

Universidade Federal de Minas Gerais  
Instituto de Ciências Biológicas  
Departamento de Biologia Geral  
Programa de Pós-Graduação em Genética

Dissertação de Mestrado

**AS COMPLEXAS ASSINATURAS GENÉTICAS DOS CASAMENTOS NÃO CASUAIS E  
DA DINÂMICA DE MISCIGENAÇÃO NO CROMOSSOMO X DOS BRASILEIROS**

Autora: Hanaisa de Plá e Sant'Anna

Orientadora: Marília de Oliveira Scliar

Co-orientador: Eduardo Tarazona-Santos

Belo Horizonte

2016

Hanaisa de Plá e Sant'Anna

**AS COMPLEXAS ASSINATURAS GENÉTICAS DOS CASAMENTOS NÃO CASUAIS E  
DA DINÂMICA DE MISCIGENAÇÃO NO CROMOSSOMO X DOS BRASILEIROS**

Dissertação apresentada ao  
Departamento de Biologia Geral do  
Instituto de Ciências Biológicas de  
Universidade Federal de Minas Gerais  
como pré-requisito parcial para obtenção  
do título de Mestre em Genética.

Orientadora: Marília de Oliveira Scliar

Co-orientador: Eduardo Tarazona-Santos

Belo Horizonte

2016

## AGRADECIMENTOS

A todos vocês que tornaram e tornam possível que eu viva intensamente a minha paixão pela biologia, pela ciência e pela genética, deixo aqui o meu singelo agradecimento por permitirem a realização dos meus sonhos. O amor, a amizade, a companhia, o incentivo, o esforço, a preocupação e o ensinamento de todos vocês foram os principais alicerces da minha formação. Espero que também se sintam parte desse trabalho e agradeço a participação de vocês na minha caminhada, na minha felicidade e no que sou.

Agradeço, em especial, aos meus orientadores, Dra. Marília Scliar e Prof. Dr. Eduardo Tarazona Santos, pelo compartilhamento de seu conhecimento e pela confiança, que permitiram a minha participação em diferentes projetos, possibilitando o meu contato com os encantos e com as dificuldades da rotina de laboratório, além de permitir o desenvolvimento de um olhar crítico e de independência. Agradeço também pela qualidade da ciência que desenvolvem, que se tornou um exemplo para mim e que me motiva para a vida acadêmica.

Agradeço também a Dra. Cleusa Fonseca, ao Dr. Jurandir Magalhães e a Dra. Maria Clara Fernandes por aceitarem a participação na banca de defesa desse trabalho, pois estou certa de que contribuirão muito para o seu aperfeiçoamento.

Agradeço aos amigos, professores e funcionários do Departamento de Genética da UFMG pela companhia e amizade. Em especial, a todos que fazem e fizeram parte do LDGH, pelo carinho e pelas discussões científicas, permitindo o meu amplo contato com diferentes áreas de conhecimento e habilidades.

Agradeço a toda a minha família, em especial à minha mãe, Elizeth, ao meu pai, Aldecir, e ao meu irmão, Henrian, pela companhia em todos os momentos, pelo incentivo, pela compreensão, pela preocupação com a minha formação e com a minha felicidade, e por serem meus exemplos de garra e força.

Ao meu grande amor e melhor amigo, Filipe, por me fazer ser uma pessoa cada vez melhor, pelo incentivo a realização dos meus sonhos e por todos os momentos inesquecíveis ao seu lado. Aos meus filhos caninos, Cristal e Loopy, pela alegria e pela companhia em todos os estudos em casa.

Gostaria também de agradecer a todos os meus amigos que, nos momentos serenos ou apreensivos, fizeram parte da minha vida, me guiaram, me incentivaram, me deram força e coragem.

A todos os participantes, colaboradores, voluntários, agências de fomento, universidades e institutos que tornaram possível a minha participação nas pesquisas e o meu desenvolvimento profissional.

Por fim, agradeço aos cuidados, aos ensinamentos, a compaixão, ao acolhimento e a solidariedade de diversas pessoas do meu dia-a-dia, agradeço por terem tornado os meus dias mais especiais e por contribuírem para os meus estudos.

## SUMÁRIO

LISTA DE FIGURAS.....	5
LISTA DE TABELAS.....	6
LISTA DE ABREVIATURAS.....	7
LISTA DE ANEXOS.....	8
RESUMO.....	9
ABSTRACT.....	10
<b>1. INTRODUÇÃO.....</b>	<b>11</b>
1.1 Viés sexual durante o processo de miscigenação.....	11
1.2 Efeitos de casamentos não casuais no Equilíbrio de Hardy-Weinberg.....	13
1.3 Casamentos preferenciais entre indivíduos de mesma origem étnica.....	15
1.4 Diferenças regionais na dinâmica de miscigenação.....	15
<b>2. OBJETIVO.....</b>	<b>19</b>
<b>3. METODOLOGIA.....</b>	<b>20</b>
3.1 Amostras.....	20
3.1.1 Populações.....	20
3.1.1.1 <i>1000 Genomes Project</i> .....	20
3.1.1.2 <i>Projeto EPIGEN-Brasil</i> .....	20
3.1.2 Verificação dos sexos dos indivíduos.....	21
3.1.3 Exclusão dos aparentados.....	21
3.1.4 Exclusão de regiões de homologia entre os cromossomos X e Y.....	23
3.1.5 Inferência de fase e imputação.....	29
3.1.6 Lidando com os cromossomos X hemizigotos dos homens.....	30
3.2 Inferência da ancestralidade local, individual e populacional.....	30
3.3 Análise da ocorrência de viés sexual através das médias de ancestralidade populacional do cromossomo X de homens e de mulheres e comparação aos autossomos.....	32
3.4 Efeitos dos casamentos não casuais no Equilíbrio de Hardy-Weinberg em SNPs do cromossomo X.....	34
3.5 Análise da ocorrência de casamentos preferenciais entre indivíduos de mesma origem étnica através da relação entre $F_{IT}$ e $F_{ST}$ .....	33
3.6 Análise das diferenças regionais na dinâmica de miscigenação através dos CSSA.....	34
<b>4. RESULTADOS E DISCUSSÃO.....</b>	<b>35</b>
4.1 Viés sexual durante o processo de miscigenação.....	35
4.2 Efeitos de casamentos não casuais no Equilíbrio de Hardy-Weinberg.....	37
4.3 Casamentos preferenciais entre indivíduos de mesma origem étnica.....	36
4.4 Diferenças regionais na dinâmica de miscigenação.....	37
<b>5. CONCLUSÃO E PERSPECTIVAS.....</b>	<b>42</b>
<b>6. REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>45</b>
ANEXOS.....	50

## LISTA DE FIGURAS

<b>Figura 1.</b> O DNA mitocondrial é herdado da linhagem materna e o cromossomo Y é restrito a linhagem dos homens.....	12
<b>Figura 2.</b> Níveis das análises de ancestralidade.....	16
<b>Figura 3.</b> Tamanhos esperados de CSSA para um cenário de miscigenação recente (painel esquerdo) e para um modelo de miscigenação antiga (painel direito).....	17
<b>Figura 4.</b> Localização das três maiores coortes populacionais brasileiras, Salvador (BA), Bambuí (MG) e Pelotas (RS), pertencentes ao Projeto EPIGEN-Brasil.....	21
<b>Figura 5.</b> Estrutura dos cromossomos sexuais X e Y e a localização das regiões PAR1 e PAR2, e não pseudoautossômica.....	23
<b>Figura 6.</b> Diferença entre MAF de 46.925 SNPs de mulheres e homens em relação à posição cromossômica (Mb) da região não pseudoautossômica do cromossomo X nas coortes brasileiras.....	24
<b>Figura 7.</b> Diferença entre MAF de mulheres e homens em relação à posição (Mb) no cromossomo X nas populações parentais nativas americanas.....	25
<b>Figura 8.</b> Diferença entre MAF de 30 mulheres e 30 homens em relação à posição cromossômica (Mb) da região não pseudoautossômica do cromossomo X.....	26
<b>Figura 9.</b> Esquema didático da inferência de fase e imputação no cromossomo X através do SHAPEIT.....	30
<b>Figura 10.</b> Fluxograma das análises.....	32
<b>Figura 11.</b> Correlação entre as diferenças nas frequências alélicas do alelo de menor frequência (MAF) de homens e de mulheres brasileiras e os valores de $F_{ST}$ entre as populações parentais europeias e africanas em SNPs do cromossomo X.....	37
<b>Figura 12.</b> $F_{it}$ (informativo da homocigosidade) versus $F_{st}$ (informativo da ancestralidade) para Salvador, Bambuí e Pelotas.....	38
<b>Figura 13.</b> Distribuição dos CSSA no cromossomo X e cromossomo 7.....	40
<b>Figura 14.</b> Modelo da dinâmica de miscigenação populacional do Brasil. Pulsos de miscigenação (1) mais antiga, 18-16 gerações atrás; (2) intermediária, 12-10 gerações atrás; e (3) recente, 6-4 gerações atrás.....	44

## LISTA DE TABELAS

- Tabela 1.** Amostras de homens e mulheres não aparentados por população e bancos de dados de origem.....22
- Tabela 2.** Regiões não pseudoautossômicas do cromossomo X com uma alta variabilidade nas frequências alélicas entre homens e mulheres.....28
- Tabela 3.** Contribuições médias das populações parentais europeias, africanas e nativo-americanas para a ancestralidade dos cromossomos autossômicos e X dos brasileiros.....35

## **LISTA DE ABREVIATURAS**

**ABC** – Computação Bayesiana Aproximada (*Approximate Bayesian Computation*)

**CEU** – residentes de Utah com ancestralidade do Norte e Oeste da Europa

**CSSA** – segmentos cromossômicos de uma ancestralidade contínua específica (*Chromosome Segments of Continuous Specific Ancestry*)

**F<sub>IT</sub>** – estatística F da variância de um indivíduo em relação à variância total

**F<sub>ST</sub>** – estatística F da variância entre populações em relação à variância total

**IBS** – população Ibérica na Espanha

**LWK** – indivíduos de Luhya em Webuey, Quênia

**MAF** – frequência alélica do alelo de menor frequência (*Minor Allele Frequency*)

**SNP** – polimorfismo de nucleotídeo único (*Single Nucleotide Polymorphism*)

**YRI** – iorubas em Ibadan, Nigéria

## LISTA DE ANEXOS

<b>Figura S1.</b> Diferença entre MAF de mulheres e homens em relação à posição (Mb) no cromossomo X nas populações parentais europeias.....	52
<b>Figura S2.</b> Diferença entre MAF de mulheres e homens em relação à posição (Mb) no cromossomo X nas populações parentais africanas.....	53
<b>Figura S3.</b> Distribuição cromossômica das estimativas de $F_{ST}$ comparativa entre estudos.....	54
<b>Tabela S1.</b> Regiões (em Mb) do cromossomo X com predomínio da ancestralidade nativa americana em análise com mulheres e homens separados e juntos.....	55
<b>Tabela S2.</b> Regiões (em Mb) do cromossomo X com predomínio da ancestralidade europeia em análise com mulheres e homens separados e juntos.....	56
<b>Tabela S3.</b> Regiões (em Mb) do cromossomo X com predomínio da ancestralidade africana em análise com mulheres e homens separados e juntos.....	57
<b>Tabela S4.</b> Distribuição cromossômica dos valores de $F_{ST}$ para os autossomos e cromossomo X.....	58
<b>Tabela S5.</b> Contribuições médias das populações parentais europeias, africanas e nativo americanas para a ancestralidade dos cromossomos autossômicos e do cromossomo X das populações de Salvador, Bambuí e Pelotas.....	59



## RESUMO

A constituição da população do Brasil é marcada por uma complexa dinâmica de miscigenação com ocorrência de casamentos não casuais relacionados à origem étnica, tanto pelo viés sexual, com casamentos entre homens de origem europeia e mulheres predominantemente africanas e nativas americanas, quanto pelos casamentos preferenciais entre indivíduos de uma mesma origem étnica. Para investigar os efeitos desses casamentos não casuais na estrutura genética da população brasileira, analisamos 37.457 Polimorfismos de Nucleotídeo Único (*Single Nucleotide Polymorphism* - SNPs) do cromossomo X de 5.823 indivíduos de três coortes populacionais do Brasil com diferentes histórias de miscigenação: Salvador (Nordeste), Bambuí (Sudeste) e Pelotas (Sul). Nossos resultados demonstram o potencial do cromossomo X em evidenciar os padrões de casamentos dos brasileiros. Primeiro, observamos um excesso de contribuição africana e nativa americana na diversidade genômica do cromossomo X, comparado aos autossomos, que apoia a ocorrência de viés sexual na história brasileira. Segundo, evidenciamos que a existência de casamentos não casuais leva a desvios nas frequências genotípicas previstas por Hardy-Weinberg que resultam em diferenças nas frequências alélicas entre homens e mulheres principalmente em SNPs do cromossomo X com alta diferenciação genética entre europeus e africanos. Esses SNPs altamente diferenciados também apresentam um excesso de homoziguidade, o que apoia a existência de casamentos preferenciais entre brasileiros com a mesma origem étnica. Além dessas assinaturas dos casamentos não casuais no cromossomo X dos brasileiros, elucidamos ainda aspectos adicionais da dinâmica de miscigenação do Brasil através da análise da distribuição de segmentos cromossômicos de uma ancestralidade contínua específica (*Chromosome Segments of Continuous Specific Ancestry* - CSSA) do cromossomo X, como a antiga contribuição dos nativos americanos para a diversidade genômica brasileira, a recente imigração europeia no sudeste e sul do país, e a contribuição mais recente de populações africanas no nordeste. Dessa forma, o presente estudo contribui para o entendimento dos eventos demográficos que moldaram a complexa estrutura genética da população brasileira.

**Palavras-chave:** viés sexual, casamentos não casuais, miscigenação brasileira, cromossomo X

## **ABSTRACT**

The Brazilian history is marked by a complex dynamic of admixture, with the occurrence of sex bias with African and Native American female bias and European male bias, and also by an ancestrally related positive assortative mating. To investigate the effects of these non-random mating in the genetic structure of Brazilians, we analyzed 37,457 Single Nucleotide Polymorphism (SNPs) on X chromosome of 5,823 individuals from three cohorts in Brazil with different admixture dynamics: Salvador (Northeast), Bambuí (Southeast) and Pelotas (South). Our results show the potential of the X chromosome to demonstrate patterns of mating in Latinos. First, we observed an excess of African and Native American contributions in genetic diversity of the X chromosome compared to the autosomes, which supports the occurrence of sex bias. Second, we noted that non-random mating leads to deviations in the genotype frequencies predicted by Hardy-Weinberg resulting in differences in allelic frequencies between male and female mainly on X chromosome SNPs with high genetic differentiation between Europeans and Africans. These highly differentiated SNPs also show an excess of homozygosity, which supports the existence of preferential mating between Brazilians with the same ethnic background. Besides these signatures of non-random mating on the X chromosome of Brazilians, we also elucidated further aspects of the dynamics of admixture in Brazil. We analyze the distribution of Chromosome Segments of Continuous Specific Ancestry (CSSA) of the X chromosome and we show the older contribution of Native Americans to the Brazilian genomic diversity, the recent European immigration in the southeast and south of the country, and the newest contribution of Africans in the northeast. Thus, this study contributes to the understanding of the demographic events that have shaped the complex genetic structure of the Brazilian population.

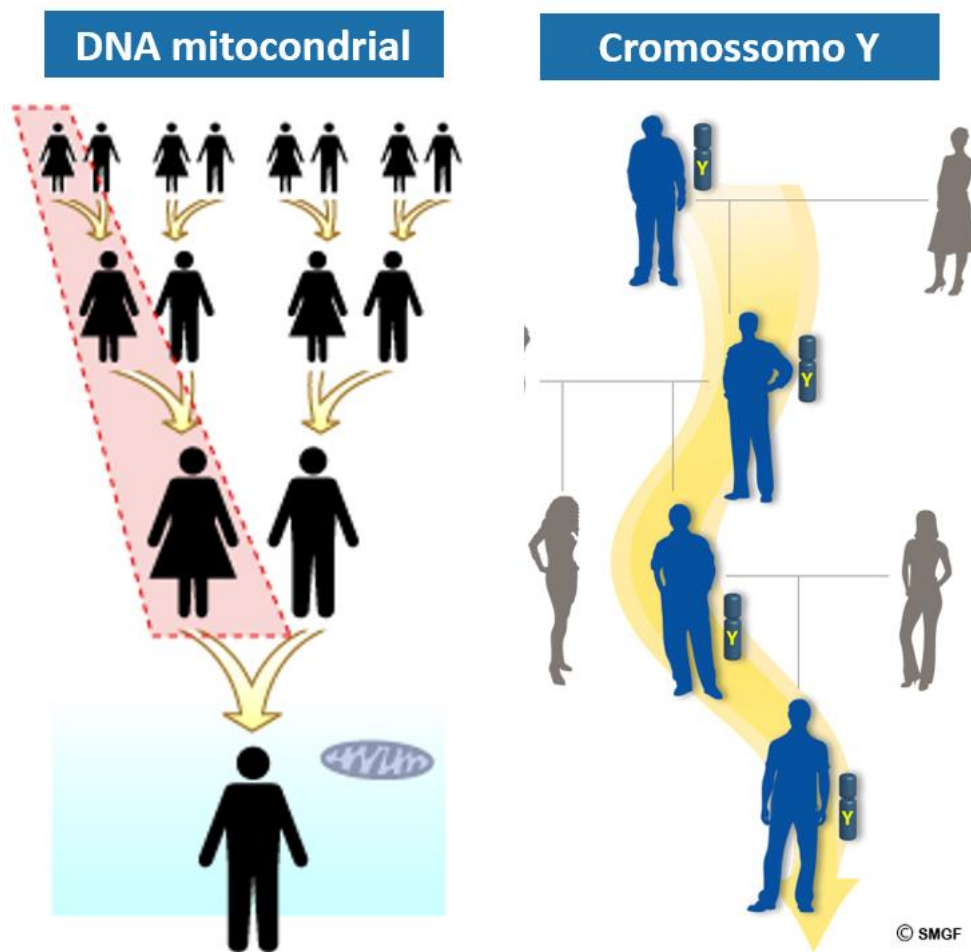
**Keywords:** sex bias, non-random mating, Brazilian admixture, X chromosome

## **1. INTRODUÇÃO**

A população do Brasil é caracterizada por uma complexa estrutura populacional, resultado da sua recente formação pela miscigenação entre colonizadores e imigrantes europeus, escravos africanos e nativos americanos. Esse processo de constituição da população brasileira foi marcado pela ocorrência de casamentos não casuais. Dados históricos reforçam a existência de um viés sexual de casamentos preferenciais entre homens predominantemente europeus e mulheres predominantemente africanas e nativas americanas (Klein e Luna, 2010). Além disso, diversos estudos demográficos indicam que, além do viés sexual durante o processo de miscigenação, há também uma tendência recente dos brasileiros em se relacionarem preferencialmente dentro de seus grupos étnicos e socioeconômicos (Petruccelli, 2001). Esses padrões de casamentos não casuais históricos e recentes deixam diversas assinaturas genéticas, possibilitando seu estudo através dos marcadores genéticos.

