

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Instituto de Biociências
Programa de Pós-Graduação em Genética e Biologia Molecular

**ANÁLISE DA ESTRUTURA GENÉTICA DA POPULAÇÃO DO RIO GRANDE DO
SUL ATRAVÉS DE MICROSSATÉLITES AUTOSSÔMICOS E DE
CROMOSSOMOS SEXUAIS**

Fábio Pereira das Neves Leite

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do grau de Doutor em Ciências

Orientadora: Profa. Dra Mara Helena Hutz
Co-orientadora: Profa. Dra. Sídia Maria Callegari-Jacques

Porto Alegre, outubro de 2006.

AGRADECIMENTOS

À Profa. Dra. Mara Helena Hutz, pela orientação competente e eficiente, além dos ensinamentos valiosos para a minha formação.

À Profa. Dra. Sídia Maria Callegari-Jacques, pela co-orientação e inestimável auxílio nas análises estatísticas.

Ao Prof. Sidney Santos e à Elzemar (UFPA), pelo apoio nas análises de X-STRs.

Aos professores e funcionários do Programa de Pós-Graduação em Genética e Biologia Molecular.

À SENASP/MJ, principalmente ao Paulo Fagundes, pelo apoio ao desenvolvimento da Genética Forense no Brasil.

À Viviane e ao pessoal do Laboratório de Perícias/IGP, pelo apoio e auxílio nas atividades administrativas; e aos peritos do Laboratório de Genética Forense, pela presteza.

Aos meus pais pelo incentivo à minha formação.

À minha esposa Carla e meu filho Eduardo, por terem paciência e compreensão nestes três anos e três meses de duração deste trabalho.

As pesquisas realizadas neste trabalho foram financiadas pelo CNPQ, IGP e SENASP/MJ.

SUMÁRIO

Lista de Abreviaturas, símbolos e unidades.....	4
Resumo.....	5
Abstract.....	7
1. Introdução.....	9
1.1 .Uso de marcadores moleculares em estudos de genética de populações.....	10
1.1.1. Microssatélites.....	11
1.1.1.1. Taxas e mecanismos de mutação em microssatélites.....	13
1.1.1.2. Mecanismos de mutação e modelos.....	15
1.1.1.3. Utilização de microssatélites em estudos genéticos.....	19
1.2 .Cromossomos Sexuais.....	21
1.2.1. O Cromossomo Y.....	22
1.2.2 O Cromossomo X.....	28
1.3 . Aspectos Históricos da população do Rio Grande do Sul.....	31
1.3.1. As populações indígenas.....	33
1.3.2. Estudos de miscigenação no Rio Grande do Sul.....	35
1.4 . Justificativa e Objetivos.....	37
2. LD patterns and genetic structure of Amerindian and Brazilian populations revealed by long-range Xq markers.....	38
3. The genetic structure of southern Brazil revealed by autosomal markers.....	65
4. Y-STR analysis in Brazilian and South Amerindian populations.....	89
5. Discussão.....	115
Referências Bibliográficas.....	122
Anexos.....	138
Anexo 1 - Número de coincidências dos haplótipos de Caingang e Guaranis a partir da busca no banco de dados de haplótipos de cromossomo Y, disponível em http://www.yhrd.org	139

LISTA DE ABREVIATURAS, SÍMBOLOS E UNIDADES

AM: “Amerindian”
AMOVA: “Analysis of Molecular Variance”
AS: “Asian”
CE: “Center-East”
CW: “Center-West”
EU: “European”
IAM: “Infinite Allele Mutation Model”
IBGE: Instituto Brasileiro de Geografia e Estatística
Kb: “Kilobases”
LD: “Linkage Disequilibrium”
MALDI-TOF: “Matrix-Assisted Laser Desorption/Ionization-Time of Flight”
MDS: “Multidimensional Scaling”
Mb: “Megabases”
MET: “Metropolitan”
MSY: “Male Specific Region”
N: “North”
NE: “Northeast”
NW: “Northwest”
PAR1/PAR2: “Pseudoautosomic Region”
PCR: “Polimerase Chain Reaction”
RFLP: Restriction Fragment Length Polymorphisms”
RJ: Rio de Janeiro
RS: Rio Grande do Sul
S: “South”
SE: “Southeast”
SMM: “Stepwise Mutation Model”
SNPs: “Single Nucleotide polymorphisms”
STRs: “Short Tandem Repeats”
SW: “Southwest”
WRS: “Western Rio Grande do Sul”
Xp: Braço curto do Cromossomo X
Xq: Braço longo do Cromossomo X

RESUMO

Os microssatélites ou *Short Tandem Repeats* (STRs) são marcadores moleculares que caracterizam-se por altos níveis de heterozigotidade, dispersão uniforme, alto nível de reprodutividade, codominância e possibilidade de genotipagem por métodos rápidos e simples. Devido a estas características, estes marcadores têm sido amplamente utilizados em genética de populações, estudos evolutivos e análises forenses.

Na presente tese, foram utilizados STRs autossômicos e de cromossomos sexuais (X e Y) para a avaliação da estrutura genética da população do Rio Grande do Sul (RS). Grupos ameríndios brasileiros foram também investigados, como uma das populações parentais importantes para a formação da atual população do RS. Para tanto, foram avaliados os sistemas D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, vWA, CSF1PO, TPOX e TH01 (autossômicos); DXS9895, DXS7132, DXS6800, DXS9898, DXS6789, DXS7133, DXS7130, HPRTB, GATA31E08, DXS7423 e DXS10011 (do cromossomo X) e DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, e DYS385 (do cromossomo Y). As amostras utilizadas foram amplificadas através da técnica de reação em cadeia da polimerase, a partir de *primers* com fluoróforos incorporados. Os produtos de amplificação foram posteriormente determinados por intermédio de separação em eletroforese capilar.

Os resultados obtidos a partir de STRs autossômicos demonstraram a ausência de subestruturação quando a população do Rio Grande do Sul foi comparada com populações

de outras regiões do país. As análises de mistura racial exibiram um padrão caracterizado pela predominância da contribuição européia, seguida por ameríndia e africana. As estimativas de miscigenação por regiões do Estado revelaram que a porção oeste do Rio Grande do Sul exibiu os maiores níveis de contribuição ameríndia, se comparados com o restante do Estado. As análises de relacionamento genético, obtidas a partir de dados produzidos na presente tese e também de outras populações, demonstraram que a população do Rio Grande do Sul apresenta maior proximidade genética com a Argentina do que com alguns Estados brasileiros. Em relação às análises realizadas a partir de microssatélites do cromossomo Y, nossos resultados, corroborando estudos prévios, demonstraram uma forte predominância da contribuição européia para a população do Estado, seguida por níveis insignificantes de contribuição africana e ausência de contribuição ameríndia. Os resultados obtidos por intermédio dos marcadores de cromossomo X, para amostras do Rio Grande do Sul e de um *pool* de amostras de ameríndios, revelaram maiores níveis de desequilíbrio de ligação para o grupo ameríndio, sendo provavelmente reflexos da estruturação populacional relacionada à ação de gargalo de garrafa e à deriva genética. Os níveis de desequilíbrio de ligação encontrados para a amostra sulina foram de menor intensidade, e provavelmente correlacionados à miscigenação. Outrossim, as análises realizadas a partir destes marcadores demonstraram ainda uma maior proximidade das distribuições alélicas da amostra do Rio Grande do Sul com populações européias; enquanto que, para o grupo ameríndio, os resultados revelaram uma maior similaridade com as populações asiáticas.

Os dados obtidos na presente Tese confirmam alguns resultados encontrados a partir de análises com outras populações brasileiras, mas também trazem novas informações, que devem ser investigadas em outras amostras populacionais, a fim de que se estabeleça o seu real significado.

ABSTRACT

Microsatellites or Short Tandem Repeats (STR) are molecular markers characterized by high levels of heterozygosity, uniform dispersal among the human genome, codominance and possibility of genotyping by fast and simple laboratory methods. Due to these characteristics, these markers have been widely used in population genetics, evolution studies and forensic genetics.

In the present work, we used autosomes, as well as X-STRs and Y-STRs, to evaluate the genetic structure of the Rio Grande do Sul (RS) population. Amerindian groups were also investigated, due to their importance, as a parental group, for the formation of the present RS population. For this purpose, markers D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, vWA, CSF1PO, TPOX and TH01 (autosomic); DXS9895, DXS7132, DXS6800, DXS9898, DXS6789, DXS7133, DXS7130, HPRTB, GATA31E08, DXS7423 and DXS10011 (from the X chromosome) and DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, and DYS385 (from the Y chromosome) were used. The samples were amplified by PCR with fluorescent primers, and the PCR products were then analyzed by capillary electrophoresis.

Regarding the autosomic STR results, no substructuring was observed between Rio Grande do Sul and other Brazilian regions. The admixture analysis revealed a pattern characterized by the predominance of European, followed by Amerindian and African contributions. The admixture estimates from the western region of Rio Grande do Sul

showed a higher Amerindian contribution, if compared with the rest of the state. Genetic relationship analyses, undertaken from the comparisons of the data obtained herein and from other populations, demonstrated that Rio Grande do Sul inhabitants are genetically closer to Argentineans than to other Brazilians. The Y-STR analysis, as in previous studies, showed a higher predominance of European contribution for the patrilineages from Rio Grande do Sul, followed by negligible levels of African contribution and absence of Amerindian contribution. The X-STR results, obtained for a sample from Rio Grande do Sul and another composed by a pool of Amerindian tribes, revealed higher levels of linkage disequilibrium for the latter, probably determined by population structure, derived from bottleneck events and random genetic drift. The levels of linkage disequilibrium found for the RS sample are probably determined by admixture events. Moreover, the X-STR analysis demonstrated a higher proximity from the allelic distributions of the RS population with European populations; whereas, for the Amerindians, our results revealed a closer similarity with the allelic distributions of Asian populations, as expected.

The data obtained in the present work corroborate previous results found in other Brazilian populations. However, they also bring new information, which should be investigated in other samples, to verify its real significance.

Capítulo 1

INTRODUÇÃO

1.1. USO DE MARCADORES MOLECULARES EM ESTUDOS DE GENÉTICA DE POPULAÇÕES

Os estudos genéticos de populações humanas são utilizados visando a compreensão da história, demografia e evolução da espécie humana, além de estudos de ancestralidade e de mapeamento de genes de relevância clínica. Devido ao fato que a variação genética responsável por uma doença constitui-se em uma porção da variação genética total (Jorde *et al.* 2001), é essencial o entendimento dos processos evolucionários responsáveis por esta variação para a identificação de genes que causam doenças complexas. Outra aplicação prática dá-se na investigação forense, tanto em testes de paternidade quanto em exames de identificação humana em casos criminais.

Os marcadores inicialmente utilizados para estes estudos, como os marcadores protéicos, são pouco elucidativos para aplicação em estudos populacionais, devido ao baixo nível de variação genética e ao fato de serem potencialmente afetados pela seleção natural. Estes marcadores foram substituídos por marcadores mais poderosos, como marcadores de DNA, que apresentam uma maior variação genética, sendo alguns destes seletivamente neutros e facilmente genotipados por técnicas rápidas e simples, como a PCR (*Polimerase Chain Reaction*).

Dentre os vários marcadores moleculares, os RFLP (*Restriction Fragment Length Polymorphisms*) foram os primeiros a serem utilizados em estudos de identificação humana e populacionais, a partir dos trabalhos de Jeffreys *et al.* (1985). Posteriormente, com o advento da PCR e de técnicas de seqüenciamento direto, novos marcadores moleculares passaram a ser utilizados. Dentre estes, podemos citar os microssatélites ou STRs (*Short tandem repeats*) e os

polimorfismos de nucleotídeo único ou SNPs (*Single nucleotide polymorphisms*) autossômicos, além de marcadores uniparentais, como os do cromossomo Y e polimorfismos de DNA mitocondrial.

1.1.1. Microsatélites

Os microsatélites consistem em pequenas unidades (motivos) de repetição de DNA geralmente de 2 a 4 nucleotídeos (Figura 1), estando dispersos por todo o genoma humano. Estima-se que estes marcadores ocorram no genoma humano a cada 10.000 nucleotídeos (Budowle e Brown, 2001), sendo considerados marcadores de evolução rápida, com taxa de mutação média de 6×10^{-4} (Goldstein et al. 1995a). Quanto às suas características, estes marcadores combinam várias propriedades desejáveis, como altos níveis de polimorfismo, dispersão uniforme, alto nível de reprodutividade, codominância e possibilidade de genotipagem por métodos rápidos e simples (Budowle e Brown, 2001). STRs são classificados em categorias diferentes baseados no tipo de motivo de repetição: Repetições simples contêm uma repetição de seqüência e comprimentos idênticos (Ex: $CA_{(10)}$). Repetições compostas consistem de duas ou mais repetições simples adjacentes com diferentes motivos de repetições (Ex: $(TA)_7(TG)_{18}$) (Butler, 2001). As características dos motivos de repetição (tipo, número e continuidade) parecem afetar a taxa de mutação e os níveis de variação alélica. Desta forma, locos com motivos de repetição maiores são geralmente mais polimórficos do que aqueles compostos por motivos menores (Weber e Wong, 1993).

Os microsatélites compreendem aproximadamente 3% do genoma humano sendo os dinucleotídeos a classe mais comum em humanos. Dentre os dinucleotídeos, as repetições contendo CA são as mais representadas (50 %) enquanto as repetições CG são menos

representadas (0,1 %), podendo esta distribuição não aleatória das frequências dos diversos tipos de STRs estar associada a diferenças nas taxas de mutação dos diferentes motivos de repetição (IHGSC, 2001). Por outro lado, um estudo de Harr *et al.* (2002) mostrou um viés na eficiência do sistema de reparo por mau pareamento, o que faria com que certas seqüências fossem super representadas.

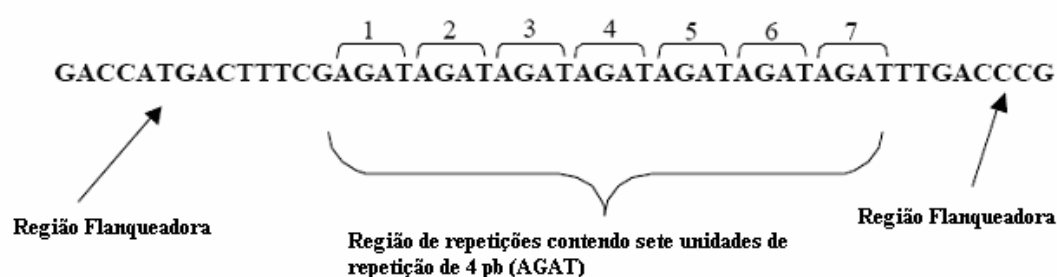


Figura 1 – Ilustração mostrando a estrutura de um *Short Tandem Repeat* (STR). O número de unidades de repetições AGAT varia entre indivíduos (modificado a partir de Schoscke, 2003).

Contudo, apesar de evidências indicarem que a grande maioria dos locos de microssatélites estarem presentes em regiões não-codificantes, sendo seletivamente neutros (Wren *et al.* 2000), uma parte significativa destas seqüências, quando presentes em regiões codificantes ou regulatórias, são funcionalmente importantes para alguns processos biológicos. Segundo Li *et al.* (2002a) os STRs podem estar associados às seguintes funções biológicas: a) organização da cromatina, b) estrutura do DNA (formação de estruturas secundárias), c) formação do centrômero, d) *hotspot* para recombinação, e) controle da replicação e ciclo celular, f) modulação da taxa de mutação, g) controle da transcrição e expressão gênica, h) ligação de proteínas regulatórias, i) inibição da tradução. Além disto,

muitos STRs estão associados a vários tipos de câncer, através de um processo de instabilidade mutacional ocasionado pela deficiência no sistema de reparo por pareamento incorreto (Jacob e Praz, 2002). Por outro lado, mesmo quando estão localizados em regiões não codificantes, os STRs podem estar ligados a locos codificantes, sofrendo então os efeitos da seleção natural.

1.1.1.1. Taxas e mecanismos de mutação em microsatélites

A fim de entendermos os mecanismos responsáveis pela mutação em microsatélites, é necessário termos uma idéia do padrão de mutação destes marcadores em humanos. Ellegren, (2000) observou em 102 eventos mutacionais que 89 % consistiam de mudanças em um motivo de repetição, enquanto que os 11 % restantes eram relacionados a mudanças em mais de um motivo de repetição. Quando consideradas as mutações com origem conhecida, 69 destas eram de origem paterna, enquanto 14 tinham fonte materna, resultando em uma razão de taxa de mutação paterna / materna (α) de 4,9. Xu *et al.* (2000) estimaram um α de 3,1 a partir de 236 mutações. Um excesso de mutações paternas foi também reportado por outros autores, como Weber e Wong, (1993), Brinkman *et al.* (1998) e Xu *et al.* (2000), podendo este fator estar relacionado, segundo Buckleton *et al.* (2005), a observações de que espermatozoides sofrem mais divisões celulares do que óvulos durante a meiose. Além dos estudos *in vivo*, muitos trabalhos têm sido realizados a partir de estudos *in vitro* com diversos organismos, como bactérias (Levinson e Gutman, 1987), leveduras (Wierdl *et al.* 1997) e linhagens celulares de mamíferos (Talbot *et al.* 1995). Apesar de algumas limitações que estes

estudos apresentam, como o fato que mutações somáticas ocorrem com maior frequência em culturas *in vitro* (Talbot *et al.* 1995), estes estudos foram fundamentais para a definição dos mecanismos moleculares que envolvem as mutações em STRs. Assim, estudos *in vitro* demonstraram os efeitos do sistema de reparo de pareamento incorreto na taxa de mutação de STRs (Farber *et al.* 1994) e também a influência de mutações pontuais no padrão de repetições destes marcadores (Petes *et al.* 1997).

A partir de observações de que STRs com mais de 30 motivos de repetição são raros (Bell e Jurka, 1997), Kruglyak *et al.* (1998) e Calabrese *et al.* (2001) sugeriram que quando as estruturas de repetição se tornam muito longas, há um incremento na probabilidade destas serem interrompidas por mutações pontuais e se dividirem em duas estruturas de repetição menores e menos mutáveis. Estudos de Ellegren, (2000), Harr e Schlötterer, (2000) e Xu *et al.* (2000), observaram uma tendência na razão de expansão/contração do motivo de repetição dependente do comprimento da estrutura. Desta forma, STRs com estruturas curtas são mais susceptíveis à mutação por expansão do que estruturas longas. À medida que a estrutura se expande, há uma tendência contrária, no sentido desta ser mais susceptível à interrupção de sua expansão (Lai e Sun, 2003). Apesar da base molecular não estar ainda esclarecida, evidências baseadas em STRs de tetranucleotídeos em humanos e dinucleotídeos em *Drosophila melanogaster* apontam para a ocorrência deste fenômeno.

1.1.1.2. Mecanismos de mutação e modelos

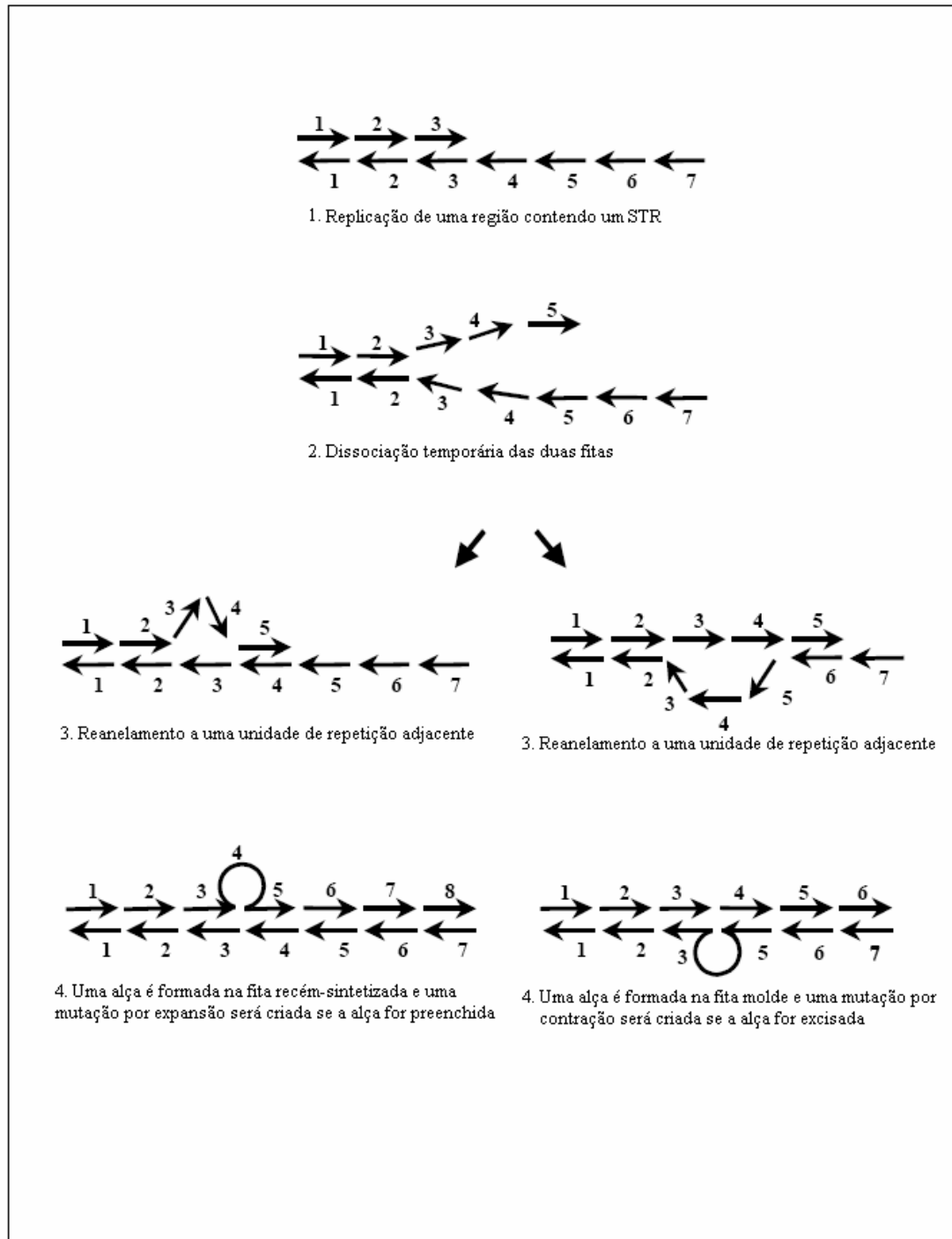
O mecanismo principal envolvido nas altas taxas de mutação dos STRs parece estar relacionado ao escorregamento da DNA polimerase durante a replicação (Levinson e Gutman, 1987, Schlotterer e Tautz, 1992). Desta forma, a fita molde se dissocia temporariamente da fita recém-sintetizada, para se re-associar após uma fração de segundos. Se este processo ocorre durante a replicação de uma região repetitiva, uma unidade de repetição na fita recém-sintetizada pode algumas vezes se re-associar a uma unidade de repetição incorreta na fita molde (escorregamento), criando uma alça em uma das fitas (figura 2). Se este pareamento incorreto não sofrer ação de reparo, a alça será excisada ou preenchida. Em qualquer uma das situações, uma nova mutação estará estabelecida se a excisão ou o preenchimento forem feitos na fita com repetição incorreta. Assim, a formação de uma alça na fita recém-sintetizada resultará em uma mutação por inserção (expansão) enquanto que a excisão da alça na fita molde resultará em uma mutação por deleção (contração). Uma estrutura de repetição mais longa forneceria mais substrato para o escorregamento ocorrer, o que explicaria a dependência do comprimento da estrutura neste mecanismo de mutação. Além disto, existe maior probabilidade que este mecanismo ocorra de forma mais freqüente em uma unidade de repetição imediatamente adjacente, o que explicaria o fato que a grande maioria das mutações ocorrem por ganho ou perda de um motivo de repetição.

Outro mecanismo inicialmente relacionado ao surgimento dos STRs estaria relacionado ao *crossing-over* desigual durante a recombinação. Contudo, evidências baseadas no fato de que cepas de leveduras deficientes em genes de recombinação não apresentaram taxas de mutação alteradas (Henderson e Petes, 1992), de que as taxas de mutação no cromossomo Y (não-recombinantes) são similares às seqüências recombinantes (Kayser *et al.*

2000) e no fato de que o padrão de mutação em STRs parece ser dependente da direção da replicação (Tran *et al.* 1995) sugerem que a replicação, e não a recombinação, é o principal mecanismo responsável pela evolução dos STRs. Contudo, como a recombinação parece ser importante para a evolução de minissatélites (Jeffreys *et al.* 1998) não podemos descartar a possibilidade de que este mecanismo esteja envolvido na geração de variabilidade em STRs.

Figura 2 – Mecanismo de mutação de STR através do escorregamento da DNA polimerase

(Modificado a partir de Goldsteind e Schlotterer, 1999).



