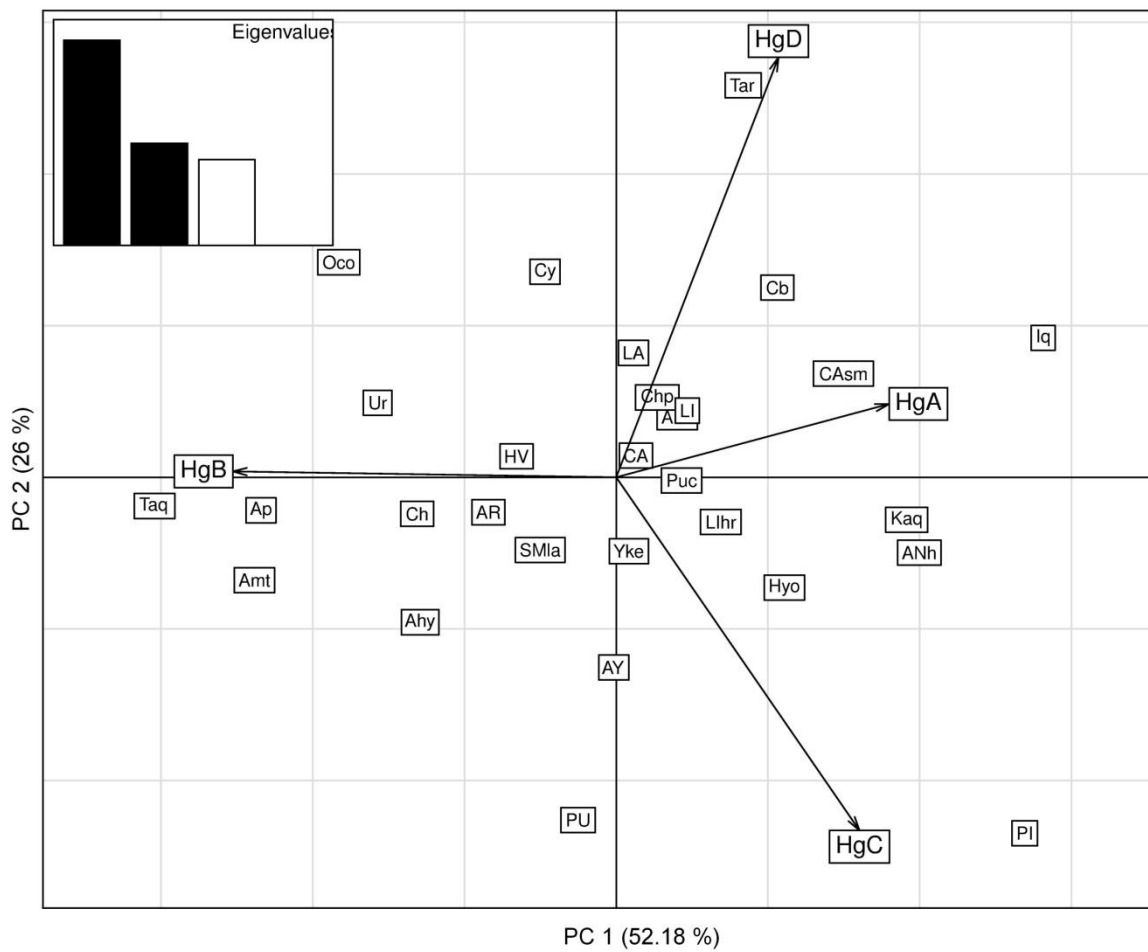


Figura 8: Gráfico bidimensional PCA com análise das frequências dos sub-haplogrupos A2, B2, C1 e D1 do mtDNA das subpopulações peruanas.

6.1.1- Polimorfismos da Região Controle do mtDNA nas populações peruanas

Através da ferramenta de predição MitoTool (Fan & Yao, 2011) foram obtidas as variantes SNPs da região controle do mtDNA em relação à rCRS e a classificação dos haplótipos mitocondriais segundo a nomenclatura “padrão” de PhyloTree (van Oven & Kayser, 2009). A nomenclatura de PhyloTree para os principais haplogrupos/sub-haplogrupos que caracterizam aos autóctones americanos são os seguintes, onde a variante ressaltada corresponde à região controle:

A: **235**-663-1736-4248-4824-8794-**16290-16319**

A2: **146-152!**-**153**-8027-12007-**16111**

B4'5: 8281-8289d

B4: **16217**

B4a: 5465-9123-**16261**

B4b: **499**-4820-13590

B2: 3547-4977-6473-9950-11177

C: 3552A-9545-11914-13263-14318-**16327**

C1: **290d-291d-16325**

D: 4883-5178A-**16362**

D4: 3010-8414-14668

D1: 2092-**16325**

Segundo essa nomenclatura, o MitoTool classificou aos haplótipos do haplogrupo A como A2, aos do B4 como B4b (equivalente ao B2 na literatura tradicional), aos do C como C1, aos de D como D1. Além disso, os haplótipos diferentes desses quatro mencionados foram classificados em outros sub-haplogrupos os quais são listados embaixo e na Tabela A5 no Anexo, em plena correspondência à nomenclatura de PhyloTree.

As variantes SNPs nas sequências de mtDNA da região controle e seus correspondentes sub-haplogrupos A2, B2, C1 e D1 determinados por MitoTool são listados por separado nas Tabelas A1, A2, A3 e A4 no Anexo e aqui somente mostramos aqueles sub-haplogrupos diferentes (o ressaltado corresponde à variante de acordo com a classificação de sub-haplogrupos por PhyloTree, e em parêntese a localidade de onde procede a amostra):

subHg B4a:

- 65G, 73, 75, 263, 310, 499, 16183C, 16189, 16217, **16261**, 16319, 16362, 16519. (Cajamarca)
- 65G, 73, 75, 94, 263, 309+CC, 315+C, 430, 455+T, 499, 16183C, 16189, 16197G, 16212, 16217, **16261**, 16319, 16519. (Huancayo)

- 65G, 73, 146, 185, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, **16261**, 16319, 16519.
(Andahuaylas)
- 65G, 73, 263, 309+C, 315+C, 499, 16183C, 16189, 16212, 16217, **16261**, 16319, 16519.
(Kaquiabamba)
- 65G, 73, 75T, 217, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, **16261**, 16319, 16519.
(Lima)
- 71+G, 73, 75, 263, 315+C, 340, 499, 16183C, 16189, 16217, **16261**, 16319, 16519.
(Chivay)
- 71+G, 73, 75, 94, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, **16261**, 16319, 16519.
(Huarochiri)
- 71+G, 73, 75, 94, 263, 315+C, 499, 16183C, 16189, 16217, **16261**, 16319, 16519.
(Huarochiri)
- 73, 263, 315+C, 499, 16168, 16183C, 16189, 16217, **16261**, 16519.
(Ayacucho)

subHg B4g:

- 51, 59, 60G, 61, 62C, 64, 65G, 73, 103T, 263, 310, 499, 16183C, 16189, 16217, 16261, **16292**, 16319, 16320, 16519. (Lima)

subHg B2h:

- 73, 263, 306, 310-312d, 499, 523-524d, 16176, 16182C, 16183C, 16189, 16217, 16319, **16468**, 16504, 16519. (Chivay)

subHg B2h:

- 73, 143, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16293, **16311**, **16468**, 16519. (Aurimac)

subHg D5:

- 67T, 97, 100-105d, **150**, 263, 309+CC, 315+C, 489, 16172, 16189, 16223, 16325, 16362. (Andoas (n=3))

subHg D4h3a:

- 73, 263, 315+C, 489, 16223, 16241, **16301**, **16342**. (Andoas)

subHg D4h3a1:

- 73, 146, 263, 309+C, 315+C, 489, 16223, 16241, **16301**, 16311, **16342**, 16362. (Iquitos)

subHg D4h2:

- 73, 263, 309+C, 315+C, 489, 16209, **16278**, **16325**, 16362. (Cajamarca)

subHg D4h3a:

- 73, **152**, 263, 309+CC, 315+C, 489, 16189, 16223, 16241, 16278, **16301**, 16316, **16342**, 16362. (San Marcos-Cajamarca)

subHg D4j2:

- 73, **152**, 263, 309+C, 315+C, 489, 16051, 16172, 16209, 16223, 16256, **16291**, 16325, 16362. (Lambayeque)

subHg D4h3a:

- 73, 103, **152**, 263, 315+C, 489, 16223, 16241, **16301**, **16342**, 16362. (Cabanaconde)

Observa-se que todos os haplótipos do subHg B4a estão associados pela variação na posição 16261 tal como assinala PhyloTree e seu compartilhamento sugere a ancestralidade comum dos indivíduos dessas regiões geográficas. O sub-haplogrupo B4a é frequente em algumas populações do Leste Asiático e da Polinésia (Soares *et al.*, 2009).

Por outro lado, o conjunto que envolve as posições 152-16301-16342 do subHg D4h3 foi reportado entre os Chachis, conhecidos também como Cayapas (Equador) e Siberianos (Perego *et al.*, 2009; Volodko *et al.*, 2008), e segundo nossas observações é possível que estejam também espalhados nas outras populações autóctones da América do Sul, apesar de sua baixa frequência.

Entre os diversos sítios polimórficos da região controle do mtDNA destaca-se o SNP 489T, variante ancestral que diferencia em dois agrupamentos os quatro sub-haplogrupos (A2, B2, C1, D1) em concordância com a nomenclatura de PhyloTree (van Oven & Kayser, 2009): em um grupo os subHgs A2 e B2 (dentro do macro-grupo N) e em outro agrupamento os C1 e D1 (macro-grupo M). Observa-se também que no sítio 499G há uma relação próxima entre os subHgs A2, C1 e D1 em contraste ao subHg B2.

Segundo a distribuição das frequências dos sub-haplogrupos A2, B2, C1 e D1 ao longo das três regiões do Peru observa-se que o subHg B2 é predominante. Considerando esta particularidade, tentamos reconstruir o nível de parentesco genético ou associações entre as

populações peruanas baseado nos SNPs desse sub-haplogrupo. As sequências nucleotídicas ou haplótipos que compartilharem mais SNPs específicos devem ser mais próximas do que aquelas que compartilharem menos, o que estaria correlacionado com seu grau de ancestralidade. Assim, o sítio 16051 (A→G) está associado mais com o norte do Peru; é frequente em Lamas, Lambayeque e Andoas, mas não foi encontrado na parte sul do país. O sítio 16168 (C→T) ocorre em vários indivíduos de Ayacucho e Andahuaylas. Esta característica é concordante com sua proximidade geográfica. O sítio 16503 (G→A) inclui alguns indivíduos de Andoas (n=4) e um de Chachapoyas (e nenhum nas outras populações), sugerindo que alguns Ashuar e Awajun compartilham em parte a linhagem materna, concordando com sua história demográfica similar (Lexus, 1998) já que pertencem à mesma família lingüística “Jívaro” (<http://www.ethnologue.tv>).

Por outro lado, na região sul do Peru observa-se outra particularidade entre as populações. Vários sítios polimórficos (73 (A); 186 (C→T); 316 (G→A); 16170 (A→G); 16188 (C→T) e 16354 (C→T)) associam as populações do Lago Titicaca, Puno e as de Arequipa (Chivay, Yanque). Estas relações das variantes mitocondriais sugerem que as populações do sul do Peru têm uma história de ancestralidade recente compartilhada e concorda com as crônicas sobre o período colonial que relatam o alto fluxo gênico entre elas desde tempos pré-colombianos (Vellard, 1959; Wachtel, 1986; Lexus, 1998).

Cabe mencionar que entre os Quéchuas-Lamistas (Lamas) a variante 16086 é comum e o mesmo SNP foi observado em algumas populações do sudeste siberiano como os Altaianos, Khakasianos, Tuvinianos, entre outros (Derenko *et al.*, 2003).

Embora a elevada homoplasia observada na região controle seja um factor que limita as inferências sobre as associações de parentesco, nossa abordagem permite a “grosso” modo ter ideia do fluxo gênico ou deriva, fatores que influem nas variações ou similaridades interpopulacionais.

6.1.2- Análises de regressão linear dos sub-haplogrupos A2, B2, C1 e D1 segundo a latitude

Para testar a hipótese de uma possível associação entre a frequência de um determinado sub-haplogrupo (variável resposta) e a latitude (variável explicativa) foi usado o método da análise de regressão linear através do programa GenGIS. Os testes por cada haplogrupo separadamente mostraram que há uma moderada correlação entre as distribuições das frequências dos sub-haplogrupos A2 e B2 com a latitude ($R^2=45\%$ e $R^2=28\%$ respectivamente). O R^2 é o coeficiente de determinação. O resultado mostrou que a frequência do subHg A2 diminui de norte a sul (Figura 9), em contraste ao subHg B2 que aumenta (Figura 10). Uma menor porcentagem da associação foi calculada para o subHg D1 ($R^2=10\%$), que é praticamente nula para o subHg C1 ($R^2=0,8\%$).

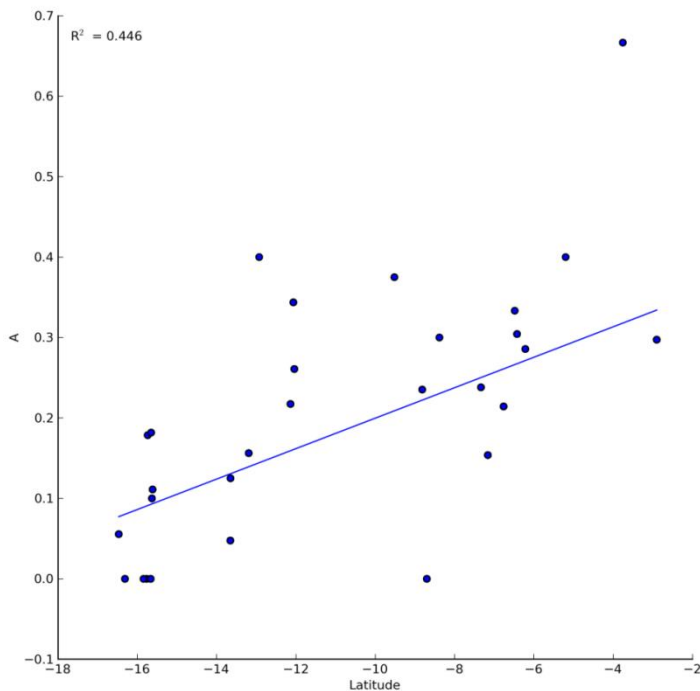


Figura 9: Análise de regressão linear entre a latitude e o sub-haplogrupo A2 (A).

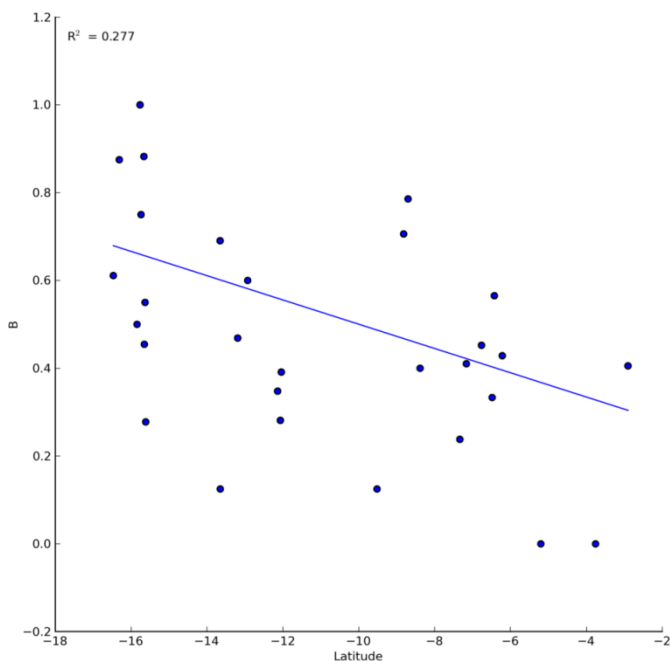


Figura 10: Análise de regressão linear entre a latitude e o sub-haplogrupo B2 (B).

6.1.3- Determinação dos agrupamentos A2, B2, C1 e D1 de mtDNA nas populações peruanas por PhyML

Para reconstruir a árvore filogenética a partir das sequências de mtDNA alinhadas usou-se o programa PhyML 3.0 baseado no algoritmo de máxima verossimilhança. O PhyML tem implementado o método aLRT (*fast approximate likelihood ratio*) para avaliar o suporte dos dados para os ramos internos ou clados de uma filogenia (similar ao *bootstrap*), que é um procedimento não paramétrico conhecido como abordagem Shimodaira-Hasegawa. O algoritmo compara a configuração atual da subárvore ou topologia em torno do ramo interno de interesse para duas alternativas definidas pelo método heurístico NNI (*Nearest-Neighbor-Interchange*), segundo algum modelo de substituição nucleotídica. Calcula-se a probabilidade destas duas alternativas e executa-se uma razão aproximada de teste (aLRT) baseado no *log-ratio* entre os valores da árvore atual e aquela com melhor alternativa. Os testes foram realizados usando 505 sequências de mtDNA sob três modelos de substituição (HKY85, F84, TN93), mas o modelo que melhor mostrou a configuração da árvore atual (comparado com a árvore de PhyloTree.org) foi utilizando o modelo TN93 (com quatro categorias segundo *default*) com árvore inicial gerada através do método BioNJ (Bio Neighbor-Joining), mostrada na Figura 11.

Através dos sítios polimórficos (S=339) das sequências de mtDNA (n=703) gerados pelo programa GenAlEx, foram feitas as análises multivariadas usando os programas ade4 e FactoMineR. O resultado dos conglomerados de haplótipos segundo sua linhagem ou sub-

haplogrupo (A2, B2, C1 e D1) correspondente é mostrado na Figura 12, que é um mapa fatorial gerado e baseado nas análises dos componentes principais (PCA), apresentando estreita concordância com a árvore gerada através do método de máxima verossimilhança.

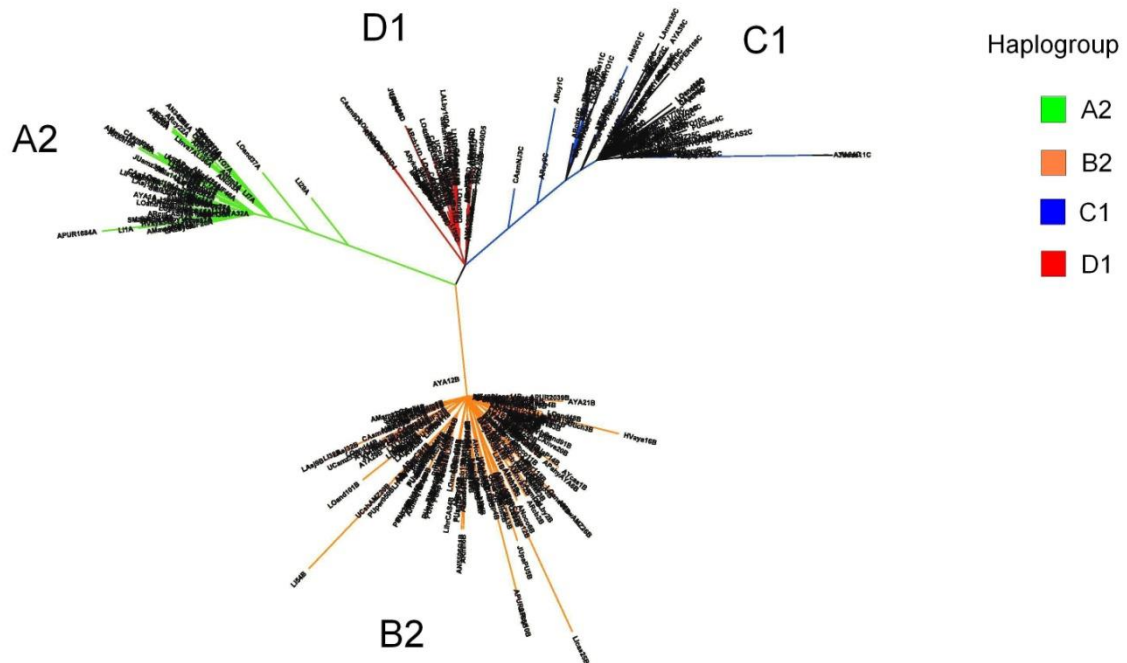


Figura 11: Árvore gerada pelo PhyML sob modelo TN93 para os sub-haplogrupos A2, B2, C1 e D1.

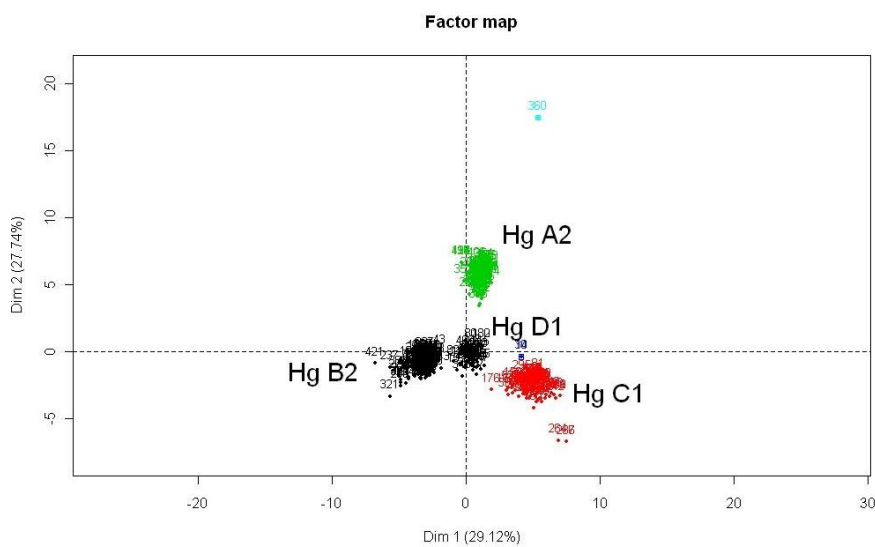


Figura 12: Mapa fatorial das análises de PCA de haplótipos dos sub-haplogrupos A2, B2, C1 e D1 nas populações peruanas baseado na região controle de mtDNA.

6.1.4- Índices de diversidade dos sub-haplogrupos A2, B2, C1 e D1 e teste de neutralidade

Através do programa DnaSP foram calculados vários índices (Tabela 2), entre eles a diversidade haplotípica (h), diversidade nucleotídica (π), número de haplótipos (k), número de sítios segregantes ou polimórficos (S). Os resultados dos testes de neutralidade de Tajima (D) sob o modelo de sítios infinitos são apresentados na Tabela 3.

O teste de Tajima é realizado para ver os efeitos demográficos e/ou seleção natural sob a hipótese nula de neutralidade. As substituições nucleotídicas são consideradas como sítios segregantes (S). A abordagem compara dois estimadores populacionais do parâmetro *Theta* (θ), um baseado no número de sítios segregantes (S , θ_w) na sequência e outro baseado na média das diferenças entre os haplótipos comparados (pairwise difference, $\theta\pi$): $\theta\pi - \theta_w$

Se o D de Tajima é igual a 0, significa que há equilíbrio entre a mutação e a deriva genética; se é positivo e significativo, sugere que há estruturação populacional, gargalo populacional (*bottleneck*) ou seleção balanceadora; e se é negativo e significativo, há um excesso de variantes raras, um sinal de expansão populacional ou de seleção natural positiva.

O resultado do teste de Tajima para todas as linhagens em conjunto ($n=700$, todos os quatro haplogrupos) ou para cada haplogrupo separadamente, mostraram valores negativos e significativos ($p<0,05$), sugerindo expansão populacional. Esta hipótese é apoiada pelos gráficos sobre as distribuições de diferenças nucleotídicas (*mismatch*) entre os haplótipos, onde para cada haplogrupo foi observado uma curva unimodal (figuras não mostradas).

Tabela 2: Índices de diversidade baseados nas sequências da Região Controle do mtDNA. As amostras ($n=700$) correspondem às 29 localidades das três regiões do Peru (Amazônia, Andes e Costa).

N	L	Índices de diversidade				
		K	Hd	S	π	$\theta\pi$
700	1129	428	0,996	306	0,0104	11,74
HgA2 (90)	1107	83	0,998	117	0,0055	6,09
HgB2 (236)	1092	196	0,998	740	0,0106	11,57
HgC1 (114)	1114	92	0,991	119	0,0053	5,90
HgD (65)	1108	53	0,988	75	0,0056	6,20

N=número de amostras

L=tamanho da sequência em pb

K=número de haplótipos

Hd=diversidade haplotípica

S=número de sítios polimórficos

π =diversidade nucleotídica (por sítio)

$\theta\pi$ =média das diferenças nucleotídicas (*pairwise differences*)

Tabela 3: Teste de neutralidade de Tajima baseado nas sequências mtDNA da Região Controle. As amostras (n=700) correspondem às 29 localidades das três regiões do Peru (Amazônia, Andes e Costa).

N	L	S	Teste de Tajima				D	p
			η	$\theta\pi$	π			
700	1088	298	323	11,21	0,0103	-2,16	<0,01	
HgA (90)	1107	117	119	6,09	0,0055	-2,46	<0,01	
HgB2 (236)	1092	740	829	11,57	0,0106	-2,88	<0,001	
HgC1 (114)	1114	119	120	5,90	0,0053	-2,40	<0,01	
HgD1 (65)	1108	75	75	6,20	0,0056	-2,07	<0,05	

N=número de amostras
L=tamanho da sequência em pb
S=número de sítios polimórficos (segregantes)
 η =eta=número total de mutações
 $\theta\pi$ =média das diferenças nucleotídicas
 π =diversidade nucleotídica
D=D de Tajima
p=Probabilidade

6.1.5- Análises de AMOVA sobre a estrutura populacional do Peru

A comparação populacional (29 localidades, n=703) de amostras agrupadas em diferentes níveis hierárquicos revela valores contrastantes de estatística F, obtidos por AMOVA (Tabela 4). Comparando entre as três regiões geográficas do Peru (Amazônia, Andes e Costa), o valor de 1,05% assinala que não há variação significativa entre as regiões, mas a diferença entre as populações (dentro dos grupos) é significativa, 9,34%. O valor de 0,104 do Fst (distância genética) envolvendo as três regiões peruanas indica que há pouca diferenciação entre todas as populações comparadas. Nas análises por região, a grande parte da variabilidade está dentro das populações sendo menor nos Andes quando comparado com às da Amazônia e Costa; em contraste, entre as populações dos Andes apresenta-se maior diferenciação (Fst=0,131) do que entre as da Amazônia (Fst=0,011, não significativo) e entre as da Costa (Fst=0,016, não significativo). Quando as populações peruanas são analisadas sem qualquer divisão, obtém-se um Fst significativo de 0,099.

Da mesma maneira foram feitas os testes de AMOVA para cada haplogrupo nas populações em diferentes agrupamentos. Nas Tabelas 5 (A2), 6 (B2), 7 (C1) e 8 (D1) são mostrados os valores em porcentagens das variações entre as populações segundo o agrupamento hierárquico em três regiões. Pelos sub-haplogrupos A2 e B2 observa-se que há maior

diferenciação entre as populações andinas do que entre as da Amazônia e da Costa. Em contraste, a análise com dados dos sub-haplogrupos C1 e D1 apresenta maior variação entre as da Amazônia. Essas diferenças observadas nos indicam que cada linhagem tem sua própria história, permitindo explorar as relações interpopulacionais e revelar algumas peculiaridades sobre suas possíveis associações ou diferenças demográficas.

Tabela 4: Resultado de AMOVA. Análises do mtDNA nas 29 subpopulações peruanas segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100

Região	Entre grupos (%)	Entre populações *WG (%)	Entre populações (%)	Dentro das populações (%)
3 regiões				
(Amazônia, Andes, Costa)	1,05	9,34		89,62
Por região				
Amazônia			1,07	98,93
Andes			13,05	86,95
Costa			1,59	98,41
Sem divisão				
1 grupo			9,96	90,04

*WG=Within Groups

- 1) 3 regiões: subpopulações=29
 $F_{sc}=0,094$ ($p=0$)
 $F_{st}=0,104$ ($p=0$)
 $F_{ct}=0,01$ ($p=0,186$)
- 2) Amazônia: subpopulações=6 (And,Iq,Puc, Chp,Tar,SMIa)
 $F_{st}=0,011$ ($p=0,214$)
- 3) Andes: subpopulações=20
(CA,CAsm,ANh,Ch,Oco,Hyo,HV,AY,Ahy,Kaq,LIhr,Taq,Amt,Ur,Ap,PU,Cb,Cy,Yke,AR)
 $F_{st}=0,131$ ($p=0$)
- 4) Costa: subpopulações=3 (PI,LI,LA)
 $F_{st}=0,016$ ($p=0,093$)
- 5) Sem divisão:
 $F_{st}=0,0996$ ($p=0$)

Tabela 5: Resultado de AMOVA. Análises do sub-haplogrupo A2 do mtDNA em 24 subpopulações peruanas (cinco não apresentam A2) segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100

Região	Entre grupos (%)	Entre populações *WG (%)	Entre populações (%)	Dentro das populações (%)
3 regiões (Amazônia, Andes, Costa)	-3,34	19,94		83,39
Amazônia			8,48	91,52
Andes			34,0	66,0
Costa			-2,1	102,1

*WG=Within Groups

- 1) 3 regiões: subpopulações=24
Fsc=0,193 (p=0)
Fst=0,166 (p=0)
Fct=-0,033 (p=0,866)
- 2) Amazônia: subpopulações=6 (Iq, And,Puc, Chp, Tar, Smla)
Fst=0,085 (p=0,019)
- 3) Andes: subpopulações=15
(CA,CAsm,ANh,Ch,Hyo,HV,AY,Ahy,Kaq,LIhr,Ur,Cb,Cy,Yke,AR)
Fst=0,34 (p=0)
- 4) Costa: subpopulações=3 (PI, LI, LA)
Fst=-0,021 (p=0,736)

Tabela 6: Resultado de AMOVA. Análises do sub-haplogrupo B2 do mtDNA em 27 subpopulações peruanas (duas não apresentam B2) segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100

Região	Entre grupos (%)	Entre populações *WG (%)	Entre populações (%)	Dentro das populações (%)
3 regiões (Amazônia, Andes, Costa)	3,04	16,47		80,49
Amazônia			14,45	85,55
Andes			19,4	80,6
Costa			6,02	93,98

*WG=Within Groups

- 1) 3 regiões: subpopulações=27
Fsc=0,17 (p=0)
Fst=0,195 (p=0)
Fct=0,03 (p=0,121)
- 2) Amazônia: subpopulações=5 (And, Puc, Chp, Tar, SMIa)
Fst=0,144 (p=0,0003)
- 3) Andes: subpopulações=20
(CA,CAsm,ANh,Ch,Oco,Hyo,HV,AY,Ahy,Kaq,LIhr,Taq,Amt,Ur,Ap,PU,Cb,Cy,Yke,AR)
Fst= 0,194 (p=0)
- 4) Costa: subpopulações=2 (LI,LA)
Fst=0,06 (p=0,015)

Tabela 7: Resultado de AMOVA. Análises do sub-haplogrupo C1 do mtDNA em 24 subpopulações peruanas (cinco não apresentam C1) segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100

Região	Entre grupos (%)	Entre populações *WG (%)	Entre populações (%)	Dentro das populações (%)
3 regiões (Amazônia, Andes, Costa)	0,74	5,06		94,2
Amazônia			8,31	91,69
Andes			5,95	94,05
Costa			-2,01	102,01

*WG=Within Groups

- 1) 3 regiões: subpopulações=24
Fsc=0,051 (p=0,007)
Fst=0,058 (p=0,0024)
Fct=0,007 (p=0,221)
- 2) Amazônia: subpopulações=5 (And, Iq, Puc, Chp, SMIa)
Fst=0,083 (p=0,167)
- 3) Andes: subpopulações=16
(CA, CAsm, ANh, Ch, Hyo, AY, Ahy, Kaq, LIhr, Amt, Ap, PU, Cb, Cy, Yke, AR)
Fst= 0,06 (p=0,006)
- 4) Costa: subpopulações=3 (PI, LI, LA)
Fst=-0,02 (p=0,67)

Tabela 8: Resultado de AMOVA. Análises do sub-haplogrupo D1 do mtDNA em 22 subpopulações peruanas (sete não apresentam D1) segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100

Região	Entre grupos (%)	Entre populações *WG (%)	Entre populações (%)	Dentro das populações (%)
3 regiões (Amazônia, Andes, Costa)	1,09	9,77		89,14
Amazônia			13,11	86,89
Andes			9,33	90,67
Costa			4,88	95,12

*WG=Within Groups

- 1) 3 regiões: subpopulações=22
Fsc=0,098 (p=0,025)
Fst=0,108 (p=0,004)
Fct=0,011 (p=0,22)
- 2) Amazônia: subpopulações=5 (And, Iq, Puc, Chp, Tar)
Fst=0,131 (p=0,136)
- 3) Andes: subpopulações=15
(CA, CAsm, ANh, Oco, Hyo, AY, Ahy, Kaq, LIhr, Ur, Ap, Cb, Cy, Yke, AR)
Fst= 0,093 (p=0,042)
- 4) Costa: subpopulações=2 (LI, LA)
Fst=0,05 (p=0,115)

O gráfico MDS das distâncias genéticas Fst de Reynolds (índices de coancestralidade) é mostrado na Figura 13. Na matriz de distâncias genéticas Fst entre as populações com sequências do sub-haplogrupo A2 (24 populações) foi observado que as subpopulações de Ur, Yke, Kaq, AR, LIhr e ANh estão mais diferenciadas quando comparadas com as demais localidades. A visualização no gráfico MDS (Figura 14) feito a partir dos índices de coancestralidade de Reynolds revela um agrupamento de 18 subpopulações e outro de 6 subpopulações (LIhr, ANh, Ur, Yke, Kaq, AR) que estão um pouco afastadas de elas.

Na matriz de distâncias genéticas Fst entre as populações com sequências do sub-haplogrupo B2 (27 populações) foi observado que as subpopulações de Ur, ANh e Kaq

estão mais diferenciadas quando comparadas com as demais localidades. O gráfico MDS (Figura 15) com os índices de coancestralidade de Reynolds revela que essas três localidades estão nos lados contrários do espaço bidimensional (Ur no extremo de ANh e Kaq). As subpopulações do Lago Titicaca (Ur, Taq, Amt, Ap, PU) estão mais próximas às do sul do Peru (Cy, AR, Yke) em correspondência à sua localização geográfica. As demais populações (exceção de ANh e Kaq) formam um grupo mais heterogêneo, embora se observe compacto. Neste MDS a localidade de SMIa também parece estar um pouco diferenciada das demais.

Na matriz de distâncias genéticas F_{st} entre as subpopulações com dados do sub-haplogrupo C1 (24 populações) foi observado que cinco destas (Ch, Kaq, Amt, Cy e SMIa) estão mais diferenciadas quando comparadas par a par com as demais localidades. O gráfico MDS (Figura 16) com os índices de coancestralidade de Reynolds revela que essas cinco subpopulações estão afastadas das demais (que formam um conglomerado compacto).

Na matriz de distâncias genéticas F_{st} entre as populações com dados do sub-haplogrupo D1 (22 populações) foi observado que três subpopulações (Ur, Ap e Yke) estão mais diferenciadas quando comparadas par a par com as demais localidades. O gráfico MDS (Figura 17) com análise dos índices de coancestralidade de Reynolds revela que essas três subpopulações junto com Cb estão afastadas das demais (que formam conglomerados heterogêneos).

Em geral, os perfis observados nas análises de AMOVA e MDS sugerem que os locais afastados correspondem a populações onde acontece menor fluxo gênico e, como consequência, tem menor variabilidade e maior distância genética com outras populações. Em contraste, nas populações conglomeradas ou mais compactas há um maior intercâmbio de migrantes que conduz à homogeneidade interpopulacional diminuindo sua distância genética, mas aumentando sua variabilidade genética intrapopulacional.

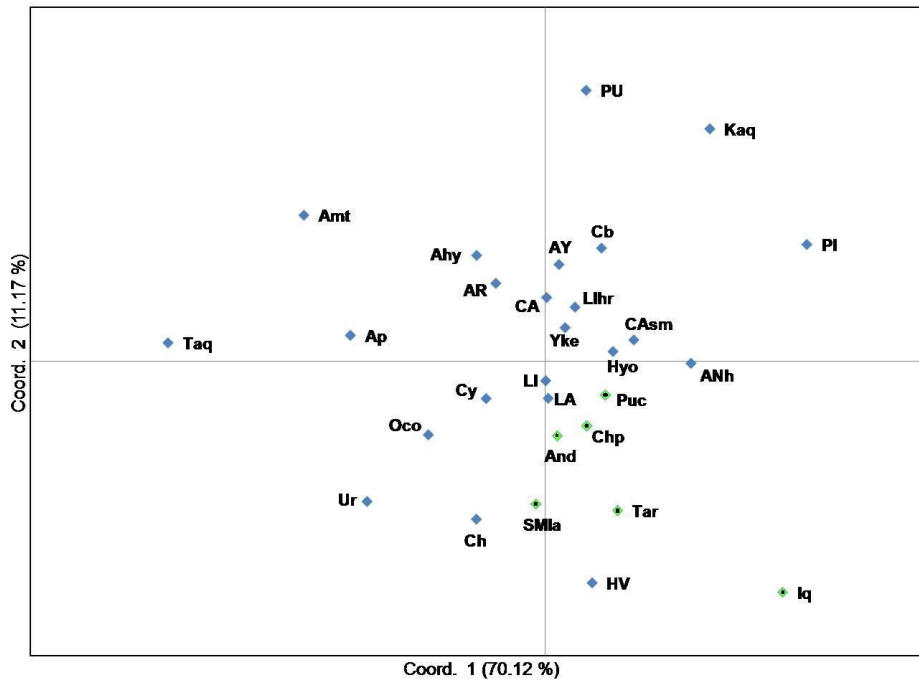


Figura 13: Plotagem MDS baseada na distância genética (Fst de Reynolds) das 29 subpopulações segundo todos os sub-haplogrupos mtDNA (A2, B2, C1 e D1).