### **1.1 Viés sexual durante o processo de miscigenação**

O viés sexual durante a miscigenação é tradicionalmente revelado pela comparação entre os haplogrupos do DNA mitocondrial e do cromossomo Y. O DNA mitocondrial é herdado somente da mãe e evidencia a linhagem materna, enquanto o cromossomo Y é restrito a homens e revela a linhagem paterna (figura 1). O estudo da diversidade desses marcadores de linhagem uniparental na população brasileira evidencia maior contribuição de mulheres nativas americanas e africanas para as linhagens de DNA mitocondrial dos brasileiros (Alves-Silva et al., 2000; Kehdy et al., 2015) e a origem europeia do cromossomo Y de grande parte dos brasileiros (Carvalho-Silva et al., 2001; Kehdy et al., 2015). Essas assinaturas genéticas são interpretadas como evidências de uma maior contribuição feminina das populações africanas e nativas americanas e maior participação masculina de populações europeias para a diversidade genética dos brasileiros.



Adaptado de University of California Museum of Paleontology's Understanding Evolution (<http://evolution.berkeley.edu>) e Sorenson Molecular Genealogy Foundation (SMGF)

**Figura 1. O DNA mitocondrial é herdado da linhagem materna e o cromossomo Y é restrito a linhagem dos homens.**

Recentemente, Goldberg, Verdu e Rosenberg (2014) demonstraram que os níveis de miscigenação em cromossomos autossômicos, por si só, também informam sobre a existência de viés sexual. Através de simulações foi evidenciado que, apesar de não haver diferenças entre os sexos na maneira de herança dos autossomos, podemos observar nesses cromossomos uma reduzida variância da ancestralidade individual na população miscigenada em um cenário com viés sexual em um evento único de miscigenação comparado a um caso sem a ocorrência de viés sexual.

Contudo, os autossomos informam apenas sobre a ocorrência de viés sexual, sendo necessária a comparação de suas estimativas de miscigenação com as médias de ancestralidade no cromossomo X para identificar qual o sexo responsável pela maior contribuição (Goldberg, Verdu e Rosenberg, 2014). O cromossomo X tem um padrão de herança diferenciado entre os sexos, em que as mulheres são portadoras de dois cromossomos X e os homens de apenas um.

Consequentemente, dois terços da história de cada cromossomo X existente ocorreram nas mulheres e, portanto, a composição genética desse cromossomo evidencia principalmente a evolução feminina (Schaffner, 2004). Diferentemente do DNA mitocondrial e da região não pseudoautosômica do cromossomo Y, o cromossomo X é sujeito à recombinação e, por isso, contém mais informação sobre a história demográfica da população e é mais adequado para a comparação com os cromossomos autossômicos. Assim, o excesso da ancestralidade africana e nativa americana no cromossomo X das mulheres da população brasileira, em relação aos autossomos, é visto como evidência de uma maior contribuição de mulheres dessas populações parentais para o processo de miscigenação, e a deficiência de ancestralidade europeia como indicativo da maior contribuição de homens de origem europeia (Resque et al., 2010; Kehdy et al., 2015).

Entretanto, Goldberg e Rosenberg (2015) demonstraram que outras interpretações podem ser plausíveis ao ser considerado apenas as diferenças de ancestralidade entre o cromossomo X e os autossomos, assim como entre o DNA mitocondrial e o cromossomo Y. Em simulações com a população afro-americana, um grande excesso de contribuição masculina europeia para a diversidade genômica dos afro-americanos, mesmo se também houve mais homens do que mulheres contribuintes das populações africanas, pode levar a uma maior fração de ancestralidade europeia nos cromossomos autossômicos (comparado ao cromossomo X) e no cromossomo Y (comparado ao DNA mitocondrial) da população miscigenada afro-americana. Assim, as simulações demonstram que essas diferenças nas ancestralidades comumente associadas a um viés masculino na população europeia e viés feminino na população africana, é também possível em um cenário com viés masculino em ambas as populações parentais.

### **1.2 Efeitos de casamentos não casuais no Equilíbrio de Hardy-Weinberg**

O padrão de cruzamentos de uma população também é evidenciado pela sua composição alélica e genotípica. Em populações com cruzamentos aleatórios, o princípio de Hardy-Weinberg demonstra que as frequências dos alelos e dos genótipos dos polimorfismos de nucleotídeo único (*Single Nucleotide Polymorphisms* - SNPs) autossômicos permanecem constantes de uma geração a outra na ausência de outras forças evolutivas, como mutação, seleção, deriva genética e fluxo gênico. Nessas populações, através da lei de Hardy-Weinberg, é possível deduzir as proporções genotípicas com base nas frequências alélicas. Supondo o caso de um locus autossômico, com dois alelos denominados "A" e "a", a frequência de "A" é representada por  $p$  e a de "a" por  $q$ , é prevista

que as frequências genotípicas sejam  $p^2$  para o genótipo homocigoto “AA”,  $2pq$  para o heterocigoto “Aa” e  $q^2$  para o homocigoto “aa”. No caso de um locus no cromossomo X, no Equilíbrio de Hardy-Weinberg, as mulheres seguem esse mesmo padrão por apresentarem dois alelos em cada locus, porém, os homens apresentam apenas um alelo e, portanto, tanto a frequência alélica quanto a genotípica de “A” será igual a  $p$  e a de “a” será igual a  $q$ . Em situação de Equilíbrio, portanto, prevê-se que a frequência alélica “A” ou “a” nas mulheres e nos homens sejam as mesmas em um locus do cromossomo X. Contudo, nas populações com casamentos não aleatórios entre indivíduos, as frequências alélicas e genotípicas podem ser alteradas a cada geração e, portanto, pode desviar-se da proporção prevista por Hardy-Weinberg. Se esses desvios são significativos, a população é descrita como estando fora do Equilíbrio de Hardy-Weinberg para o locus analisado (Hartl e Clark, 2007; Hedrick, 2011).

No cromossomo X, esse desvio leva não só a um afastamento das proporções alélicas e genotípicas previstas em Hardy-Weinberg nas mulheres, como também resulta em diferenças nas frequências alélicas entre homens e mulheres (Hedrick, 2011). Os homens herdam seu único cromossomo X da mãe, enquanto as mulheres herdam dois cromossomos X, um do pai e outro da mãe. Conseqüentemente, se em uma geração ocorre casamentos preferenciais entre homens europeus e mulheres africanas e nativas americanas, então na próxima geração a maioria das mulheres apresentarão cromossomos X de diferentes origens ancestrais, um paterno europeu e outro materno africano ou nativo americano, enquanto os homens apresentarão apenas um cromossomo X materno preferencialmente africano ou nativo americano. Essa disparidade entre os sexos na origem dos seus cromossomos X resulta em diferenças nas frequências de seus alelos, em particular para os SNPs cujas frequências alélicas se diferem entre as populações parentais (europeias, africanas e nativas americanas). Devido à distinta herança desse cromossomo, essa diferença inicial nas frequências alélicas entre homens e mulheres permanece perceptível no cromossomo X por várias gerações, mesmo após eventos de casamentos casuais em relação à ancestralidade. Portanto, se nas gerações posteriores os casamentos forem casuais em relação à origem étnica, então essas diferenças entre os sexos irão diminuir gradativamente, até que, em 8-10 gerações, as médias de ancestralidade de homens e de mulheres sejam as mesmas (Hedrick, 2011). Se permanecerem a ocorrência de alguma forma de casamentos não casuais em relação à ancestralidade, então as diferenças entre os sexos também serão mantidas (Goldberg e Rosenberg, 2015).

Portanto, se observarmos casamentos não casuais na população brasileira relacionados à origem étnica, seja historicamente pelo viés sexual quanto recentemente pelos casamentos

preferenciais entre indivíduos de uma mesma origem étnica, espera-se que os desvios nas frequências alélicas e genotípicas e, conseqüentemente, o afastamento das proporções de Hardy-Weinberg, em SNPs relacionados à ancestralidade, seja também percebidos no cromossomo X através de uma manutenção das diferenças nas frequências alélicas entre homens e mulheres em SNPs diferenciados entre europeus e africanos (com alto valor de  $F_{ST}$ ). Com isso, é possível avaliar não só a ocorrência, mas também a dinâmica do viés sexual através do estudo do cromossomo X.

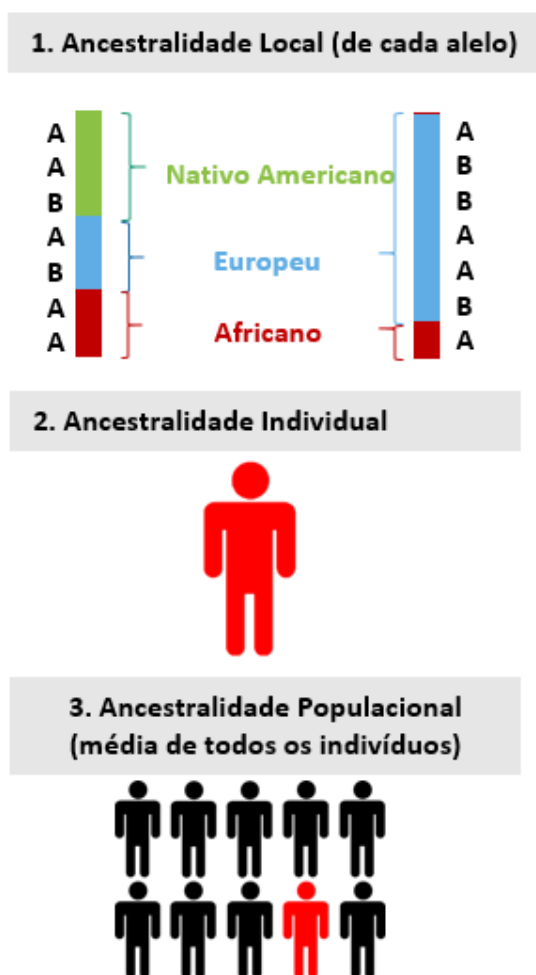
### **1.3 Casamentos preferenciais entre indivíduos de mesma origem étnica**

A existência de casamentos preferenciais gera um aumento de homozigotos nos genes (e SNPs ligados) relacionados à característica escolhida. Na população brasileira, a preferência dos indivíduos em se relacionarem dentro de seus grupos étnicos (Petruccelli, 2001) foi confirmada pelo nosso grupo em Kehdy et al. (2015) através da correlação do  $F_{ST}$  entre europeus e africanos com o  $F_{IT}$  para cada polimorfismo de nucleotídeo único (SNP). O  $F_{ST}$  foi usado como uma medida de informatividade da ancestralidade (quanto maior o  $F_{ST}$ , mais diferenciado é a frequência do SNP entre europeus e africanos) e o  $F_{IT}$  como uma medida de homozigosidade (quanto maior o  $F_{IT}$ , maior a homozigosidade e, conseqüentemente, maior o desvio das proporções previstas por Hardy-Weinberg). Portanto, se houvesse uma tendência dos brasileiros a se casarem com indivíduos de ancestralidade similar, os SNPs mais informativos para a ancestralidade (com maior  $F_{ST}$  entre africanos e europeus) evidenciarão maior homozigosidade (altos valores de  $F_{IT}$ ), determinando uma correlação positiva entre as medidas de  $F_{ST}$  e  $F_{IT}$ . Esse padrão foi exatamente o observado nos cromossomos autossômicos dos brasileiros (Kehdy et al., 2015). Espera-se que esses casamentos preferenciais também afetem SNPs do cromossomo X, gerando essa correlação positiva entre  $F_{IT}$  e  $F_{ST}$ .

### **1.4 Diferenças regionais na dinâmica de miscigenação**

A história demográfica da população brasileira ainda apresenta um nível adicional de complexidade. As populações parentais europeias, africanas e nativas americanas contribuíram com diferentes proporções para a diversidade genômica dos brasileiros e essa dinâmica de miscigenação variou ao longo das diferentes regiões do país (Kehdy et al., 2015). A maior parte dos estudos genéticos de miscigenação e do viés sexual em populações brasileiras tem se restringido a estimar a proporção de contribuição das populações parentais a nível populacional (Krieger et al., 1965; Salzano e Bortolini, 2002; Resque et al., 2010). Com a atual redução dos custos de genotipagem, é possível obter informação de um grande número de SNPs e estudar a história brasileira também a

nível local, que considera a ancestralidade inferida em pequenos trechos ao longo de todo o cromossomo (figura 2) (Winkler, Nelson e Smith, 2010).

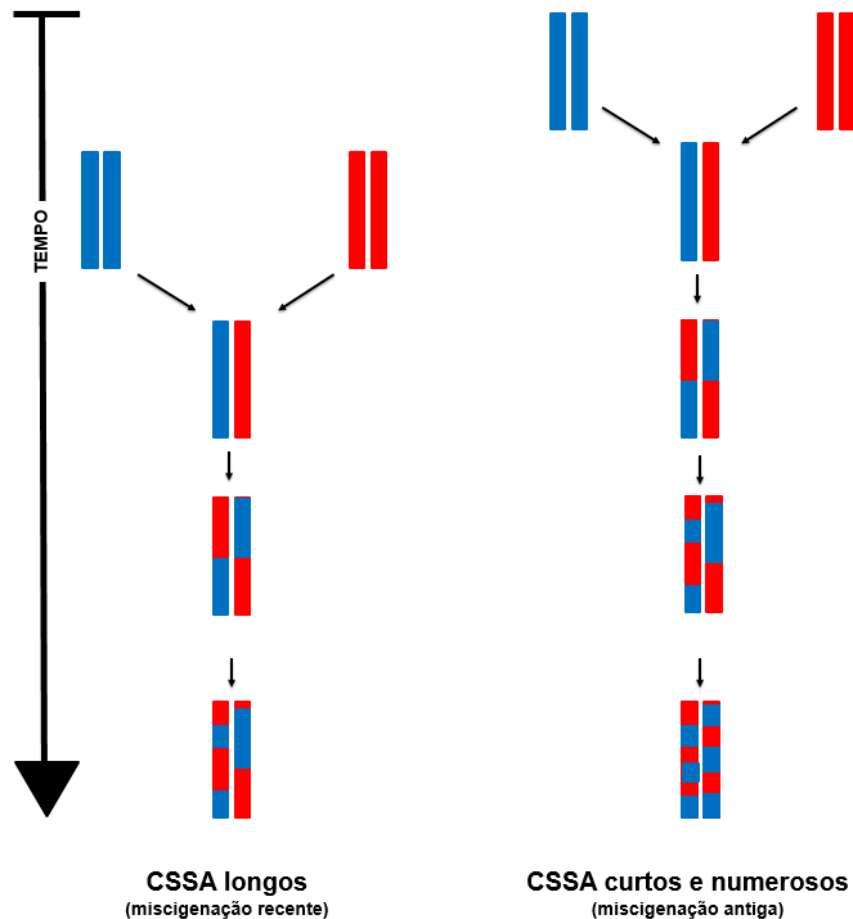


**Figura 2. Níveis das análises de ancestralidade.** (1) Local, (2) individual e (3) populacional.

Os cromossomos dos brasileiros são mosaicos de segmentos de ancestralidade europeia, africana e nativa americana que refletem o processo demográfico do Brasil. Quanto maior o tempo desde a miscigenação, maiores são as ocorrências de eventos de recombinação que quebram os segmentos cromossômicos e, conseqüentemente, menores são os segmentos cromossômicos de uma ancestralidade contínua específica (*Chromosome Segments of Continuous Specific Ancestry - CSSA*) (Liang e Nielsen, 2014). Por exemplo, na figura 3, considere duas populações miscigenadas resultantes de duas populações parentais, sendo que o processo de miscigenação foi mais recente na população representada à esquerda. Os cromossomos autossômicos de uma população parental estão representados em vermelho e os de outra população em azul. Com a ocorrência de um evento de miscigenação, em que as duas populações parentais contribuem igualmente em



proporção para a população miscigenada, espera-se que os seus cromossomos sejam formados por um homólogo de origem completamente da população vermelha e um homólogo da população azul. Então, nessa situação, observam-se dois grandes CSSA na população miscigenada, um cromossomo inteiro inferido para a ancestralidade vermelha e outro para a ancestralidade azul. Se ocorrerem casamentos aleatórios na população miscigenada, nas gerações seguintes é esperado observar uma fragmentação gradual desses segmentos cromossômicos contínuos devido à ocorrência de eventos de recombinação. Com isso, espera-se que com o passar das gerações, os cromossomos da população miscigenada sejam formados por CSSA cada vez menores e mais numerosos. Portanto, através da análise da distribuição de CSSA, é possível estimar a contribuição das populações parentais e o tempo desde os eventos de miscigenação.



