História Genética dos Gaúchos: Dinâmica

Populacional do Sul do Brasil

Andrea Rita Marrero

UFRGS

2006

Universidade Federal do Rio Grande do Sul

História Genética dos Gaúchos - Dinâmica

Populacional do Sul do Brasil

Andrea Rita Marrero

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

Orientadora: Maria Cátira Bortolini

Porto Alegre

Dezembro de 2006

Na capa, o painel intitulado "A Formação Histórica-Etnográfica do Povo Rio-Grandense" do pintor italiano Aldo Locatelli (1915 – 1962). O painel de 25 metros quadrados encontra-se no Salão Alberto Pasqualini do Palácio Piratini, em Porto Alegre. Estão representados os índios, as Entradas e Bandeiras, as Missões, a criação das fazendas (agricultura e pecuária), o gaúcho (onde o artista se retrata em primeiro plano) e o progresso (simbolizado pela energia elétrica e uma represa). Olhando com atenção, observa-se o contorno do mapa do Rio Grande do Sul no centro do painel.

Gaudério

João da Cunha Vargas

Poncho e laço na garupa
Do pingo quebrei o cacho
Dum zaino negro gordacho
Assim me soltei no pampa
Recém apontando a guampa
Pelito grosso de guacho

Fui pelechando na estrada
Do velho torrão pampeano
Já serrava sobreano
Cruzava de um pago a outro
Quebrando queixo de potro
Sem nunca ter desengano

Fui conhecendo as estâncias
O dono, a marca, o sinal
Churrasco que já tem sal
Guaiaca que tem dinheiro
Cavalo que é caborteiro
E o jujo que me faz mal

Conheço todo o Rio Grande
Qualquer estrada ou atalho
Quando me seco trabalho
Na velha lida campeira
Corro bem uma carreira
Manejo bem o baralho

Na tava sempre fui taura
Nunca achei parada feia
Quando o parceiro cambeia
Distância de nove passo
Quando espicho bem o braço
Num tiro de volta e meia

Num bolicho de campanha
De volta de uma tropeada
Botei ali uma olada
A maior da minha vida:
Dezoito sorte corrida
Quarenta e cinco clavada

E quanto baile acabei
Solito, sem companheiro
Dava um tapa no candeeiro
Um talho no mais afoito
Calçado no trinta e oito
Botava pra fora o gaiteiro

Trancava o pé no portal
Abria a porta da sala
Entre bufido de bala
E a providência divina
Só manotaços de china
Rasgando a franja do pala

Ninguém me toca por diante
Nem tampouco cabresteio
Eu me empaco e me boleio
Não paro nem com sinuelo
E tourito de outro pelo
Não berra no meu rodeio

Não quero morrer de doença
Nem com a vela na mão
Eu quero guasquear no chão
Com um balaço bem na testa
E que seja em dia de festa
De carreira ou marcação

E peço, quando eu morrer
Não me por em cemitério
Existe muito mistério
Prefiro um lugar deserto
E que o zaino paste perto
Cuidando os restos gaudério

E vou levar quando eu for
No caixão algum troféu:
Chilena, adaga, chapéu
Meu tirador e o laço
O pala eu quero no braço
Pra gauderiar lá no céu!

AGRADECIMENTOS

Bueno... e sem muitas milongas um baita agradecimento...

- O clássico e mais que merecido agradecimento ao Elmo e Ellen!!!

- Aos meus amigos do Japonês: どうも　ありがとう　ございます.

- Aos meus amigos do Origami, nada melhor que

- É muito difícil listar todos os meus amigos queridos que me acompanharam esse tempo... Wanderlan, Lessandra, Vander, Luciana, Sandson, Rafael, Tábita, Vanessa, Martin, Taiana, Jaque, Valéria, Laci, Veronica, Fabi, Shaiane, Nico, e todos aqueles que não estão nomeados, mas certamente foram lembrados;

- Ao Wilson Araújo da Silva Júnior, de quem ganhei conselhos incríveis e uma amizade muito valiosa;

- Ao Horacio Schneider e Maria Iracilda Sampaio (Irá), que me deram a primeira oportunidade na área e renovaram o incentivo em cada congresso que nos encontramos;

- À familia que agora é minha também, Maria Emília, Alfredo e Roze, Renato, Karina, Raquel, Natacha e Diana;

- Principalmente, à minha família que soube entender minhas ausências e me incentivou e sempre me deu certeza de que qualquer sacrifício valeria a pena: meu pais, Enzo e Stella, que participaram sempre da minha formação, desde cromossomos e mtDNA até o incentivo e apoio para seguir com os estudos. Sempre serão um exemplo de dedicação, coragem, respeito, sonhos e acima de tudo amor, junto com meus irmãos de quem eu tenho muito orgulho Sílvia (Mari), Cristina (Cristi) e César (Gato) e minhas sobrinhas Aline e Aylén, minhas gatinhas!

- E ao Fábio, que aprendeu a fazer PCRs, correr gel, e usar o Arlequin mesmo fazendo doutorado em Zoologia (*Splendifelix*)!!!!!! Te amo.

Gracias.

SUMÁRIO

RESUMO

Visando avaliar a extensão da diversidade genética do povo gaúcho, e com isso resgatar parte de sua história, foi realizado um estudo envolvendo 547 indivíduos, sendo 278 Nativos Americanos (Guarani e Kaingang) e 269 provenientes de populações miscigenadas do Rio Grande do Sul (RS). Foram estudados marcadores uniparentais de herança materna e paterna utilizando os seguintes sistemas: a) seqüenciamento da região hipervariável I (HVS I) do DNA mitocondrial (mtDNA), determinação de RFLPs (Restriction Fragment Length Polymorphisms) e mini-seqüenciamento da porção codificadora, envolvendo os quatro principais haplogrupos mitocondriais ameríndios (A, B, C e D); b) sete polimorfismos de base única (SNPs) (DYS199, M242, M9, 92R7, sY81, M19 e RPS4Y711), uma inserção Alu (YAP) e onze microssatélites (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439 e DYS385a/b), todos localizados na região não recombinante do cromossomo Y. Além desses marcadores uniparentais, foram obtidos dados para 16 microssatélites do cromossomo X (DXS1001, DXS1047, DXS1060, DXS1068, DXS1073, DXS1106, DXS1214, DXS1226, DXS1227, DXS8051, DXS8055, DXS986, DXS987, DXS990, DXS991 e DXS993).

Analisaram-se 200 Guarani de três parcialidades (Ñandeva, Kaiowá e M'Byá) e 78 Kaingang do Paraná e Rio Grande do Sul, visando identificar diferenças entre as duas tribos, que possam ter ocorrido ao longo do processo histórico. Dezenove linhagens mitocondriais foram detectadas e estas mostraram distribuição diferenciada. A dinâmica de mestiçagem que ocorreu com os Guarani e Kaingang ao longo do tempo foi diversa. O ingresso de genes não-nativos entre as comunidades Guarani foi marcadamente restrito a homens não-ameríndios, enquanto entre os Kaingang há evidências diretas de introdução através do lado materno. Este estudo permitiu desvendar detalhes até então não conhecidos sobre estas duas populações nativas do Rio Grande do Sul, para a história do Estado e para a formação das populações gaúchas atuais.

Já os estudos de populações não-indígenas (N=225) revelaram 94% dos cromossomos no RS como tendo origem européia, 4% ameríndia e 2% africana. Ao levar em consideração as distintas populações aqui investigadas, as quais diferem significativamente em histórias demográficas e de mistura, constatou-se que na Serra 100% das patrilinhagens são de origem européia, enquanto no Pampa há uma parcela de

contribuição ameríndia (8%) e africana (4%), embora a maior parte seja de cromossomos Y europeus. Os microssatélites (STR) dos cromossomos X e Y foram tipados apenas para a amostra do Pampa: para os Y-STRs (N=89), 81 haplótipos foram identificados, dos quais 74 deles (91%) são únicos. Comparando-se estes dados com outros previamente publicados para portugueses, espanhóis, italianos, alemães, africanos e outras populações brasileiras, observou-se a importante contribuição de ibéricos, particularmente de espanhóis, na atual formação masculina do Pampa. Para os X-STRs (N=70) o número de alelos variou de 1 a 14 e os níveis de heterozigosidade entre 0.5565 e 0.8817. Nenhum haplótipo foi encontrado mais de uma vez, indicando que existe uma grande diversidade no Pampa gaúcho.

Com relação aos resultados obtidos para o mtDNA, no RS como um todo (N=225) foram identificadas matrilinhagens européias (63%), ameríndias (30%) e africanas (7%). Porém, da mesma forma que para as patrilinhagens, a distribuição destas variou de acordo com a região estudada. Na Serra 97% dos haplogrupos mitocondriais são característicos de populações européias e apenas 3% têm origem ameríndia. Já no Pampa 51%, 38% e 11% das linhagens mitocondriais têm, respectivamente, origem ameríndia, européia e africana. Considerando-se apenas as linhagens de origem ameríndia, verificou-se que estão assim distribuídas: A – 30%, B – 31%, C – 31% e D – 8%. A marcante diferença nas distribuições destes haplogrupos, quando comparadas com os Guarani bem como com outros resultados, apontaram para a idéia de que outros grupos nativos (principalmente os Charrua), através de suas mulheres, teriam contribuído de maneira marcante para a formação das populações gaúchas contemporâneas. Foi possível verificar ainda que o legado ameríndio (Charrua e Guarani), tão marcadamente presente na cultura gaúcha tradicional, também pode ser visto em nível genômico, num exemplo extraordinário de continuidade genética e cultural entre populações nativas e miscigenadas.

ABSTRACT

To evaluate the extension of Gaucho genetic diversity of the Gauchos, and retrieve part of their history, a study with 547 individuals, of which 278 were Native Americans (Guarani and Kaingang) and 269 admixed from the state of Rio Grande do Sul, was carried out. Uniparental markers, of maternal and paternal inheritance, were studied by using the following systems: a) Mitochondrial DNA (mtDNA) Hypervariable Sequence I sequencing (HVS I), RFLP (Restriction Fragment Length Polymorphisms) determinatins and minisequencing of the coding region, involving the four major mitochondrial Amerindian haplogroups (A, B, C and D); b) seven Single Nucleotide Polymorphisms (SNPs) (DYS199, M242, M9, 92R7, sY81, M19 and RPS4Y711), one Alu insertion (YAP) and eleven microsatellites (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439 and DYS385a/b), all of them located in the nonrecombining region of the Y chromosome (NRY). In addition to these uniparental markers, data for 16 microsatellites of the X chromosome were obtained ((DXS1001, DXS1047, DXS1060, DXS1068, DXS1073, DXS1106, DXS1214, DXS1226, DXS1227, DXS8051, DXS8055, DXS986, DXS987, DXS990, DXS991 and DXS993).

Two hundred Guarani of three partialities (Ñandeva, Kaiowá e M'Byá) and 78 Kaingang from Paraná and Rio Grande do Sul were analyzed, aiming to identify differences between the two tribes, that might have occurred during their long historical process. Nineteen mitochondrial lineages were detected and these showed a distinct distribution. The admixture dynamics that occurred along the time with the Guarani and Kaingang was diverse. The introduction of non-native genes in Guarani communities was markedly restricted to non-Amerindian males, while among the Kaingang there are direct evidences of introduction by the maternal side. This study allowed to reveal details that, until now ,were not known about these two Rio Grande do Sul native populations, which had contributed to the state's history and the formation of the contemporary Gaucho population.

The studies with non-native populations (N=225) revealed that 94% of the Y chromosomes have European, 4% Amerindian and 2% African origins. When considering the distinct populations here investigated, which significantly differ in their demographic and admixture histories, it was detected that in the Serra 100% of the patrilineages have

European origin, while in the Pampa there is a fraction of Amerindian (8%) and African (4%) contributions, although the majority is of European origin. The microsatellites (STRs) X and Y were typed only in the Pampa sample: 81 haplotypes were identified for the Y-STRs (N=89), of which 74 (91%) are unique. Comparing these data with previously published results from Portuguese, Spanish, Italian, German, African and other Brazilian populations, it is noteworthy the major contribution of Iberian, particularly the Spaniards, in the present male fraction of the Pampa. For the X-STRs (N=70) the number of alleles varied from 1 to 14 and the levels of heterozygosity varied between 0.5565 and 0.8817. No haplotype was found more than once, indicating that there is a great diversity in the Gaucho.

When considering the mtDNA data, in the RS as a whole (N=225) European (63%), Amerindian (30%) and African (7%) matrilineages were identified. But in the same way as in the patrilineages, their distribution has much variation according to the studied region. In the Serra 97% of the mitochondrial haplogroups are typical of European populations and only 3% have Amerindian origin. On the other hand, in the Pampa 51%, 38% and 11% of the mitochondrial lineages have Amerindian, European and African origins, respectively. When only Amerindian origin lineages were considered, it was verified that they are distributed as follows: A – 30%, B – 31%, C – 31% and D – 8%. The marked difference in these haplogroups' distributions, when compared to Guarani and other results, pointed to the idea that other native groups (particularly the Charrua), should have contributed, through their women, to the formation of the contemporary Gaucho populations. It was also possible to verify that the Amerindian legacy (Guarani and Charrua), so markedly present in the traditional Gaucho culture, can also be seen at the genomic level, in an extraordinary example of genetic and cultural continuity between native and admixed populations

CAPÍTULO I

Introdução

I.1. POPULAÇÕES ESTUDADAS

I.1. 1. O Gaúcho

> *"Tropeiros, chasques, carreteiros ou andarengos, todos escutam do fogo a mesma voz: 'apeia e vem tomar mate'. E a nenhum deles se pede carteira de identidade. Nem mesmo o nome se pede, pois muitos guaxos não sabem se ao nascer tiveram isso."*
>
> (Barbosa Lessa, 1961).

Ainda que o Rio Grande do Sul tenha uma notável diversidade interna, existe uma tendência a associar seus habitantes a um único tipo social: o cavaleiro e peão de estância que representa a clássica figura do gaúcho[1]. Esta representação foi sendo modificada ao longo do tempo por historiadores, literatos, contadores de causos e folcloristas que foram adicionando elementos evocativos de um passado glorioso. Inicialmente pária social, esses habitantes da região do Pampa eram chamados de Changadores, Guascas e depois Gaudérios, tendo todos estes um caráter pejorativo (Oliven, 1993; Reverbel, 2002).

O desprezo inicial da sociedade foi transformado em medo a medida que estes aumentavam em número e poder. A classe gaúcha tornou-se tão poderosa que infundiu temor e mesmo admiração, uma vez que estes peões tornaram-se úteis à sociedade ao invés de prejudiciais (Nichols, 1946).

É difícil determinar a linha divisória na qual o gaúcho começa a existir como categoria social específica, assimilando essa áurea de glória, mas é certo que esta origem

---

[1] A palavra "gaúcho" neste contexto, refere-se ao conceito cultural e não ao gentílico para o sul rio-grandense, isto é, aquele nascido no Rio Grande do Sul.

gira em torno de fatores como cavalos, pastagens abundantes e rebanhos de gado. Houve, essencialmente, um fenômeno ideológico do processo de construção do gaúcho como campeador e guerreiro inserido num espaço histórico como cavaleiro arrojado, menestrel das campinas, onde atributos como coragem e virilidade são exigidos a todo momento, transportando-o ao plano de mito (Oliven, 1993).

Porém, como destacam Kern *et al.* (1993), conceituar identidade como uma construção imaginária não significa necessariamente que esta seja pura fantasia: "*A construção idealizada do gaúcho, mergulha fundo no passado histórico do Rio Grande do Sul, buscando sinais, traços e práticas que são reagrupados em torno de um novo significado: o de fornecer a uma coletividade uma imagem socialmente aceita, na qual as pessoas se reconheçam*".

A posição fronteiriça que o Rio Grande do Sul ocupa fez com que ele fosse visto como uma área limítrofe, que poderia fazer parte tanto do Brasil quanto de outros países dependendo das forças históricas em jogo (Oliven, 1997). Desta forma, a linguagem dos gaúchos é recheada de expressões espanholas e seus costumes estão muito próximos de seus vizinhos argentinos e uruguaios.

De modo geral, os gaúchos e *gauchos*[2] são reconhecidos como um grupo de homens marcados por uma visível unidade cultural. Isto porque sua origem étnico cultural é basicamente a mesma e reporta para uma herança ibérica, indígena e africana, associada a atividades pastoris nas planícies do Pampa sul-americano, uma área geográfica extensa que sobrepõe os limites de fronteiras dos países da região (Brasil, Argentina e Uruguai).

De acordo com historiadores, o gaúcho típico apareceu primeiramente nas terras do rio da Prata e seu surgimento no Rio Grande do Sul teria sido mais tardio, porém

---

[2] A grafia *gaucho*, sem acento e em itálico, refere-se à palavra em espanhol.

alicerçado em semelhanças decorrentes das peculiaridades do mesmo sistema de atividades – o pastoreio – desenvolvido dentro de um cenário mais ou menos equivalente, enriquecido em ambos os lados da experiência e dos hábitos dos campeadores nativos.

Como herança ibérica recebeu o cavalo e a faca, utensílio da maior importância que servia de arma e era o único instrumento de trabalho no abate do gado e na preparação do couro. Dos nativos americanos (Charruas e Minuanos, entre outros) vieram as boleadeiras, o poncho, o mate e o chiripá assim como diversos objetos de couro que fazem parte dos utensílios e indumentária típica gaúcha (Becker, 2003), além da técnica de dominar o cavalo, como descreve Pi Hugarte (1998): *"De los Charrua heredaron los gauchos la técnica de gobernar el caballo con un bocado de cuero unido a las riendas durante las primeras jineteadas al potro – es decir, cuando es todavía redomón – poniéndole el freno cuando el animal se había acostumbrado a ser dirigido por las riendas unidas al dicho bocado. Este sistema difiere del traído por los europeos – portugueses o españoles – puesto que estos domaban el animal embreándolo y haciendo que se volviera hacia uno u otro lado tirando separadamente de cada una de las riendas, en tanto que en el manejo del caballo al estilo indio, el animal cede fácilmente para cualquier lado moviendo las dos riendas unidas, este es el procedimiento riograndense y rioplatense"*.

Ambos tipos, tanto o *gaucho* platino como o gaúcho rio-grandense, incorporaram à sua linguagem elementos indígenas e em menor escala, elementos africanos, mesclando-os ao português e espanhol. Criaram nesta região bilíngüe entre países de língua hispânica e portuguesa, um vocabulário próprio, com arcadismos de ambos idiomas e mútuas interpretações e influências (Reverbel, 2002). Surgiu também uma literatura gauchesca, incorporando as lendas de sua tradição oral e as particularidades dialetais, exaltando sua coragem, apego à terra e seu amor à liberdade (Oliven, 1996).

A bibliografia platina sobre o *gaucho* é das mais vastas, uma vez que no Prata ele ocupa um espaço social, cultural e histórico muito abrangente, ao contrário do gaúcho rio-grandense, que ocupa um pequeno espaço correspondente a uma unidade federativa do Brasil (Reverbel, 2002).

A origem da palavra GAÚCHO é cercada de divergências, pois existem várias hipóteses etimológicas e muito debate acadêmico a respeito do assunto (Leal, 1989). Se for inserido o termo "*gaucho*", essa gama aumenta ainda mais, alternando em duas frentes de expressão: a história e a literatura. O escritor uruguaio Buenaventura Caviglia Hijo fez uma extensa pesquisa e listou trinta e seis possíveis origens da palavra, segundo autores argentinos, uruguaios e brasileiros. Entre as línguas apontadas estão o português, tupi, guarani, árabe, espanhol, araucano, charrua, latim, gitano, alemão, francês, inglês, aimará, hebráico entre outros (Reverbel, 2002).

Muitas versões podem ser encontradas, algumas curiosas como uma origem francesa a partir da palavra *gauche* que significa literalmente "esquerda", mas pode também significar "fora da lei" ou também uma que inclui a palavra inglesa *gawk* ou *gawky*[3] a qual supostamente expressa "*a deselegante maneira desses rústicos*". Alguns autores afirmam que o termo vem do Guarani e significaria "homem que canta triste", aludindo provavelmente à "*cantinela arrastada dos charruas*" (Leal, 1989). Ou ainda que a palavra é de origem espanhola, a partir da designação que os colonizadores espanhóis adotavam para referir-se aos povos nativos das Ilhas Canárias: *guanche*, que ao longo do tempo teria se transformado em *guancho* e depois *gaucho* (Barbosa Lessa, 1978). A alternativa mais plausível, sustentada pela maioria, é a de que a palavra tenha uma origem

---

[3] Gawk: *s.* (coloq) palerma, paspalhão. Gawky *adj.* desajeitado (Marques and Draper, 2001).

indígena, vinda de *guacho*, que no vocabulário gaúcho significa "órfão". *Guacho* vem do Quechua, originalmente *hauck-cha* (Leal, 1989).

A existência, ou manutenção, do gaúcho talvez seja uma das questões de maior indagação entre pensadores e folcloristas rio-grandenses e platinos, atuais e antigos. O gaúcho é senhor da fronteira, sua identidade se forma no intervalo entre ser brasileiro, argentino, uruguaio, ibérico ou nativo americano (Figueiredo, 2006), e o lugar onde ele vive, descrito por Assis Brasil (1996) *"em sua majestosa amplidão de pradarias, o Pampa chama-nos à ancestralidade, à terra, instituindo-se em território cheio de metáforas, de existência mais lírica que real"*.

I.1.2. A Grande Etnia Charrua

Como o próprio nome reporta os Pampeanos eram indígenas que habitaram as extensas planícies de gramíneas do Pampa sul-americano. Neste grupo podem ser incluídas várias tribos, sendo as mais conhecidas e importantes em termos históricos e demográficos os Charrua[4] e Minuano (Becker, 2003). Para alguns historiadores, existiu na verdade uma grande etnia (ou família) Charrua, que englobaria os Charrua, os Minuano e várias tribos pampeanas relacionadas como Yaró, Bohane, Guenoa entre outras. Isto ocorre devido à ausência de documentos históricos e outros registros, o que torna muito difícil estabelecer diferenças suficientemente marcantes para que estas possam ser tratadas como grupos separados (Alemán, 1994; Pi Hugarte, 1998). Adotando tal critério, a denominação

---

[4] Escreve-se Charrua no singular e não Charruas, no plural, pois segundo a convenção para a Grafia de Nomes Tribais, assinada pelos participantes da 1ª Reunião Brasileira de Antropologia, no Rio de Janeiro, em 1953, para uniformizar a maneira de escrever os nomes das sociedades indígenas em língua portuguesa: "os nomes tribais, quer usados como substantivos, quer como adjetivos, não terão flexão de gênero e de número, a não ser que sejam de origem portuguesa ou morficamente aportuguesados" (Derengoski, 2002). O mesmo critério foi adotado a todas as grafias tribais, exceto quando trata-se de um texto escrito em espanhol.

"grande etnia Charrua" ou simplesmente "Charrua" será utilizada neste trabalho, exceto nos casos de citações literais.

Os Charrua habitavam a antiga Banda Oriental do Uruguai, que atualmente corresponde aos Departamentos de Tacuarembó, Rivera, Artigas, Salto e Paysandu, no Uruguai, além de uma boa parte do sudoeste do Rio Grande do Sul, sendo o limite setentrional definido pelos rios Jacuí e Ibicuí. Chegaram a ocupar regiões que hoje correspondem às províncias argentinas de Entre Ríos, Santa Fé e Corrientes (Becker, 2003).

Com relação aos estudos arqueológicos das sociedades e culturas indígenas sul-americanas, o autor uruguaio Pi Hugarte (1998) discorre sobre a sucessão de vários estratos culturais anteriores ao desenvolvimento das culturas conhecidas quando da chegada dos europeus. É no estrato correspondente ao dos chamados "caçadores superiores" que ele identifica a cultura de vinculações patagônicas. O autor sugere uma relação entre estes e os Charrua. Essa ligação foi atribuída pelo emprego de técnicas avançadas, visivelmente similares, da escultura de pontas e polimento de pedras mostrando semelhanças na indústria lítica entre estes dois povos que habitaram o mesmo espaço geográfico em épocas diferentes. Desta forma, esta ligação estaria na verdade mostrando dois momentos diferentes da evolução tecnológica de um mesmo grupo indígena. Os dados arqueológicos no Rio Grande do Sul reforçam também esta postulação (Mentz-Ribeiro, 1997). A partir do sétimo milênio antes do presente, o Estado foi progressivamente ocupado por grupos populacionais de três culturas com tradições tecnológicas bem definidas, nenhuma delas possuindo artefatos de cerâmica. Estas culturas são conhecidas pelos arqueólogos como Tradição Umbu, típica de ambientes abertos e caracterizada por uma indústria lítica que fabricava entre outras coisas, bolas de boleadeiras; Tradição Humaitá, típica de florestas

meridionais, e a Tradição de Sítios Litorâneos ou Sambaquiana, típica do litoral norte do Estado. Para alguns pesquisadores haveria uma relativa continuidade na distribuição geográfica destas populações pré-históricas, caracterizada ainda pela sucessão de estratos culturais ao longo do tempo, a partir da Tradição Umbu, passando pela Tradição Vieira (ceramista) até os Charrua. Assim, estes seriam os últimos descendentes dos primeiros humanos que chegaram e colonizaram o Pampa sul-americano (Kern, 1997).

Os Charrua eram nômades. O sistema de caça foi complementado com a coleta de outros produtos alimentícios de origem animal ou vegetal. Não há dados que permitam afirmar a existência da divisão de tarefas em função do sexo, mas convém acreditar, de acordo com a etnografia geral, que a caça deveria ser uma atividade masculina, ficando a coleta a cargo de mulheres e crianças. Não há registros de que tenham utilizado cestaria e mesmo a cerâmica não era muito elaborada. A pedra e o couro foram elementos básicos com os quais confeccionaram os instrumentos necessários para a subsistência e atividades bélicas, como as boleadeiras – armas arrojadas, formadas por bolas de pedra polida, presas às extremidades de tiras de couro trançado ou retorcido (Pi Hugarte, 1998).

Historicamente comprovada a maior alteração sofrida pela cultura destes caçadores-coletores está relacionada à ação do cavalo, que foi introduzido em 1607, e à introdução do gado bovino (1634). Os Charrua modificaram seus hábitos alimentares e rapidamente tornaram-se exímios cavaleiros. Vale a pena transcrever o relato das observações recolhidas por Cezimbra Jacques: *"montaban en pelo, poniendo solamente un trozo de cuero sobre el lomo del animal. Sabían combatir en caballos alineados y cargar lanza y boleadoras. Utilizaban una estratagema que consistía en acostarse sobre su montura o en estirarse a uno de sus lados, no permitiendo así que se lo percibiera desde lejos. Amansaban a la perfección sus cabalgaduras y tanto en la paz como en la guerra, sabían*

sacarles el mayor provecho posible. Les era indiferente andar montados o acostados sobre el lomo de sus caballos y muchas veces se ocultaban debajo de ellos, por eso, sus enemigos difícilmente podían distinguir una banda de Charruas. Utilizando esa táctica obtenían ventajas en la caza y en la guerra pues así sorprendían al enemigo que no los percibía entre las tropas de animales que pastaban" (Pi Hugarte, 1998).

O amplo desenvolvimento do gado foi o motivo das incursões de estancieiros para o interior. Junto com eles os "changueadores" espanhóis, que buscavam couro para seus negócios, dando início assim ao grande tráfico de contrabando que constituiu um novo elemento de introdução ao povoamento da antiga Banda Oriental. Os Charrua não aceitaram facilmente se desfazer do seu modo de vida tradicional e rejeitaram, em grande parte, as propostas de aldeamento fixo à maneira dos colonizadores (Pi Hugarte, 1998; Becker, 2003). A população indígena, que antes ocupava o território inteiro, viu-se com espaços limitados e foi atingida de muitas formas sofrendo um violento processo de transformação. Cada uma das formas de ocupação dos colonizadores os atingiu de maneira diferente, seja pela introdução de gado ou pelas missões franciscanas e jesuítas, que tentaram incorporá-los à sociedade colonizadora, os Charrua começaram a reagir ante as frentes expansionistas e passaram a representar uma ameaça aos bens e propriedades da população colonizadora que exigiu, em contrapartida, o aniquilamento destes indígenas (Becker, 2003).

É importante lembrar que por trata-se de um povo extinto. Desta forma, as referências históricas a respeito dos Charrua são, em sua maioria, documentos referentes à conquista, com a visão dos colonizadores. A ausência quase total de elementos da sua cultura material ou representações de pouco valor etnográfico, fazem com que os fatos sejam apresentados de maneira tendenciosa, retratados de forma a justificar a

"necessidade" de eliminá-los, o que fica evidenciado nas duras palavras do naturalista francês Georges Louis Lecler – o Conde Buffon – que referiu-se a eles como "*selvagens animais de primeira categoria*" (Vidart, 2000; Pi Hugarte, 1998).

Em 1797 foi criada uma polícia de campanha (*Cuerpo de Blandengues* da Fronteira de Montevidéu) que tinha como objetivo perseguir os Charrua, os contrabandistas portugueses e os ladrões de gado que invadiam o campo.

O ataque mais significativo contra os Charrua em território uruguaio aconteceu de forma traiçoeira em 1812, quando caciques, suas mulheres e tantas pessoas quanto possível, foram convidadas para uma comemoração em Salsipuedes, Paysandu, onde lhes foi dada erva, aguardente e tabaco e, em meio às festividades, os soldados se lançaram sobre eles exterminando-os. Apenas poucos conseguiram escapar, bem como aqueles que, desconfiados, não tinham aceito o convite (Pi Hugarte, 1998; Vidart 2000; Bracco, 2004).

Os índios restantes foram definitivamente eliminados depois no Combate de Mataojos. Alguns poucos sobreviventes foram presos e repartidos nas cidades para fins vários ou entregues às estâncias como peões. Registros históricos indicam que o genocídio contra este povo pode ter sido desigual: enquanto os homens eram sacrificados ou pereceram em batalhas, as mulheres e crianças foram distribuídas entre as famílias "brancas" para o serviço doméstico, adaptando e integrando-se à vida e costume destas famílias (Alemán, 1994; Becker, 2003).

Na Argentina, os Charrua levados a Cayastá (Província de Santa Fé) para serem aldeados, não assimilaram a vida sedentária nem ficaram resignados a esta transferência imposta. Espalharam-se pelos arredores e desta maneira, após anos de sofrimento, foram extintos, desaparecendo com eles os registros da etnia Charrua, pelo menos do território santafesino (Alemán, 1994).

Além da matança direta como causadora da extinção dos Charrua não se pode excluir a ação destruidora que tiveram as epidemias trazidas pelos europeus (Pi Hugarte, 1998).

Em junho de 1833, quatro índios identificados como Charrua foram levados a Paris, para exibições exóticas: o cacique Vaimaca Peru[5], Senaqué, Laureano Tacuabé e Micaela Guyunusa. O índio Senaqué morreu pouco tempo depois de sua chegada à França, em julho do mesmo ano. Seu corpo foi transportado ao laboratório de Anatomia Humana do Museu do Homem, em Paris, mas os registros foram perdidos. Micaela Guyunusa, que estava grávida, teve uma filha no dia 20 de setembro de 1833. A única índia Charrua levada à Europa, morreu pouco tempo depois, na cidade de Lyon, em 22 de julho de 1834, quando também se perderam os registros da criança e de Laureano Tacuabé. Com relação ao Cacique Peru, também este viveu um curto período após ter saído do Uruguai. Morreu nos últimos meses de 1833 e seu corpo também foi levado ao Museu do Homem de Paris, onde permaneceu em exposição por muitos anos. Em 2002, os restos do cacique foram repatriados ao Uruguai e depositados no Panteón de la Pátria, em Montevidéu. Mediante um convênio firmado entre o *Ministério de Educación (MEC) y Cultura* e a *Facultad de Humanidades y Ciencias de la educación (FHCE), Universidad de la República*, foram realizados estudos multidisciplinares com tais restos, incluindo análises moleculares. Os resultados deste esforço coletivo na busca de dados genéticos e antropológicos sobre os Charrua, fornecidas pelas investigações detalhadas com os restos ósseos do lendário cacique, estão divulgados na rede mundial de informações (Internet), no *site* http://www.fhuce.edu.uy/antrop/cursos/abiol/vaimaca.html. Vale a ressalva que, quase

---

[5] Apesar do nome deste cacique ser bastante discutido, adotou-se aquele usado por Rivet (2002).

duzentos anos depois de sua morte, Vaimacá Peru está ajudando a resgatar a história de seu povo.

O contato com colonizadores, mestiços e outros indígenas causou um processo de deterioração econômica e cultural vertiginosa nos Charrua, colocando em desuso instrumentos originais para adaptar, na medida do possível, tais novidades. Desta forma, as armas passaram a contar com pontas metálicas e assimilaram a faca (*cuchillo*) que da mesma forma que os gaúchos era usada nas costas, atravessada no cinto. Por outro lado, o roubo e contrabando foram absorvidos como fontes de sobrevivência (Flores, 2003; Becker, 2003).

Legaram muitos termos que ainda são usados no Rio Grande do Sul, como chiripá, poncho, chasque, mate, quasca, Pampa entre outros (Flores, 2003). Palavras no vocabulário juntamente com objetos de uso, além de determinados costumes, já citados anteriormente, representam a importante influência da grande etnia Charrua para a formação da cultura gaúcha (Becker, 2003).

I.1.3. Guarani

A bibliografia relativa aos Guarani é provavelmente a mais vasta, uma vez que desde o século XVI eles têm sido objeto de inúmeras obras descritivas e que atualmente compõem um dos mais numerosos registros de povos indígenas sul-americanos (Monteiro, 1992).

Lingüisticamente definidos como parte do tronco Tupi (Campbell, 1997), dominavam as florestas subtropicais do Rio Grande do Sul, Santa Catarina, Paraná (Brasil) e Misiones (Argentina), assim como as florestas tropicais de São Paulo, Mato Grosso do Sul, Paraguai e Bolívia (Schiavetto, 2003).

Praticavam a antropofagia ritual, comendo os prisioneiros de guerra por vingança. Apenas os chefes eram polígamos, uma vez que estes precisavam de mulheres que cuidassem da comida e de objetos de seus subordinados para manterem assim a chefia. Além disso, ofereciam suas mulheres a outros homens em troca de objetos ou em penhor de uma aliança, costume que pode ter facilitado a mestiçagem com não-indígenas. Os Guarani forneceram alimentos e suas mulheres aos espanhóis, porque queriam aliados para combater seus inimigos de outras tribos indígenas (Flores, 2003).

De acordo com Pi Hugarte (1998), o estrato cultural mais tardio correspondente aos Guarani está situado em uma época bem mais recente à dos povos de "Vinculações Patagônicas". Do ponto de vista arqueológico, encontram-se restos de sua cerâmica no curso inferior do rio Uruguai e próximo ao litoral platense, bem como a oeste do Uruguai, indicando duas vias de penetração de grupos que procediam de regiões muito separadas, embora provavelmente falassem a mesma língua.

Os Guarani do Brasil Meridional podem ser divididos de acordo com diferenças lingüísticas e peculiaridades da cultura material e não-material em três grandes grupos: Ñandeva, Kaiowá e M'Byá (Schaden, 1962).

O subgrupo ou parcialidade Ñandeva ("os que somos, os que são dos nossos") é encontrado principalmente no extremo sul do Mato Grosso, a poucos quilômetros da fronteira paraguaia, nas aldeias Jacareí e Porto Lindo. Os índios destas povoações se vestem à maneira dos sertanejos mas, de modo geral, não sofreram influências muito incisivas. O subgrupo Kaiowá, também conhecido como *Teüi* (designação utilizada pelos Guarani para todos os índios, independente de qual seja a tribo, e que significa "naturais da terra") parece estar também confinado a uma série de aldeias do sul do Mato Grosso (como Dourados, Panambi, Taquapiri, Amambaí e outras) e das regiões contíguas do Paraguai. O

terceiro subgrupo, M'Byá ("gente") ocupava a maior parte do território que hoje compreende o Estado do Rio Grande do Sul (Kern *et al*., 1993; Flores, 2003). Atualmente, existe uma série de aldeias (em áreas de reservas) M'Byá no oeste de Santa Catarina e Paraná, bem como no leste paraguaio e também na porção setentrional da Argentina e Rio Grande do Sul (Schaden, 1962).

O declínio da população Guarani foi devido a causas não muito diferentes daquelas que vitimaram outras populações nativas: novas doenças de origem européia ou africana, associadas a ataques dos bandeirantes, alem da Guerra Guaranítica, da escravidão imposta pelo governo militar espanhol nas reduções depois da expulsão dos Jesuítas e da mestiçagem das mulheres com homens não indígenas (Flores, 2003).

Atualmente, muitos Guarani estão inseridos na sociedade ou confinados em reservas dos Estados da região sul, além do Mato Grosso do Sul, muitas vezes dividindo espaço com indígenas de outras etnias (Ricardo, 2000; Schiavetto, 2003).

Deixaram como herança vários termos utilizados na linguagem coloquial: araçá, caboclo, capim, capivara, capoeira, chê, cuia, goiaba e guri entre muitas outras (Flores, 2003).

I.1.4. Kaingang

De acordo com Campbell (1997), os Kaingang incluem-se no tronco lingüístico Jean e junto com os Xokleng representam o ramo mais meridional da família Jê (ou Gê).

No Rio Grande do Sul, alguns acreditam que estabeleceram-se em épocas relativamente recentes, após terem vindo do noroeste devido a lutas com os Botocudos do Brasil Central que os impeliram rumo ao sul, tendo entrado em Santa Catarina até chegar ao Rio Grande do Sul. Para outros no entanto, seriam descendentes dos nativos que em

épocas mais remotas fizeram as casas subterrâneas (Schmitz e Becker, 1997). Certo, entretanto, é que se moveram muito após a chegada dos primeiros colonizadores, indo e vindo ao longo dos Estados do sul e sudeste, sempre fugindo do contato. Antigamente eram conhecidos como Guaianá, havendo também outros subgrupos tais como Coroado, Pinaré, Ibijara, Caaguá e Gualacho, além de serem conhecidos pelos colonizadores como "bugres" (Flores, 2003; Carneiro da Cunha, 1992) no Rio Grande do Sul (Schmitz e Becker, 1997).

Viviam em pequenos grupos, formados por famílias entrelaçadas e parentes chegados; Seus alojamentos eram ranchos com tamanho proporcional ao número de ocupantes, status e hierarquia grupal. Mostravam grande respeito pelas mulheres e sua organização era, em regra geral, monogâmica. Tinham divisão de trabalho por sexo e a estrutura social era a partir das famílias que mantinham certa estabilidade. O poder de governar dentro da tribo era irrestrito (Becker, 1995).

O ambiente natural preferido eram as matas situadas nos lugares mais altos do Planalto rio-grandense, em meio aos pinheirais, de onde mantinham controle visual da vizinhança (Becker, 1995; 1999). A economia do grupo era baseada na colheita especialmente de pinhão mas também do mel e frutos silvestres. Também praticavam caça e pesca, sendo que estas eram feitas com dardo e flechas (Becker, 1995).

Foram dizimados pelas epidemias de origem européia e africana e pela ação de bandeirantes e bugreiros que recebiam pagamento por índio morto. O resultado dessa ação foi um esvaziamento demográfico nas trilhas das tropas de gado. Em 1882, Telêmaco Morocines Borba reuniu os sobreviventes não-Guarani em reservas, atribuindo-lhes o nome de Kaingang (ou Caingangue), agrupando todos aqueles que lingüística e culturalmente formavam o ramo meridional da Família Jê (Becker, 1995; Flores, 2003).

Devido ao contato com o colonizador, muitos dos seus valores e costumes caíram totalmente em desuso enquanto outros permaneceram, embora com modificações. Considerando-se as regiões geográficas do Rio Grande do Sul, a área atual dos Kaingang abrange: Litoral (norte), Campos de Cima da Serra, Encosta Inferior e Superior do Nordeste, Planalto Médio e Depressão Central, sempre ocupando as áreas mais altas (Becker, 1995).

I.2. MARCADORES GENÉTICOS

I.2.1. DNA mitocondrial (mtDNA)

O mtDNA humano constitui-se de um tipo único de DNA circular principalmente de fita dupla, cuja seqüência de 16.569 pares de bases foi estabelecida por Anderson *et al.* (1981). Esta seqüência é conhecida como CRS (*Cambridge Reference Sequence*) e foi utilizada como referência até 1999, quando foi subsitutída pela seqüência revisada por Andrews *et al.* (rCRS – *revised Cambridge Reference Sequence*). Está presente em milhares de cópias em cada célula, não no núcleo, mas sim nas mitocôndrias, organelas responsáveis pela produção de energia gerada no processo de fosforilação oxidativa (Bravi, 2005).

A alça de deslocamento (alça D ou *D-loop*) é uma pequena seção desprovida de qualquer DNA codificador conhecido, definida por uma estrutura de três fitas de DNA, devido a um curto segmento de fita pesada que é replicado uma segunda vez, gerando uma estrutura também conhecida como 7S DNA (Strachan e Read, 2002). A alça D é a região mais variável do genoma mitocondrial e a maior parte dos sítios polimórficos desta alça são concentrados em três segmentos hipervariáveis (HVS *hypervariable segment*), HVS I, HVS II e HVS III, sendo que a maioria das informações de seqüências de mtDNA publicadas até o momento são relativas à HVS I (Lutz *et al.*, 2000),

Estudos recentes comparando mtDNA de primatas e humanos indicaram que a taxa de mutação é cerca de dez vezes maior que a taxa média de mutação do DNA nuclear. Esta elevada taxa leva a um grande número de seqüências diferentes na população, uma das razões pela qual o mtDNA é uma importante ferramenta evolutiva, particularmente útil para estudos de populações proximamente relacionadas (Jobling *et al.*, 2004).

A herança do mtDNA é do tipo matrilinear, ou seja, homens e mulheres herdam mitocôndrias de suas mães mas apenas as filhas podem transmiti-las às gerações subseqüentes (Strachan e Read, 2002). O somatório de diversas características permite utilizar o mtDNA para construir filogenias moleculares precisas sem a ambigüidade causada pela recombinação, aliado à sua simples estrutura de organização, alta taxa de mutação, herança citoplasmática matrilinear e ao fato de ser essencialmente haplóide (Pakendorf e Stoneking, 2005; Jobling *et al*., 2004, Matioli, 2001). A construção de árvores filogenéticas ou *networks* usando mtDNA é muito adequado para analisar as relações evolutivas entre seqüências individuais. Além disso, as distribuições geográficas das linhagens numa árvore ou *network* podem servir como marcadores geográficos por serem específicas de certos continentes. (Pena e Bortolini, 2004).

I.2.2. Marcadores do Cromossomo Y

Agora que o seqüenciamento do genoma humano está terminado, o novo conhecimento das seqüências de DNA, seus genes e polimorfismos, forneceu uma nova geração de marcadores úteis para o estudo da diversidade do cromossomo Y em populações humanas (Jobling e Tyler-Smith, 2003).

O cromossomo Y é extremamente pobre em genes, os quais codificam apenas 27 proteínas, quando comparado aos 717 genes do cromossomo X.

Sua herança é do tipo patrilinear, isto é, passa exclusivamente do pai para os filhos homens. Possui importantes características que o diferencia dos outros cromossomos, tais como uma maior taxa mutacional e devido à falta de um cromossomo homólogo, ausência de recombinação durante a meiose. Desta forma, os haplótipos do cromosso Y usualmente

passam intactos de geração a geração. Eles mudam apenas por mutação e assim preservam registros de sua história (Joblin *et al*., 2004, Joblin e Tyler-Smith, 2003).

Desde a descoberta dos primeiros polimorfismos no cromossomo Y, o número de marcadores vem aumentando significativamente (Joblin *et al*., 2004). A porção não recombinante do cromossomo Y (NRY) possui polimorfismos como os microssatélites (ou STR - *Short Standem Repeat*) e polimorfismos bialélicos (SNPs – *Single Nucleotide Polymorphisms*).

Os polimorfismos de STR do cromossomo Y (Y-STR) são baseados em repetições em tandem, que podem ser di-, tri- ou tetranucleotídicos. São bastante utilizados em análises forenses (Silva *et al*., 2005) e apresentam como vantagens o fato de serem identificados com técnicas simples (PCR ou sequenciamento), assim como terem seus alelos descritos, podendo ser definidos sem ambiguidade pelo número exato de repetições. Numerosas publicações com populações de todas as partes do mundo têm gerado grande volume de informação sobre freqüência destes marcadores. Estes resultados, em geral, são depositados em bancos de dados especializados tais como o Y-Chromosome Haplotype Reference Database (www.yhrd.org).

Os SNPs são muito mais freqüentes que os STRs. São marcadores bialélicos e apresentam uma baixíssima taxa de mutação, tornando-os bons indicadores para estudos evolutivos e com uma crescente aplicação em genética forense, apesar do seu pequeno poder de discriminação, em determinadas situações, quando comparado aos STRs (Silva *et al*., 2005; Ridley, 2006). O grande interesse na identificação dos SNPs também pode ser atribuído ao seu potencial uso como marcadores moleculares no estudo de associação à doenças (Jobling *et al*., 2004).

O cromossomo Y fornece importantes resoluções filogenéticas e atualmente 159 haplogrupos estão definidos por marcadores bialélicos (Jobling e Tyler-Smith 2003). Inúmeros SNPs vêm sendo identificados na NRY (Underhill *et al.*, 2000), sendo que estes, conjuntamente com os STRs permitem a caracterização dos cromossomos Y. A partir daí pode-se, com relativa precisão, indicar a origem de um determinado cromossomo, e por conseqüência, fornecer dados para o resgate de história evolutiva das populações humanas (Underhill *et al.*, 2000; Bortolini *et al.*, 2002; 2003; Carvalho-Silva *et al.*, 2006).

Devido a estas características particulares, os marcadores do cromossomo Y representam a contrapartida masculina para as inferências obtidas através do estudo do mtDNA. Sendo assim, o estudo concomitante de polimorfismos de herança paterna (cromossomo Y) e materna (mtDNA) em uma mesma população tem permitido, entre outras coisas, estabelecer a natureza dos cruzamentos gênero-étnico específicos, de especial importância no estudo de populações miscigenadas (Carvajal Carmona *et al.*, 2003).

I.2.3. Marcadores do Cromossomo X

Os cromossomos sexuais humanos X e Y diferem muito entre si, porém, pareiam-se na prófase I masculina, assegurando que na anáfase I cada célula filha receba um cromossomo sexual (X ou Y). O pareamento X-Y é possível devido a uma curta região de homologia de 2,6Mb nas pontas dos braços curtos de cada cromossomo (Strachan e Read, 2002). Surgiram de um par de cromossomos inicialmente homólogos, onde seus genes, em geral, não difeririam em um indivíduo. Depois, quando cessou a troca de genes entre estes cromossomos os genes de um (cromossomo X) divergiram evolutivamente dos genes do

outro (cromossomo Y) sendo que a dimensão total desta divergência depende agora do tempo decorrido desde que cessaram as trocas gênicas (Ridley, 2006).

Os genes dos segmentos de X e Y que pareiam apresentam propriedades interessantes: (1) estão presentes, como cópias homólogas em ambos cromossomos; (2) não estão sujeitas à inativação do X e (3) devido ao *crossing over*, os alelos desses loci não apresentam padrões de herança previstos para genes ligados ao X ou ao Y mas segregam como autossomos (Jobling *et al*., 2004; Ridley, 2006)

Uma vez que nos homens o cromossomo X (ChrX) é haplóide, as tipagens de marcadores nas regiões de baixa ou nenhuma recombinação permitem o estudo de linhagens (haplótipos), o que obviamente não é possível em mulheres, uma vez que estas carregam duas cópias do ChrX (Jobling *et al*., 2004; Pereira e Pena, 2006).

O ChrX tem se mostrado um importante instrumento para estudos populacionais devido às propriedades intrínsecas tais como haplótipos acessíveis em homens e baixa taxa de recombinação quando comparado aos outros cromossomos (Lann *et al*., 2005; Schaffner, 2004). Embora existam diversos estudos com informações contidas em STRs autossômicos e Y-STRs, a utilização de microssatélites do ChrX está recém começando. Porém, tem sido demonstrado que os microssatélites do cromossomo X (X-STRs) complementam de forma eficiente as análises com outros marcadores como o mtDNA ou cromossomo Y (Szibor *et al*., 2003). Desta forma, espera-se que em curto espaço de tempo, um conjunto grande de informações com este tipo de marcador esteja disponível na literatura.

Diante do exposto acima, este trabalho visou avaliar a extensão da diversidade genética de populações do Rio Grande do Sul, bem como sua estrutura populacional e dinâmica de mestiçagem através de marcadores uniparentais de herança materna (mtDNA),

paterna (cromossomo Y), além de microssatélites localizados no cromossomo X. Uma das regiões amostradas é o Pampa, local de origem do elemento étnico/cultura conhecido como Gaúcho. Em contrapartida, uma região de mais recente colonização européia também foi amostrada. Além disso, o trabalho também buscou caracterizar populações nativas (Guarani e Kaingang) que habitavam a região antes da chegada dos colonizadores europeus. Os resultados obtidos com o presente estudo puderam ser reunidos nos seguintes artigos científicos:

1. Marrero AR, Leite FPN, Carvalho BA, Peres LM, Kommers TC, Cruz IM, Salzano FM, Ruiz-Linares A, Silva Jr WA and Bortolini MC (2005) Heterogeneity of the genome ancestry in individuals classified as white in the State of Rio Grande do Sul, Brazil. Am J Hum Biol 17: 496-506.

