



Faculdade de Medicina de São José do Rio Preto
Programa de Pós-Graduação em Ciências da
Saúde

Maria Eduarda Lopes Baitello

**Caracterização dos Dados Brasileiros de Haplótipos
Y-STR no Y - Chromosome Haplotype Reference
Database e Análise da Ancestralidade da Linhagem
Masculina Brasileira**

São José do Rio Preto
2018

Maria Eduarda Lopes Baitello

**Caracterização dos Dados Brasileiros de Haplótipos
Y-STR no Y - Chromosome Haplotype Reference
Database e Análise da Ancestralidade da Linhagem
Masculina Brasileira**

Dissertação encaminhada à Faculdade
de Medicina de São José do Rio Preto
– FAMERP para obtenção do título de
Mestre no Curso de Pós-graduação em
Ciências da Saúde, Eixo Temático:
Medicina e Ciências Correlatas.

**Orientadora: Profa. Dra. Camila Ive Ferreira
Oliveira Brancati**

**São José do Rio Preto
2018**

Baitello, Maria Eduarda Lopes

Caracterização dos dados brasileiros de haplótipos Y-STR no Y -
Chromosome Haplotype Reference Database e análise da
ancestralidade da linhagem masculina brasileira

São José do Rio Preto, 2018

86 p.

Dissertação (Mestrado) – Faculdade de Medicina de São José do Rio
Preto – FAMERP

Eixo Temático: Medicina e Ciências Correlatas

Orientadora: Profa. Dra. Camila Ive Ferreira Oliveira Brancati

1. Haplótipos 2. Y-STR 3. Genética Forense 4. Brasil

MARIA EDUARDA LOPES BAITELLO

**Caracterização dos Dados Brasileiros de
Haplótipos Y-STR no Y - Chromosome
Haplotype Reference Database e Análise da
Ancestralidade da Linhagem Masculina
Brasileira**

**BANCA EXAMINADORA
DISSERTAÇÃO PARA OBTENÇÃO DO GRAU
DE MESTRE**

**Presidente e Orientadora: Dra. Camila Ive F. O.
Brancati**

2º Examinador: Dr. Danilo Faustino Braganholi

**3º Examinador: Dra. Joyce Aparecida M. Lopes
Ferraz**

**Suplentes: Dra. Flávia Cristina Rodrigues Lisoni
Dra. Érika Cristina Pavarino**

São José do Rio Preto, 04/10/2018.

SUMÁRIO

DEDICATÓRIA	i
AGRADECIMENTOS	ii
EPÍGRAFE	iv
LISTA DE FIGURAS.....	v
LISTA DE TABELAS.....	vi
LISTA DE QUADROS	viii
LISTA DE ABREVIATURAS	ix
RESUMO.....	x
ABSTRACT.....	xi
1. INTRODUÇÃO	1
1.1 Métodos de Identificação Humana.....	1
1.2 Cromossomo Y.....	2
1.3 Aplicações dos Marcadores STR do Cromossomo Y (Y-STR).....	6
1.4 Interpretação da Evidência Genética e Banco de Dados Y-STR	9
1.5 Objetivos	12
2. MATERIAL E MÉTODOS	14
3. RESULTADOS	17
4. DISCUSSÃO	40
5. CONCLUSÃO	51
6. REFERÊNCIAS.....	53
7. APÊNDICE.....	65

DEDICATÓRIA

Dedico este trabalho à minha mãe, **Lígia**, e à minha avó, **Cirlei**, por serem exemplo de força e de amor, e a todos os professores, mestres e doutores que contribuíram para a minha formação.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter me dado o dom da vida e por olhar por mim a todo momento e nunca me abandonar. Por sempre me dar saúde e força para enfrentar os obstáculos e superar as dificuldades, e por conhecer o tempo certo de todas as coisas e sempre iluminar os meus caminhos.

À minha mãe, Lígia Eduarda Lopes Camazano, por sempre estar ao meu lado, nos momentos bons e ruins. Por sorrir minhas alegrias e chorar minhas tristezas. Por ser guerreira, lutadora e vencedora, e por ter me ensinado a nunca desistir. À você, mãe, minha eterna gratidão. Agradeço ao meu pai, Jayr Anselmo Baitello Filho, por não ter criado uma princesa, e ao meu irmão, Eduardo Augusto Lopes Baitello, pela parceria e amizade. Agradeço também à minha avó, Cirlei Camazano, por todas as orações e apoio, e às minhas tias, por serem o alicerce da nossa família. Ao meu namorado, Luis Betineli, por estar sempre ao meu lado me dando forças e por acreditar no meu potencial. Todos vocês são parte essencial da minha formação. Obrigada por acreditarem em mim!

Agradeço à minha orientadora, Dra. Camila Ive Ferreira Oliveira Brancati, pela confiança em me orientar e por todas as palavras de apoio durante essa jornada. À Professora Dra. Dorotéia Rossi Silva Souza, por ter acreditado em mim e me dado todas as oportunidades de desenvolver o meu potencial, por transmitir seus conhecimentos científicos com tanto entusiasmo e dedicação. Por ser um exemplo de caráter, de humildade, de otimismo e de fé.

Agradeço à Dra. Joyce Aparecida Martins Lopes Ferraz e à Dra. Leonor Gusmão por terem compartilhado seu exímio conhecimento com muita paciência, e por

terem me direcionado e auxiliado no meu crescimento pessoal e profissional durante essa jornada.

Aos companheiros e amigos de pesquisa do Laboratório NPBIM, agradeço pelo carinho, ensinamentos e ajuda durante meus estudos e acima de tudo pela amizade. Por todos os momentos de descontração que tornaram essa jornada muito mais leve. Esta conquista seria muito mais difícil sem o apoio de vocês.

Agradeço a todos os professores que fizeram parte da minha formação acadêmica, compartilhando seu conhecimento com tanto amor e dedicação. A todos os funcionários da pós-graduação pela paciência e pela disponibilidade em ajudar, aos funcionários do bloco U6 pela gentileza, sabedoria e amizade, e a todos os funcionários da Famerp por trabalharem juntos no desenvolvimento desta instituição.

Agradeço a todos que estiveram comigo fazendo parte desta jornada, que me incentivaram, questionaram, discordaram, apoiaram, respeitaram, que acreditaram no meu potencial, me deram condições de crescer e contribuíram para a minha formação profissional e pessoal.

Sem todos vocês, essa conquista não teria sido possível. A todos, muito obrigada!

EPÍGRAFE

“Muitas das coisas mais importantes do mundo foram conseguidas por pessoas que continuaram tentando quando parecia não haver mais nenhuma esperança de sucesso.”

(Dale Carnegie)

LISTA DE FIGURAS

- Figura 1 - Representação esquemática do processo evolutivo de formação dos cromossomos sexuais X e Y em mamíferos (Adaptado de: Bellot et al., 2009).⁽¹¹⁾ 3
- Figura 2 - Estrutura e organização das diferentes regiões do cromossomo Y humano. a) Representação esquemática do cromossomo, incluindo as regiões eucromáticas, pseudoautossômicas e heterocromáticas. b) Visualização ampliada da porção eucromática da região masculino específica (MSY), mostrando suas três sequências (sequência X-transposta, X-degenerada e amplicônica) bem como duas regiões heterocromáticas (centrômero e outra pequena região que interrompe a porção eucromática no braço longo do cromossomo Y; Modificado de: Skaletsky et al., 2003).⁽¹⁴⁾ 6
- Figura 3 - Distribuição dos haplótipos Y-STR (A) e das populações (B) por região brasileira e metapopulação, considerando-se os dados do haplótipo mínimo no *Y Chromosome Haplotype Reference Database* (YHRD). 21
- Figura 4 - Representação dos Estados brasileiros com populações disponíveis no *Y-Chromosome Haplotype Reference Database* (YHRD), considerando-se o Sistema *Yfiler*. 22
- Figura 5 - Análise em Escala Multidimensional (MDS) baseada na matriz de distância genética F_{ST} entre pares de regiões geográficas, bem como entre regiões e metapopulações brasileiras genotipadas para o sistema *Yfiler* e disponíveis no banco de dados *Y-Chromosome Haplotype Reference Database* (YHRD; Tabela 1). Stress: 0,0002. 26
- Figura 6 - Análise em Escala Multidimensional (MDS) baseada na matriz de distância genética F_{ST} entre pares de populações brasileiras miscigenadas (A) e nativo-americanas (B) genotipadas para o sistema *Yfiler* e disponíveis no banco de dados *Y-Chromosome Haplotype Reference Database* (YHRD; Tabelas 2 e 4). Stress: (A)=0,1994; (B)=0,3385. 32

LISTA DE TABELAS

Tabela 1 - Matriz de distância genética entre pares de regiões geográficas representadas por populações brasileiras do grupo miscigenado e entre tais regiões e a metapopulação nativo-americana e europeia, com valores de p (acima da diagonal) e valores de distância genética (F_{ST} ou R_{ST} ; abaixo da diagonal).	25
Tabela 2 - Matriz de distância genética entre pares de populações brasileiras do grupo miscigenado, com valores de p (acima da diagonal) e valores de distância genética (F_{ST} ; abaixo da diagonal).	27
Tabela 3 - Matriz de distância genética entre pares de populações brasileiras do grupo miscigenado, com valores de p (acima da diagonal) e valores de distância genética (R_{ST} ; abaixo da diagonal).	28
Tabela 4 - Matriz de distância genética entre pares de populações brasileiras do grupo nativo americano, com valores de p (acima da diagonal) e valores de distância genética (F_{ST} ; abaixo da diagonal).	30
Tabela 5 - Matriz de distância genética entre pares de populações brasileiras do grupo nativo americano, com valores de p (acima da diagonal) e valores de distância genética (R_{ST} ; abaixo da diagonal).	31
Tabela 6 - Matriz de distância genética entre pares de regiões geográficas representadas por populações brasileiras do grupo miscigenado e entre tais regiões e a metapopulações nativo-americana e europeia, excluindo-se as populações de Belém-PA e Boa Vista-RR, com valores de p (acima da diagonal) e valores de distância genética (F_{ST} ou R_{ST} ; abaixo da diagonal).	33
Tabela 7 - Análise de Variação Molecular (AMOVA) em diferentes grupos de populações brasileiras.	34

Tabela 8 - Origem, distribuição e frequência dos haplogrupos (HGs) das populações brasileiras de origem miscigenada, nativo-americana e europeia, com distribuição regional para populações miscigenadas, com base em haplótipos Y-STR..... 37

Tabela 9 - Frequência da ancestralidade das populações brasileiras de origem miscigenada, nativo-americana e europeia, com distribuição regional para populações miscigenadas, com base na origem dos haplogrupos. 38

LISTA DE QUADROS

- Quadro 1 - Lista das populações brasileiras com haplótipos Y-STR registrados no *Y-Chromosome Haplotype Reference Database* (YHRD), seus respectivos números de haplótipos tipados nos diferentes sistemas de marcadores do cromossomo Y, sua região geográfica, bem como a metapopulação à qual cada população pertence no YHRD. ... 17
- Quadro 2 - Populações brasileiras genotipadas para o sistema *Yfiler*, e suas respectivas regiões geográficas, metapopulações, número de haplótipos e referência na literatura. 22
- Quadro 3 - Distribuição dos haplótipos Y-STR mais comuns no Brasil. 35

LISTA DE ABREVIATURAS

AMOVA	(do inglês) <i>Analysis of molecular variance</i> - Análise de variação molecular
AS-STR	(do inglês) <i>Autosomal STR marker</i> - Marcadores STR autossômicos
DNA	(do inglês) <i>Deoxyribonucleic acid</i> - Ácido desoxirribonucleico
h	Haplótipos
ISFG	(do inglês) <i>International Society of Forensic Genetics</i> - Sociedade Internacional de Genética Forense
LD	(do inglês) <i>Linkage disequilibrium</i> – Desequilíbrio de ligação
LR	(do inglês) <i>Likelihood ratio</i> - Razão de verossimilhança
MDS	(do inglês) <i>Multidimensional scaling</i> - Escala multidimensional
MSY	(do inglês) <i>Male-specific region of the Y chromosome</i> - Região masculino específica do cromossomo Y
PAR	(do inglês) <i>Pseudoautosomal region</i> - Região pseudoautossômica
pop	Populações
PPY	<i>PowerPlex® Y</i>
SRY	(do inglês) <i>Sex-determining region Y</i> - Região determinante do sexo do cromossomo Y
STR	(do inglês) <i>Short Tandem Repeats</i> – repetições consecutivas curtas
Yfiler	<i>AmpFLSTR® Yfiler® PCR Amplification Kit</i>
YHRD	(do inglês) <i>Y-Chromosome Haplotype Reference Database</i> – Banco de dados referência de haplótipos do cromossomo Y
Y-STR	(do inglês) <i>Y Chromosome STR markers</i> - Marcadores STR do cromossomo Y

RESUMO

Introdução: Em identificação humana, o cromossomo Y possui diversas aplicações que auxiliam na rotina forense, porém é necessário um banco de dados haplotípicos de *Short Tandem Repeat* do cromossomo Y (Y-STR) para estimar a frequência de um perfil genético. A integração dos dados haplotípicos provenientes de diferentes regiões brasileiras somente é válida se não houver subestrutura populacional. **Objetivos:** Caracterizar a estrutura populacional brasileira com base em dados de haplótipos Y-STR depositados no *Y Chromosome Haplotype Reference Database* (YHRD), bem como analisar a ancestralidade da linhagem masculina do país. **Material e Métodos:** Foram recuperados na literatura os dados de populações brasileiras submetidos no banco de dados YHRD, atualização 56. As análises de distância genética (F_{ST} e R_{ST}), estrutura populacional (AMOVA) e diversidade haplotípica (Fórmula de Nei) foram realizadas utilizando populações genotipadas para o kit *Yfiler*® disponíveis na literatura. Para análise da ancestralidade brasileira foram inferidos haplogrupos com base em haplótipos Y-STR. **Resultados:** Foram efetuadas análises envolvendo 6.553 haplótipos Y-STR de populações brasileiras de origem nativo-americana, miscigenada e europeia. Entre as populações brasileiras nativo-americanas, foram encontrados altos valores de distância genética. Já nas populações miscigenadas e europeias, todos os pares apresentaram baixos valores de distância genética, com destaque para a região Norte e as populações de Belém (PA) e Boa Vista (RR), que apesar de baixa distância genética, mostraram valor-P significativo ($P < 0,00022$). A variação haplotípica considerando as populações miscigenadas separadas por região geográfica foi de 99,55% dentro das populações ($P < 0,00000$) e apenas 0,38% ($P < 0,00000$) e 0,07% ($P < 0,10119$) entre as populações dentro das regiões e entre as regiões do Brasil, respectivamente. A diversidade haplotípica das populações miscigenadas foi de 0,99993 e a capacidade de discriminação de 87,6%. Nas populações europeias os valores foram 0,99983 e 96,8%, respectivamente, enquanto em populações nativo-americanas, 0,98377 e 45,25%, respectivamente. Análise envolvendo 4.718 haplogrupos de 22 populações brasileiras de origem miscigenada e duas populações de origem europeia revelou ancestralidade essencialmente europeia no Brasil, com destaque para as regiões sul e centro-oeste. As regiões com maior contribuição africana foram o nordeste e sudeste. Entre 16 populações de origem nativo-americana, foram inferidos 342 haplogrupos, que revelaram ancestralidade ameríndia representada em 94,15% das amostras. **Conclusões:** Considerando-se o sistema *YFiler*®, as metapopulações brasileiras miscigenada e europeia não mostram subestrutura populacional, indicando que uma base nacional de haplótipos Y-STR pode ser utilizada na estimativa das frequências haplotípicas destas populações em casos forenses. Além disso, nossos achados indicam ancestralidade essencialmente europeia da linhagem masculina brasileira, confirmando a história de colonização e povoamento do Brasil.

Palavras-chave: 1. Haplótipos 2. Y-STR 3. Genética Forense 4. Brasil

ABSTRACT

Introduction: In human identification, the Y-chromosome has several applications that aid in forensic routine, but a haplotype database of Y chromosome Short Tandem Repeat (Y-STR) is needed to estimate the frequency of a genetic profile. The integration of haplotypic data from different Brazilian regions is only valid if there is no population substructure. **Objectives:** To characterize the Brazilian population structure based on data from Y-STR haplotypes deposited in the Y Chromosome Haplotype Reference Database (YHRD), as well as to analyze the ancestry of the male lineage of the country. **Material and Methods:** Data from Brazilian populations submitted to the YHRD database, update 56, were retrieved from the literature. The genetic distance (F_{ST} and R_{ST}), population structure (AMOVA) and haplotype diversity (Nei formula) analyzes were performed using populations genotyped for the Yfiler® kit and available in the literature. For analysis of Brazilian ancestry, haplogroups based on Y-STR haplotypes were inferred. **Results:** Analyzes were carried out involving 6,553 Y-STR haplotypes of Brazilian populations of Native American, Admixed and European origin. Among the Brazilian Native American populations, high genetic distance values were found. In the Admixed and European populations, all pairs presented low values of genetic distance, with emphasis on the North region and the populations of Belém (PA) and Boa Vista (RR), which, despite low genetic distance, showed significant P-value ($P < 0.00022$). The haplotypic variation within Admixed populations separated by region was 99.55% ($P < 0.00000$) and only 0.38% ($P < 0.00000$) and 0.07% ($P < 0.10119$) between populations within regions and between regions of Brazil, respectively. The haplotypic diversity of the Admixed populations was 0.99993 and the discrimination capacity was 87.6%. In European populations the values were 0.99983 and 96.8%, respectively, while in Native American populations, 0.98377 and 45.25%, respectively. Analysis involving 4,718 haplogroups of 22 Brazilian populations of Admixed origin and two populations of European origin revealed essentially European ancestry in Brazil, with emphasis on the south and center-west regions. The regions with the greatest African contribution were the northeast and southeast. Among 16 populations of Native American origin, 342 haplogroups were inferred, which revealed Amerindian ancestry represented in 94.15% of the samples. **Conclusion:** Taking into account the Yfiler® system, Brazilian Admixed and European metapopulations do not show population substructure, pointing out that a national base of Y-STR haplotypes can be used to estimate the haplotype frequencies of these populations in forensic cases. In addition, our findings indicate an essentially European ancestry of the Brazilian male lineage, establishing the history of colonization and settlement in Brazil.

Keywords: 1. Haplotypes 2. Y-STR 3. Forensic Genetics 4. Brazil

1 . INTRODUÇÃO

1. INTRODUÇÃO

1.1 Métodos de Identificação Humana

Cada indivíduo possui um conjunto de caracteres próprios e exclusivos que o individualizam perante outras pessoas, animais ou coisas, compondo o que chamamos de identidade. A identificação é um processo ou conjunto de processos destinados a estabelecer uma identidade, objetivando a diferenciação e individualização de cada pessoa. Não é apenas um reconhecimento, pois utiliza técnicas precisas, inconfundíveis e irrefutáveis para provar que uma pessoa é totalmente distinta de qualquer outra do planeta. Ao longo da história, foram empregados diversos métodos de identificação denominados de “primitivos”, desprovidos de princípios científicos e assim caracterizados: Nome, Ferrete, Mutilação, Tatuagem e Fotografia. Atualmente, no âmbito forense são utilizados alguns métodos científicos aceitos universalmente para identificação, tais como: Sistema Odontológico de Amoedo, Sistema Papiloscópico de Vucetich e Exames Genéticos.^(1,2)

Com relação aos exames genéticos, a tipagem do material genético para a identificação humana surgiu apenas em 1984 quando o geneticista Alec Jeffreys descobriu regiões repetitivas do ácido desoxirribonucleico (DNA) que variavam de um indivíduo para o outro, produzindo perfis genéticos conhecidos como “DNA *fingerprinting*”.⁽³⁾ Devido ao processo de recombinação gênica ao qual o material genético é submetido, existe alto grau de variabilidade genética entre indivíduos, de modo que cada ser humano possui um perfil genético exclusivo, com exceção dos gêmeos monozigóticos, que compartilham do mesmo perfil genético, salvo eventuais mutações.⁽⁴⁾ A molécula de DNA possui alto poder de discriminação ocasionado por

polimorfismos genéticos de sequência e de comprimento. Nesse caso, são originados principalmente por substituição de um ou mais nucleotídeos e por inserção ou deleção de dois ou mais nucleotídeos na sequência de DNA, respectivamente.⁽⁵⁾

No polimorfismo de comprimento destacam-se os marcadores de repetições consecutivas curtas (*Short Tandem Repeats*, STR) ou microsatélites, formados pela repetição de unidades com 2 a 7 pares de bases (pb) e utilizados atualmente na rotina forense.^(5,6) Marcadores STR estão localizados no DNA nuclear dos cromossomos autossomos (AS) e sexuais (X e Y). O caráter multialélico e a herança biparental, a qual permite recombinação alélica entre o par de cromossomos homólogos, conferem maior poder de discriminação aos marcadores STR autossômicos (AS-STR) e, portanto, estes são os de escolha padrão nos exames de DNA.^(5,7) No entanto, em casos específicos, marcadores STR do cromossomo Y (Y-STR) podem auxiliar a análise padrão autossômica devido ao seu padrão de herança característico.

1.2 Cromossomo Y

Estudos relatam que a origem dos cromossomos sexuais X e Y dos mamíferos deu-se há cerca de 300 milhões de anos, a partir da evolução de um par de cromossomos autossomos homólogos em cromossomos sexuais heteromórficos pela aquisição de um gene determinante do sexo masculino (*sex determining region of the Y chromosome – SRY*) no proto-cromossomo Y. Após esse evento, ocorreu uma inversão no cromossomo sexo-específico que tornou inviável a recombinação cromossômica nessa região, dando início ao processo de diferenciação do cromossomo Y (Figura 1).⁽⁸⁾

A ausência de recombinação ocorreu em diversas regiões do cromossomo Y e em tempos distintos e, portanto, foram divididas em estratos evolucionários, cada um

com seu próprio curso de degeneração do cromossomo. O estrato mais antigo em humanos (estrato 1) envolve o surgimento do SRY e data de mais de 240 milhões de anos, e o mais recente estrato identificado (estrato 5) originou-se há aproximadamente 30 milhões de anos.⁽⁹⁾

Diversos modelos evolutivos foram propostos para explicar a degeneração do cromossomo Y. Todos os modelos compartilham uma mesma premissa: a falta de recombinação é um fator chave dessa degeneração, pois reduz a eficácia da seleção natural ao impedir que atue independentemente em mutações benéficas e deletérias, forçando-a a agir, nesse caso, em cromossomos inteiros não recombinantes.^(9,10) Dois modelos amplamente citados em relação a degeneração do cromossomo Y são a “catraca de Muller” e a “carona genética”.