**Figura 3. Tamanhos esperados de CSSA para um cenário de miscigenação recente (painel esquerdo) e para um modelo de miscigenação antiga (painel direito).** Devido ao menor tempo para a ocorrência de recombinação para quebrar os CSSA na população com miscigenação recente (esquerda), espera-se que, nesse caso, os tamanhos de CSSA sejam maiores do que em populações com miscigenação antiga (direita).

Essa estimativa, contudo, deve considerar as características diferenciadas de cada cromossomo. No cromossomo X, apenas os cromossomos das mulheres recombinam-se a cada geração (Schaffner,2004). Portanto, apenas dois terços do cromossomo X ficam disponíveis para a ocorrência de quebras nesses segmentos contínuos. Com isso, comparado aos autossomos, espera-se que os CSSA estejam mais intactos no cromossomo X e, portanto, apresentando comprimentos maiores. Consequentemente, se essa característica não for considerada, pode ocorrer uma subestimação do tempo desde a ocorrência dos eventos de miscigenação.

O trabalho se realiza no contexto do projeto EPIGEN-Brasil (<https://epigen.grude.ufmg.br/>), com dados de 5.823 indivíduos das três maiores coortes brasileiras de base populacional: Salvador (Nordeste; Barreto et al., 2006), Bambuí (Sudeste; Lima-Costa, Firmo e Uchoa, 2011) e Pelotas (Sul; Victora, 2006). Primeiro, analisaremos o efeito do viés sexual durante o processo de miscigenação na estrutura genética do cromossomo X em comparação com os cromossomos autossômicos (Kehdy et al., 2015) de homens e de mulheres brasileiras. Segundo, observaremos o efeito dos casamentos não casuais nas frequências alélicas no cromossomo X de homens e de mulheres, e possíveis desvios das proporções genotípicas previstas por Hardy-Weinberg. Terceiro, estudaremos a ocorrência de casamentos preferenciais positivos para a ancestralidade através da análise de  $F_{IT}$  e  $F_{ST}$  no cromossomo X. Por fim, analisaremos diferenças regionais na dinâmica de miscigenação e nas contribuições das populações parentais através do CSSA. Com isso, espera-se contribuir para um maior conhecimento histórico e antropológico da formação da estrutura genética da população brasileira.

## 2. OBJETIVO

Verificar os efeitos do viés sexual, dos casamentos preferenciais entre indivíduos de mesma origem étnica e da dinâmica da miscigenação da população brasileira na estrutura genética do cromossomo X de homens e mulheres, além de verificar as diferenças regionais das contribuições parentais considerando três coortes populacionais do Brasil com diferentes histórias de colonização: Salvador (Nordeste), Bambuí (Sudeste) e Pelotas (Sul).

Para atingi-lo, propomos os seguintes **objetivos específicos**:

1. Estimar o efeito do viés sexual no processo de miscigenação da população brasileira através da comparação da estrutura genética do cromossomo X com os revelados pelos autossomos.
2. Testar a hipótese de que o viés sexual e os casamentos preferenciais entre indivíduos de mesma origem étnica determinam desvios das proporções previstas pela lei de Hardy-Weinberg em SNPs do cromossomo X com alta diferenciação entre populações parentais.
3. Verificar a ocorrência de casamentos preferenciais entre indivíduos de mesma origem étnica na história recente do Brasil através do estudo da relação entre  $F_{IT}$  e  $F_{ST}$  em SNPs do cromossomo X.
4. Analisar as diferenças regionais nas contribuições parentais durante o processo de miscigenação a partir da análise da ancestralidade local e dos segmentos cromossômicos de uma ancestralidade contínua específica (CSSA).

### **3. METODOLOGIA**

#### **3.1 Amostras**

##### **3.1.1 Populações**

Dados de 41.129 SNPs do cromossomo X de homens e de mulheres das populações parentais africanas (Quênia e Nigéria) e europeias (Norte e Oeste da Europa e Espanha) foram extraídos dos bancos de dados públicos *1000 Genomes Project* (seção 3.1.1.1) e dados de >45.766 SNPs das populações parentais nativas americanas (Ashaninka, Shimaá e Aymaras do Peru) foram obtidas do banco de dados do nosso Laboratório de Diversidade Genética Humana (LDGH) da Universidade Federal de Minas Gerais (UFMG) (tabela 1). Seleccionamos as mesmas populações parentais utilizadas em Kehdy et al. (2015) nas análises com os cromossomos autossômicos para manter a comparabilidade dos dados. Esses arquivos das populações parentais foram integrados com os dados de 46.925 SNPs do cromossomo X de indivíduos das coortes de Salvador (Nordeste), Bambuí (Sudeste) e Pelotas (Sul) do projeto EPIGEN-Brasil (seção 3.1.1.2) (tabela 1).

##### **3.1.1.1 *1000 Genomes Project***

O diferencial do banco de dados público *1000 Genomes Project* (McVean et al., 2012), comparado aos outros de populações humanas como o *HapMap* (Holding, 2003) e *Human Genome Diversity Project* (Cann, 2002), é a maior cobertura de seu sequenciamento das populações do leste (LWK) e oeste (YRI) africano e do norte (CEU) e sul (IBS) europeu. Com isso, ao extrairmos os dados dos SNPs da versão v3.20101123.snps\_indels\_svs.genotypes da fase 1 do *1000 Genomes Project*, mantivemos uma extensa representatividade da variabilidade genética.

##### **3.1.1.2 Projeto EPIGEN-Brasil**

O projeto EPIGEN-Brasil envolve dados genéticos de 6.485 indivíduos das três maiores coortes populacionais brasileiras, representativas de três regiões brasileiras: Salvador (Nordeste; Barreto et al., 2006), Bambuí (Sudeste; Lima-Costa, Firmo e Uchoa, 2011) e Pelotas (Sul; Victora, 2006) (figura 4). Nesse projeto, 2,5 milhões de SNPs foram genotipados através do *array* HumanOmni2.5 da Illumina, visando as variantes genéticas raras. Além disso, um total de 267 amostras (90 de Bambuí, 90 de Salvador e 87 de Pelotas) foram genotipadas para 5 milhões de SNPs, através do *array* HumanOmni5.0 da Illumina, que abrange os SNPs do HumanOmni2.5 e uma

extensa cobertura de variantes comuns em todo o genoma humano. Por fim, das 267 amostras genotipadas para os 5 milhões de SNPs, 30 (10 de Bambuí, 10 de Salvador e 10 de Pelotas) têm seus genomas completos sequenciados pela estratégia *paired-end* da Illumina, que oferece ampla cobertura do genoma humano com alta capacidade de detectar variantes genéticas. Neste trabalho, nós utilizamos os dados integrados dos *arrays* HumanOmni 2.5 e 5.0.



Fonte: The Brazilian EPIGEN Initiative

Figura 4 - Localização das três maiores coortes populacionais brasileiras, Salvador (BA), Bambuí (MG) e Pelotas (RS), pertencentes ao Projeto EPIGEN-Brasil.

### 3.1.2 Verificação dos sexos dos indivíduos

Os sexos dos brasileiros foram inicialmente obtidos de dados coletados nas coortes do EPIGEN-Brasil. Verificamos essa informação geneticamente através do Plink (Purcell et al., 2007), que se baseia nas taxas de homozigosidade em SNPs do cromossomo X para confirmar o sexo. Um indivíduo é identificado como homem se a estimativa de homozigosidade é maior que 0,8, e como mulher se for menor que 0,2.

### 3.1.3 Exclusão dos aparentados

A exclusão de aparentados foi realizada através do programa NaTora, desenvolvida no nosso Laboratório de Diversidade Genética Humana (LDGH) da Universidade Federal de Minas

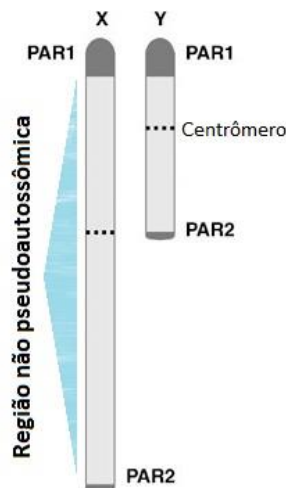
Gerais (UFMG), e que identifica os indivíduos centrais das redes de parentes e, com isso, permite retirar o parentesco presente na amostra excluindo o menor número possível de indivíduos. Em Salvador foram excluídos 63 indivíduos, em Pelotas foram 83 e em Bambuí, que apresenta uma alta estruturação familiar (Kehdy et al., 2015), 516 indivíduos foram retirados das análises. A tabela 1 descreve o *dataset* final utilizado no estudo, que apresenta dados de 5.823 indivíduos não aparentados das coortes brasileiras e de 389 indivíduos das populações parentais.

**Tabela 1. Amostras de homens e mulheres não aparentados por população e bancos de dados de origem.**

População	Número de mulheres	Número de homens	TOTAL AMOSTRAS	Banco de dados
<b>Africana</b>				
<b>Leste:</b> LWK (indivíduos de Luhya em Webuye, Quênia)	49	48	97	1000 Genomes Project (fase 1)
<b>Oeste:</b> YRI (iorubás em Ibadan, Nigéria)	45	43	88	1000 Genomes Project (fase 1)
<b>Europeia</b>				
<b>Norte:</b> CEU (residentes de Utah com ancestralidade do Norte e Oeste da Europa)	40	45	85	1000 Genomes Project (fase 1)
<b>Sul:</b> IBS (população Ibérica na Espanha)	7	7	14	1000 Genomes Project (fase 1)
<b>Nativa Americana</b>				
Ashaninka – Peru	22	22	44	LDGH – UFMG
Shimaa – Peru	24	21	45	LDGH – UFMG
Aymaras – Peru	7	9	16	LDGH – UFMG
<b>TOTAL PARENTAIS</b>	<b>194</b>	<b>195</b>	<b>389</b>	-
<b>Brasileira</b>				
<b>Nordeste:</b> Salvador	571	674	1.245	EPIGEN – Brasil
<b>Sudeste:</b> Bambuí	553	372	925	EPIGEN – Brasil
<b>Sul:</b> Pelotas	1820	1833	3.653	EPIGEN – Brasil
<b>TOTAL BRASILEIROS</b>	<b>2.944</b>	<b>2.879</b>	<b>5.823</b>	-
<b>TOTAL AMOSTRAS</b>	<b>3.138</b>	<b>3.074</b>	<b>6.212</b>	-

### 3.1.4 Exclusão de regiões de homologia entre os cromossomos X e Y

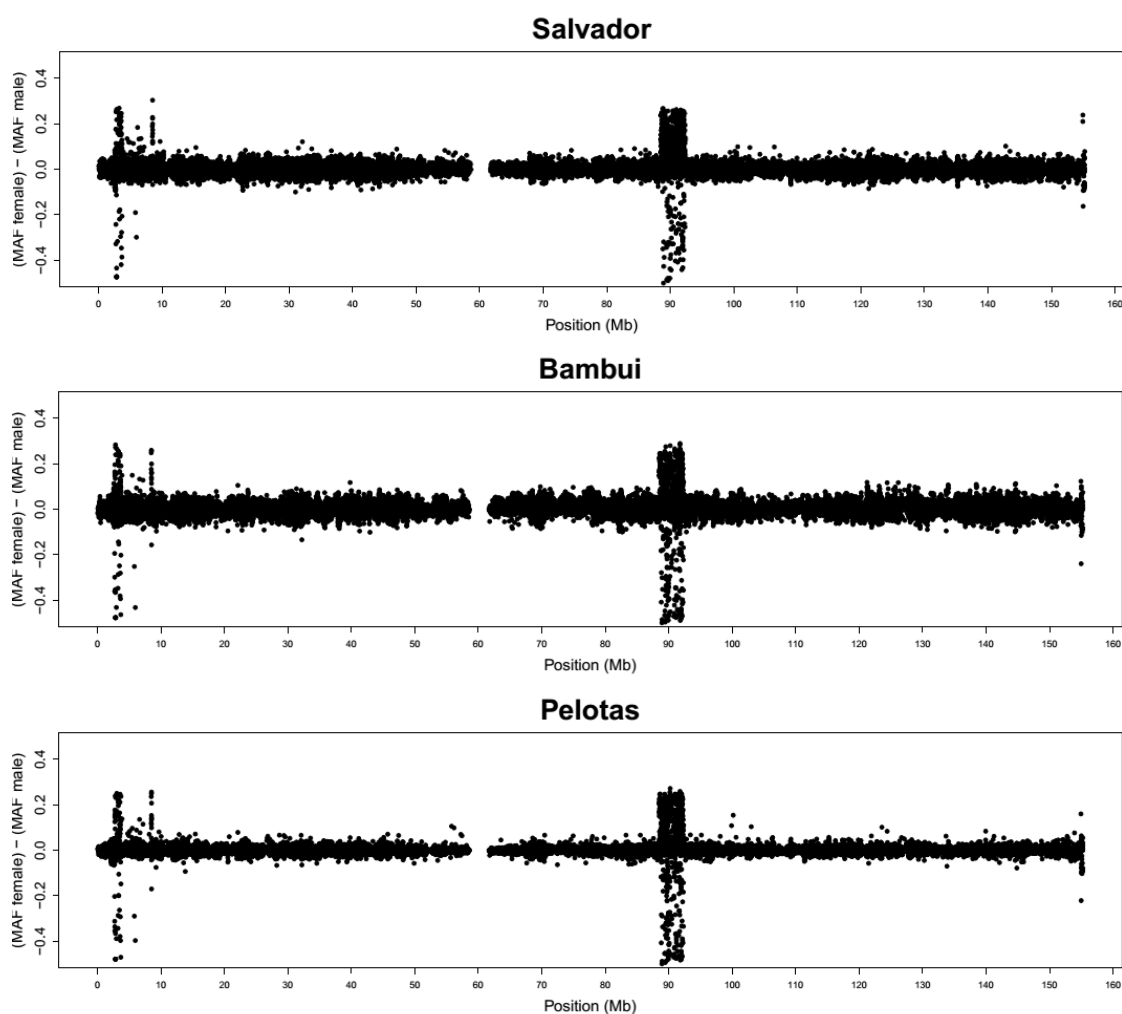
O cromossomo X é tradicionalmente dividido em uma extensa região não pseudoautossômica e duas regiões pseudoautossômicas (PAR) denominadas PAR1 e PAR2 (figura 5). As regiões PAR apresentam sequências homólogas entre os cromossomos X e Y e, por isso, exibem um padrão de herança semelhante aos dos cromossomos autossômicos, com ocorrência de eventos de recombinação em homens e mulheres (Flaquer et al., 2008). No presente estudo, assim como é visto tradicionalmente em estudos anteriores de genética de população no cromossomo X (Ramachandran et al., 2004; Casto et al., 2010; Bryc et al., 2015; Kehdy et al., 2015), incluímos apenas os dados da região não pseudoautossômica, que apresenta os padrões de herança tradicionalmente associados ao cromossomo X, em que a recombinação é prevista apenas em mulheres. Com isso, evitamos vieses nas frequências alélicas e genotípicas devido às recombinações nas regiões PAR1 e PAR2.