2. Marrero AR, Silva Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, Ruiz-Linares A, Salzano FM and Bortolini MC (2006) The demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. Am J Phys Anthropol, in press.

3. Marrero AR, Wang S, Salzano FM, Ruiz-Linares A and Bortolini MC. Population data on 17 X-chromosome short tandem repeat loci in a sample from southern Brazil. Manuscrito em preparação.

4. Marrero AR, Bravi CM, Stuart S, Long JC, Leite FPN, Kommers TC, Carvalho CMB , Pena SDJ, Ruiz-Linares A, Salzano FM and Bortolini MC. The Gaúcho genetic history – gene dynamics in southern Brazil. Manuscrito em preparação.

Além disso, estudos adicionais resultaram em três outras publicações que podem ser vistas em anexo (1, 2, 3) . Estes três trabalhos, embora não previstos nos objetivos

iniciais estão aqui apresentados pois se inserem dentro do contexto da discussão geral dos resultados.

1. Vargas AE, Marrero AR, Salzano FM, Bortolini MC and Chies JAB (2006) Frequency of CCR5Δ32 in Brazilian populations. Braz J Med Biol Res 39: 321-325.

2. Silva WA, Bortolini MC, Schneider MPC, Marrero A, Elion J, Krishnamoorthy R and Zago MA (2006) mtDNA haplogroup analysis of Black Brazilian and Sub-Saharan populations: implications for the Atlantic slave trade. Hum Biol 78: 29-41

3. Hünemeier T, Carvalho C, Marrero AR, Salzano FM, Pena SDJ and Bortolini MC. Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. Manuscrito submetido.

CAPÍTULO II

Heterogeneity of the genome ancestry in individuals

classified as White in the State of Rio Grande do Sul, Brazil

Original Research Article

Heterogeneity of the Genome Ancestry of Individuals Classified as White in the State of Rio Grande do Sul, Brazil

ANDREA RITA MARRERO,[1] FÁBIO PEREIRA DAS NEVES LEITE,[1,2] BIANCA DE ALMEIDA CARVALHO,[1,2] LEANDRO MARTINS PERES,[1,3] TRÍCIA CRISTINE KOMMERS,[2] IVANA MÂNICA DA CRUZ,[4] FRANCISCO MAURO SALZANO,[1] ANDRES RUIZ-LINARES,[5] WILSON ARAÚJO DA SILVA JÚNIOR,[6] AND MARIA CÁTIRA BORTOLINI[1]*

[1]*Departamento de Genética, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, RS, Brazil*
[2]*Instituto Geral de Perícias, 90160-093 Porto Alegre, RS, Brazil*
[3]*Universidade Luterana do Brasil, 92450-900 Canoas, RS, Brazil*
[4]*Laboratório de Geriatria e Gerontologia, Pontifícia Universidade Católica, 90610-001 Porto Alegre, RS, Brazil*
[5]*Department of Biology, University College London, London NW1 2HE, United Kingdom*
[6]*Departamento de Genética e Centro de Terapia Celular, Universidade de São Paulo, 14049-900 Ribeirão Preto, SP, Brazil*

ABSTRACT   One hundred nineteen individuals classified as White, living in different localities of the Brazilian state of Rio Grande do Sul, were studied in relation to the HVS-I region of the mitochondrial DNA (mtDNA). The male fraction of the sample ($N = 74$) was also tested for seven Y-chromosome polymorphisms. In a specific population (Veranópolis), a city characterized by a large influence of the Italian immigration of the 19th century, the results from the maternal and paternal sides indicated almost complete European ancestry. However, another sample identified as White, from different localities of Rio Grande do Sul, presented significant fractions of Native American (36%) and African (16%) mtDNA haplogroups. These results indicate that Brazilian populations are remarkably heterogeneous; while some present an overwhelming majority of transplanted European genomes, with a complete correspondence between physical appearance and ancestry, others reflect a history of extensive admixture with dissociation between physical appearance and ancestry. Am. J. Hum. Biol. 17:496–506, 2005.   © 2005 Wiley-Liss, Inc.

Brazil, a country of continental size and at present inhabited by ∼170 million persons (Brazilian Institute of Geography and Statistics (IBGE) Census 2000; http://www.ibge.gov.br), was first colonized by a wide array of Amerindian groups, who arrived there thousands of years before the first Europeans. The size of the Native American population at the time of the Portuguese discovery (1500 AD) was composed of ∼2 million of persons (Callegari-Jacques et al., 2003). Colonization of the new country involved mostly European men since the immigration of European women during the first centuries was very small (Carvalho-Silva et al., 2001). It is estimated that 500,000 Portuguese arrived in Brazil during the 1500–1808 period (Carvalho-Silva et al., 2001). Additionally, starting in the 16th century and continuing until 1855, about 4 million Africans, mainly from west-central Africa, were forced to migrate to Brazil (Bortolini et al., 2004a). These cir-

cumstances determined that the first Brazilians arose mostly by unions between Portuguese males and Amerindian and African females (Bortolini et al., 1999; Carvalho-Silva et al., 2001; Salzano and Bortolini, 2002).

A second large wave of migrants occurred in the 19th and early 20th centuries. During the 1820–1975 period, 5,686,133 immi-

grants, mostly Europeans, arrived officially in Brazil. Portuguese and Italians arrived in almost equal numbers (1.8 million and 1.6 million, respectively), followed by people from Spain, Germany, Syria, Lebanon, and Japan (Carvalho-Silva et al., 2001; Salzano and Bortolini, 2002). The distribution of these immigrants was unequal considering the several Brazilian states. For example, most Germans went to live in two states of the South region (Rio Grande do Sul and Santa Catarina), while Italians have chosen mainly São Paulo (in the Southeast) and Rio Grande do Sul.

Color is used in Brazil as an equivalent to race and is based on a complex and subjective phenotypic evaluation. In contrast to the situation in the United States, in Brazil the emphasis is on physical appearance rather than ancestry. Previous studies indicated, however, that color and other phenotype traits can be poor predictors of genomic ancestry (Parra et al., 2003). The Brazilian Institute of Geography and Statistics has adopted the criterion of classification of individuals according to the following categories: White (in Portuguese, Branco), Black (Preto), Brown (Pardo), Yellow (Amarelo), and Amerindian (Indígena). Accordingly, in Brazil as a whole, 53%, 6%, and 38% of the persons were identified as White, Black, and Brown, respectively, with the remaining 3% being distributed between the two other categories (IBGE Census 2000; http://www.ibge.gov.br).

Rio Grande do Sul is the southernmost state of Brazil. At the time the first Europeans arrived, the region was inhabited by Native Americans identified basically with three major groups: (1) Guarani (Tupian linguistic branch); (2) Kaingang (Jêan); and (3) Pampean tribes (Charrua, Minuano, Guenoas, etc.). The latter were extinct before the first decades of the 19th century.

The history of Rio Grande do Sul is peculiar because its effective colonization started in the 18th century only. Furthermore, in the colonial era, the control of the region alternated between the Spanish and Portuguese Empires (Flores, 1996). However, historical and genetic data showed that the asymmetrical matings (Portuguese/Spaniard males with Amerindian/African females) also characterized the initial demographic history of Rio Grande do Sul (Bortolini et al., 1999). Genetic studies using classical polymorphisms indicated that the Native American contribution was

about 11% in a large sample identified as White (Dornelles et al., 1999). On the other hand, historical sources suggested that the Amerindian contribution should have been basically Guarani (Kern et al., 1993), but no genetic study had been conducted to corroborate this assertion. The last Demographic Census showed that in Rio Grande do Sul (~10 million inhabitants) 87.5%, were classified as White, 5% as Black, 7% as Brown, 0.1% as Yellow, and 0.4% as Amerindian (IBGE Census 2000; http://www.ibge.gov.br).

The objective of the present study was to evaluate if there is significant heterogeneity between two samples classified as White in Rio Grande do Sul, and if this heterogeneity is reflected in both paternal (Y-chromosome) and maternal (mtDNA) lineages. It is hoped that this investigation may, in a more general way, unravel the intricacies and details that shape modern human populations and provide background material for patient/control association studies.

SUBJECTS AND METHODS

Population samples and DNA extraction

A total of 119 unrelated individuals identified as White were tested. Eighty-eight were living in Veranópolis, a city characterized by a striking influence of the Italian immigration of the 19th century; while 31 others came from casework analyses conducted in distinct places of Rio Grande do Sul (Fig. 1). Additional information about Veranópolis can be obtained in Peres et al. (2004). All donors were informed about the aims of this study and signed a written consent form. This investigation was approved by the Brazilian National Ethics Commission (CONEP number 1333/2002). The first sample will be referred as Veranópolis White and the other as General RS (from Rio Grande do Sul) White.

DNA extraction from whole blood was performed according to Lahiri and Nurnberger (1991).

Y-chromosome markers

Seven biallelic polymorphisms (M242, M3, M19, 92R7, M9, YAP, and M2) were typed in 75 men (Veranópolis White, $N = 51$; General RS White, $N = 24$) using methods described in Bortolini et al. (2003). Haplogroups defined by the mutations Q*(xQ3), Q3* (xQ3a), Q3a,
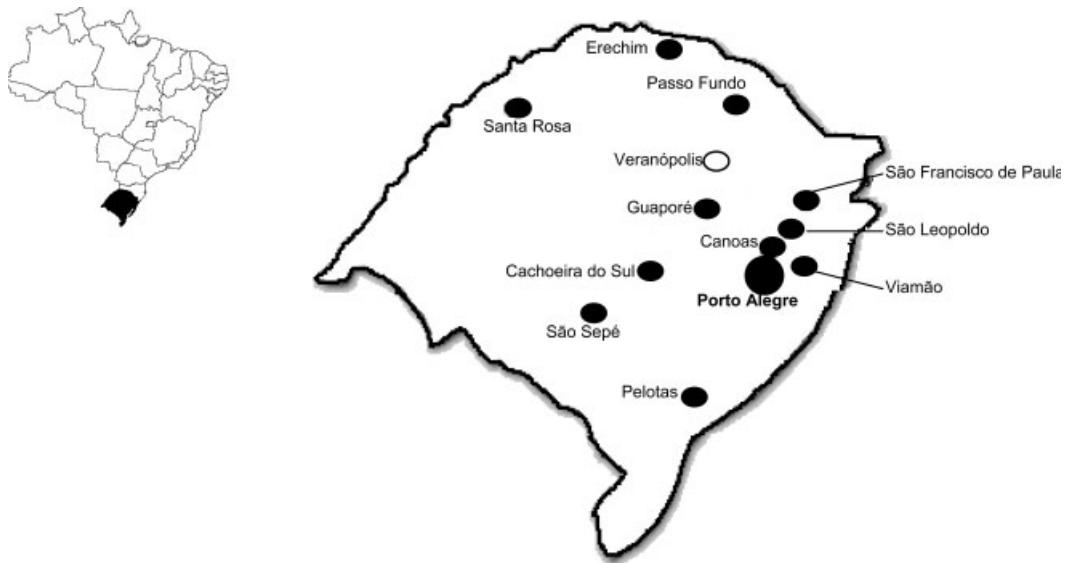
Fig. 1.   Geographic localization of the populations studied for HVS-I and Y-chromosome markers. Veranópolis and localities sampled for the General RS sample are represented respectively by open and filled circles. The shaded area of the map of Brazil (left) indicates the state Rio Grande do Sul.

P* (xQ), K* (xP), DE* (xE3a), and E3a* were designed following the nomenclature suggested by the last Y Chromosome Consortium release (Jobling and Tyler-Smith, 2003). As recommended, a designation such as Q*(xQ3) indicates the partial typing of markers in a haplogroup, in this case describing all chromosomes in clade Q except those in Q3. The M242 marker, which defined the Q clade in the present study, has been more recently identified in Amerindian and Asian populations (Bortolini et al., 2003; Seielstad et al., 2003). Haplogroup Y* denotes the presence of the ancestral alleles for the seven markers investigated here.

mtDNA amplification and sequencing

The nucleotide sequence of the first mtDNA hypervariable segment (HVS-I) was directly amplified by the polymerase chain reaction (PCR) for all sampled individuals ($N = 119$). For the Veranópolis White sample, mtDNA was amplified using the primers described by Horai et al. (1993). Thirty-five cycles of denaturation at 94°C for 1 min, annealing at 50°C for 1 min, and an extension at 72°C for 1 min were performed. For the samples of the other regions (General RS White), the primers and PCR conditions described by Alves-Silva et al. (2000) were used. In both cases, the sequence reactions were carried out using the BigDye^{TM} Terminator Cycle Sequencing Ready Reaction (AB Applied Biosystems, Foster City, CA) kit, and then the QIAquick spin columns (Qiagen, Valencia, CA) kit was used for sample purification. The run was performed in an automatic ABI Prism 3100 Genetic Analyzer sequencer. Sequence analysis was carried out with CONSED and BIOEDIT software. For all samples, both strands of DNA were sequenced.

Genetic analysis

To evaluate if systematic artifacts introduced in the course of the sequencing process could have produced "phantom" mutations compromising the accuracy of our data set, we applied the filtering process described by Bandelt et al. (2002). This analysis filters out all speedy transitions and thus scores weighty mutations only. Weight networks showing perfect star tree patterns are expected when the data are potentially free of phantom mutations. The number of weighty transitions to the number of transversions plus indels (WTTI ratio; Bandelt et al., 2002) can also be used to evaluate the

quality of the HVS-I data sets. In cases when phantom mutations abound, the WTTI ratio may strongly deviate from the corresponding ratio in comparable data sets and, thus, hint at potential anomalies.

Estimates of parental geographic contributions considering mtDNA data were obtained directly, since the major mtDNA haplogroups are geographic specific. For Y-SNP markers, however, these estimates were calculated using the weighted least-square method (Long, 1991) performed with the ADMIX program, kindly made available by Dr. J.C. Long.

RESULTS AND DISCUSSION

Y-haplogroup frequencies present in the two White samples are reported in Table 1. Haplogroup P*(xQ) was the most frequent (63% in Veranópolis and 96% in the General RS sample). These numbers are higher than that (42%) reported by Carvalho-Silva et al. (2001) for other sample of Brazilian whites from the South region (Rio Grande do Sul, Santa Catarina, and Paraná states). Table 1 also shows that this haplogroup is the most frequent in Europe (average = 55%). Because it is virtually absent in Africans and Native Americans, P*(xQ) can be considered a good marker of European ancestry. Haplogroups E3a* and Q*(xQ3)/Q3*(xQ3a)/Q3a, typical respectively of sub-Saharan Africans and Amerindians, were not observed in our samples. Haplogroup Y* (ancestral to all biallelic markers typed here) and DE*(xE3a) are less informative because they are not continental-specific.

The networks obtained for the HVS-I weighty variation are shown in Figure 2. For this analysis, two transitions (16069 C→T and 16224 T→C) were added to the speedy filter of Bandelt et al. (2002), since these mutations are common in European haplogroups J and K, respectively (Alves-Silva

et al., 2000; Maca-Meyer et al., 2003). A perfect star tree can be observed for the Veranópolis sequences. The network obtained for the General RS tree presented one reticulation that is basically due to the presence of a transversion (C→A) present in site 16183, which occurs in three different lineages associated to Amerindian haplogroup B. This mutation has already been described in the same (Ruiz-Linares et al., unpublished data) and in other backgrounds (Pereira et al., 2000; Salas et al., 2002), indicating that this may be a real phenomenon rather than an artifact introduced in the course of the sequencing process. The calculated WTTI ratio is 2.3 (Veranópolis White) and 1.8 (General RS White), which are intermediate in relation to those estimated for the European (4.8), Amerindian (3.8), and African (1.5) fractions of the Brazilian mtDNA sample studied by Alves-Silva et al. (2000). These results as a whole show that our HVS-I data sets are potentially free of phantom mutations.

Tables 2 and 3 list, respectively, the 43 and 29 lineages identified in the Veranópolis White ($N = 88$) and General RS White ($N = 31$) samples. All nucleotide changes were transitions except for five transversions at positions 16182 (A→C; General RS White), 16183 (A→C; General RS White), 16187 (C→A; Veranópolis White), 16188 (C→G; Veranópolis White), and 16318 (A→T; Veranópolis White and General RS White). The majority of the sequences (90% and 87% for the Veranópolis and General RS, respectively) could be identified with some major continental-specific mtDNA haplogroup. For Veranópolis, six typical European haplogroups (H, J, K, T, U, and V) were identified, while seven were observed in General RS White (H, J, K, T, U, pre-HV, and I). Haplogroup H is the most frequent in the Veranópolis sample (60% of sequences), as is

TABLE 1. Y-chromosome haplogroup frequencies (%) in men classified as White from the state of Rio Grande do Sul compared to those found in three major human geographical groups

Samples	Q*(xQ3)	Q3*(xQ3a)	Q3a	P*(xQ)	K*(xP)	DE*(xE3a)	E3a*	Y*
Veranópolis White ($N = 51$)				63		6		31
General RS White ($N = 24$)				96				4
Europeans ($N = 479$)[a]				55		14		31
Sub-Saharan-Africans ($N = 490$)[b]						12	80	8
South-Amerindians ($N = 390$)[c]	10	84	6					

[a]Carvajal-Carmona et al. (2003), Carvalho-Silva et al. (2001), and Bortolini et al. (2004a).
[b]Carvajal-Carmona et al. (2003) and Luis et al. (2004).
[c]Bortolini et al. (2003). In this last study, frequencies of 1–4% were found for haplogroups DE*(xE3a), P*(xQ), E3a*, and Y*. They were not considered here because they may represent recent admixture with non-Indians.
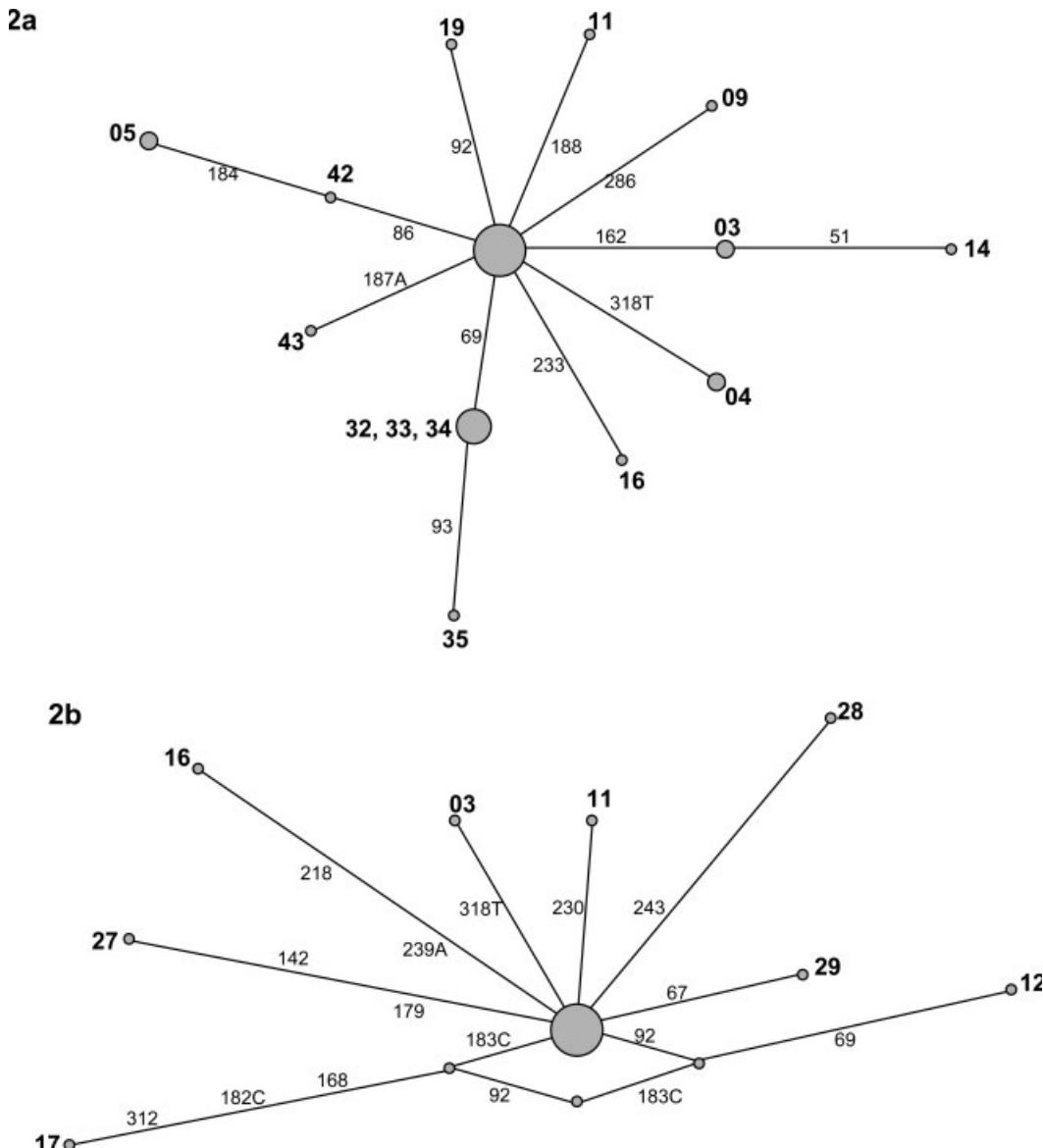
Fig. 2.   Median-joining networks: (**a**) Veranópolis White sample, and (**b**) General RS White sample. The numbers in bold represent the haplotypes listed in Tables 2 and 3. Weighty mutations are shown in the branches, where the numbers represent the nucleotide position (-16000). Transversions are indicated by letters after the numbers.

true in the European populations of four selected countries (Portugal, Spain, Italy, and Germany; Table 4). On the other hand, haplogroup H was identified in only 14% of the General RS sequences.

Table 4 shows also that the four major Amerindian haplogroups were observed in the General RS White sample (in percentages: A = 4, B = 11, C = 17, and D = 4), but only one of them (C, 2%) was detected in Veranópolis. Sub-Saharan African haplogroups L1*, L2*, and L3* are present in the General White sample with a frequency of 4% each.

TABLE 2. HVS-I sequence variation and probable major continental-specific mtDNA haplogroups observed in individuals classified as White in the Veranópolis sample

The table below lists, for 25 observed haplotypes, the variable nucleotide positions of HVS-I (shown relative to the reference sequence, with "." indicating identity to the reference and a letter indicating a substitution), together with the number of individuals (N) and the probable haplogroup assignment. The top row gives the reference-sequence base at each variable position.

Reference base →	A	C	T	T	T	T	G	T	G	A	A	T	C	C	C	C	T	C	G	T	C	A	T	C	C	C	C	C	C	C	C	T	T	T	A	G	T	C	N	Haplogroup
1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	H
2	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	H
3	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	H
4	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	2	H
5	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	H
6	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	2	H
7	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	1	H
9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	1	H
11	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
12	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
13	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
14	.	.	.	.	.	.	.	.	A	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
15	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	1	H
16	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.	.	.	.	.	.	1	H
17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.	.	.	.	A	.	.	.	1	H
18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.	.	.	.	A	A	.	.	4	V
19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	2	K
20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	C	.	.	.	.	.	.	.	.	1	K
21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	C	.	.	.	.	.	.	.	.	.	.	.	1	K
22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	1	K
23	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	U
24	.	.	.	.	.	.	.	G	.	.	.	.	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	1	U
25	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	1	U

26	1	U
27	1	U
28	1	U
29	2	T
30	2	T
31	3	J
32	1	J
33	1	J
34	1	J
35	1	J
36	2	C
37	1	Undetermined
38	1	Undetermined
39	5	Undetermined
40	1	Undetermined
41	1	Undetermined

Note: The nucleotide positions considered for the analysis were from 16051 and 16383 of the HVS-I region. Sequences were aligned with the revised reference sequence (Andrews et al., 1999) using BIOEDIT (Hall, 1999).

TABLE 3. HVS-I sequence variation and probable major continental-specific mtDNA haplogroups observed in individuals identified as White in the General RS

Variable Nucleotide Position

Position (16___)																																																					N	Haplogroup
(hundreds)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3		
(tens/units)	69	92	14	6 9	24 6	25 8	68 3	76 9	78 8	88 9	79 2	09 7	11 8	12 3	23 4	30 0	34 9	39 1	44 3	45 9	46 6	46 0	47 4	48 0	49 3	94 5	96 8	99 4	99 5	90 6	91 1	92 8	89 9	26 7	2																				
Reference	C	C	T	C	T	T	G	C	G	C	C	C	C	C	A	A	C	C	T	C	T	T	C	C	T	A	C	C	A	T	T	C	C	C	C	C	G	C	C	A	C	C	C	T	T	A	T	A	A	G	T	C	T	N	Haplogroup
1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	H
2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	1	H
4	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	T	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	1	Pre-HV
5	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	C	2	K
6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	K
7	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	?	1	U
8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	?	?	?	1	U
9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	U
10	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	I
11	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	.	T	.	T	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	T
12	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	G	.	.	.	.	T	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	J
13	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	L1b1
14	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	.	.	.	.	T	.	G	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	1	L3f
15	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	A	.	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	L2a1
16	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	C	.	.	T	T	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	C	.	.	1	A2
17	.	.	.	.	.	.	.	.	.	.	T	.	.	C	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	B
18	.	C	.	.	.	.	.	.	.	.	.	T	.	C	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	.	1	B
19	.	.	.	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	.	1	B
20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	C	T	.	1	C
21	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	C	.	.	.	.	.	.	.	C	T	.	1	C
22	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.	C	T	.	1	C
23	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.	C	T	.	1	C
24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	C	T	.	1	C
25	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	1	D
26	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	Undetermined
27	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	1	Undetermined
28	.	C	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	Undetermined
29	T	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	1	Undetermined

Note: The nucleotide positions considered for the analysis were from 16051 and 16383 of the HVS-I region. Sequences were aligned with the revised reference sequence (Andrews et al., 1999) using BIOEDIT (Hall, 1999).

TABLE 4. *mtDNA haplogroup frequencies (%) obtained in four European countries, compared with those found in the present study*

| Haplogroups | Country/population | | | | | |
| | Germany[a] | Italy[b] | Portugal[c] | Spain[d] | Whites | |
					Veranópolis[e]	General[e]
A						4
B						11
C					2	17
D						4
Pre-HV						4
H	50	43	41	52	60	14
I	2	3	1	4		4
J	9	16	7	10	10	4
K	7	5	6	7	6	11
T	9	17	10	10	7	4
U	14	5	16	4	10	11
V	5	7	7	6	5	
W	1	1	1	7		
X	2	2	2			
L1*			1			4
L2*			3			4
L3*			4			4
Others	1	<1	<1			
Total	527	117	228	52	79	27

[a]Helgason et al. (2001). In this study, frequencies <0.4% were found for haplogroups C and D. They were not considered because they probably represent recent Asian mtDNA introgression into the German sample.
[b]Montegale-Profizi et al. (2001) and Francalacci et al. (1996).
[c]Pereira et al. (2000).
[d]Côrte-Real et al. (1996).
[e]Only sequences that could be associated with some major geographic-specific mtDNA haplogroup (Tables 2 and 3) were considered.

Table 5 summarizes the frequencies (in percentage) of the major Y-chromosome and continental-specific mtDNA haplogroups in our two White samples. While for the Y-chromosome there is complete agreement between the morphological and genetic classifications, the same is not true for the mtDNA data. In both Veranópolis and General RS, the European contribution to the mtDNA gene is the most frequent. However, for the General RS sample the values associated to African (16%) and Native American (36%) ancestries are expressive, these numbers being, however, somewhat different from those described earlier for other White sample of the South region (including the states of Rio Grande do Sul, Paraná, and Santa Catarina) namely: 66% European, 12% African, and 22% Amerindian (Alves-Silva et al., 2000).

The Native American contribution observed here, particularly in the General RS sample, suggests a marked presence of mtDNA lineages identified with haplogroups C and B. Curiously, the presence of these haplogroups is low, considering the large sample of the three Guarani subgroups (Mbyá, Kaiowá, Ñandeva) studied by us (unpublished data), in percentages: C = 0, 3, and 12; B = 3, 4, and 3, respectively. As mentioned in the introduction, according to historical data only the Guarani would have contributed in a significant way to the formation of the Rio Grande do Sul population,

TABLE 5. *Parental contribution in populations identified as White in the state of Rio Grande do Sul based on Y-chromosome and mtDNA data sets*

| Population | Parental contribution (%) | | |
	European	African	Native American
Veranópolis White			
mtDNA[a]	97	0	3
Y-chromosome[b]	100	0	0
General RS White			
mtDNA[a]	48	16	36
Y-chromosome[b]	100	0	0

[a]Values obtained directly from the distributions of the major geographic-specific mtDNA haplogroups listed in Table 4 (A + B + C + D = Amerindian; H + I + J + K + T + U + V + W + X + Z = European; L1 + L2 + L3 = sub-Saharan African).
[b]Because there are some Y-haplogroups (DE* and Y*) that are not geographic-specific, the estimates of the parental contributions were obtained using the frequencies presented in Table 1 and Long's (1991) least-square method.

since they were, during all the colonial era, in contact with non-Indians (Kern et al., 1993). The present results delineate two possible scenarios: (1) Due to stochastic factors like genetic drift (Salzano and Callegari-Jacques, 1988) and a dramatic depopulation after contact with non-Indians, the contemporary Guarani gene pool can be a poor representative of that found at colonial times; or (2) Other tribes (Jêan and/or Pampean) could have made a more significant contribution, through their women, to the formation of the general population of the state.

Typical African haplogroups L1b1, L2a1, and L3f detected here in the General RS sample (Table 3) have been observed in other Brazilian populations (Alves-Silva et al., 2000; Bortolini et al., 1997). The first is a good West African (non-Bantu) marker, while the others can be observed in both Bantu and non-Bantu populations (Salas et al., 2002, 2004). African haplogroups L1c and L3e present in high frequencies in the African portion of the Brazilian mtDNA samples evaluated earlier (Alves-Silva et al., 2000; Bortolini et al., 1997, 2004b), were not observed in this study.

An additional feature present in Table 5 deserves mention. Although the number of individuals tested for the General RS sample is low, our results indicated the classical asymmetrical pattern of mating observed in other Brazilian populations identified as White (Alves-Silva et al., 2000; Carvalho-Silva et al., 2001), where the Amerindian and African gene introgression occurred exclusively through women. This indicates that some contemporary White Brazilian populations can represent an extraordinary reservoir of Amerindian and African mtDNA genomes. On the other hand, other populations, like Veranópolis, are in fact basically European (they show almost only transplanted genomes, in the nomenclature of Bortolini et al. (2004a)). They are, therefore, good candidates for association studies. Careful selection of these populations through genome markers can neutralize the frequent criticism to such studies conducted in Brazilian populations, which are assumed beforehand as mixed.

ACKNOWLEDGMENTS

LITERATURE CITED

Alves-Silva J, Santos MS, Guimarães PEM, Ferreira ACS, Bandelt H-J, Pena SDJ, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revisions of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 123:147.

Bandelt H-S, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71:1150–1160.

Bortolini MC, Weimer TA, Zago MA, Salzano FM, Silva WA Jr, Silva MCBO, Bonato SL. 1997. Evolutionary and anthropological implications of mitochondrial DNA variation in Afro-Brazilian populations. Hum Biol 69:141–159.

Bortolini MC, Silva WA Jr, Castro-de-Guerra D, Remonatto G, Mirandola R, Hutz MH, Weimer TA, Zago MA, Salzano FM. 1999. African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi- and uniparental genetic markers. Am J Hum Biol 11:551–563.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanem SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot M, Michalsky R, Nymodawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73:524–539.

Bortolini MC, Thomas MG, Chikhi L, Aguilar JA, Castro-De-Guerra D, Salzano FM, Ruiz-Linares A. 2004a. Ribeiro's typology, genomes, and Spanish colonialism, as viewed from Gran Canaria and Colombia. Genet Mol Biol 27:1–8.

Bortolini MC, Silva WA Jr, Zago MA, Elion J, Krisnamoorthy R, Gonçalves, Pena SDJ. 2004b. The phylogeography of mitochondrial DNA haplogroup L3g in Africa and the Atlantic slave trade. Am J Hum Genet 75:523–524.

Callegari-Jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Crossetti SG, Fereira ME, Hutz MH. 2003. Historical genetics: spatiotemporal analysis of the formation of the Brazilian populations. Am J Hum Biol 15:824–834.

Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimer N, Ruiz-Linares A. 2003. Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. Hum Genet 112:534–541.

Carvalho-Silva DR, Santos FR, Rocha J, Pena SD. 2001. The phylogeography of Brazilian Y-chromosome lineages. Am J Hum Genet 68:281–286.

Côrte-Real HBSM, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC. 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60:331–350.

Dornelles CL, Callegari-Jacques SM, Robinson WM, Weimer TA, Franco MHLP, Hickmann AC, Geiger CJ, Salzano FM. 1999. Genetics, surnames, grandparents' nationalities and ethnic admixture in southern

A.R. MARRERO ET AL.

Brazil: do the patterns of variation coincide? Genet Mol Biol 22:151–161.

Flores M. 1996. História do Rio Grande do Sul, 5th ed. Porto Alegre, Brazil: Nova Dimensão, 215 p.

Francalacci P, Bertranpetit J, Calafell F, Underhill P. 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.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98.

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, Sykes B. 2001. mtDNA and the Islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. Am J Hum Genet 68:723–737.

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.

Jobling M, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612.

Kern AA, Bern Z, Cohen VRA, Avancini EG, Kemel CLA, Gick PW. 1993. Rio Grande do Sul: continente múltiplo. Coleção O Continente de São Pedro, Vol V. Porto Alegre, Brazil: Riocell Marprom. 118 p.

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

Long JC. 1991. The genetic structure of admixed populations. Genetics 127:417–428.

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ. 2004. The Levant versus the horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet 74:532–544.

Maca-Meyer N, Sánchez-Velasco P, Flores C, Larruga J-M, González A-M, Oterino A, Leyva-Cobián F. 2003. Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). Ann Hum Genet 67:329–339.

Montegale-Profizi N, Chollet L, Stévanovitch A, Dubut V, Poggi C, Pradié MP, Spadoni JL, Gilles A, Béraud-Colomb. 2001. Mitochondrial DNA sequence diversity in two groups of Italian Veneto speakers from Veneto. Ann Hum Genet 65:153–166.

Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SDJ. 2003. Color and genomic ancestry in Brazilians. Proc Natl Acad Sci USA 100:177–182.

Pereira L, Prata MJ, Amorim A. 2000. Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. Ann Hum Genet 64: 491–506.

Peres A, Hünemeier T, da Cruz IM, Nardi NB, Chies JAB. 2004. T-cell receptor BV gene segment polymorphisms in healthy elderly individuals from the south of Brazil. Eur J Immunogenet 31:77–81.

Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111.

Salas A, Richards M, Larea MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465.

Salzano FM, Callegari-Jacques SM. 1988. South American Indians: a case study in evolution. Oxford, England: Oxford University Press. 260 p.

Salzano FM, Bortolini MC. 2002. Evolution and genetics of Latin American populations. Cambridge, England: Cambridge University Press. 512 p.

Seielstad M, Yuldasheva N, Singh N, Underhill P, Defner P, Shen P, Wells RS. 2003. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. Am J Hum Genet 73:700–705.

CAPÍTULO III

The Demographic and Evolutionary Trajectories

of the Guarani and Kaingang Natives of Brazil

Marrero *et al*. Am J Phys Anthropol, no prelo.

The Demographic and Evolutionary Trajectories of the Guarani and Kaingang Natives of Brazil

Andrea R. Marrero,[1] Wilson A. Silva-Junior,[2] Cláudio M. Bravi,[3] Mara H. Hutz,[1] Maria L. Petzl-Erler,[4] Andres Ruiz-Linares,[5] Francisco M. Salzano,[1] and Maria Cátira Bortolini, [1]

[1]*Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, RS, Brazil.*
[2]*Departamento de Genética e Centro de Terapia Celular, Universidade de São Paulo, 14049-900 Ribeirão Preto, SP, Brazil*
[3]*Laboratorio de Genética Molecular Poblacional, Instituto Multidisciplinario de Biología Celular (IMBICE), La Plata, Argentina.*
[4]*Departamento de Genética, Universidade Federal do Paraná, 81531-990 Curitiba, PR, Brazil*
[5]*The Galton Laboratory, University College, London, United Kingdom*

Running head: MOLECULAR GENETIC VARIATION IN AMERINDIANS

KEY WORDS: mtDNA; Y-chromosome markers; Amerindians; asymmetrical interethnic matings

_____

Correspondence to: Maria Cátira Bortolini, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: maria.bortolini@ufrgs.br.

ABSTRACT A total of 278 individuals from two Brazilian Indian tribes (Guarani and Kaingang) living in five different localities had their mitochondrial DNA sequenced for the first hypervariable segment (HVS-I), and a fraction of them was also studied for seven biallelic Y-chromosome polymorphisms. Nineteen HVS-I lineages were detected, which showed distinct distributions in the two tribes. The G_{ST} value obtained with the mtDNA data is about 5 times higher for the Guarani as compared to the Kaingang, suggesting a higher level of differentiation between the three Guarani partialities than between the two Kaingang villages. Non-Amerindian admixture varied with sex and in the Guarani was only observed through the paternal line. Using these data and those of other Tupian and Jêan tribes it was possible to make inferences about past migratory movements and the genetic differentiation of these populations.

Genetic studies have been used as powerful tools to characterize Native American populations. Schurr and Sherry (2004) showed that the mitochondrial DNA (mtDNA) and the non-recombining portion of the Y-chromosome (NRY) are at present the two genetic systems most commonly used in studies with these population groups. Investigations using mtDNA in Amerindians revealed the presence of five different haplogroups, designated A-D (Schurr et al., 1990; Torroni et al., 1992, 1993) and X (Brown et al., 1998), and the highest level of differentiation between populations considering the human major geographical groups (Bortolini and Salzano, 1996; Bortolini et al., 1997). These and other studies have also shown distinct haplogroup distributions in South America: Haplogroup A generally occurs at higher frequencies in northern regions, while haplogroups C and D are frequent in several parts of South America. Haplogroup B is only abundant in southern Peru, Andean Bolivia, northern Chile and Argentina. Haplogroup X is not found in South America (Dorneles et al., 2005).

Initial analyses with NRY markers, on the other hand, found just one haplotype at high frequencies in native populations in North and South America of all linguistic groups (Pena et al., 1995). This most common Y-chromosome was afterwards characterized by a C →T mutation at marker M3 (Underhill et al., 1996), which defines haplogroup Q3* (The Y Chromosome Consortium, 2002; Jobling and Tyler-Smith, 2003). More recently, other Asian or Native American autochthonous haplogroups have been identified (C*, Q*, Q3a), but with different distributions among populations. For example, C* (which is present in high frequencies in Asia) is only found in North, and Q3a in South America (Bortolini et al., 2003). Using microsatellite loci Tarazona-Santos et al. (2001) showed that the Andean Native populations exhibit significantly higher within-population variability than the eastern groups (Amazonian region,

4

Brazilian plateau, and the Chaco region). These authors proposed a model for the

evolution of the South Amerindian male lineages that involved differential patterns of

genetic drift and gene flow.

The origin of the Tupian linguistic family is controversial (Noelli, 1998;

Rodrigues, 2000). However, most of the authors report regions at the southern margin of

the Amazon River (Rodrigues, 1964; Migliazza, 1982; Urban, 1996, 1998;

Heckenberger et al., 1998). For example, Migliazza (1982) suggested that the probable

place of origin of the Tupian linguistic family was situated between the Jiparaná and

Aripuanã rivers, and that the postulated parental group was living there about 5,000

years before present (ybp). The diversification of this major Amerindian linguistic

family occurred, due to community isolation, concomitantly with the extraordinary and

successful dispersion of the agriculturalist Tupian speakers (Fig. 1). The Guarani speak

a language classified in the Tupi-Guarani branch (Campbell, 1997). Their split from the

other Tupi probably occurred around 1,800 ybp (Fig.1; Carneiro da Cunha, 1998). In

colonial times, the Guarani who lived in the high Paraná and Uruguay River basins were

attracted to Jesuit missions, where they remained for almost two centuries, while other

groups stayed isolated, hidden in the forests. Today, in Brazil, they generally live in

reservations and can be subdivided in three partialities, in agreement with several

aspects of their culture: Guarani Ñandeva, Guarani Kaiowá and Guarani M`byá (Vietta,

1992). Because they have been in contact with non-Indians since colonial times (Kern,

1997), they should also have contributed to the formation of the South Brazilian

admixed populations in a significant way (Marrero et al., 2005).

The term Kaingang (or Caingang) was introduced in 1882 to designate all non-

Guarani indigenous people living in South Brazil (Becker and Laroque, 1999) but at the

time of the first contact with Europeans they were known as Guaianás (16th and 17th

centuries) or Coroados (19th century). The Kaingang are Jêan speakers (Southern branch; Campbell, 1997). The distribution of the Jêan languages in Brazil suggest an origin for this linguistic family at about 3,000 ybp, between the São Francisco and Tocantins rivers (Urban, 1998; Fig. 1). The split towards the meridional region should have occurred about 3,000-2,000 ybp, whereas that in direction to the Amazonian region was more recent (2,000-1,000 ybp; Urban, 1998).

The Kaingang have been recognized as descendents of the native inhabitants of the Brazilian Central-South plateau who lived in rustic subterraneous houses (Schmitz and Becker, 1997). Their number was drastically diminished after contact with the European colonizers, but those who survived and their descendents live now in reservations in the Brazilian states of Rio Grande do Sul, Santa Catarina, Paraná, and São Paulo. Their contact with non-Indians during the colonization process was less marked than that which occurred with the Guarani, but presently the situation changed, both showing variable local interaction with the surrounding populations.

Although Guarani and Kaingang have lived next to each other since the 17th century, they are culturally distinct (Carneiro da Cunha, 1998). Genetic differences have also been reported, with blood group and protein polymorphisms (Salzano et al., 1997; Callegari-Jacques and Salzano, 1999) and different DNA data sets: Y-SNP/STR (Bortolini et al., 2003), AAB-auto-antibody (Utiyama et al., 2000), *Alu* insertions (Battilana et al., 2002), nuclear STRs (Kohlrausch et al., 2005), HLA and other MHC (major histocompatibility complex) loci (Petzl-Erler et al., 1993; Sotomaior et al., 1998; Faucz et al., 2000; Tsuneto et al., 2003), CYP-cytochrome P-450, GST-glutathione S-transferase, and the TP53 tumor-suppressor gene (Gaspar et al., 2002), and TCR-T-cell receptor and CCR5-chemokine receptor genes (Hünemeier et al., 2005).

6

The present work furnishes data related to the variation of mtDNA first hypervariable segment (HVS-I) and of markers located in the nonrecombining region of the Y-chromosome in Guarani and Kaingang, which represent the southern extremes of the population distribution of members of the Tupian and Jêan linguistic families in Brazil. Questions asked were: (a) what genetic differences can be found among them, and how are they distributed among local groups? Are their levels of diversity similar or distinct? (b) how do they correlate with independent evaluations of their history? and (c) what insights concerning the interethnic exchange which occurred along this historical process can be obtained using parentally diverse genetic markers?

SUBJECTS AND METHODS

Populations

Samples of 200 Guarani and 78 Kaingang living in reservations (Rio das Cobras, Amambaí, Limão Verde, Porto Lindo, Ivaí, Nonoai) located in central and southern states of Brazil (Mato Grosso do Sul, Paraná and Rio Grande do Sul; Fig. 2) were obtained. More details about these populations can be found in Petzl-Erler et al. (1993), Bortolini et al. (2002, 2003), Tsuneto et al. (2003) and Kohlrausch et al. (2005).

Y-chromosome markers

Thirty Guarani M`byá and 36 Kaingang from Paraná were studied for seven biallelic polymorphisms (M242, M3, M19, 92R7, M9, YAP and M2), located in the nonrecombining region of the Y-chromosome, using methods described in Bortolini et

al. (2003). The nomenclature adopted is that proposed by the last Y-chromosome

Consortium release (Jobling and Tyler-Smith, 2003). These data were then analyzed

together with those obtained by Bortolini et al. (2003) for the Guarani Ñandeva,

Guarani Kaiowá and Kaingang from Rio Grande do Sul.

mtDNA

The nucleotide sequence of the first hypervariable segment (HVS-I) of 200

Guarani (120 Guarani Kaiowá, 56 Guarani Ñandeva and 24 Guarani M`byá) and of 78

Kaingang (57 Kaingang-Rio Grande do Sul and 21 Kaingang-Paraná) was amplified

and sequenced according to conditions described in Marrero et al. (2005). Both strands

of DNA were sequenced. When low quality sequences were obtained, multiple re-

sequencing efforts were done using the same primers.

The sequences were checked manually, validated with the help of the

CHROMAS LITE 2.0 program (www.technelsyum.com.au) and aligned with the

revised Reference Sequence (rCRS, Andrews et al., 1999) using the BIOEDIT software

(Hall, 1999). Since artifacts ("phantom mutations") can be introduced during the

sequencing and editing process, we applied the filtering procedure described by Bandelt

et al. (2002) and used criteria like those of Yao et al. (2004) to check for the quality of

the sequences. After filtering a network of sequences was constructed with the

NETWORK 4.1.1.2. program (www.fluxus-engineering.com) using the median-joining

algorithm. To validate the haplogroup B result, another specific amplification to

confirm the presence of the 9-bp COII/tRNA^{Lys} deletion was performed, using primers

and conditions as described in Green et al. (2000).

8

Data analysis

Total gene diversity (H_T) and its proportion attributable to differences between populations (G_{ST}) were performed using Nei´s statistics, which can be used for any genetic system including those which are haploid like mtDNA and the non-recombining portion of the Y chromosome (Nei, 1987). Nucleotide diversities considering the mtDNA sequences for each population were also estimated using Nei´s method (Nei, 1987). DISPAN (Ota, 1993) and ARLEQUIN (Schneider et al., 2000) packages were used to obtain the results. The latter was also used to evaluate the distribution of the inter and intrapopulational genetic variations by means of an analysis of molecular variance (AMOVA, Excoffier et al., 1992).

Due to the widespread but spotted distribution of lineages carrying the 16266 C→T mutation in the continent, a medium network of these mtDNA sequences was constructed with the NETWORK 4.1.1.2. program (www.fluxus-engineering.com) using the median-joining algorithm.

Estimates of parental continental contributions for the mtDNA data were obtained directly, since the major mtDNA haplogroups are geographic specific. For the Y-SNP markers, however, these estimates were calculated using the weighted least square method (Long, 1991) performed with the ADMIX program, kindly made available by Dr. J.C. Long.

RESULTS

Y-chromosome haplogroups

The combination of the seven biallelic markers allows the identification of eight Y-SNP-haplogroups (Fig. 3), and five of them were observed in our data. Some of these markers are characteristic of European, African, or Native American populations, and thus can be informative for identifying recent admixture with non-native peoples: for example, Q3*(xQ3a) and Q3a are identified as of Amerindian origin; E3a* as of sub-Saharan African origin; P*(xQ) as of European origin. Others, however, using this level of resolution, are less informative since they do not have an identified continental-specific origin: Y* (Africa, Europe), DE*(xE3a) (Asia, Africa).

Haplogroup Q3*(xQ3a) was the most frequent in the Guarani M`byá (61%) and Kaingang-Paraná (50%). This is the most common haplogroup observed in Native American populations (Bortolini et al., 2003), and it was earlier reported with high frequencies among the Guarani Ñandeva, Guarani Kaiowá and Kaingang-Rio Grande do Sul (70%, 86%, and 86%, respectively; Bortolini et al., 2003). The other Native American/Asian haplogroup, Q*(xQ3), was detected in some (Guarani M`byá, 3%; Guarani Ñandeva, 15%; Kaingang-Paraná, 8%), but not all Guarani and Kaingang partialities/villages. The Amerindian haplogroup Q3a was not observed, corroborating the suggestion that this lineage may be restricted to northwest South America (Bortolini et al., 2003).

The presence of non-Amerindian Y-chromosomes, as indicated by the Y*, P*(xQ), and E3a* haplogroups, is important in these tribes. It ranges from 14% (Kaingang-RS and Guarani Kaiowá) to 42% (Kaingang-Paraná). Among these haplogroups, the most prevalent is Y* (Guarani M`byá, 27%; Kaingang-Paraná, 31%),

10

but it was not observed in other Guarani partialities or in the Kaingang-Rio Grande do

Sul. This result can indicate male-mediated admixture between the Guarani and

Kaingang populations of Rio das Cobras reservation, State of Paraná (Fig. 2), and/or

admixture with Euro and Afro-descendants neighbors. Haplogroup P*(xQ) has a

probable European origin, while E3a* is typical of Sub-Saharan Africans. The former

was detected in all Guarani and Kaingang samples, but the latter was only present (3%-

6%) in the three Guarani partialities

Mitochondrial DNA

The mtDNA sequence variation observed in the 278 individuals examined is

summarized in Table 1. Nineteen lineages were observed, and all nucleotides changes,

except two, were transitions involving more pyrimidines than purines, with the C$\rightarrow$T

substitution being the most frequent mutation. Transversions were identified at positions

16239 (C$\rightarrow$A) and 16114 (C$\rightarrow$A), detected in the Guarani Ñandeva lineage 6, and

Kaingang-RS lineage 18, respectively.