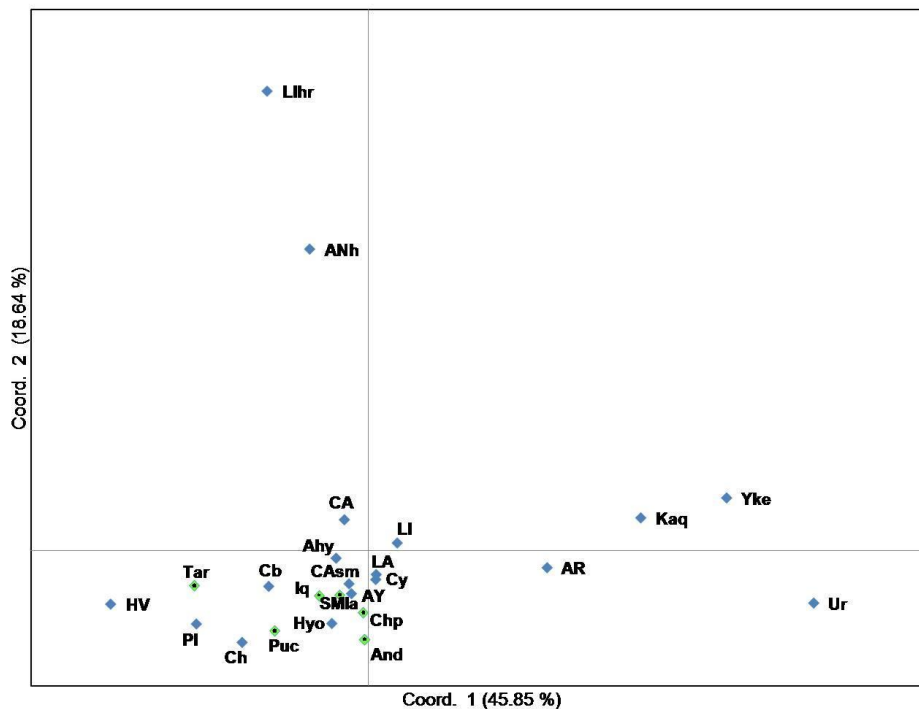


Figura 14: Plotagem MDS baseada na distância genética (Fst de Reynolds) das populações segundo o sub-haplogrupo A2.

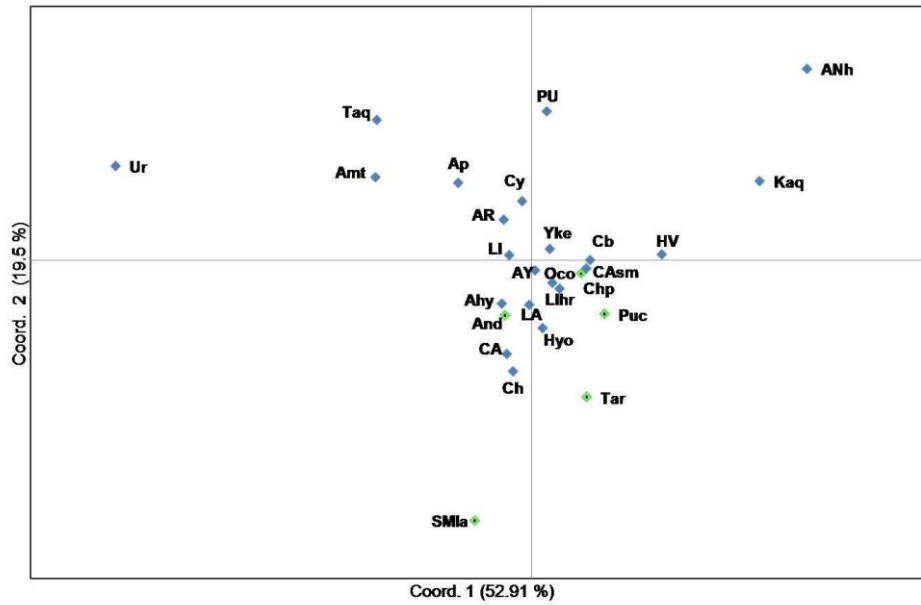


Figura 15: Plotagem MDS baseado na distância genética (Fst de Reynolds) das populações segundo o sub-haplogrupo B2.

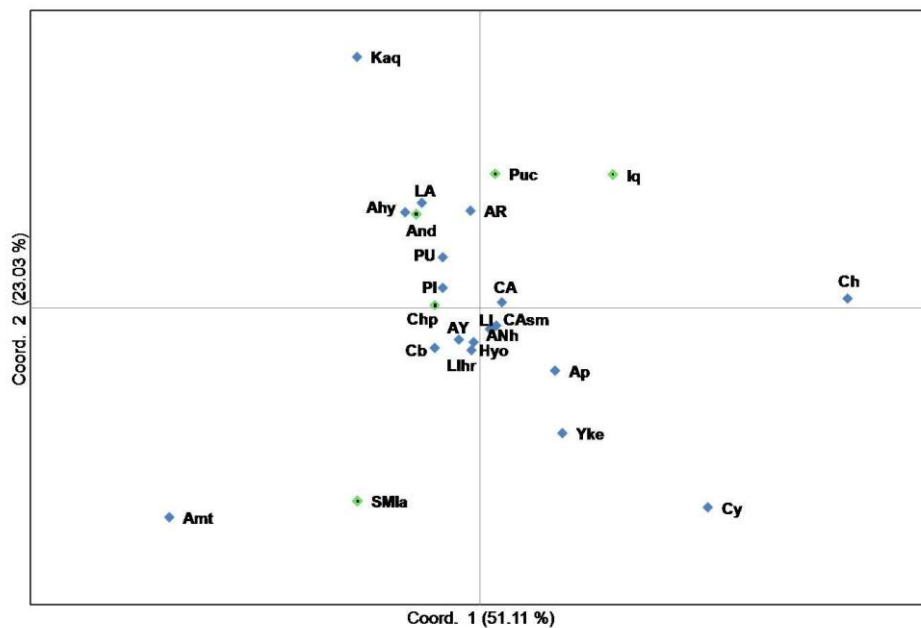


Figura 16: Plotagem MDS baseada na distância genética (Fst de Reynolds) das populações segundo o sub-haplogrupo C1.

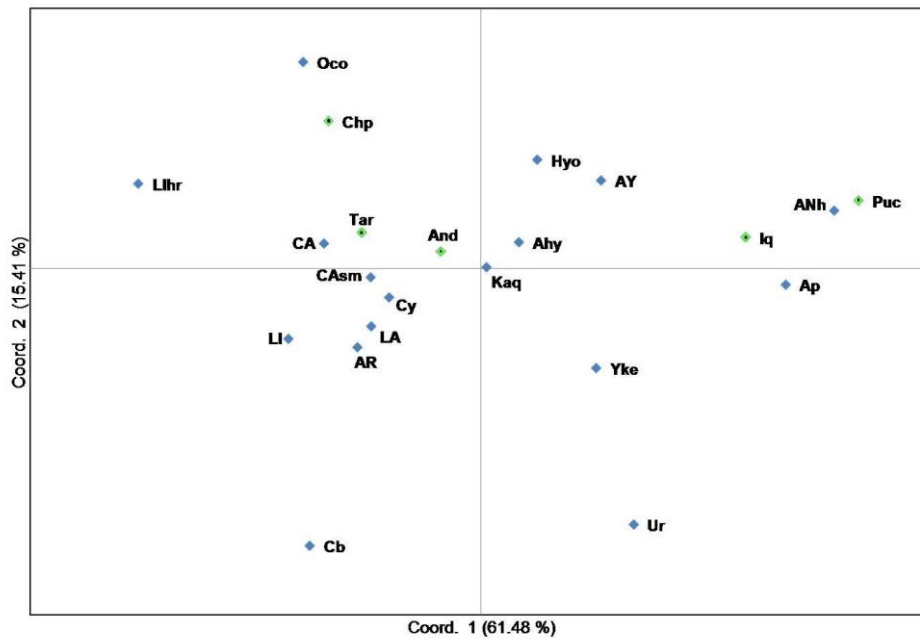


Figura 17: Plotagem MDS baseada na distância genética (Fst de Reynolds) das populações segundo o sub-haplogrupo D1.

6.1.6- Comparação das linhagens A2, B2, C1 e D1 com outras linhagens maternas das populações continentais

Com a finalidade de reconstruir os perfis genéticos mitocondriais dos peruanos contemporâneos, nossa abordagem consistiu em comparar a nível global as linhagens nativas americanas do Peru com outras populações autóctones continentais e dessa maneira desvendar os graus de parentesco, miscigenação ou de ancestralidade compartilhada entre elas. Para isso, sequências de mtDNA da Região Controle (n=333) foram obtidas através do GenBank (NCBI), a partir dos resultados de M.E. Healy e K.L. Huxley com o painel de populações HGDP-CEPH (resultados não publicados, intitulado: A simulation-based approach for comparing human male and female demographic histories. Department of Anthropology, University of New Mexico, Albuquerque, USA. Estes dados também aparecem na lista de publicações de HGDP-CEPH: University New Mexico (UNM) HGDP-CEPH mtDNA D-loop sequences supplement 6 (June 3, 2008).

As seguintes populações foram selecionadas: África (n=39): Biaka (=17) da República Central Africana; San (n=7) da Namíbia; Bantu (n=17) do Quênia e África do Sul; Oriente Médio: Druzos (n=29) de Israel; Ásia Central: Hazara (n=25) do Paquistão; Oceania: Papuas (n=17) da Papua Nova Guiné; Europa: Italianos (n=14) de Bérgamo; Franceses (n=29); Leste Asiático: Cambodianos (n=11), Han (n=32) da China, Mongólia (n=10), Yakut (n=25)

da Rússia (Sibéria), Japoneses (n=45); América: Pima (n=25) e Maias (n=25) do México, Piapoco (n=13) da Colômbia, Karitiana (n=24) e Surui (n=21) do Brasil.

Na primeira fase, para identificar os haplogrupos nessas populações foi usado o programa de predição MitoTool. Na segunda fase todas as sequências de mtDNA (incluindo as populações peruanas) foram alinhadas através do programa MEGA (MUSCLE ou CLUSTALW) seguindo as sugestões do PhyloTree (nomenclatura padrão). As comparações foram feitas com base no alinhamento da posição 16.024 até a posição 407 da Região Controle do mtDNA da amostra de referência (rCRS).

Uma rede filogenética foi construída pelo programa Network. Na Figura 18 são observados os agrupamentos segundo os haplogrupos e na Figura 19 são mostradas as regiões geográficas às quais estão associados os haplogrupos (America C=América Central; America S=América do Sul). Em geral, o perfil genético mitocondrial é consistente com a distribuição das linhagens femininas nos continentes, por exemplo, os sub-haplogrupos A2, B2, C1 e D1 se distribuem quase exclusivamente na América (exceção de algumas populações siberianas). As linhagens B4'5 do Leste Asiático são próximas ao subHg B2 e as linhagens D4 e D5 aparecem próximas ao subHg D1 em concordância com a árvore PhyloTree. As linhagens P1, Q1 e Q3a da Papua Nova Guiné formam agrupamentos indicando uma relação filogenética próxima entre eles. As linhagens associadas à Ásia Central, Leste Asiático e Europa compartilham perfis num mesmo conglomerado, sugerindo um elevado fluxo gênico na Eurásia que coincide com sua história de ondas migratórias e recolonizações (Semino *et al.*, 2000). Observou-se um indivíduo com haplogrupo L3 na população Maia com o MitoTool (HGDP00861) sugerindo que seja derivado de um fluxo pós-colombiano recente da África, provavelmente associada aos Bantu.

Com a finalidade de revelar as distâncias genéticas entre as 47 populações a nível continental foi realizado o teste AMOVA no programa Arlequin. As populações foram divididas em 9 grupos segundo suas regiões geográficas da amostragem (8 grupos são referências, n=333), e como um grupo as populações peruanas (n=700)). Segundo os resultados de AMOVA (Tabela 9), há uma diferenciação de 24,5% entre todas as 47 populações comparadas, no entanto só há 12,5% de variação entre os 9 grupos geográficos. A maior parte da variação genética está dentro das populações (75,5%).

A matriz dos coeficientes da coancestralidade (Fst de Reynolds) mostrou que as populações Pima, Karitiana e Taquile (Taq) estão fortemente diferenciadas de Biaka e San (África), da mesma forma, como Taquile e Surui da população Pima (América do Norte). Observa-se também uma moderada diferenciação de Taquile vs Iquitos (Iq) e Piura (PI). Um dos fatores que provavelmente conduz a esse perfil diferenciado seria a ação da deriva genética além do viés de amostragem (muitos eram estreitamente aparentados), já que MitoTool revelou que entre os Pima (n=25) existem quase unicamente subHg C1 (exceto um A2), entre os

Karitiana (n=24) todos são subHg D1; entre os Surui (n=21) a maioria é subHg D1 (exceto três A2); entre os de Taquile (n=51) só existe subHg B2; e na população dos Biaka (n=15) observa-se 13 indivíduos com perfil L1, um L0 e uma amostra não definida por MitoTool; entre os San (n=7) seis são L0 e só um L4.

No gráfico MDS (Figura 20), além dos contrastes mencionados anteriormente, observa-se que as populações do Lago Titicaca (Taq, Amt, Ur e Ap) são próximas entre si, sugerindo ancestrais comuns ou fluxo gênico recente, assim como afinidade entre Cy (Chivay) e AR (Arequipa); há associação entre a população do druzos, franceses e italianos que é consistente com sua história migratória (Semino *et al.*, 2000, Dubut *et al.*, 2004). Por outro lado, as populações de Puc (Pucallpa), Chp (Chachapoyas), Tar (Tarapoto) estão próximas a Yakut (Sibéria). A proximidade genética entre SMLa (Lamas) e Piapoco (Colombia), possivelmente seja por efeito das semelhantes frequências dos subHg A2, B2 e C1 e ausência do subHg D1 em ambas as populações. O conglomerado AY (Ayacucho), LA (Lambayeque), And (Andoas), LI (Lima) e CA (Cajamarca) sugere antepassados comuns ou fluxo gênico entre eles. A relação entre os Han e Hazara coincide com a história sobre a expansão da China à Ásia Central. As demais associações de populações poderiam ser também explicadas da maneira similar.

Tabela 9: Resultado de AMOVA. Análises da Região Controle mtDNA nas 47 subpopulações continentais (n=1033) segundo o agrupamento hierárquico. Iterações=10.100

Região	Entre grupos (%)	Entre populações *WG (%)	Dentro das populações (%)	Estatísticos F
9 grupos (Peru, América do Sul, América Central, Leste Asiático, Europa, Oceania, Ásia Central, Oriente Médio, África)	12,5	12,04	75,5	Fsc=0,138 (p=0) Fst=0,245 (p=0) Fct=0,125 (p=0,0001)

*WG=Within Groups

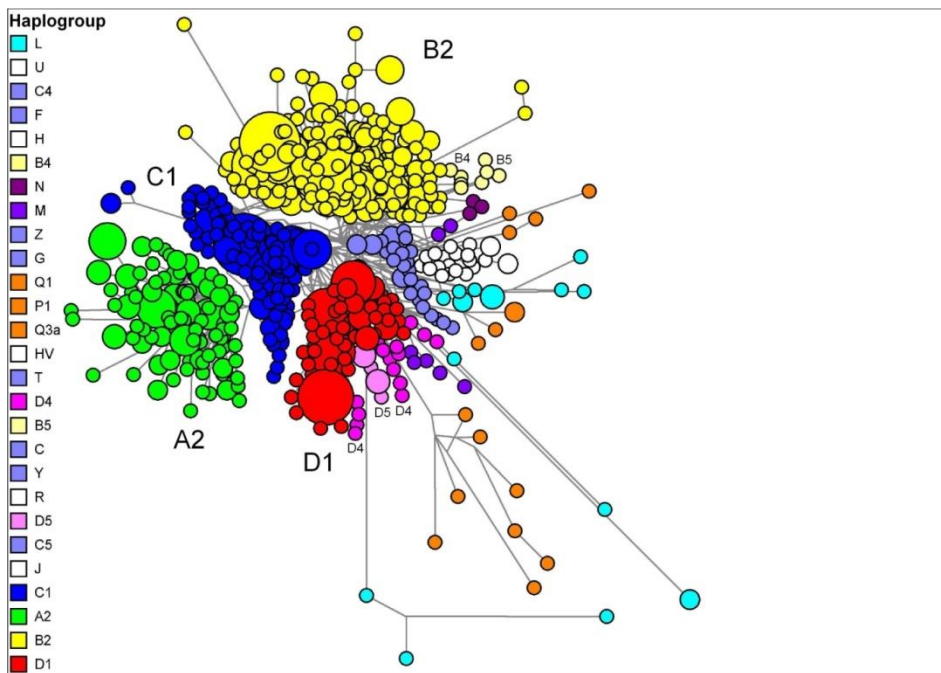


Figura 18: Rede MJ (median joining) dos haplogrupos ou sub-haplogrupos de mtDNA comparados no nível continental.

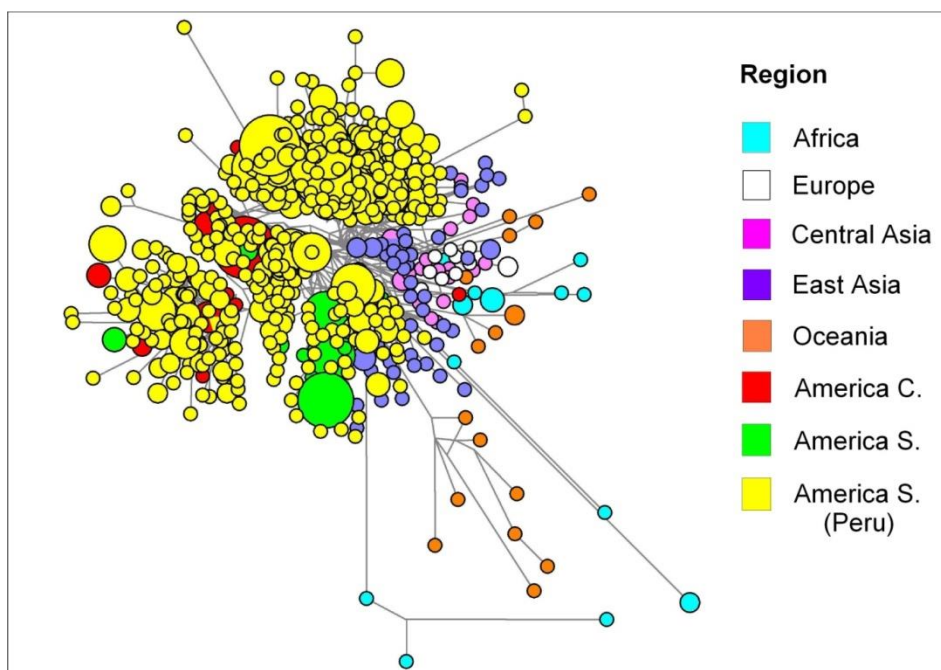


Figura 19: Rede MJ (median joining) dos haplogrupos ou sub-haplogrupos de mtDNA comparados no nível continental segundo por regiões geográficas.

MitoTool ou outro programa de conectá-los especificamente ao haplogrupo que pertencem (Tabela A5, Anexo).

Usando como referência-controle a população Cb (Cabanaconde) que contem os quatro sub-haplogrupos autóctones (A2, B2, C1 e D1), além dos outros sub-haplogrupos conhecidos, foi reconstruída uma rede median-joining (MJ) dessas linhagens através do programa Network (Figura 21), permitindo assim associar as amostras não determinadas por predição (nd). Segundo a rede MJ, as amostras “sugeridas por predição” como Hg R ou B (s38=nd14); Hg H/HV/U/B (s1=nd1); Hg H/R/B (s12=s37 ou nd6 e nd13), e o Hg H ou U (s39) estão mais relacionadas ao haplogrupo B (em macro-grupo N) do que a outros, como é esperado e consistente com PhyloTree. Em contraste, as amostras s6 (nd2); s7 (nd3); s9 (nd4); s10 (nd5); s13 (nd7); s16=s18=s20 (nd8-10), e s34=s35 (nd11 e nd12) estão associadas ao macro-grupo M. Em geral (à exceção de s14=HV), as demais amostras “preditas” nos agrupamentos estão de acordo com a árvore filogenética sugerida pelo PhyloTree. Os sub-haplogrupos J2, HV, grupo R/H são derivados do Hg R (que deriva do macro-grupo N), distribuídos em alta frequência na Europa. Por outro lado, o sub-haplogrupo G1 e o agrupamento MD derivam do macro-grupo M, sendo frequentes na Ásia. Os sub-haplogrupos L0, L1, L2 e L3 são linhagens africanas, sendo esta última (L3), a linhagem ancestral dos macro-grupos M e N que deram origem a todas as linhagens fora da África. Cabe mencionar que o sub-haplogrupo L3 é comum em toda África, com maior frequência na região sub-saariana e África Central (Soares *et al.*, 2012), de onde parecem ter vindo a maior parte dos escravos africanos à América. Isto explicaria sua distribuição (em baixa frequência) em locais como Lima e Lambayeque, em concordância com os registros históricos. Surpreendentemente, na localidade de Andoas foi detectado o haplogrupo L0 (n=1) que corresponde uma das linhagens basais da humanidade, provavelmente em decorrência de migração recente pós-colombiana.

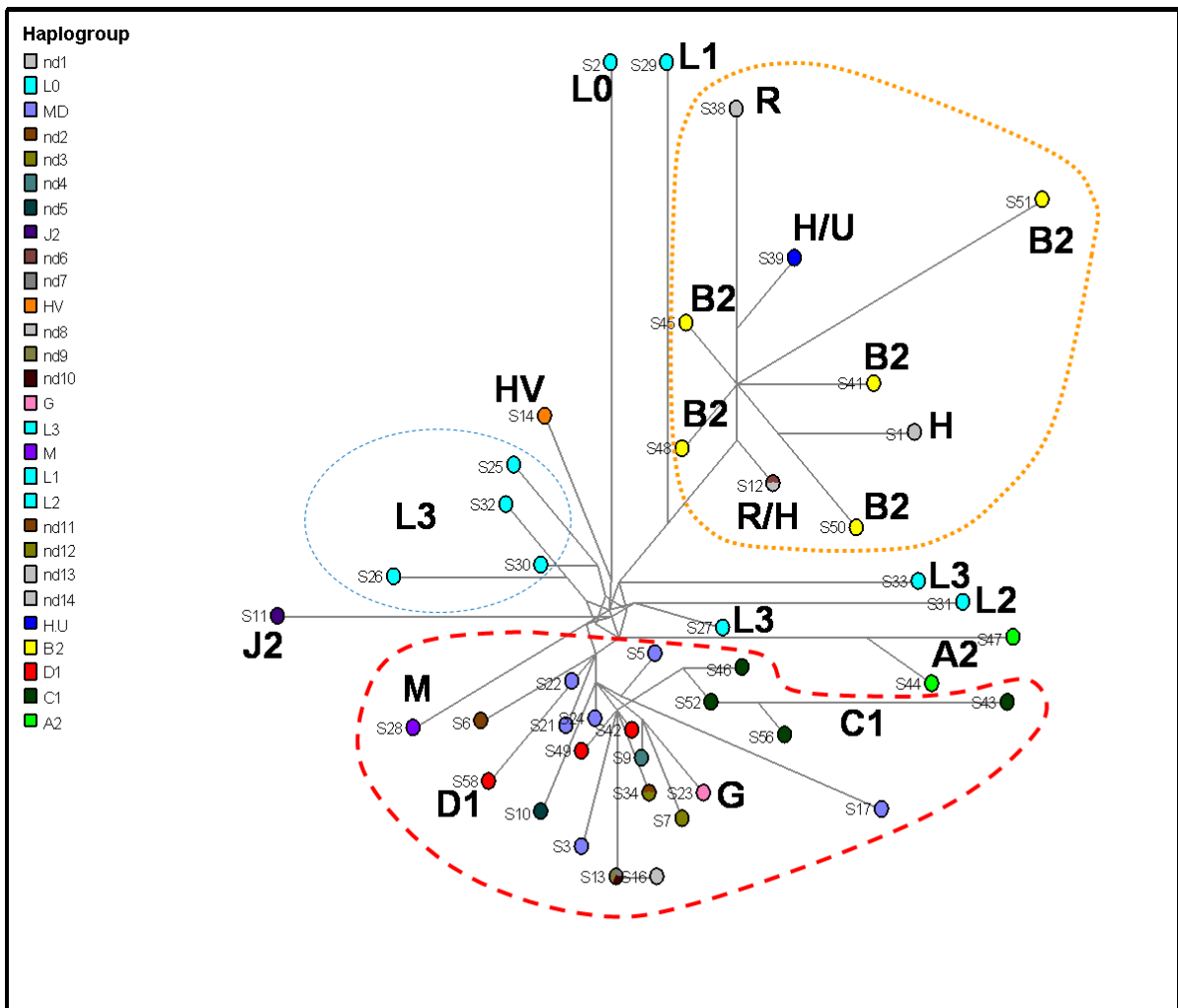


Figura 21: Rede MJ feita pelo programa Network. Os sub-haplogupos A2, B2, C1 e D1 correspondem à população de Cabanaconde (Cb). nd=não determinado por predição.

6.2- Sobre os SNPs e STRs do cromossomo Y

Em uma amostra de 464 homens das três regiões do Peru (Amazônia, Andes e Costa) foram determinadas as linhagens do cromossomo Y, usando 10 marcadores do tipo SNP da região não recombinante (NRY), segundo indicado na árvore filogenética (Figura A1, Anexo), correspondentes aos haplogrupos Q (Q1a3a=M3, Q1a3*=M346(xM3), Q*=M378 (xM346)), R-M207, F-M89 e E-M96. As frequências dos SNPs que determinam as linhagens masculinas são apresentadas na Tabela 10 onde se observa que o haplogrupo Q-M3 é prevalente.

Apesar da predominância da linhagem Q-M3 em quase todas as localidades analisadas, observa-se nas cidades de Cajamarca (Tabela 10) e San Marcos (próximo a Cajamarca), uma proporção da linhagem R-M207 (38,6%) maior do que Q-M3/M346 (24,6%), concordando com as crônicas que indicam que as primeiras “miscigenações” entre autóctones e europeus (homens) se deram nessa região. Corroborando isto, foi em Cajamarca que mulheres nativas foram entregues pelo Inca Atawualpa aos espanhóis como sinal de amizade e boas-vindas (comunicação pessoal feita pela historiadora espanhola – Prof. Berta Ares Queija- da Escuela de Estudios Hispano-Americanos-Espanha, durante conferência no dia 16/09/2010 na FAFICH/UFMG).

Considerando somente populações com 5 ou mais indivíduos amostrados (retiradas localidades de Tar, PI e PU), e perfazendo um total de 451 homens genotipados com SNPs (da Tabela 10), estas foram submetidas à análise por PCA (Figura 22). Para o plotagem dos dados populacionais com as frequências dos SNPs, estes foram divididos em quatro grupos (M207, M89, M96 e M3/M346). O panorama mostrou um gradiente de miscigenação que se apresenta muito similar ao quadro observado com 40 marcadores INDELS. Evidentemente as populações do Lago Titicaca (Taquile, Amantani, Uros, Anapia) apresentam-se como não miscigenadas em termos de linhagens paternas, já que não foram detectadas linhagens diferentes de Q nessas populações. Cabe mencionar que o Q-M378 foi excluído das análises, já que é um único indivíduo dessa linhagem revelada nas populações peruanas. Este perfil genético foi revelado também na Ásia Central. Segundo os dados dos 40 INDELS, seu genoma nuclear é de ascendência autóctone americano, de uma linhagem pré-colombiana.

Tabela 10: Frequências (absolutas) dos haplogrupos do cromossomo Y nas populações peruanas.

Local	n	Q-M346*	Q-M3	Q-M378	R-M207	F-M89*	E-M96
Amazônia							
Andoas	45	3	33	-	4	4	1
Iquitos	4	1	2	-	1	-	-
Pucallpa	6	-	5	-	1	-	-
Chachapoyas	9	1	5	-	2	1	-
Tarapoto	4	-	1	-	1	2	-
Andes							
Cajamarca	32	4	5	-	14	7	2
San Marcos	25	1	9	-	8	6	1
Chogo	16	1	11	-	-	2	2
Ocopon	14	3	6	-	5	-	-
Huancayo	25	-	12	-	6	6	1
Ayacucho	20	-	7	-	7	4	2
Andahuaylas	19	-	8	-	4	6	1
Kaquiabamba	9	1	4	-	2	-	2
Huarochari	15	1	9	-	3	2	-
Arequipa	15	-	10	-	3	-	2
Chivay	18	-	13	-	5	-	-
Yanque	10	-	8	-	-	2	-
Cabanaconde	19	-	12	1	2	3	1
Amantani	22	-	22	-	-	-	-
Taquile	36	-	36	-	-	-	-
Uros	16	-	16	-	-	-	-
Anapia	13	1	11	-	-	1	-
Puno	4	-	4	-	-	-	-
Costa							
Piura	4	1	1	-	1	1	-
Lambayeque	25	-	12	-	6	3	4
Lima	39	1	17	-	12	8	1
Total	464	19	279	1	87	58	20

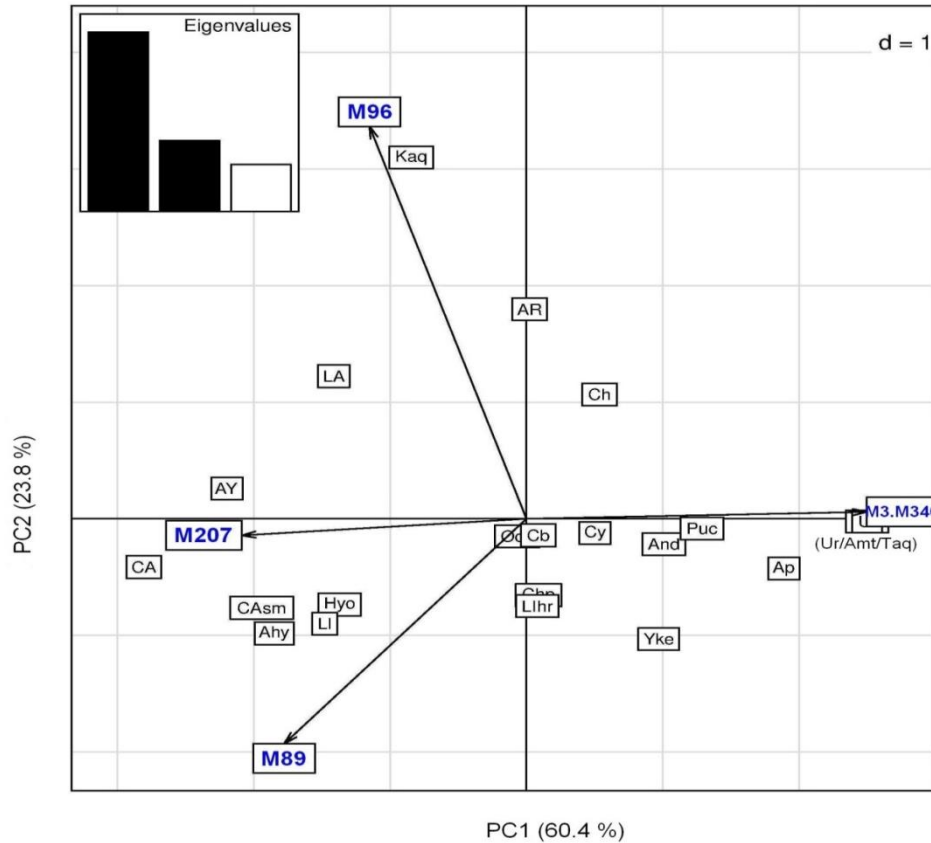


Figura 22: Plotagem da distribuição das frequências dos Y-SNPs (Tabela 10) através de componentes principais (PC).

6.2.1- Análises das linhagens paternas através do programa Network

Para as análises dos 17 Y-STRs genotipados em uma amostra de 371 indivíduos (n=58 da Amazônia, n=276 dos Andes, e n=37 da Costa), foi considerada a correspondência ao seu haplogrupo (SNP) respectivo (Q-M3, Q-M346, Q-M378, R-M207, F-M89 (ou F*) e E-M96) que previamente havia sido determinado por RT-PCR. Em algumas análises, a combinação de Y-SNPs e Y-STRs foram usadas nas redes feitas com Network (Figura 23), observando-se os agrupamentos dos haplótipos do haplogrupo correspondente, enquanto outras análises incluíram somente os haplótipos do haplogrupo Q-M3 (Figura 24) com o objetivo de buscar associações geográficas e históricas. Por exemplo, foi identificado o compartilhamento de haplótipos entre populações de Arequipa, Chivay e Cabanaconde (Figura 25-C), que estão localizadas na mesma região do país; assim como entre indivíduos de Taquile e Amantani (ilhas próximas), sendo que todas as demais apresentam uma distribuição muito heterogênea (Figura 25-A e B, da Amazônia e do centro dos Andes, respectivamente), que foi também corroborada pelo método de PCA (Figura 26). Este panorama indica que nas populações analisadas há uma grande variabilidade

intrapopulacional que pode ser devida a um fluxo gênico constante ao longo dos Andes, da Amazônia e da Costa do Peru. Alternativamente, é possível que parte da explicação seja devido à história pré e pós-colombiana sobre os deslocamentos forçados que aconteceram com alguns grupos de indivíduos de diferentes culturas, protoganizados pelos incas e depois pelos espanhóis.

A lista dos haplótipos únicos dos 17 Y-STRs do haplogrupo Q-M3, revelados nas populações peruanas, são apresentados na Tabela A6 do Anexo.

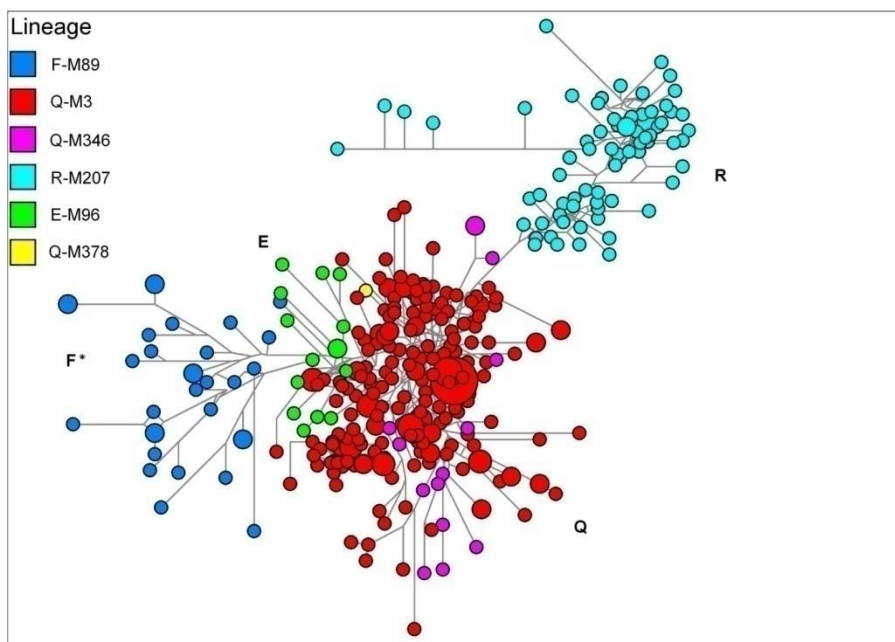


Figura 23: Rede MJ sobre as linhagens paternas e baseado nos SNPs e 17 Y-STRs (n=371).

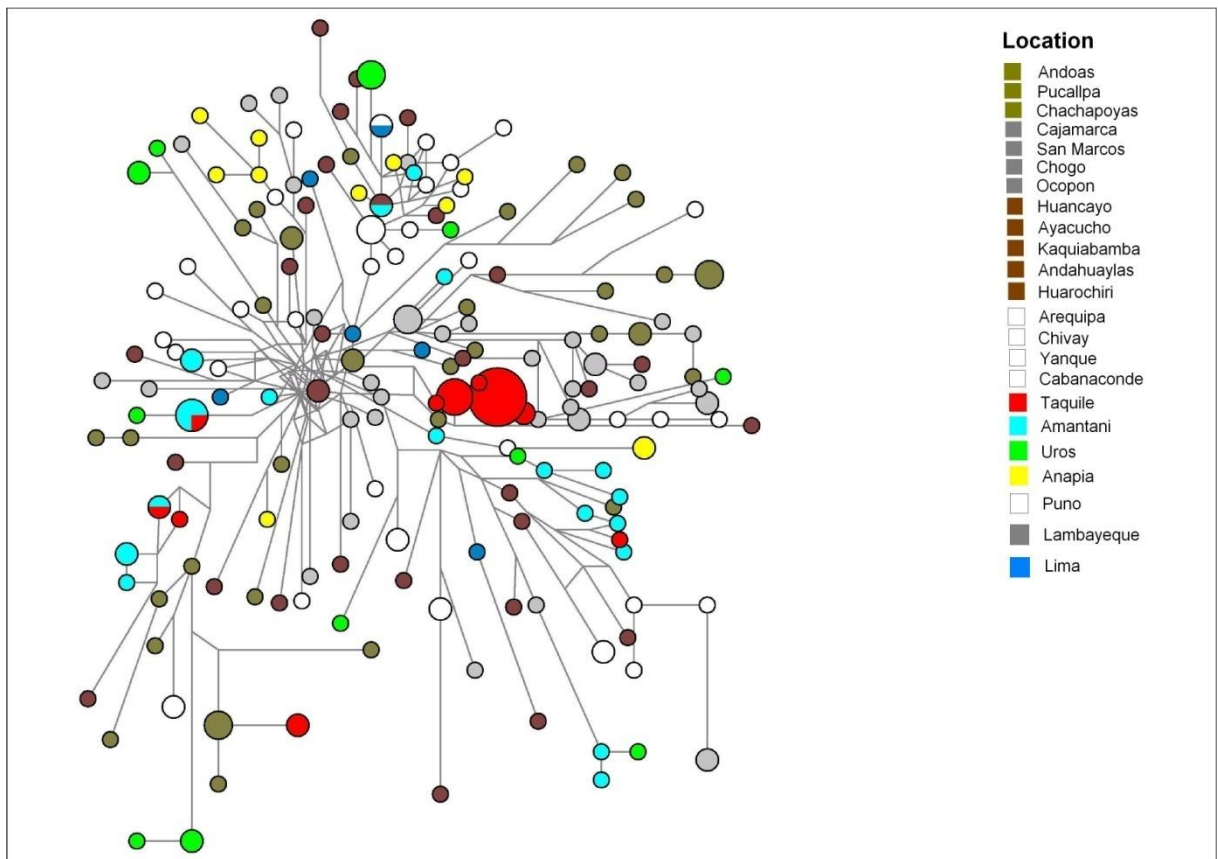


Figura 24: Rede MJ de 178 haplótipos dos 17 Y-STRs do Hg Q-M3 do cromossomo Y nas 23 populações peruanas (n=229). As cores representam as regiões: verde oliva (Amazônia); prata (norte dos Andes), marron (centro dos Andes), branco (sul dos Andes), vermelho (Taquile), celeste (Amantani), verde (Uros), amarelo (Anapia) e azul (Lima). Os círculos representam haplótipos com área proporcional à frequência de indivíduos.