Adaptado de Flaquer et al. (2008)

**Figura 5. Estrutura dos cromossomos sexuais X e Y e a localização das regiões PAR1 e PAR2, e não pseudoautossômica.**

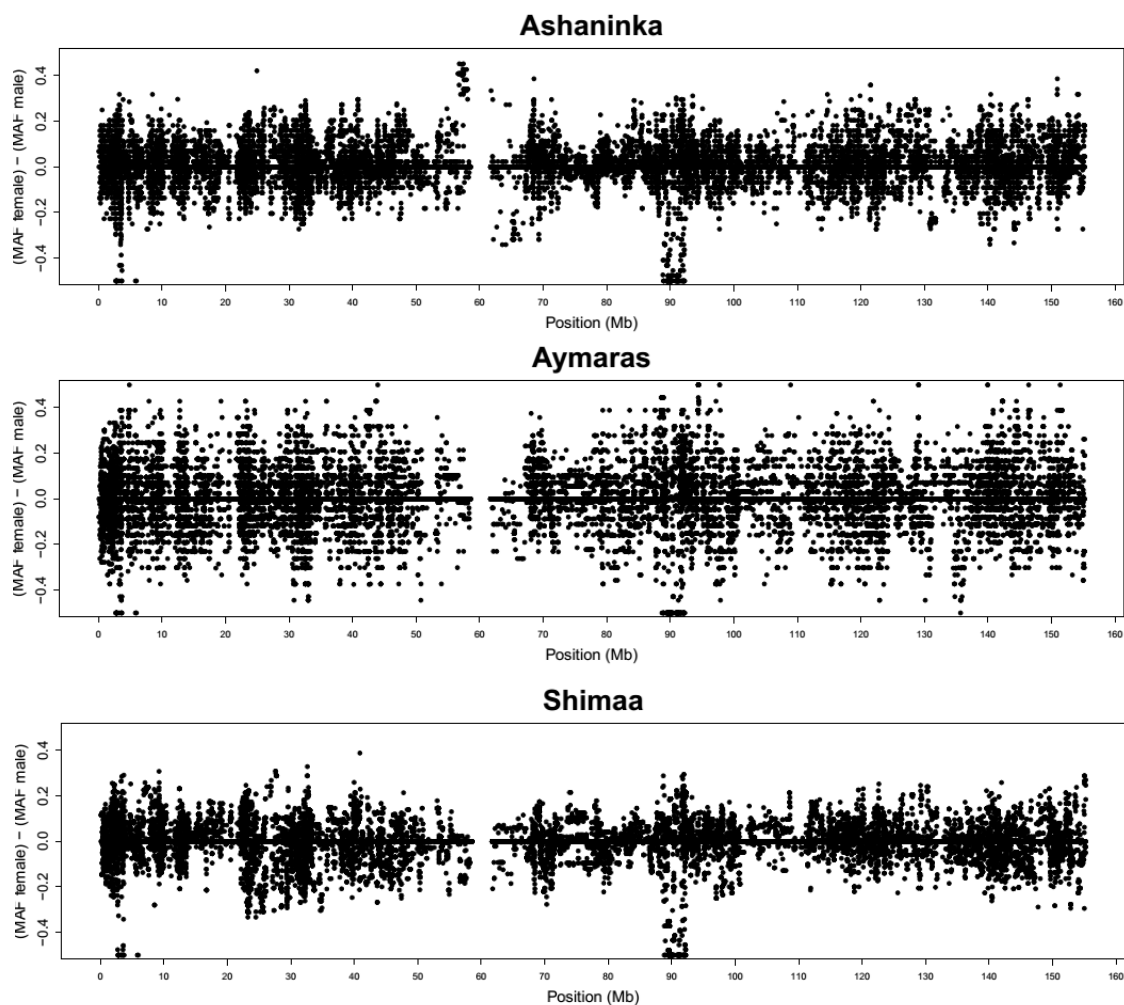
Porém, ao iniciar as análises das frequências alélicas e genotípicas na população brasileira, observamos quatro posições da região não pseudoautossômica do cromossomo X (i. 2,68-3,87Mb, ii. 5,81-8,52Mb, iii. 88,4-93,0Mb e iv. >154,93Mb) que apresentam grandes diferenças nas frequências dos alelos de menor frequência (*Minor Allele Frequency* - MAF) entre mulheres e homens (figura 6). Por esse ser um fenômeno ainda não reportado nessas regiões, realizamos análises adicionais em busca de compreender os motivos que ocasionam a essa interessante variabilidade intersexual nessas posições.



**Figura 6. Diferença entre MAF de 46.925 SNPs de mulheres e homens em relação à posição cromossômica (Mb) da região não pseudoautossômica do cromossomo X nas coortes brasileiras.**

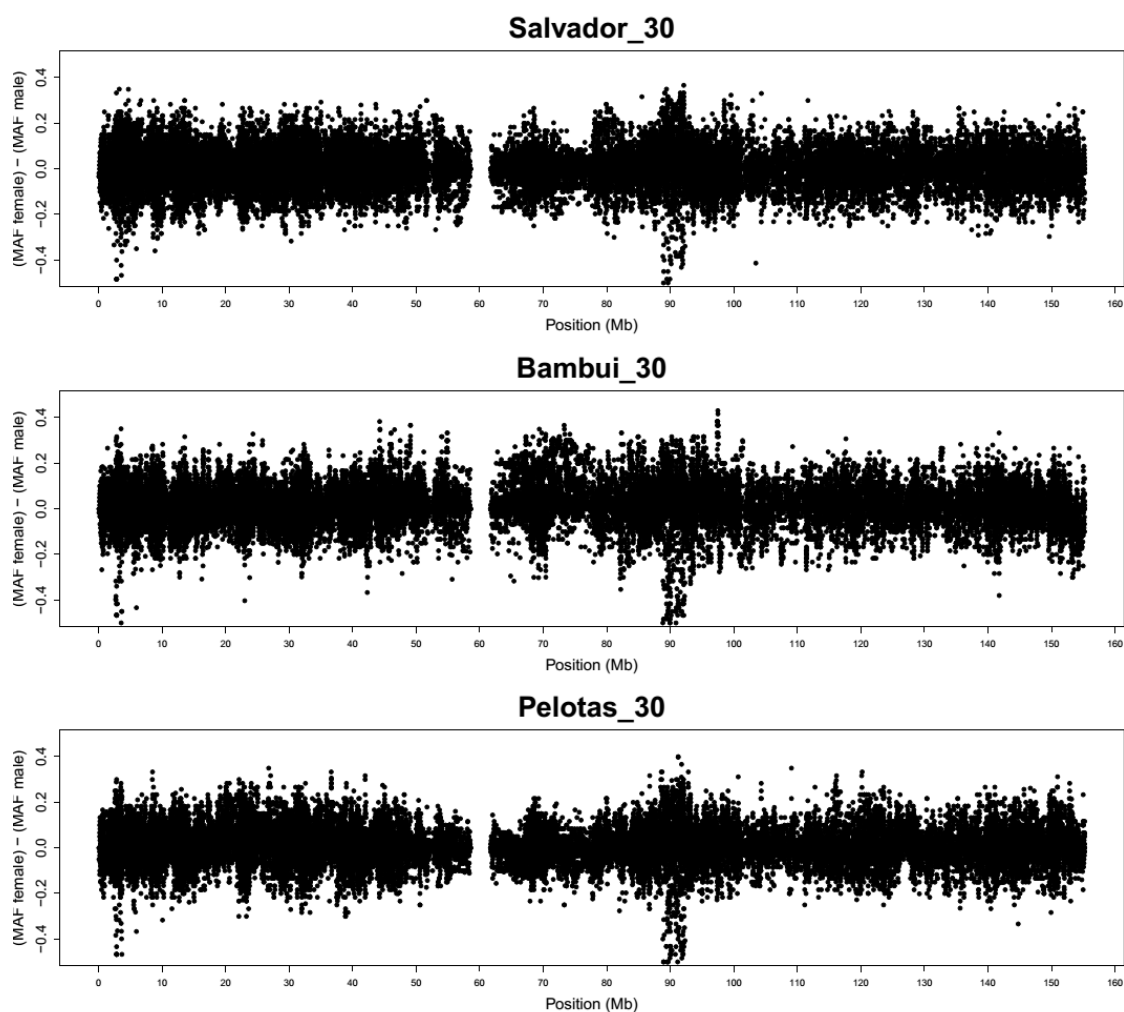
Iniciamos nossas análises verificando que essa distribuição pontuada da variabilidade intersexual das frequências alélicas não é observada nas populações parentais europeias (figura S1) e africanas (figura S2). Nas populações nativas americanas (figura 7), apesar de menos perceptível, ainda se destacam as posições (i) 2,68-3,87Mb e (iii) 88,4-93,0Mb.





**Figura 7.** Diferença entre MAF de mulheres e homens em relação à posição (Mb) no cromossomo X nas populações parentais nativas americanas.

Também verificamos se a limitada amostragem das populações parentais afetou a capacidade de detecção dessas distribuições diferenciadas das frequências dos alelos. Para isso, reanalisamos as frequências das coortes brasileiras considerando apenas 30 homens e 30 mulheres, selecionados aleatoriamente. A figura 8 demonstra que as posições de grande variabilidade entre os sexos se tornam bem menos evidentes quando diminuimos o número de amostras consideradas, demonstrando a importância de grandes amostragens para tornar detectáveis sutilezas nas frequências alélicas das populações. Contudo, apesar de menos perceptíveis, duas regiões (i. 2,68-3,87Mb e iii. 88,4-93,0Mb) ainda se destacam, o que exclui a possibilidade de que essas diferenças intersexuais sejam consequências apenas de amostragem desigual dos sexos.



**Figura 8.** Diferença entre MAF de 30 mulheres e 30 homens em relação à posição cromossômica (Mb) da região não pseudoautossômica do cromossomo X.

As duas posições (i. 2,68-3,87Mb e iii. 88,4-93,0Mb) que se mantiveram com grandes desvios intersexuais, mesmo com reduzida amostragem, são as mesmas observadas nas populações nativas americanas. Com isso, consideramos a possibilidade de que as diferenças nas frequências estejam relacionadas à regiões cromossômicas com predomínio dessa ancestralidade. Através da análise de ancestralidade local, observamos que nenhuma das quatro posições está localizada em regiões que apresentam desvios significativos, comparado a média cromossômica, das médias de ancestralidade nativa americana (tabela S1). Essas regiões também não estão associadas com regiões de predomínio de ancestralidade europeia (tabela S2) ou africana (tabela S3). Excluimos, portanto, a possível influência direta das contribuições das populações parentais para a manutenção dessas diferenças entre os sexos.

Por fim, observamos que, entre as quatro regiões detectadas, uma delas (iv. >154,93Mb) apresenta o gene *SPRY3*, um *hotspot* de recombinação próximo à região pseudoautossômica PAR2 (tabela 2). Essa região apresenta uma incomum elevada variabilidade interindividual entre os homens na taxa de crossover (Sarbjana et al., 2012). Essa taxa variável entre os homens pode, portanto, contribuir para a diversidade de frequências alélicas entre homens e mulheres nessa região.

Nas outras três posições (i. 2,68-3,87Mb; ii. 5,81-8,52Mb e iii. 88,4-93,0Mb), são reportadas regiões de homologia entre os cromossomos X e Y fora da região pseudoautossômica PAR1 e PAR2. Essa homologia pode propiciar a ocorrência ocasional de recombinação entre X e Y. Dos 37 genes e pseudogenes na região não pseudoautossômica que conhecidamente são compartilhados entre esses cromossomos sexuais (El-Mogharbel e Graves, 2008), em 16 deles observamos essa grande variabilidade intersexual nas frequências alélicas. Em 13 dessas regiões, há casos já reportados de recombinação X-Y, de não-inativação no X ou de diferenças intersexuais na expressão gênica, dosagem de produto gênico ou mutação (tabela 2), que podem contribuir para a variabilidade entre os sexos.

**Tabela 2. Regiões não pseudoautossômicas do cromossomo X com uma alta variabilidade nas frequências alélicas entre homens e mulheres.** As regiões (i) 2,68-3,87, (ii) 5,81-8,52 e (iii) 88,4-93,0 apresentam genes e pseudogenes compartilhados com o cromossomo Y. A região (iv) >154,93 contém um *hotspot* de recombinação.

Posições (Mb)	Gene (símbolo)		Características adicionais
	Cromossomo X	Cromossomo Y	
(i) 2,68-3,87	<i>GYG2</i>	Pseudogene	*
	<i>ARSD</i>	Pseudogene	Escapam da inativação do X (Meroni et al., 1996)
	<i>ARSE</i>	Pseudogene	
	<i>ARSF</i>	Pseudogene	
	<i>MRXA5</i>	Pseudogene	Mutações são mais frequentes em homens do que em mulheres (Xiong et al., 2012)
<i>PRKX</i>	<i>PRKY</i>	<ul style="list-style-type: none"> <li>• <b>Hotspot de recombinação X/Y</b></li> <li>• <b>Sequência nucleotídica que predispõe a eventos de quebra</b> (Klink et al., 1995)</li> <li>• <b>No cromossomo Y é o gene com a maior taxa de recombinação fora da região de pseudoautossômica.</b> (Schiebel et al., 1997)</li> </ul>	
(ii) 5,81-8,52	<i>NLGN4X</i>	<i>NLGN4Y</i>	*
	<i>STS</i>	Pseudogene	Escapam da inativação do X (Craig e Tolley, 1986)
	<i>VCX3A</i>	<i>VCY</i>	Expressos exclusivamente em células germinativas masculinas (Lahn, 2000)
	<i>VCX</i>	<i>VCY</i>	
	<i>VCX2</i>	<i>VCY</i>	
	<i>VCX3B</i>	<i>VCY</i>	
<i>KAL1</i>	Pseudogene	<ul style="list-style-type: none"> <li>• Escapam da inativação do X</li> <li>• Dimorfismo sexual na dosagem do produto gênico (Dodé et al., 2003)</li> </ul>	
(iii) 88,4-93,0	<i>bB348B13.2</i>	<i>bB348B13.2</i>	*
	<i>TGIF2LX</i>	<i>TGIF2LY</i>	<ul style="list-style-type: none"> <li>• Expressos especificamente nos testículos</li> <li>• <b>Maior região não pseudoautossômica compartilhada entre X/Y</b> (Blanco-Arias, Sargent e Affara, 2002)</li> <li>• <b>Proposto como sendo uma região pseudoautossômica denominado PAR3.</b> (Veerappa, Padakannaya e Ramachandra, 2013)</li> </ul>
	<i>PCDH11X</i>	<i>PCDH11Y</i>	Dimorfismo sexual no nível de expressão (Lopes et al., 2006)
(iv) > 154,93	<i>SPRY3</i>	-	<i>Hotspot</i> de recombinação (Sarbjana et al., 2012)

\*Não reportado.

As posições (i) 2,68-3,87 Mb e (iii) 88,4-93,0 Mb, que se destacam nas populações nativas americanas e também nas análises das coortes brasileiras com amostragem reduzida, apresentam alta taxa de recombinação entre os cromossomos X e Y. A primeira posição (2,68-3,87 Mb) contém o gene *PRKX*, um *hotspot* de recombinação. O seu homólogo no cromossomo Y, o *PRKY*, é o gene não pseudoautosômico do Y com as maiores taxas de recombinação. A segunda posição (88,4-93,0 Mb) apresenta homologia >98%, sendo por isso considerado a maior região não pseudoautosômica compartilhada entre X/Y (Blanco-Arias, Sargent e Affara, 2002), além de exibir frequência de recombinação semelhantes ao observado na região pseudoautosômica PAR2. Por isso, Veerappa, Padakannaya e Ramachandra (2013) propuseram esta como uma nova região pseudoautosômica denominada PAR3. Consequentemente, essas características podem propiciar a existência de grande variabilidade intersexual nas frequências alélicas nessas posições.

Em suma, evidenciamos a influência de recombinações na região não pseudoautosômica entre os cromossomos X/Y nas frequências alélicas de homens e de mulheres. Os desvios nas frequências dos alelos podem enviesar as análises de genética de populações e, portanto, essas quatro regiões foram excluídas de todas as análises do presente estudo. Com isso, o banco de dados final utilizado considera 37.457 SNPs no cromossomo X compartilhados entre os brasileiros, africanos, europeus e nativos americanos.

### **3.1.5 Inferência de fase e imputação**

Para que seja realizada a inferência de ancestralidade local (seção 3.2), foi primeiro necessário inferir a fase dos dados, ou seja, transformar os dados que estão organizados em genótipos em haplótipos. Para isso, utilizamos o algoritmo implementado no programa SHAPEIT, que considera as especificidades dos dados do cromossomo X e infere a fase nos dados de mulheres e apenas completa os dados faltantes em amostras de homens (O'Connell et al., 2014) utilizando as próprias amostras dos brasileiros como referência, em um processo denominado imputação (figura 9). As proporções totais de dados imputados foram 0,026% em Bambuí, 0,029% em Salvador e 0,024% em Pelotas.

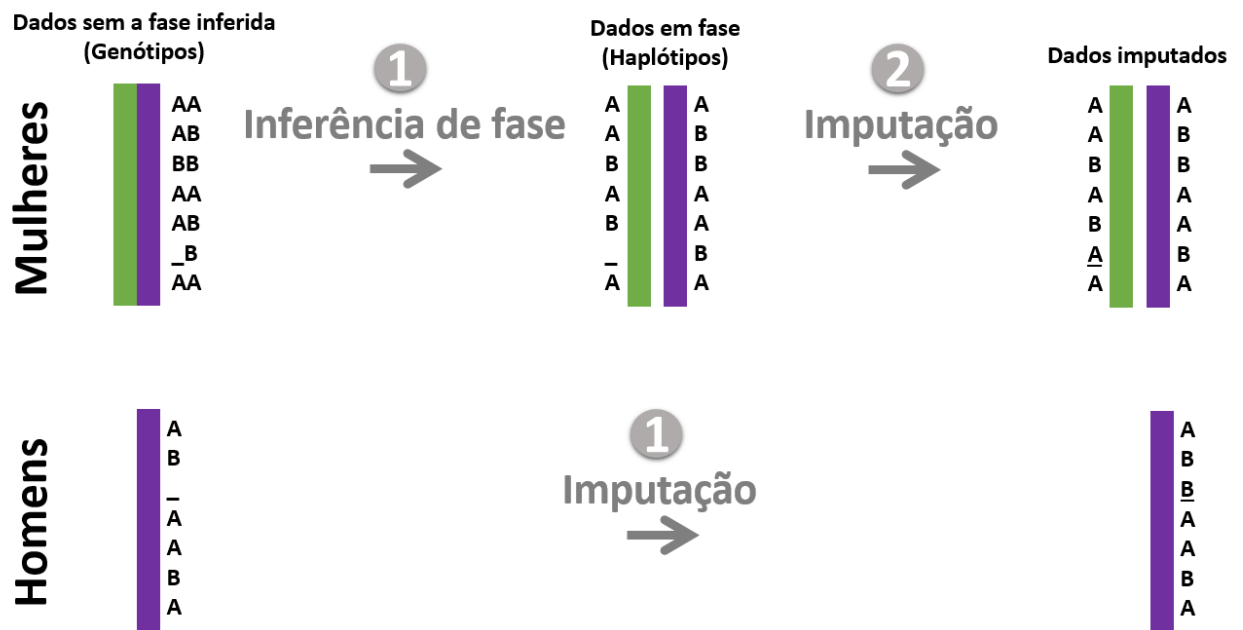


Figura 9. Esquema didático da inferência de fase e imputação no cromossomo X através do SHAPEIT.

