

Juliana Alves da Silva

**Análise da Variabilidade do mtDNA
na População Brasileira**

Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas
Departamento de Bioquímica e Imunologia
Belo Horizonte
Dezembro de 2000

Juliana Alves da Silva

Análise da Variabilidade do mtDNA na População Brasileira

Tese submetida ao Departamento de Bioquímica e Imunologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito parcial para obtenção do grau de Doutor em Ciências.

Orientadora: Profa. Dra. Vania Ferreira Prado

Co-orientador: Prof. Dr. Sérgio Danilo Junho Pena

*Dedico este trabalho a
meu pai...*

Agradecimentos

À orientadora mais brava do mundo, Vania Prado, por "tudo" que aconteceu nestes quase 5 anos, e sobretudo... pela sua amizade.

Ao melhor pesquisador que eu já conheci, meu co-orientador Sérgio Pena, pela oportunidade de ter feito parte de sua equipe; pelas valorosíssimas orientações e até mesmos pelos "puxões de orelha"...

Ao professor Hans-Jürgen Bandelt, por ter me mostrado de forma clara e muito interessante, um dos grandes valores do mtDNA.

Aos professores Walter Neves e André Prous, pelo sacrifício "doloroso" da doação das amostras fósseis.

Aos Drs. Mark Stoneking e Svante Pääbo, por terem me aberto as portas de seu laboratório, e pelas excelentes discussões durante as reuniões de laboratório e "journal clubs".

À querida Dr. Anne Stone, pelas inesquecíveis discussões sobre "ancient DNA" e seus problemas.

Aos amigos Brigitte, Silke, Hiroki e Siva, pela saudade que eu sinto todos os dias daquela terra maravilhosa, que é a Alemanha.

Aos amigos do laboratório de Genética Bioquímica, minha segunda família, por todos os momentos de alegria e de tristeza, por todas as discussões enriquecedoras e até mesmo nas besteiras ditas no dia-dia, o meu muito obrigado.

À turminha, agora muito defasada, do "gel da meia-noite"... valeu!

Aos amigos do Laboratório de Neurobiologia Molecular pela amizade, companheirismo e ajuda nos momentos mais complicados.

À Celise pelo valioso apoio sempre dispensado aos alunos de pós-graduação.

Ao Departamento de Bioquímica e Imunologia, na pessoa de seu atual Coordenador de Pós-Graduação Professor Carlos Alberto Tavares, pelo suporte dado ao longo dos anos para a realização deste trabalho.

À Neuzinha, flor que nos acaricia e anjo que nos guarda, todas as palavras de agradecimento seriam poucas...

À Kátia Barroso e Mirian pelo importante apoio técnico e auxílio nos cuidados com o laboratório.

Aos Professores Carlos Renato (Nem), Santuza e Fabrício Santos pelo apoio e pelas sugestões importantes no exame de qualificação.

Aos amigos Charles Anacleto, Eduardo Tarazona, Pedro Edson e Renato Santana pelo apoio incondicional no momento mais delicado da produção desta tese, justamente na corrida contra o tempo...

Ao Núcleo de Genética Médica pela gentileza em colaborar fornecendo as amostras para as análises.

Aos amigos do Buteco Biologia pelo companheirismo nas poucas horas de descontração.

À Magdinha, irmã, amiga e anjo da guarda.

Ao meu lindo, Maurício, pelo carinho, companhia e acima de tudo, paciência.

Ao meu pai, por estar sempre perto e nunca me deixar desanimar.

Índice

Resumo	xi
Abstract	xii
I – Introdução	01
I.1 – Os brasileiros	02
I.2 – Heterogeneidade genética da população brasileira	05
I.3 – Utilização do DNA mitocondrial em estudos da variabilidade humana	06
I.3.1 – O mtDNA humano	06
I.3.2 – Propriedades do mtDNA humano	07
I.3.2.1 – Alta variabilidade	07
I.3.2.2 – Herança materna	10
I.3.3 – Utilização do DNA mitocondrial em estudos de variabilidade	11
I.3.3.1 – Haplogrupos de mtDNA característicos das populações africanas.	12
I.3.3.2 – Haplogrupos de mtDNA característicos de populações asiáticas e nativo-Americanas	16
I.3.3.3 – Haplogrupos de mtDNA característicos de populações européias	18
I.4 – Análises da variabilidade genética da população brasileira desenvolvidas em nosso laboratório	21
I.5 – Análise de uma população pré-histórica brasileira	22
I.5.1 – mtDNA e análises de racemização de aminoácidos em estudos de arqueologia molecular.	23
II – Objetivos	26
III – Material e Métodos	28
III.1 – Os Indivíduos	29
III.1.1 – Amostras da população brasileira atual	29
III.1.2 – Amostras da população de Portugal	29
III.1.3 – Amostras de uma população brasileira pré-histórica	29
III.2 – Os Iniciadores	30
III.3 – As enzimas de restrição	30
III.4. – Análise de amostras contemporâneas	30
III.4.1 – Amplificação por PCR da Região Controle de amostras contemporâneas	30
III.4.2 – Sequenciamento da Região Controle	34
III.4.3 – Análise das sequências	34

III.4.4 – Amplificação por PCR do mtDNA total	35
III.4.5 – Análise de RFLP para polimorfismos continente-específicos de mtDNA	35
III.4.6 – Eletroforese em gel de agarose	35
III.4.7 – Eletroforese em gel de poliacrilamida corado por sais de prata	37
III.4.8 - Clonagem e sequenciamento de produto de PCR	37
III.5. – Análise de amostras arqueológicas	38
III.5.1 – Cuidados especiais no trabalho com DNA antigo	38
III.5.2 – Preparo das amostras	39
III.5.3 – Análises de racemização de aminoácidos.	39
III.5.4 – Extração de DNA de amostras fósseis	40
III.5.5. – Amplificação por PCR da Região Controle dos extratos de DNA de amostras arqueológicas	42
III.5.6 – Reamplificação de produtos de PCR	42
III.5.7 – Clonagem e sequenciamento dos produtos de PCR	43
III.5.8 – Soluções e tampões utilizados nos experimentos de extração de DNA de amostras arqueológicas.	45
IV – Resultados	46
IV.1. – Caracterização da linhagem contendo o polimorfismo de inserção de 9pb do mtDNA	47
IV.1.1 – Caracterização do polimorfismo de inserção da unidade repetitiva de 9-bp	47
IV.1.2 – Sequenciamento da região controle do mtDNA e análises de RFLP para a amostra BR13	47
IV.1.3 – Análise dos polimorfismos de tamanho na região intergênica COII/tRNA ^{Lys} em Portugal.	51
IV.2. Estudo da variabilidade do mtDNA em amostras da população branca das regiões Norte, Nordeste e Sul do Brasil	52
IV.2.1 Sequenciamento da região controle do mtDNA em 148 amostras de indivíduos brasileiros	52
IV.2.2 – Análises Filogenéticas e a origem geográfica das linhagens de mtDNA na população brasileira.	60
IV.2.2.1 – Componente ameríndio da população brasileira.	69
IV.2.2.2 – Componente africano da população brasileira	70
IV.2.2.3 – Componente europeu da população brasileira	73
IV.2.3 – Distribuição diferencial dos haplogrupos continente-específicos de mtDNA na população branca brasileira	76
IV.2.3.1 – Componente ameríndio	77

IV.2.3.2 – Componente africano	77
IV.2.3.3 – Componente europeu	79
IV.3 – Análise das amostras de Santana do Riacho	79
V – Discussão	91
V.1 – Variabilidade do DNA mitocondrial	92
V.2 – Classificação de linhagens de mtDNA	92
V.3 – Origem das linhagens mitocondriais brasileiras	94
V.3.1 – Haplótipos ameríndios na população branca brasileira	94
V.3.2 – Haplótipos africanos na população branca brasileira	97
V.3.3 – Haplótipos europeus na população branca brasileira	104
V.4 – Distribuição de linhagens ameríndias, africanas e européias nas regiões Norte, Nordeste e Sul do Brasil	110
VI – Referências Bibliográficas	115
VII – Apêndice 1	136
VIII – Apêndice 2	137
IX – Apêndice 3	138
X – Apêndice 4	139

Lista de tabelas

Tabela 1: Auto-classificação dos brasileiros de acordo com a cor de pele. Distribuição nas principais regiões geográficas	04
Tabela 2: Principais polimorfismos de sítios de restrição associados a haplogrupos Específicos de populações africanas (Afr), asiáticas (As) e nativo-americanas (NA).	15
Tabela 3: Principais polimorfismos de sítios de restrição associados a haplogrupos específicos de populações européias.	19
Tabela 4: Iniciadores utilizados nas reações de PCR e sequenciamento envolvendo amostras de DNA moderno	31
Tabela 5: Iniciadores utilizados nas reações de PCR e sequenciamento envolvendo amostras de DNA antigo	32
Tabela 6: Endonucleases utilizadas nas análises de RFLP	33
Tabela 7: Polimorfismos de RFLP analisados na identificação dos haplogrupos Continente-específicos de mtdna	36
Tabela 8: Sítios polimórficos observados na sequência de HVS-I e HVS-II do mtdna da amostra BR13.	50
Tabela 9: Análise dos principais polimorfismos de sítio de restrição associados às populações africanas (Afr), asiáticas (As), nativo-americanas (NA) e européias (Eur) nas amostras BR13, PT01 e PT02	53
Tabela 10: Sequências de região hipervariável I do mtDNA identificadas na população brasileira	54
Tabela 11: Diversidade de Ney observada nas amostras da população brasileira.	59
Tabela 12: Resultados das análises de RFLP realizadas nos mtdnas brasileiros.	65
Tabela 13: Frequência dos haplótipos de mtdna Continente-específicos no <i>pool</i> mitocondrial brasileiro	68
Tabela 14: Frequência dos haplogrupos de mtdna dentro das três contribuições continentais para o <i>pool</i> de DNA mitocondrial brasileiro.	78
Tabela 15: Razão da área sob os picos dos aminoácidos D-Asp e L-Asp observados para as amostras arqueológicas de Lagoa Santa.	87
Tabela 16: Código das amostras utilizadas nas análises do haplogrupo L3e	99
Tabela 17: Proporção dos haplogrupos europeus no Brasil (e em cada região), em Portugal e na Europa em geral	111

Lista de Figuras

Figura 1: Mapa estrutural do DNA mitocondrial humano	08
Figura 2. Diagrama da Região Controle do DNA mitocondrial humano	09
Figura 3: Origem africana do DNA mitocondrial humano	13
Figura 4: Esquema da amplificação por PCR de fragmentos sobrepostos da HVR-1	44
Figura 5: Amplificação por PCR da região intergênica COII/tRNA ^{Lys} do mtDNA humano.	48
Figura 6: Sequência da região intergênica COII/tRNA ^{Lys} do indivíduo BR01	49
Figura 7: Polimorfismos do domínio hipervariável (entre os nucleotídeos 16180 e 16193) observados nas sequências brasileiras	58
Figura 8: Árvore filogenética de Neighbor-Joining construída com as sequências de Região Hipervariável I e resultados de RFLP para as amostras da Região Norte.	61
Figura 9: Árvore filogenética de Neighbor-Joining construída com as sequências de Região Hipervariável I e resultados de RFLP para as amostras da Região Nordeste.	62
Figura 10: Árvore filogenética de Neighbor-Joining construída com as sequências de Região Hipervariável I e resultados de RFLP para as amostras da Região Sul.	63
Figura 11: Diagrama ilustrativo das relações filogenéticas entre os haplogrupos de DNA mitocondrial observados na amostra da população Brasileira analisada no presente estudo.	71
Figura 12: Cromatogramas de fluorescência representando o perfil de picos observado para (a) uma amostra padrão de mistura de aminoácidos L; (b) amostra sr-1.	80
Figura 13: Perfis de fluorescência obtidos nas análises de racemização de aminoácidos das amostras arqueológicas de Lagoa Santa	82
Figura 14: Sequências obtidas para os clones contendo insertos de produtos amplificados por PCR a partir do extrato da amostra SR-1	89
Figura 15: Network aplotípica das linhagens do haplogrupo L3e	100
Figura 16: Diagrama representativo das mutações que definem os principais subgrupos do haplogrupo L3e	101

Resumo

A ancestralidade materna de 149 indivíduos brasileiros foi avaliada no presente trabalho através de uma extensa análise de polimorfismo continente-específicos do DNA mitocondrial. As amostras foram provenientes de três regiões geográficas do Brasil (regiões Norte, Nordeste e Sul), e constituem principalmente uma amostra da população branca brasileira. Associando as metodologias de sequenciamento do segmento hipervariável I do mtDNA e análise por RFLP de marcadores informativos, nós fomos capazes de determinar com alta confiabilidade a origem ameríndia, africana ou européia das nossas amostras. Nossos resultados demonstraram claramente que existe um predomínio de linhagens ameríndias na região Norte, uma maior contribuição de linhagens africanas na região Nordeste, enquanto a região Sul apresentou maior influência dos haplótipos europeus. Nossa amostra apresentou um total de 32% de linhagens ameríndias, 24% de linhagens africanas e 44% de linhagens européias. A classificação de todas as linhagens brasileiras em haplogrupos já descritos na literatura para populações ameríndias, africanas e européias demonstrou que esses haplogrupos se apresentam diferentemente distribuídos pelo Brasil. Nosso trabalho parece refletir diferentes processos que levaram à formação da população branca brasileira, como o inter cruzamento que ocorreu preferencialmente entre homens portugueses e mulheres índias e africanas no início da colonização, bem como a influência das imigrações européias que ocorreram principalmente nos últimos dois séculos. Em uma segunda parte desse trabalho nós tentamos caracterizar as linhagens de mtDNA presentes em amostras arqueológicas brasileiras. A metodologia de análise de racemização de aminoácidos foi utilizada para avaliar o nível de degradação molecular dessas amostras antigas e os resultados indicaram que essas amostras não se apresentavam adequadas para estudos envolvendo a recuperação de DNA. Assim nós evitamos a destruição de ossos raros que constituem parte dos importantes remanescentes de uma população antepassada conhecida como "Homem de Lagoa Santa".

Abstract

In the present study we characterised the maternal ancestry of 149 Brazilian individuals through an extensive analysis of mtDNA continent-specific polymorphisms. The samples came from three different geographic regions of Brazil, namely, North, Northeast and South regions, and represent mostly the white fraction of the Brazilian population. Associating hypervariable segment I sequencing analysis and RFLP tests we were able to determine the Amerindian, African or European origin for all of our samples with a high degree of confidence. Our results clearly showed a greater contribution of Amerindian mtDNA lineages in the Northern population, while the Northeast is characterised by the predominance of African lineages, and the European lineages are by far the most frequent in the South. The total sample presented 32% of Amerindian lineages, 24% of African, and 44% of European haplotypes. All of our samples were classified in one of the so called continent-specific mtDNA haplogroups described for Amerindian, African and European populations which were not even distributed across Brazil. Finally, our results seems to reflect different processes taken place in the formation of the present day white Brazilian population, like directional mating involving European (Portuguese) men and Native-American or African women since early of the colonisation time, as well as the recent immigrant waves from Europe of the last two centuries. The second part of this study is related to the characterisation of mtDNA lineages in archaeological samples from Brazil. Unfortunately our preliminary results of aminoacid racemization analysis demonstrated that those samples were not appropriate for ancient DNA retrieval and thus we decided not to continue with DNA extraction procedures in order to save samples which constitute important human remains of the "Lagoa Santa Man" population.

Introdução

I.1 – OS BRASILEIROS

Como uma das populações mais heterogêneas do mundo, a população brasileira formou-se por uma intensa fusão de culturas, costumes e principalmente genes a partir de três *grupos étnicos* principais. Os estoques humanos que participaram desse complexo processo de mistura, que já dura mais de 500 anos, foram: o componente colonizador, vindo principalmente de Portugal a partir do ano de 1500, as populações nativo-americanas que já estavam presentes no território brasileiro há milhares de anos, e os milhões de africanos que foram deslocados para o Brasil durante mais de 3 séculos de tráfico de escravos. A expressão “grupos étnicos” aparece destacada na primeira frase porque ela expressa apenas a extensão do conceito de população, relacionado a um “grupo” que é caracterizado por uma cultura própria, ou até mesmo características genéticas distintas, muitas vezes correlacionadas à sua localização geográfica. Cabe aqui ressaltar que, do ponto de vista genético, o conceito de “raça” não faz sentido algum quando aplicado para a espécie humana (TEMPLETON, 1999).

As estimativas a cerca do número de ameríndios que habitavam a região onde é hoje o Brasil, variam entre 1 e 8,5 milhões (CARNEIRO DA CUNHA, 1992), embora os números mais aceitos pareçam apontar para uma população de 2,5 a 3 milhões de indivíduos por volta do ano 1500 (SALZANO E FREIRE-MAIA, 1970; BETHEL, 1997). Com a chegada dos Portugueses, as tribos ameríndias começaram a sofrer um intenso processo de declínio demográfico, devido principalmente a guerras que promoveram o extermínio em massa de milhares de índios e também a inúmeras doenças para as quais eles não estavam adaptados. Hoje os índios são cerca de 300 mil, de acordo com o Instituto Brasileiro de Geografia e Estatística (<http://www.ibge.gov.br>), e vivem principalmente em reservas indígenas demarcadas pelo governo federal.

Não se sabe ao certo quando desembarcou no Brasil o primeiro negro, mas é possível que houvesse algum, ou alguns, entre os tripulantes da esquadra de Cabral, uma vez que Portugal já comercializava escravos nas costas africanas antes de 1500 (CURTIN 1969). Os escravos africanos começaram a ser sistematicamente introduzidos na economia brasileira a partir da 2ª metade do século XVI, para trabalharem nas fazendas de cana de açúcar e posteriormente nas minas de ouro e diamante, e nas plantações de café. Dados históricos sugerem que entre o ano de 1551 até 1850 (quando o tráfico de escravos foi abolido) mais de 3,5 milhões de negros africanos foram importados para o Brasil (SALZANO E FREIRE-MAIA, 1967; CURTIN, 1969; RIBEIRO 1995). O pico de introdução ocorreu antes de 1820 e de acordo com a literatura os africanos foram trazidos principalmente da costa ocidental da África (SALZANO e FREIRE-MAIA, 1970; RIBEIRO, 1995).

Os Portugueses chegaram em 1500, quando então iniciaram a imigração. Nos primeiros 50 anos, os Portugueses que vieram para o Brasil foram principalmente os deportados, marinheiros, soldados, e comerciantes de pau-brasil (SALZANO e FREIRE-MAIA, 1970; RIBEIRO, 1995). Como

esses indivíduos eram principalmente homens (de acordo com dados históricos, as primeiras mulheres européias só chegaram ao Brasil por volta de 1551), o acasalamento com as índias teve início rapidamente e foi até mesmo estimulado a partir de 1755, como estratégia de crescimento populacional e ocupação colonial do país. O processo de miscigenação que se iniciou com os índios já nos primeiros anos de ocupação européia logo se estendeu também aos escravos negros. Esses três grupos viveram juntos, praticamente sem a presença de outros imigrantes até o início do século XIX, se nós desconsiderarmos, é claro, as tentativas de invasão holandesa e francesa, que ocorreram respectivamente em Pernambuco e no Rio de Janeiro. O apogeu do processo migratório europeu não ocorreu, como seria de se esperar, durante o período colonial, mas sim décadas depois da independência, quando o noroeste de Portugal, a região do Minho, tornou-se uma fonte inesgotável de trabalhadores. A origem do português imigrante é bastante diversificada, de uma próspera elite nos primeiros séculos de colonização, chega-se à segunda metade do século XIX com um número sempre crescente de imigrantes portugueses pobres, que se dirigem, preferencialmente, para as grandes cidades. Estima-se que aproximadamente 500.000 portugueses tenham chegado ao país, entre 1500 e 1808; a partir daí, com a abertura dos portos brasileiros para as nações amigas, o Brasil começou a receber um crescente número de imigrantes de diversas partes do mundo. Portugal continuou a ser a principal fonte de imigrantes europeus, seguido por Itália, Espanha e Alemanha. No século XX importantes imigrações asiáticas ocorreram, principalmente do Japão e da China, mas também do Líbano e Síria. De acordo com Callegari-Jacques e Salzano, (1999), 58% dos imigrantes que contribuíram para a heterogeneidade genética da população brasileira até o ano de 1972 eram europeus, 40% africanos e 2% eram neo-asiáticos.

Esses grupos se distribuíram diferentemente pelos 8 milhões e 500 mil quilômetros quadrados do território brasileiro. Por exemplo, São Paulo foi o estado que recebeu o maior número de imigrantes, sendo que grande parte deles vieram da Itália, Espanha, Japão e Alemanha. O Rio de Janeiro recebeu um grande número de Italianos e Espanhóis, enquanto o Rio Grande do Sul recebeu predominantemente alemães.

Na contagem populacional do IBGE de 1996 (<http://www.ibge.gov.br/estatistica/populacao/contagem/brcont96.shtm>), a população brasileira era constituída por 157.070.163 habitantes e estava distribuída em 5 regiões geográficas oficiais: Norte (11.288.259), Nordeste (44.766.851), Sudeste (67.000.738), Sul (23.513.736) e Centro-Oeste (10.500.579). O IBGE adota um critério simples de auto-classificação para os brasileiros de acordo com a cor da pele, e seus dados indicam que 51,6% da população se considera branca, 5% negra, 0,4% amarela, 42,4% parda e 0,2% indígena (TAB 1).

Tabela 1: AUTO-CLASSIFICAÇÃO DOS BRASILEIROS DE ACORDO COM A COR DE PELE. DISTRIBUIÇÃO NAS PRINCIPAIS REGIÕES GEOGRÁFICAS

População residente (Habitante)* Ano = 1991						
Região Geográfica	Cor de pele ou “raça”					
	Branca	Preta	Amarela	Parda	Indígena	sem declaração
Centro-Oeste 9425053	4418571 0,469	292943 0,031	30686 0,003	4615250 0,490	52750 0,006	14853 0,001
Nordeste 42494112	11317738 0,266	2368206 0,056	27371 0,006	28611078 0,673	55854 0,013	113865 0,003
Norte 10027373	2279173 0,227	329261 0,033	13994 0,001	7230657 0,721	124618 0,012	49670 0,005
Sudeste 62730146	39260994 0,626	3662794 0,058	471732 0,008	18985393 0,302	30584 0,005	328649 0,005
Sul 22129131	18428446 0,833	681926 0,031	86875 0,004	2873707 0,130	30342 0,001	27835 0,001
Brasil	0,516	0,050	0,004	0,424	0,002	0,004

* Os primeiros números expressam valores absolutos do número de indivíduos para cada região e em vermelho estão relacionadas as proporções de cada categoria.

FONTE: IBGE – CENSO DEMOGRÁFICO (1991).

Uma vez que diferentes processos demográficos foram responsáveis pela colonização das diferentes regiões brasileiras, espera-se que cada uma dessas regiões seja caracterizada por uma heterogeneidade genética própria. De fato, é possível observar um gradiente norte → sul na distribuição dos brasileiros de acordo com o caráter cor de pele: brancos constituem 22,7% da população na região Norte e 83,3 % na região Sul.

I.2 – HETEROGENEIDADE GENÉTICA DA POPULAÇÃO BRASILEIRA

Poucos estudos tentando avaliar a variabilidade genética da população brasileira podem ser encontrados na literatura (SANTOS e cols., 1993, 1996; HEIDRICH e cols., 1996), com exceção dos grupos indígenas, os quais vêm sendo intensamente estudados (ZAGO cols., 1996; BORTOLONI e cols., 1998a; HUTZ e cols., 1997; VALLINOTO e cols., 1998; WEIMER e cols., 1998). Os trabalhos que representaram uma tentativa de caracterizar o perfil genético dos brasileiros “não-ameríndios” (para uma revisão ver SALZANO, 1997; CALLEGARI-JACQUES & SALZANO 1999; GUERREIRO e cols., 1999) se concentraram principalmente nas regiões Norte e Sul do país (FRANCO e cols., 1982; ROSA e cols., 1984; MORAES e cols., 1993; SANTOS, E. e cols., 1996; BATISTA DOS SANTOS e cols., 1999; DORNELLES e cols., 1999). Além disso, foram predominantemente analisados grupos isolados da população brasileira, formados por descendentes dos escravos africanos (como é o caso das várias comunidades derivadas de ex-quilombos que vêm sendo estudados pelo grupo do prof. Franciso Mauro Salzano do Rio Grande do Sul) (BORTOLINI e cols., 1992, 1995, 1997a, 1997b, 1998a; ARPINI-SAMPAIO e cols., 1999), ou descendentes de imigrantes europeus (DORNELLES e cols., 1999). Os poucos estudos realizados na população branca brasileira basearam-se principalmente em marcadores clássicos, usando sistemas genéticos protéicos e demonstraram um maior ou menor grau de mistura genética, que pode ser interpretado como resultado de miscigenação, para todas as amostras analisadas. Foi observado que a influência genética de ancestrais ameríndios e africanos nos “brancos” brasileiros é amplamente variada demonstrando padrões distintos nas diferentes regiões geográficas do país (KRIEGER e cols., 1965; FRANCO e cols., 1982; CONCEIÇÃO e cols., 1987; SANTOS & GUERREIRO 1995; SANTOS e cols., 1996a; SALAZANO 1997; ARPINI-SAMPAIO e cols., 1999; BATISTA DOS SANTOS e cols., 1999; DORNELLES e cols., 1999).

Nos últimos 10 anos, um esforço muito grande tem sido destinado a pesquisas que envolvem a busca de marcadores moleculares, que sejam sobretudo específicos para cada um dos diversos grupos étnicos que compõem a espécie humana, como os africanos, europeus, asiáticos e ameríndios (para citar apenas os principais). Esses estudos encontraram seus maiores avanços, na análise de segmentos genômicos que apresentam baixos níveis (ou ausência completa) de recombinação, como algumas regiões dos cromossomos X e Y, e o DNA mitocondrial humano. Uma

série de características peculiares fazem com que esses sistemas se apresentem particularmente informativos em estudos relacionados a evolução humana e genética de populações. A seguir, serão apresentados os pontos relacionados à variabilidade do mtDNA, por ser esta a ferramenta utilizada em nossos trabalhos com a população brasileira. Aspectos relacionados aos estudos que envolvem a análise do cromossomo Y e do cromossomo X podem ser encontrados em Santos e Tyler-Smith (1996), Rosser e colaboradores (2000), Harris e Hey (1999) e Kaesmman e colaboradores (1999).

I.3 – UTILIZAÇÃO DO DNA MITOCONDRIAL EM ESTUDOS DA VARIABILIDADE HUMANA

A heterogeneidade do mtDNA vem sendo caracterizada em praticamente todas as populações mundiais, nos últimos 10 anos, desde a publicação dos trabalhos de Cann e colaboradores. (1987), Vigilant (1990), Stoneking e colaboradores. (1991) e Vigilant e colaboradores. (1991), que demonstraram que esse pequeno segmento do genoma humano era capaz de apresentar polimorfismos ou mutações que estavam relacionados à origem étnico-geográfica dos indivíduos. Essa variabilidade só é possível devido a dois aspectos intrínsecos do DNA mitocondrial, representados pela herança exclusivamente materna associada a uma alta taxa de mutação, que permitiu o surgimento de marcadores continente-específicos de população, ao longo de mais de 100.000 anos de evolução humana.

I.3.1 – O MTDNA HUMANO

O DNA mitocondrial (mtDNA) constitui o material genético extranuclear presente nas mitocôndrias, que são as organelas citoplasmáticas responsáveis pela produção da maior parte da energia necessária para o exercício das funções celulares nas células não fotossintetizantes (RANGEL, 1974). Com 16569 pares de base, o DNA mitocondrial humano foi o primeiro genoma mitocondrial a ser sequenciado na sua totalidade e interpretado em relação ao seu conteúdo gênico (ANDERSON e cols., 1981). Com 37 genes (ANDERSON e cols., 1981; PALMER, 1997) o mtDNA humano é extremamente especializado, contendo exclusivamente genes que são necessários para a síntese de componentes catalíticos do sistema de fosforilação oxidativa (WALLACE, 1995; KENYON e MORAES, 1997).

A região codificadora do mtDNA humano, que corresponde a mais de 90% da sua sequência de nucleotídeos, apresenta vinte e dois genes que são transcritos em RNAs transportadores (tRNAs), dois genes para as subunidades 12S e 16S do RNA ribossomal (rRNA), além de 13 genes que codificam para diferentes proteínas envolvidas nos processos de fosforilação oxidativa (OXPHOS), sendo eles: 7 (ND1, ND2, ND3, ND4, ND4L, ND5 e ND6) dos 42 polipetídeos do complexo I da

fosforilação oxidativa (NADH: ubiquinona oxidoredutase); 1 (citocromo b, cytb) dos 11 polipetídeos do complexo III (ubiquinol: citocromo c oxidoredutase); 3 (COI, COII, E COIII) dos 13 polipeptídeos do complexo IV (citocromo c oxidase); e 2 (ATP6 e ATP8) dos 16 polipetídeos do complexo V (Complexo F₀ ATPase) (ANDERSON e cols., 1981; WALLACE, 1995) (FIG 1).

O mtDNA humano apresenta uma enorme compactação, onde cerca de 90% de sua sequência é transcrita em algum produto gênico, contrastando com cerca de 3,5% de seqüências codificadoras presentes no genoma nuclear (MORITZ e cols., 1987). Seus genes não possuem íntrons e pouco espaço intergênico. De fato, uma característica intrigante da organização gênica mitocondrial é a enorme proximidade entre os genes codificadores de proteína com os genes de RNA transportador (tRNAs). Até mesmo a sobreposição das extremidades 5' e 3' de alguns genes é observada, isto é, a mesma base participa do códon de terminação de um gene e do códon de iniciação do gene subsequente (ANDERSON e cols., 1981; OJALA e cols., 1981a, 1981b).

Localizada entre os genes dos tRNAs para os aminoácidos prolina (tRNA^{Pro}) e fenilalanina (tRNA^{Phe}), a região controle (RC) do mtDNA humano (FIG 2) tem 1122 pb e representa aproximadamente 7% deste genoma. Neste segmento, estão presentes os sinais necessários para o início de replicação de uma das fitas de mtDNA (origem da replicação da molécula) e para o início da transcrição das duas cadeias de DNA, além de seqüências associadas com o término de replicação do mtDNA (WOLSTENHOLME, 1992). Dessa forma, a RC é alvo de numerosas proteínas e enzimas tais como RNA e DNA polimerases, fatores de transcrição e de regulação (SACCONI e cols., 1991).

No início do processo de replicação do mtDNA, ocorre a síntese de um novo segmento de uma das fitas de DNA, denominado DNA 7S. Isso provoca o deslocamento da outra fita parental resultando em uma estrutura em forma de alça que recebe o nome de alça-D (do inglês "Displacement-loop" ou "D-loop"). Como a alça-D corresponde a quase toda a RC (SACCONI e cols., 1991; CLAYTON, 1992), esta última é também comumente chamada de D-loop (ANDERSON e cols., 1981; CLAYTON, 1992; WOLSTENHOLME e JEON, 1992).

I.3.2 – PROPRIEDADES DO MTDNA HUMANO

I.3.2.1 – ALTA VARIABILIDADE

A taxa de evolução ao longo da região codificadora do mtDNA é 10 a 12 vezes maior que a observada em genes nucleares de função homóloga (BROWN e cols., 1979; NECKELMANN e cols., 1987; WALLACE e cols., 1987,1999; BIANCHI e BAILLIET, 1993; WALLACE, 1995). A região controle tem demonstrado ser ainda menos conservada. Estudos baseados em análise de restrição e no sequenciamento demonstraram que essa região apresenta uma taxa de divergência de

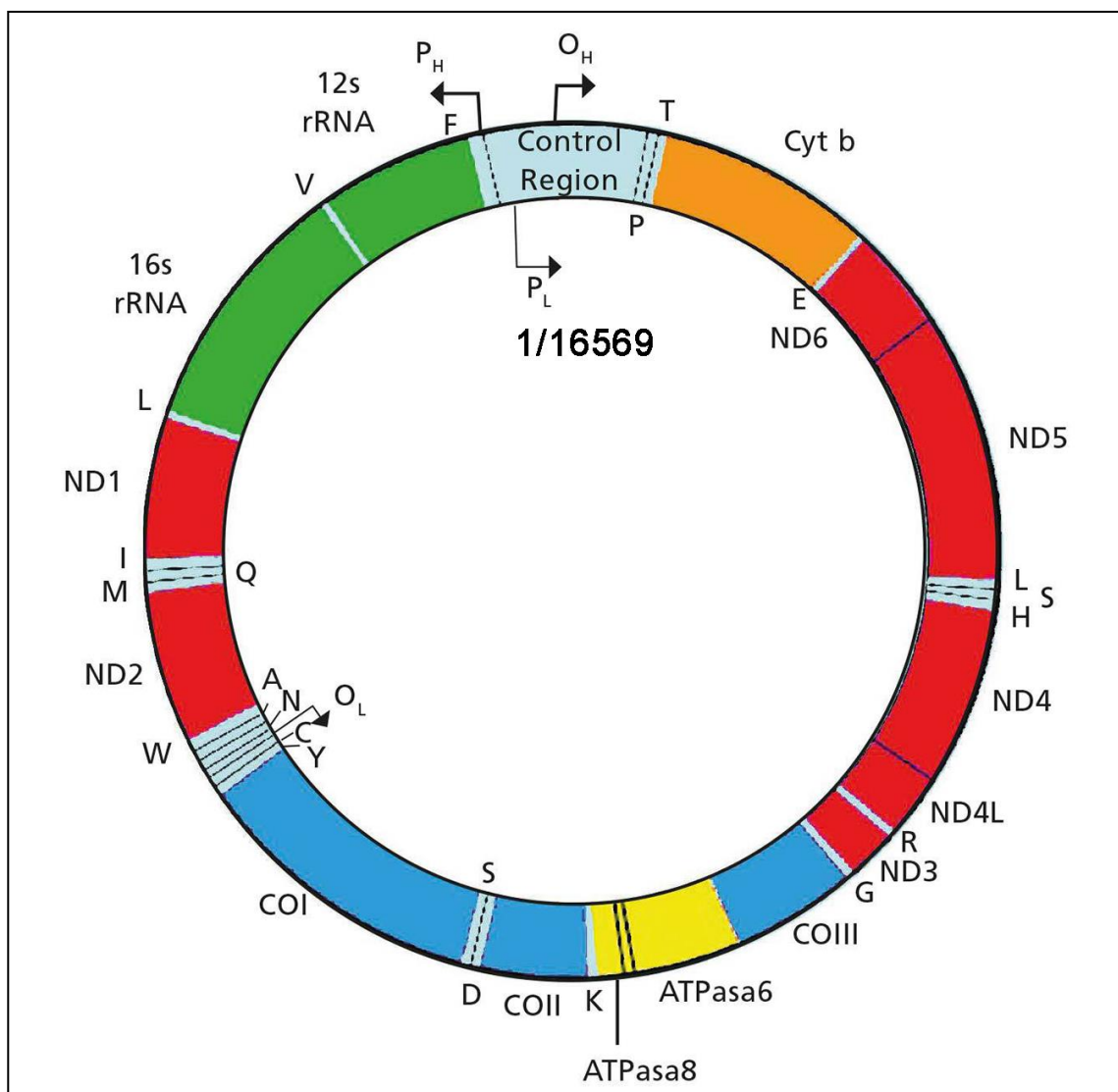


FIGURA 1: MAPA ESTRUTURAL DO DNA MITOCONDRIAL HUMANO. Os genes de proteína e rRNA estão representados pelos códigos definidos no texto. RNAs transportadores estão indicados pelo código de uma letra do aminoácido correspondente. 1/16569 define o primeiro e o último nucleotídeo da molécula circular, sendo que os demais nucleotídeos são numerados no sentido horário, de acordo com Anderson e colaboradores (1981). Figura adaptada de Wallace (1997).

Região Controle do mtDNA humano

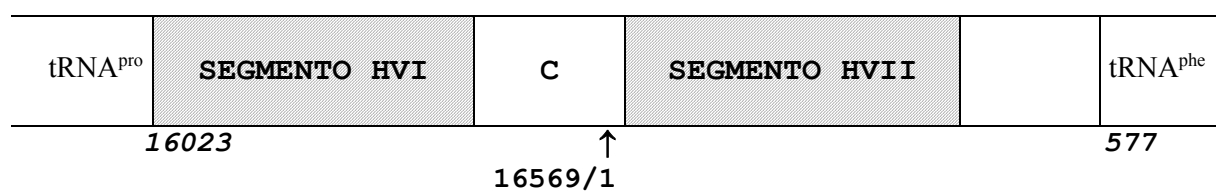


FIGURA 2. DIAGRAMA DA REGIÃO CONTROLE DO DNA MITOCONDRIAL HUMANO. Esta região é definida pelos genes dos tRNA para os aminoácidos prolina (tRNA^{pro}) e fenilalanina (tRNA^{phe}), entre os nucleotídeos 16023 e 577 (de acordo com a sequência de Anderson, 1981). 16569/1 indica o último e o primeiro nucleotídeo da molécula. A região central está indicada (C). Adaptado de Saccone e cols., 1991 e Stoneking e cols., 1991.

nucleotídeos de 3 a 5 vezes mais alta que outras partes da molécula (AQUADRO e GREENBERG, 1983; GREENBERG e cols., 1986; CANN e cols., 1987; VIGILANT e cols., 1989, 1991; HORAI e HAYASAKA, 1990; HORAI e cols., 1991, 1993; DI RIENZO e WILSON, 1991; TAMURA e NEI, 1993; WALLACE e cols., 1999). Esses estudos mostraram que a variação estava concentrada nas regiões adjacentes aos extremos 5' e 3', denominados segmentos hipervariáveis I e II, enquanto a região central não apresentava maiores variações entre as diferentes seqüências humanas analisadas (HORAI e cols., 1993; BIANCHI e BAILLET, 1993; PÄÄBO, 1996; MEYER e cols., 1999) (FIG 2). Recentemente, Hoffman e colaboradores (1997) sequenciando a região controle de 200 alemães, descreveram um terceiro segmento hipervariável (HVSIII) localizado entre os nucleotídeos 438-574.

Diversas pesquisas vêm sendo desenvolvidas na tentativa de esclarecer quais os principais mecanismos envolvidos na geração dessa diversidade. Entretanto, os principais fatores parecem ser:

- Localização: Estando localizado na matriz mitocondrial, em íntima proximidade com a membrana mitocondrial interna, o mtDNA está exposto constantemente ao elevado fluxo de agentes mutagênicos do tipo radicais livres, que resultam do metabolismo incompleto do oxigênio nos processos de transferência de elétrons da cadeia respiratória (RICHTER e cols., 1988).
- Ausência de proteção por proteínas do tipo histonas como ocorre com o DNA nuclear, o que torna o mtDNA mais susceptível ao “estresse oxidativo” (RICHTER e cols., 1988; ADELMAN e cols., 1988; RICHTER, 1992; BANDY e DAVISON, 1990).

1.3.2.2 – HERANÇA MATERNA

O mtDNA humano é herdado exclusivamente da mãe. Giles e colaboradores (1980), e Case e Wallace (1981), demonstraram que o mtDNA é transmitido da mãe para todos os seus descendentes e das filhas para a próxima geração, sem a contribuição do mtDNA paterno. Essa particularidade se deve ao fato de que o citoplasma do zigoto, formado a partir da fecundação do óvulo pelo espermatozóide, é constituído quase exclusivamente (cerca de 95%) pelo citoplasma da célula germinativa feminina. Consequentemente, a população mitocondrial do zigoto corresponde às mitocôndrias provenientes do óvulo.

Alguns autores têm sugerido que pelo menos em camundongos (GYLLENSTEN e cols., 1991) e *Drosophyla* (KONDO e cols., 1990) existe também contribuição paterna na transmissão do mtDNA. Esses autores, baseados na geração de híbridos inter-específicos envolvendo retrocruzamentos, demonstraram a acumulação de baixos níveis de mtDNA paterno, da ordem de 0,001% por geração. Essa pequena contribuição paterna mostrou-se consistente com a quantidade relativa de mtDNA materno e paterno no zigoto, imediatamente após a fertilização.

Kaneda e colaboradores (1995) demonstraram que em cruzamentos intra-específicos, praticamente todas as mitocôndrias do gameta masculino desaparecem quase imediatamente após a fertilização, ainda no estágio de pronúcleo, garantindo a fidelidade da herança exclusivamente materna. Entretanto, em experimentos envolvendo híbridos inter-específicos pode ocorrer a sobrevivência das mitocôndrias do espermatozóide. Baseado nesses dados, esses autores sugeriram a existência de um mecanismo espécie-específico, de reconhecimento seletivo e destruição das mitocôndrias provenientes do espermatozóide (KANEDA e cols., 1995; ANKEL-SIMONS e CUMMINS, 1996; OHNO, 1997).

É importante salientar que mesmo traços de herança paterna do mtDNA poderiam permitir a recombinação com o mtDNA materno, o que resultaria em um impacto significativo nas interpretações dadas aos padrões de variação do mtDNA (STONEKING e SOODYALL, 1996, AWADALLA e cols., 1999; HAGELBERG e cols., 1999; MORRIS & LIGHTOWLERS, 2000; STONEKING, 2000).

I.3.3 – UTILIZAÇÃO DO DNA MITOCONDRIAL EM ESTUDOS DE VARIABILIDADE

Devido às suas propriedades, o mtDNA tem sido extensamente utilizado em estudos de variabilidade genética. Por exemplo, várias pesquisas utilizando o mtDNA foram desenvolvidas com o objetivo de esclarecer a origem e evolução do homem moderno (*Homo sapiens sapiens*) (CANN e cols., 1987; VIGILANT e cols., 1989 e 1991; DI RIENZO e WILSON, 1991; PICKFORD, 1991; HORAI e cols., 1995; NEI, 1995; PENNY e cols., 1995; BONATO E SALZANO, 1997a; FORSTER e cols., 1996; WATSON e cols., 1996, 1997; VON HAESELER e cols., 1996; QUINTANA-MURCI e cols., 1999, CHEN e cols., 2000, JORDE e cols., 2000) e esses estudos sugerem que toda a variabilidade do mtDNA humano contemporâneo tenha sido originada na África, a partir de uma única linhagem mitocondrial cerca de 150.000 a 200.000 anos atrás. Há aproximadamente 80– 50 mil anos tiveram início as ondas migratórias das populações ancestrais que permitiram a colonização de todo o mundo. Essa hipótese é comumente conhecida como “Eva mitocondrial” ou “Eva Africana” (PICKFORD, 1991) e está representada na figura 3.

Além disso, o mtDNA passou a ser uma ferramenta importante nos estudos de taxonomia molecular, antropologia e medicina forense. Como exemplos dos resultados possíveis pelo estudo do mtDNA podemos citar a averiguação da relação filogenética do mamute com as espécies atuais de elefantes (HOSS e cols., 1994; HAGELBERG e cols., 1994a), a confirmação da origem Polinésia dos habitantes da Ilha da Páscoa (HAGELBERG e cols., 1994b) e a identificação dos restos mortais da família do Czar da Rússia (GILL e cols., 1994; IVANOV e cols., 1996). Recentemente Krings e colaboradores (1997, 1999a) sugeriram que não houve troca gênica entre os ancestrais do homem moderno (*homo sapiens sapiens*) e o Homem de Neandertal (*homo neanderthalensis*) mesmo tendo

sido comprovado que essas duas espécies coabitaram as mesmas regiões geográficas no mesmo espaço de tempo (SHREEVE, 1995; BAHN, 1998; MELLARS, 1998). Resultados semelhantes aos de Krings e colaboradores (1997) foram também observados por Ovchinnikov e colaboradores (2000). Entretanto é importante ressaltar que embora esses resultados tenham se mostrado confiáveis, as análises de homologia do mtDNA indicaram apenas a ausência de uma linhagem matrilinea “humana” no espécimen do Homem de Neandertal analisado.

Na última década, a variabilidade do mtDNA foi estudada em um grande número de indivíduos da Ásia, Europa, América e África (CANN e cols., 1987; MERRIWETHER e cols., 1994; HORAI e HAYASAKA, 1990; SCHURR e cols., 1990; TORRONI e cols., 1992, 1993a,b, 1994b, 1995, 1996, 1998; BAILLIET e cols., 1994; REED e cols., 1995; CHEN e cols., 1995; WATSON e cols., 1996, 1997; RICHARDS e cols., 1996, 1998; RANDO e cols., 1998; FRANCALACCI e cols., 1999; MACAULAY e cols., 1999; GREEN e cols., 2000; HELGASON e cols., 2000; SIMONI e cols., 2000). Duas metodologias principais foram utilizadas para se avaliar direta ou indiretamente as variações na sequência do mtDNA. Essas técnicas consistem no sequenciamento das porções mais variáveis da molécula de mtDNA, ou seja, as regiões hipervariáveis I e II da região controle, e análises de RFLP (do inglês *Restriction Fragment Length Polymorphism*). Grande parte dos estudos de RFLP têm utilizado uma metodologia de alta resolução que consiste em amplificar por PCR todo o mtDNA em 9 fragmentos parcialmente sobrepostos, que são posteriormente digeridos com 14 endonucleases (*AluI*, *Avall*, *BamHI*, *Ddel*, *Haell*, *HhaI*, *HincII*, *Hinfl*, *HpaI*, *HpaII/MspI*, *MboI*, *RsaI* e *TaqI*). Este procedimento permite vasculhar indiretamente cerca de 15 a 20% da sequência do mtDNA. Após estabelecer a natureza e a prevalência das mutações do mtDNA na população humana, análises filogenéticas permitiram a definição de conjuntos de haplótipos de mtDNA (denominados haplogrupos de mtDNA) que são utilizados como marcadores genéticos para identificar a origem étnico-geográfica dos indivíduos (HORAI e HAYASAKA, 1990; SCHURR e cols., 1990; MERRIWETHER e cols., 1994; TORRONI e cols., 1992, 1993a,b, 1994a,b, 1996; CHEN e cols., 1995; WALLACE, 1995; REED e cols., 1995; WALLACE e cols., 1999).

I.3.3.1 – HAPLOGRUPOS DE MTDNA CARACTERÍSTICOS DAS POPULAÇÕES AFRICANAS.

A primeira evidência clara de que as variações do mtDNA estavam correlacionadas com a origem étnica e geográfica dos indivíduos, surgiu com a descoberta de que 70 a 100% das populações do Sub-Saara Africano são definidas pela presença de um sítio para a enzima *HpaI* na posição 3592 do mtDNA humano, devido a uma transição T→C na posição 3594 (SCOZZARI e cols., 1988, 1994; SOODYALL e JENKINS, 1992; CHEN e cols., 1995, 2000). Fora da África, este marcador é encontrado apenas em baixa frequência e somente naquelas populações onde a contribuição

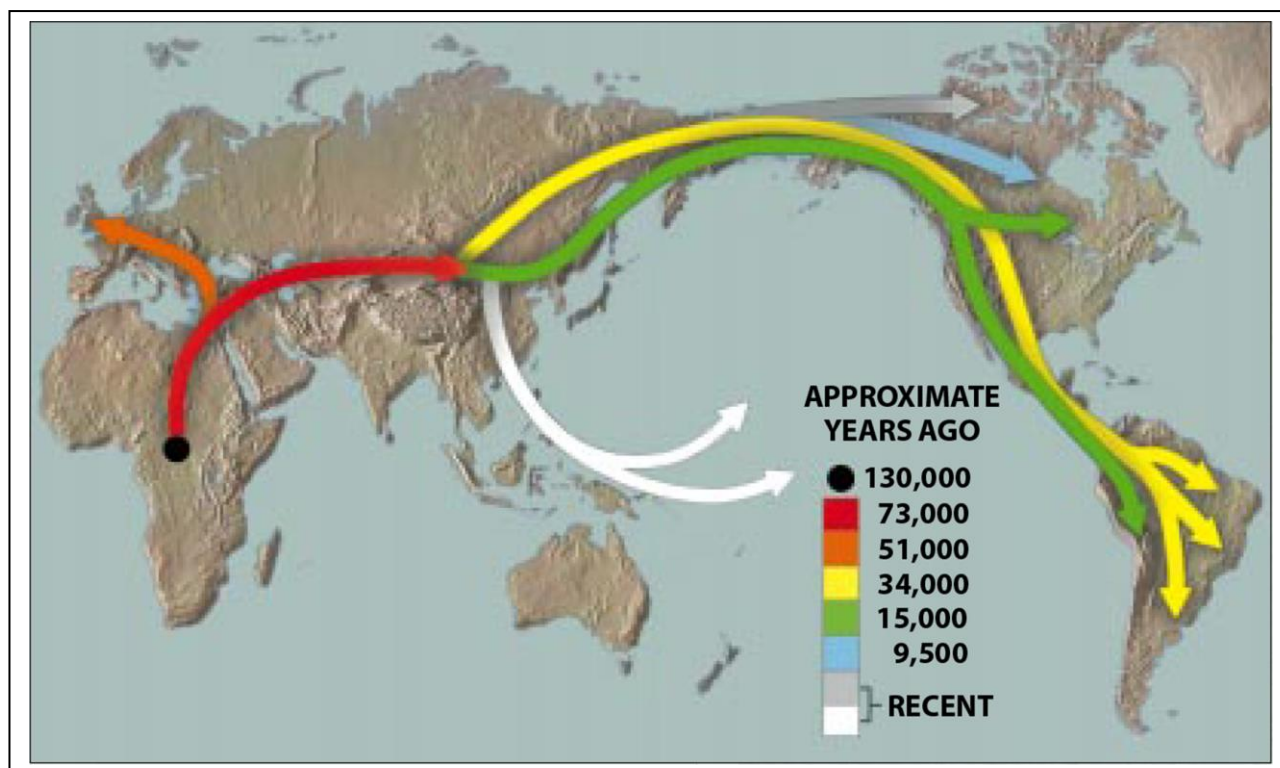


FIGURA 3: ORIGEM AFRICANA DO DNA MITOCONDRIAL HUMANO. A hipótese da “Eva Mitocondrial” ou “Eva Africana” sugere que a origem do homem moderno tenha ocorrido na África, cerca de 150.000 anos atrás. Há aproximadamente 80.000 anos tiveram início as ondas migratórias das populações ancestrais que permitiram a colonização de todo o mundo. A partir daí teriam se originado as linhagens mitocondriais específicas dos continentes com a migração das mulheres e o acúmulo sequencial de mutações. As setas indicam teorias existentes na literatura para as rotas migratórias de populações antepassadas que teriam permitido a colonização dos diferentes continentes. As cores representam o tempo passado até o presente. Figura adaptada de WALLACE (1997).

gênica de Africanos é historicamente conhecida (BONNÉ-TAMIR e cols., 1986; SEMINO e cols., 1989, CÔRTE-REAL e cols., 1996; KRINGS e cols., 1999b; WALLACE e cols., 1999).

Análises de alta resolução dos polimorfismos de restrição do mtDNA realizadas por CHEN e colaboradores (1995, 2000) demonstraram que todos os mtDNAs que apresentam um sítio para a enzima *HpaI* na posição 3592 (*HpaI* + 3592) fazem parte de um grande grupo monofilético denominado macro-haplogrupo L, que é subdividido em dois haplogrupos, L1 e L2 (TAB 2). O haplogrupo L1 é definido pela presença de um sítio de *HinfI* na posição 10806 (resultado de uma transição T→C no nucleotídeo 10810), e compreende 39% dos haplótipos L sendo encontrado em 34,3% da população africana. O haplogrupo L2 é definido pela associação de perda de um sítio de *Avall* na posição 16390 e ganho de um sítio de *HinfI* na posição 16389 (devido a uma transição G→A no nucleotídeo 16390), e compreende 61% dos haplótipos L, sendo encontrado em 42,1% dos africanos.

Outros haplogrupos africanos caracterizados pela ausência do sítio de *HpaI* na posição 3592 foram também determinados e denominados haplogrupos L3*. Embora não constituam um grupo monofilético como o macro haplogrupo L, as linhagens L3 africanas foram divididas em 4 haplogrupos principais, L3a, L3b, L3d e L3e (WATSON e cols., 1997; RANDO e cols., 1998; CHEN e cols., 2000) que são também identificados por mutações específicas, detectadas através de análises de RFLP e/ou sequenciamento de Região Controle (TAB 2). Cada um desses haplogrupos africanos menores apresenta também uma alta diversidade, que tem sido objeto de estudos mais detalhados (RANDO e cols., 1998, 1999). Atualmente existe a necessidade de uma revisão na nomenclatura dos haplogrupos africanos, uma vez que, enquanto 10 letras são utilizadas na denominação dos haplogrupos característicos das populações européias (que apresentam uma variabilidade consideravelmente menor, no nível do mtDNA, quando comparada às africanas), apenas a letra “L” é utilizada para todos os haplogrupos africanos. Isso tem dado margem a certas confusões que ocorrem ao se comparar por exemplo, os dados de Rando e colaboradores (1998) e Quintana-Murci e colaboradores (1999), `aqueles de Chen e colaboradores (2000). Os primeiros autores dão os nomes de: haplogrupo L3a ao grupo de linhagens caracterizado pelo marcador 10394 – *Ddel*; L3b ao grupo que apresenta o polimorfismo 10084 + *TaqI*; L3d ao grupo definido por 8616 – *MboI*; e L3e ao grupo caracterizado pelo polimorfismo 2349 + *MboI*. Por sua vez, Chen e colaboradores (2000) usaram a denominação de haplogrupos L3d, L3c, L3b e L3a, para os grupos de haplótipos, que apresentavam, respectivamente aqueles mesmos marcadores.

Estudos sobre a origem dos haplogrupos que caracterizam as populações euro-asiáticas sugerem que linhagens L3 foram as únicas a serem carregadas da África pelas migrações do *Homo sapiens sapiens*, que em última análise deram origem às populações modernas dos outros

Tabela 2: PRINCIPAIS POLIMORFISMOS DE SÍTIOS DE RESTRIÇÃO ASSOCIADOS A HAPLOGRUPOS ESPECÍFICOS DE POPULAÇÕES AFRICANAS (Afr), ASIÁTICAS (As) E NATIVO-AMERICANAS (NA).

Haplogrupo	Populações	Polimorfismos de mtDNA
L1	Afr	3592 (+) <i>Hpa</i> I; 11641 (+) <i>Hae</i> III
L2	Afr	3592 (+) <i>Hpa</i> I; 16389 (+) <i>Hinf</i> I; 16390 (-) <i>Ava</i> II
L3a	Afr	3592 (-) <i>Hpa</i> I; 10394 (-) <i>Dde</i> I
L3b	Afr	3592 (-) <i>Hpa</i> I ; 10084 (+) <i>Mbo</i> I
L3d	Afr	3592 (-) <i>Hpa</i> I ; 8616 (-) <i>Mbo</i> I
L3e	Afr	3592 (-) <i>Hpa</i> I ; 2349 (+) <i>Mbo</i> I
A	As e NA	663 (+) <i>Hae</i> III
B	As e NA	Deleção de 9-bp (região V)
C	As e NA	13259 (-) <i>Hinc</i> II/ 13259
D	As e NA	5176 (-) <i>Alu</i> I
X	NA	1715 (-) <i>Dde</i> I
E	As	7598 (-) <i>Hha</i> I
F	As	12406 (-) <i>Hpa</i> I/ <i>Hinc</i> II; 16517 (+) <i>Hae</i> III
G	As	4830 (+) <i>Hae</i> II; 4830 (-) <i>Hha</i> I
M	As	10394 (+) <i>Dde</i> I; 10397 (+) <i>Alu</i> I

Os números representam a posição do nucleotídeo da extremidade 5' do sítio de reconhecimento da endonuclease de acordo a sequência de referência (ANDERSON e cols., 1981; ANDREWS e cols., 1999). (+) e (-) indicam respectivamente a presença e a ausência do sítio em questão.

continentes (CHEN e cols., 1995; TORRONI e cols., 1996; KIVISILD e cols., 1999a,b; MACAULAY e cols., 1999). Esses estudos foram baseados na similaridade das linhagens africanas dos haplogrupos L3 com mtDNAs europeus e asiáticos, e na ausência do polimorfismo *HpaI* + 3592 em populações não africanas.

I.3.3.2 – HAPLOGRUPOS DE mtDNA CARACTERÍSTICOS DE POPULAÇÕES ASIÁTICAS E NATIVO-AMERICANAS

A variabilidade do mtDNA nas populações asiáticas e nativo americanas vem sendo extensivamente estudada (SCHURR e cols., 1990; CHAKRABORTY & WEISS, 1991; BALLINGER e cols., 1992; HORAI e cols., 1993, 1996; WALLACE e TORRONI, 1992; TORRONI e cols., 1993a, 1993b, 1994a,b; SZATHMARY, 1993; BATISTA e cols., 1995; BIANCHI e cols., 1995; BONNATO e cols., 1996; MERRIWETHER e cols., 1995, 1996; MERRIWETHER e FERREL, 1996; KOLMAN e cols., 1995, 1996; RIBEIRO DOS SANTOS e cols., 1996; FORSTER e cols., 1996; BONATTO e SALZANO, 1997a,b; COMAS e cols., 1998; STONE & STONEKING, 1998, 1999; OOTA e cols., 1999; RICKARDS e cols., 1999; SCHURR e cols., 1999). Esses estudos têm fornecido informações importantes sobre o povoamento das Américas e acredita-se atualmente que a migração de populações asiáticas, através do Estreito de Bering, durante a última Era Glacial, cerca de 10-30 mil anos atrás, teriam permitido o povoamento inicial das Américas (TORRONI e cols., 1992, 1993a,b; BIANCHI e cols., 1995; FORSTER e cols., 1996; BONATTO e SALZANO, 1997a,b; STONE e STONEKING, 1998, 1999).

Vários autores demonstraram que todos os mtDNAs asiáticos podem ser subdivididos em dois “macro” grupos caracterizados pela presença (+) ou ausência (-) do sítio de *Ddel* na posição 10394 do mtDNA (BALLINGER e cols., 1992; TORRONI e cols., 1993a e 1993b; WALLACE, 1995; KOLMAN e cols., 1996). Grande parte dos haplogrupos asiáticos caracterizados pelo polimorfismo *Ddel* + 10394 possuem também uma transição C para T na posição 10400, que origina um sítio adjacente de *AluI* na posição 10397 e define o haplogrupo M, presente em cerca de 55% das populações da Sibéria e do leste asiático (BALLINGER e cols., 1992; TORRONI e cols., 1993b, 1994c; CHEN e cols., 1995; WALLACE, 1995).

Na Ásia, outros haplogrupos específicos (A, B, C, D, E, F e G) (TAB 2) foram caracterizados (BALLINGER e cols., 1992; TORRONI e cols., 1993b; TORRONI e WALLACE 1995; WALLACE, 1995), e mais recentemente foram descritos também os haplogrupos Y e Z (SCHURR e cols., 1999). Os haplogrupos C, D, G, E e Z são originários do haplogrupo M, e portanto também caracterizados pelos polimorfismos *Ddel* + 10394 e *AluI* + 10397. As mutações que caracterizam cada um desses haplogrupos são as seguintes:

- O haplogrupo A é definido pelo ganho de um sítio de *HaeIII* na posição 663 e por outras quatro transições de nucleotídeos nas posições 16223 (C→T), 16290 (C→T), 16319 (G→A) e 16362 (T→C) da RC.
- O haplogrupo B é definido por uma deleção de um segmento repetitivo de 9 bp na região V e por duas transições T→C nas posições 16189 e 16217 da RC.
- O haplogrupo C é definido pela perda de um sítio de *HincII* na posição 13259 e o ganho simultâneo de um sítio de *AluI* na posição 13262 (devido a uma transição A→G em 13262), além das transições 16223 (C→T), 16298 (T→C) e 16327 (C→T) na RC.
- O haplogrupo D é definido pela perda de um sítio de *AluI* na posição 5176 (transição C→T em 5178) e pelas mutações em 16223 (C→T) e 16362 (T→C) na RC.
- O haplogrupo E é definido pela perda de um sítio de *HhaI* na posição 7598.
- O haplogrupo F é definido pela perda combinada dos sítios de *HpaI/HincII* na posição 12406 e está presente principalmente nas populações do sul asiático (TORRONI e cols., 1993b; WALLACE, 1995).
- O haplogrupo G é definido por um sítio de *HaeIII* na posição 4830 além de um sítio de *HhaI* na posição 4831.

Os haplogrupos Y e Z foram caracterizados recentemente (SCHURR e cols., 1999) apenas por análises de sequenciamento da Região Controle e seus polimorfismos característicos estão apresentados a seguir:

- O haplogrupo Y é definido pelas substituições em 16126 (T→C), 16189 (T→C), 16231 (T→C), 16266 (C→T), e 16519 (T→C).
- O haplogrupo Z é definido pelas substituições em 161289 (G→A), 16185 (C→T), 16224 (T→C), 16260 (C→T), 16298 (T→C), e 16519 (T→C).

Os haplogrupos A, B, C e D foram predominantemente encontrados nas tribos indígenas americanas já analisadas (TORRONI e cols., 1993a, 1994a; GINTHER e cols., 1993; MONSALVE e cols., 1994, 1996; FORSTER e cols., 1996; BONATTO & SALZANO, 1997a), o que suporta a hipótese de que linhagens ancestrais desses haplogrupos deram origem a mais de 95% de todos os mtDNAs nativo-americanos.

Em algumas populações nativas da América do Norte e da região Amazônica o haplogrupo X pode ser também encontrado (RIBEIRO-DOS-SANTOS e cols., 1996; SCOZZARI e cols., 1997; BROWN e cols., 1998; STONE e STONEKING 1998,1999; SMITH e cols., 1999). Esse haplogrupo não foi identificado ainda em populações asiáticas e curiosamente ele é observado na Europa tendo sido descrito por Torroni e colaboradores (1996) como característico daquelas populações (TORRONI e cols., 1996, RICHARDS e cols., 1996,1998; MACAULAY e cols., 1999; KIVISILD e cols., 1999a,b;

SIMONI e cols., 2000). Por esse motivo os primeiros trabalhos consideravam que a presença de um haplogrupo "europeu" em populações nativo-americanas representava processos de mistura gênica com os colonizadores. Entretanto, Stone e Stoneking (1998, 1999) demonstraram a presença de linhagens do haplogrupo X em populações pré-históricas do Nordeste dos Estados Unidos, ao analisarem ossadas de um antigo cemitério indígena. Esses resultados sugeriram que o haplogrupo X consistia na verdade em mais um haplogrupo fundador do pool de mtDNA nativo-americano. Finalmente, Brown e colaboradores (1998) demonstraram que as linhagens européias e nativo-americanas do haplogrupo X podem ser diferenciadas por mutações específicas, embora elas tenham se originado de um mesmo ancestral comum entre 12.000 e 36.000 anos atrás.