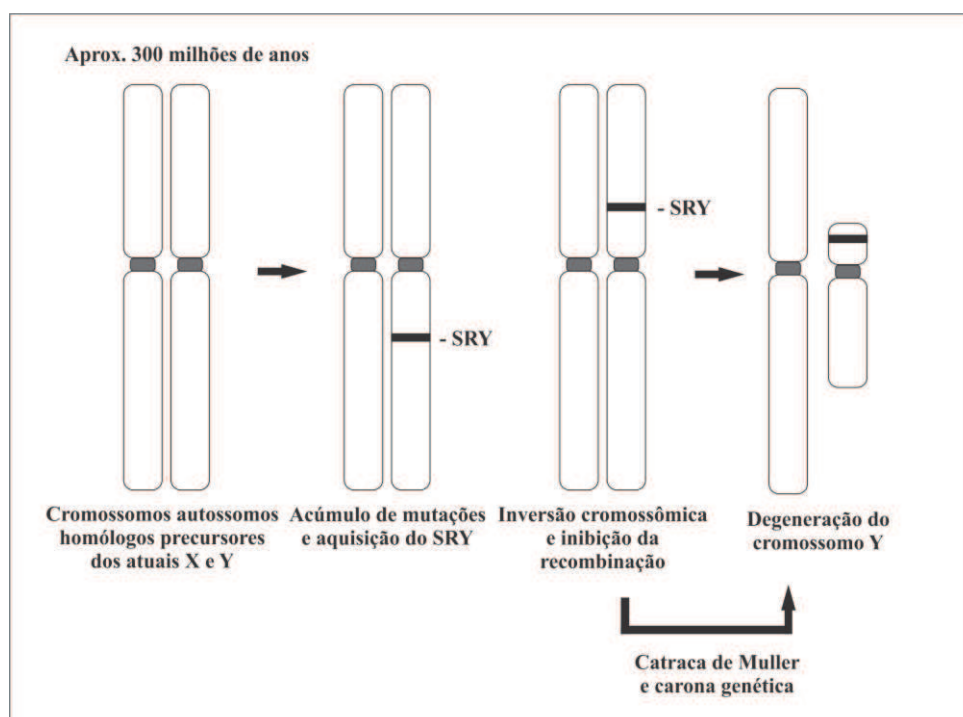


Figura 1 - Representação esquemática do processo evolutivo de formação dos cromossomos sexuais X e Y em mamíferos (Adaptado de: Bellot et al., 2009).⁽¹¹⁾

A “catraca de Muller” é uma teoria estocástica e sustenta que, em uma população finita, se uma região do cromossomo que possui locus suscetível a mutações

deletérias não sofre recombinação, e se mutações secundárias são raras, ocorrerá um acúmulo irreversível de mutações deletérias nesta população, pois a recombinação é essencial para recriar cromossomos livres de mutação. Já a “carona genética” postula que, na ausência de recombinação, a disseminação de um alelo mutante benéfico originado no cromossomo Y causa também a fixação de alelos deletérios em outros locus que estão presentes nesse mesmo cromossomo, caso a vantagem seletiva da mutação benéfica supere o efeito do alelo deletério a ela ligado. Em ambos os casos, o cromossomo X terá aumento da sua atividade nos loci homólogos aos alelos deletérios do cromossomo Y, à medida que o Y será inativado.⁽¹⁰⁾ Caso o efeito da mutação deletéria supere a vantagem da mutação benéfica, ambas mutações serão eliminadas, o que causa uma evolução menos adaptativa do cromossomo Y em relação ao cromossomo X.⁽⁹⁾ Embora esses modelos possuam teorias gerais bem explicativas e amplamente aceitas, sua contribuição relativa à degeneração dos cromossomos Y ainda é obscura.^(12,13)

O cromossomo Y é um dos menores cromossomos humanos, com tamanho médio estimado de 68 milhões de pares de base (Mb), consideravelmente menor que o cromossomo X (~155 Mb). Sua estrutura é dividida em duas regiões pseudoautossômicas (*pseudoautosomal region* – PAR 1 e 2; ~3Mb) e na região masculino-específica do cromossomo Y (*male-specific region of the Y chromosome* – MSY; ~65 Mb), a qual é dividida em uma porção heterocromática (~41 Mb) e uma porção eucromática (~24 Mb).⁽¹⁴⁾ Na Figura 2 está representada a estrutura e organização do cromossomo Y humano. As regiões PAR 1 e PAR 2 estão localizadas nas extremidades dos braços curto (Yp) e longo (Yq) do cromossomo Y,

respectivamente. Essas regiões possuem homólogos no cromossomo X, portanto estão sujeitas à recombinação nos cromossomos sexuais durante a meiose.⁽¹⁵⁾

A região masculino específica do cromossomo Y (MSY) abrange 95% do cromossomo Y e é responsável pela diferenciação dos sexos, de modo que não sofre recombinação com o cromossomo X e determina um perfil haploide masculino. Sua porção heterocromática é constituída por repetições consecutivas de baixa complexidade de sequência.⁽¹⁴⁾

A porção eucromática da MSY é quase que integralmente constituída por três classes de sequência: X-transposta, X-degenerada e amplicônica. A sequência X-transposta (~3,4 Mb) originou-se de uma transposição de X para Y que ocorreu cerca de três a quatro milhões de anos atrás. Desse modo, essa região é 99% idêntica à sua sequência homóloga no cromossomo X, no entanto não possui recombinação durante meioses masculinas,^(9,14)

A sequência X-degenerada (~8,6 Mb) é remanescente dos autossomos ancestrais a partir dos quais X e Y co-evoluíram. Consiste em uma versão deteriorada do cromossomo X e abrange o gene determinante do sexo masculino (SRY).⁽¹⁴⁾

A terceira sequência da porção eucromática, a amplicônica (~10,2 Mb), é uma sequência altamente repetitiva que possui genes com funções masculinas específicas, todos expressos nos testículos. Seus genes representam nove famílias distintas, algumas sobreviventes da degradação do cromossomo Y, outras derivadas de transposições de genes autossômicos. Geralmente arrançadas em palíndromos, as regiões amplicônicas exibem acentuada similaridade com outras sequências em MSY (mais que 99,9%), o que permite recombinação produtiva entre genes, como se fossem dois alelos de

autossomos homólogos, e possibilita a remodelação contínua do cromossomo Y, estabilizando sua degeneração.⁽¹¹⁾

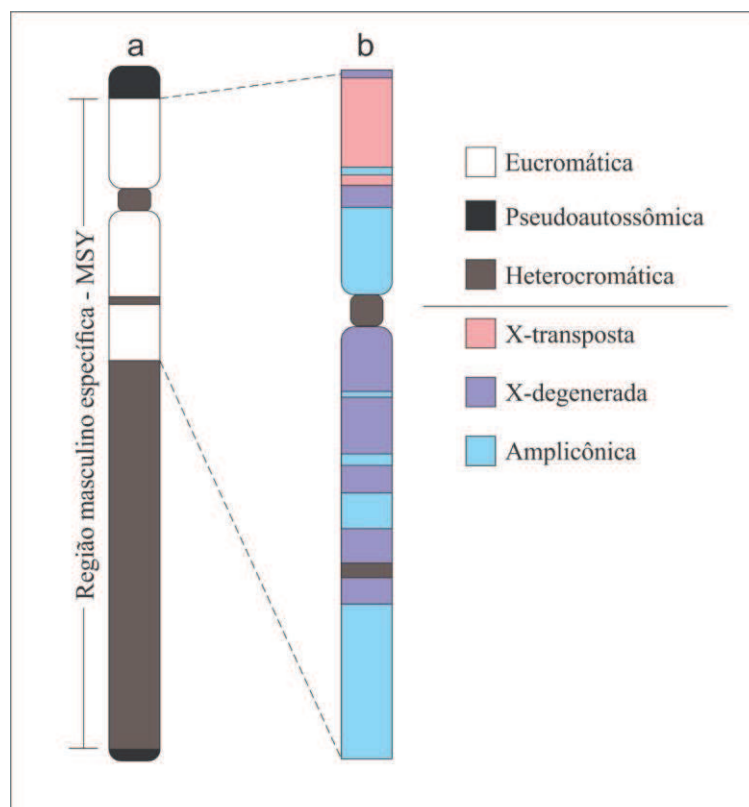


Figura 2 - Estrutura e organização das diferentes regiões do cromossomo Y humano. a) Representação esquemática do cromossomo, incluindo as regiões eucromáticas, pseudoautossômicas e heterocromáticas. b) Visualização ampliada da porção eucromática da região masculina específica (MSY), mostrando suas três sequências (sequência X-transposta, X-degenerada e amplicônica) bem como duas regiões heterocromáticas (centrômero e outra pequena região que interrompe a porção eucromática no braço longo do cromossomo Y; Modificado de: Skaletsky et al., 2003).⁽¹⁴⁾

1.3 Aplicações dos Marcadores STR do Cromossomo Y (Y-STR)

Os marcadores STR autossômicos (AS-STR) possibilitam maior discriminação entre indivíduos devido ao caráter multialélico e a herança biparental dos cromossomos

autossomos, o que permite recombinação alélica entre seus pares homólogos. Por esse motivo, eles são o padrão de escolha para exames de DNA em genética forense.⁽¹⁶⁾

No entanto, em casos específicos, os AS-STR podem não fornecer uma resposta definitiva, de modo que Y-STRs podem auxiliar eficientemente a análise padrão autossômica. Em virtude das características específicas do cromossomo Y como herança uniparental, ausência de recombinação (exceto nas regiões PAR 1 e 2) e forte desequilíbrio de ligação, filhos homens herdam um perfil haplotípico de Y-STR, o qual é transmitido idêntico de pai para filho e compartilhado por toda a linhagem masculina de uma família, salvo por eventuais mutações.⁽¹⁷⁾ Desse modo, a diversidade genética da região MSY é uma ferramenta útil para identificar homens paternalmente relacionados (inclusão), diferenciar linhagens masculinas não relacionadas (exclusão), e, ainda, traçar a história, movimentos e ancestralidade de uma população.^(18,19)

Em alguns casos, a análise dos AS-STR apresentam até três inconsistências em exames de vínculo biológico e necessitam de análise de marcadores adicionais para certificar a exclusão ou não do parentesco investigado.^(20,21) Isso ocorre, principalmente, em casos de reconstrução, em que o suposto pai está ausente e são utilizados perfis genéticos de seus parentes para o exame.⁽²²⁾ Nesse contexto, laboratórios forenses tem adotado Y-STR para auxiliar a investigação quando o filho envolvido é do gênero masculino.⁽²³⁾

Os marcadores do cromossomo Y são transmitidos de forma inalterada (perfil haplotípico) do pai para o filho masculino, salvo por mutações. Assim, a aplicação de Y-STRs em testes de parentescos limita-se aos casos com análise da descendência masculina e não permite diferenciar os indivíduos masculinos relacionados por via

paterna, no entanto, é uma ferramenta muito poderosa para excluir o alegado pai que não está relacionado ao verdadeiro pai biológico da criança.⁽²⁰⁾

Outra situação na qual os Y-STRs podem auxiliar com eficiência a análise padrão autossômica são casos de identidade forense, em especial violência sexual, em que há mistura do material da vítima e do agressor, com grande sobreposição do material feminino ao masculino. Nessas situações, a análise dos AS-STR se torna difícil ou até impossível, pois o componente majoritário é mais favorável à amplificação, podendo mascarar o perfil genético minoritário ou até suprimir a sua amplificação. Além disso, a presença de mais de um agressor também dificulta a individualização do perfil genético de cada um deles. Nesses casos, a análise de Y-STR é eficaz na identificação do perfil masculino, pois apenas o agressor possui o cromossomo Y, e a presença de mais de um alelo em cada marcador auxilia na identificação da quantidade de agressores envolvido naquele caso.⁽²⁴⁾

Em casos de identificação humana de pessoas desaparecidas, como vítimas carbonizadas, exumações e restos mortais em desastre em massa, é possível que não se tenha uma amostra referência da vítima para comparar aos restos mortais analisados. Além disso, a amostra referência da vítima, caso disponível, pode estar degradada, resultando em um perfil AS-STR incompleto. Nestas circunstâncias, quando o indivíduo for do gênero masculino, a transmissão do perfil Y-STR idêntico de pai para filho permite que qualquer ascendente ou descendente masculino paterno do indivíduo em questão sirva de amostra referência para reconstruir a relação paterna. Isso é possível mesmo que o familiar seja um parente biológico muito distante, como reportado em vários casos de identificação humana solucionados pela tipagem de Y-STRs.⁽²⁵⁻²⁷⁾

1.4 Interpretação da Evidência Genética e Banco de Dados Y-STR

Dado um conjunto específico de genótipos, a genética forense deve determinar a probabilidade de relação de parentesco entre os indivíduos testados, bem como a probabilidade de identidade entre um vestígio biológico e um determinado indivíduo. Para tanto, a Comissão de DNA da Sociedade Internacional de Genética Forense (*International Society of Forensic Genetics*, ISFG) recomenda o uso do método de Razão de Verossimilhança (*Likelihood Ratio*, LR) para avaliar o peso de uma evidência genética.⁽²⁸⁾ Nesse contexto, a Comissão de Testes de Paternidade da ISFG expõe que os laboratórios são responsáveis por estabelecer e admitir seu próprio critério de exclusão para uma relação de parentesco, porém também recomenda que seja baseado no LR.⁽²⁹⁾

O LR é um método que compara a probabilidade de uma evidência sob duas hipóteses mutuamente exclusivas, expresso pela seguinte equação:

$$LR = \frac{\Pr(E|H1)}{\Pr(E|H2)}, \quad \text{na qual:}$$

Pr = abreviação de probabilidade;

E = evidência genética, que pode ser, nesse caso, a coincidência (*match*) entre o perfil do suspeito e da amostra questionada;

H1 = hipótese de inclusão, por exemplo, a fonte da amostra questionada é o indivíduo analisado;

H2 = hipótese de exclusão, por exemplo, a fonte da amostra questionada é outra pessoa não relacionada ao indivíduo analisado.

Desse modo, considerando um caso simples de contribuidor único, e com perfil de amostra questionada idêntico ao perfil do indivíduo analisado, $\Pr(E|H1) = 1$. Já

para resolver a equação $\Pr(E|H2)$, será necessário calcular a frequência, na população, do perfil genético em questão.^(30,31)

Os Y-STR constituem um haplótipo único cujos *loci* estão em forte desequilíbrio de ligação (*Linkage Disequilibrium*, LD), desse modo não é válido calcular a frequência do haplótipo pela multiplicação da frequência alélica dos *loci* envolvidos, como é realizado para marcadores AS-STR, herdados independentemente. Desse modo, faz-se necessário obter a frequência do haplótipo Y-STR diretamente na população referência da amostra em questão. Para isso, é necessário um amplo banco de dados, representativo da população em questão e com controle de qualidade dos haplótipos depositados.^(32,33)

Nesse contexto, o banco de dados referência de haplótipos do cromossomo Y (Y-Chromosome Haplotype Reference Database – YHRD)⁽³⁴⁾ é o maior repositório forense internacional de amostras deste cromossomo. Com amostras provenientes de diversas populações das mais variadas regiões do mundo, o referido banco permite a aquisição, distribuição e interpretação de conjuntos de dados de DNA para auxiliar o processo de tomada de decisão em genética forense.⁽³⁵⁾

A região haploide de herança uniparental do cromossomo Y possui estrutura genética pronunciada devido à acentuada deriva genética ao qual está submetido. Assim, populações que se encontram em regiões diferentes ou possuem origem distinta, podem apresentar alto nível de subestrutura populacional, pois possuem alelos em frequências significativamente diferentes e algumas vezes únicos.⁽³⁶⁾

Nesse sentido, a Comissão de DNA da ISFG recomenda o uso de base de dados regionais de frequências haplotípicas. A junção de amostras provenientes de diferentes regiões é válida apenas se não houver subestrutura populacional, ou seja, diferença

estatística significativa na distribuição de haplótipos Y-STR. Caso contrário, os efeitos da subestrutura devem ser considerados nos cálculos de estimativa de frequência de haplótipos do cromossomo Y.^(29,37)

Sabe-se que quanto maior o tamanho amostral de uma população, mais discriminativo e confiável será o cálculo da estimativa de frequência de um haplótipo Y-STR.⁽³⁸⁾ Desse modo, para aplicação em genética forense no Brasil, é interessante a junção de amostras brasileiras de diferentes populações, a fim de se obter um banco de dados brasileiro amplo e representativo de haplótipos Y-STR.

No entanto, o Brasil é considerado uma das populações mais heterogêneas do mundo como consequência de cinco séculos de cruzamentos interétnicos entre os povos de três continentes: ameríndios, europeus e africanos. Originalmente povoado por milhões de ameríndios, o Brasil recebeu, nos primeiros dois séculos de colonização, cerca de 500 a cinco mil portugueses por ano, em sua maioria homens, e esses números aumentaram até 1930. A imigração de mulheres europeias durante os primeiros séculos foi insignificante. Durante o período da escravidão, quatro milhões de africanos foram enviados ao país, e posteriormente, imigrantes de diversas nacionalidades chegaram ao Brasil, principalmente portugueses e italianos, mas também espanhóis, alemães, sírios, turcos e japoneses.⁽³⁹⁾

Desse modo, faz-se necessário avaliar a estrutura da população brasileira em relação aos marcadores Y-STR, visando validar a junção de amostras de diferentes regiões a fim de obter um amplo banco de dados unificado do Brasil. Isso poderá contribuir na caracterização de frequências haplotípicas do cromossomo Y na população brasileira, com aplicação no cálculo de probabilidade de evidências em genética forense.

1.5 Objetivos

Diante do exposto, este estudo propôs os seguintes objetivos:

1. Avaliar, no Brasil, a estrutura populacional do cromossomo Y para haplótipos Y-STR, com base no maior banco de dados destes marcadores disponível atualmente; o YHRD.
2. Validar a utilização de um banco de dados unificado de haplótipos Y-STR no Brasil, com amplo número amostral, para aplicação em genética forense.
3. Analisar a ancestralidade masculina brasileira com base em haplótipos Y-STR do banco de dados YHRD.

2. MATERIAL E MÉTODOS

2. MATERIAL E MÉTODOS

Os dados utilizados neste trabalho foram obtidos no banco de dados YHRD (<http://www.yhrd.org>) e em artigos disponíveis na literatura (até junho/2018) correspondentes aos dados incluídos no banco, considerando-se a versão 56. Foram selecionadas apenas as amostras genotipadas para os 17 marcadores Y-STR incluídos no sistema AmpFLSTR® Yfiler® PCR Amplification Kit (*Yfiler*), tendo em vista sua expressão na literatura.

Foi estimada a distância genética entre pares de populações, pares de regiões geográficas e pares de metapopulações por meio das estatísticas F (F_{ST} e R_{ST}), com auxílio do programa de bioinformática Arlequin v.3.5.2.2.⁽⁴⁰⁾ Os resultados foram reproduzidos em duas dimensões com o método de Escala Multidimensional (*Multidimensional scaling* - MDS) e auxílio do programa Statistica Trial 13.0 (StatSoft Inc., <http://www.statsoft.com.br>). Os valores de P foram obtidos após 10.000 permutas, e considerados significantes após correção de Bonferroni.

A análise de estrutura hierárquica foi realizada entre populações e regiões brasileiras pelo método de Análise de Variação Molecular (*Analysis of molecular variance* - AMOVA), com o auxílio do programa de bioinformática Arlequin v.2.5.2.2.⁽⁴⁰⁾ Para todos os testes acima descritos, foi excluído o marcador DYS385a/b, subtraído o marcador DYS389I do DYS389II e removidos os alelos intermediários.

Os cálculos da diversidade haplotípica e desvio padrão foram realizados de acordo com a fórmula de Nei⁽⁴¹⁾. A capacidade de discriminação foi determinada pela razão do número de diferentes haplótipos observados pelo número total de haplótipos amostrados.

A inferência dos haplogrupos para análise de ancestralidade da população foi realizada com auxílio do programa Haplogroup Predictor (www.hprg.com/hapest5/), com aplicação de escore 40 e probabilidade 95%. ^(42,43)

3. RESULTADOS

3. RESULTADOS

Entre os 207.469 haplótipos (h) Y-STR depositados no YHRD até a versão 56 (09/04/2018), 9.706 (4,7%) haplótipos (h) corresponderam a dados do Brasil. Foram incluídas 72 populações (pop) brasileiras, classificadas em metapopulações nativo-americana (546 h de 19 pop), africana (693 h de 14 pop), europeia (1.213 h de 5 pop) e miscigenada (7.254 h de 34 pop). Todas as amostras brasileiras atenderam ao haplótipo mínimo (nove marcadores), sendo a tipagem pelo PowerPlex® Y System (Promega Corporation, Madison, WI; *PPY* - 12 marcadores) e AmpFLSTR® Yfiler® PCR Amplification Kit (Applied Biosystems, Foster City, CA; *Yfiler* - 17 marcadores) realizada, respectivamente, em 56 (8.066 h) e 53 (7.846 h) populações, enquanto apenas seis populações (996 h) foram genotipadas para o PowerPlex® Y23 System (Promega Corporation, Madison, WI; *PPY23* - 23 marcadores), uma população (258 h) para o Yfiler® Plus PCR Amplification Kit (Applied Biosystems, Foster City, CA; *Yfiler Plus* - 27 marcadores) e nenhuma para o kit Maximal (YHRD Max Loci - 29 marcadores; Quadro 1).

Quadro 1 - Lista das populações brasileiras com haplótipos Y-STR registrados no *Y-Chromosome Haplotype Reference Database* (YHRD), seus respectivos números de haplótipos tipados nos diferentes sistemas de marcadores do cromossomo Y, sua região geográfica, bem como a metapopulação à qual cada população pertence no YHRD.

População	Minimal	PPY	<i>Yfiler</i>	PPY23	<i>Yfiler Plus</i>	Maximal	Região	Meta-população
Porto Velho	139	139	139	0	0	0	N	Miscigenada
Belém	400	400	400	0	0	0	N	Miscigenada
Boa Vista	67	67	67	0	0	0	N	Miscigenada
Macapá	197	197	197	0	0	0	N	Miscigenada

Continua na próxima página

Manaus	281	281	281	204	0	0	N	Miscigenada
Palmas	30	30	30	0	0	0	N	Miscigenada
Rio Branco	31	31	31	0	0	0	N	Miscigenada
Santarém	78	78	78	0	0	0	N	Miscigenada
Bahia	205	205	205	0	0	0	NE	Miscigenada
Alagoas	247	0	0	0	0	0	NE	Miscigenada
Ceará	109	109	109	0	0	0	NE	Miscigenada
Maranhão	140	140	140	0	0	0	NE	Miscigenada
Pernambuco	30	30	30	0	0	0	NE	Miscigenada
Paraíba	300	300	300	0	0	0	NE	Miscigenada
Fortaleza	101	101	101	0	0	0	NE	Miscigenada
Brasília	502	502	502	0	0	0	CO	Miscigenada
Goiás	93	93	93	0	0	0	CO	Miscigenada
Mato Grosso	200	200	200	200	0	0	CO	Miscigenada
Mato Grosso do Sul	50	50	50	0	0	0	CO	Miscigenada
Espirito Santo	355	355	253	0	0	0	SE	Miscigenada
Minas Gerais	200	200	200	0	0	0	SE	Miscigenada
Rio de Janeiro	1685	1685	1685	198	258	0	SE	Miscigenada
Sao Paulo State	45	0	0	0	0	0	SE	Miscigenada
Sao Paulo State	50	0	0	0	0	0	SE	Miscigenada
Sao Paulo State	366	366	366	0	0	0	SE	Miscigenada
Sao Paulo	712	712	712	212	0	0	SE	Miscigenada
Paraná	77	77	77	0	0	0	S	Miscigenada
Rio Grande do Sul	243	243	243	0	0	0	S	Miscigenada
Santa Catarina North	51	51	51	0	0	0	S	Miscigenada
Santa Catarina Valley	54	54	54	0	0	0	S	Miscigenada
Santa Catarina Plateau	58	58	58	0	0	0	S	Miscigenada
Santa Catarina West	56	56	56	0	0	0	S	Miscigenada
Santa Catarina Capital	50	50	50	0	0	0	S	Miscigenada
Santa Catarina South	52	52	52	0	0	0	S	Miscigenada
Curiaú	48	0	0	0	0	0	N	Africana
Itancoã-Miri	22	0	0	0	0	0	N	Africana

Continua na próxima página

Marajó	58	0	0	0	0	0	N	Africana
Mazagão Novo	24	0	0	0	0	0	N	Africana
Mazagão Velho	36	0	0	0	0	0	N	Africana
Pitimandeuca	26	0	0	0	0	0	N	Africana
Saracura	46	0	0	0	0	0	N	Africana
Trombetas	20	0	0	0	0	0	N	Africana
Bahia	53	53	0	0	0	0	NE	Africana
Sergipe	25	25	0	0	0	0	NE	Africana
Pontal	20	0	0	0	0	0	NE	Africana
Goiás	40	40	0	0	0	0	CO	Africana
Rio de Janeiro	235	100	100	0	0	0	SE	Africana
Sao Paulo State	40	0	0	0	0	0	SE	Africana
Rio de Janeiro	245	0	0	0	0	0	SE	Europeia
Sao Paulo State	65	0	0	0	0	0	SE	Europeia
Sao Paulo	567	120	120	120	0	0	SE	Europeia
Porto Alegre	81	81	81	0	0	0	S	Europeia
Rio Grande do Sul	255	255	255	0	0	0	S	Europeia
Brazil [Yanomami]	12	0	0	0	0	0	N	Nativo-Americana
Pará [Parakanã]	38	38	38	0	0	0	N	Nativo-Americana
Amapá [Tiriyó]	35	35	35	0	0	0	N	Nativo-Americana
Amapá [Waiãpi]	13	13	13	0	0	0	N	Nativo-Americana
Pará [Zoé]	25	25	25	0	0	0	N	Nativo-Americana
São Gabriel da Cachoeira	62	62	62	62	0	0	N	Nativo-Americana
Pará [Arara]	20	20	20	0	0	0	N	Nativo-Americana
Pará [Asurini]	15	15	15	0	0	0	N	Nativo-Americana
Pará [Gaviao]	18	18	18	0	0	0	N	Nativo-Americana
Pará [Ipixuna]	21	21	21	0	0	0	N	Nativo-Americana
Rondonia [Karitiana]	34	34	34	0	0	0	N	Nativo-Americana
Pará [Kayapó]	16	14	14	0	0	0	N	Nativo-Americana
Pará [Mapuera]	5	5	5	0	0	0	N	Nativo-Americana
Roraima [Yanomami]	10	10	10	0	0	0	N	Nativo-Americana
Amazonas [Ticuna]	23	23	23	0	0	0	N	Nativo-Americana

Continua na próxima página

Rondonia [Surui]	24	24	24	0	0	0	N	Nativo-Americana
Maranhão [Awa-Guajá]	69	46	46	0	0	0	NE	Nativo-Americana
Maranhão [Urubu-Kaapor]	56	27	27	0	0	0	NE	Nativo-Americana
Mato Grosso do Sul [Terena]	50	50	50	0	0	0	CO	Nativo-Americana
Legenda: PPY: PowerPlex Y; <i>Yfiler</i> : AmpFI1STR Yfiler; PPY23: PowerPlex Y23; N: Norte; NE: Nordeste; CO: Centro-Oeste; SE: Sudeste; S: Sul.								

Com relação à distribuição regional dos haplótipos brasileiros (Figura 3), a região sudeste se apresentou como maior contribuinte (4.565 h), seguida do norte (1.874 h), nordeste (1.355 h), sul (977 h) e centro-oeste (935 h). Todas as regiões mostraram populações classificadas no grupo miscigenado, o qual correspondeu a maior porcentagem de haplótipos brasileiros no banco (74,7%), oriundos principalmente da região sudeste (47%). Apenas as regiões sul e sudeste depositaram haplótipos do grupo europeu. Por outro lado, todas as regiões do Brasil, exceto a região Sul, apresentaram registro do grupo africano. A região norte destacou-se pela contribuição com amostras do grupo nativo americano (67,9% dos haplótipos desta Metapopulação), porém observou-se que suas populações contribuíram com pequeno número amostral.