### 3.1.6 Lidando com os cromossomos X hemizigotos dos homens

Os dados de genotipagem do cromossomo X dos homens das populações parentais e brasileiras são comumente armazenados em bancos de dados congelados com os seus genótipos duplicados, já que a grande maioria dos programas atuais de genética de populações não lida com dados haploides. Dessa forma, um indivíduo hemizigoto com genótipo “A C C”, é armazenado como diploide homocigoto “AA CC CC”. Esses genótipos duplicados podem acrescentar um viés em nossas análises de genética de populações, que envolvem observações de frequências alélicas e genotípicas, inflando a quantidade de homocigotos nos nossos dados e subestimando a informação de heteroziguidade da população. Por isso, para as análises de estatísticas F (seção 3.5), cujo objetivo é avaliar precisamente a homociguidade da população brasileira, utilizamos apenas os dados das mulheres.

## 3.2 Inferência da ancestralidade local, individual e populacional

A inferência da ancestralidade individual baseada no cromossomo X e a ancestralidade local (de pequenos conjuntos de SNPs ao longo de todo o cromossomo) foi realizada pelo PCAdmix (Brisbin et al., 2012) com K=3 grupos ancestrais (europeu, africano e nativo americano). Esse programa apresenta um algoritmo que utiliza os Componentes Principais para separar as amostras considerando a estrutura genética da população. Com base nos valores dessa análise, é inferida a ancestralidade de cada indivíduo e também identificadas as variantes genéticas mais informativas

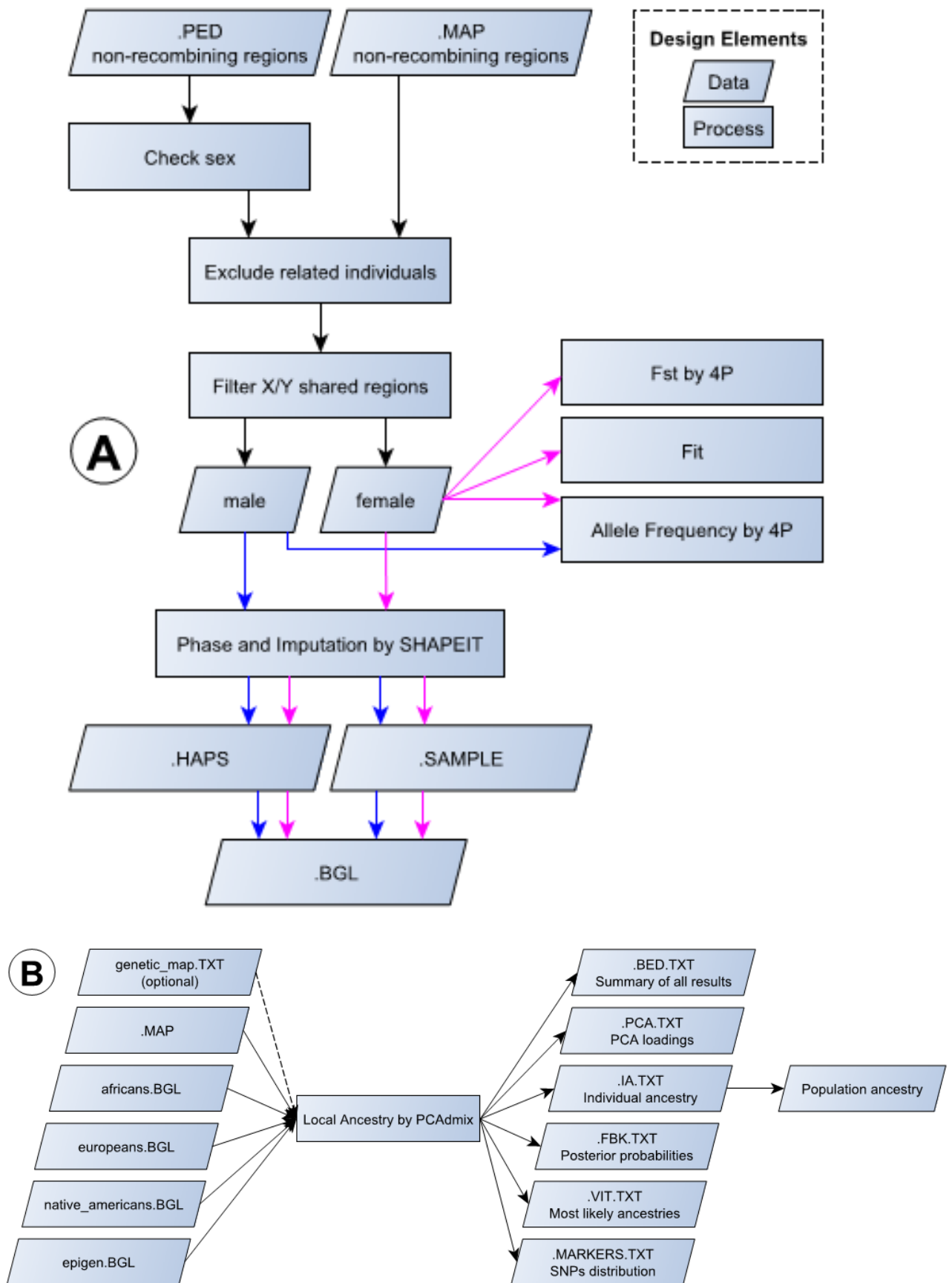
sobre a ancestralidade em cada conjunto de SNPs. Posteriormente, o programa infere estatisticamente a ancestralidade de cada conjunto de SNPs através do modelo oculto de Markov (Brisbin et al., 2012).

O uso do PCAdmix em dados do cromossomo X foi inédito, exigindo testes para definição dos parâmetros que melhor se adequam as especificidades desse cromossomo. Utilizamos o parâmetro “-wseg”, que dispensa o uso de mapas genéticos externos. Neste parâmetro, indicamos o tamanho mínimo de cada conjunto de SNPs e o programa automaticamente define o tamanho ideal para cada conjunto com base em um mapa de recombinação gerado a partir das amostras fornecidas. Esse parâmetro é usualmente utilizado para dados de exomas, mas também permitiu lidar com os densos agrupamentos de SNPs do cromossomo X relatados em Amos (2010). Em nossa análise, o limite mínimo de SNPs foi definido de acordo com o recomendado pelos autores Brisbin et al. (2012), que constataram que conjuntos com o mínimo de 20 SNPs apresentam um aumento considerável de acurácia.

Esse algoritmo do PCAdmix analisa separadamente cada cromossomo, permitindo que as inferências sejam realizadas nos cromossomos X hemizigotos dos homens, uma grande vantagem frente a métodos de estimativa de ancestralidade baseados em modelos estatísticos, como *Structure* (Pritchard, Stephens e Donnelly, 2000) e ADMIXTURE (Alexander, Novembre e Lange, 2009), que presumem diploidia dos indivíduos. Com isso, esses programas se limitam a analisar os cromossomos X diploides das mulheres, como realizado anteriormente através do ADMIXTURE (Kehdy et al., 2015).

Para a estimativa da ancestralidade populacional, calculamos as médias das ancestralidades individuais no cromossomo X inferidas para cada coorte brasileira. Validamos esses resultados comparando as proporções de ancestralidade individual e populacional inferidas nas mulheres pelo PCAdmix com as estimadas através do ADMIXTURE e publicadas em Kehdy et al. (2015). Observamos uma alta correlação em todas as inferências, sendo >99% para a ancestralidade africana, >98% para a europeia e >95% para a nativa americana. A menor correlação para os nativos americanos pode ser explicada pela baixa contribuição desse grupo parental para a proporção de ancestralidade dos brasileiros, e já foi vista em outros estudos com indivíduos de origem latina (Bryc et al., 2015).

As conversões de dados e os arquivos de entrada e de saída das análises estão resumidas na figura 10. As linhas de comando estão descritas e disponíveis em <http://1drv.ms/1SYoxS1>.



**Figura 10. Fluxograma das análises.** (A) Conversões de dados e análises de Equilíbrio de Hardy-Weinberg através das frequências alélicas,  $F_{ST}$  e  $F_{IT}$ . (B) Análises de ancestralidade local.



### **3.3 Análise da ocorrência de viés sexual através das médias de ancestralidade populacional do cromossomo X de homens e de mulheres em comparação aos autossomos**

Visando confirmar, na estrutura genética da população brasileira, a existência de viés sexual histórico de acasalamento, comparamos as médias populacionais de ancestralidade europeia, africana e nativa americana do cromossomo X de homens e de mulheres. O excesso, em relação aos autossomos, da média de uma ancestralidade no cromossomo X pode ser evidência de que essa população parental contribuiu com mais mulheres no processo de miscigenação, e a deficiência de uma ancestralidade pode indicar a maior contribuição de homens (Bryc et al., 2015).

### **3.4 Efeitos dos casamentos não casuais no Equilíbrio de Hardy-Weinberg em SNPs do cromossomo X**

Avaliamos a dimensão dos casamentos não casuais através dos desvios ocasionados no Equilíbrio de Hardy-Weinberg. Baseamos no princípio de que o Equilíbrio de Hardy-Weinberg se atinge gradualmente para SNPs no cromossomo X, com as frequências alélicas se aproximando gradativamente da igualdade em homens e mulheres sob regime de casamentos casuais. Com isso, avaliamos se esses desvios das proporções previstas por Hardy-Weinberg estão relacionados com a ancestralidade através da análise da correlação entre as diferenças entre os sexos nas frequências alélicas do alelo de menor frequência (MAF) e a diferenciação genética entre as populações parentais europeias e africanas (valores de  $F_{ST}$ ) em SNPs do cromossomo X. As frequências alélicas e os valores de  $F_{ST}$  (Weir e Cokerham, 1984) foram inferidos pelo programa 4P (Benazzo, Panziera e Bertorelle, 2014). Para validar as estimativas de  $F_{ST}$ , primeiro comparamos os seus valores aos reportados para os cromossomos autossômicos, observando o padrão já bem documentado de maiores médias para os SNPs do cromossomo X (Payseur e Nachman 2002; Akey et al. 2002; Schaffner 2004; Vicoso 2006; Lambert et al. 2010), que indicam a maior estruturação populacional no cromossomo X (tabela S4). Posteriormente, comparamos os nossos valores de  $F_{ST}$  com os estimados para as populações europeias, africanas e asiáticas do HapMap em Lambert et al. (2010), sendo observado uma grande semelhança entre as inferências (figura S3).

### 3.5 Análise da ocorrência de casamentos preferenciais entre indivíduos de mesma origem étnica através da relação entre $F_{IT}$ e $F_{ST}$

Verificamos também nos dados do cromossomo X se há alguma tendência recente de casamentos não casuais entre indivíduos com ancestralidades similares. Esses casamentos não aleatórios resultam em desvios significativos do esperado pelo princípio de Hardy-Weinberg em diversos SNPs do cromossomo X. Analisamos esse desvio através do  $F_{IT}$ , calculado para os dados das mulheres separadamente para cada SNP e para cada coorte brasileira como

$$\frac{(\textit{Heterozigosidade esperada} - \textit{Heterozigosidade observada})}{(\textit{Heterozigosidade esperada})}$$

A estimativa de cada coorte considera a média do  $F_{IT}$  de todos os 37.457 SNPs do cromossomo X analisados. Inferimos a diferenciação genética entre europeus e africanos através do  $F_{ST}$  (Weir e Cokerham, 1984) pelo programa 4P (Benazzo, Panziera e Bertorelle, 2014). Correlacionamos o  $F_{IT}$  com o  $F_{ST}$ , em cada população para cada SNP, pelo coeficiente de Spearman.

### 3.6 Análise das diferenças regionais na dinâmica de miscigenação através dos CSSA

Apuramos diferenças regionais na dinâmica de miscigenação ao nível de ancestralidade local do cromossomo X através da análise do CSSA, que são regiões cromossômicas contínuas nas quais foi inferida uma única ancestralidade. Em geral, quanto maior o tempo desde a miscigenação, maiores são as ocorrências de eventos de recombinação que quebram esses segmentos cromossômicos contínuos e, conseqüentemente, menores serão os CSSA (Liang e Nielsen, 2014). Com isso, analisamos as diferenças no tempo e nas contribuições parentais entre as distintas regiões do país, representadas pelas coortes de Salvador (Nordeste), Bambuí (Sudeste) e Pelotas (Sul).

## 4. RESULTADOS E DISCUSSÃO

### 4.1 Viés sexual durante o processo de miscigenação

A complexidade do processo de miscigenação da população brasileira é inicialmente evidenciada em nossos dados pela comparação entre a estrutura genética revelada no cromossomo X e nos autossomos de homens e de mulheres (tabela 3). Observamos o predomínio da contribuição europeia nos cromossomos autossômicos de Salvador (+14%), Bambuí (+13%) e Pelotas (+9%), e no cromossomo X das contribuições africana (+8% em Salvador, +4% em Bambuí e +3% em Pelotas) e nativo americana (+6% para Salvador, +8% para Bambuí e +6% para Pelotas). Essas diferenças podem evidenciar o padrão bem documentado de casamentos preferenciais entre homens com alta ancestralidade europeia e mulheres com predomínio da ancestralidade africana ou nativo americana durante o processo de miscigenação brasileira. Esse padrão se mantém considerando, separadamente, os cromossomos X dos homens e das mulheres, e as médias de cada cromossomo autossômico (tabela S2).

**Tabela 3. Contribuições médias das populações parentais europeias, africanas e nativo-americanas para a ancestralidade dos cromossomos autossômicos e X dos brasileiros.** Evidencia-se o excesso de ancestralidade africana e nativa americana no cromossomo X comparado aos cromossomos autossômicos.

Cromossomo	Salvador			Bambuí			Pelotas		
	Europeu	Africano	Nativo Americano	Europeu	Africano	Nativo Americano	Europeu	Africano	Nativo Americano
Autossômicos	43%	51%	6%	79%	15%	7%	76%	16%	8%
X	29%	59%	12%	66%	19%	15%	67%	19%	14%
Autossômicos – X <sup>1</sup>	↑ +14%	↓ -8%	↓ -6%	↑ +13%	↓ -4%	↓ -8%	↑ +9%	↓ -3%	↓ -6%

<sup>1</sup> Diferença entre as médias de ancestralidade dos cromossomos autossômicos e cromossomo X, sendo que as setas verdes indicam viés positivo, e as vermelhas, negativo.

Essa história demográfica diferenciada de homens e mulheres foi relatada em comparações entre marcadores genéticos do cromossomo X e dos autossomos em outros estudos com indivíduos de origem latina (Bryc et al., 2010; Bryc et al., 2015) e também foi evidenciada nos estudos do DNA mitocondrial e do cromossomo Y das coortes do projeto EPIGEN-Brasil (Kehdy et al., 2015). Observamos maior contribuição de mulheres nativas americanas para Bambuí e Pelotas, e mulheres

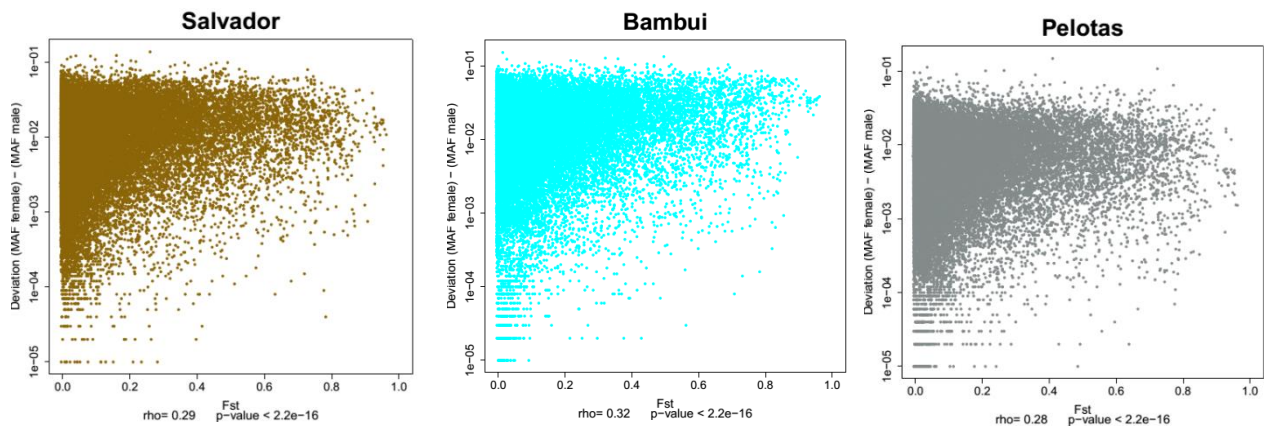
africanas para Salvador, condizente com as análises realizadas anteriormente no meu Trabalho de Conclusão de Curso, que utilizou apenas os dados do cromossomo X das mulheres (Sant Anna, 2014; Kehdy et al., 2015).