A partir do conhecimento do modo de mutação dos STRs, torna-se possível a formulação de um modelo de mutação para estes marcadores, a fim de que possamos aplicá-lo em estudos evolutivos e de genética de populações. O primeiro modelo, proposto por Kimura e Crow, (1964) denomina-se “infinite allele mutation model” (IAM) e considera cada nova mutação como originando de forma aleatória um novo alelo. De acordo com este modelo, a maior proximidade em relação ao número de repetições não significa uma relação filogenética mais próxima. Ohta e Kimura, (1973) desenvolveram um modelo denominado “stepwise mutation model” (SMM) que tem sido aplicado em estudos envolvendo STRs. De acordo com este modelo, um alelo somente pode mutar a partir do ganho/perda de um alelo. A adoção do modelo SMM implica que diferenças em tamanho entre dois alelos são informativas: quanto maior a diferença, maior o número de eventos mutacionais que ocorreram desde a ancestralidade comum destes dois alelos e menor a relação biológica entre eles. Devido ao fato que ambos os modelos envolvem a perda/ganho de motivos de repetição, homoplasia (situação na qual dois alelos são idênticos embora tenham sido originados por eventos mutacionais independentes) pode ocorrer nos modelos acima mencionados. Um terceiro modelo foi proposto por Di Rienzo *et al.* (1994), denominado “two phase model”, na qual a maioria das mutações envolvendo STRs relacionam-se à perda/ganho de uma repetição; contudo, mutações mais amplas são permitidas, mas em uma taxa menor.

Baseados nos modelos acima expostos, várias distâncias genéticas foram criadas para utilização em estudos de genética de populações. Desta forma, o parâmetro F_{st} (Wright, 1951) assume o modelo IAM, enquanto as distâncias genéticas R_{st} (Slatkin, 1995), D_{sw} (Shriver *et al.* 1995) e $(\delta\mu)^2$ (Goldstein *et al.* 1995b) assumem o modelo SMM. Outrossim, outras distâncias genéticas, como D_c (Cavalli-Sforza e Edwards, 1967), D_a (Nei *et al.* 1983) e D_{as}

(Stephens *et al.* 1992), não são baseadas nos modelos evolucionários acima expostos, mas podem ser utilizadas de forma satisfatória em análises filogenéticas envolvendo populações com relações biológicas próximas (Goldstein *et al.* 1999).

1.1.1.3. Utilização de microssatélites em estudos genéticos

Devido às características de altos níveis de heterozigidade, abundância de locos no genoma humano, e neutralidade seletiva, os microssatélites têm sido amplamente utilizados para mapeamento de genes de interesse médico e também em estudos evolutivos e de genética de populações. Quanto a aplicações dos STRs autossômicos na história evolutiva das populações humanas, os trabalhos de Goldstein *et al.* (1995a), na qual estimaram que a emigração dos humanos do continente africano ocorreu há aproximadamente 150.000 anos atrás; de Jorde *et al.* (1997) que encontraram um significativo aumento na diversidade genética em populações africanas em relação às asiáticas e européias e o de Reich e Goldstein, (1998), que detectaram sinais de expansão humana em populações africanas no período Paleolítico, serviram para dar suporte à teoria de origem do homem moderno na África. Outrossim, a estrutura genética de populações humanas em nível global foi estudada por Rosenberg *et al.* (2002). Estes autores observaram que diferenças dentro de populações contribuíram para 93-95 % da variação genética, enquanto diferenças entre grupos constituíram de 3-5 % da variação genética total. Análises filogenéticas em várias populações humanas foram realizadas por intermédio de STRs (Bowcock *et al.* 1994, Jorde *et al.* 1995, Deka *et al.* 1995, Zhivotovsky *et al.* 2003). Contudo, existem limitações desses marcadores que devem ser consideradas quando da sua utilização em estudos de estrutura de populações e análises

filogenéticas: a) devido às altas taxas de mutação, os STRs refletem divergências relativamente recentes; assim, inferências evolucionárias mais antigas devem ser obtidas a partir de marcadores bialélicos; b) como consequência do mecanismo evolucionário e de suas altas taxas de mutação, homoplasia pode ocorrer, o que pode interferir em análises genéticas. Contudo, apesar destas limitações, a convergência entre os estudos utilizando STRs e os aspectos históricos e geográficos das populações estudadas (Wilson e Balding, 1998) e também o surgimento de um número significativo de distâncias genéticas especificamente desenvolvidas para uso nestes marcadores (Bowcock *et al.* 1994; Goldstein *et al.* 1995; Shriver *et al.* 1995; Slatkin, 1995; Zhivotovsky, 1999), indicam que os STRs são uma ferramenta muito útil para a inferência de relações evolutivas e da história demográfica de populações humanas (Deka *et al.* 1995; Jorde *et al.* 1995, 1997, 2000; Rosenberg *et al.* 2002).

Microssatélites são também úteis na identificação individual e estabelecimento de bancos de dados populacionais, para aplicação na área forense. Alguns autores, como Budowle *et al.* (1999), Jin *et al.* (2000) Geada *et al.* (2000), Biondo *et al.* (2001), Ayres *et al.* (2002) e Reichenpfader *et al.* (2003) publicaram estudos visando o estabelecimento de bancos de dados regionais em diversos países. Além disto, trabalhos como os de Primorac *et al.* (1995), Geada *et al.* (2000) e Meyer *et al.* (2000) enfatizaram a aplicação destes marcadores em estudos de identificação humana.

Em relação às populações brasileiras nativas, vários estudos, como os de Pérez- Lezaun *et al.* (1997), Calafell *et al.* (1998), Harpending e Rogers, (2000), Jin *et al.* (2000), Hutz *et al.* (2002) e Kohlrausch *et al.* (2005), descreveram a variabilidade de locos de microssatélites. Quanto a populações não indígenas, Grattapaglia *et al.* (2001), Silva *et al.* (2002) e Leite *et al.* (2003) publicaram dados relativos às populações dos Estados de Alagoas e Rio Grande do Sul,

respectivamente, para fins de estabelecimento de bancos de dados populacionais e outros estudos genéticos.

Embora os marcadores de DNA sejam muito poderosos para a distinção genética ao nível individual, a maioria oferece pouco poder para permitir a distinção de etnia ou para sustentar qualquer afirmação sobre as características físicas de um indivíduo (Shriver *et al.* 1997). Isto acontece principalmente pelo fato que a maior parte (80-90 %) da variação genética entre humanos é intra-populacional, sendo que somente 10-20 % da variação total é devido à variação entre populações (Rosenberg *et al.* 2002).

Apesar destas limitações, diversos autores, como Shriver *et al.* (1997) e Bamshad *et al.* (2003) utilizaram STRs, tanto para estimar o grau de mistura ao nível populacional quanto para avaliar a ancestralidade genômica ao nível individual. Contudo, para atribuição de indivíduos em determinados grupos étnicos ou geográficos, os autores concluíram que os melhores marcadores destinados para estes fins seriam aqueles que apresentassem os maiores diferenciais de frequências alélicas entre os grupos definidos. Marcadores com esta característica são úteis em estudos envolvendo genética humana, mapeamento genético e genética forense. Contudo, para a sua correta utilização, são necessários estudos populacionais prévios, a fim de se avaliar os diferenciais de frequências dos mesmos, e conseqüentemente, o seu efetivo potencial para aplicação nestas áreas.

1.2. CROMOSSOMOS SEXUAIS

Os cromossomos sexuais (X e Y) evoluíram a partir de um par de cromossomos autossômicos, há aproximadamente 300 milhões de anos atrás (Ohno, 1967; Hughes *et al.*

2005). Durante a evolução, a ocorrência de ao menos quatro eventos de inversões no cromossomo Y levou à supressão de recombinação em certas regiões (estratos) cromossômicas, ocasionando a diferenciação dos cromossomos sexuais (Lahn e Page, 1999). Contudo, apesar destes processos levarem à divergência dos cromossomos X e Y, estes cromossomos ainda apresentam duas regiões de homologia, denominadas Regiões Pseudoautossômicas 1 e 2. Nestas regiões, ocorre o pareamento e recombinação dos cromossomos sexuais durante a meiose de indivíduos do sexo masculino.

1.2.1. O Cromossomo Y

O cromossomo Y representa aproximadamente 2% do genoma humano e tem aproximadamente 24 Mb de comprimento, (Jobling e Tyler-Smith, 2003). Embora estudos das décadas de 1950-1960 caracterizassem este cromossomo como contendo poucos genes, publicações atuais demonstraram a existência de 76 genes que codificam ao menos 27 proteínas, sendo muitos destes genes funcionalmente especializados na espermatogênese (Lahn e Page, 1997; Skaletsky *et al.* 2003). A maior parte (~95%) deste cromossomo compreende uma região denominada MSY (“Male Specific Y Region”). Skaletsky *et al.* (2003) classificaram esta região em 3 grandes classes de sequências, denominadas X degeneradas, X transpostas e sequências amplicônicas. As sequências X degeneradas são representadas por genes que sofreram a ação de eventos mutacionais, sendo muitos deles não funcionais. As sequências transpostas foram originadas a partir de eventos de transposição de genes do cromossomo X há aproximadamente 3 a 4 milhões de anos, apresentando poucos genes funcionais. As sequências amplicônicas exibem a maior densidade de genes dentro o MSY. Esta região apresenta oito extensas palíndromes, cujo comprimento varia entre 9 kb até

1,45 megabases. A origem destas seqüências deve-se a processos de conversão gênica intracromossomal durante a meiose, e tem a função de substituir genes que sofreram mutações deletérias (Rozen *et al.* 2003 e Skaletsky *et al.* 2003).

De acordo com Skaletsky *et al.* (2003), a existência das seqüências degeneradas corroboram o modelo originalmente proposto, segundo o qual, a partir da evolução dos cromossomos sexuais, o cromossomo X permaneceu com seus genes ancestrais intactos, devido à ocorrência de recombinação em mulheres; enquanto que, para o cromossomo Y, houve o acúmulo de mutações que não puderam ser removidas pela recombinação. Contudo, segundo estes mesmos autores, a ocorrência de conversão gênica nas seqüências amplicônicas do MSY trazem à tona a existência de um novo modelo evolutivo para o cromossomo Y.

O restante do cromossomo compreende as Regiões Pseudo-autossômicas 1 e 2 (PAR1 e PAR2), abrangendo 2,7 Mb e 330Kb, respectivamente. A figura 3 mostra a estrutura do cromossomo Y, evidenciando as diferentes regiões e também os STRs mais utilizados para análise forense e em genética de populações.

Tendo em vista as suas características peculiares, como herança patrilinea e ausência de recombinação, os microsátélites de cromossomo Y (Y-STRs) são analisados a partir de haplótipos, ou conjuntos de marcadores ligados. Como apresentam taxas de mutação similares às dos STRs autossômicos (Kayser *et al.* 2000), quando combinados com marcadores bialélicos (de evolução lenta) permitem obter boas discriminações em nível individual e populacional (Jobling e Tyler-Smith, 1995). Além disto, inferências histórico-geográficas acerca de eventos relacionados à migração de populações e colonização de áreas geográficas têm sido obtidas a partir destes marcadores. Assim, Hammer *et al.* (1998) analisaram a distribuição global das linhagens de cromossomo Y em diversas populações, encontrando a

linhagem ancestral como tendo origem africana. Este estudo, juntamente com o de Thomson *et al.* (2000) confirmaram o modelo *Out of Africa* para a origem humana. Outrossim, Luis *et al.* (2004) obtiveram evidências que reforçaram a utilização de dois corredores de imigração para movimentos migratórios entre a África e Eurásia. Estudos de Santos *et al.* (1999), Karafet *et al.* (1999), Lell *et al.* (2002) e Tarazona-Santos *et al.* (2002), entre outros, analisaram a origem siberiana dos povos nativos americanos e também a dinâmica de eventos migratórios que ocorreram entre a Sibéria e as Américas, enquanto que Bortolini *et al.* (2003) estimaram a colonização do continente americano há aproximadamente 14.000 anos. Roewer *et al.* (2005) em um estudo com 12.700 amostras provenientes de 19 localidades, detectaram a ocorrência de duas histórias demográficas diferentes entre as regiões oeste e leste da Europa, concluindo que Y-STRs são os marcadores moleculares mais apropriados para estudos envolvendo história demográfica recente e estrutura populacional regional.

Y-STRs são também úteis em exames forenses, principalmente em casos envolvendo agressões sexuais ou em casos de paternidade com ausência de pai (Jobling *et al.* 1997, Kayser *et al.* 1997). Contudo, nos casos onde há coincidência entre os haplótipos obtidos a partir da cena de crime e do suspeito, é necessário que se informe um parâmetro que represente a raridade deste haplótipo. Para este fim, necessita-se o estabelecimento de bancos de dados haplotípicos baseados em amostras da população de origem do suspeito. Vários pesquisadores estabeleceram um haplótipo mínimo para a inclusão de dados de Y-STR em uma base de dados central, disponível em <http://www.ystr.org>. O haplótipo mínimo consiste dos seguintes marcadores: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, e DYS385. Desta forma, a partir da busca do haplótipo de um indivíduo genotipado nos marcadores acima exibidos, é possível se obter a frequência deste haplótipo em diversos

grupos populacionais. Contudo, devido ao seu pequeno tamanho efetivo (1/4 dos marcadores autossômicos) e também a patrilocalidade, ou a tendência das mulheres, ao se casarem, de se moverem para o local de nascimento dos homens, as populações humanas são mais estruturadas no que se refere aos marcadores de cromossomo Y (Jobling e Tyler-Smith, 2003). Esta característica traz a necessidade do estabelecimento de bases de dados haplotípicas regionais, especialmente nos casos de populações isoladas.

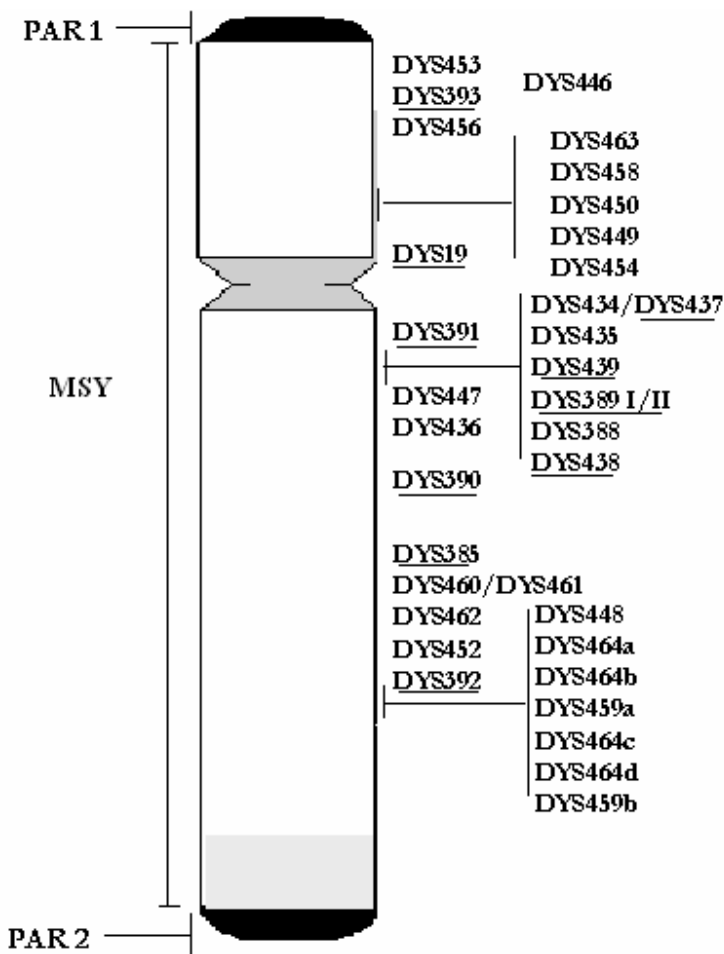


Figura 3 - Representação esquemática do cromossomo Y humano, mostrando as regiões pseudoautossômicas (PAR1 e PAR2) além da região MSY. Sublinhados estão os marcadores utilizados nesta tese.

Outra característica dos microssatélites refere-se à presença de homoplasia, ou a geração da mesma seqüência a partir de rotas independentes (Xu *et al.* 2000) e também devido às altas taxas de mutação dos STRs, a definição de haplogrupos é realizada principalmente a partir de marcadores bialélicos. Contudo, nos casos onde um haplogrupo é fortemente associado a uma população de origem, então os haplótipos definidos por STRs dentro deste

haplogrupo serão também característicos da mesma população (Jobling, 2001), podendo ser utilizados para definir haplogrupos (Bosch *et al.* 1999; Foster *et al.* 2000). Thomas *et al.* (2000) em um estudo com populações africanas e de origem judaica, classificaram as amostras analisadas em quatro haplogrupos, definidos a partir da genotipagem de 6 marcadores bialélicos. As amostras foram a seguir genotipadas usando seis Y-STRs (DYS19, DYS388, DYS390, DYS391, DYS392 e DYS393). Dos 142 haplótipos obtidos a partir dos microsátélites, somente 2 foram encontrados em mais de um grupo. Além disto, os autores observaram, nos grupos africanos, a grande frequência do haplótipo DYS19*15, DYS390*12, DYS390*21, DYS391*10, DYS392*11, DYS393*13 e daqueles que diferem deste haplótipo em um motivo de repetição. Assim, este haplótipo foi identificado como característico da população Bantu. Pereira *et al.* (2002) encontraram elevada frequência deste haplótipo em populações africanas, identificando-o como uma assinatura da expansão Bantu em diferentes populações. Desta forma, segundo Thomas *et al.* (2000) é improvável que convergência e deriva genética possam explicar os casos em que duas ou mais populações compartilham haplótipos muito frequentes, sendo estes haplótipos então indicativos da origem de algumas populações.

No Brasil, diversas publicações trataram do estabelecimento de bases de dados haplotípicas a partir de Y-STRs para as diversas regiões do país, para fins forenses e de genética de populações (Grattapaglia *et al.* 2005; Góes *et al.* 2005, Palha *et al.* 2006; Pereira *et al.* 2006).

1.2.2. O Cromossomo X

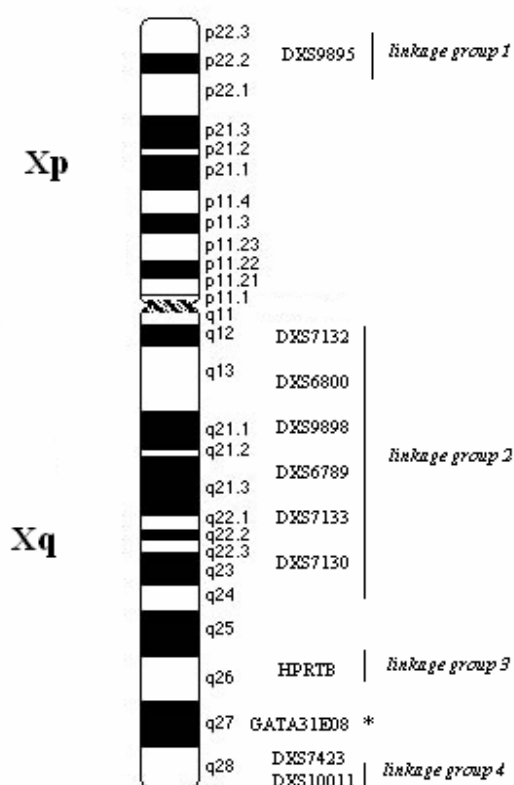
O Cromossomo X humano compreende 165 Mb e constitui aproximadamente 5% do genoma humano (Graves, 2002). Contudo, a densidade gênica neste cromossomo é uma das menores, comparando-se com os dados obtidos em outros cromossomos. Esta característica pode estar associada a uma baixa densidade gênica no cromossomo ancestral autossômico, ou mesmo à ação da seleção, favorecendo à transposição de genes do cromossomo X para os autossômicos durante a evolução (Ross *et al.* 2005).

Schaffner, (2004) realizou uma análise comparativa das características dos marcadores de cromossomos X, Y, autossômicos e de DNA mitocondrial para estudos de genética de populações. Segundo este autor, os marcadores de cromossomo X apresentam várias características peculiares, que devem ser consideradas quando utilizados nestes estudos: a) Baixas taxas de mutação, ocasionadas pelo fato que estes marcadores permanecem 2/3 de sua existência em mulheres- como as mulheres apresentam taxas de mutação menores do que homens (Li *et al.* 2002b), mutações ocorrem menos freqüentemente no cromossomo X do que em autossomos; b) Menor tamanho efetivo - o tamanho efetivo do cromossomo X equivale à $\frac{3}{4}$ dos autossomos, determinando uma ação mais rápida da deriva genética, e fazendo com que a estrutura populacional seja mais pronunciada nestes marcadores; c) Intervalos extensos de desequilíbrio de ligação – devido ao fato que o cromossomo X somente recombina em mulheres, sua taxa de recombinação corresponde à $\frac{2}{3}$ da média do genoma. Este fator, associado à idade mais jovem do cromossomo X, faz com que a recombinação tenha menos tempo de atuar para dissolver o desequilíbrio de ligação ora presente, resultando em intervalos mais extensos de desequilíbrio de ligação no cromossomo X do que em autossomos.

Szibor *et al.* (2003) estimaram as distâncias genéticas entre vários microssatélites de cromossomo X (X-STRs) utilizando dados de 128 meioses informativas de mães e sua proles. Quando foi aplicado um ponto de corte de 2.0 para o máximo valor de *LOD score*, quatro grupos de ligação foram identificados (Figura 4). Estes grupos de ligação mostram uma tendência geral para a possibilidade de ligação entre os marcadores presentes em cada grupo. Contudo, o simples fato dos marcadores estarem presentes no mesmo grupo de ligação não significa que os mesmos estejam ligados, assim como marcadores localizados em cromossomos diferentes podem estar em desequilíbrio de ligação (Buckleton *et al.* 2005). De fato, como o desequilíbrio de ligação é influenciado por fatores demográficos, populacionais e estocásticos (Ardlie, 2002) e como estes fatores atuam de forma singular em populações diferentes, o padrão de desequilíbrio de ligação é único em cada população estudada.

Outrossim, como o cromossomo X somente recombina em mulheres, a obtenção de dados haplotípicos somente é possível a partir de amostras de homens. Desta forma, baseados nas características únicas do cromossomo X, acima ressaltadas, vários autores utilizaram X-STRs para a realização de estudos sobre a extensão do desequilíbrio de ligação (LD) em diversas populações humanas, e também para analisar os efeitos de diferentes histórias demográficas sobre os níveis de LD (Laan e Paabo, 1997, Wilson e Goldstein, 2000, Zavattari *et al.* 2000, Tailon-Miller *et al.* 2000, Katoh *et al.* 2002, Laan *et al.* 2004). Assim, Laan e Paabo, (1997) estudaram marcadores da região Xq em populações com diferentes histórias demográficas, como os Saami, Finlandeses e Suecos. Estes autores encontraram maior amplitude de LD nos Saami, uma população isolada que não havia sofrido expansão demográfica recente. Outros autores, como Zavattari *et al.* (2000) e Katoh *et al.* (2002), também detectaram maior amplitude de LD em populações isoladas, como as da Sardenha e

Mongólia. Por outro lado, Wilson e Goldstein, (2000) observaram altos valores de LD gerados por processos de miscigenação em uma população Bantu. No Brasil, somente um trabalho (Pereira e Pena, 2006) avaliou a estrutura genética e filogeografia de amostras de populações ameríndias e miscigenadas, mas sem resultados conclusivos.



* Este marcador não foi incluído em qualquer grupo de ligação devido à falta de informações na literatura disponível sobre sua localização.

Figura 4 – Localização dos quatro grupos de ligação do cromossomo X proposto por Szibor *et al.* (2003), juntamente com os X-STR utilizados neste trabalho.

1.3. ASPECTOS HISTÓRICOS DA POPULAÇÃO DO RIO GRANDE DO SUL

A população do Estado do Rio Grande do Sul compreende, segundo IBGE (2000), um total de 10.187.798 indivíduos. Esta população caracteriza-se pelo seu alto nível de heterogeneidade, resultado de cinco séculos de cruzamentos interétnicos entre imigrantes europeus, escravos africanos e ameríndios. Em relação aos indígenas do Estado do Rio Grande do Sul, Flores (1990) destaca os grupos Jês, pampeanos e Guarani como ocupantes do estado antes da chegada dos primeiros imigrantes europeus.