All sequences could be identified with some continental-specific mtDNA

haplogroup. The 200 Guarani carried haplotypes belonging, according to Bandelt et al.,

(2003), to three of the four major Native American haplogroups, as follows: 84% A,

9.5% C, and 6.5% D. No lineage presented the characteristic mutations of haplogroup

B. Lineage 1, the nodal sequence for Native American haplogroup A, is the most

common in both Guarani and Kaingang. Lineage 1 and lineage 2 are the only ones

shared across the three Guarani partialities. Lineage 2 diverges from the nodal A by the

gain of 16291, a change that has already been reported for this haplogroup in one

Tayacaja Quechua (Fuselli et al., 2004). Interestingly, a single case of lineage 2 is

reported here for one Kaingang-Paraná, who share the same reservation with the Rio das

Cobras Guarani M'byá . Derived lineage 3 (A) is shared by the Kaiowá and Ñandeva, indicating the higher identity between these two subdivisions as compared to the M`byá.

The most frequent Amerindian haplogroups in the whole Kaingang sample were C (46%) and A (42%). Two individuals assignable to haplogroup A (lineage 8) carried the haplotype 16126C-16223T-16278T-16290T-16319A-16362-C previously described in one Krahó (Torroni et al., 1993) and in two non-Native Brazilians (Alves-Silva et al., 2000). More information from the coding region is needed in order to establish if this lineage belongs to an A2 subtype and has reverted the 16111 mutation, or if it represents some further founder haplotype for haplogroup A in the Americas.

Three individuals (5%) showed the 16189C-16217C combination and the 9bp COII/tRNA^{Lys} deletion, which jointly characterizes haplogroup B (Kivisild et al., 2002). Haplogroup D was not detected. Only three lineages are shared between the Kaingang-Rio Grande do Sul and Kaingang-Paraná: lineage 1 (A), 12 (C), and 14 (C). Non-Amerindian admixture in this tribe was detected through the presence of three persons with the sub-Saharan African haplogroup L2b1 and by one subject with the rCRS HVS-I motif, the single most frequent lineage in West Eurasian populations. More than 95% of the lineages carrying the rCRS HVS-I can be assigned to haplogroup H, but in the absence of further molecular information membership in other West Eurasian haplogroups (e.g. HV, U, R, etc) cannot be discarded.

Table 2 shows the Amerindian mtDNA haplogroup frequencies in the Guarani and Kaingang, as well as in other Tupian and Jêan tribes. The frequency of haplogroup A in the Guarani is very high, particularly in the Kaiowá and Ñandeva partialities, differently of what is observed in the other Tupian tribes (with the exception of the Wayampi). The absence of haplogroup B in the Guarani contrasts with its presence in

12

seven of the 10 other Tupian tribes considered. This result could indicate that the

Tupian migration from the Amazonian region to the South may have resulted in the loss

of this mtDNA haplogroup. However, more studies with other Brazilian and non-

Brazilian Guarani groups are needed to confirm this suggestion. Major differences are

also observed between the Kaingang and other Jêan-speaking tribes. Although in this

case the sample sizes are more limited, there is a clear inversion in the totals of the

Kaingang and other groups for the frequencies of two haplogroups: Kaingang 4% B,

49% C; others 67% B, 1% C.

Table 2 also furnishes two diversity parameters, one generated by the mtDNA

sequence data sets and the other by the single-site nucleotide frequencies obtained

according to Nei (1987). Since in the second case the analysis considered only the

variable sites, it could give an insight not furnished by the nucleotide diversity statistics,

that includes both variable and non-variable sites. Evaluation of the intra and

interpopulation variabilities was also done with AMOVA, but this procedures did not

yield sufficiently discriminative power in this set of data (data not shown). The

nucleotide diversity (0.0067) and gene diversity (0.0495) calculated for the Guarani are

lower than those obtained for seven of the nine other Tupian tribes. It is possible,

therefore, that in their southern route the Tupi-Guarani would have lost part of their

intrapopulational variation.

A low π value (0.0171) can also be observed when the Kaingang tribe is

compared with the mean calculated for the other Jêan (0.0379). An inverse situation is

observed when the gene diversity is calculated (respectively 0.1008 and 0.0938). But

the low sample sizes in the other Jêan populations make comparisons risky and could be

responsible for this discrepancy.

The total gene diversity estimated for the Guarani is 0.0495, and 19% of it can be attributable to differences between the three Guarani partialities, whereas for the other Tupian groups there is high intertribal differences (53%). A similar G_{ST} value (58%) was obtained for the other Jêan groups, but the differences between the two Kaingang villages represent only 4% of the total variability found in this tribe. Their total gene diversity (0.1008), however, is two times higher than the Guarani value.

Admixture analyses

Guarani and Kaingang are in an advanced stage of acculturation and previous studies using blood group and protein polymorphisms demonstrated that both tribes have some degree of admixture with non-Amerindians. These investigations also revealed that the Kaingang had a higher proportion of non-Amerindian alleles when compared to the Guarani (Salzano et al., 1997; Callegari-Jacques and Salzano, 1999). However, the specific nature of this gene flow was not known. Table 3 presents the parental contribution estimates using the present mtDNA and Y-chromosome data sets and those values published earlier considering biparental markers (Salzano et al., 1997; Callegari-Jacques and Salzano, 1999). The results revealed that 5% of the mtDNA sequences observed in the Kaingang have an African or European origin, while none was detected in the three Guarani partialities. On the other hand, Y-chromosomes of non-Amerindian origin were detected in both populations, although the typical sub-Saharan Y-chromosome haplogroup was only observed in the Guarani. The autosome estimates, as expected, presented intermediate values between the mtDNA and Y-chromosome numbers, further validating them. These results reveal that despite some specific details, the admixture present in both is influenced by gender.

DISCUSSION

Mitochondrial DNA lineages 4 and 6 observed respectively in eight and fourteen Guarani Ñandeva individuals, are connected to a series of others, spotted but widely distributed in South America, defined by the presence of a C → T transition at position 16266. As shown in the legend of Figure 4, the nodal sequence (lineage 4 in Table 1: 16111T-16223T-16266T- 16290T-16319A-16362C) was previously described in Amerindians from lowland Bolivia, one non-Native SE Brazilian, and in one non-Native Uruguayan. We report here its presence in eigth Guarani Ñandeva. One- and two-step derivatives were observed in one Gavião (Brazil; #2) and in one Quechua (Peru; #3), the former geographically located in the probable center of spread of the Tupian languages (Figure 1). Of special interest is a further branch defined by the 16239A transversion, present in Amazonian and non-native S and SE Brazilian populations, as well as in the Guarani Ñandeva. While the non-Native Brazilian lineages (#6, #7) have a transition at position 16218, the Guarani Ñandeva lineage (#8), present in 14 individuals, carries instead the private transition 16153. The Ñandeva Guarani lineages (#1 and #8) are connected to lineages from Amazonia (#2, #4) and Peru (#3), as well others (#6, #7) present in south and southeast Brasil. These relationships conform to their history of dispersion (Amazonia → S/SE Brazil).

The low intrapopulational variability observed in the Guarani suggests that they may have experienced a bottleneck in their southern migration from Amazonia. This event may have been moderate or severe, since their later population growth (the Guarani had an enormous success in this dispersion because they dominated agricultural techniques, and have been also associated with the Jesuitic missions that lasted for a

long period of time) was not enough to restore the postulated level of the pre-migration variability.

As a whole our results reveal that the Kaingang and Guarani show some marked differences. When the mtDNA data are considered, the differentiation between the three Guarani partialities is much higher ($G_{ST}= 19\%$) than that observed between the two Kaingang villages (4%). On the other hand, based on (a) the proportion of mtDNA intertribal differentiation obtained for the Tupian and Jêan groups (53% and 58%, respectively; Table 2); (b) the time of origin of these linguistic families (~5,000 and ~3,000 ybp; Schmitz, 1997; Carneiro da Cunha, 1998; Urban, 1998); and (c) the Guarani and Kaingang mtDNA G_{ST} values (respectively 19% and 4%; Table 2) it is possible to estimate through a simple proportion that the three major Guarani partialities present in Brazil (Ñandeva, Kaiowá, M`byá) have been separated during at least ~1,800 ybp (0.19 x 5,000/ 0.53 $\cong$ 1,800), while the two Kaingang populations would have split at just ~207 ybp (0.04 x 3,000/ 0.58 $\cong$ 207). Of course, these numbers should be considered with caution, since G_{ST} is a simple coefficient of interpopulation differences, with numerous assumptions about the nature of this variability, and we are estimating it just from the maternal side. But they can indicate that the separation of the Guarani groups was an ancient event, previous to contact with European colonizers and African slaves, whereas the separation of the two Kaingang populations was a more recent event.

Salzano and Callegari-Jacques (1988) using blood group and protein polymorphisms analyzed the correlation between genetic distances and linguistic affinities. They found that the average within linguistic stock genetic distances were always lower than those between stocks, except for the Tupian. The Tupian finding was explained by the migratory behavior of this group. Individuals from it covered large

16

geographical distances in their dispersion, and they had contacts with several autochthone peoples, favoring gene flow between populations with distinct gene pools, which could be responsible for the large genetic distances found within this linguistic family. The authors did not discard natural selection as another factor that would have contributed for this differentiation. The high G_{ST} value (19%) obtained with our mtDNA data considering the three Guarani partialities is in the same direction of these earlier studies, but an additional possibility is that at least a part of this genetic diversity may be due to the relatively large time of divergence between the three Brazilian Guarani subgroups. One evidence in favor of this view is the existence of a private polymorphism in this tribe, present in mtDNA lineage 6 of Table 1 (shown as #8 in the network of Figure 4) which evolved only in the Guarani Ñandeva. Other possible inference from these results is that the introduction of typical sub-Saharan Y-chromosomes (E3a*; frequencies ranging from 3% to 6%) and of other non-Amerindian Y-chromosomes likewise probably occurred independently in the three Guarani subgroups.

On the other hand, as already mentioned, the separation between the two Kaingang subgroups is probably associated with more recent occurrences. Several historical registers have described an intense migratory movement of the Kaingang along the south and southeast regions of Brazil due to contact, and their consequences, with non-Amerindian colonizers (Schmitz and Becker, 1997).

The results also revealed that non-native admixture within the Guarani communities was largely restricted to males of European and African descent, while non-native admixture within the Kaingang was more variable. In the Kaingang-Rio Grande do Sul, European and African mtDNA lineages were observed, whereas this was not seen in the Kaingang who live in Paraná; while non-Amerindian Y-chromosomes

were detected in Rio das Cobras (Paraná) only. This emphasizes the distinct cultural factors influencing the mating behavior of these two tribes.

Asymmetrical sex-mediated admixture was common during the first centuries of Brazil's colonization, and it involved mostly European men and Amerindian/African women (Bortolini et al., 1999; Alves-Silva et al., 2000; Carvalho-Silva et al., 2001; Salzano and Bortolini, 2002). The main consequence of this historical contact was the formation of a people characterized by a composite genome, since their Y-chromosomes have been mainly transplanted from Europe, while their mtDNA would derive predominantly from Amerindian and African sources. Their autosome sets, on the other hand, would have been considerably shuffled (Bortolini et al., 2004; Marrero et al., 2005). This could also explain the introduction of some non-Amerindian Y-chromosomes in the tribes through interethnic matings. In this situation the children normally stay with their mothers. Another possibility, at least considering the African contribution, would be associated to the absorption of escaped slaves (mostly men) by the tribe. More recently, two other factors may have served to increase the amount of asymmetrical gene flow between the tribal societies and the surrounding society: a) prostitution, involving Amerindian women and men who live near the border of the reservations (http://revistaeducacao.uol.com.br/; on line edition, number 96); and b) while Amerindians who live in reservations have free access to land for cultivation, this is not true for non-Native Brazilians; the latter, therefore, may marry Indian women and establish themselves in the reservations, in some cases even claiming a certain degree of Indian ancestry, to guarantee land rights (Callegari-Jacques and Salzano, 1999).

Finally, the presence of European and African mtDNA genomes among the Kaingang-Rio Grande do Sul deserves additional consideration, since it suggests the absorption of non-Indian women by a tribal community. Salzano (1961) made an

extensive demographic study in nine Kaingang reservations of Rio Grande do Sul, including that sampled in the present investigation (Nonoai). The author found that in Nonoai 14% of the matings were between Amerindians and non-Amerindians, and that of these, 84% involved an Amerindian woman and a non-Amerindian man. Conversely, in 16% of the cases the non-Amerindian partner would be a woman, therefore explaining the mtDNA findings.

CONCLUSIONS

Answer for the questions asked in the introduction can now be made. First, there are clear differences in the frequencies of the mtDNA lineages between Guarani and Kaingang, although they are less marked for the Y-chromosome haplogroups. Mitochondrial DNA nucleotide and gene diversities, and the amount of interpopulation variability found in the latter, are also diverse between the two tribes. Second, the present information and mtDNA results from other Tupian and Jêan tribes were compatible with previous linguistic and historical data which documented extensive, older Tupian migrations as compared to the Jêan more recent movements. Third, the process of interethnic exchanges that occurred in the Guarani and Kaingang along time was also diverse. Non-native admixture within the Guarani communities was largely restricted to non-Amerindian males, while among the Kaingang direct evidence of introduction through the maternal side was found. In general, our results illustrate the importance of relating information from diverse areas of knowledge to unravel the complex history of human populations.

19

ACKNOWLEDGMENTS

20

LITERATURE CITED

Alves-Silva J, Santos MS, Guimarães PEM, Ferreira ACS, Bandelt H-J, Pena SD, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67: 444 - 461.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revisions of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 123: 147.

Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71:1150-1160.

Bandelt H-J, Herrnstadt C, Yao Y-G, Kong Q-P, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, Zhang Y-P. 2003. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. Ann Hum Genet 67: 512-524.

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

Becker IIB, Laroque LFS. 1999. O índio Kaingang do Paraná: subsídios para uma etno-história. São Leopoldo: UNISINOS.

Bert F, Corellay A, Gené M, Turbón D. 2004. Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands. Ann Hum Biol 31: 9-28.

Bonatto SL, Salzano FM. 1997. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. Proc Natl Acad Sci USA 94: 1866-1871.

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, Zago MA, Salzano FM, Silva Júnior WA, Silva MCBO, Weimer TA. 1997. Genetic variability in two Brazilian ethnic groups: a comparison of mitochondrial and protein DNA. Am J Phys Anthropol 103: 147-156.

Bortolini MC, Silva Júnior WA, Castro-de-Guerra D, Remonatto G, Mirandola R, Hutz MH, Weimer TA, Zago MA, Salzano FM. 1999. African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi and uniparental genetic markers. Am J Hum Biol 11: 551-563.

Bortolini MC, Salzano FM, Bau C, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Bedoya G, Ruiz-Linares A. 2002. Y-chromosome biallelic polymorphisms and Native American population structure. Ann Hum Genet 66: 255-259.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanem SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73: 524-539.

Bortolini MC, da Silva WA JR, Zago MA, Elion J, Krishnamoorthy R, Gonçalves VF, Pena SD. 2004. The phylogeography of mitochondrial DNA haplogroup L3g in África and the Atlantic slave trade. Am J Hum Genet 75: 522-524.

Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, Cruciani F, Wallace DC. 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. Cienc Cult 51: 166-174.

Campbell L. 1997. The historical linguistics of Native America. American Indian languages. Oxford: Oxford University Press.

Carneiro da Cunha M. 1998. Historia dos índios no Brasil. São Paulo: Companhia das Letras.

Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ. 2001. The phylogeography of Brazilian Y-chromosome lineages. Am J Hum Genet 68: 281-286.

Dornelles CL, Bonatto SL, Freitas LB, Salzano FM. 2005. Is haplogroup X present in extant South American Indians? Am J Phys Anthropol 127: 439-448.

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

Faucz FR, Probst CM, Petzl-Erler ML. 2000. Polymorphism of LMP2, TAP1, LMP7 and TAP2 in Brazilian Amerindians and Caucasoids: implications for the evolution of allelic and haplotypic diversity. Eur J Immunogenet 27: 5-16.

Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. 2004. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. Mol Biol Evol 20: 1682-1691.

Gaspar PA, Hutz MH, Salzano FM, Hill K, Hurtado AM, Petzl-Erler ML, Tsuneto LT, Weimer TA. 2002. Polymorphisms of *CYP1A1, CYP2E1, GSTM1, GSTT1*, and *TP53* genes in Amerindians. Am J Phys Anthropol 119: 249-256.

Green LD, Derr JN, Knight A. 2000. mtDNA affinities of the peoples of north Central Mexico. Am J Hum Genet 66: 989-998.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98.

Heckenberger MJ, Neves EG, Petersen JB. 1998. De onde surgem os modelos? As origins e expansões Tupi na Amazônia Central. Rev Antropol (São Paulo) 41: 1-17.

Hünemeier T, Neves AG, Nornberg I, Hill K, Hurtado AM, Carnese FR, Goicoechea AS, Hutz MH, Salzano FM, Chies JAB. 2005. T-Cell and chemokine receptor variation in South Amerindian populations. Am J Hum Biol 17: 515-518.

Jobling M, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4: 598-612.

Kern AA. 1997. Arqueologia pré-histórica do Rio Grande do Sul. Porto Alegre: Mercado Aberto.

Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R. 2002. The emerging limbs and twing of th East asia mtDNA tree. Mol Biol Evol 10: 1737-1751.

Kohlrausch FB, Callegari-Jacques SM, Tsuneto LT, Petz-Erler ML, Hill M, Hurtado AM, Salzano FM, Hutz MH. 2005. Geography influences microsatellite polymorphism diversity in Amerindians. Am J Phys Anthropol 125: 463-470.

Long JC. 1991. The genetic structure of admixed populations. Genetics 127: 417-428.

Marrero AR, Leite FPN, Carvalho BA, Peres LM, Kommers TC, Cruz IM, Salzano FM, Ruiz-Linares A, Silva-Jr WA, Bortolini MC. 2005. Heterogeneity of the genome

ancestry of individuals classified as white in the state of Rio Grande do Sul, Brazil. Am J Hum Biol 17: 496-506.

Migliazza EC. 1982. Linguistic prehistory and the refuge model in Amazonia. In: Prance GT, editor. Biological diversification in the tropics. New York: Columbia University Press. p 497-519.

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

Noelli FS. 1998. The Tupi: explaining origin and expansions in terms of archaeology and of historical linguistics. Antiquity 277: 648-664.

Ota T. 1993. DISPAN: genetic distance and phylogenetic analysis. University Park: Institute of Molecular Evolutionary Genetics, Pennsylvania State University.

Pagano S, Sans M, Pimenoff V, Cantera AM, Alvarez JC, Lorente JA, Peco JM, Mone P, Sajantila A. 2005. Assessment of HV1 and HV2 mtDNA variation for forensic purposes in a Uruguayan population sample. J Forenscic Sci 50: 1-4.

Pena SD, Santos FR, Bianchi NO, Bravi CM, Carnese FR, Rothammer F, Gerelsaikhan T, Munkhtuja B, Oyunsuren T. 1995. A major founder Y-chromosome haplotype in Amerindians. Nat Genet 11:15-16.

Petzl-Erler ML, Luz R, Sotomaior VS. 1993. The HLA polymorphism of two distinctive South-American Indian tribes: the Kaingang and the Guarani. Tiss Antig 41: 227-237.

Rodrigues AD. 1964. A classificação do tronco lingüístico Tupi. Rev Antropol (São Paulo) 12: 99-104.

Rodrigues AD. 2000. Hipótese sobre as migrações dos três subconjuntos meridionais da família Tupi-Guarani. Atas do II Congresso Nacional da ABRALIN: 1596-1605. Universidade Federal de Santa Catarina, Florianópolis, SC.

Salzano FM. 1961. Studies on the Caingang Indians. I. Demography. Hum Biol 33; 110-130.

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

Salzano FM, Callegari-Jacques SM 1988. South American Indians. A case study in evolution. Oxford: Clarendon Press.

Salzano FM, Callegari-Jacques SM, Weimer TA, Franco MHLP, Hutz MH, Petzl-Erler ML. 1997. Electrophoretic protein polymorphisms in Kaingang and Guarani Indians of Southern Brazil. Am J Hum Biol 9: 505-512.

Santos SEB, Ribeiro-dos-Santos AKC, Meyer D, Zago MA. 1996. Multiple founder haplotypes of mitochondrial DNA in Amerindians revealed by RFLP and sequencing. Ann Hum Genet 60: 305-319.

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

Schmitz PI,. 1997. Migrantes da Amazônia; a tradição Tupiguarani. In: Kern A, editor. Arqueologia pré-histórica do Rio Grande do Sul. Porto Alegre: Mercado Aberto. p 295-330.

Schmitz PI, Becker IIB. 1997. Os primitivos engenheiros do planalto e suas estruturas subterrâneas: a tradição Taquara. In: Kern A, editor. Arqueologia pré-histórica do Rio Grande do Sul. Porto Alegre: Mercado Aberto. p 251-289.

Schneider S, Roessli D, Excoffier L. 2000. Arlequin (ver 2000) - a software for population genetic data analysis. Geneva: University of Geneva.

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC. 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, Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. Am J Hum Biol 16: 420-439.

Sotomaior VS, Faucz FR, Schafhauser C, Janzen-Duck M, Boldt AB, Petzl-Erler ML. 1998. HLA-DQA1 and HLA-DQB1 alleles and haplotypes in two Brazilian Indian tribes: evidence of conservative evolution of HLA-DQ. Hum Biol 70:789-797.

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

The Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12: 339-348.

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Willians RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC. 1992. Native

American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153-162.

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

Tsuneto LT, Probst CM, Hutz MH, Salzano FM, Rodriguez-Delfin LA, Zago MA, Hill K, Hurtado AM, Ribeiro-dos-Santos AKC, Petz-Erler ML. 2003. HLA class II diversity in seven Amerindian populations. Clues about the origin of the Aché. Tiss Antig 62: 512-526.

Underhill P, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc Natl Acad Sci USA 93: 196-200.

Urban G. 1996. On the geographical origins and dispersion of Tupian languages. Rev Antropol (São Paulo) 39: 61-104.

Urban G. 1998. A história da cultura brasileira segundo as línguas nativas. In: Carneiro da Cunha MC, editor. Historia dos índios no Brasil. São Paulo: Companhia das Letras. p 87-102.

Utiyama SRR, Guardiano J, Petzl-Erler ML, Mocelin V, de Messias-Reason IJT. 2000. Autoantibody profile among Kaingang and Guarani tribe Indians in Southern Brazil. Rev Panam Sal Publ 7: 371-376.

Vietta K. 1992. M`byá: Guarani de verdade. Dissertação de Mestrado. Programa de Pós-Graduação em Antropologia Social. Universidade Federal do Rio Grande do Sul, Porto Alegre, RS.

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, Santos RV. 1996. Mitochondrial DNA polymorphism in three Brazilian Indian tribes. Am J Hum Biol 8: 317-323.

Yao Y-G, Bravi CM, Bandelt H-J. 2004. A call for mtDNA data quality control in forensic science. Forens Sci Int 141:1-6.

TABLE 1. *HVS-I sequence variation and major continental-specific mtDNA haplogroups observed in the Guarani and Kaingang samples[1].*

HVS-I sequence position

```
11111111111111111111111111111111
66666666666666666666666666666666
00011111112222222222222223333333
57911225780112367779999112 2356
15314693299373960480158195 7552
```

	rCRS ATTCCTGGTTTGTCCCCGCCCCTTGTCACT	Haplogroup	Guarani				Kaingang		
			GKW	GNA	GMB	**Total**	KRS	KPR	**Total**
1	...T.........T.....T....A....C	A	64	9	6	**79**	18	12	**30**
2	...T.........T.....TT...A....C	A	3	7	6	**16**		1	**1**
3	...T......C..T.....T....A....C	A	42	8		**50**			
4	...T.........T.T...T....A....C	A		8		**8**			
5	..T.........T...A.T....A....C	A					2		**2**
6	...T...A.....TAT...T....A....C	A		14		**14**			
7	...T.........T..T..T....A....C	A	1			**1**			
8	....C.......T....TT...A....C	A					2		**2**
9	.........C...C...........C......	B					3		**3**
10	............T...........CT...	C	2	9		**11**			
11	.C..........T...........CT...	C	8			**8**			
12	............T.......C..CT...	C					9	3	**12**
13	............T.......C..CTG..	C					2		**2**
14	G.......C....T......TCC.CTG..	C					14	5	**19**
15	G...........T.......C..CT...	C					1		**1**
16	G.C.........T.......C..CT...	C					2		**2**
17	............T...........C...C	D		1	12	**13**			
18	...A.A....A.T....T.........TC	L2b1					3		**3**
19	....	H or HV or U or R					1		**1**
Total			120	56	24	**200**	57	21	**78**

[1] Abreviations are as follows: rCRS: Revised Cambridge Reference Sequence (Andrews et al., 1999); GKW= Guarani Kaiowá; GNA= Guarani Ñandeva; GMB= Guarani M´byá; KRS= Kaingang-Rio Grande do Sul; KPR= Kaingang-Paraná.

TABLE 2. *mtDNA diversity parameters in Guarani and Kaingang populations, compared to other Tupian and Jêan tribes*

Population	No. stud.	Amerindian haplogroups (%)[1]				Nucleotide diversity (π)	Gene diversity[2]	Reference
		A	B	C	D			
Guarani-Tupian							G_{ST} (19 %)	
Guarani-Ñandeva	56	82	0	16	2	0.0094 ± 0.005	0.0608 ± 0.017	Present study
Guarani-Kaiowá	120	92	0	8	0	0.0042 ± 0.003	0.0275 ± 0.009	Present study
Guarani-M´byá	24	50	0	0	50	0.0067 ± 0.004	0.0434 ± 0.176	Present study
Total	**200**	**84**	**0**	**9.5**	**6.5**	**0.0067 ± 0.004**	**0.0495 ± ND**	
Others-Tupian							G_{ST} (53 %)	
Aché[3]	63	10	90	0	0	0.0029 ± 0.002	0.0214 ± 0.007	Schmitt et al. (2004)
Cinta Larga	20	20	0	20	60	0.0558 ± 0.043	0.0818 ± 0.037	Dornelles et al. (2005), Ribeiro-dos-Santos et al. (unpublished)
Gavião	43	9	9	0	82	0.0085 ± 0.005	0.0515 ± 0.026	Ward et al. (1996), Ribeiro-dos-Santos et al. (unpublished)
Munduruku	40	13	15	10	62	0.0144 ± 0.016	0.0877 ± 0.037	Ribeiro-dos-Santos et al. (unpublished)
Parakanã	13	8	23	46	23	NE	NE	Dornelles et al. (2005), Ribeiro-dos-Santos et al. (unpublished)
Potujuara	20	45	0	25	30	0.0173 ± 0.012	0.1053 ± 0.031	Santos et al. (1996), Ribeiro-dos-Santos et al. (unpublished)
Suruí	44	11	2	0	87	0.0049 ± 0.003	0.0319 ± 0.012	Bonatto and Salzano (1997),
Urubu-Kaapor	42	22	33	14	31	0.0631 ± 0.048	0.1286 ± 0.035	Dornelles et al. (2005), Ribeiro-dos-Santos et al. (unpublished)
Wayampi	24	75	0	17	8	0.0192 ± 0.015	0.1052 ± 0.032	Santos et al. (1996), Ribeiro-dos-Santos et al. (unpublished)
Zoró	30	20	3	13	64	0.0115 ± 0.007	0.0698 ± 0.015	Ward et al. (1996)
Total	**339**	**20**	**25**	**10**	**45**	**0.0423** ± 0.021	**0.0992 ± ND**	
Kaingang-Jêan							G_{ST} (4 %)	
Kaingang-RS	57	41	6	53	0	0.0169 ± 0.009	0.1083 ± 0.024	Present study
Kaingang-PR	21	62	0	38	0	0.0148 ± 0.008	0.0959 ± 0.025	Present study
Total	**78**	**47**	**4**	**49**	**0**	**0.0171 ± 0.009**	**0.1008 ± ND**	
Others-Jêan							G_{ST} (58 %)	
Kokraimoro	2	50	50	0	0	0.0254 ± 0.027	0.1579 ± 0.048	Ribeiro-dos-Santos et al. (unpublished)
Krahó	8	50	38	12	0	0.0166 ± 0.010	0.1013 ± 0.027	Torroni et al. (1993)
Kubenkokre	4	0	100	0	0	0.0139 ± 0.010	0.0848 ± 0.026	Ribeiro-dos-Santos et al. (unpublished)
Mekranoti	1	0	100	0	0	NE	NE	
Txukahamae	2	100	0	0	0	0.0027 ± 0.004	0.0175 ± 0.017	Dornelles et al. (2005)
Xavante	25	16	84	0	0	0.0081 ± 0.008	0.0526 ± 0.016	Ribeiro-dos-Santos et al. (unpublished)
Xikrin	43	37	63	0	0	NE	NE	Dornelles et al. (2005), Ribeiro-dos-Santos et al. (unpublished)
Total	**85**	**32**	**67**	**1**	**0**	**0.0379 ± 0.019**	**0.0938 ± ND**	

28

[1]Using data from RFLP and/or sequences.

[2]Using the method presented by Nei (1987), and considering the HVS-I variant nucleotide frequencies. Nucleotide and gene diversities were not estimated for some populations (NE) because just one sequence was available, the haplogroup frequencies being estimated by RFLP.

[3]The Aché Indians from eastern Paraguay speak a language listed under the Tupi-Guarani linguistic branch, but their phylogenetic relationship with other Amerindians is unclear. Some genetic studies showed a link with the Guarani (Battilana et al., 2002; Tsuneto et al., 2003), but others indicated a higher identity with Jêan-speaking populations rather than with Guarani groups (Kohlrausch et al., 2005).

TABLE 3. *Parental contributions in Guarani and Kaingang populations based on Y-chromosome, mitochondrial and nuclear DNA data sets*

Population	Parental contribution (%)		
	European	African	Native American
Guarani M´byá			
mtDNA[1]	0	0	100
Y-Chromosome[2]	4	31	65
Guarani Ñandeva			
mtDNA[1]	0	0	100
Y-Chromosome[2]	10	5	85
Guarani Kaiowá			
mtDNA[1]	0	0	100
Y-Chromosome[2]	11	3	86
Guarani (Total)			
mtDNA[1]	**0**	**0**	**100**
Y-Chromosome[2]	**9**	**14**	**77**
Biparental[3]	**0-3**	**0-3**	**97**
Kaingang-Paraná			
mtDNA[1]	0	0	100
Y-Chromosome[2]	18	24	58
Kaingang-Rio Grande do Sul			
mtDNA[1]	2	5	93
Y-Chromosome[2]	14	0	86
Kaingang (Total)			
mtDNA[1]	**1**	**4**	**95**
Y-Chromosome[2]	**15**	**16**	**69**
Biparental[3]	**0-7**	**0-7**	**93**

[1]Values obtained directly from the distributions of the major continental-specific mtDNA haplogroups listed in Table 1 (A+B+C+D = Amerindian; H = European; L2b1 = sub-Saharan African).

[2]Since some Y-haplogroups (DE* and Y*) are not continental-specific, the estimates of the parental contributions were obtained using the frequencies presented in Figure 3 and Long´s (1991) least square method. Parental frequencies used in this analysis were those given by Marrero et al. (2005).

[3]Values compiled from Salzano et al. (1997) and Callegari-Jacques and Salzano (1999). The authors did not discriminate the European and African contributions, calculating just the non-Native component.

 Fig. 1. Probable routes of dispersion of the Tupian and Jêan speakers. Arrows indicate possible routes and the estimated dates when they may have occurred in years before present. Dots show the main archeological sites of the Tupi-Guarani culture (modified from Schmitz, 1997), while the circle represents the probable region of origin of the Jêan linguistic family (Urban, 1998).

Fig. 2. Map showing the approximate geographic location of the Native American populations studied here for the HVS-I and Y-chromosome markers. Filled and open circles represent the Tupian and Jêan villages, respectively.

Fig. 3. Phylogenetic tree of the Y-chromosome haplogroups and their distributions (%) in the Guarani and Kaingang populations studied here and those tested by Bortolini et al. (2003).

Fig. 4. Medium network of the specific A lineage carrying 16266T transition. The root haplotype is identified by #1 (16111T-16223T-16266T-16290T-16319A-16362C). Variant positions from the root are indicated as numbers (mutations from the reference sequence minus 16,000) in the branches of the network; the letter A after 239 indicates a transversion. Circle sizes are proportional to the lineage frequency. #1 – 1 Ignaciano and 1 Yuracare (Bert et al., 2004); 1 non-Native Brazilian (Alves-Silva et al., 2000); 1 non-Native Uruguayan (Pagano et al., 2005); 8 Guarani Ñandeva (present study). #2 – 1 Gavião (Ward et al., 1996). #3 – 1 Tayacaja Quechua (Fuselli et al., 2004). #4 – 3 Amazonian Amerind (Santos et al., 1996); 2 Wai Wai (Bonatto and Salzano, 1997). #5 – 1 Amazonian Amerind (Santos et al., 1996). #6 – 3 Neo Brazilian (Alves-Silva et al., 2000; Marrero et al., 2005). #7 – 1 Neo Brazilian (Alves-Silva et al., 2000). #8 – 14 Guarani Ñandeva (present study).

A- Tupian linguistic family
B- Tupi-Guarani branch
C- Tupi language
D- Guarani language
E- Jêan linguistic family
F- Jêan-Southern branch

Figure 1

33

●Guarani Kaiowá (23°06'S, 55°12'W)
●Guarani Ñandeva (23°48'S, 54°30'W)
●Guarani M`byá (25°18'S, 52°32'W)
○Kaingang Paraná (24°30'S, 55°50'W)
○Kaingang Rio Grande do Sul (27°S, 52°W)
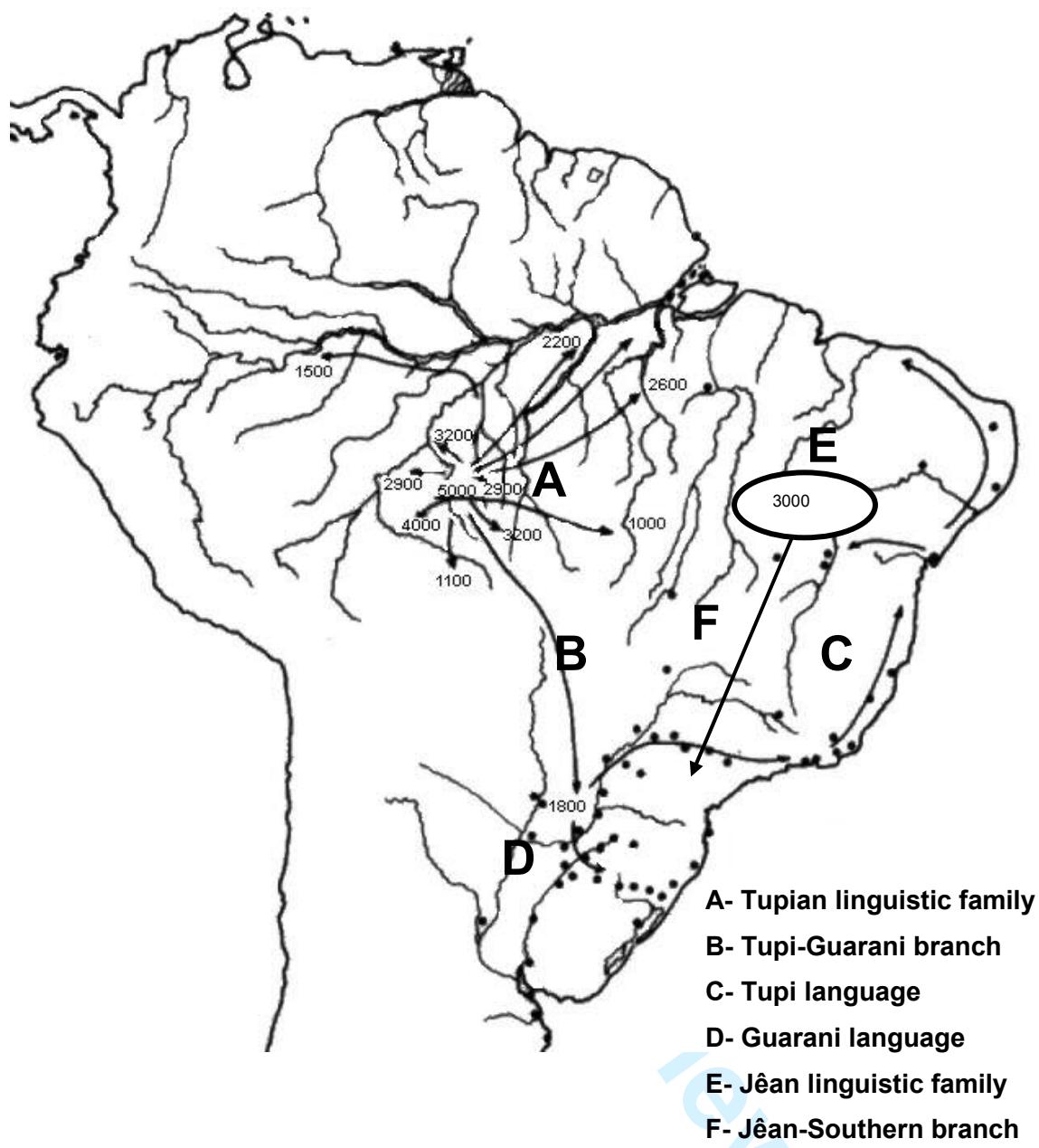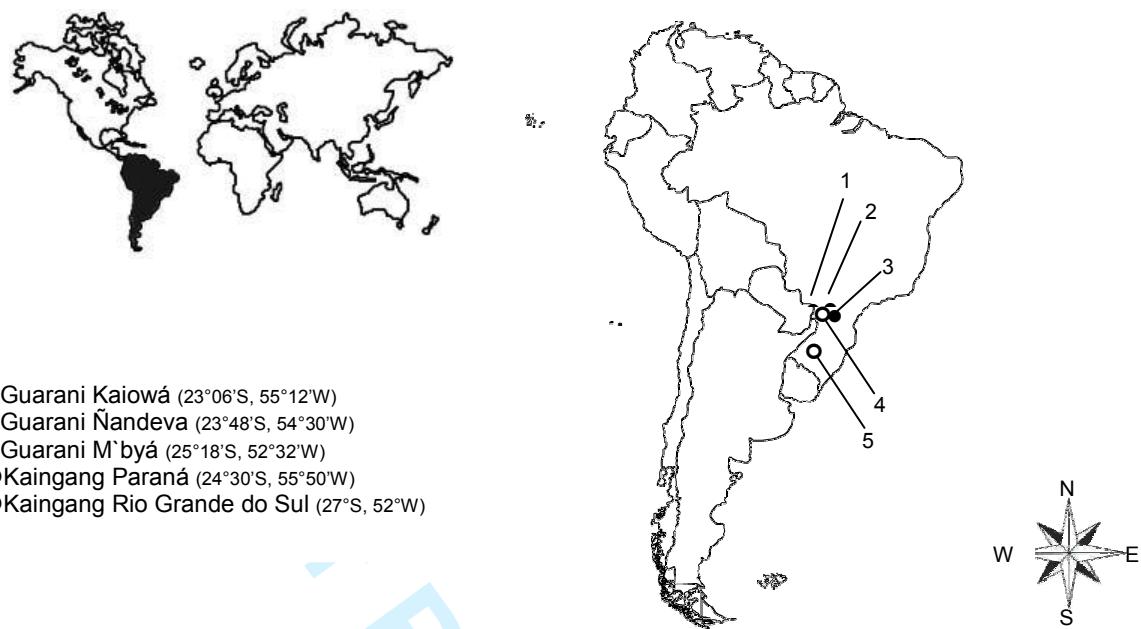
Figure 2
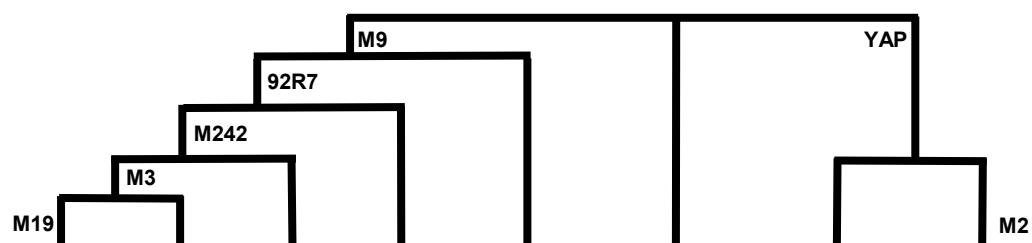
Population	Q3a	Q3*(xQ3a)	Q*(xQ3)	P*(xQ)	K*(xP)	Y*	DE*(xE3a)	E3a*
Guarani M`byá (30)		61	3	3		27		6
Guarani Ñandeva (20)[1]		70	15	10				5
Guarani Kaiowá (28)[1]		86		11				3
Guarani, Total (78)		**72**	**5**	**8**		**10**		**5**
Kaingang PR (36)		50	8	11		31		
Kaingang RS (22)[1]		86		14				
Kaingang, Total (58)		**64**	**5**	**12**		**19**		

[1]Bortolini et al. (2003).

Figure 3
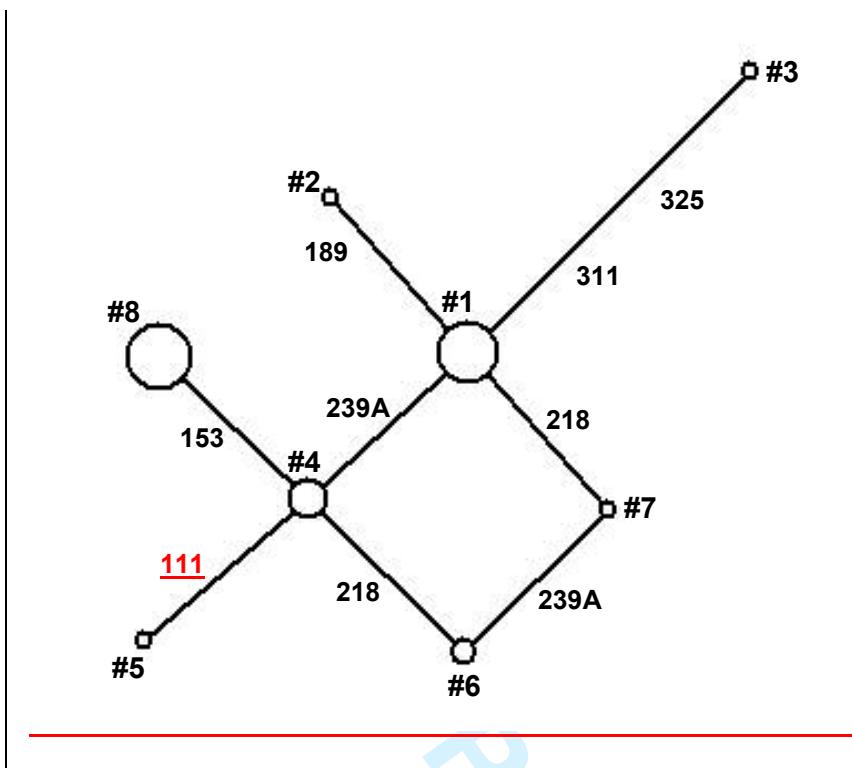
Figure 4

CAPÍTULO IV

Population data on 17 X-chromosome short tandem

repeat loci in a sample from southern Brazil

Marrero *et al*., manuscrito em preparação

Announcement of population data

Population data on 17 X-chromosome short tandem repeat loci in a sample from southern Brazil

Andrea R. Marrero[a], Siija Wang[b], Francisco M. Salzano[a], Andres Ruiz-Linares[b], Maria Cátira Bortolini[a*]

[a]*Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil*

[b]*Department of Biology, University College London, London, United Kingdom*

KEYWORDS: X-chromosome STRs; Haplotypes; Brazil; Population genetics.

* Corresponding author. Tel.: +55 51 33169844   Fax: +55 51 3316 9823

E-mail address: maria.bortolini@ufrgs.br (M.C. Bortolini)

Abstract

Genetic diversity at 16 X-chromosome short tandem repeat loci has been studied in a southern Brazilian admixed (European/African/Amerindian) population. DXS1199, DXS1047, DXS1060, DXS1069, DXS1073, DXS1106, DXS1214, DXS1226, DXS1227, DXS8051, DXS8055, DXS986, DXS987, DXS990, DXS991 and DXS993 allele frequencies were obtained from 70 unrelated males. DX8051 has the highest (0.8687) gene diversity while DXS8055 shows the lowest (0.5565). The averages of power of discrimination were estimated as 90% (female) and 75% (male).

Population: Blood samples were obtained from 70 unrelated healthy volunteer males from the Brazilian Pampa region (Fig.1). The Pampa expression designates a geographic region that includes portions of Argentina, Uruguay and Brazil, and in relation to the latter specifically the southwest of Rio Grande do Sul state. The males inhabitants of this region are best-known as *Gaúcho* (or *Gaucho* in Spanish) and they generally are known by their specific and impressing cultural unity. All donors were informed about the aims of this study and signed a written consent form. This investigation was approved by the Brazilian National Ethics Commission (CONEP number: 1333/2002).

DNA extraction: DNA was extracted from whole blood, according to Marrero et al. [1].

Typing: We typed 17 fluorescent-labelled microsatellite or STR markers from Panel 28 of the ABI PRISM Linkage panel sets V2.5 (Table 1), according to the user's manual provided by the manufacturer. An ABI 3730xl sequencer was used to detect the microsatellites, while genotyping was processed by GENEMAPPER v3.5.

Data analysis: Allele frequencies were calculated for each locus through the gene counting method. Diversities were estimated using the ARLEQUIN version 3.01 software (http://cmpg.unibe.ch/software/arleqin3/) and other parameters were estimated as previously described [2].

Access to the data: The complete data set is available through maria.bortolini@ufrgs.br.

Results: Allele distributions and statistical parameters are shown in Table 2 and haplotype frequencies in Table 3.

Other remarks: The number of alleles varied from one to 14, the most common number being seven. Heterozygosity levels ranged from 0.5565 (DXS8055) to 0.8817 (DXS1226), PIC values were generally high (0.50 - 0.85), and the PD averages were estimated as 0.8991 for females and 0.7516 for males. Since this panel set of markers is a new one, there are no published studies to compare. Table 4 shows the number of recombinations and informative meioses, i.e., the number of meioses in which the corresponding pair of markers were both informative (http://research.marshfieldclinic.org). This information should be valuable to forensic and population genetic investigations. This paper follows the guidelines for publication of population data requested by the journal [3].

References

[1] A. Marrero, F. Leite, B. Carvalho, L. Peres, T. Kommers, I. Cruz, F. Salzano, A. Ruiz-Linares, W. Silva-Junior, M. Bortolini. Heterogeneity of the genome ancestry of individuals classified as White in the state of Rio Grande do Sul, Brazil. Am.J.Hum.Biol. 17 (2005) 496-506.

[2] L. Excoffier, G. Laval, S. Schneider. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1 (2005) 47-50.

[3] P. Lincoln, A. Carracedo, Publication of population data of human polymorphisms, Forensic Sci. Int. 110 (2000) 3–5.

Table 1

Characteristics of the 17 Chromosome X short tandem repeats investigated in this study

Symbol	Locus	Localization	GenBank Access	Fragment Length(bp)
DXS1001	HS248WE5		Z17117	351
DXS1047	HS150XF10	Xq25-q26	Z23398	331
DXS1060	HS205TF2	Xp22.32	Z17254	322
DXS1068	HS238YC11		Z17268	382
DXS1073	HS276XH9	Xq28	Z51243	327
DXS1106	HS263WE1	Xq22.1-q22.2	Z17276	389
DXS1214	HS283WG9	Xp21.1	Z24023	350
DXS1226	HS316YF5	Xp22.13	Z24327	383
DXS1227	HS317YE9	Xq26	Z24335	262
DXS8051	HSB285YG9	Xp22.31	Z53331	326
DXS8055	HSB291YE5	Xq23	Z53364	370
DXS986	HS116XG1	Xq21.1	Z16606	340
DXS987	HS120XA9	Xp22.2	Z16615	303
DXS990	HS136YC7	Xq21.33	Z16647	221
DXS991	HS151XF6	Xp11.21	Z16680	316
DXS993	HS203WF4	Xp11.4	Z16898	367

Table 2

Allele frequencies and gene diversities for X-chromosome STR loci in a sample from southern of Brazil.