Contudo, é importante considerar que essa diferença nas contribuições parentais para o cromossomo X e os autossomos, assim como para o DNA mitocondrial e cromossomo Y, ao invés de resultante de viés sexual oposto entre os europeus e as populações africanas e nativas americanas, também pode ser resultado de um viés do mesmo sexo, porém em magnitudes diferentes (Goldberg e Rosenberg, 2015), entre europeus e uma ou ambas as populações dos outros grupos parentais (africanos e nativos americanos). Na história brasileira, há uma bem documentada ocorrência de casamento preferencial entre homens predominantemente europeus e mulheres predominantemente africanas ou nativas americanas, bem como de uma sobre representação masculina entre os escravos africanos, com uma relação estimada entre homens e mulheres no tráfico escravo de 4 para 1 (Ribeiro, 1995). Portanto, é ainda plausível a existência de viés masculino também em africanos, porém mais intenso em europeus.

#### **4.2 Efeitos de casamentos não casuais no Equilíbrio de Hardy-Weinberg**

Além desses casamentos históricos preferenciais, a dinâmica de miscigenação brasileira está associada também a uma tendência recente dos brasileiros em se relacionarem preferencialmente dentro de seus grupos étnicos (Petruccelli, 2001). Esses padrões de casamentos não aleatórios na população brasileira mantêm uma assinatura genética em SNPs do cromossomo X evidenciada por desvios nas proporções alélicas previstas por Hardy-Weinberg. Na figura 11, observamos a existência de diferenças nas frequências alélicas entre homens e mulheres principalmente naqueles SNPs com alta diferenciação genética entre as populações parentais europeias e africanas (evidenciada pelos altos valores de  $F_{ST}$ ) em Salvador ( $\rho=0,29$ ,  $p<2,2^{-16}$ ), Bambuí ( $\rho=0,32$ ,  $p<2,2^{-16}$ ) e Pelotas ( $\rho=0,28$ ,  $p<2,2^{-16}$ ).

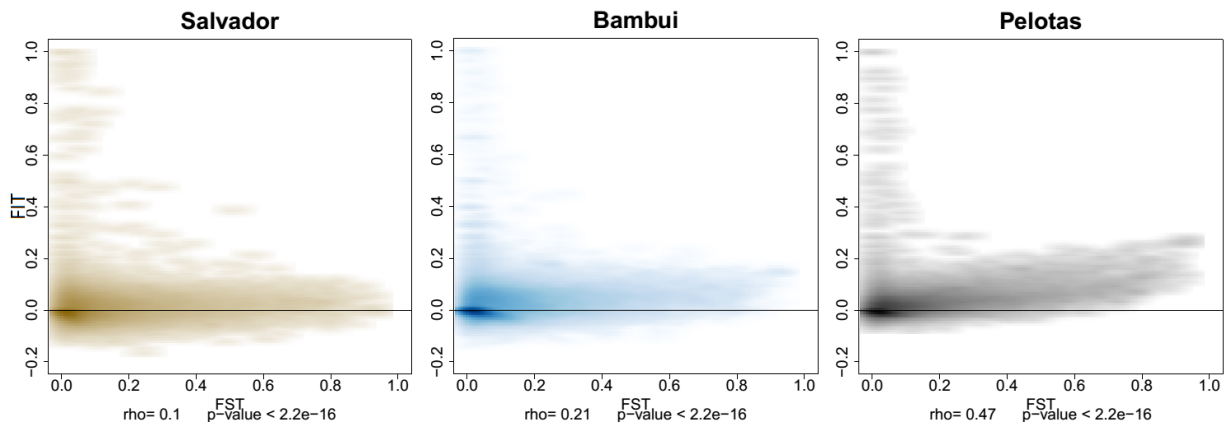
Essa relação positiva entre as diferenças intersexuais nas frequências alélicas e os valores de  $F_{ST}$  entre populações parentais corrobora com a ocorrência de casamentos não casuais na população brasileira relacionados à origem étnica. O viés sexual durante a história brasileira pode contribuir para gerar e manter diferenças nas frequências alélicas entre os sexos por várias gerações. A ocorrência de casamentos preferenciais entre indivíduos de mesma origem étnica na história recente da população brasileira pode colaborar para a existência de desvios do Equilíbrio de Hardy-Weinberg e, conseqüentemente, contribuir para manter essa distinção alélica entre os sexos.



**Figura 11.** Correlação entre as diferenças nas frequências alélicas do alelo de menor frequência (MAF) de homens e de mulheres brasileiras e os valores de  $F_{ST}$  entre as populações parentais europeias e africanas em SNPs do cromossomo X. Correlação positiva é observada nas três coortes.

### 4.3 Casamentos preferenciais entre indivíduos de mesma origem étnica

O padrão recente de casamentos preferenciais positivos para a ancestralidade também mantém um excesso de homozigotos (evidenciado pelos valores positivos de  $F_{IT}$ ) principalmente em SNPs do cromossomo X com alta diferenciação entre europeus e africanos (evidenciada pelos altos valores de  $F_{ST}$ ) (figura 12). Essa correlação positiva significativa entre  $F_{IT}$  das três coortes brasileiras e  $F_{ST}$  entre africanos e europeus, indicativo da ocorrência desses casamentos não casuais, foi menor em Salvador ( $\rho=0,10$ ,  $p < 2,2 \cdot 10^{-16}$ ), comparada a Bambuí ( $\rho=0,21$ ,  $p < 2,2 \cdot 10^{-16}$ ) e a Pelotas ( $\rho=0,47$ ,  $p < 2,2 \cdot 10^{-16}$ ). Esse padrão confirma o anteriormente observado para os cromossomos autossômicos (Salvador:  $\rho=0,08$ , Bambuí:  $\rho=0,18$  e Pelotas:  $\rho=0,38$ ) em Kehdy et al. (2015). A escolha do parceiro de acordo com a ancestralidade reflete provavelmente a escolha por variadas características fenotípicas, culturais e sociais associadas à ancestralidade, sendo que a própria existência do casamento preferencial explica em parte porque, após cinco séculos de miscigenação, ainda se observa correlação entre a ancestralidade e essas características.



**Figura 12.  $F_{it}$  (informativo da homoziguidade) versus  $F_{st}$  (informativo da ancestralidade) para Salvador, Bambuí e Pelotas.** O gráfico representa a associação positiva entre o excesso de homoziguidade e o informativo da ancestralidade calculado em indivíduos europeus e africanos.

#### 4.4 Diferenças regionais na dinâmica de miscigenação

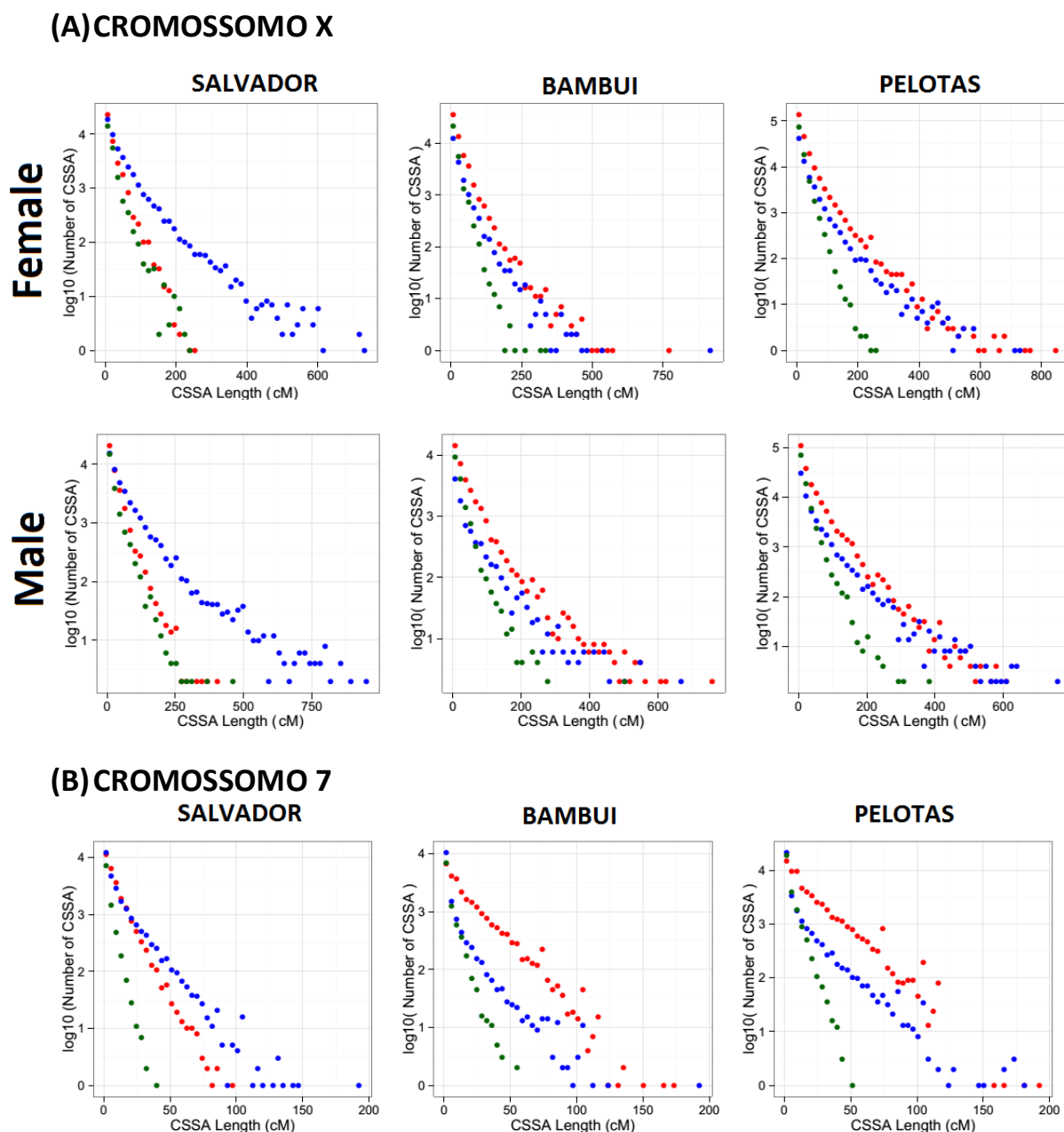
A análise do número e tamanho de segmentos cromossômicos de uma ancestralidade contínua específica (CSSA) evidenciou diferenças nas contribuições parentais para as diferentes regiões do país (figura 13). Os menores CSSA de ancestralidade nativa americana, tanto em homens quanto em mulheres, realça o maior tempo desde a sua participação na miscigenação. Desde a chegada dos primeiros colonizadores europeus no litoral da Bahia, em 1500, a população local nativa americana foi drasticamente reduzida devido a conflitos com os colonizadores e o contato com doenças europeias. Além disso, a partir de 1570, a mão-de-obra escrava indígena foi progressivamente substituída pela africana (Ferreira *et al.*, 2006). Esses acontecimentos históricos ocasionaram em uma contribuição nativa americana baixa e mais antiga que ocorreu principalmente logo após a conquista europeia das Américas (<http://brasil500anos.ibge.gov.br/>). Essas características da contribuição nativa americana também são observadas em análises de cromossomos autossômicos em diferentes regiões do país (Callegari-Jacques *et al.*, 2003; Godinho *et al.*, 2008; Lins *et al.*, 2010; Pena *et al.*, 2011; Manta *et al.*, 2013; Kehdy *et al.*, 2015).

Os nossos resultados refletem a participação das populações africanas para a formação das três populações brasileiras estudadas, mas principalmente para a coorte de Salvador (Nordeste), que apresenta um maior número de CSSA de origem africana, esperada à luz de sua história. Durante os séculos XVI e XIX, aproximadamente 3,5 milhões de africanos, oriundos principalmente do oeste, centro-oeste e sudeste da África, foram trazidos para o Brasil como mão-de-obra escrava (Ferreira *et al.*, 2006; <http://brasil500anos.ibge.gov.br/>). O desembarque ocorria principalmente nos portos de Salvador e do Rio de Janeiro, sendo posteriormente redistribuídos para o interior do

Brasil, com demanda no Sudeste e Sul do Brasil a partir de 1693-1695 (Florentino, Ribeiro e da Silva, 2004).

Devido aos menores CSSA europeus em Salvador, inferimos uma colonização europeia mais antiga em Salvador (Nordeste) e uma imigração mais recente em Bambuí (Sudeste) e Pelotas (Sul), consistente com os registros históricos. Até 1820, a economia se concentrava na região Nordeste do Brasil e atraía principalmente colonizadores de origem portuguesa. Entre 1820 e 1975, a mudança da economia do Brasil para o sudeste e sul atraiu, para essas regiões, imigrantes de origem portuguesa e italiana, seguidos por espanhóis, alemães, sírio-libaneses e japoneses. Com a proibição do tráfico de escravos em 1850, a elevada mortalidade da população negra, o forte estímulo à imigração europeia devido à expansão cafeeira e a intensa miscigenação entre brancos e negros, houve uma redução contínua da população negra e aumento progressivo da população branca (Ferreira *et al.*, 2006). Isso explica a contribuição alta e recente das populações europeias na diversidade genética de Bambuí (Sudeste) e Pelotas (Sul), também observada nos cromossomos autossômicos (Kehdy *et al.*, 2015).

Os CSSA do cromossomo 7 (figura 5), cujo número de pares de base se aproxima ao do cromossomo X, atingem tamanho máximo de 200cM, enquanto no cromossomo X os tamanhos ultrapassam 600cM. Essa diferença é esperada devido à distinta herança desse cromossomo sexual, em que apenas os cromossomos X das mulheres recombina-se em cada geração (Schaffner, 2004). Com isso, é previsto maiores segmentos cromossômicos inferidos para uma única ancestralidade (CSSA) no cromossomo X, que se mantiveram intactos com o passar do tempo devido ao menor número de eventos de quebra em relação aos autossomos.



**Figura 13. Distribuição dos CSSA no cromossomo X e cromossomo 7.** Os pontos vermelhos, azuis e verdes representam CSSA europeu, africano e nativo americano, respectivamente. Em (A) dados do cromossomo X em mulheres (superior) e homens (inferior). Em (B) as inferências do cromossomo 7 considerando tanto homens quanto mulheres.

Em suma, nossos resultados demonstram diferenças nas histórias demográficas de homens e de mulheres do Brasil e também os efeitos da miscigenação e dos casamentos não casuais na estrutura genética do cromossomo X. Primeiro, reforçamos a existência de viés sexual no processo de miscigenação brasileiro através de diferenças na estrutura genética do cromossomo X em relação aos cromossomos autossômicos de homens e de mulheres. Segundo, evidenciamos que a existência de casamentos não casuais leva a desvios nas frequências genotípicas previstas por



Hardy-Weinberg que resultam em diferenças nas frequências alélicas entre homens e mulheres principalmente em SNPs do cromossomo X com alta diferenciação genética entre europeus e africanos. Esses SNPs altamente diferenciados também apresentam um excesso de homozigosidade, o que apoia a existência de casamentos preferenciais entre brasileiros com a mesma origem étnica. Por fim, nossa análise da distribuição de segmentos cromossômicos de uma ancestralidade contínua específica (CSSA) do cromossomo X demonstra a limitada contribuição nativa americana para a população brasileira, a contribuição africana principalmente para o Nordeste brasileiro e a imigração recente europeia para o Sudeste e Sul do país.

## 5. CONCLUSÃO E PERSPECTIVAS

Em continuidade com os avanços realizados nos estudos dos cromossomos autossômicos e do cromossomo X das mulheres brasileiras realizados em Kehdy et al. (2015), o presente estudo inclui os dados de homens e reitera evidências para a ocorrência de um padrão histórico de casamento preferencial entre homens com alta ancestralidade europeia e mulheres com alta ancestralidade africana ou nativo americana. Porém, adiciona-se complexidade nessa interpretação ao propor a possibilidade da ocorrência de viés masculino também em populações parentais africanas.

Também apoiamos a existência de um padrão de casamentos entre brasileiros com ancestralidades similares, seja por questões históricas, geográficas, sociais ou aspectos econômicos e culturais. A ocorrência desses casamentos entre indivíduos com a mesma origem étnica, juntamente com a existência histórica de viés sexual, reforçam assinaturas genéticas em variantes do cromossomo X, sendo possível observar desvios nas proporções esperadas no Equilíbrio de Hardy-Weinberg que resultam em maiores diferenças nas frequências alélicas entre homens e mulheres em SNPs com alta diferenciação em europeus e africanos. Ademais, o excesso de homoziguidade nesses SNPs altamente diferenciados reforçam evidências para a existência de casamentos preferenciais relacionados à origem étnica.

Além desses padrões de casamentos não aleatórios, o presente estudo elucida também aspectos adicionais da dinâmica de miscigenação da população brasileira, como as disparidades regionais. A contribuição dos nativos americanos para a diversidade genômica brasileira se limitou ao início do processo de miscigenação, como confirmado no estudo de CSSA. Essas análises também evidenciam recente imigração europeia no Sudeste e Sul do país, e contribuição mais recente das populações africanas no Nordeste.