Os haplótipos Y-STR disponíveis no YHRD foram tipados para o sistema *Yfiler* (17 Y-STRs) em 81% das amostras brasileiras (7.846 h) registradas no referido banco (Figura 4, Quadro 1). Tendo em vista que este sistema possui considerável número de marcadores e poucos dados brasileiros não genotipados no YHRD, além da expressiva quantidade de artigos publicados que referenciam sua submissão no banco, foram recuperados na literatura os dados genotipados para *Yfiler* de 22 populações brasileiras miscigenadas (5.778 h, 85% dos dados totais do YHRD para este sistema), 16 populações nativo-americanas (400 h, 83% dos dados totais do YHRD para este

sistema) e duas populações europeias (375 h, 82% dos dados totais do YHRD para este sistema; Quadro 2). O Brasil possui apenas uma população de origem africana disponível no YHRD genotipada para o sistema *Yfiler*, porém esta população não foi disponibilizada na literatura, portanto não foi incluída neste trabalho.⁽⁴⁴⁻⁵⁵⁾

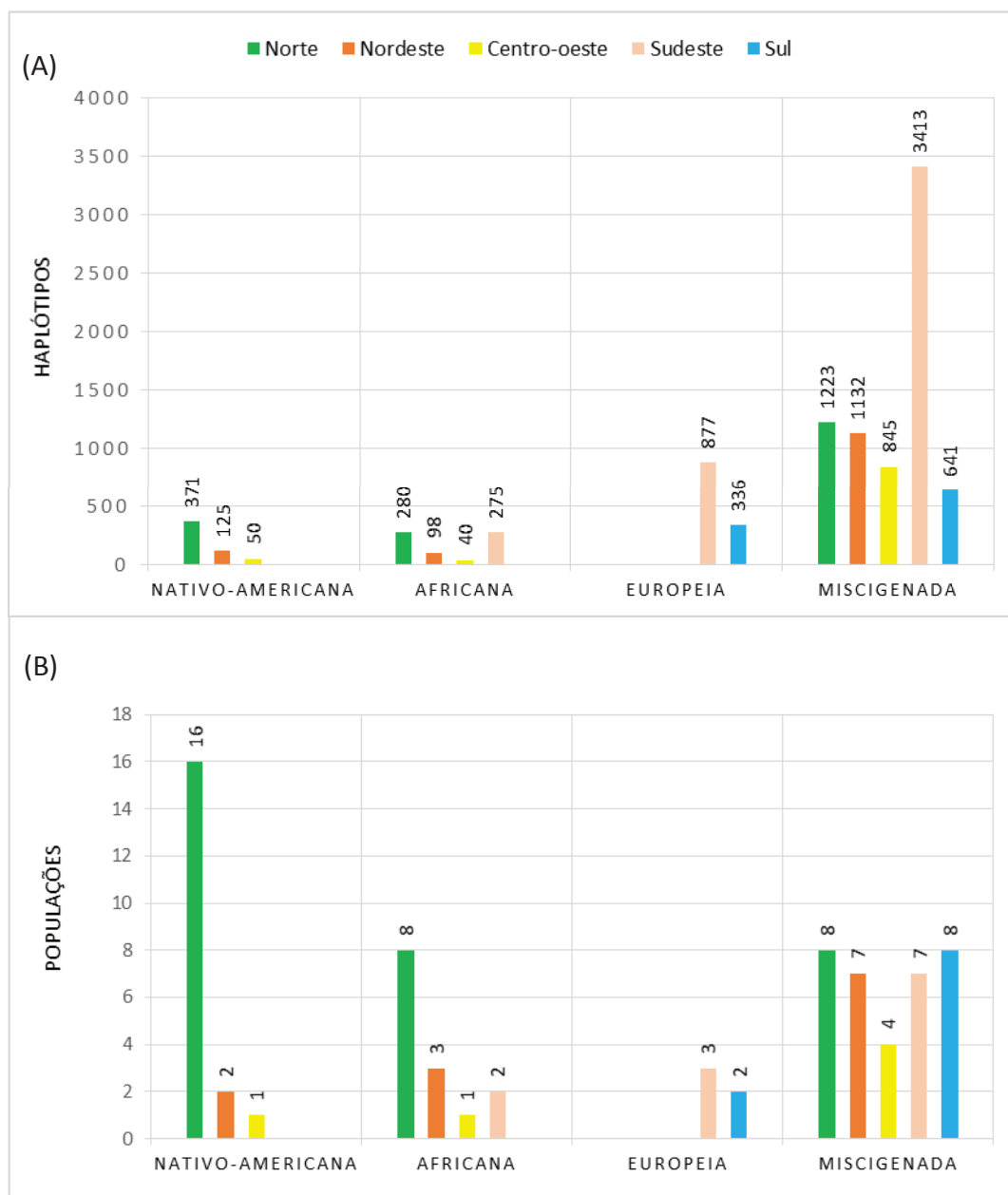


Figura 3 - Distribuição dos haplótipos Y-STR (A) e das populações (B) por região brasileira e metapopulação, considerando-se os dados do haplótipo mínimo no *Y Chromosome Haplotype Reference Database (YHRD)*.

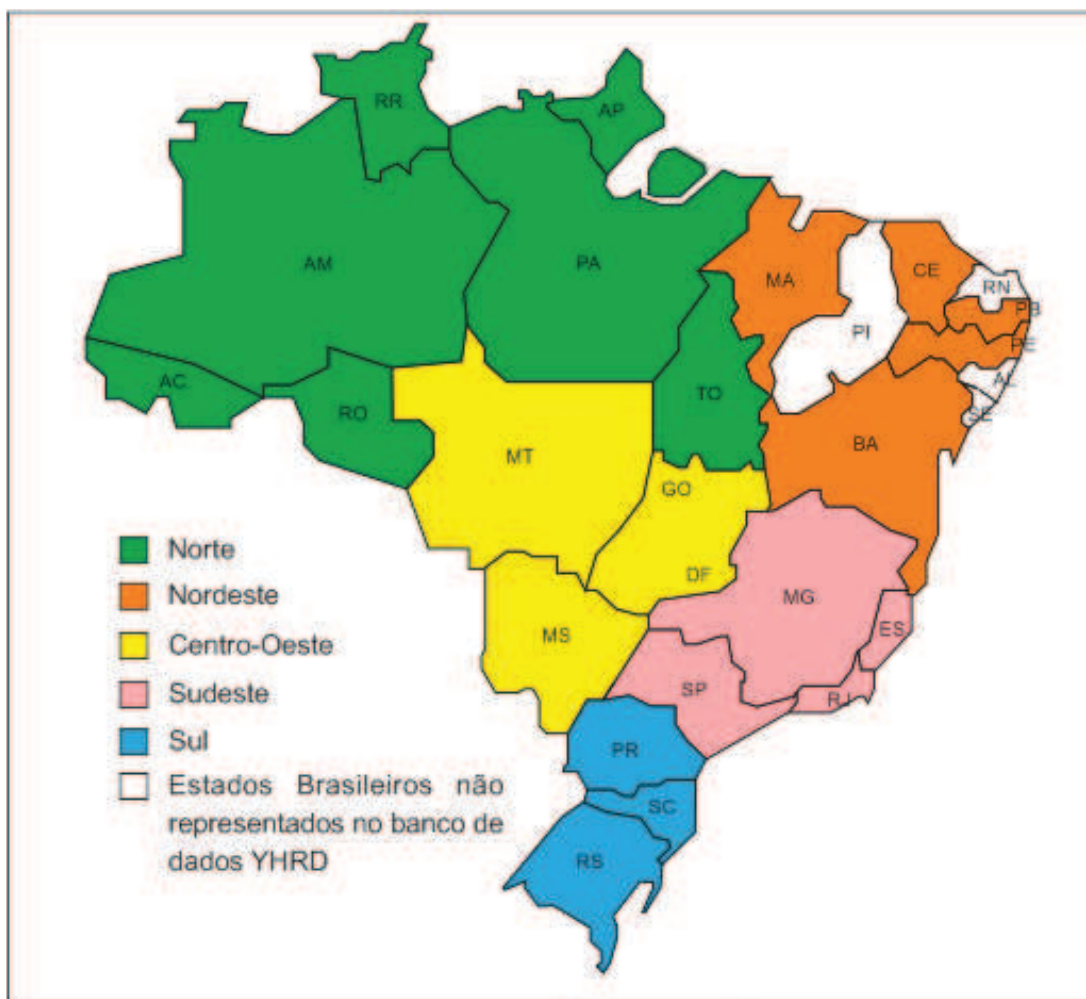


Figura 4 - Representação dos Estados brasileiros com populações disponíveis no *Y-Chromosome Haplotype Reference Database* (YHRD), considerando-se o Sistema *Yfiler*.

Quadro 2 - Populações brasileiras genotipadas para o sistema *Yfiler*, e suas respectivas regiões geográficas, metapopulações, número de haplótipos e referência na literatura.

Metapopulação	Região	População	Haplótipos	Referência
Miscigenada	Sudeste	São Paulo	500 212	Cardena et al., 2016 Disponibilizado pelo autor
		Estado de São Paulo	208 158	Palha et al., 2012 Souza et al., 2018
		Rio de Janeiro	618 850	Disponibilizado pelo autor Jannuzzi et al, 2018
		Espírito Santo	253	Figueiredo et al., 2016
		Minas Gerais	200	Palha et al., 2012

Continua na próxima página.

Miscigenada	Sul	Santa Catarina	324	Torres, 2014
		Rio Grande do Sul	243	Palha et al., 2012
		Paraná	77	Palha et al., 2012
	Centro-Oeste	Brasília	300	Chemale et al., 2014
			201	Barra et al., 2014
		Goiás	93	Palha et al., 2012
		Mato Grosso do Sul	50	Palha et al., 2012
		Mato Grosso	200	Disponibilizado pelo autor
	Norte	Belém	400	Palha et al., 2012
		Rio Branco	31	Palha et al., 2012
		Macapá	138	Francez et al., 2012
	Nordeste	Porto Velho	65	Palha et al., 2012
		Boa Vista	135	Palha et al., 2012
		Santarém	67	Palha et al., 2012
		Palmas	72	Palha et al., 2012
		Palmas	30	Palha et al., 2012
	Nordeste	Manaus	74	Palha et al., 2012
Ceará		109	Palha et al., 2012	
Pernambuco		30	Palha et al., 2012	
Nativo-Americana	Maranhão	140	Palha et al., 2012	
	Pará (Arara)	20	Roewer et al., 2013	
	Pará (Asurini)	15	Roewer et al., 2013	
	Pará (Gavião)	18	Roewer et al., 2013	
	Pará (Ipixuna)	21	Roewer et al., 2013	
	Pará (Kayapó)	17	Roewer et al., 2013	
	Pará (Mapuera)	5	Roewer et al., 2013	
	Pará (Parakanã)	38	Roewer et al., 2013	
	Pará (Zoé)	25	Roewer et al., 2013	
	Maranhão (Awa)	46	Roewer et al., 2013	
	Maranhão (Urubu)	27	Roewer et al., 2013	
	Mato Grosso do Sul (Terena)	32	Roewer et al., 2013	
	Amapá (Tiryiό)	35	Roewer et al., 2013	
	Amapá (Waiãpi)	13	Roewer et al., 2013	
	Roraima (Yanomami)	10	Roewer et al., 2013	
	Rondonia (Karitiana)	17	Roewer et al., 2013	
	São Gabriel da Cachoeira	61	Purps et al., 2014	
Europeia	Rio Grande do Sul	255	Schwengber et al., 2009	
	São Paulo	120	Purps et al., 2014	

As populações miscigenadas, separadas por região geográfica, foram submetidas à análise de distância genética entre pares de regiões brasileiras, bem como entre pares

de tais regiões e as metapopulações nativo-americana e europeia (Tabela 1). Foram realizados os cálculos utilizando-se os parâmetros F_{ST} e R_{ST} . As análises F_{ST} entre as regiões brasileiras de populações miscigenadas mostraram valores P significantes apenas entre pares envolvendo a região norte, sendo o maior valor de distância genética aquele entre as regiões norte e nordeste. Para R_{ST} , o mesmo ocorreu para a região norte, com significância em todas as comparações com as demais regiões miscigenadas do Brasil, porém, houve significância também em comparações entre outras regiões (sul-sudeste; nordeste-sul; nordeste-centro oeste). Apesar de terem apresentado valores P significantes, os valores de distância genética foram baixos em todas as análises envolvendo populações miscigenadas.

Em relação à metapopulação europeia, as comparações com populações brasileiras miscigenadas separadas por região geográfica mostrou significância apenas para a região norte, tanto na análise F_{ST} (0,00455; $P=0,00030$) quanto R_{ST} (0,01195; $P=0,0000$), a qual também apresentou o maior valor de distância genética, embora os valores ainda sejam considerados baixos. Já nas comparações envolvendo a metapopulação nativo-americana, houve diferença significativa entre todas as regiões geográficas destacando-se, ainda, altos valores de distância genética, superiores a 0,14244 para F_{ST} e 0,17595 para R_{ST} .

Tabela 1 - Matriz de distância genética entre pares de regiões geográficas representadas por populações brasileiras do grupo miscigenado e entre tais regiões e a metapopulação nativo-americana e europeia, com valores de p (acima da diagonal) e valores de distância genética (F_{ST} ou R_{ST} ; abaixo da diagonal).

F_{ST}	Sudeste	Sul	Centro-Oeste	Norte	Nordeste	Nativo Americano	Europeia
Sudeste	-	0.01148	0.00297	0.00000*	0.08415	0.00000*	0.35947
Sul	0.00113	-	0.19949	0.00109*	0.03119	0.00000*	0.06762
Centro Oeste	0.00130	0.00031	-	0.00139*	0.01832	0.00000*	0.09623
Norte	0.00357	0.00240	0.00194	-	0.00050*	0.00000*	0.00030*
Nordeste	0.00099	0.00201	0.00239	0.00510	-	0.00000*	0.30086
Nativo Americano	0.14244	0.15374	0.15800	0.14510	0.15713	-	0.00000*
Europeia	0.00005	0.00115	0.00087	0.00455	0.00032	0.15881	-
R_{ST}							
Sudeste	-	0.00010*	0.01129	0.00000*	0.02792	0.00000*	0.12929
Sul	0.00380	-	0.12276	0.00158*	0.00030*	0.00000*	0.00337
Centro Oeste	0.00153	0.00076	-	0.00129*	0.00099*	0.00000*	0.02119
Norte	0.00952	0.00339	0.00320	-	0.00000*	0.00000*	0.00000*
Nordeste	0.00266	0.00810	0.00728	0.01927	-	0.00000*	0.05485
Nativo Americano	0.19632	0.17595	0.20218	0.20125	0.26164	-	0.00000*
Europeia	0.00087	0.00475	0.00309	0.01195	0.00308	0.23631	-

*Valores de P considerados significantes após correção de Bonferroni: $\leq 0,0024$.

A análise de Escala Multidimensional (MDS) foi realizada com base nos valores F_{ST} obtidos nas análises de distância genética entre pares de regiões geográficas e entre tais regiões e as metapopulações nativo-americana e europeia (Figuras 5). Observou-se que a metapopulação nativo-americana apresentou a maior distância em relação às populações miscigenadas separadas por regiões geográficas e à metapopulação europeia. Entre as populações miscigenadas, a região Norte apresentou a maior distância em relação às demais.

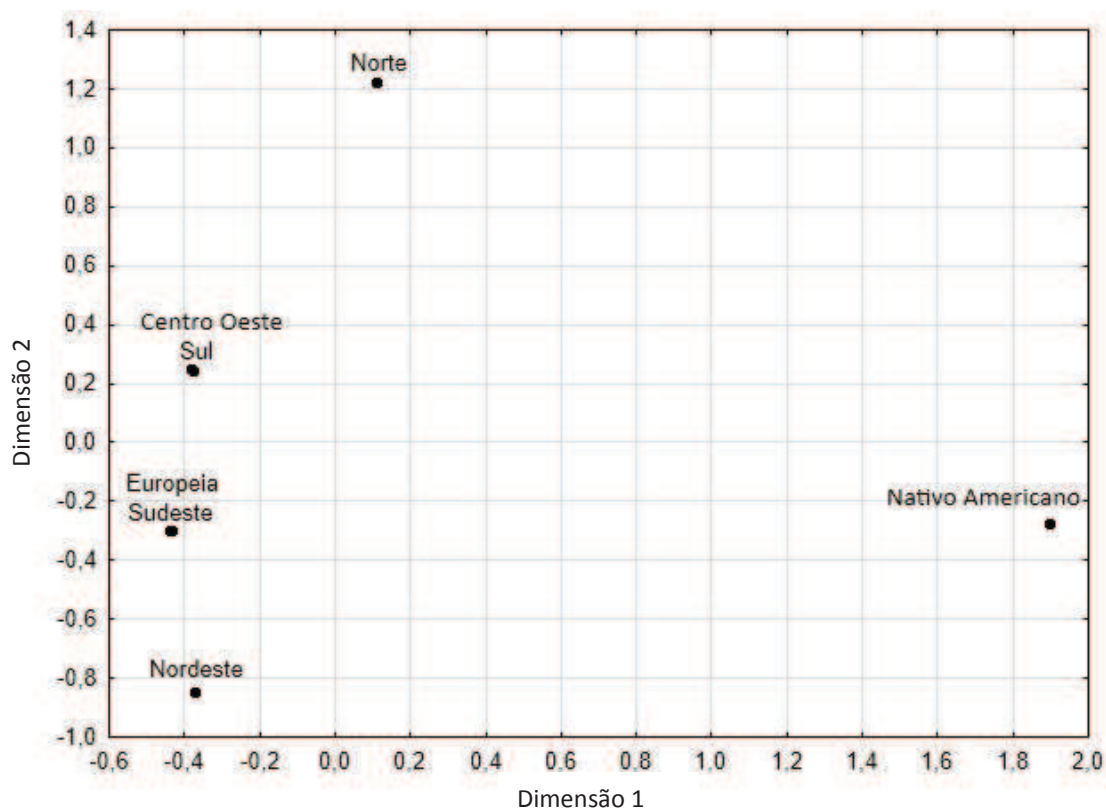


Figura 5 - Análise em Escala Multidimensional (MDS) baseada na matriz de distância genética F_{ST} entre pares de regiões geográficas, bem como entre regiões e metapopulações brasileiras genotipadas para o sistema *Yfiler* e disponíveis no banco de dados *Y-Chromosome Haplotype Reference Database* (YHRD; Tabela 1). Stress: 0,0002.

Diante do exposto, realizou-se a análise de distância genética entre as populações miscigenadas, assim como nas populações nativo-americanas, a fim de identificar a respectiva contribuição por região (Tabelas 2 a 5). Pode-se observar que, das 231 análises entre pares de populações miscigenadas, apenas 21 (9,1%) (F_{ST}) e 22 (9,5%) (R_{ST}) apresentaram valor-P significativo, das quais 18 (85,8%) (F_{ST}) e 15 (68,2%) (R_{ST}) envolveram as populações de Belém (PA) e Boa Vista (RR), ambas localizadas na região norte do Brasil. Além disso, os maiores valores de distância genética foram observados principalmente na população de Boa Vista, no entanto estes valores são considerados de baixa diferenciação genética (Tabelas 2 e 3).

Tabela 2 - Matriz de distância genética entre pares de populações brasileiras do grupo miscigenado, com valores de p (acima da diagonal) e valores de distância genética (F_{ST} ; abaixo da diagonal).

Pop	(SP)	(RJ)	(MG)	(ES)	(SC)	(RS)	(PR)	(DF)	(GO)	(MS)	(MT)	(PA-1)	(AC)	(AP)	(RO)	(RR)	(PA-2)	(TO)	(AM)	(CE)	(PE)	(MA)
(SP)	-	0.09445	0.01950	0.41253	0.19107	0.00198	0.58212	0.04831	0.00238	0.99040	0.04406	0.00000*	0.59172	0.00406	0.24176	0.00020*	0.07059	0.73369	0.01535	0.74577	0.14979	0.00812
(RJ)	0.00037	-	0.01079	0.69567	0.03099	0.00119	0.85180	0.00881	0.00010*	0.95832	0.07217	0.00000*	0.39897	0.00030	0.17870	0.00000*	0.06881	0.83903	0.01109	0.63914	0.14553	0.00208
(MG)	0.00279	0.00351	-	0.05405	0.02683	0.02990	0.15731	0.02703	0.10662	0.91823	0.01861	0.00248	0.90635	0.03821	0.04366	0.01148	0.02455	0.27532	0.03465	0.06277	0.26087	0.00208
(ES)	0.00000	-0.00050	0.00320	-	0.60024	0.01891	0.54787	0.22552	0.00139	0.95832	0.21305	0.00000*	0.78873	0.02168	0.50351	0.00000*	0.14058	0.46530	0.02119	0.64835	0.21087	0.00624
(SC)	0.00049	0.00160	0.00351	-0.00052	-	0.03148	0.26225	0.53717	0.00228	0.99069	0.06702	0.00000*	0.63370	0.07821	0.08356	0.00020*	0.06861	0.36462	0.02000	0.23196	0.18087	0.02237
(RS)	0.00451	0.00533	0.00375	0.00409	0.00281	-	0.03762	0.04188	0.00406	0.80121	0.02940	0.00000*	0.68280	0.09504	0.00931	0.00683	0.02119	0.07821	0.12425	0.04069	0.43788	0.00000*
(PR)	-0.00086	-0.00229	0.00291	-0.00097	0.00115	0.00628	-	0.24750	0.00663	0.86288	0.40600	0.00218	0.32700	0.04267	0.27631	0.00089	0.05762	0.95931	0.03524	0.67528	0.25394	0.14603
(DF)	0.00102	0.00183	0.00322	0.00060	-0.00023	0.00222	0.00122	-	0.00198	0.99891	0.35006	0.00000*	0.71349	0.10266	0.07871	0.00030	0.09455	0.39491	0.09692	0.21909	0.23859	0.00218
(GO)	0.00910	0.01347	0.00326	0.01337	0.01106	0.00907	0.01328	0.01111	-	0.19780	0.00228	0.05643	0.19820	0.02148	0.00554	0.06287	0.00109	0.03663	0.00416	0.00693	0.08861	0.00327
(MS)	-0.00579	-0.00493	-0.00523	-0.00568	-0.00651	-0.00329	-0.00563	-0.00708	0.00347	-	0.92654	0.12553	0.89526	0.89714	0.82546	0.05267	0.49104	0.79160	0.36986	0.94595	0.63855	0.33393
(MT)	0.00211	0.00160	0.00514	0.00093	0.00229	0.00376	-0.00001	0.00015	0.01289	-0.00522	-	0.00000*	0.25146	0.15177	0.19513	0.00040	0.10306	0.71468	0.04277	0.31383	0.14504	0.00198
(PA-1)	0.01051	0.01306	0.00592	0.01270	0.01165	0.00945	0.01240	0.01192	0.00382	0.00414	0.01243	-	0.20246	0.00020*	0.00010*	0.00020*	0.00040	0.08069	0.00059	0.00050	0.10375	0.00000*
(AC)	-0.00208	0.00020	-0.00708	-0.00498	-0.00275	-0.00328	0.00186	-0.00381	0.00465	-0.00970	0.00282	0.00379	-	0.37838	0.39402	0.07326	0.52817	0.22037	0.64964	0.57163	0.50045	0.05069
(AP)	0.00417	0.00608	0.00378	0.00459	0.00204	0.00196	0.00655	0.00155	0.00691	-0.00484	0.00159	0.00888	0.00042	-	0.01554	0.00366	0.02930	0.11553	0.03802	0.02247	0.20978	0.00465
(RO)	0.00069	0.00108	0.00433	-0.00043	0.00240	0.00647	0.00134	0.00244	0.01063	-0.00408	0.00141	0.01305	0.00030	0.00625	-	0.00010*	0.22552	0.58588	0.03297	0.72349	0.05702	0.01148
(RR)	0.01975	0.02354	0.01122	0.02415	0.02154	0.01117	0.02381	0.02155	0.00617	0.01036	0.02115	0.01648	0.01124	0.01458	0.02265	-	0.00020*	0.00792	0.00129	0.00079	0.01663	0.00000*
(PA-2)	0.00386	0.00421	0.00838	0.00300	0.00456	0.00789	0.00777	0.00365	0.01946	-0.00103	0.00410	0.01536	-0.00200	0.00780	0.00198	0.03294	-	0.31027	0.54826	0.22087	0.12949	0.00198
(TO)	-0.00375	-0.00522	0.00249	-0.00092	0.00074	0.00878	-0.01016	0.00027	0.01347	-0.00737	-0.00417	0.00873	0.00634	0.00750	-0.00249	0.02592	0.00238	-	0.08098	0.67766	0.32502	0.36699
(AM)	0.00720	0.00829	0.00707	0.00812	0.00766	0.00307	0.00990	0.00338	0.01514	0.00090	0.00629	0.01531	-0.00355	0.00698	0.00778	0.02176	-0.00122	0.01079	-	0.03336	0.52995	0.00010*
(CE)	-0.00118	-0.00086	0.00433	-0.00119	0.00110	0.00473	-0.00219	0.00115	0.01244	-0.00650	0.00057	0.01298	-0.00252	0.00656	-0.00175	0.02245	0.00214	-0.00414	0.00878	-	0.15474	0.03425
(PE)	0.00515	0.00573	0.00268	0.00407	0.00472	-0.00026	0.00355	0.00297	0.00942	-0.00457	0.00604	0.00718	-0.00226	0.00399	0.01170	0.02095	0.00843	0.00268	-0.00175	0.00673	-	0.03346
(MA)	0.00487	0.00713	0.01158	0.00783	0.00457	0.01509	0.00314	0.00817	0.01251	0.00090	0.01110	0.01791	0.01201	0.00903	0.00799	0.02560	0.01767	0.00071	0.02204	0.00630	0.01517	-

* Valores de P considerados significantes após correção de Bonferroni: $\leq 0,00022$. Pop = Populações. Populações consideradas na análise: (SP) = São Paulo; (RJ) = Rio de Janeiro; (MG) = Minas Gerais; (SC) = Santa Catarina; (RS) = Rio Grande do Sul; (PR) = Paraná; (DF) = Brasília; (GO) = Goiás; (MS) = Mato Grosso do Sul; (MT) = Mato Grosso; (PA-1) = Belém; (AC) = Rio Branco; (AP) = Macapá; (RO) = Porto Velho; (RR) = Boa Vista; (PA-2) = Santarém; (TO) = Palmas; (AM) = Manaus; (CE) = Ceará; (PE) = Pernambuco; (MA) = Maranhão.

Tabela 3 - Matriz de distância genética entre pares de populações brasileiras do grupo miscigenado, com valores de p (acima da diagonal) e valores de distância genética (R_{ST} ; abaixo da diagonal).