I.3.3.3 – HAPLOGRUPOS DE MTDNA CARACTERÍSTICOS DE POPULAÇÕES EUROPÉIAS

Torroni e colaboradores (1994c, 1996) demonstraram que todos os mtDNAs europeus podem ser divididos em dois grandes grupos caracterizados pela presença (1/4) ou ausência (3/4) do sítio de *Ddel* na posição 10394. Além dessa macro subdivisão, 9 haplogrupos de mtDNA distintos foram observados e nomeados Haplogrupos H, I, J, K, T, U, V, X e W (TAB 3). Os haplogrupos H, T, U, V, W e X pertencem à subdivisão caracterizada por *Ddel* – 10394, enquanto os haplogrupos I, J e K têm em comum o polimorfismo *Ddel* + 10394.

Os marcadores de RFLP que definem cada um dos haplogrupos europeus estão sumarizados na tabela 3:

- O haplogrupo H, definido pelos polimorfismos *Ddel* – 10394 e *AluI* – 7025 (devido a uma transição C→T na posição 7028) é o principal haplogrupo europeu, compreendendo cerca de 40% dos mtDNAs caucasianos (TORRONI e cols., 1994c, 1996; CÔRTE-REAL e cols., 1996).
- O haplogrupo I é definido pelas mutações: perda dos sítios de *Ddel* na posição 1715 e de *HaeIII* na posição 4529; ganho de um sítio de *AluI* na posição 10032; ganho de um sítio de *Avall* na posição 8249, perda de um sítio de *HaeIII* na posição 8250; ganho de um sítio de *BamHI/MboI* na posição 16389 e consequentemente, perda de um sítio de *Avall* na posição 16390. Foram definidas também as mutações de RC específicas do haplogrupo I, sendo elas: A na posição 16129; T na posição 16223 e C na posição 16311 (TORRONI e cols., 1994c, 1996).
- O haplogrupo J é definido pela perda de um sítio de *BstNI* na posição 13704 e pela perda de um sítio de *HinfI* na posição 16065 (TORRONI e cols., 1994c, 1996).
- O haplogrupo T é definido pela presença dos sítios de *BamHI* na posição 13366 e *AluI* na posição 15606 (TORRONI e cols., 1996).

Tabela 3: PRINCIPAIS POLIMORFISMOS DE SÍTIOS DE RESTRIÇÃO ASSOCIADOS A HAPLOGRUPOS ESPECÍFICOS DE POPULAÇÕES EUROPÉIAS.

Haplogrupo	Polimorfismos de mtDNA
H	7025 (–) <i>AluI</i> ; 10394 (–) <i>DdeI</i>
I	8249 (+) <i>AvaII</i> ; 10032 (+) <i>AluI</i> ;
J	13704 (–) <i>BstNI</i> ; 16065 (–) <i>HinfI</i>
K	9052 (–) <i>HaeII</i> ; 10394 (+) <i>DdeI</i>
T	13366 (+) <i>BamHI</i> ; 15606 (+) <i>AluI</i>
U	12308 (+) <i>HinfI</i> ; 10394 (–) <i>DdeI</i>
V	4577 (–) <i>NlaIII</i> ; 10394 (–) <i>DdeI</i>
X	1715 (–) <i>DdeI</i>
W	8249 (+) <i>AvaII</i> ; 8994 (–) <i>HaeIII</i>

Os números representam a posição do nucleotídeo da extremidade 5' do sítio de reconhecimento da endonuclease de acordo a sequência de referência (ANDERSON e cols., 1981; ANDREWS e cols., 1999). (+) e (–) indicam respectivamente a presença e a ausência do sítio em questão.

- O haplogrupo U é caracterizado pela presença de um sítio de *HinfI* na posição 12308 comum ao haplogrupo K, não associado ao polimorfismo *HaeII* + 9052 (TORRONI e cols., 1996).
- O haplogrupo K é definido pela combinação de perda do sítio de *HaeII* na posição 9052 e perda do sítio de *HhaI* na posição 9053 (TORRONI e cols., 1994c, 1996).
- O haplogrupo V é definido pela perda de um sítio de *NlaIII* na posição 4577 além do polimorfismo *DdeI* – 10394 (TORRONI e cols., 1996).
- O haplogrupo W é definido pela presença de um sítio de *AvaII* na posição 8249 e pela perda de um sítio de *HaeIII* na posição 8994 (TORRONI e cols., 1996).

Finalmente, o haplogrupo X é definido pela perda de um sítio de *DdeI* na posição 1715, também comum ao haplogrupo I, associado à presença de um sítio para a enzima de restrição *AccI* na posição 14465 (TORRONI e cols., 1996; BROWN e cols., 1998).

Os polimorfismos de sequência da RC foram também definidos para alguns desses haplogrupos (FRANCALLACCI e cols., 1996; TORRONI e cols., 1996, 1998, 2000; RICHARDS e cols., 1996, 1998; BROWN e cols., 1998; MACAULAY e cols., 1999; HELGASON e cols., 2000) o que facilita a análise da presença desses haplogrupos em outras populações. A partir daí constatou-se que os haplogrupos I, K, J e W são essencialmente específicos de Europeus, enquanto os outros haplogrupos podem ser encontrados em frequências muito baixas em populações que historicamente já estiveram em contato com populações européias. (TORRONI e cols., 1994c, 1996; COMAS e cols., 1996, 1998; PASSARINO e cols., 1996a, 1996b). Como a maioria desses 10 haplogrupos parecem estar confinados às populações da Eurásia Ocidental, acredita-se que eles surgiram em uma população caucasóide ancestral depois da sua separação dos progenitores dos asiáticos e africanos modernos (TORRONI e cols., 1996, RICHARDS e cols., 1996, 1998; KIVISILD e cols., 1999a,b; MACAULAY e cols., 1999).

Um grande número de mtDNAs europeus foi descrito na literatura, principalmente nos últimos 5 anos (TORRONI e cols., 1994c, 1996, 1997, 1998; CÔRTE-REAL e cols., 1996; RICHARDS e cols., 1996, 1998; HOFMANN e cols., 1997; MACAULAY e cols., 1999, SIMONI e cols., 2000; HELGASON e cols., 2000), provocando uma revisão na filogenia das linhagens européias e sugerindo uma nova classificação para seus haplogrupos. Por exemplo, durante muito tempo o haplogrupo **K** foi considerado como um grupo independente (TORRONI e cols., 1994, 1996), mas as análises mais recentes sugerem a sua redefinição como um sub-grupo do haplogrupo **U** (MACAULAY e cols., 1999). Da mesma forma, foi observado também que os haplótipos de vários haplogrupos podiam ser associados em sub-haplogrupos, que por sua vez, são definidos por substituições específicas de região controle e/ou polimorfismos de RFLP (RICHARDS e cols., 1998; FRANCALLACCI e cols., 1999; KIVISILD e cols., 1999a,b; MACAULAY e cols., 1999).

I.4 – ANÁLISES DA VARIABILIDADE GENÉTICA DA POPULAÇÃO BRASILEIRA DESENVOLVIDAS EM NOSSO LABORATÓRIO

Nos últimos anos, nosso grupo de pesquisas iniciou uma análise da variabilidade genética dos “caucasóides” brasileiros (ou indivíduos “brancos”) que, tendo em vista os aspectos da colonização do país, constituem os descendentes dos colonizadores e imigrantes europeus. Nós optamos por analisar as contribuições maternas dessa fração da população brasileira, determinando a origem das linhagens mitocondriais encontradas nas nossas amostras. Os inúmeros avanços na caracterização dos haplogrupos continente-específicos de populações africanas, asiáticas, ameríndias e européias (descritos acima), fazem com que a alocação de determinado haplótipo ou linhagem de mtDNA em um desses grupos permita a identificação de sua origem geográfica ancestral. Por isso, a análise da variabilidade do mtDNA tem sido bastante utilizada nos últimos anos para se avaliar a origem (materna) de populações resultantes de processos de miscigenação (TORRONI e cols., 1995; MATEU e cols., 1997; BRAVI e cols., 1997; MERRIWETHER e cols., 1997; RANDO e cols., 1998; CARVAJAL-CARMONA e cols., 2000). Nossos primeiros resultados foram referentes à análise do polimorfismo de deleção de 9-bp na região V do mtDNA em 250 indivíduos. Nós verificamos que, diferentemente das populações européias onde esse polimorfismo aparece apenas em casos isolados, nos brasileiros brancos ele pode ser observado em alta frequência (ALVES-SILVA e cols., 1999b; Apêndice 1). Além disso, esse estudo nos permitiu identificar uma linhagem de mtDNA que apresentava um polimorfismo raro de inserção da mesma unidade repetitiva de 9 pares de base, que não havia sido até então observada em populações européias (ALVES-SILVA e cols., 1999a; Apêndice 2). Quando nós utilizamos o sequenciamento da região controle do mtDNA e análises de RFLP para determinar a origem das linhagens com a deleção de 9-pb, nós verificamos que 76,2% delas eram de origem ameríndia/asiática, 19% de origem africana e apenas 4,8% de origem européia (ALVES-SILVA, 1997; ALVES-SILVA e cols., 1999b). Esses dados sugeriram que, durante a formação da população caucasóide brasileira houve uma grande contribuição materna por parte de populações ameríndias e africanas. Como nós analisamos apenas uma fração pequena da nossa amostragem dos brancos brasileiros (21 indivíduos que apresentaram a deleção de 9bp), não foi possível determinar a real contribuição de cada um daqueles grupos. Em um trabalho seguinte foram analisados 100 indivíduos brancos da região Sudeste do país, através das metodologias de sequenciamento das regiões hipervariáveis I e II da Região Controle e análises de RFLP (SANTOS, 1998). Neste estudo foi demonstrado que mais de 65% das linhagens mitocondriais analisadas apresentavam haplótipos tipicamente ameríndios e africanos, enquanto apenas 31% eram linhagens européias. No presente trabalho, portanto, nós resolvemos ampliar a análise do mtDNA para um grupo maior de indivíduos, tendo o cuidado de selecionar uma amostra mais representativa do Brasil

como um todo, com o principal objetivo de verificar se a alta variabilidade encontrada na população branca da região Sudeste se estendia também às outras regiões do país.

I.5 – ANÁLISE DE UMA POPULAÇÃO PRÉ-HISTÓRICA BRASILEIRA

Não existem dúvidas quanto à extrema importância de se estudar as populações humanas contemporâneas, em busca de marcadores genéticos que possam ajudar no entendimento dos processos evolutivos que ocorreram há centenas e até mesmo milhares de anos. Esses estudos são direcionados na maioria das vezes a questões de colonização dos diversos continentes, e à origem geográfica de populações ancestrais, através da análise filogenética de genes e/ou linhagens de DNA (TORRONI e WALLACE, 1994; WALLACE, 1995, 1997; BANDELT e cols., 1995; CAVALLI-SFORZA, 1998; HEY, 1998; CHAKRAVARTI, 1999; WALLACE, 1999). Os resultados desses estudos têm se mostrado, às vezes, bastante conclusivos, como por exemplo, a descoberta da origem asiática das populações nativo americanas atuais (para uma revisão ver TORRONI e cols., 1993a, BONATTO e SALZANO, 1997a,b; FORSTER e cols., 1996). Por outro lado, eles deram margem também ao surgimento de importantes pontos de debate, principalmente no que diz respeito às rotas de migração que teriam sido traçadas pelas populações ancestrais e ao tempo em que esses eventos teriam ocorrido (HORAI e cols., 1993; TORRONI e cols., 1993; FORSTER e cols., 1996; BONATO & SALZANO 1997). Como exemplo, podemos citar os dados recém publicados por Quintana-Murci e cols. (1999) onde os autores apresentaram evidências de que a saída do homem da África parece ter ocorrido, primeiramente, pelo sul da península arábica e não pelo Norte do Egito, como se acreditava anteriormente (STRINGER e cols., 1989).

Entretanto, a história da evolução humana foi marcada também por processos de extinção em massa de indivíduos, levando ao extermínio de populações inteiras. Com um exemplo, podemos citar a extinção do Homem de Neandertal (*Homo neanderthalensis*) (SHREEVE, 1995; BAHN, 1998; MELLARS, 1998). Baseando-se na hipótese de que essa espécie co-habitou as mesmas regiões geográficas com o *Homo sapiens* sem que ocorresse entre eles nenhum grau de mistura gênica, somente através da extração de DNA alguns destes fósseis, seria possível (e de fato foi) recuperar a informação genética do homem de Neandertal (KRINGS e cols., 1997, 1999a; OVCHINNIKOV e cols., 2000).

No Brasil, os sítios arqueológicos da região de Lagoa Santa (município de Santana do Riacho), onde foram encontrados restos humanos do que se convencionou chamar de “Homem de Lagoa Santa”, têm despertado um grande interesse na comunidade científica nos últimos anos. Estudos morfométricos e craniológicos realizados nesses espécimens indicaram uma origem diferente daquela sugerida para as demais populações nativo americanas (MUNFORD e cols., 1995; NEVES e cols., 1997, 1998, 1999). O Homem de Lagoa Santa parece não ter compartilhado o padrão

“mongolóide” de formato craniano, característico das demais populações nativas da América do Sul, apresentando por sua vez, um crânio dolicocefálio (caracterizado por queixo proeminente, aberturas nasais mais largas e face mais longitudinal), com maior semelhança às características negróides encontradas em populações ancestrais do sul da Ásia e austronésios (para maiores detalhes ver POWELL & NEVES, 1999).

Muito debate ainda existe entre os pesquisadores que investigam a origem do homem nas Américas através da análise da variabilidade genética das populações nativas americanas atuais (para uma revisão ver POWELL & NEVES 1999). É praticamente um consenso que essas migrações foram provenientes da Ásia, e ocorreram principalmente durante a última era glacial, mas diferentes pesquisadores defendem a ocorrência de uma (FORSTER e cols., 1996; MERRIWETHER e cols., 1996; BONATTO e SALZANO 1997a, 1997b; SANTOS e cols., 1999), duas (SZATHMARY 1986), três (GREENBERG e cols., 1986; GREENBERG, 1987; SCHURR e cols., 1990; TORRONI e cols., 1992, 1993a, WALLACE, 1995), e até mesmo quatro (BALLINGER e cols., 1992; HORAI e cols., 1993; NEVES e cols., 1996a,b) ondas migratórias distintas. Neves e colaboradores (1996b), que propõem uma hipótese de quatro migrações, sugeriram que uma onda migratória “extra” teria sido responsável pela origem das populações pré-históricas de Lagoa Santa. Uma vez que os grupos indígenas brasileiros (que ainda se mantêm isolados da influência de miscigenação com europeus e africanos) não apresentam as características previstas para o “homem de Lagoa Santa”, acredita-se que essa população tenha sido extinta. Por outro lado, alguns autores sugerem que tribos indígenas atuais, como os Gê Botocudos do Planalto Brasileiro, representam os descendentes vivos do Homem de Lagoa Santa (RAMOS, 1951).

Portanto, a análise direta da variabilidade genética da população pré-histórica de Lagoa Santa, através de extração e caracterização do DNA de alguns desses fósseis, pode ser a única forma (ou pelo menos a mais informativa), de se resgatar a informação genética necessária para responder questões relacionadas à sua origem.

Mais uma vez o DNA mitocondrial é a ferramenta mais adequada para estudos que envolvem a recuperação de amostras fósseis, por representar um genoma naturalmente amplificado.

I.5.1 – mtDNA E ANÁLISES DE RACEMIZAÇÃO DE AMINOÁCIDOS EM ESTUDOS DE ARQUEOLOGIA MOLECULAR.

Diferentemente do DNA nuclear, o mtDNA está presente em múltiplas cópias por célula, com o número de moléculas variando de várias centenas em fibroblastos de camundongo, até mais que 100.000 em oócitos maduros (SATO e KUROIWA, 1991; WALLACE, 1995). Assim, cerca de um terço do DNA total em um oócito maduro é mtDNA (LIGHTOWLERS e cols., 1997). Além disso, o número de mitocôndrias por célula também é altamente variado e está intimamente relacionado com

as necessidades energéticas de cada tecido, podendo chegar a até 2000 organelas por célula, como é o caso dos hepatócitos (ALBERTS e cols., 1983).

A presença de centenas ou milhares de cópias iguais de mtDNA num determinado tecido aumenta as chances de preservação de moléculas intactas em amostras altamente degradadas, que podem ser recuperadas e amplificadas por reações de PCR. Essa característica é especialmente importante em estudos de antropologia molecular e forense onde geralmente há pouco DNA, como por exemplo em manchas de sangue (GILL e cols.; 1987), em um fio de cabelo (HIGUCHI e cols., 1988; VIGILANT e cols., 1989) ou num pequeno pedaço de osso (HAGELBERG e cols., 1989, 1991a, b, 1994; PÄÄBO, 1989; PÄÄBO e cols., 1989; HORAI e cols., 1989; HAGELBERG e CLEGG, 1991; HÖSS e PÄÄBO, 1993; HAUSWIRT e cols., 1994; KRINGS e cols., 1997).

Assim, a última parte do nosso projeto está relacionada com a caracterização molecular de amostras arqueológicas da região de Lagoa Santa, e por isso, nossos objetivos principais se concentram na extração de DNA de amostras ósseas desses espécimens fósseis.

Apesar de ter se desenvolvido de forma muito acelerada, principalmente nos últimos 10 anos, as metodologias aplicadas na obtenção e caracterização de DNA de fósseis ainda se deparam com grandes obstáculos. Os problemas ao se trabalhar com amostras degradadas estão relacionados com a contaminação com DNA moderno e principalmente, com a dificuldade de se confirmar a autenticidade dos resultados, especialmente quando exemplares humanos estão envolvidos (HANDT e cols., 1996, LINDAHL, 1997; KUMAR e cols., 1999). Outro problema consiste na destruição de amostras raras, uma vez que uma quantidade considerável de material tem que ser comprometida, na maioria das vezes, para se obter mesmo baixas concentrações de DNA (POINAR e cols., 1996, 1999).

Assim, um grande esforço tem sido direcionado para o aprimoramento de técnicas mais sensíveis e eficientes de extração de DNA. Além disso, têm sido desenvolvidos métodos alternativos de análise, que permitem avaliar a probabilidade de se obter DNA autêntico em fósseis utilizando-se quantidades mínimas de amostra. Esses métodos evitam a destruição desnecessária de exemplares únicos e arqueologicamente importantes.

Com esse objetivo, Poinar e cols. (1996, 1999) demonstraram que a razão entre as formas D e L dos aminoácidos (aa) de uma amostra fóssil pode indicar o nível de degradação molecular sofrido por essa amostra. Quando um ser vivo morre, os aminoácidos constituintes de suas proteínas, que se encontram todos na forma L, começam a sofrer um processo contínuo e dinâmico de racemização para o isômero D, até que essas duas formas atinjam concentrações de equilíbrio. Observou-se que os mesmos fatores responsáveis por esse processo de racemização de aminoácidos (R-AA) estavam também relacionados à degradação do DNA nas células. Os resultados obtidos por Poinar e colaboradores (1996) para várias espécies diferentes sugeriram que níveis de até 0.10 para a razão das concentrações das formas D e L para o aminoácido Ácido Aspártico (Asp) (que apresenta maior

velocidade de racemização) são compatíveis com a obtenção de DNA autêntico em uma amostra (POINAR e cols., 1996; KRINGS e cols., 1997). Essa técnica mostrou-se ainda capaz de avaliar, de modo indireto, através dos níveis de racemização do aminoácido Alanina, o nível de contaminação dos fósseis com fontes modernas de proteína e DNA. Entretanto, o fator mais importante relacionado à análise prévia de R-AA em amostras fósseis, consiste na necessidade de pequenas quantidades de material, na ordem de 10 mg de amostra, enquanto a extração de DNA envolve no mínimo, cerca 2 a 3 gramas de tecido.

Nós portanto, decidimos padronizar e implantar a metodologia de análise de racemização de aminoácidos em amostras fósseis, e utilizar essa metodologia para verificar se nossas amostras da região de Lagoa Santa apresentam níveis de preservação molecular compatíveis com a recuperação de DNA autêntico. Essa parte do trabalho foi feita em colaboração com os professores André Prous do Museu de História Natural da UFMG e o professor Walter Neves, que nos cederam as amostras, além dos Drs. Mark Stoneking e Svante Pääbo do Max Planck Institute for Evolutionary Anthropology na Alemanha, que nos auxiliaram na padronização da metodologia de análise de racemização de aminoácidos em amostras fósseis.

Objetivos

II - OBJETIVOS

OBJETIVOS GERAIS

- ⇒ Analisar a variabilidade do mtDNA em amostras da população branca brasileira de diferentes regiões do país;
- ⇒ Tentar caracterizar o mtDNA de uma população pré-histórica do Brasil.

OBJETIVOS ESPECÍFICOS

1. Determinar a origem geográfica de uma linhagem mitocondrial caracterizada pelo polimorfismo raro de inserção de 9 pb, encontrada em um indivíduo brasileiro;
2. Determinar a sequência da região hipervariável I do mtDNA para aproximadamente 150 indivíduos brasileiros, provenientes das regiões Norte, Nordeste e Sul do país;
3. Determinar a origem geográfica das linhagens mitocondriais encontradas através de análises de filogeografia e testes de RFLP;
4. Comparar os resultados obtidos no presente estudo com aqueles previamente observados para a região Sudeste;
5. Realizar análises de racemização de aminoácidos em amostras de uma população pré-histórica brasileira;
6. Verificar se o nível de degradação molecular desses fósseis apresenta-se compatível com a recuperação de DNA;
7. Tentar recuperar DNA das amostras que se apresentarem promissoras através dos resultados de racemização de aminoácidos;

Material e Métodos

III.1 – Os INDIVÍDUOS

III.1.1 – AMOSTRAS DA POPULAÇÃO BRASILEIRA ATUAL

No presente trabalho foram analisadas 149 amostras de DNA de indivíduos brasileiros que foram gentilmente fornecidas pelo Núcleo de Genética Médica de Minas Gerais. Nós estávamos interessados em analisar uma amostra de pessoas brancas da população brasileira que apresentasse uma ampla distribuição geográfica no país, e por isso foram selecionados, entre os vários indivíduos que se submeteram a testes de paternidade naquele laboratório, 48 indivíduos da Região Norte, 50 indivíduos da Região Nordeste e 50 indivíduos da região Sul do Brasil. A princípio esses indivíduos não apresentavam nenhuma relação de parentesco entre si e em sua maioria podem ser considerados “brancos”. Uma amostra de DNA de um indivíduo da região Sudeste proveniente do mesmo laboratório foi também analisado. O consentimento formal e informado foi obtido de todos os indivíduos cujas amostras de DNA foram utilizadas neste trabalho.

III.1.2 – AMOSTRAS DA POPULAÇÃO DE PORTUGAL

Noventa e seis amostras de DNA de indivíduos portugueses da região do Porto foram analisadas. Essas amostras foram gentilmente cedidas pelo professor Jorge Rocha da Universidade do Porto - Portugal.

III.1.3 – AMOSTRAS DE UMA POPULAÇÃO BRASILEIRA PRÉ-HISTÓRICA

Vinte e duas amostras de ossadas humanas provenientes de sítios arqueológicos da região de Lagoa Santa, município de Santana do Riacho, e três amostras de outros sítios arqueológicos (Lapa do Boquete, Lapa do Malhador e Caixa d'água de Buritizeiros, foram também analisadas neste estudo. Essas amostras foram gentilmente fornecidas pelos professores Walter Neves e André Prous e constituem os remascentes de uma população extinta que habitou aquela região. Essas ossadas fazem parte de uma coleção arqueológica que se encontra hoje sob a curadoria do prof. André Prous no Museu de História Natural da UFMG. Os exemplares ósseos selecionados para o nosso estudo consistiam principalmente de fragmentos de ossos longos, como fêmur, tíbia, e costelas.

III.2 – OS INICIADORES

Todos os iniciadores utilizados nas reações de amplificação por PCR foram sintetizados pela Operon Technologies (Alameda, CA, EUA). A sequência dos iniciadores é apresentada nas tabelas 4 e 5.

III.3 – AS ENZIMAS DE RESTRIÇÃO

As enzimas utilizadas nas análises de RFLP foram adquiridas das empresas Promega, Gibco-BRL, Sigma, Stratagene e New England Biolabs®, acompanhadas de seus tampões próprios de reação. Na tabela 6 estão relacionadas as condições específicas de digestão para cada enzima.

III.4. – ANÁLISE DE AMOSTRAS CONTEMPORÂNEAS

III.4.1 – AMPLIFICAÇÃO POR PCR DA REGIÃO CONTROLE DE AMOSTRAS CONTEMPORÂNEAS

As amplificações iniciais por PCR foram realizadas utilizando-se dois pares de iniciadores: MiL15926 e MiH16498 para a região hipervariável I e; MiL029 e MiH580 para a região hipervariável II (vide sequências na tabela 4). As amplificações foram realizadas em 50µL de volume final contendo 0.8 pmol/µL (0.8µM) dos iniciadores, 200µM de cada dNTP e 0,5 unidade de *Taq* DNA polimerase (Cenbiot) em tampão de reação Tris-HCl 10 mM pH 8,5, KCl 50 mM, MgCl₂ 1,5 mM. Em cada tubo de PCR foram adicionados 40 ng de amostra de DNA. O programa de amplificação executava uma etapa inicial de desnaturação de 5 min a 94°C, seguida de 30 ciclos de desnaturação a 94°C por 1 min, anelamento a 55°C por 45 seg e extensão a 72°C por 1 min. Em cada grupo de reações de PCR foram feitos controles negativos (sem a adição de DNA) para verificar a contaminação dos reagentes. As reações de PCR foram realizadas em termocicladores PT-100 da MJ Research. Os produtos de PCR de 612 pb e 591 pb foram visualizados em géis de agarose 1% corados com brometo de etídeo (0,10µg/ml) e purificados com o kit Wizard™ PCR Preps (Promega) para serem utilizadas posteriormente em reações de sequenciamento.

Uma estratégia diferente foi adotada na amplificação da região hipervariável I de 25 amostras da região Sul. Ela consistiu na utilização do par de iniciadores MiH16498 e MiL15926-M13, possibilitando que o iniciador M13-direto fluorescente fosse usado na reação de sequenciamento. O iniciador L dessa primeira reação foi sintetizado associando-se a sequência do iniciador M13-direto à extremidade 5' do iniciador MiL15926.

Tabela 4: INICIADORES UTILIZADOS NAS REAÇÕES DE PCR E SEQUENCIAMENTO ENVOLVENDO AMOSTRAS DE DNA MODERNO

Sistema	Iniciador*	Sequência (5'-3')	Temp. de Anelamento (°C)	Tamanho do Fragmento de PCR (bp)
	MiL8209	CATCGTCCTAGAATTAATTCC	55	86 ou 95
	MiH8304	CTTTACAGTGGGCTCTAGAGG		
	MiL29	GGTCTATCACCTATTAACCAC	55	591
	MiH80	TTGAGGAGGTAAGCTACATA		
	MiL15926	TCAAAGCTTACACCAGTCTTGTA AAAACC	55	612
	MiH16498	CCTGAAGTAGGAACCAGATG		
	MiL48^F	fluoresceína-CTCACGGGAGCTCTCCATGC	55	
	MiH80^F	fluoresceína-CTGTTAAAAGTGCATACCGCCA	55	
	MiL15996^F	fluoresceína-TCCACCATTAGCACCCAAAGC	55	
	MiH16401^F	fluoresceína-TGATTTACGGAGGATGGTG	55	
S1	MiL16453	CCGGGCCCATAACTTGGG	61	1.812
	MiH1696	GGAGTGGGTTTGGGGCTAGG		
S2	MiL1562	GTAACATGGTAAGTGTACTG	51	2.096
	MiH3717	GGCTACTGCTCGCAGTG		
S3	MiL3108	TTCAAATTCCTCCCTGTACG	53	2.809
	MiH5917	CGGTCGGCGAACATCAGTGG		
S4	MiL5591	GACTGCAAAACCCCACTCTG	53	1.842
	MiH7433	GTATACGGGTTCTTCGAATG		
S5	MiL7367	CTCCATAAACCTGGAGTG	51	1.806
	MiH9172	GTGTGAAAACGTAGGCTTG		
S6	MiL8282	CCCCTCTAGAGCCCACTGTAAAGC	55	1.825
	MiH10107	GTAGTAAGGCTAGGAGGGAG		
S7	MiL9802	TAGCCACAGGCTTCCACGG	59	2.071
	MiH11873	TGGGGGGTAAGGCGAGGTTAGCG		
S8	MiL11673	CCCCCTGAAGCTTCACCGG	57	2.277
	MiH13950	GGGGATTGTGCGGTGTGTG		
S9	MiL13809	CGCTGTCACTTTCCTAGGAC	57	2.738
	MiH16547	GGAACGTGTGGGCTATTTAGG		
S10	MiL607	CACTGAAAAATGTTTAGACGGG	56	101
	MiH707	GGGATGCTTGCATGTGTAATC		
S11	MiL5150	CCTACTACTATCTCGCACCTG	58	68
	MiH5217	AGAGGAGGGTGGATGGAATTA		
S12	MiL13209	CGCCCTTACACAAAATGACATCAA	58	93
	MiH13301	GGTTGGTTGATGCCGATTGTA		
S13	MiL16475	TAGCTAAAGTGAAGTGTATCC	55	113
	MiH18	GGTGATAGACCTGTGATCCAT		
	M13-40	GTTTTCCCAGTCACGAC	60	
	M13 reverso	CAGGAAACAGCTCTGAC	55	
	M13 Universal^a	CGACGTTGTAAAACGACGGCCAGT	60	
	M13-direto^F	fluoresceína-GTAAAACGACGGCCAGTA	55	
	MiL15926-M13	GTAAAACGACGGCCAGTATCAAAGCTTACACCAGTCTTGTA AAAACC		

*A numeração dos iniciadores MiL e MiH desta tabela corresponde à posição de seus nucleotídeos 5' na molécula de mtDNA humano, de acordo com a sequência de referência CRS (ANDERSON e cols, 1981; ANDREWS e cols., 1999).

^a Iniciador fluorescente utilizado nas reações de sequenciamento do item III.4.2.

Tabela 5: INICIADORES UTILIZADOS NAS REAÇÕES DE PCR E SEQUENCIAMENTO ENVOLVENDO AMOSTRAS DE DNA ANTIGO

Sistema	Iniciador*	Sequência (5'-3')	Temp. de Anelamento (°C)	Tamanho do Fragmento de PCR (bp)
A1	MiL 16055	GAAGCAGATTTGGGTACCAC	55	123
	MiH 16139	TACTACAGGTGGTCAAGTAT		
A2	MiL16055	GAAGCAGATTTGGGTACCAC	55	126
	MiH16142	ATGTACTACAGGTGGTCAAG		
B	MiL16131	CACCATGAATATTGTACGGT	55	126
	MiH16218	TGTGTGATAGTTGAGGGTTG		
C1	MiL16209	CCATGCTTACAAGCAAGT	55	102
	MiH16271	TAGGTTTGTGTTGGTATCCTA		
C2	MiL16209	CCATGCTTACAAGCAAGT	55	131
	MiH16303	TGGCTTTATGTACTATGTAC		
D	MiL16287	CACTAGGATACCAACAAACC	55	131
	MiH16379	CAAGGGACCCCTATCTGAGG		
E1	MiL16347	GGTACATAGCACATTACAGT	55	93
	MiH16401	TGATTTACGGAGGATGGTG		
E1	MiL16347	GGTACATAGCACATTACAGT	55	102
	MiH16410	GCGGGATATTGATTTACGG		
	M13-direto	GTAAAACGACGGCCAGTA	55	
	M13-reverso	GGAAAACAGCTATGACCATG	55	

* Para esses iniciadores MiL e MiH, a numeração corresponde à posição de seus nucleotídeos 3' na molécula de mtDNA humano de acordo com a sequência de referência CRS (Anderson e cols, 1981). Os iniciadores M13-direto e M13-reverso correspondem a iniciadores que acompanham o kit de clonagem **TOPO™TA Cloning®**.

Tabela 6: ENDONUCLEASES UTILIZADAS NAS ANÁLISES DE RFLP

Enzima	Fabricante	Condições ótimas de digestão:	
		Tampão	Temperatura
<i>AluI</i>	Stratagene	Buffer # 3 _1X	37°C
<i>Avall</i>	Stratagene	Universal buffer 1X	37°C
<i>BamHI</i>	Stratagene	Universal buffer 1X	37°C
<i>BstNI</i>	New England Biolabs®	NEBuffer 2 (1X)+ BSA 100µg/ml	60°C
<i>DdeI</i>	Stratagene	Buffer # 7 _1X	37°C
<i>HaeII</i>	Promega	Buffer B _1X	37°C
<i>HaeIII</i>	New England Biolabs®	NEBuffer 2 _1X	37°C
<i>HhaI</i>	Stratagene	Universal buffer 1X	37°C
<i>HindII</i>	Promega	Buffer B _1X	37°C
<i>HinfI</i>	Stratagene	Universal buffer 1X	37°C
<i>HpaI</i>	Stratagene	Universal buffer 1X	37°C
<i>MboI</i>	Stratagene	Universal buffer 1,5 X	37°C
<i>MseI</i>	Gibco-BRL	REACT® 1 Buffer 1X	37°C
<i>NlaIII</i>	New England Biolabs®	NEBuffer 4 (1X)+ BSA 100µg/ml	37°C
<i>RsaI</i>	Stratagene	Buffer # 1 _1X	37°C
<i>TaqI</i>	Stratagene	Universal buffer 1X	65°C

III.4.2 – SEQUENCIAMENTO DA REGIÃO CONTROLE

A sequência do segmento hipervariável I da Região Controle foi determinada para todos os indivíduos e a sequência de HVS-II foi determinada apenas para algumas amostras. O método utilizado foi o de sequenciamento direto a partir de produtos de PCR. As reações de sequenciamento foram realizadas com o kit "Thermo Sequenase™ fluorescent labelled primer cycle sequencing kit with 7-deaza-dGTP" (Amersham Life Science) segundo o método descrito por Sanger e colaboradores (1977). Os iniciadores utilizados nas reações de sequenciamento foram o MiL 15996^F, MiH 16401^F, M13-direto^F (para HVSI) e MiL 048^F, MiH 480^F (para HVSII). Como pode ser observado na tabela 3 todos esses iniciadores são marcados com fluoresceína.

Os produtos do sequenciamento foram submetidos à eletroforese em gel de poliacrilamida a 6% contendo 8M, de uréia utilizando o sequenciador automático A.L.F. (Automated Labeled Fluorescent DNA sequencer) da Pharmacia Biotech. As corridas foram realizadas a 1500V, 45mA, 34W e 45°C por 420 min. As sequências foram obtidas em forma de um cromatograma e a edição das ambiguidades que apareceram algumas vezes, foi realizada de acordo com os resultados obtidos com o sequenciamento da fita complementar. Posteriormente essas sequências foram comparadas com a sequência de referência do DNA mitocondrial humano (ANDERSON e cols., 1981; ANDREWS e cols., 1999) e as mutações (ou polimorfismos) característicos de cada linhagem, e portanto de cada indivíduo, foram identificadas.

III.4.3 – ANÁLISE DAS SEQUÊNCIAS

O programa Arlequin v.1.1 (SCHNEIDER e cols., 1997) foi utilizado para cálculos de diversidade haplotípica entre as amostras. Os dados do sequenciamento da região controle foram submetidos a análises filogenéticas para avaliar a relação entre as linhagens mitocondriais utilizando-se os programas DNADIST, NEIGHBOR, PARSIMONY e SEQBOOT, do pacote de programas PHYLIP (Phylogeny Inference Package, Felsenstein, 1993). Árvores filogenéticas que combinam a informação de todos os polimorfismos observados em cada haplótipo (incluindo os dados do sequenciamento e dos testes de RFLP), foram construídas pelo método de Neighbor-Joining (SAITOU e NEI, 1987). Para a estimativa da distância filogenética entre duas linhagens utilizou-se o método de Tamura-Nei (1993), que assume taxas de mutação gama-distribuídas por posições de nucleotídeos.

As sequências foram também submetidas a pesquisas de homologia utilizando-se principalmente o servidor BLAST ("Basic Local Alignment Search Tool") (ALTSCHUL e cols, 1990) do NCBI (National Center for Biotechnology Information) da Biblioteca Nacional de Medicina do NIH (National Institutes of Health), Maryland, EUA. O banco de dados HVRbase (BURCKHARDT e cols.,

1999) também foi utilizado na identificação de sequências homólogas aos haplótipos observados na população brasileira.

III.4.4 – AMPLIFICAÇÃO POR PCR DO MTDNA TOTAL

Essa metodologia foi aplicada apenas para as amostras de indivíduos atuais, pois consiste na amplificação de grandes partes do DNA mitocondrial o que somente é possível para extratos de DNA de boa qualidade. As análises de RFLP abrangem sítios de restrição dispersos por todo o genoma mitocondrial. Por esse motivo, utilizou-se 13 sistemas de iniciadores (S1-S13, TAB. 4) que são capazes de amplificar por PCR, toda a molécula o mtDNA em fragmentos sobrepostos (TORRONI e cols., 1992). As condições de PCR foram as mesmas descritas no item III.4.1, com exceção da temperatura de anelamento que foi específica para cada par de iniciadores como está apresentado na tabela 4. Os sistemas que resultam em produtos de PCR com tamanho entre 0,62 e 2,8 Kb foram visualizados em géis de agarose 0,8% corado com brometo de etídeo. Os demais sistemas, que resultam em fragmentos pequenos de 68 a 113 bp, e foram desenvolvidos para amplificar regiões específicas de polimorfismos nativo-americanos, foram visualizados em géis de poliacrilamida 6% corados por nitrato de prata.

III.4.5 – ANÁLISE DE RFLP PARA POLIMORFISMOS CONTINENTE-ESPECÍFICOS DE MTDNA

As análises de RFLP foram realizadas a partir da digestão dos produtos de PCR descritos no item III.8 com 16 endonucleases de restrição: *AluI*, *Avall*, *BamHI*, *BstNI*, *DdeI*, *HaeII*, *HaeIII*, *HhaI*, *HincII*, *HinfI*, *HpaI*, *MboI*, *MseI*, *NlaIII*, *RsaI* e *TaqI* (TAB. 6). As digestões foram efetuadas em volume final de 30µL, contendo 6 µL de produto de PCR (~500ng), tampão de digestão próprio de cada enzima na concentração especificada pelo fabricante (1X ou 1,5X) (TAB 6) e em média 0,5 a 1 unidade de enzima para cada sítio específico presente no fragmento amplificado. As misturas de reação foram incubadas por no mínimo três horas à temperatura ótima de digestão de cada enzima (TAB 6). Os produtos de digestão foram visualizados em géis de agarose 0,8-1,0%, ou acrilamida 6-12%, de acordo com seus tamanhos e a necessidade de resolução de cada um. Esse procedimento permitiu a identificação dos polimorfismos de sítios de restrição associados aos principais haplogrupos continente-específicos apresentados nas tabelas 2, 3 e 7.

III.4.6 – ELETROFORESE EM GEL DE AGAROSE

Todos os produtos de PCR com peso molecular acima de 200pb e alguns produtos de digestão foram visualizados em géis de agarose 0,8 a 1,0%. 5 µl das reações de PCR foram

Tabela 7: POLIMORFISMOS DE RFLP ANALISADOS NA IDENTIFICAÇÃO DOS HAPLOGRUPOS CONTINENTE-ESPECÍFICOS DE MTDNA

Haplogrupo	Sítios de restrição característicos	Africanos	Nativo Americanos ^a	Europeus
L1a	+ 3592 <i>Hpa</i> I; + 11641 <i>Hae</i> III	4		
L1b	+ 3592 <i>Hpa</i> I; – 7055 <i>Alu</i> I; + 2349 <i>Mbo</i> I	4		
L1c	+ 3592 <i>Hpa</i> I; + 9070 <i>Taq</i> I, + 12810 <i>Rsa</i> I	4		
L2	+ 3592 <i>Hpa</i> I; + 16389 <i>Hinf</i> I	4		
L3a	– 3592 <i>Hpa</i> I; – 10394 <i>Dde</i> I	4		
L3b	– 3592 <i>Hpa</i> I ; + 10084 <i>Mbo</i> I	4		
L3d	– 3592 <i>Hpa</i> I ; – 8616 <i>Mbo</i> I	4		
L3e	– 3592 <i>Hpa</i> I ; + 2349 <i>Mbo</i> I	4		
A	+ 663 <i>Hae</i> III		4	
B	Deleção de 9-bp		4	
C	– 13259 <i>Hinc</i> II		4	
D	– 5176 <i>Alu</i> I		4	
H	– 7025 <i>Alu</i> I			4
V	– 4577 <i>Nla</i> III			4
HV	– 14766 <i>Mse</i> I			4
U	+ 12308 <i>Hinf</i> I			4
K	– 9052 <i>Hae</i> II			4
J	– 13704 <i>Bst</i> NI			4
T	+ 13366 <i>Bam</i> HI; + 15606 <i>Alu</i> I			4
I	+ 8249 <i>Ava</i> II; +10032 <i>Alu</i> I			4
X	– 1715 <i>Dde</i> I		4	4

^a Os haplogrupos A, B, C e D estão presentes também na Ásia;

Os números representam a posição do nucleotídeo da extremidade 5' do sítio de reconhecimento da endonuclease de acordo a sequência de referência (ANDERSON e cols., 1981; ANDREWS e cols., 1999). (+) e (–) indicam respectivamente a presença e a ausência do sítio em questão.

adicionados a igual volume de tampão da amostra 2X (azul de bromo-fenol 0.1%, xileno-cianol 0.1% em TBE 2X, 176mM Tris-borato, 4mM EDTA, pH8,0) e submetidos a eletroforese em géis de agarose 1% (em TAE 0,5X, 27,5mM Tris-acetato, 0,5mM EDTA, pH8,0) com brometo de etídio ($1,0 \times 10^{-4}$ mg/ml). A corrida foi realizada a 80 V por aproximadamente 50 min. Os fragmentos foram visualizados em um transiluminador de ultravioleta e em seguida, fotografados. Para a resolução dos produtos de digestão, adicionou-se 3 μ l de tampão da amostra 6X a 15 μ l desses produtos, utilizando-se as mesmas condições de corrida. O tempo de corrida foi determinado de acordo com a necessidade de resolução para cada teste de RFLP.

III.4.7 – ELETROFORESE EM GEL DE POLIACRILAMIDA CORADO POR SAIS DE PRATA

Os produtos de amplificação das reações de PCR de tamanho inferior a 200pb foram analisados em géis de poliacrilamida não desnaturante na concentração de 6%, e os produtos de algumas reações de digestão foram analisados em géis com concentrações de 5-12%. A escolha da concentração dos géis foi feita de acordo com a necessidade de resolução necessária para que os polimorfismos pudessem ser visualizados com clareza.

Alíquotas de 3-5 μ L dos produtos de PCR foram adicionadas a igual volume de tampão da amostra 2X e aplicadas nos géis. Os géis foram submetidos a uma voltagem média de 100V por aproximadamente duas horas até o corante xileno-cianol migrar cerca de 4 centímetros da canaleta. O tempo de corrida para resolução dos produtos de digestão também foi ajustado de acordo com a concentração do gel e a necessidade de resolução. Após a eletroforese as bandas foram visualizadas utilizando-se o método de coloração dos géis com nitrato de prata de acordo com Sanguinett e colaboradores (1994). Os géis foram fixados por 10 min à temperatura ambiente em solução de etanol 10% e ácido acético 0,5%, seguindo-se uma incubação de 10 min em solução de nitrato de prata 0,15%, etanol 0,1% e ácido acético 0,05%. A revelação foi feita em uma solução de NaOH 3% (p/v) e formaldeído 1,1% durante o tempo necessário para o aparecimento das bandas. A revelação foi interrompida com a solução fixadora, fotografando-se os géis em seguida. Nas análises de RFLP o volume aplicado nos géis era de 15 μ L do produto final das digestões.

III.4.8 - CLONAGEM E SEQUENCIAMENTO DE PRODUTO DE PCR

Experimentos de clonagem foram realizados utilizando-se o kit SureClone® da Pharmacia Biotech, e bactérias *E. coli* DH5 α competentes de acordo com os procedimentos descritos por Cohen e cols. (1972). Os fragmentos clonados foram sequenciados utilizando os iniciadores M13-40 e M13 reverso fluorescentes (TAB 4), seguindo os mesmos procedimentos descritos no item III.4.2

III.5. – ANÁLISE DE AMOSTRAS ARQUEOLÓGICAS

Essa parte do projeto foi inicialmente desenvolvida em colaboração com o Instituto Max Planck de Antropologia Evolucionária (em suas instalações em Leipzig - Alemanha) sob a orientação da Dra. Anne Stone (da Universidade do Novo México) e dos Drs. Mark Stoneking e Hendrik Poinar

III.5.1 – CUIDADOS ESPECIAIS NO TRABALHO COM DNA ANTIGO

O trabalho com amostras fósseis ou que apresentam altos níveis de degradação de DNA encontra seus maiores obstáculos na obtenção de quantidades de DNA suficientes para serem utilizadas em reações de amplificação por PCR. Além disso, o DNA autêntico das amostras deve ser obtido em total ausência de contaminação com DNA moderno, que por sua vez, é preferencialmente utilizado como molde em qualquer reação de PCR, devido ao seu melhor estado de preservação.

Assim, é extremamente importante que sejam tomadas uma série de medidas e cuidados ao se trabalhar com DNA antigo, com o objetivo de minimizar a possibilidade de contaminação com DNA moderno provenientes principalmente de: (i) contato das amostras em seus sítios arqueológicos com os mais diversos tipos de organismos e microorganismos; (ii) ausência de cuidados (como uso de luvas e máscaras, por exemplo), ao se manipular as amostras no momento de sua retirada dos próprios sítios arqueológicos. Sobretudo, devem ser tomados cuidados especiais nas etapas de extração de DNA e amplificação por PCR, para se evitar a contaminação das amostras fósseis com o DNA do próprio pesquisador.

Condições ótimas para a extração e caracterização de DNA mitocondrial de amostras fósseis, de acordo com Pääbo (1993) deveriam levar em conta as seguintes medidas:

- ⇒ Deve ser feita a separação física completa dos laboratórios onde são desenvolvidas análises com DNA moderno e DNA antigo;
- ⇒ Os reagentes (soluções de extração de DNA e soluções de PCR), aparelhos e equipamentos (centrífugas, máquinas de PCR, pipetas, tubos eppendorf, etc.), e vidraria utilizados nos procedimentos de extração de DNA devem ser totalmente separados daqueles destinados ao trabalho com DNA contemporâneo;
- ⇒ Toda vidraria deve ser previamente tratada com soluções de NaOH 0,5N e HCl 0,25N;
- ⇒ Deve ser feita uma limpeza periódica com solução de hipoclorito de sódio a 15% em todas as bancadas e equipamentos;
- ⇒ Equipamentos descartáveis de proteção como máscaras, luvas (no mínimo duas), guardapós de manga comprida e toucas de cabelo, devem ser utilizados durante todos os

procedimentos que envolvam o contato direto ou indireto com as amostras, principalmente durante as etapas de extração de DNA e preparo das misturas de PCR;

- ⇒ O laboratório destinado a manipulação das amostras e extração de DNA deve ser irradiado com luz ultravioleta à noite e no mínimo durante 24 horas antes de qualquer procedimento;

III.5.2 – PREPARO DAS AMOSTRAS

Os laboratórios e as condições de trabalho no Instituto Max Planck em Leipzig preenchem perfeitamente os requisitos listados acima. Após terem sido tomados todos estes cuidados, realizamos a primeira etapa do nosso trabalho, que consistiu na limpeza cuidadosa das amostras ósseas.

Cerca de 1mm de superfície foi retirada em toda a camada externa dos ossos utilizando-se um "dril". Todas as amostras apresentavam um revestimento "brilhante" de PVA (polímero de poli-vinil-acetato), que foi utilizado na curadoria das amostras para impedir seu desgaste pelo manuseio excessivo, sendo este completamente removido durante a primeira etapa de limpeza. De todas as amostras foi posteriormente retirado um fragmento de aproximadamente 2 centímetros para processamento e utilização nas análises de racemização de aminoácidos e extração de DNA.

A superfície interna e externa do fragmento ósseo foi novamente removida, utilizando-se um dispositivo mais fino do "drill". A amostra foi deixada sob irradiação de luz UV (nas duas superfícies), por cerca de 15-20 minutos, e então pulverizada com o auxílio de um cadinho e pistilo. Para aquelas amostras que apresentaram maior resistência ao desgaste físico, o uso de N₂ se fez necessário. Uma quantidade de N₂ suficiente para completar o volume do cadinho foi adicionado e após sua evaporação a amostra foi novamente macerada até a pulverização. As amostras pulverizadas foram aliquotadas em tubos eppendorf de 1,5 ml e armazenadas à temperatura de 4°C.

III.5.3 –ANÁLISES DE RACEMIZAÇÃO DE AMINOÁCIDOS.

A primeira etapa de uma análise de racemização de aminoácidos consiste na digestão completa das proteínas da amostra com HCl 6 N. Para isso adicionou-se a 0,010 gramas de pó de osso, 1 ml de HCl 6 N (Sigma) incubando-se à temperatura de 100°C, por 24 horas. Essa etapa é realizada em pequenos frascos de vidro previamente esterilizados a 200°C por 7 dias. A seguir essa mistura de HCl foi evaporada (removida) sob pressão de vácuo em um dessecador contendo NaOH por aproximadamente 18 a 24 horas até a evaporação completa do HCl. Adicionou-se então 1 ml de tampão borato de sódio 0,4M, pH 9.4, para neutralizar os vestígios de HCl.

A próxima etapa consistiu na derivatização das amostras com uma solução de OPA-NAC

(orto-phityldialdeído e N-acetyl-L-cisteína) que reage com as formas D e L de todos os aminoácidos livres em solução resultando em produtos fluorescentes capazes de serem separados e identificados por cromatografia líquida de alta performance (HPLC), em uma coluna de fase reversa C18 (ZHAO e BADA, 1995).

No processo de derivatização reagiu-se 10 µl da solução digerida da amostra com 5 µl de solução OPA-NAC (OPA 6 mM; NAC 12 mM; metanol 30%; tampão borato de sódio 0,1M pH 9.4) e 10 µl de tampão borato de sódio (0,4M, pH 9.4) em um tubo eppendorf de 1,5 ml, por 1 minuto à temperatura ambiente. A reação foi paralisada ao final de 60 segundos adicionando-se 475 µl de uma solução 50mM de acetato de sódio, pH 5.5.

Os tubos foram centrifugados a seguir e 20 µl deste produto de reação foram aplicados no HPLC (modelo CROMTOPAC-CR4A, da Shimadzu). Uma coluna de fase reversa C18 Econosphere (250 X 4.6 mm ID, Beckman) foi utilizada na separação dos compostos da amostra e uma corrida de 50 minutos, com o fluxo ajustado para 1ml/min, foi realizada sob o seguinte gradiente de concentração dos tampões A (metanol para análises em HPLC) e B (tampão acetato de sódio 50mM, pH5-6, na diluição de 92:8 com metanol):

1. A coluna é equilibrada com 100% de solução B por 4 minutos;
2. Aplicação da mostra no tempo de 4 minutos;
3. Aumento gradual da concentração do tampão A, tal que aos 14 minutos a concentração de B atinja 58%;
4. Concentração do tampão B é de 58% aos 24 minutos;
5. Concentração do tampão B é de 40% aos 29 minutos;
6. Concentração do tampão B se mantém em 40% até o tempo de 35 minutos;
7. Concentração do tampão B é aumentada até chegar a 100% ao final da corrida (50 minutos);

A medição da fluorescência do eluente da coluna foi feita por um detector de fluorescência da marca Shimadzu RF-530, ajustando-se os parâmetros de comprimento de onda de excitação para 340 nm e comprimento de onda de emissão para 450 nm.

III.5.4 – EXTRAÇÃO DE DNA DE AMOSTRAS FÓSSEIS

Várias metodologias de extração de DNA de fósseis têm sido desenvolvidas e aprimoradas nos últimos anos, inclusive em nosso laboratório (PRADO e cols., 1997), e demonstram que os principais problemas relacionados com a extração de DNA dessas amostras estão relacionados à presença de compostos que inibem as etapas subsequentes de análise como, por exemplo, amplificação por PCR. Assim, nós optamos por seguir os mesmos protocolos de extração de DNA

utilizados no Instituto Max Planck, que consistem de modificações no método que utiliza sílica para imobilizar e recuperar o DNA (BOOM e cols., 1990; HÖSS E PÄÄBO, 1993).

Essa metodologia consiste em um protocolo rápido e prático de extração de DNA que dispensa as etapas de digestão da amostra com proteinase K e extração fenólica. Os procedimentos foram iniciados incubando-se 250 mg de pó de osso em um tubo eppendorf de 1,5 ml contendo 820 µl de tampão L6 (para concentração e o modo de preparo das soluções utilizados na extração de DNA arqueológico ver item III.5.8) por 2 horas a 60°C, vortexando-se a mistura ocasionalmente. A seguir, centrifugou-se o eppendorf (à velocidade de 8.000 rpm) e transferiu-se 500µl da solução sobrenadante para novo tubo eppendorf de 1,5 ml contendo uma mistura de 500 µl de tampão L6 e 40 µl de ressuspensão de sílica. Após ter sido vortexada, a mistura foi incubada por 15 minutos à temperatura ambiente para permitir a ligação do DNA às partículas de sílica. Ao final desse tempo, centrifugou-se novamente o tubo eppendorf, descartou-se o sobrenadante e procederam-se cinco etapas de lavagem do complexo sílica-DNA com tampão L2 (2 vezes), solução de etanol 70% (2 vezes), e acetona (1 vez). As etapas de lavagem do *pellet* de sílica consistem em adicionar-se o solvente, agitar vigorosamente (com o uso de vórtex), centrifugar por 2 minutos a uma velocidade entre 8.000 e 10.000 rpm e descartar o sobrenadante. Após a ultima etapa de lavagem o pellet é secado em banho de areia a 95°C. A etapa final consiste em se eluir o DNA da sílica, acrescentando-se 65 µl de ddH₂O, incubando-se a 56°C por 10 minutos e tomando-se o sobrenadante após centrifugação a 10.000 rpm por 2 minutos. Esse processo é repetido, o que resulta em aproximadamente 120-130 µl de extrato de DNA. Já foi demonstrado que a sílica constitui um forte inibidor da enzima Taq DNA polimerase e portanto foi tomado um enorme cuidado no momento da recuperação da fase sobrenadante, a fim de que não fossem pipetadas grandes quantidades de sílica. 5 µl do produto de extração de DNA são utilizados em cada reação de PCR. No mínimo duas extrações de DNA devem ser realizadas para cada amostra, em dias diferentes, e se possível, por pessoas diferentes em diferentes laboratórios. A amostra era sempre centrifugada por 2 minutos a 10.000 rpm antes de ser pipetada para as reações de PCR.

Dois “controles de extração” foram também realizados para cada amostra, que consistiam em tubos “brancos”, nos quais todos os procedimentos efetuados com as amostras ósseas, durante o processo de extração de DNA são também realizados, porém sem a adição de amostra. Esses controles ou “brancos de extração” são também utilizados nas reações de PCR para se verificar a ocorrência de contaminação com DNA contemporâneo no momento da extração de DNA. A presença (confirmada por repetição da reação de PCR) de bandas nos produtos de PCR de um branco de extração, implicaria no descarte de todas os extratos de DNA que foram processados no mesmo experimento.

III.5.5. – AMPLIFICAÇÃO POR PCR DA REGIÃO CONTROLE DOS EXTRATOS DE DNA DE AMOSTRAS ARQUEOLÓGICAS

Para amplificar a região hipervariável I do mtDNA nas amostras fósseis, foram utilizados 8 pares de iniciadores, cujas sequências estão apresentadas na tabela 5. Esses sistemas permitiram a amplificação de parte da Região Controle (entre os nucleotídeos 16055 e 16410) em fragmentos parcialmente sobrepostos, que estão ilustrados na figura 4. Essa estratégia foi utilizada porque amostras fósseis geralmente apresentam elevados níveis de degradação de DNA impedindo que fragmentos maiores que 200pb sejam, na maioria das vezes, amplificados por PCR. As amplificações foram realizadas em 50µL de volume final contendo 0.5 pmol/µL (0.5µM) dos iniciadores, 200µM de cada dNTP e 0,5 unidade de AmpliTaqGolg (Perkin-Elmer, Cetus) em tampão de reação Tris-HCl 67 mM pH 8,8, MgCl₂ 2 mM, na presença de BSA a 2 mg/ml. Em cada tubo de PCR foram adicionados 5 µl de extrato de DNA. O programa de amplificação executava uma etapa inicial de desnaturação de 5 min a 94°C, seguida de 39 ou 44 ciclos de desnaturação a 94°C por 45 seg, anelamento a 55°C por 45 seg e extensão a 72°C por 1 min. Uma etapa final consistindo de uma extensão a 72°C por 7 minutos foi também realizada. Em cada grupo de reações de PCR foram feitos pelo menos 6 controles negativos de mistura de reação (sem a adição de DNA) para verificar a contaminação dos reagentes. Além disso os dois controles de extração (descritos no item anterior) de cada amostra eram também amplificados. Utilizou-se termocicladores PT 200 da MJ Research para as reações de PCR. Para os experimentos envolvendo DNA antigo, todas as reações de PCR foram visualizadas em géis de agarose NuSieve 2,6% (em TBE 0,5X, 45mM Tris-borato, 0,5mM EDTA, pH8,0), corados com brometo de etídeo (0,10µg/ml).

Todos os produtos de PCR resultantes de amplificações dos extratos dos ossos, obtidos em experimentos onde observou-se também a total ausência de bandas nos controles de amplificação e brancos de extração foram recuperados dos géis de agarose e utilizados em experimentos de clonagem.

III.5.6 –REAMPLIFICAÇÃO DE PRODUTOS DE PCR

As bandas selecionadas para análise foram cortadas dos géis utilizando-se lâminas de bisturi estéreis, e a seguir foram diluídas em 50 µl de tampão low TE. As regiões correspondentes nas canaletas dos controles de extração também foram recuperadas e tiveram o mesmo tratamento das bandas das amostras. Como as amplificações dessa primeira reação de PCR se apresentaram sempre muito fracas, foram utilizados 5 µl desses produtos em uma segunda reação para reamplificar o fragmento, utilizando-se o mesmo par de iniciadores. O objetivo era se obter quantidades de DNA

suficientes para que fosse realizada a clonagem, e as mesmas condições de PCR foram utilizadas, embora nesta etapa o programa do termociclador era ajustado para que fossem realizados apenas 25 ciclos, com temperatura de anelamento dos iniciadores acrescida de 1°C. A mistura de reação era preparada no laboratório destinado ao trabalho com amostras fósseis, aliquotada e levada ao laboratório de DNA moderno, para que fosse adicionado o DNA amplificado. Os tubinhos contendo os fragmentos de gel com a banda a ser reamplificada em solução de low TE eram brevemente aquecidos a 40°C para derreter a agarose. Os resultados desse PCR também foram visualizados em géis de agarose NuSieve 2,6% (em TBE 0,5X, 45mM Tris-borato, 0,5mM EDTA, pH8,0), corados com brometo de etídeo (0,10µg/ml).

III.5.7 – CLONAGEM E SEQUENCIAMENTO DOS PRODUTOS DE PCR

Os produtos de PCR que mais uma vez se apresentaram positivos para a reamplificação, na total ausência de bandas nos controles negativos (branco da mistura de reação e branco do controle de extração), foram clonados e transformados em células bacterianas competentes utilizando-se o kit de clonagem TOPO™ TA Cloniong® (Invitrogen). A seleção das colônias (brancas) positivas para o inserto foi possível pelo método da α -complementação com o gene da β -galactosidase presente no vetor pCR®-TOPO do kit. A presença do inserto foi confirmada por PCR de colônia utilizando os iniciadores M13-direto e M13-reverso (TAB 5) também fornecidos no kit de clonagem. As amplificações do PCR de colônia foram realizadas em 40µL de volume final contendo 0.25 pmol/µL (0.25µM) dos iniciadores, 200µM de cada dNTP e 0,2 unidade de *Taq* DNA polimerase (Promega) em tampão de reação Tris-HCl 10 mM pH 8,5, KCl 50 mM, MgCl₂ 1,5 mM, na presença de BSA a 1mg/ml. As colônias brancas eram levemente pinçadas da placa com o auxílio de uma ponteira fina e adicionadas à mistura de reação de PCR, encostando-se a ponteira no fundo do tubo eppendorf contendo a mistura. O programa de amplificação executava uma etapa inicial de desnaturação de 5 min a 94°C, seguida de 30 ciclos de desnaturação a 94°C por 30 segundos, anelamento a 55°C por 1 minuto e extensão a 72°C por 1 minuto. Em cada grupo de reações de PCR foram feitos controles negativos (sem a adição de DNA) para verificar a contaminação dos reagentes. Uma colônia azul era também utilizada como controle de amplificação na ausência de inserto. As reações de PCR foram realizadas em termocicladores PT 200 da MJ Research. Os produtos de PCR foram visualizados em géis de agarose NuSieve 2,6% (em TBE 0,5X, 45mM Tris-borato, 0,5mM EDTA, pH8,0), corados com brometo de etídeo (0,10µg/ml) e purificados com o kit Wizard™ PCR Preps (Promega) para serem posteriormente submetidos à reações de sequenciamento. Cerca de 8-12 µl de cada produto do PCR de colônia purificado foram utilizados nessas reações, com um dos iniciadores de DNA mitocondrial utilizados na primeira amplificação, e o kit ABI PRISM® BigDye™ Terminator Cycle Sequencing

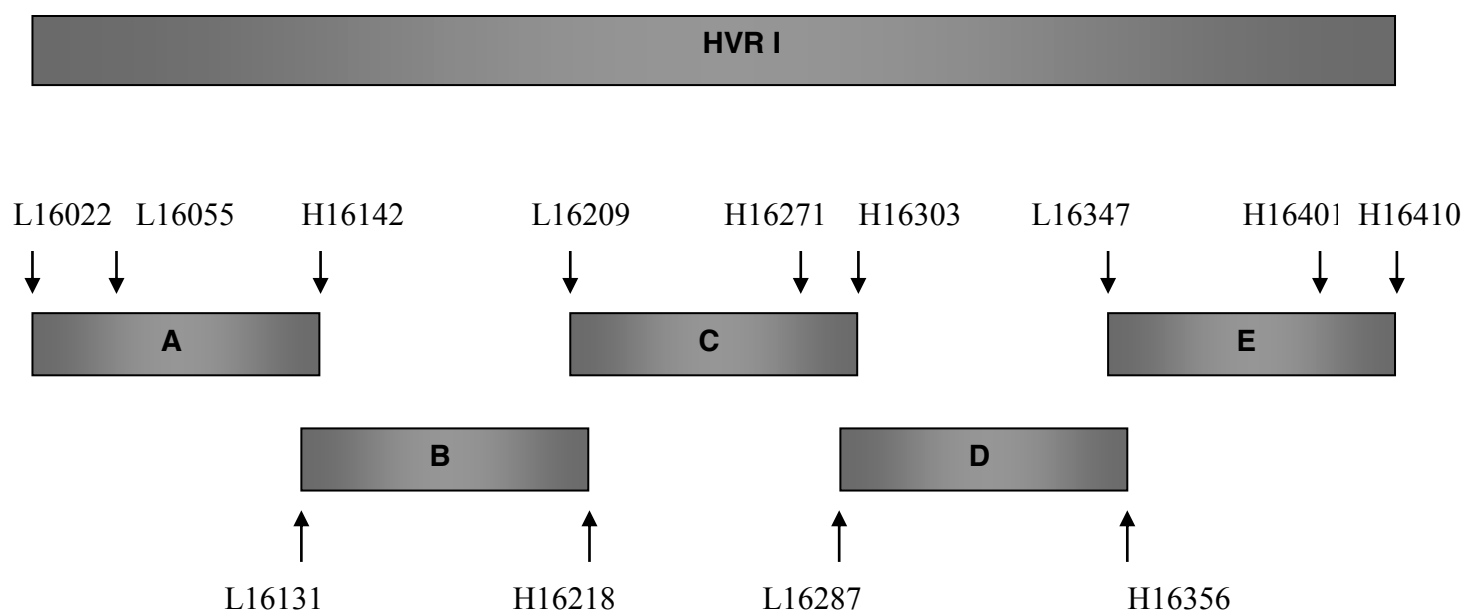


FIGURA 4: ESQUEMA DA AMPLIFICAÇÃO POR PCR DE FRAGMENTOS SOBREPOSTOS DA HVR-1. Os números indicam as posições na região controle dos nucleotídeos 3' dos iniciadores. A sequência de cada iniciador está representada na tabela 4. Para os sistemas A, C e E, as amplificações eram realizadas primeiramente com os iniciadores mais externos, mas a obtenção de resultados negativos para esses sistemas, implicava na tentativa de amplificação com os respectivos iniciadores internos.