Espanhóis e portugueses foram os primeiros europeus a chegar no Rio Grande do Sul, nos séculos XVI e XVII, como desbravadores e exploradores (Cesar, 1970). A partir da chegada dos primeiros imigrantes portugueses, o cruzamento entre homens europeus e mulheres indígenas se tornou comum e até mesmo encorajado como uma estratégia para crescimento populacional e de ocupação do país (Morner 1967). De acordo com Flores, (1990), durante o período colonial, o Rio Grande do Sul sofreu os choques das frentes coloniais lusas e espanholas pelo domínio da área. Os espanhóis tinham orientação específica para colonizar e povoar a terra através de vilas e povoados. O sistema colonial luso consistia em ocupar o litoral com cidades, para facilitar o comércio com Portugal e povoar o interior com sesmarias. A partir do século XVIII teve início um fluxo colonizador mais expressivo, em especial de casais provenientes de ilhas atlânticas, como a ilha dos Açores. As famílias enviadas foram assentadas às margens dos rios Pardo e Taquari, junto à Vila de Rio Grande e na península onde mais tarde surgiria a Vila de Porto Alegre (Flores, 1990). Até meados de 1800, estima-se que aproximadamente 500.000 portugueses migraram para o Brasil (Salzano e

Freire-Maia, 1967). Após este período, novas ondas migratórias sucederam-se, desta vez oriundas de outras nacionalidades, como Itália, Espanha e Alemanha. A primeira leva de imigrantes alemães chegou ao Rio Grande do Sul em 1824, e se estabeleceu na Colônia de São Leopoldo. Os primeiros imigrantes italianos chegaram ao Brasil para trabalharem nas fazendas de café, substituindo os negros em São Paulo. No Rio Grande do Sul, por volta de 1870, foram criadas as colônias de Conde D'Eu e Dona Isabel, localizadas entre o rio Caí e a localidade de Vacaria, além de outras em Garibaldi, Carlos Barbosa e Bento Gonçalves, habitadas por imigrantes italianos. A partir de 1875, estes imigrantes, juntamente com os alemães, foram os pioneiros na agricultura e na indústria da Encosta da Serra, expandindo-se posteriormente a outras áreas do Estado (Cesar, 1970; Flores, 1990).

Na metade do século XVII, africanos foram trazidos para o Brasil como escravos, para trabalharem na cultura da cana de açúcar e após, na mineração de ouro e diamantes (Alves-Silva *et al.* 2000). No Rio Grande do Sul, os primeiros africanos chegaram para trabalharem como tropeiros e sesmeiros (Flores, 1990). Quanto a suas regiões de origem, há indícios que tenham se originado da Guiné e do Congo. Dados históricos do início do século XX mostram que havia maiores concentrações de negros nas charqueadas e nos portos da Capital e das localidades de Rio Grande, São José do Norte, Jaguarão, Triunfo e Taquari (Cesar, 1970).

Estas sucessivas ondas migratórias de origens diversas levaram a um elevado nível de mistura étnica para a população do Brasil, o que refletiu-se também no estado do Rio Grande do Sul, o qual apresenta regiões com diferentes graus de mistura e padrões de colonização distinto entre as mesmas. Outrossim, aspectos de história populacional, como práticas culturais relacionadas às estruturas sociais, além de diferenças comportamentais entre homens e

mulheres durante eventos de colonização e migração, também influíram na formação étnica da população do Rio Grande do Sul.

1.3.1. As populações indígenas

As populações indígenas sul-americanas exibem uma notável diversidade em seus aspectos culturais, lingüísticos e biológicos. Das populações indígenas características da região sul do Brasil, podemos citar os Guarani, Caingang e Charrua/Minuano. Os Caingang eram moradores do Planalto, com extensão para os demais Estados da Região Sul e Misiones (Argentina). Eram coletores, caçadores, pescadores e horticultores. Quanto às suas características físicas, tinham aspecto mongolóide, estatura mediana, rosto oval, nariz achatado, pele bronzeada e cabelo liso. Eram conhecidos como Botocudos, Aweikoma, Bugre e Guayaná, e seu idioma é o Kaingáng, pertencente ao tronco das línguas Jê (Lazzarotto, 1978). Foram dizimados por uma peste por volta de 1610 e, sobretudo, pelos bandeirantes. Descendentes desses e alguns outros grupos foram reunidos por Telêmaco Borba, em 1882, sob uma única denominação: os Kaingangs (Kaa=mato, ingang=morador). Atualmente, os Caingang vivem em reservas indígenas nos Estados das regiões sul e centro-oeste do Brasil. Deste total, aproximadamente 11.000 vivem no Rio Grande do Sul (Ricardo, 2000).

Os Charrua e Minuano habitavam o pampa gaúcho e uruguaio, bem como grande parte da região de Entre Rios, na Argentina (Becker, 2003). Por razões históricas, ou mesmo pela falta de informações disponíveis, algumas vezes são tratados como um grupo só. Pertenciam à etnia pampeana, com estatura mediana, nariz aquilino e levemente achatado, cor morena acentuada e cabelos lisos. As diferenças entre estes grupos concentravam-se principalmente na altura, sendo os Minuano mais baixos e menos robustos do que os Charrua. Quanto a sua distribuição geográfica, os Charrua habitavam o oeste do estado, ocupando ambas as margens

do rio Uruguai e tendo mais contato com os espanhóis; os Minuano, por sua vez, localizavam-se mais para leste, nas áreas irrigadas pelas lagoas dos Patos, Mirim e Mangueira, com extensão até as proximidades de Montevideo (Lazzarotto,1978). Falavam uma língua identificada com o grupo lingüístico Ge-Pano-Caribe (Greenberg, 1987), e foram praticamente exterminados durante as colonizações portuguesa e espanhola. Dos que restaram, grande parte foi assimilada como peões nas estâncias do pampa gaúcho e uruguaio (Lazzarotto, 1978).

Os Guarani eram originários das florestas da região sudoeste da Amazônia, que migraram em direção sul durante os séculos XVI-XVIII, espalhando-se por toda a bacia do Prata, em terras brasileiras, argentinas, uruguaias e paraguaias. Falavam a língua Tupi-Guarani, e, quanto às suas características físicas, apresentavam estatura média/baixa, rosto largo, cabelos lisos e olhos pretos (Lazarotto, 1978). Em 1603, os padres jesuítas fundaram a primeira redução jesuítica Guarani, em uma região que abrangia o noroeste do Rio Grande do Sul, entrecortada pelos rios Paraná e Uruguai, e que incluía também territórios do Uruguai, Paraguai e Argentina. Nestas reduções, viveram milhares de índios Guarani catequizados. Contudo, fatores diversos, como os ataques dos bandeirantes, a guerra guaranítica, a escravidão imposta pelo governo militar durante o período após a expulsão dos jesuítas e a mestiçagem das mulheres Guarani com os homens brancos, fizeram com que os Guarani históricos desaparecessem (Flores, 1996). Atualmente, os Guarani vivem em reservas indígenas, nos estados do sul do Brasil e também em Mato Grosso do Sul e Mato Grosso.

1.3.2. Estudos de miscigenação no Rio Grande do Sul

Diversos trabalhos analisaram as proporções de mistura através de marcadores protéicos e de DNA (nucleares e uniparentais) da população do Rio Grande do Sul.

Em um estudo realizado por Schneider e Salzano, (1979), utilizando-se o sistema Gm de imunoglobulinas, a contribuição européia foi estimada em 53 % para uma amostra de negros de Porto Alegre. Franco *et al.* (1982) também em uma amostra de negros de Porto Alegre, encontraram 65% de contribuição européia utilizando doze sistemas protéicos. Bortolini *et al.* (1997) estimaram a contribuição relativa de africanos no “pool” gênico da população afro-derivada de Porto Alegre e região metropolitana em 41%, enquanto que a contribuição européia foi de 59%. Nessa estimativa, 13 sistemas genéticos foram utilizados.

Dornelles *et al.* (1999) utilizaram 17 marcadores protéicos em populações de diferentes mesoregiões do estado do Rio Grande do Sul, com o objetivo de avaliar a variação entre as mesoregiões, bem como o grau de mistura étnica de cada uma delas. Também foram obtidos dados relativos a nacionalidades e sobrenomes. Os resultados encontrados não demonstraram diferença significativa das frequências alélicas entre as mesoregiões, considerando-se os locos individualmente. No entanto, na análise multivariada, usando-se a distância D_A , as mesoregiões da metade nordeste do estado separaram-se das da outra metade. Resultados semelhantes foram obtidos usando-se a nacionalidade dos pais e os sobrenomes dos indivíduos estudados. Outrossim, os autores estimaram uma contribuição média de 82% de europeus, 7 % de africanos e 11% de ameríndios, para a população caucasóide do Rio Grande do Sul. Estimativa semelhante de contribuição africana (8%) foi obtida por Franco et al. (1982) em

outra amostra de indivíduos classificados como euro-descendentes. Zembruski *et al.* (2006) baseados em 10 marcadores informativos de ancestralidade, genotiparam 203 indivíduos com ancestralidade européia, sendo 101 com doenças cardiovasculares e 102 controles, encontrando valores similares (na faixa de 5 a 6 %) para os níveis de ancestralidade africana de casos e controles estudados.

Marrero *et al.* (2005) analisaram a contribuição parental dos grupos ameríndio, europeu e africano a partir de marcadores uniparentais (Y-SNPs e mtDNA) em uma amostra de 119 indivíduos do Rio Grande do Sul, sendo 88 da cidade de Veranópolis e os 31 restantes representando as diversas regiões do estado. Neste estudo, os autores observaram, para a amostra mais representativa do estado, uma total predominância de linhagens patrilineas européias; enquanto que, para o DNA mitocondrial (herança matrilinea) houve uma maior distribuição nos níveis de contribuição, representada por proporções de 48/36/16 % para as contribuições européia, ameríndia e africana, respectivamente. Contudo, para a amostra de Veranópolis, tanto as estimativas de contribuição para cromossomo Y quanto para mtDNA apresentaram grande predominância de linhagens européias (100% e 97 %, respectivamente); demonstrando a heterogeneidade entre regiões do Rio Grande do Sul.

1.4. JUSTIFICATIVA E OBJETIVOS

Este trabalho visa analisar a estrutura populacional e ancestralidade genômica de populações brasileiras e ameríndias, através da utilização de microsátélites localizados nos cromossomos X, Y e autossômicos. Devido às diferenças nos modos de herança, nas taxas de recombinação e mutação, bem como nas características apresentadas pelos marcadores utilizados, pretendemos obter diferentes visões sobre os aspectos históricos e demográficos destas populações. Outrossim, pretendemos ainda analisar, através de STRs autossômicos, os aspectos acima mencionados sobre um ponto de vista microgeográfico (a partir da classificação das amostras em mesoregiões) para a população do Rio Grande do Sul, a fim de avaliarmos a heterogeneidade apresentada pelas diferentes regiões do Estado.

Os resultados obtidos serão úteis para estudos de mapeamento por associação, além de servirem para o estabelecimento de bancos de dados populacionais com aplicações em exames de identificação humana, e também em estudos filogeográficos e de genética de populações.

Tendo em vista o acima apresentado, pretendemos alcançar os seguintes objetivos específicos:

- 1) Analisar a estrutura genética e a ancestralidade genômica da população do Rio Grande do Sul e de populações ameríndias, utilizando microsátélites autossômicos e de cromossomos sexuais;
- 2) Avaliar o grau de contribuição dos grupos ameríndio, europeu e africano para a constituição das populações do Estado do Rio Grande do Sul ao nível microgeográfico;
- 3) Avaliar a necessidade de estabelecimento de bases de dados regionais, para fins de análises forenses.

Capítulo 2

Linkage Disequilibrium patterns and genetic structure of Amerindian and Brazilian populations revealed by long-range Xq markers.

Manuscrito em preparação

Linkage Disequilibrium patterns and genetic structure of Amerindian and Brazilian populations revealed by long-range Xq markers.

Fabio P.N.Leite . Elzemar M.R. Rodríguez . Sidney E.B. Santos . Sidia M. Callegari-Jacques .
Luiza T. Tsuneto . M. Luiza Petzl-Erler . Francisco M. Salzano . Mara H. Hutz.

Corresponding author:

Mara H. Hutz
Departamento de Genética, UFRGS
Caixa Postal 15053
91501-970 Porto Alegre, RS, Brazil
Tel: 55 51 3316 6720
Fax: 55 51 3316 9825
E-mail: mara.hutz@ufrgs.br

Keywords: X-chromosome microsatellites. haplotypes. Brazil. Native Americans. Population diversity. Linkage disequilibrium

F. M. Salzano . M.H. Hutz (✉)
Departamento de Genética,
Universidade Federal do Rio Grande do Sul,
Caixa Postal 15053, 91501-970, Porto Alegre, RS, Brazil
Fax: 55 51 3316-9823
Tel 55 51 3316-6720
e-mail: mara.hutz@ufrgs.br

Fabio P.N.Leite
Laboratório de Perícias, Instituto Geral de Perícias, Porto Alegre, RS, Brazil

Sidia M. Callegari-Jacques
Departamento de Estatística
Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

Luiza T. Tsuneto . M. Luiza Petzl-Erler
Departamento de Genética, Universidade Federal do Paraná, Curitiba, PR, Brazil

Sidney E.B. dos Santos . Elzemar M. R. Rodrigues
Laboratório de Genética Humana e Médica, Universidade Federal do Pará, Belém, Brazil

Abstract

The extent of linkage disequilibrium (LD) was studied in a Southern Brazil population, and also in a pool of samples from Amerindian tribes. For this purpose, we analyzed 11 microsatellites, located mostly in a Xq region comprising ~86 Mb. The smaller gene diversity of the Amerindian pool, associated to the significant differences found between the populations analyzed, indicated population structure as the main cause for the higher LD values in the Amerindian pool. On the other hand, the LD levels of the Brazilian sample, although less extensive than that of the Amerindian pool, were probably determined by admixture events. Our results indicated that different demographic histories have significant effects on LD levels of human populations, and provide a first approach about the ancestry from the X chromosome of Amerindians and Brazilian populations, being valuable for future studies involving associating mapping and population genetic studies.

Introduction

The X-chromosome is a useful marker for population genetic studies due to its intrinsic properties, such as haplotype accessibility in males, lower mutation and recombination rates, and faster genetic drift due to smaller effective population size. The use of the X chromosome in population genetics is still in its beginning, although it has been widely used as a model to understand the extension and distribution of linkage disequilibrium (LD) in the human genome.

The Brazilian population was formed by successive migratory waves. The Amerindian people were occupying the Brazilian territory when the Portuguese arrived in 1500. After that, Africans were brought to Brazil as slaves, between the sixteenth and the nineteenth centuries. Besides the Portuguese, other migratory waves followed, mainly from Italy, Germany and Spain, during the nineteenth and the twentieth centuries (Salzano & Bortolini 2002). All these migratory events have contributed to the formation of a highly multiethnic admixed population. This heterogeneity was shown in several genetic studies, using either uniparental or autosomal markers, which demonstrated a typical triethnic (European+African+Amerindian) pattern for the Brazilian population gene pool. Nevertheless, several investigations showed that the contributions of these parental populations were not uniform throughout the country. Southern Brazilian populations presented smaller levels of African and higher degrees of European contributions when compared to other Brazilian groups (Alves-Silva et al. 2000, Carvalho-Silva et al. 2001, Salzano & Bortolini, 2002, Callegari-Jacques et al. 2003, Parra et al. 2003, Zembruski et al. 2006).

The extant indigenous populations of the Americas exhibit remarkable cultural, linguistic and biological diversities, but the origins and nature of this variability, as well as the processes that shaped them, are still a matter of controversy (Salzano & Callegari-Jacques, 1988; Salzano, 2002). Studies that use the X-chromosome for population genetics have so far been virtually

absent in Amerindian and Brazilian populations except for a recent investigation (Pereira & Pena, 2006) that addressed the genetic structure of Amerindians and Brazilians using five microsatellite markers, but without conclusive results.

Several investigations have used the X chromosome microsatellites to address the extension of LD in different world populations, and to analyze the effects of different demographic histories on the extension of LD. Laan & Paabo (1997) studied Xq markers in populations with different demographic histories, such as the Saami, Estonian, Finish and Swedish . They found more extensive LD in the Saami, an isolated populations that has not undergone recent expansion. Extensive LD levels were also found by Zavattari et al. (2000) and Katoch et al. (2002) in isolated populations from Sardinia and Mongolia. On the other hand, Wilson & Goldstein (2000) observed consistent LD generated by admixture in a Bantu population. Therefore, the comparison of more populations with different demographic histories may help to better understand this important issue.

The present study investigates the variability of 11 X-chromosome microsatellites markers in the admixed population of Rio Grande do Sul, Brazil's southernmost State, and also in a pool of samples from seven isolated Amerindian populations. Due to their different demographic histories (admixture and isolation), respectively, the objectives of this study are the following: (1) To estimate LD in both groups; (2) To analyze the influence of demographic factors on LD levels observed; and (3) To verify if these X chromosome microsatellites are good markers for population genetics in Brazilian populations.

Material and methods

Populations

The samples were composed of 143 unrelated males from the population of the Brazilian Southernmost State Rio Grande do Sul (RS), and 61 unrelated Amerindian males from the Caingang (n=17), Guarani (n=11), Wai Wai (n=8), Xavantes (n=7), Gavião (n=8), Zoró (n=6), and Suruí (n=4) populations. The characteristics and geographic location of these Amerindian groups are given elsewhere (Kolhrausch et al. 2005). All subjects were informed about the aims of this study and signed a written informed consent form.

Laboratory methods

Genomic DNA was extracted from blood samples using a salting out method (Miller et al. 1988). A total of 11 X-chromosome microsatellites (DXS9895, DXS7132, DXS6800, DXS9898, DXS6789, DXS7133, DXS7130, HPRTB, GATA31E08, DXS7423 and DXS10011) were studied. The assignment of these markers into linkage groups is shown in Figure 1, according to the classification proposed by Szibor et al. (2003). Primers for the markers were designed using Primer3 software (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi, based on published DNA sequences, available in Genbank. Primers were designed such that all the PCR products could be multiplexed in one PCR reaction. The multiplex PCR was performed in a volume of 12.5 uL, with a reaction mix containing: 5 ng of genomic DNA, 1X PCR buffer with 3 µM MgCl₂, 1.25 mM each dNTPs, 1U AmpliTaq DNA Polymerase (Applied Biosystems, Foster City, CA). The amplification conditions consisted of an initial denaturation at 95°C for 11 min, followed by 10 cycles at 94°C for 1 min, 60°C for 1 min, 70°C for 2 min, plus 17 cycles of 90°C

for 1 min, 60°C for 1 min and 70°C for 2 min, and a final extension of 60°C for 60 min. Allelic ladders for each marker were constructed by allele sequencing performed on purified fragments from DNA of hemizygous male subjects using the BigDye Terminator v.1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). Electrophoresis of the amplified fragments was performed in a 310 ABI PRISM Genetic Analyzer (Applied Biosystems, Foster City, CA), according to the manufacturer's recommended protocol. The GeneScan v3.2.1 analysis software was used to track lanes and measure fragment sizes, while the Genotyper v2.5.2 software was employed to automatically designate alleles by comparison with locus specific allelic ladders.

Statistical analysis

Allele frequencies were obtained with the Arlequin software (Schneider et al. 2000). Locus diversity and mean number of alleles per locus were calculated according to Nei (1987). Heterogeneity between the RS and Amerindian groups was tested using Analysis of Molecular Variance (AMOVA, Michalakis & Excoffier 1996). For LD analysis, we calculated the D' parameter, for all the possible pair of loci, with the GOLD software (Abecasis & Cookson, 2000). The respective P values (uncorrected) were obtained using the standard Chi-square test, implemented in the same software. As the Chi-square test is sensitive to rare allele frequencies, GOLD pools rare alleles at a default frequency of 7 % threshold by default. In addition, to ensure the P values were not influenced by multiple testing, the P values were corrected by the Holm-Sidak procedure, described in Lautenberger et al. (2000). The multilocus association parameter r_d was also estimated, through the use of the software Multilocus (Agapow & Burt, 2001), to assess the extent of an overall LD between the loci studied.

Results

Allele frequency distributions for the 11 loci are shown in Table 1. The average locus diversity was 0.76 in the RS sample and 0.68 in the Amerindians,. The mean number of alleles per locus, another parameter which measures genetic diversity, was also higher in RS (10.3) than Amerindians (7.3). As expected for a larger and ethnically more diversified sample, the genetic diversity values were higher for RS than for the Amerindian group.

To study the ancestry of the X chromosomes in RS and Amerindians, comparisons between allele frequencies of the Brazilian samples and those described for Asian and European populations (Zarrabeitia et al. 2005; Edelman et al. 2001; Pepinski et al. 2005; Chen & Pu 2004; Lee et al. 2004; Shin et al. 2005) were made for all markers. Mean frequency differences were estimated for four groups (Amerindian, Asian, RS and European). Because, to the best of our knowledge, no available data for these markers in African population have been described, this parental population was not included in the comparisons. The mean comparisons are shown in Figure 2. In average, the RS allele frequencies are more similar to those described in European populations whereas, for the Amerindians, the smaller differences were observed in the comparison with the Asian populations.

LD analysis

Table 2 shows the pairwise P values for the two groups studied, before and after the Holm-Sidak correction. As expected, no significant LD was found between DXS9895 (located in Xp) and the other markers located in the Xq region. In Amerindians, 5 of 55 pairs showed

significant LD before correction, but only 2 pairs (between markers DXS7132-DXS9898 and DXS9898-HPRTB) remained significant after correction. For RS, only 1 of 55 pairs showed significance after correction (between markers DXS7130-GATA31E08). In Amerindians, the largest physical distance for significant LD was 45.7 Mb (between markers DXS9898 and HPRTB), while, for RS, this value was 20.9 Mb (between markers DXS7130 and GATA31E08). The D' average was 0.186 and 0.358 in RS and Amerindian samples, respectively (Figure 3). As expected, the multilocus association parameter was higher in Amerindians ($r_d=0.126$) than in RS sample ($r_d=0.086$). When D' values were plotted as a function of physical distance (Figure 4), a lower rate of D' decline was observed in Amerindians with $D'>0.3$ even with an intermarker distance larger than 100 Mb.

Genetic structure

An AMOVA test was performed in order to address the possibility of population structure as the main cause for the elevated LD levels in Amerindians. The results showed that most of the genetic variation (96.5 %) was found within populations, while the genetic variation between populations (Amerindian and RS) comprised 3.5 % of the total variability. The F_{st} (0.035) value was statistically significant ($P<0.00001$), confirming the existence of genetic differentiation between populations.

Discussion

The present study demonstrate that the X chromosome is a useful marker to achieve the goals proposed in this study. In fact, through the use of this marker, we provide significant insights about the genetic structure and the effects of different demographic histories on the LD patterns of human populations.

Our findings also show significant genetic differences between both groups studied, determined by a population substructuring process in the Amerindian group and also ancestral parentage. The distribution of allele frequencies indicate an Asian ancestry for Amerindian X-chromosomes and an European origin for RS X-chromosomes. These results are not unexpected, since the European heritage predominates in Southern Brazil (Callegari-Jacques et al. 2003) and there is a general agreement about the Asian source of the Amerindian ancestral lineages (Cavalli-Sforza et al. 2004).

Considering that demographic factors are responsible for the presence of LD across large extensions of the genome, and because most investigations used markers restricted to short regions of the X-chromosome, which is less appropriate to measure long-range LD (Tallon-Miller et al. 2000, Katoh et al. 2002, Laan et al. 2004) we decided, firstly, to use microsatellites, because these markers provide more power to detect LD over large genome extensions than SNPs (Varilo et al. 2000). Secondly, we chose markers located mostly over large genetic distances across the long arm (Xq) region and one from the short arm (Xp) region. This Xp marker was used as a control for our results, since LD is not expected to occur between the Xp and Xq

regions, due to their independent pattern of segregation. Thus, we used markers widely distributed over 86 Mb of the Xq region, to provide a general overview of the LD patterns in this region of the X-chromosome, and to allow the characterization of these patterns in populations with different demographic histories.

One important finding of our study is the existence of more extensive LD pattern on the Amerindian (isolated) than on the RS (admixed) group. Population structure, possibly originated by ancestral parentage, endogamy, and eventual founder effects, probably produced the higher LD, and the significant P value over 46 Mb of the Xq region between markers HPRTB and DXS9898 in the Amerindian gene pool. Moreover, admixture events, restricted to the Caingang and Guarani tribes (Petzl-Erler et al. 1993), may also have contributed to the LD results in this group. In this way, our results are concordant with other studies. Laan & Paabo (1997) studying microsatellite markers in the Xq13 region, detected more extensive LD (even between markers >10 Mb apart) in the Saami, an isolated population that has not undergone recent population expansion. Using common SNPs and microsatellites, several researchers (Peterson et al. 1995; Collins & Morton 1998; Wright et al. 1999) found significant LD between markers separated by distances of 100 kb to several megabase pairs, and Launterberger et al. (2000) observed significant linkage disequilibrium-signals over a 30cM region in chromosome 1.

For the RS group, due to the fact that the markers located in closer physical distances than the DXS7130 and GATA31E08 did not show significant P-values, our results again indicate possible effects of demographic events on LD levels. For this group, the present results suggest that admixture was responsible for the significant LD values on DXS7130-GATA31E08 pair. Previous studies, using autosomic and uniparental data, have reported significant levels of admixture to the RS populations. Marrero et al. (2005) using mtDNA data, reported expressive values of Native American (36%) and African (16%) contribution for the RS population, despite

the predominance of European contribution (48%). Callegari-Jacques et al. (2003) using autosomic markers, found the proportions of 82%-11%-7% for the European-African-Amerindian parental contribution in the Brazilian southern Region. These information allow us to build a scenario, where historical admixture between differentiated populations resulted in elevated LD levels for some genomic regions, which were not totally eroded by recombination, given the short period of admixture of the RS population. In fact, LD over large genomic regions generated by admixture was described by several authors (Chackraborty & Weiss, 1988; Stephens et al. 1994; Wilson & Goldstein 2000).