Pop	(SP)	(RJ)	(MG)	(ES)	(SC)	(RS)	(PR)	(DF)	(GO)	(MS)	(MT)	(PA-1)	(AC)	(AP)	(RO)	(RR)	(PA-2)	(TO)	(AM)	(CE)	(PE)	(MA)
(SP)	-	0.02406	0.00743	0.49371	0.01554	0.00000*	0.82041	0.12088	0.00386	0.99663	0.67162	0.00000*	0.55470	0.01426	0.53203	0.00020*	0.02624	0.16018	0.01257	0.20404	0.20147	0.00267
(RJ)	0.00122	-	0.00069	0.24918	0.00040	0.00000*	0.94842	0.00208	0.00030	0.96674	0.47827	0.00000*	0.29027	0.00069	0.22107	0.00010*	0.02416	0.37056	0.00614	0.65677	0.13593	0.00723
(MG)	0.00649	0.01073	-	0.00960	0.01119	0.08811	0.12118	0.03841	0.04445	0.33422	0.11900	0.00257	0.94862	0.01475	0.01346	0.01841	0.02079	0.10643	0.10138	0.01851	0.60598	0.00010*
(ES)	-0.00028	0.00057	0.00957	-	0.11553	0.00188	0.70349	0.10266	0.00812	0.98842	0.29245	0.00000*	0.67389	0.02020	0.65469	0.00030	0.03604	0.10751	0.02010	0.69389	0.19147	0.00891
(SC)	0.00319	0.00717	0.00645	0.00222	-	0.00218	0.38442	0.28126	0.05198	0.83784	0.18701	0.00000*	0.66488	0.18028	0.18325	0.01851	0.15226	0.25295	0.23364	0.07415	0.35759	0.01396
(RS)	0.01166	0.01664	0.00341	0.01174	0.00781	-	0.02020	0.00149	0.00307	0.15890	0.01505	0.00030	0.91110	0.00446	0.00129	0.05415	0.00515	0.00842	0.03713	0.00297	0.37204	0.00000*
(PR)	-0.00318	-0.00457	0.00571	-0.00297	-0.00013	0.01324	-	0.34195	0.02247	0.89813	0.90328	0.00158	0.39679	0.04564	0.59945	0.00257	0.02099	0.53143	0.01327	0.98109	0.28522	0.54589
(DF)	0.00089	0.00421	0.00457	0.00204	0.00051	0.00916	0.00061	-	0.00604	0.90328	0.47183	0.00000*	0.65152	0.12860	0.27671	0.00030	0.06653	0.12454	0.15236	0.05742	0.15781	0.00050
(GO)	0.01422	0.02499	0.00871	0.01538	0.00794	0.01843	0.01625	0.01400	-	0.10573	0.00723	0.01010	0.60222	0.06188	0.00871	0.00069	0.00178	0.01198	0.01069	0.00238	0.13108	0.00040
(MS)	-0.00886	-0.00747	0.00107	-0.00962	-0.00634	0.00578	-0.00949	-0.00655	0.01013	-	0.90298	0.00644	0.77705	0.42669	0.96644	0.01861	0.44213	0.42055	0.23681	0.83210	0.35947	0.19632
(MT)	-0.00090	-0.00032	0.00318	0.00075	0.00166	0.00775	-0.00514	-0.00037	0.01667	-0.00721	-	0.00000*	0.52292	0.06999	0.26067	0.00178	0.08524	0.29918	0.05049	0.29264	0.25691	0.00931
(PA-1)	0.02452	0.03274	0.00998	0.02947	0.01825	0.01340	0.02417	0.01897	0.01181	0.02374	0.01983	-	0.21711	0.00010*	0.00010*	0.00188	0.00000*	0.00416	0.00069	0.00000*	0.02990	0.00000*
(AC)	-0.00285	0.00225	-0.01189	-0.00575	-0.00751	-0.01029	-0.00005	-0.00513	-0.00475	-0.01122	-0.00283	0.00537	-	0.49639	0.32878	0.20018	0.24334	0.09920	0.62469	0.34947	0.60539	0.02485
(AP)	0.00489	0.01261	0.00851	0.00698	0.00176	0.01059	0.00998	0.00194	0.00701	-0.00050	0.00445	0.01506	-0.00223	-	0.04633	0.00099	0.04762	0.01287	0.10732	0.00317	0.07791	0.00000*
(RO)	-0.00058	0.00124	0.01106	-0.00156	0.00197	0.01638	-0.00228	0.00087	0.01719	-0.00950	0.00136	0.02537	0.00224	0.00641	-	0.00020*	0.03544	0.16097	0.01634	0.27413	0.08920	0.02742
(RR)	0.03773	0.04288	0.01632	0.03851	0.02140	0.00913	0.03582	0.03163	0.04407	0.02823	0.02910	0.02296	0.00780	0.03187	0.04206	-	0.00139	0.00743	0.00653	0.00040	0.19048	0.00000*
(PA-2)	0.00983	0.01081	0.01409	0.01051	0.00334	0.01922	0.01825	0.00717	0.03067	-0.00081	0.00752	0.03858	0.00561	0.00951	0.01199	0.03799	-	0.07484	0.66389	0.01059	0.06613	0.00020*
(TO)	0.00761	0.00054	0.01349	0.01197	0.00300	0.03575	-0.00373	0.01072	0.03591	-0.00105	0.00300	0.03992	0.02389	0.03068	0.00881	0.04833	0.01775	-	0.02416	0.25710	0.16860	0.24057
(AM)	0.01194	0.01730	0.00650	0.01319	0.00167	0.00994	0.02256	0.00371	0.02029	0.00422	0.00973	0.02700	-0.00534	0.00590	0.01639	0.02723	-0.00373	0.03170	-	0.00693	0.11276	0.00010*
(CE)	0.00164	-0.00153	0.01210	-0.00219	0.00572	0.01859	-0.00802	0.00559	0.02568	-0.00717	0.00112	0.03789	0.00122	0.01797	0.00159	0.04308	0.02131	0.00439	0.02497	-	0.16652	0.08078
(PE)	0.00568	0.00942	-0.00453	0.00675	-0.00147	0.00087	0.00373	0.00889	0.01144	0.00151	0.00435	0.02221	-0.00721	0.01506	0.01406	0.00866	0.01825	0.01423	0.01357	0.00905	-	0.04604
(MA)	0.01021	0.00886	0.02688	0.01084	0.00860	0.04061	-0.00161	0.01555	0.02859	0.00506	0.01200	0.04358	0.02494	0.02509	0.00893	0.06780	0.03505	0.00521	0.04362	0.00636	0.01995	-

* Valores de P considerados significantes após correção de Bonferroni: $\leq 0,00022$. Pop = Populações. Populações consideradas na análise: (SP) = São Paulo; (RJ) = Rio de Janeiro; (MG) = Minas Gerais; (SC) = Santa Catarina; (RS) = Rio Grande do Sul; (PR) = Paraná; (DF) = Brasília; (GO) = Goiás; (MS) = Mato Grosso do Sul; (MT) = Mato Grosso; (PA-1) = Belém; (AC) = Rio Branco; (AP) = Macapá; (RO) = Porto Velho; (RR) = Boa Vista; (PA-2) = Santarém; (TO) = Palmas; (AM) = Manaus; (CE) = Ceará; (PE) = Pernambuco; (MA) = Maranhão.

Em relação às populações nativo-americanas, das 120 análises entre pares de populações, 107 (89,2%) e 97 (80,8%) (F_{ST} e R_{ST} , respectivamente) apresentaram valor P significativo após correção de Bonferroni, assim como valores de distância genética muito acima dos observados entre as populações miscigenadas brasileiras (Tabelas 4 e 5). A análise de distância genética mostrou semelhança entre as duas populações europeias (F_{ST} de 0,00079 e $P=0,26879$; R_{ST} de 0,00138 e $P=0,25403$).

Foi realizada análise de Escala Multidimensional (MDS) com base nos valores F_{ST} obtidos nas análises de distância genética entre pares populações miscigenadas e entre pares de populações nativo-americanas (Figuras 6). Entre as populações miscigenadas, observou-se discreta distância genética das populações de Belém-PA e Boa Vista-RR. Entre as populações nativo-americanas, maior distância genética pode ser observada em algumas tribos indígenas dos estados do Pará, Maranhão e Amapá.

Tabela 4 - Matriz de distância genética entre pares de populações brasileiras do grupo nativo americano, com valores de p (acima da diagonal) e valores de distância genética (F_{ST} ; abaixo da diagonal).

Pop	(AP) Tiriyo	(PA) Kayapo	(PA) Arara	(PA) (Zoe)	(PA) (Mapuera)	(PA) (Parakanã)	(AP) (Waiapi)	(MA) (Urubu)	(MA) (Awa)	(PA) (Asurini)	(PA) (Gavião)	(MS) (Terena)	(RR) (Yanomami)	(RO) (Karitiana)	(PA) (Ipixuna)	(AM) São Gabriel da Cachoeira
(AP) - Tiriyo	-	0.00050	0.00000*	0.00000*	0.04851	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(PA) - Kayapo	0.12923	-	0.00020*	0.00000*	0.06663	0.00000*	0.00000*	0.00000*	0.00000*	0.00050	0.00099	0.00000*	0.00257	0.00010*	0.00000*	0.00000*
(PA) - Arara	0.20618	0.24853	-	0.00000*	0.07197	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00079	0.00000*	0.00000*	0.00000*
(PA) - Zoe	0.42964	0.63963	0.63173	-	0.00010*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(PA) - Mapuera	0.09046	0.14375	0.13710	0.76235	-	0.00000*	0.00010*	0.00000*	0.00000*	0.01396	0.00168	0.00119	0.00188	0.00010*	0.00000*	0.20196
(PA) - Parakanã	0.50406	0.66956	0.58854	0.94524	0.78448	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(AM) - Waiapi	0.49449	0.62119	0.52862	0.97117	0.70845	0.95277	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(MA) - Urubu	0.36127	0.49053	0.51833	0.72218	0.57935	0.75522	0.86178	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(MA) - Awa	0.44817	0.59150	0.63338	0.64370	0.64193	0.82558	0.85065	0.65286	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(PA) - Asurini	0.19701	0.19314	0.28700	0.54048	0.17189	0.71439	0.69948	0.39332	0.44751	-	0.00030*	0.00000*	0.00000*	0.00030*	0.00000*	0.00000*
(PA) - Gavião	0.26117	0.20791	0.38223	0.57025	0.28419	0.77523	0.72127	0.52461	0.54238	0.16909	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(MS) - Terena	0.35658	0.39809	0.22995	0.73476	0.30411	0.73327	0.61505	0.65146	0.67826	0.40949	0.49698	-	0.00000*	0.00000*	0.00000*	0.00000*
(RR) - Yanomami	0.23251	0.22198	0.30061	0.80019	0.29232	0.79530	0.73526	0.65172	0.71449	0.34673	0.42946	0.37364	-	0.00000*	0.00000*	0.00000*
(RO) - Karitiana	0.30194	0.34217	0.38685	0.66532	0.38593	0.80789	0.74516	0.53797	0.64174	0.20074	0.29448	0.52632	0.36956	-	0.00000*	0.00000*
(PA) - Ipixuna	0.47318	0.63773	0.65196	0.95696	0.76122	0.95452	0.98534	0.81287	0.74230	0.49350	0.58386	0.70843	0.75600	0.63718	-	0.00000*
(AM) - São Gabriel da Cachoeira	0.17142	0.14806	0.12461	0.38140	0.02665	0.42911	0.38689	0.33181	0.40859	0.13919	0.16787	0.21555	0.22092	0.21169	0.39991	-

* Valores de P considerados significantes após correção de Bonferroni: $\leq 0,00042$. Pop = Populações; (AP) = Amapá; (PA) = Pará; (MA) = Maranhão; (MS) = Mato Grosso do Sul; (RR) = Roraima; (RO) = Rondônia; (AM) – Amazonas.

Tabela 5 - Matriz de distância genética entre pares de populações brasileiras do grupo nativo americano, com valores de p (acima da diagonal) e valores de distância genética (R_{ST} ; abaixo da diagonal).

Pop	(AP) Tiriyo	(PA) Kayapo	(PA) Arara	(PA) (Zoe)	(PA) (Mapuera)	(PA) (Parakana)	(AP) (Waiapi)	(MA) (Urubu)	(MA) (Awa)	(PA) (Asurini)	(PA) (Gavião)	(MS) (Terena)	(RR) (Yanomami)	(RO) (Karitiana)	(PA) (Ipixuna)	(AM) São Gabriel da Cachoeira
(AP) - Tiriyo	-	0.05158	0.01277	0.00000*	0.45827	0.00000*	0.00000*	0.00000*	0.00000*	0.00079	0.00010*	0.00000*	0.04534	0.00010*	0.00000*	0.00000*
(PA) - Kayapo	0.06474	-	0.00129	0.00000*	0.10128	0.00000*	0.00000*	0.00000*	0.00000*	0.00010*	0.00010*	0.00000*	0.03277	0.00000*	0.00000*	0.00050
(PA) - Arara	0.08807	0.21786	-	0.00000*	0.28423	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00010*	0.01841	0.00000*	0.00000*	0.00119
(PA) - Zoe	0.31116	0.65623	0.46579	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(PA) - Mapuera	-0.01196	0.11527	0.02986	0.75323	-	0.00000*	0.00040*	0.00010*	0.00000*	0.00079	0.00089	0.00960	0.14236	0.00030*	0.00000*	0.04495
(PA) - Parakanã	0.39072	0.69412	0.55569	0.96156	0.77731	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(AP) - Waiapi	0.38452	0.60578	0.52349	0.97920	0.73494	0.97004	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(MA) - Urubu	0.35626	0.51363	0.47285	0.79003	0.68436	0.82654	0.87618	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(MA) - Awa	0.48892	0.66913	0.65838	0.82751	0.74233	0.83043	0.89694	0.69974	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(PA) - Asurini	0.17783	0.26533	0.29594	0.60653	0.30595	0.61913	0.74459	0.35486	0.40998	-	0.02822	0.00000*	0.00000*	0.00020*	0.00000*	0.02831
(PA) - Gavião	0.23663	0.28159	0.31498	0.39477	0.33134	0.68782	0.69628	0.31369	0.44851	0.09008	-	0.00000*	0.00010*	0.00446	0.00000*	0.00099
(MS) - Terena	0.22195	0.31300	0.21505	0.66930	0.27830	0.68181	0.52883	0.63397	0.71140	0.46322	0.47587	-	0.00000*	0.00000*	0.00000*	0.00000*
(RR) - Yanomami	0.08354	0.12379	0.15294	0.71571	0.08836	0.75738	0.60270	0.64794	0.77487	0.37054	0.38207	0.30433	-	0.00000*	0.00000*	0.00139
(RO) - Karitiana	0.25168	0.40987	0.35393	0.66523	0.49112	0.83610	0.85340	0.45286	0.71141	0.20688	0.13845	0.56958	0.44202	-	0.00000*	0.00158
(PA) - Ipixuna	0.21766	0.48733	0.45442	0.95696	0.65231	0.95761	0.98983	0.77278	0.80247	0.47302	0.44052	0.64665	0.54379	0.61169	-	0.00000*
(AM) - São Gabriel da Cachoeira	0.15112	0.13996	0.08888	0.15416	0.11330	0.26725	0.32047	0.14721	0.25953	0.05234	0.09750	0.20595	0.14942	0.10484	0.17264	-

* Valores de P considerados significantes após correção de Bonferroni: $\leq 0,00042$. Pop = Populações; (AP) = Amapá; (PA) = Pará; (MA) = Maranhão; (MS) = Mato Grosso do Sul; (RR) = Roraima; (RO) = Rondônia; (AM) – Amazonas.

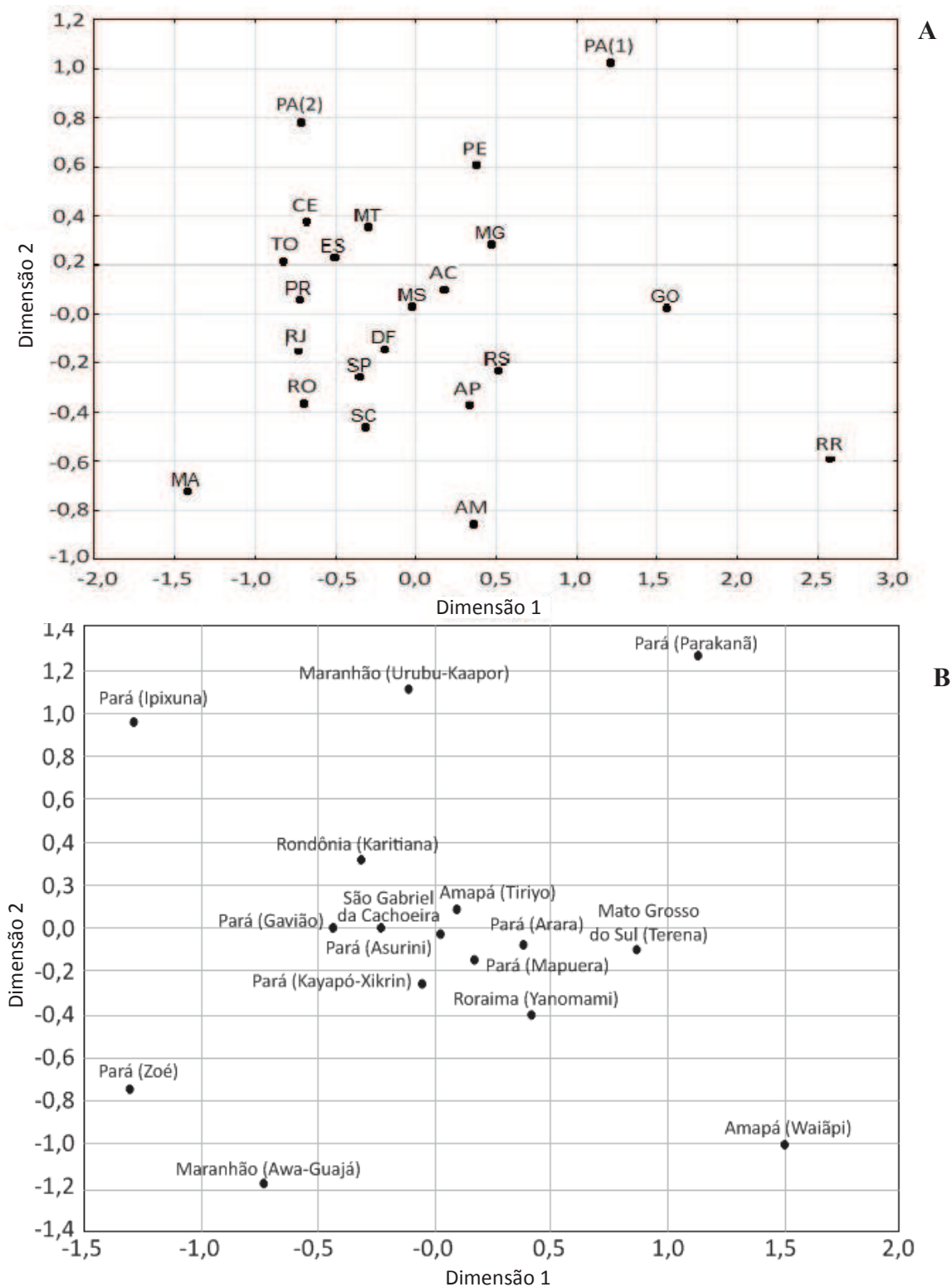


Figura 6 - Análise em Escala Multidimensional (MDS) baseada na matriz de distância genética F_{ST} entre pares de populações brasileiras miscigenadas (A) e nativo-americanas (B) genotipadas para o sistema *Yfiler* e disponíveis no banco de dados *Y-Chromosome Haplotype Reference Database* (YHRD; Tabelas 2 e 4). Stress: (A)=0,1994; (B)=0,3385.

Diante do observado, foi realizada nova análise de distância genética entre pares de regiões geográficas e entre tais regiões e as metapopulações nativo-americana e europeia, excluindo-se as populações de Belém-PA e Boa Vista-RR, a fim de verificar se a ausência de tais populações reduziria os valores de distância genética nas análises envolvendo a região norte do Brasil (Tabela 6).

Tabela 6 - Matriz de distância genética entre pares de regiões geográficas representadas por populações brasileiras do grupo miscigenado e entre tais regiões e a metapopulações nativo-americana e europeia, excluindo-se as populações de Belém-PA e Boa Vista-RR, com valores de p (acima da diagonal) e valores de distância genética (F_{ST} ou R_{ST} ; abaixo da diagonal).

F_{ST}	Sudeste	Sul	Centro-Oeste	Norte	Nordeste	Nativo Americano	Europeia
Sudeste	-	0.01188	0.00337	0.00446	0.08346	0.00000*	0.36759
Sul	0.00113	-	0.19473	0.07039	0.03188	0.00000*	0.06623
Centro Oeste	0.00130	0.00031	-	0.37828	0.01871	0.00000*	0.09286
Norte	0.00153	0.00089	0.00005	-	0.00772	0.00000*	0.12821
Nordeste	0.00099	0.00201	0.00239	0.00343	-	0.00000*	0.29977
Nativo Americano	0.14244	0.15374	0.15800	0.14894	0.15713	-	0.00000*
Europeia	0.00005	0.00115	0.00087	0.00082	0.00032	0.15881	-
R_{ST}							
Sudeste	-	0.00059*	0.01198	0.00099*	0.02673	0.00000*	0.12830
Sul	0.00380	-	0.12177	0.02554	0.00040*	0.00000*	0.00307
Centro Oeste	0.00153	0.00076	-	0.34799	0.00198*	0.00000*	0.01970
Norte	0.00356	0.00203	0.00010	-	0.00000*	0.00000*	0.01168
Nordeste	0.00266	0.00810	0.00728	0.01317	-	0.00000*	0.05495
Nativo Americano	0.19632	0.17595	0.20218	0.20024	0.26164	-	0.00000*
Europeia	0.00087	0.00475	0.00309	0.00414	0.00308	0.23631	*

*Valores de P considerados significantes após correção de Bonferroni: $\leq 0,0024$.

Observou-se que a região norte não apresentou valores P significantes após a exclusão das populações de Belém-PA e Boa Vista-RR nas análises F_{ST} entre regiões geográficas, e apresentou apenas dois valores significantes nas análises R_{ST} (norte-

sudeste; norte-nordeste). Os valores de distância genética na região norte também foram menores do que os observados na análise incluindo as populações de Belém e Boa Vista, exceto a distância genética entre esta região e a metapopulação nativo americana.

Foi realizada análise AMOVA entre as regiões geográficas do Brasil, fundamentada na distância F_{ST} e teste de significância com 10.000 permutas (Tabela 7). Nas populações miscigenada e europeia, observou-se expressiva porcentagem de variação dentro das populações (99,55%; 99,92%; respectivamente). Na nativo-americana, houve menor variação dentro das populações (50,78%).

Tabela 7 - Análise de Variação Molecular (AMOVA) em diferentes grupos de populações brasileiras.

Teste	Grupos*	Tipo de Variação	Variação (%)	Valor de P
1	Norte	Entre grupos	0,07	0,10119
	Nordeste			
	Centro-Oeste	Entre populações dentro grupos	0,38	0,00000**
	Sudeste			
	Sul			
2	Europeu	Entre populações	0,08	0,27871
		Dentro das populações	99,92	
3	Nativo Americano	Entre populações	49,22	0,00000**
		Dentro das populações	50,78	

*As populações usadas em cada região estão reportadas na Tabela 2.

** Valores de P significantes para o nível de $\leq 0,05$.

Para análise de diversidade haplotípica no grupo de origem miscigenada, considerou-se 5.778 haplótipos tipados para o sistema *Yfiler*, dos quais 5.061 (87,6%) haplótipos foram diferentes e 4.591 (90,7%) únicos. O cálculo da diversidade haplotípica foi $0,99993 \pm 0,000004$ e a capacidade de discriminação 87,6%. No grupo nativo-americano, dos 400 haplótipos analisados, 181 (45,2%) foram diferentes, dos quais 130 (71,8%) únicos, o que resulta em valores de diversidade haplotípica e

capacidade de discriminação de $0,98377 \pm 0,00137$ e 45,2%, respectivamente. No grupo europeu, dos 375 haplótipos analisados, 363 (96,8%) mostraram-se diferentes, sendo 351 (96,7%) deles únicos, com valores de diversidade haplotípica e capacidade de discriminação de $0,99983 \pm 0,00005$ e 96,8%, respectivamente.

O Quadro 3 apresenta os 10 haplótipos Y-STR mais comuns no Brasil, e sua distribuição por região. Desses haplótipos, sete foram observados apenas nas populações nativo-americanas, e os outros três em populações miscigenadas, distribuídos nas diversas regiões brasileiras.

Quadro 3 - Distribuição dos haplótipos Y-STR mais comuns no Brasil.

Haplótipo	População Miscigenada					NA	Eu	Total
	N	NE	CO	SE	S			
16-14-24-30-18-13-13/14-13-9-10-21-11-12-14-10-20	1	-	1	8	1	-	1	12
15-14-24-30-17-13-13/14-13-9-10-21-11-12-14-10-20	-	2	1	6	1	-	2	12
13-12-24-28-16-15-13/17-13-10-11-21-11-11-16-9-19	1	2	3	7	-	-	-	13
15-13-24-30-17-13-14/15-13-10-12-22-15-11-14-11-20	-	-	-	-	-	11	-	11
17-14-23-31-15-13-13/17-14-10-12-22-13-11-14-10-19	-	-	-	-	-	12	-	12
16-13-25-31-17-13-14/14-13-10-12-23-14-13-15-11-20	-	-	-	-	-	15	-	15
16-12-24-30-17-13-14/17-13-11-11-22-14-11-14-11-20	-	-	-	-	-	16	-	16
16-13-24-29-17-13-13/14-13-10-12-23-14-11-14-11-20	-	-	-	-	-	21	-	21
15-13-23-31-17-13-15/18-13-11-11-22-14-13-14-11-19	-	-	-	-	-	22	-	22
15-13-24-30-16-13-14/15-13-10-12-22-14-11-15-11-20	-	-	-	-	-	23	-	23
Ordem dos marcadores: DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, YGATAH4, DYS437, DYS438, DYS448; Legenda: N = Norte; NE = Nordeste; CO = Centro-Oeste; SE = Sudeste; S = Sul; NA = Nativo Americano; Eu = Europeu.								

Com auxílio do programa *Haplogroup Predictor*, foi realizada a inferência dos haplogrupos de populações brasileiras miscigenadas, nativo-americanas e europeias

registradas no YHRD, com base em seus haplótipos Y-STR. A análise envolveu 6.553 amostras genotipadas para o kit *Yfiler*. Após aplicação de score 40 e probabilidade de 95%, o programa inferiu 5.060 haplogrupos, correspondendo a 77,2% do total de amostras.