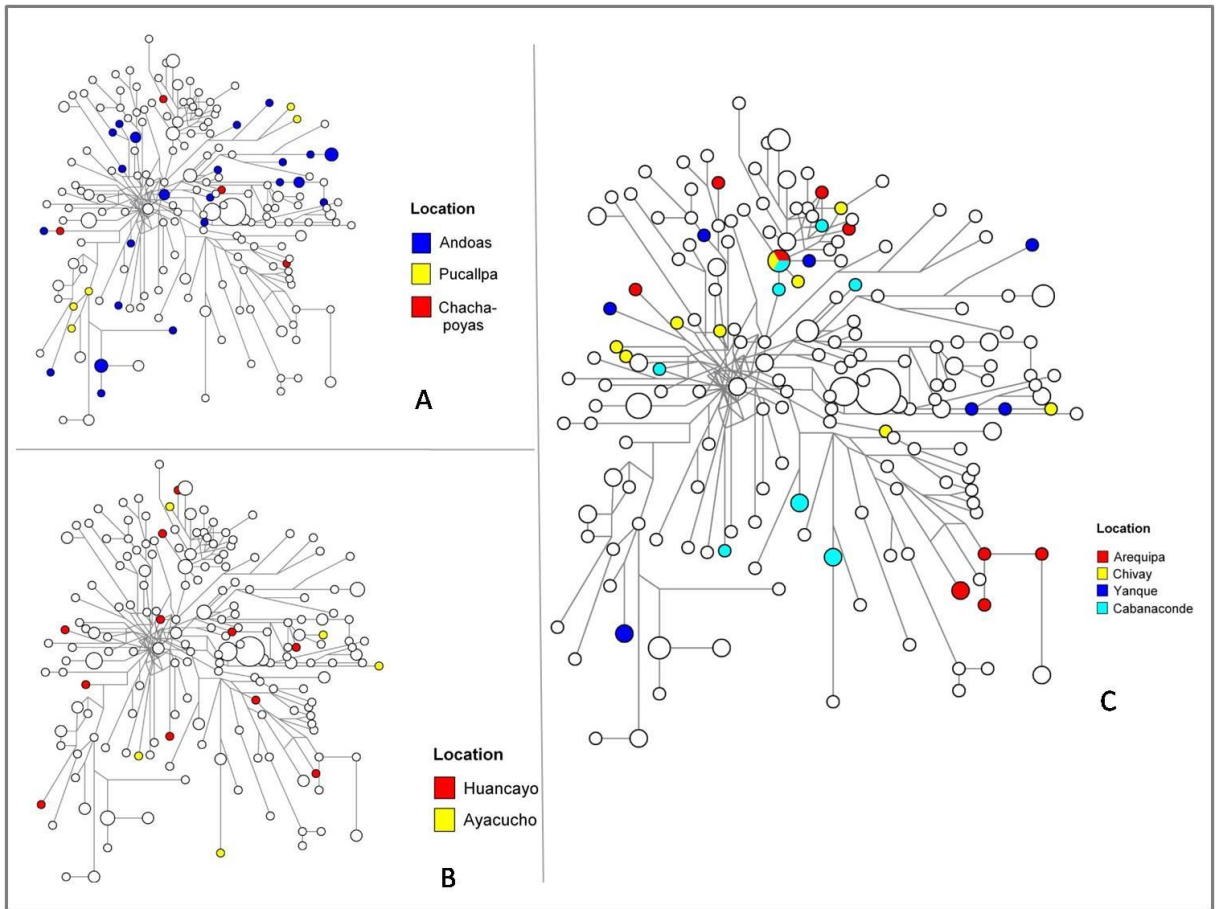


Figura 25: Rede MJ da distribuição dos haplótipos dos 17 Y-STR do subHq Q-M3 em algumas populações do Peru.

Principal Component Analysis

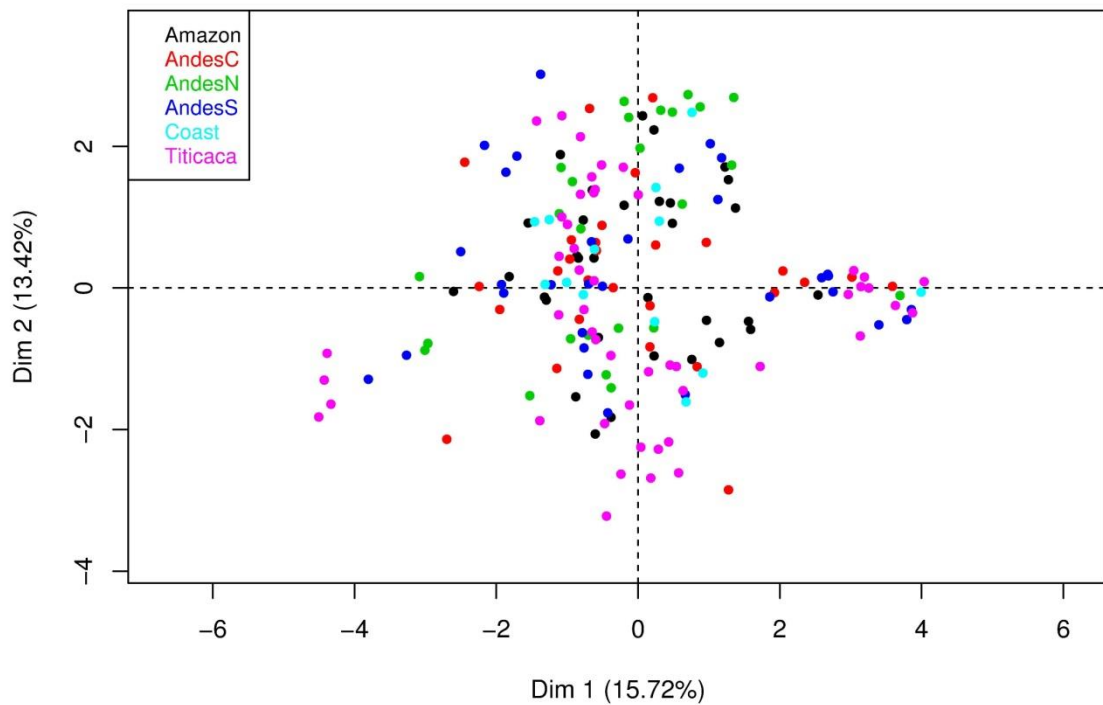


Figura 26: Gráfico bidimensional PCA sobre a distribuição dos 178 haplótipos dos 17 Y-STR da linhagem Q-M3 nas 23 populações peruanas (n=229) agrupadas por regiões geográficas. Cada cor representa indivíduos-haplótipos de sua correspondente população. AndesN (Andes do Norte); AndesC (Andes do Centro); AndesS (Andes do Sul). Titicaca (Taquile, Amantani, Uros, Anapia, Puno); Coast (Lima, Lambayeque); Amazon (Andoas, Pucallpa, Chachapoyas).

6.2.2- Variabilidade genética das populações peruanas segundo os 17 Y-STRs

Segundo as análises de AMOVA sobre os haplótipos de STR do Hg Q-M3 divididos em 3 regiões (Tabela 11), o valor de R_{st} (0,135) sugere que existe uma moderada diferenciação interpopulacional quando são comparadas as populações de todas as localidades ($n=229$; 23 locais, $n=39$ da Amazônia, $n=176$ dos Andes, $n=14$ da Costa), porém não se observa diferenças significativas entre essas as 3 regiões ($d=0$; -3,07%). O teste mostra também que há uma alta diversidade dentro das populações analisadas (86,53%).

Nas análises feitas separadamente por regiões, as populações dos Andes mostram maior diferenciação entre elas ($R_{st}=0,178$) quando comparado com as da Amazônia ($R_{st}=0,006$) e com as da Costa ($R_{st}=0,029$); com variação entre indivíduos relativamente menor nas populações dos Andes (82,24%).

No MDS baseado em distâncias genéticas (R_{st} de Reynolds) dos haplótipos de 17 Y-STRs do haplogrupo Q-M3 ($n=229$) (Figura 27), observa-se que a diferenciação entre Chogo, Taquile (andinos) e Pucallpa (amazônicos) é notória, quando comparada com as demais localidades. No entanto, as demais populações mostram relações mais próximas entre elas sugerindo um elevado fluxo gênico ao longo dos Andes.

Tabela 11: Resultado de AMOVA. Análises de 17 Y-STRs correspondentes ao haplogrupo Q-M3 nas 23 subpopulações peruanas ($n=229$, 3 regiões) segundo o agrupamento hierárquico por regiões ou separadas. Iterações=10.100

Região	Entre grupos (%)	Entre populações *WG (%)	Entre populações (%)	Dentro das populações (%)
3 regiões (Amazônia, Andes, Costa)	-3,07	16,54		86,53
Amazônia			0,56	99,44
Andes			17,76	82,24
Costa			2,93	97,07

*WG=Within Groups

- 1) 3 regiões: subpopulações=23 ($n=229$)
 $R_{sc}=0,16$ ($p=0$)
 $R_{st}=0,135$ ($p=0$)
 $R_{ct}=-0,031$ ($p=0,73$)
- 2) Amazônia: subpopulações=3 (And, Puc, Chp) ($n=39$)
 $R_{st}=0,006$ ($p=0,385$)

- 3) Andes: subpopulações=18 (CA,CAsm,Ch,Oco,Hyo,AY,Kaq, Ahy,Llhr,AR, Cy,Yke,Cb,Amt,Taq,Ur,Ap,PU) (n=176)
Rst= 0,178 (p=0)
- 4) Costa: subpopulações=2 (LI,LA) (n=14)
Rst=0,029 (p=0,272)

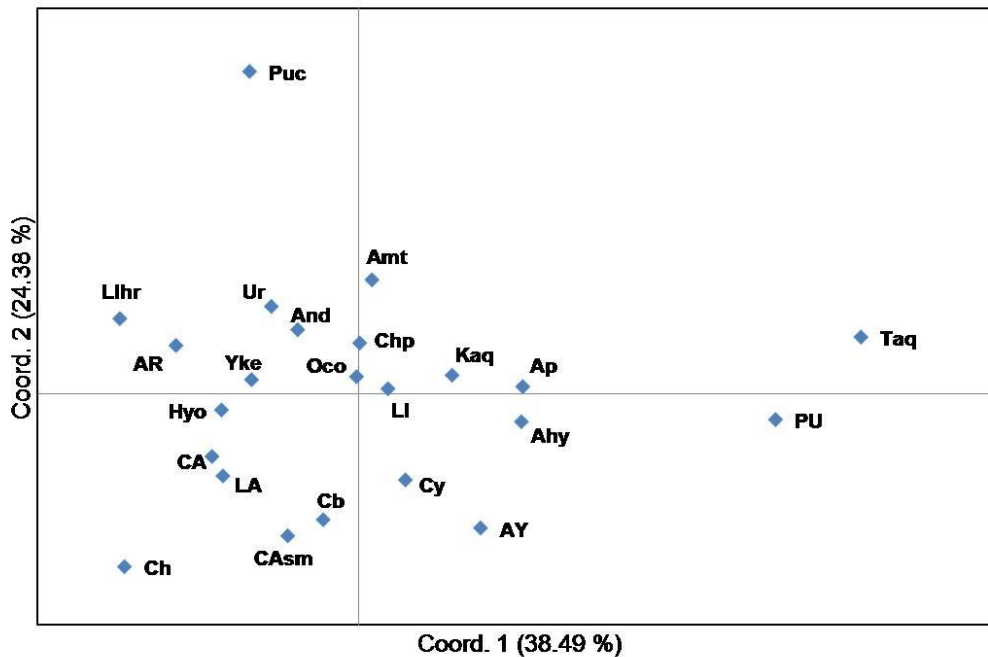


Figura 27: Plotagem MDS através da distância genética (Rst de Reynolds) entre as populações peruanas usando 17 Y-STRs.

6.3- A história genética uniparental dos Aimarás e Quéchuas do Lago Titicaca

Baseado nas crônicas e dados arqueológicos sobre o povoamento do Lago Titicaca, foi considerado analisar separadamente as populações (n=139) da região do lago: Taquile (n=48), Amantani (n=33), Uros (n=28), da Anapia (n=24) e de Puno (n=6), com o objetivo de revelar algumas peculiaridades de sua ancestralidade, já que as análises feitas através de 40 marcadores autossômicos (ver adiante) mostraram uma elevada proporção de ancestralidade autóctone americana comparada com as demais populações do Peru. Nestas análises são apresentadas as relações parentais e níveis de diferenciação entre as localidades (ilhas) de Taquile, Amantani, Los Uros (ilhas flutuantes), Amantani, Anapia, incluindo a alguns indivíduos da cidade de Puno.

Segundo as análises de AMOVA baseado em 43 haplótipos de mtDNA com 60 loci polimórficos (1.144 pb) há uma moderada diferenciação entre as 5 populações ($F_{st}=0,139$, $p=0$) quando foram consideradas como um só grupo. A variação dentro das populações é de 86,1%. Entretanto, dividindo em três grupos (Quéchuas, Aymarás e Puno), a AMOVA mostra falta de diferenciação entre os grupos (-1,66%), mas com moderada diferenciação entre todas as populações ($F_{st}=0,135$, $p=0$).

As análises (excluindo a localidade Puno) par a par dos valores F_{st} (pairwise difference) entre as quatro populações (ou análises independentes) mostraram os seguintes resultados:

Grupo 1: Taquile-Uros; $F_{st}=0,23$ ($p=0$); var. dentro das populações (77%)

Grupo 2: Taquile-Anapia: $F_{st}=0,06$ ($p=0,007$); var. dentro das populações (94,1%)

Grupo 3: Taquile-Amantani; $F_{st}=0,15$ ($p=0$); var. dentro das populações (85%)

Grupo 4: Amantani-Uros; $F_{st}=0,20$ ($p=0$); var. dentro das populações (79,6%)

Grupo 5: Amantani-Anapia; $F_{st}=0,04$ ($p=0,03$); var. dentro das populações (95,8%)

Grupo 6: Uros-Anapia; $F_{st}=0,153$ ($p=0,0002$); var. dentro das populações (84,7%)

Estes resultados indicam que: 1) os perfis genéticos de mtDNA das populações quéchua-falantes (Taquile e Amantani) são muito próximos aos de aimará-falantes (Anapia) tal como mostra suas distâncias genéticas (<6%); 2) os perfis de mtDNA da população dos Uros são mais próximos aos de Anapia, ilha na fronteira com Bolívia ($F_{st}=0,153$), do que em relação aos seus vizinhos de Taquile e Amantani ($F_{st}>0,20$); 3) as populações de Taquile e Amantani mostram diferenciação moderada entre elas ($F_{st}=0,15$) de modo similar que entre os Uros e Anapia ($F_{st}=0,153$); e 4) a maior diferenciação se observa entre os mtDNA de Taquile e os dos Uros ($F_{st}=0,23$). Este relação genética entre as quatro ilhas do Lago Titicaca é mostrada de modo similar quando se fez as comparações com todas demais populações peruanas ($n=700$) nas análises de AMOVA, tal como com outras populações de diferentes lugares da América, Eurásia, Oceania e África. As comparações intercontinentais mostraram que elas são muito próximas quanto comparadas as outras populações peruanas.

Os testes de neutralidade de Tajima (D) e de Fu (F_s) sob o modelo de sítios infinitos para cada população são mostrados na Tabela 12 (determinados por Arlequin), onde p =probabilidade. Em Anapia, os dois testes mostraram p significativos ($p<0,05$, assinalados com asterisco), sugerindo a expansão populacional.

Tabela 12: Testes de neutralidade entre as populações do Lago Titicaca.

Local	n	Teste de Tajima		Teste de Fu	
		D	p	F _s	p
Taquile	48	-0,63	0,32	-0,98	0,37
Amantani	33	-0,96	0,18	-1,27	0,31
Uros	28	0,73	0,80	6,3	0,98
Anapia	24	-1,64	0,03*	-6,2	0,007*
Puno	6	-0,33	0,44	-1,1	0,18

Com a tentativa de obter um panorama detalhado sobre conexões entre os haplótipos mitocondriais dos povoadores do Lago Titicaca, usamos duas abordagens com métodos similares de reconstrução filogenética, um baseado nos algoritmos Median Joining (Bandelt *et al.*, 1995) através do programa Network, e outro baseado no método de Máxima Parcimônia implementado no programa TCS (Clement *et al.*, 2000). No primeiro caso, o algoritmo implementado no programa Network reconstrói a “rede” usando as frequências haplotípicas as quais são transformadas previamente em um sistema binário (Reduced Median) ou caracteres (Median Joining) de vetores. Os vetores são conectados segundo a Máxima Parcimônia (MP) e ao final da reconstrução filogenética mostra a “árvore” mais provável ou árvore MP, onde os haplótipos são representados em forma de nós ou círculos e de tamanho proporcional as suas frequências. No segundo caso, o programa TCS reduz todas as sequências DNA a haplótipos únicos, e baseados nas frequências das diferenças haplotípicas (pairwise difference), o algoritmo calcula o “peso” com maior probabilidade (p=95%) para cada correlação haplotípica e os agrupa segundo suas conexões mais próximas, mostrando o haplótipo com maior peso que corresponderia possivelmente à avó “fundadora” na árvore genealógica (Clement *et al.*, 2000).

Na rede de haplótipos feita com o programa Network, foram utilizadas 139 amostras previamente analisadas por AMOVA, e para o TCS usamos 143 amostras (4 amostras adicionais).

A distribuição espacial de 42 haplótipos (n=139, 1.116 pb) se mostra através da rede MJ (Median Joining) na Figura 28. É notória a formação de conglomerados dos haplótipos em seus respectivos haplogrupos (Hg A2, B2, C1 e D1). As análises dos haplótipos mostram que os Hg A2, C1 e D1 formam um conglomerado e os haplótipos do Hg B2 outro grupo. Dentro do Hg B2, os indivíduos de Amantani, Taquile e os Uros compartilham dois haplótipos frequentes (n=18 e n=10). Apesar da maior distância geográfica entre Taquile

(em frente a Puno) e Anapia (fronteira com a Bolívia) observam-se haplótipos em comum (n=14). Em contraste, os Uros apresentam um haplótipo freqüente (n=18) somente observado neles, mas tem conexões com alguns indivíduos de Amantani, sugerindo seu “parentesco” mais próximo.

As variantes (SNPs) mitocondriais compartilhadas entre Taquile, Amantani, Uros, Anapia e Puno foi já analisado anteriormente em conjunto com as demais populações peruanas indicando ancestrais comuns.

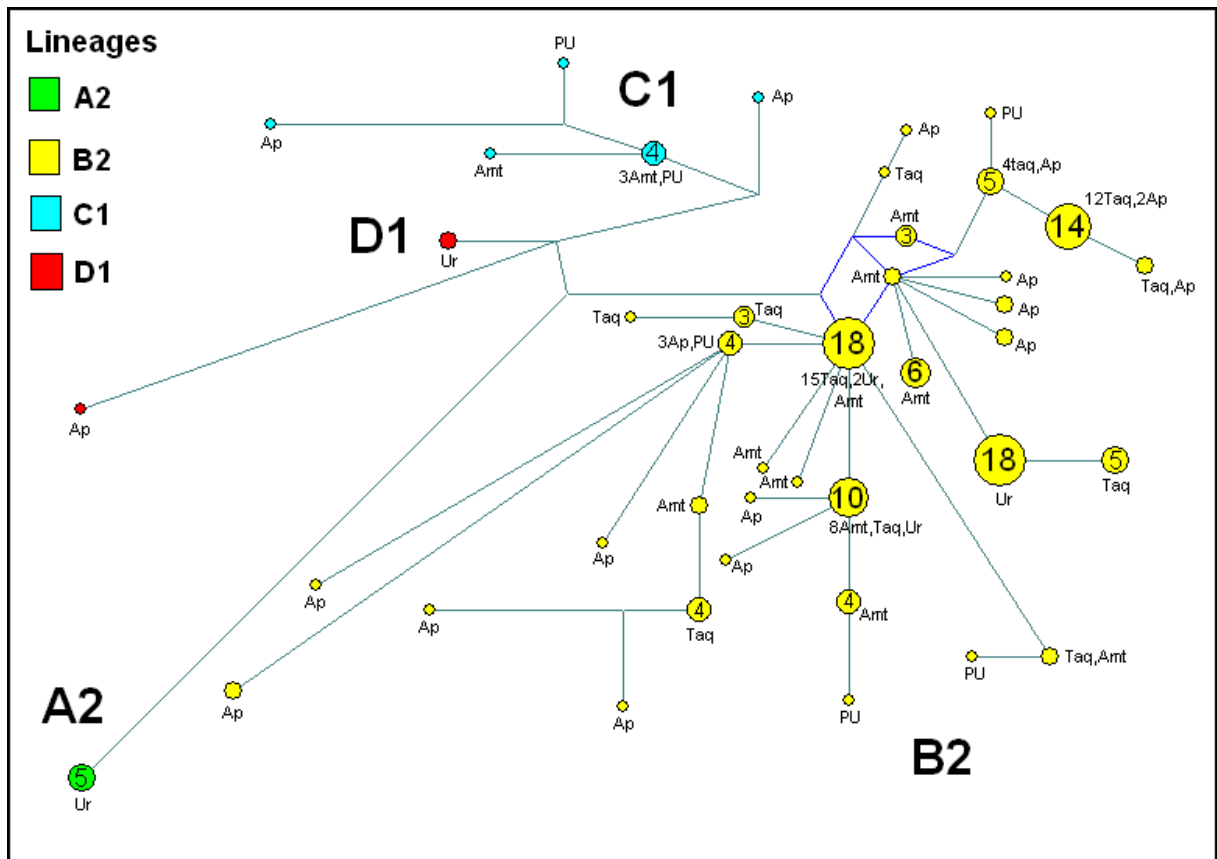


Figura 28: Rede MJ da distribuição de 42 haplótipos (65 sítios polimórficos) da Região Controle do mtDNA (1.116 pb) dos indivíduos de Taquile, Amantani, Uros, Anapia e Puno (n=139).

Por outro lado, a reconstrução genealógica de 45 haplótipos com 75 sítios polimórficos (Figura 29) dos Quéchuas e Aimarás (n=143) mostrou 10 subredes com o programa TCS, que correspondem aos conglomerados de haplótipos dentro dos haplogrupos a que pertencem (Figura 30). Segundo os cálculos probabilísticos do TCS, dentro do Hg B2 há uma maior densidade de haplótipos, sendo o haplótipo Hap_2 detentor de maior “peso” (representado pelo retângulo - haplotype with biggest outgroup weight (biggest probability)),

sugerido como um possível ancestral de grande parte dos demais haplótipos (18 indivíduos têm esta variante). Observa-se também que três haplótipos (Hap_8, Hap_42 e Hap_37) do Hg B2 são diferentes. Na rede TCS o Hap_25 (Hg A2), os Hap_30, Hap_34, Hap_44 (Hg C1), os Hap_23 e Hap_40 (Hg D1) e Hap_19, Hap_16, Hap_43 (Hg C1) conformam outros *clusters* apresentando-se como outras linhagens mitocondriais diferentes do Hg B2 confirmando sua divergência ancestral. Na filogenia gerada também se mostram as frequências (total de indivíduos-local-ao lado de cada conglomerado de haplótipos) dos haplótipos. A filogenia do Hg B2 é igual à gerada pelo programa Network. Da mesma forma observa-se que os de Taquile, Uros e Amantani compartilham os haplótipos mais frequentes que são o Hap_2 (n=18) e Hap_5 (n=10). Outro haplótipo frequente é o Hap_24 (n=18) que se apresentam somente nos Uros. Em geral, a maioria dos indivíduos do Hg B2 compartilha avós ancestrais à exceção do Hap_37 que possivelmente seja um subtipo do B2. Em contraste, devido à pouca quantidade de haplótipos dos Hg A2, C1 e D1, há ausência de conexões entre elas na configuração haplotípica, à exceção do Hap_19 (Hg C1) que mostra conexão entre os de Amantani e Puno. As inserções não foram consideradas nem as posições 16182-16183 no alinhamento inicial do mtDNA em relação à rCRS. A inclusão das posições 16182 e 16183 junto às demais variantes polimórficas não mudou a configuração de coancestralidade. O teste usando o sistema binário gerado através do programa mtDNA GeneSyn mostrou resultados iguais.

Por outro lado, pela linhagem paterna, na Figura 31 mostra-se a distribuição de 47 haplótipos STRs do cromossomo Y correspondentes ao Hg Q-M3 na rede MJ (Median Joining). Observa-se que alguns haplótipos são compartilhados entre os de Taquile e Amantani (de língua Quéchua). Entretanto, os haplótipos dos Uros e Anapia (ilha próxima à fronteira com a Bolívia, de língua Aimará) mostram maior heterogeneidade haplotípica em relação aos de Taquile (13 indivíduos com iguais haplótipos), mas com conexões próximas com os Quéchuas, e que tem concordância ao ser comparada às demais 19 supopulações peruanas (n=229) (Figura 32, indicado como *others*). Na ilha Taquile, o haplótipo frequente (n=13) sugere o efeito fundador similar aos dados obtidos pela linhagem materna. A análise componentes principais (PCA) mostrou as proximidades dos haplótipos/indivíduos no espaço bidimensional (Figura 33).

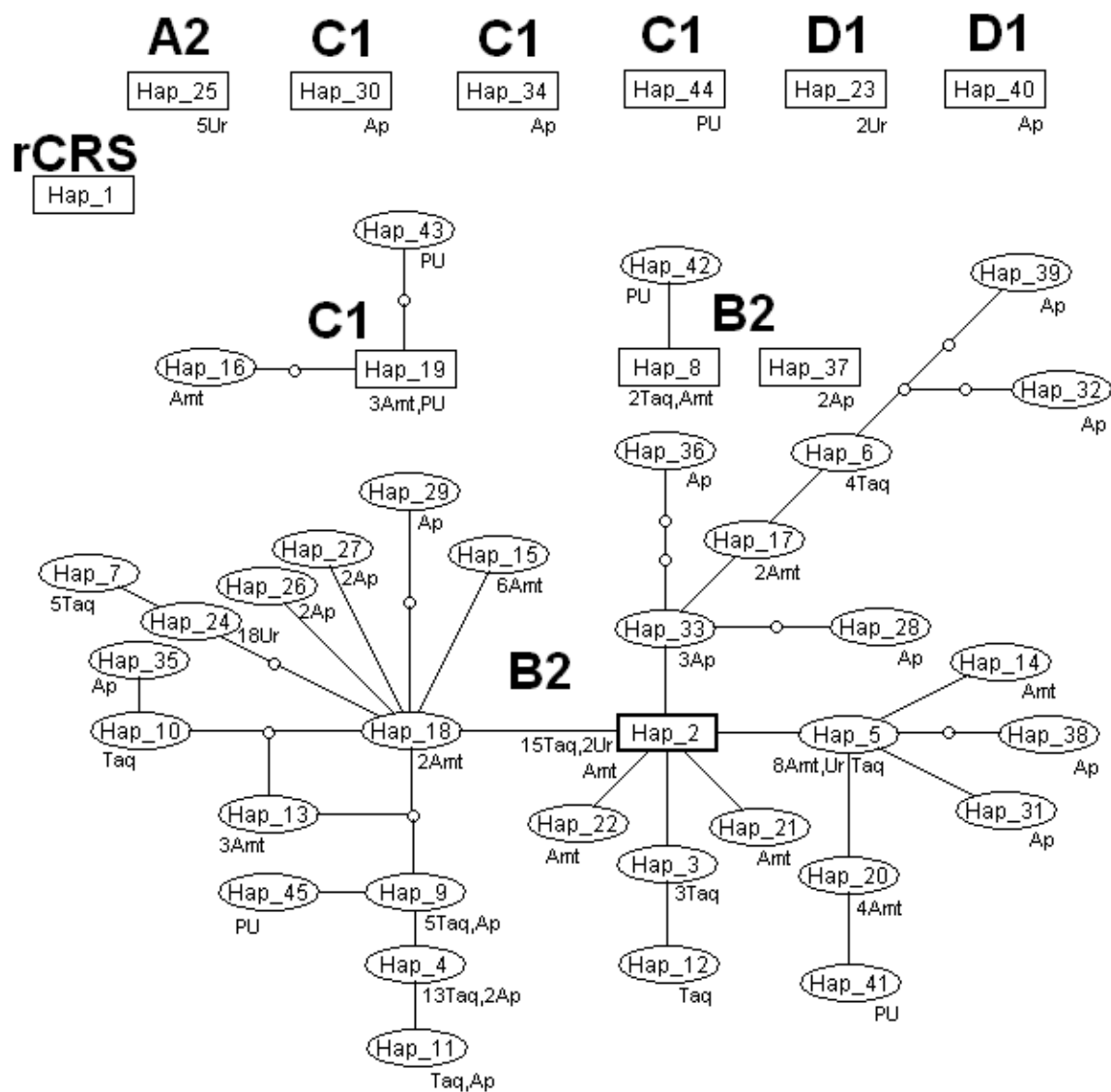


Figura 30: Rede de haplótipos gerada pelo programa TCS com haplótipos de indivíduos do Lago Titicaca. Hd (diversidade haplotípica)=0,95; haplótipos=45; S=75; tamanho das seqüências de mtDNA=1.114 pb.

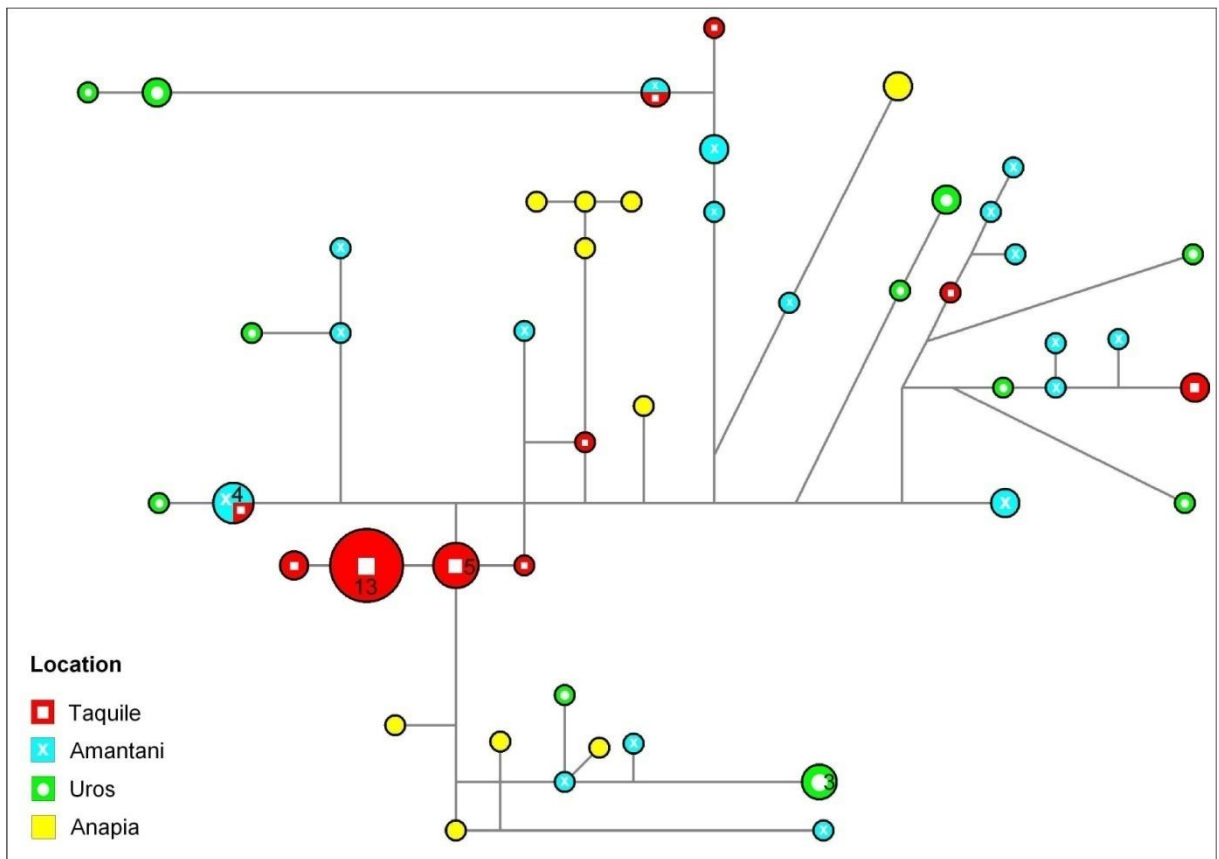


Figura 31: Rede MJ baseada na distribuição de 47 haplótipos dos 17 Y-STRs que correspondem à linhagem Q-M3 (Q1a3a) distribuída nas ilhas Taquile, Amantani, Uros e Anapia (n=76). Os círculos correspondem às proporções das frequências; a cor vermelha à Taquile; celeste à Amantani; verde à Uros, amarelo à Anapia.

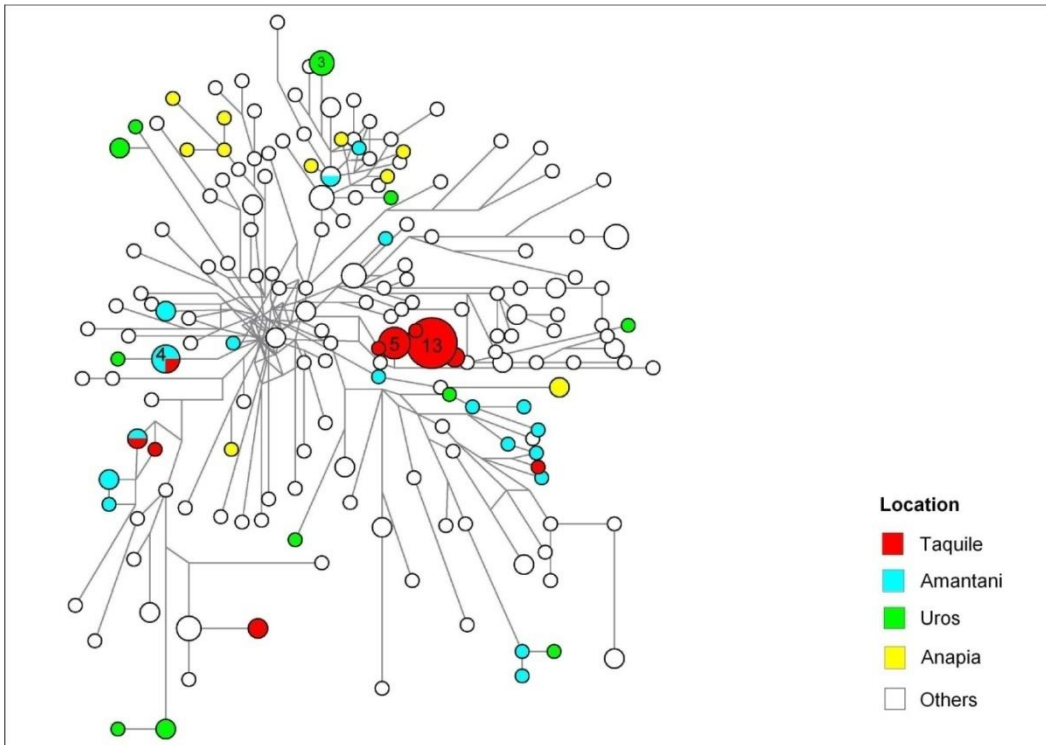


Figura 32: Rede MJ sobre a localização dos indivíduos do Lago Titicaca em relação aos demais indivíduos do Peru.

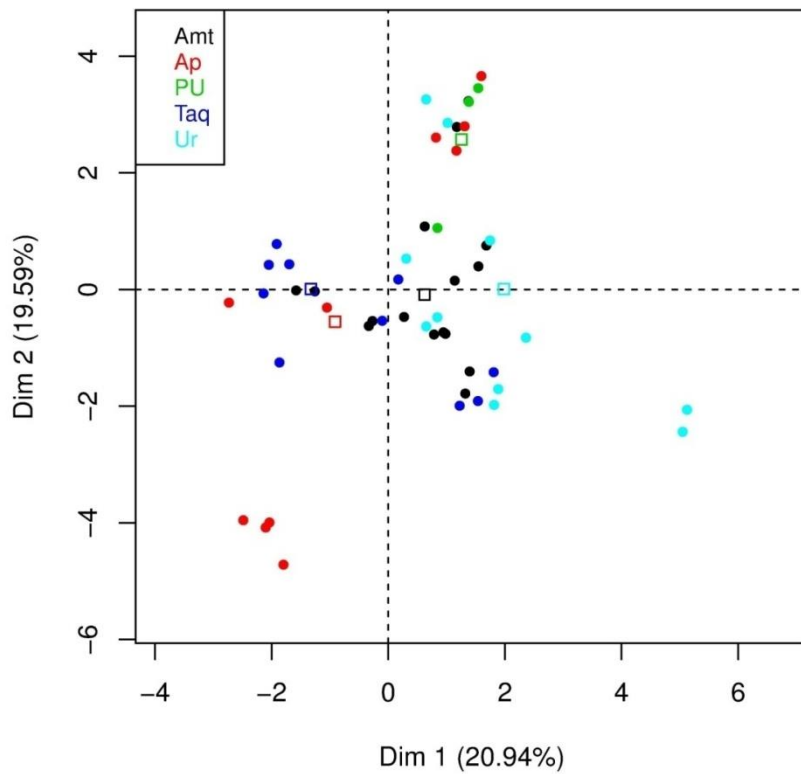


Figura 33: Gráfico bidimensional PCA dos haplótipos dos 17 Y-STRs do Haplogrupo Q-M3 das populações do Lago Titicaca (n=79, incluindo 3 de Puno).