However, despite the effects of demographic factors on the LD levels, there are other factors that need to be considered to explain the expressive LD levels found herein. Zavattari et al. (2000) and Taillon-Miller et al. (2000) demonstrated that the LD patterns of some X chromosome regions are similar even in populations with different demographic histories, indicating the LD patterns are uneven across genomic regions. Moreover, as the X Chromosome has less opportunity to recombine (only in females), associated to its smaller effective size ($2/3$ of the genome average) and younger age (less time to recombine), yield an expectation of higher LD levels on the X chromosome than in the other chromosomes. Service et al. (2001) using microsatellites densely spaced across the genome, found higher levels of background linkage disequilibrium in the X chromosome than on other regions of the genome.

In conclusion, our study demonstrates the influence of demographic factors on the LD levels of two populations with different demographic histories, indicating the possibility that these effects were more extensive in the Amerindian than in the RS population.

Acknowledgments

Financial aid was provided by Institutos do Milênio, Programa de Apoio a Núcleos de Excelência (PRONEX), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Secretaria Nacional de Segurança Pública (SENASP), and Instituto Geral de Perícias (IGP) from the State of Rio Grande do Sul. The logistic support of the Fundação Nacional do Índio in the Amerindian studies is gratefully acknowledged. This investigation was approved by the Brazilian National Ethics Commission (CONEP Resolution no. 123/98).

References

- Abecasis GR, Cookson WO (2000).GOLD-graphical overview of linkage disequilibrium. *Bioinformatics* 16:182-183.
- Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* 1:101-102.
- Alves-Silva J, Santos MS, Guimarães PEM, Ferreira ACS, Bandelt HJ, Pena SDJ, Prado VF (2000). The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67:444-461
- Callegari-Jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Crossetti SG, Ferreira ME, Hutz MH (2003) Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. *Am J Hum Biol* 15:824-834
- Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ (2001) The phylogeography of Brazilian Y-chromosome lineages. *Am. J. Hum. Genet.* 68: 281–286.
- Cavalli-Sforza LL, Piazza A, Menozzi P (1994) History and geography of human genes. Princeton University Press,Princeton, NJ.
- Chakraborty R, Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci USA* 85:9119-9123.
- Chen MY, Pu CE (2004) Population data on the X chromosome short tandem repeat loci DXS10011, DXS101, DXS6789,DXS7132, DXS8377, and DXS9895 in Taiwan. *Forensic Sci Int* 146:65–67.
- Collins A, Morton NE (1998) Mapping a disease locus by allelic association. *Proc Natl Acad Sci USA* 95:2742-2745.
- Edelmann J, Hering S, Michael M (2001) 16 X-chromosome STR loci frequency data from a German population. *Forensic Sci Int* 124:215–218.

- Katoh T, Mano S, Ikuta T, Munkhbat B, Tounai K, Ando H, Munkhtuvshin N, Imanishi T, Inoko H, Tamiya G (2002) Genetic isolates in East Asia: A study of linkage disequilibrium in the X chromosome. *Am J Hum Genet* 71:395-400.
- Kohlrausch F, Callegari-Jacques SM, Tsuneto LT, Petzl-Erler ML, Hill K, Hurtado M, Salzano FM, Hutz MR. 2005. Geography Influences Microsatellite Polymorphism Diversity in Amerindians. *Am J Ph Ant* 126:468-470.
- Laan M, Paabo S (1997) Demographic history of and linkage disequilibrium in human populations. *Nat Genet* 17:435-438.
- Laan M, Wiebe V, Khusnutdunova E, Remm M, Paabo, S (2004) X-chromosome as a marker for population history: linkage disequilibrium and haplotype study in Eurasian populations. *Eur J Hum Genet* 1-11.
- Lautenberger JA, Stephens JC, O'Brien SJ, Smith MW (2000) Significant admixture linkage disequilibrium across 30 cM around the FY locus in African Americans. *Am J Hum Genet* 66:969-978.
- Lee HY, Park MJ, Jeong CK, Lee SY, Yoo JE, Chung U, Choi JH, Kim CY, Shin KJ (2004) Genetic characteristics and population study of 4 X-chromosomal STRs in Koreans: evidence for a null allele at DXS9898. *Int J Legal Med* 118: 355–360.
- Marrero AR, Leite FPN, Carvalho BA, Peres LM, Kommers TC, Cruz IM, Salzano FM, Ruiz-Linares A, Silva-Junior 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
- Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* 142:1061-1064

- Miller SA, Dykes DD, Polesky HF (1988) A single salting out procedure for extracting DNA from human nucleated cells. *Nucl Ac Res* 16:1215
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.
- Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SJD (2003) Color and genomic ancestry in Brazilians. *Proc Nat Acad Sci USA* 100:177-182
- Pepinski W, Skawronskaa M, Niemcunowicz-Janica A, Koc-Zorawskaa, Janicaa J, Soltyszewskib I (2005) Polymorphism of four X-chromosomal STRs in a Polish population sample. *For Sci Int* 151:93–95.
- Pereira RW, Pena SD (2006) Phylogeography of haplotypes of five microsatellites located in a low-recombination region of the X chromosome: studies worldwide and in Brazilian populations. *Genetica* (2006) 126:243–250.
- Peterson AC, Di Rienzo A, Lehesjoki AE, de la Chapelle A, Slatkin M, Freimer MB (1995) The distribution of linkage disequilibrium over anonymous genome regions. *Hum Mol Genet* 4:887-94.
- Petzl-Erlor 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
- Salzano FM, Callegari-Jacques SM (1998) *South American Indians. A case study in evolution*. Oxford, UK: Clarendon press.
- Salzano FM (2002) Molecular variability in Amerindians: widespread but uneven information. *Ann Braz Acad Sci* 74:223-263.
- Salzano FM, Bortolini MC (2002) *The evolution and genetics of Latin American populations*. Cambridge University Press, Cambridge.
- Schneider S, Roessli D, Excoffier L (2000) *Arlequin (ver.2000) – a software for population genetics data analysis*. University of Geneva, Geneva

- Service SK, Ophoff RA, Freime NB (2001) The genome-wide distribution of background linkage-disequilibrium in a population isolate. *Human Molecular Genetics* 10(5):545-551.
- Shin SH, Yu JS, Park SW, Min GS, Chung KW (2005) Genetic analysis of 18 X-linked short tandem repeat markers in Korean population. *Forensic Sci Int* 147:35–41
- Stephens JC, Briscoe D, O'Brien SJ (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *Am J Hum Genet* 58:595-608.
- Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003) Use of X-linked markers for forensic purposes. *Int J Leg Med* 117:67-74.
- Tailon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen P, Cao T, Kere J, Piria G, Rice JP, Kwoj PY (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat Genet* 25:324-328.
- Varilo T, Laan M, Hovatta I, Wiebe V, Terwilliger JD, Peltonen L (2000) Linkage disequilibrium in isolated populations: Finland and a young population of Kuusamo. *Eur J Hum Genet* 8:604-612.
- Wilson JF, Goldstein DB (2000) Consistent Long-Range Linkage Disequilibrium Generated by Admixture in a Bantu-Semitic Hybrid Population. *Am J Hum Genet* 67:926-935.
- Wright AF, Carothers AD, Pirastu AM (1999) Population choice in mapping genes for complex diseases. *Nat Genet* 23:297-304.
- Zarrabeitia MT, Alonso A, Zarrabeitia A, Castro A, Fernandez I, Martinez de Pancorbo M. (2005) X-linked microsatellites in two Northern Spain populations. *Forensic Sci. Int.* 145:57-59.
- Zavattari P, Deidda E, Whalen M, Lampis R, Mulargia A, Loddo M, Eaves I, Mastio G, Todd J, Cucca F (2000) Major factors influencing linkage disequilibrium by analysis of different

chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Human molecular genetics* 20:29 47-2957.

Zembruski VM, Callegari-Jacques SM, Hutz MH (2006) Application of an African Ancestry Index as a Genomic Control Approach in a Brazilian Population. *Ann Hum Genet* 70(1-7).

Table 1. Allele frequency distributions for X-chromosome microsatellites in the populations investigated

Systems and alleles	Populations		Systems and alleles	Populations	
	RS (n=143)	Amerindian (n=61)		RS (n=143)	Amerindian (n=61)
	Frequency	Frequency		Frequency	Frequency
DXS6800			Dxs7133		
*15	0.000	0.034	*7	0.007	0.000
*16	0.367	0.655	*9	0.464	0.712
*17	0.014	0.069	*10	0.101	0.133
*18	0.158	0.138	*11	0.333	0.111
*19	0.317	0.086	*12	0.065	0.022
*20	0.022	0.000	*13	0.022	0.000
*21	0.122	0.018	*15	0.008	0.022
DXS7132			Dxs7130		
*11	0.007	0.000	*11	0.021	0.105
*12	0.079	0.078	*12	0.134	0.158
*13	0.293	0.098	*13	0.042	0.123
*14	0.364	0.275	*13.3	0.015	0.000
*15	0.200	0.294	*14	0.028	0.035
*16	0.029	0.216	*14.3	0.232	0.228
*17	0.021	0.039	*15	0.000	0.000
*18	0.007	0.000	*15.3	0.401	0.316
Dxs7423			*16	0.021	0.000
*13	0.096	0.000	*16.3	0.092	0.000
*14	0.272	0.089	*17.3	0.014	0.018
*15	0.441	0.689	*19.3	0.000	0.017
*16	0.140	0.022			
*17	0.051	0.200	Gata31E		
HPRTB			*5	0.015	0.019
*10	0.008	0.000	*6	0.007	0.000
*11	0.113	0.000	*7	0.138	0.058
*12	0.314	0.180	*8	0.065	0.096
*13	0.293	0.480	*9	0.109	0.231
*14	0.188	0.280	*10	0.275	0.538
*15	0.053	0.040	*11	0.254	0.039
*16	0.023	0.020	*12	0.123	0.019
*17	0.008	0.000	*13	0.014	0.000

Continued

System and alleles	Populations		System and alleles	Populations	
	RS n=143	Amerindian (n=61)		RS (n=143)	Amerindian (n=61)
DXS 6789			DXS 10010		
*14	0,014	0.000	*24	0.007	0.020
*15	0,056	0.016	*25	0.007	0.000
*16	0,049	0.148	*26	0.000	0.020
*18	0,007	0.000	*27	0.022	0.020
*19	0,035	0.131	*28	0.007	0.000
*20	0,392	0.361	*29	0.022	0.061
*21	0,231	0.279	*29.2	0.007	0.000
*22	0,126	0.065	*30	0.015	0.020
*23	0,049	0.000	*30.2	0.052	0.000
*24	0,035	0.000	*31	0.022	0.020
*25	0,006	0.000	*31.2	0.067	0.000
			*32	0.037	0.061
			*32.2	0.060	0.020
DXS 9898			*33	0.037	0.000
*7	0.007	0.000	*33.2	0.022	0.000
*8.3	0.210	0.000	*34	0.022	0.061
*10	0.007	0.000	*34.2	0.000	0.020
*11	0.174	0.019	*35	0.069	0.102
*12	0.326	0.264	*35.2	0.000	0.000
*13	0.188	0.509	*36	0.069	0.082
*14	0.080	0.189	*37	0.090	0.102
*15	0.008	0.000	*37.2	0.015	0.041
*16	0.000	0.019	*38	0.030	0.061
			*38.2	0.015	0.041
DXS 9895			*39	0.069	0.082
*12	0.007	0.000	*39.2	0.015	0.000
*13	0.270	0.327	*40	0.075	0.020
*14	0.299	0.545	*40.2	0.015	0.020
*15	0.248	0.091	*41	0.037	0.045
*15.2	0.007	0.000	*41.2	0.015	0.000
*16	0.102	0.018	*42	0.022	0.020
*16.2	0.067	0.000	*42.2	0.007	0.000
*18	0.000	0.018	*43	0.015	0.000
			*43.3	0.014	0.041
			*44	0.015	0.000
			*45.2	0.007	0.000
			*48	0.000	0.020

Table 2 Pairwise P values (before and after correction) of LD between loci in the groups studied

Marker Pair	P value before correction		P value after correction	
	RS	Amerindian	RS	Amerindian
DXS9895-DXS7132	0.1951	0.6113	0.9261	1.0000
DXS9895-DXS6800	0.4000	0.5281	1.0000	1.0000
DXS9895-DXS9898	0.9321	0.7704	1.0000	1.0000
DXS9895-DXS6789	0.4099	0.6834	1.0000	1.0000
DXS9895-DXS7133	0.7349	0.0214	1.0000	0.0829
DXS9895-DXS7130	0.7252	0.3092	1.0000	0.9999
DXS9895-HPRTB	0.4638	0.9009	1.0000	1.0000
DXS9895-GATA31E08	0.8062	0.1224	1.0000	0.7290
DXS9895-DXS7423	0.8892	0.0170	1.0000	0.0501
DXS9895-DXS10011	0.2231	0.1502	0.9824	0.9130
DXS7132-DXS6800	0.8378	0.3234	1.0000	1.0000
DXS7132-DXS9898	0.0696	0.0084	0.3028	0.0084
DXS7132-DXS6789	0.2900	0.2428	0.9996	0.9971
DXS7132- DXS7133	0.2294	0.6525	0.9929	1.0000
DXS7132-DXS7130	0.6515	0.2305	1.0000	0.9931
DXS7132-HPRTB	0.0627	0.1412	0.2282	0.8618
DXS7132-GATA31E08	0.5885	0.2623	1.0000	0.9991
DXS7132-DXS7423	0.3033	0.0748	0.9998	0.3728
DXS7132-DXS10011	0.1773	0.1407	0.8831	0.8379
DXS6800-DXS9898	0.0486	0.5577	0.1388	1.0000
DXS6800-DXS6789	0.3234	0.6257	1.0000	1.0000
DXS6800-DXS7133	0.1313	0.4721	0.7553	1.0000
DXS6800-DXS7130	0.8486	0.0929	1.0000	0.5416
DXS6800-HPRTB	0.7714	0.1208	1.0000	0.6861
DXS6800-GATA31E08	0.4307	0.3249	1.0000	1.0000
DXS6800-DXS7423	0.2699	0.5360	0.9986	1.0000
DXS6800-DXS10011	0.8859	0.4476	1.0000	1.0000
DXS9898-DXS6789	0.6044	0.1471	1.0000	0.8922
DXS9898-DXS7133	0.8452	0.1285	1.0000	0.7797
DXS9898-DXS7130	0.2251	0.7832	0.9869	1.0000
DXS9898-HPRTB	0.2331	0.0101	0.9950	0.0201
DXS9898-GATA31E08	0.6680	0.0762	1.0000	0.4258
DXS9898-DXS7423	0.8901	0.3728	1.0000	1.0000
DXS9898-DXS10011	0.7425	0.2942	1.0000	0.9998
DXS6789-DXS7133	0.0799	0.8898	0.3932	1.0000
DXS6789-DXS7130	0.9218	0.5405	1.0000	1.0000
DXS6789-HPRTB	0.9128	0.5540	1.0000	1.0000
DXS6789-GATA31E08	0.0466	0.1680	0.0910	0.9473
DXS6789-DXS7423	0.0995	0.3369	0.6106	1.0000
DXS6789-DXS10011	0.3502	0.7208	1.0000	1.0000

Continued

Table 2 Pairwise P values (before and after correction) of LD between loci in the groups studied

Marker Pair	P value before correction		P value after correction	
	RS	Amerindian	RS	Amerindian
DXS7133-DXS7130	0.5725	0.9388	1.0000	1.0000
DXS7133-HPRTB	0.0872	0.0321	0.4720	0.1505
DXS7133-GATA31E08	0.0980	0.2218	0.5610	0.9890
DXS7133-DXS7423	0.1961	0.7071	0.9414	1.0000
DXS7133-DXS10011	0.2228	0.9366	0.9772	1.0000
DXS7130-HPRTB	0.7562	0.3283	1.0000	1.0000
DXS7130-GATA31E08	0.0173	0.2355	0.0173	0.9953
DXS7130-DXS7423	0.8708	0.5959	1.0000	1.0000
DXS7130-DXS10011	0.5958	0.5884	1.0000	1.0000
HPRTB-GATA31E08	0.7570	0.3390	1.0000	1.0000
HPRTB-DXS7423	0.2113	0.7475	0.9640	1.0000
HPRTB-DXS10011	0.8849	0.2567	1.0000	0.9985
GATA31E08-DXS7423	0.2758	0.2173	0.9992	0.9845
GATA31E08-DXS10011	0.2277	0.7618	0.9904	1.0000
DXS7423-DXS10011	0.4414	0.3666	1.0000	1.0000

Legend for Figure 1

Figure 1. Location of the markers analyzed on the X-chromosome
(source <http://www.gdb.org/hugo>)

Legend for Figure 2

Figure 2. Average allele frequencies differences between pairs of populations

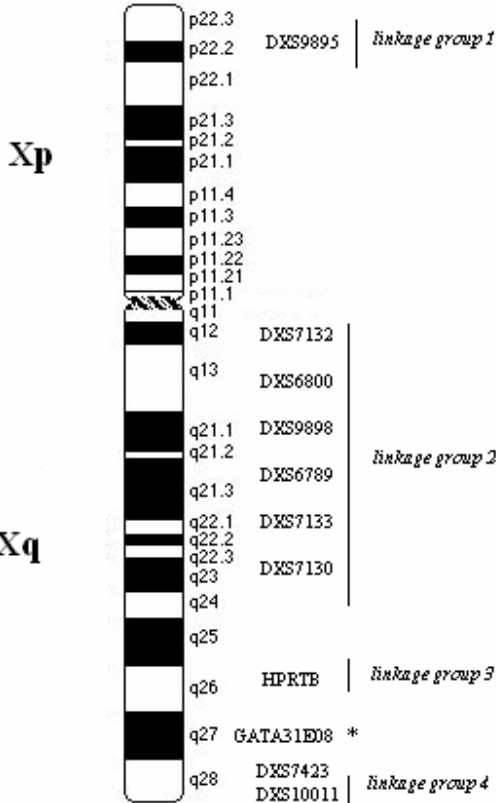
Legend for Figure 3

Figure 3. LD patterns, based on D' values from the RS groups (left) an the Amerindian (right). The scale in the axis represent the markers analyzed, and the scale of colors represent the D' values plotted in the figures.

Legend for Figure 4

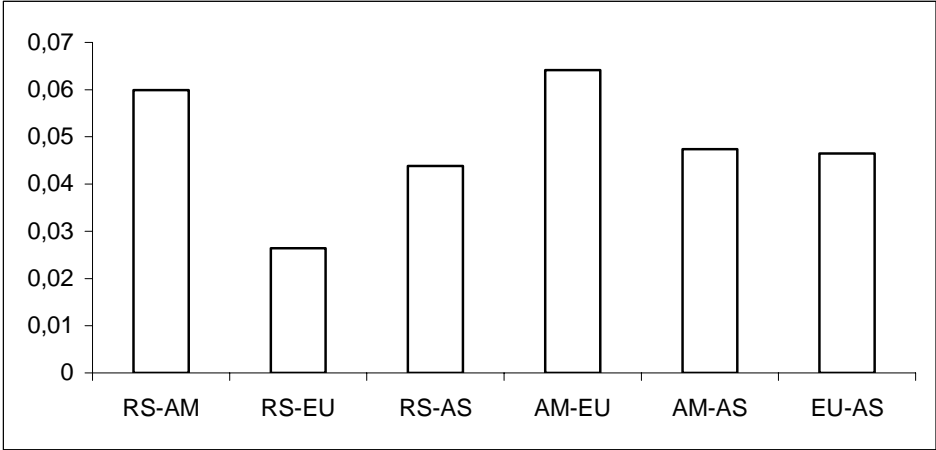
Figure 4. Relationship between mean level of LD (measured by D') and the physical distance (in Mb) in Amerindian and RS populations.

Figure 1



* This marker was not assigned to any linkage group due to the lack of available data in the literature about its genetic location

Figure 2



RS=Rio Grande do Sul; AM=Amerindian; EU=European; AS=Asian

Figure 3

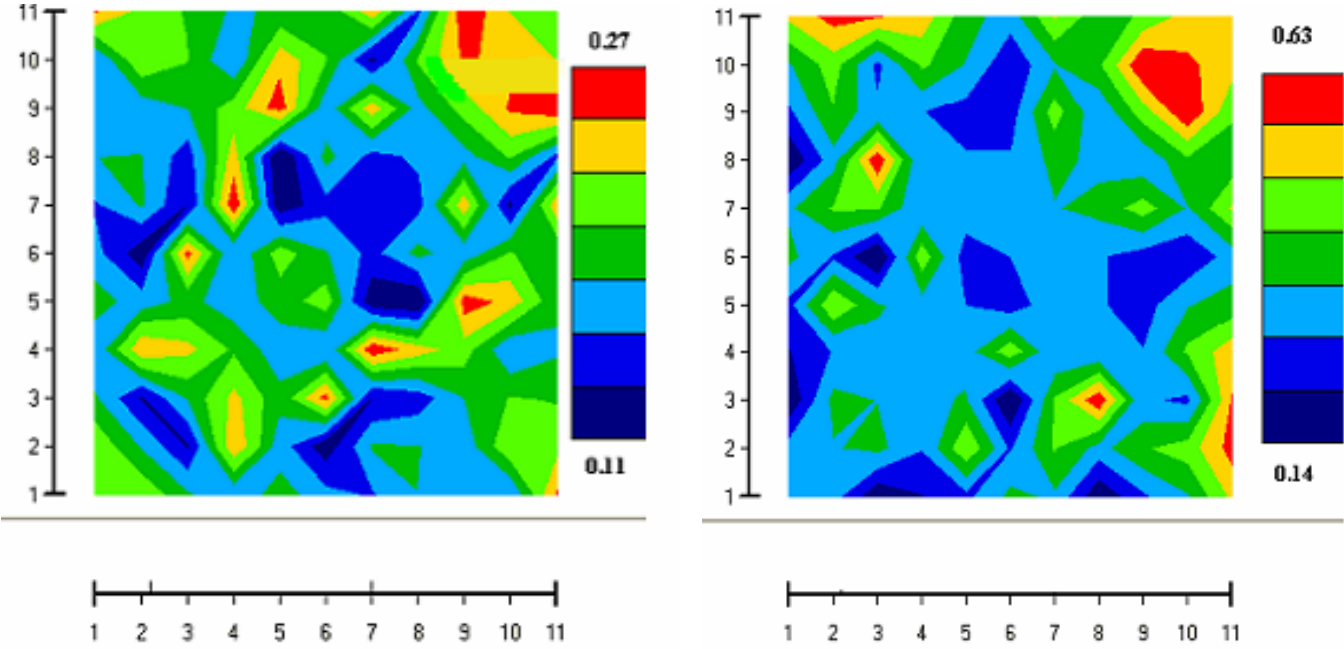
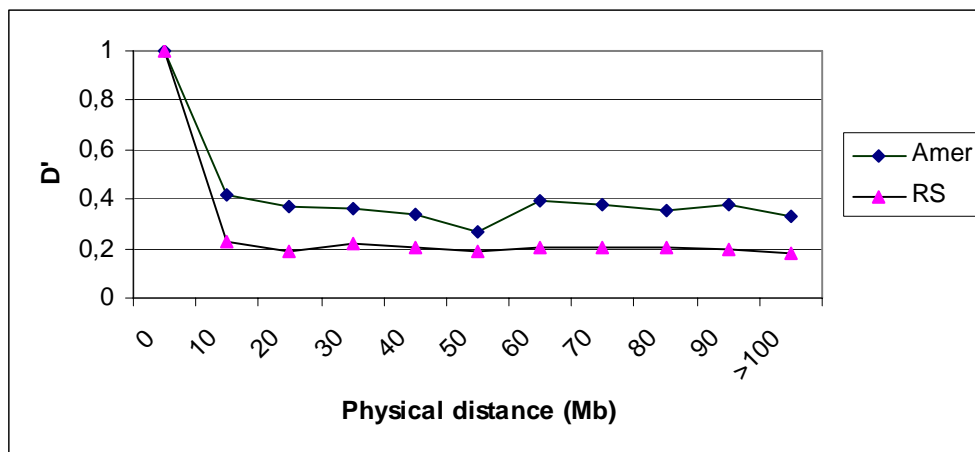


Figure 4



Capítulo 3

The genetic structure of southern Brazil revealed by autosomal markers

Manuscrito em preparação

The genetic structure of southern Brazil revealed by autosomal markers

Fabio P.N.Leite . Sidia M. Callegari-Jacques . Bianca A.Carvalho . Trícia Kommers . Cecília H. F. Matte . Paulo E. Raimann . Solange P. Schwengber . Mara H. Hutz.