A Tabela 8 apresenta a distribuição de frequências de haplogrupos, com base em haplótipos Y-STR, de 40 populações brasileiras, sendo 22 (55%) miscigenadas, (4.408 haplogrupos), separadas por região geográfica, 16 (40%) de origem nativo-americanas (342 haplogrupos) e duas (5%) de origem europeia (310 haplogrupos). Com exceção das populações nativa americanas, as quais 94% das amostras pertencem ao haplogrupo Q, em todas as demais populações brasileiras houve prevalência do haplogrupo R1b ($\approx 52\%$), seguido pelo haplogrupo E1b1b, com exceção da região sudeste, cuja segunda maior prevalência foi do haplogrupo E1b1a. Com exceção à população nativo-americana, a região norte destacou-se em relação aos haplogrupos Q, comparado às demais populações brasileiras (6,5% contra $\approx 1,8\%$). A distribuição de haplogrupos das populações europeias foi semelhante ao das populações miscigenadas brasileiras.

Tabela 8 - Origem, distribuição e frequência dos haplogrupos (HGs) das populações brasileiras de origem miscigenada, nativo-americana e europeia, com distribuição regional para populações miscigenadas, com base em haplótipos Y-STR.

HGs	Origem	Metapopulação Miscigenada (%)					NA (%)	Eu (%)	
		Haplogrupo Total [n=4.408]	N [n=674]	NE [n=213]	CO [n=620]	S [n=475]			SE [n=2.426]
R1b	Europeu	2.305 (52,3)	374 (55,5)	97 (45,5)	353 (56,9)	259 (54,5)	1.222 (50,4)	9 (2,6)	158 (51,0)
R1a	Europeu	99 (2,2)	14 (2,1)	2 (0,9)	9 (1,5)	23 (4,8)	51 (2,1)	0 (0,0)	9 (2,9)
E1b1b	Europeu	477 (10,8)	71 (10,5)	29 (13,6)	62 (10,0)	43 (9,1)	272 (11,2)	3 (0,9)	34 (11,0)
G2a	Europeu	202 (4,6)	22 (3,3)	14 (6,6)	28 (4,5)	25 (5,3)	113 (4,7)	2 (0,6)	15 (4,8)
I2a	Europeu	16 (0,4)	2 (0,3)	0 (0,0)	2 (0,3)	3 (0,6)	9 (0,4)	0 (0,0)	2 (0,6)
I2a1	Europeu	59 (1,3)	8 (1,2)	4 (1,9)	3 (0,5)	6 (1,3)	38 (1,6)	0 (0,0)	2 (0,6)
I2b	Europeu	3 (0,1)	0 (0,0)	0 (0,0)	1 (0,2)	1 (0,2)	1 (0,0)	0 (0,0)	1 (0,3)
I1	Europeu	164 (3,7)	23 (3,4)	7 (3,3)	21 (3,4)	22 (4,6)	91 (3,8)	0 (0,0)	19 (6,1)
I2b1	Europeu	123 (2,8)	23 (3,4)	9 (4,2)	18 (2,9)	14 (2,9)	59 (2,4)	0 (0,0)	7 (2,3)
J2b	Europeu	73 (1,7)	9 (1,3)	6 (2,8)	12 (1,9)	7 (1,5)	39 (1,6)	1 (0,3)	5 (1,6)
J2a1	Europeu	122 (2,8)	20 (3,0)	6 (2,8)	16 (2,6)	9 (1,9)	71 (2,9)	1 (0,3)	11 (3,5)
L	Europeu	24 (0,5)	4 (0,6)	1 (0,5)	1 (0,2)	6 (1,3)	12 (0,5)	0 (0,0)	1 (0,3)
N	Europeu	7 (0,2)	0 (0,0)	0 (0,0)	2 (0,3)	2 (0,4)	3 (0,1)	0 (0,0)	1 (0,3)
H	Asiático	1 (0,0)	1 (0,1)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	1 (0,3)
T	Asiático	75 (1,7)	10 (1,5)	2 (0,9)	12 (1,9)	10 (2,1)	41 (1,7)	2 (0,6)	9 (2,9)
J1	Africano	100 (2,3)	14 (2,1)	5 (2,3)	21 (3,4)	8 (1,7)	52 (2,1)	1 (0,3)	2 (0,6)
E1b1a	Africano	446 (10,1)	35 (5,2)	28 (13,1)	43 (6,9)	30 (6,3)	310 (12,8)	1 (0,3)	27 (8,7)
Q	Americano	112 (2,5)	44 (6,5)	3 (1,4)	16 (2,6)	7 (1,5)	42 (1,7)	322 (94,2)	6 (1,9)

Legenda: HGs: Haplogrupos; N = Norte; NE = Nordeste; CO = Centro-Oeste; S = Sul; SE = Sudeste; NA = Nativo-Americana; Eu = Europeia; n=número de amostras.

A Tabela 9 apresenta a distribuição da ancestralidade brasileira por região geográfica e por metapopulação, de acordo com a origem dos haplogrupos. Observou-se expressiva ancestralidade europeia em todas as regiões do Brasil e na metapopulação europeia. Na metapopulação nativo-americana, a ancestralidade mostrou-se essencialmente americana.

Tabela 9 - Frequência da ancestralidade das populações brasileiras de origem miscigenada, nativo-americana e europeia, com distribuição regional para populações miscigenadas, com base na origem dos haplogrupos.

Origem Haplogrupo	Haplogrupo Total [n=4.408]	Metapopulação Miscigenada (%)					NA (%)	Eu (%)
		N [n=674]	NE [n=213]	CO [n=620]	S [n=475]	SE [n=2.426]	[n=342]	[n=310]
Europeu	3.674 (83,3)	570 (84,6)	175 (82,2)	528 (85,2)	420 (88,4)	1.981 (81,7)	16 (4,7)	265 (85,5)
Asiático	76 (1,7)	11 (1,6)	2 (0,9)	12 (1,9)	10 (2,1)	41 (1,7)	2 (0,6)	10 (3,2)
Africano	546 (12,4)	49 (7,3)	33 (15,5)	64 (10,3)	38 (8,0)	362 (14,9)	2 (0,6)	29 (9,4)
Americano	112 (2,6)	44 (6,5)	3 (1,4)	16 (2,6)	7 (1,5)	42 (1,7)	322 (94,1)	6 (1,9)

Legenda: N = Norte; NE = Nordeste; CO = Centro-Oeste; S = Sul; SE = Sudeste; NA = Nativo-Americana; Eu = Europeia; n=número de amostras.

4 . DISCUSSÃO

4. DISCUSSÃO

Este trabalho consiste na mais ampla análise de dados brasileiros de haplótipos Y-STR depositados no banco de dados YHRD. O Brasil é considerado uma das populações mais heterogêneas do mundo, devido ao seu processo de colonização iniciado há cinco séculos, envolvendo populações de diversas origens ancestrais, dentre elas os ameríndios, europeus e africanos.⁽³⁹⁾ Além disso, o Brasil possui grande extensão territorial (8.515.759,090 km²) e populacional (207.660.929 indivíduos), ambos índices não estão distribuídos uniformemente pelos estados e regiões brasileiras.^(56,57)

A região Sudeste, por exemplo, a qual apresentou maior número de dados disponíveis na literatura para o cromossomo Y, possui 924.608,854 km² de extensão, a segunda menor extensão territorial entre as regiões brasileiras, perdendo apenas para a região Sul (576.783,781 km²). No entanto, a região Sudeste possui o maior índice demográfico entre as regiões geográficas do Brasil (86.949.714 indivíduos), o que justifica sua maior representação de dados de Y-STR na literatura. A região com menor índice demográfico é a Centro-Oeste (15.875.907 indivíduos), que apresenta também menos dados de Y-STR disponibilizados na literatura.^(56,57)

Estudos afirmam que o uso de um banco de dados para cálculos estatísticos em análises de parentesco e testes de identidade forense é mais confiável se os dados forem representativos da população em análise, além disso, a união de amostras de diferentes regiões é válida apenas se não houver subestrutura da população.^(29,37)

Para considerar a existência de subestrutura populacional entre duas populações, é necessário avaliar a distância genética entre elas. Essa pode ser uma tarefa difícil, já que a interpretação da distância genética, por si só, pode ser imprecisa. A interpretação dos extremos (0 e 1), é simples: o valor zero significa que as populações não possuem

diferenças nas frequências haplotípicas; o valor de um significa que as populações são fixadas para alelos diferentes. Valores entre esses dois extremos representam vários níveis de estruturação, sendo que quanto mais próximos de zero, maior a probabilidade de inexistência de subestrutura. Considera-se que valores abaixo de 0,05 indicam baixas distâncias genéticas, sendo que valores 0,05 a 0,15 sugerem diferenciação moderada; 0,15 a 0,25, grande diferenciação; e acima de 0,25, diferenciação acentuada. No entanto, pode ser difícil e enganoso conferir significado biológico a esses valores.⁽⁵⁸⁾

Os parâmetros F_{ST} e R_{ST} baseiam-se em um modelo mutacional diferente, e não há um consenso na literatura de qual modelo é ideal para estudos de Y-STR. Embora o parâmetro R_{ST} (uma variação da verdadeira estatística original, a F_{ST}) tenha sido desenvolvido especialmente para estudos com microssatélites, seu uso pode ser limitado ou até inútil caso não se comprove o modelo mutacional no qual se baseia, que é o modelo de mutação passo-a-passo.^(58,59)

Em relação à população miscigenada do Brasil, todas as análises entre pares de populações e entre regiões geográficas apresentaram valores de distância genética abaixo de 0,05, exceto uma análise R_{ST} entre as populações de Boa Vista (Roraima) e Maranhão ($R_{ST}= 0,06780$; $P=0,00000$). Embora a população de Boa Vista tenha apresentado os maiores valores de distância genética entre as populações miscigenadas, estes valores são baixos, de modo que as populações podem ser consideradas geneticamente próximas. Tais resultados demonstram ausência de subestrutura entre as populações miscigenadas do Brasil, corroborando outros estudos. Palha et al. (2012)⁽⁴⁴⁾ analisaram 23 marcadores Y-STR em cerca de 2.000 amostras miscigenadas brasileiras distribuídas nas cinco regiões geográficas e observaram ausência de subestrutura populacional evidente no país. O mesmo confirmou-se no estudo de Grattapaglia et al.

(2005)⁽⁶⁰⁾ que, embora tenham analisado apenas nove marcadores Y-STR, também encontraram semelhança entre populações brasileiras.

Por outro lado, o grupo nativo americano brasileiro apresentou altos valores de distância genética nas comparações com as regiões do Brasil (F_{ST} de 0,14244 a 0,15713; R_{ST} de 0,17595 a 0,26164), valores considerados de grande diferenciação.⁽⁵⁸⁾ Além disso, 119 das 120 análises F_{ST} entre populações nativo-americanas apresentaram valores de distância genética acima de 0,05, dos quais apenas oito valores foram menores que 1,5, o que evidencia a grande subestrutura populacional neste grupo. Nas análises R_{ST} , as populações nativo-americanas apresentaram apenas dois valores abaixo de 0,05 (-0,01196 e 0,02986, ambas envolvendo a população de Pará – Mapuera). Essa discrepância pode estar relacionada ao pequeno tamanho amostral que representa esta população (apenas cinco amostras).⁽⁵¹⁾

O gráfico de escala multidimensional destacou a grande distância genética do grupo nativo americano em relação às demais populações do Brasil. Esse achado pode estar relacionado à cultura do matrimônio entre parentes adotada pelas tribos indígenas, o que dificulta a mistura do material genético desses indivíduos com o restante da população brasileira.⁽³⁹⁾ Ressalta-se também maior distância genética da região norte em relação às demais regiões geográficas, o que pode estar relacionado à maior incidência de indígenas nessa região, quando comparada às demais regiões do Brasil.⁽³⁹⁾

A análise MDS realizada entre populações miscigenadas do Brasil evidenciou a distribuição homogênea dos grupos miscigenados em geral. Ressalta-se, porém, distância genética discreta das populações de Belém-PA e Boa Vista-RR, que pode estar relacionada à grande incidência de indígenas urbanos residentes nessas cidades. Boa Vista e Belém representam, respectivamente, a primeira e a terceira maiores incidências

de indígenas urbanos no Brasil. Estudos mostraram que os indígenas urbanos tendem a sofrer discriminação étnica, de modo que negam sua identidade a fim de viver em igualdade de condições com os demais indivíduos.^(61,62) Além disso, Boa Vista é conhecida por seu intenso fluxo migratório desde a colonização, tendo em vista que se localiza próximo à fronteira com dois países (República Cooperativa da Guiana e a Venezuela), destacando-se zonas onde os migrantes representam 85% dos residentes.⁽⁶²⁾

Com relação às populações nativo-americanas, observou-se maior distância genética entre algumas etnias nativo-americanas dos estados do Pará (Ipixuna, Parakanã e Zoé), Maranhão (Awa-Guajá e Urubu-Kaapor) e Amapá (Waiãpi). Embora todas essas tribos pertençam à mesma família linguística, o Tupi-Guarani, cada tribo possui sua língua própria, muitas vezes incompreensível para as demais.⁽⁶³⁾ Além disso, as tribos Tupis possuem o costume de realizar matrimônio avuncular (tio materno com a sobrinha) ou matrimônio entre primos cruzados (filhos das irmãs de seus pais ou dos irmãos de suas mães),⁽³⁹⁾ como por exemplo ocorre com as tribos Waiãpi, Parakanã e Yanomami.⁽⁶⁴⁻⁶⁶⁾ Isso contribui para a ausência de mistura do material genético entre as tribos indígenas, o que poderia explicar as distâncias genéticas observadas tanto na matriz quanto no MDS.

A exclusão das populações de Belém-PA e Boa Vista-RR nas análises de distância genética entre pares de regiões geográficas e as demais metapopulações brasileiras, tornou não significantes as análises F_{ST} . Além disso, diminuiu a ocorrência de análises R_{ST} significantes envolvendo a região norte. Adicionalmente, contribuiu para diminuição nos valores de distância genética de ambos os parâmetros nesta região, com exceção à análise envolvendo a região norte e a metapopulação nativo americana. Nesse caso, a distância genética foi maior quando comparada à análise da região norte

incluindo as populações de Belém-PA e Boa Vista-RR. Isso reforça a possibilidade de que a amostragem de tais populações esteja representada também por indígenas urbanos, e sugere que análises forenses envolvendo estas populações devem ser feitas com cautela, já que as mesmas podem influenciar na subestrutura populacional da região norte do Brasil.^(29,37, 61, 62)

A análise da metapopulação brasileira de origem miscigenada separada entre as regiões geográficas mostrou maior variação haplotípica dentro das populações (99,55%; $P=0,00000$), com baixa variação entre as regiões e valor P não significativo (0,07%; $P=0,10119$). Observou-se também grande variação haplotípica na análise dentro das populações europeias (99,92%). Já em relação ao grupo nativo americano, 49,22% e 50,78% da variação haplotípica foi observada entre e dentro de populações, respectivamente, o que confirma a subestrutura genética desta população. A baixa variação haplotípica dentro das populações nativo americanas, em comparação àquela nas metapopulações miscigenada e europeia, pode estar relacionada à cultura do matrimônio entre parentes, adotado pelas tribos indígenas em geral.^(39,64-66)

Ressalta-se, ainda, a maior diversidade haplotípica das populações brasileiras miscigenadas e europeias ora apresentada (0,9999 e 0,9998, respectivamente), comparado a México (0,9977), Portugal (0,9994), China (0,9995), Argentina (0,9995), Itália (0,9996) e Coréia (0,9997), considerando o mesmo sistema de marcadores.⁽⁶⁷⁻⁷²⁾ Já nas populações nativo-americanas, o baixo valor de diversidade haplotípica (0,9838) corrobora um estudo em populações nativo-americanas da Argentina (0,9769 e 0,9497).⁽⁷³⁾ A diversidade haplotípica é o principal parâmetro analítico de um conjunto de Y-STRs, pois trata-se da probabilidade de que dois indivíduos selecionados ao acaso possuam haplótipos distintos.⁽⁷⁴⁾ Este estudo destaca o alto desempenho analítico do

sistema *Yfiler* para populações brasileiras miscigenadas (99,99%) e europeias (99,98%), mas não para as nativo-americanas (97,74%).

Com relação aos 10 haplótipos mais comuns encontrados no Brasil, os sete com maior número de repetições ocorreram apenas nas populações nativo-americanas. Tal fato contribui para o baixo poder de discriminação e a baixa diversidade haplotípica observada nas populações nativo-americanas brasileiras. Além disso, justifica a grande distância genética desse grupo em relação às regiões geográficas do Brasil formadas por populações miscigenadas. Esse achado corrobora estudo em populações nativo-americanas, desde o México até a Argentina, com diversos haplótipos repetidos mais de 10 vezes em uma mesma área geográfica ou mesma população nativo-americana.⁽⁷⁵⁾

Considerando todos os dados disponíveis no banco de dados YHRD, os outros três haplótipos mais comuns no Brasil (populações miscigenadas e europeias) foram encontrados também no restante do mundo e classificados principalmente nas metapopulações Europeia e Miscigenada, com raras ocorrências em outras metapopulações.⁽³⁴⁾ Adicionalmente, esses três haplótipos quando submetidos também à análise de polimorfismos de nucleotídeo único no cromossomo Y (Y-SNP) foram classificados nos haplogrupos E1b1b e J2b, de acordo com dados do YHRD. Tais haplogrupos destacam-se como alguns dos mais comuns no continente Europeu,^(76,77) o que reflete a influência europeia na população brasileira, e corrobora estudos em diversas populações miscigenadas do Brasil que mostram uma linhagem paterna essencialmente europeia (85,9% a 95,6%), com pouca influência africana e nativo-americana.^(45,46,78-83)

Costumeiramente, haplogrupos do cromossomo Y são definidos com auxílio de Y-SNPs, devido à baixa taxa de mutação destes marcadores comparados aos Y-STRs.

No entanto, em razão da diferença nas taxas de mutação entre Y-SNPs e Y-STRs, um haplogrupo poderia incluir vários haplótipos Y-STR, uma vez que cada mutação Y-STR ocorre em algum cromossomo Y pertencente a um determinado haplogrupo. Desse modo, haplótipos Y-STR podem ser usados para inferir haplogrupos de cromossomo Y, com precisão variável.⁽¹⁹⁾

Em recente diretriz para publicação de dados genéticos populacionais reconheceu-se que a predição de haplogrupos fundamentada em Y-STR pode ser usada para elucidar a ancestralidade paterna de uma população. No entanto, a precisão desse tipo de teste pode ser ineficiente no caso de haplogrupos com baixa representatividade na maioria das populações.⁽⁸⁴⁾ Isso ocorre pois o programa utiliza, para comparação, bancos de dados públicos e artigos publicados com amostras genotipadas para Y-SNP e Y-STR, a fim de realizar a predição de haplogrupos. Desse modo, amostras pertencentes a haplogrupos pouco representados nos bancos de dados e artigos podem ser rejeitadas pelo programa.^(42, 43)

Objetivando verificar a precisão do programa *Haplogroup Predictor* para inferência de haplogrupos por meio de Y-STR, estimamos haplogrupos para 605 perfis haplotípicos da população miscigenada do Rio de Janeiro e para 339 perfis de 15 populações nativo-americanas brasileiras, os quais já possuíam haplogrupos definidos pelo método tradicional (SNPs).^(51,85) Com base em 17 marcadores Y-STR do kit *YFiler* para ambas populações, e aplicação de escore 40 e probabilidade 95%, o programa definiu haplogrupos em 84% e 85% dos perfis haplotípicos do Rio de Janeiro e da população nativo-americana, respectivamente. Destes, houve compatibilidade em 100% dos haplogrupos definidos pelo programa para a população nativo-americana, e em 99% para o Rio de Janeiro, comparado ao obtido pelo modo tradicional. Desse modo,

consideramos a precisão do programa confiável para a estimativa de haplogrupos na população brasileira por meio da análise de marcadores Y-STR.

Foram inferidos haplogrupos em 77,2% (5.060 h) dos haplótipos Y-STR brasileiros incluídos neste trabalho. O haplogrupo R1b destacou-se, seguido pelo haplogrupo E1b1b, em todas as populações brasileiras de origem miscigenada e europeia, com apenas discreta variação regional. Reconhecidamente, o haplogrupo R1 corresponde a mais de 50% da população masculina europeia, distribuídos entre os sub-grupos R1a e R1b, mais frequentes no leste e oeste da Europa, respectivamente.⁽⁸⁶⁾ O haplogrupo E1b1b possui origem na África Subsaariana e se expandiu para o oriente e norte da África no final do pleistoceno. Introduzido no sul da Europa por agricultores imigrantes durante a expansão neolítica, atualmente E1b1b é o único haplogrupo mediterrânico encontrado de forma consistente em toda a Europa.⁽⁸⁷⁾ Desse modo, a linhagem paterna da Europa é dividida em dois principais grupos: os Velhos Europeus, que formam a maior parte da população masculina europeia atual, composto pelos haplogrupos R1b, R1a, I e N; e os Primeiros Agricultores, últimas populações a povoar a Europa com a chegada da agricultura, composto pelos haplogrupos E1b1b, J2 e G.⁽⁷⁶⁾

Diante do exposto, este estudo mostrou notável representatividade de haplogrupos europeus em 77,2% da amostragem inferida nas populações brasileiras, indicando ancestralidade essencialmente europeia no país. Os referidos achados corroboram estudos que relatam a origem da população brasileira como um cruzamento de homens europeus e mulheres ameríndias e africanas,⁽⁷⁹⁾ além de confirmar a história de povoamento do Brasil desde o seu descobrimento.

O haplogrupo E1b1a, que representa aquele de maior frequência na África Ocidental, sugerindo origem essencialmente africana,⁽⁸⁸⁾ destacou-se nas regiões sudeste

e nordeste do Brasil. Desse modo, corrobora a história da escravidão no Brasil, cujas estatísticas de povoamento em 1864 indicavam que 45% e 43% da população escrava de origem africana estavam no nordeste e sudeste, respectivamente, e em 1887, à véspera da abolição da escravatura, a frequência era de 24% e 67% nas respectivas regiões.⁽³⁹⁾

Os haplogrupos Q e C possuem origem atribuída ao nordeste da Ásia, sendo os únicos que colonizaram a América.⁽⁸⁹⁾ No entanto, estudos relatam que o haplogrupo C é encontrado apenas no norte da América do Norte.^(89, 90) Nesse contexto, a população nativo-americana do Brasil representaria apenas o haplogrupo Q. A alta frequência de haplogrupo Q na região norte em populações miscigenadas pode estar relacionada à ampla extensão de terras indígenas localizadas nesta região (48%) em relação ao restante do país, de acordo com o último Censo Demográfico no Brasil.⁽⁶⁴⁾

Destacaram-se neste estudo haplogrupos europeus nas populações miscigenadas e europeias brasileiras, assim como haplogrupos americano (região norte) e africano (regiões nordeste e sudeste), concordante a outra casuística com 1.217 homens das cinco regiões geopolíticas do Brasil, utilizando 41 Y-SNPs.⁽⁷⁸⁾ Nossos achados corroboram outro estudo em população norte-americana,⁽⁹¹⁾ com colonização semelhante ao Brasil, expressiva ancestralidade europeia e baixa contribuição americana e africana.

Adicionalmente, neste estudo, a maior ancestralidade nativo-americana observada na região norte em relação às demais regiões do Brasil, inferida por haplogrupos a partir de Y-STR, é semelhante àquela observada em população da cidade de Macapá-AP (7,2%)⁽⁴⁵⁾, situada à norte da região norte. Além disso, análise de haplogrupos por Y-SNPs mostrou representatividade semelhante à observada neste estudo para ancestralidade nativo-americana na região centro-oeste (2,8%)⁽⁹²⁾. Tal

achado pode estar relacionado ao modo como estas regiões foram povoadas. A região norte possui o maior contingente indígena do Brasil e a região centro-oeste, antes território não povoado, foi formada por uma mistura de imigrantes e povos nativo-americanos da região norte.^(39,74,92)

Em relação à ancestralidade africana, com destaque nas regiões sudeste e nordeste com base nos haplótipos Y-STR, há concordância com estudo em população da região nordeste, com maior frequência de haplogrupo africano no estado da Bahia (19%) comparado a outras regiões do Brasil.⁽⁹³⁾ Nossos achados também corroboram estudos em populações da região sudeste, com notável representatividade de haplogrupos africanos nas cidades do Espírito Santo (11,4%) e Rio de Janeiro (15,7%).^(46,85) Isso reflete a história da escravidão brasileira, cuja população africana povoou principalmente as regiões nordeste e sudeste.⁽³⁹⁾

A ancestralidade europeia prevaleceu em populações distribuídas nas diversas regiões do Brasil (85,9% a 95,6%),^(45, 46, 80-83) assim como no presente estudo (81,7% a 88,4%), que contribui para o esclarecimento da ancestralidade patrilinear essencialmente europeia da população brasileira, com pouca influência africana e nativo-americana. No entanto, a rejeição de 22,8% de amostras brasileiras pelo programa de predição alerta para as limitações da estimativa de haplogrupos por Y-STR para aplicação em genética forense,⁽⁸⁴⁾ tendo em vista que podem influenciar a caracterização e distribuição dos haplogrupos.

5. CONCLUSÃO

5. CONCLUSÃO

Este estudo permite concluir que:

1. As metapopulações miscigenada e europeia do Brasil não apresentam subestrutura populacional, assim é possível a utilização de um banco de dados unificado desses grupos populacionais.

2. A acentuada subestrutura das populações nativo-americanas brasileiras impossibilita um banco de dados único deste grupo, bem como sua junção às demais metapopulações brasileiras. Desse modo, torna-se impossível a utilização de um banco de dados único de todas as amostras brasileiras disponíveis no YHRD para aplicação em genética forense. É necessário um banco de dados exclusivo de amostras miscigenadas e europeias.