6.4- Análises de genomas completos de mtDNA dos haplogrupos A2, B2, C1 e D1 encontrados no Peru

Para reconstruir a história demográfica das linhagens maternas através do tempo, foi usado o programa BEAST. Ele é baseado no algoritmo bayesiano MCMC e faz a reconstrução filogenética pelo método da coalescência e a *posteriori* calcula o tempo de divergência das linhagens, neste caso dos sub-haplogrupos A2, B2, C1 e D1, aproximando o tempo do ancestral comum recente ou TMRCA para as linhagens. Quanto maior a população, dois indivíduos quaisquer levarão mais tempo em traçar sua ancestralidade de volta ao passado até encontrar o ancestral comum de todos os indivíduos. A reunião de linhagens em uma ancestral é conhecida como evento de coalescência (Heled & Drummond, 2008).

Para simular os eventos de coalescência das linhagens A2, B2, C1 e D1 do genoma mtDNA dos povoadores do Lago Titicaca (Uros=Ur9_A2; Yanaque=YQ1_B2; Anapia=Ap22_C1; Anapia=Ap9_D1), nós usamos o método denominado E BSP. Este método assume uma função linear constante (piecewise linear) dos eventos demográficos já que a dinâmica populacional tem tendência a ser contínua ao longo das gerações (Heled & Drummond, 2008) em contraste ao modelo usado no BSP (Bayesian Skyline Plot). Através do E BSP infere-se o tamanho efetivo populacional em relação ao tempo usando as sequências DNA. Essa abordagem permitiu, por exemplo, detectar alguns eventos de gargalos populacionais (bottleneck) acontecidos nas populações autóctones americanas como efeito da colonização européia, especialmente no início do período da exploração espanhola na América do Sul conduzindo a uma redução do tamanho populacional no Peru (O'Fallon & Fehren-Schmitz, 2011; Hunefeldt, 2004) comparado ao período pré-colombiano quando se assume que nos Andes, por exemplo, sustentava entre 3-30 milhões de habitantes (Hunefeldt, 2004).

Para simular e comparar nossas quatro sequências de mtDNA completas (sub-haplogrupos A2, B2, C1 e D1) foram selecionadas 16 sequências-referências do banco de dados GenBank do site de National Center for Biotechnology Information (HGDP01412_Bantu-HgL3; HGDP01040_Pima-HgC1; GU733726.1_Mam4-HgR; GU733719.1_Mam10-HgM; GU733740.1_Mam60-HgN; EF153803.1_Chukchi1-HgA2a; EF153779.1_Buryat324-HgC1; EF153831.1_Yakut44-HgD2a; EF153778.1_Buryat312-HgN1a1; EF153782.1_Buryat365-HgM7c2a; GU903056.1_ROS2-HgB5b2; AF347001_Pima-HgB2; EU095200 (Y655)_Yanomama-HgA2; EU095230 (Y650)_Yanomama-HgC1; EU095210 (XAV04)_Xavante-HgB2; EU095236 (Zor23)_Zoro23-HgD1). Adicionalmente, foram usadas também sequências parciais de: EU095218 (KKT01)_Kayapo/Kriketun-HgB2; EU095212 (1876)_Quéchua-HgB2; DQ112872 (AM08)_Pima-HgD1.

Através dos programas MitoTool e mtPHYL foram determinados ou confirmados os haplogrupos / sub-haplogrupos em relação à rCRS.

Para modelar a simulação bayesiana foi usado o BEAUti com os seguintes parâmetros: HKY85 (Hasegawa-Kishino-Yano, 1985) como modelo de substituição nucleotídica; quatro categorias que descrevem a distribuição γ (gamma) com sítios invariáveis ($\gamma + 1$) para os sítios que se desenvolvem com diferentes taxas de mutação; UC LogNormal (uncorrelated LogNormal) como modelo de relógio molecular relaxado com a taxa de mutação de $1,665 \times 10^{-8}$ (uma substituição cada 3624 anos) para genomas mtDNA completos e a taxa de mutação $1,708 \times 10^{-8}$ para a região codificadora (uma substituição cada 3533 anos) (Soares *et al.*, 2009) e EBSP (Extended Bayesian Skyline Plot) como modelo de coalescência. As iterações foram feitas com MCMC=80.000.000; Burn-in=8.000.000; 5 corridas cada. Em outros casos com parâmetros MCMC=100.000.000; Burn-in=10.000.000; 5 corridas cada. A medida da “qualidade” da distribuição a *posteriori* de todos os parâmetros é o parâmetro denominado ESS (effective sample size) e somente se consideraram aquelas simulações que superaram o nível mínimo (>200) tal como é sugerido pelos autores de BEAST.

Os resultados das corridas MCMC feitas por BEAST foram visualizadas através dos programas TRACER, LogCombiner, e processadas através de TreeAnnotator com 0.5 de limite de probabilidade a *posteriori*. Este programa determina o valor denominado *highest Log clade credibility* que corresponde à árvore com a máxima soma de probabilidades a *posteriori* (similar ao *majority-rule consensus tree* no programa CONSENSE de Phylip) a qual foi visualizada pelo programa FigTree.

Os tempos de coalescência para cada haplogrupo vão ser diferentes dependendo de considerar o genoma completo de mtDNA ou somente a região codificadora. Os resultados obtidos após de um total de 80 “corridas” ou simulações MCMC são apresentados na Tabela 13. A inferência bayesiana através da região codificadora mtDNA sugere que os sub-haplogrupos A2, B2, C1 e D1 têm idades entre 16.000 e 23.000 anos e que coalescem ao ancestral comum recente (MRCA) *circa* 65.000 anos atrás. Estes valores são aproximados aos obtidos por Fagundes *et al.* (2008) nas simulações bayesianas e concordantes com os obtidos através do parâmetro estatístico *Rho* (Perego *et al.*, 2009; Achilli *et al.*, 2008; Volodko *et al.*, 2008; Fagundes *et al.*, 2008). Apesar de que as comparações foram feitas com estudos que usaram diferentes taxas de mutação, os resultados independentes apresentam idades similares para os sub-haplogrupos A2, B2, C1 e D1, e mostra a consistência de nossa abordagem. Em contraste, nossos resultados bayesianos através dos genomas completos apresentaram idades entre 24.000 e 41.000 anos atrás para os diferentes Hg, que coalescem ao ancestral comum recente (MRCA) *circa* 91.000 anos atrás. Estes valores sugerem que as idades dos sub-haplogrupos A2, B2, C1 e D1 (e de outras linhagens) são mais antigas do que sugere a região codificadora. Na Figura 34 apresenta-se a árvore filogenética dos 21 mtDNA completos segundo EBSP e usando o modelo HKY85 que é consistente com o modelo da árvore mtDNA humano sugerido pelo PhyloTree (van

Oven & Kayser, 2009). Na Figura 35 mostram-se resultados da coalescência das 21 sequências de mtDNA completas, e os valores são segundo a mediana (entre parêntese o desvio padrão) em anos. Na Figura 36 mostram-se os resultados da coalescência das 21 sequências mtDNA da região codificadora onde os valores são segundo a mediana (entre parêntese o desvio padrão) em anos. Para comparar a árvore obtida por BEAST foi feita outra reconstrução filogenética das 21 sequências mtDNA de forma independente através de PhyML (Figura 37) usando o modelo TN93, NNI (topologia de busca), BioNJ (tipo de árvore), e 4 categorias para a distribuição gamma. Observa-se que as duas árvores são praticamente iguais com elevado suporte estatístico (*bayesian posterior probabilities* e *bootstraps*) de modo tal que se complementam na inferência filogenética.

Tabela 13: Resultado de BEAST usando o EBSP (Extended Bayesian Skyline Plot) segundo o modelo de substituição nucleotídica HKY85. MCMC=100.000.000 de iterações para a corrida separadamente por haplogrupos, e MCMC=80.000.000 de iterações para a corrida de 21 sequências de mtDNA completas. Foram feitas mínimo 5 corridas para cada teste. HPD=Highest Posterior Density; T=Tempo; s.e.=standard error

Haplogrupo	n	Para cada haplogrupo				Incluído em 21 mtDNA genomas	
		T_média (anos)	s.e.	95%_HPD_lower	95%_HPD_upper	T_média (anos)	s.e. (anos)
Genoma Completo							
A2	3	22401,3	83,3	9133	37236	23761	21
B2	3	41911	599,5	22089	63086	40901	25,4
C1	4	31796	64,3	16993	48609	29803	20,8
D1	2	35883,1	897,2	14278	60100	30091	26,7
	21			74385	109000	90948,5	33,7 (TMRCA)
Região Codificadora							
A2	3	14724	54,4	4253	26870	16212,4	13,8
B2	3	24124	105,7	9345	41166	22974,8	23,4
C1	4	18427,6	139,3	6903	31322	17655,8	21,5
D1	3	20658,7	769,4	6745	35040	18966,8	31,6
	21			50844	79732	64924,3	21,7 (TMRCA)

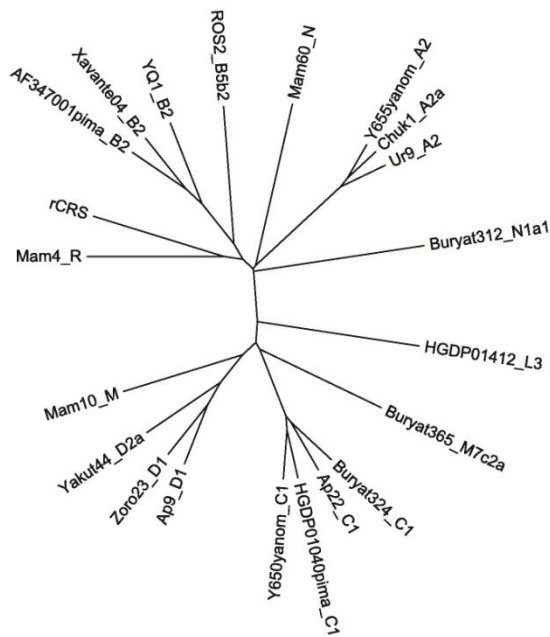


Figura 34: Árvore filogenética dos 21 mtDNA completos segundo EBSF e o modelo HKY85 que está de acordo com o modelo padrão sobre a árvore mtDNA humano sugerido pelo PhyloTree (van Oven & Kayser, 2009).

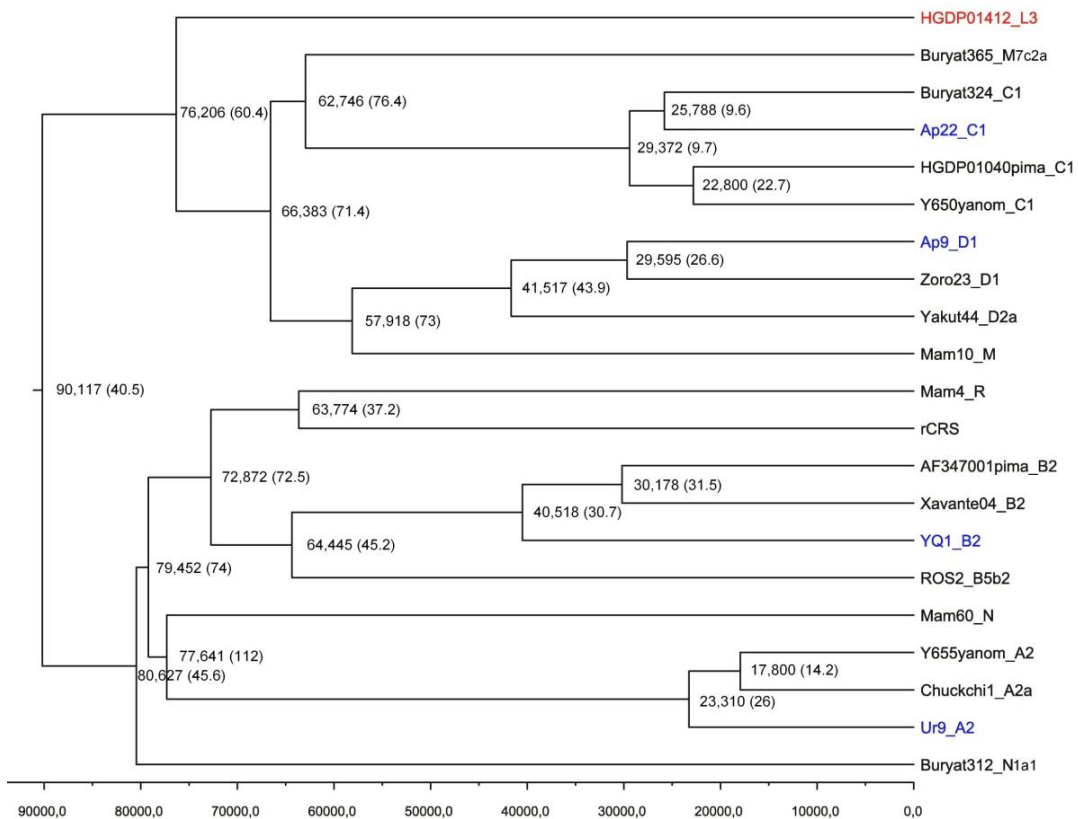


Figura 35: Resultados da coalescência das 21 seqüências mtDNA completas. Valores segundo a mediana (entre parêntese o desvio padrão) em anos.

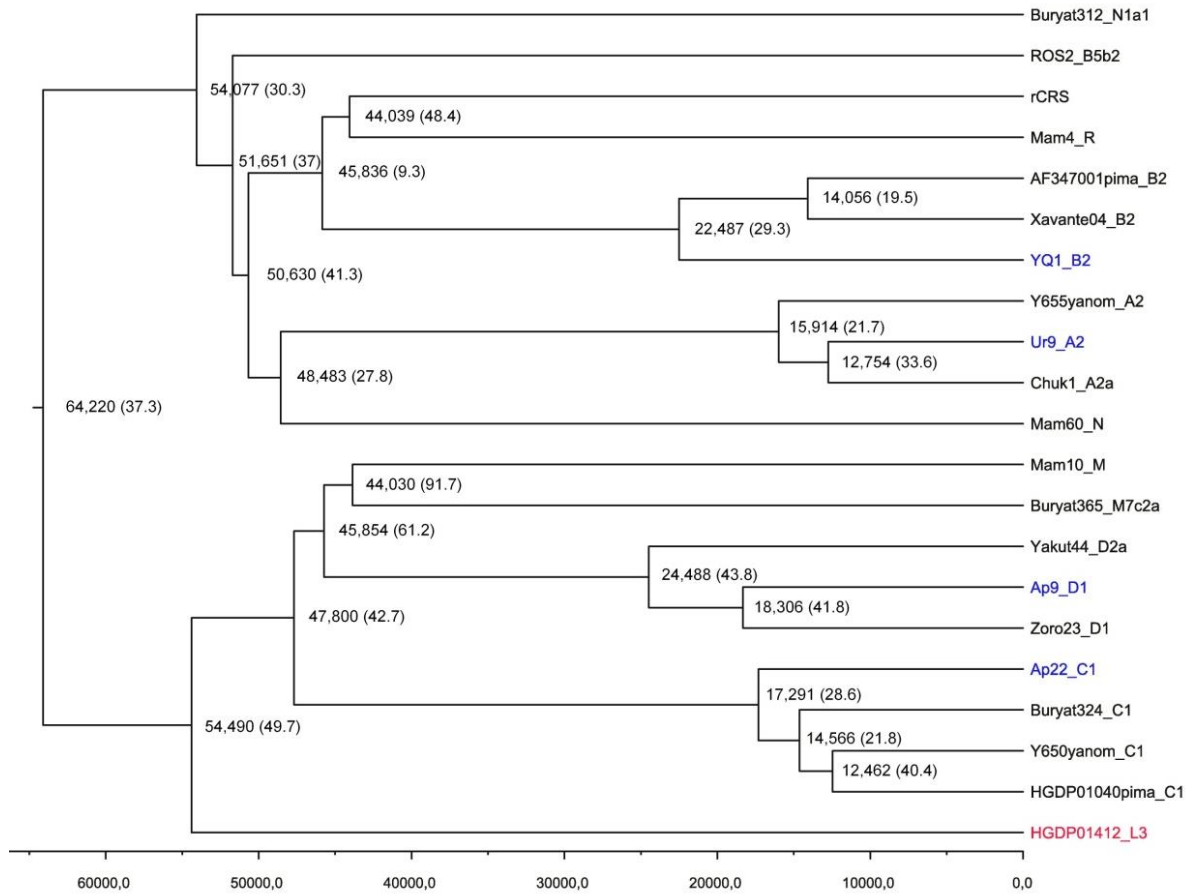


Figura 36: Resultados da coalescência das 21 seqüências mtDNA da região codificadora. Valores segundo a mediana (entre parêntese o desvio padrão) em anos.

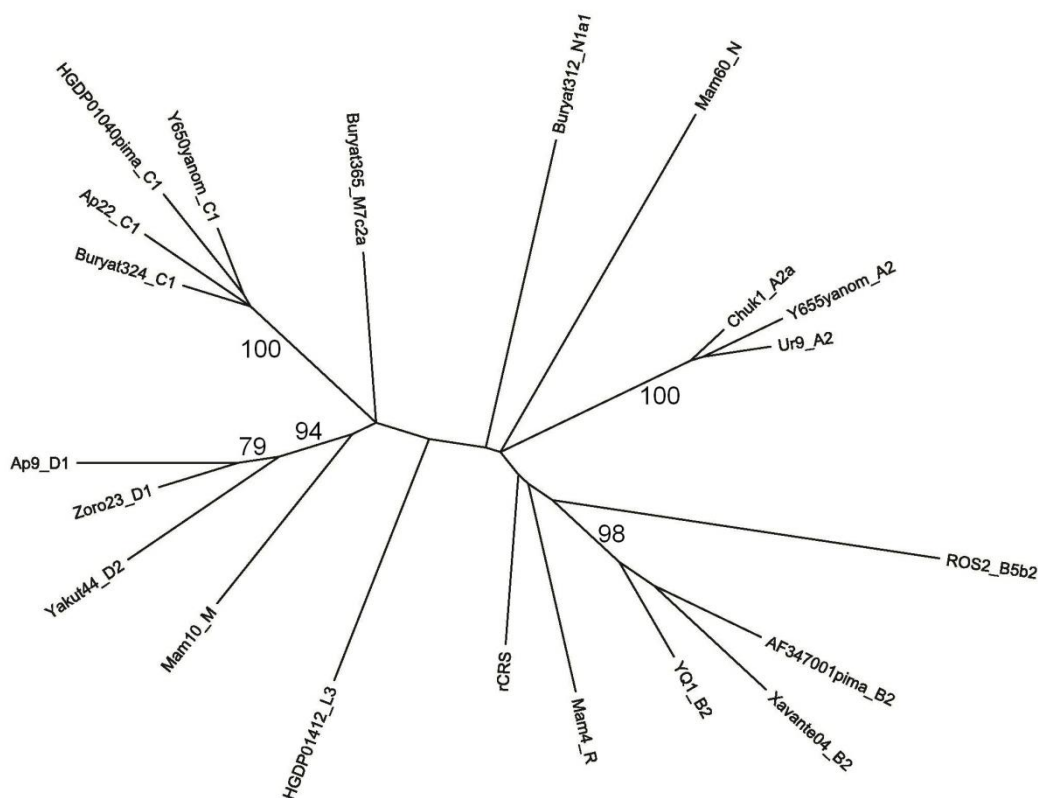


Figura 37: Reconstrução filogenética das 21 sequências de genomas de mtDNA completos usando o programa PhyML segundo o modelo TN93, aLRT, NNI, BioNJ, bootstrap=100. Bootstraps: HgA2=100%; HgB2=98%; HgC1=100%; HgD1=79%.

6.4.1- Determinação da divergência genética das linhagens A2, B2, C1 e D1 por mtPHYL

O algoritmo do programa mtPHYL permite reconstruir as árvores filogenéticas por máxima parcimônia (MP) e calcula os tempos de coalescência para cada agrupamento (ou dentro do grupo) segundo os tipos de haplogrupos tendo como referência o ancestral comum recente (MRCA) ao haplogrupo L3. As idades aproximadas estão baseadas no parâmetro estatístico *Rho* (ρ), o qual é definido como o número médio de diferenças ou polimorfismos observados para as linhagens reconstruídas a partir do nó ancestral de uma árvore genealógica (Cox, 2008). O nó ancestral é o TMRCA, mas pode ser um conjunto de sequências predefinidas, que assumindo uma taxa da mutação *a priori* (conhecida) é possível estimar suas divergências ao longo do tempo. O método assume que as mutações ao longo da árvore seguem uma função de distribuição de acordo a um modelo de “relógio molecular” evolutivo (estrito ou relaxado).

Para os testes foi usada a taxa de mutação de $1,665 \times 10^{-8}$ (uma substituição a cada 3624 anos) para genomas mtDNA completos e a taxa de mutação $1,708 \times 10^{-8}$ para a região codificadora (uma substituição a cada 3533 anos) (Soares *et al.*, 2009). Os resultados são apresentados na Tabela 14 e sugerem que os sub-haplogrupos A2, B2, C1 e D1 coalescem aproximadamente entre 19.000 e 33.000 anos atrás para genomas completos ou entre 11.000 e 20.000 anos atrás através da região codificadora. Estudos prévios também obtiveram resultados similares (Tabela 19) para a região codificadora apesar do uso de diferentes taxas de mutação.

A árvore reconstruída para os quatro genomas completos de mtDNA é mostrada na Figura 38, e para as mesmas 21 sequências que foram usadas anteriormente para as simulações bayesianas, na Figura 39, que é concordante à obtida pelo método bayesiano de BEAST, assim como à árvore gerada por máxima verossimilhança implementada no PhyML. Essas árvores são consistentes à sugerida pelo PhyloTree (van Oven & Kayser, 2009).

Tabela 14: Determinação do tempo de coalescência dos sub-haplogrupos mtDNA através do parâmetro estatístico *Rho* (ρ) no programa mtPHYL. σ =desvio padrão

Haplogrupo	n	<i>Rho</i> (ρ)	<i>Sigma</i> (σ)	T(anos)	ΔT (anos)
Genoma completo					
A2	3	5,33	1,41	19300	5100
B2	3	9	1,73	32600	6300
C1	4	6,5	1,32	23600	4800
D1	2	9	2,12	32600	7700
Região codificadora					
A2	3	3	1	10600	3533
B2	3	5	1,29	17700	4560
C1	4	5,33	1,33	18900	4700
D1	2	5,5	1,66	19500	5900

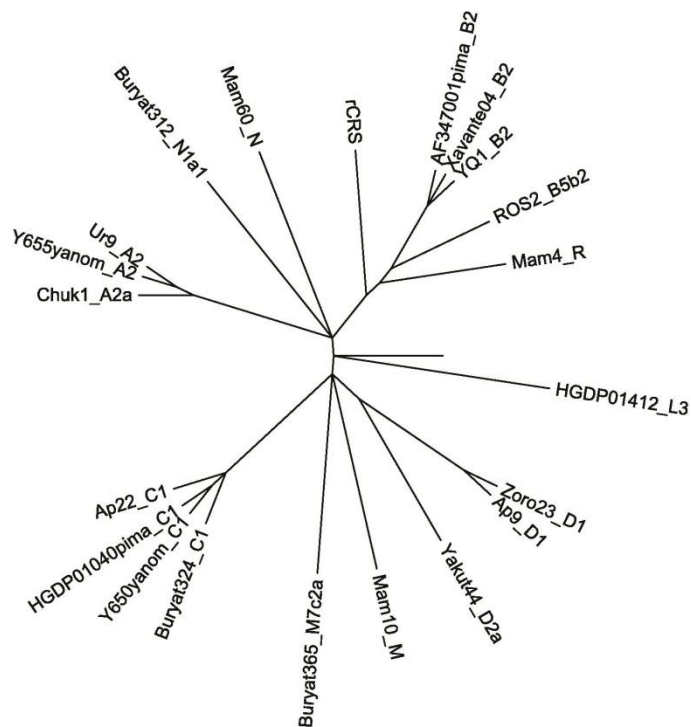


Figura 39: Filogenia das 21 sequências de genomas completos de mtDNA por Máxima Parcimônia através de mtPHYL.

6.5- Sobre os INDELS autossômicos

6.5.1- Determinação das proporções de ancestralidade das populações peruanas através de STRUCTURE

Para determinar a ancestralidade genômica e a subestrutura populacional peruana utilizou-se o método bayesiano Markov Chain Monte Carlo (MCMC) implementado no programa STRUCTURE. Este calcula as frequências alélicas dos *loci* e estima a probabilidade de pertencer a um determinado grupo (*membership's coefficient*), mostrando a partição em grupos ($K > 1$) segundo a correlação de ancestralidade genômica dessas frequências, sejam de indivíduos ou populações, assumindo que sua distribuição esteja de acordo a Hardy-Weinberg.

Baseado na história do Peru, de miscigenações das populações na época pós-colombiana, usou-se o modelo “*Admixture*”, com frequências alélicas correlacionadas, com parâmetros MCMC=200.000; burn-in=50.000; MCMC= 2×10^6 , burn-in=100.000. Primeiro foram testadas só as amostras do Peru que consistiram de 355 indivíduos dos Andes, 122 da Amazônia e 74 da Costa (10 vezes cada, até $K=10$). Posteriormente, foram adicionadas as amostras selecionadas do painel de HGDP-CEPH para determinar as proporções de correspondência ou ancestralidade a um grupo parental (Europa=161, Leste Asiático=251, América=108). O

procedimento foi feito independentemente das duas formas e mostrou os mesmos resultados de agrupamento em dois grupos.

Os *outputs Q-membership* dos indivíduos e das populações geradas por STRUCTURE foram processados através do programa STRUCTURE Harvester que inclui o método de Evanno (Earl & vonHoldt, 2011), que gerou o valor $\Delta K=2$, o valor modal dessa distribuição (Figura 40). A média dos valores Q se determinou via programa CLUMPP (Jakobsson & Rosenberg, 2007) e a representação do gráfico através do programa DISTRUCT (Figura 41).

Após obter os “*outputs*” ou resultados das 100 simulações foi feito um teste de comparação das 100 simulações das 551 amostras como também das populações em forma separada (geradas pelo programa STRUCTURE) através do programa Simco. Foi deduzido o coeficiente de similaridade (SimCoef) das probabilidades (10 vezes cada K) mostrando o SimCoef=0,974 (SE=0,002) em uma ordem de 0,941 a 0,988 para os indivíduos, e o SimCoef=0,978 (SE=0,002) em uma ordem de 0,945 a 0,996 para as populações. Quando foram incluídas as populações de referência foi obtido o SimCoef=0,995 (SE=0,0004) para as populações em conjunto. Procedeu-se da mesma forma em outros testes. Esses valores SimCoef indicam que os agrupamentos MCMC da partição $\Delta K=2$ feitos por STRUCTURE foram similares e muito elevados para as populações (98% e 99% respectivamente).

Para desvendar as proporções de ancestralidade das populações peruanas em forma completa nós usamos como referência todo o painel de 52 populações do HGDP-CEPH (n=1064) que foi dividido *a priori* em quatro macro-regiões (África; Eurásia (Europa, Oriente Médio, Ásia Central, Leste Asiático); Oceania e América). As simulações MCMC no STRUCTURE foram feitas desde K=2 até K=10. Os resultados K=3-6 são apresentados na Figura 42.

Os resultados mostraram que o agrupamento que melhor explica a estrutura dos dados é a partição K=5 (Figura 43) que coincide geograficamente com os lugares de origem das amostras. Baseado no marco referencial K=5 foram obtidas as proporções de ancestralidade das populações peruanas (Tabela 15). Para ilustrar graficamente o gradiente de miscigenação das populações peruanas apresenta-se a plotagem em 3D na Figura 44.

Em geral, nas populações peruanas há uma elevada proporção de ancestralidade autóctone americana (Q=0,538-0,965) e o grau de miscigenação das mesmas é heterogêneo. Nas localidades de San Marcos, Characato, Cajamarca, Chogo, Lambayeque e Lima apresentam-se com as mais elevadas proporções de correspondência à região da Eurásia (Europa_ME_CA) (31,2%; 24,4%; 20,5%; 14,6%; 14,5%; e 14,3%, respectivamente), sendo as proporções intermediárias em Lamas, Ayacucho e Huancayo (8,7%; 8,1%; e 6,1%, respectivamente). Além disso, essas mesmas populações mostram pequena proporção de ancestralidade da África (<3,4%). Também se observam diferentes graus de ancestralidade

associadas ao Leste Asiático em Chachapoyas (8,2%), Mollebaya (8%), e Iquitos (6%). Em Pucallpa apresenta-se a 5,2% relacionado ao Leste Asiático, 9% à Oceania, e 8% a Europa, Oriente Médio e Ásia Central.

A representação dos agrupamentos segundo as proporções de ancestralidade populacional com maior grau de miscigenação (San Marcos) e sem miscigenação detectada (Taquile) em relação às populações selecionadas (n=126 da África, n=161 da Europa, n=108 da América) do painel HGDP-CEPH são mostrados na Figura 45.

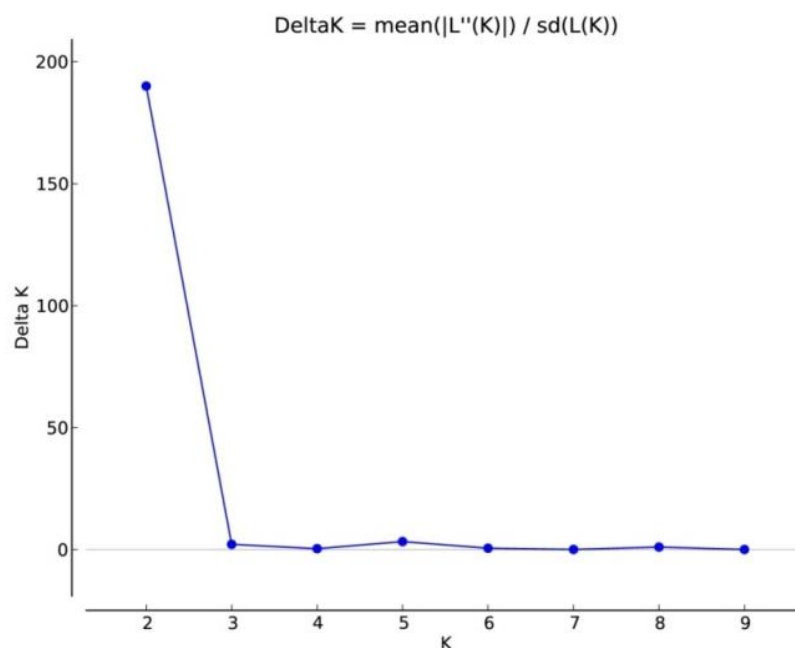


Figura 40: Resultado de $\Delta K=2$ (valor da moda=190.05) segundo o método de Evanno que foi obtida através do programa STRUCTURE HARVESTER. Os 551 indivíduos do Peru são agrupados em dois conglomerados. $\ln P(K)=-24225.1$; $SD=2.32$; $\ln'(K)=213.33$; $|\ln'(K)|=440.45$; 10 réplicas.

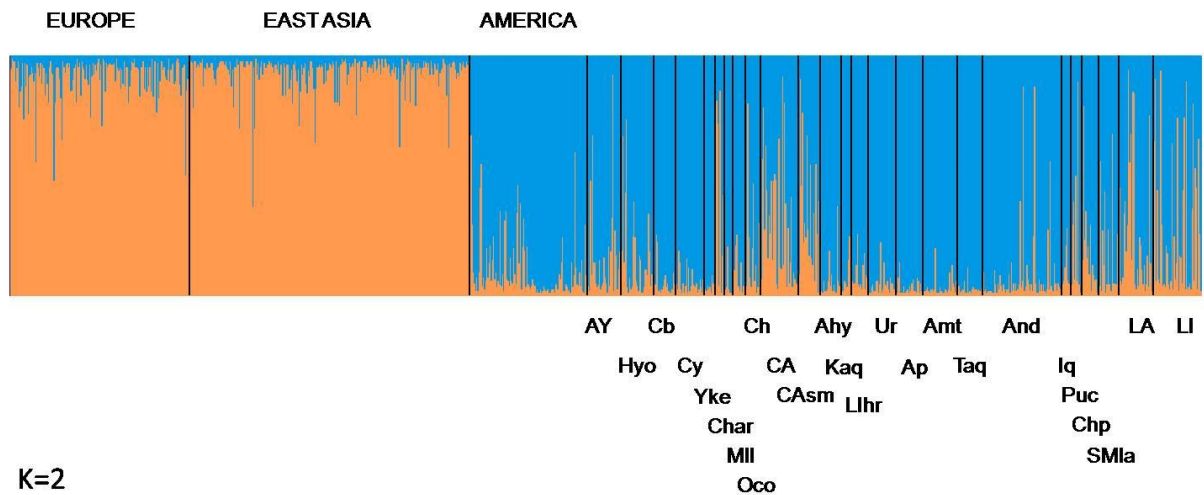


Figura 41: Plotagem por DISTRUCT da partição K=2 das populações peruanas (códigos em Material e Métodos) usando como referência as populações da Europa (n=161), Leste Asiático (n=251) e da América (n=108) do painel do HGDP-CEPH. As populações estão representadas por segmentos delimitados e os indivíduos pelas linhas verticais coradas segundo sua proporção de ancestralidade.

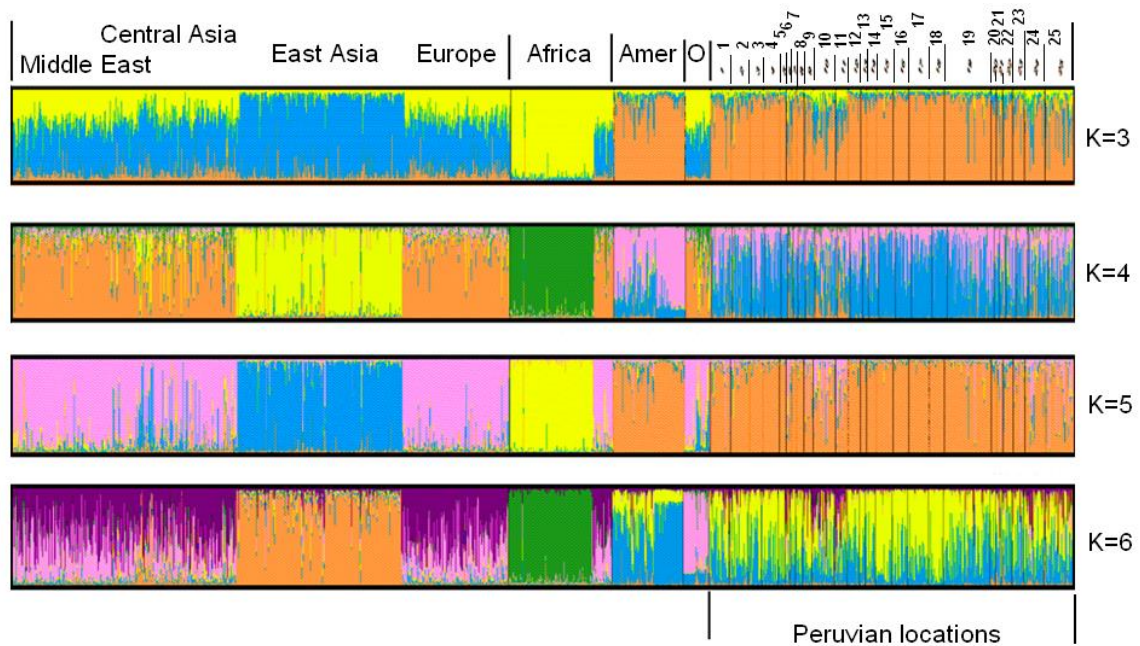


Figura 42: Plotagem por DISTRUCT das partições desde K=3 até K=6 obtido por STRUCTURE. As populações de referência do painel HGDP-CEPH foram divididas *a priori* em quatro macro-regiões (África; Eurásia (Europa, Oriente Médio, Ásia Central, Leste Asiático); Oceania; América). As populações estão representadas por segmentos delimitados e os indivíduos pelas linhas verticais coradas segundo sua proporção de ancestralidade. Amer=América; O=Oceania.

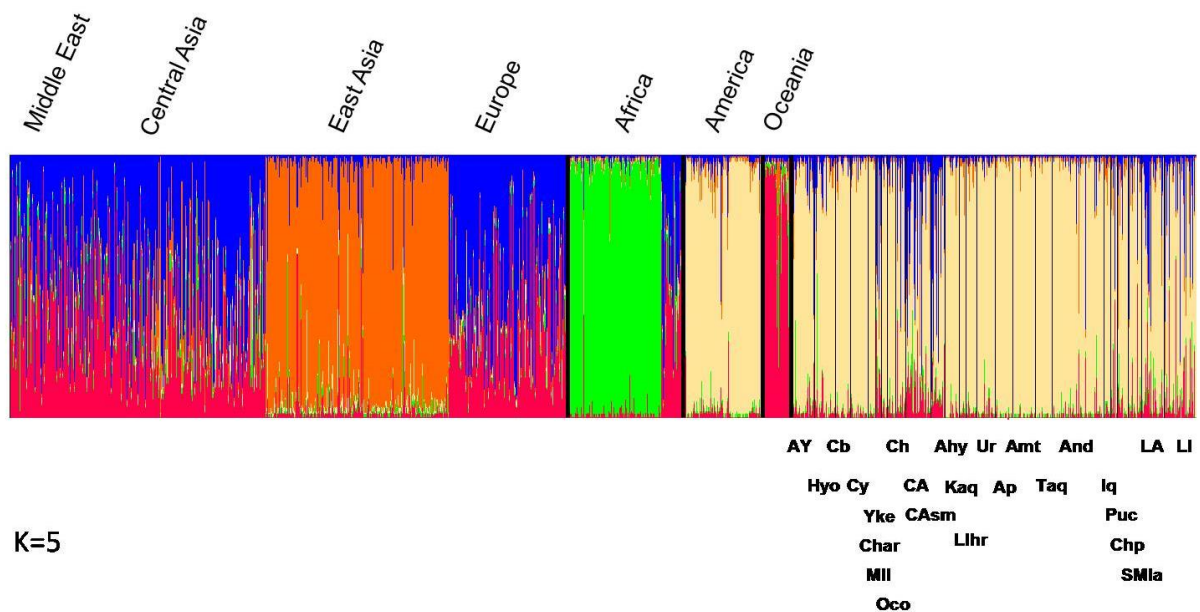


Figura 43: Plotagem por DISTRUCT da partição K=5 das populações peruanas (códigos em Material e Métodos) usando como referência as 52 populações do painel HGDP-CEPH. As populações estão representadas por segmentos delimitados e os indivíduos pelas linhas verticais coradas segundo sua proporção de ancestralidade.