(Perkin-Elmer Applied Biosystems). Os produtos da reação de sequenciamento foram precipitados com isopropanol 80%, ressuspensos em tampão de corrida contendo formamida e EDTA, e submetidos à eletroforese em gel de poliacrilamida a 6% contendo 8M de uréia utilizando o sequenciador automático ABI PRISM 377 (Perkin-Elmer Applied Biosystems). Os resultados obtidos em forma de um cromatograma, foram processados utilizando-se programas do próprio sequenciador automático, para se obter a leitura da sequência de nucleotídeos. Em média 10 a 12 clones foram sequenciados para cada produto de PCR clonado originalmente.

III.5.8 – SOLUÇÕES E TAMPÕES UTILIZADOS NOS EXPERIMENTOS DE EXTRAÇÃO DE DNA DE AMOSTRAS ARQUEOLÓGICAS.

⇒ Preparo da solução de sílica:

Dissolveu-se 12 g de dióxido de sílica (Sigma) em 100 ml de ddH₂O estéril em uma proveta coberta com folhas de alumínio. Essa mistura foi agitada vigorosamente com o uso de vórtex e deixada em repouso por 24 horas à temperatura ambiente protegida de luminosidade. A seguir, 86 ml do sobrenadante foram cuidadosamente descartados (onde estão as partículas menores de sílica) e completou-se novamente para o volume de 100 ml com ddH₂O estéril. Agitou-se vigorosamente para ressuspender o pellet de sílica e essa mistura foi deixada em repouso no escuro, e à temperatura ambiente por mais 5 horas. Finalmente, 88 ml do sobrenadante foram descartados e adicionou-se 120 µl de HCl (32% vol/vol) para se ajustar o pH a 2. A solução de ressuspensão de sílica foi alíquotada em tubos de 1,5 ml e guardada a 4°C protegida de luz. Nessas condições, essa solução se mantém estável por até 6 meses.

⇒ Tampões de extração L2 e L6.

O tampão L6 constitui uma solução de Isotiocianato de Guanidina (Sigma) 5 M, tris-HCl 0,1M (pH 7.4), EDTA 0,02 M, Triton X-100 1,3%; e representa o tampão de lise celular de vários protocolos de extração de DNA que utilizam sílica. O tampão L2 apresenta a mesma composição de Isotiocianato de Guanidina (5M) em tris-HCl 0,1 M, na ausência de EDTA e do detergente. Ao final do preparo destas duas soluções é adicionado cerca 1,5 ml de solução de sílica (em um volume final de 25 ml) para que seja capturado qualquer partícula de DNA que possa estar presente na solução. As soluções são alíquotadas em tubos Falcon de 12 ml na ausência de luz e nessas condições se mantêm estáveis por até 3 semanas. Antes de serem utilizados na extração de DNA os tubos foram centrifugados por 3 minutos a 12.000 g para garantir a precipitação da sílica.

Resultados

IV.1 – CARACTERIZAÇÃO DA LINHAGEM CONTENDO O POLIMORFISMO DE INSERÇÃO DE 9PB DO MTDNA

IV.1.1 – CARACTERIZAÇÃO DO POLIMORFISMO DE INSERÇÃO DA UNIDADE REPETITIVA DE 9-BP

Na região V do DNA mitocondrial, entre os genes da enzima citocromo oxidase II e o RNA transportador de lisina, (COII/tRNA^{Lys}) estão presentes duas repetições em tandem de um segmento de 9 pares de base (9pb), que podem apresentar-se polimórficas em alguns indivíduos, principalmente em populações do sudeste asiático, ameríndios e africanos. O polimorfismo encontrado mais frequentemente é denominado de *deleção de 9pb*, e como o próprio nome indica, é representado pela presença de apenas uma sequência de 9pb.

Ao analisarmos o polimorfismo de deleção de 9pb em uma amostra de 245 indivíduos brasileiros foi identificado um polimorfismo diferente, definido pela expansão da região repetitiva (FIG 5). Experimentos de clonagem e sequenciamento da região intergênica polimórfica foram realizados, e em quatro clones independentes foi demonstrada a presença de uma repetição adicional de 9pb (ALVES-SILVA e cols., 1999a; Apêndice 2). A sequência de nucleotídeos entre as posições 8262 e 8304 do mtDNA determinada para essa linhagem, definida aqui como BR-3R, é apresentada na figura 6, onde se observa a presença das três repetições em tandem do segmento de 9pb.

IV.1.2 – SEQUENCIAMENTO DA REGIÃO CONTROLE DO MTDNA E ANÁLISES DE RFLP PARA A AMOSTRA BR-3R

A linhagem mitocondrial BR-3R foi analisada através de sequenciamento das regiões hipervariáveis I e II da região controle, para se estabelecer sua classificação de acordo com os haplogrupos continente-específicos de mtDNA, característicos de populações nativas americanas, asiáticas, européias e africanas. Determinou-se a sequência de nucleotídeos de um segmento de 302 pb entre as posições 16061 e 16362 (HVS-I); e de um segmento de 266 pb entre as posições 72 e 337 (HVS-II). As sequências obtidas foram alinhadas e comparadas com a sequência de referência de Anderson (ANDERSON e cols., 1981; ANDREWS e cols., 1999) mostrando que essa linhagem é definida por transições nas posições 16192, 16274, 16362 (HVS-I), além de 239 e 263 (HVS2) (tabela 8). A princípio estas mutações não constituem marcadores específicos para nenhum haplogrupo de DNA mitocondrial, embora a ausência de uma transição de A para G (A→G) na posição 73 do mtDNA seja característica de linhagens dos haplogrupos europeus H e V. A fim de classificar precisamente o haplótipo BR-3R foi realizada uma extensa análise de RFLPs, onde foram testados todos os polimorfismos de sítio de restrição característicos dos principais haplogrupos de

mtDNA até então descritos em populações ameríndias (haplogrupos A, B, C e D) africanas (haplogrupos L1, L2 e L3) e

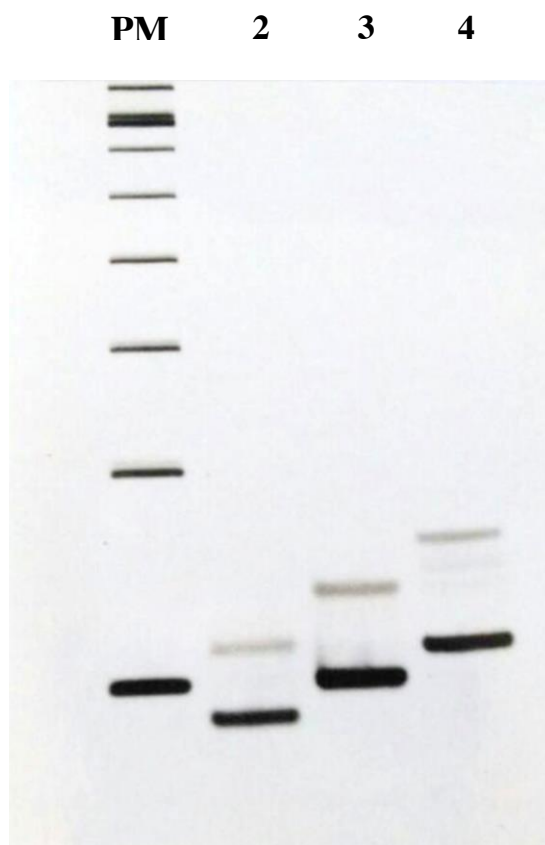


FIGURA 5: AMPLIFICAÇÃO POR PCR DA REGIÃO INTERGÊNICA COII/TRNA^{LYS} DO MTDNA HUMANO.

Utilizando os iniciadores L8209 e H8304 amplifica-se um fragmento de 95pb para mtDNAs com 2 repetições de 9pb (canaleta 3). A banda correspondente às linhagens com a deleção de 9pb tem 86pb (canaleta 2) e a produto de PCR resultante da linhagem BR-3R apresentou uma banda de 104pb (canaleta 4). PM representa o padrão de peso molecular 100pb Ladder (Life Technologies, Gaithersburg, MD, EUA).

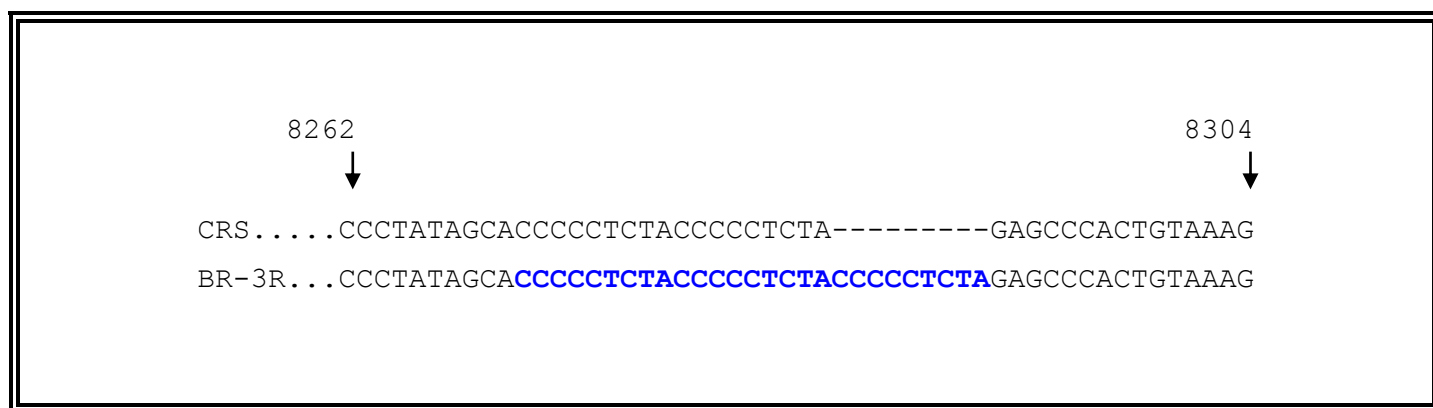


FIGURA 6: SEQUÊNCIA DA REGIÃO INTERGÊNICA COII/TRNA^{Lys} DO INDIVÍDUO BR-3R. O produto de PCR da região V do mtDNA do indivíduo BR-3R foi clonado e 4 clones independentes foram sequenciados. Três repetições perfeitas do segmento de 9pb (mostrado em azul) foram observadas em todos os clones. CRS representa a região correspondente na sequência de referência de Anderson (1981).

Tabela 8: SÍTIOS POLIMÓRFICOS EM RELAÇÃO À CRS OBSERVADOS NA SEQUÊNCIA DE HVS-I E HVS-II DO MTDNA DA AMOSTRA BR-3R.

	HVS-I	HVS-II
	1 1 1	
	6 6 6	
Posição	1 2 3	1 2 2
no mtDNA	9 7 6	5 3 6
↓	2 4 2	2 9 3
CRS	C G T	T T A
BR-3R	T A C	. C G
PT01	T A C	. C G
PT02	T A C	C C G

As posições dos nucleotídeos na molécula de mtDNA estão representadas na parte superior da figura de acordo com a numeração de ANDERSON e colaboradores (1981). A sequência de Anderson (CRS de *Cambridge Reference Sequence*) é mostrada na primeira linha e apenas os nucleotídeos variáveis estão indicados para as demais sequências. BR-3R é o haplotipo brasileiro; PT01 e PT02 correspondem aos haplotipos portugueses.

HVS-I: segmento hipervariável I

HVS-II: segmento hipervariável II

européias (haplogrupos H, I, J e K) (TAB 9). Nessas análises a linhagem BR-3R apresentou os polimorfismos de perda do sítio para a enzima *Ddel* na posição 10394 (10394 – *Ddel*), e perda do sítio para a enzima *AluI* na posição 7025 (7025 – *AluI*), que permitiu a classificação desse haplótipo como pertencente ao haplogrupo H europeu (TAB 9).

IV.1.3 – ANÁLISE DOS POLIMORFISMOS DE TAMANHO NA REGIÃO INTERGÊNICA COII/tRNA^{Lys} EM PORTUGAL.

Eventos de inserção como a *triplicação de 9pb* na região repetitiva COII/tRNA^{Lys}, constituem polimorfismos raros e haviam sido descritos na literatura, em linhagens de mtDNA asiáticas (WRISCHNIK e cols., 1987; BALLINGER e cols., 1992; SHIELDS e cols., 1992; PASSARINO e cols., 1993; MERRIWETHER e cols., 1995), e africanas (MERRIWETHER e cols., 1995). A associação dessa mutação a uma linhagem mitocondrial europeia foi descrita apenas recentemente, por Thomas e colaboradores (1998), e na maioria desses mtDNAs o polimorfismo de inserção de 9pb estava associado a linhagens de diferentes haplogrupos, indicando que se tratava portanto, de eventos mutacionais independentes (BALLINGER e cols., 1992; PASSARINO e cols., 1993; MERRIWETHER e cols., 1995). Como é provável que a grande maioria das linhagens mitocondriais européias no Brasil constituam uma herança materna por parte dos colonizadores Portugueses, nós decidimos tentar identificar em Portugal, a origem geográfica da linhagem brasileira BR-3R.

Utilizando os iniciadores MiL8209 e MiH8304 é possível analisar, de acordo com o tamanho do fragmento de PCR, tanto o polimorfismo de deleção de 9pb quanto o polimorfismo de inserção da terceira unidade repetitiva. Dessa forma, 96 amostras de DNA de indivíduos portugueses da região da cidade do Porto foram analisadas por PCR quanto à presença dessas mutações de tamanho na região intergênica COII/tRNA^{Lys}. Nenhuma dessas amostras, apresentou a deleção de 9pb, enquanto o polimorfismo de triplicação foi encontrado em duas linhagens (2% dos indivíduos). Estas amostras, identificadas como PT01 e PT02 também foram submetidas ao sequenciamento das regiões hipervariáveis I e II da região controle e análises de RFLP, para se confirmar a presença de marcadores europeus.

O resultado do sequenciamento da região controle se mostrou bastante interessante, uma vez que as amostras PT01 e PT02 foram ambas caracterizadas pelas mesmas transições que definiram o haplótipo BR-3R brasileiro, ou seja, transições nas posições 16192 – 16274 – 16362 (HVS I); e 239 – 263 (HVS II). No entanto, a amostra PT02 apresentou ainda uma transição T→C na posição 152 do segmento hipervariável II (TAB 9).

Uma vez que as linhagens européias portuguesas apresentaram mutações de região controle idênticas à linhagem BR-3R, e esta já havia sido classificada como haplótipo europeu **H**, apenas os

polimorfismos de RFLP característicos deste haplogrupo foram testados para PT01 e PT02.

Novamente os resultados de RFLP se apresentaram idênticos àqueles observados para a linhagem BR-3R, ou seja, as duas linhagens portuguesas apresentaram os marcadores 10394 – *Ddel* e 7025 – *AluI*, confirmando a classificação de mtDNAs do tipo "H" (TAB 9).

Assim, analisando um polimorfismo raro para se encontrar linhagens filogeneticamente relacionadas, nós conseguimos localizar em Portugal a origem de uma linhagem mitocondrial brasileira. Demonstramos também a importância das análises de sequenciamento e RFLP para verificar a relação filogenética dos haplótipos. A partir desses resultados, nós sugerimos que o polimorfismo de triplicação de 9pb observado nos três mtDNAs é o resultado de um evento único de inserção que ocorreu no passado em uma linhagem de mtDNA do haplogrupo H.

IV.2 – ESTUDO DA VARIABILIDADE DO MTDNA EM AMOSTRAS DA POPULAÇÃO BRANCA DAS REGIÕES NORTE, NORDESTE E SUL DO BRASIL

IV.2.1 – SEQUENCIAMENTO DA REGIÃO CONTROLE DO MTDNA EM 148 AMOSTRAS DE INDIVÍDUOS BRASILEIROS

A determinação da origem geográfica da linhagem BR-3R no norte de Portugal, constitui apenas um exemplo de como a análise do DNA mitocondrial de uma população pode identificar pelo menos em parte, a contribuição genética materna de seus indivíduos. O mtDNA já foi utilizado por exemplo, para avaliar a participação de matrilineagens ameríndias em uma amostra da população uruguaia (BRAVI e cols., 1997) e da região Amazônica (BATISTA-DOS-SANTOS e cols., 1999). O estudo da origem geográfica dos haplótipos de mtDNA encontrados em 100 indivíduos da região Sudeste do Brasil, realizados por Santos (1998), demonstrou que aquela amostra apresentava uma alta variabilidade de linhagens, sendo formada por: 33% de DNAs mitocondriais de origem ameríndia, 34% de origem africana, e apenas 31% de linhagens européias. Para verificar se esse predomínio de linhagens ameríndias e africanas em indivíduos brancos brasileiros poderia ser observado também nas outras regiões do país, nós decidimos analisar os mesmos parâmetros estudados por Santos (1998), em indivíduos das regiões Norte (48 indivíduos), Nordeste (50 indivíduos) e Sul (50 indivíduos) do país.

Determinou-se para tanto a sequência de nucleotídeos do mesmo segmento de 302 pb da região hipervariável I do mtDNA (entre as posições 16061 e 16362) para o mtDNA destes 148 brasileiros. As sequências apresentadas na tabela 10 correspondem às sequências consenso que foram obtidas através do sequenciamento direto das duas fitas do mtDNA, utilizando os iniciadores fluorescentes MiL15996 e MiH16401. A análise de todos os 148 mtDNAs permitiu identificar 107

Tabela 9: ANÁLISE DOS PRINCIPAIS POLIMORFISMOS DE SÍTIO DE RESTRIÇÃO ASSOCIADOS ÀS POPULAÇÕES AFRICANAS (AFR), ASIÁTICAS (AS), NATIVO-AMERICANAS (NA) E EUROPEIAS (EUR) NAS AMOSTRAS BR-3R, PT01 E PT02

Haplogrupo	Polimorfismos	Amostras		
		BR-3R	PT01	PT02
A (As e NA)	<i>Hae</i> III (+) 663	0		
B (As e NA)	<i>Hae</i> III (+) 16517	0		
C (As e NA)	<i>Hinc</i> II (-) 13259	0		
D (As e NA)	<i>Alu</i> I (-) 5176	0		
F (As)	<i>Hpa</i> I (-) 12406	0		
M (As)	<i>Alu</i> I (+) 10397	0		
	<i>Dde</i> I (+) 10394	0		
L (Afr)	<i>Hpa</i> I (+) 3592	0		
L1 (Afr)	<i>Hinf</i> I (+) 10806	0		
L2 (Afr)	<i>Hinf</i> I (+) 16389	0		
H (Eur)	<i>Alu</i> I (-) 7025	1	1	1
	<i>Dde</i> I (-) 10394	1	1	1
I (Eur)	<i>Dde</i> I (-) 1715	0		
	<i>Alu</i> I (+) 10028	0		
	<i>Hae</i> II (-) 4529	0		
	<i>Ava</i> II (+) 8249	0		
	<i>Bam</i> HI (+) 16389	0		
J (Eur)	<i>Bst</i> NI (-) 13704	0		
	<i>Hinf</i> I (-) 16065	0		
K (Eur)	<i>Hae</i> II (-) 9052	0		

Os polimorfismos estão representados pela presença (+) ou ausência (-) dos sítios para as endonucleases de restrição. Os números definem a extremidade 3' do sítio de reconhecimento dessas enzimas. "1" indica a presença do alelo e "0" indica a ausência.

[illegible]

Distribuição	SE	NE	N	CRS	Haplogrupo
Regional				CCTTTCCCTTGACTGCCGTTAACTCCTCCCGCTCTGCTACCGCTCACCACCTCCCGCACCTCCCGCACCTTTATAAGCTCCCGCTTCT	
				Africanas	
	2			BR55	L1a
	1			BR56	L1a
	1			BR57	L1a
	1			BR59	L1b
	1			BR60	L1b
	1			BR62	L1c
	1			BR63	L1c
	1			BR65	L1c
	1			BR68	L1c
	1			BR72	L1c
	1			BR74	L2
	1			BR75	L2
	1			BR76	L2
	1			BR77	L2
	1			BR82	L2
	1			BR83	L2
	1			BR86	L3d
	1			BR87	L3d
	1			BR88	L3d
	1			BR89	L3d
	1			BR90	L3e
	3			BR91	L3e
	1			BR92	L3e
	1			BR94	L3e
	1			BR99	L3e
	1			BR100	L3e
	1			BR101	L3e
	1			BR104	L3e
	1			BR109	L3*
	1			BR111	U6

Distribuição Regional		S NE N CRS		CCTTTCCCTTGACTGCCGTTAACTCCTCCCTCTGCTCACCTCCACCTCCCTCCACCTCCCTCCACCTCCCTCCCTTCT		Haplogrupo	
Europeias							
7	2	4	BR112				H
1			BR113				H
1			BR114				H
	1		BR116				H
2			BR117				H
1			BR118				H
1			BR119				H
1			BR121				H
1			BR122				H
1			BR123				H
1			BR125				H
1			BR126				H
1			BR127				H
1			BR128				H
1			BR129				H
1		1	BR131				pre*V
1			BR132				pre*V
1		1	BR135				V
1			BR136				V
1			BR138				HV*
1		1	BR139				U2
1			BR140				U4
1			BR142				U5b*
1			BR143				U5b1
1			BR144				U7
1		1	BR146				K
1		1	BR147				K
1			BR149				K
1			BR150				K
1			BR151				pre*HV
2	1	2	BR152				J*
1			BR153				J*
1		1	BR154				J*
1			BR155				J1*
2			BR156				J1b1
1		1	BR157				T*
1		1	BR158				T*
1			BR159				T*
1			BR162				T*
1		1	BR163				T*
1		1	BR164				T*
1		1	BR166				T1
1		1	BR167				I
1			BR168				X
50	50	48	Total				

linhagens, ou sequências diferentes, definidas por 89 sítios polimórficos quando comparadas à sequência de referência de Anderson (ANDERSON e cols., 1981; ANDREWS e cols., 1999).

Várias sequências foram identificadas em mais de um indivíduo e também em indivíduos de regiões diferentes. Por exemplo, os haplótipos BR112, BR17, BR152 e BR1 que foram identificados nas três regiões analisadas, estavam presentes em 13, 7, 5 e 3 indivíduos respectivamente (TAB 10).

A maioria das mutações observadas corresponderam a transições. A proporção das substituições entre pirimidinas é maior que entre purinas e reflete a composição de pirimidinas/purinas na região estudada. Esse tipo de desvio é comumente reportado na literatura (COMAS e cols., 1996; FRANCALACCI e cols., 1996; BORTOLINI e cols., 1997a; KOLMAN e cols., 1996), e de acordo com WAKELEY (1993), essa tendência é devido à natureza química dos nucleotídeos.

A transição de T para C na posição 16189, presente em 35 das 107 linhagens observadas cria um homopolímero de 10 citosinas entre as posições 16184-16193, também denominado domínio hipervariável de Cs (HORAI e cols., 1993), na ausência de outras mutações nos nucleotídeos adjacentes (FIG 7). A mutação em 16189 torna essa região instável, sendo alvo de eventos de inserções de citosinas, que dão origem a mtDNAs com tamanhos diferentes (heteroplasmia). A heteroplasmia prejudica o sequenciamento direto dos produtos de PCR e dessa forma não foi possível identificar o número exato de Cs desse domínio, para alguns mtDNAs. Portanto, polimorfismos de inserção na região entre os nucleotídeos 16184-16193 não foram considerados em nenhuma de nossas análises. Essa dificuldade tem sido frequentemente descrita nos trabalhos onde o mesmo polimorfismo foi identificado (HORAI e cols., 1993; TORRONI e cols., 1994c, 1996, BENDALL & SYKES, 1995).

Índices de diversidade foram utilizados para quantificar e comparar a variabilidade do mtDNA na população brasileira analisada e os resultados demonstraram uma alta diversidade. A diversidade haplotípica (h), ou heterozigosidade, calculada pelo método de NEI (1987), foi de $0,9992 \pm 0,0016$, considerando-se todos os 148 indivíduos. Isto indica que duas sequências selecionadas ao acaso, têm uma probabilidade 99,92% de serem diferentes. Quando a variabilidade foi analisada independentemente para cada região geográfica, os resultados apresentaram-se semelhantes, sendo que as três regiões apresentaram índices de diversidade haplotípica de $1,00000 \pm 0,0040$ (Regiões Sul e Nordeste) e $1,00000 \pm 0,0043$ (Região Norte). Os resultados estão resumidos na tabela 11 em comparação com aqueles obtidos por Santos (1998) para a região Sudeste.

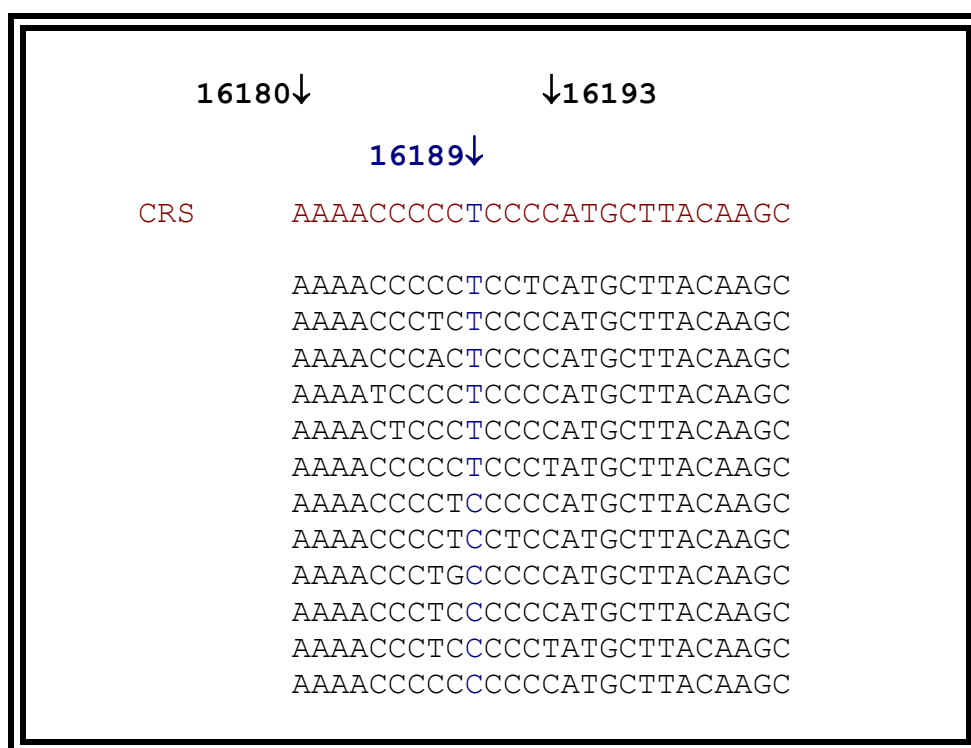


FIGURA 7: POLIMORFISMOS DO DOMÍNIO HIPERVARIÁVEL (ENTRE OS NUCLEOTÍDEOS 16180 E 16193) OBSERVADOS NAS SEQUÊNCIAS BRASILEIRAS. CRS representa a sequência de referência de Anderson (Anderson *et al*, 1981). Algumas sequências com a transição T→C na posição 16189 (em azul) apresentam variações no número de Cs.

Tabela 11: DIVERSIDADE DE NEI OBSERVADA NAS AMOSTRAS DA POPULAÇÃO BRASILEIRA.

Região Geográfica	Genetic Diversity (h) ± DP
	Total
Brasil	0.9982 +/- 0.0010
Norte	1.0000 +/- 0.0043
Nordeste	1.0000 +/- 0.0040
Sul	1.0000 +/- 0.0040
Sudeste	0.9889 +/- 0.0057

IV.2.2 – ANÁLISES FILOGENÉTICAS E A ORIGEM GEOGRÁFICA DAS LINHAGENS DE mtDNA NA POPULAÇÃO BRASILEIRA.

Todas as sequências identificadas na nossa amostra foram submetidas a análises filogenéticas e filogeográficas para verificarmos a origem de suas linhagens. Árvores filogenéticas foram construídas pelo método de Neighbor- Joining (NJ), que permitiu o agrupamento das sequências brasileiras a partir do compartilhamento de mutações específicas (FIGs 8, 9 e 10). É interessante notar que esses grupos, ou ramos das árvores filogenéticas, corresponderam exatamente aos haplogrupos continente-específicos de mtDNA conforme será descrito a seguir, e a maioria das mutações que os definem (e estão representadas em negrito na tabela 10), constituem as mutações características desses haplogrupos.

A determinação da origem étnico geográfica de determinada linhagem mitocondrial e sua classificação em um haplogrupo continente-específico é possível a partir de análise que envolvem: (i) análise de homologia com sequências de mtDNA descritas nas diversas populações mundiais através de comparação em bancos de dados; (ii) presença de mutações específicas na sequência da Região Controle; (iii) análises por RFLP de polimorfismos específicos na região codificadora do mtDNA.

Nós utilizamos essas três metodologias no presente trabalho, o que permitiu a classificação de todos os haplótipos de mtDNA encontrados em haplogrupos continente-específicos. Nossa estratégia consistiu em analisar, primeiramente, a sequência obtida para cada amostra, quanto à presença das mutações que definem esses haplogrupos. Todas as sequências de HVS-I foram então comparadas com sequências do mesmo segmento descritas na literatura e também submetidas a pesquisas de homologia em bancos de dados. Os bancos de dados utilizados nessas análises foram principalmente o Genbank do NCBI, utilizando o programa BLAST (ALTSCHUL e cols., 1990) e o HVRbase (BURCKHARDT e cols., 1999). Isso permitiu identificar a presença de sequências idênticas ou com alta homologia aos haplótipos brasileiros em populações nativas da África e América, bem como em várias populações européias. Isso constituiu uma forte evidência para a determinação da origem geográfica das linhagens brasileiras.

Analisando-se apenas os resultados do sequenciamento, já foi possível inferir a alocação em um haplogrupo continente-específico para quase todas as sequências brasileiras, embora a análise por RFLP dos marcadores de região codificadora seja muito importante para confirmar essa classificação. Portanto, para cada amostra foram realizadas análises de RFLP à procura dos principais polimorfismos continente-específicos descritos na literatura para o haplogrupo inferido pela análise de sequência (TORRONI e cols., 1996; BONATTO E SALZANO, 1997a, 1997b; WATSON e

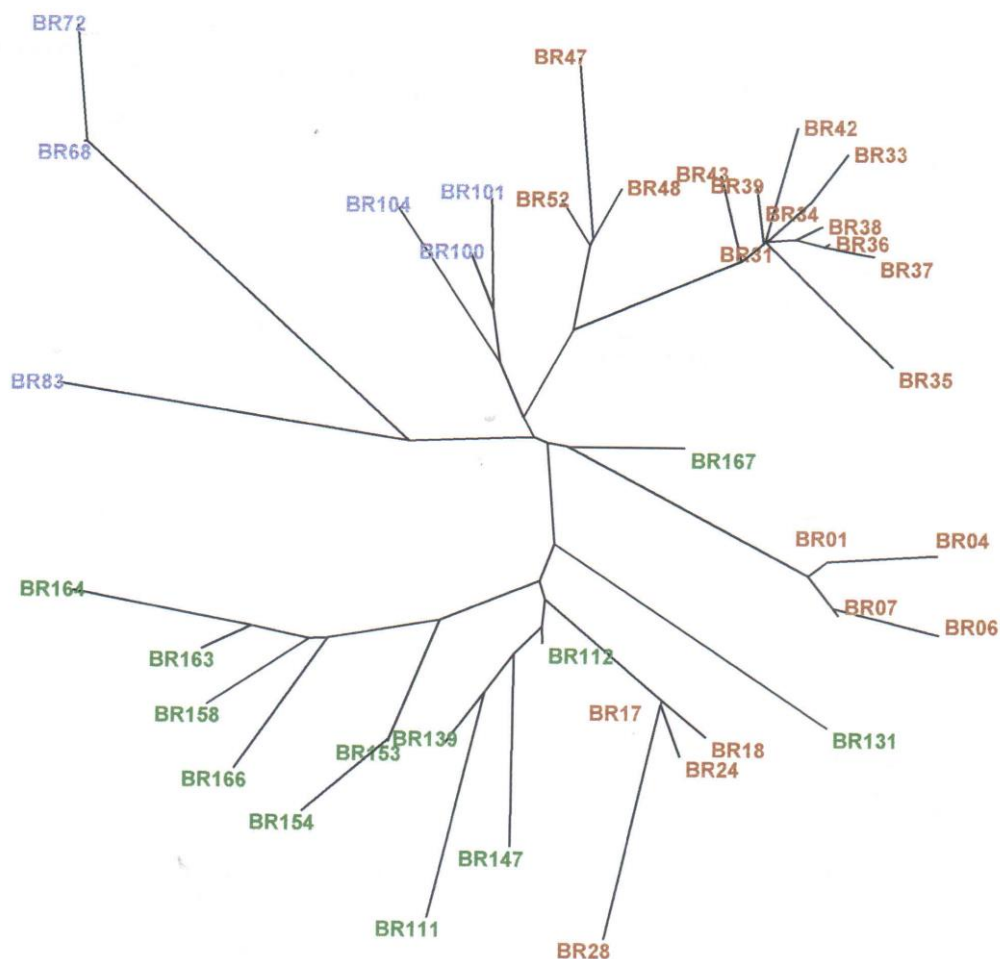


FIGURA 8: ÁRVORE FILOGENÉTICA DE NEIGHBOOR-JOINING CONSTRUÍDA COM AS SEQUÊNCIAS DE REGIÃO HIPERVARIÁVEL I E RESULTADOS DE RFLP PARA AS AMOSTRAS DA REGIÃO NORTE. Em azul estão representadas as linhagens africanas, em verde as européias e em vermelho as ameríndias. A árvore foi construída utilizando-se o pacote de programas PHYLIP e as distâncias filogenéticas foram calculadas pelo método Tamura-Nei (1993).

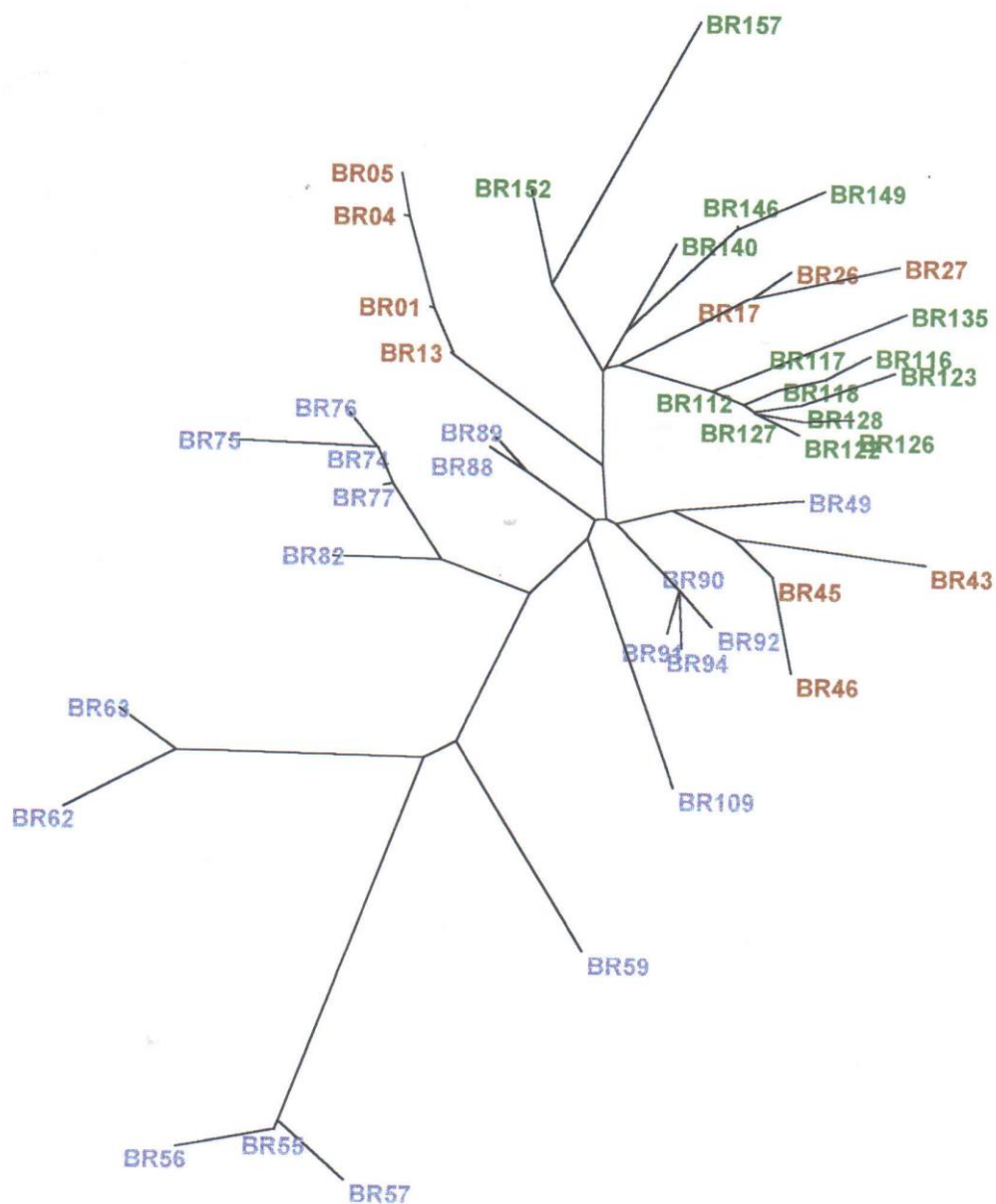


FIGURA 9: ÁRVORE FILOGENÉTICA DE NEIGHBOOR-JOINING CONSTRUÍDA COM AS SEQUÊNCIAS DE REGIÃO HIPERVARIÁVEL I E RESULTADOS DE RFLP PARA AS AMOSTRAS DA REGIÃO NORDESTE. Em azul estão representadas as linhagens africanas, em verde as européias e em vermelho as ameríndias. A árvore foi construída utilizando-se o pacote de programas PHYLIP e as distâncias filogenéticas foram calculadas pelo método Tamura-Nei (1993).

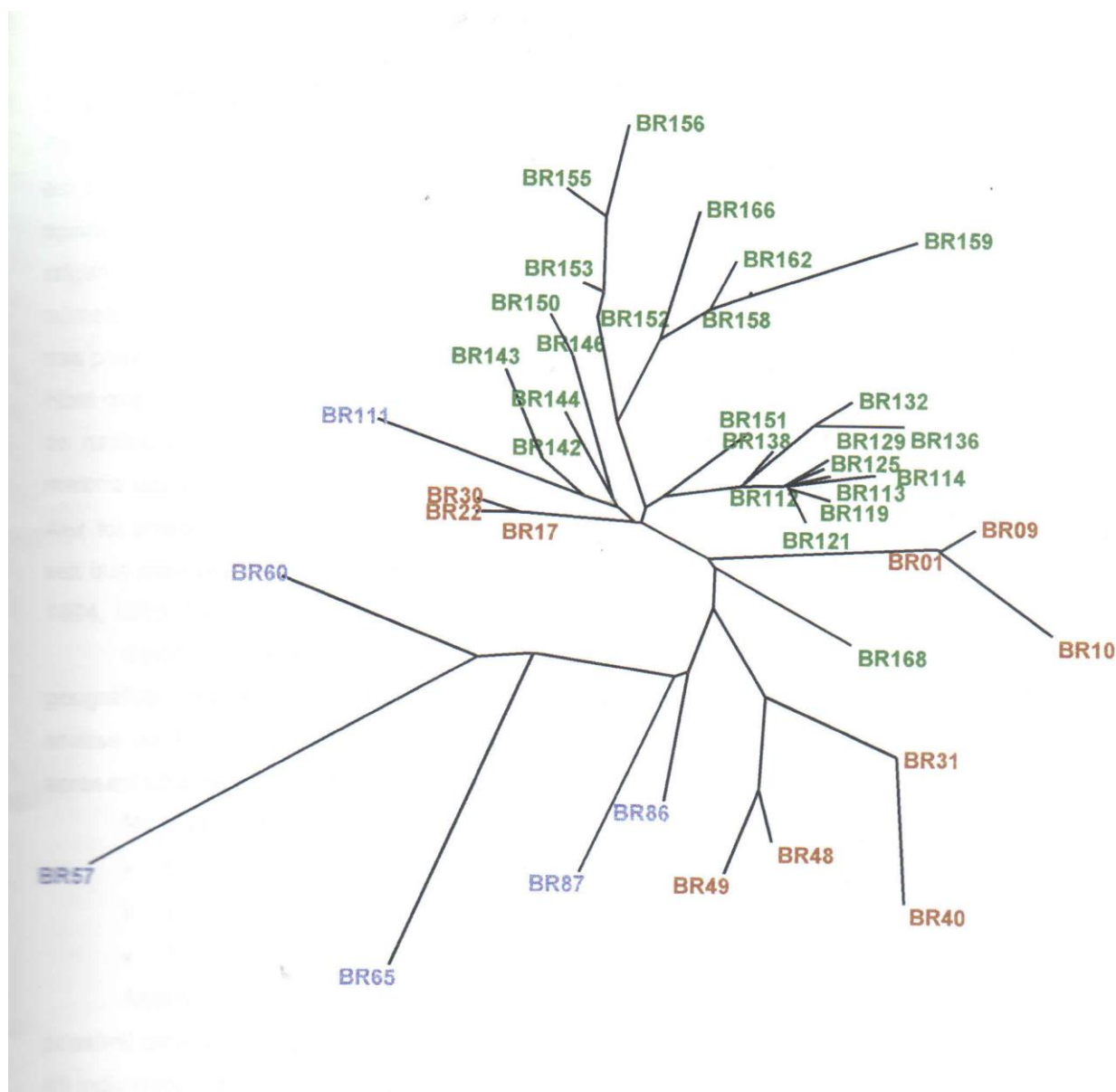


FIGURA 10: ÁRVORE FILOGENÉTICA DE NEIGHBOOR-JOINING CONSTRUÍDA COM AS SEQUÊNCIAS DE REGIÃO HIPERVARIÁVEL I E RESULTADOS DE RFLP PARA AS AMOSTRAS DA REGIÃO SUL. Em azul estão representadas as linhagens africanas, em verde as européias e em vermelho as linhagens ameríndias. A árvore foi construída utilizando-se o pacote de programas PHYLIP e as distâncias filogenéticas foram calculadas pelo método Tamura-Nei (1993).

cols., 1997; MACAULAY e cols., 1999). Como os mtDNAs que apresentaram sequências idênticas na Região Controle podem, na realidade, pertencer a linhagens diferentes (devido ao caráter altamente polimórfico do mtDNA e à ocorrência de homoplasias), analisou-se por RFLP todos os 148 indivíduos. Para a maior parte das linhagens, apenas os RFLPs específicos do haplogrupo e/ou do grupo étnico em questão foram analisados. Por exemplo, os mtDNAs classificados como europeus foram testados apenas para os polimorfismos europeus, e o mesmo foi realizado para as linhagens de provável origem africana e/ou ameríndia (ou asiática). No entanto, algumas análises foram estendidas a um número maior de amostras para aumentar a confiabilidade da classificação. Para confirmar a origem das possíveis linhagens africanas, todas elas foram testadas primeiramente para o marcador 3592 + *HpaI* que de acordo com Chen e colaboradores (1995) e Watson e colaboradores (1997) caracteriza os haplogrupos **L1** e **L2**, que em conjunto estão presentes em cerca de 70% dos indivíduos na maioria das populações africanas da região do sub-Saara. Da mesma forma, o polimorfismo 7025 – *AluI* foi analisado para todas as amostras que apresentaram possíveis sequências européias, uma vez que esse marcador caracteriza o haplogrupo **H**, principal haplogrupo europeu (TORRONI e cols., 1994, 1996, 1998; RICHARDS e cols., 1996, 1998; WALLACE e cols., 1999).

Cinco linhagens (BR109, BR131, BR132, BR139 e BR151) não puderam ter sua origem geográfica inferida a partir dos dados do sequenciamento. Para essas amostras foi realizada uma análise de RFLP de alta resolução onde todos os polimorfismos de RFLP continente-específicos apresentados na tabela 12 foram testados.

No total, 30 sítios polimórficos foram analisados sendo assim distribuídos:

- 10 polimorfismos característicos de populações africanas
- 14 polimorfismos característicos de populações européias
- 17 polimorfismos característicos de populações asiáticas e nativo-americanas.

Assim, a partir da análise das 107 linhagens encontradas nas 148 amostras brasileiras foi possível determinar que 32,43% da nossa amostra (representada por 33 linhagens encontradas em 48 indivíduos) apresentaram origem ameríndia, enquanto 23,65% (presentes em 35 indivíduos) eram de origem africana, e 43,92% (65 indivíduos) eram de origem européia (TAB 13). Considerando-se que a amostra analisada era formada por indivíduos que se auto classificam “brancos” e pertencem principalmente às classes sociais média e média-alta, nossos resultados se mostraram bastante interessantes, uma vez que mais de 55% da amostra é definida por matrilineagens nativo americanas ou africanas.

Tabela 12: RESULTADO DAS ANÁLISES DE RFLP REALIZADAS NOS HÁPLÓTIPOS BRASILEIROS

Sítios de Restrição																																
Amostras																													Haplogrupo			
	HaeIII 663	Ddel 1715	Mbol 2349	Hpal 3592	HaeII 4529	NlaIII 4577	HaeII 4830	AluI 5176	AluI 7025	AluI 7055	HhaI 7598	Avall 8249	Mbol 8616	HaeIII 8994	HaeII 9052	TaqI 9070	AluI 10032	TaqI 10084	Ddel 10394	AluI 10397	HaeIII 11641	HinfI 12308	Hpal 12406	Rsa I 12810	HincII 13259	BamHI 13366	BstNI 13704	Mse I 14766		AluI 15606	HinfI 16389	
BR01-N	+																															A
BR01-NE	+																															A
BR01-S	+																															A
BR04-N	+																															A
BR04-NE	+																															A
BR05	+																															A
BR06	+	+	-	-	+		-		+			-	+	+			-	-	-	-		-	+			-	+		-	-		A
BR07	+																															A
BR09	+																															A
BR10	+																															A
BR13	+																															A
BR31-N																										-						C
BR31-S (a)																										-						C
BR31-S (b)																										-						C
BR33																										-						C
BR34					+	+	-																			-						C
BR35																										-						C
BR36																										-						C
BR37																			+							-						C
BR38																										-						C
BR39																										-						C
BR40																										-						C
BR42																										-						C
BR43-N																										-						C
BR43-NE																										-						C
BR45-NE (a)								-																								D
BR45-NE (b)								-																								D
BR46								-																								D
BR47								-																								D
BR48-S								-																								D
BR48-N (a)								-																								D
BR48-N (b)								-																								D
BR49								-				-	+	+	+		-	-	+	+		-				-	+					D
BR52			-	-	+	+	-	-	+			-	+				-	-	+	+		-	+									D
BR55-NE (a)				+																		+										L1a
BE55-NE (b)			-	+														-	+			+										L1a
BR56				+																		+										L1a
BR57-S				+																		+										L1a
BR57-NE				+																		+										L1a
BR59		+	+								-						-									-						L1b
BR60		+	+								-																					L1b
BR62				+							-						+								+							L1c
BR63				+							-						+								+							L1c
BR65				+							-						+								+							L1c
BR68				+							-						+								+							L1c
BR72				+							-						+								+							L1c

Amostras	HaeIII 663	Ddel 1715	Mbol 2349	HpaI 3592	HaeIII 4529	NlaIII 4577	HaeIII 4830	AluI 5176	AluI 7025	AluI 7055	HhaI 7598	Avall 8249	Mbol 8616	HaeIII 8994	HaeIII 9052	TaqI 9070	AluI 10032	TaqI 10084	Ddel 10394	AluI 10397	HaeIII 11641	HinfI 12308	HpaI 12406	Rsa I 12810	HincII 13259	BamHI 13366	BstNI 13704	Mse I 14766	AluI 15606	HinfI 16389	Haplogrupo
BR74				+														-	+												L2
BR75		+	-	+										+				-	+		-						+			+	L2
BR76				+														-	+											+	L2
BR77				+														-	+											+	L2
BR82				+														-	+											+	L2
BR83				+															+											+	L2
BR86			-	-									-					-	+												L3d
BR87			-	-									-					-	+												L3d
BR88			-	-	+		-		+									-	+												L3d
BR89			-	-					+				-					-	+												L3d
BR90			+	-									+					-	+												L3e
BR91-NE (a)			+	-									+					-	+												L3e
BR91-NE (b)			+	-									+					-	+												L3e
BR91-NE (c)			+	-									+					-	+												L3e
BR92			+	-									+					-	+												L3e
BR94			+	-									+					-	+												L3e
BR99			+	-					+									-	+												L3e
BR100			+	-	+	+	-		+				+					-					+					+			L3e
BR101			+	-									+					-	-												L3e
BR104			+	-									+																		L3e
BR109	-	+	-	-	+	+	-	+	+			-	+	+	+		-	-	+	-		-	+	+	+	-	+	-	-	-	L3*
BR111-S			-	-									+					-	-			+									U6
BR111-NE	-	+	-	-	+	+	-	+	+			-	+	+	+		-	-	-	-		+	+	+	+	-	+	-	-	-	U6
BR112-N (a)										-										-											H
BR112-N (b)										-										-											H
BR112-N (c)										-										-											H
BR112-N (d)										-										-											H
BR112-NE (a)										-										-											H
BR112-NE (b)										-										-											H
BR112-S (a)										-										-											H
BR112-S (b)										-										-											H
BR112-S (c)										-										-											H
BR112-S (d)										-										-											H
BR112-S (e)										-										-											H
BR112-S (f)	+									-			+							-			-								H
BR112-S (g)										-										-											H
BR113										-										-											H
BR114										-										-											H
BR116										-										-											H
BR117-NE (a)										-										-											H
BR117-NE (b)										-										-											H
BR118										-										-			-								H
BR119										-										-											H
BR121										-										-											H
BR122			-	-						-			+					-	-												H
BR123										-										-											H
BR125				-						-										-											H
BR126										-										-											H
BR127										-										-											H

Amostras	HaeIII 663	Ddel 1715	Mbol 2349	Hpal 3592	HaeIII 4529	NlaIII 4577	HaeIII 4830	AluI 5176	AluI 7025	AluI 7055	HhaI 7598	Avall 8249	Mbol 8616	HaeIII 8994	HaeIII 9052	TaqI 9070	AluI 10032	TaqI 10084	Ddel 10394	AluI 10397	HaeIII 11641	HinfI 12308	Hpal 12406	Rsa I 12810	HincII 13259	BamHI 13366	BstNI 13704	Mse I 14766	AluI 15606	HinfI 16389	Haplogrupo
BR128									-										-												H
BR129									-										-												H
BR131	-	+	-	-	+	+	-	+	+			-	+	+	+		-	-	-	-	-	-	+	+	+	-	+	-	-	-	pre-V
BR132	-	+	-	-	+	+	-	+	+			-	+	+	+		-	-	-	-	-	-	+	+	+	-	+	-	-	-	pre-V
BR135						-			+																						V
BR136						-			+																						V
BR138	-	+	-	-	+	+	-	+	+			-	+	+	+		-	-	-	-	-	-	+	+	+	-	+	-	-	-	HV
BR140		+	-	-	+	+	-		+			-	+	+	+		-	-	-	-	-	+	+	+	+	-	+		-	-	U2
BR141	-	+	-		+	+	-	+	+		+	-	+	+	+		-	-	-	-	-	+	+	+	+	-	+		-	-	U4
BR143	-	+	-	-	+	+	-	+	+			-	+	+	+		-	-	-	-	-	+	+	+	+	-	+		-	-	U5b
BR144	-	+	-	-	+	+	-	+	+			-	+	+	+		-	-	-	-	-	+	+	+	+	-	+		-	-	U5b1
BR145	-	+	-	-	+	+	-	+	+			-	+	+	+		-	-	-	-	-	+	+	+	+	-	+		-	-	U7
BR147-NE															-																K
BR147-S															-																K
BR148				-											-												+				K
BR150															-																K
BR151															-																K
BR153-N (a)									+																			-			J
BR153-N (b)																												-			J
BR153-NE																						-						-			J
BR153-S (a)																		-	+									-			J
BR153-S (b)																						-						-			J
BR154																												-			J
BR155																						-						-			J
BR156																												-			J1
BR157-S (a)																												-			J1b1
BR157-S (b)																												-			J1b1
BR158																										+			+		T
BR159-N																										+			+		T
BR159-S																										+			+		T
BR160																										+			+		T
BR163																										+			+		T
BR164																										+			+		T
BR165																										+			+		T
BR167-N									+																	+			+		T1
BR167-S																										+			+		T1
BR168				-	-		-				+	+	+				+	-	-	-	-	+				-	+		-	-	I
BR169	-	-	-										+				-	-													X

Nota: As letras em parênteses correspondem às amostras que apresentaram haplótipos idênticos de HVS-I dentro da mesma região geográfica.

Tabela 13: FREQUÊNCIA DOS HAPLÓTIPOS DE MTDNA CONTINENTE-ESPECÍFICOS NO *POOL* MITOCONDRIAL BRASILEIRO

Componente	Frequência por Região					
	Brasil*	Norte	Nordeste	Sul	Brasil (total)	Sudeste
Ameríndio	.32	.54	.22	.22	.33	.33
Africano	.24	.15	.44	.12	.28	.34
Europeu	.44	.31	.34	.66	.39	.31

* os dados da região Sudeste não foram incorporados nos valores dessa coluna.

IV.2.2.1 – COMPONENTE AMERÍNDIO DA POPULAÇÃO BRASILEIRA.

Nossa estratégia de análises filogeográficas permitiu que 48 amostras fossem classificadas como linhagens ameríndias, pela presença de marcadores descritos para os 4 haplogrupos (**A**, **B**, **C** e **D**) característicos das populações Nativo-americanas (TORRONI e cols., 1992, 1993a; FORSTER e cols., 1996; BONATTO & SALZANO, 1996; STONE & STONEKING, 1999).

As sequências BR1, BR4, BR5, BR6, BR7, BR9, BR10 e BR11 que apresentaram, em comum, as transições C→T nas posições 16111, 16223 e 16290; e G→A na posição 16319, permaneceram agrupadas nas árvores de NJ e foram classificadas como linhagens do haplogrupo **A**. Para todas essas amostras a presença de um sítio pra enzima de restrição *HaeIII* na posição 663 do mtDNA confirmou a classificação de haplótipos **A** (TAB 12).

As linhagens BR17 a BR30 da tabela 10 também aparecem agrupadas nas árvores filogenéticas e são caracterizadas por transições T→C nas posições 16189 e 16217, que definem o haplogrupo **B** de populações ameríndias/asiáticas. A classificação de haplótipo **B** das 14 amostras que apresentaram essas 8 sequências, foi confirmada pelo polimorfismo de deleção de 9pb, que associado àquelas duas mutações identifica os mtDNAs do grupo B.

As linhagens BR31 a BR43 (TAB 10) apresentaram as transições C→T na posição 16223 e 16327; e T→C na posição 16298 e 16325, e foram por isso classificadas como mtDNAs do haplogrupo **C** ameríndio/asiático. Todas as 14 amostras que apresentaram essas sequências, foram caracterizadas também pelo polimorfismo 13259 – *HincII* que caracteriza as linhagens do haplogrupo **C** (TAB 12).

Finalmente, as linhagens BR45 a BR52 (TAB 10), que compartilham as transições em 16223 e 16325 com as linhagens do haplogrupo C, e apresentam ainda a transição T→C na posição 16362, se apresentaram agrupadas em todas as árvores filogenéticas (sempre em um ramo intimamente relacionado ao ramo dos haplótipos C) e foram classificadas como haplótipos do grupo **D** Ameríndio/Asiático. O polimorfismo 5176 – *AluI* foi confirmado para essas 9 amostras (TAB 12).

A disposição sempre associada dos ramos dos haplogrupos C e D se deve ao fato de que estes dois haplogrupos constituem grupos derivados do macro-haplogrupo asiático **M** (TORRONI e cols., 1996, WALLACE e cols., 1999), com o qual compartilham mutações em comum (identificadas pelos polimorfismos de RFLP 10394 + *Ddel*, e 10397 + *AluI*).

É importante mencionar que os quatro principais haplogrupos ameríndio/asiáticos estão representados em todas as três regiões geográficas brasileiras analisadas no presente trabalho, e foram também observados na amostra da Região Sudeste (SANTOS, 1998). As principais linhagens fundadoras dos haplogrupos **A**, **B**, **C** e **D** (FORSTER e cols., 1996; SMITH e cols., 1999) foram identificados em nossa amostra, e correspondem às linhagens BR1, BR17, BR31 e BR45, que são

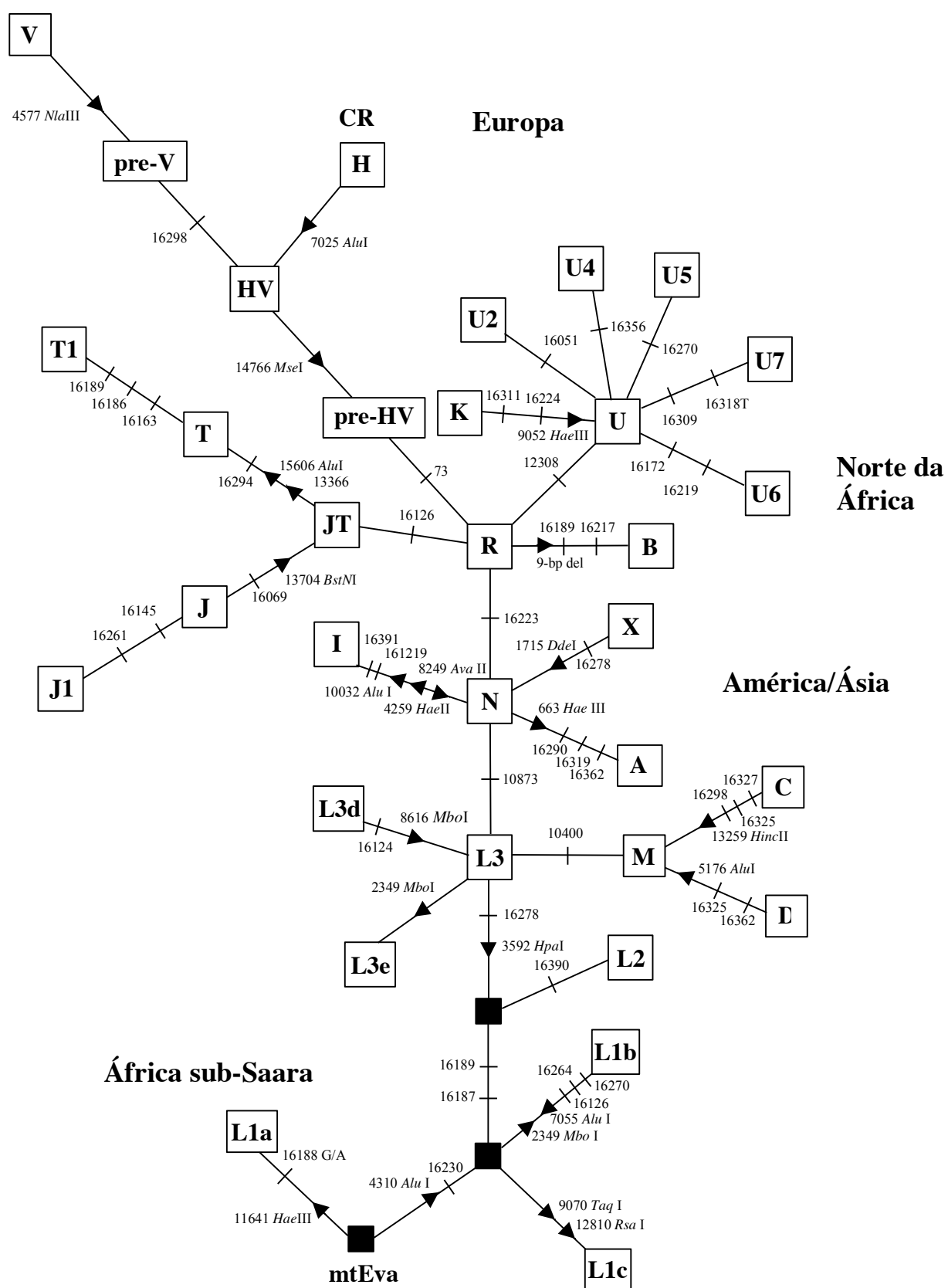
também compartilhadas com várias populações Nativas Americanas. Além disso, as linhagens BR26, BR27, BR43 e BR47 também já haviam sido previamente descritas por Santos, S. e cols. (1996) em populações da Amazônia; e o haplótipo BR49 fora anteriormente identificado em um indivíduo da tribo Mapuche na Argentina (GINTHER e cols., 1993).