3. A ancestralidade patrilinear brasileira essencialmente europeia na metapopulação miscigenada e europeia confirma a história de povoamento do Brasil desde o seu descobrimento. Já a ancestralidade quase que exclusivamente americana em populações nativo-americanas pode refletir seu isolamento reprodutivo, dificultando o compartilhamento desse patrimônio genético na população brasileira.

4. A criação de um banco de dados brasileiro unificado de haplótipos Y-STR, contendo amostras de origem miscigenada e europeia, pode ser uma ferramenta muito útil para cálculo de frequência haplotípica e aplicação em genética forense. No entanto, é necessário considerar as baixas, porém significantes distâncias genéticas nas populações de Belém-PA e Boa Vista-RR.

6. REFERÊNCIAS

6. REFERÊNCIAS

- [1] Almeida Jr A, Costa Jr JBO. Lições de Medicina Legal. 21ª ed. São Paulo (SP): Biblioteca Universitária- Companhia Editora Nacional, 1996.
- [2] França GV. Medicina Legal. 6ª ed. Rio de Janeiro (RJ): Guanabara-Kogan, 2001.
- [3] Jeffreys AJ, Wilson V, Thien SL. Individual-specific ‘fingerprints’ of human DNA. *Nature* 1985;316:76-9.
- [4] Dumache R, Ciocan V, Muresan C, Enache A. Molecular Genetics and its Applications in Forensic Sciences, Forensic Analysis - From Death to Justice, Dr. B Suresh Kumar Shetty (Ed.), InTech, 2016. DOI: 10.5772/63530. <<https://www.intechopen.com/books/forensic-analysis-from-death-to-justice/molecular-genetics-and-its-applications-in-forensic-sciences>>.
- [5] Bodner M, Bastisch, I, Butler, JM, Fimmer R, Gill P, Gusmão L, et al. Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). *Forensic Sci. Int. Genet* 2016;24:97-102.
- [6] Roewer L. DNA fingerprinting in forensics: past, present, future. *Investigative Genetics* 2013;4:22.
- [7] Lin S, Li C, Ip SCY. A selection guide for the new generation 6-dye DNA profiling systems. *Forensic Sci. Int. Genetic* 2017;30:34-42.
- [8] Marais G, Galtier N. Sex chromosomes: how X-Y recombination stops. *Current Biology* 2003;13:R641–3. DOI: 10.1016/S0960-9822(03)00570-0.
- [9] Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics* 2013;14:113-24.

- [10] Charlesworth B. The evolution of sex chromosomes. *Science* 1991;251:1030-3.
- [11] Bellot DW, Page DC. Reconstructing the evolution of vertebrate sex chromosomes. *Cold Spring Harb Symp Quant Biol* 2009;74:345-53.
- [12] Charlesworth B, Charlesworth D. The degeneration of Y chromosomes. *Phil Trans R Soc Lond B* 2000;355:1563–72.
- [13] Crowson D, Barret SCH, Wright SI. Purifying and Positive Selection Influence Patterns of Gene Loss and Gene Expression in the Evolution of a Plant Sex Chromosome System. *Mol Biol Evol* 2017;34:1140-54. DOI: 10.1093/molbev/msx064.
- [14] Skaletsky H, Kuroda-Kawagushi T, Minx PJ, Cordum HS, Hillier L, Brown LG, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003;423:825–37.
- [15] Otto SP, Oannel JR, Peichel CL, Ashman TL, Charlesworth D, Chippindale AJ, et al. About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet* 2011;27:358–67.
- [16] Prinz M, Carracedo A, Mayr WR, Morling N, Parsons TJ, Sajantila A, et al. DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Sci Int Genet* 2007;1:3-12.
- [17] Diegoli TM. Forensic typing of short tandem repeat markers on the X and Y chromosomes. *Forensic Sci Int Genet* 2015;18:140-51.
- [18] Kayser M. Forensic use of Y-chromosome DNA: a general overview. *Hum Genet* 2017;136:621–35. DOI: 10.1007/s00439-017-1776-9.

- [19] Calafell F, Larmuseau MHD. The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. *Hum Genet* 2017;136:559-73. DOI:10.1007/s00439-016-1740-0.
- [20] Lee JC, Tsai LC, Chu PC, Lin YY, Huang TY, Linacre A, et al. The risk of false inclusion of a relative in parentage testing - an in silico population study. *Croat Med J* 2013;54:257-62.
- [21] González-Andrade F, Sanchez D, Penacino G, Martínez Jarreta B. Two fathers for the same child: A deficient paternity case of false inclusion with autosomic STRs. *Forensic Sci Int Genet* 2009;3:138–40.
- [22] Builes JJ, Manrique A, Aguirre D, Puerto Y, Bravo ML, Gusmão L. Utility of Y and X-STRs in the research of complex biological relationship. *Forensic Sci Int Genet Supl Serie* 2009;2:236–7.
- [23] Thomsen AR, Hallenberg C, Simonsen BT, Langkjaer RB, Morling N. A report of the 2002-2008 paternity testing workshops of the English speaking working group of the International Society for Forensic Genetics. *Forensic Sci Int Genet* 2009;3:214-21.
- [24] Thompson JM, Ewing MM, Frank WE, Pogemiller JJ, Nolde CA, Koehler DJ, et al. Developmental validation of the PowerPlex Y23 system: a single multiplex Y-STR analysis system for casework and database samples. *Forensic Sci Int Genet* 2013;7:240–50.
- [25] Definis Gojanović M, Sutlović D. Skeletal remains from World War II mass grave: from discovery to identification. *Croat Med J* 2007;48:520-7.
- [26] Haas C, Shven N, Ruhli FJ, Papageorgopoulou C, Purps J, Geppert M, et al. Y-chromosomal analysis identifies the skeletal remains of Swiss national hero Jörg Jenatsch (1596-1639). *Forensic Sci Int Genet* 2013;7:610-7.

- [27] Ambers A, Gill-King H, Dirkmaat D, Benjamin R, King J, Budowle B. Autosomal and Y-STR analysis of degraded DNA from the 120-year-old skeletal remains of Ezekiel Harper. *Forensic Sci Int Genet* 2014;9:33-41.
- [28] Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci Int* 2006;160:90-101.
- [29] Gjertson DW, Brenner CH, Baur MP, Carracedo A, Guidet F, Luge JA, et al. ISFG: Recommendations on biostatistics in paternity testing. *Forensic Sci Int Genet* 2007;1: 223-31.
- [30] Gill P, Gusmão L, Haned H, Mayr WR, Morling N, Parson W, et al. 'DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods'. *Forensic Sci Int Genet* 2012;6:679-88.
- [31] National Research Council (DC). *The Evaluation of Forensic DNA Evidence*. Washington: National Academy Press, 1996, p. 128.
- [32] Roewer L, Willuweit S. Y-Chromosomale STR-Analyse in der forensischen Praxis. *Rechtsmedizin* 2018;28:149-64.
- [33] Krawczak M. Forensic evaluation of Y-STR haplotype matches: a comment. *Forensic Sci Int* 2001;118:114-5.
- [34] Y-Chromosome Haplotype Reference Database. YHRD © 2000-2018 Sascha Willuweit & Lutz Roewer. Licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. <https://yhrd.org/>.
- [35] Willuweit S, Anslinger K, Babler G, Eckert M, Fimmers R, Hohoff C, et al. Gemeinsame Empfehlungen der Projektgruppe "Biostatistische DNA-

Berechnungen" und der Spurenkommission zur biostatistischen Bewertung von Y-chromosomalen DNA Befunden. Rechtsmedizin 2018;28: 138-42.

- [36] Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H. Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 2001;18:1189–203.
- [37] Gusmao L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, et al. 'DNA Commission of the International Society of Forensic Genetics. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis.'. *Forensic Sci Int* 2006;157:187-97.
- [38] Budowle B, Adamowicz M, Aranda XG, Barna C, Chakraborty R, Cheswick D, et al. Twelve short tandem repeat loci Y chromosome haplotypes: Genetic analysis on populations residing in North America. *Forensic Sci Int* 2005;150:1–15.
- [39] Instituto Brasileiro de Geografia e Estatística - IBGE, Centro de Documentação e Disseminação de Informações. Brasil: 500 anos de povoamento. IBGE, Rio de Janeiro (RJ), 2007.
- [40] Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010;10:564-7. DOI: 10.1111/j.1755-0998.2010.02847.x.
- [41] Nei M. *Molecular Evolutionary Genetics*. Columbia University Press, New York, 1987.
- [42] Athey TW. Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach. *J Genet Geneal* 2005;1:1-7.
- [43] Athey TW. Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach. *J Genet Geneal* 2006;2:34-9.

- [44] Palha T, Gusmão L, Ribeiro-Rodrigues E, Guerreiro JF, Ribeiro-dos-Santos Â, Santos S. Disclosing the Genetic Structure of Brazil through Analysis of Male Lineages with Highly Discriminating Haplotypes. *PLoS ONE* 2012;7:e40007. DOI: 10.1371/journal.pone.0040007.
- [45] Francez PAC, Ramos LPV, Palha TJBF, Santos SEB. Haplotype diversity of 17 Y-str loci in an admixed population from the Brazilian Amazon. *Genet Mol Biol* 2012;35:45-52.
- [46] Figueiredo RF, Ambrosio IB, Branganholi DF, Chemale G, Martins JA, Gomes V, et al. Male-specific contributions to the Brazilian population of Espírito Santo. *Int J Legal Med* 2016;120:679-81. DOI: 10.1007/s00414-015-1214-2.
- [47] Cardena MM, Mansur AJ, Pereira AC, Fridman C. Y haplotype variability in a population of São Paulo state, Brazil [letter]. *Forensic Sci Int Genet* 2016;23:e22-e23. DOI: 10.1016/j.fsigen.2016.05.003.
- [48] Torres SRR. Avaliação Da Estrutura Genética Da População Atual De Santa Catarina Com Diferentes Marcadores Moleculares Para Aplicação Na Genética Forense. [Tese Doutorado em Biologia Celular e do Desenvolvimento] – Florianópolis (SC): Universidade Federal de Santa Catarina; 2014.
- [49] Chemale G, Freitas JM, Badaraco JL, Rosa LS, Martins JA, Martins AD, et al. Y-Chromosomal STR haplotypes in a sample from Brasília, Federal District, Brazil [letter]. *Forensic Sci Int Genet* 2014;9:e3-e4. DOI: 10.1016/j.fsigen.2013.06.005.
- [50] Barra GB, Chianca CF, Rita THS, Velasco LFR, Souza CF. Haplotype diversity of 23 Y-chromosomal STRs in a population sample from the Federal District (Brazil)—a territory that arose from nothing. *Int J Legal Med* 2014;128:945-7. DOI: 10.1007/soo414-013-0960-2.
- [51] Roewer L, Nothnagel M, Gusmão L, Gomes V, Gonzales M, Corach D, et al. Continent-Wide Decoupling of Y-Chromosomal Genetic Variation from Language and Geography in Native South Americans. *PLoS Genet* 2013;9:e1003460. DOI: 10.1371/journal.pgen.1003460.

- [52] Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, et al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci Int Genet* 2014;12:12-23.
- [53] Schwengber SP, Kommers T, Matte C, Raimann PE, Carvalho BA, Leite FPN, et al. Population data of 17 Y-STR loci from Rio Grande do Sul state (South Brazil). *Forensic Sci Int Genet* 2009;4:e31–e33.
- [54] Souza LF, Motta CHAS, Moura-Neto RSM. 17 Y-STR haplotype diversity in São Paulo state (southeast of Brazil). *Int J Legal Med* 2018:1-3.
- [55] Jannuzzi J, Domingues P, Simão F, Chaves ACS, Carvalho EF, Gusmão L. Genetic characterization of Rio de Janeiro for different Y-STR sets. *Int J Legal Med* 2018:1-3.
- [56] Instituto Brasileiro de Geografia e Estatística - IBGE. Aprova os valores de áreas territoriais do Brasil, Estados e Municípios. Resolução N° PR-02 de 29 de Junho de 2017. <https://www.ibge.gov.br/home/geociencias/cartografia/default_territ_area.shtm>. 2017a.
- [57] Instituto Brasileiro de Geografia e Estatística - IBGE, Coordenação de Populações e Indicadores Sociais. Estimativas da população residente para os municípios e para as unidades da federação brasileiros com data de referência em 1° de julho de 2017: [notas metodológicas]. IBGE, Rio de Janeiro, 2017b.
- [58] Balloux F, Lugon-Moulin N. The estimation of population differentiation with microsatellite markers. *Mol Ecol* 2002;11:155–65. DOI: 10.1046/j.0962-1083.2001.01436.x
- [59] Gymrek M, Willems T, Reich D, Erlich Y. Interpreting short tandem repeat variations in humans using mutational constraint. *Nature Genetics* 2017;49:1495-501.

- [60] Grattapaglia D, Kalupniek S, Guimarães CS, Ribeiro MA, Diener OS, Soares CN. Y-chromosome STR haplotype diversity in Brazilian populations. *Forensic Sci Int* 2005;149:99–107. DOI: 10.1016/j.forsciint.2004.06.003.
- [61] Ponte LASX. A população indígena da cidade de Belém, Pará: alguns modos de sociabilidade. *Bol. Mus. Para. Emílio Goeldi* 2009;4:261-75.
- [62] Staevie PM. Expansão urbana e exclusão social em Boa Vista – Roraima. *Oculum Ensaio* 2011; 13:68-87.
- [63] IBGE. Censo Demográfico 2010. Características gerais dos indígenas: resultados do universo. IBGE, Rio de Janeiro, 2011. <http://biblioteca.ibge.gov.br/visualizacao/periodicos/95/cd_2010_indigenas_universo.pdf>. Acesso em: 18 de abril de 2018.
- [64] Pagliaro H, Azevedo MM, Santos RV. Demografia dos povos indígenas do Brasil [online]. Rio de Janeiro: Editora FioCruz, 2005, p.192.
- [65] Gallois DT. Organização das aldeias e formas de casamento entre os Wajãpi: algumas informações básicas. Museu do Desenvolvimento Sustentável do Amapá. 2001.
- [66] Instituto Socioambiental. Povos Indígenas no Brasil: Parakanã. 2018. <https://pib.socioambiental.org/pt/Povo:Parakanã>.
- [67] Ramos-González B, Aguilar-Velázquez JA, de Lourdes Chávez-Briones M, Del Rocío Escareño-Hernández M, Alfaro-Lopez E, Rangel-Villalobos H. Genetic population data of three Y-STR genetic systems in Mexican-Mestizos from Monterrey, Nuevo León (Northeast, Mexico). *Forensic Sci Int Genet.* 2017;29:e21-e22. DOI: 10.1016/j.fsigen.2017.04.016.
- [68] Alves C, Gomes V, Prata MJ, Amorim A, Gusmão L. Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFISTR Yfiler) in Portugal. *Forensic Sci Int* 2007;171:250-5. DOI: 10.1016/j.forsciint.2006.10.022

- [69] Wang Z, Du W, He G, Liu J, Hou Y. Forensic characteristics and phylogenetic analysis of Hubei Han population in central China using 17 Y-STR loci. *Forensic Sci Int Genet* 2017;29:e4-e8. DOI: 10.1016/j.fsigen.2017.04.013
- [70] Marino M, Furfuro S. Haplotype frequencies and mutation rates for 17 Y-STRs in a sample from Mendoza province (Argentina). *Forensic Sci Int Genet Suppl Series* 2011;3:e65-e66. DOI: 10.1016/j.fsigss.2011.08.032.
- [71] Brisighelli F, Blanco-Verea A, Boschi I, Garagnani P, Pascali VL, Carracedo, et al. Patterns of Y-STR variation in Italy. *Forensic Sci Int Genet* 2012;6:834-9. DOI: 10.1016/j.fsigen.2012.03.003.
- [72] Kim SH, Han M., Kim W, Kim W. Y chromosome homogeneity in the Korean population. *Int J Legal Med* 2010;124:653-7. DOI: 10.1007/s00414-010-0501-1.
- [73] Toscanini U, Gusmão L, Berardi G, Amorim A, Carracedo A, Salas A, Raimondi E. Genetic variability of 17 Y chromosome STRs in two Native American populations from Argentina. *Int Congress Series* 2006;1288:154-5. DOI: 10.1016/j.ics.2005.09.169.
- [74] Chianca CF. Diversidade Haplotípica de 23 Y-STR em uma amostra da população do Distrito Federal (Brasil) – um território que surgiu do nada a realidade [Dissertação Mestrado em Ciências da Saúde] – Brasília (DF): Universidade de Brasília; 2013.
- [75] Di Corcia D, Mellado CS, Francia TJD, Ferri, Sarno S, Luiselli D, Rickards O. East of the Andes: The genetic profile of the Peruvian Amazon populations. *Am J Phys Anthropol* 2017;163:328–38. DOI: 10.1002/ajpa.23209.
- [76] Wiik K. Where did european men come from? *J Genet Geneal* 2008;4:35-85.
- [77] Rey-González D, Gelabert-Besada M, Cruz R, Brisighelli F, Lopez-Soto M, Rasool M, et al. Micro and macro geographical analysis of Y-chromosome lineages in South Iberia. *Forensic Sci Int Genet* 2017;29:e9-e15. DOI: 10.1016/j.fsigen.2017.04.021.

- [78] Resque R, Gusmão L, Geppert M, Roewer L, Palha T, Alvarez L, et al. Male Lineages in Brazil: Intercontinental Admixture and Stratification of the European Background. *PLoS ONE* 2016;11:e0152573. DOI: 10.1371/journal.pone.0152573.
- [79] Carvalho-Silva DR, Santos FR, Rocha J, Pena SDJ. The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet* 2001;68:281-6. DOI: 10.1086/316931.
- [80] Silva DA, Carvalho E, Costa G, Tavares L, Amorim A, Gusmão L. Y-chromosome genetic variation in Rio de Janeiro population. *Am J Hum Biol* 2006;18:829-37. DOI: 10.1002/ajhb.20567.
- [81] Leite FP, Callegari-Jacques SM, Carvalho BA, Kommers T, Matte CH, Raimann PE, et al. Y-STR analysis in Brazilian and South Amerindian populations. *Am J Hum Biol* 2008;20:359-63. DOI: 10.1002/ajhb.20702.
- [82] Azevedo DA, Silva LAF, Gusmão L, Carvalho EF. Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil. *Forensic Sci Int Genet Suppl Series* 2009;2:421-2. DOI: 10.1016/j.fsigs.2009.08.166.
- [83] Carvalho M, Brito P, Lopes V, Andrade L, Anjos MJ, Real FC, Gusmão L. Analysis of paternal lineages in Brazilian and African populations. *Genet Mol Biol* 2010;33:422-7. DOI: 10.1590/S1415-47572010005000067.
- [84] Gusmão L, Butler JM, Linacre A, Parson W, Roewer L, Schneider PM et al. Revised guidelines for the publication of genetic population data. *Forensic Sci Int Genet* 2017;30:160-3. DOI: 10.1016/j.fsigen.2017.06.007.
- [85] Oliveira AM, Domingues PM, Gomes V, Amorim A, Jannuzzi J, Carvalho EF, Gusmão L. Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs. *Forensic Sci Int* 2014;13:264-8. DOI: 10.1016/j.fsigen.2014.08.017.
- [86] Myres NM, Rootsi S, Lin AA, Järve M, King RJ, Kutuev I, et al. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet* 2001;19:95-101. DOI: 10.1038/ejhg.2010.146.

- [87]Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, et al. Phylogeographic Analysis of Haplogroup E3b (E-M215) Y Chromosomes Reveals Multiple Migratory Events Within and Out Of Africa. *Am J Hum Genet* 2004;74:1014–22. DOI: 10.1086/386294.
- [88]Rosa A, Ornelas C, Jobling MA, Brehm A, Villems R. Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. *BMC Evol Biol* 2007;7:124. DOI: 10.1186/1471-2148-7-124.
- [89]Chiaroni J, Underhill PA, Cavalli-Sforza LL. Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad Sci* 2009;106:20174–9. DOI: 10.1073/pnas.0910803106/.
- [90]Zhong H, Shi H, Qi XB, Xiao CJ, Jin L, Ma RZ et al. Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. *J Hum Genetics* 2010;0:1-8.
- [91]Beaty KG, Mosher MJ, Crawford MH, Melton P. Paternal Genetic Structure in Contemporary Mennonite Communities from the American Midwest. *Hum Biol* 2016;88:95-108.
- [92]Barcelos R, Ribeiro G, Silva Jr W, Abe-Sandes K, Godinho N, Marinho-Neto F, et al. Male contribution in the constitution of the Brazilian Centro-Oeste population estimated by Y-chromosome binary markers. *Int Congr Ser* 2006;1288:228–230. DOI: doi:10.1016/j.ics.2005.11.109.
- [93]Nascimento E, Cerqueira E, Azevedo E, Freitas V, Azevedo D. The Africa male lineages of Bahia's people – Northeast Brazil: A preliminar SNPs study. *Forensic Sci Int Genet Suppl Series* 2009;2:349–50. DOI: 10.1016/j.fsigss.2009.07.010.

7. APÊNDICE

7. APÊNDICE

Brazilian population structure of Y-STR haplotypes deposited in the Y-Chromosome Haplotype Reference Database (YHRD)

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Abstract

In the forensic routine, the integration of the haplotypic data of Short Tandem Repeats of the Y chromosome (Y-STR) from different Brazilian regions is valid only in the absence of population substructures. Thus, the aim of this study was to evaluate the Brazilian population structure based on data of Y-STR haplotypes available in the literature and deposited in the Y-Chromosome Haplotype Reference Database (YHRD). Data from the YHRD (update 56) were considered. Pairwise genetic distance analysis (F_{ST}) was performed between pairs and groups of populations typed for the Yfiler® kit employing the AMOVA method. Pairwise analyzes were performed involving 6178 Y-STR haplotypes of Brazilian populations of Admixed and Native American metapopulations. High F_{ST} values ($0.02665 \leq F_{ST} \leq 0.98543$) were found among the Brazilian Native American populations. Between the Admixed populations, almost all the pairs of populations showed $F_{ST} < 0.05$ with the majority of p-values not significant (90.9%). The haplotypic variation of Admixed populations stratified by geographic region was 99.55% within the populations ($p < 0.00001$) and the variations were only 0.38% ($p < 0.00001$) and 0.07% ($p < 0.10119$) among populations within regions and among the regions of Brazil, respectively. In the Native American group, the haplotypic variation was 50.78% and 49.22% within and among populations, respectively ($p < 0.00001$). The haplotype diversity of the Admixed populations was 0.99993 and the discrimination capacity was 87.6%, while for the Native American populations the values were 0.98377 and 45.2%, respectively. Considering the Yfiler® system, the Admixed Brazilian metapopulation does not show population substructure, which allows the use of a national database of Y-STR haplotypes to estimate haplotype frequencies of this population in forensic cases.

Keywords: Haplotypes, Y-STR, Forensic Genetics, Brazil.

Introduction

The analysis of Y chromosome Short Tandem Repeat markers (Y-STRs) is a powerful tool in kinship and forensic identity tests, as well as contributing to ancestral studies [1]. However, because Y-STRs are transferred from father to son in the absence of recombination, these markers are in strong Linkage Disequilibrium (LD), that is, they have an allelic association with each other, making it impossible to obtain haplotype frequencies of Y-STRs by multiplying the respective allelic frequencies of the involved loci [2]. Therefore, the calculation of the probability of matching of a given haplotype Y-STR profile should follow a genealogical approach, where the haplotype frequency is determined in groups of geographically distributed populations that share a common ancestry [3].

Due to the smaller effective population size of the Y chromosome in relation to autosomes ($1/4$), a more pronounced genetic drift and, consequently, a more accentuated genetic structure is expected for these chromosomes [4]. Thus, the International Society of Forensic Genetics (ISFG) recommends the use of a regional haplotype frequency database; and is valid to include samples from different regions only when there are no population substructures between the populations. Otherwise, when present, the substructure should be included in Likelihood Ratio (LR) calculations [2,5].

Because of five centuries of interethnic miscegenation, mainly between Amerindians, Europeans and Africans, Brazil is considered one of the most heterogeneous populations in the world. Originally populated by millions of Amerindians, Brazil received, in the first two centuries of colonization, between 500 and five thousand Portuguese per year, mostly men, with these numbers increasing until 1930. Immigration of European women during the first centuries was insignificant. During the period of slavery, four million Africans were sent to Brazil, and subsequently immigrants of different nationalities arrived in the country, mainly Portuguese and Italian, but also Spanish, German, Syrian, Turkish and Japanese [6].

Importantly, Brazil has regions with different ancestral contributions, with the integration of genetic data of these regions being valid only in the absence of population substructures [2]. Thus, a haplotype database of Y-STRs with a large number of

individuals genotyped for a high number of these markers is necessary in order to obtain a reliable genetic structure of the Brazilian population.

Several databases of Y-STRs are available, such as the Y-Chromosome Haplotype Reference Database (YHRD - <http://www.yhrd.org>), the US Y-STR Database (<http://www.usystrdatabase.org>), and the Yfiler Haplotype Database (<http://www6.appliedbiosystems.com/yfilerdatabase>). The YHRD became accessible in February 2000 with data from European populations only, but today it is the largest database of Y-STR haplotypes with a total of 207,469 haplotypes from 1152 population samples, which integrate 134 national databases with 33 metapopulations (update 56).

Thus, in order to characterize the Brazilian haplotypic profile of Y-STRs and evaluate possible population substructures in the country, all Brazilian Y-STR haplotype data deposited in the YHRD (update 56) and available in the literature were used in this study making it the most extensive work performed with these markers in the Brazilian population.

Materials and Methods

The data used in this work were obtained from the YHRD (<http://www.yhrd.org>) and articles available in the literature corresponding to the data included in the database.

Pairwise genetic distance (F_{ST}) was estimated by the Analysis of Molecular Variance (AMOVA) method and these results were visualized in two dimensional space with the Multidimensional Scaling (MDS) analysis employing the Arlequin bioinformatics program (version 3.5.2.2)[7] and Statistica Trial 13.0 software (StatSoft Inc., <http://www.statsoft.com.br>), respectively. The AMOVA analysis within and between the different regional groups was realized with the Arlequin software. For both tests was excluded the DYS385a/b marker, subtracted the DYS389I marker from DYS389II and removed the intermediate alleles. The haplotypic diversity and standard deviation were calculated according to the Nei formula [8] and the discrimination capacity was determined by the ratio of the number of different haplotypes observed by the total number of haplotypes sampled.