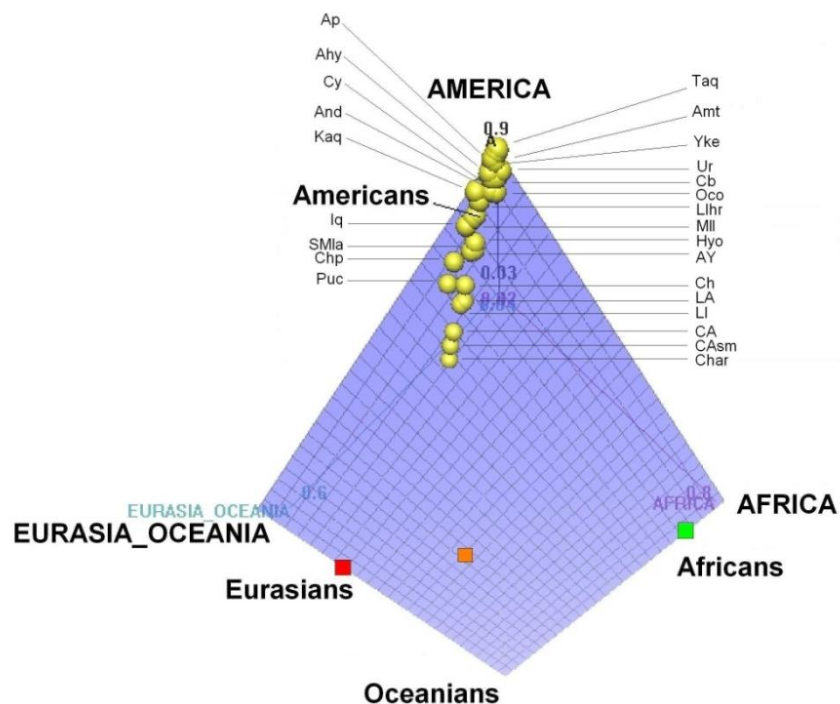


Figura 44: Plotagem 3D das populações peruanas segundo os valores Q (*ancestry proportion*) na partição K=3 (Eurásia_Oceania; África, e América). Referência: HGDP-CEPH: n=157 da África; n=760 da Eurásia; n=108 da América; n=39 da Oceania.

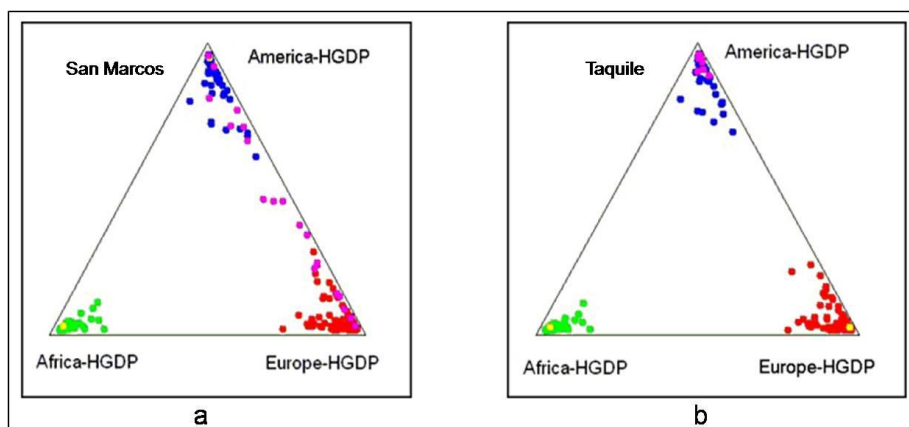


Figura 45: Plotagem por STRUCTURE da partição K=3 (África, Europa e América). a) miscigenação em San Marcos (pontos de cor rosa); b) não miscigenação em Taquile (pontos de cor rosa).

Tabela 15: Coeficientes (média) de proporção de ancestralidade (Q) segundo a partição K=5 obtida por STRUCTURE em relação às 52 populações de referência do painel HGDP-CEPH. Também são apresentados os resultados da partição K=2 para a comparação das proporções de ancestralidade entre dois grupos: not-America e America. ME=Middle East; CA=Central Asia.

Local	n	K=5					K=2	
		America	Oceania	East Asia	Europe_ME_CA	Africa	Not America	America
Andes								
Ayacucho	31	0,808	0,058	0,031	0,081	0,022	0,228	0,772
Huancayo	29	0,822	0,040	0,049	0,061	0,028	0,213	0,787
Cabanaconde	20	0,923	0,022	0,019	0,026	0,01	0,156	0,844
Chivay	25	0,930	0,019	0,019	0,023	0,009	0,151	0,849
Characato	8	0,548	0,136	0,041	0,244	0,031	0,346	0,654
Mollebaya	8	0,852	0,026	0,080	0,031	0,011	0,179	0,821
Ocopon	11	0,903	0,024	0,019	0,037	0,017	0,171	0,829
Chogo	14	0,725	0,068	0,033	0,146	0,028	0,267	0,733
Cajamarca	34	0,619	0,093	0,049	0,205	0,034	0,311	0,689
San Marcos	19	0,538	0,093	0,037	0,312	0,02	0,336	0,664
Andahuaylas	19	0,934	0,017	0,022	0,018	0,008	0,131	0,869
Kaquiabamba	9	0,918	0,018	0,041	0,016	0,007	0,149	0,851
Huarochari	15	0,911	0,020	0,037	0,02	0,012	0,175	0,825
Yanque	10	0,955	0,011	0,015	0,011	0,008	0,124	0,876
Uros	25	0,936	0,020	0,014	0,018	0,012	0,145	0,855
Anapia	24	0,958	0,009	0,017	0,009	0,007	0,112	0,888
Amantani	31	0,961	0,009	0,013	0,01	0,006	0,107	0,893
Taquile	23	0,965	0,009	0,011	0,009	0,006	0,105	0,895
Amazônia								
Andoas	71	0,903	0,026	0,022	0,037	0,012	0,153	0,847
Iquitos	8	0,882	0,022	0,060	0,018	0,018	0,196	0,804
Pucallpa	10	0,764	0,090	0,052	0,080	0,014	0,242	0,758
Chachapoyas	15	0,806	0,034	0,082	0,054	0,024	0,218	0,782
Lamas	18	0,800	0,060	0,035	0,087	0,018	0,222	0,778
Costa								
Lambayeque	31	0,710	0,067	0,046	0,145	0,033	0,280	0,720
Lima	43	0,690	0,099	0,044	0,143	0,023	0,283	0,717
Total	551	0,830	0,044	0,035	0,074	0,017	0,200	0,800

6.5.2- Teste de ancestralidade das populações peruanas através do método Mínimos Quadrados

Com o objetivo de comparar nossos resultados prévios de K=2 obtidos pelo método Bayesiano de STRUCTURE, foi usado outro método conhecido como WLS (weighted-least square), o qual está implementado no programa ADMIX (www.genetica.fmed.edu.uy/software.htm). A abordagem está baseada nas probabilidades de identidade genética e considera como matriz as frequências alélicas dos diferentes *loci* para estimar a proporção de miscigenação ou ancestralidade entre as populações (Chakraborty *et al.*, 1992). A comparação é feita em relação às populações conhecidas *a priori* como “parentais”. Para isso foi usado como referências as mesmas populações da Europa e América que foram usadas para o STRUCTURE (n=161 e n=108 respectivamente). Os resultados dos testes mostraram valores de miscigenação similares aos obtidos por STRUCTURE como é mostrado na Tabela 16.

Tabela 16: Comparação dos coeficientes ou proporções de ancestralidade em agrupamento K=2 obtido pelos programas STRUCTURE e ADMIX.

Local	por STRUCTURE (K=2)		por ADMIX	
	América	Europa	América	Europa
Ayacucho	0,778	0,222	0,787	0,213
Huancayo	0,798	0,202	0,793	0,207
Cabanaconde	0,887	0,113	0,916	0,085
Chivay	0,910	0,090	0,901	0,099
Yanque	0,945	0,055	0,948	0,052
Characato	0,539	0,461	0,594	0,406
Mollebaya	0,847	0,153	0,929	0,071
Ocopon	0,861	0,139	0,853	0,147
Chogo	0,711	0,289	0,692	0,308
Cajamarca	0,617	0,383	0,661	0,339
San Marcos	0,521	0,479	0,576	0,424
Andahuaylas	0,911	0,089	0,958	0,042
Kaquiabamba	0,918	0,082	0,943	0,057
Huachochiri	0,905	0,095	0,856	0,144
Uros	0,909	0,091	0,893	0,107
Anapia	0,955	0,045	0,962	0,038
Amantani	0,953	0,047	0,975	0,025

Taquile	0,956	0,044	0,945	0,056
Andoas	0,886	0,114	0,942	0,059
Iquitos	0,904	0,096	0,914	0,086
Pucallpa	0,762	0,238	0,770	0,230
Chachapoyas	0,802	0,198	0,840	0,160
Lamas	0,788	0,212	0,807	0,193
Lambayeque	0,696	0,304	0,688	0,312
Lima	0,675	0,325	0,711	0,289

6.5.3- Agrupamento das populações peruanas segundo BAPS

Para contrastar o desempenho do algoritmo MCMC do STRUCTURE, foi realizada uma análise da miscigenação das populações peruanas usando o BAPS que tem implementado um modelo estocástico MCMC diferente.

Os testes foram feitos a níveis de agrupamento individuais e populações usando *a priori* o mesmo valor do número de populações (K=25) com 10 réplicas. As iterações automatizadas obtiveram K=2 como o valor ótimo da distribuição *a posteriori*, sendo:

Cluster 1: Cabanaconde, Chivay, Yanque, Mollebaya, Ocopon, Andahuaylas, Kaquiabamba, Huarochiri, Uros, Anapia, Amantani, Taquile, Andoas, Iquitos.

Cluster 2: Ayacucho, Huancayo, Characato, Chogo, Cajamarca, San Marcos, Pucallpa, Chachapoyas, Lamas, Lambayeque, Lima.

Os valores de Log (marginal likelihood) = -24471.4867 e da probabilidade = 0.986.

Esta análise convergiu à mesma divisão (K=2) que foi obtida pelo programa STRUCTURE. Além disso, o modelo da miscigenação usando as amostras de “referência” do HGDP-CEPH e agrupando *a priori* K=4 (Figura 46) mostrou resultados similares aos obtidos com o modelo de “não miscigenação, frequências alélicas não correlacionadas” (*a priori* K=5), testado no STRUCTURE (Figura 47).

K=4

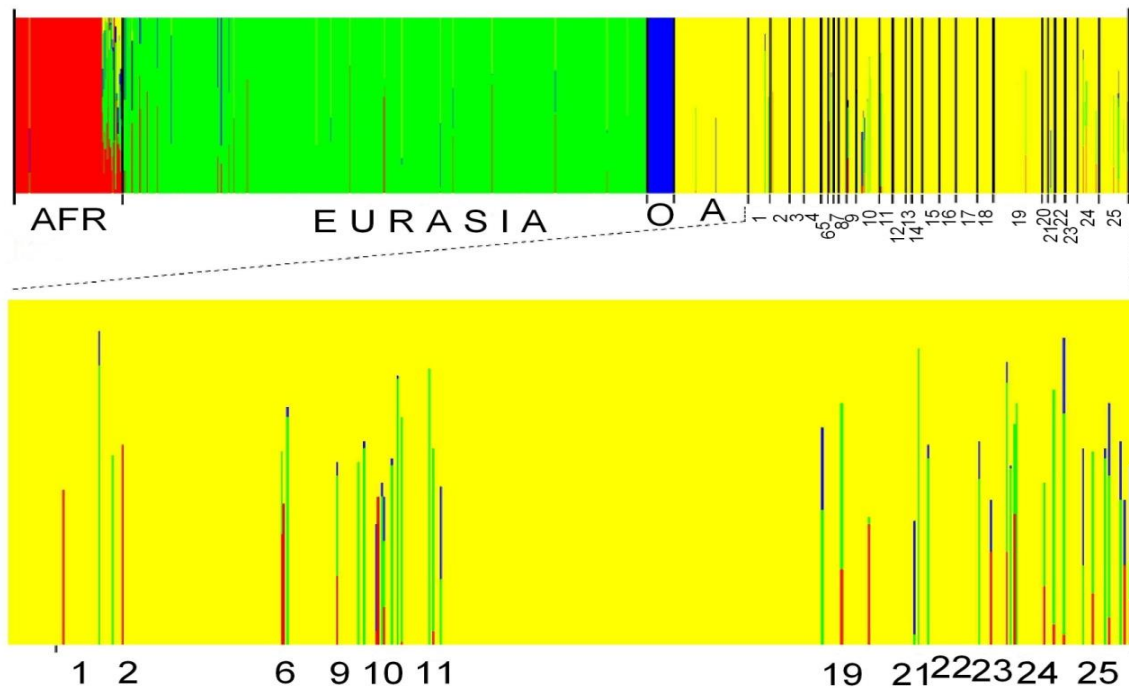


Figura 46: “Trained” admixture clustering por BAPS assumindo *a priori* K=4. AFR=África, O=Oceania; A=América; Eurásia (Europa, Oriente Médio, Ásia Central, Leste Asiático). Veja os códigos das localidades peruanas em Material e Métodos. As análises estão baseadas em 10.000 iterações das probabilidades *a posteriori* das frequências alélicas e 10 réplicas.

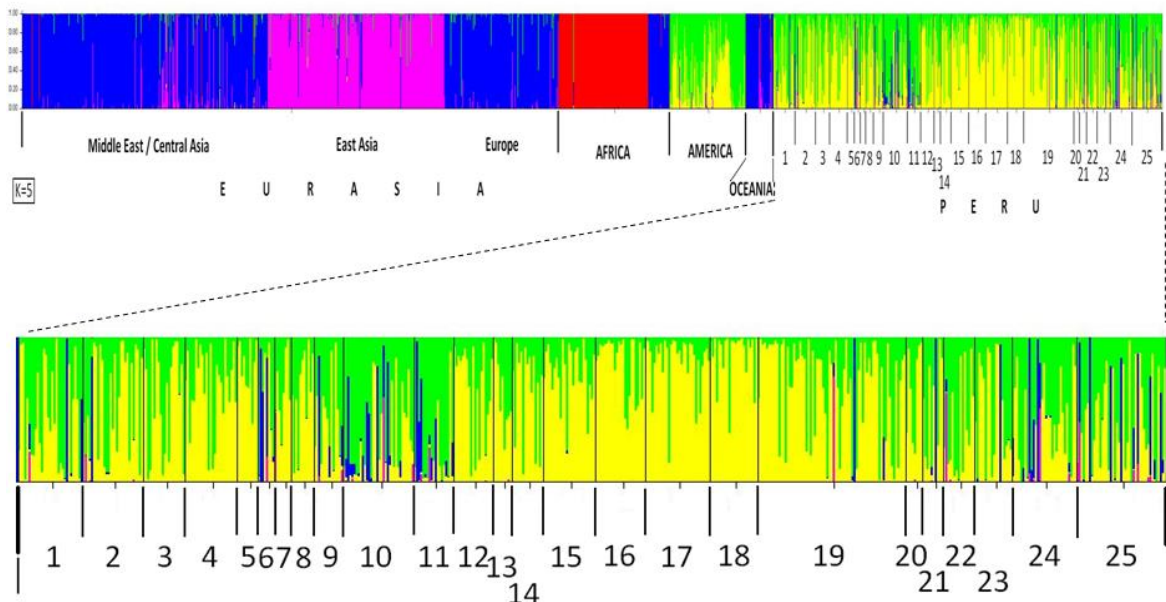


Figura 47: Plotagem dos indivíduos segundo as populações em agrupamento K=5 por STRUCTURE, segundo o modelo “não miscigenação” e “frequências alélicas não correlacionadas”. MCMC= 2×10^6 , Burn-in=100.000.

6.5.4- Agrupamento das populações peruanas por Componentes Principais

Foram executadas as análises dos componentes principais (PCA) com os dados dos 40 INDELS para correlacionar as frequências alélicas e genótípicas entre todos os indivíduos ou populações. O pacote PCAGEN foi usado para estimar a porcentagem de cada valor de inércia nos dois componentes principais e para calcular o valor P de cada valor *eigen* foram usadas 10.000 iterações (*randomizations*) dos genótipos.

Os resultados mostraram que os valores *eigen* para os dois primeiros componentes (PC1= 23,8%; PC2= 11,3%) são altamente significativos ($p\text{-value}=0.000$), também que o F_{st} (0,01) do PC1 apresenta maior diferenciação entre as populações do que o F_{st} do PC2 (0,05). As análises apresentaram em global, o $F_{st}=0,044$ e a heterozigidade total, $H_t=37,3\%$.

Na Figura 48 observa-se que os locais CAsm (San Marcos_Cajamarca), CA (Cajamarca), Char (Characato_Arequipa), LA (Lambayeque), LI (Lima) e Ch (Chogo) estão em lados opostos ou são os mais distantes em relação aos locais Taq (Taquile), Amt (Amantani), Ap (Anapia), Ur (Uros) e Yke (Yanque). A mesma distribuição espacial das populações foram observadas ao usar o método de Análise por Correspondência Fatorial (FCA) implementado no programa GENETIX. Por outro lado, é interessante observar que há um gradiente de distribuição dos genótipos dos indivíduos em nível continental, consistente à noção sobre a estrutura populacional em forma de gradiente da população humana global (Figura 49).

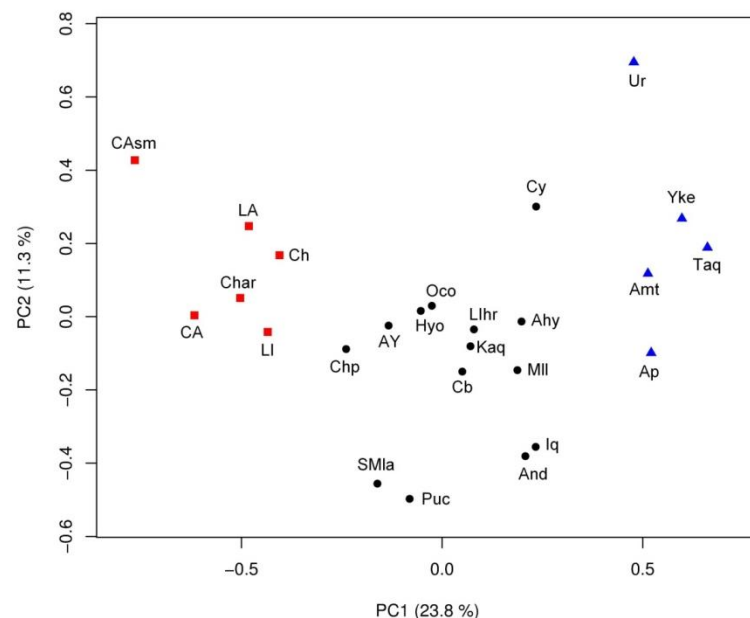


Figura 48: Plotagem das populações pelo método de PCA. Os quadrados representam as localidades mais miscigenadas que estão do lado contrário daquelas representadas por triângulos (menos miscigenadas).

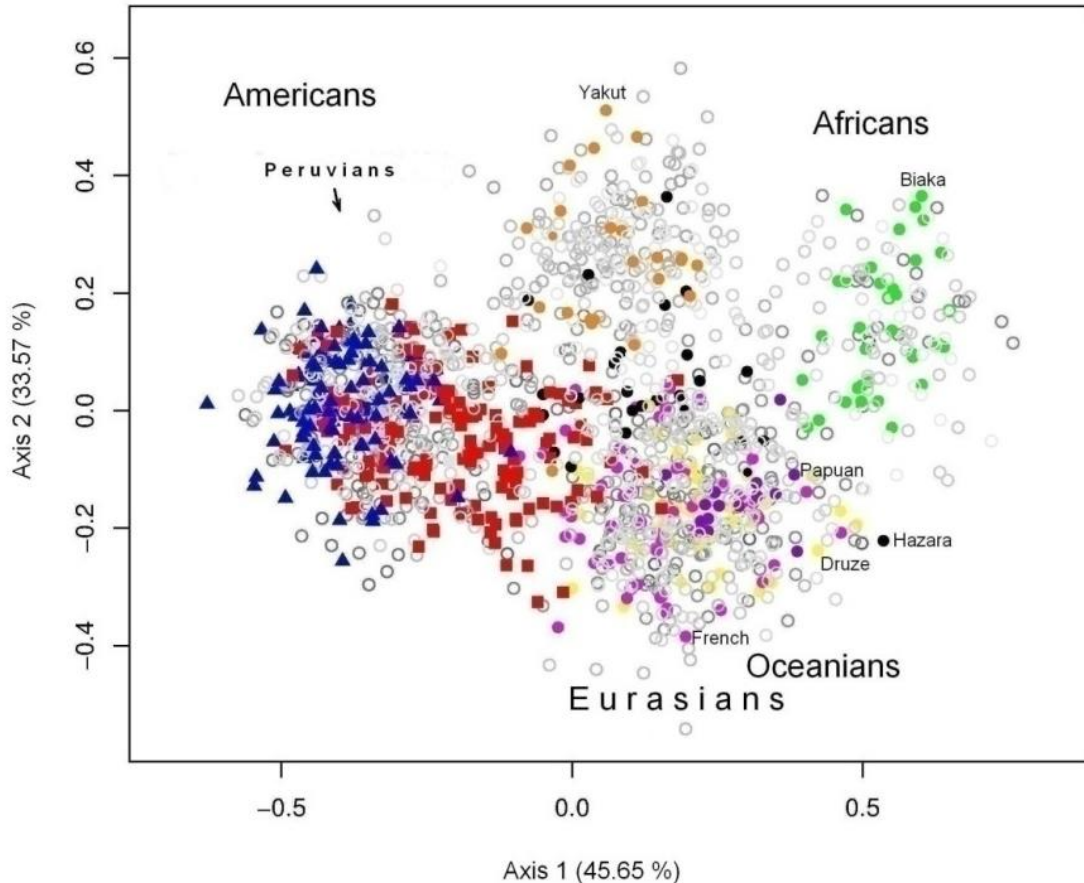


Figura 49: Plotagem pelo método FCA (Factorial Correspondence Analysis) dos indivíduos que são agrupados segundo seus genótipos nos dois principais eixos (Axis1 e Axis2). Os quadrados representam aos indivíduos (de CA, CAsm, Char, Ch, LA, LI) formando um gradiente próximo à Eurásia que corresponde ao grau de miscigenação genética. Os triângulos representam indivíduos com menor grau de miscigenação (Taq, Amt, Ap, Ur e Yke).

6.5.5- Análises da variabilidade genética nas populações peruanas

Para revelar as diferenças interpopulacionais e intrapopulacionais das 25 localidades peruanas em níveis hierárquicos, se realizou o teste AMOVA (Tabela 17). A comparação entre as três regiões (Amazônia, Andes, Costa) apresentou 0,33% indicando que não há diferenciação entre os grupos comparados. Quando a comparação é feita entre as populações dentro das regiões (WG=within groups) apresenta-se uma pequena diferenciação (2,182%), e menor ainda quando comparada entre indivíduos dentro das populações (WP=within populations) que foi de 1,373%. A maior diferenciação foi observada entre os indivíduos (96,12%).

Por outro lado, a comparação feita em forma separada (Amazônia, Andes e Costa), indica que há maior variabilidade entre as localidades dos Andes ($F_{st}=0,0274$) do que entre as da Amazônia e Costa ($F_{st}=0,00995$ e $F_{st}=0,00365$, respectivamente). Tomando em conjunto todas as subpopulações como um só grupo, obtiveram-se 2,37 % de diferenciação entre as populações, 1,37 % entre os indivíduos dentro das populações e 96,26 % entre os indivíduos.

As análises multidimensionais (MDS) das distâncias genéticas (F_{st} de Reynolds) mostraram que os agrupamentos realizados através de AMOVA estão de acordo com aqueles determinados pelos programas STRUCTURE, BAPS, PCAGEN e GENETIX, com exceção de Oco que se observa um pouco afastado. As distâncias genéticas F_{st} entre as populações estão correlacionadas entre elas, de tal modo que foi observado um gradiente molecular das mesmas. A configuração espacial no MDS mostrou que grande parte da variabilidade genética (33,5%) é explicada pela partição na coordenada 1, e que por outro lado, 11 locais (CAsm, CA, Char, LA, LI, Ch, Chp, SMIa, AY, Puc e Hyo) apresentam ser um grande conglomerado. Os locais Taq, Amt, Ap e Ur apresentam estar mais próximos entre si; Iq, Yke e Oco estão separados dos conglomerados, e os 7 locais restantes (LIhr, Kaq, Cb, MII, And, Ahy e Cy) sugerem manter alguma relação entre si.

Com o objetivo de obter maior resolução nos agrupamentos populacionais no nível continental, foram feitas as comparações com 18 populações selecionadas ($n=386$) do painel de referência HGDP-CEPH (África ($n=55$), Oceania ($n=17$), Oriente Médio ($n=29$), Europa ($n=43$), Ásia Central ($n=25$), Leste Asiático ($n=109$), América ($n=108$)). O MDS das distâncias genéticas F_{st} de Reynolds mostrou um alto grau de estruturação populacional (Figura 50), e observando-se que as afinidades genéticas dependem do nível de variabilidade e da miscigenação. Entre as populações peruanas é interessante observar a proximidade entre os de AY (Ayacucho), Hyo (Huancayo_Junin) e os indivíduos de SMIa (Lamas_San Martin), que poderia sugerir ancestrais comuns ou história de miscigenação pós-colombiana entre elas. Segundo a história do Peru essas populações compartilhariam descendentes de culturas pré-incas como os Chancas e Pocras ("originárias" de Ayacucho e Apurimac), os quais foram submetidos com dificuldade pelos incas, sendo que os que conseguiram fugir chegaram às regiões de San Martin (Lamas, Centro Povoado Wayku), Amazonas e parte de Loreto (Lexus, 1998; Pardo, Doherty & Sangama, 2001).

No entanto observa-se também a proximidade entre os de LI (Lima), CA e CAsm (Cajamarca), LA (Lambayeque), Ch (Chogo) e Char (Characato), o que sugeriria ser efeito das miscigenações pós-colombianas dessas subpopulações com as da Eurásia, correspondendo à história conhecida da colonização do Peru. Por outra parte, Taquile e Amantani estão muito próximas, corroborando as histórias de ancestralidade comum, já que são Quéchuas-falantes e provavelmente se deslocaram da mesma península chamada

Capachica às ilhas onde hoje habitam, que também são relativamente próximas à ilha de Anapia (Aimará-falantes). Nestas comunidades observa-se uma elevada ancestralidade genética da América, sem miscigenação detectada, e junto com as de Yke (Yanque_Arequipa) são as mais afins à etnia de Karitiana da Amazônia brasileira. As comunidades amazônicas de Puc (Pucallpa), Iq (Iquitos) e Chp (Chachapoyas) têm vinculação com as de Suruí (Brasil) e de Piapoco (Colômbia). Por outro lado, Cb (Cabanaconde) e Cy (Chivay) além da proximidade geográfica entre elas (Abrupto do Colca_Arequipa), estão correlacionadas com Ur (Uros) e And (Andoas), assim como sendo próximas às de Mll (Mollebaya_Arequipa) e as de Oco (Ocopon_Ancash).

Para testar as diferenciações populacionais foi usado o programa GENEPOP 4.0. Os testes sobre a deficiência ou excesso de heterozigotos segundo o modelo Hardy-Weinberg foram feitos por meio do Global H-W test (score (U) test) com parâmetros iguais de Markov Chains (*dememorization*=100.000; *batches*=100; *iterations per batch*=10.000). No caso do excesso de heterozigotos (*multi-population test*) os loci MID1352 e MID1642 mostraram valores P significativos ($p < 0,05$), mas em conjunto para todos os *loci* e populações não foi significativo (p -value= 0,9832, SE=0,0002). No caso do teste sobre deficiência de heterozigotos (*multiloci test*) para cada população, valores P significativos foram encontrados para San Marcos, Lamas e Lima; e no teste por *locus*, os marcadores MID258, MID1, MID1092, MID420 e o MID1100 apresentaram valores P significativos ($p < 0,05$). Em conjunto, para todas as populações e *loci* o p -value foi de 0,0169 (SE=0,0002).

Os testes de diferenciação genética (*exact G test*) através das frequências genotípicas ou alélicas entre todos os pares de localidades peruanas mostraram uma grande diferenciação (*P-value highly significant*) entre os de CA, CAsm comparados com os de Taq, Amt, Ap e Ur; os de And comparados com os de CA, CAsm e Ur; entre os de LA, LI comparados com os de Taq, Amt e Ur. Chama a atenção que oito locais têm marcadas diferenças genéticas entre elas, tanto mostradas pelos valores F_{st} quanto pelos agrupamentos nas análises de componentes principais (PCA), onde CAsm, CA, LI, LA estão muito distantes (marginal) comparados com os de Taq, Amt, Ur e Ap (marginal). Os parâmetros de Markov Chain foram: *dememorization*=100.000; *batches*=500; *iterations per batch*=10.000. A média de heterozigosidade observada para todos os 40 *loci* ($H_o=36\%$) mostrou o valor muito similar ao esperado ($H_e=37\%$), da mesma maneira quando todas as populações foram consideradas como um só grupo, a proporção foi a mesma ($H_o=36\%$; $H_e=36\%$), similar ao valor obtido pelo programa PCAGEN ($H_{total}=37,3\%$).

Um elevado grau de heterozigosidade esperada (H_e) foi revelado nas populações de Characato, San Marcos, Cajamarca, Lambayeque, Lima e Chogo (0,425, 0,418, 0,41, 0,408, 0,407, 0,40, respectivamente), em contraste, um baixo grau nas populações de Taquile, Anapia, Amantani, Yanque e Uros (0,307, 0,308, 0,319, 0,332 e 0,333 respectivamente).

Assim, ao comparar esses valores de H_e com os graus de miscigenação das populações peruanas em relação ao $K=2$ (Not-America) da Tabela 15, através do teste de Pearson, foi obtida uma elevada correlação linear ($r=0.975$, $p=2.20e-16$) como é apresentado na Figura 51.

Para determinar se há correlação ou não entre a distância genética (F_{st} de Reynolds) e geográfica (Km), foram feitos testes de Mantel através dos programas GenAlEx e IBDWS. Os resultados mostraram que há pouca correlação entre os dois parâmetros ($R^2=0,069$).

Tabela 17: Resultado de AMOVA em níveis hierárquicos sobre 25 populações peruanas através dos 40 marcadores INDELS.

Região	Entre grupos (%)	Entre populações WG (%)	Entre populações (%)	Entre indivíduos WP (%)	Entre indivíduos (%)
3 regiões (Amazônia, Andes, Costa)	0,330	2,182		1,373	96,12
Amazônia			0,995	2,088	96,917
Andes			2,740	0,816	96,444
Costa			0,365	2,652	96,982
1 região (Peru)			2,366	1,375	96,259

WG=Within Groups, WP=Within populations

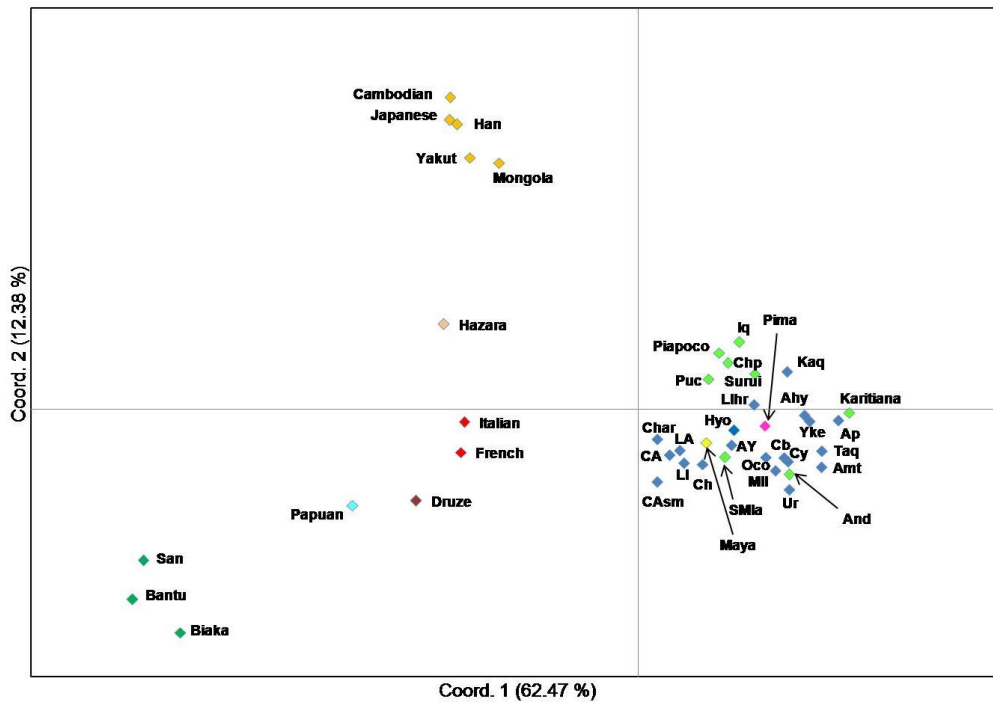


Figura 50: Plotagem MDS das distâncias genéticas (F_{st} de Reynolds) entre as 25 populações peruanas (códigos em Material e Métodos) e as 18 populações selecionadas do painel HGDP-CEPH.

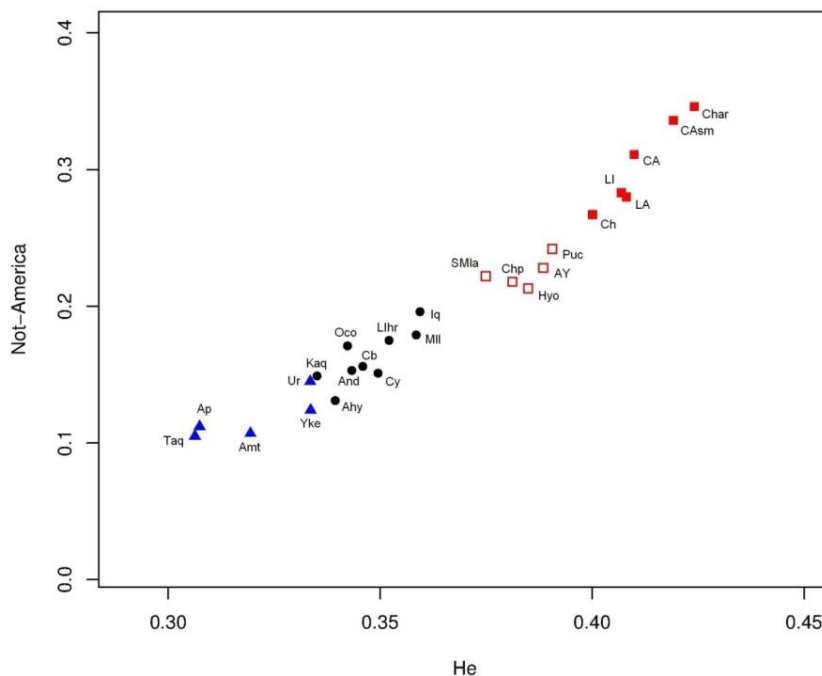


Figura 51: Plotagem dos índices de heterozigosidade esperada (H_e) e dos coeficientes da proporção de ancestralidade (Q -membership ancestry proportions) de Not-America quando as populações peruanas são agrupadas em $K=2$ (da Tabela 15) usando como referência o painel HGDP-CEPH.

7 – DISCUSSÃO

7.1- Sobre o mtDNA

Nos estudos prévios sobre as linhagens maternas autóctones das populações peruanas atuais e de amostras antigas (arqueogenética) foi sugerido que o nível de variabilidade genética das populações do oeste é similar ao do leste da América do Sul (Fehren-Schmitz *et al.*, 2010; Kemp *et al.*, 2009; Lewis *et al.*, 2008; Lewis *et al.*, 2007) em contraste às hipóteses preliminares sobre uma marcada diferenciação entre as duas regiões (Fuselli *et al.*, 2003; Rodriguez-Delfin *et al.*, 2001).

Nossos resultados mostram que o nível da variabilidade de cada linhagem e suas frequências varia dependendo da história demográfica de cada população assim como da região em estudo, sendo o fluxo gênico e a deriva como os fatores que influenciam, na configuração das relações de F_{st} (diferenciação) entre as populações. Por exemplo, foi observado que ao longo dos Andes há predominância do sub-haplogrupo B2 (Bisso-Machado *et al.*, 2012; Kemp *et al.*, 2009; Shinoda *et al.*, 2006; Lewis *et al.*, 2004), mas no litoral norte-Lambayeque- a frequência do sub-haplogrupo D1 é maior ou próxima à do B2 (Shinoda *et al.*, 2006; Shimada *et al.*, 2004). Por outro lado, algumas populações próximas apresentam uma diferenciação maior do que outras que estão mais distantes entre si.

Em nossos resultados, entre as populações andinas comparadas ($F_{st}=0,1305$) há menor diferenciação do que entre as do Lago Titicaca ($F_{st}=0,135-0,139$, dependendo de como é feito o agrupamento). O teste feito sem as populações do Lago Titicaca (Taq, Amt, Ur, Ap, PU) mostrou ainda menor diferenciação ($F_{st}=0,041$) entre as populações dos Andes comparadas as do Lago Titicaca. Esse resultado coincide com outro estudo prévio (Barbieri *et al.*, 2011) que mostra que a diferenciação entre 21 populações andinas ($F_{st}=0,119$) ou 8 populações de Quéchua-falantes ($F_{st}=0,106$) é menor do que entre aquelas do Lago Titicaca ($F_{st}=0,154$, divididas em Quéchuas e Aimarás). Nossas observações são apoiadas por uma revisão recente feita sobre os estudos das linhagens uniparentais na América do Sul (Bisso-Machado *et al.*, 2012), na qual se demonstra o impacto da geografia e da linguagem nas diferenciações interpopulacionais.

Desde um ponto de vista histórico pré-colombiano, o império incaico atingiu todo o Peru, Equador, Bolívia, parte da Colômbia, Argentina, e Chile, homogeneizando não somente a língua (Quéchua) se não também algumas culturas pré-incas conquistadas ou aliadas (Chimu, Pachacamac, Chincha, Chopka, Chanka, Wanka, Collagua, Cañari (Equador), Chachapoya, entre outros) através das deslocamentos forçados. Este sistema é conhecido como “mitimae” e aconteceu ao longo do *Qhapaq Ñan* (Trilha Inca) e de suas vertentes. O padrão genético revelado em alguns estudos prévios, aparentemente seria o efeito disso, a maior homogeneidade nas populações andinas em relação aos amazônicos, já que na

Amazônia os grupos nativos ficaram “isolados” e diferenciados por deriva posteriormente. No entanto, no período da colonização espanhola também aconteceram migrações forçadas através das denominadas “encomendas” e “reduções” por causa da exploração das minas e da evangelização, além do “boom” da borracha na Amazônia. Todo esse panorama mostra a grande complexidade das interações demográficas que nos sugere considerar a maior parte das regiões ou populações do Peru nas inferências históricas através da genética, sendo esta a nossa abordagem principal.

