Universidade do Estado do Rio de Janeiro

Centro Biomédico

Instituto de Biologia Roberto Alcântara Gomes

Rodrigo Fabián Flores Espinoza

Avaliação da estrutura genética das populações miscigenadas do Equador com base em marcadores do cromossomo X

Rio de Janeiro

2022

Rodrigo Fabián Flores Espinoza

Avaliação da estrutura genética das populações miscigenadas do Equador com base em marcadores do cromossomo X

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Biociências, da Universidade do Estado do Rio de Janeiro.

Orientadora: Prof.ª Dra. Maria Leonor Rodrigues de Sousa Botelho de Gusmão

Rio de Janeiro

2022

CATALOGAÇÃO NA FONTE
UERJ/REDE SIRIUS/BIBLIOTECA CB-A

E77 Espinoza, Rodrigo Fabián Flores.
Avaliação da estrutura genética das populações miscigenadas do Equador com base em marcadores do cromossomo X / Rodrigo Fabián Flores Espinoza. - 2022.
83 f.

Orientadora: Prof.ª Dra. Maria Leonor Rodrigues de Sousa Botelho de Gusmão

Mestrado (Dissertação) - Universidade do Estado do Rio de Janeiro, Instituto de Biologia Roberto Alcântara Gomes. Pós-graduação em Biociências.

1. Genética de populações – Teses. 2. Cromossomo x – Teses. 3. Variedade genética. 4. Equador - Teses. I. Gusmão, Maria Leonor Rodrigues de Sousa Botelho de. II. Universidade do Estado do Rio de Janeiro. Instituto de Biologia Roberto Alcântara Gomes. III. Título.

CDU 575.1(8)

Bibliotecária: Ana Rachel Fonseca de Oliveira
CRB7/6382

Autorizo apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

_____ _____
Assinatura Data

Rodrigo Fabián Flores Espinoza

Avaliação da estrutura genética das populações miscigenadas do Equador com base em marcadores do cromossomo X

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Biociências, da Universidade do Estado do Rio de Janeiro.

Aprovada em 22 de fevereiro de 2022.

Orientadora: Prof.ª Dra. Maria Leonor Rodrigues de Sousa Botelho de Gusmão

Instituto de Biologia Roberto Alcântara Gomes – UERJ

Banca Examinadora: _____

Prof.ª Dra. Cíntia Barros Santos-Rebouças

Instituto de Biologia Roberto Alcântara Gomes – UERJ

_____

Prof.ª Dra. Vania Pereira

Universidade de Copenhague

_____

Prof.ª Dra. Márcia Holsbach Beltrame

Universidade Federal do Paraná

Rio de Janeiro

2022

AGRADECIMENTOS

À minha orientadora Leonor Gusmão por ter acreditado em mim e me orientado da melhor forma para o meu crescimento académico, profissional e pessoal.

Ao meu Professor Germán Burgos pela confiança depositada em mim. Obrigado pelo apoio prestado durante a realização deste trabalho.

À Direção Geral de Investigação da Universidade das Américas (UDLA), Quito-Ecuador, pelo apoio prestado durante a execução deste projeto de pesquisa.

Aos meus amigos e colegas do Laboratório de Diagnostico por DNA pelo excelente acolhimento e disposição para me ajudar sempre que precisei.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão de bolsa de estudos durante o curso de mestrado.

À Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) pela concessão de Bolsa nota 10 no último período do curso de mestrado.

À minha família pelo apoio incondicional nesta fase da minha vida.

RESUMO

FLORES, Rodrigo Fabián Espinoza. **Avaliação da estrutura genética das populações miscigenadas do Equador com base em marcadores do cromossomo X.** 2022. 83 f. Dissertação (Mestrado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2022.