Corresponding author:

Mara H. Hutz
Departamento de Genética, UFRGS
Caixa Postal 15053
91501-970 Porto Alegre, RS, Brazil
Tel: 55 51 3316 6720
Fax: 55 51 3316 9825
E-mail: mara.hutz@ufrgs.br

M.H. Hutz (✉)
Departamento de Genética,
Universidade Federal do Rio Grande do Sul,
Caixa Postal 15053, 91501-970, Porto Alegre, RS, Brazil
Fax: 55 51 3316-9823
Tel 55 51 3316-6720
e-mail: mara.hutz@ufrgs.br

Fabio P.N.Leite . Bianca A.Carvalho . Trícia Kommers . Cecília H. F. Matte . Paulo E. Raimann . Solange P. Schwengber .
Laboratório de Perícias, Instituto Geral de Perícias, Porto Alegre, RS, Brazil

Sidia M. Callegari-Jacques
Departamento de Estatística
Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

ABSTRACT

Rio Grande do Sul, the southernmost State from Brazil, has a typical tri-ethnic pattern of genetic contribution. However, previous studies showed a greater European contribution to the gene pool of this State, if compared with other Brazilian regions. Genetic analysis for 15 microsatellites were performed in a sample of 316 individuals from Rio Grande do Sul (RS). No significant genetic differences were detected among RS and other Brazilian regions. The overall admixture analysis showed a pattern of European>Amerindian>African contribution. When we considered only the Western region of the State, several methods of analysis showed a higher Amerindian contribution if compared with the whole State, implicating this region as a reservoir of Amerindian genes. When our data was compared with other Brazilian and world populations, the results showed that Rio Grande do Sul as a whole is genetically closer with Argentina than Bahia and Rio de Janeiro (two Brazilian States). Furthermore, our data demonstrated that, despite the presence of a tri-ethnic pattern of genetic contribution for the RS population, there are differences in the levels of contribution of each parental population within the State. These differences reflect historical differences in the pattern of settlement and migration in South America admixed populations, and highlight the importance of population genetics studies based on smaller geographic divisions for the understanding of the genetic structure of human populations.

Introduction

The study of ancestry patterns and genetic structure of human populations have important applications for the knowledge of historical and geographic aspects of colonization processes, forensic analysis and also the prevalence of certain diseases in different populations. Admixture mapping has been proposed as an efficient approach to localize disease related variants that differ in frequency between historically separated populations . However, for the correct utilization of this tool, the admixture patterns of populations should be fully understood (Parra et al., 2004). The present Brazilian population is the result of centuries of multiethnic crossings between Amerindian, African and European people. Due to different patterns of distribution of immigrants and the continental size of the country, several authors have observed a great heterogeneity in the ancestry of the Brazilian populations, using different markers, like protein markers (Dornelles et al.,1999), SNPs (Parra et al., 2003), Y microsatellites (Carvalho-Silva et al., 2001), mtDNA (Alves-Silva et al., 2000) and autosomal microsatellites (Callegari-Jacques et al., 2003). Dornelles et al., (1999) undertook admixture analysis in several regions from Rio Grande do Sul in a sample classified as European-derived. These authors estimated levels of Amerindian admixture higher than 10% in the Northwest (NW) and Southwest (SW). Moreover, Marrero (2003) using uniparental markers in a sample from the Pampa (SW) of Rio Grande do Sul, described a high proportion of Amerindian mtDNA haplogroups, and suggested that this region could constitute a reservoir of Amerindian genes. The present study investigates the variability in 15 autosomal microsatellites in the population of Rio Grande do Sul. Due to the desirable characteristics of these markers for population genetics studies, such as high mutations rates and high heterozygosity values, our aims were the following: 1) To obtain a more accurate comprehension of the genetic structure of this southern Brazilian population as well as the relation of this population with other Brazilian and South American groups, 2) To verify the extent of heterogeneity of the admixture patterns at the microgeographic level and 3) To address the hypothesis that the West of this State is a source of Amerindian genes.

Subjects and Methods

Populations

The samples consisted of 316 unrelated individuals, from the southern Brazilian State of Rio Grande do Sul (Fig. 1). Based on skin color and morphological characteristics, the individuals were classified as of European ancestry (82 %), African Brazilians (13 %) or has a mixed appearance (5 %). All donors were informed about the aims of this study and signed a written consent form.

Laboratory analysis

DNA was extracted from whole blood samples through a salting-out method (Miller et al., 1988). A total of 15 STR loci (all repetitions of a 4-base-pair motif), D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, vWA, CSF1PO, TPOX and TH01 were determined, using the AmpFISTR Identifiler amplification kit (Applied Biosystems, Foster City, CA). Amplifications were performed following the instructions provided in the kit user manual, with the recommended DNA amount (1.0-2.0 ng). Electrophoresis of the amplified fragments was performed in a 310 ABI PRISM Genetic Analyzer (Applied Biosystems, Foster City, CA), according to the manufacturer recommended protocol. The GeneScan™ v3.2.1 analysis software was used to track lanes and measure fragment sizes, while the Genotyper™ v2.5.2 software was employed to automatically designate alleles by comparison with locus specific allelic ladders.

Statistical analysis

Allele frequencies were obtained by gene counting. To address the possibility of genetic structuring, the RS population was compared with published data, from central-west (CW), south

(S) southeast (SE), northeast (NE) and north (N) Brazilian regions (Callegari-Jacques et al., 2003). For this purpose Pairwise F_{ST} values were estimated, their statistical significance being assessed using 1000 permutations (Schneider et al., 2000) and corrected using the classical Bonferroni method (Wright, 1992). These analyses were conducted with the Arlequin software (Schneider et al., 2000). To disclose the degree of individual Amerindian ancestry of our study population, the software Structure (Pritchard et al., 2000) was employed for the estimation of the number (k) of clusters. For this purpose, we added the genotypes obtained from subjects from the five Brazilian regions (Callegari-Jacques et al., 2003) plus those of Caingang and Guarani individuals (Kohlrausch et al., 2005) to our data. For these analyses, five independent replicates of each value of k (number of clusters) were ran, with a burnin period of 10000, 100.000 Markov Chain Monte Carlo iterations and with information about the population of origin of each individual. The Structure software provides, for each individual, the posterior probability of belonging to each of the clusters identified. We considered one individual as assigned to a cluster when its posterior probability of belonging to it was higher than 50 % (Romualdi et al., 2002).

For the purposes of this article, we organized our data in two sets. The first one, denominated WRS, was composed by 115 individuals whose place of birth included the western region of the State (Northwest plus southwest). The second one, named CERS, contained samples of 201 individuals born in the Center and East of the State. Estimates of parental contributions were obtained using Chackraborty's (1985) gene identity method, implemented in the program Admix (provided by B. Bertoni and available at <http://www.genetica.fmed.edu.uy/software.htm>). Allele frequencies for the parental populations were the same used by Callegari-Jacques et al (2003). The genetic relationship of populations was obtained by the construction of a genetic distance matrix through the software Poptree (Takezaki, 2001), using the modified Cavalli-Sforza D_A distance (Nei, 1983). For this purpose, we compared our data with those available in the

literature from the following populations: Brazil: Rio de Janeiro (Góes et al., 2004) and Bahia (Santos et al., 2004); Argentina: Buenos Aires (Berardi et al., 2003) and Misiones Province (Marino et al., 2006); Africa: Angola (Beleza et al., 2004) and Guinea Equatorial (Alves et al., 2005); Europe: Azores (Velosa et al., 2002), North Portugal (Pinheiro et al., 2005), Northeast Spain (Paredes et al., 2003) and Italy (Garofano et al., 1998); and Amerindian: Caingang and Guarani (Kohlrausch et al., 2005). The genetic distance matrix was then used to perform a Multidimensional Scaling analysis using the software SPSS version 8.0. For all the comparisons among our data with others, we used 12 microsatellite loci (D3S1358, D5S818, D7S820, D8S1179, D13S317, D18S51, D21S11, FGA, VWA, CSF1PO, TPOX and TH01), due to the unavailability of data for the D16S539, D2S1338 and D19S433 markers for some populations used as references.

Results and discussion

Allele frequencies for the 15 microsatellite analyzed are shown in Table 1. The most variable loci were D18S51 and D21S11, which contributed for the mean number of alleles of 10.6 and the average gene diversity value of 0.80, for the RS population. In these and other loci differences are observed when comparing the frequencies of the most common alleles in this samples with regional data from previously published Brazilian populations (Callegari-Jacques et al., 2003). To evaluate the significance of these differences, we calculated pairwise F_{ST} values among these data sets (Table 2). No significant values (after the Bonferroni correction) were found in the comparisons between Rio Grande do Sul (RS) with the samples from the other Brazilian regions. These results are in accordance with previous studies of Brazilian populations (Silva et al., 2004; Góes et al., 2004) and have important implications for forensic purposes, since they support the possibility of developing a single DNA database for the country. However, some

comparisons undertaken herein presented P values close to the threshold for the Bonferroni correction. Additionally, there are not many genetic studies concerning the presence of substructuring in Brazilian populations, if the size and population diversity of the country is considered. Therefore, further studies need to be performed to make a decision about the need of a global or regional databases.

Because Caingang and Guarani Indians were in Rio Grande do Sul when Europeans arrived, genotypes from these settled groups (Kohlrausch et al., 2005) were used for population comparisons with data obtained herein, using the Structure software. The highest log probability of data was that for K=2 ($Ln = -73801.7$ for K=2; -73942.6 for K=1; -74166.3 for K=3 and -74835.8 for K=4), indicating the presence of two clusters. When we analyzed the inferred ancestry of the individuals studied, we observed that 92 % of the Caingang and 100 % of the Guarani persons were assigned to cluster 2, implicating this cluster as related to Amerindian Ancestry (Figure 2). On the other hand, for the RS data, we observed that individual ancestry estimates varied widely. Moreover, when we compared WRS with CERS, we observed that, whereas in the WRS, 62 % of the individuals were assigned to cluster 2, in the CERS sample, only 39 % of the individuals were assigned to this cluster (Figure 2). These results suggest that the WRS gene pool is more similar to that of the Caingang/Guarani group than that from CERS.

Table 3 shows admixture estimates obtained for RS, WRS and CERS data, together with values obtained from other Brazilian regions and Argentina, which shares its East border with the Rio Grande do Sul Western region (WRS). Firstly, our results show that, despite the fact that all Brazilian regions present a predominance of European ancestry, the proportions of each parental contribution is highly variable, with the RS data showing the highest levels of European combined with the smallest levels of African contributions. However, despite the smaller levels of African contribution, in the CERS data the African ancestry levels are slightly higher than

Amerindian, contrasting with the European > Amerindian > African pattern demonstrated in the RS as a whole and, with more intensity, with the WRS results, which showed negligible levels of African contribution. The Structure results taken together with the admixture estimates confirmed the high heterogeneity of the Rio Grande do Sul people ancestry, even when smaller geographic divisions were compared, and are in concordance with the results of Marrero et al., (2005), which detected different ancestry patterns within different regions of the Rio Grande do Sul. Moreover, the results found for the WRS data corroborated the proposed classification of the Southwest of the State as a reservoir of Amerindian genes (Marrero, 2003). In fact, our results indicate a wider geographical amplitude, implicating the whole Western region of the State (SW plus NW) as a source of Amerindian genes.

The higher levels of Amerindian ancestry for the WRS data can be explained by several factors. First of all, historical data indicate the SW as originally populated by the Charruas/Minuanos tribes, that, during the European colonization, suffered a huge population decline. However, these groups mated extensively with the European (mainly Spanish) colonizers of this region (Becker, 2003). For the NW, it was widely populated by the Guarani tribes, who were gathered there by the Jesuit missionaries, that settled several missions throughout this region. Despite the fact that Marrero (2003) had already found considerable mtDNA Amerindian contribution in the SW of Rio Grande do Sul, a remarkable finding was the expressive levels of Amerindian ancestry using autosomal markers. As these markers have high rates of mutation and recombination, which could erode the prevalence of these genes in the European-derived people of that region, our findings indicated that the Amerindian contribution for the WRS people was of greater magnitude than expected. All these information, plus the fact that the majority of our sample is composed by individuals classified as White, lead us to confirm the conclusion of Bortolini et al., (2004), that the process of colonization of the Western region of the State must

have followed the traditional trend observed in the Spanish and Portuguese colonies, characterized by the collapse of Native populations and the introgression of European genes in Native women .

The graphic representation of the Multidimensional Scaling results for the RS, WRS and CERS are depicted in Figures 3A, 3B and 3C, respectively. In general, a clear separation of African, European and Native American populations is observed in all representations. However, despite the fact that the RS data always occupy intermediate positions, CERS (Fig. 3C) is closer to other Brazilian States (Bahia and Rio de Janeiro), while RS and WRS are more related to the Argentinean than to the Brazilian samples (Figs. 3A and 3B, respectively). This closer relationship between RS/WRS data and Argentina seems to be related to a closer geographical distance with that country than with Bahia and Rio de Janeiro States. Moreover, the admixture results (Table 3) indicate that the pattern of parental contribution of RS and WRS samples, represented by high levels of European contribution, followed by considerable levels of Amerindian and smaller levels of African contributions, is more similar to the Argentinian levels of admixture than to those of other Brazilian geographic regions.

In summary, our results demonstrated that, despite the high levels of admixture for the Brazilian populations, there are considerable differences in the ancestry patterns of these populations, resultant of their different demographic histories and geographical localizations, causing genetic structuring processes and genetic differences between populations near the significance level. All these results have considerable implications in forensic exams, association studies and phylogenetic analysis, and point to the importance of population analysis based on smaller geographic divisions.

Acknowledgments

Financial aid was provided by Institutos do Milênio, Programa de Apoio a Núcleos de Excelência (PRONEX), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Secretaria Nacional de Segurança Pública (SENASP), and Instituto Geral de Perícias (IGP) from the State of Rio Grande do Sul.

References

- Alves C, Gusmão L, López-Parra AM, Soledad M, Amorim A, Arroyo-Pardo E. 2005. Allelic frequencies for an African population sample (Equatorial Guinea) using AmpFlSTR Identifiler and Powerplex 16 kits. *For Sci Int* 148:239-242.
- Alves-Silva J, Santos MS, Guimarães PEM, Ferreira ACS, Bandelt HJ, Pena SDJ, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67:444-461
- Becker IB. 2003. Os índios charrua e minuano na antiga banda oriental do Uruguai. São Leopoldo. Editora Unisinos.
- Beleza B, Alves C, Reisc F, Amorim A, Carracedo A, Gusmão L. 2004. 17 STR data (AmpF/STR Identifiler and Powerplex 16 System) from Cabinda (Angola). *For Sci Int* 114:193-196.
- Berardi G, Toscanini U, Raimondi E. 2003. STR data for PowerPlex116 System from Buenos Aires population, Argentina. 2003. *For Sci Int* 134:222–224.
- Bortolini MC, Thomas MG, Chikhi L, Aguilar JA, Castro-De-Guerra D, Salzano FM, Ruiz-Linares A. 2004. Ribeiro's topology, genomes, and Spanish colonialism, as viewed from Gran Canaria and Colombia. *Genetics and Molecular Biology* 1:1-8.

- Callegari-Jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Crosseti SG, Ferreira ME, Hutz MH. 2003. Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. *Am J Hum Biol* 15:824-834.
- Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ. 2001. The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet* 68:281–286.
- Chackraborty R. 1985. Gene identity in racial hybrids and estimation of admixture rates. In: Neel JV, Ahuja Y. *Genetic microdifferentiation in man and other animals*. New Dehli: Indian Anthropological Association. p.171-180.
- Dornelles CL, Callegari-Jacques SM, Robinson WM, Weimer TA, Franco MHLP, Hickmann AC, Geiger CJ, Salzano FM. 1999. Genetic, surnames, grandparents nationalities, and ethnic admixture in southern Brazil-do the patterns of variation coincide?. *Genet Mol Biol* 22:151-161.
- Etcheverry MA. 1949. Grupo sanguineo e fator Rh en los Vascos . *Revista de la Sociedad Argentina de Hematologia e Hemoterapia*, 1, 114-18.
- Garofano L , Pizzamiglio M , Vecchio C, Lago G , Floris T, D'Errico G, Brembilla G , Romano A, Budowle B. 1998. Italian population data on thirteen short tandem repeat loci: HUMTH01, D21S11, D18S51, HUMVWFA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820, D13S317, D5S818, D3S1358. *For Sci Int* 97:53-60.
- Góes ACS, Silva DA, Gil EHF, Silva MTD, Pereira RW, Carvalho EF. 2004. Allele frequencies data and statistic parameters for 16 STR loci—D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX, D3S1358—in the Rio de Janeiro population, Brazil. *For Sci Int* 140:131-132.

- Kohlrausch F, Callegari-Jacques SM, Tsuneto LT, Petzl-Erler ML, Hill K, Hurtado M, Salzano FM, Hutz MR. 2005. Geography Influences Microsatellite Polymorphism Diversity in Amerindians. *Am J Ph Ant* 126:468-470.
- Marino M, Sala A, Corach D. 2006. Genetic analysis of the populations from Northern and Mesopotamian provinces of Argentina by means of 15 autosomal STRs. *For Sci Int* 160(1):184-188.
- Marrero A. 2003. Os gaúchos : sua história evolutiva revelada a partir de marcadores genéticos. Dissertação de Mestrado. Programa de Pós-Graduação em Genética e Biologia Molecular, 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-Junior 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.
- Martinez-Marignac VL, Bertoni B, Parra EJ, Bianchi N. 2004. Characterization of Admixture in an Urban Sample from Buenos Aires, Argentina, Using Uniparentally and Biparentally Inherited Genetic Markers. *Hum Biol* 76(4):543-557.
- Miller SA, Dykes DD, Polesky HF. 1988. A single salting out procedure for extracting DNA from human nucleated cells. *Nucl Ac Res* 16:1215.
- Nei M. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* 19:153-170.
- Paredes M, Crespillo M, Luque JA, Valverde JL. 2003. STR frequencies for the PowerPlex16 System Kit in a population from Northeast Spain. *For Sci Int* 135:75-78.
- Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SJD. 2003. Color and genomic ancestry in Brazilians. *Proc Nat Acad Sci USA* 100:177-182.

- Parra EJ, Kittles RA, Shriver MD. 2004. Implications of correlations between skin color and genetic ancestry for biomedical research. *Nature Genet* 36(11):554-559.
- Pinheiro MF, Caine L, Pontes L, Abrantes D, Lima G, Pereira MJ, Rezende P. 2005. Allele frequencies of sixteen STRs in the population of Northern Portugal. 2005. *For Sci Int* 148:221-223.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 105:767-779.
- Romualdi C, Balding D, Nasidze IS et al. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res* 12:602-612.
- Santos MV, Santos A, Anjos MJ, Andrade L, Vide MC, Corte-Real F, Vieira DN. 2004. Population genetic data for the STR loci using the AmpFlSTR Identifiler Kit in Bahia, Brazil. *International Congress Series* 1261: 219– 222.
- Schneider S, Roessli D, Excoffier L. 2000. Arlequin (ver.2000) – a software for population genetics data analysis. University of Geneva, Geneva.
- Silva DA, Crouse CA, Chakraborty R, Góes ACS, Carvalho EF. 2004. Statistical analyses of 14 short tandem repeat loci in Brazilian populations from Rio de Janeiro and Mato Grosso do Sul states for forensic and identity testing purposes. *For Sci Int* 139:173-176.
- Takezaki N. 2001. <http://www.bio.psu.edu/People/Faculty/Nei/Lab/Programs.html>
- Velosa RG, Fernandes AT, Brehm A. 2002. Genetic profile of the Açores Archipelago population the new PowerPlex16 System Kit. *For Sci Int* 129:68-71.
- Wright SP. 1992. Adjusted P-values for simultaneous inference. *Biometrics* 48:1005-1013

Table 1 Allele frequency distributions for the microsatellites analyzed.

Systems and alleles	RS population	Systems and alleles	RS population
	(n=316)		(n=316)
	Frequency		Frequency
CSF1PO		D7S820	
*7	0.008	*7	0.014
*8	0.017	*8	0.129
*9	0.019	*9	0.107
*10	0.301	*10	0.241
*11	0.285	*11	0.257
*12	0.301	*12	0.216
*13	0.060	*13	0.033
*14	0.005	*14	0.003
*15	0.004		
		D8S1179	
D2S1338		*8	0.016
*14	0.002	*9	0.008
*15	0.002	*10	0.082
*16	0.061	*11	0.090
*17	0.251	*12	0.144
*18	0.075	*13	0.261
*19	0.121	*14	0.239
*20	0.117	*15	0.117
*21	0.034	*16	0.033
*22	0.059	*17	0.006
*23	0.124	*18	0.004
*24	0.071		
*25	0.067	D13S317	
*26	0.011	*8	0.107
*27	0.005	*9	0.094
		*10	0.046
D3S1358		*11	0.307
*11	0.003	*12	0.263
*12	0.002	*13	0.110
*13	0.005	*14	0.073
*14	0.095		
*15	0.287	D16S539	
*16	0.275	*8	0.014
*17	0.187	*9	0.152
*18	0.136	*10	0.089
*19	0.008	*11	0.289
*20	0.002	*12	0.260
		*13	0.180
		*14	0.016

Table 1 (continued)

Systems and alleles	RS population	Systems and alleles	RS population
	(n=316)		(n=316)
	Frequency		Frequency
D18S51		FGA	
*11	0.016	*18	0.009
*12	0.120	*18.2	0.003
*13	0.106	*19	0.081
*14	0.199	*20	0.113
*15	0.163	*21	0.141
*15.2	0.002	*22	0.140
*16	0.120	*22.2	0.005
*17	0.108	*23	0.152
*18	0.068	*24	0.149
*18.2	0.002	*25	0.121
*19	0.063	*26	0.065
*20	0.019	*27	0.013
*21	0.005	*28	0.006
*22	0.006	*31.2	0.002
*22.2	0.002		
*23	0.001	TPOX	
		*5	0.002
		*6	0.013
D19S433		*7	0.005
*9	0.002	*8	0.476
*10	0.002	*9	0.087
*11	0.004	*10	0.048
*11.2	0.002	*11	0.297
*12	0.087	*12	0.070
*12.2	0.013	*13	0.002
*13	0.230		
*13.2	0.033	vWA	
*14	0.336	*13	0.002
*14.2	0.025	*14	0.085
*15	0.172	*15	0.096
*15.2	0.040	*16	0.258
*16	0.037	*17	0.280
*16.2	0.015	*18	0.188
*17.2	0.002	*19	0.076
		*20	0.013
		*21	0.002

Table 1 (continued)

Systems and alleles	RS populations	Systems and alleles	RS population
	(n=316)		(n=316)
	Frequency		Frequency
TH01		D21S11	
*5	0.002	*24.2	0.002
*6	0.233	*25	0.002
*7	0.236	*27	0.037
*8	0.134	*28	0.151
*9	0.129	*29	0.217
*9.3	0.249	*30	0.224
*10	0.017	*30.2	0.031
		*31	0.070
D5S818		*31.2	0.110
*7	0.033	*32	0.006
*8	0.008	*32.2	0.115
*9	0.057	*33	0.003
*10	0.051	*33.2	0.024
*11	0.335	*34.2	0.006
*12	0.338	*35	0.002
*13	0.163		
*14	0.013		
*15	0.002		

Table 2 Pairwise F_{ST} values (above diagonal) and P-values (below diagonal)^a for the comparisons between Brazilian populations^b.

Populations	RS	S	NE	CW	N	SE
RS		0.0013	0.0010	0.0004	0.0009	-0.0008
S	0.0059		0.0005	0.0010	0.0010	-0.0011
NE	0.0293	0.1494		0.0004	0.0001	-0.0021
CW	0.1416	0.0096	0.2100		0.0008	-0.0017
N	0.0098	0.0146	0.4033	0.0166		-0.0010
SE	0.9336	0.9463	0.9990	0.9980	0.9512	

^a Statistically significant threshold after Bonferroni correction: 0.0033

^b S= South, NE= Northeast, CW=Center-West, N=North, SE= Southeast

Table 3 Admixture estimates from different Brazilian and Argentinean populations, and the number (n) and type of markers used in each study.

Country or region	% of parental contribution			Markers (n)
	European	African	Amerindian	
North Brazil ^a	68±0	14±0	18±0	STR (12)
Northeast Brazil ^a	75±10	15±5	10±0	STR (12)
Center-West Brazil ^a	71±0	18±0	11±0	STR (12)
Southeast Brazil ^a	75±0	18±0	7±0	STR (12)
South Brazil ^a	81±1	11±1	8±1	STR (12)
Rio Grande do Sul ^b	82±1	7±1	11±1	Protein (17)
Rio Grande do Sul (RS) ^c	80±1	8±1	12±1	STR (12)
Center/East Rio Grande do Sul (CERS) ^c	81±0	12±0	7±0	STR (12)
West Rio Grande do Sul (WRS) ^c	79±0	2±0	19±0	STR (12)
Argentina ^d	82±10	3±5	15±7	Blood groups/Rh (1)
Argentina (La Plata) ^e	68±3	6±6	26±4	Autosomal SNPs (5)

^a Callegari-Jacques et al (2003), ^b Dornelles et al (1999), ^c Present study, ^d Etcheverry et al (1949), ^e Martinez-Marignac et al (2004)

Legend for Figure 1

Figure 1. Geographical localization of the Brazilian and Argentinean populations used for the comparisons, identified as follows: N (North Brazil), NE(Northeast Brazil), CW (Centerwest Brazil), S (South Brazil), RJ (Rio de Janeiro), GUA (Guarani), CAI (Caingang), MIS (Misiones), SWAR (Southwest Argentina), BUE (Buenos Aires), LP (La Plata)

Legend for Figure 2

Figure 2. Estimates of individual ancestry for individuals of the CERS (1), WRS (2), Guarani (3) and Caingang (4) samples using the Structure program.