Em geral, segundo a distribuição das frequências dos haplogrupos A, B, C e D de mtDNA observadas nas 29 localidades peruanas das três regiões geográficas (Amazônia, Andes e Costa), em conjunto, a frequência do sub-haplogrupo B2 é predominante (52,7%), seguida dos sub-haplogrupos A2 e C1 (18,1%, ambos com igual porcentagem), um pouco menos pelo sub-haplogrupo D1 (11%).

Considerando em conjunto todos os haplogrupos revelados nas populações analisadas, a distribuição de suas frequências indica que a contribuição matrilinear é em grande parte de proporção autóctone (97,1%). Por outro lado, a pequena porcentagem da contribuição dos “outros” haplogrupos assinala baixo grau de “miscigenação” das mulheres nativas europeias, asiáticas ou africanas com os homens nativos peruanos após o colonialismo europeu, em concordância com os registros deixados pelos cronistas na época da colonização espanhola. Este cenário é um pouco diferente aos de outros países da América do Sul, em particular da história do Brasil, onde foram revelados elevados graus de “miscigenação” pós-colombiana, sendo predominantes as linhagens maternas europeias (39%), seguida por linhagens nativas americanas (33%) e em menor porcentagem pelas linhagens africanas (28%) (Alves-Silva *et al.*, 2000).

Por outro lado, nas análises sobre as diferenciações populacionais com os testes AMOVA (só com os haplogrupos A, B, C e D), a comparação entre as populações da Amazônia mostrou uma menor variação do que entre as populações andinas. Por sua vez, só considerando a distribuição do haplogrupo B, também se observou o mesmo resultado ($F_{st}=0,1445$ entre as populações amazônicas e $F_{st}=0,194$ entre as andinas respectivamente).

Por outro lado, as análises das linhagens uniparentais das populações do Lago Titicaca mostraram que tanto os Quéchuas quanto os Aimarás são muito próximos, quando comparados às demais populações peruanas. Além disso, os resultados mostram o efeito fundador tal como foi sugerido em nosso prévio estudo (Sandoval *et al.*, 2004).

Adicionalmente, através do método MCMC Bayesiano de BEAST foram estimadas as idades aproximadas da coalescência das quatro linhagens autóctones (A2, B2, C1 e D1) dos habitantes do Lago Titicaca usando como referência 17 sequências completas do genoma mtDNA. Segundo os resultados obtidos através dos genomas completos, os quatro sub-

haplogrupos teriam idades entre 24.000 a 41.000 anos (TMRCA de 91.000 anos) e somente através da região codificadora entre 16.000 a 23.000 anos (TMRCA de 65.000 anos). As simulações feitas para cada sub-haplogrupo por separado mostraram idades similares quando incluídas as 17 sequências (ver Tabela 13). Entretanto, quando as análises foram abordadas pela estatística *Rho* (ρ), apresentaram menores valores (19.000-33.000 para genomas completos; e 11.000-20.000 para a região codificadora). As idades dos quatro subhaplogrupos autóctones oscilam dentro dos limites obtidos por outros grupos de pesquisadores seja usando o método estatística *Rho* (Tabela 18) ou da simulação bayesiana de Fagundes *et al.* (2008). Apesar das diferenças dos valores observados (pelo uso de diferentes taxas de mutação, além de outros fatores ou parâmetros específicos da metodologia) nossa abordagem mostrou ser eficaz. O modelo usado sugere os eventos da divergência de cada linhagem (incluindo de todas as outras linhagens) desde o MRCA, como o ancestral comum de todas as sequências mtDNA consideradas, e que é consistente com a árvore consenso de mtDNA sugerida pelo PhyloTree (van Oven & Kayser, 2009) e Mitomap (www.mitomap.org).

Tabela 18: Comparação dos tempos de coalescência determinados através do parâmetro estatístico *Rho* (ρ) usando diferentes taxas de mutação para a região codificadora do mtDNA.

Hg	n	<i>Rho</i> (ρ)	<i>Sigma</i> (σ)	T(anos)	ΔT (anos)	^a Soares (1 substituição cada 3533 anos); mean (sigma)
Achilli <i>et al.</i>, 2008:						
$\mu=1,26 \times 10^{-8}$; (1 substituição cada 5140 anos)						
A2	97	3,34	0,322	17200	1700	11800 (1140)
B2	43	4,116	0,463	21200	2400	14550 (1640)
C1	55	4,636	0,836	23800	4300	16400 (2950)
D1	34	3,618	0,441	18600	2300	12800 (1560)
Volodko <i>et al.</i>, 2008:						
$\mu=1,26 \times 10^{-8}$ segundo mtPHYL						
A2	44	3,1	0,41	16200	2100	11000 (1450)
B2	-	-	-	-	-	-
C1	24	4,67	0,75	24000	3800	16500 (2650)
D1	20	3,35	0,55	17200	2800	11840 (1950)
Perego <i>et al.</i>, 2009:						
$\mu=3,5 \times 10^{-8}$ (1 substituição cada 4610 anos), média de Mishmar e Kivisild.						
A2	100	3,42	0,289	15800	1300	12100 (1000)
B2	62	4,065	0,371	18700	1700	14360 (1300)
C1	68	4,25	0,702	19600	3200	15000 (2500)
D1	46	3,674	0,431	16900	2000	13000 (1520)
Nossos dados:						
$\mu=1,708 \times 10^{-8}$ (1 substituição cada 3533 anos) segundo mtPHYL						
A2	3	3	1	10600	3533	
B2	3	5	1,29	17700	4560	
C1	4	5,33	1,33	18900	4700	
D1	2	5,5	1,66	19500	5900	

^a Soares *et al.*, 2009.

7.2- Sobre o cromossomo Y

No caso das linhagens masculinas, usando os 10 marcadores SNPs (Figura A1, Anexo) foi possível avaliar a distribuição dos haplogrupos Q-M3, R, E e F nas populações peruanas analisadas (n=464). A frequência do haplogrupo Q-M3 é maior (60,1%) do que dos outros haplogrupos, em especial quase três vezes a frequência da linhagem europeia R-M207 (18,8%), ou cinco vezes a frequência do Hg F* (12,5%). Como no caso da contribuição maior de linhagens maternas nativas da América, os dados obtidos sobre as linhagens paternas também indicam uma maior proporção de contribuição autóctone americana. Os eventos de “miscigenação”, em sua maioria, aconteceram entre homens de linhagens euroasiáticas com mulheres nativas americanas, situações já descritas pelos cronistas da época pós-colombiana, tal como corroborado aqui pelos marcadores uniparentais.

O gradiente de miscigenação das populações peruanas revelado através dos SNPs do cromossomo Y mostra um quadro muito similar ao obtido com os 40 marcadores de ancestralidade do tipo INDELS, com tendência de diminuição de norte a sul. As subpopulações da região do Lago Titicaca (Taquile, Amantani, Uros, Anapia) se apresentam sem miscigenação detectável, enquanto as subpopulações do norte (Cajamarca, San Marcos, Lambayeque) e do centro do país (Ayacucho, Andahuaylas, Lima, Huancayo) aparecem como miscigenadas. Coincidentemente, um gradiente similar de miscigenação foi revelado nas populações do México em um estudo de 896 indivíduos “mestiços”, no qual as frequências de ancestralidade europeia variam de 36% até 85%, de sul a norte, respectivamente (Rangel-Villalobos *et al.*, 2009). Este cenário mostra um padrão similar de miscigenações entre homens europeus e mulheres autóctones americanas, evento que aconteceu ao longo das Américas durante colonização espanhola tanto em Peru quanto em México.

Por outro lado, as análises sobre a estrutura populacional através de AMOVA, considerando os haplótipos dos 17 Y-STRs que correspondem ao haplogrupo Q-M3, e agrupadas segundo a região geográfica (Amazônia, Andes e Costa) mostraram que entre as populações amazônicas (n=39; Andoas, Pucallpa e Chachapoyas) há menor diferenciação ($R_{st}=0,0056$) quando comparado entre as andinas (18 localidades, n=176) ($R_{st}=0,178$). Este panorama está em acordo com os resultados obtidos através do mtDNA e dos 40 INDELS.

As análises espaciais com redes MJ mostram que os haplótipos Q-M3 não estão subgrupados especificamente e há uma grande variabilidade entre os dos Andes, Amazônia e Costa. Este panorama é concordante com a história sobre o fluxo gênico entre os habitantes dessas regiões desde a época pré-colombiana. O cenário sugere influência das deslocamentos dos “mitimaes” pelos incas, além da colonização europeia do Peru, incluindo as “ondas laborais” do período da exploração do “ciclo da borracha” nos séculos

XIX e XX. Além disso, eventos migratórios motivados pelo Estado peruano que levou “colonos” andinos à Amazônia, inclusive para a fronteira com Brasil, Colômbia, Equador e Bolívia. Um exemplo peculiar são os grupos denominados Kichwas (à diferença dos Quéchuas andinos típicos) que habitam diferentes lugares da Amazônia peruana. Estes movimentos migratórios internos recentes, somados ao processo de miscigenação poderiam explicar algumas diferenças observadas para os Rst entre as populações dos Andes e da Amazônia.

7.3- Sobre os INDELS

Nas décadas recentes, o uso de diferentes conjuntos de AIMs (Galanter *et al.*, 2012; Pereira *et al.*, 2012; Santos *et al.*, 2010; Bastos-Rodrigues *et al.*, 2006) na estimativa de ancestralidade dos indivíduos ou populações se mostrou eficaz nos estudos evolutivos.

Neste estudo, usamos um conjunto de 40 INDELS com o objetivo de desvendar o impacto demográfico do povoamento pré e pós-colombiano nos atuais habitantes do Peru. Estudos prévios em populações do Brasil (n=934) usando estes mesmos 40 marcadores mostraram que a proporção de miscigenação dos indivíduos é uniforme, tanto na parte norte quanto na parte sul do país (Pena *et al.*, 2011), embora com a relativa predominância da ancestralidade europeia (ordem de 60,6%-77,7%). Em contraste, nossos resultados mostram uma elevada proporção de ancestralidade autóctone americana (ordem de 53,8%-96,5%). Esta discrepância reflete a diferença na história demográfica de ambos os países. Por exemplo, vários cronistas e pesquisadores descreveram de diferentes formas a maior quantidade relativa da população efetiva ao longo dos Andes, em contraste a parte do leste da América do Sul (Cavalli-Sforza *et al.*, 1994).

Um estudo anterior baseado nos padrões dermatoglíficos revelou uma similaridade entre as populações do norte e centro dos Andes do Peru e uma diferenciação dessas com as da região sul, especialmente de Puno (Ramirez *et al.*, 2001). Por outra parte, usando um sistema de marcadores STRs (Iannaccone *et al.*, 2011) com amostras de diferentes regiões do Peru sugeriu que as populações peruanas estariam agrupadas segundo suas correspondentes regiões geográficas (norte, centro e sul), além disso, reportam que em conjunto há ao redor de 30% de miscigenação com populações não autóctones americanas. Em ambos os estudos notou-se uma significativa estruturação populacional, mas nossos resultados indicam que esta ocorre em uma forma de gradiente de miscigenação, sendo particularmente maior em San Marcos e Cajamarca, Characato, Chogo, Lambayeque e Lima e sem miscigenação ou muito pequena nas subpopulações do Lago Titicaca (Taquile, Amantani, Anapia, Uros) e Yanque (Arequipa).

Segundo as análises feitas no STRUCTURE há similaridade entre os resultados obtidos em K=5 ao usar o modelo “*admixture, allelic frequencies correlated*” e o modelo “*not-admixture, allelic frequencies not correlated*”, mas o primeiro modelo explica melhor a estrutura de nossos dados, quer dizer, é consistente com o cenário de colonização do Peru. Cabe mencionar que na partição K=2, as populações da Europa e do Leste Asiático agrupam-se em um só “módulo” e as populações da América em outro, em concordância com o quadro obtido por Wang *et al.* (2007), mas em contraste com os resultados obtidos por outros autores (Rosenberg *et al.*, 2002; Li *et al.*, 2008; Bastos-Rodrigues *et al.*, 2006). Em geral, os agrupamentos sugeridos por STRUCTURE desde K=3 até K=6 são coincidentes com esses estudos prévios.

Nos resultados de AMOVA, observa-se que há maior variabilidade entre as populações dos Andes do que entre as da Amazônia e da Costa, de maneira similar quando foram comparadas através dos marcadores uniparentais, mtDNA e cromossomo Y, em contraste aos estudos prévios sobre as populações peruanas como foi mencionado anteriormente. Esta diferença de resultados sugere que há um grande impacto do fluxo gênico pós-colombiano no grau de miscigenação entre as populações atuais, em pleno acordo com os estudos recentes realizados pelo grupo de Reich *et al.* (2012). Isto reflete por um lado a elevada variabilidade observada nas populações mais miscigenadas e por outro lado a menor variabilidade nas menos ou não miscigenadas, consistente com a correlação dos níveis de heterozigosidade e a proporção de miscigenação que foi mostrado na Figura 51.

Em geral, o nível de miscigenação nas populações analisadas varia de acordo com a história demográfica de cada local e segundo a estruturação em K=2, o Peru se compõe de 80% de ancestralidade autóctone (legado pré-colombiano) e 20% de ancestralidade não autóctone, principalmente da Eurásia, que seria o legado recente ou pós-colombiano. Pelos resultados em K=5, as populações apresentam-se em gradiente de miscigenação segundo suas proporções de ancestralidade. Em ambos os casos, há que considerar que os valores ou proporções obtidos através das simulações bayesianas são somente aproximações *ad hoc* e podem mudar dependendo dos tipos de marcadores, número de marcadores ou de amostras, assim como do grau de diferenciação entre as populações (Latch *et al.*, 2006).

8 – CONSIDERAÇÕES FINAIS

Através das linhagens uniparentais (mtDNA e Y-DNA) e de marcadores autossômicos (INDELs), caracterizamos o perfil de ancestralidade genômica de algumas subpopulações peruanas de diferentes regiões do Peru (Andes, Amazônia, Costa).

Os resultados mostram que há uma presença menor de haplogrupos não autóctones maternos em contraste aos obtidos pela linhagem paterna, quando foi observada uma moderada frequência de haplogrupos euro-asiáticos, em particular da região europeia. O padrão de miscigenação dos marcadores uniparentais foi consistente com os resultados gerados com marcadores autossômicos através das simulações bayesianas sobre a ancestralidade dos indivíduos.

Os testes genéticos nos permitiram obter um panorama geral sobre os eventos demográficos correlacionados à história pré e pós-colombiana do Peru. Considerando as estimativas de miscigenações intrapopulacionais foi possível comparar com os registros históricos, sendo que ambas as abordagens, história e genética, se mostraram complementares. De forma geral foi evidenciado um gradiente norte-sul de distribuição da variabilidade genética nas populações e indivíduos segundo o seu nível de heterozigosidade.

9 – Referências bibliográficas

1. Achilli A, Perego UA, Bravi CM, Coble MD, Kong Q-P, Woodward SR, Salas A, Torroni A, Bandelt H-J. (2008). The phylogeny of the four pan-American mtDNA haplogroups: implications for evolutionary and disease studies. *PLoS ONE* **3**(3):e1764.
2. Aldenderfer M, Craig NM, Speakman RJ & Popelka-Filcoff R. (2008). Four-thousand-year-old gold artifacts from the Lake Titicaca basin, southern Peru. *Proc Natl Acad Sci USA*. **105**:5002–5005.
3. Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VF. (2000). The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet*. **67**(2):444-61.
4. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. (1981). Sequence and organization of human mitochondrial genome. *Nature* **290**:457-465.
5. Baca M, Doan K, Sobczyk M, Stankovic A, Wegléński P. (2012). Ancient DNA reveals kinship burial patterns of a pre-Columbian andean community. *BMC Genet*. **13**:30 doi:10.1186/1471-2156-13-30.
6. Bandelt H-J, Forster P, Sykes BC & Richards MB. (1995). Mitochondrial portraits of human populations using median networks. *Genetics* **141**:743-753.
7. Barbieri C, Heggarty P, Castri L, Luiselli D & Pettener D. (2011). Mitochondrial DNA variability in the Titicaca Basin: matches and mismatches with linguistics and ethnohistory. *Am J Hum Biol*. **23**:89-99.
8. Bastos-Rodrigues L, Pimenta JR, Pena SD. (2006). The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Ann Hum Genet*. **70**:658-65.
9. Bisso-Machado R, Bortolini MC & Salzano FM. (2012). Uniparental genetic markers in South Amerindians. *Genet Mol Biol*. **35**(2):365-387.
10. Bouysse-Cassagne T. (1991). Poblaciones humanas antiguas y actuales. *ORSTOM, HISBL, Bolivia, p.* 481-498.
11. Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, 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.

12. Builes JJ, Hau J, Rodríguez JI, Montoya AE, Izarra F, Ochoa O, Pérez L, de Pancorbo M, Gusmão L, Bravo MLJ. (2006). Peruvian population study with 16 Y-STR loci. *Int Congress Series* **1288**:216-218.
13. Cann R, Stoneking M, Wilson AC. (1987). Mitochondrial DNA and human evolution. *Nature* **325**:31-36.
14. Carbajal-Caballero N, Núñez S, Narvaiza M, Aguirre C, Villanueva C, Muro J & Rodríguez-Delfín L. (2005). Polimorfismos del cromosoma Y humano en poblaciones de la región norte del Perú. *Rev peru biol.* **12**(3):341-348.
15. Cavalli-Sforza LL, Menozzi P, Piazza A. (1994). The history and geography of human genes. *Princeton University Press, Princeton, New Jersey, USA*.
16. Chakraborty R., Kamboh MI, Nwankwo M & Ferrell RE. (1992). Caucasian genes in American blacks: new data. *Am J Hum Genet.* **50**:145–155.
17. Chiaroni J, Underhill PA & Cavalli-Sforza LL. (2009). Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad Sci. USA.* **106**(48):20174-20179.
18. Clement M, Posada D & Crandall KA. (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol.* **9**:1657-1659.
19. Corander J, Marttinen P, Sirén J & Tang J. (2008). Enhanced bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**, 539.
20. Cox MP. (2008). Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum Biol.* **80**(4):335-357.
21. Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miscicka-Sliwka D, Wozniak M & Zakharov IA. (2003). Diversity of mitochondrial DNA lineages in south Siberia. *Ann Hum Genet.* **67**:391-411.
22. Drummond A & Rambaut A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* **7**:214.
23. Dubut V, Chollet L, Murail P, Cartault F, Béraud-Colomb E, Serre M & Mogentale-Profizi N. (2004). mtDNA polymorphisms in five French groups: importance of regional sampling. *Eur J Hum Genet.* **12**:293-300.
24. Earl DA & vonHoldt BM. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* doi. 10.1007/s12686-011-9548-7.
25. Easton RD, Merriwether AD, Crews DE & Ferrell RE. (1996). mtDNA variation in the Yanomami: evidence for additional New World founding lineages. *Am J Hum Genet.* **59**:213-225.
26. Espinoza Avendaño, J. (1989). Taquile. Historia-Economía-Artesanía. *Instituto de Investigaciones para el desarrollo social del altiplano (convenio UNA-NUFFIC)*, 148 p.

27. Excoffier L & Lischer HEL. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* **10**:564–567.
28. Fagundes N, Kanitz R, Eckert R, Valls A, Bogó M, Salzano F, Smith D, Silva Jr W, Zago M, Ribeiro-dos-Santos A. (2008). Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet.* **82**(3):583–592.
29. Falush D, Stephens M & Pritchard JK. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.
30. Fan L, Yao Y-G. (2011). MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations. *Mitochondrion* **11**:351-356.
31. Fehren-Schmitz L, Reindel M, Cagigao ET, Hummel S & Herrmann B. (2010). Pre-Columbian population dynamics in coastal southern Peru: a diachronic investigation of mtDNA patterns in the Palpa region by ancient DNA analysis. *Am J Phys Anthropol.* **141**:208-221.
32. Fehren-Schmitz L, Warnberg O, Reindel M, Seidenberg V, Tomasto-Cagigao E, Isla-Cuadrado J, Hummel S & Herrmann B. (2010). Diachronic investigations of mitochondrial and Y-chromosomal genetic markers in pre-Columbian andean highlanders from south Peru. *Ann Hum Genet.* doi:10.1111/j.1469-1809.2010.00620.x
33. Fuselli S, Tarazona-Santos E, Dubanloup I, Soto A, Luiselli D & Pettener D. (2003). Mitochondrial DNA diversity in South America and the genetic history of andean highlanders. *Mol Biol Evol.* **20**(10):1682-1691.
34. Galanter JM, Fernandez-Lopez JC, Gignoux CR, Barnholtz-Sloan J, Fernandez-Rozadilla C, Via M, *et al.* (2012). Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet.* **8**:e1002554.
35. Gonzáles-José R, Bortolini MC, Santos FR & Bonatto SL. (2008). The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am J Phys Anthropol. (on-line)*. doi:10.1002/ajpa.20854.
36. Hasegawa M, Kishino H & Yano T-a. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* **22**:160-174.
37. Heled J & Drummond AJ. (2008). Bayesian inference of population size history from multiple loci. *BMC Evol Biol.* **8**:289.
38. Herrera IL. (1986). Los inmigrantes chinos en la amazonia peruana. *Bull.Inst.Fr.Et.And.* XV, N3-4, pp. 49-60.

39. Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodrigues-Botigue L, Ramachandran S, Hon L, Brisbin A, Lin AA, Underhill PA, Comas D, Kidd KK, Norman PJ, Parham P, Bustamante CD, Mountain JL & Feldman MW. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci. USA.* **108**(13):5154-5162.
40. Hunefeldt C. (2004). A brief history of Peru. Lexington Associates, New York, USA.
41. Iannaccone GC, Parra R, Bermejo M, Rojas Y, Valencia C, Portugues L. (2011). Peruvian genetic structure and their impact in the identification of andean missing persons: a perspective from Ayacucho. *Forensic Sci Int.: Genet Supp Series* **3**:e291–e292.
42. Iannaccone GC, Tito RY, López PW, Medina ME & Lizárraga B. (2005). Y-chromosomal haplotypes for the PowerPlex® Y for twelve STRs in a Peruvian population sample. *J Forensic Sci.* **50**(1):1-4.
43. Jakobsson M & Rosenberg NA. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**:1801–1806.
44. Jobling MA & Tyler-Smith C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nature Genet.* **4**:598-612.
45. Jota MS, Lacerda DR, Sandoval JR, Vieira PPR, Santos-Lopes SS, Bisso-Machado R, et al. (2011). A new subhaplogroup of Native American Y-chromosomes from the Andes. *Am J Phys Anthropol.* **146**:553–559.
46. Karafet TM, Mendez FL, Meilerman MB, et al. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* **18**:830-838.
47. Kemp BM, Tung TA & Summar ML. (2009). Genetic continuity after the collapse of the Wari empire: mitochondrial DNA profiles from Wari and post-Wari populations in the ancient Andes. *Am J Phys Anthropol.* **140**:80-91.
48. Latch EK, Dharmarajan G, Glaubitz JC & Rhodes OE. (2006). Relative performance of bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conserv Genet.* **7**:295–302.
49. León Canales E. (2007). Orígenes humanos en los Andes del Perú. *Ed. USMP, Lima-Perú.* p.328.
50. Lewis CM Jr & Long JC. (2008). Native south American genetic structure and prehistory inferred from hierarchical modeling of mtDNA. *Mol Biol Evol.* **25**(3):478-486.
51. Lewis CM Jr, Lizarraga B, Tito RY, Lopez PW, Iannaccone GC, Medina A, Martinez R, Polo SI, De La Cruz AF, Caceres AM & Stone AC. (2007). Mitochondrial DNA and the peopling of South America. *Hum Biol.* **79**(2):159-178.

52. Lewis CM Jr, Tito RY, Lizarraga B & Stone AC. (2004). Land, languages, and loci mtDNA in Native Americans and the genetic history of Peru. *Am J Phys Anthropol.* **127**:351-360.
53. Lexus Edit. (1998). Gran enciclopedia del Perú. Barcelona, España.
54. Librado P & Rozas J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451-1452.
55. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, *et al.* (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**:1100–1104.
56. Lumbreras L. (1981). Los orígenes de la civilización en el Perú. 5ta. Ed. Milla Batres, Lima, Perú.
57. Lynch TA & Kennedy KAR. (1970). Early human cultural and skeletal remains from Guitarrero Cave, northern Peru. *Science* **169**:1307–1309.
58. MacNeish RS, Berger R, Protscha R. (1970). Megafauna and man from Ayacucho, highland Peru. *Science* **168**:975–977.
59. Macpherson JM, Ramachandran S, Diamond L & Feldman MW. (2004). Demographic estimates from Y chromosome microsatellite polymorphisms: analysis of worldwide sample. *Hum Genomics* **1**(5):345-354.
60. Manelis de Klein HE. (1973). Los Urus: el extraño pueblo del altiplano. *Estudios Andinos*, *N7, V3, N1*:129-150.
61. Merriwether DA, Hall WW, Vahne A, Ferrell RE. (1996). mtDNA variation indicates Mongolia may have been the source for the founding population for the New World. *Am J Hum Genet.* **59**: 204-212.
62. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A & Wallace DC. (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci. USA.* **100**(1):171-176.
63. Mulero JJ, Chang CW, Calandro LM, Green RL, Li Y, Johnson CL & Hennessy LK. (2006). Development and validation of the AmpF/STR Yfiler PCR Amplification kit: a male specific, single amplification 17 Y-STR multiplex system. *J Forensic Sci.* **51**(1). doi:10.1111/j.1556-4029.2005.00016.x
64. O'Fallon BD & Fehren-Schmitz L. (2011). Native Americans experienced a strong population bottleneck coincident with European contact. *Natl Acad Sci. USA.* **108**(51):20444-20448.
65. Pardo CM, Doherty VJ, Sangama SI. (2001). Los Kechuas lamistas y la educación bilingüe intercultural: historia y razón de un compromiso. Ed. SAC PT, Tarapoto, San Martín, Perú.

66. Parks DH, Porter M, Churcher S, *et al.* (2009). GenGIS: a geospatial information system for genomic data. *Genome Res.* **19**:1896-1904.
67. Peakall R & Smouse PE. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol.* **6**:288–295.
68. Pena SDJ, Di Pietro G, Fuchshuber-Moraes M, Genro JP, Hutz MH, Kehdy FDSG, *et al.* (2011). The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS ONE* **6**:e17063.
69. Pena SDJ, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothhammer F, Gerelsaikhan T, Munkhtuja B, Oyunsuren T. (1995). A major founder Y-chromosome haplotype in Amerindians. *Nature Genetics* **11**:15-16.
70. Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Kashani BH, Ritchie KH, Scozzari R, Kong Q-P, Myres NM, Salas A, Semino O, Bandelt H-J, Woodward SR & Torroni A. (2009). Distinctive paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol.* **19**:1-8.
71. Pereira R, Phillips C, Pinto N, Santos C, Santos SEBd, Amorim A, Carracedo A, Gusmao L. (2012). Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS ONE* **7**(1):e29684. doi:10.1371/journal.pone.0029684.
72. Pimenta JR & Pena SDJ. (2010). Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms. *Genet Mol Res.* **9**:601–607.
73. Pritchard JK, Stephens M & Donnelly P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.
74. Ramirez BO, Del Valle ML & Arzola GN. (2001). Dermatoglyphics of a high altitude Peruvian population and interpopulation comparisons. *High Alt Med. & Biol.* **2**:31–40.
75. Rangel-Villalobos H, Salazar-Flores J, Dondiego R, Anaya-Palafox M, Nuño-Arana I, Canseco-Avila LM, Flores-Flores G, Romero-Rentería O, Morales-Vallejo ME, Muñoz-Valle JF, Rubi-Castellanos R. (2009). South to north increasing gradient of paternal European ancestry throughout the Mexican territory: evidence of Y-linked short tandem repeats. *Forensic Sci Int.: Genet Supp Series* **2**:448-450.
76. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, *et al.* (2012). Reconstructing Native American population history. *Nature* **1–6**. doi:10.1038/nature11258.
77. Rodriguez-Delfin LA, Rubin-de-Celis VE & Zago MA. (2001). Genetic diversity in an Andean population from Peru and regional migration patterns of Amerindians in South America: data from Y chromosome and mitochondrial DNA. *Hum Hered.* **51**:97–106.

78. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, *et al.* (2002). Genetic structure of human populations. *Science* **298**:2381–2385.
79. Rosenberg NA. (2003). Distruct: a program for the graphical display of population structure. *Mol Ecol.* **4**:137–138.
80. Rousset F. (2008). GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour.* **8**:103–106.
81. Sandoval J, Delgado B, Rivas L, Bonilla B, Nugent D, Fujita R. (2004). Variantes del ADNmt em isleños del Lago Titicaca: máxima frecuencia del haplotipo B1 y evidencia de efecto fundador. *Rev per biol.* **11**:161-168.
82. Sandweiss DH, McInnis H, Burger RL, Cano A, Ojeda B, Paredes O, Sandweiss MC, Glascock MD. (1998). Quebrada Jaguay: early south American maritime adaptations. *Science* **281**, 5384:1830-1832.
83. Santos FR, Hutz MH, Coimbra-JR CEA, Santos RV, Salzano FM, Pena SDJ. (1995b). Further evidence for the existence of major founder Y chromosome haplotype in Amerindians. *Braz J Genet* **18**:669-672.
84. Santos FR, Pandya A, Tyler-Smith C, Pena SD, Schanfield M, Leonard WR, Osipova L, Crawford MH & Mitchell RJ. (1999a). The central Siberian origin for native American Y chromosomes. *Am J Hum Genet.* **64**:619-628.
85. Santos NPC, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AKC, Pereira R, Gusmão L, Amorim A, Guerreiro JF, Zago MA, Matte C, Hutz MH, *et al.* (2010). Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum Mut.* **31**:184–190.
86. Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA. (2000). The genetic legacy of paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* **290**:1155-1159.
87. Schlecht J, Kaplan ME, Barnard K, Karafet T, Hammer MF, Merchant NC. (2008). Machine-learning approaches for classifying haplogroup from Y chromosome STR data. *PLoS Comput Biol.* **4**(6):e1000093.
88. Shi W, Ayub Q, Vermeulen M, Shao R-g, Zuniga S, van der Gaag K, de Kniff P, Kayser M, Xue Y & Tyler-Smith C. (2010). A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. *Mol Biol Evol.* **27**(2):385-393.
89. Shimada I, Shinoda K-i, Farnum J, Corruccini R & Watanabe H. (2004). An integrated analysis of pre-Hispanic mortuary practices: a Middle Sicán case study. *Curr Anthropol.* **45**(3):369-402.

90. Shinoda K-i, Adachi N, Guillen S & Shimada I. (2006). Mitochondrial DNA analysis of ancient Peruvian highlanders. *Am J Phys Anthropol.* **131**:98-107.
91. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V & Richards MB. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet.* **84**:740-759.
92. Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilová E, Macaulay V, Richards MB, Cerny V & Pereira L. (2012). The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol Biol Evol.* **29**(3):915-927.
93. Sukernik RI, Schurr TG, Starikovskaya EB & Wallace DC. (1996). Mitochondrial DNA variation in native Siberians, with special reference to the evolutionary history of American Indians. Studies on restriction polymorphism. *Genetika.* T.32, N3:432-439.
94. Stone A & Stoneking M. (1998). mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *Am J Hum Genet.* **62**:1153-1170.
95. Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, *et al.* (2007). Beringian standstill and spread of Native American founders. *PLoS ONE* **2**(9):e829.
96. Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* **28**(10):2731-2739.
97. Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, Santos FR. (2001). Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet.* **68**:1485-1496.
98. Tito RY, López PW, Iannaccone GC & Lizárraga B. (2004). Genetic diversity of sixteen STRs in the Peruvian Mesa Redonda Lima population. *J Forensic Sci.* **49**(4):1-2.
99. Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC. (1993). Asian affinities and continental radiation of the four founding native American mtDNAs. *Am J Hum Genet.* **53**:563-590.
100. Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL. (1996). A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci. USA.* **93**:196-200.
101. Underhill PA, Sheng P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonnét-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL & Oefner PJ. (2000). Y chromosome sequence variation and the history of human populations. *Nature Genet.* **26**:358-361.
102. Vellard J. (1959). Poblaciones indígenas del Lago Titicaca. Boletín Sociedad Geográfica Lima, Perú. 76 (3-4 trimestre):17-40.

103. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. (1991). African populations and the evolution of human mitochondrial DNA. *Science* **253**:1503-1507.
104. Volodko NV, Starikovskaya EB, Mazunin IO, Eltsov NP, Naidenko PV, Wallace DC & Sukernik RE. (2008). Mitochondrial genome diversity in arctic Siberians, with particular reference to the evolutionary history of Beringia and pleistocenic peopling of the Americas. *Am J Hum Genet.* **82**:1084-1100.
105. Wachtel N. (1986). Men of the water: the Uru problem (XVI and XVII centuries). Cambridge University Press & Editions de la Maison des Sciences de L' Homm.
106. Walsh B. (2001). Estimating the time to the most recent common ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals. *Genetics* **158**:897-912.
107. Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G. *et al.* (2007). Genetic variation and population structure in native Americans. *PLoS Genet.* **3** (11):e185.
108. Wolfgang H, Balanovsky O, Sanchez J, Koshel S, Zaporozhchenko V, Adler C, Der Sarkissian C, Brandt G, Schwarz C, Nicklisch N, Dresely V, Fritsch B, Balanovska E, Villems R, Meller H, Alt K, Cooper A. & The Genographic Consortium (2010). Ancient DNA from European early neolithics farmers reveals their near eastern affinities. *PLoS Biol.* **8**(11):e1000536.
109. Y Chromosome Consortium (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**:339-348.
110. Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV & Tyler-Smith C. (1997). Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet.* **60**:1174-1183.

10 – ANEXOS

Tabela A1: Haplótipos (Hp) com seus respectivos SNPs / variantes na Região Controle mtDNA do sub-haplogrupo A2 nas populações peruanas (n=90). Os sítios ou posições nucleotídicas são anotadas segundo o rCRS (revised Cambridge Reference Sequence). +=inserção; d=deleção; as transversões são anotadas depois do sítio com seu variante, com letra.