IV.2.2.2 – COMPONENTE AFRICANO DA POPULAÇÃO BRASILEIRA

Sete haplogrupos: L1a, L1b, L1c, L2, L3d, L3e e U6, característicos de populações africanas foram identificados em um total de 35 indivíduos da nossa amostra. Por constituírem as populações mais antigas do mundo, as populações africanas possuem uma alta variabilidade genética, que tem atraído a atenção de um grande número de pesquisadores, e estimulado os estudos mais recentes que propuseram a atual classificação de suas linhagens mitocondriais. A classificação por nós adotada para os haplótipos africanos foi baseada nos estudos de Chen e colaboradores (1995), Watson e colaboradores (1997) e Rando e colaboradores (1998,1999). Nós desconsideramos entretanto, a denominação dos haplogrupos L3 apresentada por Chen e colaboradores (2000) (ver item 1.3.3.1). Com relação aos estudos de RFLP, 10 sítios polimórficos de restrição africano-específicos foram analisados, de acordo com o haplogrupo inferido pelos marcadores de HVSI. Como citado anteriormente, o polimorfismo 3592 + *HpaI* foi testado para todas as possíveis linhagens africanas e em seguida, os haplótipos positivos para este sítio (3592 + *HpaI*) foram analisados para os polimorfismos característicos dos (sub)haplogrupos **L1a**, **L1b**, **L1c** e **L2** descritos por Watson e colaboradores (1997), de acordo com a classificação inferida pela sequência de HVS-I. Da mesma forma, as linhagens 3592 – *HpaI* foram testadas para os polimorfismos característicos dos haplogrupos **L3a**, **L3b**, **L3d** e **L3e** (WATSON e cols., 1997; RANDO e cols., 1998,1999), apresentados na tabela 7.

As amostras classificadas como linhagens dos haplogrupos **L1a**, **L1b** e **L1c** constituem os ramos mais longos (e distanciados das demais amostras) em todas as árvores filogenéticas. Isso reflete a natureza ancestral desses haplogrupos e o maior tempo de divergência dessas linhagens em relação às outras. Foi a partir da alta diversidade observada nos haplogrupos africanos que surgiu a teoria mais aceita atualmente sobre a origem do homem moderno, que sustenta a hipótese da “EVA mitocondrial” (PICKFORD, 1991). Acredita-se que a raiz do mtDNA humano apresentava uma relação filogenética mais íntima com os haplogrupos africanos e principalmente o haplogrupo L1a (QUINTANA-MURCI e cols., 1999; KIVISILD e cols., 1999a;b; CHEN e cols., 2000). Essa relação filogenética está representada no diagrama da figura 11.

Nossas análises permitiram identificar como haplótipos **L1** todas as amostras que apresentaram, em comum, as transições C→T nas posições 16187 e 16223; e T→C nas posições 16189 e 16311 (WATSON e cols., 1997).



As amostras BR55, B56 e BR57 aparecem agrupadas na figura 9 e apresentam, além das mutações descritas acima, a transversão T→G na posição 16188 e as transições: A→T em 16129; C→T em 16148, 16168 e 16320; T→C em 16172; e finalmente A→G em 16230. Esse padrão de substituições permitiu a classificação dessas 3 linhagens no (sub)haplogrupo **L1a**, para as quais foi confirmado também o polimorfismo de sítio de restrição 11641+ *HaeIII*, característico desse haplogrupo.

As amostras BR59 e BR60 foram classificadas como haplótipos **L1b** por apresentarem além das mutações específicas do haplogrupo **L1**, transições para C, T e G nas posições 16126, 16270 e 16293, respectivamente, em associação com os marcadores 7055 – *AluI* e 2349 + *MboI*, do haplogrupo **L1b**.

Cinco amostras, BR62 a BR72 (TAB 10) também apresentaram os polimorfismos característicos do haplogrupo **L1** na sequência da Região Controle, associados a transições C→T em 16278, 16294 e 16360, e foram por isso classificadas como linhagens **L1c**. Nessas amostras foram confirmados os marcadores 9070 + *TaqI* e 12810 + *RsaI*, característico desse haplogrupo.

A associação de transições C→T nas posições 16223 e 16278, que caracteriza o haplogrupo **L2** Africano (WATSON e cols., 1997), bem como o ganho de um sítio para a enzima *HinfI* na posição 16389, foi observada em 6 linhagens (BR74 a BR83, na tabela 10). Dentre estas, quatro amostras apresentaram ainda, em comum, as transições C→T em 16294 e A→G em 16309.

Todas as 18 amostras identificadas pelos haplótipos BR55 a BR83 (TAB 10) apresentaram também o polimorfismo 3592 + *HpaI* característica dos haplogrupos **L1** e **L2**.

Pesquisas de homologia em bancos de dados foram particularmente importantes na classificação de 15 linhagens brasileiras pertencentes aos haplogrupos africanos L3d, L3e e L3e*. Esses haplogrupos são caracterizados por poucas mutações de região controle, e muitas dessas ocorrem em sítios hipervariáveis da molécula de mtDNA, fazendo com que para sua identificação sejam necessárias análises de RFLP (RANDO e cols., 1998).

Por exemplo, as amostras BR86, BR87, BR88 e BR89 foram classificadas como linhagens **L3d** por apresentarem transições T→C nas posições 16124 e 16311. O nucleotídeo 16311 constitui uma das posições de HVSI que apresenta um dos índices mais altos de recorrência de mutações (KRINGS e cols., 1999; MEYER e cols., 1999), tendo sido encontrado em linhagens de 12 haplogrupos diferentes, apenas na nossa amostra de brasileiros. A confirmação da classificação foi possível através do polimorfismo 8616 – *MboI* característico do haplogrupo **L3d**, que foi observado em todos esses 4 mtDNAs.

A classificação de linhagens **L3e** para os haplótipos BR91 a BR104 (tabela 10) também não foi baseada apenas nos polimorfismos de HVS-I, e sim pelo sítio de RFLP polimórfico 2349 + *MboI*. A transição C→T na posição 16223 que está presente em quase todos os haplogrupos africanos, é a

única mutação na região controle que caracteriza praticamente todas as linhagens L3e (com exceção de alguns haplótipos que apresentam mutação reversa nesta posição).

A classificação da amostra BR109 como haplótipo L3e* foi baseada em análises de homologia em bancos de dados. As 6 mutações na região hipervariável I, presentes nessa amostra não constituem polimorfismos específicos de nenhum haplogrupo em particular, embora as análises de homologia tenham apontado uma origem africana para esta linhagem. A amostra BR109 foi submetida a uma análise de RFLP de alta resolução (TAB 12), e mais uma vez, não apresentou nenhum marcador que permitisse a sua classificação em um dos haplogrupos de mtDNA já descritos na literatura. Para esta amostra foi inferida portanto a definição de haplótipo africano **L3***. A nomenclatura de “*” , identifica um haplogrupo *inespecífico*, que não pertence a nenhuma subdivisão já descrita de linhagens **L3**. A classificação como linhagem *africana* se deve ao fato de já terem sido encontradas outras linhagens que compartilham com o haplótipo BR109 as transições de HVS-I nas posições 16209 e 16311 (T→C) e 16292 (C→T) em indivíduos de africanos (Toomas Kivisild, comunicação pessoal).

Finalmente o sétimo haplogrupo africano identificado em dois indivíduos foi representado por uma linhagem **U6**. O haplogrupo **U6**, caracterizado por transições C→T em 16172, e A→G na posição 16219, foi descrito como característico de populações do Noroeste Africano, tendo sido identificado na região de Marrocos, Argélia e Ilhas Canárias (MACAULAY e cols., 1999; RANDO e cols., 1998, 1999). O fato de encontrarmos esse haplótipo em brasileiros torna-se interessante, uma vez que não existem muitos indícios, ou registros, da vinda de escravos dessa região da África para o Brasil. Por outro lado, o haplogrupo **U6** foi observado também em Portugal (CÔRTE-REAL e cols., 1996; MACAULAY e cols., 1999), em frequências de até 6%. Isso poderia de certa forma explicar a sua presença na população branca brasileira. Para essa linhagem **U6** foi confirmada a presença do polimorfismo 12308+*Hinf*I, que caracteriza todos os (sub)haplogrupos **U**, inclusive aqueles específicos de populações européias (MACAULAY e cols., 1999; RANDO e cols., 1998,1999).

IV.2.2.3 – COMPONENTE EUROPEU DA POPULAÇÃO BRASILEIRA

Quarenta e quatro linhagens (BR112 a BR168; TAB 10) caracterizadas por sequências homólogas àquelas comumente encontradas em populações européias, foram identificados em 65 indivíduos da nossa amostra. As linhagens européias estavam distribuídas em 8 grupos diferentes, e apenas alguns (sub)haplogrupos menos frequentes na Europa (como por exemplo o haplogrupo W, ou algumas variações do haplogrupo U) não foram observados na nossa amostra.

Quando foram alinhadas às outras sequências na tabela 10, as linhagens européias representaram os haplótipos com menor número de sítios polimórficos em comparação com a sequência de referência de Anderson. Isso ocorre porque a CRS constitui uma linhagem do

haplogrupo europeu **H**, apresentando portanto uma relação filogenética mais estreita com as outras linhagens européias do que com os haplótipos de outros continentes (ver diagrama da figura 11).

O haplogrupo europeu observado em maior frequência na nossa amostra foi o haplogrupo **H**, que é também o mais frequente em praticamente todas as populações européias (TORRONI e cols., 1994, 1996, 1998; RICHARDS e cols., 1996, 1998). Da mesma forma, o haplótipo mais frequente na amostra brasileira, BR112, corresponde à CRS e foi observado em 13 indivíduos. O haplogrupo **H** não pode ser definido por nenhuma mutação característica de HVS-I, e sua classificação é feita principalmente por análises de RFLP (RICHARDS e cols., 1998). Quinze linhagens **H** distintas (que correspondem aos haplótipos BR112 a BR129 na tabela 10), foram observadas nesta amostra, e todos os 28 mtDNAs que apresentaram essas sequências foram caracterizados pelos polimorfismos 7025 – *AluI* e 10394 – *DdeI* que em associação, definem o haplogrupo **H** europeu (TORRONI e cols., 1994c, 1996, 1998).

Nas árvores filogenéticas as linhagens **H** aparecem sempre associadas ao grupo dos haplótipos **V** e **pre*V**, representados na nossa amostra pelas linhagens BR131 a BR136. Embora não compartilhem nenhuma mutação de HVS-I com o haplogrupo **H**, a associação entre estes 3 grupos é estabelecida pela mutação T→C em comum na posição 14766 (que anula um sítio para a enzima de restrição *MseI* nesta posição), e por uma mutação A→G na posição 73 do mtDNA (região hipervariável II, não analisada neste trabalho). Essas mutações caracterizam também o ancestral comum de todas essas linhagens representado na figura 11 por **HV**. As linhagens **V** e **pre*V** representadas na nossa amostra pelos haplótipos BR131 e BR132 (**pre*V**); BR135 e BR136 (**V**) são definidas pela transição T→C em 16298 (não associada às mutações do haplogrupo **C** ameríndio). Para essas amostras o polimorfismo 4577 – *NlaIII*, característico do haplogrupo **V** (TORRONI e cols., 1996, 1998) foi analisado, e apenas as linhagens BR135 e BR136 apresentaram esse marcador, confirmando suas classificações como haplótipos **V**. As linhagens BR131 e BR132 por sua vez, apresentaram apenas o polimorfismo 14766 – *MseI*. A partir desses resultados, nós decidimos realizar uma análise de RFLP de alta resolução também para essas duas amostras. A ausência de outros marcadores continente-específicos sugeriu a classificação dessas linhagens como haplótipos **pre*V** (de acordo com a nomenclatura proposta por MACAULAY e cols., 1999). Este haplogrupo está indicado na figura 11 e representa as linhagens pertencentes ao grupo “ancestral” que teria dado origem ao haplogrupo **V** ou mais especificamente, ao haplogrupo **V***;

As sequências européias não apresentaram uma grande tendência à formação de grupos em nossas árvores filogenéticas, como foi observado, por exemplo, para as sequências africanas e ameríndias. Isso ocorre devido a dois motivos principais: primeiramente, os haplogrupos europeus são caracterizados por poucos sítios polimórficos na HVS-I e em segundo lugar, um número muito pequeno de amostras foi observado para cada um dos haplogrupos europeus em particular. Dessa

forma, apenas para as linhagens dos halogrupos **T** e **J** (FIGs 8 e 10), e **U** (sempre associadas a linhagens **K**, nas figuras 8, 9 e 10), foi observada uma correspondência entre os ramos das árvores e os respectivos haplogrupos.

O haplogrupo **T** foi o segundo haplogrupo europeu mais frequente na nossa amostra e suas linhagens, BR157 a BR166 (presentes em 9 indivíduos; TAB 10), foram caracterizadas pelas transições T→C na posição 16126, C→T nas posições 16294; e pelos polimorfismos 13366 + *Bam*HI e 15606 + *Al*ul que caracterizam esse haplogrupo (TORRONI e cols., 1996, MACAULAY e cols., 1999).

As sequências do haplogrupo **J** (BR152 a BR156) também são caracterizadas pela substituição T→C na posição 16126, embora aqui ela esteja associada a outra transição C→T, dessa vez na posição 16069. A transição em 16126 representa um caráter ancestral dos haplogrupos **J** e **T**, que provavelmente estava presente nas linhagens ancestrais **JT** que deram origem a estes dois haplogrupos (FIG 11) (MACAULAY e cols 1999). Para todas as 10 amostras que apresentaram as linhagens **J** foi confirmado o polimorfismo 13704 – *Bst*MI.

As linhagens **K** (BR146 a BR150) foram agrupadas pelas transições T→C nas posições 16224 e 16311 e aparecem sempre associadas às linhagens **U** (BR139 a BR144) em todas as árvores filogenéticas. Essa associação ocorre porque estes dois haplogrupos compartilham a transição A→G na posição 12308 que pode ser identificada por RFLP pelo polimorfismo 12308 + *Hinf*I (confirmado em todas as linhagens classificadas como haplótipos **U** e **K**). Entretanto, as linhagens **K** são também definidas pelo marcador 9052 – *Ha*ell.

Baseando-nos na filogenia dos haplogrupos europeus proposta por Macaulay e colaboradores (1999) foi possível refinar a classificação das seguintes amostras:

- BR139 → linhagem **U2**: transversão G→C na posição 16129;
- BR140 → linhagem **U4**: transição para C na posição 16356;
- BR142 → linhagem **U5b***: transição C→T em 16270 e T→C em 16189;
- BR143 → linhagem **U5b1**: apresentou as mesmas substituições de BR142 em associação com a transição T→C em 16144;
- BR155 → linhagem **J1**: transições de G→A em 16145 e C→T em 16261, associadas àquelas já descritas para o haplogrupo **J**;
- BR156 → linhagem **J1b1** por apresentar as mesmas substituições de BR155 associadas à transição C→T em 16222;
- BR166 → linhagem **T1**: transições A→G em 16163, C→T em 16186 e T→C em 16189;

Linhagens dos haplogrupos **I** e **X** foram observadas em apenas um indivíduo das regiões Sul e Norte, respectivamente. Esses haplogrupos são encontrados em baixas frequências nas populações européias e podem ser observados principalmente em populações do leste europeu.

Nosso haplótipo BR167 foi classificado como linhagem **I**, por ter apresentado as transições $G \rightarrow A$ em 16129 e $C \rightarrow$ em 16223. O haplogrupo **I** é definido também por uma transição $T \rightarrow C$ na posição 16311, ausente na linhagem BR167, mas as análises de RFLP, que confirmaram a presença dos 4 principais marcadores deste haplogrupo (10028 + *AluI*, 8249 + *Avall*, 4259 – *HaeIII* e 16389 + *BamHI*), foram suficientes para garantir essa classificação.

BR168 foi classificado como linhagem do haplogrupo **X**, por apresentar as transições $T \rightarrow C$ em 16189, e $C \rightarrow T$ em 16223 e 16278. O haplogrupo **X** em particular, representa um grupo interessante, uma vez que ele encontra-se compartilhado entre populações européias e nativas americanas da América do Norte (BROWN e cols., 1998). A classificação da amostra BR168 como haplótipo “europeu” foi possível, porque ela não apresentou a mutação em 16213 característica das linhagens **X** nativas-americanas, e além disso, foi definida por uma transição $C \rightarrow T$ em 16248, que já havia sido previamente indentificada em linhagens **X** européias.

Mais uma vez, a simples análise dos polimorfismos de sequência não foram suficientes para inferir uma classificação para as amostras BR138 e BR151. As pesquisas em bancos de dados por outro lado, sugeriram uma origem européia para essas duas linhagens e assim nós decidimos submetê-las também às análises de RFLP de alta resolução. A amostra BR138, que diferiu da CRS apenas por uma transição $C \rightarrow A$ na posição 16192 seria classificada como haplótipo **H** europeu, uma vez que não apresentou nenhuma substituição característica de HVS-I. No entanto, os resultados de RFLP não confirmaram essa classificação porque a amostra BR138 apresentou-se positiva para o sítio de *AluI* na posição 7025. Essa amostra foi então classificada como linhagem **HV**, devido à presença do polimorfismo 14766 – *MseI*, e ausência de outros marcadores continente-específicos (FIG 11, TAB 12).

Apesar de ter sido definida pela transição $T \rightarrow C$ na posição 16126 que define os haplogrupos europeus **J** e **T**, a amostra BR151 não apresentou os polimorfismos de RFLP característicos desses haplogrupos (TAB 12). Por outro lado, nenhum polimorfismo de RFLP informativo foi encontrado nessa amostra e por isso, nós decidimos analisar a sequência da região hipervariável II, que também é importante na definição de alguns haplogrupos continente específicos (TORRONI e cols., 1996, 1998, RICHARDS e cols., 1998). De fato, a ausência da transição $G \rightarrow A$ (de acordo com a CRS) na posição 73 determinou a classificação de BR151 como linhagem **pre-HV** (FIG 11).

IV.2.3 – DISTRIBUIÇÃO DIFERENCIAL DOS HAPLOGRUPOS CONTINENTE-ESPECÍFICOS DE MTDNA NA POPULAÇÃO BRANCA BRASILEIRA

Nossos resultados demonstraram claramente as linhagens ameríndias, africanas e européias não estão homogeneamente distribuídas nas diferentes regiões brasileiras analisadas. Nós observamos que a região Norte é caracterizada pelo predomínio de haplótipos ameríndios, que foram

identificados em 54% das amostras; enquanto na região Nordeste 44% dos indivíduos apresentaram linhagens africanas, e finalmente, mais de 65% das amostras da região Sul eram constituídas por linhagens européias. Nas regiões Norte e Nordeste o componente europeu foi o segundo mais frequente, e na região Sul, as linhagens ameríndias foram observadas em maior proporção que as africanas (TAB 13).

Da mesma forma que foram encontradas diferenças na frequência de linhagens ameríndias, africanas e européias entre as amostras das regiões Norte, Nordeste e Sul do país, nós observamos também que a proporção dos haplogrupos continente-específicos apresentou-se diferentemente distribuída nessas 3 regiões.

IV.2.3.1 – COMPONENTE AMERÍNDIO

Entre as linhagens ameríndias observou-se uma grande diferença na frequência dos quatro haplogrupos principais (**A**, **B**, **C** e **D**). Na amostra total os haplogrupos B e C foram encontrados em maior frequência, onde cada um correspondeu a 29% dos haplótipos. Essa proporção foi fortemente influenciada pelos resultados da região Norte (onde foram observadas a maioria das linhagens ameríndias), uma vez que esta foi a única região a apresentar uma predominância do haplogrupo C (38% das linhagens ameríndias). Na região Nordeste o haplogrupo A foi o mais frequente (37%) e apenas um haplótipo C (que representou 9% das linhagens ameríndias) foi encontrado, enquanto a região Sul foi caracterizada por frequências iguais dos haplogrupos A, B e C. As linhagens D foram observadas em menor frequência nas regiões Sul e Norte (TAB 14). A análise de uma amostra da região Sudeste (SANTOS, 1998), identificou proporções diferentes dos haplogrupos ameríndios, registrando um predomínio de linhagens **A** (TAB 14). Ao serem analisados em conjunto os dados das regiões Norte, Nordeste, Sul e Sudeste, a contribuição dos haplogrupos ameríndios na população brasileira apresentou-se também diferenciada (TAB 14) sugerindo a participação de diferentes contingentes indígenas na formação da população das diversas regiões do Brasil. Entretanto, o pequeno número de linhagens ameríndias observadas nas regiões Nordeste e Sul pode ser o principal responsável por esses resultados divergentes. Nossos dados apresentam-se em concordância com a literatura, que aponta uma maior frequência de haplótipos **C** nas tribos indígenas da região Norte (SANTOS e cols., 1996; TORRONI e cols., 1993a; MERRIWETHER e cols., 1997).

IV.2.3.2 – COMPONENTE AFRICANO

As diferenças na distribuição das linhagens africanas nas amostras das regiões Norte, Nordeste e Sul, estavam principalmente relacionadas à presença de um ou outro haplogrupo nessas

Tabela 14: FREQUÊNCIA DOS HAPLOGRUPOS DE MTDNA DENTRO DAS TRÊS CONTRIBUIÇÕES CONTINENTAIS PARA O POOL DE DNA MITOCONDRIAL BRASILEIRO.

Frequência por região						
Haplogrupos	Norte	Nordeste	Sul	Brasil*	Sudeste	Brasil Total
Nativo Americanos						
A	.15	.37	.27	.23	.39	.30
B	.31	.27	.27	.29	.30	.29
C	.38	.09	.27	.29	.18	.24
D	.15	.27	.18	.19	.12	.16
Total	1.00	1.00	1.00	1.00	1.00	1.00
Africanos						
L1a	-	.18	.17	.14	.06	.10
L1b	-	.05	.17	.06	.03	.04
L1c	.29	.09	.17	.14	.23	.19
L2	.14	.23	-	.17	.23	.20
L3d	-	.09	.33	.11	-	.06
L3e	.43	.32	-	.29	.32	.30
L3*	-	.05	-	.03	.06	.04
U6	.14	-	.17	.06	.06	.06
Total	1.00	1.00	1.00	1.00	1.00	1.00
Europeus						
H	.27	.65	.39	.43	.45	.44
Pre*V	.07	-	.03	.03	.03	.03
V	-	.06	.03	.03	.13	.06
HV*	-	-	.03	.01	-	.01
U	.07	.12	.06	.08	.09	.08
K	.07	.06	.09	.08	.06	.07
Pre*HV	-	-	.03	.01	-	.01
J	.20	.06	.18	.17	.03	.11
T	.27	.06	.12	.14	.13	.14
I	.07	-	-	.01	-	.01
X	-	-	.03	.01	.06	.03
Total	1.00	1.00	1.00	1.00	1.00	1.00

* os dados da região Sudeste não foram incorporados nos valores dessa coluna.

regiões. As frequências observadas para a amostra total foram influenciadas principalmente pelo resultados da Região Nordeste, que foi a única região a apresentar quase todos os haplogrupos africanos (com exceção do haplogrupo **U6**) identificados nos indivíduos brasileiros. Por exemplo, o haplogrupo mais frequente, L3e, que representa 29% das linhagens africanas, não foi sequer observado na região Sul (TAB 14), e apenas dois haplogrupos (**L1c** e **U6**) foram compartilhados entre as regiões Sul e Norte. Em geral, a proporção dos haplogrupos encontrada na amostra brasileira total se assemelha àquela observada para a região Nordeste (TAB 14), e mais uma vez, o pequeno número de linhagens africanas encontradas nas regiões Sul (apenas 6 haplótipos) e Norte (7 mtDNAs) parece ser o responsável por essas diferenças.

IV.2.3.3 – COMPONENTE EUROPEU

A região Sul foi a principal responsável pela frequência dos haplogrupos europeus na amostra total de brasileiros analisada. Várias linhagens européias foram observadas em todas as três regiões, e os haplogrupos mais frequentes na amostra total (haplogrupos **H**, **J** e **T**) foram também observados em maior frequência nas regiões Norte, Nordeste e Sul (TAB 14). As diferenças observadas na distribuição das linhagens européias de uma região para outra, estava portanto relacionada aos haplogrupos menos frequentes (**pre*V**, **V**, **HV*** e **pre*HV**), representados por apenas um indivíduo, que não foram encontrados em todas as regiões.

Os resultados aqui apresentados foram publicados no periódico *American Journal of Human Genetics* juntamente com os dados da região Sudeste (ALVES-SILVA e cols., 2000; Apêndice 3).

IV.3 – ANÁLISE DAS AMOSTRAS DE SANTANA DO RIACHO

Conforme foi apresentado na seção 1.5.1, nós nos propusemos a analisar também a variabilidade do mtDNA em amostras de uma população pré-histórica brasileira, e essa análise seria realizada a partir da extração de DNA de 18 amostras ósseas de espécimens fósseis da região de Lagoa Santa.

Por constituírem amostras raras e de grande importância arqueológica, nós decidimos aplicar primeiramente a técnica de Racemização de Aminoácidos (R-AA), nas 18 amostras, para avaliar o seu grau de degradação no nível molecular. Posteriormente, somente aquelas amostras que apresentassem valores de racemização para o aminoácido ácido aspártico inferiores a 10% seriam submetidas a experimentos de extração de DNA. As análises de R-AA por HPLC foram realizadas para todas as 18 amostras. A figura 12A mostra o cromatograma obtido para uma corrida de HPLC de um padrão comercial de aminoácidos e os picos de fluorescência estão relacionados às formas

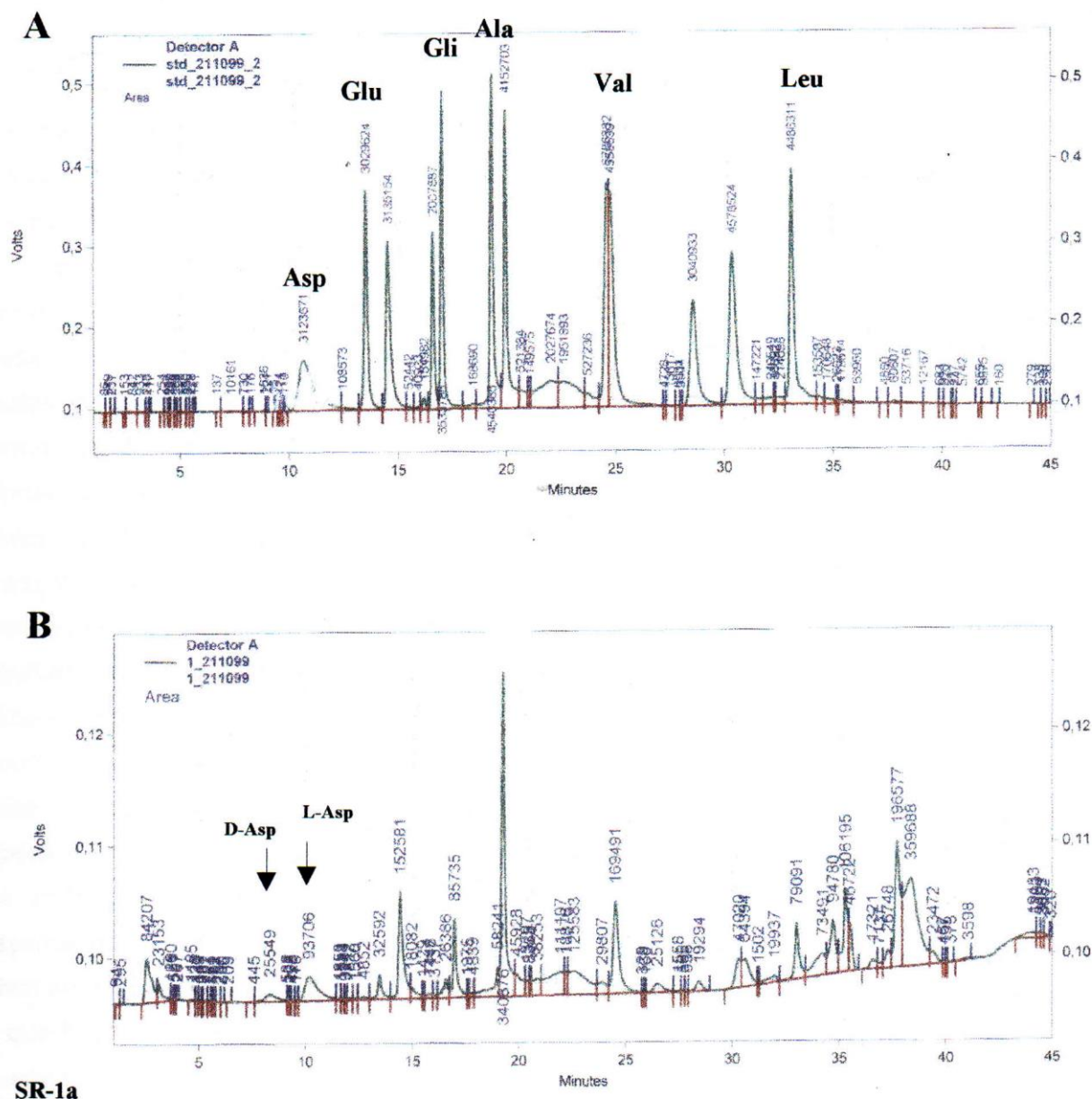
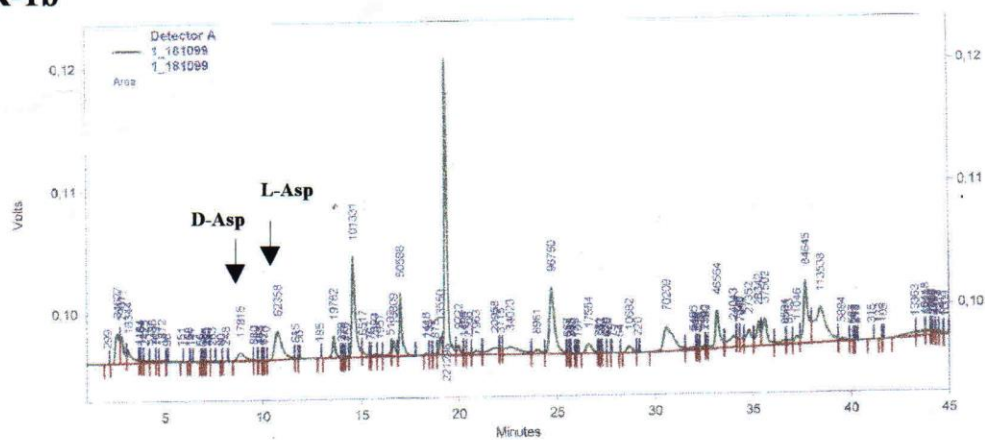


Figura 12: CROMATOGRAMAS DE FLUORESCÊNCIA REPRESENTANDO O PERFIL DE PICOS OBSERVADO PARA (A) UMA AMOSTRA PADRÃO DE MISTURA DE AMINOÁCIDOS L; (B) AMOSTRA SR-1. As proteínas presentes na amostra SR-1 foram degradadas com HCL 6 N e seus aminoácidos constituintes foram derivatizados por uma mistura de OPA e NAC. Dessa reação resultam produtos enantioméricos fluorescentes das formas D e L de cada aminoácido possibilitando sua separação em uma coluna de fase reversa por HPLC. Para o padrão comercial de aminoácidos (Sigma) a etapa de digestão com HCL 1N não foi realizada. Os picos estão identificados na figura **R-A pelo código de três letras do aminoácido correspondente.**

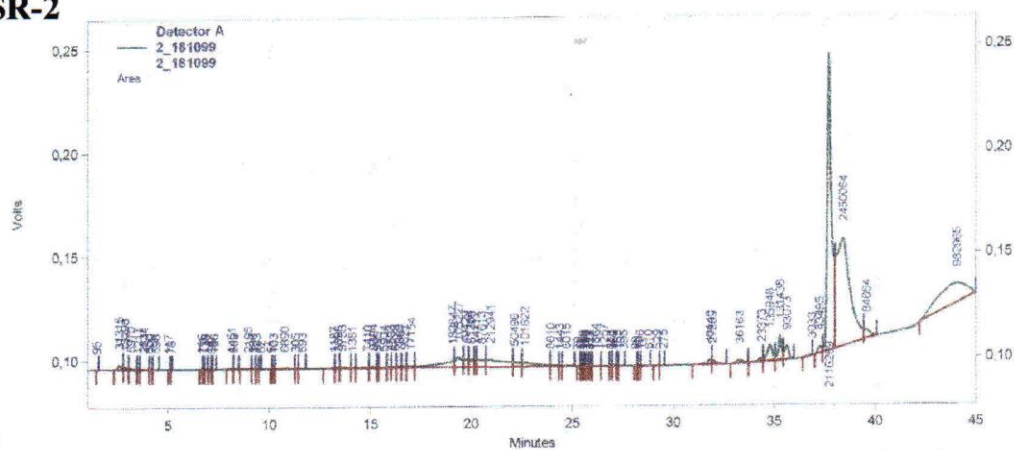
L desses aminoácidos. O perfil esperado para uma amostra fóssil pode também ser observado na figura 12B onde está representado o resultado da amostra SR-1 que apresentou os picos dos aminoácidos D e L. Essa amostra foi utilizada como controle positivo de corrida em todos os experimentos. As formas D e L da maioria dos aminoácidos podem ser facilmente identificadas por essa técnica, permitindo que a partir do cálculo da área sob seus picos seja determinada a razão entre as concentrações dos aminoácidos de interesse.

Os resultados iniciais das análises de racemização foram obtidos na Alemanha, onde a técnica foi inicialmente padronizada. Eles estão apresentados na figura 13. A maioria das amostras apresentou um grande “pico” de interferência, registrado entre os tempos de retenção de 15 e 25 minutos de corrida no HPLC (semelhante àquele identificado na amostra SR-3). Como essa interferência foi observada apenas nas amostras brasileiras, e não aparecia nos padrões de aminoácido e em amostras controle de outras espécies (nas mesmas corridas), ela foi atribuída ao polímero de polivinil-acetato (PVA) utilizado na curadoria dos ossos de Lagoa Santa. Por outro lado, um enorme cuidado foi tomado para se remover completamente a superfície externa dos ossos, antes da retirada de material para análise. Para averiguar se o PVA era realmente incompatível com os reagentes utilizados no processo de derivatização das análises de R-AA, novos experimentos foram realizados, desta vez, no nosso departamento, utilizando como controles novas amostras de sítios arqueológicos da mesma região (Santana do Riacho) não tratadas com essa resina. Nós repetimos a análise de duas amostras BR-1 e BR-9 para testar a padronização da nossa técnica, e analisamos também amostras ósseas de sítios arqueológicos de outras regiões (Caixa d'água de Buritizeiros, Lapa do Boquete, Lapa do Malhador). Os gráficos de fluorescência das novas análises não estão apresentados, mas eles também registraram os mesmos picos de interferência nas amostras que haviam sido analisadas na Alemanha. Nós repetimos a análise de todas as amostras para as quais não que haviam sido possível determinar anteriormente os picos das formas D-Asp e L-Asp e, desta vez, nós fomos capazes de encontrar valores para as amostras SR-7, SR15 e SR18. Os resultados obtidos com as amostras não tratadas com PVA apresentaram-se contudo mais claros e sem forte interferência. A estimativa direta da razão entre as concentrações das formas D e L do ácido aspártico foi calculada dividindo-se a área sob os picos inferidos nos gráficos para esses dois aminoácidos e os resultados estão sumarizados na tabela 15. Nesta tabela é possível observar que valores acima de 0.10 para a razão D/L do aa Asp foram obtidos para quase todas as amostras sugerindo um alto nível de degradação molecular. Apenas duas amostras, SR-20i e SR21i apresentaram valores de D/L Asp inferiores a 0.10. Entretanto, essas amostras, assim como a amostra SR-19i, correspondem a amostras *indeterminadas* da coleção arqueológica de Santana do Riacho, para as quais não existem dados estratigráficos precisos. Além disso, essas amostras foram selecionadas para as análises de R-AA apenas como controle, por não apresentarem o revestimento

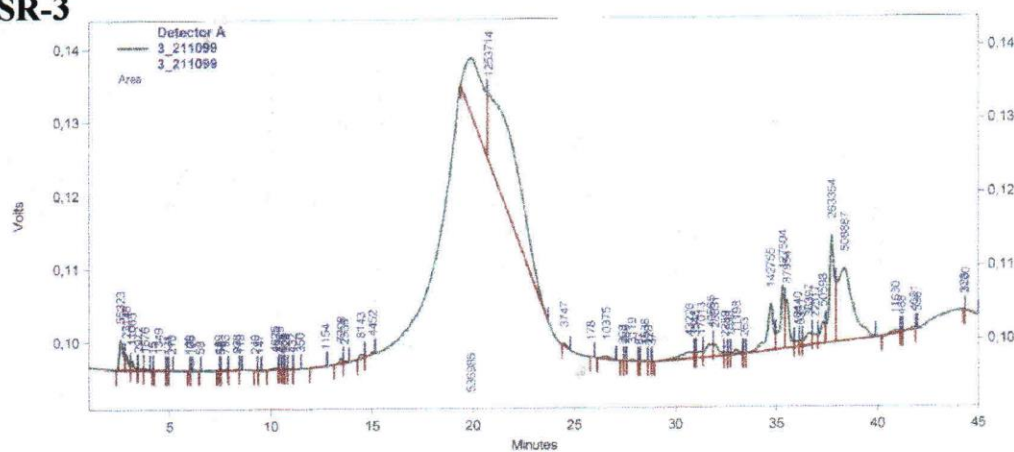
SR-1b



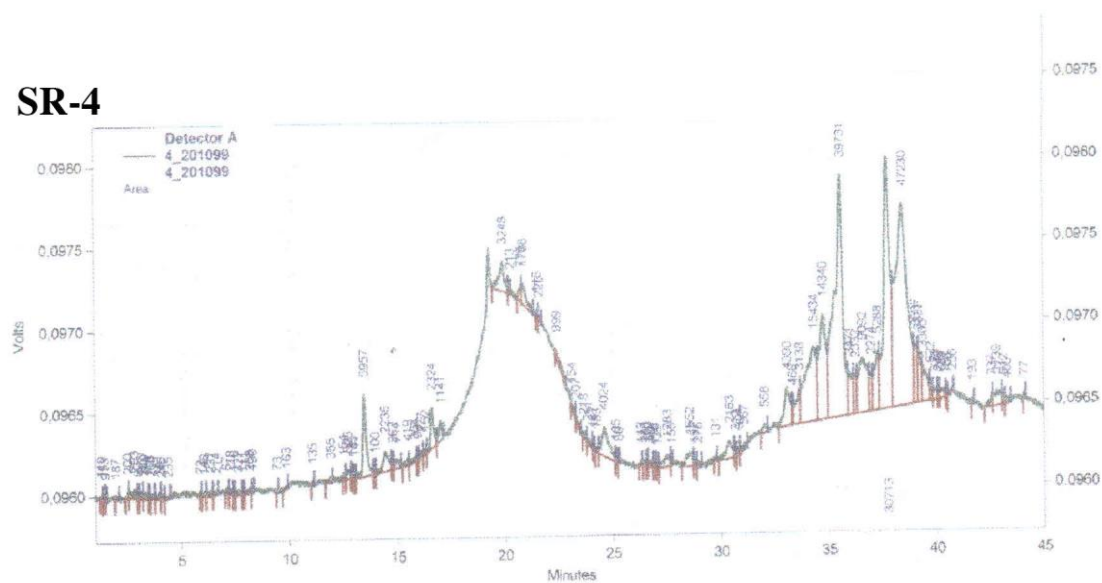
SR-2



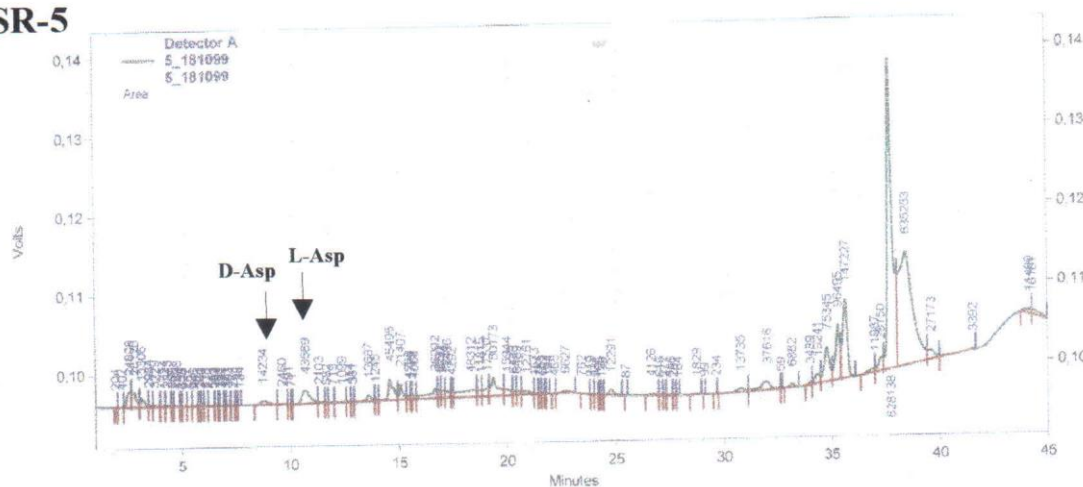
SR-3



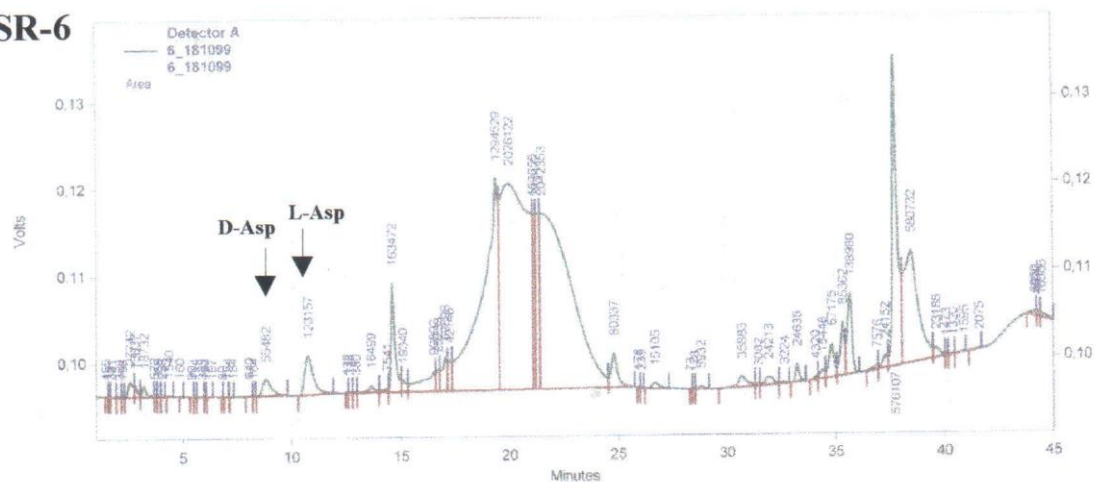
SR-4



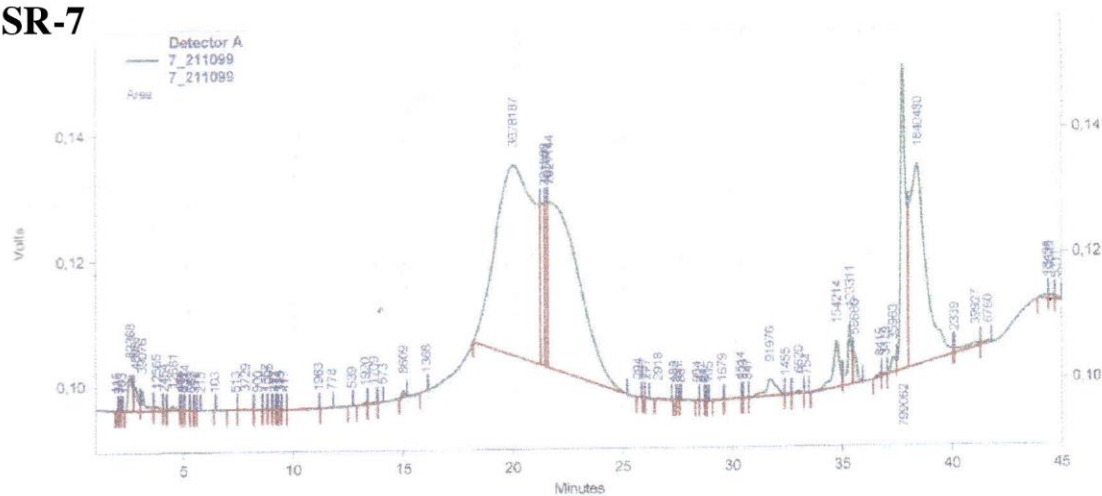
SR-5



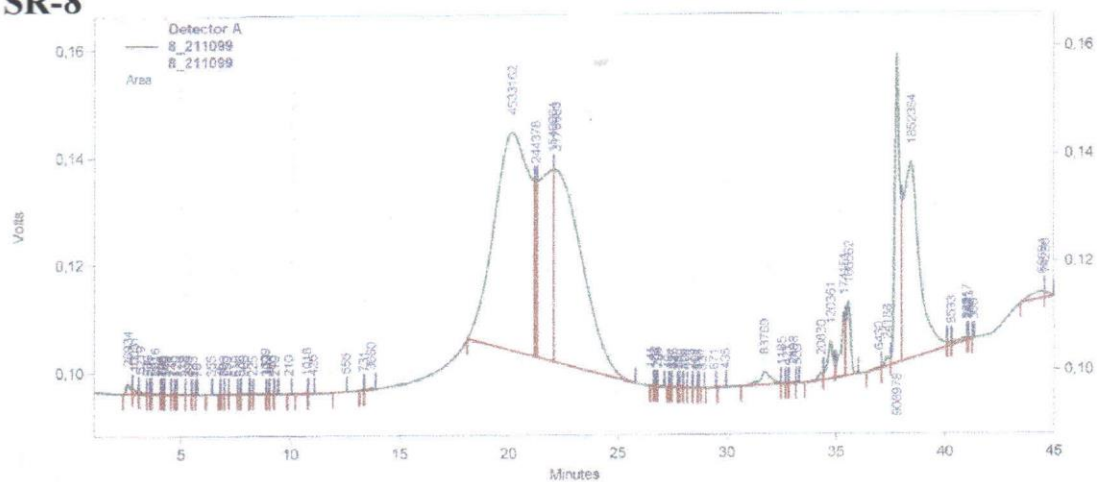
SR-6



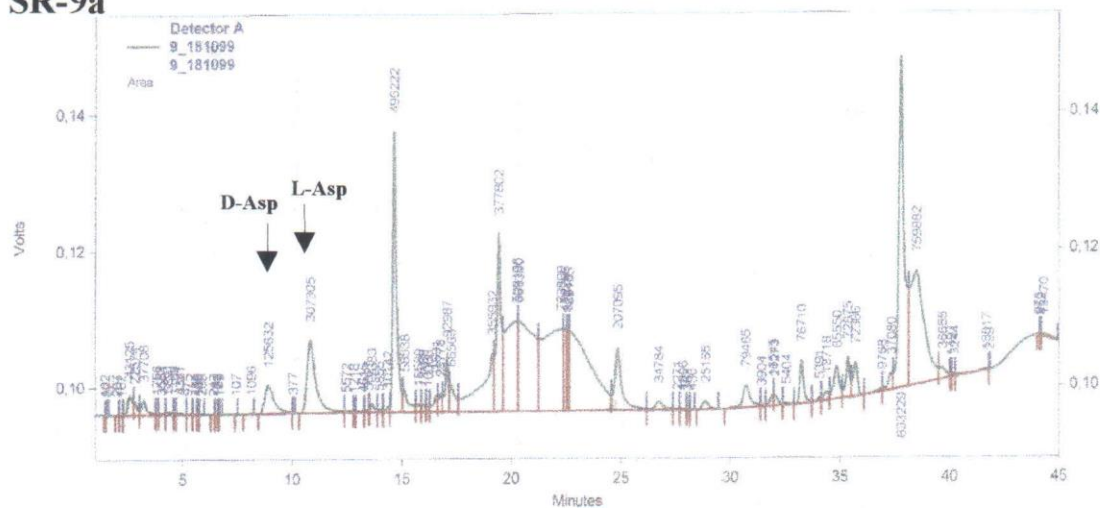
SR-7



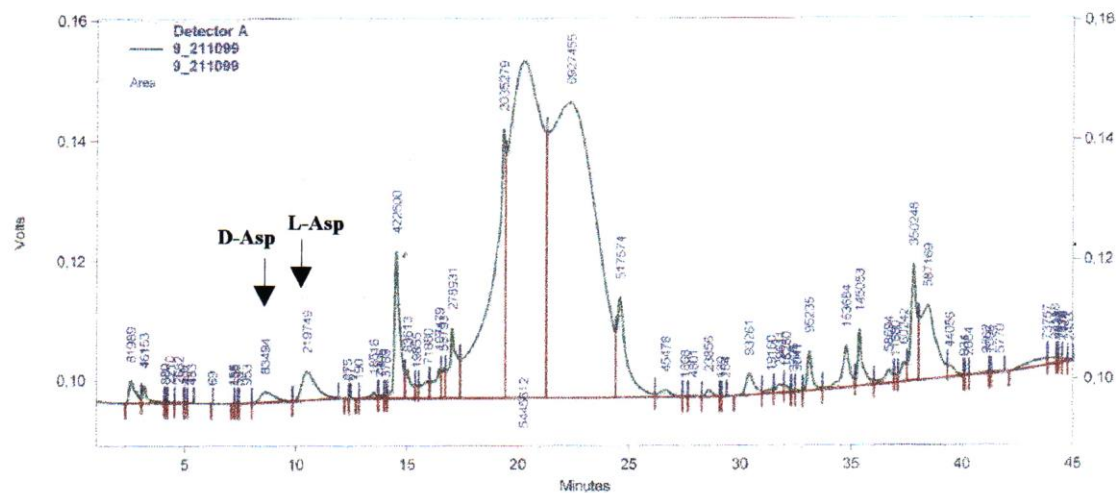
SR-8



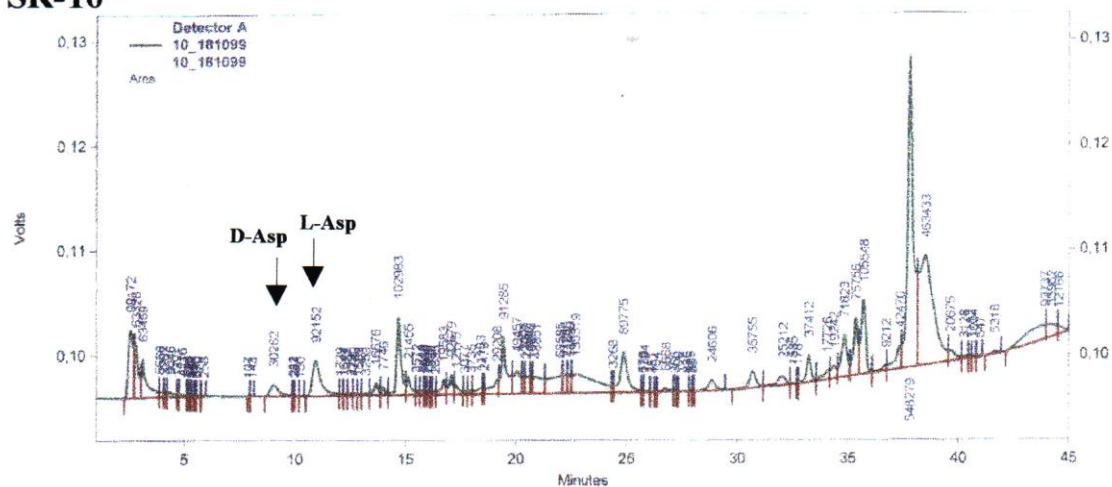
SR-9a



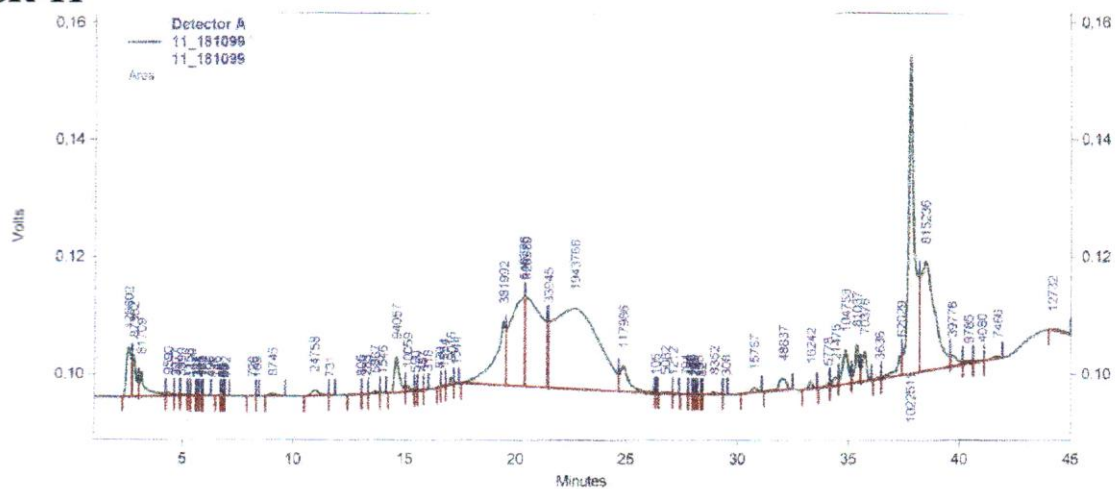
SR-9b



SR-10



SR-11



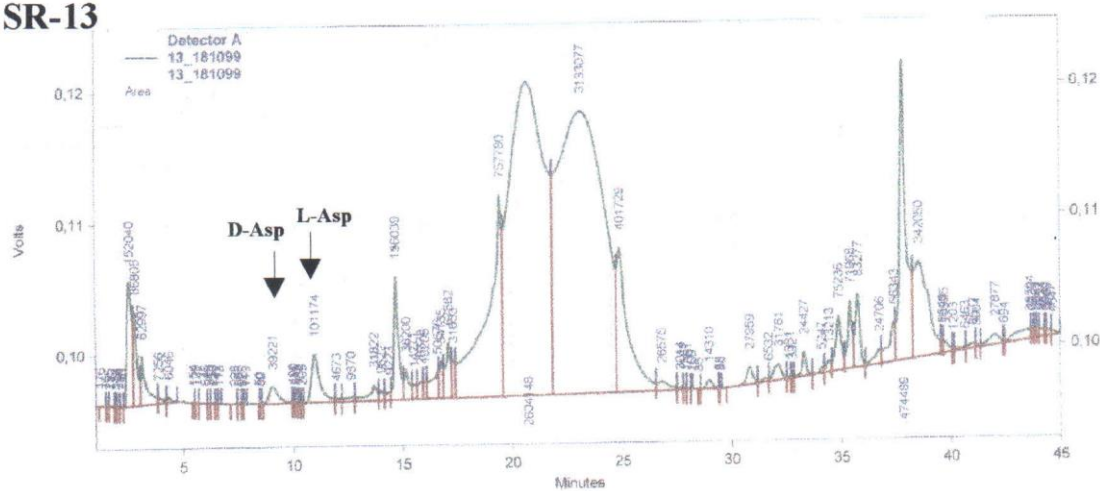
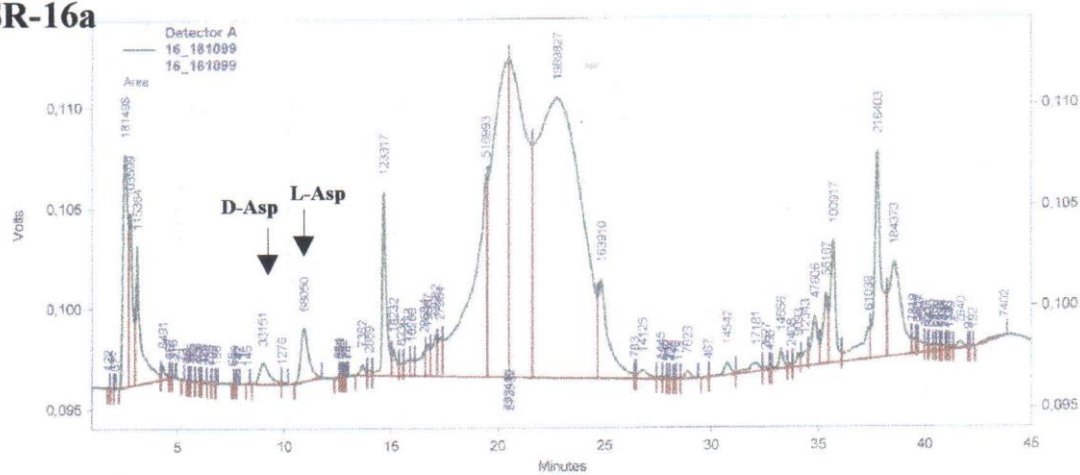
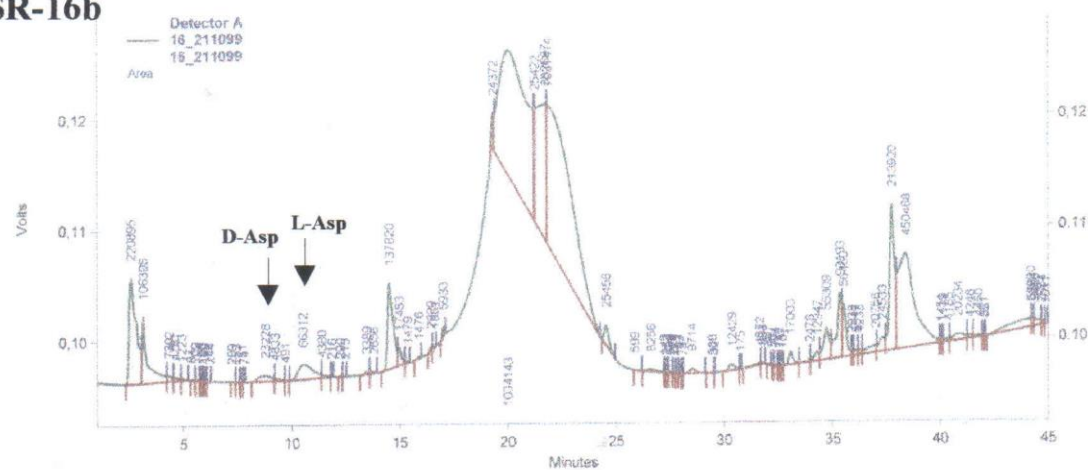
SR-13**SR-16a****SR-16b**

Tabela 15: RAZÃO DA ÁREA SOB OS PICOS DOS AMINOÁCIDOS D-ASP E L-ASP OBSERVADOS PARA AS AMOSTRAS ARQUEOLÓGICAS BRASILEIRAS.

Amostra	Área sob o pico		Razão D/L-Asp
	D-Asp	L-Asp	
SR-1a	17815	62358	0.2856
SR-1b	25549	93706	0.2726
SR-1c*	453353	1523927	0.2975
SR-5	14234	43589	0.3265
SR-6	55482	123157	0.4504
SR-7*	247541	489178	0.5060
SR-9a	125632	307305	0.4088
SR-9b	83494	219749	0.3799
SR-9c*	199158	503687	0.3954
SR-10	30262	92152	0.3284
SR-11	8745	24758	0.3532
SR-13	39221	101174	0.3876
SR-15*	86336	195701	0.4412
SR-16a	33151	68050	0.4871
SR-16b	23228	663120	0.3502
SR-18*	24229	60162	0.4027
SR-19i*	581118	3557178	0.1634
SR-20i*	2926249	32070044	0.0912
SR-21i*	1381183	16499099	0.0837
Lapa do Boquete*	4972828	27567638	0.1804
Lapa do Malhador*	37792	239486	0.1578
Caixa d'agua de Buritizeiros*	572509	2962378	0.1932

*Análises que foram realizadas no Departamento de Bioquímica e Imunologia – ICB – UFMG;

de PVA, e não havia quantidade suficiente de material para realização dos experimentos de extração de DNA.

Apesar dos resultados da racemização de aminoácidos sugerir que o DNA presente nas nossas amostras de Lagoa Santa estava altamente degradado, nós decidimos prosseguir com os experimentos de extração e análise de DNA para duas amostras. As amostras escolhidas foram, SR-1 e SR-10, porque apresentaram aparentemente um melhor aspecto de conservação, identificado pela dureza dos ossos e resistência ao desgaste físico no momento da limpeza. Além disso, para essas duas amostras foram encontrados alguns dos valores mais baixos para a razão D/L Asp (0.27 e 0.32, respectivamente), entre as amostras de Lagoa Santa. A principal finalidade desses experimentos foi, além de tentar recuperar DNA, assimilar as técnicas de extração de DNA desenvolvidas pelo grupo de pesquisas do Instituto Max Planck, bem como a familiarização com os cuidados extremos necessários no trabalho com DNA antigo ou altamente degradado.

A possibilidade de contaminação durante as etapas de extração de DNA foi monitorada por dois tubos “brancos de extração” que foram tratados exatamente da mesma forma que os tubos contendo as amostras, em cada procedimento de extração de DNA (ver item III.5.4). O produto da extração de DNA da amostra SR-10 foi submetido a 3 tentativas de amplificação por PCR, com cada um dos sistemas de iniciadores representados na figura 4, que são capazes de amplificar a HVS-I em 5 segmentos sobrepostos. Resultados negativos foram obtidos em todas essas tentativas de amplificação, bem como para os dois controles de extração realizados em conjunto para essa amostra.

Para a amostra SR-1, todas as amplificações por PCR realizadas com as duplicatas dos controles de extração também foram negativas. No entanto, o extrato de DNA dessa amostra apresentou amplificações positivas em três dos cinco pares de iniciadores utilizados (FIG 4). Esses fragmentos de PCR, que correspondem aos segmentos de Região Controle entre os nucleotídeos 16209-16271; 16287-16356; e 16347-16410 (fragmentos C, D e E da figura 4), foram inseridos em vetores bacterianos e transformados em bactérias *E. coli* competentes. Um mínimo de 10 colônias positivas diferentes foram selecionadas para sequenciamento para cada produto de PCR clonado. A ausência de contaminação nestes experimentos era avaliada através da reproducibilidade de sequência nos diferentes clones de um mesmo produto de PCR.

Isso permitiu que fossem determinadas sequências de aproximadamente 55-70 pares de base para cada produto de PCR, que podem ser visualizadas na figura 14. Nesta figura é possível verificar que a mesma sequência, definida pela transição C→T na posição 16223 (quando comparada com a sequência de referência de Anderson), foi obtida para todos os clones do produto de PCR resultante da reação com os iniciadores MiL16209 e MiH16271. Para os outros dois experimentos de clonagem o mesmo não foi observado, uma vez que mais de uma sequência foi obtida para os clones de cada um daqueles produtos de PCR.