	DXS1001	DXS1047	DXS1060	DXS1068	DXS1073	DXS1106	DXS1214	DXS1226	DXS1227	DXS8051	DXS8055	DXS986	DXS987	DXS990	DXS991	DXS993
1	0.0435	0.0139	0.0580	0.1143	0.0442	0.1213		0.0435	0.0151		0.2253	0.0156	0.0149	0.0606	0.0145	0.0141
2		0.0278	0.2464	0.0286	0.5588	0.1818	0.0704		0.0151	0.0147	0.6197	0.1719	0.1791	0.4848		0.3803
3	0.0435	0.0833	0.0869	0.2429	0.3382	0.5454	0.2817		0.4091	0.0147	0.1268	0.0781	0.1045	0.1212	0.0145	0.1549
4	0.0435	0.4167	0.2609	0.0714		0.0909	0.1549		0.1364	0.1325	0.0141	0.4375	0.0597	0.1818		
5	0.2754	0.1250	0.0435	0.0429	0.0147	0.0303	0.2957	0.0290	0.0909	0.2059	0.0141	0.0781	0.1342	0.0758		
6	0.1159	0.1528	0.2898	0.4428		0.0303	0.1127	0.0725	0.1061	0.0735			0.1194	0.0758	0.0290	0.0986
7	0.1304	0.0972	0.0145	0.0571	0.0147		0.0423	0.1159	0.1818	0.0588		0.0313	0.2537		0.0580	
8	0.1304	0.0555			0.0147		0.0423	0.1594		0.0441		0.0938	0.0448		0.2319	
9	0.2174	0.0139			0.0147			0.2029	0.0455	0.1176		0.0781	0.0300		0.1304	0.0141
10		0.0139						0.1884		0.1912		0.0156	0.0448		0.1304	0.1831
11								0.0435		0.0882			0.0149		0.2319	0.1408
12								0.1014		0.0588					0.1159	
13								0.0435							0.0435	
14																0.0141
HET	0.8367	0.7774	0.7856	0.7313	0.5790	0.6615	0.7992	0.8817	0.7706	0.8687	0.5565	0.7624	0.8643	0.7128	0.8508	0.7787
PIC	0.8032	0.7428	0.7386	0.6849	0.4966	0.6201	0.7573	0.8487	0.7306	0.8529	0.4951	0.7524	0.8350	0.6695	0.8190	0.7369
PD^F	0.9475	0.9219	0.9134	0.8862	0.7417	0.8471	0.9246	0.9666	0.9135	0.9683	0.7428	0.9129	0.9616	0.8788	0.9544	0.9152
PD^M	0.8251	0.7663	0.7742	0.7207	0.5704	0.6519	0.7876	0.8634	0.7590	0.8669	0.5485	0.7500	0.8513	0.7017	0.8385	0.7678

HET - observed heterozygosity; PIC - polymorphism information content; PD^F - Power of Discrimination for females; PD^M - Power of Discrimination for males.

Table 3

Haplotypes frequencies of 16 X-STRs in 70 Gaúcho males.

	DXS1001	DXS1047	DXS1060	DXS1068	DXS1073	DXS1106	DXS1214	DXS1226	DXS1227	DXS8051	DXS8055	DXS986	DXS987	DXS990	DXS991	DXS993
H1	5	6	2	6	3	2	8	9	5	5	1	5	5	4	9	14
H2	9	7	5	7	2	3	3	8	4	5	2	2	2	1	6	11
H3	5	5	4	3	3	3	5	13	7	4	1	9	2	2	8	11
H4	9	6	6	2	3	3	4	5	3	11	2	4	5	2	10	3
H5	9	4	2	6	3	3	4	12	3	5	2	4	7	4	12	3
H6	7	4	7	4	3	3	5	11	7	4	3	8	7	2	12	10
H7	9	7	2	6	3	3	4	9	5	10	2	8	6	2	11	2
H8	3	6	2	6	2	6	5	9	7	6	2	9	7	4	10	2
H9	5	4	1	2	2	3	5	10	7	6	2	7	7	2	11	1
H10	8	4	2	3	2	3	3	10	5	12	2	2	7	4	11	10
H11	5	4	6	1	2	3	6	11	4	4	1	5	4	1	10	11
H12	6	4	4	6	2	4	3	8	3	9	3	8	10	3	9	9
H13	5	4	4	6	2	2	6	8	9	5	2	9	2	2	10	3
H14	9	8	2	6	3	3	5	7	4	8	2	4	7	2	11	2
H15	6	4	6	6	1	3	3	9	4	6	2	4	7	2	9	3
H16	6	4	3	6	2	2	5	7	7	7	2	4	7	4	12	2
H17	4	5	6	3	2	3	4	10	5	4	1	2	7	2	8	3
H18	9	5	6	6	2	3	4	9	5	4	2	4	6	5	11	2
H19	8	6	6	5	2	1	5	10	3	6	2	3	7	6	10	2
H20	7	8	4	6	2	3	3	10	3	10	2	3	2	2	8	2
H21	9	4	4	5	2	0	5	11	3	5	1	5	7	2	8	10
H22	5	10	3	6	3	4	8	0	4	5	2	4	6	2	11	2
H23	4	3	6	6	3	1	8	0	3	11	4	2	7	3	13	6
H24	1	5	6	3	3	3	3	7	2	5	1	4	7	4	8	3
H25	8	1	6	3	3	3	7	7	7	11	2	2	2	5	7	11
H26	1	7	1	6	2	3	5	10	3	9	3	2	3	2	12	2
H27	5	4	2	6	3	3	3	12	3	2	2	4	5	2	11	6
H28	8	2	4	6	2	4	7	7	3	5	2	4	4	2	11	2
H29	5	4	4	3	2	3	5	10	6	12	2	2	7	4	11	11
H30	5	6	6	3	3	3	6	8	3	9	2	2	3	2	11	2
H31	9	8	2	6	3	3	3	9	9	5	1	9	8	6	9	2
H32	5	6	4	6	2	3	3	7	3	10	3	4	6	2	10	2
H33	9	4	4	1	2	5	6	9	3	9	2	4	2	2	11	11
H34	7	4	6	6	2	3	6	12	9	10	1	4	3	4	8	10
H35	5	4	3	3	7	2	3	7	6	5	2	4	9	3	6	10
H36	5	2	6	6	2	2	3	10	3	10	3	4	2	4	8	11
H37	5	5	6	1	8	4	5	8	4	9	2	3	6	2	8	3
H38	7	7	6	6	2	3	6	13	3	3	2	4	5	1	10	3
H39	5	7	2	3	9	3	3	8	3	9	2	9	2	2	7	2
H40	5	4	3	3	2	2	5	10	6	12	3	4	10	2	13	10
H41	7	5	2	6	2	2	5	9	3	11	2	4	5	5	8	2
H42	6	3	6	3	3	3	5	12	3	10	2	4	3	5	8	2
H43	7	7	1	4	3	3	4	9	7	12	2	4	8	2	8	11
H44	3	5	5	1	2	6	2	7	4	10	2	3	7	6	8	3
H45	7	4	3	6	2	2	2	12	6	9	5	2	4	2	11	2
H46	9	5	2	4	2	3	4	8	6	5	1	4	7	3	12	11

H47	5	4	5	3	2	3	3	6	3	4	2	4	7	2	8	10
H48	8	4	1	4	1	4	2	1	3	5	3	5	9	2	13	10
H49	6	4	6	6	3	3	4	8	3	8	1	4	8	2	8	10
H50	5	3	2	7	2	4	5	8	7	11	2	3	6	3	10	2
H51	0	6	2	6	2	3	3	10	1	5	2	4	7	2	7	2
H52	8	4	4	1	5	2	5	9	3	5	3	2	2	4	11	6
H53	7	3	4	1	2	3	5	6	6	8	2	1	3	4	10	2
H54	8	6	6	6	3	2	3	10	3	5	1	7	2	2	11	2
H55	9	4	4	3	1	3	3	9	7	7	2	8	10	6	9	2
H56	5	4	6	1	2	3	4	6	7	10	2	4	7	2	8	10
H57	6	8	4	3	3	5	5	9	7	10	2	4	7	2	11	2
H58	6	5	6	6	3	1	3	8	3	7	1	4	5	3	1	6
H59	9	4	3	5	2	2	2	6	4	10	2	4	11	5	8	11
H60	8	4	2	6	2	1	5	10	3	6	2	4	6	6	8	6
H61	9	7	4	3	2	3	3	9	3	5	2	4	3	3	11	3
H62	5	3	4	7	3	2	4	9	3	9	1	2	2	2	11	2
H63	9	4	4	6	2	3	2	13	5	10	2	4	3	2	9	2
H64	5	4	2	6	3	3	5	9	3	5	1	10	6	4	9	2
H65	9	6	2	4	2	0	7	9	7	11	1	8	5	2	9	10
H66	7	4	4	7	2	0	5	6	3	4	2	4	1	2	12	10
H67	6	6	6	1	2	3	3	10	3	10	3	4	7	3	9	6
H68	5	4	6	6	2	3	5	9	3	4	2	4	5	2	12	10
H69	8	4	6	6	2	1	3	8	3	4	2	5	5	2	12	6
H70	3	9	2	3	3	3	4	12	4	10	2	8	4	2	7	2

Table 4

Proportion between the Number of recombinations and informative meioses. The highest and lower values are underlying.

	DXS1001	DXS1047	DXS1060	DXS1068	DXS1073	DXS1106	DXS1214	DXS1226	DXS1227	DXS8051	DXS8055	DXS986	DXS987	DXS990	DXS991	DXS993
DXS1001	0															
DXS1047	11	0														
DXS1060	47	49	0													
DXS1068	44	46	36	0												
DXS1073	38	35	49	47	0											
DXS1106	23	37	43	48	51	0										
DXS1214	45	41	32	**_4_**	47	48	0									
DXS1226	54	46	28	23	60	55	15	0								
DXS1227	19	10	46	45	29	43	41	47	0							
DXS8051	50	46	5	32	49	41	28	24	44	0						
DXS8055	7	22	48	49	38	13	51	60	28	51	0					
DXS986	39	45	50	35	52	18	43	54	52	46	32	0				
DXS987	48	42	16	26	48	51	21	13	40	10	53	53	0			
DXS990	24	32	44	48	46	9	50	**_61_**	37	46	15	8	56	0		
DXS991	38	40	43	27	56	24	32	48	53	40	31	10	45	17	0	
DXS993	51	49	33	9	54	38	16	27	48	29	52	29	35	37	21	0

Figure 1. Maps showing in different levels the geographical localization of the Pampa region in Brazil.

Figure 2. Ideogram of the X chromosome with the localization of the STR used in this study.

Figure 1

DXS1060 - Xp22.32

DXS8051 - Xp22.31

DXS987 - Xp22.22

DXS1226 - Xp22.13

DXS986 and DXS1214 - Xp21.1

DXS1068 - Xp21-p11.4

DXS993 - Xp11.4

DXS 991 - Xp11.21

DXS990 - Xq21.33

DXS1106 - Xq22.1-q22.2

Xq22.2

DXS8055 - Xq23

DXS 1001 - Xq24

DXS1047 - Xq25-q26

DXS1227 - Xq26

DXS1073 - Xq28
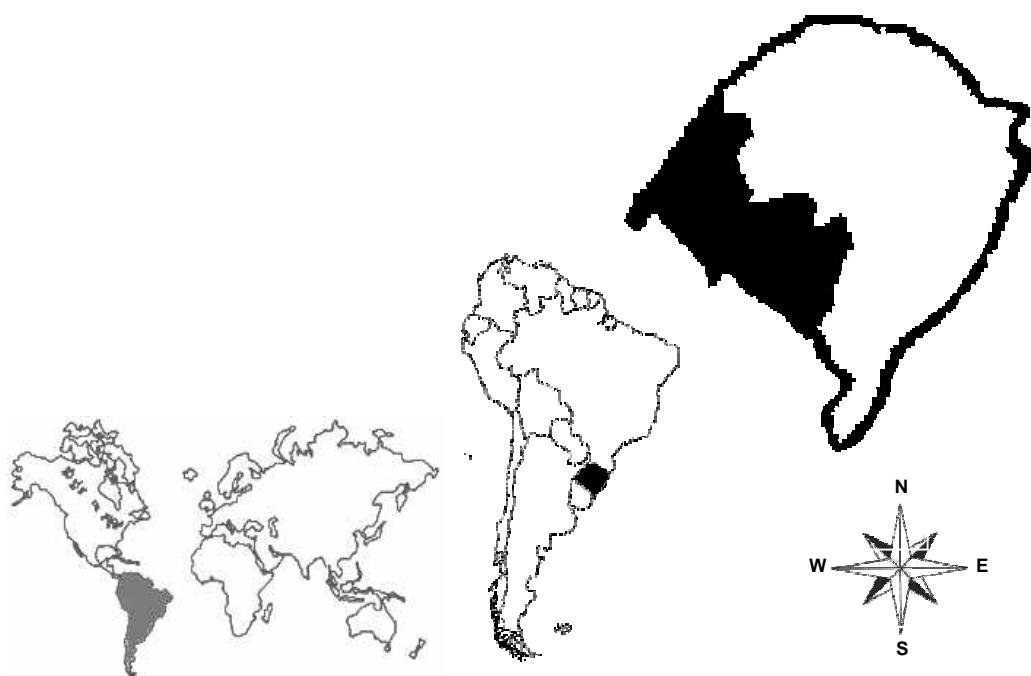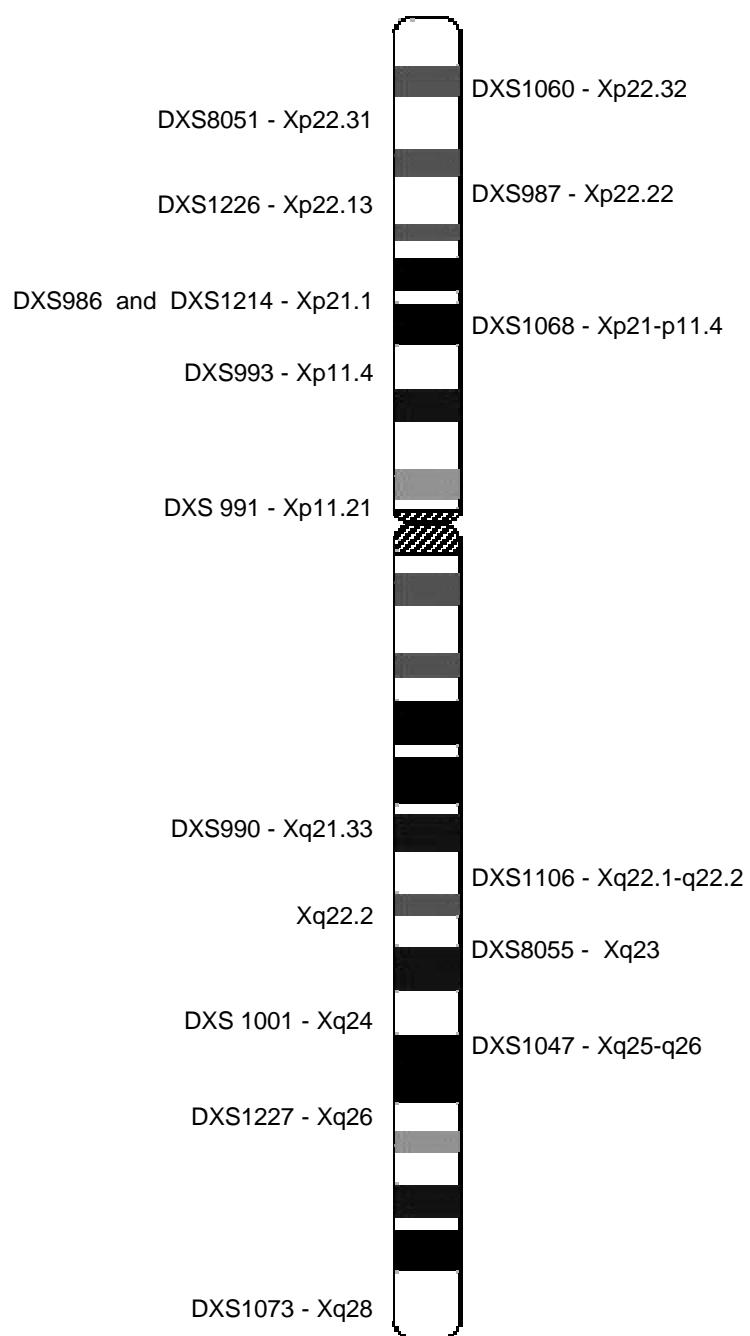
Figure 2

CAPÍTULO V

Genetic History of Gaúchos – Gene Dynamics in Southern Brazil

Marrero *et al*., manuscrito em preparação

The *Gaucho* Genetic History – Gene Dynamics in Southern Brazil

AR MARRERO[1], CM BRAVI[2] , S STUART[3*], JC LONG[4], FPN LEITE[1,5], TC KOMMERS[5], CMB CARVALHO[6], SDJ PENA[6], A RUIZ-LINARES[3], FM SALZANO[1], MC BORTOLINI[1**]

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

[2] *Laboratorio de Genética Molecular Poblacional, Instituto Multidisciplinario de Biología Celular (IMBICE), 1900 La Plata, Argentina*

[3] *The Galton Laboratory, University College, London NW1 2HE, United Kingdom*

[4] *Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109-061,USA*

[5] *Instituto Geral de Perícias, 90610-001 Porto Alegre, RS, Brazil*

[6] *Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil*

*In memoriam

**Corresponding author: Maria Cátira Bortolini, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. Fax: +55 51 3316 9844. E-mail: maria.bortolini@ufrgs.br.

Summary

A total of 150 unrelated males born in the Pampa region of Rio Grande do Sul, southern Brazil (who are known as *Gaucho)* were studied in relation to Y-chromosome and mitochondrial DNA (mtDNA) markers. This information provided data about their genetic variability and ancestry. Comparison with other Brazilian and Uruguayan populations, as well as with their putative ancestors indicated a stronger male Spanish influence than that observed elsewhere in Brazil, a former Portuguese colony. Extensive mtDNA analyses of their Amerindian component gave clear indications of the presence there of material from extinct (Charrua), as well as extant (Guarani) tribes. The genetic analyses contributed in a significant way to reveal that the known cultural continuity between pre and post-Columbian Pampa populations was also accompanied by an extraordinary genetic continuity.

Introduction

When Charles Darwin visited the Pampa region (which corresponds to parts of Argentina, Uruguay and southern Brazil) in 1833 he has written in his diary: "*During the evening a great number of Gauchos came into drink spirits and smoke cigars: their appearance is very striking; they are generally tall and handsome, but with a proud and dissolute expression. They frequently wear their moustaches and long black hair curling down their necks. With their bright coloured robes, great spurs clanking on their heels, and knifes stuck as daggers (and often so used) at their waists, they look a very different race of men what might be expected from simple countrymen. [...] There is a high enjoyment in independence of the Gaucho´s life - to be able at any moment to pull up your horse and say: Here I will pass the night. The death-like stillness of the plain, the dogs keeping watch, the gypsy group of Gauchos making their bed around the fire, has left in my mind a strongly marked picture of this night, which will never be forgotten...*"(Darwin 1996).

Darwin´s account of his encounter with the Gauchos over 173 years ago provides us with a sharp picture of them. It is hard to determine the precise point at which the Gaucho arose as an ethnic and cultural element, but it is certainly connected with the presence of some specific factors: cattle, horses, vast plains in the Pampa, and admixture between Native American women and Iberian colonizer men (Leal 1989; Flores 2003; Bracco 2004). In some moment of this history the African element was also introduced since slaves were brought to the region (Weimer 1991; Vidart 2000; Reverbel 2002).

There are various other early narratives (between the end of the 18th and beginning of the 19th centuries) where the word Gaucho appeared; differently of Darwin´s opinion, most of them presented accusations that they were persons without manners, religion or

moral that wandered along the Pampa. However, all reports were unanimous in distinguishing the Gaucho as a peculiar countryman with a cultural homogeneity which transcended national borders (Leal 1989; Bracco, 2004). Later, they became important to the regional economy and politics due to their ability as workers in the estancias (ranches of thousand of acres, dedicated to extensive cattle breeding) and bravery as soldiers in the frontier wars (Nichols 1946; Oliven 1993). Certainly, the Gaucho from the Colonial Period were different from the contemporary Gaucho, but basically they live in the same area, respecting their cultural traditions. Today in Brazil the word Gaucho (Gaúcho in Portuguese) is used to refer to anyone born in the southernmost State of Rio Grande do Sul. In this paper we will be employing the word to refer only to people that were born and live in the Pampa region of Rio Grande do Sul (Fig. 1).

At the time the first Europeans arrived the Pampa was traditionally inhabited by nomad hunter-gatherer Amerindian groups named Charrua, Minuano, Guenoa, Chaná, Iaró, Mboanes, etc. It was difficult to culturally distinguish these several tribes or bands from each other, and due to this some researchers prefer to use the general term "Charrua major ethnic group" or only "Charrua" to name them (Alemán 1994; Pi Hugarte 1998; Abella 2000). This simplification will be adopted here.

The main Charrua, as well as Guarani and Kaingang (who also inhabited the Rio Grande do Sul later) archeological sites show absence or few overlap in their geographic distributions since they were traditional enemies and culturally very different (Becker 1997; Schmitz 1997; Bracco 2004; Fig. 1). The linguistic affiliations of the Guarani (Tupían, Tupí-Guarani branch; Campbell 1997) and Kaingang (Jêan stock; Campbell 1997) are well-known, but the language spoken by the Charrua is object of controversy because very little linguistic material has survived, since the Charrua disappearance in the

19th century. Campbell (1997) proposed the existence of a Charruan stock, with no relation to other South Amerindian linguistic groups. The archeological data, on the other hand, show suggestions of a connection between the Charrua and the Tierra del Fuego and Patagonia aborigines (Schmitz 1997; Pi Hugarte 1998). These same data also suggest a probable connection between the first inhabitants of the Pampa (who colonized the area around 10,000-9,000 years before present) with the Charrua who lived in the region at the time of the European arrival, therefore indicating a cultural continuity (Mentz-Ribeiro, 1997).

The Charrua influence in Gaucho culture is well-known (Pi Hugarte 1998; Vidart 2000; Reverbel 2002; Flores 2003; Bracco 2004). Although historical sources mention the presence of Charrua women into Colonial families and their relationship with non-Indian men (Flores 2003; Bracco 2004), the Charrua contribution to the formation of the Gaucho is always neglected when compared to that of the Guarani Indians, because while the first were considered bellicose and irreducible, the latter were pacific, having contact with non-Indians during all the Colonial Era (Kern *et al.* 1993; Christensen 2001; Becker 2003; Bracco 2004).

The objective of the present study was to uncover the evolutionary and demographic history of the Gaucho through the investigation of uniparental markers (Y-chromosome and mitochondrial or mtDNA). With this information and those of recent Guarani and Kaingang genetic characterization (Marrero *et al.* 2006) it was also possible to evaluate if the Gaucho are a reservoir of mtDNA Charrua lineages and if the Pampa region is a place where both pre- and post-Columbian cultural and genetic continuity have occurred. This investigation can be viewed as a contribution to the promising field of historical genetics.

Subjects and Method

Population Samples and DNA Extraction

A total of 150 unrelated males born in the Pampa region of Rio Grande do Sul (Fig. 1) were studied. All donors were previously informed about the objectives of the investigation, and samples were collected with their consent. Information about birth place, parents and grandparents was obtained.

DNA was extracted from whole blood according to Lahiri and Nurnberger (1991) and from saliva using the QIAamp® kit (QIAGEN), following the manufacturer's instructions.

Y- chromosome Markers

The samples were genotyped for six binary NRY markers (M242, M3, M19, 92R7, M9 and M2) and the Yap *Alu* insertion polymorphism following methods described in Bortolini *et al.* (2003). Genotyping was done according to the hierarchical order of the markers described by Underhill *et al.* (2000). Haplogroups Q*(xQ3), Q3*(xQ3a), Q3a, P*(xDE), K*(xPDE), DE*(xEa*) were established following the recommendations of Jobling & Tyler-Smith (2003). Eighty nine of these individuals were also genotyped for 12 Y-chromosome short tandem repeat loci (DYS391, DYS389 I, DYS439, DYS389II, DYS438, DYS437, DYS19, DYS392, DYS393, DYS390, and DYS385a/b). Typing was done using fluorescently labeled primers using the commercial PowerPlex® Y System kit (Promega) in a GeneAmp® PCR System 9600 (Perkin Elmer) thermocycler after amplification using the manufacturer's instructions. Electrophoresis and typing were performed on an ABI PRISM 310 DNA sequencer (Applied Biosystems). GeneScan

packages were utilized for data collection and size estimation of the fluorescent-labeled DNA fragments. Allele designation was established according to the International Forensic Y-User Group database (http://www.yhrd.org).

Mitochondrial DNA

The first hypervariable segment (HVI) of the mitochondrial DNA (mtDNA) control region was sequenced from 105 individuals of the same set genotyped for the Y-chromosome markers, using primers and conditions previously indicated in Marrero *et al.* (2005, 2006). The sequence reactions were run in automatic (ABI Prism 3100 and MegaBace) sequencers. Both DNA strands were sequenced. The nucleotide positions considered were from 16050 to 16391, since this is the stretch for which most information is available. Individuals with the "C-stretch" between positions 16184-16193, which is caused by the 16189C substitution, were sequenced again in each direction, so that each base was determined twice.

The sequences were checked manually, validated with the help of the CHROMAS LITE 2.0 program (www.technelsyum.com.au) and aligned with the revised Cambridge Reference Sequence (rCRS, Andrews *et al.* 1999) using the BIOEDIT software (Hall 1999). Since artifacts ("phantom mutations") can be introduced during the sequencing and editing process, we applied the filtering procedure described by Bandelt *et al.* (2003) to check for the quality of the sequences. After filtering, the relationships between the sequences were examined with the NETWORK 4.2.0.0. program (www.fluxus-engineering.com) using the median-joining algorithm (Bandelt *et al.* 1999). Weight

networks showing star tree patterns, associated with other criteria like those suggested by Yao *et al*. (2004), guarantee that the data are potentially free of phantom mutations.

The information provided by HVS-I was used to classify the lineages into haplogroups according to Salas *et al*. (2002, 2004), Bandelt *et al*. (2002), Kivisild *et al*. (2002) and Torroni *et al*. (2006).

Due to our particular interest in the characterization of the mtDNA lineages of Amerindian origin, all sequences identified with haplogroups A, B, C and D through sequencing, as well as those which had lost some diagnosis sites, were additionally analyzed to check for informative mutations in the mitochondrial coding region, using the minisequencing method developed by Carvalho & Pena (2005), and/or RFLP tests described by Green et al. (2000) and Alves-Silva et al.(2000).

Data Analysis

Y-SNP haplogroup frequencies were obtained by counting. Y-SNP and Y-STR haplogroup and haplotype diversities were estimated using the Arlequin version 3.01 software (Excoffier *et al.* 2005). Genetic F_{ST} distances were estimated and the population relationships visualized in two-dimensional space using the Multi Dimensional Scaling (MDS) analysis included in the SPSS software package (version 10.0). The statistical significance of the F_{ST} values was estimated by permutation using 10,000 runs.

The relationships among populations, considering only the Amerindian portion of the mtDNA sequences were also evaluated using Hunley & Long´s method (2005). According to them, the basic unit of this analysis is a matrix composed of the average number of nucleotide substitutions between pairs of mtDNA sequences. Model trees are fitted to the matrix of observed pairwise differences using maximum likelihood. Each

fitted tree produces a matrix of expected average pairwise nucleotide substitutions contingent on the assumption that the model tree accurately represents the true relationships among populations (Hunley & Long 2005).

We also investigated the pattern of relationships among individual Amerindian mtDNA lineages using median joining networks (Bandelt *et al*. 1999) constructed for haplogroups A, B, C and D with the NETWORK 4.2.0.0. software.

Results

Y-Chromosome Biallelic Polymorphisms (SNPs)

The Y-SNP haplogroup distributions (Table 1) show that most of the Y-chromosomes found in the Gaucho sample has a probable European origin (P*xQ, 58%) in accordance with studies from other admixed Brazilian populations. Haplogroup Y* was the second most frequent (32%), but since it indicates the presence of the ancestral alleles for the seven markers investigated, it is less informative and no specific continental origin can be inferred. On the other hand, haplogroups E3a* and Q3*(xQ3a), typical respectively of sub-Sahara Africans and Amerindians, were observed with distributions (3% and 5%) higher than those described for other regions of Brazil (0-2%), including other populations from Rio Grande do Sul (0-1%; Table 1).

The presence of this distinct Y-chromosome background in the Gaucho sample resulted in the highest level of haplogroup diversity (0.6557   0.0254) when compared with all other populations listed in Table 1.

To test the hypothesis of random distribution of the haplogroups among populations, we computed F_{ST} values as implemented by the Arlequin software, using the

frequencies shown in Table 1. The estimates (data not shown) indicated that the Gaucho have 4 times less differentiation with Spaniards ($F_{ST} = 0.026$; $p<0.05$) than with Portuguese ($F_{ST} = 0.104$; $p<0.05$), which contrasts with the values observed for the other Brazilian populations, that showed some level of differentiation with Spaniards ($F_{ST} = 0.017$; $p<0.05$), but not with the Portuguese ($F_{ST} = 0.006$; $p>0.05$). As expected, high levels of differentiation were observed between the Gaucho and African or Amerindian groups ($F_{ST} = 0.288$; $p<0.05$; $F_{ST} = 0.457$; $p<0.05$, respectively).

Y-Chromosome Microsatellite Polymorphisms (STRs)

To better characterize the Y-chromosome set present in our sample, haplotypes were constructed for each individual using the 12 microsatellite loci. Overall, 83 different haplotypes were observed in the 89 chromosomes tested, 78 of which were unique (Table 2).

The present Gaucho Y-STR data set was then compared with the Y-Chromosome Haplotype Reference Database (YHRD-http://www.yhrd.org), which includes 13,085 haplotypes observed in 90 populations from different continents typed according to the ISFG (International Society for Forensic Genetics) and SWGDAM (Scientific Working Group on DNA Analysis Methods) recommended core haplotype profile (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, DYS438, and DYS439). To compare our results with this large database, we excluded the DYS437 results from our data, but the number of observed haplotypes remained as 83. Of these, 64 are present in the worldwide database. Most matches were obtained with haplotypes

described in European populations, but H52 is identical with one observed in Cabinda, located in the Atlantic Central-West Africa.

Considering only the YHRD minimal haplotype (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS393) the number of observed haplotypes in our sample is reduced to 60, but the comparative analyses could be done with a larger number of populations of particular interest, such as Iberians, Africans and South Amerindians.

We compared these minimal haplotypes of the present study with those from Portugal (N=637) and Spain (N=1743), both data sets downloaded from YHRD and reported in Roewer *et al.* (2005); with Africans (N=265), compiled from Pereira *et al.* (2002); and Lecerf *et al*. (2006); South Amerindians (N=72) from Argentina (Wichi, Susque, Tehuelche, Toba, Mapuche, Chorote and Humahuaqueño) and Paraguay (Ayoreo and Lengua), all studied by Bianchi *et al.* (1998); Amazonian Amerindians (N=169) from Ecuador (Cayapa), Peru (Tayacaya and Arequipa) and Brazil (Gavião, Zoró, Suruí, Xavante, WaiWai, Karitiana and Ticuna) presented by Tarazona-Santos *et al.* (2001); and other Brazilian populations, divided in two sets: southern Brazilians (N=197), compiled from Cainé *et al.* (2005), Grattapaglia *et al.* (2005) and Carvalho-Silva *et al.* (2006); and other Brazilian regions (N=682), obtained from De Souza Góes *et al.* (2005), Grattapaglia *et al.* (2005), Palha *et al.* (2006), Silva *et al.* (2006) and Carvalho-Silva *et al.* (2006).

The most frequent haplotype in the Gaucho sample (DYS19 – **14**, DYS389I – **13**, DYS389II – **29**, DYS390 – **24**, DYS391- **11**, DYS393 – **13**) has 205 matches, with 128 Spaniards, 50 other Brazilians, 14 South Brazilians, 11 Portuguese and 2 Africans.

There are five haplotypes present in the Gaucho, that are exclusively shared with Spain, and one with Portugal. The comparison with the African dataset revealed five

haplotypes that also occur in Cabinda, Cameroon, Guinea Bissau and South Africa, all located in the African region from where slaves were forcibly transported to Brazil.

The H46 unique Native American arrangement (Table 2) observed in one Gaucho did not match any other already identified in South or North Amerindians.

The Gaucho also present 16 haplotypes (27%), which did not match any other considered in this analysis. This result could indicate that this population has a significant number of private haplotypes; but despite the high microsatellite mutation rates, the short period since migration from Europe seems insufficient to account for that. More plausible explanations are that drift can increase the frequency, in the derived population, of lineages that are rare in the parental group(s) (Wright 1931; Crow & Kimura 1970), and/or that the Iberian populations considered here are not accurately representing the Gaucho male ancestral groups.

The MDS population representation obtained with this STR (minimal haplotype) data set using F_{ST} statistic pairwise distances are seen in Fig. 2. The Gaucho are closer to Spaniards (F_{ST} = 0.0107; $p<0.05$) than to Portuguese (0.0275; $p<0.05$), differently from what occurs with other Brazilians. Additionally, the P*(xQ) haplotype diversity ($\pm$ SE) (considering the minimal haplotype) observed in the Gaucho, as well as in Iberian populations were estimated as follows: Gaucho = 0.996 $\pm$ 0.005; Spain = 0.970 $\pm$ 0.012; Portugal = 0.897 $\pm$ 0.044.

Mitochondrial DNA

The mtDNA sequence variation observed in the 105 Gaucho examined is summarized in Table 3. Multialignnment with the hypervariable sequence reference (rCRS,

Andrews *et al.* 1999) allowed the identification of 54 different lineages, that could be almost perfectly identified with some of the known continental-specific mtDNA haplogroups. Comparison of the HVS-I sequence data with those generated from the RFLP and/or minisequencing method did not reveal any disagreement in the haplogroup identifications. A discrepancy of this nature, related to a reverse mutation which occurred in the mtDNA coding region was recently described in Amerindian haplogroup C (Torres *et al.* 2006).

The haplotype diversity was calculated as 0.9584 ± 0.0101, and all nucleotide changes, except six, are transitions involving more pyrimidines than purines, with the C T substitution being the most frequent mutation. Transversions were identified at positions 16129 (G→C), 16182, 16183 (both A→C), 16188 (C→G), 16265 (A→C) and 16286 (C→G), the last three diagnosing mutations that define the specific African L0a* and L1c2 haplogroups.

Of all mtDNA lineages observed 24 (44%) can be classified in Native American haplogroups, while 19 (35%) and 11 (21%) are identified with European and African haplogroups, respectively. Taking into consideration their frequencies (Table 3) we estimate that 52%, 37% and 11% of the mtDNA sequences found in the Gaucho sample have an Amerindian, European and African origin (Table 4).

Table 4 also summarized the continental origin of the mitochondrial genomes of other admixed populations from Rio Grande do Sul, other Brazilian regions, and Uruguay. The European contribution is the most frequent in the South, except for the Gaucho and Tacuarembó. This European influence is not surprising, since southern Brazil and Uruguay had important and massive European migratory waves that occurred mostly during the second half of the 19th and beginning of the 20th century (Salzano and Bortolini 2002;

Bonilla *et al.* 2004). Differently of what occurred in the earlier centuries of colonization, when almost only European men migrated, this most recent major migratory movement involved couples and families (Salzano & Bortolini 2002). But, in the Gaucho and Tacuarembó samples the most important parental contribution is Amerindian (52% and 79%, respectively). The city of Tacuarembó is located in the Pampa region of Uruguay, which was also originally inhabited by Native American tribes, including Charrua. On the other hand, the African influence can be detected by the presence of 11% of sequences derived from this continent in the Gaucho sample. This number is intermediate to those observed in individuals morphologically identified as white (~3%) or black (~80%) in distinct regions of Rio Grande do Sul (Marrero *et al*. 2005; Hünemeier 2006). Tacuarembó received fugitive slaves from nearby Rio Grande do Sul, which can explain the presence of some African chromosomes in this Uruguayan population (Bonilla *et al.* 2004).

Restricting the attention to the Amerindian portion of the mtDNA sequences, it is possible to observe (Table 5) that in the Gaucho sample haplogroups A, B and C have a similar distribution (~30%), while B is less frequent (9%). These numbers are diverse from those observed in a sample from other regions of Rio Grande do Sul, but the difference is not significant (χ^2=5.24; p>0.05). Curiously, the presence of haplogroups C and B is respectively low or absent, considering a large Guarani sample. Haplogroup C, however, has an important frequency in the Kaingang, as well in the Patogonian and Fuegian tribes. The Gaucho nucleotide diversity values (0.0142; Table 6), is the lowest of all admixed populations considered here, but this value is also two times higher than that estimated for the Guarani (0.0067) and intermediate among those observed in other South Amerindians (range: 0.0102-0.0171).

Using again F_{ST} and MDS representation, but now on the mtDNA Amerindian data, we obtain the picture shown in Fig. 3. The most notable difference is between the samples from southern Brazil, Uruguay, Patagonia and Terra del Fuego with other Brazilian regions. No significant differences were found between the Gaucho, Uruguayan and southern Brazilian samples, but high F_{ST} values (~0.74; p < 0.001) were observed when the Gaucho were compared with Brazilians from the North, Northeast and Southeast regions, or with the Guarani (F_{ST} = 0.379; p< 0.001).

Figure 4 displays the population tree that presented the best fit with the Amerindian mtDNA sequence data set considered here. Two clusters can be identified, one defined by branch (**a**) which includes the Gaucho and other southern Brazilian admixed populations and Uruguayans; and the other defined by branch (**b**) which involves other admixed Brazilian populations. The most notable pattern in the tree is, however, the clear North-South geographic division.

To further investigate the possible origin of the Gaucho Amerindian mtDNA sequences, additional analyses using all lineages available in the literature and medium joining networks were performed. Only those lineages with matches with the Gaucho sequences, as well as others directly related to them were considered in the networks. Figs. 5 to 8 show the networks found.

Mitochondrial DNA haplogroup A lineage #3 was observed in two Gaucho and in several other populations (Fig. 5). It was also observed in 50 Guarani, supporting the idea that its presence in the Gaucho can be attributed to Guarani inheritance. The same can be suggested for lineage #2, which was also detected in admixed populations from the Southeast and Uruguay, areas with some influence of Tupi-Guarani tribes.

Marrero *et al*. (2006) suggested that the low level of mtDNA nucleotide diversity and the absence of haplogroup B in the Guarani are characteristics of this population and would be associated with a bottleneck in their southern migration from Amazonia about 2,000 years before present. However, haplogroup B lineage #20 (Fig. 6) present in one Gaucho is in a cluster with others detected in Amazonian populations, area of origin of the Tupian linguistic family. Thus, it is probable that this lineage has also a Guarani derivation. This result is showing that the bottleneck can have reduced the frequency of B in the Guarani, but its absence in the contemporary Guarani can be attributable to more recent demographic phenomena (post-contact depopulation) and/or sampling error. Haplogroup B lineages #2, #6 and #16 present matches with admixed populations and clustering with other Jêan and Karib tribes, preventing their precise origin identification.

As was mentioned in the Introduction, the Charrua were extinct since the 19th century. But recently, preliminary results obtained from bones and teeth of its legendary chief (Vaimaca Peru) were furnished (http://www.fhuce.edu.uy/antrop/cursos/abiol/vaimaca.html). He had a mitochondrial lineage associated to haplogroup C, with a rare T➔C transition at position 16288 (#36; Fig. 7). Although this mutation was not identified in any other of the lineages evaluated, the haplogroup C identification in the only Charrua sample known indicates that this haplogroup was also present in this Native people. The distribution pattern of C lineages shown in Fig. 7, however, is not very clear. The exception is the cluster with the Gaucho lineages #21 and #29. Both seem to have a Guarani origin, because they are connected to a series of others, spotted but widely distributed in Amazonian populations. The same can be said of Gaucho lineage #9, which presents a C➔T transition at position 16256, which was also observed in one Zoró (tribe geographically located in the probable center of spread of

the Tupian languages). For the other Gaucho C lineages (#2, #11, #14, #31 and #32) the Amerindian specific inheritance is not clear. Equally difficult is to define the origin of Gaucho D lineage #4 (Fig. 8), which carries three private mutations (G$\rightarrow$A at position 16145; C$\rightarrow$T at 16179; C$\rightarrow$T at 16295).

Discussion

Demographic and historical circumstances related to Brazil´s colonization determined that the first Brazilians were born mostly from the union between European males and Amerindian or African females. This gender- specific gene flow resulted in that most of the Y-chromosomes of contemporary members of Brazilian populations have an European origin (Bortolini *et al*, 1999; Carvalho-Silva *et al*. 2001; Salzano & Bortolini 2002; Marrero *et al*. 2005). The results presented here indicate that this fact is also true considering the Gaucho.

Recent investigation showed that the Brazilian Y-chromosomes are almost indistinguishable from those of Portuguese men (Carvalho-Silva *et al*. 2006). Our Y-SNP/STR data globally suggest, however, that the Gaucho males have more similarity with the Spaniards than with the Portuguese. The history of Rio Grande do Sul is peculiar because, in the Colonial Era, the political control of the region alternated between the Spanish and Portuguese Empires (Flores 2003). These historical events can be associated to our findings, but some caution is needed since differentiation between Iberian Peninsula populations, as well as between them and their derived Latin American populations, at the Y-chromosome level, was not observed in other investigations (De Souza Góes *et al*. 2005; Martin *et al*. 2004; Carvalho-Silva *et al*. 2006).

The female counterpart derived from the asymmetrical unions that occurred in the Colonial Period has mostly Amerindian and African mitochondrial genomes (Alves-Silva *et al*. 2000), independently of the demographic, cultural and other changes that occurred afterwards (Salzano & Bortolini 2002). The notable finding related to the Gaucho, however, is that the mtDNA Amerindian proportion (52%) is similar to those observed in

northern Brazilian/Amazonian populations (54%-64%; Alves-Silva *et al*. 2000; Feio-dos-Santos *et al*. 2006). These results permit to suggest that the contemporary Gaucho have the most important reservoir of Amerindian mtDNA lineages in Brazil outside the Amazonian region.

Hybrid groups can arise from distinct admixture dynamics (Long 1991; Parra *et al*. 2001) and this applies to these two sets. Despite some local interaction between Kaingang/Guarani and the populations that surround their reservations (Marrero *et al*. 2006), extensive admixture between Indians and non-Indians ended in southern Brazil (Pampa included; Flores 2003) at least two centuries ago, whereas in northern Brazil/Amazonia the introduction of Amerindian genes into non-Native urban and rural populations is probably occurring until now. The results of these continuous and non-continuous patterns of gene flow can be observed comparing the mtDNA results mentioned above with the biparental loci admixture values (~13% and ~42% of Amerindian component for samples from the Pampa of Rio Grande do Sul and northern Brazilian/Amazonian region, respectively; Dornelles *et al*. 1999; Salzano & Bortolini 2002).

Another question is the origin of this Amerindian component. As mentioned in the Introduction, according to archeological and historical data the Guarani Indians (basically of the M´biá sub-group) were suggested as more likely contributors than those from other tribes to the formation of the Gaucho people, since they were considered "faithful Indians" (Kern *et al*. 1993; Bracco 2004). The Charrua were " "untrustful" Indians" (Bracco 2004), while the Kaingang inhabited the pine forests located in the North/Northeast of Rio Grande do Sul (Fig. 1), their contact with non-Indians during the colonization process being much more restricted. However, when we considered the mtDNA Amerindian portion, no

connection was detected between the Gaucho and the Guarani, while a strong relationship between the Gaucho and Uruguayans was detected (Figs. 3 and 4). Uruguay is the only South American country where independent Native American populations no longer exist (Bonilla *et al*. 2004). Recent studies, however, revealed that people from this country experienced sex-biased unions during the Colonial Period, similar to those which occurred in other Latin American populations (Bonilla *et al*. 2004; Sans *et al*. 2006). Because the Uruguayan historiography is replete of reports involving Gaucho, Charrua and Guarani persons (Pi Hugarte 1998; Bracco 2004), it was expected that the Amerindian mtDNA portion of contemporary Uruguayans would have mainly Guarani and/or Charrua origin. Our results suggest that the Charrua maternal inheritance may have been the most important.

The relative mtDNA affinity between the Gaucho and aborigines from Tierra del Fuego/Patagonia (Fig. 3) could indicate a relationship between the Pampean tribes with those that inhabited the extreme South of the continent, in accordance with archeological data (Schmitz *et al*. 1997; Pi Hugarte 1998).

Curiously, the Guarani are also statistically distant from the other Rio Grande do Sul populations (Figs. 3 and 4), besides the Gaucho. These results, associated with those observed from the lineage analyses, delineate two possible scenarios not mutually exclusive: One indicates that the present Guarani mitochondrial genome may be a poor representative of that found at colonial times; and the other that the Charrua tribe could have made a more significant contribution, through their women, to the formation of the Gaucho people. After this initial formation, these persons expanded from the Pampa to colonize, with strong success, other regions of southern Brazil as suggested by the cluster of Fig. 3. However, only additional data would provide a more precise identification of

Charrua and Guarani mtDNA lineages, as well as clues about the relative contribution of these tribes to the formation of the Gaucho.

Acknowledgments

REFERENCES

Abella, G. (2000) *Nuestra raiz charrúa: los pueblos originarios de la Banda Oriental y su imagen en los uruguayos de hoy*. BetumSan ediciones, Montevideo.

Alemán, B.E. (1994) *Santa Fe y sus aborigenes – I parte.* Junta Provincial de Estudios Históricos, Santa Fe.

Alves-Silva, J., Santos, M.S., Guimarães, P.E.M., Ferreira, A.C.S., Bandelt, H.-J., Pena, S.D.J. & Prado,V.F. (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* **67**, 444-461.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & Howell N. (1999) Reanalysis and revisions of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **123**, 147.

Bandelt, H-J., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraespecific phylogenies. *Mol Biol Evol* **16**, 37-48.

Bandelt, H-J., Herrnstadt, C., Yao, Y-G., Kong, Q-P., Kivisildt, T., Rengo, C., Scozzari, R., Richards, M., Villems, R., Macaulay, V., Howell, N., Torroni, A. & Zhang, Y-P. (2003) Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *AnnHum Genet* **67**, 512-527.

Bandelt, H-J., Quintana-Murci, L., Salas, A. & Macaulay, V. (2002) The fingerprint of phanton mutations in mitochondrial DNA data. *Am J Hum Genet* **71**, 1150-1160.

Becker, I.I.B. (1997) *O índio Kaingang no Rio Grande do Sul*. UNISINOS, São Leopoldo.

Becker, I.I.B. (2003) *Os índios Charrua e Minuano na antiga Banda Oriental do Uruguay*. UNISINOS, São Leopoldo.

Bianchi, N.O., Catanesi, C.I., Baillet, G., Martínez-Marignac, V.L., Bravi, C.M., Vidal-Roja, L.B., Herrera, R.J. & Lopez-Camelo, J.S. (1998) Characterization of ancestral and derived Y-chromosome haplotypes of New World Native populations. *Am J Hum Genet* **63**, 1862-1871.

Bonilla, C., Bertoni, B., González, S., Cardoso, H., Brum-Zorrilla, N. & Sans, M. (2004) Substantial Native American female contribution to the population of Tacuarembó, Uruguay, reveals past episodes of sex-biased gene flow. *Am J Hum Biol* **16**, 289-297.

Bortolini, M.C., Salzano, F.M., Thomas, M.G., Stuart, S., Nasamem, S.P.K., Bau, C.H.D., Hutz, M.H., Torres, M.M., Groot, M., Michalsky, R., Nymodowa, P., Bedoya, G., Bradman, N., Labuda, D.& Ruiz-Linares, A. (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* **73**, 524-539.

Bortolini, M.C., Silva Júnior, W.A., Castro de Guerra, D., Remonatto, G., Mirandola, R., Hutz, M.H., Weimer, T.A., Silva, M.C.B.O., Zago, M.A. & Salzano, F.M. (1999) African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. *Am J Hum Biol* **11**, 551-563.

Bortolini, M.C., Zago, M.A., Salzano, F.M., Silva Júnior, W.A., Boniato, S.L., Silva, M.C.B.O. & Weimer, T.A. (1997) Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. *Hum Biol* **69**, 141-159.

Bracco, D. (2004) *Charrua, Genoas y Guaraníes. Interacción y destrucción: indígenas en el Rio de la Plata*. Lunardi y Risso, Montevideo.

Bravi, C.M., Sans, M., Baillet, G., Martínez-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.

Brion, M., Salas, A., Gonzáles-Neira, A., Lareu, M.V. & Carracedo, A. (2003) Insights into Iberian population origins through the consturction of highly informative Y-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite. *Am J Phys Anthropol* **122**, 147-161.

Cainé, L., Corte-Real, F., Vieira, D.N., Carvalho, M., Serra, A., Lopes, V. & Vide, M.C. (2005) Allele frequencies and haplotype of Y-chromosomal STRs in Santa Catarina population of southern Brazil. *Forensic Sci Int* **148**, 75-79.

Campbell, L. (1997) *American Indian languages. The historical linguistics of Native America.* Oxford Univesity Press, Oxford.

Carvalho, C.B.M. & Pena, S.D.J. (2005) Optimization of a multiplex minisequencing protocol for population studies and medical genetics. *Genet Mol Research* **4**, 115-125.

Carvalho-Silva, D.R., Santos F.R., Rocha, J. & Pena, S.D.J. (2001) The Phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet* **68**, 281-286.

Carvalho-Silva, D.R., Tarazona-Santos, E., Rocha, J., Pena, S.D.J. & Santos, F.R. (2006) Y chromosome diversity in Brazilian: switching perspectives from slow to fast evolving markers. *Genetica* **126**, 251-260.

Christensen, T.N.S. (2001) *História do Rio Grande do Sul em suas origens missioneiras*. Editora Unijuí, Ijuí.

Crow, J.F. & Kimura, O. (1970) *An introduction to population genetics theory*. Harper and Row, New York.

Darwin, C. (1996) *The voyage of the Beagle*. Penguin Books, Harmondsworth.

De Souza Góes, A.C., Carvalho, E.F., Gomes, I., Silva, D.A., Gil, E.H.F., Amorim, A. & Gusmão, L. (2005) Population and mutational analysis in 17 Y-STR loci from Rio de Janeiro (Brazil). *Int J Legal Med* **119,** 70-76.