Results and Discussion

Brazil has 9706 Y-STR haplotypes (h) deposited in the YHRD from 72 populations (pop) and classified as Native American (546 h from 19 pop), African (693 h from 14 pop), European (1213 h of 5 pop) and Admixed (7254 h of 34 pop) metapopulations. All samples meet the minimum haplotype. The typing by the PowerPlex® Y System (Promega Corporation, Madison, WI) and AmpFLSTR® Yfiler® PCR Amplification Kit (Applied Biosystems, Foster City, CA) were performed for 56 (8066 h) and 53 (7846 h) populations, respectively. Only six populations (996 h) were typed employing the PowerPlex® Y23 System (Promega Corporation, Madison, WI) and one population (265 h) were typed employing the Yfiler® Plus PCR Amplification Kit (Applied Biosystems, Foster City, CA). No Brazilian population is typed for the Maximal kit (YHRD Max Loci) (Supplementary Table 1).

In relation to the regional distribution of Brazilian haplotypes, the southeastern region is the highest contributor (4565 h), followed by the north (1874 h), northeast (1355 h), south (977 h) and central-west (935 h), Figure 1. All regions have populations classified in the Admixed group, which corresponds to the highest percentage of Brazilian data in the database (74.7%), in particular from the southeastern region (47%), Figure 1. Only the southern and southeastern regions deposited haplotypes of the European group, while populations of the African group are registered for all regions of Brazil, except for the southern region, Figure 1. The northern region highlight by contribution from Native American haplotypes (67.9% of haplotypes of this metapopulation), Figure 1.

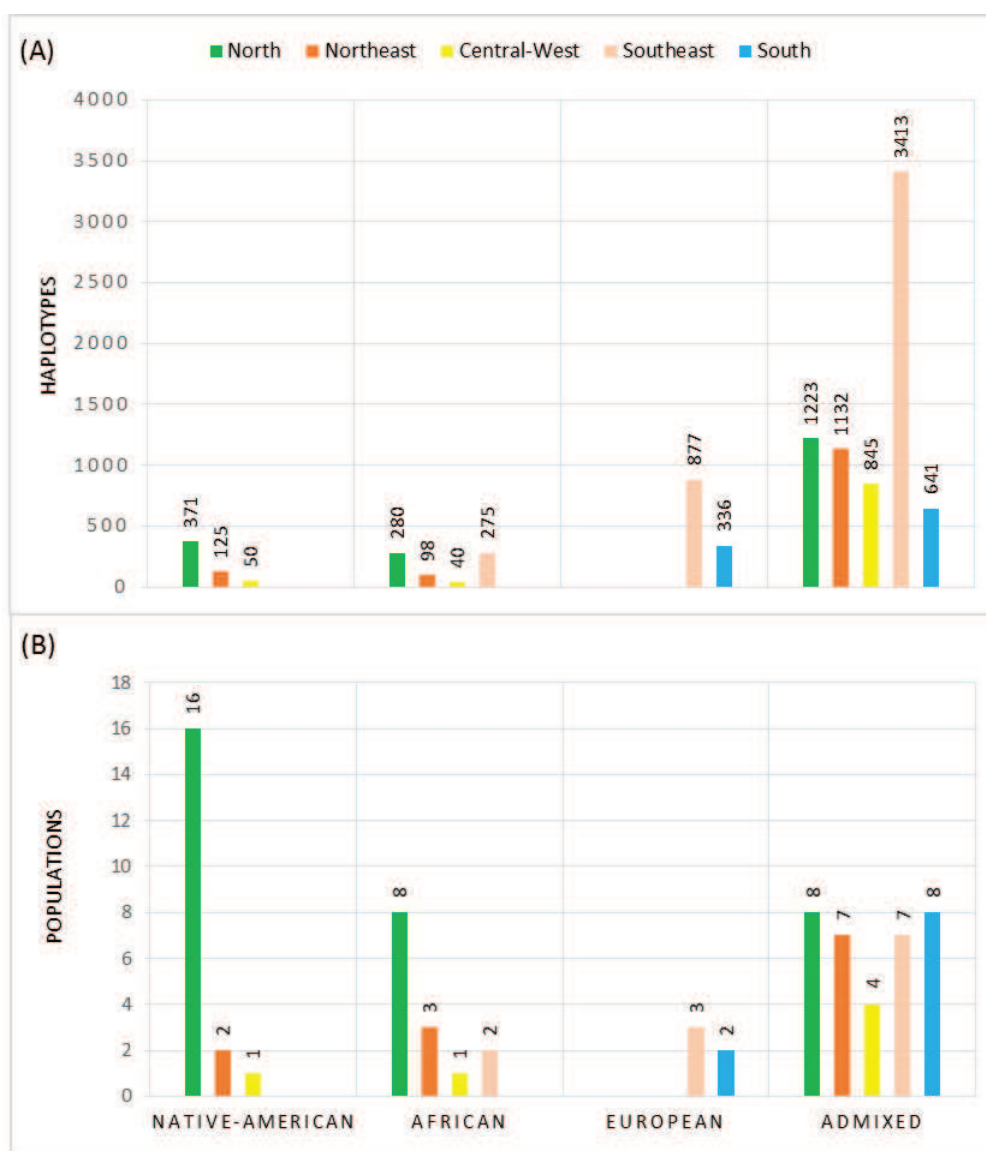


Figure 1 - Distribution of Y-STR haplotypes (A) and populations (B) by Brazilian region and metapopulation, considering the Brazilian haplotype minimum data in the Y-Chromosome Haplotype Reference Database (YHRD)

In order to analyze the Brazilian population structure of Y-STR haplotypes, the YHRD populations typed using the Yfiler system (17 Y-STRs) were selected, since it includes a high number of markers and was used in 81% of the Brazilian Y-STR data of the database (Figure 2, Supplementary Table 1).

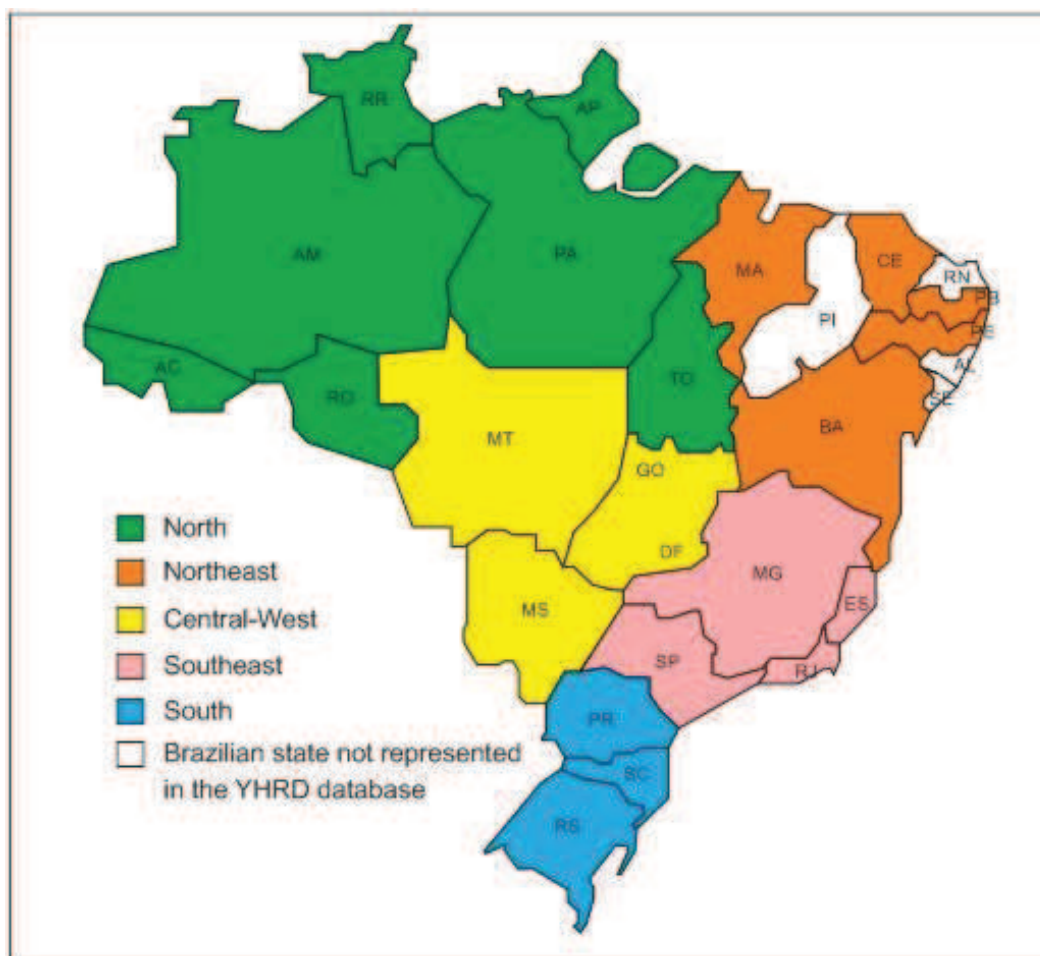


Figure 2 – Representation of Brazilian states with populations available in the Y-Chromosome Haplotype Reference Database (YHRD), considering the Yfiler system. List of the populations and haplotypes number of each population is available in Supplementary Table 1.

Thus, the Y-STR haplotype data from 22 mixed populations (5,778 h, 85% of total YHRD data) and 16 Native American populations (400 h, 83% of total YHRD data), submitted in the YHRD and typed to the Yfiler system (Supplementary Table 2) were recovered from the literature [10,28-37]. Populations were stratified by geographic region, in the case of the admixed population, and genetic distance analyzes were performed between pairs of Brazilian regions and between Brazilian regions and the Native American population (F_{ST}). Significance tests were calculated based on 10,000 permutations, and P values were considered significant after Bonferroni correction (Table 1).

Table 1 - Pairwise genetic distance matrix between the Brazilian populations of the Admixed group grouped in geographical regions and the Native-American population, with values of p (above the diagonal) and values of genetic distance (F_{ST}) (below the diagonal).

Regions	Southeast	South	Central-West	North	Northeast	Native-American
Southeast	-	0.01148	0.00297	0.00000*	0.08415	0.00000*
South	0.00113	-	0.19949	0.00109*	0.03119	0.00000*
Central-West	0.00130	0.00031	-	0.00139*	0.01832	0.00000*
North	0.00357	0.00240	0.00194	-	0.00050*	0.00000*
Northeast	0.00099	0.00201	0.00239	0.00510	-	0.00000*
Native-American	0.14244	0.15374	0.15800	0.14510	0.15713	-

*Significant p-value considered after Bonferroni correction: $\leq 0,0033$

In the pairwise analyzes between the Brazilian regions, significant P values were obtained only between pairs involving the northern region, and the highest genetic distance value was between northern and northeastern regions of Brazil. In relation to the Native American population, all analysis presented significant P values ($P = 0.00000$), with high values of pairwise genetic distance in relation to all geographic regions of Brazil (F_{ST} ranging from 0.14244 to 0.15800).

Thus, the pairwise genetic distance analysis was performed among the Brazilian populations, in order to verify which populations contribute to the distances observed in the analysis between regions (Supplementary Table 3).

Of the 231 pairwise analyzes, only 21 presented significant P values (9.1%), of which 18 involve the populations of Belém (PA) and Boa Vista (RR), both located in the north region of Brazil. In addition, the highest values of genetic distance (F_{ST}) are mainly in the population of Boa Vista (F_{ST} ranging from 0.00617 to 0.03294) and the population of Belém also presented higher F_{ST} values when compared to the other populations (0.01791 being the highest value, with the population of Maranhão). Although the populations of Goiás and Maranhão presented some values of F_{ST} similar to those observed in the population of Belém, the analysis did not return significant P values, except for two analyzes involving the population of Maranhão (Supplementary Table 3.1).

In relation to the Native American populations, of the 120 pairwise analysis, 107 showed a significant P value after Bonferroni correction (89.2%). All analysis showed values of F_{ST} much higher than the values observed among the Admixed populations of Brazil (F_{ST} ranging from 0.09046 to 0.98534), except for one value ($F_{ST}=0.02665$) but with non-significant P value (Supplementary Table 3.2).

The use of a database for statistical calculations in kinship analysis and forensic identity tests is more reliable if the data are representative of the population [3]. It is worth noting that the joining of different regions is valid only if there is no population substructure between them [2]. In order for two populations to be considered genetically close, the genetic distance should be less than or equal to 0.05, but interpreting F_{ST} values *per se* can be a dangerous task. [9]. In relation to the Admixed populations of Brazil, all the population pairs had $F_{ST} < 0.05$ with p-values that were not significant in most cases (90.9%), Supplementary Table 3.1, which demonstrates the absence of population substructures in the Admixed metapopulation of Brazil. This finding corroborates the study by Palha et al. [10], which analyzed 23 Y-STR markers in about 2000 Admixed samples from the five Brazilian regions, and found no regional substructure in the country when considering this group. Moreover, another study analyzed samples from the five regions of Brazil, but with nine Y-STR markers, and did not find any significant difference between the populations of the different regions [11].

On the other hand, evident population substructures was observed in relation to the Brazilian Native American populations, both in relation to Brazilian Admixed populations grouped by region, and within the Native American population, which makes it impossible to use a common database for this metapopulation.

Multidimensional scaling (MDS) analysis was performed based on the F_{ST} values obtained between pairs of Brazilian populations. This analysis showed the population substructure of the Native American group and the homogeneous distribution of the Admixed group of the country, confirming the results obtained by the analysis of pairwise genetic distance (Figure 3).

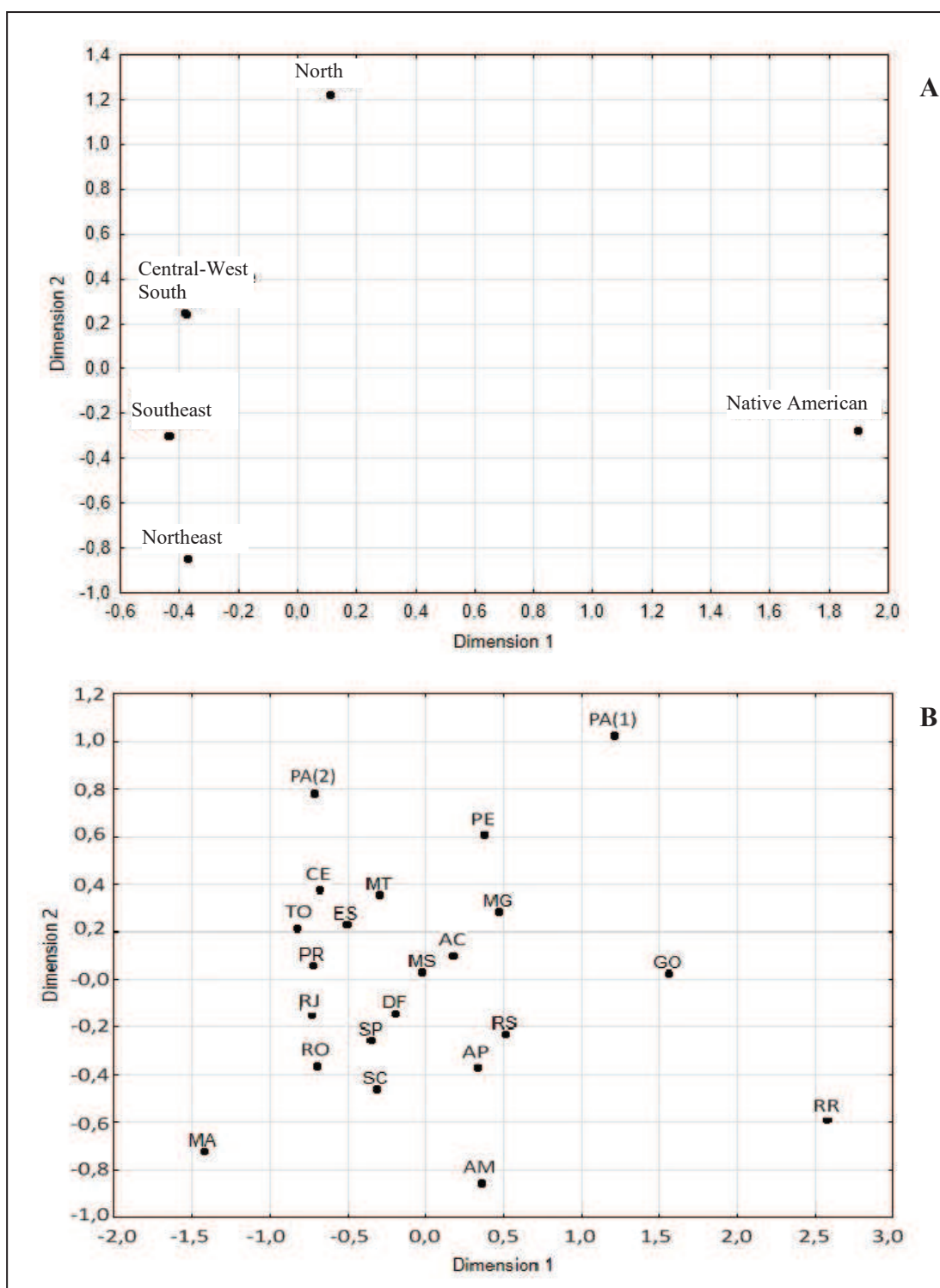


Figure 3 - Multidimensional scaling (MDS) analysis based on the pairwise genetic distance matrix (F_{ST}) of Brazilian populations typed using the Yfiler system and available in the Y-Chromosome Haplotype Reference Database (YHRD) (Table 1; Supplementary Table 3). Stress: (A)=0,0002; (B)=0,1994.

AMOVA analysis was performed between the geographical regions of Brazil, based on distance F_{ST} and test of significance with 10,000 permutations (Table 2). When analyzing the Brazilian admixed metapopulation stratified by region of Brazil (Test 1), it was observed that the highest haplotype variation was obtained within populations (99.55%, $p = 0.00000$), with low variation among Brazilian regions (0.07%, $p = 0.10119$), confirming the results of the pairwise and MDS analysis. Although in this test a significant p -value was obtained in the haplotypic variation among populations within groups, the percentage of variation was low (0.38%, $p = 0.00000$). The findings in this study corroborate those of Palha and collaborators (2012) who did not find significant variation within or between groups in the AMOVA test including 17 populations distributed in the 5 Brazilian geographic regions.

With respect to the Native American group (Test 2), 49.22% of the haplotype variation was observed among populations and only 50.78% within the populations, confirming the data obtained in the analysis of pairwise genetic distance and MDS, which demonstrated high values of F_{ST} and evident genetic distance between populations in this group.

Table 2 - Analysis of Molecular Variance (AMOVA) in different Brazilian population groups.

Test	Groups*	Variation study	Variation (%)	<i>p</i> -value
1	North	Among groups	0.07	0.10119
	Northeast	Among populations	0.38	0.00000**
	Central-west	within groups		
	Southeast	Within populations	99.55	0.00000**
	South			
2	Native American	Among populations	59.22	0.00000**
		Within populations	50.78	

* Populations used in each region are reported in Supplementary Table 2.

** Significant p values for a level of 0.05.

In the Admixed group, of the 5778 haplotypes typed using the YFiler system (Supplementary Table 2), 5061 haplotypes were different, 4591 of which were unique. The calculated haplotype diversity was 0.99993 ± 0.000004 and the discrimination capacity was 87.6%. In the Native American group, only 181 different haplotypes were found in the 400 haplotypes analyzed, only 130 of which were unique, resulting in values of haplotype diversity and discrimination capacity of 0.98377 ± 0.00137 and 45.2%, respectively. The high haplotype diversity of the Brazilian Admixed populations is higher than those reported in studies with samples from other countries such as Mexico (0.9977), Portugal (0.9994), China (0.9995), Argentina (0.9995), Italy (0.9996) and Korea (0.9997) for the same system of Y-STR markers [12-17]. The low haplotype diversity of the Native American populations of Brazil corroborates a study with Native American populations from Argentina (haplotype diversity of 0.9769 and 0.9497) [18].

Table 3 presents the 10 most common haplotypes in Brazil and their distribution by region. It is observed that seven of the most common haplotypes, classified by the YHRD in the Q1a2-M3 haplogroup (Amerindian), belong to the group of Native American populations and are absent in the other Brazilian populations studied. This contributes to the low power of discrimination and haplotype diversity observed in the Brazilian Native American group and to the genetic differentiation of this group in relation to other Brazilian metapopulations. This finding corroborates another study of Native American populations that also found several haplotypes repeated more than ten times in the same geographical area or Native American population, from Mexico to Argentina [19].

Taking into account all the data available in the YHRD database, the other most common haplotypes in the Brazilian population were also found in the rest of the world and are classified, according to the YHRD, mainly in the European and Admixed metapopulations with rare occurrences in the other metapopulations. They were classified in the haplogroups E1b1b and J2b, in cases that were also analyzed by Y-SNP. Considering that these haplogroups are some of the most common on the European continent [20,21], these findings exemplify the important European influence on the Brazilian population and corroborate studies in several Brazilian Admixed populations that reveal a paternal lineage that is essentially European (85.9% - 95.6%) with little African and Native American influence [22-29].

Table 3 - Distribution of the most common Y-STR haplotypes in Brazil

Haplotype	Admixed					NA	Total
	N	NE	C W	SE	S		
16-14-24-30-18-13-13/14-13-9-10-21-11-12-14-10-20	1	-	1	8	1	-	11
15-14-24-30-17-13-13/14-13-9-10-21-11-12-14-10-20	-	2	1	6	1	-	10
13-12-24-28-16-15-13/17-13-10-11-21-11-11-16-9-19	1	2	3	7	-	-	13
15-13-24-30-17-13-14/15-13-10-12-22-15-11-14-11-20	-	-	-	-	-	11	11
17-14-23-31-15-13-13/17-14-10-12-22-13-11-14-10-19	-	-	-	-	-	12	12
16-13-25-31-17-13-14/14-13-10-12-23-14-13-15-11-20	-	-	-	-	-	15	15
16-12-24-30-17-13-14/17-13-11-11-22-14-11-14-11-20	-	-	-	-	-	16	16
16-13-24-29-17-13-13/14-13-10-12-23-14-11-14-11-20	-	-	-	-	-	21	21
15-13-23-31-17-13-15/18-13-11-11-22-14-13-14-11-19	-	-	-	-	-	22	22
15-13-24-30-16-13-14/15-13-10-12-22-14-11-15-11-20	-	-	-	-	-	23	23

Marker order: DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385, DYS393, DYS391, DYS439, DYS635, DYS392, YGATAH4, DYS437, DYS438, DYS448; N = North; NE = Northeast; CW = Central-West; SE = Southeast; S = South; NA = Native American.

Conclusion

This work consists of the most extensive analysis of Brazilian Y-STR haplotype data deposited in the YHRD. These results demonstrated that it is possible to use a unified database of the Admixed Brazilian metapopulation in the calculation of the haplotype frequency (for application in forensic cases). On the other hand, it is evident the impossibility of joining the Native American samples to use a single database, due to the marked substructure of this population. The YHRD database only offers a tool to calculate the haplotype frequency considering all Brazilian haplotypes, regardless of metapopulation. Thus, and in view of the above, we concluded that there is a need to create an exclusive database of Brazilian Admixed samples, for application in forensic cases.

Conflicts of interest

None.

References

- [1] Kayser M. Forensic use of Y-chromosome DNA: a general overview. *Hum Genet.* 136 (2017) 621–635. DOI: 10.1007/s00439-017-1776-9.
- [2] Gusmao L., Butler JM., Carracedo A., Gill P., Kayser M., Mayr WR., Morling N., Prinz M., Roewer L., Tyler-smith C., Schneider PM, DNA Commission of the International Society of Forensic Genetics. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci Int.* 157 (2006) 187-197. DOI: 10.1016/j.forsciint.2005.04.002.
- [3] Willuweit S, Roewer L. The new Y chromosome haplotype reference database (YHRD). *Forensic Sci Int Genet.* 15 (2015) 43-48. DOI: 10.1016/j.fsigen.2014.11.024.
- [4] Hammer MF., Karafet TM., Redd AJ., Jarjanazi H., Santachiara-Benerecetti S., SoodyallH., ZeguraSL. Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol.* 18 (2001) 1189–1203. DOI: 10.1093/oxfordjournals.molbev.a003906.
- [5] Gjertson DW., Brenner CH., Baur MP., Carracedo A., Guidet F., Luque JA., Lessig R., Mayr WR., Pascali VL., Prinz M., Schneider PM., Morling N. ISFG: Recommendations on biostatistics in paternity testing. *Forensic Sci Int Genetics* 1 (2007) 223-231. DOI: 10.1016/j.fsigen.2007.06.006.
- [6] IBGE, Centro de Documentação e Disseminação de Informações. Brasil: 500 anos de povoamento. IBGE, Rio de Janeiro, 2000.
- [7] Excoffier L., Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10 (2010) 564-567. DOI: 10.1111/j.1755-0998.2010.02847.x.
- [8] Nei M. *Molecular Evolutionary Genetics*. Columbia University Press, New York, 1987.