(A)

CRS

(B)

CRS

(C)

CRS

(D)

CRS

Duas sequências distintas, caracterizadas respectivamente por transições C→T na posição 16320; e T→C em 16325; além de uma terceira sequência idêntica à CRS, foram encontradas para o produto de PCR amplificado com os iniciadores MiL16287 e MiH16356. Os clones sequenciados para o terceiro produto de PCR (iniciadores MiL16347 e MiH16410) apresentaram dois tipos de sequência sendo uma delas caracterizada pela transição T→C na posição 16362 e a outra exatamente igual à CRS.

Ao serem analisadas individualmente, as sequências encontradas não são muito informativas, uma vez que correspondem a regiões muito pequenas do mtDNA e não são suficientes para a alocação de determinada linhagem em um dos haplogrupos de mtDNA continente específicos. A transição em 16223 está presente em vários haplogrupos africanos e ameríndios como pode ser observado na tabela 10, enquanto a transição em 16325 caracteriza os haplogrupos **C** e **D** de populações ameríndias, e a transição em 16320 define o haplogrupo L3e2 africano (BANDELT e cols., submetido). A posição 16362 é considerada altamente hipervariável, mas em associação com as mutações em 16223 e 16325 define o haplogrupo **D** ameríndio.

A amostra SR-1 foi submetida a uma segunda extração de DNA para a qual apenas uma amplificação por PCR foi obtida (utilizando-se o par de iniciadores MiL16287 e MiH16356). Pelo menos 3 tentativas de amplificação para todos os outros sistemas de iniciadores foram também realizadas para esse extrato de DNA. Experimentos de clonagem e sequenciamento foram realizados com esse produto de PCR e a única sequência encontrada nos 10 clones selecionados para o sequenciamento, apresentou-se idêntica à sequência caracterizada pela transição C→T na posição 16320, obtida para o primeiro produto de extração (FIG 14). A observação de sequências múltiplas de mtDNA em um mesmo produto de PCR testemunha contra a autenticidade dos nossos resultados, e juntamente com a dificuldade de se amplificar outros sistemas de PCR (igualmente sensíveis), sugerem a contaminação com fontes modernas de DNA durante os procedimentos de amplificação por PCR. Além disso, a razão D/L do aminoácido Ácido Aspártico encontrada para SR-1 (0,27), também sugeriu a incompatibilidade com a recuperação de DNA autêntico para essa amostra.

Portanto, os experimentos de racemização de aminoácidos e extração de DNA realizados nas amostras de Lagoa Santa sugerem que as moléculas de DNA e proteína desses exemplares se encontram em um elevado nível de degradação, que provavelmente reflete o processo de fossilização avançado que estas amostras se encontram. Fatores como elevada umidade e baixos valores de pH do solo parecem estar intimamente relacionados aos processos de degradação de espécimens fósseis. Neste sentido, é importante mencionar que os sítios arqueológicos de Santana do Riacho são constituídos principalmente por grutas calcárias caracterizadas predominantemente por alta umidade.

Discussão

V.1 – VARIABILIDADE DO DNA MITOCONDRIAL

No presente trabalho nós caracterizamos a variabilidade do mtDNA em indivíduos que representam a fração “branca” da população brasileira, e apresentamos dados referentes à análise de três regiões geográficas do país. Nossos resultados nos permitiram avaliar a diversidade genética das regiões Norte, Nordeste e Sul do país, assim como descrever as contribuições relativas de matrilineagens européias, africanas e ameríndias (ou asiáticas) para o *pool* genético dessas regiões. Foi muito interessante notar que as diferenças físicas e sobretudo culturais que diferenciam “as populações” de cada região geográfica brasileira, podem ser também determinadas sob uma perspectiva genética.

Desta forma, os resultados aqui apresentados em associação àqueles obtidos anteriormente, também em nosso laboratório por Santos (1998), nos permitiram retratar a variabilidade do DNA mitocondrial de Norte a Sul do Brasil. Mais importante, nós também fomos capazes de reconhecer “geneticamente” os distintos processos demográficos que originaram a população branca das diferentes regiões brasileiras, através de uma análise da distribuição de linhagens ameríndias, africanas e européias nas nossas amostras. Nós demonstramos ainda, que a alta diversidade observada nas três regiões estudadas (identificada através de parâmetros matemáticos e apresentada na tabela 11), é o resultado da participação em diferentes proporções desses três componentes étnicos.

V.2 – CLASSIFICAÇÃO DE LINHAGENS DE MTDNA

Um dos pontos importantes do nosso trabalho constituiu na realização para cada amostra, de uma extensa análise de polimorfismos continente-específicos combinando metodologias de sequenciamento da região hipervariável I do mtDNA e análises de RFLP. Mutações recorrentes são muito frequentes em sítios da molécula que apresentam altas taxas de mutação, e podem provocar o desaparecimento desses polimorfismos fazendo com que a utilização de apenas uma ou outra metodologia possa resultar em uma classificação errada das linhagens (MACAULAY e cols., 1999; RANDO e cols., 1999; TORRONI e cols., 2000). Estas mutações ocorrem principalmente na Região Controle e geram ambigüidades nas análises filogenéticas, que podem ser resolvidas adicionando-se a informação de sequências de outras partes do mtDNA ou através de testes de RFLP (STARIKOVSKAYA e cols., 1998; MACAULAY e cols., 1999; QUINTANA-MURCI e cols., 1999). Essa associação de metodologias diferentes na análise da sequência do mtDNA tem se mostrado muito importante na determinação da origem geográfica de alguns haplótipos (TORRONI e cols., 1997; BROWN e cols., 1998; SCHURR e cols., 1999) e isso é particularmente importante nos estudos que envolvem populações onde existem intensos processos de miscigenação.

Nossa primeira comprovação da importância de se associar as metodologias de sequenciamento e análises de RFLP ocorreu ao determinarmos a origem européia da amostra BR-3R. Essa amostra apresentou o polimorfismo de deleção de 9-pb que é bastante raro na Europa, mas ao final das análises nós comprovamos a sua classificação como linhagem do haplogrupo **H** europeu. A sequência de HVS-I identificada inicialmente na amostra BR-3R, não indicava uma origem na Europa, uma vez que a associação de transições em 16192, 16274 e 16362 que definiram esse haplótipo não são frequentemente observadas naquelas populações. Além disso, a posição 16362 está entre os sítios mais hipervariáveis dentro da Região Controle (HASEGAWA e cols., 1993; MEYER e cols., 1999; STONEKING, 2000). Nas análises seguintes a ausência da transição G→A na posição 73 detectada pelo sequenciamento da HVS-II, constituiu um marcador para linhagens dos haplogrupos **H** e **V**, e finalmente, os testes de RFLP identificaram o polimorfismo *AluI* – 7025, permitindo a classificação precisa da amostra BR-3R como linhagem do haplogrupo **H**. Mesmo assim, a ampla distribuição do haplogrupo **H** por praticamente toda a Europa, Oriente Médio, Ásia Ocidental (Região do Cáucaso), e até mesmo Norte e Nordeste africanos (RICHARDS e cols., 1998, 2000; KRINGS e cols., 1999; MACAULAY e cols., 1999; RANDO e cols., 1999) colocava em dúvida a nossa inferência de uma origem européia para essa amostra. Por isso, a descoberta de linhagens idênticas e/ou homólogas a este haplótipo brasileiro na região da cidade do Porto em Portugal foi de extrema importância, permitindo a confirmação da origem européia da linhagem BR-3R em um nível (sub-) continental.

Os primeiros trabalhos envolvendo análises de DNA mitocondrial já demonstravam que a Região Controle apresenta a maior taxa de mutação dessa molécula, e vários estudos mais recentes sugerem que algumas posições em particular comportam-se ainda como “hotspots” mutacionais por serem altamente susceptíveis a eventos recorrentes de substituição nucleotídica (Meyer e cols., 1999; Stoneking, 2000). Os estudos realizados com as amostras da região Sudeste (SANTOS, 1998) mostraram que a Região Hipervariável II não apresenta um grande número de polimorfismos continente-específicos, e vários haplogrupos não podem ser diferenciados pela sequência dessa região. Alguns trabalhos da literatura também reportam dificuldades ao se tentar correlacionar os dados de HVS-II em análises filogenéticas (HORAI e HAYASAKA e cols., 1990; VIGILANT e cols., 1991; CHEN e cols., 2000). Além disso, já foi demonstrado por Meyer e colaboradores (1999) que nesta região se localizam posições com taxas de mutação extremamente altas, da ordem de 6,2 eventos mutacionais por sítio, em um milhão de anos (6,2/sítio/Ma), enquanto a média para a toda a Região Controle é de 0.54/sítio/Ma (MEYER e cols., 1999, STONEKING, 2000). Cabe aqui ressaltar que estes valores foram obtidos através de análises filogenéticas de um banco de dados composto por 1229 sequências de HVS-I e 385 sequências de HVS-II, de diversas populações. Entretanto, na literatura existe ainda muita controvérsia no que diz respeito às velocidades relativas de mutação dentro da Região Controle e também quanto à forma de avaliar essas taxas de mutação. Por

exemplo, estudos que envolvem a análise do perfil de mutações em famílias ou pedigrees apresentaram valores altamente divergentes como 2,5/sítio/Ma (PARSONS e cols., 1997) e 0,32/sítio/Ma (SIGURÐARDÓTTIR e cols., 2000).

Dessa forma, devido ao caráter altamente polimórfico e por isso mesmo “menos” informativo de alguns sítios específicos da HVS-II, nós decidimos analisar nossas 148 amostras brasileiras subsequentes através de sequenciamento da Região Hipervariável I e testes de RFLP de sítios relevantes. Entretanto, em um caso específico (como foi descrito para a amostra BR151) a análise da posição 73 do mtDNA por sequenciamento da HVS-II se fez necessária.

V.3 – ORIGEM DAS LINHAGENS MITOCONDRIAS BRASILEIRAS

Nossos resultados do sequenciamento de HVS-I e testes de RFLP nos permitiram demonstrar que entre as 148 amostras brasileiras analisadas, 32% constituíram haplótipos ameríndios, 24% apresentaram origem africana e 44% eram linhagens européias. Como foi apresentado na descrição dos resultados, a classificação de todas as amostras foi inferida e/ou confirmada pela combinação das duas metodologias. Entretanto, é importante considerar a distribuição dos haplogrupos continente-específicos nas populações nativas em que eles foram descritos, para confirmar a origem geográfica das linhagens encontradas na população brasileira.

V.3.1 – HAPLÓTIPOS AMERÍNDIOS NA POPULAÇÃO BRANCA BRASILEIRA

Apesar de terem se originado de linhagens ancestrais provenientes da Ásia há milhares de anos atrás, os haplogrupos característicos de populações nativo americanas adquiriram algumas mutações “americano-específicas” durante sua evolução (TORRONI e cols., 1992, 1993a, 1993b; FORSTER e cols., 1996), e por isso suas linhagens podem ser facilmente diferenciadas das sequências homólogas dos haplogrupos asiáticos. Essa afirmativa é válida para os haplogrupos A, C e D nativo americanos, como será discutido a seguir, e pode ser aplicada especialmente às linhagens observadas em populações urbanas não-ameríndias.

Sequências do haplogrupo **A** foram determinadas em 23% das linhagens ameríndias brasileiras e apenas um haplótipo (BR13) dentre os oito identificados (TAB 10) não foi caracterizado por uma transição C→T na posição 16111. Essa mutação determina o haplogrupo **A2** ameríndio e é muito frequente em haplótipos A nativo-americanos (FORSTER e cols., 1996; SAILLARD e cols., 2000), sendo em contrapartida, raramente observada nas linhagens **A** asiáticas (HORAI E HAYASAKA, 1990; HORAI e cols. 1993; TORRONI e cols., 1993a; KOLMAN e cols., 1996). A ampla distribuição de sequências **A2** em inúmeras tribos Norte, Centro e Sul-Americanas sugere que essa

mutação teria ocorrido na(s) linhagem(s) fundadora(s) do haplogrupo **A** Ameríndio, logo após sua chegada às Américas pelo Estreito de Bering (FORSTER e cols., 1996).

A distribuição do haplogrupo **A** entre os povos nativo-americanos já analisados apresenta um claro padrão de frequência decrescente no sentido Norte-Sul do continente. Esse haplogrupo está presente em altas frequências (muitas vezes superiores a 90%) em grupos Esquimós (da Beríngia e Siberianos) e populações Nadene e Ameríndias da América do Norte (SHIELDS e cols., 1992; MERRIWETHER e cols., 1995; STARIKOVSKAYA e cols., 1998; SAILLARD e cols., 2000), enquanto pode ser encontrado em baixas frequências em algumas populações Sul-americanas ou até mesmo estar ausente em outras populações (EASTON e cols., 1996; MERRIWETHER e cols., 1995; WALLACE, 1995; LALUEZA e cols., 1997; MORAGA e cols., 2000).

Como foi descrito anteriormente, nossos primeiros resultados indicando a presença de linhagens do haplogrupo **B** ameríndio na população branca brasileira foram obtidos com a análise de 250 indivíduos, onde observou-se uma frequência de 6,4% para estes haplótipos (ALVES-SILVA e cols., 1999b; Apêndice 1). No presente trabalho nós confirmamos a prevalência de linhagens **B** que totalizaram 29% das sequências ameríndias observadas. Quanto à distribuição e frequência do haplogrupo **B** em populações nativo-americanas, alguns trabalhos mostraram que ele está ausente ou é bastante raro no extremo Norte das Américas (STARIKOVSKAYA e cols., 1998; SAILLARD e cols., 2000), mas pode ser encontrado em frequências altamente variadas desde o Canadá até populações ameríndias sul-americanas da Argentina e do Chile (LORENS e SMITH 1996, MERRIWETHER e cols., 1996; MORAGA e cols., 2000). As linhagens **B** ameríndias não apresentam nenhuma mutação específica de HVS-I que as diferencie das sequências **B** asiáticas, embora algumas pequenas variações regionais já tenham sido identificadas (MORAGA e cols., 2000). Por exemplo, a presença de transições em 16249, 16312 e 16344 associadas aos marcadores do haplogrupo **B** (16189C e 16217C) parece representar uma “assinatura” típica de um grupo de sequências ameríndias brasileiras, que foi identificado apenas em tribos da região Amazônica (SANTOS, S. e cols., 1996), amostras pré-históricas daquela mesma região (RIBEIRO-DOS-SANTOS., 1996), e em três haplótipos BR27, BR28 e linhagem “BR05” de ALVES-SILVA e colaboradores (1999b, [Apêndice 1]) da nossa amostra. É importante mencionar que no sudeste asiático esse haplogrupo é encontrado em frequências muito altas correspondendo, por exemplo, a 100% das linhagens mitocondriais em várias populações da Polinésia. Entretanto, as linhagens **B** daquelas populações apresentam mutações específicas reconhecidas como “motivo polinésio”, que as diferenciam de outros haplótipos **B**, e essas mutações não foram observadas nas linhagens brasileiras.

Haplótipos ameríndios dos grupos **C** e **D** podem ser facilmente identificados e diferenciados de seus homólogos asiáticos pela transição T→C na posição 16325 do mtDNA. Essa mutação é praticamente diagnóstica de linhagens nativo-americanas (assim como aquela que caracteriza o

haplogrupo **A2**) e está ausente, ou pode ser raramente encontrada em haplótipos asiáticos (FORSTER e cols., 1996).

Todas as 12 linhagens do haplogrupo **C** (29% dos haplótipos ameríndios) que nós identificamos no presente trabalho foram definidas pela transição em 16325 confirmando sua classificação como linhagens ameríndias. Diferentemente do que é observado para o haplogrupo **A**, a distribuição das linhagens **C** nos povos nativo-americanos atuais parece apresentar um padrão crescente no sentido norte→sul, e suas maiores frequências foram observadas em populações sul-americanas (EASTON e cols., 1996; LORENZ e SMITH, 1996., MORAGA e cols., 2000). Para a maioria das nossas linhagens não foi possível encontrar na literatura haplótipos com sequências idênticas de HVS-I, embora as mutações 16292T e 16362C (presentes no haplótipo BR43 identificado em indivíduos das regiões Norte e Nordeste) já foram descritas anteriormente em amostras da região amazônica (SANTOS,S. e cols., 1996). Uma das amostras descritas por Santos e colaboradores (1996) (haplótipo S18/R13) apresentou todas as mutações de HVS-I características do haplogrupo **C**, mas foi entretanto classificada como haplótipo **B** naquele trabalho, por apresentar a deleção de 9-pb. A análise da região hipervariável II pode ser bastante informativa na identificação de sequências **C** ameríndias uma vez que nessa região as linhagens **C** apresentam deleções nas posições 248, e 286-287 (GINTHER e cols., 1993; KOLMAN & BERMINGHAM 1997; ALVES-SILVA e cols., 2000 [Apêndice 3]; MORAGA e cols., 2000) que não foram encontradas em linhagens asiáticas (LEE e cols., 1997).

O haplogrupo **D** apresenta uma ampla distribuição de Norte a Sul em populações nativo-americanas, mas nas populações da América Central ele está ausente ou pode ser encontrado em baixas frequências (MERRIWETHER e cols., 1996). Sua divisão nos subgrupos **D1** e **D2** foi proposta por Forster e colaboradores (1996) a partir de polimorfismos específicos de região controle. O haplogrupo **D1** é caracterizado pela transição em 16325 e parece ser específico de populações ameríndias, enquanto o haplogrupo **D2** é caracterizado por uma transição T→C na posição 16271 e pode ser encontrado nas populações Nadene da América do Norte. Em um estudo realizado recentemente em populações Chilenas (MORAGA e cols., 2000), os autores demonstraram que a maioria das linhagens **D1** identificadas naquelas populações formavam um grupo definido por transições em 16187 e 16189. A mutação em 16187 foi observada também em indivíduos da tribo argentina Mapuche (GINTHER e cols., 1993) e por isso Moraga e colaboradores sugeriram que ela seja utilizada como um marcador de linhagens **D1** provenientes de populações do sul da América do Sul. Por outro lado, haplótipos com o motivo 16187-16189 parecem ter uma distribuição maior do que aquela proposta por esses autores, e podem ser encontrados também em ameríndios da Região Amazônica (SANTOS,S. e cols., 1996). Nenhum dos 6 haplótipos **D** identificados em nossas amostras apresentou essas duas mutações e 5 deles foram classificados como haplótipos **D1**, o que confirmou a sua classificação como linhagens ameríndias. A ausência da mutação em 16325 na

linhagem BR52 poderia indicar uma possível origem asiática para esse haplótipo (identificado na Região Norte). No entanto, é mais provável que essa linhagem tenha sofrido uma mutação reversa na posição 16325, uma vez que várias outras linhagens **D** sem essa mesma transição já foram descritas na literatura em populações ameríndias (RICKARDS e cols., 1999; ALVES-SILVA e cols., 2000 [Apêndice 3] GREEN e cols., 2000).

Finalmente, a última estratégia que nós utilizamos para evidenciar a origem ameríndia e não asiática dos haplótipos brasileiros pertencentes aos haplogrupos **A**, **B**, **C** e **D** consistiu na busca de sequências idênticas a essas linhagens (ou que apresentavam alta homologia), entre os haplótipos ameríndios já descritos na literatura. Essa identificação foi possível para várias das nossas linhagens que foram também encontradas em populações ameríndias atuais e/ou pré-colombianas, principalmente nas tribos da região Amazônica (GINTHER e cols., 1993; WARD e cols., 1993; SANTOS, S. e cols., 1996; RIBEIRO-DOS-SANTOS e cols., 1996). É importante mencionar que nós não fomos capazes de detectar na nossa amostra a contribuição de linhagens asiáticas que com certeza estão integradas na população brasileira. Isto pode ser evidenciado pela ausência de linhagens dos haplogrupos **M**, **F**, **G** (e outros) que são muito frequentes no Japão e na China por exemplo, de onde nós sabemos vieram numerosas migrações para o Brasil principalmente durante este último século.

V.3.2 – HAPLÓTIPOS AFRICANOS NA POPULAÇÃO BRANCA BRASILEIRA

As populações africanas foram os primeiros grupos humanos a serem analisados quanto à variabilidade do DNA mitocondrial (DENARO e cols., 1981; SCOZZARI e cols., 1988, 1994; CANN e cols., 1987; VIGILANT, 1990; VIGILANT e cols., 1991). As primeiras análises de sequências de HVS-I africanas foram apresentadas por Vigilant e colaboradores (1991), enquanto outros pesquisadores se concentraram principalmente em identificar padrões de RFLP (utilizando poucas enzimas de restrição), que também foram capazes de refletir a imensa diversidade genética que caracteriza aquelas populações (SCOZZARI e cols., 1988; CANN e cols., 1987). As primeiras tentativas de se agrupar linhagens através de mutações comuns de Região Controle e/ou RFLP (determinando assim os haplogrupos africano-específicos) ocorreu somente há pouco mais de 5 anos com a definição dos “haplogrupos” L1, L2 e algumas subdivisões do “haplogrupo L3” (CHEN e cols., 1995; GRAVEN e cols., 1995; SOODYALL e cols., 1996). Uma descrição mais detalhada dos polimorfismos de HVS-I característicos de linhagens africanas, bem como a reestruturação de alguns haplogrupos foi proposta por Watson e colaboradores (1997) em um trabalho que descreve ainda alguns aspectos da distribuição desses haplogrupos na África. Estudos mais recentes apresentaram uma análise bastante refinada de haplogrupos característicos de populações da África ocidental (RANDO e cols., 1998), do leste africano (QUINTANA-MURCI e cols., 1999) e sul-africanas (CHEN e cols., 2000).

Entretanto, a diversidade do mtDNA na África ainda está muito pobremente caracterizada, quando comparada aos estudos realizados em populações nativo-americanas e européias. Além disso, técnicas diferentes (sequenciamento e/ou RFLP) foram utilizadas em diferentes níveis de resolução (CHEN e cols., 1995), o que torna difícil a comparação dos dados disponíveis na literatura para distintas populações. Outro ponto a ser considerado é o excesso de amostras do Oeste africano em relação a outras regiões, como por exemplo, o sudoeste do continente (região de Angola, em particular). Assim, a origem dos haplótipos africanos encontrados na população branca brasileira foi determinada de acordo com os padrões de classificação propostos por Watson e colaboradores (1997) e Rando e colaboradores (1998, 1999).

O haplogrupo africano mais frequente observado no nosso estudo foi o haplogrupo **L3e**, que correspondeu a 29% das linhagens africanas identificadas. Até o presente esse haplogrupo não havia sido bem caracterizado na literatura. Ele foi descrito em várias populações nativas da África onde o polimorfismo + 2349 *Mbol* foi observado (CHEN e cols., 1995, 2000; RANDO e cols, 1998), mas não foi identificado nos estudos que não utilizaram análises por RFLP. As linhagens africanas brasileiras constituem o maior banco de dados de sequências do haplogrupo **L3e** já descrito na literatura, e por isso nós decidimos realizar novas análises com o objetivo de melhor caracterizar a filogeografia desse haplogrupo. Para isso, nós estabelecemos uma colaboração com outros grupos de pesquisa, e estendemos nossas análises a linhagens de várias populações diferentes. Uma análise dos polimorfismos de HVS-I de 42 haplótipos **L3e** descritos na literatura, incluindo os haplótipos brasileiros (SOODYALL, 1993; RANDO e cols., 1998; ALVES-SILVA e cols., 1999b; 2000), demonstrou que praticamente todos os haplótipos **L3e** podiam ser distribuídos em 4 sub-grupos principais definidos por mutações (sempre associadas à transição 16223T) em: (i) 16327; (ii) 16320; (iii) 16264 (transições); e (iv) 16265 (transversão A→T). Esses quatro subgrupos foram denominados, respectivamente, de haplogrupos **L3e1**, **L3e2**, **L3e4** e **L3e3**. Observando esse padrão de mutações foi possível identificar na literatura 126 outras sequências **L3e** (em potencial) que foram também utilizadas em nossas análises filogenéticas (TAB 16; FIG 15).

A partir dos resultados de análises de RFLP de alta resolução obtidos por Chen e colaboradores (1995, 2000), e Torroni e colaboradores (comunicação pessoal dos dados da análise de 17 amostras de São Domingos), nós identificamos 8 sítios de restrição que se apresentavam polimórficos nas linhagens **L3e**. Uma análise desses sítios (5260 *Avall*, 5584 *Alul*, 9253 *HaeIII*, 9553 *HaeIII*, 13100 *MspI*, 13803 *Alul*, 14869 *Mbol*, e 15812 *RsaI*) em amostras dos 4 subgrupos propostos, nos mostrou que 5 deles podiam ser utilizados para subdividir o haplogrupo **L3e** nos subgrupos **L3e2**, **L3e3** e **L3e4** (FIG 16).

Tabela 16: CÓDIGO DAS AMOSTRAS UTILIZADAS NAS ANÁLISES DO HAPLOGRUPO L3E

Código dos haplótipos	Origem	Referência
AFB	Afro-brasileiro	Bortolini e cols. (1997 ^a)
BAL	Berberes da Argélia	Côrte-Real e cols. (1996)
BAM	Bambara	Rando e cols. (1998)
BMO	Berberes de Marrocos	Rando e cols. (1998)
AMBR	“Ameríndio” brasileiro	Horai e cols. (1993)
BR90-BR106	Brasileiro	Alves-Silva e cols. (2000),
BR	Brasileiro com deleção de 9pb	Alves-Silva e cols. (1999b),
BUB	Bubi	Mateu e cols. (1997)
CAV	Cabo Verde	Bandelt e cols., (submetido)
DAM	Dama	Soodyall (1993)
DIO	Diola	Rando e cols. (1998)
EGI	Egípcio	Krings e cols. (1999b)
FRA	Norte da França	Richards e cols. (2000)
FUL	Fulbe	Watson e cols. (1997)
HAU	Hausa	Watson e cols. (1997)
HER	Herero	Vigilant (1990), Vigilant e cols. (1991), Soodyall (1993)
KAN	Kanuri	Watson e cols. (1997)
KIK	Kikuyu	Watson e cols. (1997)
!KU	(Sekele, Vasikela) !Kung	Soodyall (1993), Chen e cols. (2000)
KWE	Barakwena/Khwe	Soodyall (1993), Chen e cols. (2000)
MAN	Mandenka	Graven e cols. (1995)
MEX	Mexicano	Green e cols. (2000)
PAL	Palestino	Richards e cols. (2000)
POR	Português	Bandelt e cols., (submetido)
SÃO	São Tomé	Mateu e cols. (1997)
SDO	Santo Domingo (Rep. Dominicana)	Bandelt e cols., (submetido)
SER	Serer	Rando e cols. (1998)
SOT	Sotho	Soodyall (1993)
SUD	Sudanês do Sul	Krings e cols. (1999b)
SYR	Sírio	Richards e cols. (2000)
TUA	Tuareg	Watson e cols. (1997)
WOL	Wolof	Rando e cols. (1998)
YEM	Yemenites	Richards e cols. (2000)
YOR	Yoruba	Vigilant (1990), Vigilant e cols. (1991), Watson e cols. (1997)
ZUL	Zulu	Soodyall (1993)

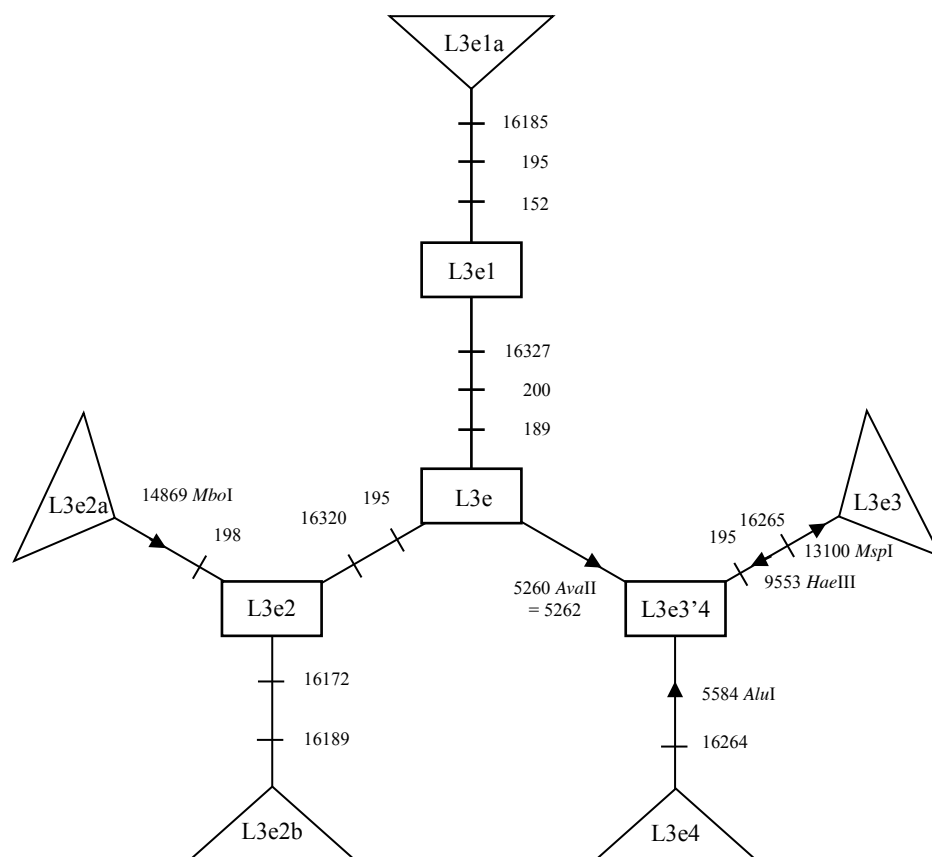


Figura 16: DIAGRAMA REPRESENTATIVO DAS MUTAÇÕES QUE DEFINEM OS PRINCIPAIS SUBGRUPOS DO HAPLOGRUPO L3E. Esta filogenia foi estabelecida por posições polimórficas de HVS-I e II, e sítios restrição na região codificadora do mtDNA. A raiz mais provável do haplogrupo **L3e** (com transições em 16223, 73, 150 e 263 relativas a CRS) e as linhagens ancestrais dos subgrupos **L3e1**, **L3e2**, e **L3e3'4** estão indicadas por retângulos, enquanto os tipos ancestrais dos grupos **L3e1a**, **L3e2a**, **L3e2b**, **L3e3** e **L3e4** estão indicados por triângulos. A ordem das mutações que diferenciam os grupos é aleatória.

Todas as análises realizadas para caracterizar o haplogrupo **L3e** estão relacionadas no manuscrito apresentado no Apêndice 4. Neste trabalho nós avaliamos a distribuição do haplogrupo **L3e** (e seus subgrupos) na África, e verificamos que: (i) o subgrupo **L3e1** (aparentemente mais velho) ocorre principalmente no sul e leste africanos; (ii) o grupo **L3e2** prevalece nos grupos Bantos de toda a África, e (iii) o haplogrupo **L3e4** está essencialmente restrito ao Oeste africano. Análises de tempo de divergência foram realizadas para as linhagens **L3e** (e seus subgrupos) e nós estimamos que esse haplogrupo surgiu possivelmente em uma região da África central há aproximadamente 46.000 (+/- 14.000) anos atrás. Dessa forma, analisando apenas as linhagens brasileiras pertencentes a um único haplogrupo, nós fomos capazes de identificar a contribuição genética de várias partes da África para a nossa população “branca”.

É importante ressaltar que nós identificamos linhagens **L3e** semelhantes ou idênticas aos nossos haplótipos, no trabalho descrito por Bortolini e colaboradores (1997a), e da Silva Jr (1999) para populações brasileiras remanescentes de antigos quilombos.

O segundo haplogrupo africano mais frequente nas nossas análises, haplogrupo **L2**, está amplamente distribuído por toda África e é um dos principais daquele continente (CHEN e cols., 1995, 2000; RANDO e cols., 1998). Ele foi descrito praticamente em todas as populações africanas analisadas (CHEN e cols., 1995, 2000; GRAVEN e cols., 1995; WATSON e cols., 1997; RANDO e cols., 1998; KRINGS e cols., 1999b), e a divisão de suas linhagens em 3 sub-haplogrupos, denominados de **L2a**, **L2b** e **L2c**, foi proposta recentemente por Chen e colaboradores (2000). Esses autores se basearam em análise de RFLP de alta resolução para confirmar, o que já havia sido observado anteriormente (WATSON e cols., 1997; RANDO e cols., 1998), que algumas linhagens desses novos haplogrupos apresentam uma distribuição bastante intrínseca e podem ser correlacionadas com regiões africanas específicas, ou até mesmo grupos populacionais. Por exemplo, as linhagens classificadas como haplogrupo **L2c** (polimorfismo – 322 *HaeIII*), parecem ser diagnósticas de populações da África ocidental (Senegal principalmente), enquanto algumas sequências do grupo **L2a** (definidas por + 11776 *RsaI* e – 13065 *DdeI*) parecem ser exclusivas de populações de pigmeus “Mbuti” da África Central. Infelizmente os polimorfismos de HVS-I característicos desses subgrupos não foram claramente definidos naquele estudo e por isso os haplótipos brasileiros não puderam ser classificados dentro desses subgrupos. As regiões africanas de onde provavelmente se originaram algumas linhagens **L2** brasileiras foram, portanto, determinadas através da comparação a sequências idênticas descritas na literatura. Sequências idênticas (ou semelhantes) aos haplótipos BR74, BR76 e BR77 são muito comuns na África e podem ser encontrados desde a Mauritânia (populações Tuareg) até o Sudão (WATSON e cols., 1997; KRINGS e cols., 1999b). Análises de RFLP adicionais foram realizadas nessas amostras e elas foram identificadas como haplótipos **L2a** (+ 13803 *HaeIII*). Nós não identificamos na literatura nenhuma linhagem com o mesmo padrão de substituições de HVS-I da amostra BR75, enquanto um haplótipo

que difere em apenas uma mutação (na posição em 16266) do haplótipo BR82 foi descrito no Senegal entre os Mandenka (WATSON e cols., 1996, 1997). Uma linhagem homóloga a BR83 (diferença em 16354) foi descrita na Costa do Marfim (HORAI e HAYASAKA, 1990).

Os haplogrupos **L1a** e **L1c** estavam igualmente representados na nossa amostra e foram encontrados em 4 indivíduos cada um. É consenso na literatura que essas linhagens, em conjunto com as do haplogrupo **L1b** são as linhagens filogeneticamente mais próximas da raiz do mtDNA (QUINTANA-MURCI e cols., 1999; CHEN e cols., 2000). O haplogrupo **L1a** pode ser encontrado em toda a África, principalmente na região do Sub-Saara, mas atinge suas maiores frequências no centro, sul e leste africanos (WATSON e cols., 1997, CHEN e cols., 2000). Linhagens **L1a** correspondem por exemplo, a 79% dos haplótipos observados na população dos Vasikela Kung da África do Sul e um subgrupo dessas linhagens, caracterizado pelo polimorfismo de deleção de 9 pb, é característico de populações sul-africanas (SOODYALL e cols., 1996; WATSON e cols., 1997; CHEN e col., 2000). Duas das 3 linhagens **L1a** que nós identificamos já foram descritas na literatura: o haplótipo BR55 é frequente no Senegal (Mandenka), mas já foi descrito também no Sudão (Turkana) (WATSON e cols., 1997), enquanto o haplótipo BR57 foi descrito na região do centro-leste africano em povos Kikuyo do Kênia.

O haplogrupo **L1c** apresenta uma distribuição mais limitada nas populações africanas tendo sido descrito como característico do centro e oeste daquele continente. Nós não encontramos na literatura nenhuma sequência idêntica aos nossos haplótipos **L1c**, e de certa forma isso pode sugerir uma origem no Centro-sul da África, na Região de Angola por exemplo, de onde se sabe saiu grande parte do escravos trazidos para o Brasil, e para a qual praticamente não existem estudos.

Os haplogrupos **L1b** (identificado em 6% das nossas linhagens africanas) e **L3d** (11%) foram descritos como específicos da África Ocidental (WATSON e cols., 1997; RANDO e cols., 1998) onde esses dois haplogrupos apresentam uma distribuição geográfica semelhante. Em conjunto as linhagens **L1b** e **L3d** estão presentes em cerca de 25% da população do Senegal e 17% dos povos Kanuri, Songhai, Hausa e Ioruba que habitam pontos diferentes das margens do Rio Niger. É possível que as linhagens **L1b** e **L3d** sejam de fato originadas desses povos, uma vez que dados históricos registram que africanos desta região geográfica foram trazidos como escravos predominantemente para a Bahia, onde a religião Ioruba originou o Candomblé.

Haplótipos **L1b** idênticos à linhagem BR59 são muito frequentes na região do Senegal e Mauritânia, ao passo que para o haplótipo BR60 não foi possível identificar nos bancos de dados uma sequência de HVS-I idêntica àquela que define essa linhagem. Duas sequências que diferem desse haplótipo em apenas uma mutação foram observadas no Senegal (RANDO e cols., 1998).

Os haplotipos **L3d** brasileiros provavelmente também se originaram no Oeste africano embora não tenha sido possível identificar na literatura linhagens idênticas aos haplótipos BR86, BR87, BR88 e BR90. Uma sequência de HVS-I que difere em apenas uma mutação do haplótipo BR88 (em

16086) foi descrito em um indivíduo Tuareg da Mauritânia (RANDO e cols., 1998). É importante notar entretanto que linhagens **L3d** foram descritas recentemente também no Sul da África (CHEN e cols., 2000).

Finalmente, o último haplogrupo africano a ser identificado nas amostras brasileiras, haplogrupo **U6**, foi descrito por Rando e colaboradores (1998) e Macaulay e colaboradores (1999) como específico do Norte e Noroeste da África. Linhagens desse grupo já haviam sido descritas anteriormente em povos africanos, mas sob a classificação de grupo “**L3c**” (WATSON e cols., 1997). Esse haplogrupo está presente em 28% dos indivíduos em populações de Berberes da Argélia e em 20% na Mauritânia. Na literatura não foram encontrados haplótipos idênticos à linhagem brasileira BR111, embora algumas sequências de HVS-I com alta homologia tenham sido descritas em Marrocos (RANDO e cols., 1998). Este haplogrupo foi descrito também na Península Ibérica, em frequências de 7% na Espanha (CÔRTE-REAL e cols., 1996) e 6% em Portugal (CÔRTE-REAL e cols., 1996; PEREIRA e cols., 2000) onde pode representar uma herança genética da ocupação bárbara que ocorreu por mais de 7 séculos nessa região. Haplótipos **U6** também penetraram o Saara e podem ser encontrados esporadicamente em populações do oeste (Senegal) e até mesmo no leste africano (Kênia). Uma vez que esse haplogrupo está presente em populações do tipo Wolof e Mandenka no Senegal nós consideramos mais plausível que a linhagem BR111 tenha vindo dessa região (RANDO e cols., 1998).

V.3.3 – HAPLÓTIPOS EUROPEUS NA POPULAÇÃO BRANCA BRASILEIRA

As extensivas análises que vêm sendo desenvolvidas em várias populações da Europa, fazem com que este seja o continente mais amostrado e melhor representado nos bancos de dados de mtDNA, de todo o mundo. Entretanto, algumas regiões na Europa central (p. ex.: Polônia), ocidental (p. ex.: França) e do sudeste europeu ainda não foram analisadas (METSPALU e cols., 1999). Os estudos se estendem também à região do Cáucaso, extremo oeste asiático e Oriente Médio que juntamente com a Europa constituem a “Eurásia Ocidental” (MACAULAY e cols., 1999). Os primeiros trabalhos que correlacionaram polimorfismos de HVS-I e análises de RFLP numa tentativa de se caracterizar haplogrupos de mtDNA específicos de populações européias foram apresentados por Torroni e colaboradores (1994c, 1996), onde os principais haplogrupos europeus (**H, J, K, T, U, V**) e outros menos frequentes (**I, W, X**) foram descritos. A partir daí, a filogenia do DNA mitocondrial vem sendo exaustivamente analisada na Europa, o que possibilitou uma grande resolução da origem e evolução desses haplogrupos que respondem por pelo menos 95% das linhagens mitocondriais em todas as populações européias e também algumas do Oriente Médio (MACAULAY e cols., 1999; METSPALU e cols., 1999; RICHARDS e cols., 2000). Richards e colaboradores (2000) apresentaram recentemente uma extensa análise de linhagens européias e do Oriente Médio, numa tentativa de

caracterizar a origem geográfica e o tempo de divergência dos diversos haplogrupos europeus dentro das várias migrações que ocorreram do Oriente Médio para a Europa durante os períodos Paleolítico Superior e Neolítico (tempo compreendido entre 50.000 e 5.000 anos atrás).

Os principais haplogrupos europeus foram encontrados em praticamente todas as populações européias analisadas e, de certa forma, eles se apresentam homogeneamente distribuídos pelo continente (RICHARDS e cols., 1996, 1998; SIMONI e cols., 2000). Por outro lado, alguns haplogrupos podem se apresentar muito frequentes em determinadas regiões da Europa, estando entretanto ausentes em outras regiões (RICHARDS, e cols 1998; TORRONI e cols., 1998). É importante ressaltar que certas populações européias como os Saami (norte da Escandinávia), os Ladinos (região dos Alpes) e os Bascos (norte da Espanha) são caracterizados por uma distribuição de linhagens mitocondriais muito diferente do restante da Europa (CÔRTE-REAL e cols., 1996; TORRONI e cols., 1998, RICHARDS e cols., 1998; SIMONI e cols., 2000). Além disso, alguns sub-haplogrupos podem ser diagnósticos de determinadas regiões européias, como por exemplo o grupo de linhagens **U5b1** que é encontrado predominantemente em populações do norte da Europa e entre os Saami (SAJANTILLA e cols., 1996; MACAULAY e cols., 1999). Outros haplogrupos menos frequentes, como os haplogrupos **X**, **I**, **W** e algumas subdivisões do haplogrupo **U**, também parecem apresentar um discreto padrão de frequência decrescente no sentido leste-oeste.

O haplogrupo mais frequente em praticamente todas as populações européias é o haplogrupo **H** que ocorre em média em 46% dos europeus (TORRONI e cols., 1996, 1998, RICHARDS e cols., 1996, 1998, 2000; KIVISILD e cols., 1999b; METSPALU e cols., 1999; HELGASON e cols., 2000). Este haplogrupo é comum em quase todas as populações analisadas por RFLP (particularmente nas populações que também apresentam o haplogrupo **V**), e sua frequência na Europa apresenta um gradiente no sentido leste → oeste. Foi observado que suas frequências mais altas (40%-60%) são encontradas no Oeste e Norte da Europa; frequências intermediárias (20%-40%) são encontradas no Sul da Espanha, Norte da África, Itália central, leste da Europa, Turquia e Cáucaso; e frequências menores que 20% são encontradas no Oriente Próximo, Índia e Sibéria central (TORRONI e cols., 1998). Em Portugal, esse haplogrupo foi observado em uma frequência média de 40% (CÔRTE-REAL e cols., 1996; PEREIRA e cols., 2000), que é muito semelhante à frequência encontrada no Brasil (43%). Como esse haplogrupo é bastante distribuído e frequente em toda a Europa, fica difícil tentar estabelecer a origem geográfica precisa das linhagens **H** observadas na nossa amostra de brasileiros.

Entre as linhagens do haplogrupo **H**, o haplótipo BR112 (que apresenta a sequência de HVS-I idêntica à CRS) foi observada em 13 indivíduos brasileiros e pode ser encontrada em todas as populações da Europa em frequências médias de 16%. Em Portugal essa sequência foi descrita em 26% (CÔRTE-REAL e cols., 1996), e 18% dos indivíduos analisados (PEREIRA e cols., 2000). Cabe aqui ressaltar que sequências de HVS-I idênticas à CRS já foram descritas também em linhagens do

haplogrupo **U** e por isso são chamadas de haplótipos U-CRS, em distinção às linhagens H-CRS (KIVISILD e cols., 1999b; RICHARDS e cols., 2000). Isso reforça a necessidade da análise de polimorfismos de RFLP ou da posição 73 da HVS-II, para uma correta diferenciação desses dois tipos de haplótipos. O estudo realizado por Côrte-Real e colaboradores (1996) constitui um exemplo da importância dessas análises, uma vez que pelo menos um haplótipo U-CRS foi erroneamente classificado entre as linhagens do haplogrupo **H** (grupo 1 daquele estudo). Além disso, os outros haplótipos **U** que aparentemente eram derivados dessa linhagem permaneceram como "linhagens não classificadas".

Entre os outros 14 haplótipos **H** brasileiros, oito (BR113, BR117, BR118, BR122, BR125, BR127, BR128 e BR129) já foram também observados em indivíduos da Península Ibérica, principalmente em Portugal (CÔRTE-REAL e cols., 1996; PEREIRA e cols., 2000), e isso sugere que a maioria das linhagens **H** encontradas no Brasil sejam provenientes dessa região. Entretanto, alguns destes haplótipos (BR117, BR118, BR125, BR128 e BR129) são comuns também em outros países da Europa como Suíça, Alemanha, Reino Unido, Bélgica e Finlândia (PULT e cols., 1994; RICHARDS e cols., 1996, 1998), o que ilustra a dificuldade de se estabelecer uma origem geográfica regional para as linhagens de mtDNA européias. Por outro lado, sequências de HVS-I idênticas ou homólogas às dos haplótipos BR114 e BR121 parecem estar restritas a regiões mais ao norte e oeste da Europa, e foram identificadas apenas na Inglaterra, Bélgica, e norte da Alemanha (RICHARDS e cols., 1998).

Linhagens do haplogrupo **V** foram encontradas em dois indivíduos brasileiros e este é um haplogrupo que se apresenta em frequências altamente heterogêneas pela Europa (TORRONI e cols., 1998). Ele corresponde a 4% dos mtDNAs europeus em geral e suas maiores frequências são encontradas entre os Saami (40,9%), na Catalunha (26,7%) e no País Basco (20%) (TORRONI e cols., 1998; LAHERMO e cols., 1996; IZAGUIRRE e DE LA RÚA, 1999). Análises de diversidade de sequências de linhagens do haplogrupo **V** em diferentes partes da Europa e do Oriente Médio sugerem que a Península Ibérica, onde ele pode ser encontrado em média em 10,6% dos indivíduos, seja o local onde esse haplogrupo se originou por volta de 13.000 anos atrás (TORRONI e cols. 1998). Em Portugal, as linhagens **V** foram observadas em frequências que variam entre 3,7% (CÔRTE-REAL e cols., 1996) e 8% dos indivíduos (PEREIRA e cols., 2000). Nenhuma sequência de HVS-I idêntica aos haplótipos brasileiros BR135 e BR136 foi observada nos indivíduos portugueses. Essas duas linhagens já foram entretanto observadas no norte da África, na Algeria e em Marrocos (CÔRTE-REAL e cols., 1996). Além disso, BR135 foi também identificada na Alemanha, Finlândia e entre os Saami (SAJANTILA e cols., 1995; RICHARDS e cols., 1996; TORRONI e cols., 1998). Portanto, embora não seja possível descartar a origem Ibérica das linhagens **V** no Brasil, é possível que esse haplogrupo tenha chegado aqui através de outros imigrantes europeus.

Considerando que cerca de 30% dos imigrantes europeus (incluindo os portugueses) que vieram para o Brasil eram provenientes da Itália (IBGE 2000, SALZANO e FREIRE MAIA, 1967;

CALLEGARI-JACQUES e SALZANO, 1999), nós poderíamos esperar uma considerável contribuição de linhagens mitocondriais Italianas para a população branca brasileira. De fato, foi possível confirmar a origem italiana para uma linhagem **V** com sequência de HVS-I idêntica à do haplótipo BR136, em um indivíduo da região Sudeste (SANTOS 1998; Apêndice 3). Infelizmente os estudos que tentaram avaliar a variabilidade do mtDNA nas populações da Itália (FRANCALACCI e cols., 1996; STENICO e cols., 1996, 1998; TORRONI e cols., 1996;) não foram capazes de detectar linhagens diagnósticas, ou perfis de haplogrupos específicos dessa região, com os quais seria possível avaliar a contribuição de mtDNAs italianos para a nossa amostra da população brasileira.

O haplogrupo **U** foi descrito inicialmente, por análises de RFLP, como específico de populações da européias (TORRONI e cols., 1996), mas as análises mais recentes demonstraram que ele apresenta uma alta diversidade e pode ser dividido em linhagens específicas da Europa, Oriente Médio ou até mesmo do norte africano (RANDO e cols., 1998; RICHARDS e cols., 1998; 1999; KIVISILD e cols., 1999b; MACAULAY e cols., 1999; RICHARDS e cols., 2000). Vários autores propuseram a divisão das linhagens **U** em 8 grupos (**U1-U7** e **K**), e observaram que alguns desses grupos apresentam ainda suas próprias subdivisões (RANDO e cols., 1998; RICHARDS e cols., 1998; MACAULAY e cols., 1999). A origem do macro-haplogrupo **U** parece ter ocorrido há mais de 50.000 anos atrás no Oriente Médio e aparentemente essas foram as únicas linhagens levadas para a Europa, durante a primeira migração de humanos modernos para esse continente, ainda no início do período Paleolítico Superior (RICHARDS e cols., 2000).

Linhagens do haplogrupo **U2** podem ser encontradas em toda a Europa e Oriente Médio (MACAULAY e cols., 1999). Uma linhagem **U2** com HVS-I idêntica ao haplótipo BR139 foi observado no Norte da Espanha (CÔRTE-REAL cols. 1996). O haplogrupo **U4** ocorre principalmente no Sul da Europa. Sequências **U4** de HVS-I idênticas à do haplótipo BR140 podem ser frequentemente encontradas em Portugal (PEREIRA e cols., 2000). O haplogrupo **U5** parece ser o único grupo das linhagens **U** que se originou em populações européias e sua presença no Oriente Médio em 2% dos mtDNAs foi atribuída a migrações mais recentes da Europa para essa região. O (sub)haplogrupo **U5b** está presente em Portugal em apenas 1% dos indivíduos analisados onde foi observado uma sequência de HVS-I idêntica à do haplótipo brasileiro BR142. Esse haplótipo é entretanto, também muito frequente em várias outras populações da Europa (RICHARDS e cols., 1998). Interessantemente, o grupo de linhagens **U5b1** parece estar restrito aos Saami na Escandinávia (onde ocorre em até 50% dos indivíduos) e regiões adjacentes (SAJANTILA e cols., 1995), o que sugere que a origem do haplótipo brasileiro BR145 seja o Norte da Europa. O haplogrupo **U7** foi descrito recentemente por Richards e colaboradores (2000) na Europa e Oriente Médio onde pode ser encontrado em baixas frequências. Ele também foi identificado ao sul de Portugal em um haplótipo definido por uma sequência de HVS-I idêntica à da linhagem BR144 (PEREIRA e cols., 2000).

O haplogrupo **K**, que de acordo com análises filogenéticas também constitui um subgrupo de **U** (MACAULAY e cols., 1999), corresponde a aproximadamente 7% das linhagens de mtDNA de toda a Europa (RICHARDS e cols., 1998). Em Portugal ele também foi observado em uma frequência média de 7%, e uma sequência de HVS-I igual à linhagem BR144 foi identificada (PEREIRA e cols., 2000). Essa linhagem é idêntica à sequência fundadora do haplogrupo **K** e está amplamente distribuída pela Europa. Outros haplótipos que se diferenciam das linhagens **K** brasileiras por apenas uma mutação também foram observados em portugueses, enquanto uma linhagem idêntica ao haplótipo BR150 foi descrita na Inglaterra (RICHARDS e cols., 1998).

O haplogrupo **J** está amplamente distribuído por toda a Europa e representa cerca de 11% dos mtDNAs europeus em geral (RICHARDS e cols., 1998). Este é um dos haplogrupos que também apresenta uma grande diversificação regional, com vários subgrupos específicos de determinadas regiões européias. (RICHARDS e cols., 1998, 2000; METSPALU e cols., 1999). As linhagens **J*** (que correspondem aos haplótipos **J** que não se classificam em nenhum dos subgrupos já descritos para o haplogrupo **J**), estão amplamente distribuídas por toda a Europa e são muito frequentes nas populações do centro e do oeste europeu. As maiores frequências deste haplogrupo foram contudo observadas entre os árabes (25%), onde este haplogrupo parece ter se fixado (RICHARDS e cols., 2000). Em Portugal as linhagens **J** foram encontradas em frequências mais baixas, em torno de 3-5% (CÔRTE-REAL e cols., 1996; PEREIRA e cols., 2000). O haplótipo BR152 mostrou-se idêntico a uma das linhagens portuguesas descritas. O grupo **J1** também foi esporadicamente observado em Portugal, mas as linhagens **J1b1** parecem ser específicas do Noroeste da Europa, principalmente da Inglaterra, onde foi identificada uma sequência de HVS-I idêntica ao haplótipo BR156 (RICHARDS e cols., 1998).

O haplogrupo **T** é um grupo “irmão” do haplogrupo **J** com o qual compartilhou uma linhagem ancestral comum (linhagem **JT**) provavelmente no Oriente Médio há mais de 50.000 anos atrás (RICHARDS e cols., 2000). Na Europa o haplogrupo **T** corresponde a cerca de 8% das linhagens de mtDNA e pode ser encontrado em praticamente todas as populações européias (RICHARDS e cols., 1996, 1998, 2000). Em um estudo recente foi proposta uma subdivisão das linhagens **T** em 5 grupos (**T1-T5**) sendo o grupo **T1** o segundo mais frequente na Europa, depois das linhagens **T***. Em Portugal o haplogrupo **T** foi encontrado em 11% dos indivíduos, e nessa população foram observadas várias linhagens com sequências de HVS-I idênticas à do haplótipo **T1** BR166 (CÔRTE-REAL e cols., 1996; PEREIRA e cols., 2000). Sequências semelhantes aos outros haplótipos **T** brasileiros também foram descritos em Portugal (BR157, BR158 e BR162), mas essas linhagens são muito frequentes em outros países mais ao norte da Europa, como Inglaterra por exemplo (PIERCY e cols., 1993; RICHARDS e cols., 1998).

Os haplogrupos **I** e **X** juntamente com o haplogrupo **W** (não identificado no nosso estudo) são os haplogrupos menos frequentes em todas as populações européias analisadas (TORRONI e cols.,

1996, RICHARDS e cols., 1998,2000; MACAULAY e cols., 1999). O haplogrupo **I** pode ser encontrado em média em 2% dos europeus e parece estar distribuído principalmente no oeste e no norte da Europa (RICHARDS e cols, 1998), sendo que uma linhagem idêntica ao haplótipo BR167 foi observada no norte da Alemanha e na Dinamarca. As análises realizadas em Portugal demonstraram que as linhagens **I** estão presentes em apenas 1% dos indivíduos e nenhuma sequência com alta homologia com o haplótipo BR167 foi observada (PEREIRA e cols., 2000). Linhagens do haplogrupo **X** também são raras na Europa e correspondem a apenas 2% dos mtDNAs europeus em geral (RICHARDS e cols., 1998). Suas frequências são um pouco maiores no leste da Europa, e ele está presente no Cáucaso em cerca de 5% dos indivíduos (KIVISILD e cols., 1999a). A maior frequência do haplogrupo **X** já observada em uma população foi descrita no norte de Israel, onde as linhagens **X** estavam presentes em 26% dos indivíduos analisados em uma população Druza. Sequências homólogas ao haplótipo **X** brasileiro BR168 (com apenas uma mutação de HVS-I diferente) foram descritas na Espanha, País Basco e Itália. Em Portugal as sequências **X** foram identificadas em 3,5 % dos indivíduos, mas as linhagens encontradas diferem do haplótipo brasileiro em pelo menos duas mutações de HVS-I, e (CÔRTE-REAL e cols., 1996; PEREIRA e cols., 2000). Assim, é provável que os haplogrupos **I** e **X** não foram trazidos para o Brasil com os colonizadores portugueses.

Linhagens **HV***, **pre-*V** e **pre-*HV** foram apenas recentemente caracterizadas e correspondem a linhagens ancestrais aos haplogrupos **H** e **V** que também foram levadas do Oriente Médio para a Europa com as diferentes ondas migratórias que colonizaram esse último continente (RICHARDS e cols., 2000). A descoberta dessas linhagens só foi possível com a realização de extensas análises de RFLP uma vez que elas não apresentam mutações específicas de HVS-I e são definidas principalmente pela ausência de polimorfismo de RFLP que caracterizam os haplogrupos mencionados acima (RICHARDS e cols., 1998, 2000; MACAULAY e cols., 1999; METSPALU e cols., 1999; KIVISILD e cols., 1999b). Isso dificulta portanto a determinação dessas linhagens nas populações européias que foram analisadas somente por sequenciamento de HVS-I. Linhagens **HV*** foram identificadas no sul e leste da Europa, e no Oriente Médio (MACAULAY e cols., 1999). Essas linhagens são muito frequentes na região do Cáucaso e na Índia, onde foram classificadas como haplótipos **Ö** (METSPALU e cols., 1999; KIVISILD e cols., 1999b). Linhagens **pre-*V** podem ser encontradas na Europa apenas esporadicamente (Bandelt comunicação pessoal). Finalmente, linhagens **pre-*HV** foram encontradas em 22% das indivíduos analisados nas populações árabes e são provavelmente o resultado de um grande efeito fundador nessa região (RICHARDS e cols., 2000). Um grupo dessas linhagens foi descrito em Israel (MACAULAY e cols., 1999), e pode ser identificado pelas mutações de HVS-I em 16126 e 16362. Essas linhagens parecem estar amplamente dispersas também no sul da Europa e no Oriente Médio. A linhagem brasileira BR151 apresentou as mutações de HVS-I em 16126 e 16362. Os dados descritos acima para os haplogrupos europeus estão resumidas na tabela 17, onde a frequência observada no Brasil para

esses haplogrupos é comparada com a que foi descrita em Portugal e na Europa em geral. A distribuição dessas frequências no Brasil e em Portugal, além da presença de vários haplótipos com sequências de HVS-I idênticas nas populações desses dois países parece indicar que a maioria das linhagens européias observadas na população brasileira constituem uma herança lusitana.

Por outro lado, é importante destacar que o predomínio de linhagens européias no sul do país parece ser um reflexo das imigrações de outras partes da Europa que vieram para o Brasil e chegaram principalmente nessa região. Com certeza os alemães representam a maior parte desses imigrantes e sua herança genética deve estar representada entre as linhagens de mtDNA européias que nós identificamos nas amostras da região Sul. Além disso, nesta região foi identificada a maior parte dos nossos haplótipos europeus, para os quais nenhuma sequência de HVS-I idêntica foi descrita em Portugal.

V.4 – DISTRIBUIÇÃO DE LINHAGENS AMERÍNDIAS, AFRICANAS E EUROPÉIAS NAS REGIÕES NORTE, NORDESTE E SUL DO BRASIL

Grande parte dos estudos que tentaram caracterizar a variabilidade genética dos brasileiros foram realizados em populações ameríndias nativas, principalmente da Região Amazônica (SCHURR e cols., 1990; TORRONI e cols., 1993a; BAILLIET e cols., 1994; BIANCHI e cols., 1995; BORTOLINI e SALZANO 1996; EASTON e cols., 1996; SANTOS e cols., 1996; WARD e cols., 1996; RODRIGUEZ-DELFIN e cols., 1997); ou em isolados populacionais remanescentes de antigos quilombos, formados por indivíduos descendentes de africanos (BORTOLINI e cols., 1992, 1995, 1997a, 1997b; GUERREIRO e cols., 1999; ARPINI-SAMPAIO e cols., 1999).

A variabilidade genética de brasileiros “brancos”, que correspondem aos descendentes dos imigrantes europeus (incluindo os colonizadores portugueses) foi abordada em trabalhos que se concentraram na análise de sistemas protéicos convencionais e foram realizados principalmente nas regiões Norte e Sul do país (FRANCO e cols., 1982; ROSA e cols., 1984; MORAES e cols., 1993; SANTOS, E. e cols., 1996; ARPINI-SAMPAIO e cols., 1999; uma revisão pode ser encontrada em DORNELLES e cols., 1999). Esses estudos demonstraram que as contribuições ameríndias para o *pool* genético desses brasileiros eram bastante variadas dependendo da região geográfica analisada. A maior influência de “genes” ou marcadores genéticos ameríndios foi observada na região Norte, onde estão localizadas a maior parte das populações indígenas brasileiras atuais (de acordo com a última contagem populacional realizada pelo IBGE). Um estudo publicado por Santos e Guerreiro (1995) que analisou 11 populações urbanas daquela região através de marcadores nucleares foi capaz de detectar ancestralidade ameríndia em 41% dos indivíduos. Um trabalho mais recente, que analisou linhagens de mtDNA por RFLP, também em amostras da Região Norte descreveu uma contribuição ameríndia, em 59% dos indivíduos (BATISTA-DOS-SANTOS e cols., 1999).

Tabela 17: PROPORÇÃO DOS HAPLOGRUPOS EUROPEUS NO BRASIL (E EM CADA REGIÃO), EM PORTUGAL E NA EUROPA EM GERAL

Haplogrupo	Região Norte ^a	Região Nordeste ^a	Região Sul ^a	Brasil Total	Portugal ^b	Europa total ^c
H	0,34	0,65	0,45 ^d	0,45	0,40	0,46
V	0,07 ^e	0,06	0,06 ^e	0,06	0,07	0,04
U^f	0,07	0,12	0,06	0,08 ^f	0,15 ^f	0,16 ^f
K	0,07	0,06	0,09	0,08	0,05	0,06
J	0,20	0,06	0,18 ^g	0,17 ^g	0,06 ^g	0,09 ^g
T	0,27 ^h	0,06	0,12 ^h	0,14 ^h	0,11 ^h	0,08 ^h
I	0,07	—	—	0,01	< 0,01	0,02
X	—	—	0,03	0,01	0,02	0,01

^a Essas frequências se referem apenas às linhagens européias obtidas no presente estudo;

^b Dados de PEREIRA e colaboradores (2000);

^c Dados de RICHARDS e colaboradores (2000);

^d Inclui linhagens HV* e pre-*HV;

^e Inclui linhagens pre-*V;

^f Inclui todas as subdivisões do haplogrupo U com exceção do haplogrupo U6;

^g Inclui todas as subdivisões do haplogrupo J;

^h Inclui linhagens T1;

Nossas análises para a região Norte corroboraram os resultados dos trabalhos descritos acima, uma vez que nessa região foi observada a maior frequência (54%) de linhagens ameríndias. Entretanto, as frequências de cada um dos quatro haplogrupos ameríndios que nós observamos foi diferente das apresentadas por Santos, S e colaboradores (1996) e Batista dos Santos e colaboradores (1999) para as populações indígenas daquela região. Nós observamos um predomínio de linhagens dos haplogrupos **C** e **B**, enquanto esses autores encontraram frequências maiores do haplogrupo **A**, nessas duas populações.