Dornelles, C.L., Callegari-Jacques, S.M., Robinson, W.M., Weimer, T.A., Franco, M.H.L.P., Hickmann, A.C., Geiger, C.J. & Salzano, F.M. (1999) Genetics, surnames, grandparents' nationalities and ethnic admixture in southern Brazil – do the patterns of variation coincide?. *Genet Mol Biol* **22**, 151-161.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Feio-dos-Santos, A.C., Carvalho, B.M., Santos, S.E.B. & Ribeiro-dos-Santos, A.K. (2006) Nucleotide variability of HV-I in admixed population of the Brazilian Amazon Region. *Forens Sci Int (in press)*.

Flores, M. (2003) *História do Rio Grande do Sul*. Ediplat, Porto Alegre.

García-Bour, J., Pérez-Peres, A., Álvarez, S., Fernández, E., López-Parra, A.M., Arroyo-Pardo, E. & Turbón, D. (2004) Early population differentiation in extinct aborígenes from Tierra del Fuego – Patagonia: ancient mtDNA sequences and Y-chromosome STR characterization. *Am J Phys Anthropol* **123**, 361-370.

Gascue, C., Mimbacas, A., Sans, M., Gallino, J.P., Bertoni, B., Hidalgo, P.C. & Cardoso, H. (2005) Frequencies of the four major Amerindian mtDNA haplogroups in the population of Montevideo, Uruguay. *Hum Biol* **77**, 873-878.

Grattapaglia, D., Kalupnieck, S., Guimarães, C.S., Ribeiro, M.A., Diener, P.S. & Soares, C.N. (2005) Y-chromosomal STR haplotype diversity in Brazilian populations. *Forensic Sci Int* **149**, 99-107.

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.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**, 95-98.

Hünemeier, T. (2006) Filogeografia dos cromossomo Y e das linhagens mitocondriais de origem africana em populações negras brasileiras. Tese de mestrado, Porto Alegre: Universidade Federal do Rio Grande do Sul.

Hunley, K. & Long, J.C. (2005) Gene flow across linguistic boundaries in Native North American populations. *Proc Natl Acad Sci USA* **102**, 1312-1317.

Jobling, M. & Tyler-Smith, C. (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* **4**, 598-612.

Kern, A.A., Bernd, Z., Cohen, V.R.A., Avancini, E.G., Kemel, C.L.A. & Gick, P.W. (1993) *Rio Grande do Sul: continente múltiplo.* Riocell Marprom, Porto Alegre.

Kivisild, T., Tolk, H.V., Parik, J., Wang, Y., Papiha, S.S., Bandelt, H-J. & Villems, R. (2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol* **19**, 1737-1751.

Lahiri, D.K. & Nurnberger, J.I. (1991) A rapid non-enzimatic method for preparation of HMW DNA from blood for RFLP studies. *Nucl Acids Res* **19**, 5444.

Lalueza-Fox, C. (1996) Mitochondrial DNA haplogroups in four tribes from Tierra del Fuego – Patagonia: inferences about the peopling of the Americas. *Hum Biol* **68**, 853-871.

Leal, O.F. (1989) Gauchos: male culture and identity in the Pampas. Ph.D. Thesis, University of California, Berkeley.

Lecerf, M., Filali, M., Grésenguet, G., Ndjoyi-Mbiguino, A., Le Goff, J., Mazancourt, P. & Bélec, L. (2006) Allele frequencies and haplotypes of eight Y-short tandem repeats in Bantu population living in Central Africa. *Forensic Sci Int (in press).*

Long, J.C. (1991) The genetic structure of admixed populations. *Genetics* **127**, 417-428.

Marrero, A.R., Leite, F.P.N., Carvalho, B.A., Peres, L.M., Kommers, T.C., Cruz, I.M., Salzano, F.M., Ruiz-Linares, A., Silva Junior, W.A. & Bortolini, M.C. (2005) Heterogeneity of the genome ancestry of individuals classified as white in the state of Rio Grande do Sul, Brazil. *Am J Hum Biol* **17**, 496-506.

Marrero, A.R., Silva Junior, W.A., Bravi, C.M., Hutz, M.H., Petzl-Erler, M.L., Ruiz-Linares, A., Salzano, F.M. & Bortolini, M.C. (2006) The demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. *Am J Phys Anthropol (in press).*

Martin, P., Garcia-Hirschfeld, J., Garcia, O., Gusmão, L., Garcia, P., Albarran, C., Sancho, M. & Alonso, A. (2004) A Spanish population study of 17 Y-chromosome STR loci. *Forensic Sci Int* **139**, 231-235.

Mentz-Ribeiro, P.A. (1997) Os caçadores pampeanos e a arte rupestre. In: *Arqueologia pré-histórica do Rio Grande do Sul* (ed. Jacoby, R.), pp. 103-134. Mercado Aberto, Porto Alegre.

Moraga, M.L., Rocco, P., Miquel, J.F., Nervi, F., Llop, E., Chakraborty, R., Rothhammer, F. & Carvalho, P. (2000) Mitochondrial DNA polymorphisms in Chilean aboriginal population: implications for the peopling of Southern Cone of the continent. *Am J Phys Anthropol* **113**, 19-29.

Nichols, M.W. (1946) *O gaúcho: caçador de gado – cavaleiro – ideal de romance*. Livraria Editora Zélio Valverde, Rio de Janeiro.

Oliven, R.G. (1993) Identidade nacional e etnicidade: diversidade étnica no Rio Grande do Sul. In: *Encontros com a Antropologia* (eds. Costa, M.C.S. & M.T. Pino), pp. 23-38, SESC, Curitiba.

Pagano, S., Sans, M., Pimeroff, V., Cantera, A.M., Alvarez, J.C., Lorente, J.A., Peco, J.M., Mones, P. & Sajantila, A. (2005) Assessment of HV1 and HV2 mtDNA variation to forensic purposes in an Uruguayan population sample. *J Forensic Sci* **50**, 1-4.

Palha, T.J.B.F., Rodríguez, E.M.R. & Santos, S.E.B. (2006) Y-chromosomal STR haplotypes in a population from the Amazon region, Brazil. *Forensic Sci Int (in press)*.

Parra, E.J., Kittles, R.A., Argyropoulos, G., Pfaff, C.L., Hiester, K., Bonilla, C., Silvestre, N., Parrish-Gause, D., Garvey, W.T., Jin, L., McKeigue, P.M., Kamboh, M.I., Ferrel, R.E., Pollitzer, W.S. & Shriver, M.D. (2001) Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol* **114**, 18-29.

Pereira, L., Gusmão, L., Alves, C., Amorim, A. & Prata, M.J. (2002) Bantu and European Y-lineages in Sub-Saharan África. *Ann Hum Genet* **66**, 369-378.

Pi Hugarte, R. (1998) *Los índios del Uruguay*. Ediciones de la Banda Oriental, Montevideo.

Reverbel, C. (2002) *O gaúcho*. L&PM, Porto Alegre.

Roewer, L., Croucher, P.J.P., Willuweit, S., Lu, T.T., Kayser, M., Leissig, R., De Kniff, P., Jobling, M.A., Tyler-Smith, C. & Krawczak, M. (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet* **116**, 279-291.

Salas, A., Richards, M., De la Fe, T., Lareu, M.V., Sobrino, B., Sanchez-Diaz, P., Macaulay, V. & Carracedo, A. 2002. The making of the African mtDNA landscape. *Am J Hum Genet* **71**, 1082-1111.

Salas, A., Richards, M., Lareu, M.V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V. & Carracedo, A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet* **74**, 454-465.

Salzano, F.M. & Bortolini, M.C. (2002) *Evolution and genetics of Latin American populations*. Cambridge: Cambridge University Press.

Sans, M., Merriwether, D.A., Hidalgo, P.C., Bentancor, N., Weimer, T.A., Franco, M.H.L.P., Alvarez, I., Kemp, B.M. & Salzano, F.M. (2006) Population structure and admixture in Cerro Largo, Uruguay, based on blood markers and mitochondrial DNA polymorphisms. *Am J Hum Biol* **18**, 513-524.

Schmitz, P.I. (1997)Migrantes da Amazônia: a tradição Tupi-Guarani. In: *Arqueologia Pré-histórica do Rio Grande do Sul* (ed. Jacoby, R.), pp. 295-330. Mercado Aberto, Porto Alegre.

Silva, D.A., Carvalho, E., Costa, G., Tavares, L., Amorim, A. & Gusmão, L. (2006) Y-chromosome genetic variation in Rio de Janeiro population. *Am J Hum Biol* **18**, 829-837.

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

Torres, M.M., Bravi, C.M., Bortolini, M.C., Duque, C., Callegari-Jacques, S.M., Ortiz, D., Bedoya, G., Respetro, H.G. & Ruiz-Linares, A. (2006) A revertant of the major founder Native American haplogroup C common in populations from Northern South America. *Am J Hum Biol* **18**, 59-65.

Torroni, A., Achilli, A., Macaulay, V., Richards, M. & Bandelt, H-J. (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet* **22**, 339-345.

Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang,W. H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Qasim Mehdi, S., Seielstad, M. T., Wells, S., Piazza, A., Davis, R. W., Feldman, M. W., L. Cavalli-Sforza, L. & Oefner, P. J. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* **26**, 358–361.

Vidart, D. (2000) *La trama de la identidad nacional. Tomo I: indios, negros, gauchos*. Ediciones de La Banda Oriental, Montevideo.

Weimer, G. (1991) *O trabalho escravo no Rio Grande do Sul*. Editora da UFRGS/SAGRA, Porto Alegre.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* **28**, 114-138.

Yao, Y.-G., Bravi, C.M. & Bandelt, H.-J. (2004). A call for mtDNA data quality control in forensic science. *Forens Sci Int* **141**, 1-6.

Table 1 Y-chromosome haplogroup distributions (%) in the Gaucho sample and in other admixed and parental populations

Population/Region	Y-SNP haplogroup							
	Q3a	Q3*(xQ3a)	Q*(xQ3)	P*	K*(xP)	Y*	DE*(xE3a)	E3a*
Admixed								
Brazil								
North (49)[1]				58	2	16	24	
Northeast (49)[1]				71			27	2
Southeast (177)[1,2]		1		55	4	18	22	>1
South (52)[1]				52		15	33	
Rio Grande do Sul								
Gaucho (150)[3]		5		58	2	32		3
Other regions (75)[4]				73		23	4	
Native								
Europe								
Portuguese (93)[1]				68		8	23	1
Spanish (84)[5]				59	7	24	10	
Africa (312)[6]				2	1	32	14	52
America (390)[6]	9	77	6	4		2	1	1

Note: Data compiled from [1]Carvalho-Silva *et al.*, 2006; [2]Silva *et al*., 2006; [3]Present study; [4]Marrero *et al*., 2005; [5]Brion *et al*., 2003 (minus Basques), [6] Bortolini *et al*., 2003. Sample sizes in parentheses.

Table 2 Y-SNP haplogroup and Y-STR haplotypes observed in the Gaucho sample

Y-SNP Haplogroup		DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS393	DYS439	DYS438	DYS437	DYS392	DYS385a/b	n
						Y-STR Haplotype							
K* (xP)													
	H1	14	13	29	23	11	13	12	12	14	13	11,15	1
	H2	14	13	29	24	11	13	12	12	15	13	11,14	1
	H3	15	14	31	24	11	14	12	13	15	13	11,14	1
P* (xQ)													
	H4	13	13	29	24	9	13	10	10	14	11	13,15	1
	H5	15	13	30	24	9	13	11	11	15	13	13,14	1
	H6	16	13	30	25	10	13	10	11	14	11	11,14	2
	H7	15	13	29	24	10	13	12	12	15	12	11,15	1
	H8	14	13	29	24	10	13	13	12	14	13	11,14	1
	H9	14	13	29	24	10	13	12	12	14	15	11,14	1
	H10	15	12	28	23	10	13	11	10	16	11	13,17	2
	H11	13	13	29	24	10	13	13	12	14	13	12,14	1
	H12	13	13	29	24	10	13	11	10	14	11	17,18	1
	H13	15	13	30	21	10	14	12	11	15	10	17,18	1
	H14	14	13	29	24	10	13	11	12	14	13	11,15	2
	H15	14	13	29	23	10	13	12	12	16	13	11,13	1
	H16	14	13	29	25	10	14	11	12	15	11	12,14	1
	H17	14	13	30	24	10	13	12	12	15	13	11,15	1
	H18	14	13	28	24	10	13	12	12	15	14	11,14	1
	H19	14	14	30	24	10	12	12	12	14	13	11,15	1
	H20	15	13	30	24	10	14	12	12	14	13	11,15	1
	H21	14	13	29	24	10	13	11	12	15	13	11,14	1
	H22	14	12	27	24	10	12	12	12	14	14	11,14	1
	H23	14	13	29	23	10	13	13	9	15	11	13,17	1
	H24	14	13	27	24	11	13	11	12	15	13	11,16	1
	H25	14	13	29	23	11	13	13	12	15	13	12,14	1
	H26	14	13	29	24	11	13	13	12	15	13	11,14	1
	H27	15	13	29	23	11	13	12	12	15	13	11,14	1
	H28	14	13	29	24	11	13	11	12	15	13	11,14	1
	H29	15	13	29	24	11	13	12	12	15	13	11,14	2
	H30	14	14	30	25	11	14	12	13	15	13	11,14	1
	H31	14	14	30	24	11	13	12	12	15	13	11,14	1
	H32	14	13	29	23	11	13	12	12	14	13	11,14	1
	H33	15	13	29	23	11	14	12	13	14	13	11,15	1
	H34	15	13	31	24	11	13	12	12	15	13	11,14	1
	H35	16	13	30	23	11	13	13	12	15	13	11,14	1
	H36	14	13	29	25	11	13	13	12	16	13	11,13	1

H37	13	13	30	25	11	13	12	12	15	13	11,14	1
H38	14	13	29	24	11	13	12	12	15	14	11,14	1
H39	14	13	30	25	11	13	12	11	15	13	11,14	1
H40	14	13	29	24	11	13	13	12	14	13	11,14	1
H41	14	13	29	24	11	13	12	12	15	13	14,17	1
H42	14	14	30	24	11	14	12	13	15	13	11,14	1
H43	14	13	29	23	11	13	13	12	15	13	11,14	1
H44	14	13	29	23	11	12	12	12	15	13	11,15	1
H45	14	13	29	24	12	13	11	12	15	13	11,12	1

Q3* (xQ3a)

H46	14	13	29	24	11	13	14	12	14	13	11,14	1

Y*

H47	13	14	30	24	6	13	10	10	14	11	13,14	1
H48	15	13	29	24	9	12	13	9	14	11	13,17	1
H49	14	13	28	24	10	13	11	12	14	13	11,16	1
H50	15	13	30	21	10	13	12	11	14	11	15,18	1
H51	11	12	28	22	10	13	11	10	16	11	13,14	1
H52	15	13	29	24	10	13	12	12	15	13	11,14	1
H53	15	12	29	24	10	13	12	9	16	11	13,17	1
H54	18	12	26	23	10	13	12	10	15	11	12,14	1
H55	16	13	29	25	10	13	11	10	14	11	11,14	1
H56	13	13	30	24	10	12	12	10	14	11	16,18	1
H57	14	13	29	23	10	13	11	12	14	13	11,14	1
H58	13	13	30	24	10	14	12	10	14	11	16,18	1
H59	14	14	31	23	10	12	11	10	14	11	14,18	1
H60	15	12	29	22	10	13	11	10	16	11	14,18	1
H61	14	14	33	24	10	13	13	10	14	11	11,14	1
H62	13	13	30	24	10	12	12	10	14	11	16,17	1
H63	15	13	32	21	10	13	12	11	14	11	16,18	1
H64	17	13	28	23	10	14	12	10	15	11	12,14	1
H65	15	13	31	22	10	14	10	10	16	11	14,15	1
H66	14	13	29	24	10	12	12	9	15	11	14,15	1
H67	15	14	31	23	10	14	11	9	14	12	14,15	1
H68	14	13	29	24	10	13	11	11	15	13	11,14	1
H69	15	14	31	21	10	15	14	11	14	11	15,19	1
H70	14	13	29	23	10	12	11	10	14	11	14,17	1
H71	14	12	28	22	10	13	11	10	16	11	13,14	1
H72	13	13	31	24	10	12	12	10	14	11	16,18	1
H73	15	13	29	22	10	13	12	9	14	11	13,14	1
H74	14	13	30	23	11	12	10	9	15	11	13,15	1
H75	14	14	30	24	11	15	12	13	15	13	11,14	1
H76	14	13	30	24	11	13	12	12	15	13	11,14	**3**
H77	13	13	30	24	11	13	12	12	16	14	11,15	1
H78	14	13	30	23	11	12	11	10	14	11	13,19	1
H79	15	14	29	25	11	11	12	10	14	11	10,12	1
H80	14	13	29	24	11	13	11	11	15	13	11,14	1
H81	16	13	31	23	11	13	12	10	15	11	13,16	1
H82	14	13	29	23	11	13	12	12	14	13	11,15	1
H83	14	13	29	24	11	13	12	12	15	13	11,14	1

Table 3 HVS-I sequence variation and major continental-specific mtDNA haplogroups observed in 106

Gaucho

Sequence	N	Variable sites	Haplogroup
Gaucho 01	7	111 223 290 319 362	A2
Gaucho 02	1	111 189 223 290 319 362	A2
Gaucho 03	2	111 183C 189 223 290 319 362	A2
Gaucho 04	1	111 182C 183C 189 223 290 319 362	A2
Gaucho 05	2	111 209 223 290 319 362	A2
Gaucho 06	3	126 223 278 290 319 362	A2
Gaucho 07	3	189 217	B2
Gaucho 08	2	183C 189 217	B2
Gaucho 09	7	182C 183C 189 217	B2
Gaucho 10	1	182C 183C 189 217 241	B2
Gaucho 11	1	189 217 249 312	B2
Gaucho 12	2	183C 189 217 311 319	B2
Gaucho 13	1	173 182C 183C 189 217 223	B2
Gaucho 14	9	223 298 325 327	C1
Gaucho 15	1	223 325 327	C1
Gaucho 16	1	223 256 298 325 327	C1
Gaucho 17	1	223 295 298 325 327	C1
Gaucho 18	1	126 223 298 325 327	C1
Gaucho 19	1	126 270 298 325 327	C1
Gaucho 20	1	209 223 234 298 325 327	C1
Gaucho 21	1	051 172 223 298 325 327	C1
Gaucho 22	1	051 184 223 287 298 311 325 327	C1
Gaucho 23	3	223 325 362	D1
Gaucho 24	1	142 145 179 223 295 325 362	D1
Gaucho 25	16	rCRS	H / HV* / U* / R*
Gaucho 26	1	162	H
Gaucho 27	1	248	H
Gaucho 28	1	093 129 316	H
Gaucho 29	1	182C 183C 189 357	H?
Gaucho 30	1	069 126 193 278	J2b
Gaucho 31	2	224 311	K
Gaucho 32	1	168 224 311	K1a4
Gaucho 33	1	168 224 311 320	K1
Gaucho 34	1	126 234 248 292 294	T3
Gaucho 35	1	126 153 182C 183C 189 294 296	T5 or T2?
Gaucho 36	1	126 129 294 296 304	T2b
Gaucho 37	2	126 294 296 304	T2b
Gaucho 38	1	189 249 311	U1a
Gaucho 39	1	051 129C 209 260 362	U2e
Gaucho 40	1	189 197 209 356	U4
Gaucho 41	1	192 256 270	U5a1
Gaucho 42	1	172 183C 189 219 278 311 362	U6a1

Gaucho 43	5	298	H / HV0
Gaucho 44	1	129 148 172 187 188G 189 223 230 311 320	L0a*
Gaucho 45	2	126 187 189 223 264 270 278 311	L1b
Gaucho 46	1	129 187 189 223 278 293 294 311 360	L1c1
Gaucho 47	1	093 129 183C 187 189 223 265C 278 286G 294 311 360	L1c2
Gaucho 48	1	086 223 278 294 309 390	L2a1 β1
Gaucho 49	1	124 223 278 311 362	L3b
Gaucho 50	1	124 223 319	L3d1
Gaucho 51	1	223 320	L3e1b
Gaucho 52	1	172 183C 189 223 320	L3e2b
Gaucho 53	1	172 189 223 320	L3e2b
Gaucho 54	1	051 223 264	L3e4

Table 4 Parental contributions (%) in admixed populations from southern Brazil and Uruguay considering the mitochondrial DNA variation

Population/Region	Amerindian	African	European	Reference
Brazil				
North (48)	54	15	31	Alves-Silva *et al.* (2000)
Northeast (50)	22	44	34	Alves-Silva *et al.* (2000)
Southeast (99)	33	34	31	Alves-Silva *et al.* (2000)
South (50)	22	12	66	Alves-Silva *et al.* (2000)
Rio Grande do Sul				
Gaucho (105)	52	11	37	Present study
Other regions (106)	11	3	86	Marrero *et al.* (2005)
Uruguay				
Montevideo (115)	37	2	62	Pagano *et al. (*2005)
Tacuarembó (19)	79	5	16	Bonilla *et al,* (2004)
Cerro Largo (43)	30	21	49	Sans *et al,* (2006)

Table 5 Amerindian mtDNA haplogroup distributions (%) in South American Admixed and Native populations

	Haplogroup				
Population/Region	A	B	C	D	References[1]
Admixed					
Brazil					
North (27)	15	31	39	15	Alves-Silva *et al.* (2000)
Northeast (11)	37	27	9	27	Alves-Silva *et al.* (2000)
Southeast (34)	40	30	18	12	Alves-Silva *et al.* (2000)
South (10)	27	27	27	19	Alves-Silva *et al.* (2000)
Rio Grande do Sul					
Gaucho (54)	30	31	30	9	Present study
Other regions (13)	8	25	59	8	Marrero *et al.* (2005)
Uruguay					
Tacuarembó (62)	21	34	32	13	Bonilla *et al.* (2004)
Montevideo (52)	19	38	26	17	Pagano *et al.* (2005); Gascue *et al.* (2005)
Melo (13)	31	38	31	0	Bravi *et al.* (1997)
Cerro Largo (80)	22	38	32	8	Sans *et al.* (2006)
Native					
Guarani (200)	84	0	9	6	Marrero *et al.* (2006)
Kaingang (78)	47	4	49	0	Marrero *et al.* (2006)
Mapuche (111)[2]	0	7	44	49	Moraga *et al.* (2000)
Pehuenche (105)[2]	3	10	41	46	Moraga *et al.* (2000)
Yaghan (21)[2]	0	0	48	52	Moraga *et al.* (2000)
Aonikenk (5)[3]	0	0	75	25	Garcia-Bour *et al.* (2004); Lalueza-Fox *et al.* (1996)
Kaweskar (11)[3]	0	0	36	64	Garcia-Bour *et al.* (2004)
Selknam (3)[3]	0	0	100	0	Garcia-Bour *et al.* (2004)
Yamana (7)[3]	0	0	71	29	Garcia-Bour *et al.* (2004); Lalueza-Fox *et al.* (1996)

Note:[1]Data obtained using sequencing and RFLP methods. [2]Patagonian tribes. [3]Fuegian tribes. Sample sizes in parentheses

Table 6 Nucleotide diversity values for the Amerindian portion of the mitochondrial DNA in South American Admixed and Native populations.

Population/Region	Nucleotide diversity (π)	References
Admixed		
Brazil		
South (10)	0.0261 ± 0.0147	Alves-Silva *et al*. (2000)
North (27)	0.0209 ± 0.0114	Alves-Silva *et al*. (2000)
Northeast (11)	0.0204 ± 0.0118	Alves-Silva *et al*. (2000)
Southeast (34)	0.0223 ± 0.0119	Alves-Silva *et al*. (2000)
Rio Grande do Sul		
Gaucho (54)	0.0142 ± 0.0076	Present study
Other regions (11)	0.0178 ± 0.0102	Marrero *et al*. (2005)
Uruguay (64)	0.0177± 0.0270	Bonilla *et al*. (2004); Pagano *et al*. (2005); Sans *et al*. (2006)
Native		
Guarani (200)	0.0067 ± 0.0040	Marrero *et al*. (2006)
Kaingang (74)	0.0171 ± 0.0090	Marrero *et al*. (2006)
Patagonian (73)	0.0102 ± 0.0058	Moraga *et al*. (2000)
Fuegian (24)*	0.0129 ± 0.0069	Garcia-Bour *et al*. (2004)

Note: The Fuegian sample was obtained from ancient DNA.

Figure 1 Geographic location of the Pampa region, Brazil. The map are not drawn to scale.

Figure 2 Multidimensional scaling plot using Y-chromosome STR data based on pairwise F_{ST} values, showing the relationships among the Gaucho (triangle), Amerindian (stars), European (squares), African (diamond) and other Brazilian (circles) groups. The stress value for the MDS plot is 0.0744.

Figure 3 Multidimensional scaling plot using pairwise F_{ST} values based on Amerindian mtDNA sequence variation, showing relationships among the Gaucho (triangle), other admixed populations (circles) and Amerindians (stars). The stress value for the MDS plot is 0.0348.

Figure 4 Population relationships with the best fit for the Amerindian mitochondrial sequence data set. Letters **a** and **b** indicate two clearly defined branches.

Figure 5 Haplogroup A medium network. The root haplotype is identified by *1 (16111T-16223T-16290T-16319A-16362C) and is represented by only 7 Gaucho sequences. Variant positions from the root are indicated as numbers. Circles are not proportional to lineage frequencies. The matches do not take into account the 182C and 183C mutations. #2 – 4 Gaucho (present study); 1 Admixed southeastern Brazilian (Alves Silva *et al*., 2000); 1 Admixed Uruguayan (Bonilla *et al*., 2004); 1 Admixed USA Hispanic (FBI Database). #3 – 2 Gaucho (present study); 1 Admixed southeastern Brazilian and 1 Admixed southern Brazilian (Alves Silva *et al*., 2000); 1 Trinitario (Bolivia; Bert *et al*., 2004); 8 Guarani Ñandeva and 42 Guarani Kaiowá (Brazil; Marrero *et al*., 2006); 2 Admixed USA Hispanic (FBI Database). #4 - 3 Pehuenche (Chile; Moraga *et al*., 2000). #5 - 11 Guahibo (Venezuela; Vona *et al*., 2005). #6 - 3 Gaucho (present study); 1 Kraho (Brazil; Torroni *et al*., 1993); 2 Admixed southeastern Brazilian (Alves Silva *et al*., 2000) 1 Txukahamãe (Brazil; Dornelles *et al*., 2005); 2 Kaingang (Brazil; Marrero *et al*., 2006); 1 Admixed Afro-Brazilian (Silva Jr *et al*., 2006). #7 – 1 Txukahamãe (Brazil; Dornelles *et al*., 2005).

Figure 6 Haplogroup B medium network of the haplogroup B. The root haplotype is identified by *1 (16189C-16217T), represented by 12 Gaucho sequences. Variant positions from the root are indicated as numbers. Circles are not proportional to lineage frequencies. The matches do not take into account the 182C

and 183C mutations. #2 – 1 Gaucho (present study); 1 Amazon Amerind (Santos *et al.*, 1996); 1 Admixed northeastern Brazilian and 1 Admixed southeastern Brazilian (Alves Silva *et al.*, 2000). #3 – 9 Xavante (Brazil; Ward *et al.*, 1996) #4 – 1 Admixed Afro-Amazonian (Ribeiro dos Santos *et al.*, 2006). #5 – 1 Admixed Uruguayan (Sans *et al.*, 2006). #6 – 1 Gaucho (present study). #7 – 1 Admixed Rio Grande do Sul (Marrero *et al.*, 2005); 1 Kubenkokre (Brazil; Santos *et al.*, 1996). #8 – 1 Arara (Brazil; Santos *et al.*, 1996; Ribeiro dos Santos *et al.*, 2001). #9 - 1 Arara (Brazil; Santos *et al.*, 1996; Ribeiro dos Santos *et al.*, 2001); 1 Sambaqui Pirabas (Ribeiro dos Santos *et al.*, 1996); 1 Admixed northeastern Brazilian (Alves Silva *et al.*, 2000). #10 – 1 Admixed Dominican (Tajima *et al.*, 2004); 1 Admixed USA Hispanic (FBI Database). #11 – 1 Admixed North Brazilian (Alves Silva *et al.*, 2000). #12 – 1 Kubenkokre (Brazil; Santos *et al.*, 1996); #13 – 3 Kaingang (Brazil; Marrero *et al.*, 2006); 4 Kuna (Panamá; Batista *et al.*, 1995); 2 Admixed USA Hispanic (FBI Database); 1 Maya Quiché (Guatemala; Boles *et al.*, 1995). #14 – 1 Yanomame (Venezuela; Williams *et al.*, 2002). #15 – 2 Embera (Panamá; Kolman *et al.*, 1997). #16 – 2 Gaucho (present study). #17 – 1 Admixed Rio Grande do Sul (Marrero *et al.*, 2005). #18 – 1 Admixed Uruguayan (Pagano *et al.*, 2005). #19 – 1 Yuracare (Bolivia; Bert *et al.*, 2004). #20 – 1 Gaucho (present study); 2 Admixed Afro-Brazilian (Ribeiro dos Santos *et al.*, 2002). #21 - 1 Admixed Afro-Brazilian (Ribeiro dos Santos *et al.*, 2002). #22 - 1 Admixed Afro-Brazilian (Ribeiro dos Santos *et al.*, 2002); 1 Admixed Uruguayan (Pagano *et al.*, 2005). #23 – 1 Admixed Afro-Brazilian (Ribeiro dos Santos *et al.*, 2002). #24 - 1 Admixed Afro-Brazilian (Ribeiro dos Santos *et al.*, 2002). #25 - 1 Admixed Uruguayan (Pagano *et al.*, 2005). #26 – 1 Pehuenche (Chile; Moraga *et al.*, 2000); #27 - 1 Pehuenche (Chile; Moraga *et al.*, 2000). #28 - 1 Pehuenche and 5 Huilliche (Chile; Moraga *et al.*, 2000). #29 – 3 Huilliche (Chile; Moraga *et al.*, 2000) #30 - 1 Pehuenche (Chile; Moraga *et al.*, 2000)

Figure 7 Haplogroup C medium network. The root haplotype is identified by *1 (16223T-16298C-16325C-16327T-16362C), represented by 9 Gaucho sequences. Variant positions from the root are indicated as numbers. Circles are not proportional to the lineage frequencies. The matches do not take into account the 182C and 183C mutations. #2 – 1 Gaucho (present study); 1 Fuegian (Argentina; Garcia-Bour *et al.*, 2004); 1 Admixed southeastern Brazilian (Alves Silva *et al.*, 2000); 1 Admixed Uruguayan (Bonilla et a., 2004); 1 ancient Ciboney (Cuba; Lalueza-Fox *et al.*, 2003); 2 Guarani Kaiowá (Brazil; Marrero *et al.*, 2006); 1 Admixed Rio Grande do Sul (Marrero *et al.*, 2005); 9 Guarani Ñandeva (Brazil; Marrero *et al.*, 2006); 1

Admixed Brazilian (Feio dos Santos *et al*., 2006); 4 Admixed USA Hispanic (FBI Database). #3 – 1 Admixed Afro-Amazonian (Ribeiro dos Santos *et al*., 2006). #4 – 8 Guarani Kaiowá (Brazil; Marrero *et al*., 2006). #5 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004). #6 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004). #7 – 1 Admixed Rio Grande do Sul (Marrero *et al*., 2005). #8 – 1 Admixed Brazilian (Feio dos Santos *et al*., 2006). #9 – 1 Gaucho (present study). #10 – 1 Zoró (Brazil; Ward *et al*., 1996). #11 – 1 Gaucho (present study); 5 Admixed USA Hispanic (FBI Database). #12 – 2 Admixed USA Hispanic (FBI Database). #13 – 1 Admixed Afro-Amazonian (Ribeiro dos Santos *et al*., 2006). #14 – 1 Gaucho (present study). #15 – 1 Pehuenche (Chile; Moraga *et al*., 2000). #16 – 1 Yaghan (Chile; Moraga *et al*., 2000). #17 – 2 Yaghan (Chile; Moraga *et al*., 2000). #18 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004). #19 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004). #20 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004). #21 – 1 Gaucho (present study); 1 Taino (Dominican Republic; Lalueza-Fox *et al*., 2001); 1 Admixed northern Brazilian (Alves Silva *et al*., 2000); 1 Admixed Uruguayan (Bonilla *et al*., 2004); 1 Admixed Rio Grande do Sul (Marrero *et al*., 2005); 1 Admixed Afro-Brazilian (Silva Jr *et al*., 2006); 3 Admixed USA Hispanic (FBI Database). #22 – 1 Admixed northern Brazilian (Alves Silva *et al*., 2000). #23 – 1 Admixed Brazilian (Feio dos Santos *et al*., 2006). #25 – 1 Admixed Afro-Amazonian (Ribeiro dos Santos *et al*., 2006); 1 Admixed Afro-Brazilian (Bortolini *et al*., 1997). #26 – 1 Admixed Afro-Amazonian (Ribeiro dos Santos *et al*., 2006). #27 – 1 Admixed USA Hispanic (FBI Database). #28 – 3 Parakanã (Brazil; Dornelles *et al*., 2005). #29 – 1 Gaucho (present study). #30 – 1 Ciboney (Cuba; Lalueza-Fox *et al*., 2003). #31 – 1 Gaucho (present study). #32 – 19 Kaingang (Brazil; Marrero *et al*., 2006). #33 – 1 Admixed USA Hispanic (FBI Database). #34 – 1 Admixed Afro-Brazilian (Hünemeier, 2006). #35 – 1 Admixed Uruguayan (Pagano *et al*., 2005). #36 – Vaimaca Peru (http://www.fhuce.edu.uy/antrop/cursos/abiol/ifvaimaca.pdf).

Figure 8 Haplogroup D medium network. The root haplotype is identified by *1 (16223T-16325C-16362C), represented by 3 Gaucho sequences. Variant positions from the root are indicated as numbers. The letters after some mutations indicate a transversion. Circles are not proportional to the lineage frequencies. The matches do not take into account the 182C and 183C mutations. #2 – 3 Piaroa (Venezuela; Ghose *et al*., 2002). #3 - 1 Piaroa (Venezuela; Ghose *et al*., 2002). #4 – 1 Gaucho (present study). #5 – 1 Huitoto (Colombia; Torres *et al*., 2006). #6 – 1 Admixed Afro-Amazonian (Ribeiro dos Santos *et al*., 2006). #7 – 1 Admixed Colombian (Horai *et al*., 1993). #8 - 1 Admixed USA Hispanic (FBI Database). #9 – 1 Tiriyó

(Brazil; Santos *et al*., 1996). #10 - 1 Tiriyó (Brazil; Santos *et al*., 1996); 1 Admixed northern Brazilian (Alves Silva *et al*., 2000); 1 Admixed Venezuelan (Ghose *et al*., 2002); 2 Admixed Uruguayan (Bonilla *et al*., 2004; Pagano *et al*., 2005); 1 Admixed USA Hispanic (FBI Database); 4 Admixed Afro-Brazilian (Ribeiro dos Santos *et al*., 2002); 5 Admixed Afro-Amazonian (Ribeiro dos Santos *et al*., 2006). #11 – 2 Katuena (Brazil; Santos *et al*., 1996); 1 Admixed Afro-Amazonian (Ribeiro dos Santos *et al*., 2006). #12 – 1 Admixed Afro-Brazilian (Ribeiro dos Santos *et al*., 2002). #13 – 3 Waiãpi (Brazil; Santos *et al*., 1996); 1 ancient Amazonian (Ribeiro dos Santos *et al*., 1996). #14 – 1 Tiriyó (Brazil; Santos *et al*., 1996). #15 – 1 Admixed Rio Grande do Sul (Marrero *et al*., 2005); 1 Afro-Brazilian (Ribeiro dos Santos *et al*., 2002). #16 - 1 Afro-Brazilian (Ribeiro dos Santos *et al*., 2002). #17 – 2 Huilliche (Chile; Moraga *et al*., 2000). #18 – 4 Yaghan (Chile; Moraga *et al*., 2000). #19 – 1 Yaghan (Chile; Moraga *et al*., 2000). #20 – 2 Pehuenche (Chile; Moraga *et al*., 2000); 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004). #21 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004); #22 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004). #23 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004). #24 – 1 Pehuenche (Chile; Moraga *et al*., 2000). #25 – 3 Huilliche (Chile; Moraga *et al*., 2000). #26 – 1 Huilliche (Chile; Moraga *et al*., 2000). #27 - 3 Huilliche (Chile; Moraga *et al*., 2000). #28 – 2 Huilliche (Chile; Moraga *et al*., 2000). #29 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004). #30 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004). #31 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004).#32 - 1 Fuegian (Argentina; Garcia-Bour *et al*., 2004).
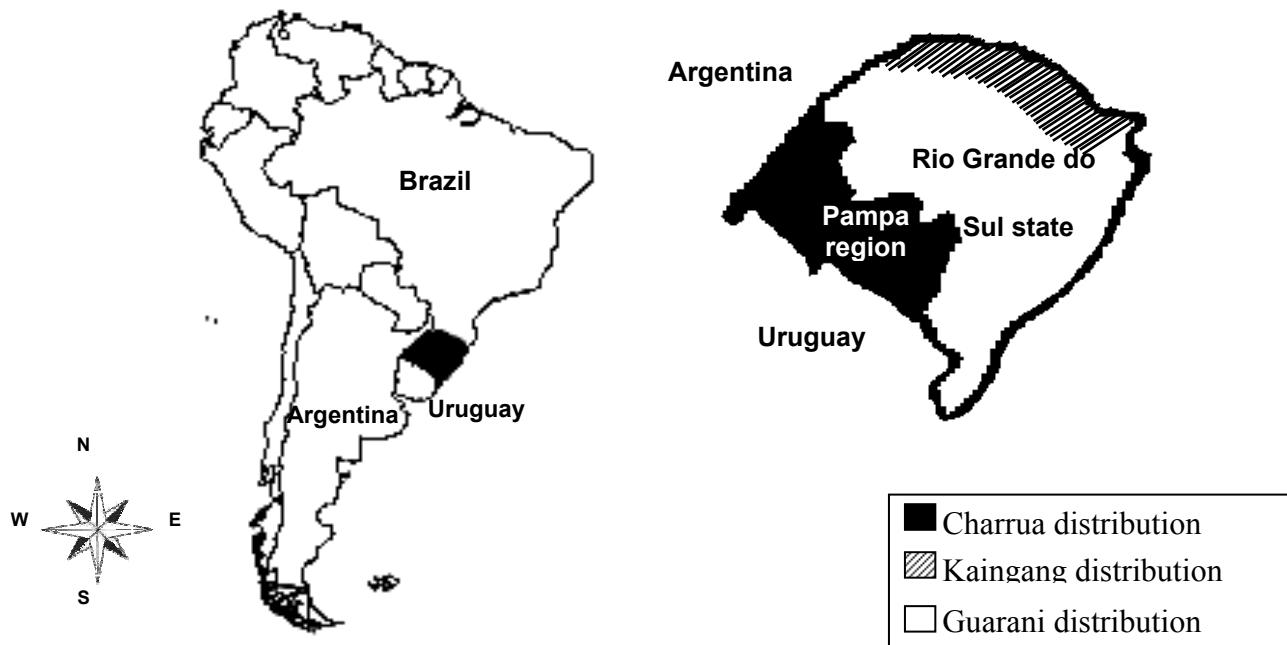
Figure 1
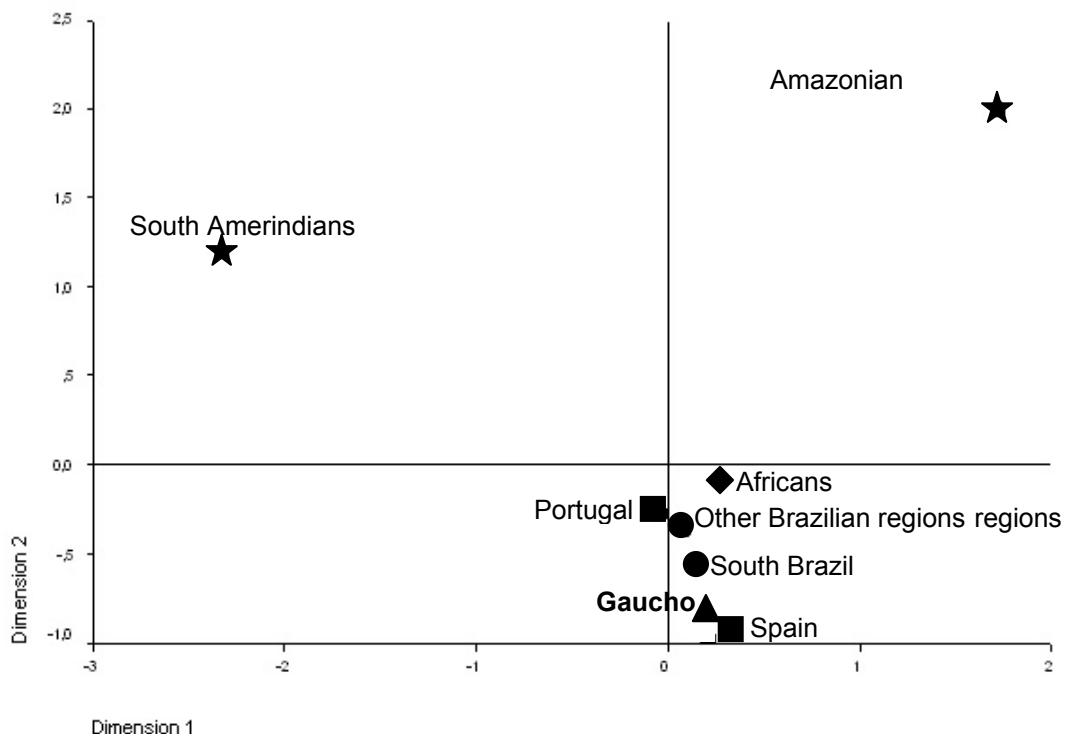
Figure 2
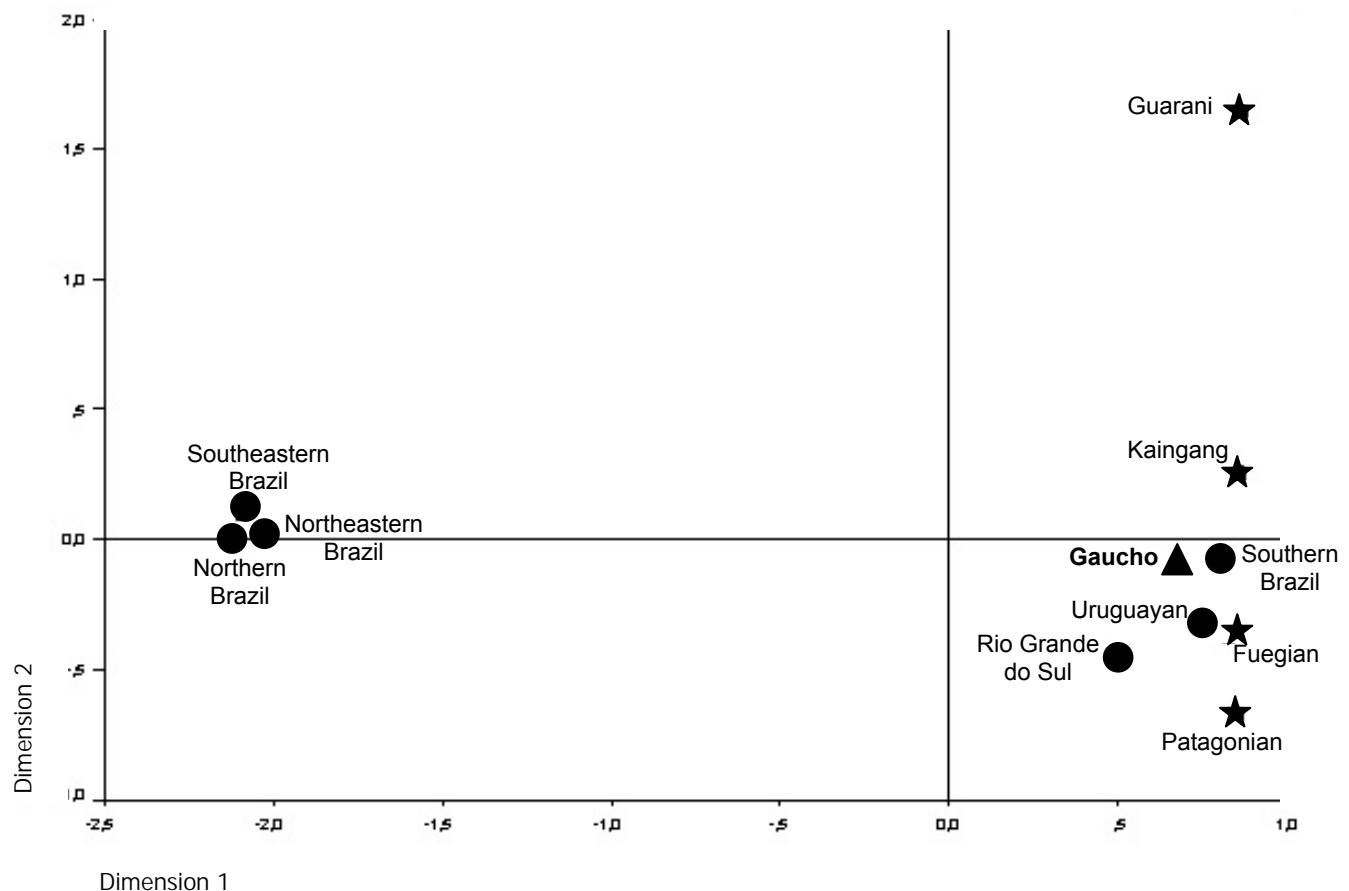
Figure 3
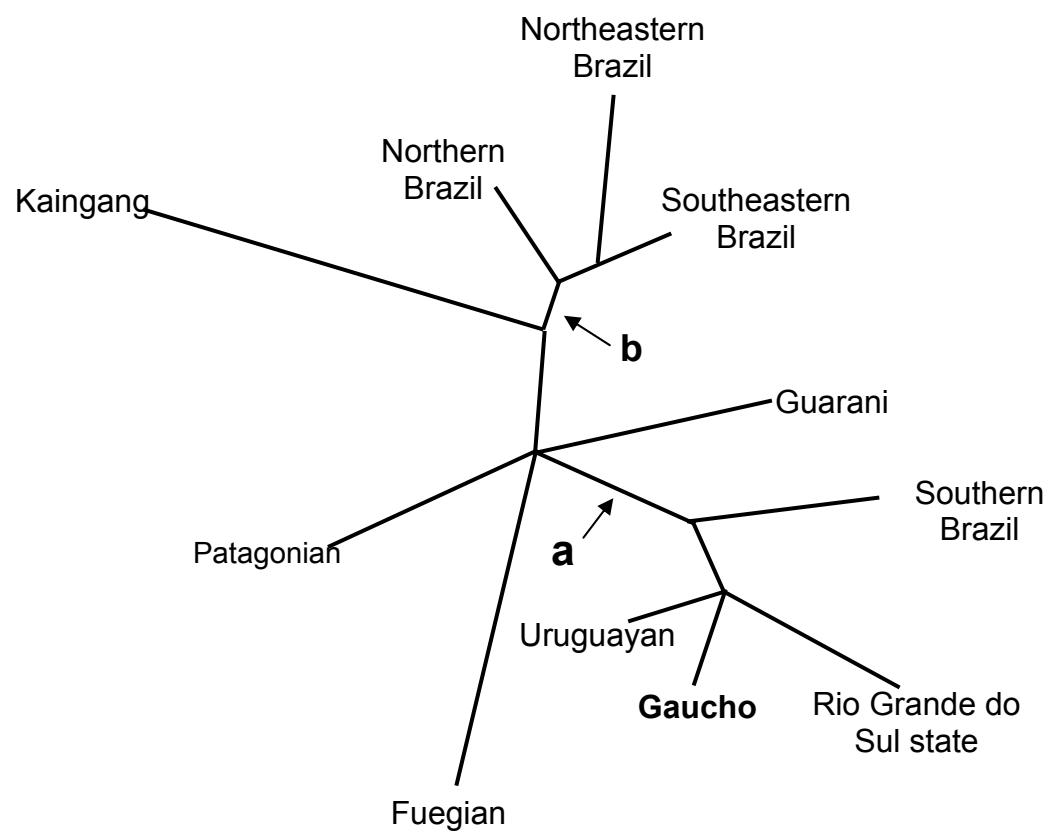
Figure 4
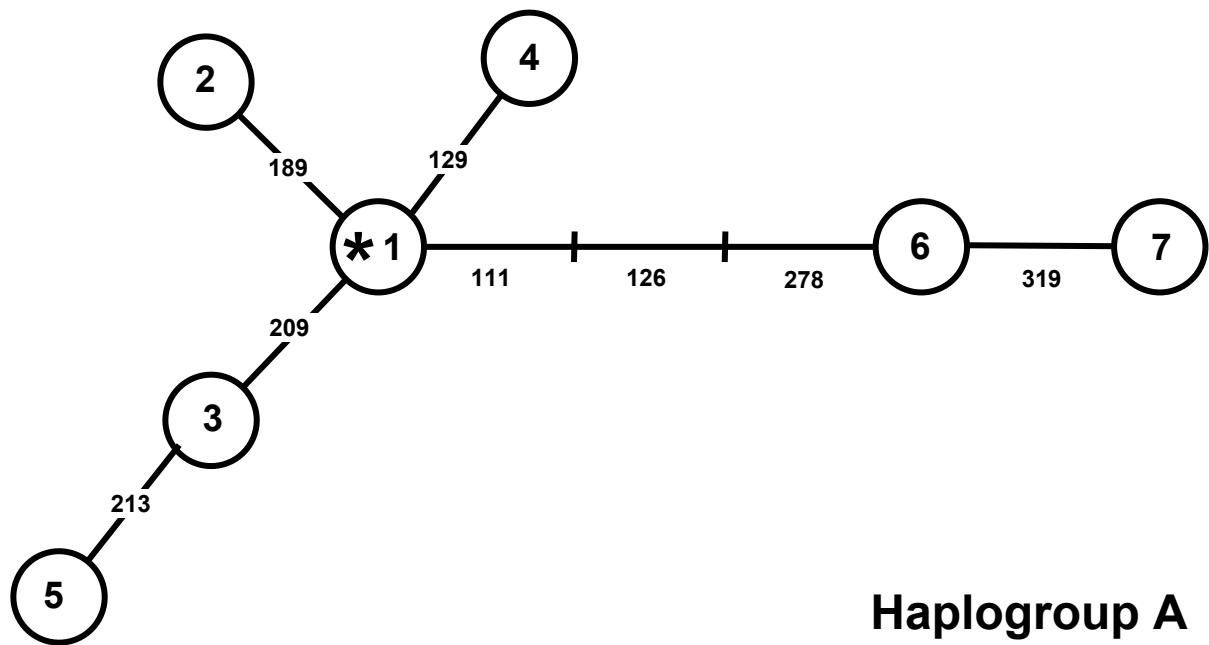
Haplogroup A
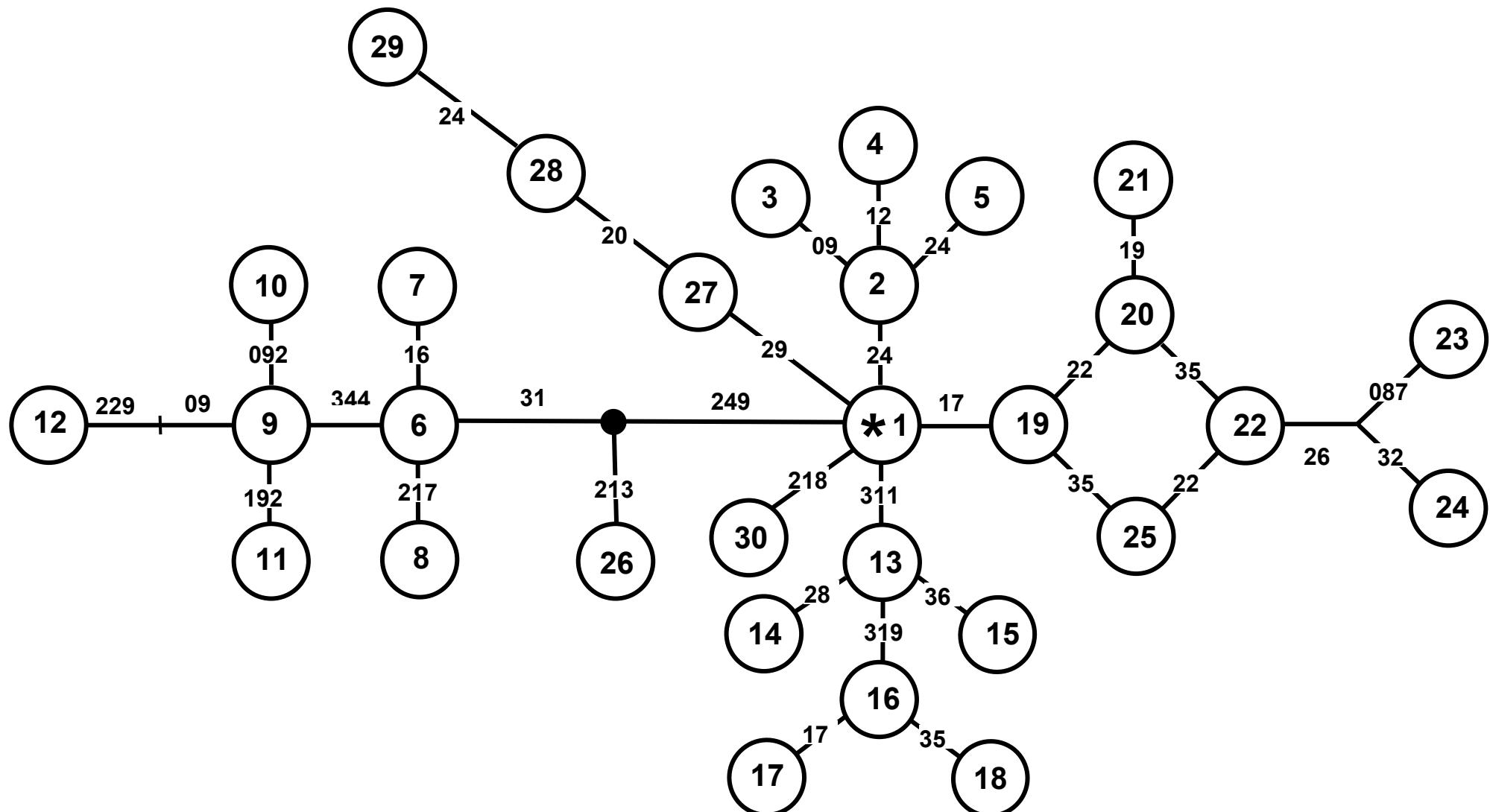
Figure 5