Local	Hp	Site/variante
And	1	64, 73, 146, 153, 235, 263, 309+CC, 315+C, 523-524d, 16111, 16223, 16264, 16290, 16319, 16362, 16519
And	2	64, 73, 146, 153, 235, 263, 309+CC, 315+C, 356+C, 499, 523-524d, 16111, 16223, 16290, 16319, 16362, 16477
And	3	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16264, 16290, 16319, 16362, 16519
And	4	73, 146, 153, 189, 207, 235, 263, 315+C, 523-524d, 16223, 16290, 16293, 16319, 16362
And	5	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16086, 16111, 16223, 16290, 16319, 16362, 16519
And	6	64, 73, 146, 153, 215, 235, 263, 309+C, 315+C, 523-524d, 16092, 16111, 16223, 16290, 16311, 16319, 16362
And	7	64, 73, 103, 146, 152, 153, 183, 235, 263, 309+C, 315+C, 455+T, 523-524d, 16111, 16223, 16290, 16319, 16362
And	8	64, 73, 146, 153, 215, 235, 263, 309+C, 315+C, 523-524d, 16092, 16111, 16193+C, 16223, 16290, 16311, 16319, 16362
And	9	64, 73, 139, 146, 153, 194, 235, 263, 315+C, 523-524d, 16111, 16223, 16290, 16291, 16319, 16362, 16519
And	10	64, 73, 146, 153, 215, 235, 263, 315+C, 374, 523-524d, 16111, 16223, 16290, 16319, 16362
And	11	64, 73, 146, 153, 215, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16290, 16305, 16319, 16362, 16519
lq	12	64, 73, 146, 153, 235, 263, 309+CC, 315+C, 485, 523-524d, 573+C, 16111, 16223, 16290, 16319, 16362, 16519
lq	13	64, 73, 139, 146, 153, 235, 263, 315+C, 523-524d, 16092, 16111, 16189, 16193+C, 16223, 16290, 16319, 16362
lq	14	64, 73, 114G, 146, 153, 235, 263, 310d, 373, 523-524d, 16111, 16223, 16290, 16319, 16362, 16519
Tar	15	64, 73, 114G, 146, 153, 235, 263, 310, 373, 523-524d, 16111, 16223, 16290, 16319, 16362, 16519
Puc	16	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16183C, 16189, 16223, 16239A, 16290, 16319, 16362
Puc	17	64, 73, 146, 153, 235, 263, 309+CC, 315+C, 523-524d, 16111, 16223, 16239A, 16266, 16290, 16319, 16362

Puc	18	64, 73, 146, 153, 235, 263, 315+C, 477, 523-524d, 16111, 16223, 16290, 16311, 16319, 16362, 16519
Chp	19	64, 73, 143, 146, 152, 153, 215, 235, 263, 309+C, 315+C, 523-524d, 16092, 16129, 16223, 16290, 16311, 16319, 16362
Chp	20	64, 73, 146, 153, 235, 263, 315+C, 523-524d, 16111, 16221, 16223, 16263, 16290, 16311, 16319, 16362
Chp	21	64, 73, 146, 153, 235, 263, 315+C, 523-524d, 551, 16092, 16111, 16223, 16290, 16319, 16362, 16448
Chp	22	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16290, 16319, 16362, 16519
Hyo	23	64, 73, 146, 153, 235, 263, 315+C, 523-524d, 16111, 16223, 16290, 16319, 16362, 16519
Tar	24	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16278, 16290, 16319, 16362, 16519
SMLa	25	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16274, 16290, 16319, 16362, 16519
SMLa	26	64, 73, 146, 153, 185, 217, 235, 263, 315+C, 523-524d, 16111, 16214, 16223, 16290, 16319, 16362, 16519
SMLa	27	64, 73, 146, 153, 207, 235, 263, 315+C, 523-524d, 16111A, 16172, 16223, 16266, 16290, 16319, 16362
CA	28	64, 73, 146, 153, 235, 263, 309+CC, 315+C, 523-524d, 16111, 16172, 16188+C, 16223, 16290, 16319, 16325, 16362, 16519
CA	29	64, 73, 146, 153, 195, 215, 235, 263, 315+C, 523-524d, 16111, 16223, 16247T, 16290, 16295, 16319, 16362, 16519
CA	30	64, 73, 146, 153, 235, 263, 315+C, 523-524d, 16111, 16223, 16241, 16290, 16311, 16319, 16362, 16468
LA	31	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16172, 16188+C, 16223, 16290, 16319, 16325, 16362, 16519
CAsm	32	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16092, 16111, 16172, 16223, 16274, 16286, 16290, 16311, 16319, 16362, 16519
CAsm	33	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16095, 16111, 16223, 16290, 16305, 16319, 16362, 16519
CAsm	34	64, 73, 235, 263, 309+C, 315+C, 523-524d, 16092, 16111, 16223, 16290, 16311, 16319, 16362, 16519
ANh	35	64, 73, 125, 127, 146, 153, 235, 263, 315+C, 16092, 16111, 16172, 16188+C, 16223, 16290, 16319, 16362, 16519
LI	36	64, 73, 125, 127, 146, 153, 235, 263, 309+CC, 315+C, 16092, 16111, 16172, 16188+C, 16223, 16290, 16319, 16362, 16519
ANh	37	64, 73, 93, 146, 153, 235, 263, 309+CC, 315+C, 523-524d, 573d, 16111, 16223, 16260, 16290, 16319, 16362
Ch	38	64, 73, 146, 153, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16290, 16311, 16319, 16362, 16519
Ch	39	73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16290, 16305, 16319, 16362, 16519
Hyo	40	64, 73, 146, 153, 235, 263, 315+C, 472, 523-524d, 16111, 16223, 16249, 16290, 16319, 16362
Hyo	41	64, 73, 162, 235, 263, 309+C, 315+C, 523-524d, 16092, 16111, 16223, 16290, 16319, 16362, 16519
Hyo	42	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16188, 16223, 16290, 16319, 16357, 16362, 16519
Hyo	43	64, 73, 146, 153, 235, 263, 315+C, 523-524d, 16104, 16111, 16223, 16290, 16319, 16362,

		16519
Hyo	44	64, 73, 146, 153, 235, 263, 309+CC, 315+C, 523-524d, 16111, 16145, 16223, 16290, 16319, 16362
PI	45	64, 73, 146, 153, 235, 263, 315+C, 523-524d, 16111, 16145, 16223, 16290, 16319, 16362
Hyo	46	64, 73, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16189, 16223, 16241, 16290, 16319, 16362, 16519
Hyo	47	64, 73, 146, 153, 217, 235, 263, 315+C, 523-524d, 16111, 16223, 16290, 16319, 16324, 16362, 16519
Hyo	48	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16239A, 16266, 16290, 16319, 16362, 16519
Hyo	49	64, 73, 146, 153, 235, 263, 309+CC, 315+C, 356+C, 499, 523-524d, 16111, 16144, 16223, 16266, 16290, 16319, 16362, 16437, 16519
HV	50	64, 73, 153, 235, 263, 297, 309+C, 315+C, 523-524d, 16086, 16111, 16129, 16223, 16290, 16319, 16362
HV	51	64, 73, 153, 235, 263, 309+CC, 315+C, 356+C, 523-524d, 16111, 16129, 16223, 16290, 16319, 16362
AY	52	64, 73, 146, 152, 153, 235, 263, 309+CC, 315+C, 523-524d, 16104, 16111, 16223, 16248, 16278, 16290, 16319, 16362, 16519
AY	53	64, 73, 146, 153, 235, 263, 309+C, 315+C, 340, 523-524d, 16111, 16192, 16223, 16290, 16319, 16362
AY	54	64, 73, 153, 195, 235, 263, 297, 309+C, 315+C, 523-524d, 16111, 16223, 16290, 16319, 16362
AY	55	73, 119, 125, 127, 146, 153, 235, 263, 315+C, 523-524d, 16111, 16223, 16261, 16290, 16304, 16319, 16362
AY	56	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16290, 16319, 16362
Ahy	57	64, 73, 146, 153, 200, 235, 309+CC, 315+C, 523-524d, 16092, 16111, 16129, 16179, 16223, 16290, 16291, 16319, 16362, 16468, 16519
Ahy	58	64, 73, 146, 153, 171, 195, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16290, 16311, 16319, 16362
Kaq	59	64, 73, 114G, 146, 153, 200, 235, 263, 310, 523-524d, 16111, 16217, 16223, 16290, 16319, 16362
LI	60	64, 73, 146, 153, 234, 235, 263C, 315+C, 498d, 523-524d, 16111, 16129, 16223, 16290, 16311, 16319, 16352, 16362, 16519
LI	61	106-111d, 146, 153, 235, 263, 309+CC, 315+C, 385, 523-524d, 16111, 16172, 16186, 16223, 16290, 16319, 16360, 16362
LI	62	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16223, 16260, 16290, 16319, 16362
LI	63	73, 146, 153, 263, 309+C, 315+C, 523-524d, 16111, 16290, 16319, 16325, 16362, 16512, 16547, 16551+G
LI	64	64, 73, 139, 146, 153, 235, 263, 309+CC, 315+C, 523-524d, 16111, 16223, 16290, 16319, 16362
LI	65	73, 146, 153, 235, 263, 315+C, 523-524d, 16111, 16189, 16223, 16290, 16319, 16362, 16526
LI	66	73, 146, 153, 235, 263, 315+C, 523-524d, 16111, 16223, 16290, 16319, 16362, 16519
LI	67	64, 73, 153, 235, 263, 309+C, 315+C, 523-524d, 16092, 16111, 16223, 16290, 16319, 16362
LI	68	73, 146, 153, 198, 200, 235, 263, 309+CC, 315+C, 523-524d, 16042C, 16111, 16124, 16189, 16223, 16290, 16319, 16362, 16438, 16519
LI	69	64, 73, 146, 153, 235, 263, 309+C, 315+C, 356+C, 523-524d, 16111, 16129, 16223, 16262, 16290, 16319, 16362
LI	70	64, 73, 146, 153, 215, 235, 263, 315+C, 523-524d, 16095, 16111, 16209, 16223, 16290, 16311, 16319, 16362

LI	71	64, 73, 146, 153, 195, 235, 263, 309+C, 315+C, 523-524d, 16223, 16258, 16290, 16311, 16319, 16362
LI	72	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16266, 16319, 16362, 16390
LI	73	64, 73, 146, 153, 212, 215, 235, 263, 309+C, 315+C, 390, 523-524d, 16111, 16127, 16217, 16223, 16290, 16319, 16362, 16519
LI	74	64, 73, 146, 153, 235, 263, 309d, 315+C, 523-524d, 16111, 16223, 16260, 16290, 16319, 16362
LI	75	64, 73, 146, 153, 235, 263, 309+CC, 315+C, 523-524d, 16111, 16239A, 16266, 16290, 16319, 16362
LI	76	73, 146, 153, 235, 263, 309+CC, 315+C, 523-524d, 16111, 16140, 16183C, 16189, 16223, 16264, 16290, 16319, 16362
Llhr	77	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16104, 16111, 16223, 16290, 16294, 16319, 16362, 16519
LA	78	64, 73, 146, 152, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16274, 16290, 16319, 16362, 16390
LA	79	64, 73, 146, 153, 235, 263, 264, 315+C, 523-524d, 16093, 16111, 16223, 16240, 16290, 16319, 16362, 16519
LA	80	64, 73, 114G, 146, 153, 200, 235, 263, 310d, 523-524d, 16111, 16223, 16290, 16319, 16362
LA	81	64, 73, 146, 153, 235, 263, 315+C, 523-524d, 16086, 16189, 16223, 16274, 16290, 16319, 16362, 16519
LA	82	64, 73, 146, 153, 235, 263, 315+C, 523-524d, 16111, 16223, 16290, 16311, 16319, 16362, 16468
LA	83	64, 73, 151, 153, 235, 263, 309+C, 315+C, 523-524d, 16092, 16111, 16223, 16290, 16319, 16362
Ur	84	64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16223, 16290, 16311, 16319, 16362
Cb	85	64, 71+G, 73, 75, 146, 153, 183, 235, 263, 309+C, 315+C, 523-524d, 16111, 16223, 16290, 16319, 16362
Cb	86	64, 73, 146, 153, 235, 263, 315+C, 523-524d, 16092, 16111, 16213, 16223, 16290, 16319, 16362, 16526
Cy	87	64, 73, 146, 153, 235, 260, 263, 315+C, 523-524d, 16111, 16217, 16223, 16290, 16319, 16343T, 16362
Cy	88	64, 73, 143, 146, 153, 199, 235, 263, 309+C, 315+C, 368, 523-524d, 16041, 16111, 16157, 16223, 16290, 16311, 16319, 16362, 16533
Yke	89	73, 146, 153, 235, 260, 263, 315+C, 523-524d, 16111, 16217, 16223, 16290, 16319, 16343T, 16362, 16519
AR	90	62T, 64, 73, 146, 153, 235, 263, 309+C, 315+C, 523-524d, 16111, 16178, 16192, 16223, 16290, 16319, 16362

Tabela A2: Haplótipos (Hp) com seus respectivos SNPs / variantes na Região Controle mtDNA do sub-haplogrupo B2 nas populações peruanas (n=234). Os sítios ou posições nucleotídicas são anotadas segundo o rCRS (revised Cambridge Reference Sequence). +=inserção; d=deleção; as transversões são anotadas depois do sítio com seu variante, com letra.

Local	Hp	Site/variante
And	1	73, 152, 263, 309+CC, 315+C, 499, 523-524d, 16092, 16183C, 16189, 16217, 16241, 16278, 16324, 16503, 16519
Chp	2	73, 152, 263, 310, 312, 314+TCC, 499, 523-524d, 16092, 16183C, 16189, 16217, 16241, 16278, 16324, 16503, 16519
And	3	73, 263, 309+C, 315+C, 499, 574, 16183C, 16189, 16217, 16221A, 16519
And	4	73, 263, 309+CC, 315+C, 499, 574, 16183C, 16189, 16217, 16221A, 16519
And	5	73, 143, 189, 240, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16234G
And	6	73, 115, 200, 234, 263, 315+C, 499, 16183C, 16189, 16217, 16519
And	7	55, 73, 199, 207, 263, 315+C, 499, 16173, 16183C, 16189, 16217, 16519
And	8	73, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16290, 16381, 16519
CA	9	73, 263, 309+C, 315+C, 499, 16182C, 16183C, 16189, 16217, 16290, 16381, 16519
And	10	73, 146, 263, 309d, 315+C, 499, 16051, 16092, 16182C, 16183d, 16189, 16217, 16519
And	11	73, 146, 152, 263, 309+CC, 315+C, 373, 499, 523-524d, 16051, 16129, 16182C, 16183C, 16189, 16217, 16316, 16360, 16519
And	12	73, 152, 263, 309+CC, 315+C, 499, 523-524d, 16051, 16183C, 16189, 16217, 16360, 16519
Ch	13	73, 152, 263, 309+C, 315+C, 499, 523-524d, 16051, 16183C, 16189, 16217, 16360, 16519
And	14	73, 153, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16221, 16519
And	15	73, 263, 283, 309+C, 315+C, 499, 574, 16183C, 16189, 16217, 16221A, 16519
And	16	73, 153, 195, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16260, 16437, 16519
And	17	73, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
Ch	18	73, 263, 309d, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
Ahy	19	73, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16519
LI	20	73, 263, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
LI	21	60, 73, 263, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
Ap	22	73, 263, 309d, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
Puc	23	73, 152, 263, 309+C, 315+C, 499, 16051, 16129, 16183C, 16189, 16217, 16519
Puc	24	73, 93, 143, 152, 263, 309+C, 315+C, 499, 16092, 16145, 16183C, 16189, 16217, 16357, 16519
Puc	25	73, 200, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16297, 16519, 16566
Puc	26	64, 73, 198, 263, 309+C, 315+C, 499, 16129, 16183C, 16189, 16217, 16357, 16519
Chp	27	73, 152, 263, 309+CC, 315+C, 499, 16156, 16183C, 16189, 16217, 16519
Chp	28	73, 139, 263, 309+C, 315+C, 499, 16129, 16182C, 16183C, 16189, 16217, 16278, 16519, 16527
Chp	29	73, 146, 263, 309+CC, 315+C, 499, 16093, 16183C, 16189, 16217, 16266, 16485, 16519

Chp	30	73, 263, 309+C, 315+C, 499, 16129, 16183C, 16189, 16217, 16290, 16381, 16519
Chp	31	73, 200, 263, 309+CC, 315+C, 499, 16102, 16183C, 16189, 16217, 16297, 16519, 16566
Tar	32	73, 209, 263, 315+C, 499, 16183C, 16189, 16217
Tar	33	73, 263, 309+C, 315+C, 499, 16182C, 16183T, 16189, 16217, 16356, 16390, 16519
SMLa	34	73, 263, 315+C, 499, 16051, 16086, 16183C, 16189, 16217, 16271, 16519
SMLa	35	73, 263, 315+C, 499, 16051, 16086, 16183C, 16189, 16217, 16519
SMLa	36	73, 263, 309+C, 315+C, 499, 16051, 16086, 16183C, 16189, 16217, 16519
SMLa	37	73, 217, 263, 309+C, 315+C, 499, 16168, 16183C, 16189, 16217, 16270, 16278, 16428, 16519
CA	38	73, 263, 309+C, 315+C, 499, 16051, 16183C, 16189, 16217, 16519
LA	39	73, 263, 315+C, 499, 16051, 16182C, 16183C, 16189, 16217, 16519
CA	40	73, 146, 263, 309+C, 315+C, 499, 16092, 16183C, 16189, 16217, 16519
CA	41	65G, 73, 75, 263, 310, 499, 16183C, 16189, 16217, 16261, 16319, 16362, 16519
Cy	42	71+G, 73, 75, 263, 315+C, 340, 499, 16183C, 16189, 16217, 16261, 16319, 16519
CA	43	73, 195, 263, 309+CC, 315+C, 499, 16037, 16182C, 16183d, 16189, 16217, 16293, 16366, 16519
CA	44	73, 263, 309+C, 315+C, 482, 499, 16183C, 16189, 16217, 16266, 16519
CA	45	73, 143, 263, 309+C, 315+C, 499, 16183C, 16189, 16217
CA	46	73, 152, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
Ch	47	73, 152, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16519
AY	48	73, 152, 263, 309+CC, 315+C, 499, 524+ACAC, 16183C, 16189, 16217, 16519
LA	49	73, 152, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16519
CA	50	73, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217
CA	51	73, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16233, 16290, 16519
CA	52	10, 73, 251, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16319, 16519, 16524
CA	53	73, 263, 309+CC, 315+C, 499, 16037, 16168, 16172, 16183C, 16189, 16217, 16357, 16519
CA	54	73, 263, 307-356d, 499, 16182C, 16183C, 16189, 16217, 16352, 16399, 16519
CA	55	73, 243, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16519
CAsm	56	73, 263, 315+C, 499, 16183C, 16189, 16217, 16290, 16519
CAsm	57	73, 146, 183, 263, 310, 312, 499, 16093, 16183C, 16189, 16193+C, 16217, 16287, 16519
CAsm	58	73, 103, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16266, 16278
CAsm	59	73, 139, 189, 215, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16357, 16399, 16519
CAsm	60	73, 251, 263, 271, 279, 309+C, 315+C, 499, 16183C, 16189, 16217, 16218+T, 16304, 16319, 16519
ANh	61	73, 106, 146, 214, 215, 263, 315+C, 499, 16182C, 16183C, 16189, 16217, 16242, 16324, 16516, 16519
Ch	62	73, 139, 263, 309+C, 315+C, 499, 16092, 16182C, 16183C, 16189, 16217, 16422, 16438, 16519
Ch	63	73, 90C, 98, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16310, 16519
Ch	64	73, 103, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16266
Ch	65	73, 239, 263, 309d, 315+C, 499, 16183C, 16189, 16217, 16519
Ch	66	73, 103, 263, 309+C, 315+C, 318, 499, 523-524d, 16183C, 16189, 16217, 16266, 16519
Ch	67	73, 146, 263, 309+CC, 315+C, 454, 499, 16092, 16183C, 16189, 16217, 16519
Oco	68	73, 152, 199, 263, 309+C, 315+C, 499, 16092, 16183C, 16189, 16217, 16234, 16456,

		16519
Oco	69	73, 152, 263, 309+CC, 315+C, 499, 16092, 16145, 16182C, 16183C, 16189, 16217, 16519
Oco	70	73, 146, 263, 309d, 315+C, 499, 16183C, 16189, 16217, 16456
Oco	71	55+T, 57, 59, 73, 146, 203C, 263, 309+C, 315+C, 499, 522, 16183C, 16189, 16217, 16316, 16519
Oco	72	73, 195, 263, 309+CC, 315+C, 499, 524+ACAC, 16140, 16182C, 16183d, 16189, 16217, 16519
Oco	73	73, 125, 127, 152, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16519
Oco	74	73, 146, 215, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16242, 16324, 16519
Llhr	75	73, 146, 215, 263, 309+C, 315+C, 499, 16182C, 16183C, 16189, 16217, 16242, 16324, 16519
Oco	76	73, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16256, 16265C, 16519
Hyo	77	73, 263, 309+C, 315+C, 499, 16051, 16182C, 16183C, 16189, 16217, 16391, 16519
Hyo	78	73, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16294, 16519
Hyo	79	65G, 73, 75, 94, 263, 309+CC, 315+C, 430, 455+T, 499, 16183C, 16189, 16197G, 16212, 16217, 16261, 16319, 16519
Hyo	80	73, 204, 263, 309+CC, 315+C, 499, 16093, 16183C, 16189, 16217, 16519
Hyo	81	73, 228T, 263, 315+C, 499, 16051, 16182C, 16183C, 16189, 16217
Hyo	82	73, 263, 309+C, 315+C, 499, 533, 573+C, 16182C, 16183C, 16189, 16217, 16519
Hyo	83	73, 143, 263, 315+C, 499, 16183C, 16189, 16217, 16288, 16519
Hyo	84	73, 263, 309+CC, 315+C, 499, 523-524d, 16140, 16182C, 16183C, 16189, 16217, 16318, 16437, 16519
Hyo	85	73, 263, 309+C, 315+C, 499, 16168, 16182C, 16183C, 16189, 16217, 16362, 16519
AY	86	73, 263, 315+C, 499, 16168, 16182C, 16183C, 16189, 16217, 16519
AY	87	73, 263, 309d, 315+C, 499, 16168, 16183C, 16189, 16217, 16519
Ahy	88	73, 263, 309+CC, 315+C, 499, 16168, 16183C, 16189, 16217, 16519
LI	89	73, 263, 310, 499, 16168, 16182C, 16183d, 16189, 16217, 16519
HV	90	73, 151, 152, 178, 200, 204, 263, 309+C, 315+C, 499, 524+AC, 16183C, 16189, 16217, 16223, 16311, 16356, 16519
HV	91	73, 152, 199, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16234, 16456, 16519
HV	92	73, 263, 309+C, 315+C, 499, 524+AC, 16183C, 16189, 16217, 16304, 16390, 16519
AY	93	73, 94, 146, 263, 309+CC, 315+C, 499, 524+ACAC, 16183C, 16189, 16217, 16269, 16519
AY	94	73, 151, 195, 215, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16311, 16519
AY	95	65+T, 73, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16325, 16519
AY	96	73, 146, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16519
LI	97	73, 146, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
Amt	98	73, 146, 263, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
AY	99	60, 73, 309+CC, 315+C, 499, 523-524d, 16124, 16182C, 16183C, 16189, 16217, 16512, 16516, 16519
AY	100	73, 103, 143, 263, 310-311d, 499, 16093, 16182C, 16183C, 16189, 16217, 16258, 16357, 16519
AY	101	44, 73, 263, 309+CC, 315+C, 499, 16147, 16183C, 16189, 16217, 16258T, 16270, 16519
AY	102	73, 263, 309d, 315+C, 499, 573+C, 16168, 16183C, 16189, 16217, 16357, 16502, 16519
AY	103	73, 263, 309+C, 315+C, 499, 524+ACAC, 16129, 16183C, 16189, 16217, 16519

AY	104	42, 73, 143, 210, 263, 309+C, 315+C, 499, 16066, 16183d, 16186, 16189, 16217, 16519
AY	105	73, 263, 315+C, 499, 16168, 16183C, 16189, 16217, 16261, 16519
AY	106	73, 234, 263, 309+CC, 315+C, 499, 16168, 16182C, 16183C, 16189, 16217, 16245G, 16258C, 16265C, 16269C, 16519
Ahy	107	36, 73, 178, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16362, 16519
Ahy	108	72, 73, 263, 309+CC, 315+C, 334, 499, 16183C, 16189, 16217, 16519
Ahy	109	73, 186, 204, 207, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16519
Ahy	110	73, 204, 207, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16290, 16519
Ahy	111	73, 143, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16289, 16519
Ahy	112	73, 263, 309+C, 315+C, 499, 16168, 16183C, 16189, 16217, 16357, 16519
Ahy	113	60, 73, 150, 236, 263, 315+C, 499, 16183C, 16189, 16217, 16519
Ahy	114	65G, 73, 146, 185, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16261, 16319, 16519
Ahy	115	65G, 73, 146, 185, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16261, 16319, 16519
Ahy	116	73, 263, 315+C, 499, 16093, 16182C, 16183C, 16189, 16217, 16247, 16325, 16519
Ahy	117	73, 263, 315+C, 499, 16093, 16182C, 16183d, 16189, 16217, 16247, 16325, 16519
Ahy	118	56T, 57, 58G, 60, 61, 65G, 73, 103, 200, 263, 309+CC, 315+C, 499, 16086, 16183C, 16189, 16217, 16220, 16519
Ahy	119 (B2h)	73, 143, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16293, 16311, 16468, 16519
Ahy	120	73, 204, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
Llhr	121	73, 204, 263, 309+C, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
Ahy	122	73, 146, 263, 315+C, 499, 545, 16183C, 16189, 16217, 16519
Ahy	123	73, 263, 309+CC, 315+C, 499, 16168, 16183C, 16189, 16217, 16295, 16519
Ahy	124	73, 185, 263, 309+C, 315+C, 499, 16168, 16183C, 16189, 16217, 16519
Ahy	125	73, 263, 309+C, 315+C, 499, 16182C, 16183C, 16189, 16217, 16258T, 16270, 16311, 16519, 16545
Ahy	126	73, 263, 315+C, 499, 16093, 16182C, 16183C, 16189, 16217, 16247, 16519
Ahy	127	73, 263, 309+C, 315+C, 499, 573+C, 16183C, 16189, 16217, 16310, 16519
Ahy	128	73, 103, 143, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16317, 16519
Ahy	129	73, 204, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16390, 16519
Ahy	130	60+T, 73, 195, 263, 309+CC, 315+C, 499, 513, 523-524d, 16182C, 16183C, 16189, 16217, 16519
Kaq	131	65G, 73, 263, 309+C, 315+C, 499, 16183C, 16189, 16212, 16217, 16261, 16319, 16519
Llhr	132	71+G, 73, 75, 94, 263, 315+C, 499, 16183C, 16189, 16217, 16261, 16319, 16519
Llhr	133	71+G, 73, 75, 94, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16261, 16319, 16519
LI	134	63, 64, 66, 73, 185, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
LI	135	73, 195, 263, 309+CC, 315+C, 499, 524+AC, 16182C, 16183C, 16189, 16217, 16519
LI	136	73, 263, 309+CC, 315+C, 499, 16086, 16182C, 16183C, 16189, 16217, 16519
LI	137	73, 195, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16437, 16519
LI	138	73, 103, 143, 199, 207, 263, 309+CC, 315+C, 499, 524+ACAC, 16129, 16183C, 16189, 16217, 16266
LI	139	73, 139, 263, 315+C, 499, 16182C, 16183C, 16189, 16217, 16422, 16519
LI	140	73, 263, 309d, 315+C, 499, 16129C, 16182C, 16183C, 16189, 16217, 16519, 16526
LI	141	73, 185, 263, 315+C, 499, 522, 16183C, 16189, 16217, 16316, 16519

LI	142	61, 62C, 63G, 64, 65, 66T, 73, 263, 309+CC, 315+C, 499, 513, 523-524d, 16129, 16182C, 16183C, 16189, 16217, 16519
LI	143	73, 186, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
LI	144	46G, 73, 146, 263, 309+C, 315+C, 499, 16093, 16140, 16183d, 16189, 16194C, 16195, 16217, 16505, 16519
LI	145	65G, 73, 75T, 217, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16261, 16319, 16519
LI	146	73, 228T, 263, 309d, 315+C, 499, 16182C, 16183d, 16189, 16217, 16474C, 16519
LI	147	73, 263, 309+C, 315+C, 499, 16168, 16183C, 16189, 16192, 16217, 16519
LI	148	71d, 73, 143, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16519
LI	149	73, 263, 309+CC, 315+C, 499, 16124, 16182C, 16183C, 16189, 16217, 16516, 16519
LI	150	73, 146, 263, 309+C, 315+C, 499, 16093, 16183C, 16189, 16217, 16311
LI	151	73, 152, 263, 309+C, 315+C, 499, 16168, 16183C, 16189, 16217, 16519
LI	152	73, 139, 152, 263, 315+C, 499, 16092, 16126, 16182C, 16183C, 16189, 16217, 16519
LI	153	73, 146, 263, 309+C, 315+C, 499, 16092, 16095, 16182C, 16183C, 16189, 16217, 16519
LI	154 (B4g)	51, 59, 60G, 61, 62C, 64, 65G, 73, 103T, 263, 310, 499, 16183C, 16189, 16217, 16261, 16292, 16319, 16320, 16519
Llhr	155	73, 94, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16519
Llhr	156	73, 263, 309d, 315+C, 499, 16182C, 16183C, 16189, 16217, 16274, 16437, 16519
Llhr	157	73, 143, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16289, 16512, 16519
Llhr	158	41, 73, 146, 207, 263, 315+C, 393, 499, 16182C, 16183C, 16189, 16217, 16221, 16519
LA	159	55, 56C, 73, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
LA	160	73, 146, 263, 309+CC, 315+C, 499, 16173, 16182C, 16183C, 16189, 16217
LA	161	73, 103, 199, 207, 263, 309+CC, 315+C, 499, 524+ACAC, 16129, 16183C, 16189, 16217, 16266
LA	162	73, 263, 315+C, 499, 523-524d, 16182C, 16183C, 16189, 16217, 16422, 16519
LA	163	73, 103, 152, 199, 207, 225, 263, 309+CC, 315+C, 499, 524+ACAC, 16129, 16183C, 16189, 16217, 16266
LA	164	73, 143, 263, 309+C, 315+C, 499, 16093, 16183C, 16189, 16217, 16288, 16519
LA	165	73, 131, 204, 263, 309+C, 315+C, 499, 16182C, 16183C, 16189, 16217, 16311, 16519
LA	166	73, 226, 263, 309+C, 315+C, 499, 16051, 16183C, 16189, 16217, 16355, 16519
LA	167	73, 139, 185, 195, 214, 263, 309+C, 315+C, 499, 16183C, 16189, 16217, 16357, 16399, 16519
Taq	168	73, 263, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Taq	169	73, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Taq	170	73, 241, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Taq	171	65, 73, 241, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Taq	172	73, 146, 186, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16319, 16354, 16519
Ap	173	73, 146, 186, 263, 309+CC, 315+C, 499, 16183C, 16188, 16189, 16217, 16319, 16354, 16519
Taq	174	60+T, 263, 309+CC, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Amt	175	263, 309+CC, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Amt	176	263, 309+C, 315+C, 499, 16183d, 16217, 16519
Amt	177	263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Taq	178	73, 146, 215, 263, 309+CC, 315+C, 455+T, 499, 16183C, 16189, 16217, 16519
Taq	179	73, 146, 215, 263, 309+C, 315+C, 455+T, 499, 16183C, 16189, 16217, 16519

Taq	180	73, 146, 186, 263, 309+C, 315+C, 316, 499, 16170, 16183C, 16188, 16189, 16217, 16519
Taq	181	73, 146, 186, 263, 309+CC, 315+C, 499, 16183C, 16188, 16189, 16217, 16354, 16519
Taq	182	73, 146, 186, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16354, 16519
Ap	183	73, 146, 186, 263, 315+C, 499, 16183C, 16188, 16189, 16217, 16354, 16519
Taq	184	73, 152, 186, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16362, 16519
Taq	185	73, 146, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16319, 16354, 16519
Ap	186	73, 146, 263, 309+CC, 315+C, 499, 16183C, 16188, 16189, 16217, 16319, 16354, 16519
Amt	187	73, 186, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16354, 16362, 16519
Amt	188	73, 186, 263, 309+CC, 315+C, 499, 16183C, 16188, 16189, 16217, 16295, 16519
Amt	189	73, 186, 263, 309+CC, 315+C, 499, 524+AC, 16183C, 16188, 16189, 16217, 16295, 16519
Amt	190	263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16381, 16519
Amt	191	263, 309+CC, 315+C, 499, 16183C, 16188, 16189, 16217, 16381, 16519
Amt	192	73, 189, 263, 309+C, 315+C, 499, 524+AC, 16183C, 16188, 16189, 16217, 16519
Amt	193	73, 120, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Ur	194	73, 186, 263, 309+C, 315+C, 316, 499, 16170, 16183C, 16188, 16189, 16217, 16519
Ap	195	73, 186, 263, 309+C, 315+C, 499, 16093, 16183C, 16188, 16189, 16217, 16519
Ap	196	73, 151, 263, 309+C, 315+C, 499, 16165, 16183C, 16189, 16217, 16519
Ap	197	73, 186, 194, 263, 309+C, 315+C, 499, 16145, 16183C, 16188, 16189, 16217, 16519
Ap	198	152, 263, 309+CC, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Ap	199	73, 143, 146, 215, 263, 309+C, 315+C, 455+T, 499, 16042, 16111, 16183C, 16189, 16217, 16519
Ap	200	69, 73, 152, 186, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16362, 16519
Ap	201	73, 185, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16256, 16288, 16519
Ap	202	57, 57+G, 73, 103, 200, 263, 309+CC, 315+C, 480, 499, 16182C, 16183C, 16189, 16217, 16519
Ap	203	64, 146, 263, 309+CC, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Ap	204	63, 64, 73, 143, 146, 215, 263, 309+CC, 315+C, 455+T, 499, 16183C, 16189, 16217, 16519
PU	205	200, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16381, 16519
PU	206	63, 64, 66, 73, 185, 263, 309+C, 315+C, 499, 16092, 16183C, 16188, 16189, 16217, 16519
PU	207	73, 146, 186, 263, 309+C, 315+C, 499, 16129, 16183C, 16188, 16189, 16217, 16354, 16519
Cb	208	73, 239, 263, 315+C, 499, 524+AC, 16168, 16183C, 16189, 16217, 16390, 16519
Cb	209	73, 143, 146, 215, 263, 309+CC, 315+C, 455+T, 499, 16183C, 16189, 16217, 16519
Cb	210	73, 139, 263, 309+C, 315+C, 499, 16183C, 16189, 16214, 16217, 16519
Cb	211	73, 93, 195, 263, 309+C, 315+C, 499, 16172, 16183C, 16189, 16217, 16422, 16519
Cb	212	73, 112, 173, 217, 263, 315+C, 499, 16129, 16178, 16182C, 16183d, 16189, 16217, 16249, 16519
Cy	213	56T, 57, 58G, 60, 61, 65G, 73, 103, 200, 263, 309+C, 315+C, 499, 16086, 16182C, 16183C, 16189, 16217, 16220, 16311, 16519
Cy	214	73, 146, 152, 215, 263, 315+C, 455+T, 499, 16182C, 16183C, 16189, 16217, 16291,

		16519
Cy	215	73, 146, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Cy	216	63, 64, 263, 309+CC, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Yke	217	63, 64, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Cy	218	64, 73, 146, 182, 215, 263, 309+C, 315+C, 455+T, 499, 16182C, 16183C, 16189, 16217, 16519
Cy	219 (B2h)	73, 263, 306, 310-312d, 499, 523-524d, 16176, 16182C, 16183C, 16189, 16217, 16319, 16468, 16504, 16519
Cy	220	10, 26, 73, 263, 315+C, 499, 16092, 16183C, 16189, 16217, 16311, 16320, 16519
Yke	221	73, 204, 207, 263, 309+CC, 315+C, 499, 16183C, 16189, 16217, 16519
Yke	222	63, 64, 66, 73, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
AR	223	63, 64, 73, 263, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
Yke	224	73, 263, 309+C, 315+C, 499, 16182C, 16183C, 16189, 16217, 16270, 16438, 16519
Yke	225	73, 143, 263, 309+C, 315+C, 499, 524+AC, 16122, 16145, 16183C, 16189, 16217, 16357, 16519
AR	226	73, 263, 309+CC, 315+C, 499, 16182C, 16183C, 16189, 16217, 16512, 16519, 16544
AR	227	73, 195, 263, 309+CC, 315+C, 499, 16168, 16183C, 16189, 16217, 16301A, 16519
AR	228	146, 263, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
AR	229	73, 204, 263, 309+C, 315+C, 499, 16168, 16183C, 16189, 16217, 16301A, 16519
AR	230	64, 73, 198, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16519
AR	231	73, 200, 263, 293, 309+CC, 315+C, 499, 16176, 16183C, 16189, 16217, 16319, 16422, 16519
AR	232	73, 103, 143, 263, 315+C, 499, 16182C, 16183C, 16189, 16217, 16519
AR	233	73, 146, 153, 215, 263, 309+C, 315+C, 499, 16086, 16182C, 16183C, 16189, 16217, 16242, 16249, 16324, 16519
AR	234	73, 146, 186, 263, 309+C, 315+C, 499, 16183C, 16188, 16189, 16217, 16256, 16354, 16519

Tabela A3: Haplótipos (Hp) com seus respectivos SNPs / variantes na Região Controle mtDNA do sub-haplogrupo C1 nas populações peruanas (n=110). Os sítios ou posições nucleotídicas são anotadas segundo o rCRS (revised Cambridge Reference Sequence). +=inserção; d=deleção; as transversões são anotadas depois do sítio com seu variante, com letra.