Legend for Figure 3

Figure 3. Multidimensional scaling obtained from data of the RS (A), WRS(B) and CERS (C) populations studied, together with data of world populations available in the literature. Each point represents one population, identified as follows: (1a) RS, (1b) WRS, (1c) CERS, (2) Angola, (3) Guinea Bissau, (4) SW Argentina, (5) Açores, (6) North Portugal, (7) Misiones (Argentina), (8) Buenos Aires (Argentina), (9) North Spain, (10) Italia, (11) Rio de Janeiro (Brazil), (12) Bahia (Brazil), (13) Caingang, (14) Guarani.

Figure 1



Figure 2

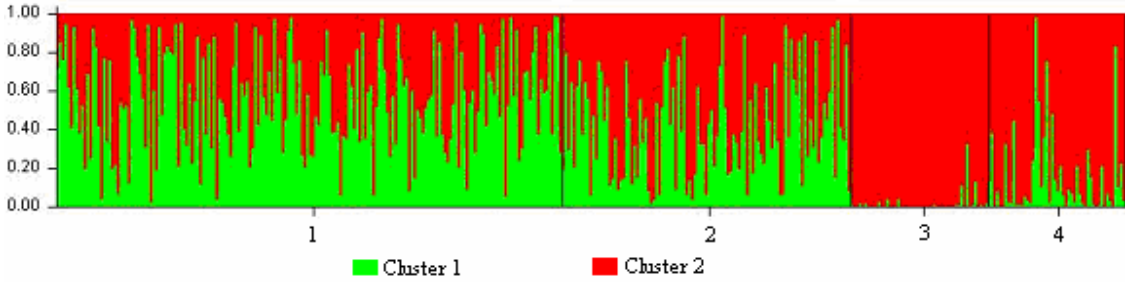
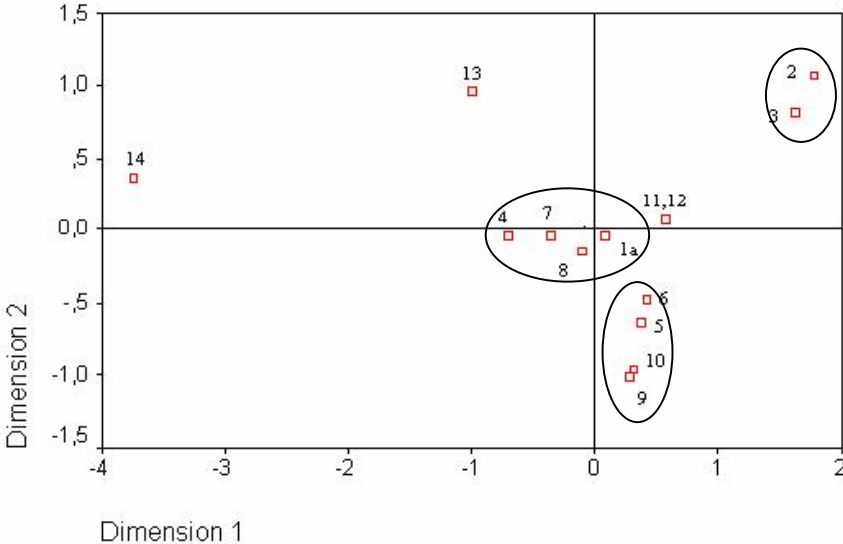
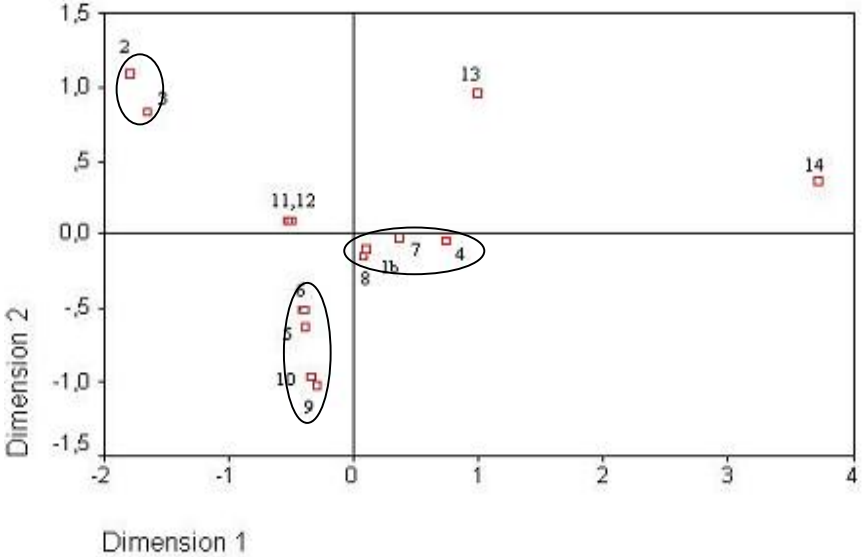


Figure 3

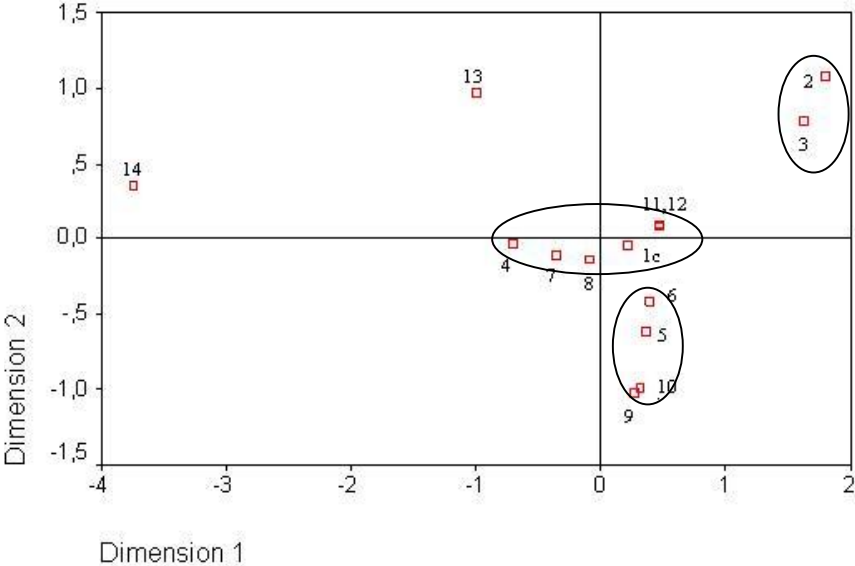


A



B

Figure 3 (continued)



C

Capítulo 4

Y-STR analysis in brazilian and South Amerindian populations

Manuscrito em preparação

Original Research Article

Y-STR analysis in Brazilian and South Amerindian populations

Fabio P.N.Leite,¹ Sidia M. Callegari-Jacques,³ Bianca A.Carvalho,¹ Trícia Kommers,¹ Cecília H. F. Matte,¹ Paulo E. Raimann,¹ Solange P. Schwengber,¹ Vinicius A. Sortica,¹ Luiza T. Tsuneto,⁴ M. Luiza Petzl-Erler,⁴ Francisco M. Salzano,² Mara H. Hutz²

¹Laboratório de Perícias, Instituto Geral de Perícias, Porto Alegre, RS, Brazil

²Departamento de Genética, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970, Porto Alegre, RS, Brazil

³Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

⁴Departamento de Genética, Universidade Federal do Paraná, Curitiba, PR, Brazil

Corresponding author:

Mara H. Hutz
Departamento de Genética, UFRGS
Caixa Postal 15053
91501-970 Porto Alegre, RS, Brazil
Tel: 55 51 3316 6720
Fax: 55 51 3316 9825
E-mail: mara.hutz@ufrgs.br

ABSTRACT

A sample of 203 Brazilian males from Rio Grande do Sul (RS), the Brazilian southernmost state, was typed for 11 Y-STR markers (DYS19, *DYS389I/II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, *DYS385*, *DYS437*, *DYS438*, *DYS439*). We also typed 42 individuals from two South Amerindian tribes (Kaingang and Guarani) to use the data as parental Amerindian contribution to our analyses. Gene and haplotypic diversities were estimated, with the South Amerindian samples showing smaller values for these parameters than Brazilians. To obtain a more comprehensive picture of the genetic structure of the Brazilian population as a whole, the Y-STR data from the RS sample was compared with those already published. No genetic substructuring was observed in the comparisons performed. Multidimensional Scaling confirmed the proposed European source of the majority of the majority of the Y-chromosome Brazilian patrilineages .

The Y-STR markers from the Male Specific region of the Y chromosome (MSY) have become an important tool in the study of human evolution, forensics and population genetics. The peculiar characteristics of the Y-chromosome markers, such as paternal inheritance and haploidy, make it possible to infer about population histories of paternal lineages and the contribution of these lineages to the gene pool of the extant populations (Kayser et al., 1997, Jobling and Tyler-Smith, 2003).

Due to their mode of inheritance and lack of recombination, these markers have been successfully used in forensic cases involving mixed samples, such as rape cases and paternity exams (Jobling et al., 1997). In these cases, autosomal STR markers sometimes fail to identify the male profile, and the Y-STR analysis can improve the probability to detect small amounts of the perpetrator DNA in a high background of female DNA. Another characteristic of the Y-STR markers is the pronounced population structure, which is determined by its small effective population size and patrilocality (Jobling and Tyler-Smith 2003). This pattern has important implications for the development of DNA database in forensics, and also for the study of diseases, since there are evidence demonstrating geographic variation in disease patterns within ethnic groups (Jackson, 2000) . Although there is some controversy whether Y-STR can define the population source of individuals and groups, due to a high mutation rate and the occurrence of homoplasy, which allow haplotypes be shared by identity by state (IBS) without being identical by descent, several studies have demonstrated the utility of these markers to trace the history of populations and to assign individuals within populations (Thomas et al., 2000, Jobling, 2001, Kayser et al., 2001).

The Brazilian population was formed by successive migratory waves. The Amerindian people were occupying the Brazilian territory when the Portuguese arrived in 1500. After that,

Africans were brought to Brazil as slaves, between the 15th and the 19th century. Besides the Portuguese, other migratory waves followed, mainly from Italy, Germany and Spain, during the 19th and the 20th centuries (Salzano and Bortolini 2002). All these migratory events have contributed to the formation of a highly multiethnic admixed population. This heterogeneity was shown in several genetic studies, using either uniparental (Alves-Silva et al., 2000; Carvalho-Silva et al., 2001) or autosomal markers (Callegari-Jacques et al., 2003), which demonstrated a typical triethnic (European+African+Amerindian) pattern for the Brazilian population gene pool. Nevertheless, Callegari-Jacques et al., (2003) using STRs, and Parra et al., (2003) using biallelic markers, showed that the contributions of these parental populations were not uniform along the country. Southern Brazilian populations presented smaller levels of African and higher degrees of European contributions when compared to other Brazilian groups. However, few studies of Y-STR were performed in Brazil (Vallinoto et al., 1999; Góes et al., 2005, Pereira et al., 2006) if the continental size and ethnic diversity of the country is considered.

The present study investigates the variability of 11 Y-STR markers in the admixed population of Rio Grande do Sul, the Brazilian southernmost state, and also in pooled samples from south Amerindian tribes. To obtain a more comprehensive picture of the genetic structure of the Brazilian population as a whole, the Y-STR data obtained herein was compared with those already published for other regions of this country. We also performed database searches, Multidimensional Scaling (MDS) and admixture analyses to consider the proposed European source of the majority of the Y-chromosome Brazilian patrilineages (Carvalho-Silva et al., 2001).

MATERIAL AND METHODS

Population Samples

A sample composed of 203 Brazilian individuals from the Brazilian state of Rio Grande do Sul (RS), and 42 individuals from South Amerindian (AM) tribes (22 Kaingang and 20 Guarani) was investigated. The first group will be referred as RS whereas the South Amerindian group will be designated as AM. The Kaingang and Guarani are the two major Amerindian populations living in southern Brazil, and speak languages of the Gê and Tupi stocks, respectively. Despite living side by side for many centuries, these groups still differ in many aspects of their culture and biology (Petzl-Erler et al., 1993). Both groups are in an advanced stage of acculturation. All individuals enrolled in the present investigation were informed about the aims of this study and signed a written consent form.

Laboratory methods

Genomic DNA was extracted from blood samples using the salting out method according to Miller et al., (1988). A total of 11 STR loci (DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, and DYS385) were studied using the PowerplexY-System (Promega). Amplifications were performed following the instructions provided in the kit user manual, with the recommended DNA amount (1.0-2.0 ng). Electrophoresis of the amplified fragments was performed in a 310 ABI PRISMTM Genetic Analyzer (Applied Biosystems, Foster City, CA), also according to the manufacturer's recommended protocol. The GeneScanTM v3.2.1 analysis software was used to track lanes and

measure fragment sizes, while the GenotyperTM v2.5.2 software was employed to automatically designate alleles by comparison with locus specific allelic ladders.

Statistical analysis

Allele frequencies were obtained with the Arlequin software (Schneider et al., 2000). Mean Gene Diversity (D), and Haplotype Diversity (HD) values were calculated according to Nei (1987). Heterogeneity between populations was tested comparing our RS data with information from samples collected in five Brazilian regions, namely central-west (CW), southeast (SE), northeast (NE) and north (N) (Pereira et al., 2006), as well as in Rio de Janeiro state (Góes et al., 2005), which is also located in the southeast region of the country. For this purpose, pairwise F_{ST} and R_{ST} values were estimated, their statistical significance being assessed using 1000 permutations (Schneider et al., 2000) and corrected using the classical Bonferroni method (Wright 1992). All analyses were conducted with the Arlequin software (Schneider et al., 2000). Estimates of parental contributions were obtained using Long's (1991) weighted least squares method, calculated with the ADMIX program, kindly provided by the author. Allele frequencies for the parental populations were based on Rosa et al., (2006), and Arroyo-Pardo et al., (2005) for the African (Guinea-Bissau and Guinea Equatorial), and Biondo et al., (2004), Martin et al., (2004) and Gusmão et al., (2002), for European (Italy, Spain and Portugal) populations. The Amerindian parental allele frequencies used in this analysis were obtained by typing the AM samples of the present study. Population relationships on the basis of R_{ST} values were determined by use of Multidimensional Scaling (MDS) analysis, as implemented in SPSS 8.0, using the RS data, plus those used for the admixture analyses added to data from Rio de Janeiro (Góes et al., 2005). In the population comparisons DYS385 was not considered, and the number of repeats in DYS389II was subtracted from DYS389I.

RESULTS AND DISCUSSION

Haplotypic distributions for the Rio Grande do Sul (RS) and Amerindian (AM) samples are presented in Table 1. From a total of 197 different haplotypes, the RS and AM samples shared only two: H24 and H46. As expected (larger and ethnically more diversified sample), haplotypic and mean gene diversities values were higher for RS (0.999 and 0.632, respectively) than for AM (0.949 and 0.529, respectively), resulting in a combined power of individual discrimination of 84 % for RS and 61 % for the AM samples. Allele and haplotype frequencies for the RS and AM samples are shown in Tables 2 and 3, respectively.

Population comparisons

Pairwise F_{ST} values between the populations compared are displayed in Table 4. For RS in general, no significant values (after the Bonferroni correction) were found in the comparisons with other Brazilian data. These results agree with those of previous investigations (Carvalho-Silva et al., 2001, Pereira et al., 2006), in which the absence of significant differences between Brazilian populations was also observed. The R_{ST} analysis yielded similar results (data not shown).

Admixture values, genetic relationship and database search

Admixture estimates obtained for RS (0.920±0.055 for European, 0.052±0.039 for African, and 0.028±0.050 for Native American contributions) indicate that, despite the high percentage of European participation, there is a non-negligible African and a residual Amerindian

contribution to this population as well. Therefore, our results differ from those of Carvalho-Silva et al., (2001), Marrero et al., (2005), who found only Y chromosomes classifiable as European in 52 southern Brazilians and 75 inhabitants of Rio Grande do Sul, respectively. The African contribution detected in our sample is probably due to the presence of four haplotypes (H19, H42, H166 and H168) that have the Bantu modal haplotype (DYS19*15; DYS390*21; DYS391*10; DYS392*11; DYS393*13) described by Thomas et al., (2000), and four with one-step neighbor arrangements (H70, H90, H96 and H117). These so-called “Bantu haplotypes” were searched against two Y-STR databases, namely the Y-hrd database (<http://www.yhrd.org>), comprising 9,632 haplotypes, and the Promega database (<http://www.promega.com/techserv/tools/pplexy/>), containing 4,004 haplotypes. The matches observed for these haplotypes are displayed in Table 5. None of the eight haplotypes were found in individuals of European populations. H19, H42, H90, H96 and H117 were observed in African or African-American individuals, while H70, H166 and H168 showed no match in both databases. The Bantu contribution to the African Brazilian gene pool was already known from beta-globin haplotype investigations (Salzano and Bortolini 2002). The Amerindian contribution can be represented by the presence of the haplotype H46, which was shared between RS and AM populations, and constitute the modal Amerindian haplotype (DYS19*13; DYS390*24; DYS391*10; DYS392*14; DYS393*13) of the Q-M242* haplogroup, according to Bortolini et al., (2003).

The MDS plot is depicted in Figure 1. As expected, the RS and RJ populations were grouped with Portugal and Spain, while the South Amerindians remained isolated and Guinéa Bissau and Guinéa Equatorial formed a third cluster. These results agree with historical data, since the early European colonization of Rio Grande do Sul state was mainly of Spanish and Portuguese individuals (Marrero et al., 2005). In addition, we observed a large predominance of

matches with European samples in the database searches done for the most frequent haplotypes observed in our RS data (data not shown).

In summary, our results showed firstly, the lack of substructuring among several Brazilian regions in relation to the Y-chromosome. Secondly, we observed that the majority (but not all) of the Brazilian Y-chromosomes studied here are of European origin, which must have obvious consequences regarding the absence of significant differences for most of the populations analyzed. Thus, despite the pronounced population structure of the Y chromosome, it seems that the homogenization of the Brazilian patrilineages have neutralized possible substructuring effects for most of the Brazilian populations, as proposed by Carvalho-Silva et al., (2006). All these results have considerable implications for forensic exams, since they point out that is unnecessary to incorporate geographic structuring into DNA databases based on Y-STRs, at least for most of the admixed populations. Moreover, the absence of population stratification and the homogeneity of Y-chromosome ancestry in Brazil suggest that variation in susceptibility for diseases related to this chromosomes are not associated to genetic differences in the Brazilian populations. However, further studies, based on a higher number of markers using samples from smaller sociogeographic regions, need to be performed to better clarify this question.

ACKNOWLEDGMENTS

Financial aid was provided by Institutos do Milênio, Programa de Apoio a Núcleos de Excelência (PRONEX), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Secretaria Nacional de Segurança Pública (SENASP), and Instituto Geral de Perícias (IGP) from the State of Rio Grande do Sul. The logistic support of the Fundação Nacional do Índio in the Amerindian studies is gratefully acknowledged. This investigation was approved by the Brazilian National Ethics Commission (CONEP Resolution no. 123/98).

LITERATURE CITED

- Alves-Silva J, Santos MS, Guimarães PEM, Ferreira ACS, Bandelt HJ, Pena SDJ, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. *Am. J. Hum Genet* 67:444-461.
- Arroyo-Pardo EL, Lopez-Parra AM, Gusmão L, Baeza C, Mesac MS, Amorim A. 2005. Genetic variability of 16 Y-chromosome STRs in a sample from Equatorial Guinea (Central Africa). *Forens Sci Int* 149: 109–113.
- Biondo R, Caglia A, D'Agostaro G, D'Agostini G, Spinella A. 2004. Typing of 20 Y-chromosome STRs in the Italian population. *Forens Sci Int* 146: S135-S138.
- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73:524–539.
- Callegari-Jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Crosseti SG, Ferreira ME, Hutz MH (2003) Historical genetics: spatiotemporal analysis of the formation of the Brazilian population. *Am J Hum Biol* 15:824-834
- Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ (2001) The phylogeography of Brazilian Y-chromosome lineages. *Am. J. Hum. Genet.* 68: 281–286.
- Carvalho-Silva DR, Tarazona-Santos E, Rocha J, Pena SDJ, Santos FR (2006) Y chromosome diversity in Brazilians: Switching perspectives from slow to fast evolving markers. *Genetica* 126:251-260.

- Góes ACS, Carvalho EF, Gomes I, Silva DA, Gil EHF, Amorim A, Gusmão L. 2005. Population and mutation analysis of 17 Y-STR loci from Rio de Janeiro (Brazil). *Int. J. Legal Med.* 119: 70–76.
- Gusmão, L., C. Alves, S. Beleza et al. 2002. Forensic evaluation and population data on the new Y-STRs DYS434, DYS437, DYS438, DYS439 and GATA A10. *Int J Legal Med* 116 : 139–147.
- Jackson F. 2000. Anthropological measurement: The mismeasure of African Americans. *Ann. Am Acad Pol Soc Sci* 568: 154–171.
- Jobling MA, Pandya A, Tyler-Smith C. 1997. The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 110:118–124.
- Jobling, MA. 2001. In the name of the father: surnames and genetics. *Trends Genet* 17, 353–357.
- Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nature Rev Genet* 4: 598-612.
- Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, K, Jobling MA, Krawczak M, Leim K, Meuser S, Meyer Oesterreich W, Pandya A, Parson W, Penacino G, Lezaun A, Piccinini A, Prinz M, Schmitt C, Roewer L P. 1997. Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* 18:1602-1607.
- Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, de Knijff P. 2001. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am. J. Hum. Genet.* 68:990–1018.
- 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 Júnior 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.
- Martin P, Garcia-Hirschfeld J, Garcia O, Gusmão L, Garcia P, Albarrana C, Sancho M, Alonso A, Martin P, Garcia-Hirschfeld J, Garcia O. 2004. A Spanish population study of 17 Y-chromosome STR loci. *Forens Sci Int* 139:231-235.
- Miller SA., Dykes DD, Polesky HF. 1988. A single salting out procedure for extracting DNA from human nucleated cells. *Nucl Ac Res* 16:1215.
- Nei M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York
- Parra FC, Amado RC, Lambertucci JR, Pena SDJ. 2003. Color and genomic ancestry in Brazilians. *Proc Nat Acad Sci USA* 100:177-182
- Pereira RW, Monteiro EHJ, Hirschfeld GCR. 2006. Haplotype diversity of 17 Y-chromosome STRs in Brazilians. *For. Sci. Int.* (in press)
- 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.
- Rosa A, Ornelas A, Brehm A, Villems R. 2006. Population data on 11 Y-chromosome STRs from Guiné-Bissau. *Forens Sci Int* 157: 210-217.
- Salzano FM, Bortolini MC. 2002. *The evolution and genetics of Latin American populations*. Cambridge University Press, Cambridge.
- Schneider S, Roessli D, Excoffier L. 2000. *Arlequin (ver.2000) – a software for population genetics data analysis*. University of Geneva, Geneva.

- Thomas MG, Parfitt T, Weiss DA. 2000. Y chromosomes traveling south: the Cohen modal haplotype and the origins of the Lemba—the “Black Jews of Southern Africa”. *Am. J. Hum. Genet.* 66:674–686.
- Vallinoto AC, Cayres-Vallinoto IM, Ribeiro dos Santos AKA, Zago MA, Santos SE, Guerreiro JF. 1999. Heterogeneity of Y chromosome markers among Brazilian Amerindians. *Am J Hum Biol* 11:481–487
- Wright SP. 1992. Adjusted P-values for simultaneous inference. *Biometrics* 48:1005-1013

Table 1. Y chromosome haplotype distribution in 203 Brazilian individuals from Rio Grande do Sul (RS), and 42 individuals from South Amerindian tribes (AM).