De acordo com a história demográfica, o Equador passou por mudanças em suas populações nativas causadas pela colonização europeia e o tráfico de escravos africanos. Os eventos de miscigenação contínua entre indivíduos europeus, nativos americanos e africanos ocorreram de forma diferente em cada região do país, produzindo uma população heterogênea e subestruturada. Portanto, o objetivo deste estudo foi investigar o nível de subestrutura genética na população miscigenada do Equador. A presença de uma única cópia do cromossomo X no sexo masculino, que só é transmitida às filhas, reduz o tamanho efetivo e a taxa de recombinação na população em relação aos autossomos. Portanto, esperamos que o desequilíbrio de ligação seja maior e a subestrutura da população mais pronunciada para marcadores do cromossomo X do que para marcadores autossômicos. Neste estudo, um total de 377 amostras masculinas e 209 amostras femininas foram genotipadas para dois conjuntos de marcadores localizados no cromossomo X (32 X-Indels e 12 X-STRs). Como foi possível determinar a fase gamética das amostras femininas através da genotipagem dos pais, o número total de amostras usadas em cada análise, variou dependendo do teste realizado. Desta forma, 377 haplótipos masculinos e 418 haplótipos femininos foram analisados. Níveis significativos de LD foram observados não apenas entre marcadores separados por distâncias menores que 1 cM, mas também entre marcadores separados por distâncias variando de 10,891 a 163,53 cM. Entre as regiões equatorianas, a Amazônia apresentou o maior valor médio de R^2. A comparação dos valores de diferenciação entre marcadores do cromossomo X e autossômicos revelou a ocorrência de acasalamentos preferenciais entre homens europeus e mulheres nativas e africanas, bem como entre homens africanos e mulheres nativas. Além disso, uma ancestralidade nativa americana distinta foi perceptível na população da Amazônia e um fluxo gênico enviesado por sexo entre a Amazônia e as regiões dos Andes e da Costa do Pacífico. Em conjunto, os resultados deste estudo demonstram a importância de integrar as informações do cromossomo X para uma visão mais abrangente da história genética e demográfica das populações miscigenadas da América do Sul.

Palavras-chave: Cromossomo X. STRs. Indels. Populações miscigenadas. Desequilíbrio de ligação

ABSTRACT

FLORES, Rodrigo Fabián Espinoza. **Evaluating genetic structure in admixed populations from Ecuador based on X-chromosomal markers.** 2022. 83 f. Dissertação (Mestrado em Biociências) – Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2022.

According to the demographic history, Ecuador has experienced shifts in its Native populations caused by European colonization and the African slave trade. The continuous admixture events among European, Native American, and African individuals occurred differently in each region of the country, producing a heterogeneous and substructure population. Thus, the aim of this study was to investigate the level of genetic substructure in the Ecuadorian Mestizo population. The presence of a single copy of the X chromosome in males that is only transmitted to the daughters reduces the recombination rate and the effective population size in relation to the autosomes. Therefore, we expect linkage disequilibrium to be greater and the population structure more pronounced for X-chromosomal than for autosomal markers. In this study, a total of 377 male and 209 female samples were genotyped for two sets of X-chromosomal markers (32 X-Indels and 12 X-STRs). Since it was possible to determine the gametic phase of female samples through parental genotyping, the total number of samples used in each analysis varied depending on the test performed. Thus, 377 male and 418 female haplotypes were analyzed. Significant levels of LD were observed not only between markers separated by distances less than 1 cM, but also between markers separated by distance varying from 10.891 to 163.53 cM. Between Ecuadorian regions, Amazonia showed the highest average value of R^2. When X-chromosomal and autosomal differentiation values were compared, a sex-biased admixture between European men and Native and African women was revealed, as well as between African men and Native women. Moreover, a distinct Native American ancestry was discernible in the Amazonian population, and a sex-biased geneflow between Amazonia and the Andes and the Pacific coast regions. Altogether, the results from this study underline the importance of integrating X chromosome information for a more comprehensive view of the genetic and demographic history of South American admixed population.

Keywords: X chromosome. STRs. Indels. Admixed populations. Linkage disequilibrium

LISTA DE FIGURAS

Figura 1 – Mudança nas frequências alélicas para um *locus* ligado ao X..... 12

Figura 2 – Interpretação das medidas de LD entre pares de *loci* bialélicos..... 16

Figura 3 – Estrutura da repetição curta em *tandem* (STR)..... 18

Figura 4 – Mecanismo mutacional causado pelo *slippage*..... 19

Figura 5 – Variação de sequência do tipo inserção-deleção..... 20

Figura 6 – Esquema dos cromossomos sexuais X e Y..... 20

Figura 7 – Herança do cromossomo X..... 22

Figura 8 – Distribuição dos doze *loci* STR analisados no *kit* Investigator Argus X-12 QS..... 25