Local	Hp	Site/variante
And	1	73, 146, 195, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16309, 16325, 16327
And	2	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16311, 16325, 16327
And	3	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
And	4	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16278, 16298, 16325, 16327, 16519
And	5	73, 146, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 512, 523-524d, 16192, 16223, 16227, 16298, 16325, 16327, 16526
And	6	73, 146, 249d, 263, 290-291d, 309+CC, 315+C, 489, 493, 523-524d, 16146, 16223, 16294, 16298, 16325, 16327
And	7	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 16223, 16260, 16298, 16325, 16327, 16519
And	8	73, 182, 222, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16399, 16519
lq	9	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16192, 16223, 16298, 16325, 16327
Puc	10	73, 146, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16092, 16223, 16256, 16298, 16319, 16325, 16327
Puc	11	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16188, 16223, 16298, 16325, 16327
Chp	12	73, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
Chp	13	73, 150, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16092, 16223, 16274, 16298, 16325, 16327, 16519
SMIa	14	73, 249d, 263, 290-291d, 309+CC, 315+C, 489, 493, 523-524d, 16199, 16223, 16254, 16298, 16325, 16327, 16519
SMIa	15	73, 249d, 263, 290-291d, 309+CC, 315+C, 316, 489, 493, 523-524d, 16178, 16223, 16298, 16325, 16327, 16519
CA	16	60+T, 73, 249d, 263, 290-291d, 309+C, 315+C, 489, 523-524d, 16051, 16086, 16147, 16223, 16298, 16325, 16327
CA	17	73, 131, 195, 249d, 263, 290-291d, 315+C, 316, 489, 16078, 16223, 16298, 16325, 16327
CA	18	73, 152, 195, 249d, 263, 290-291d, 315+C, 489, 16223, 16298, 16325, 16327
CA	19	73, 195, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 508, 523-524d, 16092, 16223, 16261, 16298, 16325, 16327, 16474T
CA	20	73, 195, 249d, 263, 290-291d, 309+CC, 315+C, 489, 493, 508, 523-524d, 16092, 16223, 16261, 16298, 16325, 16327, 16474T
CA	21	73, 249d, 263, 290-291d, 315+C, 489, 16223, 16298, 16325, 16327, 16362
CA	22	73, 249d, 263, 290-291d, 315+C, 316, 489, 493, 523-524d, 16178, 16223, 16298, 16325, 16327
CA	23	73, 185, 249d, 263, 290-291d, 309+C, 315+C, 489, 16223, 16298, 16325, 16327, 16519
CA	24	73, 195, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
CAsm	25	73, 249d, 263, 290-291d, 310+T, 315+C, 489, 16093, 16126, 16189, 16223, 16325, 16327, 16362, 16519
CAsm	26	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16319, 16325, 16327
CAsm	27	73, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16438
CAsm	28	60+T, 73, 249d, 263, 290-291d, 309+C, 315+C, 489, 523-524d, 16051, 16086, 16147, 16223, 16298, 16325, 16327

CAsm	29	73, 189, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16092, 16095, 16223, 16261, 16298, 16325, 16327, 16474T
ANh	30	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
ANh	31	73, 152, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16129, 16172, 16223, 16256, 16298, 16325, 16327, 16361
ANh	32	73, 194, 249d, 263, 290-291d, 309+C, 315+C, 489, 523-524d, 16051, 16223, 16293, 16298, 16325, 16327, 16519
Ch	33	73, 195, 249d, 263, 290-291d, 309+C, 315+C, 335, 489, 16223, 16298, 16325, 16327
Hyo	34	73, 249d, 290-291d, 315+C, 489, 493, 523-524d, 16223, 16293, 16298, 16325, 16327, 16519
Hyo	35	73, 185, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16126, 16183C, 16189, 16223, 16298, 16325, 16327, 16344, 16519
Hyo	36	73, 146, 195, 249d, 263, 290-291d, 309+C, 315+C, 489, 523-524d, 16131, 16223, 16298, 16325, 16327
Hyo	37	73, 214, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16145, 16192, 16223, 16298, 16325, 16327, 16519
Hyo	38	73, 200, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16094, 16223, 16249, 16298, 16325, 16327, 16362
Hyo	39	73, 201, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16192, 16223, 16274, 16298, 16311, 16325, 16327
Hyo	40	73, 146, 195, 249d, 263, 290-291d, 309+CC, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16537
Hyo	41	73, 200, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16094, 16223, 16249, 16298, 16325, 16327, 16362
Hyo	42	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 16131, 16223, 16291, 16298, 16325, 16327
Hyo	43	55, 57, 73, 143, 150, 201, 249d, 263, 290-291d, 309+C, 315+C, 325, 377, 489, 493, 523-524d, 16189, 16192+T, 16223, 16266, 16298, 16316, 16325, 16327, 16519
AY	44	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16311, 16325, 16327
AY	45	73, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 573+C, 16189, 16223, 16298, 16325, 16327
AY	46	73, 152, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16519
AY	47	55, 57, 73, 143, 150, 201, 249d, 263, 290-291d, 309+C, 315+C, 325, 377, 489, 493, 523-524d, 16189, 16192, 16223, 16266, 16298, 16316, 16325, 16327, 16519
AY	48	73, 249d, 290-291d, 315+C, 489, 493, 519T, 523-524d, 16180, 16223, 16298, 16325, 16327
AY	49	73, 150, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16311, 16325, 16327
AY	50	73, 146, 249d, 263, 290-291d, 309+C, 315+C, 489, 523-524d, 16051, 16209, 16223, 16298, 16311, 16325, 16327, 16519
AY	51	71, 73, 234, 235, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 16086, 16148, 16179, 16223, 16298, 16325, 16327, 16519
AY	52	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16215, 16223, 16298, 16325, 16327
AY	53	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16390
Ahy	54	72, 73, 150, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16368, 16519
Ahy	55	63, 64, 66, 73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
Ahy	56	73, 249d, 263, 290-291d, 309+CC, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
Ahy	57	73, 195, 201, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16192, 16223, 16274, 16298, 16311, 16325, 16327
Ahy	58	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16293, 16298, 16325, 16327
Ahy	59	73, 152, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
Ahy	60	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16129, 16223, 16286, 16298, 16325, 16327
Ahy	61	73, 150, 249d, 263, 290-291d, 309d, 315+C, 489, 493, 523-524d, 16223, 16298, 16311, 16325, 16327
Kaq	62	73, 79T, 146, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16124, 16223, 16298, 16325, 16327

Kaq	63	73, 146, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
Kaq	64	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16384
Kaq	65	73, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
LI	66	73, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16051, 16223, 16291, 16298, 16325, 16327
LI	67	73, 249d, 263, 289, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16298, 16325, 16327, 16390
LI	68	73, 194, 249d, 263, 290-291d, 315+C, 489, 523-524d, 16048, 16051, 16219, 16223, 16298, 16311, 16325, 16327, 16519
LI	69	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16261, 16298, 16325, 16327
LI	70	73, 195, 235, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16519
LI	71	73, 143, 152, 249d, 263, 290-291d, 309+C, 315+C, 489, 16092, 16223, 16298, 16311, 16325, 16327
LI	72	73, 194, 249d, 263, 290-291d, 309+C, 315+C, 489, 523-524d, 16051, 16223, 16298, 16301, 16325, 16327, 16519
LI	73	61, 73, 199, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
LI	74	73, 146, 212, 249d, 263, 290-291d, 309+C, 315+C, 489, 16223, 16298, 16325, 16327, 16343
LI	75	73, 152, 249d, 263, 290-291d, 309+C, 315+C, 489, 16092A, 16223, 16256, 16298, 16325, 16327
LI	76	73, 195, 249d, 263, 290-291d, 315+C, 489, 16189, 16223, 16298, 16325, 16327
LI	77	73, 76, 79T, 198, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16093, 16189, 16223, 16298, 16325, 16327
LI	78	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
Llhr	79	73, 249d, 263, 290-291d, 309+CC, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
Llhr	80	73, 249d, 263, 290-291d, 309+C, 315+C, 482, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16519
Llhr	81	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 16094, 16223, 16256, 16298, 16319, 16325, 16327, 16465, 16519
Llhr	82	73, 129, 150, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16146, 16223, 16298, 16325, 16327, 16468
Llhr	83	73, 146, 152, 214, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16145, 16223, 16270, 16298, 16325, 16327, 16519
PI	84	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 523-524d, 16051, 16086, 16147, 16223, 16298, 16325, 16327
PI	85	73, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16519
PI	86	73, 146, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16092, 16223, 16298, 16325, 16327
LA	87	73, 152, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16066, 16209, 16223, 16298, 16325, 16327
LA	88	73, 143, 249d, 263, 290-291d, 315+C, 477, 489, 493, 523-524d, 16223, 16298, 16325, 16327
LA	89	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16311, 16325, 16327
LA	90	73, 215, 235, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16428, 16430
LA	91	73, 76, 79T, 146, 198, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16093, 16223, 16298, 16325, 16327
Amt	92	73, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16147, 16223, 16257, 16298, 16325, 16327, 16519
Amt	93	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16519
Amt	94	73, 249d, 263, 290-291d, 309+CC, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16519
Ap	95	73, 150, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16223, 16298, 16311, 16325, 16327, 16526
Ap	96	10, 73, 195, 249d, 263, 290-291d, 309+CC, 315+C, 489, 16223, 16298, 16325, 16327, 16519
PU	97	73, 146, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327
PU	98	73, 195, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16129, 16189, 16223, 16294, 16298, 16325, 16327, 16390
Cb	99	73, 201, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16192, 16209, 16223, 16274, 16298,

		16311, 16325, 16327
Cb	100	73, 152, 249d, 263, 290-291d, 309+C, 315+C, 489, 16223, 16298, 16325, 16327, 16381, 16519
Cb	101	73, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16223, 16298, 16325, 16327, 16390
Cy	102	73, 194, 207, 249d, 263, 290-291d, 315+C, 489, 523-524d, 16051, 16093, 16183d, 16223, 16298, 16325, 16519
Cy	103	73, 249d, 263, 290-291d, 315+C, 489, 16093, 16189, 16193+C, 16223, 16298, 16325, 16327, 16381, 16519
Yke	104	64, 73, 194, 249d, 263, 290-291d, 315+C, 489, 523-524d, 16051, 16223, 16240, 16298, 16325, 16327, 16362, 16519
Yke	105	73, 195, 204, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16198, 16223, 16274, 16298, 16311, 16325, 16327
Yke	106	73, 194, 249d, 263, 290-291d, 315+C, 385, 489, 523-524d, 16051, 16223, 16298, 16325, 16327, 16519
AR	107	73, 152, 212, 235, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16145, 16189, 16223, 16298, 16325, 16327
AR	108	73, 212, 235, 249d, 263, 290-291d, 309+C, 315+C, 489, 493, 523-524d, 16145, 16189, 16193+C, 16223, 16298, 16325, 16327
AR	109	73, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16103, 16223, 16298, 16325, 16327
AR	110	73, 152, 249d, 263, 290-291d, 315+C, 489, 493, 523-524d, 16037, 16093, 16223, 16298, 16325, 16327

Tabela A4: Haplótipos (Hp) com seus respectivos SNPs / variantes na Região Controle mtDNA do sub-haplogrupo D1 nas populações peruanas (n=65). Os sítios ou posições nucleotídicas são anotadas segundo o rCRS (revised Cambridge Reference Sequence). +=inserção; d=deleção; as transversões são anotadas depois do sítio com seu variante, com letra. Entre parêntese a linhagem diferente de D1.

Local	Hp	Site/variante
And	1 (D5)	67T, 97, 100-105d, 150, 263, 309+CC, 315+C, 489, 16172, 16189, 16223, 16325, 16362
And	2	73, 195, 263, 309+C, 315+C, 489, 16192, 16223, 16268, 16325, 16362
And	3	73, 263, 315+C, 489, 16142, 16223, 16325, 16362, 16456, 16519
And	4	64A, 71d, 73, 263, 309+C, 315+C, 489, 16223, 16301, 16304, 16325, 16362
And	5	73, 263, 315+C, 489, 16142, 16223, 16325, 16362, 16456, 16519
And	6 (D4)	73, 263, 315+C, 489, 16223, 16241, 16301, 16342
And	7	73, 263, 309+C, 315+C, 489, 16325, 16362
And	8 (D4)	73, 146, 263, 309+C, 315+C, 489, 16223, 16241, 16301, 16311, 16342, 16362
Puc	9	73, 203, 204, 263, 309+CC, 315+C, 489, 524+ACAC, 16172, 16223, 16263, 16325, 16362
Chp	10	73, 152, 179, 263, 309+C, 315+C, 489, 524+AC, 16223, 16325, 16362
Chp	11	73, 263, 309+CC, 315+C, 489, 16183C, 16189, 16223, 16249, 16325, 16362
	12	
Tar	(D4)	73, 143, 195, 196, 263, 315+C, 489, 16179, 16223, 16325, 16362d
Tar	13	73, 263, 315+C, 489, 16223, 16250, 16325, 16362, 16519
CA	14	64A, 73, 200, 263, 309+C, 315+C, 489, 16223, 16325, 16362
	15	
CA	(D4)	73, 263, 309+C, 315+C, 489, 16209, 16278, 16325, 16362
CA	16	64A, 66, 73, 200, 263, 315+C, 489, 16192, 16223, 16325, 16362
CA	17	73, 143, 226, 263, 315+C, 318, 489, 16140, 16189, 16223, 16310, 16325, 16362
CA	18	73, 263, 309+TC, 315+C, 489, 523-524d, 16223, 16325, 16362, 16519
CA	19	73, 143, 263, 309+C, 315+C, 489, 16192, 16223, 16325, 16362
CA	20	73, 263, 309+C, 315+C, 489, 16093, 16209, 16325, 16362
CAsm	21	64A, 73, 200, 263, 309+C, 315+C, 489, 16223, 16325, 16362
CAsm	22	73, 263, 309+C, 315+C, 489, 523-524d, 16223, 16319, 16325, 16362, 16519
	23	
CAsm	(D4)	73, 152, 263, 309+CC, 315+C, 489, 16189, 16223, 16241, 16278, 16301, 16316, 16342, 16362
ANh	24	73, 143, 189, 256, 263, 309+C, 315+C, 489, 523-524d, 16093, 16183C, 16189, 16223, 16292, 16325, 16362
Oco	25	73, 204, 263, 309+CC, 315+C, 489, 573+C, 16092, 16183C, 16189, 16209, 16325, 16362
Oco	26	73, 263, 315+C, 456, 489, 16209, 16227, 16325, 16362
Oco	27	73, 146, 263, 315+C, 489, 16051, 16209, 16223, 16291, 16325, 16362
Hyo	28	73, 143, 189, 263, 315+C, 489, 16185A, 16189d, 16223, 16325, 16362
Hyo	29	73, 146, 228, 263, 309+C, 315+C, 489, 16142, 16148, 16223, 16278, 16325, 16356, 16362, 16519
AY	30	73, 143, 263, 315+CC, 489, 523-524d, 16182C, 16183C, 16189, 16223, 16325, 16362
Ahy	31	73, 263, 315+C, 489, 16223, 16325, 16362, 16519, 16527
Kaq	32	55+T, 57, 73, 143, 189, 204, 263, 315+C, 489, 16183C, 16189, 16223, 16292, 16325, 16362
Kaq	33	73, 263, 315+C, 489, 16223, 16325, 16362, 16519, 16527
LI	34	53, 55, 73, 191+A, 263, 315+C, 489, 16145, 16209, 16325, 16362
LI	35	64A, 73, 200, 207, 263, 309+C, 315+C, 481, 489, 16223, 16325, 16362
LI	36	73, 182, 263, 315+C, 489, 16209, 16325, 16362

LI	37	73, 263, 315+C, 489, 16223, 16325, 16362
LI	38	73, 263, 309+C, 315+C, 489, 16223, 16325, 16362, 16519
LI	39	64A, 73, 200, 263, 309+C, 315+C, 489, 16223, 16301, 16304, 16325, 16362, 16400
LI	40	73, 146, 228, 263, 309+CC, 315+C, 489, 16142, 16148, 16223, 16278, 16325, 16356, 16362, 16519
LI	41	73, 263, 315+C, 456, 489, 16209, 16223, 16301, 16325, 16362
LI	42	73, 143, 204, 263, 309+C, 315+C, 489, 16223, 16325, 16362
Llhr	43	73, 195, 263, 309+C, 315+C, 489, 16223, 16325, 16362
Llhr	44	73, 263, 315+C, 385, 489, 16145, 16209, 16325, 16362
Llhr	45	73, 195, 263, 309+C, 315+C, 489, 16223, 16301, 16325, 16362
LA	46	73, 143, 263, 309+C, 315+C, 489, 524+AC, 16145, 16223, 16304, 16325, 16362
LA	47	73, 195, 263, 309+C, 315+C, 489, 523-524d, 16223, 16325, 16362, 16519
LA	48	73, 263, 309+C, 315+C, 489, 523-524d, 16223, 16325, 16362, 16519
LA	49 (D4)	73, 152, 263, 309+C, 315+C, 489, 16051, 16172, 16209, 16223, 16256, 16291, 16325, 16362
LA	50	73, 103, 263, 315+C, 489, 513, 16223, 16325, 16362
LA	51	73, 146, 263, 315+C, 489, 16223, 16325, 16362, 16519, 16527
Ur	52	73, 263, 315+C, 489, 16223, 16325, 16362, 16519, 16527
Ap	53	73, 146, 228, 263, 309+C, 315+C, 489, 16142, 16148, 16223, 16278, 16325, 16356, 16362, 16519
Cb	54	73, 263, 315+C, 489, 16223, 16325, 16362, 16519, 16527
Cb	55	73, 263, 315+C, 489, 16223, 16290+A, 16325, 16362, 16519, 16527
Cb	56 (D4)	73, 103, 152, 263, 315+C, 489, 16223, 16241, 16301, 16342, 16362
Cb	57	73, 263, 315+C, 489, 16142, 16223, 16325, 16362, 16519
Cy	58	73, 143, 189, 263, 309+C, 315+C, 420, 489, 16183C, 16189, 16223, 16325, 16362
Cy	59	73, 263, 315+C, 489, 16174, 16223, 16325, 16362, 16519
Cy	60	71+G, 73, 146, 263, 315+C, 489, 16223, 16325, 16362, 16519, 16527
Cy	61	73, 143, 226, 263, 315+C, 318, 489, 16183C, 16189, 16223, 16310, 16325, 16362
Cy	62	9, 73, 85, 263, 309+C, 315+C, 489, 502, 16189, 16223, 16325, 16352, 16362
Yke	63	73, 94, 263, 315+C, 489, 16223, 16292, 16325, 16362, 16519, 16527
AR	64	73, 143, 189, 207, 263, 309+C, 315+C, 489, 16183C, 16189, 16223, 16325, 16362
AR	65	73, 263, 309+C, 315+C, 489, 16223, 16325, 16362, 16519, 16527

Tabela A5: Predição de haplogrupos para 34 amostras para as quais foram determinados seus correspondentes *clusters* na árvore filogenética mitocondrial. HPT=Haplogroup Prediction Tool do projeto Genográfico; n.d.=não determinado.

Amostra	Ferramentas de predição de haplogrupos			Cluster/Hg determinado
	HPT	MitoTool	mtPHYL	
Chp_1	M;D1	M;D1	M	MD
And_2	L0	L0	L0	L0
ANh_3	M;D1	M;G;D	M	MD
Ahy_4	G1a1;D1	G1a1	G1a1	G1a1
AR_5	H	U4	H	HU
AR_6	L3	L3	L3	L3
CA_7	HV	HV	HV	HV
CA_8	J2	J2	J2	J2
CA_9	M;D1	M;D1	D1a1	MD
Iq_10	M;D1	M;D1	M	MD
Hyo_11	M;G;D	M;G;D1	M	MD
Kaq_12	M;D1	M;D1	M	MD
LA_13	L3	L3	L3	L3
LA_14	L3	L3	L3	L3
LI_15	L3	L3	L3	L3
LI_16	L3	L3	L3	L3
LI_17	L3	L3	L3	L3
LI_18	M	M	M	M
LI_19	L3	L3	L3	L3
LI_20	L3	L3	L3	L3
LI_21	L1	L1	L1	L1
LI_22	L2	L2	L2	L2
Tar_23	M;D1	M;D1	M	MD
Chp_24	H	R;M;G;D	H	n.d (s6)
Chp_25	H	G1a1	H	n.d (s7)
And_26	H	B4b;B2;U2;U9	HV	n.d. (s1)
AR_27	R;B4b	R;B4b;B2	R	n.d. (s38)
CA_28	H	P;M;D1;D4	P	n.d. (s9)
CA_29	G1a1	G1a1	N	n.d. (s10)
CA_30	M	G1a1;D5	D1	n.d. (s13)
CA_31	M	G1a1;D5	D1	n.d. (s20)
CA_32	M	G1a1;D5	D1	n.d. (s16)
LA_33	M;D1	M;D1;D4	L3	n.d. (s35)
LA_34	H	H;R11'B6;B4'5;	-	n.d. (s37)

Tabela A6: Lista dos haplótipos únicos de STR do haplogrupo Q-M3 do cromossomo Y. A ordem dos 17 Y-STRs corresponde da esquerda à direita: DYS393-DYS390-DYS19-DYS391-DYS385a-DYS385b-DYS439-DYS389a-DYS392-DYS389b-DYS458-DYS437-DYS448-GATAH4-DYS456-DYS438-DYS635.

Local/amostra	Número de haplótipos	Haplótipo
And2	3	13 22 14 10 15 16 12 13 15 18 17 14 21 12 15 12 22
And3	1	13 24 13 10 14 18 13 12 15 17 20 14 20 12 17 11 22
And4	1	14 23 13 10 11 17 13 14 14 17 17 14 19 12 16 10 22
And7	1	13 24 13 10 17 18 13 14 14 18 15 14 20 12 15 11 22
And8	1	13 24 13 10 17 18 13 14 14 17 16 14 20 12 15 11 22
And9	1	14 23 13 10 13 18 13 14 14 17 16 14 19 12 18 11 22
And11	1	13 23 13 10 14 14 10 13 14 16 15 14 19 12 16 11 25
And13	1	13 23 13 10 13 18 11 13 14 18 18 15 20 12 15 11 22
And14	2	13 23 13 10 14 18 12 14 14 18 16 14 20 12 15 11 22
And15	1	13 25 13 10 14 20 11 13 14 17 16 14 19 12 15 11 22
And16	3	13 23 13 11 14 14 12 13 14 16 15 14 20 12 16 11 22
And17	1	13 23 14 10 14 17 11 14 14 18 17 14 20 11 15 11 22
And18	1	13 23 13 11 14 14 12 12 14 16 15 14 20 12 16 11 22
And20	2	13 24 13 10 14 18 11 13 15 19 18 14 20 12 17 12 22
And21	1	13 22 14 10 15 17 12 13 15 18 17 14 21 12 15 12 22
And26	1	13 23 14 11 14 17 12 14 14 16 15 14 20 12 16 9 22
And29	1	14 23 12 10 14 17 13 14 14 18 18 14 20 11 16 11 22
And30	2	14 23 12 10 14 17 13 14 14 18 18 14 20 11 17 11 22
And31	1	14 23 13 10 14 16 12 12 15 17 15 14 20 11 15 11 22
And38	1	13 24 13 10 17 18 12 13 14 18 17 14 20 11 15 11 22
And39	1	14 23 14 10 14 17 12 13 14 16 14 14 21 11 16 12 22
And40	1	14 24 12 10 14 19 13 14 14 18 19 14 20 11 17 11 22
And42	1	13 23 14 10 13 18 12 12 14 17 15 14 20 12 15 11 22
Puc2	1	13 23 14 10 14 16 12 13 14 16 15 14 20 12 17 12 22
Puc3	1	15 23 13 10 13 17 12 14 14 18 15 14 19 12 16 10 23
Puc4	1	13 23 13 10 14 16 12 13 14 16 15 14 20 12 17 12 22
Puc5	1	13 24 13 10 14 16 12 13 14 16 16 14 20 12 18 12 22
Puc6	1	15 23 13 11 13 16 13 13 14 17 16 14 19 12 15 10 22
Chp1	1	13 23 14 11 14 17 12 13 14 16 15 14 20 12 16 9 22
Chp4	1	13 25 13 10 14 16 13 13 14 17 16 14 19 12 15 12 22
Chp6	1	13 22 14 10 14 17 12 13 14 18 17 14 20 11 15 11 23
Chp7	1	14 24 13 10 15 18 13 13 16 17 16 14 20 12 15 11 22
CA11	1	13 23 14 10 15 17 12 13 15 17 15 15 21 11 16 11 22
CA12	1	13 23 13 10 15 18 12 13 14 17 14 14 20 12 15 11 22
CA14	1	13 24 13 10 13 16 9 13 13 18 18 14 20 12 15 11 22
CA22	1	14 23 12 10 14 18 13 14 14 18 19 14 20 11 17 11 22
CA25	1	13 25 14 10 14 15 12 14 14 17 16 14 20 12 15 11 22
CA38	1	13 23 15 10 15 19 11 13 14 17 18 14 20 11 15 11 22
CA41	3	13 23 14 10 14 18 11 13 14 18 17 14 20 11 15 11 22
CA44	1	13 21 13 10 15 17 12 13 14 18 18 14 20 11 15 11 22
CAsm30	1	13 25 13 10 15 18 11 13 14 17 15 14 20 12 15 11 22
CAsm35	1	14 23 13 10 14 18 12 13 14 15 19 14 21 11 15 11 22
CAsm36	1	13 21 13 10 16 18 12 13 14 18 19 14 20 11 15 11 22
Ch1	2	13 21 13 10 15 17 12 13 14 18 19 14 20 11 15 11 22

Ch3	2	13 21 13 10 15 17 13 14 14 18 17 14 21 11 15 11 22
Ch7	1	13 21 13 10 15 17 12 14 14 18 17 14 21 11 15 11 22
Ch8	2	13 21 13 10 14 17 13 14 15 17 17 14 22 11 14 11 22
Ch10	1	13 26 13 10 14 16 11 13 14 17 18 14 19 13 15 12 22
Ch11	1	13 21 13 10 14 17 13 13 15 18 17 14 21 11 14 11 22
Ch27	1	13 24 13 10 13 16 9 13 13 18 16 14 19 12 15 11 22
Ch29	1	13 24 13 10 14 19 12 13 14 17 17 14 20 12 15 11 22
Oco15	1	13 23 14 10 14 16 12 13 14 18 17 14 20 11 16 11 22
Oco16	1	13 24 13 10 13 16 9 13 13 18 17 14 20 13 15 11 22
Oco18	1	14 23 13 10 15 19 14 14 16 17 16 14 20 12 15 11 22
Oco21	1	13 24 13 10 11 19 13 13 14 17 17 14 20 12 15 11 22
Oco24	1	13 23 14 10 14 17 12 13 14 18 17 14 20 11 16 11 22
Oco25	1	13 24 15 10 15 18 12 13 14 17 15 14 20 13 15 11 22
Hyo4	1	13 24 13 10 14 15 13 13 14 17 18 14 20 12 15 11 22
Hyo5	1	14 24 13 9 14 19 11 12 14 17 17 14 20 11 14 11 22
Hyo6	1	14 23 13 10 15 17 14 14 14 17 17 14 20 12 15 11 22
Hyo8	1	13 23 15 10 15 18 12 13 14 17 17 14 20 11 15 11 22
Hyo11	1	13 21 13 10 15 16 11 13 14 18 18 14 20 11 15 11 22
Hyo13	1	14 23 13 10 15 18 11 13 16 17 15 14 20 12 15 11 22
Hyo14	1	13 23 13 10 13 18 12 13 14 17 15 14 20 12 16 12 22
Hyo15	1	13 24 13 10 13 13 12 13 13 17 18 14 20 11 15 12 22
Hyo16	1	13 23 13 10 15 18 11 13 14 18 16 14 19 11 17 11 22
Hyo19	1	13 23 13 10 14 18 12 14 14 17 17 14 20 12 15 11 22
Hyo20	1	13 23 13 10 14 17 11 13 14 17 18 14 21 13 15 12 22
AY2	1	13 25 14 11 15 16 13 12 14 18 18 14 19 12 15 11 24
AY4	1	14 23 13 10 15 17 13 14 15 17 16 14 20 12 15 11 22
AY8	1	13 25 13 10 15 22 14 13 14 17 18 14 19 13 15 10 22
AY9	1	13 21 13 10 15 17 12 14 14 19 19 14 20 11 15 11 22
AY12	1	13 25 12 10 14 19 13 14 14 17 18 14 20 11 15 9 22
Kaq1	1	13 24 13 10 13 17 9 13 13 17 16 14 19 13 15 11 22
Kaq2	1	14 23 13 10 15 18 13 14 16 17 16 14 20 12 15 11 22
Kaq5	1	13 24 13 10 14 16 14 13 14 18 15 14 20 13 14 11 22
Kaq8	1	14 22 13 10 15 19 12 14 16 17 16 14 20 12 15 11 22
Ahy10	1	13 22 14 10 14 18 14 12 14 16 17 14 20 13 15 11 23
Ahy16	1	13 25 13 11 15 17 12 13 14 16 16 14 19 11 15 11 23
Ahy17	1	13 25 13 10 15 18 13 13 15 17 16 14 22 12 16 11 22
Ahy18	1	14 23 13 10 15 19 14 14 16 17 16 14 19 11 15 11 22
Ahy23	1	13 23 13 10 14 17 12 13 15 18 16 14 20 11 15 12 22
Llhr2	1	13 24 13 10 14 15 12 13 14 17 17 14 20 12 16 11 22
Llhr4	2	13 24 13 10 14 18 11 13 14 17 17 14 20 12 15 11 22
Llhr8	1	13 23 13 10 13 17 11 13 14 16 16 14 20 11 15 11 22
Llhr9	1	13 25 14 11 14 14 12 13 14 17 17 14 20 12 15 11 22
AR3	1	14 23 13 10 15 19 14 14 16 16 16 14 20 12 15 11 22
AR6	2	13 24 13 10 13 14 12 13 14 17 17 14 23 11 16 11 22
AR8	1	13 25 13 10 13 14 12 13 14 18 19 14 20 11 16 11 22
AR9	1	14 23 13 10 15 18 13 13 16 16 17 14 20 12 15 11 22
AR10	1	13 24 13 10 13 14 11 13 14 18 20 14 20 11 15 11 22
AR11	1	14 23 13 10 15 18 12 14 16 17 16 14 20 12 15 11 22
AR12	1	13 24 13 10 13 14 12 13 14 18 19 14 20 11 15 11 22
AR13	1	13 24 13 10 13 16 9 13 13 18 15 14 19 12 15 11 23
AR14	1	14 24 13 11 14 17 11 13 14 16 17 14 20 13 14 11 22
Cy16	1	13 24 13 10 13 17 11 13 14 17 16 14 21 12 15 11 22
Cy17	1	13 24 13 10 13 17 11 14 14 17 18 14 20 12 15 9 22
Cy18	1	13 24 13 10 15 18 13 13 14 16 16 14 20 12 11 11 22

Cy19	1	14 23 13 10 15 20 14 14 16 17 16 14 20 12 15 11 22
Cy20	1	13 24 13 11 13 17 11 13 14 17 17 14 20 11 15 11 22
Cy21	1	13 21 13 10 15 18 13 14 14 17 17 14 20 11 14 11 22
Cy22	1	13 24 13 11 16 17 11 14 14 18 17 14 20 12 15 12 22
Cy33	1	14 23 13 10 15 17 12 14 16 17 16 14 20 12 15 11 22
Cy34	1	14 23 13 10 15 18 12 14 16 17 16 14 20 12 15 11 22
Yke24	1	13 21 13 10 15 17 13 14 14 17 17 14 21 11 14 11 22
AR25	2	13 23 13 10 14 16 12 14 14 16 16 15 19 12 17 13 22
Yke27	1	12 24 13 10 13 16 9 13 13 17 16 14 19 12 15 11 22
Yke28	1	13 23 13 10 17 17 12 13 15 17 16 14 21 12 15 12 24
Yke29	1	14 24 13 10 14 17 11 14 13 16 17 14 20 12 15 11 24
Yke31	1	14 23 13 10 15 18 12 14 16 17 17 14 20 12 15 11 22
Yke32	1	13 21 13 10 15 18 13 14 14 17 17 14 21 11 14 11 22
Cb36	1	14 23 13 10 15 19 13 14 16 16 16 14 20 12 15 11 22
Cb37	1	13 24 13 11 14 20 12 14 13 16 19 14 21 12 15 11 22
Cb39	1	12 23 14 11 14 18 11 13 14 17 18 14 20 11 15 11 22
Cb40	2	13 24 13 11 14 16 13 13 13 18 17 14 20 12 14 11 23
Cb41	1	13 24 13 10 15 17 11 13 14 17 18 14 20 12 15 11 22
Cb45	1	14 23 13 10 15 18 12 14 16 17 16 14 20 12 15 11 22
Cb47	1	13 23 13 10 15 18 12 14 16 17 16 14 20 12 15 11 22
Cb48	2	13 24 13 11 15 16 13 14 14 18 18 14 21 12 15 11 22
Amt1	1	14 24 14 10 14 15 11 14 14 15 16 14 20 12 16 10 22
Amt2	1	13 23 13 10 14 18 12 13 14 17 15 14 20 12 17 11 23
Amt3	1	13 24 13 10 15 18 11 13 14 17 15 14 20 12 17 11 23
Amt4	2	13 24 13 10 14 18 11 13 14 17 15 14 20 12 17 11 23
Amt5	1	13 24 13 10 14 19 11 14 14 18 18 14 19 12 15 11 22
Amt6	1	13 24 13 10 15 18 11 13 14 16 16 14 20 12 11 11 22
Amt7	2	13 24 13 10 15 16 11 13 14 16 17 14 20 12 15 10 22
Amt8	1	13 23 13 10 15 15 12 13 14 19 16 14 20 12 16 11 22
Amt9	1	13 24 13 10 14 14 13 13 14 18 18 14 19 12 15 12 22
Amt10	1	14 24 13 10 14 14 13 13 14 18 18 14 20 12 16 11 22
Amt11	1	13 24 13 10 14 14 13 13 14 17 18 14 19 12 15 12 22
Amt13	1	14 23 13 10 15 18 13 14 16 17 16 14 20 12 15 11 22
Amt14	1	14 24 14 10 14 15 12 14 14 15 16 14 20 12 16 10 22
Amt15	1	13 24 13 10 16 19 13 13 15 17 13 14 20 12 15 11 22
Amt16	1	14 23 13 10 15 18 14 14 16 16 16 14 20 12 15 11 22
Amt17	3	13 25 14 10 14 18 12 14 14 16 15 14 20 12 15 10 22
Amt20	1	13 23 13 10 14 15 13 13 14 19 17 14 20 12 16 11 22
Amt22	1	13 23 13 10 14 15 13 13 14 20 18 14 20 12 16 11 22
Taq23	2	13 26 13 10 14 20 13 14 14 17 16 14 19 12 15 11 22
Taq24	13	13 26 13 10 14 20 12 14 14 17 16 14 19 12 15 11 22
Taq26	5	13 25 13 10 14 20 12 14 14 17 16 14 19 12 15 11 22
Taq28	1	13 25 13 11 14 19 11 13 14 18 16 14 19 12 15 11 22
Taq31	1	13 23 13 10 14 18 11 13 14 17 15 14 20 12 17 12 23
Taq38	1	13 23 13 10 14 18 12 13 14 17 15 14 20 12 17 11 23
Taq40	1	13 25 13 11 14 20 12 14 14 17 16 14 19 12 15 11 22
Taq43	1	13 24 13 10 14 14 13 13 14 18 17 14 19 12 15 11 22
Taq44	1	13 25 14 10 14 18 12 14 14 16 15 14 20 12 15 10 22
Taq48	2	13 23 13 11 14 14 13 13 14 17 15 14 20 11 16 11 22
Ur51	3	14 23 13 10 15 17 14 14 16 17 18 14 19 12 15 11 22
Ur52	2	13 22 14 10 13 13 12 13 15 16 15 14 22 12 17 11 23
Ur53	1	13 25 14 10 14 18 13 14 14 16 15 14 20 12 15 10 22
Ur54	1	13 25 13 10 16 17 12 12 14 17 17 14 21 11 15 11 23
Ur55	1	13 21 13 10 14 16 12 14 14 17 17 14 21 12 15 11 23

Ur56	2	13 25 13 10 16 17 13 12 14 18 17 14 21 11 15 11 23
Ur58	1	14 24 14 10 14 15 12 14 14 14 16 14 20 12 16 10 22
Ur60	1	14 23 13 10 16 18 12 14 16 17 17 14 20 12 15 11 22
Ur61	1	13 22 14 10 12 13 12 13 15 16 15 14 22 12 17 11 23
Ur63	1	14 24 14 10 14 16 12 14 13 18 16 14 19 12 15 11 23
Ur65	1	13 23 13 10 14 15 13 13 14 19 17 14 20 12 15 11 22
Ap67	1	12 24 13 11 13 16 9 13 13 17 16 14 19 12 15 11 22
Ap68	1	12 24 13 11 13 16 9 13 13 18 16 14 19 12 15 11 22
Ap69	2	13 26 13 12 15 19 13 13 14 16 16 14 20 12 11 11 22
Ap70	1	15 23 13 10 15 18 13 14 16 17 16 14 20 12 15 11 22
Ap71	1	14 23 13 10 15 19 13 14 16 17 16 14 20 12 16 11 22
Ap73	1	14 23 13 10 15 20 12 14 16 17 15 14 20 12 15 11 22
Ap74	1	14 23 13 10 15 20 13 13 16 17 16 14 20 12 15 11 22
Ap76	1	12 25 13 11 13 16 9 13 13 17 17 14 19 12 15 11 22
Ap77	1	12 24 13 11 13 16 9 13 13 17 16 14 19 11 15 11 22
Ap78	1	13 24 13 11 15 18 12 13 14 15 17 14 19 12 15 11 22
PU66	1	14 23 13 10 15 18 15 14 16 17 15 14 20 12 15 11 22
PU79	1	14 23 13 10 15 20 14 14 16 17 15 14 20 12 16 11 22
PU80	1	13 24 14 10 16 17 13 13 14 17 15 14 20 13 15 11 22
LA17	1	13 21 13 10 15 18 12 13 14 17 16 14 20 11 15 11 22
LA1	2	13 24 13 11 14 15 11 15 14 18 20 14 20 11 14 11 23
LA6	1	13 24 13 10 13 18 14 14 14 18 17 14 21 12 16 12 22
LA8	1	13 24 13 10 17 18 13 13 14 16 16 14 20 13 15 11 23
LA14	1	13 24 13 11 15 15 12 14 14 18 16 14 21 12 16 11 22
LA15	1	14 21 13 10 15 17 12 13 14 18 18 14 20 11 15 11 22
LA16	1	13 24 13 10 14 18 13 13 13 18 17 14 20 12 15 11 22
LI3	1	13 24 14 10 14 16 13 12 14 16 17 14 20 12 15 11 22
LI5	1	13 23 13 10 14 18 12 14 14 17 16 14 20 12 15 11 22
LI7	1	13 23 13 10 13 17 12 13 14 18 17 15 20 12 15 11 22
LI9	1	12 24 13 10 14 17 11 13 15 17 18 14 20 12 15 11 22
LI10	1	14 23 13 10 15 18 15 14 16 17 15 14 20 12 15 11 22
LI11	1	13 24 13 10 14 18 12 15 14 16 15 14 19 11 15 11 22

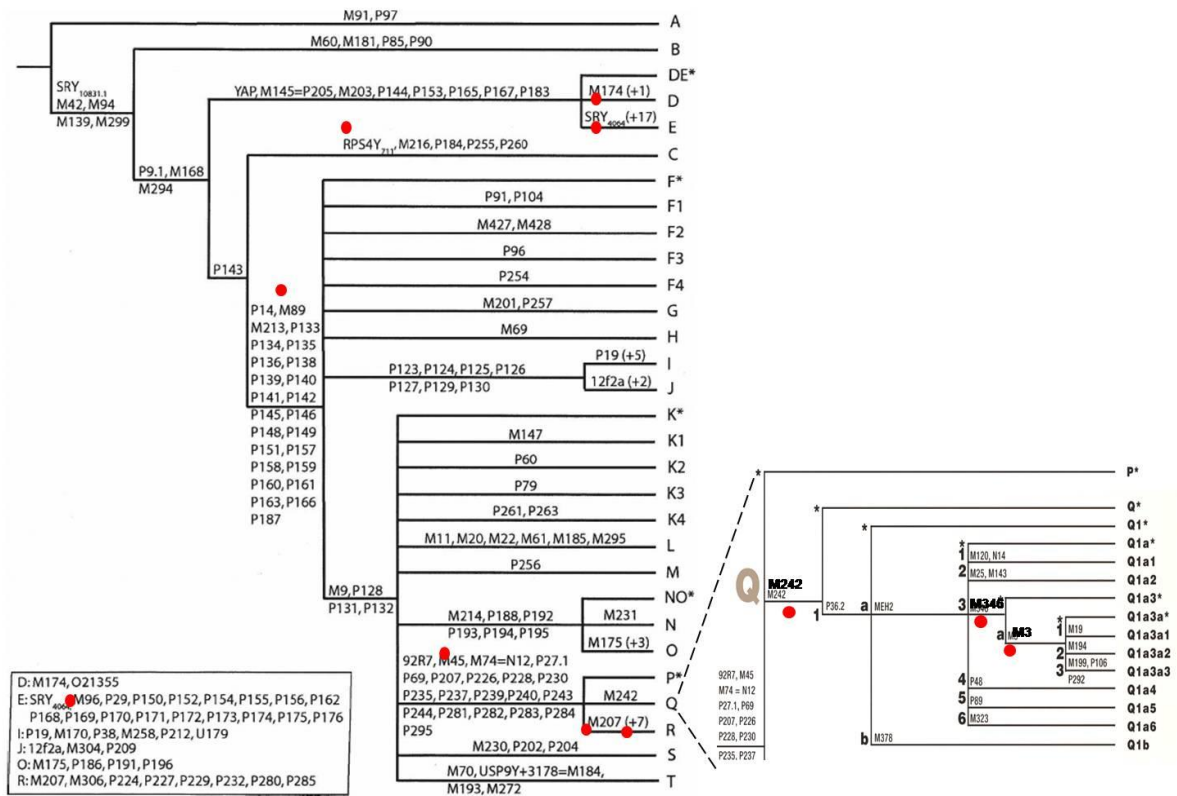


Figura A1: Painel de 10 marcadores SNPs usados (círculos cheios) no estudo da determinação das linhagens do cromossomo Y nas populações peruanas. (Figura modificada de Karafet *et al.*, 2008 e de YCC, 2002)