A continuidade das análises de CSSA permitirá usufruir ainda mais das potencialidades do cromossomo X em evidenciar a dinâmica sexo-específica da miscigenação da população brasileira. O cromossomo X possibilitará avaliar a magnitude do viés sexual ao longo dos cinco séculos de história brasileira, através da inferência da proporção de homens e mulheres contribuintes de cada população parental em diferentes momentos históricos. Para isso, adaptaremos o modelo evolutivo considerado em Kehdy et al. (2015), cuja história demográfica a ser avaliada considera três pulsos de miscigenação entre africanos, europeus e nativos americanos (i. 18-16, ii. 12-10 e iii. 6-4 gerações atrás), como mostrado na figura 14. Esse modelo é um compromisso entre a viabilidade

computacional e a complexidade da história da população brasileira (Gouveia et al., não publicado). Nesse cenário, estimaremos dezoito parâmetros  $M_{S,K,P}$ , onde  $M$  é a proporção de imigrantes do sexo  $S$  (homens e mulheres) da população ancestral  $K$  (africana, europeia ou nativa americana) contribuintes para a formação da população miscigenada no pulso  $P$  (i. 18-16, ii. 12-10 e iii. 6-4 gerações atrás). As distribuições *a posteriori* desses parâmetros serão estimadas utilizando a Computação Bayesiana Aproximada (*Approximate Bayesian Computation* - ABC). O método do ABC baseia-se na comparação de estatísticas sumárias estimadas dos dados observados com as estatísticas sumárias estimadas para uma série de simulações considerando um modelo evolutivo (no nosso caso, o modelo da figura 14). Aproveitaremos aqueles parâmetros  $M_{S,K,P}$  estimados que resultam em valores próximos as médias de ancestralidade populacional observada nas coortes brasileiras (tabela 3) como entrada para simular a distribuição dos CSSA (informativos sobre o tempo desde os eventos de miscigenação) e a relação entre os valores de  $F_{ST}$  entre as populações parentais e as diferenças de frequências alélicas entre homens e mulheres na população miscigenada (informativos sobre a magnitude do viés sexual) dos SNPs do cromossomo X. Os resultados dessas simulações serão comparados estatisticamente com os dados observados (figuras 11 e 13) na população brasileira e aqueles mais similares aos resultados reais serão utilizados para construir as distribuições *a posteriori* dos 18 parâmetros do modelo. Por fim, os valores dessas distribuições *a posteriori* representarão as proporções estimadas de imigrantes homens e imigrantes mulheres de cada população parental em três momentos da história brasileira (i. 18-16, ii. 12-10 e iii. 6-4 gerações atrás). Proporemos futuramente adaptações ao modelo do ABC, para incluir a informação de fatores evolutivos que afetam diferentemente o cromossomo X e os autossomos e são potenciais complicadores das estimativas, e também permitir a análise de mais cenários de miscigenação, que permitam considerar as peculiaridades da história da população brasileira.

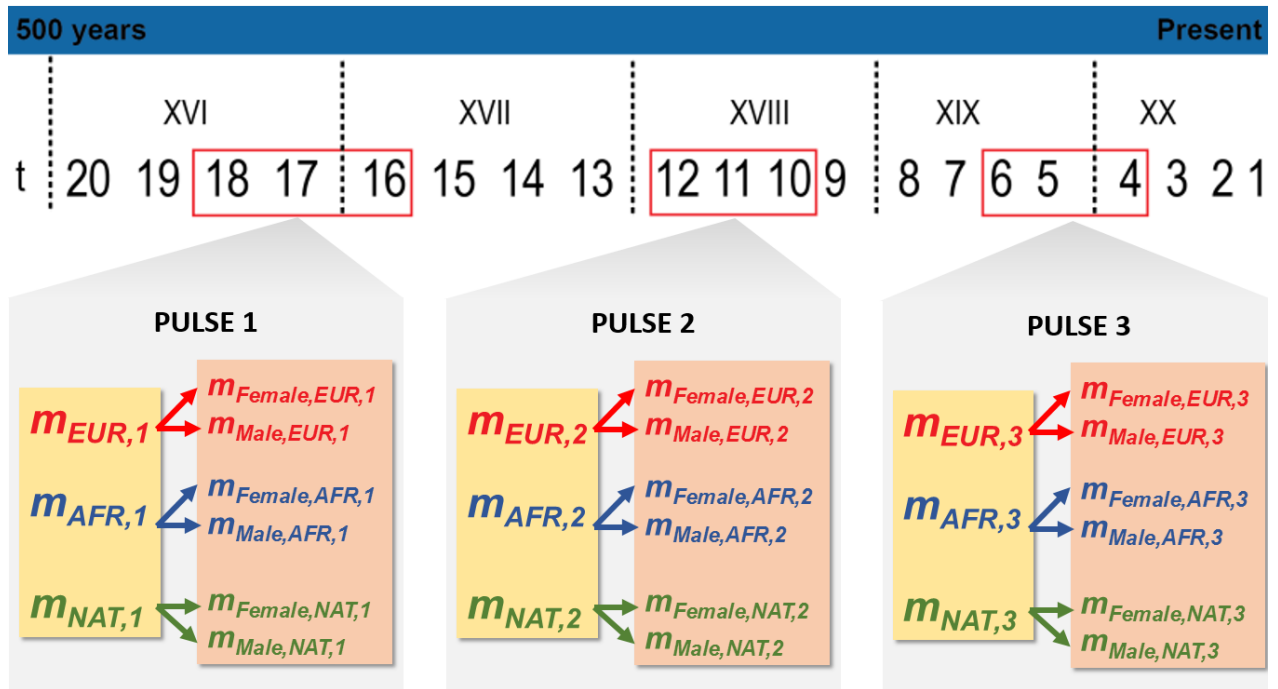


Figura 14. Modelo da dinâmica de miscigenação populacional do Brasil. Pulsos de miscigenação (1) mais antiga, 18-16 gerações atrás; (2) intermediária, 12-10 gerações atrás; e (3) recente, 6-4 gerações atrás. Em blocos amarelos, os 9 parâmetros  $M_{K,p}$  do modelo de Kehdy et al. (2015). Em blocos laranjas, os 18 parâmetros  $M_{S,K,p}$  propostos no projeto. t: gerações, na qual cada geração equivale a 25 anos.

Em suma, o projeto vem contribuindo para o entendimento dos eventos demográficos que moldaram a complexa estrutura genética da população brasileira. Espera-se que sua continuidade amplie o conhecimento sobre as complexas assinaturas do viés sexual de acasalamento no cromossomo X, obtendo estimativas das contribuições sexo-específicas europeias, africanas e nativas americanas para a população miscigenada brasileira em diferentes momentos de sua história.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- Akey, J. Interrogating a High-Density SNP Map for Signatures of Natural Selection. *Genome Research*, v. 12, n. 12, p. 1805-1814, 2002.
- Alexander, D.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, v. 19, n. 9, p. 1655-1664, 2009.
- Alves-Silva, J. et al. The Ancestry of Brazilian mtDNA Lineages. *The American Journal of Human Genetics*, v. 67, n. 2, p. 444-461, 2000.
- Amos, W. Even small SNP clusters are non-randomly distributed: is this evidence of mutational non-independence?. *Proceedings of the Royal Society B: Biological Sciences*, v. 277, n. 1686, p. 1443-1449, 2010.
- Barreto, M. et al. Risk factors and immunological pathways for asthma and other allergic diseases in children: background and methodology of a longitudinal study in a large urban center in Northeastern Brazil (Salvador-SCAALA study). *BMC Pulmonary Medicine*, v. 6, n. 1, p. 15, 2006.
- Benazzo, A.; Panziera, A.; Bertorelle, G. 4P: fast computing of population genetics statistics from large DNA polymorphism panels. *Ecol Evol*, v. 5, n. 1, p. 172-175, 2014.
- Blanco-Arias, P.; Sargent, C.; Affara, N. The human-specific Yp11.2/Xq21.3 homology block encodes a potentially functional testis-specific TGIF-like retroposon. *Mamm Genome*, v. 13, n. 8, p. 463-468, 2002.
- brasil500anos.ibge.gov.br. IBGE | Brasil: 500 anos de povoamento. Disponível em: <<http://brasil500anos.ibge.gov.br/>>. Acesso em: 18 jan. 2016.
- Brisbin, A. et al. PCAdmix: Principal Components-Based Assignment of Ancestry Along Each Chromosome in Individuals with Admixed Ancestry from Two or More Populations. *Human Biology*, v. 84, n. 4, p. 343-364, 2012.
- Bryc, K. et al. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proceedings of the National Academy of Sciences*, v. 107, n. Supplement\_2, p. 8954-8961, 2010.
- Bryc, K. et al. The Genetic Ancestry of African Americans, Latinos, and European Americans across the United States. *The American Journal of Human Genetics*, v. 96, n. 1, p. 37-53, 2015.

- Bustamante, C.; Ramachandran, S. Evaluating signatures of sex-specific processes in the human genome. *Nature Genetics*, v. 41, n. 1, p. 8-10, 2009.
- Callegari-Jacques, S. et al. Historical genetics: Spatiotemporal analysis of the formation of the Brazilian population. *Am. J. Hum. Biol.*, v. 15, n. 6, p. 824-834, 2003.
- Cann, H. A Human Genome Diversity Cell Line Panel. *Science*, v. 296, n. 5566, p. 261b-262, 2002.
- Carvalho-Silva, D. et al. The Phylogeography of Brazilian Y-Chromosome Lineages. *The American Journal of Human Genetics*, v. 68, n. 1, p. 281-286, 2001.
- Casto, A. et al. Characterization of X-Linked SNP genotypic variation in globally distributed human populations. *Genome Biol*, v. 11, n. 1, p. R10, 2010.
- Craig, I.; Tolley, E. Steroid sulphatase and the conservation of mammalian X chromosomes. *Trends in Genetics*, v. 2, p. 201-204, 1986.
- Dodé, C. et al. Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. *Nature Genetics*, v. 33, n. 4, p. 463-465, 2003.
- El-Mogharel, N.; AM Graves, J. X and Y Chromosomes: Homologous Regions. *Encyclopedia of life sciences*, p. 6, 2008.
- Ferreira, L. et al. Genomic ancestry of a sample population from the state of São Paulo, Brazil. *Am. J. Hum. Biol.*, v. 18, n. 5, p. 702-705, 2006.
- Flaquer, A. et al. The human pseudoautosomal regions: a review for genetic epidemiologists. *Eur J Hum Genet*, v. 16, n. 7, p. 771-779, 2008.
- Florentino, M., Ribeiro, A. e da Silva, D. (2004). Aspectos comparativos do tráfico de africanos para o Brasil. *Afro-Ásia*, 31, pp.83--126
- Godinho, N. et al. Regional patterns of genetic admixture in South America. *Forensic Science International: Genetics Supplement Series*, v. 1, n. 1, p. 329-330, 2008.
- Goldberg, A.; Rosenberg, N. Beyond 2/3 and 1/3: The Complex Signatures of Sex-Biased Admixture on the X Chromosome. *Genetics*, v. 201, n. 1, p. 263-279, 2015.
- Goldberg, A.; Verdu, P.; Rosenberg, N. Autosomal Admixture Levels Are Informative About Sex Bias in Admixed Populations. *Genetics*, v. 198, n. 3, p. 1209-1229, 2014.

- Gouveia et al. Approximate Bayesian Computation (ABC) Framework Based on Local Ancestry to Infer Admixture Events. (não publicado)
- Hartl, D.; Clark, A. Principles of population genetics. Tradução. Sunderland, Mass.: Sinauer Associates, 2007.
- Hedrick, P. Genetics of populations. Sudbury, Mass.: Jones and Bartlett Publishers, 2011.
- Holding, C. HapMap Project launched. *Genome Biology*, v. 4, p. spotlight-20031219-01, 2003.
- Kehdy, F. et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proceedings of the National Academy of Sciences*, v. 112, n. 28, p. 8696-8701, 2015.
- Klein, H.; Luna, F. Escravismo no Brasil. Tradução. São Paulo: EDUSP, 2010.
- Klink, A. et al. The human protein kinase gene PKX1 on Xp22.3 displays Xp/Yp homology and is a site of chromosomal instability. *Human Molecular Genetics*, v. 4, n. 5, p. 869-878, 1995.
- Krieger, H. et al. Racial admixture in north-eastern Brazil. *Annals of Human Genetics*, v. 29, n. 2, p. 113-125, 1965.
- Lahn, B. A human sex-chromosomal gene family expressed in male germ cells and encoding variably charged proteins. *Human Molecular Genetics*, v. 9, n. 2, p. 311-319, 2000.
- Lambert, C. et al. Highly Punctuated Patterns of Population Structure on the X Chromosome and Implications for african Evolutionary History. *The american Journal of Human Genetics*, v. 86, n. 1, p. 34-44, 2010.
- Liang, M.; Nielsen, R. The Lengths of Admixture Tracts. *Genetics*, v. 197, n. 3, p. 953-967, 2014.
- Lima-Costa, M.; Firmo, J.; Uchoa, E. Cohort Profile: The Bambui (Brazil) Cohort Study of Ageing. *International Journal of Epidemiology*, v. 40, n. 4, p. 862-867, 2011.
- Lins, T., Vieira, R., Abreu, B., Grattapaglia, D. e Pereira, R. (2010). Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. *american Journal of Human Biology*, 22(2), pp.187--192.
- Lopes, A. et al. Inactivation status of PCDH11X: sexual dimorphisms in gene expression levels in brain. *Hum Genet*, v. 119, n. 3, p. 267-275, 2006.

- Manta, F.N., Pereira, R., Vianna, R., de Araújo, A., Gitaí, D., da Silva, D., de Vargas Wolfgramm, E., da Mota Pontes, I., Aguiar, J., Moraes, M. e outros (2013). Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels. *PloS one*, 8(9), p.75145.
- McVean, G. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*, v. 491, n. 7422, p. 56-65, 2012.
- Meroni, G. Characterization of a cluster of sulfatase genes on Xp22.3 suggests gene duplications in an ancestral pseudoautosomal region. *Human Molecular Genetics*, v. 5, n. 4, p. 423-431, 1996.
- O'Connell, J. et al. A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genetics*, v. 10, n. 4, p. e1004234, 2014.
- Payseur, B.; Nachman, M. Natural selection at linked sites in humans. *Gene*, v. 300, n. 1-2, p. 31-42, 2002.
- Pena, S. et al. The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected. *PLoS ONE*, v. 6, n. 2, p. e17063, 2011.
- Petrucelli, J. Seletividade por cor e escolhas conjugais no Brasil dos 90. *Estudos Afro-Asiáticos*, v. 23, n. 1, 2001.
- Pritchard, K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics*, v. 155, p. 945-959, 2000.
- Purcell, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, v. 81, n. 3, p. 559-575, 2007.
- Ramachandran, S. et al. Robustness of the inference of human population structure: A comparison of X-chromosomal and autosomal microsatellites. *Hum Genomics*, v. 1, n. 2, 2004.
- Resque, R. et al. Estimates of interethnic admixture in the Brazilian population using a panel of 24 X-linked insertion/deletion markers. *Am. J. Hum. Biol.*, v. 22, n. 6, p. 849-852, 2010.
- Ribeiro, D. O povo brasileiro. Tradução. [São Paulo, Brazil]: Companhia das Letras, 1995.
- Salzano, F.; Bortolini, M. The evolution and genetics of Latin American populations. Tradução. Cambridge: Cambridge University Press, 2002.
- Sant Anna, H. Estrutura Genética da População Brasileira Inferida Pelo Cromossomo X. Graduado



[s.l.] Universidade Federal de Minas Gerais, 2014.

- Sarbajna, S. et al. A major recombination hotspot in the XqYq pseudoautosomal region gives new insight into processing of human gene conversion events. *Human Molecular Genetics*, v. 21, n. 9, p. 2029-2038, 2012.
- Schaffner, S. The X chromosome in population genetics. *Nat Rev Genet*, v. 5, n. 1, p. 43-51, 2004.
- Schiebel, K. et al. FISH localization of the human Y-homolog of protein kinase PRKX (PRKY) to Yp11.2 and two pseudogenes to 15q26 and Xq12&rarr;q13. *Cytogenetic and Genome Research*, v. 76, n. 1-2, p. 49-52, 1997.
- Sorenson Molecular Genealogy Foundation. SMGF. Disponível em: <<http://www.smgf.org/>>. Acesso em: 3 fev. 2016.
- University of California Museum of Paleontology's Understanding Evolution. <http://evolution.berkeley.ed>. Disponível em: <[http://evolution.berkeley.edu/evolibrary/news/071101\\_genealogy](http://evolution.berkeley.edu/evolibrary/news/071101_genealogy)>. Acesso em: 3 fev. 2016.
- Veerappa, A.; Padakannaya, P.; Ramachandra, N. Copy number variation-based polymorphism in a new pseudoautosomal region 3 (PAR3) of a human X-chromosome-transposed region (XTR) in the Y chromosome. *Functional & Integrative Genomics*, v. 13, n. 3, p. 285-293, 2013.
- Vicoso, B.; Charlesworth, B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet*, v. 7, n. 8, p. 645-653, 2006.
- Victoria, C. Cohort Profile: The 1982 Pelotas (Brazil) Birth Cohort Study. *International Journal of Epidemiology*, v. 35, n. 2, p. 237-242, 2006.
- Weir, B.; Cockerham, C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, v. 38, n. 6, p. 1358, 1984.
- Winkler, C.; Nelson, G.; Smith, M. Admixture Mapping Comes of Age. *Annual Review of Genomics and Human Genetics*, v. 11, n. 1, p. 65-89, 2010.
- Xiong, D. et al. Exome sequencing identifies MXRA5 as a novel cancer gene frequently mutated in non-small cell lung carcinoma from Chinese patients. *Carcinogenesis*, v. 33, n. 9, p. 1797-1805, 2012.