As análises das regiões Nordeste e Sul também demonstraram uma maior contribuição genética ameríndia do que já havia sido descrito anteriormente na literatura. Os dados publicados para a região Nordeste apontavam uma média de 13% para os marcadores de ancestralidade ameríndia (KRIEGER, e cols., 1965; FRANCO e cols., 1982; CONCEIÇÃO e cols., 1987; LIMA & AZEVEDO, 1991) enquanto nós identificamos uma contribuição quase duas vezes maior (em 22% das nossas amostras). Um amplo estudo foi realizado recentemente na região Sul, em quase 3000 indivíduos que eram predominantemente descendentes de europeus (DORNELLES e cols., 1999). Esses autores demonstraram, através da análise de 17 sistemas protéicos, que os valores máximos para a presença de marcadores ameríndios eram de 11%. Também nesta região nós fomos capazes de avaliar uma contribuição ameríndia muito maior, na qual as linhagens **A**, **B**, **C** e **D** em conjunto estavam presentes em 22% dos indivíduos.

As diferenças que nós observamos de uma região para outra na proporção dos haplogrupos ameríndios (TAB 15), podem estar retratando a frequência desses haplogrupos nas populações nativas das mesmas regiões. Não existe na literatura dados de frequência para essas linhagens nas populações nativas das regiões Nordeste e Sul, e isso dificulta a determinação da origem para as nossas linhagens.

Com relação à ancestralidade africana da população branca brasileira, os trabalhos que tentaram avaliar a presença de marcadores africanos utilizando sistemas genéticos (protéicos) clássicos (KRIEGER e cols., 1965; FRANCO e cols., 1982; LIMA & AZEVEDO, 1991; ROSA e cols., 1994; SANTOS, E. e cols., 1996; DORNELLES e cols., 1999), apresentaram resultados semelhantes para as regiões Norte e Sul, onde esses autores apontaram a presença de marcadores africanos em aproximadamente 12% das amostras para ambas as regiões (SANTOS, E. e cols., 1996; DORNELLES e cols., 1999; PROBST e cols., 2000), enquanto nós encontramos 15% de linhagens africanas nos indivíduos da região Norte e 12% para a região Sul. Na região Nordeste foi observamos a maior diferença, onde outros autores identificaram marcadores africanos em 36% dos indivíduos (FRANCO e cols., 1982; ARPINI-SAMPAIO e cols., 1999) enquanto a nossa estimativa de contribuição africana para a população chegou a 43%. Dados históricos indicam que diferentes portos do país (Salvador, Belém e Rio de Janeiro por exemplo) receberam escravos de distintas regiões e

isso pode estar determinando a diferença na composição das linhagens africanas nas três regiões analisadas (CURTIN., 1969).

Estudos de variabilidade do mtDNA já foram realizados na população de alguns países da América do Sul, Central e do Norte, entre eles, Cuba (TORRONI e cols., 1995), Uruguai (BRAVI e cols., 1997) e México (GREEN e cols., 2000). Esses trabalhos foram concentrados principalmente na avaliação da presença de linhagens ameríndias nas populações estudadas. Um trabalho descrito recentemente por Carvajal-Carmona e colaboradores (2000) analisando amostras da população de Antioquia, na Colômbia, apresentou um grande predomínio de linhagens ameríndias, presentes em 95% dos indivíduos. Esses autores compararam seus resultados aos nossos e sugeriram que as diferenças na proporção de linhagens ameríndias e européias entre essas duas populações, é devido ao fato de que a Colômbia, ao contrário do Brasil, não foi fortemente influenciada por migrações européias nos últimos séculos.

A análise de marcadores continente-específicos do cromossomo Y também foram realizadas em nosso laboratório para as amostras das regiões Norte, Nordeste e Sul, e demonstraram um grande predomínio de linhagens européias, ao lado de uma contribuição mínima de cromossomos africanos (2,5%) e ausência de linhagens ameríndias (CARVALHO-SILVA e cols., 2000). Em conjunto, os resultados obtidos com as análises do mtDNA e cromossomo Y refletem claramente os processos de mistura étnica que ocorreram desde o início da formação da população brasileira, e podem ser facilmente evidenciados por relatos históricos da literatura. É sabido que no início da colonização do Brasil, e durante um longo período, os contingentes humanos vindos de Portugal eram constituídos apenas por homens (SALZANO e FREIRE-MAIA, 1970). Ao chegarem aqui esses homens se acasalavam com as índias, tomando tantas quantas pudessem, o que resultou na primeira geração de brasileiros, denominados de mamelucos. Isso foi possível graças ao *cunhadismo* (velho costume indígena de incorporar estranhos à sua sociedade), que fazia com que os portugueses fossem automaticamente aceitos como membros da tribo, ao desposarem as índias. Com o início do tráfico de escravos para o Brasil, por volta do ano 1550, teve início a transferência de linhagens mitocondriais africanas para a população "brasileira" nascente, principalmente através da miscigenação entre mulheres africanas e, novamente, homens portugueses. Esses processos ocorreram intensamente no Brasil durante os três primeiros séculos de colonização, e parecem ser os responsáveis pelo predomínio das linhagens africanas e ameríndias (que em conjunto foram encontradas em 56% dos indivíduos analisados) na atual população do país. Casamentos inter-étnicos foram até mesmo encorajados oficialmente durante o período colonial como estratégia de povoamento do país (MÖRNER, 1967; CRUZ, 1973). É importante ressaltar entretanto, que a herança mitocondrial das inúmeras imigrações européias que ocorreram principalmente durante os dois últimos séculos, pôde também ser resgatada na nossa análise da população branca brasileira.

Concluindo, nossos dados de mtDNA sugerem que populações ameríndias e africanas são responsáveis por uma contribuição genética muito maior para os “brancos” brasileiros do que havia sido descrito anteriormente na literatura. Nossa amostra revelou grandes contribuições de linhagens africanas (24%) e ameríndias (32%), que na verdade podem estar refletindo apenas valores subestimados, uma vez que os indivíduos aqui analisados pertencem principalmente às classes média e média alta da sociedade, onde nós poderíamos esperar uma predominância de linhagens mitocondriais de origem européia.

Referências Bibliográficas

- ADELMAN, R.; SAUL, R.L.; AMES, B.N. (1988) Oxidative damage to DNA: relation to species metabolic rate and life span. *Proc Natl Acad Sci U.S.A.* **85**:2706-2708.
- ALBERTS, B.; DENNIS, B.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WATSON, J.D. Molecular Biology of the Cell. New York: Garland Publishing Inc, 1983.
- ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403-410.
- ALVES-SILVA, J. (1997) Caracterização do polimorfismo de deleção de 9 pb do mtDNA na população caucasóide brasileira. Dissertação de mestrado. Departamento de Bioquímica e Imunologia, ICB-UFMG. Belo Horizonte.
- ALVES-SILVA, J.; GUIMARAES, P.E.; ROCHA, J.; PENA, S.D.; PRADO, V.F.; (1999a) Identification in Portugal and Brazil of a mtDNA lineage containing a 9-bp triplication of the intergenic COII/tRNALys region. *Hum Hered.* **49**:56-58.
- ALVES-SILVA, J.; SANTOS, M.S.; PENA, S.D.; PRADO, V.F.; (1999b) Multiple geographic sources of region V 9-bp deletion haplotypes in Brazilians. *Hum Biol.* **71**(2):245-59.
- ALVES-SILVA, J.; SANTOS, M.S.; GUIMARAES, P.E.; FERREIRA, A.C.S.; BANDELT, H.-J.; PENA, S.D.; PRADO, V.F. (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* **67**:444-461.
- ANDERSON, S.; BANKIER, A.T.; BARRELL, B.G.; DE BRUIJN, M.H.L.; COULSON, A.R.; DROUIN, J.; EPERON, I.C.; NIERLICH, D.P.; ROSE, B.A.; SANGER, F.; SCHREIER, P.H.; SMITH, A.J.H.; STADEN, R.; YOUNG, I.G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**:457-465.
- ANDREWS, R.M.; KUBACKA, I.; CHINNERY P.F.; LIGHTOWLERS, R.N.; TURNBULL, D.M.; HOWELL, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147.
- ANKEL-SIMONS F., CUMMINS J.M. (1996) Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution. *Proc Natl Acad Sci USA* **93**(24): 13859-13863.
- AQUADRO, C.F.; GREENBERG, B.D. (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics* **103**(2):287-312.
- ARPINI-SAMPAIO, Z.; COSTA, M.C.; MELO, A.A.; CARVALHO, M.F.; DEUS, M.S.; SIMÕES, A.L. (1999). Genetic polymorphisms and ethnic admixture in African-derived black communities of northeastern Brazil. *Hum Biol* **71**:69-85
- AWADALLA, P.; EYRE-WALKER, A.; SMITH, J.M. (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* **24**:2524-2525.
- BAHN, P.G. (1998). Neanderthals emancipated. *Nature* **394**: 719-721.

- BAILLIET, G.; ROTHHAMMER, F.; CARNESE, F.R.; BRAVI, C.M.; BIANCHI, N.O. (1994) Founder mitochondrial haplotypes in Amerindian populations. *Am J Hum Genet* **55**(1):27-33.
- BALLINGER, S.W.; SCHURR, T.G.; TORRONI, A.; GAN, Y.Y.; HODGE, J.O.; HASSAN, K.; SHEN, K.H.; WALLACE D.C. (1992). Southeast Asian Mitochondrial DNA analyses reveals genetic continuity of ancient mongoloid migrations. *Genetics* **130**: 139-152.
- BANDELT, H.J.; FORSTER, P.; SYKES, B.C.; RICHARDS, M.B. (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141**(2):743-753.
- BANDELT, H.-J.; ALVES-SILVA, J.; GUIMARÃES, P.E.M.; SANTOS, M.S.; BREHM, A.; PEREIRA, L.; COPPA, A.; LARRUGA, J.M.; RENGÓ, C.; SCOZZARI, R.; TORRONI, A.; PRATA, A.; AMORIM, A.; PRADO, V.F.; PENA, S.D.J. Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Artigo submetido ao periódico *Amer J Hum Genet*
- BANDY, B.; DAVISON, A.J.; (1990) Mitochondrial mutations may increase oxidative stress: implications for carcinogenesis and aging? *Free Radic Biol Med* **8**(6):523-539.
- BATISTA, O.; KOLMAN, C.J.; BERMINGHAM, E. (1995) Mitochondrial DNA diversity in the Kuna Amerinds of Panama. *Hum Mol Genet* **4**(5):921-929.
- BATISTA DOS SANTOS, S.E.; RODRIGUES, J.D.; RIBEIRO-DOS-SANTOS, A.K.C.; ZAGO, M.A. (1999) Differential contribution of indigenous men and women to the formation of an urban population in the Amazon region as revealed by mtDNA and Y- DNA. *Am J Phys Anthropol* **109**:175-180.
- BENDALL KE, SYKES BC. (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet* **57**:248-56.
- BETHELL, L. (1997) Nota sobre as populações americanas às vésperas das invasões européias. In: Bethel L (ed) *América Latina Colonial*. Editora da Universidade de São Paulo, São Paulo, pp 129-131.
- BIANCHI, N.O.; BAILLIET, G. (1993) Powers and Pitfalls of the use of mitochondrial DNA in anthropology. *J Genet Mol Biol* **4**: 33-45
- BIANCHI, N.O.; BAILLIET, G.; BRAVI, C.M. (1995) Peopling of the Americas as Inferred Through the Analysis of Mitochondrial DNA. *Brazilian Journal of Genetics*. **18**, 4: 661-668.
- BIANCHI, N.O.; BAILLIET, G. (1997) Further comments on the characterization of founder Amerindian mitochondrial haplotypes. *Am J Hum Genet*. **61**:244-247.
- BONATTO, S.L.; REDD, A.J.; SALZANO, F.M.; STONEKING, M. (1996) Lack of ancient Polynesian-Amerindian contact. *Am J Hum Genet* **59**(1):253-258.
- BONATTO, S.L.; SALZANO, F.M. (1997a) A single and Early Migration for the Peopling of the Americas Supported by Mitochondrial DNA sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **94**(5):1866-1871.

- BONATTO, S.L.; SALZANO, F.M. (1997b) Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *Am J Hum Genet* **61**(6):1413-1423.
- BONEN, L. (1991) The Mitochondrial Genome: So Simple yet so Complex. *Current Opinion in Genetics and Development*. **1**: 515-522.
- BONNÉ-TAMIR, B.; JOHNSON, M.J.; NATALI, A.; WALLACE, D.C.; CAVALLI-SFORZA, L.L. (1986) Human mitochondrial DNA types in two Israeli populations-a comparative study at the DNA level. *Am. J. Hum. Genet.* **38**:341-351.
- BOOM, R.; SOL, C.J., SALIMANS, M.M., JANSEN, C.L., WERTHEIM-VAN DILLEN, P.M., VAN DER NOORDAA, J. (1990) Rapid and simple method for purification of nucleic acids. *J Clin Microbiol* **28**(3):495-503.
- BORTOLINI MC, WEIMER TA, FRANCO MH, SALZANO FM, LAYRISSE Z, SCHNEIDER H, SCHNEIDER MP, BONATTO SL (1992) Genetic studies in three South American black populations. *Gene Geogr* **6**:1-16.
- BORTOLINI MC, WEIMER TDA, SALZANO FM, CALLEGARI-JACQUES SM, SCHNEIDER H, LAYRISSE Z, BONATTO SL (1995) Evolutionary relationships between black South American and African populations. *Hum Biol* **67**:547-559.
- BORTOLINI, M.C.; SALZANO, F.M. (1996) mtDNA diversity analysis in Amerindians and other human populations - how different are they? *Braz J Genet* **19**:527-534.
- BORTOLINI, M. C.; ZAGO, M. A., SALZANO, F. M.; SILVA-JUNIOR, W. A.; BONATTO, S. L.; DA SILVA, M. C.; WEIMER, T. A. (1997a) Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. *Hum Biol* **69**(2):141-159.
- BORTOLINI, M.C.; SALZANO, F.M.; ZAGO, M.A.; DA SILVA JUNIOR, W.A.; WEIMER, T.A. (1997b) Genetic variability in two Brazilian ethnic groups: a comparison of mitochondrial and protein data. *Am J Phys Anthropol* **103**(2):147-156.
- BORTOLINI, M.C.; BAPTISTA, C.; CALLEGARI-JACQUES, S.M.; WEIMER, T.A.; SALZANO, F.M. (1998a) Diversity in protein, nuclear DNA, and mtDNA in South Amerinds - agreement or discrepancy? *Ann Hum Genet* **62**:133-145.
- BORTOLINI, M.C.; DA SILVA-JUNIOR, W.A.; WEIMER, T.A.; ZAGO, M.A.; DE GUERRA, D.C.; SCHNEIDER, M.P.; LAYRISSE, Z.; CASTELLANO, H.M.; SALZANO, F.M. (1998b). Protein and hypervariable tandem repeat diversity in eight African-derived South American populations: inferred relationships do not coincide. *Hum Biol* **70**:443-461.
- BRAVI, C.M.; SANS, M.; BAILLIET, G.; MARTINEZ-MARIGNAC, V.L.; PORTAS, M.; BARRETO, I.; BONILLA, C.; BIANCHI, N.O. (1997) Characterization of mitochondrial DNA and Y-chromosome haplotypes in a Uruguayan population of African ancestry. *Hum Biol* **69**:641-652.
- BROWN, W.M.; GEORGE, M.JR.; WILSON, A.C. (1979) Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. sci. USA* **76**: 1967-1971.

- BROWN, M.D.; HOSSEINI, S.H.; TORRONI, A.; BANDELT, H.J.; ALLEN, J.C.; SCHURR, T.G.; SCOZZARI, R.; CRUCIANI, F.; WALLACE, D.C. (1998) mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *Am J Hum Genet* **63**:1852-1861.
- BURCKHARDT F, VON HAESELER A, MEYER S. (1999) HvrBase: compilation of mtDNA control region sequences from primates. *Nucleic Acids Res* **27**(1):138-142.
- CALLEGARI-JACQUES, S.M.; SALZANO, F.M. (1999) Brazilian Indian/non-Indian interactions and their effects. *Ciência e Cultura* **51**:166-174
- CANN R. L., STONEKING, M.; WILSON A. C.; (1987) Mitochondrial DNA and Human Evolution. *Nature* **325**: 31-36.
- CARNEIRO DA CUNHA, M. (1992) Introdução a uma história indígena. Em: História dos índios no Brasil. M Carneiro da Cunha ed, Companhia das Letras, São Paulo.
- CARVALHO-SILVA, D.R.; SANTOS, F.R.; ROCHA, J.; PENA, S.D. (2000) The Phylogeography of Brazilian Y-Chromosome Lineages. *Am J Hum Genet* **68**:000-000 (publicado eletronicamente)
- CARVAJAL-CARMONA, L.G.; SOTO, I.D.; PINEDA, N.; ORTIZ-BARRIENTOS, D.; DUQUE, C.; OSPINA-DUQUE, J.; MCCARTHY, M.; MONTROYA, P.; ALVAREZ, V.M.; BEDOYA, G.; RUIZ-LINARES, A. (2000) Strong Amerind/White sex bias and a possible sephardic contribution among the founders of a population in northwest colombia. *Am J Hum Genet* **67**:1287-95.
- CASE, J.T.; WALLACE, D.C. (1981) Maternal inheritance of mitochondrial DNA polymorphisms in cultured human fibroblasts. *Somatic Cell Genet* **7**(1):103-108.
- CAVALLI-SFORZA, L.L. (1998) The DNA revolution in population genetics. *Trends Genet* **14**(2):60-65.
- CHAKRABORTY, R.; WEISS, K.M. (1991) Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium. *Am J Phys Anthropol* **86**(4):497-506.
- CHAKRAVARTI, A. (1999) Population Genetics — making sense out of sequence. *Nature Genet suppl* **21**: 56-60.
- CHEN, Y.S.; TORRONI, A; EXCOFFIER, L.; SANTACHIARA-BENERECETTI, A.,S.; WALLACE, D.C. (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* **57**:133-149.
- CHEN, Y.S., OLCKERS, A., SCHURR, T.G., KOGELNIK, A.M., HUOPONEN, K. & WALLACE, D.C. (2000) mtDNA variation in the South African Kung and Khwe- and their genetic relationships to other African Populations. *Am J Hum Genet* **66**:1362-1383.
- CLAYTON, D.A. (1992) Transcription and replication of animal mitochondrial DNAs. *Int Rev Cytol* **141**:217-232.
- COHEN, S. N.; CHANG, A. C. Y.; HSU, L. (1972) Nonchromosomal antibiotic resistance in bacteria: Genetic transformation of *Escherichia coli* by R factor. *Proc. Natl. Acad. Sci. U.S.A.* 69, 2110.

- COMAS, D.; CALAFELL, F.; MATEU, E.; PÉREZ-LEZAUN, A.; BERTRANPETI, J. (1996) Geographic Variation in Human Mitochondrial DNA Control Region Sequence: The Population History of Turkey and its Relationship to the European Populations. *Mol. Biol. Evol.* **13**(8): 1067-1077.
- COMAS, D.; CALAFELL, F.; MATEU, E.; PEREZ-LEZAUN, A.; BOSCH, E.; MARTINEZ-ARIAS, R.; CLARIMON, J.; FACCHINI, F.; FIORI, G.; LUISELLI, D.; PETTENER, D.; BERTRANPETIT, J. (1998) Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet.* **63**:1824-38.
- CONCEIÇÃO, M.M.; SALZANO, F.M.; FRANCO, M.H.L.P.; WEIMER, T.A.; KRIEGER, H. (1987) Demography, genetics, and race admixture in Aracaju, Brazil. *Braz J Genet* **10**:313-331.
- CORTE-REAL, H. B.; MACAULAY, V. A.; RICHARDS, M. B.; HARITI, G.; ISSAD, M. S.; CAMBON-THOMSEN, A.; PAPIHA, S.; BERTRANPETIT, J.; SYKES, B. C. (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* **60**:331-350.
- CRUZ, E. (1973) História do Pará. Belém. Governo do Estado do Pará.
- CURTIN, P.D. (1969) The Atlantic slave trade : a census. University of Wisconsin Press, Madison.
- DA SILVA Jr (1999). Diversidade Genética de Populações Afro-Brasileiras. Um estudo antropogenético de populações remanescentes de Quilombos, com base em marcadores genéticos do DNA nuclear e mitocondrial. Tese de doutorado; FMRP-USP; Ribeirão Preto.
- DENARO M, BLANC H, JOHNSON MJ, CHEN KH, WILMSEN E, CAVALLI-SFORZA LL, WALLACE DC. (1981) Ethnic variation in Hpa 1 endonuclease cleavage patterns of human mitochondrial DNA. *Proc Natl Acad Sci U S A.* **78**(9):5768-5772.
- DI RIENZO, A.; WILSON, A.C. (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* **88**:1597-1601.
- DORNELLES, C.L.; CALLEGARI-JACQUES, S.M.; ROBINSON, W.M.; WEIMER, T.A.; FRANCO, M.H.L.P.; HICKMANN, A.C.; GEIGER, C.J. (1999) Genetics, surnames, grandparents' nationalities, and ethnic admixture in Southern Brazil: do the patterns of variation coincide? *Genet Mol Biol* **22**:151-161.
- EASTON, R.D.; MERRIWETHER, D.A.; CREWS, D.E; FERRELL, R.E. (1996) mtDNA variation in the Yanomami: evidence for additional New World founding lineages. *Am J Hum Genet* **59**:213-225.
- FORSTER, P.; HARDING, R.; TORRONI, A.; BANDELT, H. J. (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* **59**(4):935-945.
- FRANCALACCI, P.; BERTRANPETIT, J.; CALAFELL, F.; UNDERHILL P.A. (1996) Sequence Diversity of the Control Region of Mitochondrial DNA in Tuscany and its Implications for the Peopling of Europe. *Am J of Physical Anthropol.* **100**: 443-460.
- FRANCALACCI, P.; MONTIEL, R.; MALGOSA, A. (1999) A mitochondrial DNA database – Applications to Problems of Nomenclature and Population Genetics. Em: Papiha D, and

- Chakraborty (eds) Genomic diversity: applications in human population genetics. Kluwer Academic/Plenum Publishers, New York, pp 103-119
- FRANCO, M.H.; WEIMER, T.A.; SALZANO, F.M. (1982) Blood polymorphisms and racial admixture in two Brazilian populations. *Am J Phys Anthropol* **58**:127-132.
- GILES, R.E.; BLANC, H.; CANN, H.M.; WALLACE, D.C. (1980) Inheritance of human mitochondrial DNA. *Proc Natl. Acad. Sci. USA* **77**: 6715-6719.
- GILL, P.; LYGO, J.E.; FOWLER, S.J.; WERRET, D.J. (1987) An evaluation of DNA fingerprinting for forensic purposes. *Electrophoresis* **8**: 38-44.
- GILL P, IVANOV PL, KIMPTON C, PIERCY R, BENSON N, TULLY G, EVETT I, HAGELBERG E, SULLIVAN K. (1994) Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* **6**(2):130-135.
- GINTHER, C.; CORACH, D.; PENACINO, G.A.; REY, J.A.; CARNESE, F.R.; HUTZ, M.H.; ANDERSON, A.; JUST, J.; SALZANO, F.M.; KING, M.C. (1993) Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. Em Pena S. D. J.; Chakraborty, R.; Epplen, J.T.; Jeffreys, A. J. (1995) DNA fingerprint: state of the Science. Birkhauser, Basel, 211-219.
- GRAVEN L, PASSARINO G, SEMINO O, BOURSOT P, SANTACHIARA-BENERECETTI S, LANGANEY A, EXCOFFIER, L. (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* **12**:334-345
- GREEN, L.D.; DERR, J.N.; KNIGH, A. (2000) mtDNA affinities of the peoples of north-central Mexico. *Am J Hum Genet* **66**:989-998.
- GREENBERG, J.H.; TURNER C.G.; ZEGURA, S.L.; (1986) The settlement of the Americas: A comparison of the linguistic, dental and genetic evidence. *Curr Anthropol* **27**:477-487.
- GREENBERG, J.H. (1987) Language in the Americas. Palo Alto, CA: Stanford University Press.
- GUERREIRO, J.F.; RIBEIRO-DOS-SANTOS, A.K.C.; SANTOS, E.J.M.; VALLINOTO, A.C.R.; CAYRES-VALLINOTO, I.M.V.; AGUIAR, G.F.S.; SANTOS, S.E.B. (1999) Genetical-demographic data from two Amazonian populations composed of descendants of African slaves: Pacoval and Curiau. *Genet Mol Biol* **22**:163-167.
- GYLLENSTEN, U.; WHARTON, D.; JOSEFFSON, A.; WILSON, A.C. (1991) Paternal Inheritance of Mitochondrial DNA in Mice. *Nature* **352**: 255-257.
- HAGELBERG, E.; CLEGG, J.B. (1991) Isolation and characterization of DNA from archaeological bone. *Proc R Soc Lond B Biol Sci* **244**(1309):45-50.
- HAGELBERG, E., SYKES, B.; HEDGES, R. (1989) Ancient bone DNA amplified. *Nature* **342**(6249):485.

- HAGELBERG, E.; GRAY, I.C.; JEFFREYS, A.J. (1991a) Identification of the skeletal remains of a murder victim by DNA analysis. *Nature* 352(6334):427-429.
- HAGELBERG, E.; BELL, L.S.; ALLEN, T.; BOYDE, A.; JONES, S.J.; CLEGG, J.B. (1991b) Analysis of ancient bone DNA: techniques and applications. *Philos Trans R Soc Lond B Biol Sci.* **333**(1268):399-407.
- HAGELBERG E, THOMAS MG, COOK CE JR, SHER AV, BARYSHNIKOV GF, LISTER AM. (1994a) DNA from ancient mammoth bones. *Nature* **4**:333-334.
- HAGELBERG, E.; QUEVEDO, S.; TURBON, D.; CLEGG, J.B. (1994b) DNA from ancient Easter Islanders. *Nature* **369**(6475):25-26.
- HAGELBERG, E.; GOLDMAN, N.; LIO, P.; WHELAN, S.; SCHIEFENHOVEL, W.; CLEGG, J.B.; BOWDEN, D.K. (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc R Soc Lond B Biol Sci.* **7**:485-492.
- HANDT, O.; KRINGS, M.; WARD, R.H.; PÄÄBO S. (1996) The retrieval of ancient human DNA sequences. *Am J Hum Genet* **59**(2):368-76.
- HARRIS, E.E.; HEY, J. (1999) X chromosome evidence for ancient human histories. *Proc Natl Acad Sci U S A.* **16**;96(6):3320-3324.
- HASEGAWA M, DI RIENZO A, KOCHER TD, WILSON AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. *J Mol Evol* **37**:347-354.
- HAUSWIRT, W. W.; DICKEL, C.D.; ROWOLD, D.J.; HAWSWIRT, N.A. (1994). Inter- and intra population studies of ancient humans. *Experientia* **50**: 585-591.
- HEIDRICH, E.M.; HUTZ, M.H.; SALZANO, F.M.; COIMBRA, C.E. JR.; SANTOS, R.V. (1996) D1S80 locus variability in three Brazilian ethnic groups. *Hum Biol* **67**:311-9.
- HELGASON, A.; SIGURARDÓTTIR, S.; GULCHER, J.R.; WARD, R.; STEFÁNSSON, K. (2000) mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* **66**:999-1016.
- HEY, J. (1998) Population genetics and human origins--haplotypes are key! *Trends Genet* **14**(8):303-305.
- HIGUCHI, R.; VON BEROLDINGEN, C.H.; SENSABAUGH, G.F.; ERLICH, H.A. (1988) DNA typing from single hairs. *Nature* **332**(6164):543-546.
- HOFMANN, S.; JAKSCH, M.; BEZOLD, R.; MERTENS, S.; AHOLT, S.; PAPROTTA, A.; GERBITZ, K.D. (1997) Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. *Hum Mol Genet* **6**:1835-1846.
- HORAI, S.; HAYASAKA, K. (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.* **46**:828-842.

- HORAI, S.; KONDO, R.; MURAYAMA, K.; HAYASHI, S.; KOIKE, H.; NAKAI, K. (1991) Phylogenetic affiliation of ancient and contemporary humans inferred from mitochondrial DNA. *Philos. Trans. R. Soc. Lond. [Biol.]* **333**:409-417.
- HORAI, S.; KONDO, R.; NAKAGAWA-HATTORI, Y.; HAYASHI, S.; SONODA, S.; TAJIMA, K. (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol. Biol. Evol.* **10**:23-47.
- HORAI, S.; HAYASAKA, K.; KONDO, R.; TSUGANE, K.; TAKAHATA, N. (1995) Recent African Origin of Modern Humans Revealed by Complete Sequences of Hominoid Mitochondrial DNAs. *Proc. Natl. Acad. Sci. USA* **92**: 532-536.
- HORAI, S.; MURAYAMA, K.; HAYASAKA, K.; MATSUBAYASHI, S.N.; HATTORI, Y.; FUCHAROEN, G.; HARIHARA, S.; PARK, K.S.; OMOTO, K.; PAN, I.H. (1996) mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *Am J Hum Genet* **59**(3):579-590.
- HÖSS, M. & PÄÄBO, S. (1993). DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Res* **21**(16):3913-3914.
- HÖSS, M.; PÄÄBO, S.; VERESHCHAGIN, N.K. (1994) Mammoth DNA sequences. *Nature*. **4** 370(6488):333.
- HUTZ, M.H.; MATTEVI, V.S.; CALLEGARI-JACQUES, S.M.; SALZANO, F.M.; COIMBRA JUNIOR, C.E.; SANTOS, R.V.; CARNESE, R.F.; GOICOECHEA, A.S.; DEJEAN, C.B. (1997) D1S80 locus variability in South American Indians. *Ann Hum Biol.* **24**:249-55.
- IVANOV, P.L.; WADHAMS, M.J.; ROBY, R.K.; HOLLAND, M.M.; WEEDN, V.W.; PARSONS, T.J. (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet* **12**:417-420.
- IZAGIRRE N & DE LA RUA C. (1999) An mtDNA analysis in ancient Basque populations: implications for haplogroup V as a marker for a major paleolithic expansion from southwestern europe. *Am J Hum Genet* **65**(1):199-207.
- JORDE, L.B.; BAMSHAD, M.J.; WATKINS, W.S.; ZENGER, R.; FRALEY, A.E.; KRAKOWIAK, P.A.; CARPENTER, K.D.; SOODYALL, H.; JENKINS, T.; ROGERS, A.R. (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am J Hum Genet* **57**(3):523-538.
- JORDE, L.B.; WATKINS, W.S.; BAMSHAD, M.J.; DIXON, M.E.; RICKER, C.E.; SEIELSTAD, M.T.; BATZER, M.A. (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet.* **66**:979-988.
- KAESSMANN, H.; HEISSI, F.; VON HAESELER, A.; PAABO, S. (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat Genet.* **22**:78-81.

- KANEDA, H.; HAYASHI, J.I.; TAKAHAMA, H.S.; TAYA, C.; LINDAHL, K.F.; YONEKAWA, H. (1995) Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. *Proc. Natl. Acad. Sci. USA* **92**: 4542-4546.
- KENYON, L.; MORAES, C.T. (1997) Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids. *Proc Natl Acad Sci U S A* **94**(17):9131-9135.
- KIVISILD, T.; BAMSHAD, M.J.; KALDMA, K.; METSPALU, M.; METSPALU, E.; REIDLA, M.; LAOS, S.; PARIK, J.; WATKINS, W.S.; DIXON, M.E.; MASTANA, S.S.; MIR, M.R.; FERAK, V.; VILLEMS, R. (1999a) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* **9**:1331-1334.
- KIVISILD, T.; KALDMA, K.; METSPALU, E.; PARIK, J.; PAPIHA, S.S.; VILLEMS, R. (1999b) The place of the Indian mitochondrial DNA variation in the global network of maternal lineages and peopling of the old world. Em: Papiha D, and Chakraborty (eds) *Genomic diversity: applications in human population genetics*. Kluwer Academic/Plenum Publishers, New York, pp 135-152.
- KOLMAN, C.J.; BERMINGHAM, E.; COOKE, R.; WARD, R.H.; ARIAS, T.D.; GUIONNEAU-SINCLAIR, F. (1995) Reduced mtDNA diversity in the Ngobe Amerinds of Panama. *Genetics* **140**(1):275-283.
- KOLMAN, C.J.; SAMBUUGHIN, N.; BERMINGHAM, E. (1996) Mitochondrial DNA Analysis of Mongolian Populations and Implications for the Origin of the New World Founders. *Genetics*. **142**: 1321-1334.
- KOLMAN CJ & BERMINGHAM E.(1997) Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. *Genetics* **147**(3):1289-1302.
- KONDO, R.; SATTA, Y.; MATSUURA, E.T.; ISHIWA, H.; TAKAHATA, N.; CHIGUSA, S.I. (1990) Incomplete Maternal Transmission in Mitochondrial DNA in *Drosophila*. *Genetics* **126**: 657-653.
- KRIEGER, H.; MORTON, N.E.; MI, M.P.; AZEVEDO, E.; FREIRE-MAIA, A.; YASUDA, N. (1965) Racial admixture in north-eastern Brazil. *Ann Hum Genet* **29**:113-125.
- KRINGS, M.; STONE, A.; SCHMITS, R. W.; KRAINITZKI, H.; STONIEKING, M.; PÄÄBO, S. (1997) Neandertal DNA Sequences and the Origin of Modern Humans. *Cell* **90**: 19-30.
- KRINGS, M.; GEISERT, H.; SCHMITZ, R.W.; KRAINITZKI, H.; PÄÄBO, S. (1999a) DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proc Natl Acad Sci USA* **96**:5581-5585.
- KRINGS, M.; SALEM, A.E.; BAUER, K.; GEISERT, H.; MALEK, A.K.; CHAIX, L.; SIMON, C.; WELSBY, D.; DI RIENZO, A.; UTERMANN, G.; SAJANTILA, A.; PAABO, S.; STONEKING, M. (1999b) mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am J Hum Genet*. **64**:1166-76.
- KUMAR. S.S.; SUBRAMANIAN. V.; WALIMBE, S.R.; SINGH, L. (1999). Current trends in 'ancient DNA studies' – A review. *Current Science*, **76**:879-885.

- LAHERMO P, SAJANTILA A, SISTONEN P, LUKKA M, AULA P, PELTONEN L, SAVONTAUS ML. (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. *Am J Hum Genet* **58**(6):1309-1322.
- LALUEZA C, PEREZ-PEREZ A, PRATS E, CORNUDELLA L, TURBON D. (1997) Lack of founding Amerindian mitochondrial DNA lineages in extinct aborigines from Tierra del Fuego-Patagonia. *Hum Mol Genet* **6**(1):41-46.
- LEE SD, SHIN CH, KIM KB, LEE YS, LEE JB. (1997) Sequence variation of mitochondrial DNA control region in koreans. *Forensic Sci Int.* **87**(2):99-116.
- LIGHTOWLERS, R.N.; CHINNERY, P.F.; TURNBULL, D.M.; HOWELL, N. (1997) Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends Genet* **13**(11):450-455.
- LIMA AM & AZEVEDO ES (1991) Factor VIII:C, ABO blood groups, and black admixture in a Brazilian sample. *Hum Biol.* **63**(1):77-83.
- LINDAHL, T. (1997) Facts and artifacts of ancient DNA. *Cell* **90**(1):1-3.
- LORENZ JG & SMITH DG. (1996) Distribution of four founding mtDNA haplogroups among Native North Americans. *Am J Phys Anthropol.* **101**(3):307-323.
- MACAULAY, V.; RICHARDS, M.; HICKEY, E.; VEGA, E.; CRUCIANI, F.; GUIDA, V.; SCOZZARI, R.; BONNE-TAMIR, B.; SYKES, B.; TORRONI, A. (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control- region sequences and RFLPs. *Am J Hum Genet* **64**:232-249.
- MATEU, E.; COMAS, D.; CALAFELL, F.; PEREZ-LEZAUN, A.; ABADE, A.; BERTRANPETIT, J. (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann Hum Genet* **61**:507-518.
- MELLARS, P. (1998) The fate of the Neanderthals. *Nature* **395**: 539-540
- MERRIWETHER, D. A.; ROTHHAMMER, F.; FERRELL, R.E. (1994) Genetic variation in the New World: ancient teeth bone and tissue as sources of DNA. *Experientia* **50**: 592-601.
- MERRIWETHER, D.A.; ROTHHAMMER, F.; FERRELL, R.E. (1995) Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. *Am J Phys Anthropol* **98**(4):411-430.
- MERRIWETHER, D.A.; HALL, W.W.; VAHLNE, A.; FERRELL, R.E. (1996) mtDNA variation indicates Mongolia may have been the source for the founding population for the New World. *Am J Hum Genet* **59**(1):204-212.
- MERRIWETHER, D.A.; HUSTON, S.; IYENGAR, S.; HAMMAN, R.; NORRIS, J.M.; SHETTERLY, S.M.; KAMBOH, M.I.; FERRELL, R.E. (1997) Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. *Am J Phys Anthropol* **102**:153-159.
- MEYER, S.; WEISS, G.; VON HAESELER, A. (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* **152**:1103-1110.

- METSPALU, E.; KIVISILD, T.; KALDMA, K.; PARIK, J.; REIDLA, M.; TAMBETS, K.; VILLEMS, R. (1999) The trans-Caucasus and the expansion of the caucasoid-specific human mitochondrial DNA. Em: Papiha D, and Chakraborty (eds) Genomic diversity: applications in human population genetics. Kluwer Academic/Plenum Publishers, New York, pp 121-133
- MONSALVE, M.V.; GROOT DE RESTREPO, H.; ESPINEL, A.; CORREAL, G.; DEVINE, D.V. (1994) Evidence of mitochondrial DNA diversity in South American aboriginals. *Ann. Hum. Genet.* **58**:265-273.
- MONSALVE, M.V.; CARDENAS, F.; GUHL, F.; DELANEY, A.D.; DEVINE, D.V.; (1996) Phylogenetic analysis of mtDNA lineages in South American mummies. *Ann Hum Genet* 60:293-303
- MONTEIRO, J.M. (1994). Negros da terra, índios e bandeirantes nas origens de São Paulo. Companhia da Letras, São Paulo
- MORAES, M.E.; FERNANDEZ-VINA, M.; SALATIEL, I.; TSAI, S.; MORAES, J.R.; STASTNY, P. (1993) HLA class II DNA typing in two Brazilian populations. *Tissue Antigens* **41**:238-242.
- MORAGA ML, ROCCO P, MIQUEL JF, NERVI F, LLOP E, CHAKRABORTY R, ROTHHAMMER F, CARVALLO P. (2000) Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. *Am J Phys Anthropol.* **113**:19-29.
- MORITZ, C.; DOWLING, T.E.; BRWON, W. (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev.Ecol. Syst.* **18**: 269-292.
- MÖRNER, M. (1967) Race mixture in the history of Latin America. Little, Brown, Boston.
- MORRIS, A.A., LIGHTOWLERS, R.N. (2000) Can paternal mtDNA be inherited? *Lancet.* **15**:1290-1291.
- MUNFORD, D.; ZANINI, M.C.; NEVES, W.A. (1995) Human cranial variation in South America; Implications for the settlement of the New World. *Braz J Genet* **18**:673-688.
- NECKELMANN, N.; LI, K.; WADE, R.P.; SHUSTER, R.; WALLACE, D.C. (1987) cDNA sequence of a human skeletal muscle ADP/ATP translocator: lack of a leader peptide, divergence from a fibroblast translocator cDNA, and coevolution with mitochondrial DNA genes. *Proc Natl Acad Sci U S A* **84**(21):7580-7584.
- NEI, M. (1995) Genetic Support for the out-of-Africa theory of Human Evolution. *Proc Natl Acad Sci U S A* **92**(21):6720-6722.
- NEVES, W.A.; MEYER, D.; PUCCIARELLI, H.M (1996a) Early skeletal remains and the peopling of the Americas. *Revista Anthropol* **392**:121-139.
- NEVES, W.A.; MUNFORD, D.; ZANINI, M.C. (1996b) Cranial morphological variation and the colonization of the New World: Towards a four migration model. *Am J Phys Anthropol Suppl* **22**:176.

-
- NEVES, W.A.; ZANINI, M.C.; MUNFORD, D.; PUCCIARELLI, H.M. (1997) O povoamento da América à luz da morfologia craniana. *Rev USP* **34**:96-105.
- NEVES, W.A.; POWELL, J.F.; PROUS, A.; OZOLINS, E.G. (1998) Lapa Vermelha IV, Hominid: Morphological affinities of the earliest known American. *Am J Phys Anthropol Suppl* B26: 169
- NEVES, W.A.; POWELL, J.F.; OZOLINS, E.G. (1999) Modern human origins as seen from the peripheries. *J Hum Evol* **37**(1):129-133.
- OHNO, S. (1997) The One Ancestor per Generation Rule and Three Other Rules of Mitochondrial Inheritance. *Proc. Natl. Acad. Sci. USA*. **94** 8033-8035.
- OJALA, D.; MONTOYA, J.; ATTARDI, G. (1981a) tRNA punctuation model of RNA processing in human mitochondria. *Nature* **290**(5806):470-474.
- OJALA, D.; CREWS, S.; MONTOYA, J.; GELFAND, R.; ATTARDI, G. (1981b) A small polyadenylated RNA (7 S RNA), containing a putative ribosome attachment site, maps near the origin of human mitochondrial DNA replication. *J Mol Biol* **150**(2):303-314.
- OOTA, H.; SAITOU, N.; MATSUSHITA, T.; UEDA, S. (1999) Molecular genetic analysis of remains of a 2,000-year-old human population in China-and its relevance for the origin of the modern Japanese population. *Am J Hum Genet*. **64**:250-258.
- OVCHINNIKOV, I.V.; GOTHERSTROM, A.; ROMANOVA, G.P.; KHARITONOV, V.M.; LIDEN, K.; GOODWIN, W. (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* **30**:490-493.
- PÄÄBO, S. (1989). Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci* **86**:1939-1943
- PÄÄBO, S.; HIGUCHI, R.G.; WILSON, A.C. (1989) Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *J Biol Chem* 264(17):9709-9712.
- PÄÄBO, S.; IRWIN, D.M.; WILSON, A.C. (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem* **265**(8):4718-4721.
- PÄÄBO, S. (1993) Ancient DNA. *Sci Am* **269**(5):86-92.
- PÄÄBO, S. (1996) Mutation Hots Pots in the Mitochondrial Microcosm. *Am. J. Hum. Genet*. **59**: 493-396.
- PALMER, D.J. (1997) The mitochondrion that time forgot. *Nature*. **387**: 454-455.
- PARSONS TJ, MUNIEC DS, SULLIVAN K, WOODYATT N, ALLISTON-GREINER R, WILSON MR, BERRY DL, HOLLAND KA, WEEDN VW, GILL P, HOLLAND MM. (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* **15**:363-8.
- PASSARINO G, SEMINO O, MODIANO G, SANTACHIARA-BENERECETTI AS (1993) COII/tRNA(Lys) intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have Oriental affinities. *Am J Hum Genet* **53**(3):609-18.
-

-
- PASSARINO, G.; SEMINO, O.; BERNINI, L.F.; SANTACHIARA-BENERECETTI, A.S. (1996a) Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. *Am J Hum Genet* **59**(4):927-934.
- PASSARINO, G.; SEMINO, O.; MODIANO, G.; BERNINI, L.F.; SANTACHIARA BENERECETTI, A.S. (1996b) mtDNA provides the first known marker distinguishing proto-Indians from the other Caucasoids; it probably predates the diversification between Indians and Orientals. *Ann Hum Biol Mar* **23**(2):121-126.
- PENNY, D.; STEEL, M.; WADDELL, P.J.; HENDY, M.D. (1995) Improved Analyses of Human mtDNA Sequences Support a Recent African Origin for *Homo sapiens*. *Mol. Biol. Evol.* **12**(5): 863-882.
- PEREIRA, P.; PRATA, M.J.; AMORIM, A. (2000) Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann Hum Genet* (no prelo).
- PICKFORD, M. (1991) Paradise Lost: Mitochondrial Eve Refuted. *Human Evolution*. **6**(3): 263-268.
- PIERCY R, SULLIVAN KM, BENSON N, GILL P. (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med*. 106(2):85-90.
- POINAR, H.N.; HOSS, M.; BADA, J.L.; PAABO, S. (1996) Amino acid racemization and the preservation of ancient DNA. *Science* **272**(5263):864-866.
- POINAR, H.N.; STANKIEWICZ, B.A. (1999) Protein preservation and DNA retrieval from ancient tissues. *Proc Natl Acad Sci U S A* **96**(15):8426-8431.
- POWELL, J.F.; NEVES, W.A. (1999) Craniofacial Morphology of the First Americans: Pattern and Process in the Peopling of the New World. *Yearbook of Phys. Anthropol.* **42**:153-188.
- PRADO, V.F.; CASTRO, A.K.; OLIVEIRA, C.L.; SOUZA, K.T.; PENA, S.D. (1997) Extraction of DNA from human skeletal remains: practical applications in forensic sciences. *Genet Anal.* **14**:41-44.
- PROBST, C.M.; BOMPEIXE, E.P.; PEREIRA, N.F.; DE O DALALIO, M.M.; VISENTAINER, JE.; TSUNETO, L.T.; PETZL-ERLER, M.L.; (2000) HLA polymorphism and evaluation of European, African, and Amerindian contribution to the white and mulatto populations from Parana, Brazil. *Hum Biol* **72**:597-617.
- PULT I, SAJANTILA A, SIMANAINEN J, GEORGIEV O, SCHAFFNER W, PAABO S. (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. *Biol Chem Hoppe Seyler* **375**(12):837-840.
- QUINTANA-MURCI, L.; SEMINO, O.; BANDELT, H.J.; PASSARINO, G.; MCELREAVEY, K.; SANTACHIARA-BENERECETTI, A.S. (1999) Genetic evidence of an early exit of *Homo sapiens* from Africa through eastern Africa. *Nat Genet.* **23**:437-441.
- RAMOS, A. (1951) Introdução à Antropologia Brasileira. Vol1- As culturas não européias. Rio de Janeiro: :Livraria Editora da Casa do estudante do Brasil.
- RANDO JC, PINTO F, GONZÁLEZ AM, HERNÁNDEZ M, LARRUGA JM, CABRERA VM, BANDELT H-J (1998). Mitochondrial DNA analysis of northwest African populations reveals genetic
-

- exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet* **62**:531-550.
- RANDO, J.C.; CABRERA, V.M.; LARRUGA, J.M.; HERNÁNDEZ, M.; GONZÁLEZ, A.M.; PINTO, F.; BANDELT, H.J. (1999) Phylogeographic patterns of mtDNA reflecting the colonisation of the Canary Islands. *Ann Hum Genet* **63**:413-428.
- RANGEL, N.M. Fundamentos de Citologia Geral. Rio de Janeiro: Editora Guanabara Koogan S.A. 1974.
- REDD, A.; VALENCIA, D.; FRAZIER, B.; PÄÄBO, S. (1993) Genetic and linguistic differentiation in the Americas. *Proc. Natl. Acad. Sci. U.S.A.* **90**:10663-10667.
- REDD, A.J.; TAKEZAKI, N.; SHERRY, S.T.; MCGARVEY, S.T.; SOFRO, A.S.M.; STONEKING, M. (1995) Evolutionary history of the COII/tRNA^{Lys} intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol. Biol. Evol.* **17**(4): 604-615.
- RIBEIRO DOS SANTOS, A.K.; SANTOS, S.E.; MACHADO, A.L.; GUAPINDAIA, V.; ZAGO, M.A. (1996) Heterogeneity of mitochondrial DNA haplotypes in Pre-Columbian Natives of the Amazon region. *Am J Phys Anthropol* **101**(1):29-37.
- RIBEIRO, D. (1995) O povo brasileiro. *A formação e o sentido do Brasil*. São Paulo: Editora Scwarcz LTDA.
- RICHARDS, M.; CORTE-REAL, H.; FORSTER, P.; MACAULAY, V.; WILKINSON-HERBOTS, H.; DEMAINE, A.; PAPIHA, S.; HEDGES, R.; BANDELT, H.J.; SYKES, B. (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet.* 1996 Jul; **59**(1): 185-203.
- RICHARDS, M.B.; MACAULAY, V.A.; BANDELT, H.J.; SYKES, B.C. (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* **62**:241-260.
- RICHARDS, M.; MACAULAY, V.; HICKEY, E.; VEJA, E.; SYKES, B.; GUIDA, V.; RENGO, C. e cols (2000) Tracing european founder lineages in the near eastern mtDNA pool. *Am J Hum Genet* **67**:1251-1276.
- RICHTER, C.; PARK, J.W.; AMES, B.N. (1988) Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proc Natl Acad Sci U S A* **85**(17):6465-6467.
- RICHTER, C. (1992) Reactive oxygen and DNA damage in mitochondria. *Mutat. Res.* **275**: 249-255.
- RICKARDS, O.; MARTINEZ-LABARGA, C.; LUM, J.K.; DE STEFANO, G.F.; CANN, R.L. (1999) mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. *Am J Hum Genet* **65**:519-530
- ROSA, V.L.; SALZANO, F.M.; FRANCO, M.H.; FREITAS, M.J. (1984) Blood genetic studies in five Amazonian populations. *Braz J Genet* **75**:69-82.

-
- ROSSER ZH, ZERJAL T, HURLES ME, ADOJAAN M, ALAVANTIC D, AMORIM A, AMOS W, e cols., (2000) Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, Rather than by Language. *Am J Hum Genet* **67**:1526-1543.
- SACCONI, C.; PESOLE, G.; SBISA, E. (1991) The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *J Mol Evol* **33**(1):83-91.
- SAILLARD, J.; FORSTER, P.; LYNNERUP, N.; BANDELT, H.-J.; NØRBY, S. (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* **67**:718-726.
- SAITOU, N.; NEI, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**(4):406-425.
- SAJANTILA A, LAHERMO P, ANTTINEN T, LUKKA M, SISTONEN P, SAVONTAUS ML, AULA P, BECKMAN L, TRANEBJAERG L, GEDDE-DAHL T, ISSEL-TARVER L, DIRIENZO A, PAABO S. (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res.* **5**(1):42-52.
- SAJANTILA A, SALEM AH, SAVOLAINEN P, BAUER K, GIERIG C, PAABO S.(1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc Natl Acad Sci U S A.* **93**(21):12035-9.
- SALZANO, F.M.; FREIRE-MAIA, N. (1967) Populações brasileiras, aspectos demográficos, genéticos e antropológicos. São Paulo. Companhia Editora Nacional.
- SALZANO, F.M.; FREIRE-MAIA, N. (1970) Problems in human biology: *a study of brazilian populations*. Detroit: Wayne State University Press.
- SALZANO, F.M. (1997). Human races: myth, invention or reality? *Interciência* **22**:221-227.
- SANGER, F.; NICKLEN, S.; COULSON, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-7.
- SANGUINETTI, C.J., DIAS NETO, E. E SIMPSON, A.J. (1994) Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques* **17**: 914-921.
- SANTOS FR, PENA SD, EPPLEN JT. (1993) Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. *Hum Genet* **90**:655-656.
- SANTOS, S.E.B.; GUERREIRO, J.F. (1995) The indigenous contribution to the formation of the population of the Brazilian Amazon region. *Braz J Genet* **18**:311-315.
- SANTOS, F.R.; TYLER-SMITH, C. (1996). Reading the human Y chromosome: the emerging DNA markers and human genetic history. *Braz J Genet* **19**: 665-670.
- SANTOS EJM, RIBEIRO-DOS-SANTOS AKC, GUERREIRO JF, AGUIAR GFS, SANTOS SEB (1996). Migration and ethnic change in an admixed population from the Amazon region (Santarém, Pará). *Braz J Genet* **19**:511-515.
-

- SANTOS, FR, RODRIGUEZ-DELFIN, LA., PENA, SDJ; MOORE, J., WEISS, K. (1996) North And South Amerindians May Have The Same Major Founder Y Chromosome Haplotype. *Am J Hum Genet*.**58**:1369-1370.
- SANTOS, S.E.; RIBEIRO-DOS-SANTOS, A.K.; MEYER, D.; ZAGO, M.A.. (1996) Multiple founder haplotypes of mitochondrial DNA in Amerindians revealed by RFLP and sequencing. *Ann Hum Genet* **60**:305-319.
- SANTOS, M.S. (1998) Estudo da variabilidade do DNA mitocondrial na População Brasileira. Dissertação de mestrado. Departamento de Bioquímica e Imunologia, ICB-UFMG. Belo Horizonte.
- SANTOS, FR, PANDYA, A, TYLER-SMITH, C, PENA, SDJ, SCHANFIELD, M, LEONARD, WR., OSIPOVA, L., CRAWFORD, HAMER M, MITCHELL, R. J. (1999) The Central Siberian Origin For Native American Y-Chromosomes. *Am J Hum*
- SATOH, M.; KUROIWA, T. (1991) Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Exp Cell Res* 196(1):137-140.
- SCHENEIDER S, KUEFFER J-M, ROESSLI D, EXCOFFIER L (1997) Arlequin ver. 1.1: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- SCHURR, T.G.; BALLINGER, S.W.; GAN, Y.Y.; HODGE, J.A.; MERRIWETHER,D.A.; LAWRENCE, D.N.; KNOWLER, W.C.; WEISS, K.M.; WALLACE, D.C. (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am. J. Hum. Genet.* **46**:613-623.
- SCHURR TG, SUKERNIK RI, STARIKOVSKAYA YB, WALLACE DC. (1999).Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Bering Sea region during the Neolithic. *Am J Phys Anthropol* **108**:1-39.
- SCOZZARI, R.; TORRONI, A.; SEMINO, O.; SIRUGO, G.; BREGA, A.; SANTACHIARA BENERECETTI, A.S. (1988) Genetic studies on the Senegal population. I.Mitochondrial DNA polymorphisms. *Am. J. Hum. Genet.* **43**:534-544.
- SCOZZARI, R.; TORRONI, A.; SEMINO, O.; CRUCIANI, F.; SPEDINI, G.; SANTACHIARA BENERECETTI, S.A. (1994) Genetic studies in Cameroon: mitochondrial DNA polymorphisms in Bamileke. *Hum Biol* **66**(1):1-12.
- SCOZZARI, R.; CRUCIANI, F.; SANTOLAMAZZA, P.; SELLITTO, D.; COLE, D.E.; RUBIN, L.A.; LABUDA, D.; MARINI, E.; SUCCA, V.; VONA, G.; TORRONI, A. (1997) mtDNA and Y chromosome-specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. *Am J Hum Genet* **60**:241-244.

- SEMINO, O.; TORRONI, A.; SCOZZARI, R.; BREGA, A; DE BENEDICTIS, G.; SANTACHIARA BENERECETTI, A.S. (1989). Mitochondrial DNA polymorphisms in Italy. III. Population data from Sicily: a possible quantitation of maternal African ancestry. *Ann Hum Genet* **53**(Pt 2):193-202.
- SHIELDS GF, HECKER K, VOEVODA MI, REED JK. (1992) Absence of the Asian-specific region V mitochondrial marker in Native Beringians. *Am J Hum Genet* **50**(4):758-765.
- SHREEVE, J. (1995) The Neanderthal enigma. W. Morrow Co. (Penguin Books, 1997).
- SIMONI, L.; CALAFELL, F.; PETTENER, D.; BERTRANPETIT, J.; BARBUJANI, G. (2000) Geographic patterns of mtDNA diversity in Europe. *Am J Hum Genet.* **66**:262-78.
- SIGURÐARDÓTTIR S, HELGASON A, GULCHER JR, STEFANSSON K, DONNELLY P. (2000) The mutation rate in the human mtdna control region. *Am J Hum Genet* **66**(5):1599-1609.
- SMITH, D.G.; MALHI, R.S.; ESHLEMAN, J.; LORENZ, J.G.; KAESTLE, F.A. (1999) Distribution of mtDNA haplogroup X among Native North Americans. *Am J Phys Anthropol* **110**:271-284
- SOODYALL, H.; JENKINS, T. (1992) Mitochondrial DNA polymorphisms in Khoisan populations from southern Africa. *Ann Hum Genet* **56**:315-324.
- SOODYAL, H. (1993) Mitochondrial DNA polymorphisms in southern African populations. Tese de doutorado, University of Witwatersrand, Johannesburg.
- SOODYALL H, VIGILANT L, HILL AV, STONEKING M, JENKINS T. (1996) mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. *Am J Hum Genet* **58**(3):595-608.
- STARIKOVSKAYA YB, SUKERNIK RI, SCHURR TG, KOGELNIK AM, WALLACE DC. (1998) mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World. *Am J Hum Genet* **63**:1473-1491
- STENICO M, NIGRO L, BARBUJANI G. (1998) Mitochondrial lineages in Ladin-speaking communities of the eastern Alps. *Proc R Soc Lond B Biol Sci.* **7**;265(1396):555-561.
- STONE, A.C.; STONEKING, M. (1998) mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *Am J Hum Genet.* **62**:1153-70.
- STONE, A.C.; STONEKING, M. (1999) Analysis of ancient DNA from a prehistoric Amerindian cemetery. *Philos Trans R Soc Lond B Biol Sci.* **29**:153-159.
- STONEKING, M.; HEDGECOCK, D.; HIGUCHI, R.G.; VIGILANT, L.; ERLICH, H.A. (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am J Hum Genet* **48**(2):370-382.
- STONEKING, M.; SOODYALL (1996) Human evolution and the mitochondrial genome. *Current Opinion in Genetics & Development* **6**: 731-736.
- STONEKING, M. (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. *Am J Hum Genet* **67**:1029-1032

- STRINGER, C.B.; GRUN, R.; SCHWARCZ, H.P.; GOLDBERG, P. (1989) ESR dates for the hominid burial site of Es Skhul in Israel. *Nature* **27**(6218):756-758.
- SZATHMARY, E. J. E. (1986) Comment on "The settlement of the Americas: A comparison of the linguistic, dental and genetic evidence." *Curr Anthropol* **27**:490-491.
- SZATHMARY, E.J.E. (1993) mtDNA and the peopling of the Americas. *Am J Hum Genet* **53**(4):793-799.
- TAMURA, K.; NEI, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**(3):512-526.
- TEMPLETON, A. R. (1999) 'Human races: a genetic and evolutionary perspective', in *American Anthropologist*, v. 100, p. 632. 3:
- THOMAS MG, COOK CE, MILLER KW, WARING MJ, HAGELBERG E. (1998) Molecular instability in the COII-tRNA(Lys) intergenic region of the human mitochondrial genome: multiple origins of the 9-bp deletion and heteroplasmy for expanded repeats. *Philos Trans R Soc Lond B Biol Sci.* **29**:955-965.
- TORRONI A., SCHURR, T.G.; YOUNG, C.C.; SZATHMARY, E.J.E.; WILLIAMS, R.C.; SCHANFIELD, M.S.; TROUP, G.A.; KNOWLER, W.C.; LAWRENCE, D.N.; WEISS, K.M.; WALLACE D.C. (1992) Native American mitochondrial DNA analyses indicates that Amerind and the Nadene populations were founded by two independent migrations. *Genetics* **130**: 153-162.
- TORRONI, A.; SCHURR, T.G.; CABELL, M.F.; BROWN, M.D.; NEEL, J.V.; LARSEN, M.; SMITH, D.G.; VULLO, C.M.; WALLACE, D.C; (1993a) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* **53**:563-590.
- TORRONI, A.; SUKERNIK, R.I.; SCHURR, T.G.; STARIKOVSKAYA, Y.B.; CABELL, M.F.; CRAWFORD, M.H.; COMUZZIE, A.G.; WALLACE, D.C. (1993b) MtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* **53**:591-608.
- TORRONI. A.; CHEN, Y.S.; SEMINO, O.; SANTACHIARA-BENECERETTI, A.S.; SCOTT, C.R.; LOTT, M.T.; WINTER, M.; WALLACE, D.C. (1994a) mtDNA and Y-chromosome polymorphisms in four Native American populations from Southern Mexico. *Am J Hum Genet* **54**(2):303-318.
- TORRONI, A.; NEEL, J.V.; BARRANTES, R.; SCHURR, T.G.; WALLACE, D.C. (1994b) Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. *Proc Natl Acad Sci U S A* **91**(3):1158-1162.
- TORRONI, A.; LOTT, M.T.; CABELL, M.F.; CHEN, Y.; LAVERGE, L.; WALLACE, D.C. (1994c) MtDNA and the origin of Caucasians. Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am. J. Hum. Genet.* **55**:760-776.

- TORRONI, A. e WALLACE, D.C. (1994) Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. *J Bioenerg Biomembr* **26**(3):261-271.
- TORRONI A, BROWN MD, LOTT MT, NEWMAN NJ, WALLACE DC. (1995). African, Native American, and European mitochondrial DNAs in Cubans from Pinar del Rio Province and implications for the recent epidemic neuropathy in Cuba. *Hum Mutat.* **5**:310-317.
- TORRONI, A.; HUOPONEN, K.; FRANICALACCI, P.; PETROZZI, M.; MORELLI, L.; SCOZZARI, R.; OBINU, D.; SAVONTAUS, M.L.; WALLACE, D.C. (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**(4):1835-1850.
- TORRONI, A.; PETROZZI, M.; D'URBANO, L.; SELLITTO, D.; ZEVIANI, M.; CARRARA, F.; CARDUCCI, C.; LEUZZI, V.; CARELLI, V.; BARBONI, P.; DE NEGRI, A.; SCOZZARI, R. (1997) Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. *Am J Hum Genet* **60**:1107-1121.
- TORRONI, A.; BANDELT, H.J.; D'URBANO, L.; LAHERMO, P.; MORAL, P.; SELLITTO, D.; RENGO, C.; FORSTER, P.; SAVONTAUS, M.L.; BONNE-TAMIR, B.; SCOZZARI, R. (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet.* **62**(5): 1137-1152.
- TORRONI, A.; RICHARDS, M.; MACAULAY, V.; FORSTER, P.; VILLEMS, R.; NORBY, S.; SAVONTAUS, M.L.; HUOPONEN, K.; SCOZZARI, R.; BANDELT, H.J. (2000) mtDNA haplogroups and frequency patterns in Europe. *Am J Hum Genet.* **66**:1173-1177.
- VALLINOTO, A.C.; CAYRES-VALLINOTO, I.M.; ZAGO, M.A.; SANTOS, S.E.; GUERREIRO, J.F.; (1998) D1S80 polymorphism in Amerindians from the Amazon region of Brazil. *Hum Biol.* **70**:507-516.
- VIGILANT, L.; PENNINGTON, R.; HARPENDING, H.; KOCHER, T.D.; WILSON, A.C. (1989) Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci U S A.* **86**:9350-9354.
- VIGILANT, L. (1990) Control region sequences from African population and the evolution of human mitochondrial DNA. PhD thesis, University of California, Berkeley.
- VIGILANT, L.; STONEKING, M.; HARPENDING, H.; HAWKES, K.; WILSON, A.C. (1991) African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503-1507.
- VON HAESLER, A.; SAJANTILA, A.; PAABO, S. (1996) The genetical archaeology of the human genome. *Nat Genet* **14**:135-140.
- WAKELEY, J. (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J Mol Evol* **37**(6):613-623.