Figure 6

Haplogroup B

Figure 7
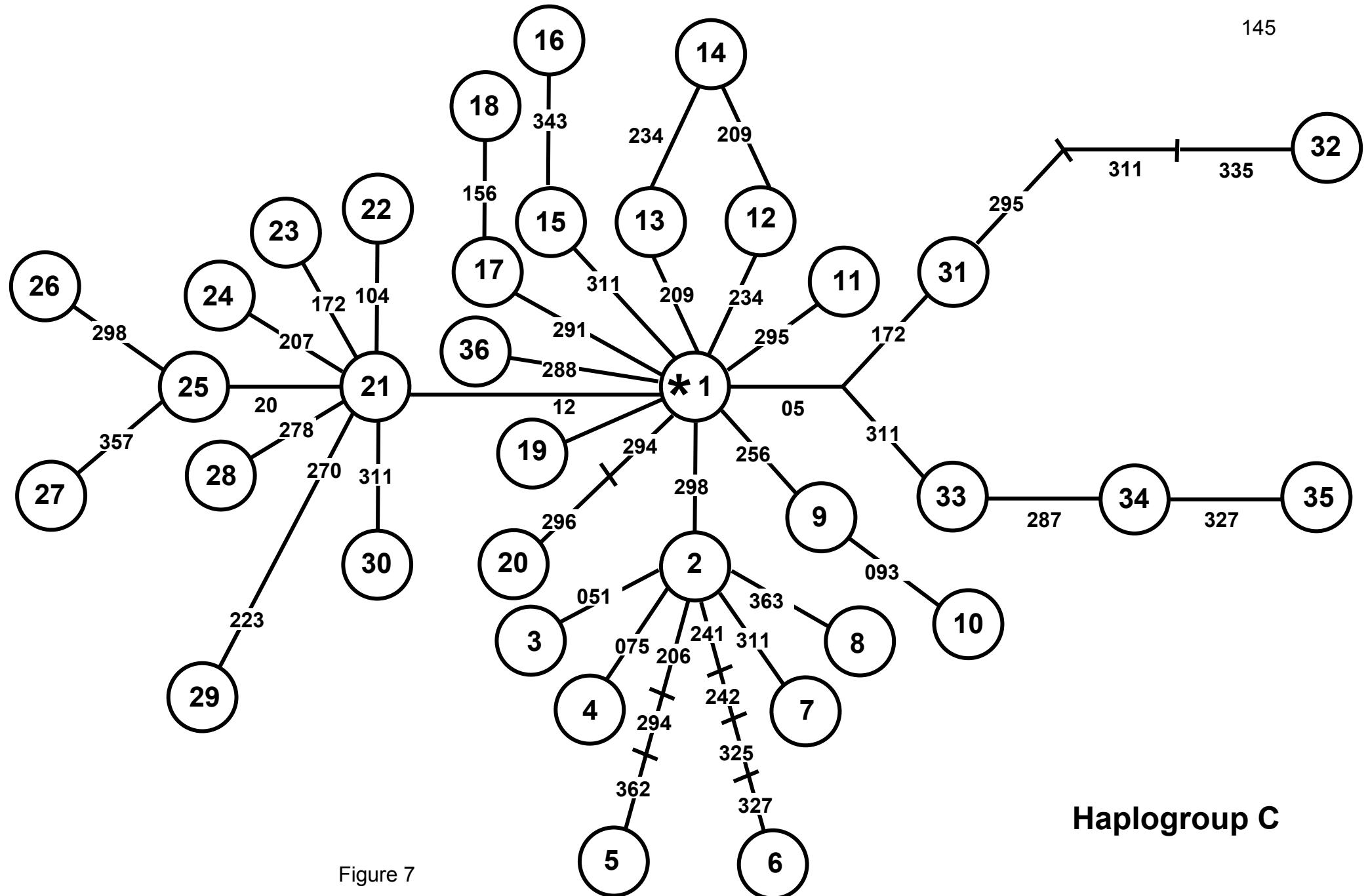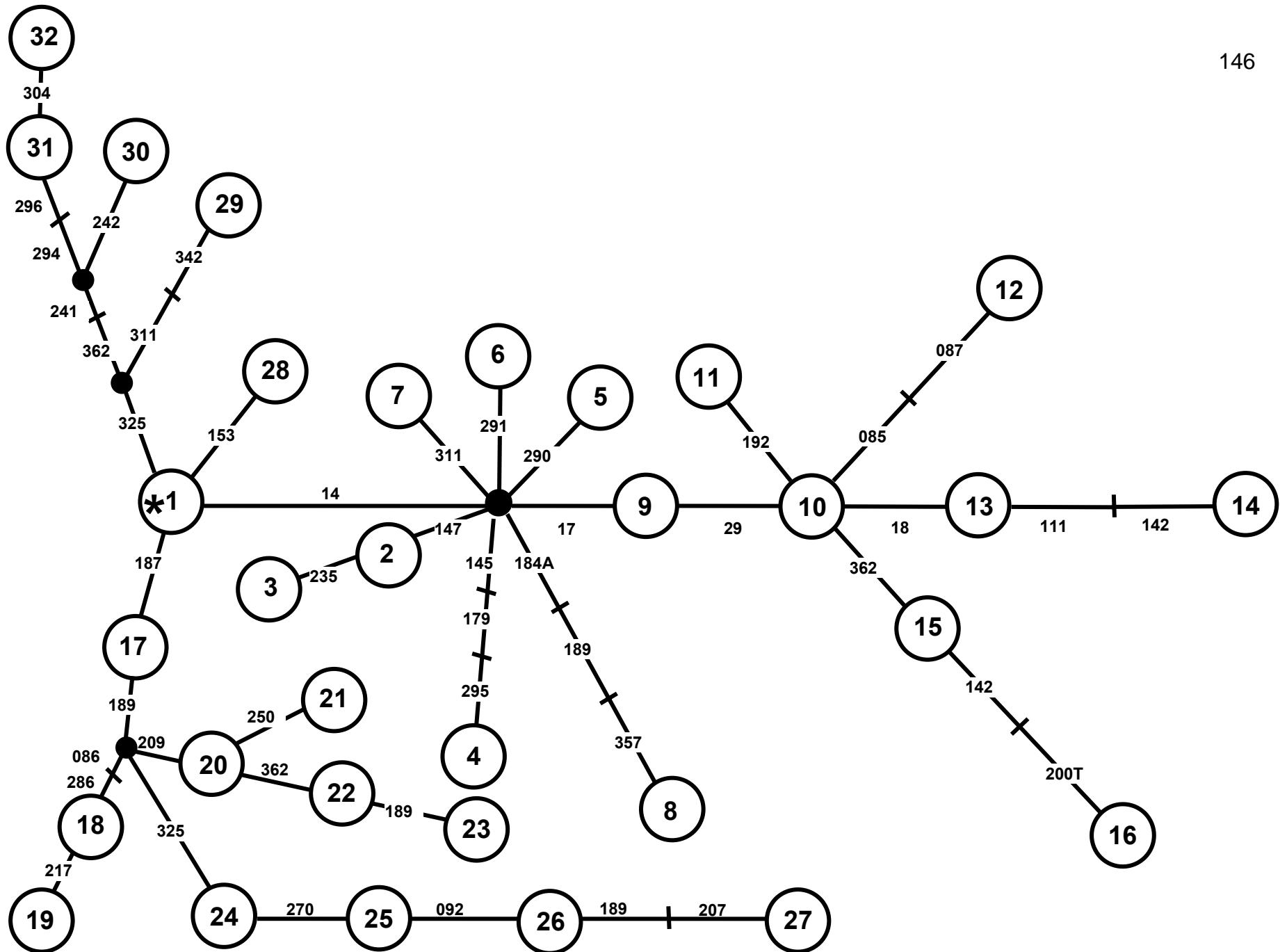
Haplogroup C

Figure 8

Haplogroup D

CAPÍTULO VI

Discussão

Os artigos apresentados nos capítulos 2, 3, 4 e 5 descrevem e discutem em detalhes os resultados obtidos neste trabalho. Desta forma, este capítulo tem a intenção de apenas agrupar as informações e discutir de maneira mais convergente os resultados apresentados em cada artigo em particular. Entretanto, alguma repetição e sobreposição de idéias é inevitável. Além disso, para que o texto flua de forma mais prazerosa para a leitura, buscou-se evitar repetir citações que, exaustivamente, já foram feitas em cada um dos artigos anteriormente apresentados.

Este trabalho envolveu 547 indivíduos, sendo 269 provenientes de populações híbridas do Rio Grande do Sul e 278 nativos americanos das tribos Guarani e Kaingang. A investigação envolveu diversos sistemas genéticos, de herança uniparental materna: mtDNA (N = 503); paterna: Y-SNPs (N = 291), Y-STRs (N = 89); e biparental: X-STRs (N = 70). Com isso buscou-se desvendar a história evolutiva e demográfica das populações aqui investigadas.

O Rio Grande do Sul tem uma história de colonização tardia quando comparado a outros Estados brasileiros. Além disso, sempre houve alternância entre o domínio do império português e espanhol sobre a região (Flores, 2003). Os escravos africanos também chegaram mais tardiamente, só no final do século XVIII, com o crescimento da indústria do charque (Flores, 2003). A partir do século XIX, aportou no Estado uma grande quantidade de imigrantes europeus (especialmente italianos e alemães) maior que aquela registrada para outros Estados. Foram atraídos pelas terras e vieram para substituir o então abolido trabalho escravo, estes imigrantes se fixaram nos vales e montanhas do Estado.

Apesar de ser considerado um Estado de influência essencialmente européia, diversos estudos realizados vêm mostrando que existem perfis genéticos diferentes quando se leva em consideração as distintas regiões do Rio Grande do Sul. Por exemplo, Leite

(2006) estudando marcadores autossômicos, dos cromossomos X e Y, verificou que existe uma marcada diferenciação entre a região centro-leste e oeste do Estado, sendo que os níveis de ancestralidade ameríndia são maiores nesta última, onde se localiza o Pampa rio-grandense.

O primeiro artigo apresenta resultados do estudo com 119 indivíduos fenotipicamente identificados como brancos, oriundos de várias regiões do Rio Grande do Sul. A amostras foi investigada com relação ao mtDNA (seqüenciamento da HVS-I), e a porção masculina da mesma (N = 74) foi também tipada para Y-SNPs. A amostra total foi dividida em duas: Serra, representada por indivíduos da cidade de Veranópolis, que tem uma forte influência da imigração italiana no século XIX e RS-Geral com indivíduos provenientes de diversas cidades do Rio Grande do Sul. Se por um lado as matrilinhagens e patrilinhagens identificadas na Serra apontam para uma quase completa ancestralidade européia, no RS Geral enquanto 100% dos cromossomos Y são europeus, significantes frações de genomas mitocondriais de origem ameríndia (36%) e africana (16%) foram encontradas. Embora o tamanho da amostra RS-Geral seja pequeno, estes resultados indicaram o clássico padrão assimétrico de cruzamentos observado em populações brasileiras miscigenadas, onde a introgressão ameríndia e africana ocorreu preferencialmente através das mulheres. A particularidade deste estudo, entretanto, foi quantificar a marcante diferença na ancestralidade genômica que pode existir entre amostras igualmente identificada como "brancas" dentro de um Estado brasileiro. Tal diferença, que certamente está relacionada com as distintas histórias de colonização, chama a atenção para a necessidade de critérios na seleção das populações usadas em determinados tipos de estudos: enquanto a amostra do RS-Geral representa uma população notavelmente heterogênea, a Serra comporta-se como um caso de genoma europeu

transplantado. Além disso, fica evidente que em algumas circunstâncias cor e outros traços fenotípicos são pobres indicadores de ancestralidade, como já havia sido postulado anteriormente (Parra *et al*., 2003), mas em outras indicam, com boa precisão, a herança geográfica ancestral da população derivada que está sob investigação. Uma boa correlação entre cor e ancestralidade também pode ser encontrada no trabalho de Vargas *et al*. (2006; anexo1), que só reforça a cautela que se deve ter ao se fazer afirmativas que dizem respeito a características gerais sobre as populações brasileiras.

O segundo artigo apresentado discute especificamente duas populações nativas de interesse para este estudo como um todo: Guarani e Kaingang. Estas são as duas únicas tribos indígenas que ainda existem no Estado, uma vez que os Charrua são considerados extintos. As amostras tiveram a HVS-I sequenciada e a fração masculina desta foi também genotipada para marcadores bialélicos localizados no cromossomo Y. Este estudo foi direcionado para verificar diferenças entre Guarani e Kaingang, seja na dinâmica de mestiçagem com populações vizinhas, seja com relação a identidades com grupos indígenas relacionados (outros Tupi e Jê). Os Guarani, que representam a porção mais austral da notável e bem sucedida expansão Tupi, apresentam 85%, 9% e 6% das linhagens mitocondriais associadas aos haplogrupos ameríndio A, C e D, respectivamente. A baixa variabilidade intrapopulacional observada nos Guarani, bem como a ausência do haplogrupo B sugere que eles possam ter experimentado um *bottleneck* na migração que protagonizaram a partir da Amazônia. Foi sugerido que os Guarani contemporâneos poderiam não ser representativos daqueles da época do contato com colonizadores, sugestão esta que foi corroborada com estudo posterior com a amostra de Gaúchos da região do Pampa. O valor do coeficiente de diferença interpopulacional (G_{ST}) obtido com dados de mtDNA foi cerca de cinco vezes maior para Guarani que para Kaingang,

sugerindo um alto nível de diferenciação entre as três parcialidades (Ñandeva, Kaiowá e M'Byá). Este achado permitiu inferir que a separação destes três sub-grupos Guarani foi um evento anterior ao contato com colonizadores europeus e escravos africanos. Já a separação entre os subgrupos Kaingang estaria associada a eventos mais recentes (ver item I.1.4).

Ambas tribos mostraram importantes graus de mistura com não-ameríndios, mas foi somente com este estudo que foi possível verificar que esta é influenciada pelo gênero; nos Guarani mistura com não-índios só chegou através da linhagem paterna. Além disso, diante do fato das parcialidades estarem separadas desde tempos pré-contato, a introdução de cromossomos Y não-ameríndios ocorreu de forma independente nos Ñandeva, Kaiowá e M'Byá. Já nos Kaingang, mistura tanto com mulheres quanto com homens não ameríndios foi detectada.

O terceiro manuscrito (em preparação) apresenta os haplótipos encontrados para 70 Gaúchos do Pampa, com relação aos 16 X-STRs analisados. Trata-se de um conjunto de marcadores recentemente identificados (*Panel* 28). Desta forma, poucas publicações com eles puderam ser encontradas, inviabilizando comparações pertinentes. Para se ter uma idéia, no banco de dados do CEPH (*Centrè d'Etude du Polymorphisme Humain*) existem até o momento 65 homens tipados com estes X-STRs, sendo 30 chineses, 10 quenianos, 9 basco-franceses, 5 paquistaneses e outros 11 cuja procedência não foi indicada. Como nenhuma destas populações teve importante contribuição na formação da população gaúcha, comparações não foram feitas. Desta forma, optou-se por apresentar os dados de maneira descritiva. O manuscrito será ricamente acrescido quando forem publicados os resultados para várias outras populações Latino-americanas e ancestrais, num estudo colaborativo que está sendo coordenado pelo Dr. Andrés Ruiz-Linares, da *University*

College of London. Vale ressaltar que estudos com marcadores do cromossomo X, de forma geral, têm se mostrado muito úteis para aplicações forenses e pesquisas populacionais (Pereira e Pena, 2006). É importante ressaltar que nestes 70 indivíduos tipados, nenhum haplótipo foi encontrado mais de uma vez, o que indica uma grande diversidade no Pampa.

Finalmente, no quarto artigo, 150 Gaúchos do Pampa foram investigados. Os resultados dos artigos anteriores, particularmente aqueles que descrevem a população geral do RS, bem como aquele que caracterizou pela primeira vez as tribos Guarani e Kaingang quanto a um conjunto de marcadores paternos e maternos, foram de particular importância para que o cenário da história evolutiva e demográfica do Gaúcho típico fosse delineado. Observou-se por exemplo, que o cromossomo Y, P*(xQ) era o mais comum (58%) indicando a origem européia dos mesmos. Entretanto, 5% dos cromossomos são do haplogrupo Q3*(xQ3a), o que indica que uma pequena contribuição ameríndia veio pelo lado paterno. O mesmo não havia sido encontrado em amostras de outras regiões do Estado, indicando aqui a primeira particularidade de populações do Pampa. Por outro lado, a presença do cromossomo E3a*, típico de povos do sul do Saara também foi observado com freqüência de 3%, presença marcadamente inferior aquela encontra em uma amostra de indivíduos identificados como negros de Porto Alegre e região metropolitana, onde a presença de E3a* foi de 25% (Hünemmeir *et al.*, 2006; anexo1). Os estudos com Y-SNPs foram refinados com o uso de Y-STRs e novas comparações puderam ser feitas, as quais apontaram para a significativa contribuição ibérica para a formação da população masculina de região, e talvez para o Estado como um todo. Entretanto, diferentemente do que foi visto para outras populações brasileiras, parece que a contribuição espanhola foi

marcante, o que está de acordo com os dados históricos, que apontam para a "mobilidade" das fronteiras no extremo do Império Português.

Combinando metodologias de seqüenciamento, análises de RFLP e mini-seqüenciamento foi possível identificar 106 seqüências de mtDNA, que confirmaram que o padrão clássico de cruzamentos assimétricos também esteve presente na formação do Gaúcho típico. A presença de linhagens mitocondriais de origem européia e africanas foi de 38% e 11% respectivamente. Os genomas mitocondriais de origem africana (L0a*, L1b, L1c1, L1c2, L2a1β1, L3b, L3d1, L3e1b, L3e2b e L3e4) mostram heterogeneidade como aquelas encontradas em outras populações brasileiras identificadas como negros, estudadas por Silva *et al*., 2006 (Ribeirão Preto – São Paulo; Cametá, Trombetas - ambas no Pará; anexo 2) e Hünemeier *et al*., 2006 (Porto Alegre e região metropolitana - Rio Grande do Sul; anexo 3). O haplogrupos L3e4, por outro lado, presentes nos Gaúchos, não foi encontrado em nenhuma destas populações brasileiras investigadas nos dois estudos citados, incluindo a outra amostra do Rio Grande do Sul. L3e4 tem maiores freqüências em populações de Camarões (Hünemeier *et al*., 2006). Somente adicionais estudos, entretanto, poderão discriminar melhor se este achado representa alguma particularidade dos escravos que chegaram ou Pampa ou se está relacionada a desvios devido a amostragem.

Entretanto, o notável nos resultados do estudo com o mtDNA na amostra do Pampa referem-se a expressiva contribuição materna ameríndia (51%) encontrada, das quais 31% pertencem ao haplogrupo B e outros 31% ao C. Nível tão elevado de contribuição materna nativa só foi anteriormente detectado em populações do norte do país, o que faz do Pampa o mais extraordinário reservatório de linhagens ameríndias fora da Amazônia. Ao comparar as distribuições de haplogrupos mitocondriais nos Guarani, considerados historicamente como os nativos amerindíos que tiveram a maior contribuição

para a formação das populações gaúchas de modo geral, verifica-se que além da ausência do B, o C não é o mais frequente, inclusive na parcialidade M'Byá. Apesar do haplogrupo C estar presente nos Kaingang (49%), fontes históricas e arqueológicas mostram que este grupo indígena nunca esteve no Pampa, sendo sua contribuição para a formação das populações gaúchas, pouco relevante. Além disso, dados arqueológicos revelaram que os sítios da Tradição Tupi-Guarani não aparecem na área da distribuição Charrua. De acordo com Schmitz (1997) onde se encontram bolas de pedra (boleadeiras) não se localizam restos ceramistas e vice-versa.

Também se observou uma alta variabilidade de linhagens do mtDNA ameríndias encontradas no Pampa, o que contrasta com a baixa variabilidade mitocondrial dos Guarani.

Análises considerando apenas a porção ameríndia das seqüências foram também utilizadas para averiguar relações populacionais. Uma notável diferenciação entre os Gaúchos e os Guarani é encontrada. Em contrapartida identidade é encontrada com uruguaios, outros rio-grandenses, bem como com Patagones e Fueguinos. Este conjunto de resultados permite inferir que a região possa ser um importante reservatório de mtDNA de outros indígenas, particularmente dos Charrua (ver item I.1.2). Diversos estudos históricos e arqueológicos relacionam os grupos Charrua com Patagones e Fueguinos.

Entretanto, diferenciação entre os Guarani e outras populações do RS também foi observada. Adicionalmente se buscou identificar a relação entre cada linhagem ameríndia encontrada. As chamadas *networks* claramente mostram a presença Guarani, mas outras linhagens não tiverem sua origem definida.

Conjuntamente, estes resultados associados os elementos e evidências históricas levaram a algumas conclusões, como por exemplo, que o estoque genético mitocondrial

Guarani atual não é um bom representante daquele dos tempos da colonização. Uma boa parte das seqüências mitocondriais Guarani pode ser encontrada nas populações gaúchas não indígenas. A presença Charrua parece, de fato, ir além do legado cultural, embora não se tenha conseguido, com este nível de resolução, identificar seqüências específicas oriundas desta tribo extinta. Além disso, o homem Gaúcho parece ter identidade genética, via linhagem paterna, maior com espanhóis que com portugueses diferentemente do que acontece com outros brasileiros.

Finalmente, por ocasião das coletas realizadas na região do Pampa, foi preenchido um formulário constando, entre outras coisas, a auto-classificação do indivíduo amostrado com relação à sua cor/ ancestralidade. Comparando os dados de mtDNA com essa auto-classificação, verificou-se que dos 14 indivíduos que se auto-classificaram como "mestiços", 10 apresentaram haplogrupos ameríndios, 3 haplogrupos europeus e 1 com haplogrupo mitocondrial característico de africanos. Entre aqueles 28 auto-classificados como "negros", a distribuição dos haplogrupos de origem africana, ameríndia e européia foi de 5, 15 e 8, respectivamente. Dos 45 auto-classificados como "brancos", 20 apresentaram haplogrupos mitocondriais de populações européias, 18 tiveram haplogrupos ameríndios e 7 africanos.

Com relação ao cromossomo Y, nem todos os haplogrupos são continente-específicos, mas considerando o P*(xQ), DE*(xE3a) e Q3*(xQ3a) que são encontrados em maiores freqüências nas populações européias, africanas e ameríndias, respectivamente, constatou-se que daqueles 30 indivíduos auto-classificados como "brancos", onze apresentaram haplogrupos mitocondriais e do cromossomo Y de origem européia, porém catorze tiveram mtDNA ameríndio e cromossomo Y europeu, quatro com mtDNA africano

sendo dois com cromossomo Y europeu e outros dois ameríndios, e um indivíduo apresentou tanto o mtDNA quanto o cromossomo Y de origem ameríndia.

Para os 15 auto-classificados como "negros", sete deles possuem mtDNA, ameríndio sendo 6 com cromossomo Y europeu e um ameríndio. Três tinham mtDNA africano e cromossomo Y europeu e outros cinco com mtDNA e cromossomo Y ambos europeus.

Por outro lado, dos oito que se auto-classificaram como "mestiços", sete deles tinham mtDNA ameríndio; 6 e 1 cromossomo Y europeu e ameríndio, respectivamente; e um indivíduo com mtDNA e cromossomo Y europeu.

Esta compilação reforçar a idéia de que pode existir uma fraca correlação entre cor e ancestralidade no Brasil, como já foi apontado por Parra *et al*. (2003). Porém como mencionado anteriormente, nossos estudos (Marrero *et al*., 2005, Capítulo II; Vargas *et al*., 2006, Anexo 1) aqui apresentados também chamam a atenção que esta afirmativa não pode ser extrapolada para todas as populações brasileiras. No caso daquelas do Rio Grande do Sul, uma marcante heterogeneidade genômica entre regiões foi encontrada.

CAPÍTULO VII

Referências Bibliográficas

Alemán BE (1994) Santa Fe y sus Aborigenes – I parte. Junta Provincial de Estudios Históricos, Santa Fe, 282 pp.

Anderson S, Bankier AT, Barrel BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon JC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Sataden R and Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457 – 465.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM and Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23: 147.

Assis Brasil LA (1996) A cidade cercada de pampa. In Continente Sul/Sur. Revista do Instituto Estadual do Livro, 1: 23-26.

Barbosa Lessa LC (1961) Os Guaxos. Francisco Alves, Rio de Janeiro, 349 pp.

Barbosa Lessa LC (1978) Rodeio dos Ventos. Mercado Aberto, Porto Alegre, 203 pp.

Becker IIB (1995) O índio Kaingang no Rio Grande do Sul. UNISINOS, São Leopoldo, 334 pp.

Becker IIB (1999) O índio Kaingang no Paraná. Subsídios para uma etno-história. UNISINOS, São Leopoldo, 344 pp.

Becker IIB (2003) Os Índios Charrua e Minuano na Antiga Banda Oriental do Uruguay. Editora da UNISINOS, São Leopoldo, 248 pp.

Bortolini MC, Salzano FM, Bau C, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro de Guerra D, Bedoya G and Ruiz-Linares A (2002) Y chromosome biallelic polymorphisms and Native America population structure. Ann Hum Genet 66: 255 – 259.

Bortolini MC, Salzano FM, Thomas MG, Nasanem SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro de Guerra D, Torres MM, Groot H, Michalski R, Nimadawa P, Bedoya G, Bradman N, Labuda D and Ruiz-Linares A (2003) Y chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73: 524 – 539.

Bracco D (2005) Charrúas, Guenoas y Guaraníes. Interacción y destrucción: indígenas en el Río de la Plata. Lunardi y Risso, Montevidéu, 398 pp.

Bravi CM (2005) Análisis de lineajes maternos en poblaciones indígenas Americanas. Tese de doutorado. Programa de Pós Graduação em Genética. La Plata, Universidad Nacional de La Plata.

Campbell L (1997) The historical linguistic of Native America. American Indian languages. Oxford University Press: Oxford, 481 pp.

Carneiro da Cunha M (1992) História dos índios no Brasil. Companhia das Letras, São Paulo, 611 pp.

Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimar N and Ruiz-Linares A (2003) Genetic demography of Antioquia (Colombia) and the central valley of Costa Rica. Hum Genet 112: 534 – 541.

Carvalho-Silva DR, Tarazona-Santos E, Rocha J, Pena SDJ and Santos FR (2006) Y chromosome diversity in Brazilians: switching perspectives from slow to fast evolving markers. Genetica 126: 251 – 260.

Derengoski, PR (2002) A saga dos Guarani. Insular, Florianópolis, 64pp.

Figueiredo JB (2006) Tradução da Tradição: gaúchos, guaxos e sombras. O regionalismo revisitado de Luiz Carlos Barbosa Lessa e Ricardo Güiraldes. Tese de Doutorado.

Programa de Pós Graduação em Letras. Porto Alegre, Universidade Federal do Rio Grande do Sul.

Flores M (2003) História do Rio Grande do Sul. 7th edition . Ediplat, Porto Alegre.

Hünemeier T (2006) Filogeografia dos cromossomos Y e das linhagens mitocondriais de origem africana em populações negras brasileiras. Dissertação de mestrado. Programa de Pós Graduação em Genética e Biologia Molecular, Porto Alegre, Universidade Federal do Rio Grande do Sul.

Hünemeier T, Carvalho C, Marrero AR, Salzano FM, Pena SDJ and Bortolini MC. Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. Manuscrito submetido.

Jobling MA and Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4: 598 – 610.

Jobling MA, Hurles ME and Tyler-Smith C (2004) Human evolutionary genetics. Garland Publishing, New York, 523 pp.

Kern AA (1997) Paleopaisagens e povoamento pré-histórico no Rio Grande do Sul. In Jacoby R (ed) Arqueologia Pré-histórica do Rio Grande do Sul, 2ª edition. Mercado Aberto, Porto Alegre, pp 13-62.

Kern AA, Bernd Z, Cohen VRA, Avancini EG, Kemel CLA and Gick PW (1993) Rio Grande do Sul: Continente Múltiplo. Coleção O continente de São Pedro, volume V. Riocell Marprom, Porto Alegre, 159 pp.

Laan M, Wiebe V, Khusnutdinova E, Remm M and Pääbo S (2005) X-chromosome as a marker for population history: linkage disequilibrium and haplotype study in Eurasian populations. Eur J Hum Genet 13: 452 – 462.

Leal OF (1989) Gaúchos: male culture and identity in the Pampas. Tese de Doutorado. Programa de Pós Graduação em Antopologia. Berkeley, University of California.

Leite FPN (2006) Análise da estrutura genética da população do rio Grande do Sul através de microssatelites autossômicos e de cromossomos sexuais. Tese de doutorado. Programa de Pós Graduação em Genética e Biologia Molecular, Porto Alegre, Universidade Federal do Rio Grande do Sul.

Lutz S, Wittig H, Weisser HJ, Heizmann J, Junge A, Dimo-Simonin N, Parsons W, Edelmann J, Anslinger K, Jung S and Augustin C (2000) Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? Forensic Sci Inter 113: 97-101.

Marques A and Draper D (2001) Dicionário Inglês/ Português, Português/ Inglês. 22th edition. Ática, São Paulo, 560 pp.

Marrero AR (2003) Os Gaúchos: sua história evolutiva revelada a partir de marcadores genéticos. Dissertação de mestrado. Programa de Pós Graduação em Genética e Biologia Molecular, Porto Alegre, Universidade Federal do Rio Grande do Sul.

Marrero AR, Leite FPN, Carvalho BA, Peres LM, Kommers TC, Cruz IM, Salzano FM, Ruiz-Linares A, Silva Jr WA and Bortolini MC (2005) Heterogeneity of the genome ancestry of individuals classified as while in the state of Rio Grande do Sul, Brazil. Am J Hum Biol 17: 496 – 506.

Matioli SR (2001) Biologia molecular e evolução. Holos, Ribeirão Preto, 202 pp.

Mentz-Ribeiro PA (1997) Os caçadores Pampeanos e a Arte Rupestre. In Jacoby R (ed) Arqueologia Pré-histórica do Rio Grande do Sul, 2nd edition. Mercado Aberto, Porto Alegre, pp 103-134.

Monteiro JN (1992) Os Guarani e a história do Brasil Meridional, séculos XVI-XVII. In Carneiro da Cunha M (org) História dos índios no Brasil. Companhia das Letras, São Paulo, pp 475 – 498.

Nichols MW (1946) O Gaúcho: caçador de gado – cavaleiro - ideal de romance. Livraria Editora Zélio Valverde, Rio de Janeiro, 232 pp.

Oliven RG (1993) Identidade Nacional e Etnicidade: diversidade étnica no Rio Grande do Sul. In Costa MCS and Pinto MT (ed). Encontros com a Antropologia (Identidade, Migração e Memória). SESC, Curitiba, pp 23-38.

Oliven RG (1996) Tradition Matters: Modern Gaúcho Identity in Brazil. Columbia University Press, New York, 151 pp.

Oliven RG (1997) O Gaúcho brasileiro e o *Gaucho* platino, como vistos do Brasil. In Fronteras culturales y Ciudadania, II Reunión de Antropologia del MERCOSUR. Piriápolis, pp 31-40.

Pakendorf B and Stoneking M (2005) Mitochondrial DNA and Human Evolution. Annu Rev Genomics Hum Genet 6: 165-183.

Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM and Pena SDJ (2003) Color and genomic ancestry in Brazilians. Proc Natl Acad Sci USA 100: 177 – 182.

Pena SDJ and Bortolini MC (2004) Pode a genética definir quem deve se beneficiar das cotas universitárias e demais ações afirmativas? Estudos Avançados 18: 1-20.

Pereira RW and Pena SDJ (2006) Phylogeograhy of haplotypes of five microsatellites located a low-recombination region of the X-chromosome studies worldwide and in Brazilian populations. Genetica 126: 243 – 250.

Pi Hugarte R (1998) Los Índios del Uruguay. Ediciones de La Banda Oriental, Montevidéu, 240 pp.

Reverbel C (2002) O Gaúcho. L&PM, Porto Alegre, 150 pp.

Ricardo AC (2000) Povos indígenas no Brasil. 1996/ 2000. Instituto Socioambiental, São Paulo, 166 pp.

Ridley M (2006) Evolução. Artmed, Porto Alegre, 752 pp.

Rivet P (2002) Los últimos Charruas. 2nd edition. Ediciones de La Plaza, Montevidéu, 83pp.

Schaden E (1962) Aspectos fundamentais da cultura Guarani. Difusão Européia do Livro, São Paulo, 190 pp.

Schaffner SF (2004) The X chromosome in population genetics. Nat Rev Genet 5: 43 – 51.

Schiavetto SNO (2003) A arqueologia Guarani: construção e descontrução da identidade indígena. Annablume/ Fapesp, São Paulo, 138 pp.

Silva FG, Ewald GM, Jobim MRSL; Jobim MSL (2005) Investigação laboratorial. In Identificação humana: identificação pelo DNA; Identificação Médico-Legal; Perícias Odontológicas. Volume II. Millennium Editora, Campinas, 304 pp.

Silva WA, Bortolini MC, Schneider MPC, Marrero A, Elion J, Krishnamoorthy R and Zago MA (2006) mtDNA haplogroup analysis of Black Brazilian and Sub-Saharan populations: implications for the Atlantic slave trade. Hum Biol 78: 29-41.

Schmitz PI (1997)Migrantes da Amazônia: a Tradição Tupi-Guarani. In Kern AA (org) Arqueologia pré-histórica do Rio Grande do Sul. 2nd edition. Mercado Aberto, Porto Alegre, 356 pp.

Strachan T and Read AP (2002) Genética humana molecular. 2nd edition, Artmed, Porto Alegre, 576 pp.

Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlish E and Krause D (2003) Use of X-linked markers for forensic purposes. Int J Legal Med 117: 67 – 74.

Torroni A, Achilli A, Macaulay V, Richards M and Bandelt H-S (2006) Harvesting the fruit of the human mtDNA tree. Trends Genet 22: 339 – 345.

Underhill P, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldmann MW, Cavalli-Sforza LL and Oefner PJ (2000) Y chromosome sequence variation and the history of human population. Nat Genet 26: 358 – 361.

Vargas AE, Marrero AR, Salzano FM, Bortolini MC and Chies JAB (2006) Frequency of CCR5Δ32 in Brazilian populations. Braz J Med Biol Res 39: 321-325.

Vidart D (2000) La trama de la Identidad Nacional. Tomo I: Indios, Negros, Gauchos. Ediciones de la Banda Oriental, Montevideu, 171 pp.

ANEXOS

ANEXO 1

Frequency of CCR5Δ32 in

Brazilian populations

Vargas *et al*. (2006) Braz J Med Biol Res 39: 3215-325

Frequency of CCR5Δ32 in Brazilian populations

A.E. Vargas, A.R. Marrero,
F.M. Salzano, M.C. Bortolini
and J.A.B. Chies

Departamento de Genética, Instituto de Biociências,
Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

Abstract

Correspondence

J.A.B. Chies
Departamento de Genética, UFRGS
Caixa Postal 15053
91501-970 Porto Alegre, RS
Brasil
Fax: +55-51-3316-7311
E-mail: jabchies@terra.com.br

A sample of 103 randomly chosen healthy individuals from Alegrete, RS, Brazil, was tested for the CCR5Δ32 allele, which is known to influence susceptibility to HIV-1 infection. The CCR5Δ32 allele was identified by PCR amplification using specific primers flanking the region of deletion, followed by electrophoresis on a 3% agarose gel. The data obtained were compared to those reported for other populations and interpreted in terms of Brazilian history. The individuals studied came from a highly admixed population. Most of them were identified as white (N = 59), while blacks and browns (mulattoes) were N = 13 and N = 31, respectively. The observed frequencies, considering the white, black and brown samples (6.8, 3.8, and 6.4%, respectively), suggest an important European parental contribution, even in populations identified as black and brown. However, in Brazil as a whole, this allele shows gradients indicating a relatively good correlation with the classification based on skin color and other physical traits, used here to define major Brazilian population groups.

One of the most interesting characteristics of the Brazilian population is its heterogeneity. When Brazil was "discovered" by the Portuguese in 1500, the land was already inhabited by Amerindians (estimated at 2 million people). Since then, emigration of individuals from different countries and continents with diverse ethnic backgrounds has contributed to the establishment of the genetic pool of the contemporary Brazilian population. These parental contributions included a constant influx of Portuguese, 4 million Africans (mainly from West-Central Africa) and 3.9 million Europeans (other than Portuguese), who arrived here in the 19th and 20th centuries (1).

However, the distribution of these immigrants was unequal in the various Brazilian regions. In the North, the populations were formed mainly by Europeans and Amerindians; in the Southeast and Northeast, Europeans, Africans and Amerindians had different degrees of influence, while in the South, the European heritage prevails (1).

CCR5 is a chemokine receptor present mainly in cells of the immune system, such as macrophages and T lymphocytes, playing a major role in the migration of these cells to sites of inflammation. The gene encoding CCR5 (*CKR5*) is located in the p21.3 region of the human chromosome 3, forming a cluster with other chemokine receptor genes (2). Deng et al. (3) demonstrated that CCR5 serves as a co-receptor for human immunodeficiency vi-

rus-1 (HIV-1). The variant allele CCR5Δ32 described by Liu et al. (2) contains a 32-bp deletion that generates a truncated protein, which confers relative resistance to HIV-1 infection.

The study of the allelic frequency of CCR5Δ32 in 18 European populations revealed an interesting North-South gradient, with the highest frequencies of the variant allele being observed in Finnish and Mordvinian populations (16%) and the lowest in Sardinia (4%) (4). The last investigators also proposed that the CCR5Δ32 allele originated from a single mutation event in Northeastern Europe a few thousand years ago. The high frequencies of CCR5Δ32 found in Europeans have been attributed to a strong selective pressure, possibly exerted by pathogens such as *Yersinia pestis* (the bubonic plague agent), *Shigella*, *Salmonella*, and *Mycobacterium tuberculosis*, all of which target macrophages, or by some other infectious diseases such as syphilis, smallpox and influenza (5).

Thus, the prevalence of this allele is of obvious medical importance. We have investigated its distribution in a random sample of individuals from Alegrete, a town located in the western region of Rio Grande do Sul (29º53' S; 55º57' W) where the population was basically established from a mixture of Spanish, Portuguese and African individuals and native Amerindians, and compared it to those already reported for other populations, interpreting the data in terms of Brazilian history.

A sample of 103 randomly chosen unrelated healthy individuals from Alegrete, RS, Brazil, was analyzed in the present study. Most of the individuals studied were identified as white (N = 59), while blacks and browns (mulattoes) were N = 13 and N = 31, respectively. This classification was based on physical appearance as judged by the researcher at the time of blood collection, and on data about the ethnicity of parents/grandparents reported by the participants.

The investigation was approved by the Brazilian National Ethics Committee (CONEP No. 1333/2002) and all donors were informed about the aims of this study and signed a written consent.

DNA was extracted from saliva or blood samples using the Nucleon DNA Extraction kit (Nucleon Bioscience, Coatbridge, UK) or a salting-out method, respectively.

Genotyping was performed by PCR amplification with specific primers. PCR samples were prepared to a final volume of 25 μL as follows: 1 μL DNA (0.2-0.5 μg), 2.5 μL 10X PCR buffer (200 mM Tris-HCl, pH 8.4, 500 mM KCl), 1 μL 50 mM MgCl$_2$, 1 μL 3 mM dNTP mix, 1 μL 10 pmol primer mix, and 0.2 μL *Taq* DNA polymerase, 5 U/μL (Invitrogen Corporation, San Diego, CA, USA). Samples were submitted to 40 cycles of 1 min at 94ºC, 1 min at 55ºC, and 1 min at 72ºC. The set of specific primers used to amplify the CCR5 gene segment was described by Chies and Hutz (6). It yields a 137-bp fragment for the wild-type allele and a 105-bp fragment for the CCR5Δ32 variant. PCR products were plotted on 3% agarose gel containing ethidium bromide and submitted to electrophoresis. Fragments were visualized under UV irradiation.

Skin color is used in Brazil as the equivalent of race, and is based on a complex and subjective phenotypic evaluation. In Brazil, the emphasis is on physical appearance rather than ancestry, which is in contrast to the situation in the United States. The Brazilian Institute of Geography and Statistics (IBGE) adopts the criterion of classification of individuals into the following categories: white (in Portuguese, branco), black (preto), brown (pardo), yellow (amarelo), and Amerindian (indígena). Accordingly, in Brazil as a whole, 53, 6, and 38% of the persons are identified as white, black, and brown, respectively, the remaining 3% being distributed among yellow and Amerindian persons. In Rio Grande do Sul (~10 million inhabitants), the numbers are 87.5, 5, 7, and 0.5% for white, black,

brown, and yellow + Amerindian individuals, respectively (7). More recently, the expression Afro-descendent has been incorporated into this ethnic semantic definition (8). However, the last investigators have estimated that about 148 million Brazilians present more than 10% of African nuclear genome ancestry, and that at least 89 millions of individuals have mtDNA lineages of African origin (8). This illustrates the extension of admixture in Brazil and supports the suggestion that skin color and other phenotypic traits can be poor predictors of genomic ancestry. These results reinforce the idea that, independently of the chosen criteria, it is problematic to classify people. To facilitate reading and comprehension, the word "black" will be used here to refer to any person (or population) identified and/or self-identified with some term that reports African ancestry according to physical appearance, whereas "white" will be used to define those that, according to their physical traits,

do not report admixture with non-Europeans. Brown will be used to refer to individuals with intermediate physical appearance between white and black.

It has been widely observed that most populations share alleles at any given locus and that those alleles that are most frequent in one population are also found at high frequency in others, reflecting the recent dispersion of *Homo sapiens* into continental groups (9). Due to this fact, there are few classical or DNA markers that have been demonstrated either to be population-specific or to have large frequency disparities among geographically and ethnically defined populations (9).

In the present study, no CCR5Δ32/CCR5Δ32 homozygotes were detected. The presence of the CCR5/CCR5Δ32 genotype among whites, blacks and browns was 14, 8, and 13%, indicating a CCR5Δ32 allele frequency of 6.8, 3.8, and 6.4%, respectively. The CCR5Δ32 distributions observed in

Table 1. CCR5 genotype and CCR5Δ32 allele frequencies in Brazilian populations and in their putative parental groups.

Brazilian populations	No. individuals	Genotype frequencies (%)		Δ32 allele frequency (%)[a]	References
		CCR5/Δ32	Δ32/Δ32		
Brown/unclassified urban/semi-urban					
North	203	15	1	4.2	14
Southeast	539	57	0	5.3	17-19
South	31	4	0	6.4	Present study
Black urban/semi-urban					
Northeast	549	29	0	2.6	6
Southeast	54	4	0	1.9	6
South	71	3	0	0.7	6, present study
Black					
Rural	296	11	0	1.9	14,15
White urban/semi-urban					
South	158	19	1	6.6	10, present study
Parental groups					
Amerindians	1071	5	0	0.2	10,12,14,20
Europeans	2668	492	23	10.1	4
Africans	251	0	0	0	11

[a]Weighted average allele frequencies were obtained when more than one study was considered.

whites and browns in Alegrete are similar to those reported for other white Brazilians (German descendants: 6.5% (10) and Portuguese: 6.4% (4).

For a better understanding of the scenarios of the CCR5 genotype distribution and the CCR5Δ32 allele frequency in Brazil, we grouped our data with those obtained by others, considering the sample classification into three major population groups: white, black and brown/unclassified (Table 1). Additionally, Table 1 provides information about the numbers described for the three putative parental groups (European, African, and Amerindian). The CCR5Δ32 allele frequency in Europe is ~10%, whereas in Sub-Saharan African populations this allele is absent (11). Therefore, its presence among native Americans (0.2%) is probably influenced by admixture with non-Indians, as observed in the Pataxó and Kaingang tribes (12,13).