## ANEXOS

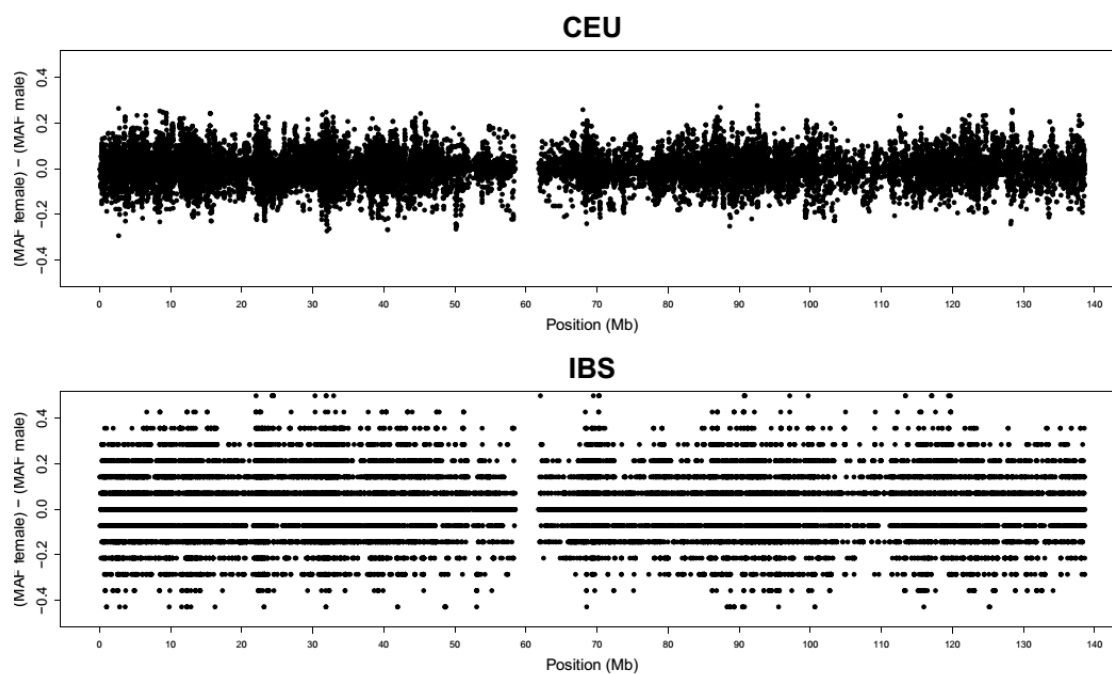


Figura S1. Diferença entre MAF de mulheres e homens em relação à posição (Mb) no cromossomo X nas populações parentais europeias.

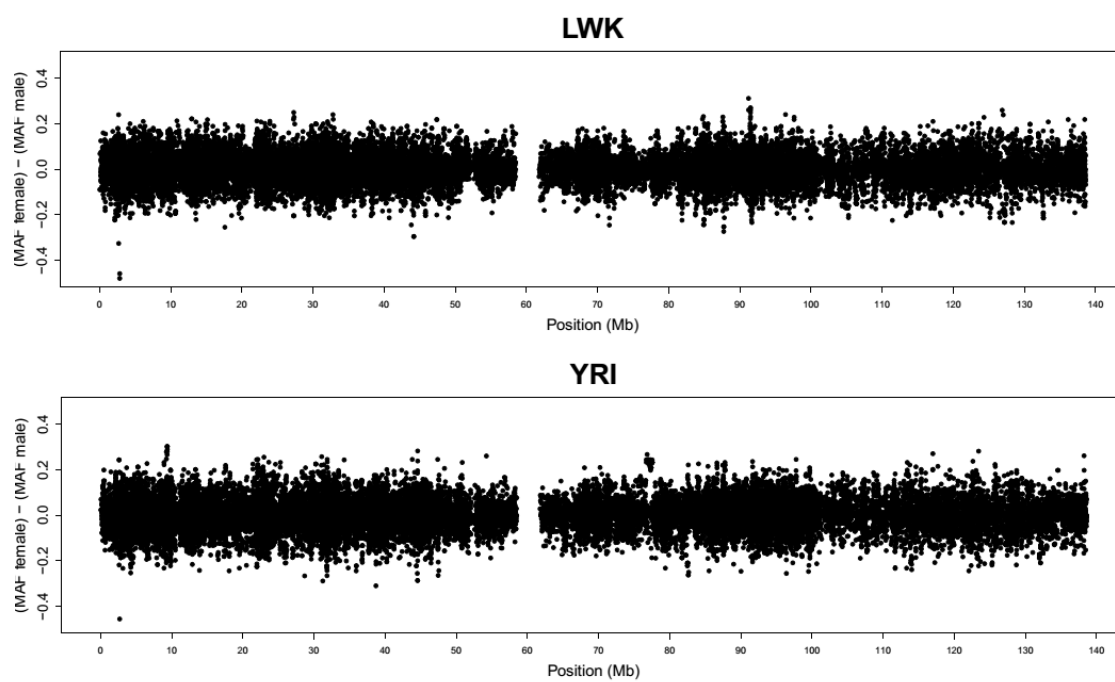
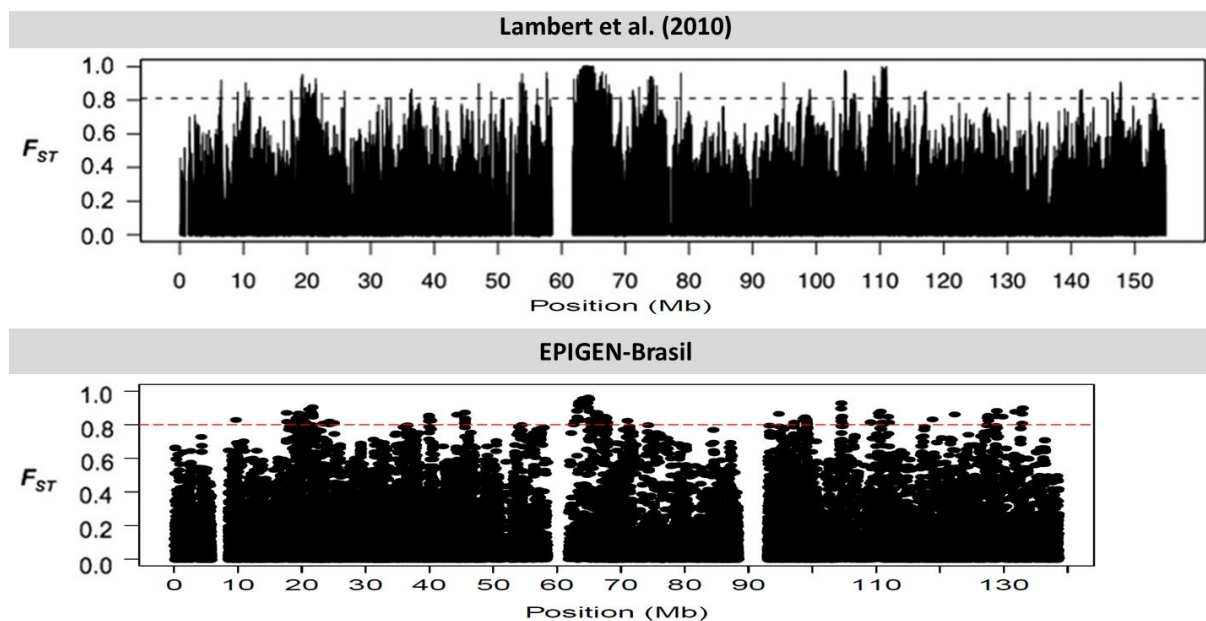


Figura S2. Diferença entre MAF de mulheres e homens em relação à posição (Mb) no cromossomo X nas populações parentais africanas.



Adaptado de Lambert et al. (2010)

**Figura S3. Distribuição cromossômica das estimativas de  $F_{ST}$  comparativa entre estudos.** Superior, inferências para populações europeias (CEU), africanas (YRI) e Asiáticas (CHB+JPT) do HapMap. Inferior, estimativas das populações europeias (CEU+IBS) e africanas (LWK+YRI) do 1000 Genomes Project utilizadas no presente estudo. Após filtragem de qualidade, nossos dados se limitam a 0-140Mb. A linha pontilhada representa o percentil 99 da distribuição empírica de  $F_{ST}$ . Há grande semelhança entre as inferências.

#### REFERÊNCIA BIBLIOGRÁFICA

Lambert, C. et al. Highly Punctuated Patterns of Population Structure on the X Chromosome and Implications for african Evolutionary History. *The American Journal of Human Genetics*, v. 86, n. 1, p. 34-44, 2010.

**Tabela S1. Regiões (em Mb) do cromossomo X com predomínio da ancestralidade nativa americana em análise com mulheres e homens separados e juntos.**

	Salvador			Bambuí			Pelotas		
	Mulher	Homem	Mulher e Homem	Mulher	Homem	Mulher e Homem	Mulher	Homem	Mulher e Homem
(1)	10,4	10,4	10,9	17,7	17,6	17,5	17,6	17,6	17,7
(2)	17,5	17,5	17,5	19,1	17,7	17,7	17,7	17,7	55,0
(3)	17,7	17,6	17,7	48,6	19,0	48,9	48,7	19,0	74,2
(4)	19,0	19,0	18,9	48,8	48,6	54,8	55,0	48,7	74,5
(5)	48,7	20,0	52,8	54,8	48,8	55,0	74,0	55,0	84,6
(6)	55,0	52,8	54,8	54,9	48,9	74,2	74,4	74,0	84,8
(7)	74,1	54,8	54,9	64,1	54,8	87,4	86,2	74,3	110,9
(8)	74,3	55,0	74,2	73,3	55,0	110,8	87,4	135,0	153,7
(9)	82,2	73,3	74,5	74,2	74,2	153,4	136,2	135,4	154,2
(10)	82,3	74,2	75,1	74,5	74,5	153,8	136,5	136,2	154,6
(11)	84,9	74,6	75,9	87,4	129,4	154,2	153,6	136,4	
(12)	136,7	75,2	153,7	136,3	135,4		154,0	153,9	
(13)	153,7	82,2	154,1	136,6	153,3		154,5	154,4	
(14)	154,2	86,2	154,6	153,8	153,7		154,8	154,7	
(15)	154,6	153,4		154,2	154,1				
(16)				154,6	154,6				

Tabela S2. Regiões (em Mb) do cromossomo X com predomínio da ancestralidade europeia em análise com mulheres e homens separados e juntos.

	Salvador			Bambuí			Pelotas		
	Mulher	Homem	Mulher e Homem	Mulher	Homem	Mulher e Homem	Mulher	Homem	Mulher e Homem
(1)	17,6	17,5	17,6	17,8	17,6	14,4	17,6	17,7	17,8
(2)	17,7	17,7	17,7	18,9	17,8	17,6	17,8	17,8	18,9
(3)	43,5	48,7	48,7	48,7	19,4	17,7	48,8	48,8	48,7
(4)	55,2	48,9	48,9	48,8	19,5	48,7	55,0	55,1	48,9
(5)	66,8	74,3	55,1	55,0	48,7	48,9	74,1	74,1	55,0
(6)	97,6	87,4	110,6	74,25	48,9	54,9	74,4	74,4	74,6
(7)	110,6	87,5	111,0	74,6	49,0	55,0	87,3	135,1	84,7
(8)	111,0	110,6	153,8	87,4	54,9	74,2	87,4	135,5	111,0
(9)	153,8	111,0	154,2	136,3	55,0	87,3	87,4	136,2	153,8
(10)	154,2	153,4	154,7	136,7	74,2	87,4	136,3	136,5	154,2
(11)	154,7			153,8	74,5	110,9	136,6	153,6	154,7
(12)				154,2	153,4	153,4	153,6	154,0	
(13)				154,7	153,7		154,0	154,4	
(14)					154,1		154,5	154,8	
(15)					154,6		154,9		

Tabela S3. Regiões (em Mb) do cromossomo X com predomínio da ancestralidade africana em análise com mulheres e homens separados e juntos.

	Salvador			Bambuí			Pelotas		
	Mulher	Homem	Mulher e Homem	Mulher	Homem	Mulher e Homem	Mulher	Homem	Mulher e Homem
(1)	10,5	19,0	17,7	20,9	19,0	20,9	62,9	19,0	18,9
(2)	18,9	19,1	18,8	21,1	20,7	48,7	74,1	63,0	48,7
(3)	19,1	52,8	19,0	21,5	20,9	48,9	74,4	74,1	48,9
(4)	48,6	74,3	52,8	52,8	21,1	53,0	86,3	74,4	51,8
(5)	48,8	74,6	55,0	55,0	21,3	63,0	135,2	74,7	63,0
(6)	55,0	82,2	74,2	74,2	48,7	74,2	135,6	94,9	67,0
(7)	74,1	86,3	74,6	129,4	48,9	110,9	154,0	135,0	74,2
(8)	74,4	135,4	76,0	135,4	51,9	135,3	154,5	135,5	85,3
(9)	74,7	153,8	135,4	135,7	129,5	135,7		153,6	111,0
(10)	135,4	154,2	135,7	136,3	135,4	154,2		154,0	129,5
(11)	135,7	154,7	154,2	136,7	135,8			154,4	135,4
(12)	154,2		154,7	153,8					135,7
(13)	154,7			154,2					

**Tabela S4. Distribuição cromossômica dos valores de  $F_{ST}$  para os autossomos e cromossomo X.** Comparativamente, o cromossomo X apresenta maiores médias e medianas de  $F_{ST}$ .

<b>Cromossomo</b>	<b>1000 Genomes Project</b>	
	<b>Autossômico</b>	<b>X</b>
<b>Número de SNPs analisados</b>	2.061.478	37.457
<b>Média</b>	0.092	0.122
<b>Mediana</b>	0.046	0.055
<b>Desvio Padrão</b>	0.118	0.161
<b>3º Quartil</b>	0.124	0.165



**Tabela S5. Contribuições médias das populações parentais europeias, africanas e nativo americanas para a ancestralidade dos cromossomos autossômicos e do cromossomo X das populações de Salvador, Bambuí e Pelotas.** Realçado em amarelo estão as médias encontradas no cromossomo 7 (tamanho em pares de bases: 159.138.663), que apresenta tamanho semelhante ao do cromossomo X (tamanho em pares de bases: 155.270.560). Evidencia-se o excesso de ancestralidade africana e nativa americana no cromossomo X comparado aos cromossomos autossômicos.

Cromossomo	Tamanho (mm)	Pares de bases	Salvador			Bambuí			Pelotas		
			Europeu	Africano	Nativo Americano	Europeu	Africano	Nativo Americano	Europeu	Africano	Nativo Americano
1	85	249.250.621	0.385	0.520	0.094	0.760	0.146	0.095	0.733	0.162	0.106
2	83	243.199.373	0.404	0.512	0.084	0.764	0.156	0.080	0.745	0.163	0.092
3	67	198.022.430	0.399	0.525	0.076	0.768	0.156	0.076	0.746	0.168	0.087
4	65	191.154.276	0.397	0.524	0.078	0.761	0.157	0.082	0.733	0.168	0.098
5	62	180.915.260	0.399	0.518	0.083	0.763	0.151	0.086	0.741	0.161	0.098
6	58	171.115.067	0.378	0.533	0.090	0.749	0.158	0.093	0.726	0.166	0.108
7	54	159.138.663	0.394	0.524	0.082	0.765	0.15	0.085	0.741	0.165	0.094
8	50	146.364.022	0.394	0.521	0.085	0.749	0.158	0.093	0.725	0.165	0.110
9	48	141.213.431	0.398	0.519	0.083	0.768	0.154	0.078	0.742	0.164	0.093
10	46	135.534.747	0.394	0.526	0.079	0.751	0.165	0.084	0.737	0.168	0.095
11	46	135.006.516	0.391	0.522	0.087	0.756	0.155	0.090	0.738	0.165	0.097
12	45	133.851.895	0.394	0.520	0.086	0.757	0.152	0.091	0.736	0.163	0.102
13	39	115.169.878	0.404	0.522	0.074	0.758	0.156	0.087	0.745	0.163	0.093
14	36	107.349.540	0.400	0.517	0.083	0.756	0.159	0.085	0.735	0.167	0.099
15	35	102.531.392	0.406	0.513	0.081	0.762	0.147	0.090	0.743	0.158	0.099
16	31	90.354.753	0.408	0.514	0.078	0.770	0.143	0.087	0.739	0.165	0.096
17	28	81.195.210	0.380	0.519	0.102	0.744	0.157	0.100	0.724	0.162	0.115
18	27	78.077.248	0.402	0.521	0.077	0.763	0.155	0.081	0.738	0.172	0.090
19	20	59.128.983	0.390	0.505	0.105	0.748	0.151	0.101	0.727	0.161	0.113
20	21	63.025.520	0.396	0.525	0.078	0.757	0.149	0.094	0.737	0.165	0.098
21	16	48.129.895	0.412	0.501	0.088	0.735	0.163	0.102	0.733	0.168	0.099
22	17	51.304.566	0.411	0.505	0.084	0.750	0.156	0.094	0.729	0.167	0.104
Autossômicos	-	-	0.429	0.507	0.064	0.786	0.147	0.067	0.761	0.159	0.080
X	53	155.270.560	0.292	0.584	0.123	0.658	0.192	0.150	0.671	0.188	0.141
X das mulheres	53	155.270.560	0.290	0.585	0.125	0.658	0.191	0.151	0.670	0.189	0.141
X dos homens	53	155.270.560	0.293	0.585	0.121	0.689	0.163	0.148	0.673	0.186	0.141