Figura 9 – Distribuição dos 32 marcadores de tipo X-Indel analisados..... 26

Figura 10 – Expansão do império inca..... 30

Figura 11 – Mapa representando as principais regiões de onde os escravos africanos foram trazidos para a América..... 31

Figura 12 – Localização do Equador..... 32

Figura 13 – Valores médios de R^2 medidos na amostra total e nas três regiões equatorianas. (A) Valores médios de R^2 (eixo Y) calculados para pares de marcadores agrupados em intervalos de 10 cM (eixo X). (B) Valores médios de R^2 (eixo Y) calculados para pares de marcadores agrupados em intervalos de 1 cM (eixo X) para distâncias abaixo de 9 cM..... 50

Figura 14 – Análise de escala multidimensional (MDS) com base nas distâncias genéticas F_{ST} calculadas para (A) 32 X-Indels (*Stress* = 0.0678) e (B) 12 X-STRs (*Stress* = 0.0308)..... 57

Figura 15 – Análise de componentes principais (PCA)..... 60

Figura 16 – Análise de componentes principais entre: populações de referência equatorianas (este estudo) e nativas americanas (IBARRA et al., 2014), africanas e europeias (PROJETO 1000 GENOMAS, 2015) medidas por X-Indels..... 61

Figura 17 – Análise de escala multidimensional (MDS) com base nas distâncias genéticas de Nei calculadas para os 32 X-Indels..... 63

LISTA DE TABELAS

Tabela 1 – Descrição dos conjuntos de marcadores X-STRs usados em genética populacional e forense..... 24

Tabela 2 – Descrição dos conjuntos de marcadores X-Indels descritos na literatura forense..... 26

Tabela 3 – Conjunto de marcadores Indel estudados..... 36

Tabela 4 – Conjunto de marcadores STR estudados..... 37

Tabela 5 – Condições de amplificação para o *kit* Argus X12 QS..... 38

Tabela 6 – Condições de amplificação para o sistema *Multiplex* X-Indels..... 38

Tabela 7 – Número total e características das amostras/haplótipos incluídos em cada uma das análises realizadas durante este estudo..... 41

Tabela 8 – Valores observados e esperados dos heterozigotos e o valor de *p* para o teste de equilíbrio de Hardy-Weinberg, para os doze X-STRs..... 42

Tabela 9 – Resultados do teste exato de desequilíbrio de ligação realizado em uma amostra de 558 haplótipos..... 45

Tabela 10 – Taxa média de mutação e diversidade genética por *locus*, número de haplótipos diferentes, diversidade de haplótipos e distância genética entre os marcadores mais distantes em cada grupo de ligação..... 46

Tabela 11 – Resultados do teste exato de desequilíbrio de ligação para todos os pares de marcadores que apresentaram valores de *p* significativos (em vermelho) ou menores que 0,009 (em verde), em uma amostra de 318 haplótipos..... 48

Tabela 12 – Valores médios de heterozigosidade esperada calculados para o conjunto completo de marcadores do cromossomo X, doze X-STRs e 32 X-Indels, nas três regiões equatorianas..... 51

Tabela 13 – Distâncias genéticas (F_{ST}) entre as três regiões do Equador (diagonal inferior) e os correspondentes valores de não diferenciação *p* (diagonal superior), para todos os marcadores do cromossomo X..... 52

Tabela 14 – Distâncias genéticas (F_{ST}) entre as populações estudadas (diagonal inferior) e os valores correspondentes de probabilidade de não diferenciação (diagonal superior), para os 23 X-indels..... 54

Tabela 15 – Distâncias genéticas (F_{ST}) entre as populações estudadas (diagonal inferior) e os valores correspondentes de probabilidade de não diferenciação, *p* (diagonal superior), para os doze X-STRs..... 55

LISTA DE ABREVIATURAS E SIGLAS

AIM	Marcador Informativo de Ancestralidade
AMOVA	Análise de variância molecular
ANOVA	Análise de variância
DNA	Ácido desoxirribonucleico
F	Coeficiente de endogamia
FST	Índice de fixação
DG	Distância genética
HW	*Hardy-Weinberg*
Indel	Polimorfismo de Inserção-deleção
LD	Desequilíbrio de ligação *(Linkage disequilibrium)*
LG	Grupo de ligação *(Linkage group)*
MDS	Escalonamento multidimensional *(Multidimensional scaling)*
mtDNA	DNA mitocondrial
PAR	Região pseudoautossômica
PCA	Análise de Componentes Principais
R^2	Quadrado do coeficiente de correlação de Pearson
RFLP	Polimorfismo de comprimento de fragmentos de restrição
SNP	Polimorfismo de nucleotídeo único
STR	Repetição curta em *tandem*
VNTR	Repetição em *tandem* de número variável