- WALLACE, D.C.; YE, J.H.; NECKELMANN, S.N.; SINGH, G.; WEBSTER, K.A.; GREENBERG, B.D. (1987) Sequence analysis of cDNAs for the human and bovine ATP synthase beta subunit: mitochondrial DNA genes sustain seventeen times more mutations. *Curr Genet* **12**(2):81-90.
- WALLACE, D.C.; TORRONI, A. (1992) American Indian prehistory as written in the mitochondrial DNA: a review. *Hum. Biol.* **64**:403-416.
- WALLACE, D. C. (1995) Mitochondrial DNA Variation in Human Evolution, Degenerative Disease, and Aging. *Am. J. Hum. Genet.* **57**: 201-223.
- WALLACE DC, SHOFFNER JM, TROUNCE I, BROWN MD, BALLINGER SW, CORRAL-DEBRINSKIM, HORTON T, JUN AS, LOTT MT (1995) Mitochondrial DNA mutations in human degenerative diseases and aging. *Biochim Biophys Acta* **24**:141-151.
- WALLACE, D.C. (1997) Mitochondrial DNA in aging and disease. *Sci Am* **277**(2):40-47.
- WALLACE, D.C.; BROWN, M.D.; LOTT, M.T. (1999) Mitochondrial DNA variation in human evolution and disease. *Gene*. **30**:211-30.
- WARD, R.H.; SALZANO, F.M.; BONATTO, S.L.; HUTZ, M.H.; COIMBRA, J.R; SANTOS, R.V. (1996) Mitochondrial DNA polymorphism in three Brazilian Indian tribes. *Am. J. Hum. Biol* **8**:317-323.
- WATSON, E., BAUER, K., AMAN, R., WEISS, G.; HAESELER, A. V.; PÄÄBO, S. (1996) mtDNA Sequence Diversity in África. *Am. J. Hum. Genet.* **59**, 437-444.
- WATSON, E.; FORSTER, P.; RICHARDS, M.; BANDELT, H.J. (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet.*; **61**(3): 691-704.
- WEIMER, T.A.; SALZANO, F.M.; WESTWOOD, B.; BEUTLER, E. (1998) G6PD variants in three South American ethnic groups: population distribution and description of two new mutations. *Hum Hered.* **48**:92-96.
- WOLSTENHOLME, D.R. (1992) Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* **141**:173-216.
- WOLSTENHOLME, D.R.; JEON, W.K. Mitochondrial Genomes. San Diego: Academic Press Inc., 1992.
- WRISCHNIK, L.A.; HIGUCHI, R.G.; STONEKING, M.; ERLICH, H.A.; ARNHEIM, N.; WILSON, A.C. (1987) Length Mutations in Human Mitochondrial DNA: Direct Sequencing of Enzymatically Amplified DNA. *Nucleic Acid Research* **15**, 2:529-542.
- ZAGO, M.A.; SILVA JUNIOR, W.A.; TAVELLA, M.H.; SANTOS, S.E.; GUERREIRO, J.F.; FIGUEIREDO, M.S. (1996) Interpopulational and intrapopulational genetic diversity of Amerindians as revealed by six variable number of tandem repeats. *Hum Hered* **46**(5):274-289.
- ZHAO, M. e BADA, J.L. (1995) Determination of α -dialkylamino acids and their enantiomers in geological samples by high-performance liquid chromatography after derivatization with a chiral adduct of *o*-phthalaldehyde. *J Chromatogr A* **690**:55-63.

Apêndice 1

Identification in Portugal and Brazil of a mtDNA Lineage Containing a 9-bp Triplication of the Intergenic COII/tRNA^{Lys} Region

Juliana Alves-Silva^a Pedro E.M. Guimarães^a Jorge Rocha^{b, c}
Sérgio D.J. Pena^a Vania F. Prado^a

^aDepartamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil;

^bInstituto de Patologia e Imunologia Molecular, ^cDepartamento de Zoologia-Antropologia, Universidade do Porto, Porto, Portugal

Key Words

MtDNA · 9-bp triplication · Brazilian Caucasians · Portuguese Caucasians · RFLP · Sequencing

Portugal and that it has apparently differentiated into at least two lineages defined by the mutation in HVS2 suggests that it probably occurred a long time ago.

Abstract

Although the deletion of one of the 9-bp repeats in region V of mitochondrial DNA is very common in Asians, Asian-derived populations and Africans, the triplication of the 9-bp segment was described only a few times, mostly on individuals from Asian origin. Here, we report for the first time the presence of the 9-bp triplication in Europeans. The triplication was initially found in one Brazilian individual. Sequencing of the hypervariable segments I (HVS1) and II (HVS2) of the control region and RFLP analysis of the coding region classified the mtDNA as belonging to the European haplogroup H. Since white Brazilians are predominantly of Portuguese descent, we screened 96 unrelated Northern Portuguese for the 9-bp triplication and found its presence in two of them (2.1%). One of these had an mtDNA haplotype identical to that of the Brazilian individual, while the other differed in a single base change in HVS2. The fact that the 9-bp triplication has reached polymorphic frequencies in Northern

Introduction

The intergenic region between the cytochrome oxidase II and lysine tRNA genes (COII/tRNA^{Lys}) contains two tandemly repeated copies of the 9-bp sequence CCCCCTCTA in the mtDNA reference sequence [1]. Several reports have demonstrated that the loss of one of the repeats is very common in Asians and populations of Asian origin, including Polynesians and Native Americans [2] and also in Sub-Saharan Africans [3]. These studies have made the 9-bp deletion one of the most informative continent-specific polymorphisms [2, 3]. On the other hand, insertion events in this segment seem to be rare. The complete triplication of the nononucleotide was described only a few times, namely by Shields et al. [2] in one individual from Siberia, by Passarino et al. [4] in one person from eastern Tharu and by Merriwether et al. [5] in Alaskan Eskimos and in Ivory Coast Africans. Partial

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 1999 S. Karger AG, Basel
0001-5652/99/0491-0056\$17.50/0

Accessible online at:
<http://BioMedNet.com/karger>

Dr. Vania F. Prado
Departamento de Bioquímica e Imunologia
Instituto de Ciências Biológicas, UFMG, Caixa Postal 486
Belo Horizonte, MG 30161-970 (Brasil)
Fax +55 31 227 3792, E-Mail vprado@mono.icb.ufmg.br

insertions of approximately 4-bp were reported by Wrischnik et al. [6] and Ballinger et al. [7] in Asians. Most of these insertions were associated with different haplotypes and it has been suggested that they originated from independent mutational events [4, 5, 7].

While screening Brazilians for the presence of the 9-bp deletion [13], we observed one individual with the 9-bp triplication. Sequencing of the control region and RFLP analysis of the coding region classified the mtDNA as belonging to the European haplogroup H. Since white Brazilians are predominantly of Portuguese descent, we screened 96 unrelated Northern Portuguese for the 9-bp triplication and found its presence in two of them (2.1%). One of these had an mtDNA haplotype identical to that of the Brazilian individual, while the other differed in a single base change in HVS2. This is the first time that the 9-bp triplication is reported in Europeans; it has apparently occurred a long time ago by a new mutational event on a haplogroup H background.

Materials and Methods

Samples

Two hundred and forty-five unrelated white Brazilians and 96 unrelated individuals from Northern Portugal were studied. DNA was prepared from blood cells using standard procedures.

9-bp Polymorphism

Amplifications were performed in 10-μl volumes. Each tube contained 10 pmol of primers L8209 5'-CATCGTCCTAGAAT-TAATTCC-3' and H8304 5'-CTTTACAGTGGGCTCTAGAGG-3', 200 μM dNTP and 0.5 U of *Taq* DNA polymerase (Promega Corporation, Madison, Wisc., USA). Thirty cycles of denaturation at 94°C for 1 min, annealing at 55°C for 30 s and extension at 72°C for 1 min were carried out. PCR products were visualized by 6% acrylamide gel electrophoresis after silver staining.

mtDNA Control Region Amplification and Sequencing

Primers L15926 5'-TCAAAGCTTACACCAGTCTTGTAAC-ACC-3' and H16498 5'-CCTGAAGTAGGAACCA GATG-3' were used, and the amplification conditions were the same as described above. PCR products were purified using Magic™ PCR Preps (Promega) and dideoxy sequencing was carried out with Thermo Sequenase Sequencing Kit (Amersham Life Science, Little Chalfont, Buckinghamshire, England) using the fluorescent labeled primers L15996 5'-CTCCACCATTAGCACCCAAAGC-3' and H16401 5'-TGATTTACGGAGGATGGTG-3'.

Analysis of the Main Continent-Specific RFLP

The entire mtDNA of the individuals with the 9-bp triplication was amplified in 9 overlapping fragments by PCR using primers and conditions described by Torroni et al. [9]. The PCR segments were digested by 15 restriction endonucleases (*AluI*, *AvaII*, *BamHI*, *BstNI*, *DdeI*, *HaeII*, *HaeIII*, *HhaI*, *HincII*, *HinfI*, *HpaI*, *MboI*, *MseI*, *RsaI*

		1	1	1			
		6	6	6			
Position		1	2	3	1	2	2
		9	7	6	5	3	6
	↓	2	4	2	2	9	3

CRS		C	G	T	T	T	A

BR13		T	A	C	.	C	G
PT01		T	A	C	.	C	G
PT02		T	A	C	C	C	G

Fig. 1. Polymorphic sites of the mtDNA hypervariable segment I (HVS1) (positions 16061–16362) and hypervariable segment II (HVS2) (positions 72–328) found in the Brazilian and the two Portuguese individuals with the 9-bp triplication. Nucleotide positions are given in comparison to the Cambridge Reference Sequence (CRS [1]) and only positions with sequence divergence are shown. Dots indicate identity with the CRS.

and *TaqI*), following the conditions specified by the manufacturer (GibcoBRL Life Technologies, Gaithersburg, Md., USA). The resulting fragments were resolved through electrophoresis in 1% agarose gels and visualized by UV-induced fluorescence after ethidium bromide staining.

Results and Discussion

We recently screened the mtDNA of 245 Brazilian Caucasians for the presence of the 9-bp deletion [13] and observed that one individual (BR13) showed a PCR product larger than the 95-bp band expected from the reference sequence COII/tRNA^{Lys} region. This segment was cloned and sequenced, showing a third copy of the 9-bp repeat. Unfortunately, BR13 was an anonymous DNA sample, and family studies were not possible. However, in an effort to establish the geographic origin of this mtDNA, we undertook the complete sequencing of hypervariable segments I (HVS1; bases 16061–16362) and II (HVS2; bases 72–328) of the control region (fig. 1) as well as analysis of the most continent-specific polymorphisms of the coding region [8]. In addition to the 9-bp triplication, the mtDNA haplotype showed, compared to the Anderson sequence, three transitions at HVS1 (C → T at 16192, G → A at 16274, T → C at 16362), two transitions at HVS2 (T → C at 239 and A → G at 263) a *DdeI* site loss at 10394 and an *AluI* site loss at 7025. Although the control region sequence was unspecific, the RFLP analysis classi-

fied the mtDNA as belonging to haplogroup H, which is typical of European populations [9].

Since white Brazilians are predominantly of Portuguese descent [10], we decided to look for the possible presence of other instances of 9-bp triplication in the Portuguese population. We studied 96 unrelated individuals from Northern Portugal and indeed could find the triplication in two of them. By sequencing of the mtDNA control region and RFLP analysis of these individuals we discovered that one of them (PT01) had a haplotype identical to that of the Brazilian individual, while the other (PT02) differed in only one position in HVS2 (T → C at position 152) (fig. 1). This is the first time that the 9-bp triplication is reported in Europeans: it has apparently occurred by a new mutational event on a haplogroup H background. Although we cannot rule out the possibility of two independent triplication events, it is much more likely that the different haplotypes observed in the two Portuguese individuals were generated by a mutation in

HSV2 after a previous single 9-bp triplication. Position 152 is known to be highly polymorphic: the same T → C transition at this site was present in 14 out of 47 haplotypes from Tuscany [11] and in 34 out of 69 haplotypes from several other populations [12]. Unfortunately, we do not have any family information on the Brazilian individual to find out when women carrying this mtDNA lineage came to Brazil. However, the fact that the triplication can be seen in at least two different haplotypes in Portugal suggests that it may have occurred a long time ago. A more comprehensive study of the Portuguese population is being undertaken to shed light on the age and genealogy of this interesting mtDNA lineage.

Acknowledgment

We would like to thank Katia Barroso and Neuza Antunes Rodrigues for technical assistance. This work was supported by grants from CNPq, FAPEMIG and PRPq-UFMG.

References

- 1 Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Rose BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG: Sequence and organization of the human mitochondrial genome. *Nature* 1981;290:457–465.
- 2 Shields GF, Hecker K, Voevoda MI, Redd JK: Absence of the Asian specific region V mitochondrial marker in Native Beringians. *Am J Hum Genet* 1992;50:758–765.
- 3 Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T: mtDNA control-region sequence variation suggests multiple independent origins of an 'Asian-Specific' 9-bp deletion in sub-Saharan Africans. *Am J Hum Genet* 1996;58:595–608.
- 4 Passarino G, Semino O, Modiana G, Santachiara-Benerecetti AS: COII/tRNA^{Lys} intergenic 9-bp deletion and other mtDNA markers clearly reveals that Tharus Southern Nepal have oriental affinities. *Am J Hum Genet* 1993;53:609–618.
- 5 Merriwether DA, Rothhammer F, Ferrel RE: Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. *Am J Phys Anthropol* 1995;98:411–430.
- 6 Wrischnik LA, Higuchi RG, Stoneking M, Ehrlich HA, Arnheim N, Wilson AC: Length mutations in human mitochondrial DNA: Direct sequencing of enzymatically amplified DNA. *Nucleic Acid Res* 1987;15:529–542.
- 7 Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JO, Hassan K, Shen KH, Wallace DC: Southeast Asian mitochondrial DNA analyses reveals genetic continuity of ancient Mongoloid migrations. *Genetics* 1992;130:139–152.
- 8 Kogelnik AM, Lott MT, Brown MD, Navathe SB, Wallace DC: MITOMAP: A human mitochondrial genome database. *Nucleic Acids Res* 1996;24:177–179.
- 9 Torroni A, Huoponen K, Francalacci P, Petrosz M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC: Classification of European mtDNAs from an analysis of three European populations. *Genetics* 1996;144:1835–1850.
- 10 Salzano FM, Freire-Maia N: Problems in Human Biology: A Study of Brazilian Populations. Detroit, Wayne State University Press, 1970.
- 11 Francalacci P, Bertranpetit J, Calafell F, Underhill PA: Sequence diversity of the control region mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am J Phys Anthropol* 1996;100:443–460.
- 12 Bamshad M, Fraley AE, Crawford MH, Cann RL, Busi AR, Naidu JM, Jorde LB: mtDNA variation in caste populations of Andhra Pradesh, India. *Hum Biol* 1996;68:1–28.
- 13 Alves-Silva J, Santos MS, Pena SDJ, Prado VF: Multiple geographic sources of region V 9-bp deletion haplotypes in Brazilians. *Hum Biol*, in press.

Apêndice 2

The Ancestry of Brazilian mtDNA Lineages

Juliana Alves-Silva,¹ Magda da Silva Santos,¹ Pedro E. M. Guimarães,¹
Alessandro C. S. Ferreira,¹ Hans-Jürgen Bandelt,² Sérgio D. J. Pena,¹ and Vania Ferreira Prado¹

¹Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brazil; and ² Fachbereich Mathematik, Universität Hamburg, Hamburg

We have analyzed 247 Brazilian mtDNAs for hypervariable segment (HVS)–I and selected restriction fragment-length–polymorphism sites, to assess their ancestry in different continents. The total sample showed nearly equal amounts of Native American, African, and European matrilineal genetic contribution but with regional differences within Brazil. The mtDNA pool of present-day Brazilians clearly reflects the imprints of the early Portuguese colonization process (involving directional mating), as well as the recent immigrant waves (from Europe) of the last century. The subset of 99 mtDNAs from the southeastern region encompasses nearly all mtDNA haplogroups observed in the total Brazilian sample; for this regional subset, HVS-II was analyzed, providing, in particular, some novel details of the African mtDNA phylogeny.

Introduction

Brazilians form one of the most heterogeneous populations in the world, the result of 5 centuries of inter-ethnic crosses between peoples from three continents: the European colonizers, represented mainly by the Portuguese; African slaves; and the autochthonous Amerindians. When the Portuguese arrived, exactly 500 years ago, there were ~2.5 million indigenous people living in the area of what is now Brazil (Salzano and Freire-Maia 1970; Bethell 1997). The Portuguese-Amerindian admixture started soon after the arrival of the first colonizers. Mating between European men and indigenous women became commonplace and later (after 1755) was even encouraged as a strategy for population growth and colonial occupation of the country (Mörner 1967). The Amerindian tribes underwent a drastic demographic decline due to conflicts with the European colonizers and diseases to which they were not adapted (Salzano and Freire-Maia 1967, 1970; Monteiro 1994; Ribeiro 1995). Today there are ~326,000 Amerindians in Brazil, living on land set aside for them by the federal government. Africans were introduced beginning in the middle of the 16th century, brought to Brazil as slaves to work on sugarcane farms and, later, in the gold and diamond mines and on coffee plantations. Historical records suggest that between 1551 and 1850 (when the slave trade

was abolished), ~3.5 million Africans arrived in Brazil (Salzano and Freire-Maia 1967; Curtin 1969; Ribeiro 1995). As to the European immigration, it is estimated that ~500,000 Portuguese arrived in the country between 1500 and 1808 (Salzano and Freire-Maia 1967). From then on, after the Brazilian ports were legally opened to all friendly nations, Brazil received increasing numbers of immigrants from several parts of the world. Portugal remained by far the most important source of migrants, followed by Italy, Spain, and Germany. In the 20th century, Asian immigration took place, mainly from Japan, as well as from Lebanon and Syria. According to Callegari-Jacques and Salzano (1999), 58% of the immigrants who arrived in Brazil between 1500 and 1972 were Europeans, 40% were Africans, and 2% were Asians. The question that arises is, How much did these different groups actually contribute to the gene pool of present-day Brazilians?

Several studies were performed during recent decades, in an attempt to characterize the genetic background of the non-Amerindian Brazilian population (for a review, see Salzano 1997; Callegari-Jacques and Salzano 1999; Dornelles et al. 1999; Guerreiro et al. 1999). These studies included mainly samples from the northern and southern regions of the country. On the basis of classical genetic markers, these studies demonstrated that all analyzed groups show some degree of admixture and that the extent of admixture varied, depending on the region analyzed. Recently, some Brazilian population samples have been analyzed for mtDNA (Bortolini and Salzano 1996; Santos et al. 1996b; Ward et al. 1996; Bortolini et al. 1997b; Batista dos Santos et al. 1999) and Y-chromosome polymorphisms (Batista dos Santos et al. 1999). The mtDNA studies have shown that Amerindian and African contributions to northern Brazilians

Received March 20, 2000; accepted for publication May 22, 2000; electronically published June 28, 2000.

Address for correspondence and reprints: Prof. Sérgio D. J. Pena, Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, UFMG, Caixa Postal 486, Belo Horizonte, MG, Brazil 30161-970. E-mail: spena@dcc.ufmg.br

© 2000 by The American Society of Human Genetics. All rights reserved.
0002-9297/2000/6702-0021\$02.00

are larger than those previously observed on the basis of classical markers (Batista dos Santos et al. 1999; Santos et al. 1999).

mtDNA analysis has been used extensively during the past 10 years, since the pioneering works of Vigilant (1990), Stoneking et al. (1991), and Vigilant et al. (1991). Phylogeographic analysis of mtDNA lineages from all over the world has led to the identification of mtDNA haplogroups that are specific to either Africans, Europeans, or Asians/Amerindians (Torroni et al. 1993, 1994, 1996, 1998; Chen et al. 1995; Richards et al. 1996; Watson et al. 1997; Kivisild et al. 1999a, 1999b; Macaulay et al. 1999; Metspalu et al. 1999). Haplogroup allocation of a given mtDNA lineage allows the assessment of its (sub)continental origin, so that the matrilineal ancestry of admixed populations can be evaluated well (Torroni et al. 1995; Bravi et al. 1997; Green et al. 2000; Rando et al. 1999).

In the present article, we follow this approach by sequencing part of the control region and by screening specific RFLP sites, to better understand the extent of the matrilineal genetic contribution of Europeans, Africans, Amerindians, and Asians to the gene pool of present-day Brazilians.

Subjects and Methods

Samples

We analyzed 247 unrelated Brazilian individuals (mainly classified as "white" in Brazil and belonging to the middle and upper-middle classes) who came from four of the five major geographic regions of the country (fig. 1). According to the Instituto Brasileiro de Geografia Estatística, responsible for the census in Brazil, 51.6% of Brazilians in 1996 classified themselves as white. In detail, there were 99 individuals from the southeastern (mostly from the state of Minas Gerais), 50 from the southern (states of Rio Grande do Sul, Santa Catarina, and Paraná), 48 from the northern (states of Amazonas, Pará, Rondônia, and Acre), and 50 from the northeastern (state of Pernambuco) regions. Thirty-seven individuals were students or staff in our laboratory, whereas 210 were randomly chosen unrelated participants in paternity-testing studies. Written consent was obtained from all participants, and all analyses were performed anonymously.

mtDNA Control-Region Amplification and Sequencing

The nucleotide sequence of mtDNA hypervariable segment I (HVS-I), between nucleotide positions (np) 16060 and 16362, was determined for all the individuals in the study (see GenBank accession numbers in the Electronic-Database Information section). PCR amplification of the mtDNA control region was performed in a 45- μ l vol-



Figure 1 Five major geographic regions of Brazil: N = northern, NE = northeastern, SE = southeastern, S = southern, and CW = central west. Brazilian states from which the mtDNA lineages of the present study have mainly been sampled are indicated by name.

ume. Each tube contained 0.8 μ M of primers L15926 5'-TCAAAGCTTACACCAGTCTTGTAACC-3' and H16498 5'-CCTGAAGTAGGAACCAGATG-3', 200 μ M dNTP, and 0.5 U of *Taq* DNA polymerase (Promega). Thirty cycles of denaturation at 94°C for 1 min, annealing at 55°C for 30 s, and an extension at 72°C for 1 min were done. Negative controls were run simultaneously, to detect reagent contamination. PCR products were visualized in 1% agarose-gel electrophoresis with ethidium bromide. Amplified segments were purified using Magic PCR Preps (Promega), and dideoxy sequencing was done with the Thermo Sequenase Sequencing Kit (Amersham Life Science) and a fluorescent-labeled primer L15996 5'-CTCCACCATTAGCACCC-AAAGC-3' or H16401 5'-TGATTTCACGGAGGATGGTG-3'. For the samples from the southeastern region, the HVS-II sequence between np 72 and 337 was also determined. Primers L29 5'-GGTCTATCACCCCTATTAACCAC-3' and H580 5'-TTGAGGAGGTAAGCTACATA-3' were used in PCR reactions, in the same conditions described above, and a fluorescent-labeled primer L48 5'-CTCACGGGAGCTCTCCATGC-3' or H408' 5'-CTGTTAAAGTGCATACCGCCA-3', was used in sequencing reactions.

Partial Restriction-Site Analysis

Several amplified segments, mainly in the mtDNA coding regions, were analyzed by RFLP tests, according

to the method described by Chen et al. (1995) and Torroni et al. (1996), to screen haplogroup-specific sites (table 1). PCR amplifications were performed using the primers and conditions described by Torroni et al. (1992, 1993, 1996). Digestions were carried out according to the conditions specified by the manufacturer (Gibco BRL). The resulting fragments were resolved by electrophoresis in 1% agarose gels and were visualized by UV-induced fluorescence after ethidium bromide staining. Depending on the number and length of resulting fragments, they were visualized in 8% acrylamide gels after silver staining. The 12308 *HinfI* polymorphic site was analyzed using the mismatched primer described by Torroni et al. (1996).

Phylogeographic Analysis

We build on the phylogenetic analyses of European (Richards et al. 1998; Macaulay et al. 1999) and African (Rando et al. 1998) mtDNA, which combine HVS-I and RFLP information. According to the nomenclature of those analyses, human mtDNAs are divided into three supergroups—L1 (+3592 *HpaI*, –10806 *HinfI*), L2 (+3592 *HpaI*, –16390 *HinfI*), and L3 (–3592 *HpaI*). L1 and L3 are further subdivided into haplogroups, which can be recognized by specific restriction sites (table 1). L1 and L2 are African specific, whereas L3 is ubiquitous but encompasses several haplogroups that are (nearly) continent specific. From HVS-I sequences

alone, the fine-grained haplogroup status can be read off only to some extent, and, therefore, their characteristic restriction site(s) need to be tested for confirmation. In the case of haplogroups that are shared between continents, HVS-I motifs or exclusive matches could further suggest the most plausible geographic origin. Figure 2 displays the hierarchy of haplogroups that is relevant for the present study. Note that haplogroup U includes haplogroup K. In the fine classification of mtDNA lineages, we employ the “asterisk notation” (Richards et al. 1998): an mtDNA lineage belongs to some haplogroup labeled with an asterisk if it is a member of that group but not of any otherwise-highlighted subgroup.

Results

mtDNA Composition of the Brazilian Population

The 247 Brazilian mtDNA lineages, yielding 170 different HVS-I haplotypes, can be perfectly allocated to the known haplogroups (tables 2 and 3 and fig. 2). Altogether, 82 mtDNA lineages fall into the Native American/Asian haplogroups, A–D (with one A lineage of confirmed western-Asian ancestry), whereas 69 belong to various African haplogroups and 96 belong to European haplogroups. The relative frequencies of these continental fractions of the mtDNA pool, though, vary considerably over the four Brazilian regions analyzed. In the northern region, the majority of the mtDNA line-

Table 1

RFLP Polymorphisms Used to Identify mtDNA Haplogroups and Geographic Origin

HAPLOGROUP	CHARACTERISTIC RESTRICTION SITE(S)	STATUS ^a		
		Sub-Saharan African	Native American	European
L1a	+3592 <i>HpaI</i> , +11641 <i>HaeIII</i>	+	–	–
L1b	+3592 <i>HpaI</i> , –7055 <i>AluI</i> , +2349 <i>MboI</i>	+	–	–
L1c	+3592 <i>HpaI</i> , +9070 <i>TaqI</i> , +12810 <i>RsaI</i>	+	–	–
L2	+3592 <i>HpaI</i> , +16389 <i>HinfI</i>	+	–	–
L3b	+10084 <i>TaqI</i>	+	–	–
L3d	–3592 <i>HpaI</i> , –8616 <i>MboI</i>	+	–	–
L3e	–3592 <i>HpaI</i> , +2349 <i>MboI</i>	+	–	–
A	+663 <i>HaeIII</i>	–	+	–
B	9-bp deletion	–	+	–
C	–13259 <i>HincII</i>	–	+	–
D	–5176 <i>AluI</i>	–	+	–
H	–7025 <i>AluI</i>	–	–	+
V	–4577 <i>NlaIII</i>	–	–	+
HV	–14766 <i>MseI</i>	–	–	+
U	+12308 <i>HinfI</i>	–	–	+
K	–9052 <i>HaeII</i>	–	–	+
J	–13704 <i>BstNI</i>	–	–	+
T	+13366 <i>BamHI</i> , +15606 <i>AluI</i>	–	–	+
I	–4529 <i>HaeII</i> , +8249 <i>AvaII</i> , +16389 <i>BamHI</i> , +10032 <i>AluI</i>	–	–	+
W	+8249 <i>AvaII</i> , –8994 <i>HaeIII</i>	–	–	+
X	–1715 <i>DdeI</i>	–	+	+

^a A plus sign (+) denotes that the haplogroup is indigenous; a minus sign (–) denotes that it is not.

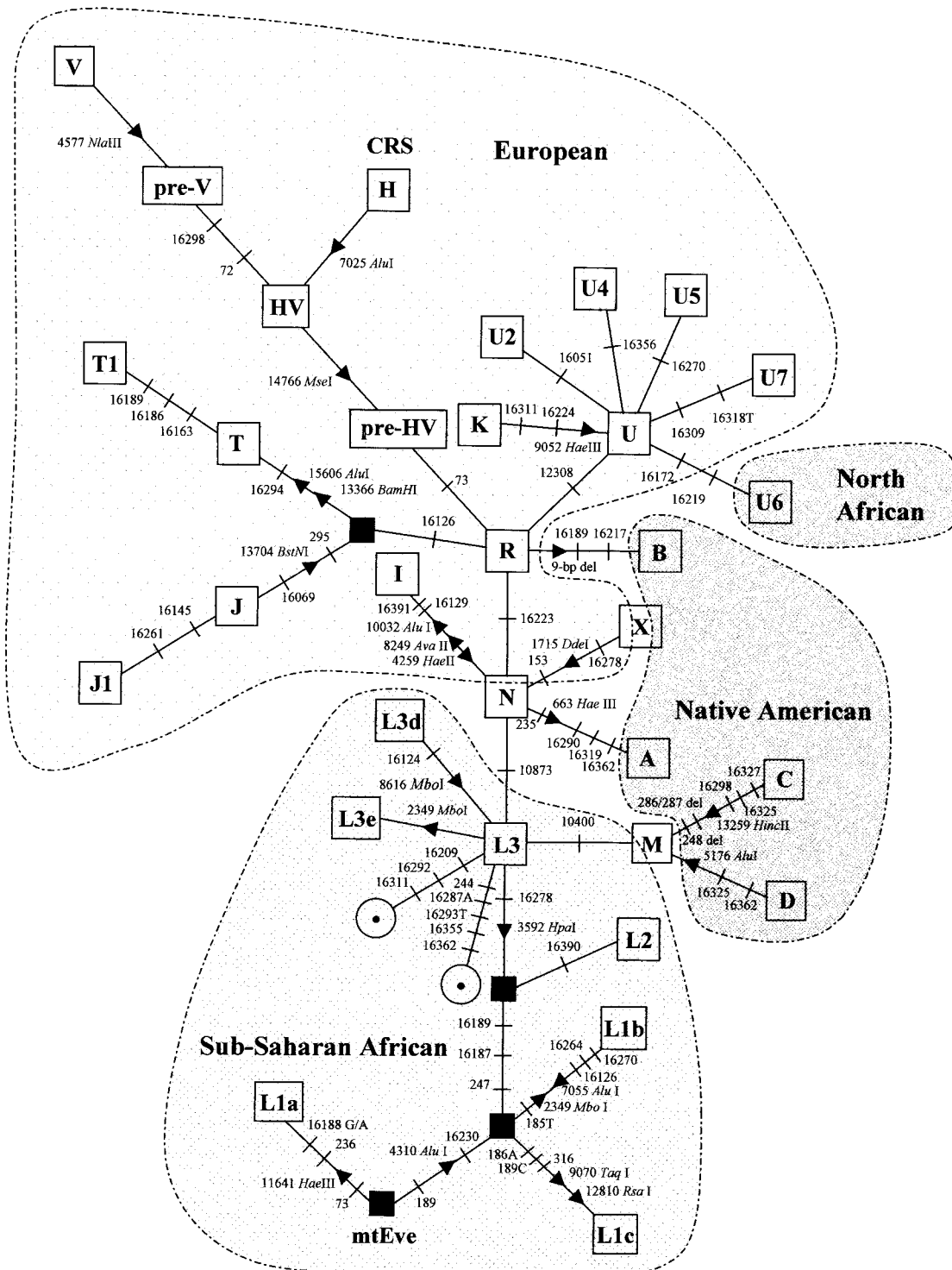


Figure 2 Classification tree highlighting selected diagnostic sites and positions for haplogroups present in the Brazilian sample (see tables 1 and table 6). Each square represents the root node of the respective haplogroup, with the acronym inscribed; two central/eastern-African haplogroups, represented by circles, are only partially characterized (T. Kisivild, personal communication). “CRS” indicates the revised reference sequence (Andrews et al. 1999). Numbers along links refer to RFLP sites (with arrows pointing to presence of sites) or transitions, unless a single-letter suffix indicates a transversion. Note that some diagnostic sites and positions, especially in the control region, have undergone recurrent mutations. The root of the tree, labeled “mtEve,” is inferred by employing the Neanderthal HVS-I and HVS-II sequences (Krings et al. 1997, 1999) and the coding-region sequences of bonobo and common chimpanzee (Horai et al. 1995) as outgroups; this corroborates the rooting of the Vigilant (1990) tree.

BR41	1	T.....	T.....G.....	C.....CT.....	C
BR42	1	A.....	T.....A.....	C.....CT.....T.	C
BR43	1	A.....	T.....T.....	C.....CT.....	C
BR44	2	A.....	T.....T.....	C.....CT.....	C
BR45	2	A.....	T.....T.....	C.....C.....	D
BR46	1	T.....	T.....T.....	C.....C.....	D
BR47	1	T.....	T.....T.....	C.....C.....	D
BR48	1	A.....	T.....T.....	C.....C.....	D
BR49	1	A.....	T.....T.....	C.....C.....	D
BR50	1	C.....	T.....T.....	C.....C.....	D
BR51	1	C.....	T.....T.....	C.....C.....	D
BR52	1	T.....	T.....G.....	C.....C.....	D
BR53	1	T.....	T.....G.....	C.....C.....	D
African:					
BR54	1	T.....	T.....G.....	C.....T.....	L1a
BR55	2	A.....	T.....G.....	C.....T.....	L1a
BR56	1	A.....	T.....G.....	C.....T.....	L1a
BR57	1	A.....	T.....G.....	C.....T.....	L1a
BR58	1	C.....	T.....G.....	C.....T.....	L1a
BR59	1	C.....	T.....T.....	C.....C.....	L1b
BR60	1	T.....	T.....T.....	C.....C.....	L1b
BR61	1	C.....	T.....T.....	C.....C.....	L1b
BR62	1	A.....	T.....T.....	C.....C.....	L1c
BR63	1	C.....	T.....T.....	C.....C.....	L1c
BR64	1	T.....	T.....T.....	C.....C.....	L1c
BR65	1	A.....	T.....T.....	C.....C.....	L1c
BR66	2	A.....	T.....T.....	C.....C.....	L1c
BR67	1	T.....	T.....T.....	C.....C.....	L1c
BR68	1	T.....	T.....T.....	C.....C.....	L1c
BR69	1	A.....	T.....T.....	C.....C.....	L1c
BR70	1	A.....	T.....T.....	C.....C.....	L1c
BR71	1	C.....	T.....T.....	C.....C.....	L1c
BR72	1	A.....	T.....T.....	C.....C.....	L1c
BR73	1	C.....	T.....T.....	C.....C.....	L1c
BR74	1	T.....	T.....T.....	C.....G.....	L2
BR75	1	C.....	T.....T.....	G.....	L2
BR76	1	CT.....	T.....T.....	G.....	L2
BR77	1	T.....	T.....T.....	G.....	L2
BR78	1	T.....	T.....T.....	G.....	L2
BR79	1	T.....	T.....T.....	G.....	L2
BR80	1	A.....	T.....T.....	C.....T.....	L2
BR81	2	T.....	T.....T.....	C.....	L2
BR82	1	T.....	T.....T.....	C.....	L2
BR83	1	A.....	T.....T.....	T.....	L2
BR84	1	A.....	T.....T.....	T.....	L2
BR85	1	TC.....	T.....T.....	T.....	L2
BR86	1	C.....	T.....T.....	C.....	L3d
BR87	1	C.....	T.....T.....	C.....	L3d
BR88	1	C.....	T.....T.....	C.....	L3d
BR89	1	C.....	T.....T.....	T.....	L3d
BR90	1	T.....	T.....T.....	T.....	L3e

BR131	1	1	1	T.....	C...C.....	pre*V
BR132					C.....	pre*V
BR133	1	1		T.....	C.....	V
BR134	1			C.....	C.....	V
BR135		1		A.....	C.....	V
BR136	1			C.....	C.....	V
BR137	2				C.....T.....	V
BR138		1		A.....		HV*
BR139	1		1	C.....	C.....	U2
BR140	1		1	T.....	C.....C.....	U4
BR141	1			C.....	T...A.....	U5b*
BR142		1		C.....	T.....	U5b*
BR143	1		1	C.....	T.....	U5b1
BR144	1		1	C.....	T.....	U7
BR145					G...T.....	K
BR146	1	1	1		C.....	K
BR147				C.....	C.....	K
BR148	1		1	A.....	C.....	K
BR149			1	C.....	C...A.....	K
BR150		1		C.....	C...A.....	K
BR151		1		C.....	C.....	pre*HV
BR152	1	2	1	T.....	J.....	J*
BR153		1		T.....	T.....	J*
BR154		1		C.....	T.....	J*
BR155	1		1	C...A.....	T.....	J1*
BR156		2		C...A.....	T.....	J1b1
BR157			1	C.....	T...C.....	T*
BR158	1	1	1	C.....	T...C.....	T*
BR159			1	C.....	T...C.....	T*
BR160				C.....	T...C.....	T*
BR161	1			C.....	T...C.....	T*
BR162		1		C...A.....	T...C.....	T*
BR163			1	C...A.....	T...C.....	T*
BR164			1	T...C...A.....	T...T.....	T*
BR165	1			C.....	T...T.....	T*
BR166		1	1	C.....	T...T.....	T1
BR167			1	A.....	T.....	I
BR168			1	C.....	T.....	X
BR169	1			C.....	G...T.....	X
BR170	1			C.....	T.....	X
Total	99	50	48			

^a BR13 and BR14 share the loss of 3534 *Ddel*, whereas controls BR4, BR5, BR15, and BR16 were +3534 *Ddel*. BR132 has C at np 72, and for BR139 the U2-characteristic transition at np 16051 has been confirmed.

^b SE = southeastern, S = southern, NE = northeastern, N = northern.

^c The prefix "16" has been deleted from all (three-digit) numbers. Length polymorphisms in the C run (np 16184 and 16193) are disregarded.

^d Confirmed by testing required restriction sites, listed in table 1, for each lineage (except for sites +9070 *TaqI* and +12810 *RsaI* in BR69 and BR70); more-extensive screening was performed for the lineages of status L3e*, pre*V, and pre*HV (see table 3).

^e Source: Andrews et al. (1999).

Table 3

RFLP Sites Screened in Some mtDNAs

POLYMORPHISM	STATUS OF SAMPLE ^a						
	BR107	BR108	BR109	BR131		BR132	BR151
				Southeastern	Northern		
9-bp deletion	—	—	—	—	—	—	—
663 <i>Hae</i> III	—	—	—	—	—	—	—
1715 <i>Dde</i> I	+	+	+	+	+	+	+
2349 <i>Mbo</i> I	—	—	—	—	—	—	—
3592 <i>Hpa</i> I	—	—	—	—	—	—	—
4216 <i>Nla</i> III	ND	ND	—	—	—	—	—
4529 <i>Hae</i> II	+	+	+	+	+	+	+
4577 <i>Nla</i> III	ND	ND	+	+	+	+	+
5176 <i>Alu</i> I	+	+	+	+	+	+	+
7025 <i>Alu</i> I	+	+	+	+	+	+	+
8249 <i>Ava</i> II	—	—	—	—	—	—	—
8616 <i>Mbo</i> I	+	+	+	+	+	+	+
8994 <i>Hae</i> III	ND	+	+	+	+	+	+
9052 <i>Hae</i> II	+	+	+	+	+	+	+
10032 <i>Alu</i> I	—	—	—	—	—	—	—
10084 <i>Taq</i> I	—	—	—	—	—	—	—
10394 <i>Dde</i> I	+	+	+	—	—	—	—
10397 <i>Alu</i> I	ND	—	—	—	—	—	—
12308 <i>Hinf</i> I	—	—	—	—	—	—	—
12406 <i>Hpa</i> I	—	—	+	—	+	+	+
13259 <i>Hinc</i> II	+	+	+	+	+	+	+
13366 <i>Bam</i> HI	—	—	—	—	—	—	—
13704 <i>Bst</i> NI	+	+	+	+	+	+	+
14766 <i>Mse</i> I	ND	+	ND	—	—	—	+
15606 <i>Alu</i> I	—	—	—	—	—	—	—
16389 <i>Bam</i> HI	—	—	—	—	—	—	—
Haplogroup	L3*	L3*	L3*	pre*V	pre*V	pre*V	pre*HV

^a A plus sign (+) denotes presence; a minus sign (—) denotes absence; and ND = not done.

ages are of Native American ancestry, whereas African ancestry is most prominent in the northeastern region, with the southeastern region being intermediate between the two former regions in this regard; finally, the southern region stands out, with a great majority of European mtDNA lineages (table 4).

Native American Fraction of the Brazilian mtDNA Pool

Haplogroup A is the most frequent haplogroup within the Native American fraction, closely followed by haplogroup B, with C coming next and D last (table 5). As to the regional distribution, the same order of frequencies is observed in the southern and southeastern regions, but C is the leading haplogroup in the northern region, whereas it is the least frequent in the northeastern region.

The major founder haplotypes of the Native American haplogroups A–D (Forster et al. 1996; Smith et al. 1999) are all present in Brazil and are shared with many Native American populations. Interestingly, there are several matches with derived haplotypes in other South American populations: haplotypes BR26, BR27, BR43, and BR47 have been observed elsewhere, in the Amazonian sample of Santos et al. (1996b); haplotype BR14 has

also been identified, with the absence of the site 3534 *Dde*I (table 2, footnote a), in the Krahó from Goiás, Brazil (Torroni et al. 1993); BR51 occurs in the Zoró from Mato Grosso, Brazil, and in the Gavião from Rondônia, Brazil (Ward et al. 1996); haplotype BR39 is found in Colombia (Horai et al. 1993); and BR49 is found in the Mapuche from Argentina (Ginther et al. 1993). Remarkably, one D lineage (BR53) constitutes a true outlier in our sample, since it differs by as many as seven transitions from the D founder haplotype, four of which are shared with a haplotype found in the Cayapa from Ecuador (Rickards et al. 1999).

African Fraction of the Brazilian mtDNA Pool

Haplogroups L3e and L1c together constitute approximately one-half (49%) of the African fraction (table 5). Nowhere in western Africa has such a high percentage been observed so far for this haplogroup pair: 7% is seen in the mtDNA pool of several Senegalese populations (Graven et al. 1995; Rando et al. 1998), and 17% has been seen in the joint mtDNA pool of the Songhai, Yoruba, Hausa, and Kanuri (Watson et al. 1997, table A1). The Bubi, from the island of Bioko in

Table 4**Frequency of Continent-Specific mtDNA Haplotypes in the Brazilian mtDNA Pool**

CONTINENTAL FRACTION	FREQUENCY				
	Brazil	Northern	Northeastern	Southeastern	Southern
Native American	.33	.54	.22	.33 ^a	.22
African	.28	.15	.44	.34	.12
European	.39	.31	.34	.31	.66

^a Excludes the single lineage of confirmed Asian ancestry.

the Gulf of Guinea (central Africa), however, have an mtDNA contribution of 33% L3e and 22% L1c, yielding a joint percentage of 56% (as inferred from table 2 of Mateu et al. 1997). This exceeds even the corresponding relative frequency in our Brazilian sample. On the other hand, haplogroups L3d and L1b, which are quite specific to western Africa, are absent in the Bubi but together occur at a frequency of 10% in the African fraction of the Brazilian mtDNA pool. For this haplogroup pair, much higher frequencies are found in western Africa: 25% in Senegal and 17% among the Songhai, Yoruba, Hausa, and Kanuri. This suggests that the majority of the mtDNA lineages of African ancestry in the Brazilian sample had their origin in central Africa (which would include Cameroon, as well as Angola), although a substantial number must have come from western Africa. Only few mtDNA haplotypes in the Brazilian sample could potentially testify to southeastern-African origin. BR55 would be a good candidate since it perfectly matches the most frequent 9-bp-deletion L1a haplotype, found in Malawi and in southeastern Bantu speakers (Soodyall et al. 1996): the matching involves the 9-bp deletion, HVS-I (table 2), and HVS-II (table 6), even extending to position 64 (which, in this case, could be read with only one primer). The same HVS-I type (with the 9-bp deletion) has been found in São Tomé (Mateu et al. 1997), which served as an entrepôt for the Atlantic slave trade (Curtin 1969).

European Fraction of the Brazilian mtDNA Pool

The Brazilian sample includes mtDNA lineages from almost all the familiar European haplogroups (Torroni et al. 1996; Macaulay et al. 1999), except for some marginal ones, such as W and other quite-rare haplogroups related to haplogroup I (Kivisild et al. 1999b). The frequency of the dominant haplogroup H (44%; table 5) in the European fraction is somewhat higher, on average, than that observed in Europe but is well within the range of western-European H frequencies (Torroni et al. 1998). In particular, the relatively high frequency of the Cambridge reference sequence (CRS [Andrews et al. 1999]) haplotype (24%) and of haplogroup pre-V (9%) suggests predominantly western-European ancestry.

The majority of haplotypes from the European frac-

tion already have been recorded in the European mtDNA pool (Richards et al. 1996, 1998; Helgason et al. 2000) and, in many instances, match mtDNA lineages from the Iberian Peninsula. Nevertheless, there are a few sequences for which one would not predict southwestern-European ancestry; the most striking example is the U5b1 haplotype BR143, which bears the “Saami motif” (Sajantila et al. 1995) and is thus of northern Fennoscandian origin.

HVS-II Motifs for Classification of Brazilian mtDNA Lineages

To see to what extent a first sorting into haplogroups could be based on HVS-II sequences, we have sequenced all but one of the mtDNA lineages of the southeastern-Brazilian sample, for HVS-II (table 6). It turns out, by comparison with the data of Vigilant (1990), that haplogroups L1a, L1b, L1c, and an unnamed haplogroup (within L3*) are readily identified individually by one to three positions. L1 is separated from L2 and L3 by a transition at np 247 (fig. 2), which is also evident from the data of Graven et al. (1995). Curiously, np 247 was sorted into the class of most highly mutable positions in HVS-II, by Meyer et al. (1999), although it is known to be virtually unvaried in European mtDNA (see the data sets of Piercy et al. [1993] and Helgason et al. [2000]). The two major African haplogroups, L2 and L3e, represented in the southeastern-Brazilian data set exhibit only diagnostic positions that seem to be hyper-variable (np 146, 150, 152, and 195), although HVS-II certainly offers some information for internal classification of L2. Among the Native American haplogroups, only A and C can be detected using HVS-II alone: A is recognized by a transition at np 235, and C is recognized by two deletions (fig. 2). The major HVS-II polymorphism for European mtDNA is at np 73 (Côrte-Real et al. 1996): 73A characterizes pre-HV. Note that 73A is also characteristic of the African haplogroup L1a, but that, in general, np 73 seems to be fairly stable: no further parallel mutation from G to A has been observed, either in our data or in those of Graven et al. (1995), Brown et al. (1998), Macaulay et al. (1999), and Helgason et al. (2000), a result that is at variance with the findings of Salas et al. (1998), who reported J, K, and

Table 5**Haplogroup Frequencies within the Three Continental Fractions of Brazilian mtDNA Pool**

HAPLOGROUP	FREQUENCY IN BRAZIL				
	Overall	Northern	Northeastern	Southeastern	Southern
Native American: ^a					
A	.30	.15	.37	.39	.27
B	.29	.31	.27	.30	.27
C	.24	.38	.09	.18	.27
D	.16	.15	.27	.12	.18
Total	1.00	1.00	1.00	1.00	1.00
African:					
L1a	.10	...	.18	.06	.17
L1b	.04	...	.05	.03	.17
L1c	.19	.29	.09	.23	.17
L2	.20	.14	.23	.23	...
L3d	.06	...	.09	...	.33
L3e	.30	.43	.32	.32	...
L3*	.04	...	.05	.06	...
U6	.06	.14	...	.06	.17
Total	1.00	1.00	1.00	1.00	1.00
European:					
H	.44	.27	.65	.45	.39
pre*V	.03	.07	...	.03	.03
V	.06	...	.06	.13	.03
HV*	.01	...	...	...	.03
U	.16	.13	.18	.16	.15
pre*HV	.01	...	...	...	.03
J	.11	.20	.06	.03	.18
T	.14	.27	.06	.13	.12
I	.01	.07	...	...	...
X	.03	...	...	.06	.03
Total	1.00	1.00	1.00	1.00	1.00

^a Excludes the single A lineage (from the southeastern region) of confirmed Asian ancestry.

T lineages with 73A. The few back mutations from A to G that are observed in haplogroup H (Helgason et al. 2000) and L1a (Soodyall et al. 1996) may testify to a directional bias in substitution rates at np 73. Furthermore, only haplogroups pre-V, J, and X can be recognized by single positions in HVS-II; otherwise, HVS-II is not of much help for allocation to European haplogroups. In summary, even this rough HVS-II classification scheme would permit allocation of the majority of the Brazilian mtDNA lineages to the three continental fractions.

Discussion

We examined individuals from four different regions in Brazil (fig. 1), in an attempt to establish a portrait of the mtDNA variation throughout the country and to determine the relative matrilineal contributions of Europeans, Amerindians/Asians, and Africans to present-day white Brazilians. The total sample revealed as much as 33% Amerindian and 28% African contribution to the total mtDNA pool. In fact, these values are probably

minimum percentages, because, since our study group is primarily composed of middle- and upper-middle-class Brazilians, a bias toward a higher contribution of European mtDNA is to be expected.

Most of the studies involving the Brazilian population have been performed on Amerindian tribes, mainly from the Amazonian region (Schurr et al. 1990; Torroni et al. 1993; Bailliet et al. 1994; Bianchi et al. 1995; Bortolini and Salzano 1996; Easton et al. 1996; Santos et al. 1996b; Ward et al. 1996; Bortolini et al. 1998a), and predominantly African-derived groups (Schneider et al. 1987; Bortolini et al. 1992, 1995, 1997a, 1997b, 1998b; Arpini-Sampaio et al. 1999; Guerreiro et al. 1999). A small number of studies performed on white Brazilians (deemed to be mostly of European descent) that used mainly protein genetic systems focused on populations from the southern or northern regions (Franco et al. 1982; Rosa et al. 1984; Moraes et al. 1993; Santos et al. 1996a; Batista dos Santos et al. 1999; Dornelles et al. 1999). These studies showed that the amount of Amerindian ancestry in the white Brazilians varies widely and has distinct patterns in the different

regions of the country. The highest Amerindian influence was observed in populations of the Amazonian region, where a study analyzing 11 urban populations by use of nuclear markers observed an average of 41% Amerindian ancestry (Santos and Guerreiro 1995). Recent mtDNA analysis of another population from the northern region showed an even higher Amerindian contribution (59% [Batista dos Santos et al. 1999]). In our sample, we also observed a high Amerindian influence in the northern region (54%), corroborating the mtDNA data obtained by Batista dos Santos et al. (1999). In the other regions of Brazil, the genetic contribution from Amerindians is also markedly higher for mtDNA than for nuclear DNA: 22% (table 4) versus 13% (Krieger et al. 1965; Franco et al. 1982; Conceição et al. 1987) in the northeastern region and 22% (table 4) versus 11% (Dornelles et al. 1999) in the southern region. For the southeastern region, where we have detected 33% frequency of mtDNA lineages of Amerindian ancestry, not a single study with nuclear markers has yet been performed, probably because one did not anticipate measurable Amerindian genetic influence on urban populations (Salzano 1997). As for African admixture in the white Brazilian population, the picture is similar to what we have seen for the Amerindian genetic input: mtDNA analysis (table 4) suggests a higher contribution than that by nuclear markers, for which 12% in the northern region (Santos et al. 1996a), 36% in the northeastern region (Franco et al. 1982; Arpini-Sampaio et al. 1999), and 10% in the southern region (Dornelles et al. 1999) were reported; again, no nuclear data are available for the southeastern region.

The allocation of haplogroups to continents (as indicated in table 1) is, of course, not absolutely clear-cut. For instance, the European haplogroups H and U5 do occur in the sub-Saharan mtDNA pool, albeit in only two founder types (bearing transitions at np 16145 and 16222 for the H type and transitions at np 16189, 16192, 16270, and 16320 for the U5 type), possibly transmitted by Berbers or, even earlier, during the Saharan Neolithic age (Rando et al. 1999). None of these particular mtDNA lineages occur in our Brazilian sample. Similarly, northern-African U6 haplotypes have penetrated the Sahara and are found sporadically from the west (Senegal) to the east (Kenya). We consider it most plausible that the four U6 lineages in our sample have come from western Africa. On the other hand, African haplotypes were also transmitted, in low numbers, to Europe, especially to the Mediterranean area. African mtDNA lineages, then, constitute erratic outliers in the respective mtDNA samples, for instance, such as the L1c lineage in the British data of Piercy et al. (1993).

There is one caveat with regard to the distinction between European mtDNA haplotypes and Native

American ones: haplogroup X is shared by western Eurasia and North America (Brown et al. 1998; Smith et al. 1999), although there is as yet no compelling evidence for the occurrence of haplogroup X in Central or South America. The three X haplotypes that we detected in the Brazilian sample are certainly of European ancestry, since BR169 and BR170 do not bear the np-200 transition that is characteristic of (most of) the Native American haplogroup X (Brown et al. 1998), whereas BR168 (for which no HVS-II information is available) bears a transition (namely, at np 16248) already observed in Europe (Richards et al. 1998).

The distinction between Asian and Native American mtDNA haplotypes is more intricate inasmuch as haplogroups A–D are of Asian origin. Fortunately, the Native American A, C, and D founder HVS-I and HVS-II types can be distinguished from Asian haplotypes by mutations that are virtually absent, or at least rare, in Asia. The transition at np 16325 is (almost) diagnostic for Native American C and D haplotypes; the 2-bp deletion in HVS-II seems to be characteristic of Native American C (table 6; also see Ginther et al. 1993; Kolman and Bermingham 1997), since it has not been reported in Asian mtDNAs so far (Lee et al. 1997). The “Beringian” transition at np 16111 is seen in most Native American A lineages but is virtually absent in Asia (Horai and Hayasaka 1990; Horai et al. 1993; Torroni et al. 1993; Kolman et al. 1996; Lee et al. 1997). Thus, only haplotype BR16, which, incidentally, matches an mtDNA lineage from Hokkaido (Horai et al. 1996), would be a clear suspect for potential Asian ancestry. In fact, the lineage BR16 (from a student in our laboratory) turned out to be of Lebanese matrilineal ancestry, but its ultimate ancestry would be central/eastern Asian. Taking into account that the mtDNA haplotypes of seemingly Native American ancestry constitute only a small minority in the Asian mtDNA pool, we would realistically assume that no more than, say, one additional mtDNA lineage in our Brazilian sample was actually of eastern-Asian ancestry. The same could be said for western-Asian ancestry, inasmuch as, in our sample, we did not observe any U1, U3, R*, N*, M*, or non-Amerindian C haplotypes that would occur at considerable frequency in western Asia (Macaulay et al. 1999).

One lineage classified as haplogroup D, BR53, lacks the characteristic transition at np 16325 (very possibly because of back mutation) and is considerably different from most other Native American D lineages (Torroni et al. 1993; Horai et al. 1996; Santos et al. 1996b; Ward et al. 1996; Stone and Stoneking 1998); however, exactly as does the unique D lineage of the Mexican sample (Green et al. 2000), it shares three distinguishing mutations (at np 16241, 16301, and 16342) with a haplotype detected in the Cayapa from Ecuador (Rickards et al. 1999). The Cayapa haplotype also bears the

HVS-II Haplotypes in Southeastern Brazil

[illegible]

[illegible]

^b Haplogroup-diagnostic nucleotide positions are underlined.

In principle, it should be possible to narrow the matrilineal ancestry of Brazilians to a geographic scale narrower than that of the (sub)continents (e.g., see Alves-Silva et al. 1999). This, in general, requires extensive phylogeographic studies of the populations from the potential source areas, which, at present, are not avail-

able for Africa or most parts of Europe—in particular, Italy. Considering that 30% of the European immigrants (including the Portuguese colonizers) to Brazil came from Italy (Salzano and Freire Maia 1967; Callegari-Jacques and Salzano 1999), one can expect that a considerable number of mtDNA lineages in the Brazilian sample have Italian ancestry. In one case (a student in our laboratory), we could indeed confirm Italian matrilineal ancestry (BR137). Given the paucity of published Italian HVS-I mtDNA data (Francalacci et al. 1996; Torroni et al. 1996), we have practically no diagnostic haplotypes or characteristic haplogroup profiles at hand from which one could extrapolate the Italian mitochondrial input to the regional mtDNA pools

of Brazil. It needs to be emphasized that genetic distances, although trivial to compute, between the Brazilian mtDNA sample and mtDNA samples representing potential source populations would not allow the calculation of reliable admixture proportions, as demonstrated by Rando et al. (1999) in the case of the mixed population from the Canary Islands.

One could also use a reverse approach and infer the mtDNA profile of a source population from that of the target mixed population (given sufficient information on the other participating source populations). For instance, no mtDNA data for Angola are available—yet, since Angola was the major source of African slaves brought to Brazil, we can make inferences on how the mtDNA pool of Angolans would look: we should expect (i) a considerable number of L3e lineages—in particular, those bearing the np-16327 transition, also observed in the Herero and in other southern African populations (Vigilant 1990; Soodyall 1993); (ii) a possibly equal amount of L1c lineages, several of which will show the transversions at np 16265 and 16286, detected in Equatorial Guinea (Pinto et al. 1996) as well as in Namibia (Soodyall 1993); and, furthermore, (iii) some L2 lineages (which seem to be omnipresent in Bantu populations) but probably no Khoisan-specific mtDNA haplotypes from the L1 subgroup described by Bandelt and Forster (1997). The seven northeastern-Brazilian lineages (except for the U6 lineage) could therefore represent a typical Angolan minisample. The white Brazilian population, paradoxically, seems to be an excellent resource with which to study the phylogeny of western- and central-African mtDNA.

In conclusion, our mtDNA study of a random sample of white Brazilians has revealed an astonishingly high matrilineal contribution of Amerindians and Africans. Present-day Brazilians thus still carry the genetic imprint of the early-colonization phase: the pioneer-colonial population typically had Amerindian ancestry—and, after few generations, increasingly African ancestry—in the maternal line but Portuguese ancestry in the paternal line (as is reflected by Y-chromosome markers [D. R. Carvalho-Silva, F. R. Santos, and S. D. J. Pena, unpublished results]).

Acknowledgments

We would like to thank Katia Barroso and Neuza Antunes Rodrigues for technical assistance and Toomas Kivisild (Tartu) for helpful information on mtDNA classification. This research was supported by grants from Conselho Nacional de Pesquisa, Fundação de Amparo à Pesquisa do Estado de Minas Gerais, and Pró-Reitoria de Pesquisa da Universidade Federal de Minas Gerais and a travel grant from Deutscher Akademischer Austauschdienst and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (to H.-J.B. and S.D.J.P.).