Considering the frequencies in the major human geographical groups shown in Table 1, CCR5Δ32 can be referred to as a private European allele, and could be used along with other genetic markers in studies of genome ancestry in admixed populations.

The frequency of the CCR5Δ32 allele in the Brazilian populations identified as brown/unclassified ranges from 4.2 in the North region (14) to 6.4 in the South (present study). Our data contribute to the establishment of a scenario that shows a North-South gradient. This view is compatible with the colonization of Brazil, since the Southeast and South regions received the highest numbers of European immigrants during the 19th and early 20th centuries (1). The numbers are lower when we consider the black Brazilian populations, ranging from 0.7 to 2.6% (14,15,

present study), and are indicative of an inverse gradient. Curiously, rural black populations, which are usually the descendants of "quilombos" (communities founded by fugitive slaves at colonial times), show average CCR5Δ32 allele frequency values similar to those observed in urban black samples (14,15). These results suggest that the level of introgression of European genes in these small rural communities is not very different from that observed in urban groups. Alternatively, random drift could be evoked to explain an increase in the frequency of formerly rare alleles, a phenomenon that has been frequently described in other quilombo communities (16).

Finally, our investigation concerning a typical European private allele revealed that the CCR5Δ32 frequencies in Brazil as a whole (Table 1) show a relatively good correlation with the classification based on skin color and other physical traits, used here to define major Brazilian population groups.

The implication of these results for public health is that the contemporary Brazilian populations may be differentially susceptible to infectious diseases, and that caution is needed since, sometimes, depending on the population and genetic markers used, the association between phenotypic characters and genome ancestry is not found, whereas in others, as demonstrated here, it is.

Acknowledgments

References

1. Callegari-Jacques SM, Grattapaglia D, Salzano FM et al. (2003). Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. *American Journal of Human Biology*, 15: 824-834.

2. Liu R, Paxton WA, Choe S et al. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell,* 86: 367-377.

3. Deng H, Liu R, Ellmeier W et al. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature,* 381: 661-666.

4. Libert F, Cochaux P, Beckman G et al. (1998). The deltaccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. *Human Molecular Genetics*, 7: 399-406.

5. Stephens JC, Reich DE, Goldstein DB et al. (1998). Dating the origin of the CCR5-del32 AIDS-resistance allele by the coalescence of haplotypes. *American Journal of Human Genetics,* 62: 1507-1515.

6. Chies JA & Hutz MH (2003). High frequency of the CCR5delta32 variant among individuals from an admixed Brazilian population with sickle cell anemia. *Brazilian Journal of Medical and Biological Research*, 36: 71-75.

7. IBGE - Instituto Brasileiro de Geografia e Estatística (2000). Censo Demográfico 2000. Características gerais da população. Resultado da amostra. Available at: http://www.ibge.gov.br/home/estatistica/populacao/censo2000/populacao/censo2000_populacao.pdf. Accessed March, 2004.

8. Bortolini MC & Pena SD (2004). Pode a genética definir quem pode se beneficiar das cotas universitárias e demais ações afirmativas? *Estudos Avançados*, *USP*, 18: 1-22.

9. Shriver MD, Smith MW, Jin L et al. (1997). Ethnic-affiliation estimation by use of population-specific DNA markers. *American Journal of Human Genetics*, 60: 957-964.

10. Grimaldi R, Shindo N, Acosta AX et al. (2002). Prevalence of the CCR5Delta32 mutation in Brazilian populations and cell susceptibility to HIV-1 infection. *Human Genetics*, 111: 102-104.

11. Martinson JJ, Hong L, Karanicolas R et al. (2000). Global distribution of the CCR2-64I/CCR5-59653T HIV-1 disease-protective haplotype. *AIDS*, 14: 483-489.

12. Barbosa AAL, Sousa SMB & Simões AL (2004). Análise da freqüência do alelo CCR5Δ32 nos indígenas Pataxó indica contribuição européia. *50º Congresso Brasileiro de Genética*, Florianópolis, SC, Brazil, September 7-10.

13. Hünemeier T, Neves AG, Nornberg I et al. (2005). T-cell and chemokine receptor variation in South Amerindian populations. *American Journal of Human Biology*, 17: 515-518.

14. Carvalhaes FAPL, Carvalho MIM & Guerreiro JF (1999). A mutação CCR5-delta 32 em populações humanas da Amazônia. *45º Congresso Nacional de Genética*, Gramado, RS, Brazil, October 3-6.

15. Carvalho MWP, Leboute APM, Oliveira SF et al. (2004). CCR5Δ32 mutation in three Brazilian populations of predominantly Sub-Saharan African ancestry. *Genetics and Molecular Biology*, 27: 321-325.

16. Bortolini MC, Weimer T de A, Salzano FM et al. (1995). Evolutionary relationships between black South American and African populations. *Human Biology*, 67: 547-559.

17. Passos Jr GA & Picanco VP (1998). Frequency of the delta ccr5 deletion allele in the urban Brazilian population. *Immunology Letters*, 61: 205-207.

18. Pereira RW, Pires RE, Duarte APM et al. (2000). Frequency of the CCR5Δ32 allele in Brazilians: a study in colorectal cancer and in HTLV-I infection. *Genetics and Molecular Biology*, 23: 523-526.

19. Munerato P, Azevedo ML, Sucupira MC et al. (2003). Frequency of polymorphisms of genes coding for HIV-1 co-receptors CCR5 and CCR2 in a Brazilian population. *Brazilian Journal of Infectious Diseases*, 7: 236-240.

20. Leboute AP, de Carvalho MW & Simoes AL (1999). Absence of the deltaccr5 mutation in indigenous populations of the Brazilian Amazon. *Human Genetics*, 105: 442-443.

ANEXO 2

mtDNA Haplogroup Analysis of Black Brazilian and Sub-Saharan Populations: inplications for the Atlantic Slave Trade

Silva Jr *et al*. (2006)Hum Biol 78: 29-41

mtDNA Haplogroup Analysis of Black Brazilian and Sub-Saharan Populations: Implications for the Atlantic Slave Trade

WILSON ARAÚJO SILVA,[1] MARIA CÁTIRA BORTOLINI,[2] MARIA PAULA CRUZ SCHNEIDER,[3] ANDREA MARRERO,[2] JACQUES ELION,[4] RAJAGOPAL KRISHNAMOORTHY,[4] AND MARCO ANTONIO ZAGO[5]

Abstract   Seventy individuals from two African and four black Brazilian populations were studied for the first hypervariable segment of mtDNA. To delineate a more complete phylogeographic scenario of the African mtDNA haplogroups in Brazil and to provide additional information on the nature of the Atlantic slave trade, we analyzed our data together with previously published data. The results indicate different sources of African slaves for the four major Brazilian regions. In addition, the data revealed patterns that differ from those expected on the basis of historical registers, thus suggesting the role of ethnic sex differences in the slave trade.

From the 15th to the 19th century, 9 million sub-Saharan Africans were brought to the Americas as slaves; about 40% of them were probably brought to Brazil (Klein 2002). This forced migration had a tragic impact on some African societies and determined that part of the history of Africans began to be written outside Africa. At present, genetic studies of Brazilians and other New World descendents of Africans have provided data that have been used to rescue part of this particular history (Zago et al. 1992; Figueiredo et al. 1994; Bortolini et al. 1997, 1999, 2004; Silva et al. 1999; Salzano and Bortolini 2002). Some of these investigations used *HBB*S* haplotypes to define the origin of Africans that arrived in Brazil, because historical records about slavery contain many gaps (Zago et al. 1992; Figueiredo et al. 1994). In an extensive review, Salzano and Bortolini (2002) estimated that 61%, 34%, and 3% of the *HBB*S* haplotypes found in

[1]Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil.
[2]Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.
[3]Departamento de Genética, Laboratório de Polimorfismo de DNA, Universidade Federal do Pará, Belém, Brazil.
[4]INSERM, UMR 458, Biochimie, Génétique, Hôspital Robert Débre, Paris, France.
[5]Departamento de Clínica Médica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil.

KEY WORDS: MTDNA, MTDNA HAPLOGROUPS, HVS-I, SUB-SAHARAN POPULATIONS, BLACK BRAZILIANS, AFRICANS, BANTU POPULATIONS, SLAVE TRADE, BRAZIL, CAMEROON, CONGO, ANGOLA, MOZAMBIQUE, NETWORK ANALYSIS, PHYLOGEOGRAPHY.

Brazil are of the Bantu, Benin, and Senegal types, respectively. These values are different from those observed in other New World countries, such as Venezuela, Cuba, and Jamaica, where the Benin haplotype is the most frequent.

More recently, however, lineage markers such as mtDNA have been used for this purpose. Alves-Silva et al. (2000) furnished an initial landscape of the phylogeography of African mtDNA haplogroups in Brazil. Together, haplogroups L3e and L1c constitute approximately 49% of the African fraction of sequences identified by Alves-Silva et al. (2000). Although these analyses have been limited so far by the lack of mtDNA data on important African slave sources to Brazil, such as Angola, Congo, and Mozambique, Alves-Silva et al. (2000) suggested that most of the mtDNA lineages of African ancestry in their Brazilian sample had an origin in central Africa, although a substantial number must have come from western Africa. Bandelt et al. (2001) evaluated the phylogeography of the L3e mtDNA haplogroup, which is omnipresent in Africa but virtually absent in Eurasia, and concluded that the distributions of haplogroup L3e in Brazil and in the Caribbean area still reflect the different African slave sources to the New World.

Salas et al. (2004a) compared the distribution of all the main mtDNA haplogroups and their derived lineages in Africa with available data from the Americas. They estimated that 65%, 41%, and 28% of mtDNA types found in South, North, and Central America, respectively, had a Central-West African origin. Earlier the same research group had identified a new African haplogroup (L3g), which is frequent in Tanzania and Kenya (Salas et al. 2002). The presence of haplogroup L3g in three Brazilians (among the 92 African mtDNA haplotypes that were characterized; Bortolini et al. 1997; Alves-Silva et al. 2000) was interpreted as either direct slave trade from eastern Africa to the New World or hitherto undetected gene flow from eastern Africa into western or southwestern Africa and then into the Americas. Bortolini et al. (2004) evaluated this proposal and, from the identification of L3g in several Cameroon ethnic groups, concluded that the Camaroonese L3g lineages originated from eastern Africa by transcontinental gene flow and that the L3g lineages in the Americas probably have their immediate origin in Cameroon or in neighboring regions and not in eastern Africa. On the basis of the extensive amount of new data that could be added to the L3g phylogeny, Salas et al. (2004b) corroborated this proposal.

Color is used in Brazil as an equivalent to "race" and is based on a subjective phenotypic evaluation (Parra et al. 2003). In contrast to the situation in the United States, in Brazil the emphasis is on physical appearance rather than ancestry. The Brazilian Institute of Geography and Statistics (IBGE) has adopted the criterion of classification of individuals according to the following categories: white (in Portuguese, *branco*), black (*preto*), brown (*pardo*), yellow (*amarelo*), and Amerindian (*Indígena*). Accordingly, in Brazil as a whole, 90 million, 10 million, and 65 million of the people were identified as white, black, and brown, respectively; the remaining 5 million people are distributed between the two other categories (IBGE Census 2000; available at http://www.ibge.gov.br). However,

according to Telles (2003), two other major systems, beyond that adopted by the IBGE, are associated with "racial classification": (1) The popular discourse, which uses a large and variable nomenclature, includes several ambiguous terms, such as mulatto; and (2) the political discourse of the black organized social movements lumps together as "black" all the variations, such as black, brown, and mulatto. More recently, the expression *Afro-descendant* has been incorporated into this ethnic semantics (Pena and Bortolini 2004). Pena and Bortolini (2004) estimated that 148 million Brazilians present more than 10% African nuclear genome ancestry and that at least 89 million individuals have mtDNA lineages of African origin. This illustrates the extent of admixture in Brazil and corroborates the suggestion that color and other phenotypic traits can be poor predictors of genomic ancestry (Parra et al. 2003). These results also strengthen the opinion that classification of individuals within a population is always difficult and subject to error, whatever the basis for the "ethnic" or "racial" classification. In this paper we use the word *black* to refer to any person (or population) identified and/or self-identified with some term that reports African ancestry according to physical appearance.

Here, we provide information about the distribution of the mtDNA haplogroups in three rural and one urban Brazilian black communities and in two Bantu populations from Africa (Cameroon and Democratic Republic of Congo). Our data from African populations furnish information about an until now largely uncharacterized region, which is known as the birthplace of and an important route for the major Bantu expansion. The importance of Cameroon and Congo as sources of slaves to Brazil is also well known. In addition, we analyzed our data with respect to other recently published data, including data from Angola and Mozambique, and provide new considerations about the nature of the Atlantic slave trade to Brazil.

Materials and Methods

Population Samples and DNA Extraction. The African samples were obtained from 20 Bantu-speaking subjects living in two African countries: (1) the Democratic Republic of Congo (formerly Zaire) (samples from 10 individuals were collected in Lubumbashi city, in Shaba province); and (2) Cameroon (samples from 10 individuals were collected in Yaoundé city from the Boulou, Bamileke, Bene, Eton, Nweh, Sonaga, and Etongo ethnic groups).

The black Brazilian samples consist of 30 individuals from 3 rural communities: (1) Cametá ($N = 10$; 2°3′ S, 59°55′ W), in the region of the lower Tocantins River, state of Pará, northern Brazil; (2) Trombetas ($N = 10$; 1°8′–1°46′ S, 55°51′–57° W), at the margins of the Trombetas and Cuminá rivers, state of Pará, northern Brazil; and (3) Cajueiro ($N = 10$; 2°25′ S, 44°20′ W), located in the county of Alcantara, state of Maranhão, northeastern Brazil. These rural black communities are recognized as *quilombos*, because their founders were probably

fugitive slaves. One urban sample was also investigated. This sample was obtained from 20 individuals living in Ribeirão Preto (20°10′ S, 40°75′ W), located in the northern part of the state of São Paulo. Additional information about these populations can be obtained from Bortolini et al. (1999, 2004) and Silva et al. (1999).

DNA extraction from whole blood was performed according to the method of Lahiri and Nurnberger (1991).

mtDNA Amplification and Sequencing. The nucleotide sequence of the first mtDNA hypervariable segment (HVS-I) was directly amplified using the polymerase chain reaction (PCR) with the primers and PCR protocol described by Ward et al. (1991). Reaction products were then purified and sequenced according to the conditions described or referenced by Bortolini et al. (1997). For all samples both strands of DNA were sequenced.

Genetic Analysis. Nucleotide positions 16020 to 16365 were considered for the analysis. To evaluate whether artifacts were generated (phantom mutations) during the sequencing process, we applied the method described by Bandelt et al. (2002). The first analysis filtered out all speedy transitions and thus scored weighty mutations only. After filtering for speedy transitions, we constructed a network of sequences with the program Network 3.1 (available at http://www .fluxus-engineering.com) using a median-joining algorithm (Bandelt et al. 1995, 1999). Weight networks showing perfect star tree patterns are expected when the data are potentially free of phantom mutations.

The information provided by HVS-I was used to classify the lineages into haplogroups, according to Salas et al. (2002, 2004a). However, studies of the coding regions have revealed novel parsimony, informative polymorphisms, or previously unidentified splits in the inner branches of the mtDNA phylogeny. Considering this recent information, Kivisild et al. (2004) defined new haplogroups that extend the framework of the existing classification scheme. For example, haplogroup L3g shares motifs, within HVS-I and HVS-II and at positions 769 and 1018, with L4a. This information led Kivisild et al. (2004) to suggest that haplogroup L3g is actually a sister cluster of haplogroup L4a and therefore to propose that L3g be renamed L4g. Haplogroup L1e, previously characterized on the basis of HVS-I motifs (Salas et al. 2002), has been recently redefined as L5a because it occupies an intermediate phylogenetic position between the L1 and L2′L3 major haplogroups (Shen et al. 2004).

Although the hierarchical relation among the human mtDNA lineages is well known, the terminology to define them remains confusing. *Haplogroup*, *clade*, *subhaplogroup*, and *subclade* are words frequently used synonymously. In this paper the term *major haplogroup* is used to define the major lineages (A, B, C, D, J, L0, L1, L2, L3, L4, L5, etc.), whereas *haplogroup* is used to identify their first derivations (L2a, L2b, L3e, etc.). *Subclade* and *subhaplogroup* are used

equally to define any derived lineage from the haplogroups (L2a1, L1c1, L3e1, L3e2, etc.).

Because most sub-Saharan mtDNA haplogroups are not region specific, we estimated the parental contributions using the haplogroup frequencies and Long's (1991) least-squares method.

Results and Discussion

The networks obtained for the HVS-I weighty variation showed perfect star tree configurations, indicating that our HVS-I data sets are potentially free of phantom mutations.

Table 1 shows the mtDNA lineages and the haplogroups or subhaplogroups identified in our black Brazilian and African samples. The higher non-African fraction was observed in Cametá, because 60% of sequences can be associated with the major Amerindian haplogroups A, B, C, and D. The large Amerindian component in Cametá is not surprising, because the community is located in the Amazonian region. The native American component was also detected in Cajueiro and Trombetas but in lower proportions (30% and 10%, respectively). No European mtDNA sequence was observed in these populations. These results probably reflect the introduction of native American women into the *quilombos*, particularly during the slavery era, because the number of men who escaped was larger than the number of women who escaped.

European presence was detected only in the urban sample of Ribeirão Preto, but in low frequency (5%; major haplogroup J). A more significant presence of non-African lineages would be expected in the urban black Brazilian population; for instance, Bortolini et al. (1997) estimated that 17% of the mtDNA sequences in their urban black sample had an Amerindian or European origin. However, we had selected the Ribeirão Preto urban sample so that it would include only individuals who did not report any nonblack ancestry (Silva et al. 1999).

The African sequences show large diversity, with several haplogroups or subhaplogroups normally found in the sub-Saharan region detected; all these sequences could be assigned to the major African haplogroups L0, L1, L2, L3, and L4 (Salas et al. 2002, 2004a; Bortolini et al. 2004; Kivisild et al. 2004; Plaza et al. 2004). The subclade L2a1 is unique and was found in all four black Brazilian populations. Haplogroup L4g (formerly L3g), which has been the target of recent analyses (Bortolini et al. 2004; Salas et al. 2004b), is present in Trombetas. This last result indicates that the distribution of haplogroup L4g is not restricted to the southern and southeastern regions of Brazil (Bortolini et al. 1997, 2004; Alves-Silva et al. 2000). Both subclades L4g1 and L4g2 (Salas et al. 2004b) are found in Brazil, with subclade L4g2 showing a higher frequency (75% of the Brazilian L4g sequences).

Table 1. mtDNA Haplogroups and Lineages in Black Brazilian and African Populations

Haplogroup	HVS-I (16000+)	Africans		Black Brazilians			
		Congo	Cameroon	Cajueiro	Cametá	Ribeirão Preto	Trombetas
African							
L0a1	129 148 168 171 172 187 188G 189 223 230 256A 278 291 311 320	1					
L0a1	129 148 168 172 187 188G 189 223 230 311 320						
L0a1	129 148 168 172 187 188G 189 223 230 278 293 311 320					2	
L1b	126 187 189 193 213 223 264 278 293 311					1	
L1b	111 126 187 189 223 239 270 278 287 293 311		1				
L1c1	129 163 187 189 209 223 278 293 294 298 311 360		1				
L1c1	086 129 187 189 223 241 278 291 293 294 311 360					1	
L1c1	038 187 189 223 278 293 294 311 360					1	
L1c2	093 129 187 189 223 265C 278 286G 294 311 358G 360	1				1	
L1c2	129 145 187 189 213 223 265C 278 286G 294 311 360	1				1	
L1c2	129 187 189 223 234 265C 278 286G 294 311 360					1	
L1c2	129 187 189 223 265C 278 286G 294 311 359 360						
L1c2	129 187 189 223 265C 278 286G 294 311 343T 360			1			
L1c2	187 189 223 265C 278 286G 294 311 343T 360			1			
L2a	111A 223 234 249 278 294 295					1	
L2a	223 234 235C 249 278 294 295	1					
L2a	095G 096G 223 234 249 278 294 295	1	1				
L2a1	223 278 294 309		1				
L2a1	223 225 234 278 294 309						1

Haplogroup	HVS-I/II variants							
L2a1	223 278 291 294 309 360						1	1
L2a1	042 086 092 223 278 294 309 356						1	1
L2a1	223 278 294 309 363							1
L2a1	189 223 278 294 363							1
L2a1	131 189 223 278 294 309 363				1			
L2a1	189 223 278 294 309			1	2			
L2a1	189 192 223 278 294 309		1					
L2b	114A 129 213 223 278 354				1			
L2b	093 129 167 189 278 300 311 354							2 1
L2c	223 278							
L3b	124 189 223 278 362		1					
L3b	145 223 278 362						1	
L3d	069 124 192 223 242						1	
L3d	124 223 266 319							1
L3d1	111 124 223		1					
L3e1	223 327		1	1				
L3e1	223 260 327							
L3e1	126 169 180 223 255 327						2 1	
L3e1a	185 223 327							
L3e1a	185 209 223 327		1					
L3e4	051 223 264							
L3e2b	172 223 278 320		1	1				
L3e2b	172 223 311 320			1				
L3e2b	223 234 311 320							

Table 1. (Continued)

Haplogroup	HVS-I (16000+)	Africans		Black Brazilians			
		Congo	Cameroon	Cajueiro	Cametá	Ribeirão Preto	Trombetas
L3e2b	172 189 223 320					1	
L3e2b	164 172 189 320						
L3f	209 223 311			1			
L3f	148 209 223 311					1	
L3f1	129 209 223 292 311		1				
L3f1	111 209 218 223 292 311		1				
L4g1	051 114 189 192 223 293T 311 316 355 360		1				1
L4g2	093G 223 287A 293T 301 311 355 362		1				1
Amerindian							
A	126 223 278 290 319 362						
A	111 223 290 319 362				1		
B[a]	217						
C	051 093 223 298 325 327				1		
C	179 223 298 325 327				1		
C	223 298 325 327 354				1		
C	129 223 294 298 325 327 360			1			
C	223 298 311 325 327			1			
C	126 223 298 325 327			1			
D	223 293 325 362				1		
European							
J	069 126					1	
Total		10	10	10	10	20	10

a. CoII/tRNA^{Lys} 9-bp deletion was also detected in this sample, beyond the transition 16217C → T, which characterizes Amerindian haplogroup B. The 9-bp deletion has been described in several other backgrounds, including the African haplogroup L0a2 (Bortolini et al. 1999).

Twelve mtDNA haplogroups were identified in the two Bantu populations. Only two subclades are shared: L3e1 and L3e2. This result shows important differences between the Bantu from Congo and those from Cameroon, but caution is necessary because this result can also reflect sampling, as the number of individuals investigated is low.

Table 1 shows that in Ribeirão Preto two L1c2 and L2a1 sequences are the same as two other sequences observed in Congo and Cameroon, respectively. Among the black Brazilian populations, only Cametá and Cajueiro shared one identical L2a1 sequence.

To better understand the phylogeographic scenarios of the mtDNA sub-Saharan haplogroups and their respective subclades found in Brazil, we grouped our data with data obtained by Bortolini et al. (1997) and Alves-Silva et al. (2000), according to the origin of the sampled individuals into four main geographic regions of country: north, northeast, southeast, and south. Table 2 shows these distributions and the estimates considering three major sub-Saharan groups: West-Central Bantu, East Bantu, and West Africa.

Haplogroups L2c and L1b are the most common among the West Africans (19% and 17%, respectively), but both have low frequencies among the Central-West and East Bantu speakers. Haplogroups L3b and L3d are also mainly found in West Africans. On the other hand, subclades L0a1 and L0a2 can be considered reliable Bantu markers, because they are not found in West Africans (in Africa the COII/tRNA^{lys} 9-bp deletion has been associated with subclade L0a2; Soodyal et al 1996; Bortolini et al. 1999; Kivisild et al. 2004; Plaza et al. 2004). The phylogeography of subclades L0a1 and L0a2 in Africa has been associated with the Bantu expansion from the Cameroon plateau, 3,500 years ago (Cavalli-Sforza et al. 1994; Salas et al. 2002; Plaza et al. 2004).

The origin of haplogroup L1c was postulated to be in Central Africa toward the Atlantic coast with a reasonable diffusion to the east (Salas et al. 2002). The virtual absence of this haplogroup and of its subclades, L1c1, L1c2, and L1c3, in West Africans makes them reliable Bantu markers. Haplogroup L3e is the most widespread, frequent, and ancient of the L3 haplogroups, comprising most of the L3 subtypes in sub-Saharan Africa (Salas et al. 2002). Subclade L3e1 is common among Central-West (11%) and East Bantu (9%) speakers, but it is rare in non-Bantus from West Africa; subclade L3e2 is found with a significant frequency only among the Central-West Bantus (12%).

Additional haplogroups and subhaplogroups are rare, and others are amply distributed in both Bantu and non-Bantu speakers. For example, haplogroup L2a is the most frequent and widespread mtDNA cluster in Africa (nearly one-fourth of all natives types; Salas et al. 2002). Subclade L2a1 has a similar distribution among Central-West Bantus (16%) and West Africans (15%), but in East Bantus its frequency is twice as high (34%). A West African origin of subclade L2a1 has been postulated, and its phylogeographic picture is also compatible with the earliest demographic Bantu dispersal (Pereira et al. 2001; Salas et al. 2002; Plaza et al. 2004).

Table 2.   Sub-Saharan mtDNA Haplogroups and Their Distributions (%)[a] in Four Brazilian Regions and in Bantu and Non-Bantu-Speaking Populations

| Haplogroup or Subclade | Africa[b] | | | Brazil[c] | | | | |
	Central-West Bantu (111)	East Bantu (416)	West Africa (348)	Southeast (51)	South (28)	Northeast (33)	North (19)	Total (131)
L0a		<1	1					
L0a1	4	10		6	14	19		10
L0a2	7	15		2	4			2
L0d		5						
L1b	4	1	17	4	4	3		3
L1c	4	<1						
L1c1	4	2	1	16			10	8
L1c2	6	2		8	14	12		9
L1c3	2	1						
L2	4	2	5	10	11	6		8
L2a1	16	34	15	8		18	40	13
L2b	4	1	6	2	7		10	4
L2c		1	19	4			10	3
L2d	6	1	2					
L3		1						
L3b	2	3	12		7		5	2
L3d	1	5	9	4	7	6	5	5
L3e1	11	9	1	14		21	11	
L3e2	12	1	5	12	24	9	10	14
L3e3	4	3	1	4			5	2
L3e4		<1	2			3		1
L3f	5	2	4	2	4	3		2
L4g	3			4	4		5	3
L5a	1	<1						

a. Total number of individuals studied is shown in parentheses.
b. West-Central Bantu: Mbundu and Bakongo (Angola) (Plaza et al. 2004); Bubi and Fang (Guinea Equatorial) (Salas et al. 2002); Congolese (Democratic Republic of Congo) (present study); and Bamileke, Bene, Eton, Nweh, Sonaga, and Etongo (Cameroon) (present study). East Bantu: Yao, Tonga, Shangaan, Chopi, Chwabo, Lomwe, Makonde, Makhuwa, Ndau, Nguni, Nyungwe, Nyanja, Ronga, Shona, Sena, and Tswa (Mozambique) (Salas et al. 2002); and other nondefined Bantu-speaking people from Mozambique (Pereira et al. 2001). West Africa: Hausa, Kanuri, Fulbe, Songhai, Yoruba, Senegalese, Serer, Wolof, and Mandenka (from Nigeria, Niger, Benin, Cameroon, Burkina Faso, and Senegal) (Salas et al. 2002).
c. Southeast: White, brown, and black Brazilians (Alves-Silva et al. 2000; present study). South: White, brown, and black Brazilians (Alves-Silva et al. 2000; Bortolini et al. 1997). Northeast: White, brown, and black Brazilians (Alves-Silva et al. 2000; Bortolini et al. 1997; present study). North: White, brown, and black Brazilians (Alves-Silva et al. 2000; present study).

Most of the African haplogroups within L0–L5 are present in at least one of the four Brazilian regions. Exceptions are represented by haplogroups L0a (frequency in East Bantu, <1%; West Africa, 1%), L0d (East Bantu, 5%), L1c (Central-West Bantu, 4%; East Bantu, <1%), L5 (Central-West Bantu, 1%; East Bantu, <1%), and L3 (East Bantu, 1%). However, the pattern of sub-Saharan mtDNA type distributions is clearly different in the four main geographic Brazilian regions. For example, subclade L2a1 was detected in three Brazilian regions; distributions range from 8% in the southeast to 40% in the north, but curiously subclade L2a1 does not appear in the southern region. In contrast, subclade L0a1, which can be taken as an East Bantu marker, apparently is not present in the northern region. Subclade L1c1 shows substantial frequencies in the southeastern (16%) and northern (10%) regions, but it is absent in the southern and northeastern mtDNA pools, whereas haplogroup L3b, with a predominantly West African distribution, is present only in the southern (7%) and northern (5%) populations. Although sampling errors cannot be ruled out, it is possible that our results reflect the different African sources that supplied slaves to the Brazilian regions. There are, however, some differences between the patterns revealed by our results and those expected on the basis of the historical register (Klein 2002), and these differences deserve consideration.

Because of geographic proximity and other factors, the northeastern and northern regions of Brazil received the largest number of West Africans who were forcibly moved to Brazil during the Atlantic slave trade. Interestingly, the typical West African markers—haplogroups L1b, L2c, and L3b (together they represent about 50% of the mtDNA haplogroups observed in region)—have a relatively low distribution frequency in the northeastern and northern regions (L1b + L2c + L3b = 3% and 15%, respectively). Several hypotheses could explain these last results: (1) Early (when laws prohibited the direct slave trade from Africa) and recent internal migrations between Brazilian regions camouflaged the original phylogeographic landscape; (2) the West Africa group of Senegalese (≅70% of 348 sequences shown in Table 1), particularly the Mandenka, who did not come to Brazil in higher numbers than other groups (such as the Yoruba, who have been less studied), is overrepresented; (3) there are ethnic sex-specific differences in the Atlantic slave trade, so that more West African men than women would have been brought to Brazil.

Finally, Table 3 provides a general view of the origins of the Africans who came to Brazil. The numbers show that West-Central Africa provided most of the African slaves to Brazil, as the historical sources indicate. However, the differences observed between the West-Central Bantu and the West African contributions (80% and 65%, respectively, according to historical data, and 15% and 30%, respectively, according to mtDNA data) reinforce scenario 3 suggested in the previous paragraph. Only additional studies using geography-specific African Y-chromosome haplogroups (Cruciani et al. 2004) can provide a more complete picture of the origin of African slaves and can answer other questions related to the Atlantic slave trade to Brazil.

Table 3. Origin of Slaves (%) Who Arrived in Brazil During the Atlantic Slave Trade Considering Genetic (mtDNA) and Historical Sources

Source	Central-West Bantu	East Bantu	West Africans
mtDNA[a]	80 ± 1.3	5 ± 1.5	15 ± 8.9
Historical[b]	$\cong 65$	$\cong 5$	$\cong 30$

a. Because most sub-Saharan mtDNA haplogroups and subhaplogroups are not geography specific, the estimates of the African contributions were obtained using the frequencies presented in Table 2 and Long's (1991) least-squares method.

b. Values obtained according to data presented by Klein (2002).

Literature Cited

Alves-Silva, J., M. S. Santos, P. E. M. Guimarães et al. 2000. The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.* 67:444–461.

Bandelt, H.-J., J. Alves-Silva, P. E. M. Guimarães et al. 2001. Phylogeography of the human mitochondrial haplogroup L3e: A snapshot of African prehistory and Atlantic slave trade. *Ann. Hum. Genet.* 65:549–563.

Bandelt, H.-J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37–48.

Bandelt, H.-J., P. Forster, B. C. Sykes et al. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753.

Bandelt, H.-J., L. Quintana-Murci, A. Salas et al. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. *Am. J. Hum. Genet.* 71:1150–1160.

Bortolini, M. C., W. A. Silva Jr., D. Castro-de-Guerra et al. 1999. African-derived South American populations: A history of symmetrical and asymmetrical matings according to sex revealed by bi- and uniparental genetic markers. *Am. J. Hum. Genet.* 11:551–563.

Bortolini, M. C., W. A. Silva Jr., M. A. Zago et al. 2004. The phylogeography of mitochondrial DNA haplogroup L3g in Africa and the Atlantic slave trade. *Am. J. Hum. Genet.* 75:522–524.

Bortolini, M. C., M. A. Zago, F. M. Salzano et al. 1997. Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. *Hum. Biol.* 69:141–159.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1994. *History and Geography of Human Genes.* Princeton, NJ: Princeton University Press.

Cruciani, F., R. La Fratta, P. Santolamazza et al. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and outside Africa. *Am. J. Hum. Genet.* 74:1014–1022.

Figueiredo, M. S., M. C. Silva, J. F. Guerreiro et al. 1994. The heterogeneity of the Beta S cluster haplotypes in Brazil. *Gene Geogr.* 8:7–12.

Kivisild, T., M. Reidla, E. Metspalu et al. 2004. Ethiopian mitochondrial DNA heritage: Tracking gene flow across and around the Gate of Tears. *Am. J. Hum. Genet.* 75:752–770.

Klein, H. S. 2002. As origens africanas dos escravos brasileiros. In *Homo brasilis*, S. D. J. Pena, ed. Ribeirão Preto, Brazil: FUNPEC, 93–112.

Lahiri, D. K., and J. I. Nurmberger. 1991. A rapid nonenzymatic method for preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res.* 19:5444.

Long, J. C. 1991. The genetic structure of admixed populations. *Genetics* 127:417–428.

Parra, F. C., R. C. Amado, J. R. Lambertucci et al. 2003. Color and genomic ancestry in Brazilians. *Proc. Natl. Acad. Sci. USA* 100:177–182.

Pena, S. D. J., and M. C. Bortolini. 2004. Pode a genética definir quem deve se beneficiar das cotas universitárias e demais ações afirmativas? *Estud. Avançados* 18:31–50.

Pereira, L., V. Macaulay, A. Torroni et al. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: Insights into the Bantu expansions and the slave trade. *Ann. Hum. Genet.* 65:439–458.

Plaza, S., A. Salas, F. Calafell et al. 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Hum. Genet.* 115:439–447.

Salas, A., M. Richards, T. De la Fe et al. 2002. The making of the African mtDNA landscape. *Am. J. Hum. Genet.* 71:1082–1111.

Salas, A., M. Richards, M. V. Larea et al. 2004a. The African diaspora: Mitochondrial DNA and the Atlantic slave trade. *Am. J. Hum. Genet.* 74:454–465.

Salas, A., A. Torroni, M. Richards et al. 2004b. Reply to Bortolini et al. *Am. J. Hum. Genet.* 75:524–526.

Salzano, F. M., and M. C. Bortolini. 2002. *Evolution and Genetics of Latin American Populations.* Cambridge: Cambridge University Press.

Shen, P., T. Lavi, T. Kivisild et al. 2004. Reconstruction of patri- and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Hum. Mut.* 24:248–260.

Silva, W. A., Jr., M. C. Bortolini, D. Meyer et al. 1999. Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable loci. *Am. J. Phys. Anthropol.* 109:425–437.

Soodyall, H., L. Vigilant, A. V. Hill et al. 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.

Telles, E. 2003. *Racismo à brasileira: Uma nova perspectiva sociológica.* Rio de Janeiro: Relume Dumará.

Ward, R. H., B. L. Frazier, K. Dew-Jager et al. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. USA* 88:8720–8724.

Zago, M. A., M. S. Figueiredo, and S. H. Ogo. 1992. Bantu Beta S cluster haplotype predominates among Brazilian blacks. *Am. J. Phys. Anthropol.* 88:295–298.

ANEXO 3

Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data

Hünemeier *et al*., manuscrito submetido à Am J Phys Anthropol

1

Niger-Congo speaking populations and the formation of the Brazilian gene pool:

mtDNA and Y-chromosome data

Tábita Hünemeier,[1] Cláudia Carvalho,[2] Andrea Rita Marrero,[1] Francisco Mauro Salzano,[1]

Sérgio Danilo Juno Pena,[2] and Maria Cátira Bortolini[1*]

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

[2]*Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil*

Number of text pages plus bibliography: 23; number of Tables: 4; number of figures: 3. One Appendix.

RUNNING HEADLINE: Y-SNP AND MTDNA VARIATION IN BLACK BRAZILIANS

KEY WORDS: uniparental genetic markers; migrant origins, gender-specific dispersal; African diaspora

*Correspondence to: Maria Cátira Bortolini, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: maria.bortolini@ufrgs.br.

ABSTRACT We analyzed sequence variation in the mitochondrial DNA (mtDNA) hypervariable segment I (HVS-I) from 201 Black individuals from two Brazilian cities (Rio de Janeiro and Porto Alegre), and compared these data with published information from 21 African populations. A subset of 187 males of the sample was also characterized for 30 Y-chromosome biallelic polymorphisms, and the data compared with those from 48 African populations. The mtDNA data indicated that 69% and 82% of the matrilineages found in Rio de Janeiro and Porto Alegre originated from West-Central/South-East Africa, respectively. These estimates are in close agreement with historical records showing that most of the Brazilian slaves who arrived in Rio de Janeiro were from West-Central Africa. In contrast to mtDNA, Y-chromosome haplogroup analysis did not allow discrimination between places of origin in West or West-Central Africa. Thus, when comparing these two major African regions, there seems to be higher genetic structure with mtDNA than with Y-chomosome data, suggesting a higher migration rate of the Niger-Congo speaking males than females.

3

The massive forced African migration to the Americas that occurred from the 15th to the 19th centuries (involving not less than about 10 million persons) included Brazil in a fraction estimated as 40% (Klein, 2002). Most of these individuals were men, since it was supposed that they would be more able to support the hard work in the farms and mineral mines (Bergmann, 1977; Conrad, 1985). Additionally, the slave trade to Middle East preferentially involved sub-Saharan women, and this resulted in a relatively reduced number of available slave women plus an increased price for them (Klein, 2002).

The Africans that were brought to Brazil as slaves originated mainly from West-Central/Southeast Africa and West Africa (Klein, 2002). The first region included basically the area represented presently by Angola, Republic of Congo, Democratic Republic of Congo and Mozambique, whereas the second coversthe region located at the north of the Gulf of Guinea (Fig. 1, which shows three other subdivisions). These regions have continental dimensions and are populated by very distinct peoples and cultures. There is, however, a relative linguistic unity since they are inhabited by speakers of languages belonging to the Niger-Congo linguistic sub-phylum, with exceptions like the Hausa, who speak an Afro-Asiatic language family belonging to the Chadic branch (Greenberg, 1963; Ruhlen, 1987). The Niger-Congo sub-phylum comprises the large Bantu branch, which includes about ~500 languages spoken in virtually all Central-South Africa, except for the area occupied by the Khoisan-speaking groups (Cavalli-Sforza et al., 1994).

The location of the Bantu languages origins has been identified as most likely being between Cameroon and Nigeria (Newman, 1995). The Bantu expansion age (about 3,400 ± 1,100 years before present) coincided with the end of the Neolithic and was apparently related to the diffusion of iron metallurgy and grain cultivation throughout sub-Saharan Africa (Murdock, 1959; Curtin et al., 1991; Phillipson, 1993; Cavalli-Sforza et al., 1994; Diamond and Bellwood 2003; Plaza et al., 2004; Zhivotovsky et al., 2004;Rexová et al., 2006). In contrast, in

the West African branch of the Niger-Congo sub-phylum, the native inhabitants speak several

non-Bantu languages (Greenberg, 1963; Ruhlen, 1987; Cavali-Sforza et al., 1994). Genetic

studies have demonstrated that Niger-Congo speaking populations are more related to each

other than to other Africans. These same investigations have also shown that Bantu speaking

groups show a higher level of genetic homogeneity than the non-Bantu populations (Cavalli-

Sforza et al., 1994; Poloni et al., 1997; Cruciani et al., 2002; Salas et al., 2002, 2005; Wood et

al., 2005).

Because historical records on slavery contain many gaps, genetic studies with South

Americans of African descent have been used to trace the ancestry of Niger-Congo speakers

back into Africa. Salzano and Bortolini (2002) showed that 61%, 34% and 3% of the $Hb\beta^S$

haplotypes found in Brazil as a whole are of the types named Bantu (or Central African

Republic-CAR), Benin and Senegal, respectively. These results are in good agreement with the

historical sources which indicate that ~68%, and ~32% of the African slaves were brought from

West-Central/Southeast and West Africa, respectively (Klein, 2002).

Recently, lineage markers [mitochondrial DNA (mtDNA) and the non-recombining

portion of the Y-chromosome] have been used to try to unravel the history of human

populations, since they are uniparentally transmitted, and escape recombination. These markers

allow the reconstruction of unequivocal haplotype phylogenies, which can be related to

geographic distributions, in an approach known as phylogeography (Avise, 2000).

Alves-Silva et al., (2000) reported an initial landscape about the phylogeography of the

African mtDNA haplogroups in Brazil as a whole. Haplogroups L3e and L1c together

constituted approximately 49% of the African fraction of mtDNAs analyzed by these authors.

These results suggested that the majority of the mtDNA lineages of African ancestry in their

Brazilian sample would have a origin in West-Central Africa, with a minor contribution from

the Southeast, although a substantial number could also have come from West Africa through non-Bantu speaking carriers.

Salas et al., (2004) estimated for the first time the quantitative contribution of the different African regions to the formation of the New World mtDNA gene pool. According to their estimate, 65% of the types found in South America would have a West-Central African origin, its complementary value indicating a West African contribution. These numbers are particularly different from those obtained for Central America (41% West-Central, 59% West), and North America (28% West-Central, 72% West), in agreement with the historical data of these regions (Curtin 1969;Thomas 1998). Using the same kind of approach but substantially more data, Salas et al., (2005) estimated that > 55% of the U.S. mtDNA lineages have a West African ancestry, with < 41% coming from West-Central or Southwestern Africa, results which are close to the historical record (McMillin, 2004).

However, the most recent investigation of mtDNA-HVS I variation in Brazilian populations has yielded discrepancies between the patterns obtained with the mtDNA haplogroup distributions and the historical sources (Silva et al., 2006). These findings raised the suggestion of a possible geographical-gender specific difference, with a proportionally larger number of West-African men than women compulsorily migrating to Brazil (Silva et al., 2006). The authors mentioned that only research with Y-chromosome markers could provide a more complete picture about this and other questions related to the Atlantic slave trade to Brazil.

Several studies of Y-chomosome phylogeographical landscape in Africa are now available (Cruciani et al., 2004, Luis et al., 2004, Beleza et al., 2005, Wood et al., 2005), but up to now no investigation has evaluated the same set of markers in males from the three Americas.

Here we provided information about the distribution of the mtDNA and Y-chromosome haplogroups in two Brazilian Black populations, and compared these results with those

6

published for populations of several African regions. The questions asked were: (a) Can these two sets of data furnish information about possible regions of origin of the African slaves who arrived in Brazil? (b) Do they show the same distribution pattern both for the Brazilian and African populations? and (c)In which way do possible differences throw light on gender - specific patterns of migration?

SUBJECTS AND METHODS

Populations

After appropriate informed consent, samples of 201 individuals classified as Black according to their physical appearance and originating from two Brazilian cities, Rio de Janeiro (N=94), the capital of Rio de Janeiro state, and Porto Alegre (N= 107) the capital of Rio Grande do Sul, the southernmost state of Brazil were studied. Rio de Janeiro, plus the northeastern cities of Salvador (state of Bahia) and Recife (state of Pernambuco) were the most important ports of arrival of slaves in Brazil. From these centers these persons would be distributed to the other provinces. For example, according to historical data, 88% of the Rio Grande do Sul slave population was brought from Rio de Janeiro, with the complementary number of slaves being brought from other Brazilian provinces and Uruguay, not directly from Africa (Maestri-Filho, 1993; Berute, 2006).

mtDNA

The nucleotide sequence of the first hypervariable segment (HVS-I) of 213 individuals was amplified and sequenced according to conditions described in Marrero et al., (2005). Both strands of DNA were sequenced.

The information provided by HVS-I was used to classify the lineages into haplogroups according to Salas et al., (2002, 2004), with two exceptions: (a) Haplogroup L3g shares motifs, within HVS-I and HVS-II and at positions 769 and 1018 with L4a. This information led Kivisild et al., (2004) to suggest that L3g is actually a sister cluster of L4a; therefore, they proposed to rename it L4g; and (b) Haplogroup L1e has been recently redefined as L5a because it occupies an intermediate phylogenetic position between L1 and L2'L3 major haplogroups (Shen et al., 2004).

Although the hierarchical relation between the human mtDNA lineages is well known, the terminology to define them remains confusing. Haplogroup, clade, sub-haplogroup and sub-clade are words frequently used as synonymous. In this paper, to facilitate reading and comprehension, the term haplogroup will be used to define the major lineages (L0, L1, L2, L3, L4, L5, etc.), as well as their derivations (L2a, L3e, L2a1, L3e1, etc.).

Y-chromosome markers

The male fraction of our sample (N = 187) was studied for thirty biallelic Y-chromosome polymorphisms (92R7, M9, M3, M19, M242, RPSY711, M17, M173, SRY2627, PN2, M2, M174, M145, M33, M35, M75, M58, M191, M149, M116.2, M10, M78, M154, M155, M281, M123, M81, M213, M60, V6) using hierarchical strategies plus RFLP and mini-sequencing methods as described in Bortolini et al. (2003) and developed by Carvalho and Pena (2005), respectively. These markers define the major European, Amerindian and African haplogroups, but identify especially well sub-types of the haplogroup E, the most common and widespread Y chromosome in Africa.

The haplogroup nomenclature adopted is that proposed by the last Y-chromosome Consortium release (Jobling and Tyler-Smith, 2003). Here also the term haplogroup will be

8

used to define the major lineages (E, etc.), as well as their derivations (E1, E2, E3, E3a, E3a7, etc.).

Data analyses

The mtDNA sequences were checked manually, validated with the help of the CHROMAS LITE 2.0 program (www.technelsyum.com.au) and aligned with the revised Reference Sequence (rCRS, Andrews et al., 1999) using the BIOEDIT software (Hall, 1999). Since artifacts ("phantom mutations") can be introduced during the sequencing and editing process, we applied the filtering procedure described by Bandelt et al., (2002) and used criteria like those of Yao et al. (2004) to check for the quality of the sequences. After filtering a network of sequences was constructed with the NETWORK 4.1.1.2. program (www.fluxus-engineering.com) using the median-joining algorithm. Weight networks showing perfect star tree patterns are expected when the data are potentially free of phantom mutations. However, other criteria as phylogenetic analysis in comparisons with closely related sequences from other databases must be observed to guarantee the quality of the data (Yao et al., 2004).

Estimates of geographic parental contributions considering our mtDNA and Y-SNP data were calculated using the weighted least square method (Long, 1991) performed with the ADMIX program, kindly made available by Dr. J.C. Long. The relationships among the populations were examined using the D_A distance and the neighbor-joining method (Nei et al., 1983; Saitou and Nei, 1987; Nei and Roychoudhoury, 1993). The trees were visualized using the TreeView program(Page, 1996), version1.6.6 (http://taxonomy.zoology.gla.ac.uk/rod/rod.html). Analyses of molecular variance (AMOVA), implemented in the Arlequin program (Excoffier et al., 1992), version 3.01 (http://cmpg.unibe.ch/software/arleqin3/), were performed to test the hypothesis of differentiation between populational groups of particular interest observed in the phylogenetic trees.

RESULTS

mtDNA

The networks obtained for the HVS-I weighty variation showed perfect star trees. Associated with other criteria (see Subjects and methods), they indicate that our HVS-I data sets are potentially free of artifacts (data not shown).

Ninety percent and 79% of the mtDNA sequences found in Blacks from Rio de Janeiro and Porto Alegre, respectively, are estimated as having an African origin. The remaining sequences were identified with haplogroups of European (2% and 6%, respectively) or Amerindian (8% and 15%) origin (data not shown). Table 1 presents the mtDNA haplogroup distributions for the two Brazilian Black samples and for 21 African populations. About 70% of the haplogroups present in these African groups can also be seen in Brazil, while all haplogroups observed in these two Brazilian Black samples can be found in Africa. Table 1 also shows that there are similarities of haplogroup frequencies between the West and West-Central regions of Africa in comparison with other major regions of the continent (Salas et al., 2005), probably reflecting genetic similarity within the Niger-Congo linguistic sub-phylum. However, some haplogroups are present only in West-Central and/or Southeast Africa (L3e1a, L5a1, L0d, L0d1, L0d2), whereas others seem to be exclusive of West Africa (L2a-α2, L2c1, L2d2, L3b1). Many haplogroups show striking differences in their distributions. For example, the cumulative frequency of L1b1 in the West (12.3%) is about 7 times higher than that found in West-Central/Southeast Africa (1.7%). Ancient or more recent (but not less complex) demographic events have been related to these particular mtDNA haplogroup distributions across Africa (Salas et al., 2002).

Of special interest is the presence of haplogroup L0d1 in Rio de Janeiro. This and other related haplogroups (L0d, L0d2) are characteristic of southern African Khoisan-speaking

10

groups, but are also present in Mozambique, probably due to admixture between Khoisan women and Bantu Southeast men (Salas et al., 2002). The sequence observed in Rio de Janeiro is the same as that described by Salas et al., (2002) in Mozambique, suggesting that the occurrence of L0d1 in Brazil is probably due to the direct slave trade from this former African Portuguese colony to Brazil.