LISTA DE SÍMBOLOS

%	Porcentagem
µL	Microlitros
Cm	Centimorgan
Mb	Mega base
Min	Minutos
Ŋg	Nanograma
S	Segundos

SUMÁRIO

	INTRODUÇÃO	11
1	**OBJETIVOS**	34
1.1	**Objetivo Geral**	34
1.2	**Objetivos Específicos**	34
2	**MATERIAIS E MÉTODOS**	35
2.1	**Seleção das amostras**	35
2.2	**Análise de marcadores moleculares do cromossomo X**	35
2.2.1	Reação de amplificação com o kit Investigator® Argus X-12 QS	37
2.2.2	Reação de amplificação com o sistema *Multiplex* X-Indels	38
2.2.3	Eletroforese capilar e detecção dos produtos de amplificação para o *kit* Argus X-12 QS	39
2.2.4	Eletroforese capilar e detecção dos produtos de amplificação para o sistema *Multiplex* X-Indels	39
2.3	**Análise estatística**	39
3	**RESULTADOS E DISCUSSÃO**	42
3.1	**Análise de equilíbrio de Hardy-Weinberg**	42
3.2	**Análise de desequilíbrio de ligação entre X-STRs**	43
3.3	**Análise de desequilíbrio de ligação entre o conjunto completo de *loci***	46
3.4	**Parâmetros de diversidade, distâncias genéticas e AMOVA**	51
3.5	**Comparação com outras populações**	52
3.6	**Diferenciação populacional para indels de cromossomo X e autossômicos**	58
	CONCLUSÕES	65
	REFERÊNCIAS	66
	APÊNDICE – Artigo *"Investigating genetic diversity in admixed populations from Ecuador"*	73

INTRODUÇÃO

O estudo da genética consiste em três subdisciplinas principais: (1) genética molecular, que estuda a natureza química própria dos genes, ou seja, como a informação genética é codificada, replicada e expressa; (2) genética de transmissão, que estuda a variação genética dos organismos e a hereditariedade, ou seja, a forma como os organismos recebem e transmitem as características biológicas de geração para geração; (3) genética populacional, que estuda a composição genética de indivíduos da mesma espécie, e como essa composição muda geograficamente com o passar do tempo, criando variação dentro e entre as populações. (GILLESPIE, 1998).

A genética populacional é baseada na lei de Hardy-Weinberg (HW), que foi formulada por G. H. Hardy e W. Weinberg em 1908. Esta lei consiste em um modelo matemático que estima o efeito das forças evolutivas nas frequências alélicas e genotípicas de uma população. Para um *locus* autossômico com dois alelos, a lei de HW afirma que: se uma determinada população for grande, o acasalamento for aleatório e não for afetada por outras forças evolutivas como seleção, mutação e migração, as frequências alélicas permanecerão constantes de geração em geração (HARDY, 1908). Portanto, a lei de HW indica que, quando esses pressupostos são atendidos, a reprodução por si só não altera as frequências alélicas ou genotípicas, apresentando uma relação matemática entre as frequências genotípicas de uma geração e as frequências alélicas da próxima (CROW, 2001).

A lei de HW tem várias implicações para a estrutura genética de uma população. A primeira implicação é que, uma população não pode evoluir se atender às suposições de HW, porque a evolução consiste na mudança nas frequências alélicas de uma população. Processos, como seleção, mutação, deriva genética ou migração, são necessários para que as populações evoluam (PIERCE, 2012).