- [9] Balloux F, Lugon-Moulin N. The estimation of population differentiation with microsatellite markers. *Mol Ecol.* 11 (2002) 155–165. DOI: 10.1046/j.0962-1083.2001.01436.x.
- [10] Palha T., Gusmão L., Ribeiro-Rodrigues E., Guerreiro JF., Ribeiro-dos-Santos Â., Santos S.. Disclosing the Genetic Structure of Brazil through Analysis of Male Lineages with Highly Discriminating Haplotypes. *PLoS ONE.* 7 (2012) e40007. DOI: 10.1371/journal.pone.0040007.
- [11] Grattapaglia D., Kalupniek S., Guimarães CS., Ribeiro MA., DienerOS., Soares CN. Y-chromosome STR haplotype diversity in Brazilian populations. *Forensic Sci Int.* 149 (2005) 99–107. DOI: 10.1016/j.forsciint.2004.06.003.
- [12] Ramos-González B., Aguilar-Velázquez JA., de Lourdes Chávez-Briones M., Del Rocío Escareño-Hernández M., Alfaro-Lopez E., Rangel-Villalobos H. Genetic population data of three Y-STR genetic systems in Mexican-Mestizos from Monterrey, Nuevo León (Northeast, Mexico). *Forensic Sci Int Genet.* 29 (2017) e21-e22. DOI: 10.1016/j.fsigen.2017.04.016.
- [13] Alves C., Gomes V., Prata MJ., Amorim A., Gusmão L. Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFI STR Yfiler) in Portugal. *Forensic Sci Int.* 171 (2007) 250-255. DOI: 10.1016/j.forsciint.2006.10.022.
- [14] Wang Z., Du W., He G., Liu J., Hou Y. Forensic characteristics and phylogenetic analysis of Hubei Han population in central China using 17 Y-STR loci. *Forensic Sci Int Genet.* 29 (2017) e4-e8. DOI: 10.1016/j.fsigen.2017.04.013.
- [15] Marino M., Furfuro S. Haplotype frequencies and mutation rates for 17 Y-STRs in a sample from Mendoza province (Argentina). *Forensic Sci Int Genet Suppl Series.* 3 (2011) e65-e66. DOI: 10.1016/j.fsigss.2011.08.032.
- [16] Brisighelli F., Blanco-Verea A., Boschi I., Garagnani P., Pascali VL., Carracedo A., Capelli C., Salas A. Patterns of Y-STR variation in Italy. *Forensic Sci Int Genet.* 6(2012) 834-939. DOI: 10.1016/j.fsigen.2012.03.003.

- [17] Kim SH., Han MS., Kim W., Kim W. Y chromosome homogeneity in the Korean population. *Int J Legal Med.* 124 (2010) 653-657. DOI: 10.1007/s00414-010-0501-1.
- [18] Toscanini U., Gusmão L., Berardi G., Amorim A., Carracedo A., Salas A., Raimondi E. Genetic variability of 17 Y chromosome STRs in two Native American populations from Argentina. *International Congress Series.* 1288 (2006) 154-155. DOI: 10.1016/j.ics.2005.09.169.
- [19] Di Corcia D., Mellado CS., Francia TJD., Ferri F., Sarno S., Luiselli D., Rickards O. East of the Andes: The genetic profile of the Peruvian Amazon populations. *Am J Phys Anthropol.* 163 (2017) 328–338. DOI: 10.1002/ajpa.23209.
- [20] Wiik K. Where did european men come from? *Journal of Genetic Genealogy.* 4 (2008) 35-858.
- [21] Rey-González D., Gelabert-Besada M., Cruz R., Brisighelli F., Lopez-Soto M., Rasool M., Naseer MI., Sanchez-Diz O., Carracedo A. Micro and macro geographical analysis of Y-chromosome lineages in South Iberia. *Forensic Sci Int Genet* (2017). DOI: 10.1016/j.fsigen.2017.04.021.
- [22] Resque R., Gusmão L., Geppert M., Roewer L., Palha T., Alvarez L., Ribeiro-dos-Santos Â., Santos S. Male Lineages in Brazil: Intercontinental Admixture and Stratification of the European Background. *PLoS ONE* 11 (2016) e0152573. DOI: 10.1371/journal.pone.0152573.
- [23] Carvalho-Silva DR., Santos FR., Rocha J., Pena SDJ. The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet.* 68 (2001) 281-286. DOI: 10.1086/316931.
- [24] Silva DA., Carvalho E., Costa G., Tavares L., Amorim A., Gusmao L. Y-chromosome genetic variation in Rio de Janeiro population. *Am J Hum Biol.* 18 (2006) 829-837. DOI: 10.1002/ajhb.20567.
- [25] Leite FP., Callegari-Jacques SM., Carvalho BA., Kommers T., Matte CH., Raimann PE., Schwengber SP., Sortica VA., Tsuneto LT., Petzl-Erler ML., Salzano

- FM., Hutz MH. Y-STR analysis in Brazilian and South Amerindian populations. *Am J Hum Biol.* 20 (2008) 359-363. DOI: 10.1002/ajhb.20702.
- [26] Azevedo DA., Silva LAF., Gusmão L., Carvalho EF. Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil. *Forensic Sci Int Genet Suppl Series.* 2 (2009) 421-422. DOI: 10.1016/j.fsigss.2009.08.166.
- [27] Carvalho M., Brito P., Lopes V., Andrade L., Anjos MJ., Real FC., Gusmão L. Analysis of paternal lineages in Brazilian and African populations. *Genet Mol Biol.* 33 (2010) 422-427. DOI: 10.1590/S1415-47572010005000067.
- [28] Francez PAC., Ramos LPV., Palha TJB., Santos SEB. Haplotype diversity of 17 Y-str loci in an admixed population from the Brazilian Amazon. *Genet Mol Biol.* 35 (2012) 45-52.
- [29] Figueiredo RF., Ambrosio IB., Braganholi DF., Chemale G., Martins JA., Gomes V., Gusmão L., Cicarelli RMB. Male-specific contributions to the Brazilian population of Espírito Santo. *Int J Legal Med.* 120 (2016) 679-681. DOI: 10.1007/s00414-015-1214-2.
- [30] Cardena MM., Mansur AJ., Pereira AC., Fridman C. Y haplotype variability in a population of São Paulo state, Brazil [letter]. *Forensic Sci Int Genet.* 23 (2016) e22-e23. DOI: 10.1016/j.fsigen.2016.05.003.
- [31] Oliveira AM., Domingues PM., Gomes V., Amorim A., Jannuzzi J., Carvalho EF., Gusmão L. Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs. *Forensic Sci Int.* 13 (2014) 264-268. DOI: 10.1016/j.fsigen.2014.08.017.
- [32] Torres SRR. Avaliação Da Estrutura Genética Da População Atual De Santa Catarina Com Diferentes Marcadores Moleculares Para Aplicação Na Genética Forense. Florianópolis. Tese [Doutorado em Biologia Celular e do Desenvolvimento] – Universidade Federal de Santa Catarina; 2014.
- [33] Chemale G., Freitas JM., Badaraco JL., Rosa LS., Martins JA., Martins AD., Jacques GS., Paranaíba RT. Y-Chromosomal STR haplotypes in a sample from

Brasília, Federal District, Brazil [letter]. *Forensic Sci Int Genet.* 9 (2014) e3-e4.
DOI: 10.1016/j.fsigen.2013.06.005.

- [34] Roewer L., Nothnagel M., Gusmão L., Gomes V., Gonzales M., Corach D., et. al. Continent-Wide Decoupling of Y-Chromosomal Genetic Variation from Language and Geography in Native South Americans. *PLoS Genet.* 9 (2013) e1003460. DOI: 10.1371/journal.pgen.1003460.
- [35] Januzzi J., Domingues P., Simão F., Chaves ACS., Carvalho EF., Gusmão L. Genetic characterization of Rio de Janeiro for different Y-STR sets. *Int J Legal Med.* (2018) 1-3. DOI: 10.1007/s00414-018-1776-x.
- [36] Souza LF, Motta CHAS, Moura-Neto RSM. 17 Y-STR haplotype diversity in São Paulo state (southeast of Brazil). *Int J Legal Med* 2018;1-3.
- [37] Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, et al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci Int Genet* 2014;12:12-23.

Supplementary Table 1: List of Brazilian populations with Y-STR haplotypes registered in the Y Chromosome Haplotype Reference Database (YHRD), their respective numbers of haplotypes typed in the different Y-STR markers system, as well as the Brazilian geographic region and metapopulation to which each belong in the YHRD.

Population	Minimal	PPY	Yfiler	PPY23	Yfiler Plus	Maximal	Region	Meta-population
Porto Velho	139	139	139	0	0	0	N	Admixed
Belém	400	400	400	0	0	0	N	Admixed
Boa Vista	67	67	67	0	0	0	N	Admixed
Macapá	197	197	197	0	0	0	N	Admixed
Manaus	281	281	281	204	0	0	N	Admixed
Palmas	30	30	30	0	0	0	N	Admixed
Rio Branco	31	31	31	0	0	0	N	Admixed
Santarém	78	78	78	0	0	0	N	Admixed
Bahia	205	205	205	0	0	0	NE	Admixed
Alagoas	247	0	0	0	0	0	NE	Admixed
Ceará	109	109	109	0	0	0	NE	Admixed
Maranhão	140	140	140	0	0	0	NE	Admixed
Pernambuco	30	30	30	0	0	0	NE	Admixed
Paraíba	300	300	300	0	0	0	NE	Admixed
Fortaleza	101	101	101	0	0	0	NE	Admixed
Brasília	502	502	502	0	0	0	CO	Admixed
Goiás	93	93	93	0	0	0	CO	Admixed
Mato Grosso	200	200	200	200	0	0	CO	Admixed
Mato Grosso do Sul	50	50	50	0	0	0	CO	Admixed
Espirito Santo	355	355	253	0	0	0	SE	Admixed
Minas Gerais	200	200	200	0	0	0	SE	Admixed
Rio de Janeiro	1685	1685	1685	198	258	0	SE	Admixed
Sao Paulo State	45	0	0	0	0	0	SE	Admixed
Sao Paulo State	50	0	0	0	0	0	SE	Admixed
Sao Paulo State	366	366	366	0	0	0	SE	Admixed
Sao Paulo	712	712	712	212	0	0	SE	Admixed
Paraná	77	77	77	0	0	0	S	Admixed
Rio Grande do Sul	243	243	243	0	0	0	S	Admixed
Santa Catarina North	51	51	51	0	0	0	S	Admixed
Santa Catarina Valley	54	54	54	0	0	0	S	Admixed
Santa Catarina Plateau	58	58	58	0	0	0	S	Admixed
Santa Catarina West	56	56	56	0	0	0	S	Admixed
Santa Catarina Capital	50	50	50	0	0	0	S	Admixed
Santa Catarina South	52	52	52	0	0	0	S	Admixed
Curiaú	48	0	0	0	0	0	N	African
Itançoã-Miri	22	0	0	0	0	0	N	African
Marajó	58	0	0	0	0	0	N	African

Mazagão Novo	24	0	0	0	0	0	N	African
Mazagão Velho	36	0	0	0	0	0	N	African
Pitimandeuá	26	0	0	0	0	0	N	African
Saracura	46	0	0	0	0	0	N	African
Trombetas	20	0	0	0	0	0	N	African
Bahia	53	53	0	0	0	0	NE	African
Sergipe	25	25	0	0	0	0	NE	African
Pontal	20	0	0	0	0	0	NE	African
Goiás	40	40	0	0	0	0	CO	African
Rio de Janeiro	235	100	100	0	0	0	SE	African
Sao Paulo State	40	0	0	0	0	0	SE	African
Rio de Janeiro	245	0	0	0	0	0	SE	European
Sao Paulo State	65	0	0	0	0	0	SE	European
Sao Paulo	567	120	120	120	0	0	SE	European
Porto Alegre	81	81	81	0	0	0	S	European
Rio Grande do Sul	255	255	255	0	0	0	S	European
Brazil [Yanomami]	12	0	0	0	0	0	N	Native-American
Pará [Parakanã]	38	38	38	0	0	0	N	Native-American
Amapá [Tiriyó]	35	35	35	0	0	0	N	Native-American
Amapá [Waiãpi]	13	13	13	0	0	0	N	Native-American
Pará [Zoé]	25	25	25	0	0	0	N	Native-American
Sao Gabriel da Cachoeira	62	62	62	62	0	0	N	Native-American
Pará [Arara]	20	20	20	0	0	0	N	Native-American
Pará [Asurini]	15	15	15	0	0	0	N	Native-American
Pará [Gaviao]	18	18	18	0	0	0	N	Native-American
Pará [Ipixuna]	21	21	21	0	0	0	N	Native-American
Rondonia [Karitiana]	34	34	34	0	0	0	N	Native-American
Pará [Kayapó]	16	14	14	0	0	0	N	Native-American
Pará [Mapuera]	5	5	5	0	0	0	N	Native-American
Roraima [Yanomami]	10	10	10	0	0	0	N	Native-American
Amazonas [Ticuna]	23	23	23	0	0	0	N	Native-American
Rondonia [Surui]	24	24	24	0	0	0	N	Native-American
Maranhão [Awa-Guajá]	69	46	46	0	0	0	NE	Native-American
Maranhão [Urubu-Kaapor]	56	27	27	0	0	0	NE	Native-American
Mato Grosso do Sul [Terena]	50	50	50	0	0	0	CO	Native-American

Legend: PPY: PowerPlex Y; PPY23: PowerPlex Y23; N:North; NE: Northeast; CO: Central-West; SE: Southeast; S: South.

Supplementary Table 2 - List of the populations considered in this study, typed for the Yfiler system, and their respective geographic regions, metapopulations, n° of haplotypes sampled and reference in the literature.

Metapopulation	Region	Population	Haplotypes	Reference
Admixed	Southeast	São Paulo	500	Cardena et al., 2016
			212	Provided by the author
		São Paulo State	208	Palha et al., 2012
			158	Souza et al., 2018
		Rio de Janeiro	618	Provided by the author
			850	Jannuzzi et al., 2018
		Espírito Santo	253	Figueiredo et al., 2016
	Minas Gerais	200	Palha et al., 2012	
	South	Santa Catarina	324	Torres, 2014
		Rio Grande do Sul	243	Palha et al., 2012
		Paraná	77	Palha et al., 2012
	Central-West	Brasília	300	Chemale et al., 2014
			201	Barra et al., 2014
		Goiás	93	Palha et al., 2012
		Mato Grosso do Sul	50	Palha et al., 2012
		Mato Grosso	200	Provided by the author
	North	Belém	400	Palha et al., 2012
		Rio Branco	31	Palha et al., 2012
		Macapá	138	Francez et al., 2012
			65	Palha et al., 2012
		Porto Velho	135	Palha et al., 2012
		Boa Vista	67	Palha et al., 2012
		Santarém	72	Palha et al., 2012
		Palmas	30	Palha et al., 2012
		Manaus	74	Palha et al., 2012
		Northeast	Ceará	109
	Pernambuco		30	Palha et al., 2012
Maranhão	140		Palha et al., 2012	
Native-American	Pará (Arara)	20	Roewer et al., 2013	
	Pará (Asurini)	15	Roewer et al., 2013	
	Pará (Gavião)	18	Roewer et al., 2013	
	Pará (Ipixuna)	21	Roewer et al., 2013	
	Pará (Kayapó)	17	Roewer et al., 2013	
	Pará (Mapuera)	5	Roewer et al., 2013	
	Pará (Parakanã)	38	Roewer et al., 2013	
	Pará (Zoé)	25	Roewer et al., 2013	
	Maranhão (Awa)	46	Roewer et al., 2013	
	Maranhão (Urubu)	27	Roewer et al., 2013	
	Mato Grosso do Sul (Terena)	32	Roewer et al., 2013	
	Amapá (Tiryiô)	35	Roewer et al., 2013	
	Amapá (Waiãpi)	13	Roewer et al., 2013	
	Roraima (Yanomami)	10	Roewer et al., 2013	
	Rondonia (Karitiana)	17	Roewer et al., 2013	
	São Gabriel da Cachoeira	61	Purps et al., 2014	

Supplementary Table 3.1 – Pairwise genetic distance matrix among Brazilian populations of the Admixed group, with p-values (above the diagonal) and values of genetic distance (F_{ST}) (below the diagonal).

Pop	(SP)	(RJ)	(MG)	(ES)	(SC)	(RS)	(PR)	(DF)	(GO)	(MS)	(MT)	(PA-1)	(AC)	(AP)	(RO)	(RR)	(PA-2)	(TO)	(AM)	(CE)	(PE)	(MA)
(SP)	-	0.09445	0.01950	0.41253	0.19107	0.00198	0.58212	0.04831	0.00238	0.99040	0.04406	0.00000*	0.59172	0.00406	0.24176	0.00020*	0.07059	0.73369	0.01535	0.74577	0.14979	0.00812
(RJ)	0.00037	-	0.01079	0.69567	0.03099	0.00119	0.85180	0.00881	0.00010*	0.95832	0.07217	0.00000*	0.39897	0.00030	0.17870	0.00000*	0.06881	0.83903	0.01109	0.63914	0.14553	0.00208
(MG)	0.00279	0.00351	-	0.05405	0.02683	0.02990	0.15731	0.02703	0.10662	0.91823	0.01861	0.00248	0.90635	0.03821	0.04366	0.01148	0.02455	0.27532	0.03465	0.06277	0.26087	0.00208
(ES)	0.00000	-0.00050	0.00320	-	0.60024	0.01891	0.54787	0.22552	0.00139	0.95832	0.21305	0.00000*	0.78873	0.02168	0.50351	0.00000*	0.14058	0.46530	0.02119	0.64835	0.21087	0.00624
(SC)	0.00049	0.00160	0.00351	-0.00052	-	0.03148	0.26225	0.53717	0.00228	0.99069	0.06702	0.00000*	0.63370	0.07821	0.08356	0.00020*	0.06861	0.36462	0.02000	0.23196	0.18087	0.02237
(RS)	0.00451	0.00533	0.00375	0.00409	0.00281	-	0.03762	0.04188	0.00406	0.80121	0.02940	0.00000*	0.68280	0.09504	0.00931	0.00683	0.02119	0.07821	0.12425	0.04069	0.43788	0.00000*
(PR)	-0.00086	-0.00229	0.00291	-0.00097	0.00115	0.00628	-	0.24750	0.00663	0.86288	0.40600	0.00218	0.32700	0.04267	0.27631	0.00089	0.05762	0.95931	0.03524	0.67528	0.25394	0.14603
(DF)	0.00102	0.00183	0.00322	0.00060	-0.00023	0.00222	0.00122	-	0.00198	0.99891	0.35006	0.00000*	0.71349	0.10266	0.07871	0.00030	0.09455	0.39491	0.09692	0.21909	0.23859	0.00218
(GO)	0.00910	0.01347	0.00326	0.01337	0.01106	0.00907	0.01328	0.01111	-	0.19780	0.00228	0.05643	0.19820	0.02148	0.00554	0.06287	0.00109	0.03663	0.00416	0.00693	0.08861	0.00327
(MS)	-0.00579	-0.00493	-0.00523	-0.00568	-0.00651	-0.00329	-0.00563	-0.00708	0.00347	-	0.92654	0.12553	0.89526	0.89714	0.82546	0.05267	0.49104	0.79160	0.36986	0.94595	0.63855	0.33393
(MT)	0.00211	0.00160	0.00514	0.00093	0.00229	0.00376	-0.00001	0.00015	0.01289	-0.00522	-	0.00000*	0.25146	0.15177	0.19513	0.00040	0.10306	0.71468	0.04277	0.31383	0.14504	0.00198
(PA-1)	0.01051	0.01306	0.00592	0.01270	0.01165	0.00945	0.01240	0.01192	0.00382	0.00414	0.01243	-	0.20246	0.00020*	0.00010*	0.00020*	0.00040	0.08069	0.00059	0.00050	0.10375	0.00000*
(AC)	-0.00208	0.00020	-0.00708	-0.00498	-0.00275	-0.00328	0.00186	-0.00381	0.00465	-0.00970	0.00282	0.00379	-	0.37838	0.39402	0.07326	0.52817	0.22037	0.64964	0.57163	0.50045	0.05069
(AP)	0.00417	0.00608	0.00378	0.00459	0.00204	0.00196	0.00655	0.00155	0.00691	-0.00484	0.00159	0.00888	0.00042	-	0.01554	0.00366	0.02930	0.11553	0.03802	0.02247	0.20978	0.00465
(RO)	0.00069	0.00108	0.00433	-0.00043	0.00240	0.00647	0.00134	0.00244	0.01063	-0.00408	0.00141	0.01305	0.00030	0.00625	-	0.00010*	0.22552	0.58588	0.03297	0.72349	0.05702	0.01148
(RR)	0.01975	0.02354	0.01122	0.02415	0.02154	0.01117	0.02381	0.02155	0.00617	0.01036	0.02115	0.01648	0.01124	0.01458	0.02265	-	0.00020*	0.00792	0.00129	0.00079	0.01663	0.00000*
(PA-2)	0.00386	0.00421	0.00838	0.00300	0.00456	0.00789	0.00777	0.00365	0.01946	-0.00103	0.00410	0.01536	-0.00200	0.00780	0.00198	0.03294	-	0.31027	0.54826	0.22087	0.12949	0.00198
(TO)	-0.00375	-0.00522	0.00249	-0.00092	0.00074	0.00878	-0.01016	0.00027	0.01347	-0.00737	-0.00417	0.00873	0.00634	0.00750	-0.00249	0.02592	0.00238	-	0.08098	0.67766	0.32502	0.36699
(AM)	0.00720	0.00829	0.00707	0.00812	0.00766	0.00307	0.00990	0.00338	0.01514	0.00090	0.00629	0.01531	-0.00355	0.00698	0.00778	0.02176	-0.00122	0.01079	-	0.03336	0.52995	0.00010*
(CE)	-0.00118	-0.00086	0.00433	-0.00119	0.00110	0.00473	-0.00219	0.00115	0.01244	-0.00650	0.00057	0.01298	-0.00252	0.00656	-0.00175	0.02245	0.00214	-0.00414	0.00878	-	0.15474	0.03425
(PE)	0.00515	0.00573	0.00268	0.00407	0.00472	-0.00026	0.00355	0.00297	0.00942	-0.00457	0.00604	0.00718	-0.00226	0.00399	0.01170	0.02095	0.00843	0.00268	-0.00175	0.00673	-	0.03346
(MA)	0.00487	0.00713	0.01158	0.00783	0.00457	0.01509	0.00314	0.00817	0.01251	0.00090	0.01110	0.01791	0.01201	0.00903	0.00799	0.02560	0.01767	0.00071	0.02204	0.00630	0.01517	-

* Significant p-value considered after Bonferroni correction: $\leq 0,00022$. Pop = Population; (SP) = São Paulo; (RJ) = Rio de Janeiro; (MG) = Minas Gerais; (SC) = Santa Catarina; (RS) = Rio Grande do Sul; (PR) = Paraná; (DF) = Brasília; (GO) = Goiás; (MS) = Mato Grosso do Sul; (MT) = Mato Grosso; (PA-1) = Belém; (AC) = Rio Branco; (AP) = Macapá; (RO) = Porto Velho; (RR) = Boa Vista; (PA-2) = Santarém; (TO) = Palmas; (AM) = Manaus; (CE) = Ceará; (PE) = Pernambuco; (MA) = Maranhão.

Supplementary Table 3.2 – Pairwise genetic distance matrix among Brazilian populations of the Native American group, with p-values (above the diagonal) and values of genetic distance (F_{ST}) (below the diagonal).

Pop	(AP) Tiriyo	(PA) Kayapo	(PA) Arara	(PA) Zoe	(PA) (Mapuera)	(PA) (Parakanã)	(AP) (Waiapi)	(MA) (Urubu)	(MA) (Awa)	(PA) (Asurini)	(PA) (Gaviao)	(MS) (Terena)	(RR) (Yanomami)	(RO) (Karitiana)	(PA) (Ipixuna)	(AM) São Gabriel da Cachoeira
(AP) - Tiriyo	-	0.00050	0.00000*	0.00000*	0.04851	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(PA) - Kayapo	0.12923	-	0.00020*	0.00000*	0.06663	0.00000*	0.00000*	0.00000*	0.00000*	0.00050	0.00099	0.00000*	0.00257	0.00010*	0.00000*	0.00000*
(PA) - Arara	0.20618	0.24853	-	0.00000*	0.07197	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00079	0.00000*	0.00000*	0.00000*
(PA) - Zoe	0.42964	0.63963	0.63173	-	0.00010*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(PA) - Mapuera	0.09046	0.14375	0.13710	0.76235	-	0.00000*	0.00010*	0.00000*	0.00000*	0.01396	0.00168	0.00119	0.00188	0.00010*	0.00000*	0.20196
(PA) - Parakanã	0.50406	0.66956	0.58854	0.94524	0.78448	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(AM) - Waiapi	0.49449	0.62119	0.52862	0.97117	0.70845	0.95277	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(MA) - Urubu	0.36127	0.49053	0.51833	0.72218	0.57935	0.75522	0.86178	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(MA) - Awa	0.44817	0.59150	0.63338	0.64370	0.64193	0.82558	0.85065	0.65286	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(PA) - Asurini	0.19701	0.19314	0.28700	0.54048	0.17189	0.71439	0.69948	0.39332	0.44751	-	0.00030*	0.00000*	0.00000*	0.00030*	0.00000*	0.00000*
(PA) - Gavião	0.26117	0.20791	0.38223	0.57025	0.28419	0.77523	0.72127	0.52461	0.54238	0.16909	-	0.00000*	0.00000*	0.00000*	0.00000*	0.00000*
(MS) - Terena	0.35658	0.39809	0.22995	0.73476	0.30411	0.73327	0.61505	0.65146	0.67826	0.40949	0.49698	-	0.00000*	0.00000*	0.00000*	0.00000*
(RR) - Yanomami	0.23251	0.22198	0.30061	0.80019	0.29232	0.79530	0.73526	0.65172	0.71449	0.34673	0.42946	0.37364	-	0.00000*	0.00000*	0.00000*
(RO) - Karitiana	0.30194	0.34217	0.38685	0.66532	0.38593	0.80789	0.74516	0.53797	0.64174	0.20074	0.29448	0.52632	0.36956	-	0.00000*	0.00000*
(PA) - Ipixuna	0.47318	0.63773	0.65196	0.95696	0.76122	0.95452	0.98534	0.81287	0.74230	0.49350	0.58386	0.70843	0.75600	0.63718	-	0.00000*
(AM) - São Gabriel da Cachoeira	0.17142	0.14806	0.12461	0.38140	0.02665	0.42911	0.38689	0.33181	0.40859	0.13919	0.16787	0.21555	0.22092	0.21169	0.39991	-

* Significant p-value considered after Bonferroni correction: $\leq 0,00042$. Pop = Population; (AP) = Amapá; (PA) = Pará; (MA) = Maranhão; (MS) = Mato Grosso do Sul; (RR) = Roraima; (RO) = Rondônia; (AM) – Amazonas.