Using the haplogroup distributions presented in Table 1 we constructed a tree which showsthree well defined clusters (Fig 2). One of them (A) groups all West Africans; another (B) clusters the West-Central/Southeast Africans with the two Brazilian Black populations. Note the proximity of the latter with the two former Portugal colonies, Angola and Mozambique. A third, intermediate and more restricted cluster (C), is represented by three populations from Cameroon (Bassa, Bakaka and Fulbe). Cameroon is geographically located in the probable center of spread of the Bantu languages and is positioned exactly between Western and Western-Central African regions. As a consequence it contains both Bantu (Bakaka and Bassa) and non-Bantu (Fulbe) speaking populations. Figure 2 also suggests a genetic differentiation within the Niger-Congo sub-phylum, separating the West-Central/Southeast Bantu speakers (Fang, Cabinda, Bubi, Angola, Mozambique) from the Western non-Bantu speakers (Yoruba, Kanuri, Fulbe, Shongai, Senegalese, Limba, Temne, Mende, Loko, Wolof, Mandenka, Serer).

Using an analysis of molecular variance (AMOVA) we tested the hypothesis of differentiation between these two major geographical groups (excluding the Brazilian samples). The value obtained, although low, is significant: $\Phi_{CT} = 0.025$, $P < 10^{-4}$.

Y-chromosome biallelic markers

Table 2 shows that 56% and 36% of the Y-chromosomes from Rio de Janeiro and Porto Alegre respectively have an African origin. All the other Y-chromosomes (44%) found in Rio

de Janeiro have a probable European origin, while for Porto Alegre 5% and 59% have a probable Amerindian and European origins, respectively (data not shown).

E3a* is the most frequent African chromosome found in our Brazilian sample, followed by E3a7. With the exception of E3b2, all African haplogroup E chromosomes found in Brazil are also present in sub-Saharan Africans. E3b2 has been described in high frequencies in North African populations, particularly among the Berber (Cruciani et al., 2002; Luis et al., 2004; Semino et al., 2004). However, the presence of the E3b2 chromosome in Brazil is most likely related to Iberian men, since typical Berber Y-chromosomes have been reported in Portuguese/Spanish populations. The existence of a common genetic background between Berbers and Iberians probably reflects the genetic impact of the Islamic occupation of the Iberian Peninsula for 7 centuries (Carvalho-Silva et al., 2001; Lucotte et al., 2001; Bortolini et al., 2004b; Cruciani et al., 2004; Semino et al., 2004; Gonçalves et al., 2005).

Since few African populations have been studied with the same set of Y-SNPs used here, we assembled the haplogroups according to a hierarchical strategy. This procedure allowed the comparison of our results with those from 48 African populations, including 36 Niger-Congo speaking groups (Table 3). Afterwards, this information was used to obtain a distance matrix and a neighbor-joining tree (Fig. 3), which shows a clear split separating the Niger-Congo speakers (cluster B) from the other Africans (Afro-Asiatic and Nilo-Saharan speakers; cluster A). But there are some exceptions (the Massai and Luo from Kenya clustered together with Niger-Congo speakers, whereas Mixed-Adamawa, Fulbe-Cameroon and Tupuri grouped with the Afro-Asiatic speakers). The two Black Brazilian populations are closely related to each other and with the Niger-Congo speaking-populations. The Niger-Congo cluster, however, does not show internal structure in accordance with geography or language, a pattern which differs from that observed with mtDNA. The same tendency was observed when just Niger-Congo populations were considered in the analysis (data not shown).

Using the populations from West and West-Central/Southeast Africa given in Table 3 (excluding those from Cameroon, see comment above) we obtained a value of $\Phi_{CT} = 0.006$; P > 5%, *i.e.*, no Y-chromosome differentiation between West-Central/Southeast (Bantu) and West (non-Bantu) men.

Admixture analysis

The two major population groups observed in our mtDNA phylogenetic tree (West-Central/Southeast Bantu speakers and Western non-Bantu speakers), which showed significant differences in their mtDNA haplogroup distributions, were used as parental stocks in the admixture analysis, using the data presented in Table 1 and a least squares approximation. The West-Central and Southeast African maternal contribution was majority (69% for Rio de Janeiro; 82% for Porto Alegre), whereas the complementary numbers can be attributable to the West African contribution (Table 4). These admixture values are very similar to those suggested by the historical records. These findings could reflect the absence of major geographic gender-specific differences in the Atlantic slave trade (as mentioned in the introduction) in disagreement with another data set (Silva et al., 2006). Although sampling error cannot be discarded, the discrepancy between the two studies may be due to the different African sources that supplied slaves to the several Brazilian regions, and/or to different patterns of the slave trade for each of them.

DISCUSSION

As expected, estimates of the African contribution to the Black Brazilian mtDNA gene pool (79%-90%) are larger than those obtained for populations identified as White in the different Brazilian regions, where the proportion of African mtDNA lineages ranged from 0 to 44% (Alves-Silva et al., 2000; Marrero et al., 2005). From these results a picture emerges, that the contemporary Brazilian population presents the most important reservoir of African mtDNA lineages out of Africa. It has been estimated that at least 90 million persons in Brazil, independently of their physical appearance, show mtDNAs of sub-Saharan African origin (Pena and Bortolini 2004). This particularity allowed inferences not only about the probable mtDNA lineage composition of populations from African regions hitherto not studied (Alves-Silva et al., 2000), but also about possible evolutionary and demographic events mediated by women, which who should have occurred in Africa (Bortolini et al., 2004a).

On the other hand, some demographic and historical circumstances related to Brazil´s colonization determined that the first Brazilians arose mostly by the union between Portuguese males and Amerindian or African females (Bortolini et al., 1997; Carvalho-Silva et al., 2001; Salzano and Bortolini, 2002). These asymmetrical matings determined that most of the Y-chromosomes of contemporary Brazilian populations have an European origin. The present results indicated that, although the proportion of the typical African chromosomes in the Black samples (36%-56%) are much higher than those obtained for Brazilian populations identified as White (0 to ~5%; Carvalho-Silva et al., 2001; Abes-Sandes et al., 2004; Marrero et al., 2005), the amount of Y-chromosomes of European origin is striking.

Restricting the attention to Africa, the clear geographic or language structure observed within the Niger-Congo cluster with mtDNA, but not with the Y-SNP data, deserves additional investigation.

14

Cruciani et al., (2002) suggested that the absence of Y-chromosome differentiation in Africa could be due to relatively recent range expansion(s). E3a* chromosomes could have been already present along the Western region and then spread to South Africa through the Bantu expansion. This haplogroup was also observed in high frequencies among hunter-gatherer populations, like the Biaka/Mbuti and Khoisan-speaking people, probably due to admixture between Bantu-speaking men and Biaka/Mbuti/Khoisan women. The M191 mutation, which defines haplogroup E3a7, probably arose in West-Central Africa. A later demic expansion should have brought E3a7 chromosomes from West-Central to Western Africa (Cruciani et al., 2002). Studies with Y-SNPs associated with the fast-evolving Y microsatellite loci (STRs) revealed that the STR variation is structured within SNP-haplogroups. However, this variation is also not well correlated with geography (Scozzari et al., 1999; Cruciani et al., 2002). These results reinforce the idea that the present differences between HVS-I mtDNA and Y-SNP data is not an artifact related to different mutation rates.

A first important implication of these findings is that E3a* should be interpreted as a Niger-Congo marker. Although the presence of E3a* in Central to South Africa can be associated with the Bantu expansion, this chromosome may have existed for at least ~11,000 years before the spread of the carriers of the Bantu languages (Scozzari et al., 1999). Probably E3a* was the most common chromosome in West Africa at the time of the Niger-Congo language emergence. A second implication is that these demic expansions in Africa, including the Bantu dispersion, probably did not involve a higher migration rate of Niger-Congo speaking women than men, but maybe the opposite, or at least the same female/male migration rate.

Seielstad et al., (1998) suggested that due mainly to the widespread practice of patrilocality (in which women move into their husband´s residences after marriage) the rate of human migration among populations could be nearly eight times higher for females than males. Mesa et al. (2000), however, demonstrated that this situation is not universal, and their findings

15

were later confirmed (Wilder et al., 2004). Actually Hammer et al., (2001) suggested that sub-Saharan Africans might represent a case in which the genetic structure of human populations has been shaped by greater male mobility. Of course, the absence of any evidence for a higher migration rate for females compared to males on a global scale does not contradict the evidence for patrilocality effects at local scales (Wilder et al., 2004), which have been described in several agriculturalist sub-Saharan groups (Destro-Bisol et al., 2004). Recently, Wilkins and Marlowe (2006) proposed a model in which female-biased migration would be a recent phenomenon over most of human history, associated to changes due to the transition from a forager mobile to a sedentary agricultural lifestyle.

CONCLUSIONS

The questions asked in the introduction can now be answered: (a) The mitochondrial DNA results basically confirmed the historical evidence that the main source of African migration to Brazil originated from West-Central and Southeast Africa; (b) The Y-chromosome results, however, were inconclusive; (c) Although other explanations can be advanced, the absence of structure observed in the Y-chromosome pattern of distribution in Africa suggests that males, not females had higher migration rates in the past, at least among Niger-Congo-speaking populations.

ACKNOWLEDGMENTS

LITERATURE CITED

Abe-Sandes K, Silva-Jr WA, Zago MA. 2004. Heterogeneity of the Y chromosome in Afro-Brazilian populations. Hum Biol 76: 77-86.

Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt H-J, Pena SD, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461.

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.

Avise JC. 2000. Phylogeography: the history and formation of species. Cambridge: Harvard University Press.

Bandelt H-S, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71:1150–1160.

Beleza S, Gusmão L, Amorim A, Carracedo A, Salas A. 2005. The genetic legacy of western Bantu migrations. Hum Genet 117: 366-375.

Bergmann M. 1977. Nasce um povo. Petrópolis: Editora Vozes.

Berute GS. 2006. Características dos escravos traficados para o Rio Grande de São Pedro (1788-1802). Humanas (in press).

Bortolini MC, Zago MA, Salzano FM, Silva-Junior WA, Bonatto SL, Silva MC da, Weimer TA. 1997. Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. Hum Biol 69: 141-159.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse, Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra, D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73: 524-539.

Bortolini MC, Silva-Junior WA, Zago MA, Elion J, Krishnamoorthy R, Gonçalves VF, Pena SD. 2004a. The phylogeography of mitochondrial DNA haplogroup L3g in Africa and the Atlantic slave trade. Am J Hum Genet 75: 522-524.

Bortolini MC, Thomas MG, Chikhi L, Aguilar JA, Castro-de-Guerra D, Salzano FM, Ruiz-Linares A. 2004b. Ribeiro´s typology, genomes, and Spanish colonialism, as viewed from Gran Canária and Colombia. Genet Mol Biol 27: 1-8.

Carvalho CM, Pena SD. 2005. Optimization of a multiplex minisequencing protocol for population studies and medical genetics. Genet Mol Res 4: 115-125.

Carvalho-Silva DR, Santos FR, Rocha J, Pena SD. 2001. The phylogeography of Brazilian Y-chromosome lineages. Am J Hum Genet 68: 281-286.

Cavali-Sforza LL, Menazzi P, Piazza A. 1994. The history and geography of human genes. Princeton: Princeton University Press.

Coia V, Destro-Bisol G, Verginelli F, Battagglia C, Boschi I, Cruciani F, Spedini G, Comas D, Callafel F. 2005. mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. Am J Phys Anthropol 128: 678-681.

Conrad RE. 1985. Tumbeiros, o tráfico de escravos para o Brasil. São Paulo: Ed. Brasiliense.

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am J Hum Genet 70: 1197-1214.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R. 2004. Phylogeographic analysis of haplogroup E3b

18

(E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet 74: 1014-1022.

Curtin P. 1969. The Atlantic slave trade: a census. Madison: University of Wisconsin Press.

Curtin P, Feierman S, Thompson L, Vansina J. 1991. African history. London and New York, Longman.

Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Cagliá A, Tofanelli S, Spedini G, Capelli C. 2004. Variation of female and male lineages in Sub-Saharan populations: the importance of sociocultural factors. Mol Biol Evol 21: 1673-1682.

Diamond J, Bellwood P. 2003. Farmers and their languages: the first expansion. Science 300: 597-603.

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

Gonçalves R, Freitas A, Branco M, Rosa A, Fernandes AT, Zhivotovsky LA, Underhill PA, Kivisild T, Brehm A. 2005. Y-chromosome lineages from Portugal, Madeira and Açores record elements of Sephardim and Berber ancestry. Ann Hum Genet 69: 443-454.

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier, L. 1995. Evolutionary correlation between control region and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12: 334-345.

Greenberg JH. 1963. The languages of Africa. Bloomington: Indiana University Press.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41: 95-98.

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL. 2001. Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol 18: 1189-1203.

Jackson BA, Wilson JL, Kirbah S, Sidney SS, Rosenberg J, Bassie N, Alie JA, McLean DC, Garvey WT, Ely B. 2005. Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. Am J Phys Anthropol 128: 156-163.

Jobling M, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 75: 752-770.

Klein HS. 2002. As origens africanas dos escravos brasileiros. In: Pena SDJ, editor. *Homo brasilis*. Aspectos genéticos, lingüisticos, históricos e socioantropológicos da formação do povo brasileiro. Ribeirão Preto: FUNPEC Editora. p 93-112.

Long JC. 1991. The genetic structure of admixed populations. Genetics 127: 417-428.

Lucotte G, Gerard N, Mercier G. 2001. North African genes in Iberia studied by Y-chromosome DNA haplotype V. Hum Immunol 62: 885-888.

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ. 2004. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet 74: 532-544.

Maestri-Filho M. 1993. O escravo gaúcho: persistência e trabalho. Porto Alegre: Editora da UFRGS.

Marrero AR, Das Neves Leite FP, De Almeida Carvalho B, Peres LM, Kommers TC, Da Cruz IM, Salzano FM, Ruiz-Linares A, Silva-Junior WA, Bortolini MC. 2005. Heterogeneity

20

of the genome ancestry of individuals classified as White in the state of Rio Grande do

Sul, Brazil. Am J Hum Biol 17: 496-506.

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.

McMillin JA. 2004. The final victims: foreign slave trade to North America 1783-1810.

Columbia: University of South Carolina Press.

Mesa NR, Mondragon MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, Garcia LF, Velez

ID, Bravo ML, Munera JG, Bedoya G, Bortolini MC, Ruiz-Linares A. 2000. Autosomal,

mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of

gene flow in South America. Am J Hum Genet 67: 1277-1286.

Murdock GP. 1959. Africa. Its peoples and their culture history. New York: McGraw-Hill.

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular

data. J Mol Evol 19: 153-170.

Nei M, Roychoudhury AK. 1993. Evolutionary relationships of human populations on a global

scale. Mol Biol Evol 10: 927-943.

Newman J. 1995. The peopling of Africa: a geographic interpretation. New Haven: Yale

University Press.

Page RDM. 1996. TREEVIEW: an application to display phylogenetic trees on personal

computers. Comp Appl Bios 12: 357-358.

Pena SD, Bortolini MC. 2004. Pode a genética definir quem deve se beneficiar das cotas

universitárias e demais ações afirmativas? Est Avanç (São Paulo) 18: 1-20.

Phillipson DW. 1993. African archaeology. Cambridge: Cambridge University Press.

Pinto F, González AM, Hernández M, Larruga JM, Cabrera VN. 1996. Genetic relationship between the Canary islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann Hum Genet 60: 321-330.

Plaza S, Salas A, Calafell F, Corte-Real F, Bertranpetit J, Carracedo A, Comas D. 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. Hum Genet 115: 439-447.

Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L. 1997. Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. Am J Hum Genet 61: 1015-1035.

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 65: 531-550.

Rexová K, Bastin Y, Frynta D. 2006. Cladistic analysis of Bantu languages: a new tree based on combined lexical and grammatical data. Naturwissenschaften 93: 189-194.

Ruhlen M. 1987. A guide to the world´s languages. Standford: Standford University Press.

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

Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diaz P, Macaulay V, Carracedo A. 2002. The making of the African mtDNA landscape. Am J Hum Genet 71: 1082-1111.

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74: 454-465.

Salas A, Carracedo A, Richards M, Macaulay V. 2005. Charting the ancestry of African Americans. Am J Hum Genet 77: 676-680.

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

Scozzari R, Cruciani F, Santonamazza P, Malaspina P, Torroni A, Sellitto D, Arredi B, Destro-Bisol G, De Stefano G, Richards O, Martinez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novelletto A. 1999. Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. Am J Hum Genet 65: 829-846.

Seielstad MT, Minch E, Cavalli-Sforza LL. 1998. Genetic evidence for a higher female migration rate in humans. Nat Genet 20: 278-280.

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS. 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74: 1023-1034.

Shen P, Lavi T, Kivisild T, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman M, Oefner P. 2004. Reconstruction of patri- and matri-lineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. Hum Mut 24: 248-260.

Silva WA, Bortolini MC, Schneider MPC, Marrero AR, Elion J, Krishnamoorthy R, Zago MA. 2006. mtDNA haplogroup analysis of Black Brazilian and sub-Saharan populations: implications for the Atlantic slave trade. Hum Biol 78: 29-41.

Thomas H. 1998. The slave trade - the history of the Atlantic slave trade: 1440-1870. London: Macmillan.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC. 1991. African populations and the evolution of mitochondrial DNA. Science 253: 1503-1507.

Watson E, Forster P, Richards M, Bandelt H-J. 1997. Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61: 691-704.

Wilder JA, Mobasherm Z, Hammer MF. 2004. Genetic evidence for unequal effective population sizes of human females and males. Mol Biol Evol 21: 2047 – 2057.

Wilkins JF, Marlowe FW. 2006. Sex-biased migration in humans: what should we expected from genetic data? BioEssays 28: 290-300.

Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, Louie L, Bamshad M, Strassmann BI, Soodyall H, Hammer MF. 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. Eur J Hum Genet 13: 867-876.

Yao YG, Bravi CM, Bandelt H-J. 2004. A call for mtDNA data quality control in forensic science. Forensic Sci Int 141:1-6.

Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Young KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population divergence time. Am J Hum Genet 74: 50-61.

24

TABLE 1. Major sub-Saharan African mtDNA haplogroups and their distributions in two Brazilian and twenty-one African populations

Column groupings: **Brazil[1]** = POA, RJ. **Africa[2]** — Niger-Congo speakers: West-Central = columns 1–7, Southeast = column 8, West = columns 9–20; Afro-Asiatic speakers West = column 21.

Haplogroups	POA	RJ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
L0									0.022														
L0a										0.006		0.125	0.033					0.018					
L0a1	**0.059**	**0.071**	0.045		0.111	0.068	0.023	0.022		0.097									0.018	0.033	0.015	0.051	
L0a2		**0.048**	0.068	0.111		0.029	0.111			0.177													
L0d										0.038													
L0d1		**0.012**								0.013													
L0d2										0.003													
L1b	**0.047**	**0.095**	0.045			0.020				0.006		0.026		0.067	0.048		0.027	0.027	0.075	0.100	0.046	0.017	
L1b1					0.111	0.010	0.066	0.067	0.129	0.006	0.100	0.102	0.125	0.135	0.024	0.210	0.163	0.178	0.205	0.133	0.015	0.103	
L1c		**0.048**			0.111	0.029			0.064	0.003										0.067		0.026	
L1c1	**0.070**	**0.048**	0.045	0.028	0.111	0.098	0.067	0.111		0.019			0.125	0.033	0.048				0.018	0.068	0.046	0.043	
L1c2	**0.059**	**0.071**	0.093		0.111	0.126	0.088	0.133	0.032	0.019													
L1c3	**0.047**		0.023	0.028		0.010				0.013													
L2							0.023	0.022	0.165														
L2a-α1		**0.024**	0.045			0.020				0.006	0.200	0.051		0.033			0.027	0.054	0.018	0.033			
L2a-α2							0.023	0.022	0.032						0.024		0.054			0.100	0.031	0.043	
L2a-α3										0.010						0.052		0.027	0.037	0.068	0.077	0.043	
L2a1a	**0.047**	**0.059**	0.068							0.091	0.100	0.026	0.125	0.033	0.024								0.067
L2a1b										0.129								0.018					
L2a1-β1	**0.119**	**0.117**	0.023	0.222		0.010	0.066	0.044		0.031		0.051		0.100	0.024	0.158	0.108	0.036	0.111	0.033	0.077	0.008	
L2a1-β2	**0.012**	**0.012**	0.045				0.023	0.111	0.064	0.010				0.067	0.048	0.052	0.081		0.018	0.033	0.046	0.051	0.067
L2a1-β3	**0.023**		0.068			0.010				0.013		0.102	0.125	0.033	0.071		0.027	0.009	0.018			0.034	0.196
L2b	**0.070**	**0.024**	0.045	0.028		0.010	0.023	0.022	0.032	0.013					0.071		0.081		0.018	0.033	0.015		
L2b1		**0.048**				0.059		0.044	0.032					0.033	0.048	0.215	0.108	0.027			0.031	0.008	
L2c		**0.024**										0.051		0.071	0.052		0.027	0.293	0.037		0.124	0.120	0.067
L2c1																		0.027	0.062				0.017
L2c2						0.020				0.006					0.024			0.027	0.018	0.033	0.062	0.017	
L2d1				0.194						0.010							0.048					0.026	0.067
L2d2											0.100						0.081		0.056	0.033	0.015	0.034	
L3	**0.012**	**0.012**					0.132	0.180	0.193	0.013									0.037				
L3b	**0.047**	**0.036**	0.023			0.039	0.023	0.022	0.064	0.025	0.100	0.051	0.250		0.165		0.054	0.027	0.094	0.100	0.124	0.094	0.067
L3b1									0.032		0.100	0.102		0.033	0.071	0.105		0.027	0.018		0.046	0.008	0.067
L3b2	**0.023**								0.032									0.009					0.067
L3d	**0.012**	**0.024**	0.023			0.010				0.025	0.200	0.102		0.133	0.024		0.054	0.045	0.056	0.033	0.046	0.068	0.067
L3d1	**0.023**	**0.012**	0.023			0.010				0.038		0.051											
L3d2		**0.048**				0.010									0.024	0.052		0.036	0.018	0.100	0.015	0.043	
L3d3			0.023			0.030				0.003								0.026					
L3e1	**0.059**	**0.024**	0.068	0.083		0.049	0.088	0.089		0.028				0.033							0.031		
L3e1a	**0.059**	**0.012**	0.045			0.039				0.038													
L3e1b		**0.012**		0.028		0.039				0.028													
L3e2	**0.023**		0.023	0.250	0.111	0.078	0.044		0.129	0.010	0.100	0.157		0.100	0.024		0.054	0.009			0.015	0.008	0.067
L3e2b	**0.023**																		0.056			0.078	
L3e3	**0.047**	**0.071**	0.068	0.028		0.020	0.066	0.022		0.038				0.067		0.052		0.009					0.067
L3e4						0.010	0.088								0.024	0.052	0.027	0.036			0.077		
L3f	**0.036**	**0.024**	0.023		0.334	0.137	0.023	0.067		0.025	0.100	0.128	0.125	0.067	0.095								0.067
L3f1	**0.047**	**0.024**								0.010									0.056		0.046	0.060	
L3g (L4g)	**0.036**		0.045			0.010	0.023												0.018				
L5a1 (L1e)			0.023																				

25

[1] POA: Porto Alegre, N= 85; RJ : Rio de Janeiro, N= 84.

[2] The numbers correspond to the following African populations: 1-Angola, N=44, Plaza et al. (2004); 2- Bubi, N=36 (Equatorial Guinea) Mateu et al. (1997); 3-Fang, N=9 (Equatorial Guinea), Pinto et al. (1996); 4-Cabinda, N= 101 (Cabinda, former Portuguese protectorate), Beleza et al. (2005); 5- Bakaka, N=44 (Cameroon), Coia et al. (2005); 6-Bassa, N= 45 (Cameroon), Coia et al. (2005); 7- Fulbe, N= 31 (Cameroon), Coia et al. (2005); 8- Mozambique, N=307, Salas et al. (2002); 9- Kanuri, N=10 (Niger, Nigeria), Watson et al. (1997); 10- Fulbe, N= 39 (Nigeria, Niger, Benin, Cameroon, Burkina Faso), Watson et al. (1997); 11-Songhai, N=8 (Nigeria, Niger, Mali), Watson et al. (1997); 12-Yoruba, N= 30 (Nigeria), Watson et al. (1997), Vigilant et al. (1991); 13- Senegalese, N= 42 (Senegal), Rando et al. (1998); 14- Serer, N= 19 (Senegal), Rando et al. (1998); 15-Wolof, N= 37 (Senegal), Rando et al. (1998); 16- Mandenka, N= 112 (Senegal), Graven et al. (1995); 17- Mende, N=54 (Sierra Leone), Jackson et al. (2005); 18- Loko, N=30 (Sierra Leone), Jackson et al. (2005); 19- Limba, N=65 (Sierra Leone), Jackson et al. (2005); 20- Temne, N=117 (Sierra Leone), Jackson et al. (2005); 21-Hausa, N=15 (Niger, Nigeria), Watson et al. (1997).

26

TABLE 2. Distributions (in %) of the B*, D* and E* Y-chromosome haplogroups in two Brazilian and twenty-one African populations[1]

Population (country)	N	Haplogroup[2]																				
		E3* (PN2)	E3a* (M2)	E3a1 (M58)	E3a2 (M116.2)	E3a3 (M149)	E3a4 (M154)	E3a5 (M155)	E3a6 (M10)	E3a7 (M191)	E1* (M33)	E2* (M75)	E3b* (M35)	E3b1* (M78)	E3b2* (M81)	E3b3 (M123)	E3b4 (M281)	E-V6[3]	YAP(xDE) (M145)	D* (xE) (M174)	B* (M60)	Others
Porto Alegre (Brazil)	57		16							9			3	3	2						3	64
Rio de Janeiro (Brazil)	130		33	1						12	2	2	4		1						1	44
Niger-Congo speakers																						
West																						
Mossi (Burkina Faso)	49	2	68							22		4	2				ND	ND				2
Rimaibe (Burkina Faso)	37	3	52	5						8	5	27					ND	ND				
Fulbe (Burkina Faso)	20		90								10						ND	ND				
Fon (Benin)	100		38		ND	ND	ND	ND		57		5						ND				
West-Central																						
Fulbe (Cameroon)	17	6									53						ND	ND				41
Ewondo (Cameroon)	29		66							21							ND	ND			10	3
Fali (Cameroon)	39		26							33							ND	ND			18	23
Tali (Cameroon)	15		46						7	20	20						ND	ND				7
Mixed Adamawa (Cameroon)	18		17						11								ND	ND			12	60
Bakaka (Cameroon)	12		67				8			25							ND	ND				
Bamileke (Cameroon)	48		25				15			56							ND	ND			4	
Bamileke (Cameroon)	85		59		ND	ND	ND	ND		41								ND				
Bantu (Cameroon)	14		57		ND	ND	ND	ND		21								ND				22
Central-East																						
Bantu (Kenya)	29		21		ND	ND	ND	ND		31		17	14					ND			3	14
Wairak (Tanzanya)	43		21		ND	ND	ND	ND	2	19		2	37					ND				19
Hutu (Ruanda)	69		22	10	ND	ND	ND	ND		51		8	3					ND				6
Tutsi (Ruanda)	94		32		ND	ND	ND	ND		48		4	1					ND				15
Nilo-Saharian speakers																						
West-Central																						
Mixed-Nilo-Saharian (Cameroon)	9		11							22		11					ND	ND			22	34
Afro-Asiatic-Speakers																						
West-Central																						
Mixed-Chadic (Cameroon)	15		7							7			7				ND	ND			7	72
Daba (Cameroon)	18		28									22	6				ND	ND				44
Ouldeme (Cameroon)	21																ND	ND			5	95

[1]The African data were compiled from Cruciani et al. (2002) and Luis et al (2004). ND = not determined (marker was not investigated).

[2]Nomenclature according the The International Y–Chromosome Consortium revised by Jobling and Tyler-Smith (2003).

[3]This haplogroup showed frequencies ranging from 4% to 17% in populations from Kenya and Ethiopia (Cruciani et al., 2004).

27

TABLE 3. *Distributions (in %) of the B* and E* Y-chromosome haplogroups in two Brazilian and in 48 African populations*[1]

Population (country)	N	E3* (xE3a)	E3a* (xE3a7)	E3a7 (M191)	E1* (M33)	E2* (M75)	E3b* (xE3b1,xE3b2)	E3b1* (M78)	E3b2* (M81)	B* (M60)	Others
Porto Alegre (Brazil)	57		16	9			3	3	2	3	64
Rio de Janeiro (Brazil)	130		34	12	2	2	4		1	1	44

Niger-Congo speakers

<u>West</u>

Population (country)	N	E3* (xE3a)	E3a* (xE3a7)	E3a7 (M191)	E1* (M33)	E2* (M75)	E3b* (xE3b1,xE3b2)	E3b1* (M78)	E3b2* (M81)	B* (M60)	Others
Wolof (Gambia/Senegal)	34	3	68		12	3	6		6		2
Mandinka (Gambia/Senegal)	39		79		3			5	3	3	7
Ewe (Ghana)	30	3	73	23							1
Ga (Ghana)	29		62	34	3						1
Fante (Ghana)	32	3	44	41	3		3				6
Fon (Benin)	100		38	57		5					
Mossi (Burkina Faso)	49	2	68	22		4	2			2	
Rimaibe (Burkina Faso)	37	3	57	8	5	27					
Fulbe-I (Burkina Faso)	20		90		10						

<u>West-Central</u>

Population (country)	N	E3* (xE3a)	E3a* (xE3a7)	E3a7 (M191)	E1* (M33)	E2* (M75)	E3b* (xE3b1,xE3b2)	E3b1* (M78)	E3b2* (M81)	B* (M60)	Others
Mixed-Adamawa (Cameroon)	18		28							12	60
Fali (Cameroon)	39		26	33						18	23
Tali (Cameroon)	15		53	20	20						7
Fulbe-II (Cameroon)	17	6				53					41
Tupuri (Cameroon)	21									11	89
Ewondo (Cameroon)	29		66	21						10	3
Bakaka-I (Cameroon)	12		75	25							
Bakaka-II (Cameroon)	17		47	53							
Bamileke-I (Cameroon)	48		40	56						4	
Bamileke-II (Cameroon)	85		59	41							
Bantu (Cameroon)	14		57	21							22
Bassa (Cameroon)	11		55	36							9
Ngoumba (Cameroon)	31		39	32	6					23	
Nande (Democratic Republic of Congo)	18		33	37							30
Hema (Democratic Republic of Congo)	18		17	11		39	28				5
Cabinda (Democratic Republic of Congo)	74		46	32	ND	ND	ND	ND	ND	9	13

Cont.

28

TABLE 3 (Cont.)

Central-East

Bantu (Kenya)	29		21	31	17	14		3	14
Wairak (Tanzanya)	43		22	19	2	37			20
Hutu (Ruanda)	69		32	51	8	3			6
Tutsi (Ruanda)	94		32	48	4	1			15
Ganda (Uganda)	26		31	46	16				7

Southwest

Herero (Namibia)	24		38	33					29
Ambo (Namibia)	22	5	50	32	5	5			3

Southeast

Shona (Zimbabwe)	49		51	37	2			10	

South

Sotho-Tswana (South Africa)	28	4	36	21	4	7		18	10
Zulu (South Africa)	29	3	34	21	21			17	4
Xhosa (South Africa)	80	4	34	20	28	5		5	4

Nilo-Saharian speakers

West-Central

Mixed (Cameroon)	9		11	22	11			22	34

Central-East

Massai (Kenya)	26		12	4		35	15	8	26
Luo (Kenya)	9		22	44				22	12

Afro-Asiatic speakers

West-Central

Mixed-Chadic (Cameroon)	15		7	7				7	79
Podokwo (Cameroon)	19						5		95
Mandara (Cameroon)	28		11	4	7			4	74
Uldeme (Cameroon)	13							31	69
Ouldeme (Cameroon)	21							5	95
Daba (Cameroon)	18		28		22		6		44

Central-East

Amhara (Ethiopia)	18	6				11	33		50

TABLE 3 (Cont.)

Mixed Semitic (Ethiopia)	20	10		20	35	35
Oromo (Ethiopia)	9	11		11	22	56

[1]The African population data were compiled from Cruciani et al. (2002), Luis et al. (2004), Wood et al. (2005), and Beleza et al. (2005). ND = not determined (marker was not investigated).

[2]Nomenclature according to the International Y–Chromosome Consortium revised by Jobling and Tyler-Smith (2003).

30

TABLE 4. OriginS of Africans (in %) who arrived in Rio Grande do Sul and Rio de Janeiro at

the time of the slave trade considering genetic and historical sources

	West-Central and Southeast Africa[1]	West Africa[2]
Porto Alegre (POA)		
mtDNA[3]	82 ± 14	18 ± 14
Historical[4]	~80	~20
Rio de Janeiro		
mtDNA[3]	69 ± 13	31 ± 13
Historical[5]	~70	~30

[1]Major geographical regions characterized by the presence of people who speak languages identified with the Bantu branch, Niger-Congo subphylum. Two important previous Portuguese colonies were located in this region: Angola and Mozambique.

[2]Major geographical region characterized by the presence of people who speak languages identified with several non-Bantu linguistic groups of the Niger-Congo subphylum (except Hausa, see text).

[3]Some sub-clades with low frequencies in the derived populations (RJ and POA) were grouped in their respective haplogroups.

[4]According to estimates presented by Klein (2002).

[5]According to Maestri-Filho (1993).

31

Fig.1. Map of Africa showing the regions/countries considered in the text.

Fig. 2. Unrooted tree based on the mtDNA haplogroup distributions presented in Table 1.
A: Western non-Bantu cluster; B: West-Central/Southeastern Bantu cluster;
C: Cameroon populations.

Fig. 3. Unrooted tree based on Y-SNP haplogroup distributions presented in Table 3.
Since all "other" haplogroups in Rio de Janeiro and Porto Alegre had an European or
Amerindian origin, this category was excluded of the analyses for these two populations.
A: Afro-Asiatic speaker cluster; B: Niger-Congo speaker cluster.

Figure 1

Figure 2

34
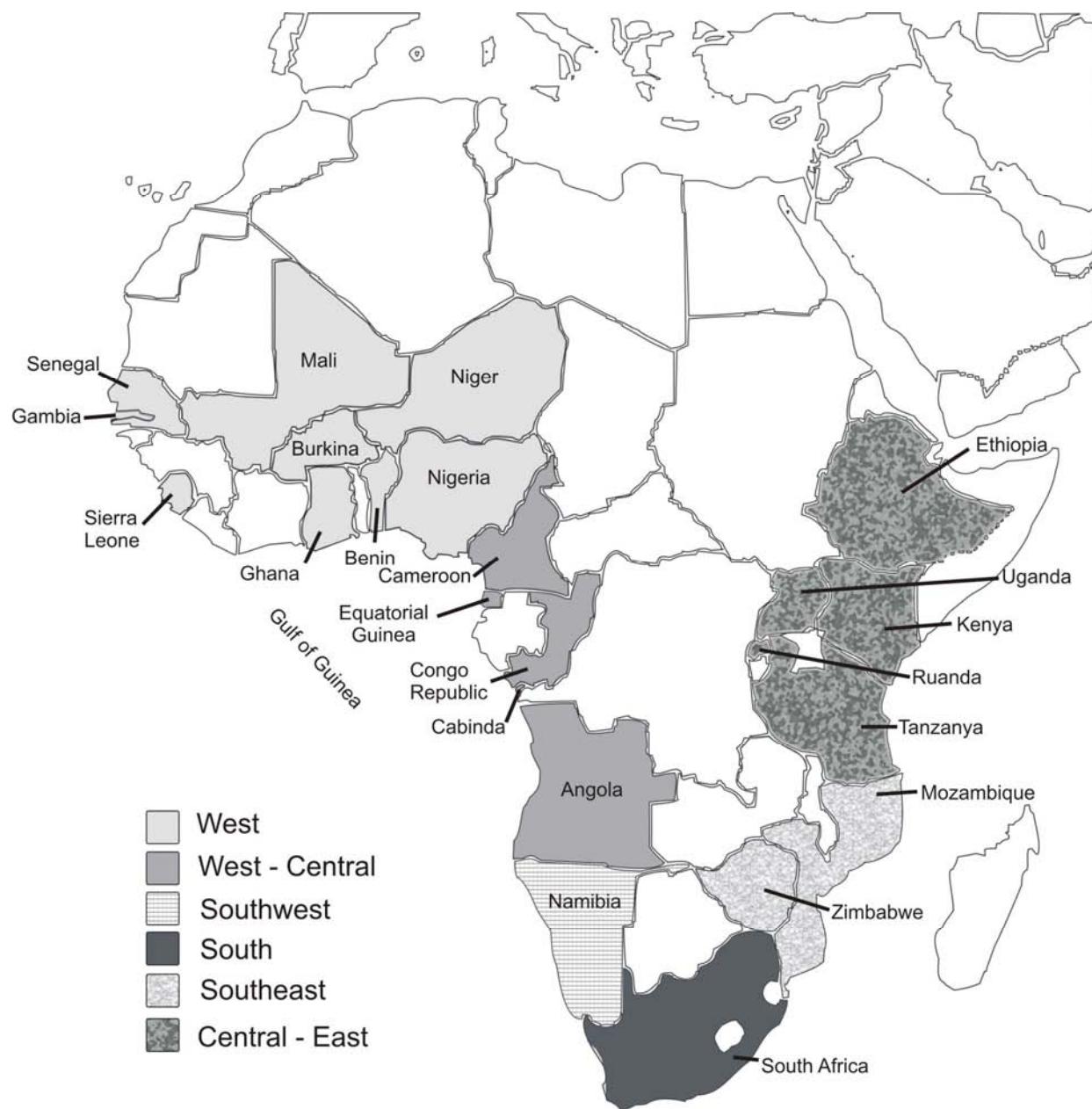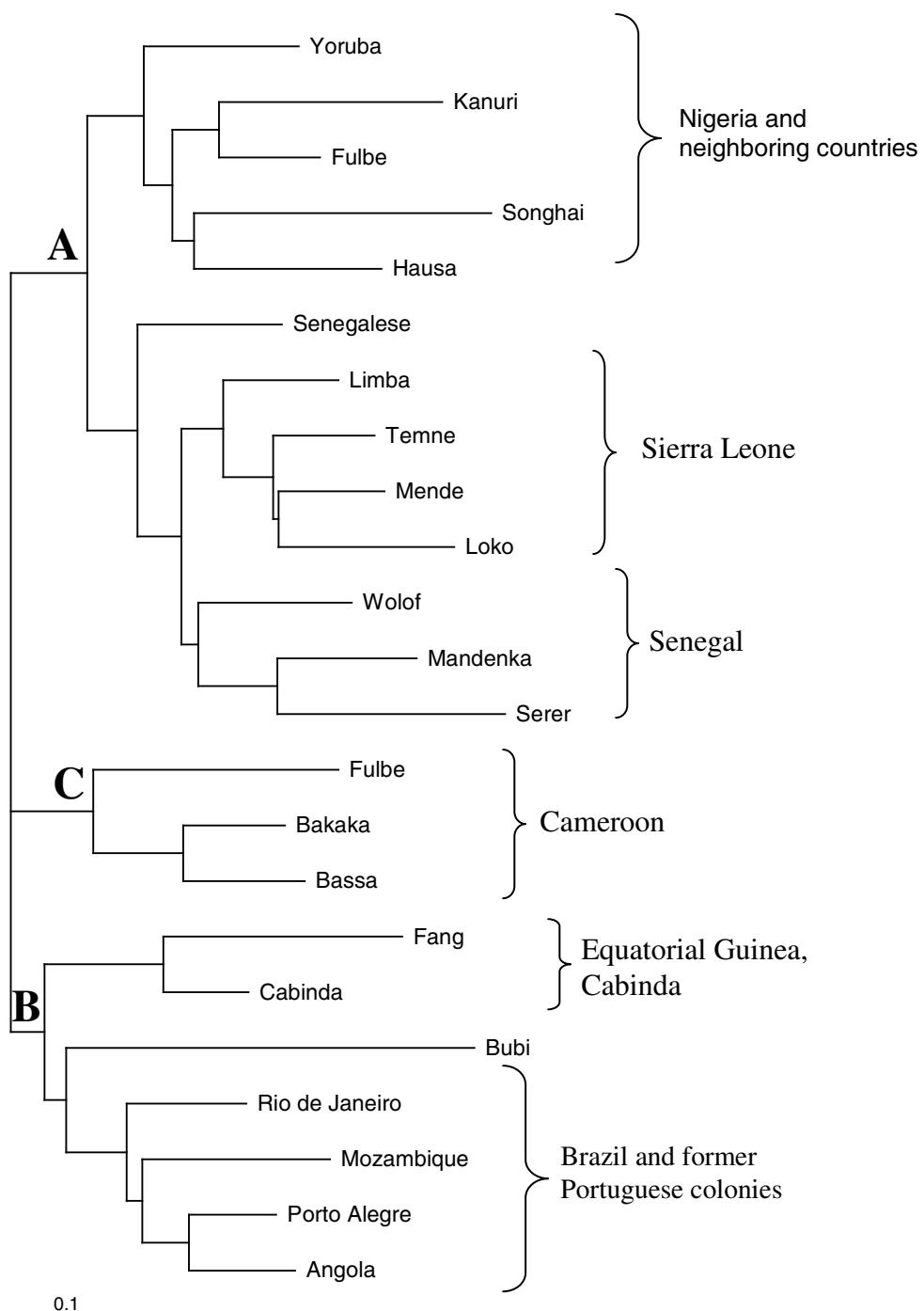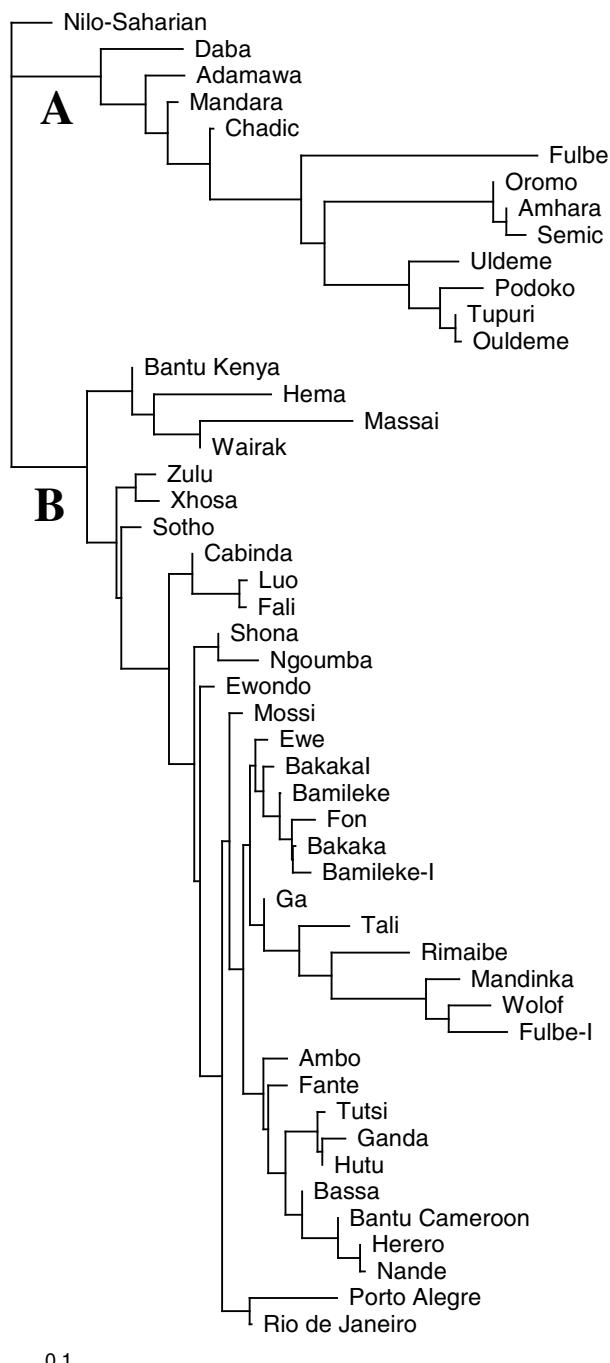
Figure 3

APPENDIX. List of the mtDNA haplogroups and haplotypes observed in Rio de Janeiro (RJ)

and/or Porto Alegre (POA)

| Haplotype	Haplogroup	HVS1 Mutations[1]	RJ	POA
1	L0a1	129 148 168 172 187 188G 189 223 230 278 311 320	1	1
2	L0a1	129 148 168 172 187 188G 189 223 230 278 293 311 320	3	3
3	L0a1	093 129 148 168 172 187 188G 189 223 230 278 293 311 320	2	
4	L0a2	148 172 187 188A 189 223 230 311 320	1	
5	L0a2	148 172 187 188G 189 223 230 311 320	3	
6	L0d1	129 187 189 223 230 239 243 294 311	1	
7	L1a1a	093 129 148 168 172 187 188A 189 223 230 278 293 311 320		1
8	L1b	111 126 187 189 223 264 270 278 293 311	1	
9	L1b	126 187 189 223 264 270 278 293 311	2	2
10	L1b	126 148 187 189 223 264 270 278 311	1	
11	L1b	126 187 189 223 264 270 278 311	4	2
12	L1c	129 187 189 223 278 294 311 360	1	
13	L1c	129 187 189 223 261 278 311 360	1	
14	L1c	129 187 189 223 274 278 287 294 311 320 360	1	
15	L1c	129 187 189 223 278 294 311 355 360 362	1	
16	L1c1	129 187 189 223 278 293 294 311 360	2	2
17	L1c1	129 187 189 223 274 278 293 294 311 360	1	
18	L1c1	093 129 187 189 223 263 278 293 294 311 360	1	1
19	L1c1	129 163 187 189 209 223 278 293 294 311 360		2
20	L1c1	093 129 187 189 223 278 293 294 311 360		1
21	L1c2	129 187 189 223 265C 278 286G 294 311 320 360	3	
22	L1c2	129 187 189 223 265C 278 286G 294 311 355 360	1	
23	L1c2	129 187 189 223 265C 278 286G 294 311 360	1	3
24	L1c2	129 145 187 189 223 234 265C 278 286G 294 311 360	1	
25	L1c2	129 187 189 223 265C 278 286A 294 311 320 360		1
26	L1c2	129 187 189 213 223 234 265C 278 286G 294 311 360		1
27	L1c3	129 189 215 223 278 294 311 360		3
28	L1c3	129 189 215 223 278 294 311 354 360		1
29	L2a α1	223 234 249 278 294	2	
30	L2a1 β1	223 278 294 309	2	2
31	L2a1 β1	193 213 223 239 278 294 309	1	
32	L2a1 β1	093 223 256 278 292 294 309	1	
33	L2a1 β1	223 256 278 294 309	2	1
34	L2a1 β1	093 223 256 278 294 309	4	2
35	L2a1 β1	223 278 291 294 309		1
36	L2a1 β1	129 223 278 294 309		2
37	L2a1 β1	092 223 278 294 309		2
38	L2a1 β2	189 193 223 245 278 294 309	1	
39	L2a1 β2	189 223 278 294 309		1
40	L2a1 β3	189 192 223 278 294 309		2
41	L2a1a	092 223 278 286 294 309	2	
42	L2a1a	223 278 286 294 309	3	4
43	L2b	114A 129 213 223 278 354	2	1
44	L2b	114A 129 213 223 274 278		3
45	L2b	114A 223 264 274 278		1
46	L2b	223 264 274 278		1
47	L2b1	114A 129 213 223 278 355 362	3	

APPENDIX (Cont.)

Haplotype	Haplogroup	HVS1 Mutations[1]	RJ	POA
48	L2b1	114A 129 213 223 278 311 362	1	
49	L2c	223 264 278	2	
50	L3	223	1	1
51	L3b	124 223 278 362	1	2
52	L3b	124 145 223 278 362	2	
53	L3b	223 278 294 362		2
54	L3b2	124 223 278 311 362		2
55	L3d	124 223 319	2	
56	L3d	124 223 278 290 292 312 362		1
57	L3d1	124 223	1	
58	L3d1	124 145 223 278 290 319 362		1
59	L3d1	124 223 278 290 319 362		1
60	L3d2	124 223 256	4	
61	L3e1	223 327	1	4
62	L3e1	176 223 327	1	1
63	L3e1a	185 223 327	1	1
64	L3e1a	185 223 311 327		2
65	L3e1a	185 209 223 327		2
66	L3e1b	223 325D 327	1	
67	L3e2	093 192 223 320		1
68	L3e2	192 223 320		1
69	L3e2b	172 189 223 320		2
70	L3e3	223 265T	3	2
71	L3e3	223 265T 355	1	
72	L3e3	223 265T 316	1	
73	L3e3	223 265T 288	1	
74	L3e3	189 223 265T		2
75	L3f	209 223 311	2	2
76	L3f	192 209 223 311		1
77	L3f1	129 209 223 292 295 311	1	2
78	L3f1	093 129 209 223 292 295 311	1	
79	L3f1	209 223 292 311		2
80	L3g	093 223 287 293T 301 311 355 362		2
81	L3g	093 223 293T 301 311 355 362		1
Total			84	85

[1]The nucleotide positions (less 16,000) considered for the analyses were from 16051 to 16384.
Sequences were aligned with the revised reference sequence (Andrews at al., 1999).