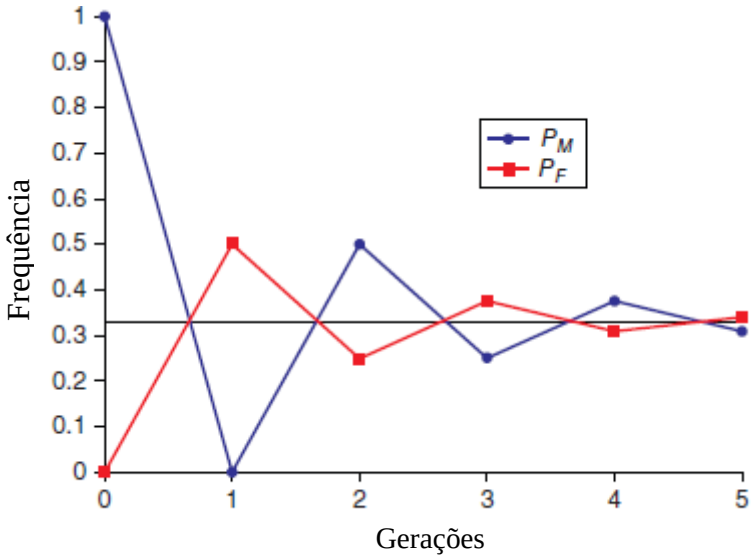
Uma segunda implicação é que, quando uma população está em equilíbrio de HW, as frequências genotípicas são determinadas através das frequências alélicas. Por outro lado, quando uma população não está em equilíbrio, não é possível prever as frequências genotípicas a partir das frequências alélicas. Embora seja possível determinar as frequências alélicas a partir das frequências genotípicas (ver Equação 1), o inverso só é possível quando a população está em equilíbrio de HW (PIERCE, 2012).

$$p = f(A)\frac{2n_{AA} + n_{AA}}{2N};\ q = f(a)\frac{2n_{aa} + n_{aa}}{2N} \quad (1)$$

onde n_{AA}, n_{Aa} e n_{aa} representam o número de indivíduos AA, Aa e aa, e N representa o número total de indivíduos na amostra, para um *locus* com dois alelos “A” e “a”.

Uma terceira implicação do equilíbrio de HW é que, na ausência de forças evolutivas, um único evento de acasalamento aleatório produz as frequências de equilíbrio (p^2, 2pq, q^2), sem importar quão diferentes sejam as frequências observadas das frequências de equilíbrio. No entanto, quando se trata de genes ligados ao cromossomo X, há uma exceção a essa terceira implicação do equilíbrio de HW. Isso acontece porque as mulheres são diploides para o cromossomo X, tendo duas cópias, mas os homens são haploides, tendo apenas uma cópia do cromossomo X. Como os homens herdam unicamente o cromossomo X por via materna, a frequência de um alelo ligado ao X nos homens da próxima geração é a mesma das mulheres na geração atual. Portanto, para um gene ligado ao X, com frequências de alelos desiguais em homens e mulheres, várias gerações são necessárias para se atingir o equilíbrio de HW (Figura 1) (RELETHFOR, 2011).

Figura 1 - Mudança nas frequências alélicas para um *locus* ligado ao X



Legenda: Mudança nas frequências alélicas para um *locus* ligado ao X em função do número de gerações. P_M é a frequência em homens e P_F é a frequência em mulheres.
Fonte: Adaptado de STONEKING (2017).

Os princípios estabelecidos pelo equilíbrio de HW fornecem uma base para a compreensão da mudança evolutiva. Ao comparar as condições reais com os pressupostos teóricos do equilíbrio de HW, podemos desenvolver uma visão do processo evolutivo.

Um pressuposto teórico do equilíbrio de HW é o **acasalamento aleatório** em relação ao genótipo. No entanto, quando esse pressuposto não é atendido, os alelos se combinam de forma não aleatória para formar genótipos, observando-se uma alteração nas frequências genotípicas de uma população. Podemos distinguir entre dois tipos de acasalamento não aleatório: acasalamento seletivo e consanguinidade. O acasalamento seletivo significa que existe um acasalamento preferencial, e pode ocorrer tanto como seletivo positivo, o que significa que acasalamentos entre pessoas que são semelhantes em algum aspecto de seu fenótipo ocorrem mais frequentemente do que o esperado ao acaso, ou como seletivo negativo, também chamado de heterogamia, o que significa que os acasalamentos entre pessoas que diferem em algum aspecto de seu fenótipo ocorrem com mais frequência do que o esperado ao acaso (STONEKING, 2017).