Haplotype	DYS 391	DYS 389I	DYS 439	DYS 389II	DYS 438	DYS 437	DYS 19	DYS 392	DYS 393	DYS 390	DYS 385	RS	AM
H1	11	13	12	28	12	13	15	13	13	24	12,14	1	
H2	12	13	12	29	12	15	14	14	13	24	11,14	1	
H3	11	13	12	30	12	15	15	13	14	24	11,14	1	
H4	10	13	11	30	10	14	13	11	13	23	16,16	1	
H5	10	12	11	28	9	15	16	11	12	24	13,17	1	
H6	10	12	12	29	10	16	15	11	14	22	12,16	1	
H7	10	14	10	32	9	15	14	11	12	23	15,15	1	
H8	11	13	12	29	12	14	14	13	13	24	11,14	4	
H9	10	13	12	29	12	15	15	13	13	24	10,14	1	
H10	10	12	11	28	10	16	14	11	13	22	13,14	2	
H11	11	13	12	29	12	15	14	13	13	24	11,14	5	
H12	9	14	10	30	10	14	13	11	13	24	13,14	7	
H13	11	13	11	29	12	15	14	13	13	24	10,15	1	
H14	10	13	11	29	12	15	14	13	13	24	12,14	1	
H15	10	13	11	28	10	15	16	11	13	23	11,12	1	
H16	10	14	11	30	12	14	14	13	12	24	11,14	1	
H17	11	13	13	29	12	15	15	13	13	24	12,14	1	
H18	9	14	10	30	10	14	14	11	13	24	13,14	1	
H19	10	13	12	31	11	14	15	11	13	21	15,15	1	
H20	11	12	12	28	10	16	16	11	14	21	15,16	2	
H21	10	14	12	31	12	11	14	13	13	24	12,14	1	
H22	11	13	12	29	12	15	14	13	13	24	11,15	2	
H23	11	13	12	29	12	15	14	13	13	24	11,16	1	
H24	11	13	12	29	12	15	15	13	13	24	11,14	3	1
H25	10	13	13	29	12	15	15	13	13	26	11,13	1	
H26	10	12	11	29	10	16	15	11	14	23	13,13	1	
H27	10	12	11	30	10	16	15	11	15	23	13,14	1	
H28	10	13	11	31	10	14	16	11	14	24	11,13	1	
H29	11	13	13	29	13	15	14	13	13	24	11,14	1	
H30	11	14	12	30	11	15	14	13	13	26	11,14	1	
H31	11	13	12	29	12	15	14	13	13	24	10,14	1	
H32	10	13	13	29	12	15	14	13	13	24	11,13	1	
H33	10	12	11	29	9	15	14	11	12	23	13,17	1	
H34	10	14	12	30	12	15	14	13	13	24	11,14	1	
H35	11	13	11	29	10	15	16	12	14	23	13,13	1	
H36	11	12	11	28	10	15	16	12	14	23	13,13	1	
H37	9	15	10	31	10	14	14	11	12	20	8,9	1	

Table 1. (cont)

Haplotype	DYS 391	DYS 389I	DYS 439	DYS 389II	DYS 438	DYS 437	DYS 19	DYS 392	DYS 393	DYS 390	DYS 385	RS	AM
H38	9	13	11	29	10	14	13	11	13	24	13,14	3	
H39	10	13	11	30	10	14	15	12	14	23	15,16	1	
H40	11	13	12	29	12	15	14	11	13	24	11,11	1	
H41	10	13	10	28	9	15	14	11	12	23	13,16	1	
H42	10	13	13	31	12	14	15	11	13	21	16,16	1	
H43	10	13	11	29	12	15	14	13	13	23	11,14	2	
H44	9	13	10	29	10	14	13	11	13	24	13,14	2	
H45	11	13	12	29	13	15	14	11	13	24	10,15	1	
H46	10	13	11	30	11	14	13	14	13	24	14,16	1	
H47	11	12	11	28	11	14	14	11	14	26	14,20	1	
H48	11	12	13	28	12	14	15	14	13	23	11,14	1	
H49	10	13	12	30	10	14	13	11	13	24	16,18	1	
H50	11	13	13	29	12	15	15	13	13	24	11,14	1	
H51	11	13	12	29	12	15	14	13	13	24	12,16	1	
H52	12	13	11	29	12	15	15	13	13	24	11,15	1	
H53	11	13	12	30	12	15	14	13	14	23	11,14	1	
H54	10	14	11	29	9	15	13	13	13	23	13,16	1	
H55	10	12	11	29	10	16	15	11	13	23	13,14	1	
H56	11	13	14	30	12	14	14	13	14	24	10,14	1	
H57	11	13	13	29	12	14	14	13	13	23	12,14	1	
H58	11	13	12	30	12	15	14	13	13	26	11,14	1	
H59	11	13	11	29	12	15	14	13	13	25	11,14	2	
H60	10	13	12	29	11	13	14	14	13	24	12,14	1	
H61	10	13	13	30	10	14	13	11	13	24	16,18	1	
H62	11	13	12	30	12	15	15	13	14	24	11,15	1	
H63	10	13	12	29	12	15	14	13	13	24	11,14	1	
H64	10	14	13	31	11	14	17	10	15	21	17,19	1	
H65	11	13	11	29	9	15	17	10	12	23	12,16	1	
H66	10	12	11	28	12	15	14	12	13	24	11,14	1	
H67	11	13	12	31	11	13	15	10	13	21	16,17	1	
H68	10	12	12	28	12	16	14	12	13	24	11,13	1	
H69	11	13	12	29	12	15	15	12	13	24	11,14	1	
H70	10	12	11	30	11	14	15	10	13	21	17,17	1	
H71	10	12	12	28	10	16	15	11	13	23	12,13	1	
H72	11	13	11	29	12	15	14	13	13	23	11,14	3	
H73	10	14	12	30	9	15	15	11	11	24	17,18	1	
H74	11	13	12	29	12	15	15	13	13	26	11,14	1	
H75	11	12	12	28	12	14	14	13	13	24	11,15	1	
H76	10	13	12	29	10	14	13	11	13	24	17,18	1	
H77	11	14	12	29	12	15	14	14	12	24	13,14	1	

Table 1. (cont)

Haplotype	DYS 391	DYS 389I	DYS 439	DYS 389II	DYS 438	DYS 437	DYS 19	DYS 392	DYS 393	DYS 390	DYS 385	RS	AM
H78	10	13	12	29	12	16	14	13	13	23	11,13	2	
H79	11	13	13	30	12	14	14	13	13	24	12,14	1	
H80	11	13	13	29	13	15	14	13	12	24	11,12	1	
H81	11	13	12	29	12	15	15	13	12	24	11,14	1	
H82	10	13	11	29	12	15	14	11	14	25	12,14	1	
H83	10	12	11	29	10	14	14	11	13	23	13,18	1	
H84	11	13	13	30	12	15	14	13	12	24	11,14	1	
H85	12	13	11	29	13	15	14	13	13	24	11,14	1	
H86	11	13	12	29	12	15	15	13	13	25	11,14	1	
H87	10	13	11	29	9	14	14	13	14	23	13,15	1	
H88	11	12	14	28	12	15	14	13	13	24	11,14	1	
H89	10	14	14	31	10	16	15	11	14	22	13,15	1	
H90	11	13	11	31	11	14	15	11	13	21	16,18	1	
H91	11	14	12	30	12	14	14	13	13	25	11,14	1	
H92	11	13	13	29	12	14	14	13	13	25	11,14	1	
H93	10	13	13	29	12	15	13	13	13	24	11,15	1	
H94	9	13	13	29	9	14	15	11	12	24	13,17	1	
H95	11	14	13	31	10	16	15	11	14	22	14,14	1	
H96	11	13	11	30	11	14	15	11	13	21	16,16	1	
H97	10	13	11	30	10	14	13	11	13	24	16,19	1	
H98	10	13	11	30	11	14	16	11	13	26	11,13	1	
H99	11	13	11	32	12	14	14	13	13	24	11,14	1	
H100	10	13	13	29	12	16	14	13	13	23	11,13	3	
H101	10	15	12	31	12	14	14	13	13	24	11,15	1	
H102	10	13	13	29	9	15	14	11	13	24	13,17	1	
H103	11	13	12	29	11	15	14	13	12	24	11,14	1	
H104	11	13	11	29	12	15	14	13	13	24	11,14	2	
H105	11	14	12	30	12	15	14	13	12	23	11,14	1	
H106	11	13	12	29	12	14	15	13	13	23	11,14	1	
H107	10	13	13	30	12	15	14	13	13	24	11,14	1	
H108	9	14	11	30	10	14	13	11	13	24	13,14	1	
H109	11	12	11	29	8	17	15	11	13	22	15,16	1	
H110	10	12	10	30	11	14	16	11	14	25	11,14	1	

Table 1. (cont)

Haplotype	DYS 391	DYS 389I	DYS 439	DYS 389II	DYS 438	DYS 437	DYS 19	DYS 392	DYS 393	DYS 390	DYS 385	RS	AM
H111	10	12	12	30	10	15	14	11	13	23	10,14	1	
H112	10	13	11	29	10	15	14	11	13	22	13,15	1	
H113	11	13	13	31	12	15	14	13	13	24	11,14	1	
H114	10	12	12	28	10	16	14	11	13	22	12,14	1	
H115	10	12	10	29	11	14	16	11	14	25	11,14	1	
H116	11	13	13	29	12	15	15	13	13	23	11,13	1	
H117	10	13	12	30	11	14	15	11	14	21	15,19	1	
H118	10	13	13	30	10	14	14	11	13	22	18,21	1	
H119	10	13	13	29	13	15	14	13	13	23	11,15	1	
H120	10	13	13	29	12	14	15	13	13	24	10,14	1	
H121	11	12	11	28	10	16	14	11	13	22	13,14	1	
H122	10	13	11	31	10	14	15	12	14	23	15,16	1	
H123	10	13	10	29	11	14	17	11	13	25	10,14	1	
H124	11	14	12	30	12	14	14	13	13	24	11,14	1	
H125	11	13	13	29	12	14	14	13	13	24	11,14	1	
H126	10	12	12	29	9	16	15	11	12	24	14,18	1	
H127	11	13	11	29	9	15	14	11	13	23	12,17	1	
H128	11	13	12	28	12	15	16	13	13	24	11,14	2	
H129	10	13	11	30	10	16	15	11	13	22	15,15	1	
H130	9	14	10	30	10	15	13	11	13	24	13,15	1	
H131	10	12	11	28	12	11	14	13	13	23	11,13	1	
H132	11	13	12	30	12	14	14	13	13	23	11,16	1	
H133	10	13	11	29	10	14	14	10	12	23	13,13	1	
H134	10	13	12	30	9	14	14	11	13	23	14,15	1	
H135	11	14	11	31	10	14	15	12	15	23	15,15	2	
H136	11	13	10	31	11	14	15	11	13	25	11,14	1	
H137	11	13	11	30	11	14	15	11	14	21	17,17	1	
H138	10	12	11	28	10	16	14	11	13	23	13,14	1	
H139	11	14	12	30	12	14	14	13	13	24	11,15	1	
H140	11	13	11	29	12	15	14	12	11	23	11,14	1	
H141	11	14	10	32	10	16	15	10	12	21	13,14	1	
H142	9	14	9	30	10	14	13	10	11	23	12,14	1	
H143	10	12	11	28	10	16	15	11	13	23	13,17	2	

Table 1. (cont)

Haplotype	DYS 391	DYS 389I	DYS 439	DYS 389II	DYS 438	DYS 437	DYS 19	DYS 392	DYS 393	DYS 390	DYS 385	RS	AM
H144	10	12	12	30	10	14	13	10	11	23	18,18	1	
H145	10	13	11	29	9	15	14	11	12	26	13,13	1	
H146	10	14	10	31	10	14	16	11	14	25	10,14	1	
H147	10	11	11	28	12	15	14	14	11	23	11,13	1	
H148	11	13	12	29	12	15	15	13	12	22	11,14	1	
H149	11	12	12	30	10	16	14	11	13	23	14,14	1	
H150	10	14	11	32	9	14	15	11	13	24	12,14	1	
H151	11	13	12	30	12	15	14	13	13	24	11,14	1	
H152	11	13	11	29	9	15	15	11	12	22	12,16	1	
H153	10	13	12	30	12	15	14	13	13	24	11,14	1	
H154	11	13	11	30	12	15	15	14	13	24	11,14	1	
H155	10	13	12	30	11	14	16	11	14	21	14,20	1	
H156	10	12	12	28	12	14	14	13	13	24	11,14	1	
H157	11	13	12	30	12	15	14	13	13	23	11,14	1	
H158	10	14	12	30	12	15	14	13	13	25	11,14	1	
H159	11	12	11	29	12	15	15	13	13	24	10,11	1	
H160	10	14	12	30	12	15	14	14	13	25	11,14	1	
H161	10	12	12	29	9	14	15	11	13	23	13,15	1	
H162	11	12	12	28	12	15	14	13	13	24	11,14	1	
H163	11	13	11	28	12	15	14	13	13	23	11,15	1	
H164	11	12	12	27	12	15	14	13	13	24	11,14	1	
H165	11	13	11	30	9	13	16	13	13	23	15,15	1	
H166	10	13	12	32	11	14	15	11	13	21	14,14	1	
H167	9	13	11	29	9	14	15	11	12	23	13,16	1	
H168	10	13	12	32	11	14	15	11	13	21	16,18	1	
H169	10	13	12	29	13	15	14	13	13	24	11,13	1	
H170	11	14	11	30	12	15	13	13	13	23	11,14	1	
H171	11	13	12	29	12	15	14	14	13	24	11,14	1	
H172	10	12	13	29	11	14	13	15	12	24	14,14		4
H173	11	13	12	29	12	15	15	13	13	24	11,14		1
H174	10	13	12	28	12	15	14	13	13	24	11,14		2
H175	10	13	12	29	12	15	14	13	13	24	12,14		1

Table 1. (cont)

Haplotype	DYS 391	DYS 389I	DYS 439	DYS 389II	DYS 438	DYS 437	DYS 19	DYS 392	DYS 393	DYS 390	DYS 385	RS	AM
H176	11	13	12	30	11	14	15	14	14	24	14,17	3	
H177	11	13	12	29	11	14	15	14	14	24	14,17	1	
H178	11	13	12	29	12	15	14	13	13	24	11,11	3	
H179	10	13	11	31	11	14	13	14	14	24	12,15	1	
H180	11	13	12	30	12	15	14	14	13	24	11,14	1	
H181	10	13	11	30	11	14	13	14	13	24	14,16	3	
H182	10	13	12	29	12	15	15	13	13	23	11,14	2	
H183	10	12	12	28	12	15	14	13	13	24	12,14	1	
H184	10	13	14	30	11	14	16	11	14	21	15,19	1	
H185	11	12	10	26	12	15	15	13	13	24	12,14	1	
H186	10	12	12	28	10	16	14	11	13	23	13,15	1	
H187	10	12	12	29	11	14	14	14	13	24	13,13	1	
H188	11	13	12	29	12	15	14	13	13	24	14,14	2	
H189	10	12	12	29	11	14	14	14	13	23	13,17	3	
H190	10	12	12	30	11	14	13	14	13	24	13,17	1	
H191	10	12	12	30	11	14	13	14	13	24	14,17	3	
H192	10	12	13	28	12	14	13	14	12	24	14,14	1	
H193	10	12	11	30	11	14	13	14	13	24	14,17	1	
H194	10	12	12	29	11	14	14	14	13	23	13,16	1	
H195	11	12	11	29	11	14	16	11	13	21	16,17	1	
H196	10	12	13	30	11	14	13	15	12	24	14,14	1	
H197	10	12	12	28	10	16	13	11	13	23	14,15	1	

Table 2. Allele frequencies and DYS385I/II haplotype frequencies for the Rio Grande do Sul (RS) sample (N = 203)

Allele	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 437	DYS 438	DYS 439	DYS385 Haplotype	
8									0.004		11, 11	0.005
9					0.094				0.094	0.005	15, 16	0.025
10					0.424	0.039			0.276	0.099	12, 14	0.054
11		0.005			0.468	0.394	0.025	0.010	0.103	0.345	11, 14	0.335
12		0.187			0.014	0.049	0.103		0.493	0.398	16, 16	0.015
13	0.123	0.631				0.474	0.734	0.020	0.030	0.138	13, 17	0.030
14	0.512	0.167				0.044	0.118	0.374		0.015	12, 16	0.020
15	0.276	0.010					0.020	0.463			15, 15	0.030
16	0.074							0.128			10, 14	0.034
17	0.015							0.005			13, 14	0.108
20				0.005							10, 15	0.010
21				0.074							11, 15	0.049
22				0.064							11, 16	0.010
23				0.276							11, 13	0.069
24				0.483							13, 15	0.025
25				0.064							16, 18	0.020
26				0.034							13, 13	0.025
27			0.005								8, 9	0.005
28			0.138								13, 16	0.011
29			0.467								14, 16	0.005
30			0.276								17, 19	0.005
31			0.084								16, 17	0.005
32			0.030								17, 17	0.010
											12, 13	0.005
											17, 18	0.010
											13, 18	0.005
											14, 14	0.015
											11,12	0.010
											14, 20	0.010
											16, 19	0.005
											15, 19	0.005
											18, 18	0.005
											10, 11	0.005
											18,21	0.005
											14,18	0.005
											12, 17	0.005
											14, 15	0.005

Table 3. Allele frequencies and DYS385I/II haplotype frequencies for the South Amerindian (AM) group (n=42)

Allele	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 437	DYS 438	DYS 439	DYS385 Haplotype	
10					0.690				0.048	0.024	14, 14	0.190
11					0.310	0.095			0.595	0.143	11, 14	0.143
12		0.500					0.143		0.357	0.667	12, 14	0.071
13	0.381	0.500				0.310	0.714			0.143	14, 17	0.190
14	0.381					0.476	0.143	0.619		0.024	11, 11	0.071
15	0.190					0.119		0.333			12, 15	0.025
16	0.048							0.048			14, 16	0.071
21				0.048							15, 19	0.024
23				0.190							13, 15	0.024
24				0.762							13,17	0.095
26			0.024								13,13	0.024
28			0.143								13,16	0.024
29			0.476								16, 17	0.024
30			0.333								14,15	0.024
31			0.024									

Table 4. Pairwise F_{ST} values (above diagonal) and P-values (below diagonal) for the comparisons between Brazilian populations

Populations	RS	CW	N	NE	S	SE	RJ
RS		0.00007	0.00079	0.00065	0.00204	-0.00062	0.00037
CW	0.373		-0.00007	-0.00021	0.00149	-0.00089	0.00026
N	0.107	0.486		-0.00068	-0.00009	-0.00105	0.00042
NE	0.198	0.504	0.725		0.00122	-0.00102	0.00065
S	0.005	0.036	0.508	0.106		0.00078	0.00139
SE	0.736	0.815	0.853	0.805	0.213		-0.00049
RJ	0.208	0.272	0.248	0.206	0.025	0.702	

Statistically significant threshold after Bonferroni correction: 0.0024

RS= Rio Grande do Sul state, CW=Central-West region, N=North region, S=South region, SE=Southeast region, RJ=Rio de Janeiro state.

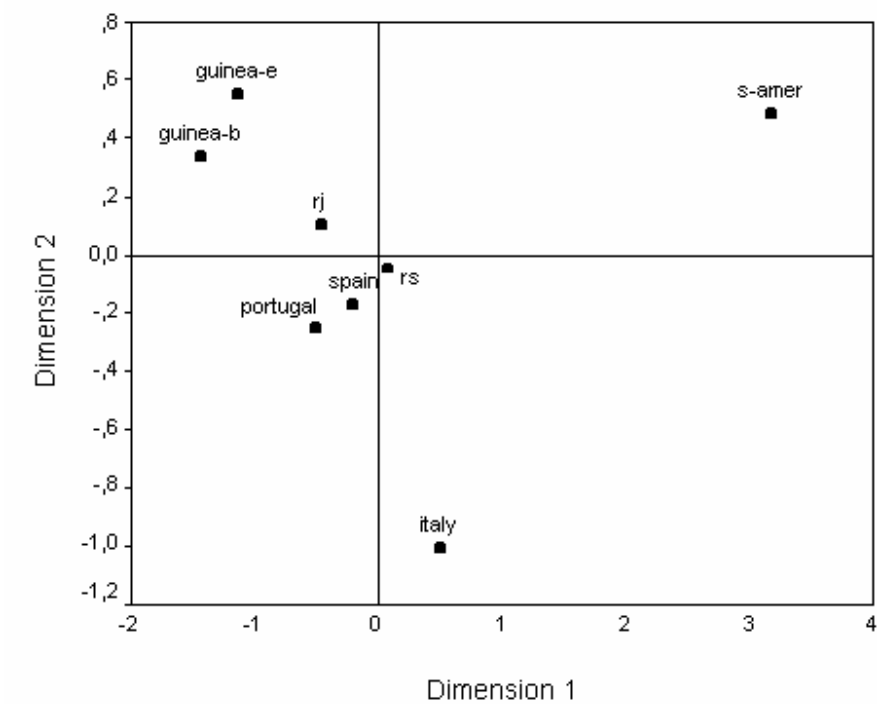
Table 5. Number of matches of the Bantu haplotypes from the RS population with the Promega Y database and the Y-STR database

Haplotype	Database and population			
	Y-hrd		Promega	
	African	Latin American	African American	Caucasian
H19	0	1	6	0
H42	1	0	0	0
H70	0	0	0	0
H90	3	0	2	0
H96	4	1	4	0
H117	2	0	0	0
H166	0	0	0	0
H168	0	0	0	0

Legend for Figure 1

Figure 1. Multidimensional Scaling of 11 Y-STR loci, obtained from data of the populations studied, together with data of world populations available in the literature. Each point represents one population, identified as followed: rs= Rio Grande do Sul, rj=Rio de Janeiro-RJ, guinea-b=Guinea Bissau, guinea-e=Guinea Equatorial, s-amer= South Amerindians.

Figure 1.



Capítulo 5

DISCUSSÃO

5. DISCUSSÃO

Os aspectos específicos relativos aos resultados obtidos para as populações estudadas constam nos três artigos, que são apresentados nos capítulos 2, 3 e 4 do presente trabalho. Este capítulo trata de uma discussão mais ampla acerca dos aspectos comuns e divergentes referentes aos marcadores investigados na presente tese.

Um aspecto a ser salientado diz respeito à predominância da contribuição européia na população do Rio Grande do Sul, detectada através do uso de STRs em diversas regiões cromossômicas, a partir de diferentes métodos estatísticos. Contudo, se analisarmos de forma mais aprofundada os resultados obtidos, observaremos o que segue:

Para o cromossomo Y, os resultados demonstraram níveis mais elevados de contribuição européia para a população gaúcha, se comparados com os dados obtidos a partir dos marcadores autossômicos. Esta característica pode ser evidenciada pelos resultados das análises de miscigenação, onde detectamos níveis de 92 % de contribuição européia com os Y-STRs e 80% com os STR autossômicos; e, de forma indireta, pelas análises de diferenciação genética entre populações, que mostraram a ausência de diferenças significativas entre a população do Rio Grande do Sul e outras populações brasileiras, além de menores valores de F_{st} . Estes resultados estão relacionados ao processo de colonização do Brasil, caracterizado pela imigração predominante de homens de origem de européia. Este processo histórico causa uma homogeneização das linhagens patrilíneas brasileiras, provavelmente neutralizando os possíveis efeitos de subestruturação para os marcadores de cromossomo Y, que deveriam ser mais pronunciados devido ao seu menor tamanho efetivo e patrilocalidade, característicos deste cromossomo. Apesar desta grande predominância européia, as análises de miscigenação obtidas a partir destes marcadores também demonstraram um pequeno percentual de

contribuição africana e níveis insignificantes de contribuição ameríndia para a população do Rio Grande do Sul, o que contrasta com os dados de Carvalho-Silva et al. (2001) e Marrero et al. (2005), que reportaram 100 % de contribuição européia para o cromossomo Y das regiões sociogeográficas do Brasil e do Rio Grande do Sul, respectivamente. Tais divergências nas estimativas devem estar relacionadas com diferenças nas abordagens metodológicas, uma vez que as amostras utilizadas por Carvalho-Silva et al. (2000) e Marrero et al. (2005) pertenciam a indivíduos classificados etnicamente como brancos; enquanto que as utilizadas neste trabalho apresentavam classificações étnicas variadas (vide capítulo 4). Além disto, os marcadores utilizados também foram de classes distintas – enquanto Carvalho-Silva et al. (2000) e Marrero et al. (2005) usaram Y-SNPs, no presente trabalho foram utilizados Y-STRs. A ausência de contribuição ameríndia nos cálculos de mistura para o cromossomo Y da população do Rio Grande do Sul indica a possibilidade da contribuição deste grupo parental ser predominantemente de origem materna. De fato, o estudo de Marrero et al. (2005) aponta para esta direção, demonstrando que 36 % das linhagens mitocondriais do Rio Grande do Sul são de origem ameríndia; enquanto que, para o cromossomo Y, não foi detectada contribuição deste grupo parental. A fim de analisarmos mais detalhadamente esta questão, realizamos uma busca no endereço eletrônico <http://www.yhrd.org> por haplótipos coincidentes aos encontrados para os Caingang e Guaraní estudados neste trabalho. Dos 28 haplótipos encontrados para estas populações indígenas, 18 não apresentaram coincidência na base de dados haplotípicos; dos 10 restantes, 9 apresentaram haplótipos coincidentes com amostras de indivíduos europeus e apenas um com um indivíduo sul-americano (vide Anexo 1). O fato de 63% dos haplótipos identificados não apresentarem correspondência com a base de dados sugere que esses sejam privados de ameríndios e, conseqüentemente, estejam sendo

identificados pela primeira vez no presente trabalho. Na realidade, a ação da deriva genética em populações endogâmicas e isoladas poderia ocasionar o surgimento de haplótipos únicos, que não estariam presentes nas outras populações pesquisadas. Em relação aos haplótipos restantes, alguns fatores poderiam ser elencados para explicar o fato de serem compartilhamentos com populações européias: a) primeiramente, tanto os Caingang quanto os Guaraní apresentam um elevado estágio de aculturação, esta característica deve também se refletir em seus genomas, fazendo com que haja uma grande miscigenação nestas populações. b) devido a natureza destes marcadores e suas altas taxas de mutação, que permitiria a ocorrência de haplótipos comuns entre ameríndios e europeus por homoplasia, isto é, os haplótipos seriam idênticos por estado e não por descendência. Certamente, mais estudos, com outras populações ameríndias com menor grau de miscigenação, são necessários para elucidar esta questão.