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GenBank Overview, <http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html> (for HVS-I [accession numbers AF243627–AF243796] and HVS-II [accession numbers AF243539–AF243626])

Instituto Brasileiro de Geografia Estatística, <http://www.ibge.gov.br/>

References

- Alves-Silva J, Guimarães PE, Rocha J, Pena SD, Prado VF (1999) Identification in Portugal and Brazil of a mtDNA lineage containing a 9-bp triplication of the intergenic COII/tRNA_{Lys} region. *Hum Hered* 49:56–58
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147
- Arpini-Sampaio Z, Costa MC, Melo AA, Carvalho MF, Deus MS, Simões AL (1999) Genetic polymorphisms and ethnic admixture in African-derived black communities of northeastern Brazil. *Hum Biol* 71:69–85
- Bailliet G, Rothhammer F, Carnese FR, Bravi CM, Bianchi NO (1994) Founder mitochondrial haplotypes in Amerindian populations. *Am J Hum Genet* 55:27–33
- Bandelt H-J, Forster P (1997) The myth of bumpy hunter-gatherer mismatch distributions. *Am J Hum Genet* 61:980–983
- Batista dos Santos SE, Rodrigues JD, Ribeiro-dos-Santos AKC, Zago MA (1999) Differential contribution of indigenous men and women to the formation of an urban population in the Amazon region as revealed by mtDNA and Y-DNA. *Am J Phys Anthropol* 109:175–180
- Bethell L (1997) Nota sobre as populações americanas às vésperas das invasões européias. In: Bethell L (ed) *América Latina colonial*. Editora da Universidade de São Paulo, São Paulo, pp 129–131
- Bianchi NO, Bailliet G, Bravi CM (1995) Peopling of the Americas as inferred through the analysis of mitochondrial DNA. *Braz J Genet* 18:661–668
- Bortolini MC, Baptista C, Callegari-Jacques SM, Weimer TA, Salzano FM (1998a) Diversity in protein, nuclear DNA, and mtDNA in South Amerinds—agreement or discrepancy? *Ann Hum Genet* 62:133–145
- Bortolini MC, da Silva-Junior WA, Weimer TdA, Zago MA, de Guerra DC, Schneider MP, Layrisse Z, et al (1998b) Protein and hypervariable tandem repeat diversity in eight African-derived South American populations: inferred relationships do not coincide. *Hum Biol* 70:443–461
- Bortolini MC, Salzano FM (1996) mtDNA diversity analysis in Amerindians and other human populations—how different are they? *Braz J Genet* 19:527–534
- Bortolini MC, Salzano FM, Zago MA, Da Silva Junior WA, Weimer TdA (1997a) Genetic variability in two Brazilian ethnic groups: a comparison of mitochondrial and protein data. *Am J Phys Anthropol* 103:147–156

- Bortolini MC, Weimer TA, Franco MH, Salzano FM, Layrisse Z, Schneider H, Schneider MP, et al (1992) Genetic studies in three South American black populations. *Gene Geogr* 6: 1–16
- Bortolini MC, Weimer TDA, Salzano FM, Callegari-Jacques SM, Schneider H, Layrisse Z, Bonatto SL (1995) Evolutionary relationships between black South American and African populations. *Hum Biol* 67:547–559
- Bortolini MC, Zago MA, Salzano FM, Silva-Junior WA, Bonatto SL, da Silva MC, Weimer TA (1997b) Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. *Hum Biol* 69:141–159
- Bravi CM, Sans M, Bailliet G, Martinez-Marignac VL, Portas M, Barreto I, Bonilla C, et al (1997) Characterization of mitochondrial DNA and Y-chromosome haplotypes in a Uruguayan population of African ancestry. *Hum Biol* 69: 641–652
- Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, et al (1998) mtDNA haplogroup X: an ancient link between Europe/western Asia and North America? *Am J Hum Genet* 63:1852–1861
- Callegari-Jacques SM, Salzano FM (1999) Brazilian Indian/non-Indian interactions and their effects. *Ciênc Cult* 51: 166–174
- Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 57:133–149
- Conceição MM, Salzano FM, Franco MHL, Weimer TA, Krieger H (1987) Demography, genetics, and race admixture in Aracaju, Brazil. *Braz J Genet* 10:313–331
- Côrte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, et al (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* 60:331–350
- Curtin PD (1969) *The Atlantic slave trade: a census*. University of Wisconsin Press, Madison
- Dornelles CL, Callegari-Jacques SM, Robinson WM, Weimer TA, Franco MHL, Hickmann AC, Geiger CJ, et al (1999) Genetics, surnames, grandparents' nationalities, and ethnic admixture in Southern Brazil: do the patterns of variation coincide? *Genet Mol Biol* 22:151–161
- Easton RD, Merriwether DA, Crews DE, Ferrell RE (1996) mtDNA variation in the Yanomami: evidence for additional New World founding lineages. *Am J Hum Genet* 59:213–225
- Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59:935–945
- Francalacci P, Bertranpetit J, Calafell F, Underhill PA (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am J Phys Anthropol* 100:443–460
- Franco MH, Weimer TA, Salzano FM (1982) Blood polymorphisms and racial admixture in two Brazilian populations. *Am J Phys Anthropol* 58:127–132
- Ginther C, Corach D, Penacino GA, Rey JA, Carnese FR, Hutz MH, Anderson A, et al (1993) Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. *EXS* 67:211–219
- Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* 12:334–345
- Green LD, Derr JN, Knight A (2000) mtDNA affinities of the peoples of north-central Mexico. *Am J Hum Genet* 66: 989–998
- Guerreiro JF, Ribeiro-dos-Santos AKC, Santos EJM, Vallinoto ACR, Cayres-Vallinoto IMV, Aguiar GFS, Santos SEB (1999) Genetical-demographic data from two Amazonian populations composed of descendants of African slaves: Pacoval and Curiau. *Genet Mol Biol* 22:163–167
- Helgason A, Sigurðardóttir S, Gulcher JR, Ward R, Stefánsson K (2000) mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* 66:999–1016
- Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am J Hum Genet* 46:828–842
- Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci USA* 92:532–536
- Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol Biol Evol* 10:23–47
- Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, et al (1996) mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *Am J Hum Genet* 59:579–590
- Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, et al (1999a) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9:1331–1334
- Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha S, Villems R (1999b) The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the old world. In: Papiha G, Deka R, Chakraborty R (eds) *Genomic diversity: applications in human population genetics*. Kluwer Academic/Plenum, New York, pp 135–152
- Kolman CJ, Bermingham E (1997) Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. *Genetics* 147:1289–1302
- Kolman CJ, Sambuughin N, Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142:1321–1334
- Krieger H, Morton NE, Mi MP, Azevedo E, Freire-Maia A, Yasuda N (1965) Racial admixture in north-eastern Brazil. *Ann Hum Genet* 29:113–125
- Krings M, Geisert H, Schmitz RW, Krainitzki H, Pääbo S (1999) DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proc Natl Acad Sci USA* 96:5581–5585
- Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19–30

- Lee SD, Shin CH, Kim KB, Lee YS, Lee JB (1997) Sequence variation of mitochondrial DNA control region in Koreans. *Forensic Sci Int* 87:99-116
- Macauley V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, et al (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64:232-249
- Mateu E, Comas D, Calafell F, Perez-Lezaun A, Abade A, Bertranpetit J (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann Hum Genet* 61:507-518
- Metspalu E, Kivisild T, Kaldma K, Parik J, Reidla M, Tambets K, Vilems R (1999) The trans-Caucasus and the expansion of the caucasoid-specific human mitochondrial DNA. In: Papiha G, Deka R, Chakraborty R (eds) *Genomic diversity: applications in human population genetics*. Kluwer Academic/Plenum, New York, pp 121-133
- Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152:1103-1110
- Monteiro JM (1994) Negros da terra, índios e bandeirantes nas origens de São Paulo. Companhia da Letras, São Paulo
- Moraes ME, Fernandez-Vina M, Salatiel I, Tsai S, Moraes JR, Stastny P (1993) HLA class II DNA typing in two Brazilian populations. *Tissue Antigens* 41:238-242
- Mörner M (1967) *Race mixture in the history of Latin America*. Little, Brown, Boston
- Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* 106:85-90
- Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet* 60:321-330
- Rando JC, Cabrera VM, Larruga JM, Hernández M, González AM, Pinto F, Bandelt H-J (1999) Phylogeographic patterns of mtDNA reflecting the colonisation of the Canary Islands. *Ann Hum Genet* 63:413-428
- Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt H-J (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet* 62:531-550
- Ribeiro D (1995) *O povo brasileiro: a formação e o sentido do Brasil*. Companhia da Letras, São Paulo
- Richards M, Côrte-Real H, Forster P, Macauley V, Wilkinson-Herbots H, Demaine A, Papiha S, et al (1996) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59:185-203
- Richards MB, Macauley VA, Bandelt H-J, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62:241-260
- Rickards O, Martinez-Labarga C, Lum JK, De Stefano GF, Cann RL (1999) mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. *Am J Hum Genet* 65:519-530
- Rosa VL, Salzano FM, Franco MH, Freitas MJ (1984) Blood genetic studies in five Amazonian populations. *Braz J Genet* 75:69-82
- Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, et al (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res* 5:42-52
- Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* 6:365-375
- Salzano FM (1997) Human races: myth, invention or reality? *Interdisciplinaria* 22:221-227
- Salzano FM, Freire-Maia N (1967) *Populações brasileiras, aspectos demográficos, genéticos e antropológicos*. Companhia Editora Nacional, São Paulo
- (1970) *Problems in human biology: a study of Brazilian populations*. Wayne State University Press, Detroit
- Santos EJM, Ribeiro-dos-Santos AKC, Guerreiro JF, Aguiar GFS, Santos SEB (1996a) Migration and ethnic change in an admixed population from the Amazon region (Santarém, Pará). *Braz J Genet* 19:511-515
- Santos SEB, Guerreiro JF (1995) The indigenous contribution to the formation of the population of the Brazilian Amazon region. *Braz J Genet* 18:311-315
- Santos SE, Ribeiro dos Santos AKC, Meyer D, Zago MA (1996b) Multiple founder haplotypes of mitochondrial DNA in Amerindians revealed by RFLP and sequencing. *Ann Hum Genet* 60:305-319
- Santos SEB, Ribeiro dos Santos AKC, Santos EJM, Guerreiro JF (1999) The Amazonian microcosm. *Ciênc Cult* 51:181-190
- Schneider H, Guerreiro JF, Santos SEB, Weimer TA, Schneider MPC, Salzano FM (1987) Isolate breakdown in Amazonia: the blacks of the Trombetas river. *Braz J Genet* 10:565-574
- Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, et al (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet* 46:613-623
- Smith DG, Malhi RS, Eshleman J, Lorenz JG, Kaestle FA (1999) Distribution of mtDNA haplogroup X among Native North Americans. *Am J Phys Anthropol* 110:271-284
- Soodyall H (1993) *Mitochondrial DNA polymorphisms in southern African populations*. PhD thesis, University of Witwatersrand, Johannesburg
- Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. *Am J Hum Genet* 58:595-608
- Stone AC, 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
- Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am J Hum Genet* 48:370-382
- Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Selitto D, Rengo C, et al (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62:1137-1152
- Torroni A, Brown MD, Lott MT, Newman NJ, Wallace D, Cuba Neuropathy Field Investigation Team (1995) African,

- Native American, and European mitochondrial DNAs in Cubans from Pinar del Rio Province and implications for the recent epidemic neuropathy in Cuba. *Hum Mutat* 5: 310–317
- Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, et al (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835–1850
- Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC (1994) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet* 55:760–776
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, et al (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53:563–590
- Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, et al (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153–162
- Vigilant L (1990) Control region sequences from African population and the evolution of human mitochondrial DNA. PhD thesis, University of California, Berkeley
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507
- Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA Jr, Santos RV (1996) Mitochondrial DNA polymorphism in three Brazilian tribes. *Am J Hum Biol* 8:317–323
- Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691–704

Apêndice 3

Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade

H.-J. BANDELT¹, J. ALVES-SILVA², P. E. M. GUIMARÃES², M. S. SANTOS², A. BREHM³,
L. PEREIRA^{4,5}, A. COPPA⁶, J. M. LARRUGA⁷, C. RENGO^{8,9}, R. SCOZZARI⁸,
A. TORRONI^{8,10}, M. J. PRATA^{4,5}, A. AMORIM^{4,5}, V. F. PRADO² AND S. D. J. PENA²

¹*Fachbereich Mathematik, Universität Hamburg, 20146 Hamburg, Germany*

²*Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte-MG, Brazil 30161-970*

³*Centro de Ciências Biológicas e Geológicas, Universidade da Madeira, 9000 Funchal, Portugal*

⁴*IPATIMUP (Instituto de Patologia e Imunologia Molecular da Universidade do Porto), 4200 Porto, Portugal*

⁵*Faculdade de Ciências da Universidade do Porto, 4099-002 Porto, Portugal*

⁶*Dipartimento di Biologia Animale e dell'Uomo, Università 'La Sapienza', 00185 Roma, Italy*

⁷*Departamento de Genética, Universidad de La Laguna, 38271 La Laguna, Tenerife, Spain*

⁸*Dipartimento di Genetica e Biologia Molecolare, Università 'La Sapienza', 00185 Roma, Italy*

⁹*Istituto di Medicina Legale, Università Cattolica del Sacro Cuore, 00168 Roma, Italy*

¹⁰*Dipartimento di Genetica e Microbiologia, Università di Pavia, 27100 Pavia, Italy*

(Received 11.5.01. Accepted 31.8.01)

SUMMARY

The mtDNA haplogroup L3e, which is identified by the restriction site +2349 *Mbo*I within the Afro-Eurasian superhaplogroup L3 (–3592 *Hpa*I), is omnipresent in Africa but virtually absent in Eurasia (except for neighbouring areas with limited genetic exchange). L3e was hitherto poorly characterised in terms of HVS-I motifs, as the ancestral HVS-I type of L3e cannot be distinguished from the putative HVS-I ancestor of the entire L3 (differing from the CRS by a transition at np 16223). An *Mbo*I screening at np 2349 of a large number of Brazilian and Caribbean mtDNAs (encompassing numerous mtDNAs of African ancestry), now reveals that L3e is subdivided into four principal clades, each characterised by a single mutation in HVS-I, with additional support coming from HVS-II and partial RFLP analysis. The apparently oldest of these clades (transition at np 16327) occurs mainly in central Africa and was probably carried to southern Africa with the Bantu expansion(s). The most frequent clade (transition at np 16320) testifies to a pronounced expansion event in the mid-Holocene and seems to be prominent in many Bantu groups from all of Africa. In contrast, one clade (transition at np 16264) is essentially restricted to Atlantic western Africa (including Cabo Verde). We propose a tentative L3e phylogeny that is based on 197 HVS-I sequences. We conclude that haplogroup L3e originated in central or eastern Africa about 46,000 ($\pm 14,000$) years ago, and was a hitchhiker of much later dispersal and local expansion events, with the rise of food production and iron smelting. Enforced migration of African slaves to the Americas translocated L3e mitochondria, the descendants of which in Brazil and the Caribbean still reflect their different regional African ancestries.

Correspondence: Prof. Dr. Hans-Jürgen Bandelt, Fachbereich Mathematik der Universität, Bundesstr.55, 20146 Hamburg, Germany. Fax: +49-40-42838-5190.

E-mail: bandelt@math.uni-hamburg.de, bandelt@yahoo.com

INTRODUCTION

African mtDNAs seem to fall into a restricted number of basal haplogroups having different ages and geographic specificity (Scozzari *et al.* 1988; Chen *et al.* 1995; Watson *et al.* 1997). In regard to the enormous complexity and time depth of the prehistory of the African continent (Phillipson, 1993; Shillington, 1995), the phylogeographic information (Avise, 2000) hitherto derived from mtDNA variation is rather unsatisfactory. The major cause of this dilemma may be seen in the apparent shift of scientific paradigms towards global population comparisons, mimicking studies of nuclear DNA polymorphisms (Simoni *et al.* 2000), that seek to establish a correlation between genetic variation and cultural markers, such as subsistence pattern and language (Watson *et al.* 1996; Barbujani, 1997; Poloni *et al.* 1997). Another reason is the unbalanced sampling of mtDNA across Africa, where Atlantic western Africa (Senegal in particular) is clearly over-represented. A third shortcoming is that different studies focussed on different parts of the molecule, with little or no attention paid to combining these data, despite the fact that the same population samples have been employed repeatedly (e.g. from the Biaka, Mbuti, and Mandenka): Scozzari *et al.* (1988) studied six enzyme RFLPs; and Cann *et al.* (1987) used 12-enzyme restriction mapping; Vigilant (1990) and Vigilant *et al.* (1991) sequenced the segments HVS-I & II of the control region (which subsequently became a popular enterprise, especially the compilation of HVS-I sequences); Nachman *et al.* (1996) provided ND3 information; and Watson *et al.* (1997) studied HVS-I sequences jointly with two key restriction sites which distinguish the mtDNA superhaplogroups L1, L2, and L3. Notably, Soodyall (1993) obtained both six enzyme RFLPs and HVS-I & II sequences for southern African mtDNAs, but without striving to link the systems phylogenetically. The work of Graven *et al.* (1995) established a connection between HVS-I & II and six enzyme RFLPs for Senegalese mtDNAs, but this information was

not linked with the parallel study of Chen *et al.* (1995) on 14-enzyme RFLPs. Later, Chen *et al.* (2000) reported 14-enzyme RFLPs combined with HVS-I & II from South Africa, which, however, are partly in conflict with each other as well as with earlier results. Finally, Ingman *et al.* (2000) presented a worldwide set of 53 complete mtDNA sequences but ignored the previously published information about the mtDNA phylogeny. The *a posteriori* linking of the information and the resolution of conflicts in the data is not yet completed, but some progress was made by Rando *et al.* (1998) and Quintana-Murci *et al.* (1999). However, one of the African-specific haplogroups, L3e, that emerged from these attempts is still poorly understood in its geographic distribution and HVS-I variation across Africa. As a clade within L3 it is characterised only by the presence of the *Mbo*I site at np 2349, which is not identified by the six enzyme system. A recent screening of 69 Brazilian mtDNAs of African ancestry for the combination of -3592 *Hpa*I and $+2349$ *Mbo*I indicated that L3e is the dominant African haplogroup (compared to L1a, L1b, L1c, L2, L3b, L3d, and U6) in the Brazilian sample (Alves-Silva *et al.* 2000). The advantage of using Brazilian data for phylogenetic mtDNA studies is the geographically diverse origin of the corresponding ancestral lineages, *viz.*, from western Africa (coast of Guinea), central Africa (Angola), and possibly even eastern Africa (Mozambique). In addition to 37 confirmed and 5 inferred L3e mtDNAs from the literature, a further 29 mtDNAs mainly from new samples of Santo Domingo (Dominican Republic) and Cabo Verde (Cape Verde) were found to harbour the characteristic RFLP markers for L3e. The corresponding HVS-I sequences then turned out to fall into four principal clades, each of which can be identified by one mutation in HVS-I. The thus discovered HVS-I motifs allowed the screening of the database for potential members of L3e: altogether 126 additional sequences can be allied with one of the four clades. The phylogenetic analysis of the total collection of 197 HVS-I sequences, leaving of course some ambiguities, is rather straightforward because the whole set

connects up quite well with 1- or 2-mutational steps. Available HVS-II information on 56 of these sequences and the screening of a few restriction sites assisted in establishing a nested hierarchy of haplotypes. Some of the more peripheral clades in L3e show a geographically localized provenance on the one hand and long-distance links on the other hand, which reflect the imprint of major dispersal or diffusion events that took place in the Holocene. L3e as a whole has a considerable age of 46,000 ($\pm 14,000$) years and, in regard to age and continental ubiquity, it could be seen as an African counterpart of the European haplogroup U5.

MATERIALS AND METHODS

1. mtDNA classification

We employ the classification of European and African mtDNA established in Rando *et al.* (1998), Richards *et al.* (1998), and Macaulay *et al.* (1999), which combine HVS-I and RFLP information. Following the nomenclature of those papers, human mtDNAs are divided into the (paraphyletic) cluster L1 and the clades L2 and L3. Whereas L1 and L2 are essentially African-specific, L3 (characterised by the loss of the *HpaI* site at np 3592) is ubiquitous. L3 can be further subdivided into basal clades: haplogroup M (with C at np 10400), haplogroup N (with T at np 10873), L3b (with gain of the 10084 *TaqI* site), L3d (with loss of the 8616 *MboI* site), L3e (with gain of the 2349 *MboI* site), and further central/eastern African clades that await characterisation. Haplogroups M and N together encompass virtually all Eurasian mtDNAs (Quintana-Murci *et al.* 1999), in particular the familiar European and Asian/American haplogroups, whereas haplogroups L3b, L3d, and L3e are African-specific (Chen *et al.* 1995, 2000; Watson *et al.* 1997; Rando *et al.* 1998, 1999).

2. Population samples

We employed a sample of 247 Brazilian mtDNAs analysed for HVS-I and in part for HVS-II (those with code BR-SE) and a number

of restriction sites, so that L3e membership could reliably be assessed (Alves-Silva *et al.* 2000). This provided 21 L3e mtDNAs, all of which have also been analysed for HVS-II. MtDNAs of 127 subjects from Santo Domingo (Dominican Republic) were analysed by high resolution RFLP analysis according to Torroni *et al.* (1999), and 17 were found to harbour the L3e RFLP motif. HVS-I sequence variation of these 17 mtDNAs was also determined. Two L3e HVS-I sequences were also found in a sample of 109 Palestinians, one in a sample of 71 French, and one in a sample of 86 Yemenites. These four mtDNAs as well as the L3e mtDNAs from the study of Rando *et al.* (1998) were further typed by partial RFLP analysis. A data set from Cabo Verde of size 292 yielded an additional 42 L3e mtDNAs, for which HVS-I and some diagnostic restriction sites were analysed. From a new Portuguese sample of 298 HVS-I sequences we further extracted five L3e sequences.

3. Control region sequencing and RFLP screening

HVS-I sequences were determined according to Torroni *et al.* (1999) and Alves-Silva *et al.* (2000). HVS-II sequence between positions 72 and 337 was determined for all Brazilian L3e samples (with codes BR-N and BR-NE) not previously analysed for HVS-II. The pair of primers L29 5'-GGTCTATCACCTATTAACCAAC-3' and H580 5'-TTGAGGAGGTAAGCTACATA-3' was used to amplify a 570 bp fragment in a 45 μ l volume PCR reaction. Each tube contained 0.8 μ M of each primer, 200 μ M dNTP and 0.5 U of *Taq* DNA polymerase (Promega Corporation, USA). Thirty cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 30 sec and extension at 72 °C for 1 min were carried out. PCR products were visualized in 1 % agarose gel electrophoresis with ethidium bromide. Amplified segments were purified using Magic[®] PCR Preps (Promega Corporation, USA) and dideoxy sequencing was carried out with Thermo Sequenase Sequencing Kit (Amersham LIFE SCIENCE, Inc., USA) using fluorescently

labelled primers L48 5'-CTCACGGGAGCTCTCC-ATGC-3' or H408 5'-CTGTTAAAAGTGCATAC-CGCCA-3'. Employing the information from Chen *et al.* (1995, 2000) and that obtained from the high resolution 14-enzyme RFLP analysis of the 17 L3e samples from Santo Domingo (A. Torroni, unpub. data), we identified a number of polymorphic restriction sites which potentially could be used to subdivide L3e into subclades. Screening of these sites (5260 *Ava*II, 5584 *Alu*I, 9253 *Hae*III, 9553 *Hae*III, 13100 *Msp*I, 13803 *Alu*I, 14869 *Mbo*I, and 15812 *Rsa*I) in selected samples was performed by PCR using pairs of primers and conditions described by Torroni *et al.* (1992), Chen *et al.* (1995), and Macaulay *et al.* (1999). Digestions were carried out following the conditions specified by the manufacturer (GibcoBRL Life Technologies, USA). Resulting fragments were resolved by electrophoresis in 8% acrylamide gels after silver staining. For the Northwest African and Senegalese samples shorter fragments were amplified.

4. Database search

We searched the mtDNA database (including all published HVS-I sequences, up to the year 2000) for additional sequences that match a sequence with confirmed (or inferred) L3e status or that bear a mutation (relative to CRS) at np 16223 and one of the four positions 16327, 16320, 16265 (transversion) and 16264. Only those sequences were retained as potential members of L3e which do not bear motifs of other known haplogroups or other clusters of yet unknown status. It was taken into consideration whether the competing haplogroup/cluster occurs in the area where the sequence in question was sampled. We then searched for 1-step neighbours of the candidate sequences that might have undergone a back mutation at the corresponding motif position but share other mutations with the potential members of L3e.

Some of the sequences from the database were adjusted as follows. We assumed that one HVS-I sequence from the Kikuyu (Watson *et al.* 1997)

should bear the transition at np 16185 rather than at np 16186 as published (base shift error). Note that Table A1 of that publication lists only mutations within the scoring frame (thus not displaying e.g. a transition at np 16067), and in one case (p. 700) erroneously recorded the A to T transversion at np 16265 as a transition. The position scored as 196 in HVS-II by Vigilant (1990) was reconstructed as 195. We took care of the multiple enumeration shifts (relative to CRS) in HVS-I & II as listed by Soodyall (1993), but otherwise retained the raw sequences (in which evidently some variant nucleotides were not recorded).

5. Phylogenetic analysis

We have chosen the MJ network method with parameter $\epsilon = 0$ (Bandelt *et al.* 1999) as the initial stage of the analysis since the collection of HVS-I sequences under study is almost 2-step connected (with large 1-step components), except for one potential outlier. In order to enhance the search for plausible trees within the MJ network we propose the following heuristic 'thinning' procedure (tailored to networks with large 1-step components and only a few cycles). To this end, we appreciate potential differences in the positional mutation rates, by adopting the mutational scores inferred by Hasegawa *et al.* (1993) as rough estimates for the relative positional rates. The mutational score $m(i)$, being a number between 0 and 15 in their test data, of each position i from the HVS-I reading frame 16042–16400 is translated into a weight w_i by the following (*ad hoc*) scaling:

$$w_i = 1/[3 + m(i)]$$

We then screen all 4-cycles with 1-step links in the MJ network. Assume that C is such a cycle, in which the two pairs of opposite links are labelled by positions i and j in HVS-I, respectively. For each node x of C we determine the total number $F(x)$ of individuals from the sample located at node x or at 1-step neighbours of x outside C . The sum of the F -values for the four

Table 1. *Population codes*

Code	Population	Reference
AFB	African Brazilian	Bortolini <i>et al.</i> (1997)
BAL	Berbers from Algeria	Côrte-Real <i>et al.</i> (1996)
BAM	Bambara	Rando <i>et al.</i> (1998)
BMO	Berbers from Morocco	Rando <i>et al.</i> (1998)
BRA	Brazilian ‘Amerindian’	Horai <i>et al.</i> (1993)
BR-	Brazilian from North (N), Northeast (NE), Southeast (SE), South (S)	Alves-Silva <i>et al.</i> (1999), Alves-Silva <i>et al.</i> (2000)
BUB	Bubi	Mateu <i>et al.</i> (1997)
CAV	Cabo Verde	This study
DAM	Dama	Soodyall (1993)
DIO	Diola	Rando <i>et al.</i> (1998)
EGY	Egyptian	Krings <i>et al.</i> (1999)
FRA	Northern French	Richards <i>et al.</i> (2000)
FUL	Fulbe	Watson <i>et al.</i> (1997)
HAU	Hausa	Watson <i>et al.</i> (1997)
HER	Herero	Vigilant (1990), Vigilant <i>et al.</i> (1991), Soodyall (1993)
KAN	Kanuri	Watson <i>et al.</i> (1997)
KIK	Kikuyu	Watson <i>et al.</i> (1997)
!KU	(Sekele, Vasikela)!Kung	Soodyall (1993), Chen <i>et al.</i> (2000)
KWE	Barakwena/Khwe	Soodyall (1993), Chen <i>et al.</i> (2000)
MAN	Mandenka	Graven <i>et al.</i> (1995)
MEX	Mexican	Green <i>et al.</i> (2000)
PAL	Palestinian	Richards <i>et al.</i> (2000)
POR	Portuguese	Pereira <i>et al.</i> (2000), this study
SÃO	São Toméan	Mateu <i>et al.</i> (1997)
SDO	Santo Domingan	This study
SER	Serer	Rando <i>et al.</i> (1998)
SOT	Sotho	Soodyall (1993)
SUD	Southern Sudanese	Krings <i>et al.</i> (1999)
SYR	Syrian	Richards <i>et al.</i> (2000)
TUA	Tuareg	Watson <i>et al.</i> (1997)
WOL	Wolof	Rando <i>et al.</i> (1998)
YEM	Yemenite	Richards <i>et al.</i> (2000)
YOR	Yoruba	Vigilant (1990), Vigilant <i>et al.</i> (1991), Watson <i>et al.</i> (1997)
ZUL	Zulu	Soodyall (1993)

nodes of C is denoted by $F(C)$. We then delete the link xy between two neighbouring nodes x and y of C , which is labelled by position j , provided that the following three requirements are met:

- (1) $w_j \geq 2 \cdot w_i$;
- (2) $F(x) + F(y) < F(C)/2$;
- (3) the link xy belongs to no other cycle than C .

Criterion (1) in combination with the proposed scaling requires that the mutational score of position j is considerably smaller than that of position i (thus allowing for some stochastic error and systematic bias in the scoring by Hasegawa *et al.* 1993): e.g. (1) is met when $m(i) \geq 9$ and $m(j) \leq 3$. Criterion (2) breaks ties by favouring

the more frequent side of the rectangle C and its 1-step neighbourhood for the inferred evolutionary pathway. Finally, criterion (3) blocks any potential decision to delete a link that would create a larger cycle in the network.

Most parsimonious reconstruction of HVS-II evolution on a tree inferred from HVS-I sequences (in conjunction with partial RFLP information) is hampered by the following circumstances: (i) HVS-II sequences are available only for a small subset of the data; (ii) several positions in HVS-II are subject to particularly high mutation rates (Aris-Brosou & Excoffier, 1996; Bandelt *et al.* 2000); (iii) the quality of the HVS-II sequences collected from the (early) literature is suboptimal, reflecting obvious errors

(such as notorious misscoring of np 263). To shield against highly unreliable reconstructions, we adopt a hierarchical consensus approach for finding HVS-II motifs, which operates on a collapsed tree focussing only on the deepest branches of the phylogeny. This tree, referred to as a skeleton of the phylogeny, defines the principal clades and pronounced subclades. Then our approach simply amounts to most parsimonious reconstruction (Maddison, 1989) along the highly polytomous tree in which the tips represent the different HVS-I & II haplotypes and the interior part is determined by the skeleton.

6. Age estimation

The coalescence time of a group of individual HVS-I sequences is estimated via the averaged mutational distance, ρ , to the reconstructed most recent common ancestor. The calibration for converting ρ into time scores only transitions in the segment np 16090–16365 such that $\rho = 1$ corresponds to 20,180 years (Forster *et al.* 1996). A lower bound for the standard deviation of ρ is obtained as $\sqrt{(\rho/n)}$ where n is the sample size (by assuming a perfect star genealogy). A direct estimation for the standard deviation σ uses a reconstructed phylogeny as a hypothetical genealogy (which is thus not fully resolved, being locally a star at each branching node). Then, if there are altogether m nested clades, each of which is defined by a branch with d_i scored mutations and carries n_i individuals ($i = 1, \dots, m$), we obtain the following formulae (assuming independent Poisson processes along the branches with parameters $d_1, \dots, d_m$):

$$\rho = (n_1 d_1 + n_2 d_2 + \dots + n_m d_m) / n,$$

$$\sigma^2 = (n_1^2 d_1 + n_2^2 d_2 + \dots + n_m^2 d_m) / n^2$$

(Saillard *et al.* 2000). Age estimates are presented in the form $\rho \pm \sigma$ converted to time.

7. Regional profiles

The African samples are allocated to five major regions (Senegambia, non-Atlantic western, eastern, central, and southern Africa) and

two islands (Cabo Verde and São Tomé). These are compared to the samples from Santo Domingo, Brazil, and Portugal. Thus only five (with population codes BAL, BMO, FRA, and MEX) out of 197 mtDNAs are left unassigned. Haplogroup L3e is subdivided into six clusters (five principal subclades and one paraphyletic cluster). For each region we record the absolute sample frequencies of these clusters; the resulting vectors (in 6-dimensional Euclidean space) are then referred to as the regional L3e profiles. Angles θ between profile vectors measure dissimilarity and could be transformed into the familiar chord length measure by taking the square root of $2(1 - \cos \theta)$.

RESULTS

With two exceptions, all of the 66 HVS-I sequences recognised as L3e members by the *MboI* site at position 2349 bear exactly one of the following four mutations: a transition at np 16327, or 16320, or 16264, or an A $\rightarrow$ T transversion at np 16265. The four principal clades within L3e defined by these mutations are named L3e1, L3e2, L3e4, and L3e3, respectively. Five additional sequences from the Mandenka belong to the 6-enzyme RFLP haplogroup 52-2 (Graven *et al.* 1995), which coincides with the superclade L3e3'4 (see below). The database search offers an additional 126 sequences as L3e candidates. The minimum spanning network (which is the superposition of all minimum spanning trees; see Excoffier & Smouse, 1994; Bandelt *et al.* 1999) for this collection of 197 sequences is nearly 2-step connected, with only one outlier requiring 3 steps to link up. Large parts of the principal clades are even 1-step connected. The MJ network (not shown) has 10 cycles, six of which are within L3e1, one in L3e2, one in L3e3, and two in L3e4. Position 16311 is involved in six of these cycles, which testifies to its extreme variability in this haplogroup. The thinning procedure (see Materials and Methods) resolves five cycles and thus yields the network shown in Fig. 1, where a plausible phylogeny (employed for time estimates) is highlighted. In this phy-

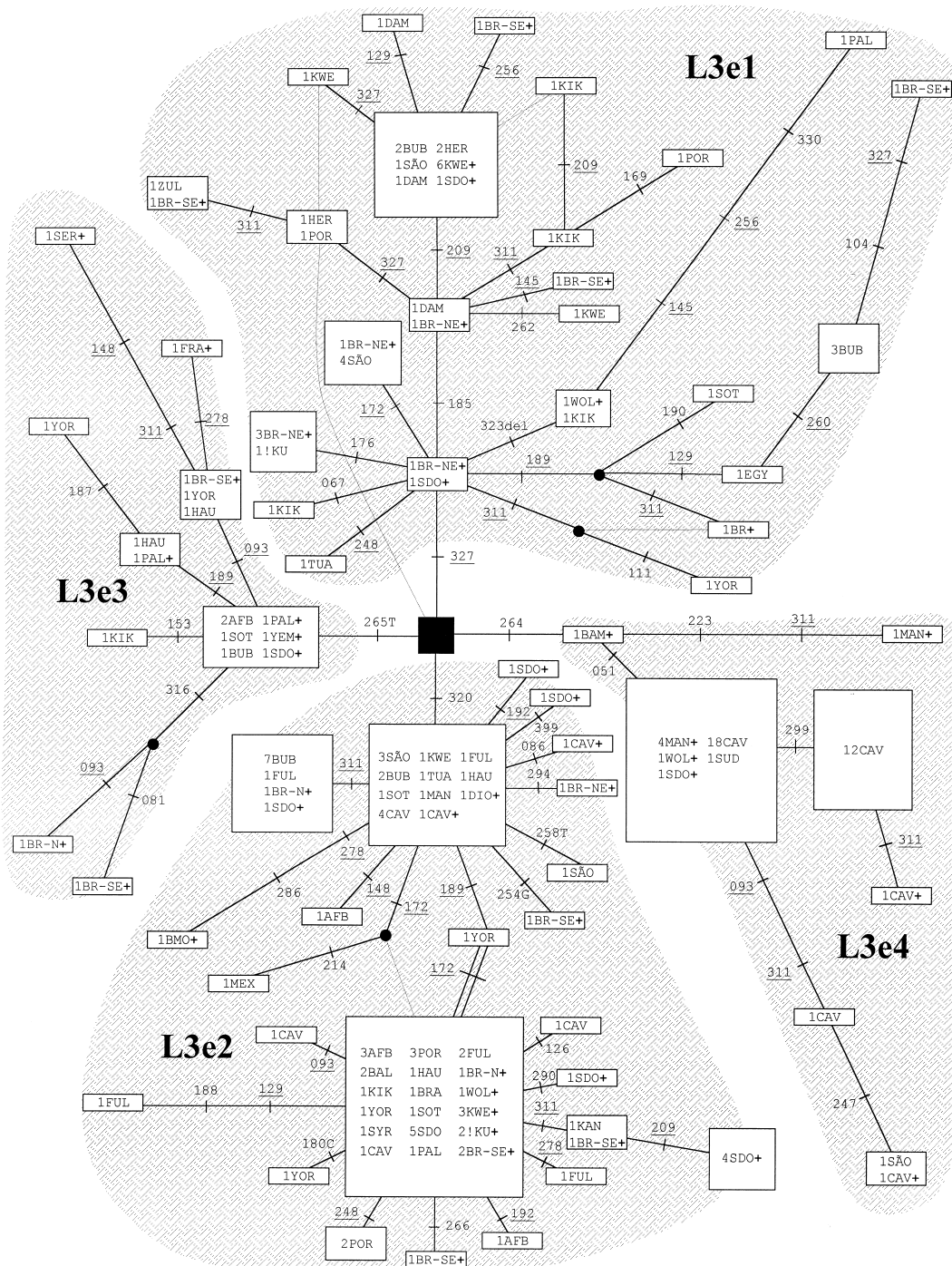


Fig. 1. Inferred HVS-I network of 197 potential L3e lineages, which is derived from the MJ network by applying the thinning procedure. A plausible phylogeny is indicated by solid lines (assuming one additional back mutation indicated by a double-line). The central node (black square) constitutes the root of L3e, distinguished from CRS (Andrews *et al.* 1999) by a transition at np 16223. Numbers preceding the population codes (see Table 1) are numbers of sequences of the same type sampled from the population in question; suffix + indicates whether the RFLP status at 2349 *Mbo*I has been confirmed (or inferred in the case of the Mandenka lineages). Numbers along links refer to nucleotide positions in HVS-I minus 16000, suffixes indicate a transversion or a deletion; underlining highlights recurrent mutations. Length polymorphisms in the A-C run are ignored.

Table 2. *Restriction sites tested for subclassification of confirmed L3e sequences (including published !Kung/Khwe data)*

Clade/Population code	HVS-I	5260 <i>Ava</i> II	5584 <i>Alu</i> I	9253 <i>Hae</i> III	9553 <i>Hae</i> III	13100 <i>Msp</i> I	13803 <i>Alu</i> I	14869 <i>Mbo</i> I	15812 <i>Rsa</i> I
L3e1:									
SDO	16223 16327	—	+	—	+	—	—	+	+
BR-NE	16223 16327			—	+				
WOL	16223 16323del 16327	—	+						
BR-NE	16176 16223 16327			—	+	—		+	+
L3e1a:									
BR-NE	16185 16223 16327			—	+				
BR-SE	16185 16223 16311			—	+				
BR-SE	16145 16185 16223 16327			+	+				
6 KWE, SDO	16185 16209 16223 16327	—	+	+	+	—	—	+	+
BR-SE	16185 16209 16223 16256 16327			+	+			+	
L3e2a:									
DIO, CAV	16223 16320							—	
CAV	16086 16223 16320							—	
SDO	16192 16223 16320	—	+	—	+	—	—	—	+
BR-SE	16223 16254G 16320							—	
BR-NE	16223 16294 16320			—	+	—		—	+
SDO	16223 16311 16320	—	+	—	+	—	—	—	+
BR-N	16223 16311 16320							—	
SDO	16223 16320 16399	—	+	—	+	—	—	—	+
BMO	16223 16278 16286 16320	—	+					—	
L3e2b:									
WOL	16172 16189 16223 16320	—	+					+	
2!KU, 3 KWE, 5 SDO	16172 16189 16223 16320	—	+	—	+	—	—	+	+
BR-SE	16172 16189 16223 16320			—	+	—		+	
BR-SE	16172 16189 16223 16320			—	+	—		+	+
SDO	16172 16189 16223 16290 16320	—	+	—	+	—	—	+	+
BR-SE	16172 16189 16223 16311 16320							+	
4 SDO	16172 16189 16209 16223 16311 16320	—	+	—	+	—	—	+	+
L3e3:									
YEM	16223 16265T	+	+		—	+	—		+
PAL	16223 16265T	+	+		—	+	—		—
SDO	16223 16265T	+	+	—	—	+	+	+	—
BR-SE	16093 16223 16265T			—	—	+			—
FRA	16093 16223 16265T 16278	+	+		—	+	—		—
BR-N	16093 16223 16265T 16316	+		—	—	+		+	—
SER	16093 16148 16223 16265T 16311							+	
PAL	16189 16223 16265T	+	+		—	+	—		—
L3e4:									
BAM	16223 16264	+	—					+	
SDO	16051 16223 16264	+	—	—	+	—	—	+	+
WOL	16051 16223 16264	+	—						

Table 3. *HVS-I & II sequence types from L3e, listed by motifs relative to CRS (confirmed L3e status indicated by +)*

Clade/Population code	HVS-I	HVS-II
L3e1:		
BR-NE+	16223 16327	73 150 189 200 263
BR-NE+	16172 16223 16327	73 150 189 200 263
3 BR-NE+	16176 16223 16327	73 150 152 189 200 263
!KU	16176 16223 16327	73 150 152 189 200
SOT	16189 16190 16223 16327	73 150 185 189
BR-SE+	16104 16129 16189 16223 16260	73 150 200 263
L3e1a		
BR-NE+	16185 16223 16327	73 150 189 200 263
DAM	16185 16223 16327	73 189 195 200 207 263
HER	16185 16223	73? 150 152 189 195 200 263
KWE	16185 16209 16223	73 150 152 189 195 200 207 263
KWE	16185 16223 16262 16327	73 150 152 195
BR-SE+	16185 16223 16311	73 150 185 189 263
POR	16169 16185 16223 16311 16327	73 150 185 189 200 263
BR-SE+	16145 16185 16223 16327	73 150 152 189 195 200 207 263
ZUL	16185 16223 16311	73 150 189 200 263
2 HER, DAM, 6 KWE+	16185 16209 16223 16327	73 150 152 189 195 200 207 263
BR-SE+	16185 16209 16223 16256 16327	73 150 152 189 195 263
DAM	16129 16185 16209 16223 16327	73? 189 195 200 263
L3e2a		
KWE	16223 16320	73 150 195 198
SOT	16223 16320	150 195 263
BR-SE+	16223 16254G 16320	73 150 195 198 263
BR-NE+	16223 16294 16320	73 150 195 198 263
BR-N+	16223 16311 16320	73 150 195 198 263
L3e2b		
SDO+	16172 16189 16223 16320	73 150 195 263
BR-N+, BR-SE+, 2!KU+, 3 KWE+	16172 16189 16223 16320	73 150 152 195 263
BR-SE+	16172 16189 16223 16320	73 150 152 263
SOT	16172 16189? 16223 16320	73? 195
POR	16172 16189 16223 16248 16320	73 150 195 263
BR-SE+	16172 16189 16223 16266 16320	73 150 195 263
BR-SE+	16172 16189 16223 16311 16320	73 150 195 263
L3e3		
SOT	16223 16265T	73 150 195 263
YOR	16093 16223 16265T	73 150 195 263
BR-SE+	16093 16223 16265T	73 150 189 195 263
BR-N+	16093 16223 16265T 16316	73 150 195 263
BR-SE+	16081 16223 16265T 16316	73 150 195 263
L3e4		
MAN+	16264 16311	73 150 263
2 MAN+	16051 16223 16264	73 150 263
2 MAN+	16051 16223 16264	73 150 263 316

logeny as many as 10 parallel events at np 16311 are postulated, followed by positions 16093, 16172, and 16327 with 4 recurrent mutations each.

Table 2 presents the results of our screening of RFLP sites that are polymorphic in L3e and potentially basal in the L3e phylogeny. The gain of the *Ava*II site at np 5260 (corresponding to a transition at np 5262) defines L3e3'4 (the

smallest clade embracing L3e3 and L3e4). Note that Graven *et al.* (1995) chose to score this *Ava*II site at np 5164 (which would require a C to G transversion at np 5165) in the low resolution six-enzyme RFLP analysis. L3e3 is defined by the loss of the *Hae*III site at np 9553 and the gain of the *Msp*I site at np 13100. L3e4 appears to be characterised by the loss of the *Alu*I site at np 5584. It is quite conceivable that L3e4 is also

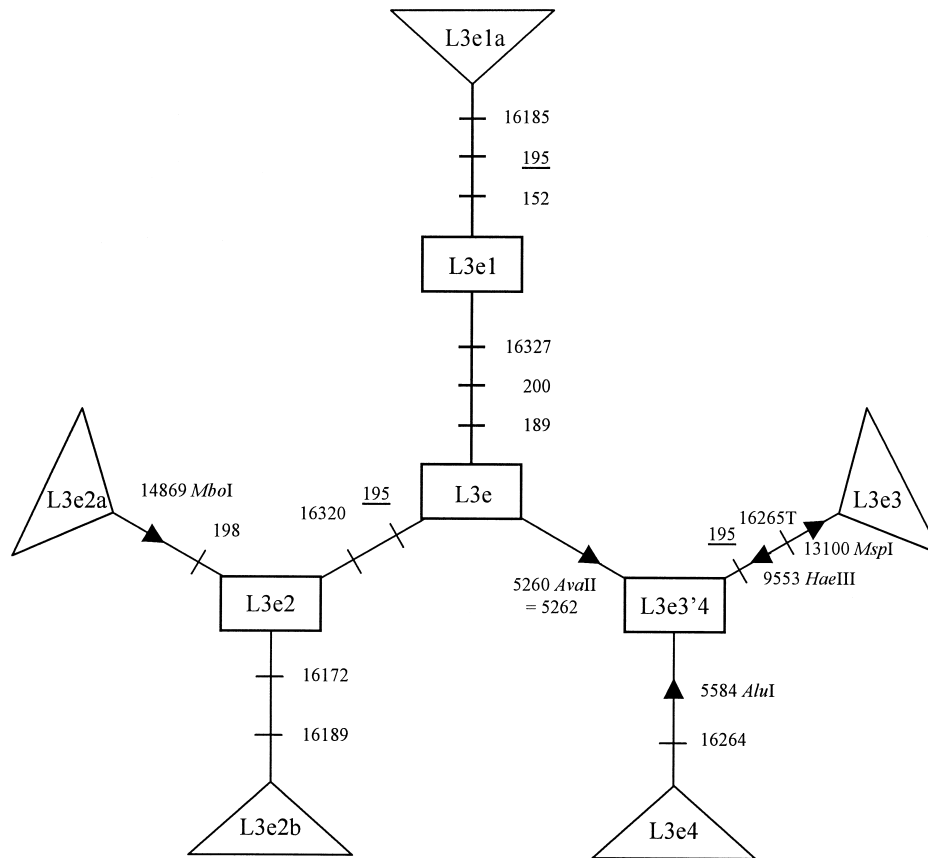


Fig. 2. The skeleton of major branching nodes in the L3e phylogeny, based on HVS-I & II sequences and restriction sites in the coding region (Tables 2 and 3). The potential root type of L3e (with motif 16223, 73, 150, 263 relative to CRS) and the inferred ancestral types of L3e1, L3e2, and L3e3'4 are indicated by rectangles, whereas the ancestral types of the peripheral clades are indicated by triangles. The order of mutations along single links is undetermined.

characterised by the loss of the *Rsa*I site at np 16049 (caused by a transition at np 16051) since the two exceptional L3e4 sequences (Fig. 1) may well have back-mutated at this site. The *Mbo*I site at np 14869 seems to partition L3e2 into two subclades: the loss characterises L3e2a, while the complementary clade, L3e2b, appears to be defined by transitions at nps 16172 and 16189. The site 9253 *Hae*III is polymorphic within L3e1 but the site may be present in all those sequences which bear the transition at np 16209. Within L3e3 the sites 13803 *Alu*I and 15812 *Rsa*I are polymorphic but may constitute private mutations in single individuals. Since the data set from Santo Domingo has been checked for 14-enzyme RFLPs, we can conclude that no further RFLP sites exist that would characterise the distinguished subclades of L3e.

HVS-II information is available for 56 sampled mtDNAs (Table 3). The skeleton of the L3e phylogeny, for which we reconstruct ancestral HVS-II types (see Material and Methods), comprises the ancestral nodes of L3e, L3e1, L3e1a, L3e2, L3e2a, L3e2b, L3e3'4, L3e3, and L3e4. Figure 2 displays the reconstructed HVS-II evolution along the skeleton, where the ancestral HVS-II type of L3e is distinguished from CRS by transitions at nps 73, 150, and 263, but the ancestral state at the highly variable position 195 cannot be determined unambiguously. HVS-II information may help to decide whether a mtDNA with an ambiguous HVS-I motif likely belongs to L3e or not. For instance, a Yoruban mtDNA with transitions at nps 16124, 16223, and 16327 in HVS-I bears transitions only at nps 73, 146, 152, and 263 in HVS-II (Vigilant, 1990)

Table 4. *Estimated coalescence times of clades within L3e*

Clade	Characteristic sites	Sample size	ρ	$\sqrt{(\rho/n)}$	σ	Age (ky)
L3e	+ 2349 <i>Mbo</i> I	197	2.28	0.11	0.68	46 ± 14
L3e1	16327	51	1.84	0.19	0.67	37 ± 13
L3e1a	16185	27	1.19	0.21	0.63	24 ± 13
L3e2	16320	85	1.60	0.14	0.83	32 ± 17
L3e2a	− 14869 <i>Mbo</i> I	36	0.47	0.11	0.29	10 ± 6
L3e2b	16172, 16189	49	0.43	0.09	0.16	9 ± 3
L3e3'4	+ 5260 <i>Ava</i> II	61	1.36	0.15	0.75	27 ± 15
L3e3	16265T, − 9553 <i>Hae</i> III, + 13100 <i>Msp</i> I	18	0.89	0.22	0.37	18 ± 7
L3e4	16264, − 5584 <i>Alu</i> I	43	0.56	0.11	0.32	11 ± 7

Table 5. *L3e profiles in different regions of Africa, America, and Europe*

Region	Number of HVS-I sequences per cluster					
	L3e1*	L3e1a	L3e2a	L3e2b	L3e3	L3e4
Africa (except BAL, BMO)						
Senegambia:						
BAM, DIO, MAN, SER, WOL	1	—	2	1	1	7
Cabo Verde: CAV	—	—	6	3	—	33
Non-Atlantic western Africa:						
FUL, HAU, KAN, TUA, YOR	2	—	4	9	4	—
(North)eastern Africa and Near East:						
EGY, KIK, PAL, SUD, SYR, YEM	4	2	—	3	4	1
Central Africa: BUB	3	2	9	—	1	—
São Tomé: SÃO	4	1	4	—	—	1
Southern Africa:						
DAM, HER, !KU, KWE, SOT, ZUL	2	15	2	6	1	—
America (except MEX)						
Caribbean: SDO						
	1	1	3	10	1	1
Brazil: AFB, BRA, BR	7	4	4	10	5	—
Europe (except FRA)						
Portugal: POR	—	2	—	5	—	—

and therefore does not qualify as an L3e1 member but rather belongs to L3d (see Fig. 4 of Rando *et al.* 1998), where similar HVS-II motifs have been observed (Graven *et al.* 1995).

We can now compare our L3e analysis with the information offered by the complete mtDNAs of Ingman *et al.* (2000). Although their aphylogenetic sampling strategy missed some African haplogroups (even a major one, L3b) their data set happens to contain four L3e sequences, *viz.*, one L3e1, one L3e3, and two L3e2b mtDNAs. No new HVS-I & II mutations are seen in these sequences. But we can infer that L3e as a whole has two further characteristic sites (besides np 2352) in the coding region: transitions at nps 10819 and 14212. The basal trichotomy at the root of L3e between L3e1, L3e2, and L3e3'4, however, cannot be resolved further. This con-

firms that the view on the L3e phylogeny obtained via RFLPs and HVS-I & II is quite satisfactory.

We estimate the coalescence time of haplogroup L3e via ρ (see Material and Methods) as 46,000 years (Table 4). This estimate, however, has a very large variance due to the structure of the inferred phylogeny (Fig. 1), which is not starlike near the root. The three deepest clades (L3e1, L3e2, and L3e3'4) of L3e are not really starlike either, and their coalescence times should not be regarded as very different (in view of the high variances): they all could actually be of the order of 30,000 years, the oldest clade possibly being L3e1.

There is a sharp contrast in the geographic distribution of the clades L3e1a and L3e4 (Table 5). The former is restricted to southern and

Table 6. *Affinities between regional L3e profiles expressed by the cosine of the angle between profile vectors*

BAM, DIO, MAN, SER, WOL									
.98	CAV								
.28	.14	FUL, HAU, KAN, TUA, YOR							
.35	.18	.70	EGY, KIK, PAL, SUD, SYR, YEM						
.30	.16	.44	.30	BUB					
.44	.29	.38	.48	.88	SÃO				
.11	.05	.39	.54	.34	.32	DAM, HER, IKU, KWE, SOT, ZUL			
.31	.23	.94	.57	.32	.29	.48	SDO		
.28	.11	.90	.88	.50	.57	.62	.84	AFB, BRA, BR	
.12	.08	.77	.52	.08	.06	.68	.91	.75	POR

eastern Africa whereas the latter seems to be confined to Atlantic western Africa (with only sporadic incursions in Sudan). The clades L3e2a, L3e2b, and the cluster L3e1* (that is, L3e1 without L3e1a) in contrast are almost omnipresent but apparently at relatively different frequencies per region.

Comparing the L3e profiles of the four regions (Cabo Verde, Santo Domingo, São Tomé, and Brazil), which were settled by slave trade, with the potential source areas in Africa, it is apparent that Cabo Verde represents a faithful image of Senegambia, Santo Domingo is very similar to non-Atlantic western Africa, São Tomé is remarkably akin to Bioko of Central Africa (although one should bear in mind that no haplotypes are shared between these two populations other than two L3e types), whereas Brazil resembles all areas to some extent except Senegambia. Southern Africa comes closest to Portugal and Brazil. These affinities are numerically expressed by the angles θ between the corresponding profile vectors (see Table 6 for the values $\cos \theta$, where the highest values corresponding to angles $< 30^\circ$ are highlighted). The sharpest contrast is between Cabo Verde and southern Africa, for which the profile vectors are almost perpendicular. This nicely parallels the contrasts found in the β^S haplotype distribution for the beta globin gene (Bortolini & Salzano, 1999).

DISCUSSION

The present analysis of L3e assists in directing the search for correct haplogroup assignment by

HVS-I motifs. It demonstrates that a first sorting of HVS-I sequences from L3 as performed in Watson *et al.* (1997) led to an unsatisfactory coarse grouping that failed to identify L3e (and other clades). Standard phylogenetic analyses of HVS-I sequences alone, which do not anticipate the correlation with putative RFLP status, have no chance to identify a set of diverse L3e sequences as a clade in an MP or NJ or any other tree. The two NJ trees (Bioko vs São Tomé) displayed in Fig. 3 of Mateu *et al.* (1997) constitute an illustrative case in question; also the NJ tree for the entire data set does not support L3e as a clade since sequences from other haplogroups are interspersed among the L3e sequences. Even with the information provided in the present study, there is an obvious risk in inferring L3e status just from HVS-I motif comparison. The positions 16264, 16320, and 16327 have evidently undergone multiple hits in Africa. The transition at np 16264 in conjunction with +2349 *MboI* is also part of the recognition HVS-I & RFLP motif for haplogroup L1b, but inasmuch as L1b can be distinguished from L3 by several mutations in HVS-I, HVS-II, as well as 14-enzyme RFLPs, there is no danger of confusion. The presence of the *MboI* site at np 2349 has also been observed in a sequence type from U6 (Torroni *et al.* 1999). To date we have no positive evidence for any further independent mutational events at np 2352 that create the *MboI* site at np 2349 (except possibly another event in haplogroup M scored by Ballinger *et al.* 1992).

The precise location and time of origin of

haplogroup L3e cannot yet reliably be reconstructed. A central African (or southern Sudanic) origin is certainly plausible considering the geographic distribution, especially of its presumably oldest and most diverse clade L3e1. Even the seemingly southern African subclade L3e1a harbours sequences from the Bantu-speaking Kikuyu of Kenya (which otherwise possess a typical east African mtDNA profile, akin to the Turkana mtDNA pool; Watson *et al.* 1997). If this subclade had really come from the south, one would expect to find sequences in central/eastern Africa from the ancient 'Khoisan' haplogroups (Chen *et al.* 2000). It rather seems likely that the southern African mtDNA pool received a package of L2/L3 mtDNAs (of limited diversity) through Bantu migrations. In particular, L3e1 must have been prominent in this 'southern Bantu package', although also L3e2 and L3e3 participated to some extent. The relatively high frequency of L3e in the Brazilian mtDNA pool may then be explained by the fact that the majority of the slaves that arrived in Brazil came from Bantu groups, mainly from Angola. The sporadic occurrences of L3e (and other sub-Saharan haplogroups; Pereira *et al.* 2000) in Portugal is then not surprising in view of early slave trade and back migration from the colonies.

The most frequent and widespread type of L3e is the ancestral type of the subclade L3e2b. Interestingly, this type (also found in our Brazilian sample) was previously reported as a Brazilian Amerindian sequence by Horai *et al.* (1993), but Forster (1997) emphasised that this sequence should rather be regarded as an African Brazilian sequence. The two dominant L3e2 types could have been successful hitchhikers of population movements in the Sahara during the Great Wet Phase (of the early Holocene) and the subsequent Neolithic Wet Phase (Muzzolini, 1993). Members of L3e2 are not only found in northern Africa but even in the Near East sporadically. In Egypt and Sudan there are also incursions of western African L1b and north-west African U6a sequences, which seem to be out of place. This may testify to long-distance

contacts in the Sahel zone and beyond, especially with pastoralists movements (such as the recent 'Fulbe diaspora'; Newman, 1995).

It is evident that our investigation of a single, though widespread, haplogroup can offer no more than a snapshot of African prehistory from the genetic perspective. In the future, detailed phylogeographic analyses of the other African mtDNA haplogroups as well as the African-specific Y-chromosome haplotypes and types of autosomal genes will provide a rich mosaic, which will eventually allow a more faithful picture of the genetic landscape to be sketched than was hitherto possible with classical genetic markers (Cavalli-Sforza *et al.* 1994).

We thank Martin Richards for providing some samples for RFLP testing. This research was supported by grants from CNPq, FAPEMIG, PRPq-UFMG (Brazil), a travel grant from DAAD and CAPES to H.-J. B. and S. D. J. P., and grants from Fondazione Telethon (E.0890) to A. T., Fondo d'Ateneo per la Ricerca dell'Università di Pavia to A. T., Istituto Pasteur Fondazione Cenci Bolognietti, Università di Roma "La Sapienza" to R. S., Progetti Ricerca Interesse Nazionale 1999 and 2001 to R. S., A. T., A. C., and Facoltà 60% to A. C.; J. M. L. was supported by grant PB96-1034 from DGICT (Spain); A. B. was supported by a grant from ICCTI (Portugal) and acknowledges the support of the Cape Verde Army Chief of Staff; L. P. was supported by a Ph.D. grant from Fundação para a Ciência e a Tecnologia (PRAXIS XXI/BD/13632/97).

REFERENCES

- Alves-Silva, J., Santos, M. S., Guimaraes, P. E. M., Ferreira, A. C. S., Bandelt, H.-J., Pena, S. D. J., *et al.* (2000). The ancestry of Brazilian mitochondrial DNA lineages. *Am. J. Hum. Genet.* **67**, 444–461.
- Alves-Silva, J., Santos, M. S., Pena, S. D. & Prado, V. F. (1999). Multiple geographic sources of region V 9-bp deletion haplotypes in Brazilians. *Hum. Biol.* **71**, 245–259.
- Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147.
- Aris-Brosou, S. & Excoffier, L. (1996). The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol. Biol. Evol.* **13**, 494–504.
- Avice, J. C. (2000). *Phylogeography: the history and formation of species*. Cambridge, MA: Harvard University Press.
- Ballinger, S. W., Schurr, T. G., Torroni, A., Gan, Y. Y., Hodge, J. A., Hassan, K., *et al.* (1992). Southeast Asian mitochondrial DNA analysis reveals genetic

- continuity of ancient mongoloid migrations. *Genetics* **130**, 139–152.
- Bandelt, H.-J., Forster, P. & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48.
- Bandelt, H.-J., Macaulay, V. & Richards, M. (2000). Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol. Phylogenet. Evol.* **16**, 8–28.
- Barbujani, G. (1997). DNA variation and language affinities. *Am. J. Hum. Genet.* **61**, 1011–1014.
- Bortolini, M. C., Zago, M. A., Salzano, F. M., Silva-Junior, W. A., Bonatto, S. L., Da Silva, M. C., *et al.* (1997). Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. *Hum. Biol.* **69**, 141–159.
- Bortolini, M. C. & Salzano, F. M. (1999). β^s haplotype diversity in Afro-Americans, Africans, and Euro-Asiatics – an attempt at a synthesis. *Ciência e Cultura* **51**, 175–180.
- Cann, R. L., Stoneking, M. & Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature* **325**, 31–36.
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994). *History and geography of human genes*. Princeton, NY: Princeton University Press.
- Chen, Y. S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. S. & Wallace, D. C. (1995). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* **57**, 133–149.
- Chen, Y. S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K. & Wallace, D. C. (2000). mtDNA variation in the South African Kung and Khwe – and their genetic relationships to other African populations. *Am. J. Hum. Genet.* **66**, 1362–1383.
- Côrte-Real, H. B., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A., *et al.* (1996). Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann. Hum. Genet.* **60**, 331–350.
- Excoffier, L. & Smouse, P. E. (1994). Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics* **136**, 43–59.
- Forster, P., Harding, R., Torroni, A. & Bandelt, H.-J. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* **59**, 935–945.
- Forster, P. (1997). Wandertrieb im Blut. Interview in 'Der Spiegel' (3/1997), pp. 152–153.
- Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A., *et al.* (1995). Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol. Biol. Evol.* **12**, 334–345.
- Green, L. D., Derr, J. N. & Knight, A. (2000). mtDNA affinities of the peoples of north-central Mexico. *Am. J. Hum. Genet.* **66**, 989–998.
- Hasegawa, M., Di Rienzo, A., Koehler, T. D. & Wilson, A. C. (1993). Toward a more accurate time scale for the human mitochondrial DNA tree. *J. Mol. Evol.* **37**, 347–354.
- Horai, S., Kondo, R., Nakagawa-Hattori, Y., Hayashi, S., Sonoda, S. & Tajima, K. (1993). Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol. Biol. Evol.* **10**, 23–47.
- Ingman, M., Kaessmann, H., Pääbo, S. & Gyllenstein, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713.
- Krings, M., Salem, A. E., Bauer, K., Geisert, H., Malek, A. K., Chaix, L. *et al.* (1999). mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am. J. Hum. Genet.* **64**, 1166–1176.
- Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., *et al.* (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**, 232–249.
- Maddison, W. P. (1989). Reconstructing character evolution on polytomous cladograms. *Cladistics* **5**, 365–377.
- Mateu, E., Comas, D., Calafell, F., Perez-Lezaun, A., Abade, A. & Bertranpetit, J. (1997). A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann. Hum. Genet.* **61**, 507–518.
- Muzzolini, A. (1993). The emergence of a food-producing economy in the Sahara. In *The archaeology of Africa: food, metals and towns* (eds. T. Shaw *et al.*), pp. 227–239. London: Routledge.
- Nachman, M. W., Brown, W. M., Stoneking, M. & Aquadro, C. F. (1996). Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* **142**, 953–963.
- Newman, J. L. (1995). *The peopling of Africa*. New Haven: Yale University Press.
- Pereira, L., Prata, M. J. & Amorim, A. (2000). Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann. Hum. Genet.* **64**, 491–506.
- Phillipson, D. W. (1993). *African archaeology* (2nd ed.). Cambridge: Cambridge University Press.
- Poloni, E. S., Semino, O., Passarino, G., Santachiara-Benerecetti, A. S., Dupanloup, I., Langaney, A., *et al.* (1997). Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am. J. Hum. Genet.* **61**, 1015–1035.
- Quintana-Murci, L., Semino, O., Bandelt, H.-J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A. S. (1999). Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat. Genet.* **23**, 437–441.
- Rando, J. C., Cabrera, V. M., Larruga, J. M., Hernández, M., González, A. M., Pinto, F., *et al.* (1999). Phylogeographic patterns of mtDNA reflecting the colonisation of the Canary Islands. *Ann. Hum. Genet.* **63**, 413–428.
- Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M., *et al.* (1998). Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann. Hum. Genet.* **62**, 531–550.
- Richards, M. B., Macaulay, V. A., Bandelt, H.-J. & Sykes, B. C. (1998). Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.* **62**, 241–260.
- Richards, M., Macaulay, V. A., Hickey, E., Vega, E., Sykes, B., Guida, V., *et al.* (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* **67**, 1251–1276.

- Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.-J. & Nørby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am. J. Hum. Genet.* **67**, 718–726.
- Scozzari, R., Torroni, A., Semino, O., Sirugo, G., Brega, A. & Santachiara-Benerecetti, A. S. (1988). Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. *Am. J. Hum. Genet.* **43**, 534–544.
- Shillington, K. (1995). *History of Africa* (2nd ed.). London: Macmillan.
- Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J. & Barbujani, G. (2000). Reconstruction of prehistory on the basis of genetic data. *Am. J. Hum. Genet.* **66**, 1177–1179.
- Soodyall, H. (1993). Mitochondrial DNA polymorphisms in southern African populations. PhD thesis, University of Witwatersrand, Johannesburg.
- Torroni, A., Cruciani, F., Rengo, C., Sellitto, D., Lopez-Bigas, N., Rabionet, R., *et al.* (1999). The A1555G mutation in the 12S rRNA gene of human mtDNA: recurrent origins and founder events in families affected by sensorineural deafness. *Am. J. Hum. Genet.* **65**, 1349–1358.
- Torroni, A., Schurr, T. G., Yang, C.-C., Szathmary, E. J. E., Williams, R. C., Schanfield, M. S., *et al.* (1992). Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* **130**, 153–162.
- Vigilant, L. (1990). Control region sequences from African population and the evolution of human mitochondrial DNA. PhD thesis, University of California, Berkeley.
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. (1991). African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503–1507.
- Watson, E., Bauer, K., Aman, R., Weiss, G., von Haeseler, A. & Pääbo, S. (1996). mtDNA sequence diversity in Africa. *Am. J. Hum. Genet.* **59**, 437–444.
- Watson, E., Forster, P., Richards, M. & Bandelt, H.-J. (1997). Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.* **61**, 691–704.