Para os marcadores autossômicos, observamos uma maior heterogeneidade nas distribuições alélicas, resultado da influência dos diversos grupos populacionais que contribuíram para a formação da população do Rio Grande do Sul. Este padrão traduziu-se de forma consistente em nossas análises, tanto nas estimativas de miscigenação, que demonstraram heterogeneidade na distribuição das contribuições parentais dentro das regiões do Estado, quanto nas análises de diferenciação genética, que indicaram diferenças próximas do nível de significância, quando comparamos nossos dados com os de outras regiões brasileiras. Outros métodos estatísticos utilizados no presente trabalho permitiram um melhor entendimento sobre a ancestralidade das populações estudadas. O Programa Structure (Pritchard et al. 2000) permitiu o cálculo de estimativas de ancestralidade ameríndias individuais para as amostras do Rio Grande do Sul, a partir dos dados autossômicos. As

estimativas coletivas obtidas por tal método podem ser comparadas com os cálculos de mistura, a fim de obtermos uma maior certeza acerca dos resultados alcançados. Neste caso, se compararmos os resultados do programa Structure com as estimativas de mistura obtidas pelo programa Admix (Chackraborty, 1985) para a população do Rio Grande do Sul, e também para as regiões agrupadas como oeste e centro-leste do Estado (ver capítulo 3), observamos uma notável coerência nos resultados alcançados em ambas as estimativas. Tal coerência pode ser observada pelo fato de que, em ambas estimativas, o oeste do Estado apresentou os maiores níveis de ancestralidade ameríndia, enquanto que o Centro-Leste exibiu os menores valores. Esta heterogeneidade nos níveis de ancestralidade exibida nas diferentes regiões do Rio Grande do Sul já havia sido demonstrada em estudos anteriores (Dornelles et al. 1999 e Marrero et al. 2005) utilizando marcadores proteicos e uniparentais, respectivamente. Desta forma, nosso estudo serviu para sedimentar os resultados previamente obtidos e também para definir a região oeste do Estado como um reservatório genético de genes ameríndios.

Especificamente em relação aos marcadores de cromossomo X e sua utilização para análise do desequilíbrio de ligação, este estudo é pioneiro na utilização de marcadores espaçados para verificar a extensão do desequilíbrio de ligação em grandes extensões do genoma de populações brasileiras e ameríndias. Os resultados encontrados indicaram a presença de altos níveis de desequilíbrio de ligação, principalmente no grupo indígena, trazendo possibilidades animadoras para estudos futuros, como por exemplo, para o mapeamento e a descoberta de genes de interesse médico. Neste aspecto, o cromossomo X cumpre um papel de fundamental importância, principalmente se considerarmos que as conseqüências fenotípicas de uma mutação deletéria são reveladas diretamente em homens, para qualquer gene sem homologia no cromossomo Y. Assim, embora contenha somente 4 %

do total de genes humanos, quase 10 % de todas as doenças mendelianas são atribuídas ao cromossomo X (Ross et al. 2005), traduzindo a importância de estudarmos o desequilíbrio de ligação através deste cromossomo. Além disto, obtivemos uma inferência indireta acerca da ancestralidade do cromossomo X da população do Rio Grande do Sul e de ameríndios, a partir da análise das diferenças médias das distribuições alélicas destas populações com dados de populações européias e asiáticas, devido ao fato de não termos dados de populações africanas disponíveis na literatura para os marcadores utilizados.

Os resultados obtidos no presente trabalho trazem perspectivas promissoras para o estudo da ancestralidade da população do Rio Grande do Sul, tanto em nível de pesquisa básica como para as suas aplicações futuras, tais como em análises forenses e estudos de associação.

Em relação às análises forenses, a ausência, para os Y-STRs, de diferenças significativas entre os dados haplotípicos do Rio Grande do Sul com outras regiões brasileiras indica que, sob um ponto de vista geral, bases de dados de maior amplitude populacional, como bases de dados regionais ou até mesmo nacionais, podem ser aplicadas para estes marcadores em populações miscigenadas. Contudo, estudos futuros devem ser realizados, a fim de sedimentarmos nossas conclusões. Para os marcadores autossômicos, devido à maior heterogeneidade encontrada em nossos dados populacionais, que indicaram diferenças próximas do nível de significância, quando comparamos nossos dados com os de outras regiões brasileiras, recomenda-se a adoção de bases de dados estaduais, ao invés de uma base de dados nacional. Para os X-STRs, a falta de dados na literatura de populações brasileiras para eventuais comparações não nos permitiu inferir sobre a presença de subestruturação ou não em populações miscigenadas.

Para os estudos de associação, a heterogeneidade das estimativas de miscigenação encontradas a partir dos STRs autossômicos indica a necessidade de verificar-se a presença ou não de estratificação por miscigenação significativa na população do Rio Grande do Sul, pois sua não consideração pode levar a associações espúrias entre marcadores e fenótipos. Contudo, os marcadores utilizados nesta tese não são os mais apropriados para este tipo de avaliação, por serem escolhidos especificamente para a identificação de indivíduos, sendo mais variáveis dentro de populações e produzindo menores valores de diferenciação gênica entre as mesmas (Barnholtz-Sloan et al. 2005). Assim, trabalhos futuros, baseados em marcadores bialélicos (SNPs e indels) devem ser efetuados na população do Rio Grande do Sul. Estes marcadores, devido às suas baixas taxas de mutação, tornam-se mais facilmente fixados em uma população. Dentre estes, deve-se concentrar esforços futuros na utilização de marcadores de ancestralidade, que apresentam grandes diferenças entre populações e/ou grupos étnicos. Outrossim, novas abordagens de amostragem devem ser utilizadas, consistindo, por exemplo, na genotipagem simultânea de amostras parentais (provenientes de indivíduos africanos, ameríndios e europeus) e de amostras de indivíduos brasileiros. Estas amostras parentais poderiam servir para a calibração das estimativas de ancestralidade, possibilitando a correta atribuição de ancestralidade ao nível individual. Por outro lado, uma das limitações para a genotipagem de marcadores bialélicos refere-se aos seus baixos níveis de heterozigotidade (Brookes, 1999). Assim, estudos de genética de populações envolvendo estes marcadores devem ser realizados com uma maior quantidade de loci, se comparados com os microsátélites, que são multialélicos. Para tanto, deve-se buscar novas metodologias que aliem a rapidez no processamento de amostras com a genotipagem simultânea de um grande número de marcadores, como por exemplo, os microarranjos, MALDI-TOF e outras.

REFERÊNCIAS BIBLIOGRÁFICAS

- Alves-Silva J, Santos MS, Guimarães PEM, Ferreira ACS, Bandelt HJ, Pena SDJ, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67:444-461.
- Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299-310.
- Ayres KL, Chaseling J, Balding DJ (2002) Implications for DNA identifications arising from analysis of Australian forensic databases. *For Sci Int* 129:90-98.
- Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB (2003) Human population genetic structure and inference of group membership. *Am J Hum Genet* 72:578-589.
- Barnholtz-Sian JS, Pfaff CL, Chackraborty R, Long JC (2005) Informativeness of the CODIS STR loci for admixture analysis. *J For Sci* 50 (6):1322-1326.
- Becker IIB (2003) Os índios charrua e minuano na antiga banda oriental do Uruguai. São Leopoldo: Editora da UNISINOS.
- Bell GI, Jurka J (1997) The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. *J Mol Evol* 44: 414-421.
- Biondo R, Spinella A, Montagna P, Walsh OS, Holt C, Budowle B (2001) Regional Italian allele frequencies at nine short tandem repeat loci. *For Sci Int* 115:95-98.
- Bortolini MC, Weimer TA, Salzano FM, Moura LB, Silva MCBO (1997) Genetic structure of two urban Afro-Brazilian populations. *Int J Anthropol* 12:5-16.

- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erlor 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.
- Bosch E, Calafell F, Santos FR, Pérez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. *Am J Hum Genet* 65:1623-1638.
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368: 455-457.
- Brinkmann B, Klitschar M, Neuhuber F, Huhne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* 62: 1408-1415.
- Buckleton J, Triggs CM, Walsh SJ (2005) *Forensic DNA Evidence Interpretation*. CRC Press, 534 p.
- Budowle B, Brown BL (2001). The use of DNA typing for forensic identification. *Forensica* 1:23-37.
- Budowle B, Moretti AL, Baumstark AL (1999) Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S.Caucasians, Hispanics Bahamians, Jamaicans and Trinidadians. *J For Sci* 44:1277-1286.
- Butler JM, *Forensic DNA Typing*, Academic Press, San Diego, CA, (2001).

- Calabrese PP, Durrett RT, Aquadro CF (2001) Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. *Genetics* 159: 839-852.
- Calafell F, Shuster A, Speed WC (1998) Short tandem repeat polymorphism evolution in humans. *Europ J Hum Genet* 6:38-49.
- Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ (2001) The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet* 68:281–286.
- Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Am J Hum Genet* 19:233-257.
- Cesar G (1970) *História do Rio Grande do Sul*. Editora Globo, Porto Alegre.
- Chakraborty R (1985) Gene identity in racial hybrids and estimation of admixture rates. In: Aluja YR & Neel JV (eds). *Genetic Differentiation in Human and other Animal Populations*. Indian Anthropological Association, Dehli, pp. 171-180.
- Deka R, Jin L, Shriver MD, Yu LM, DeCruo S, Hundrieser J, Bunker CH, Ferrell RE, Chakraborty R (1995) Population genetics of dinucleotide (dC-dA)_n(dG-dT)_n polymorphisms in world populations. *Am J Hum Genet* 56:461–474.
- Di Rienzo AA, Peterson C, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA* 91: 3166-3170.
- Dornelles CL, Callegari-Jacques SM, Robinson WM, Weimer TA, Franco MHL, Hickmann AC, Geiger CJ (1999) Genetics, surnames, grandparents, nationalities, and ethnic admixture in Southern Brazil: do the patterns of variation coincide? *Gen Mol Biol* 22:151-161.

- Ellegren H (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* 16: 551-558.
- Farber RA, Petes TD, Dominska M, Hudgens SS, Liskay RM (1994) Instability of simple sequence repeats in a mammalian cell line. *Hum Mol Genet* 3: 253-256.
- Flores M (1990). *História do Rio Grande do Sul*. Ed. Nova Dimensão, Porto Alegre.
- Foster P, Rohl A, Lunnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A short tandem repeat based phylogeny for the human Y chromosome. *Am J Hum Genet* 67:182-196.
- Franco MHL, Weimer TA, Salzano FM (1982) Blood polymorphisms and racial admixture in two Brazilian populations. *Am J Phys Anthropol* 58:127-132.
- Géada H, Rito RM, Ribeiro T, Espinheira R (2000) Portuguese population and paternity investigation studies with a multiplex PCR- the AmpFISTR Profiler Plus. *For Sci Int* 108:31-37.
- Góes ACS, Carvalho EF, Silva DA (2005). Population and mutation analysis of 17 Y-STR loci from Rio de Janeiro (Brazil). *Int J Leg Med* 119: 70–76.
- Goldstein D, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995a) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Nat Acad Sci* 92:6723-6727.
- Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW (1995b) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 139: 463-471.
- Goldstein DB, Schlotterer C (1999) *Microsatellites: Evolution and Applications*. Oxford University Press, New York, 343 pp.

Grattapaglia D, Schimidt AB, Costa e Silva C, Stringher C, Fernandes AP, Ferreira ME (2001) Brazilian population database for the 13 STR loci of AmpFISTR Profiler Plus™ multiplex kits. *For Sci Int* 118:91-94.

Grattapaglia D, Kalupniek S, Guimarães CS, Ribeiro MA, Diener OS, Soares CN (2005) Y-chromosome STR haplotype diversity in Brazilian populations. *Forens Sci Int* 149:99–107 .

Graves JM (2002) The rise and fall of SRY. *Trend Genet* 18 (5):259-264.

Greenberg JH (1987) *Language in the Americas*. California: Stanford University Press.

Hammer MF, Karafet T, Rasanayagam A et al (1998). Out of Africa and back again: Nested cladistic analysis of human Y chromosome variations. *Mol Biol Evol* 15:427-441.

Harpending H, Rogers A (2000). Genetic perspectives on human origins and differentiation. *Annu Rev Gen Hum Genet* 1:361-385.

Harr B, Schlötterer C (2000) Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics* 155: 1213- 1220.

Harr B, Todorova J, Schlötterer C (2002) Mismatch repair-driven mutational bias in *D. melanogaster*. *Mol Cell* 10: 199-205.

Henderson, ST, Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Mol Cell Biol* 12: 2749-2757.

- Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, Rozen S, Wilson RK, Page DC (2005). Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* 437:101-104.
- Hutz MH, Callegari-Jacques SM, Almeida SEM, Armbrorst T, Salzano FM (2002) Low levels of STRP variability are not universal in American Indians. *Hum Biol* 74:791-806.
- IBGE (2000) Censo Demográfico 2000. Características Gerais da População e Instrução. Rio Grande do Sul. Instituto brasileiro de Geografia e Estatística, Rio de Janeiro.
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
- Jacob S, Praz F (2002) DNA mismatch repair defects: role in colorectal carcinogenesis. *Biochimie* 84: 27-47.
- Jeffreys A, Wilson V, Thein SL (1985). Hypervariable “minisatellite” regions in human DNA. *Nature* 314(6006): 67-73.
- Jeffreys A, Neil D, Neumann R (1998) Repeat instability at human minisatellites arising from meiotic recombination. *EMBO J* 17:4147-4157.
- Jin L, Baskett LL, Cavalli-Sforza LL, Zhivotovsky LA, Feldman MW, Rosenberg NA (2000) Microsatellite evolution in modern humans: a comparison of two data sets from the same populations. *Ann Hum Genet* 62:117-134.
- Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 11:449-456.

Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 110:118–124.

Jobling MA (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. *For Sci Int* 118:158-162.

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

Jorde L, Bamshad M, Watkins W, Zenger R, Fraley A, Krakowiak P, Carpenter K, Soodyall H, Jenkins T, Rogers A (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am J Hum Genet* 57:523-538.

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krayowiak P, Sung S, Kere J, Harpending HC (1997) Microsatellite diversity and the demographic history of modern humans. *Proc Natl Acad Sci USA* 94:3100-3103.

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hm Genet* 66:979-988.

Jorde LB, Watkins WS Bamshad MJ (2001) Population genomics: a bridge from evolutionary history to genetic medicine. *Hum Mol Genet* 10:2199-2207.

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817–831.

- Katoh T, Mano S, Ikuta T, Munkhbat B, Tounai K, Ando H, Munkhtuvshin N, Imanishi T, Inoko H, Tamiya G (2002) Genetic isolates in East Asia: A study of linkage disequilibrium in the X chromosome. *Am J Hum Genet* 71:395-400.
- Kayser M, De Knijff PD, Djeltjes P, Krawczak M, Nagy M, Zerjal T (1997) Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* 18:1602-1607.
- Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observations in father/son pairs. *Am J Hum Genet* 66:1580–1588.
- Kimura M & Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* 49:725-738.
- Kohlrausch F, Callegari-Jacques SM, Tsuneto LT, Petzl-Erler ML, Hill K, Hurtado M, Salzano FM, Hutz MR (2005) Geography influences microsatellite polymorphism diversity in Amerindians. *Am J Ph Ant* 126:468-470.
- Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc Natl Acad Sci USA* 95: 10774- 10778.
- Laan M, Paabo S (1997) Demographic history of and linkage disequilibrium in human populations. *Nat Genet* 17:435-438.
- Laan M, Wiebe V, Khusnutdunova E, Remm M, Paabo, S (2004) X-chromosome as a marker for population history: linkage disequilibrium and haplotype study in Eurasian populations. *Eur J Hum Genet* 1-11.

- Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. *Science* 286:964-967.
- Lai Y, Sun F (2003) The Relationship Between Microsatellite Slippage Mutation Rate and the Number of Repeat Units. *Mol Biol Evol* 20 (12) 2123-2131.
- Lazzarotto P (1978). *História do Rio Grande do Sul*. Ed. Sulina, 3º edição. 154 p.
- Leite FPN, Menegassi FJ, Schwengber SP, Raimann PE, Albuquerque TK (2003) STR data for 09 autosomal STR markers from Rio Grande do Sul (Southern Brazil). *For Sci Int* 132:223-224.
- Levinson G, Gutman GA (1987) Slipped strand mispairing: A major mechanism for DNA sequence evolution. *Molecular and Biology Evolution* 4:203-221.
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002a) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* 11:2453-2465.
- Li WH, Yi S, Makova K (2002b) Male-driven evolution. *Curr Opin Genet Dev* 12:650-656.
- Luis JR, Rowold DJ, Rugeley 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.
- Marrero AR, Leite FPN, Carvalho BA, Peres LM, Kommers TC, Cruz IM, Salzano FM, Ruiz-Linares A, Silva-Junior 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.

- Meyer E, Wiese M, Bruchhaus H, Claussen M, Klein A (2000) Extraction and amplifications of authentic DNA from ancient human remains. *For Sci Int* 113:97-90.
- Morner M (1967) *Race Mixture in the History of Latin America*. Little Brown, Boston.
- Nei M, Tajima F, Tatenos Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *J Mol Evol* 19:153-170.
- Ohno S (1967) *Sex Chromosomes and Sex-linked Genes*. Springer, Berlin.
- Palha TJBF, Rodrigues EMB, Santos SEB (2006) Y-chromosomal STR haplotypes in a population from the Amazon region, Brazil. (in press).
- Pereira L, Gusmão L, Alves C, Amorim A, Prata MJ (2002). Bantu and European Y-lineages in Sub-Saharan Africa. *Ann Hum Genet* 66:369-378.
- Pereira RW, Monteiro EHG, Hirschfeld GCR (2006) Haplotype diversity of 17 Y-chromosome STRs in Brazilians. *For Sci Int* (in press).
- Pereira RW, Pena SD (2006) Phylogeography of haplotypes of five microsatellites located in a low-recombination region of the X chromosome: studies worldwide and in Brazilian populations. *Genetica* (2006) 126:243–250.
- Pérez-Lezaun A, Calafell F, Mateu E (1997) Allele frequencies for 20 microsatellites in a worldwide population survey. *Hum Hered* 47:189-196.
- Petes, TD, Greenwell PW, Dominska M (1997) Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. *Genetics* 146: 491-498.

- Primorac D, Andelinovic S, Definis-Gojanovic M, Drmic I, Rezic B, Baden MM, Kennedy MA, Schanfield MS, Skakel SB, Lee HC (1995) Identification of war victims from mass graves in Croatia, Bosnia and Herzegovina by the use of standard forensic methods and DNA typing. *J For Sci* 41:891-894.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. *Proc Natl Acad Sci USA* 95:8119-8123.
- Reichenpfader B, Immel U, Lintschar M (2003) Population data on the AmpFISTR SGM plus PCR amplification kit in Germans and Austrians. *For Sci Int* 132:84-86.
- Ricardo AC (2000) Povos indígenas do Brasil 1996/2000. São Paulo: Instituto Socioambiental.
- Roewer L, Croucher PJP, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet* 116: 279–291.
- Rosenberg N A, Pritchard J K, Weber J L, Cann H M, Kidd K K, Zhivotovsky L A, Feldman M W (2002) Genetic structure of human populations. *Science* 298: 2381–2385.
- Ross MT, Grafham DV, Coffey AJ, Scherer S, Bentley DR (2005) The DNA sequence of the human X chromosome. *Nature* 434:325-337.
- Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423:873-876.

- Salzano FM, Freire-Maia N (1967) *Populações Brasileiras: Aspectos Demográficos, Genéticos e Antropológicos*. Companhia Editora Nacional, São Paulo.
- Santos FR, Pandya A, Tyler-Smith C, Pena SD, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ (1999) The central Siberian origin for Native American Y chromosomes. *Am J Hum Genet* 64:619–628.
- Schaffner SF (2004) The X chromosome in population genetics. *Nature Rev Genet* 5:43-51.
- Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20: 211-215.
- Schneider H, Salzano FM (1979) Gm allotypes and racial admixture in two Brazilian populations. *Hum Genet* 53:101-105.
- Schoske R (2003) The design, optimization and testing of Y chromosome repeat megaplexes. PhD Thesis, American University, Washington.
- Shriver, MD, Jin L, Boerwinkle E, Deka E, Ferrell RE, Chakraborty R (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol. Biol. Evol.* 12: 914-920.
- Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrel R (1997). Ethnic-affiliation by use of population-specific DNA markers. *Am J Hum Genet* 60:957-964.
- Silva LAF, Pimentel BJ, Azevedo DA, Silva ENP, Santos SS (2002). Allele frequencies of nine STR loci-D16S539, D7S820, D13S317, CSF1PO, TH01, F13A01, FESFPS and vWA- in the population from Alagoas, northeastern Brazil. *For Sci Int* 130:187-188.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier HD, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell D, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong B, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825-837.

Slatkin, M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.

Stephens JC, Bilbert DA, Yuhki N, O'Brien SJ (1992) Estimation of heterozygosity for single-probe multilocus DNA fingerprints. *Mol Biol Evol* 9:729-743.

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

Tailon-Miller P, Bauer-Sardina I, Saccone NL, Putzel J, Laitinen P, Cao T, Kere J, Piria G, Rice JP, Kwoj PY (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat Genet* 25:324-328.

Talbot, CC, Avramopoulos D, Gerken S, Chakravarti A, Armour JA, Matsunami N, White R, Antonarakis SE (1995) The tetranucleotide repeat polymorphism D21S1245 demonstrates hypermutability in germline and somatic cells. *Hum Mol Genet* 4: 1193-1199.

Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB (2000). Y chromosome traveling south: the Cohen modal haplotype and the origins of the Lemba- the black Jews of southern Africa. *Am J Hum Genet* 66: 674-686.

- Thomson R, Pritchard JK, Shen PD, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes : evidence from DNA sequence data. *Proc Natl Acad Sci USA* 97:7360-7365.
- Tran HT, Degtyareva NP, Koloteva NN, Sugino A, Masumoto H, Gordenin DA, Resnick MA (1995) Replication slippage between distant short repeats in *Saccharomyces cerevisiae* depends on the direction of replication and the RAD50 and RAD52 genes. *Mol. Cell. Biol.* 15: 5607-5617.
- Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Hum Mol Genet* 2: 1123-1128.
- Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* 146: 769-779.
- Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* 150:499-510.
- Wilson JF, Goldstein DB (2000) Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. *Am J Hum Genet* 67:926-935.
- Wren, JD, Forgacs E, Fondon JW, Pertsemlidis A, Cheng SY, Gallardo T, Williams RS, Shohet RV, Garner HR (2000) Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. *Am J Hum Genet* 67: 345-356.
- Wright S (1951) The genetical structure of populations. *Annual Eugenics* 15:323-354.
- Xu X, Peng M, Fanget Z (2000) The direction of microsatellite mutations is dependent upon allele length. *Nat Gen* 24:396-399.

Zavattari P, Deidda E, Whalen M, Lampis R, Mulargia A, Loddo M, Eaves I, Mastio G, Todd J, Cucca F. (2000) Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum Mol Genet* 20:29 47-2957.

Zembruski VM, Callegari-Jacques SM, Hutz MH (2006). Application of an African Ancestry Index as a genomic control approach in a Brazilian population. *Ann Hum Genet* 70:1-7.

Zhivotovsky LA (1999) A new genetic distance with application to constrained variation at microsatellite loci. *Mol Biol Evol* 16:467-471.

Zhivotovsky LA, Rosenberg NA, Feldman MW (2003) Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am J Hum Genet* 72:1171–1186.

ANEXOS

Anexo 1 Número de coincidências dos haplótipos de Caingang e Guarani a partir da busca no banco de dados de haplótipos de cromossomo Y, disponível em <http://www.yhrd.org>.

Haplótipo	Y-hrd		
	Europa	América Latina	América do Norte
H24	21	5	
H175	6	1	
H178	8		1
H180	2		
H182	2		
H183	2		
H184	1		
H185		1	
H186	1		
H188	3		

Obs: Os haplótipos H46, H172, H173, H174, H176, H177, H179, H181, H187, H189, H190, H191, H192, H193, H194, H195, H196 e H197 não apresentaram coincidência na base de dados pesquisada.