A consanguinidade envolve o acasalamento entre indivíduos aparentados geneticamente. Portanto, pode ser considerada como acasalamentos entre indivíduos com genótipos que são mais semelhantes do que o esperado no acasalamento aleatório (PIERCE, 2012).

Qualquer tipo de acasalamento não aleatório (exceto o seletivo negativo) implica em um desvio nas frequências de equilíbrio de HW, levando a um aumento na proporção de homozigotos e uma diminuição na proporção de heterozigotos em uma população. Como consequência, a população passa a estar subestruturada, ou seja, dividida em subgrupos que diferem em sua composição genética. A diminuição da heterozigosidade é uma função da variância das frequências alélicas entre as populações e, portanto, pode ser usada como medida do grau de subestrutura genética em uma população (WAHLUND, 1928).

A seguinte suposição da lei de HW a considerar envolve o tamanho da população. A lei assume um tamanho populacional infinitamente grande. No entanto, não existe uma população infinitamente grande. Portanto, com base nesse fato, iremos analisar duas consequências associadas ao tamanho da população.

A primeira consequência tem a ver com as flutuações aleatórias nas frequências dos alelos. Se o tamanho da população for suficiente pequeno, a frequência dos alelos pode mudar de geração para geração, simplesmente devido aos efeitos aleatórios. Esses efeitos aleatórios nas frequências alélicas são conhecidos como **deriva genética**. A deriva genética é o processo de mudança evolutiva que envolve a amostragem aleatória de alelos da geração parental para

produzir a geração descendente, o que torna a composição dos descendentes e das gerações parentais diferentes (RELETHFOR, 2011).

A deriva genética resulta do pequeno tamanho contínuo da população, do efeito fundador (estabelecimento de uma população por alguns fundadores) e do efeito gargalo (redução dramática no tamanho da população). Esta força evolutiva causa mudança nas frequências alélicas dentro de uma população, uma perda de variação genética por meio da fixação de alelos e divergência genética entre as populações (PIERCE, 2012).

A segunda consequência tem a ver com o aumento de **consanguinidade**, porque em uma população pequena muitos indivíduos estarão geneticamente relacionados. Assim, mesmo que os acasalamentos sejam aleatórios provavelmente envolvem parentes e, portanto, esperamos um aumento na proporção de alelos que sejam idênticos por descendência ao longo do tempo.

Outra força evolutiva que pode provocar mudanças nas frequências alélicas é a **mutação**. Esta força evolutiva é de grande importância, devido ao fato de que é responsável por criar variação. Sem variação, não pode haver evolução e nenhuma das outras forças evolutivas (deriva genética, migração ou seleção) tem algo sobre o que agir. Embora a mutação seja a fonte principal da variação genética, devido à raridade de ocorrência, seu efeito como um desvio para o equilíbrio de HW é quantitativamente desprezível, a menos que outros fatores, como seleção natural e deriva genética, contribuam (RELETHFOR, 2011).

O impacto da seleção natural na mutação depende de como uma nova mutação afeta a probabilidade de um indivíduo sobreviver e se reproduzir. Se um novo alelo mutante for prejudicial, sua frequência será reduzida. Por outro lado, se um alelo mutante é útil em um determinado ambiente e aumenta a probabilidade de sobrevivência e reprodução, a frequência dessa mutação aumentará com o tempo. No entanto, o que acontece quando a seleção natural não é um fator de impacto, ou seja, quando um alelo mutante é neutro? Como foi descrito na secção anterior, o efeito esperado da deriva genética é maior em populações pequenas. Levando este fato em consideração e que qualquer mutação específica (para um determinado *locus*) é relativamente rara, a deriva genética geralmente resulta na perda do alelo mutante. No entanto, ocasionalmente, a frequência de um alelo mutante, por causa da deriva genética continuará a aumentar, até que a nova mutação se espalhe pela população. Desse modo, o alelo se tornará mais comum na próxima geração, atingindo uma frequência maior (RELETHFOR, 2011).

A seguinte força evolutiva que afeta o equilíbrio de HW é a **seleção natural**. A seleção natural ocorre quando indivíduos com características adaptativas específicas produzem um número maior de descendentes do que os produzidos por outros na população. As características

que oferecem uma vantagem reprodutiva aumentam com o passar do tempo, permitindo que as populações se tornem mais adequadas aos seus ambientes para se tornarem mais adaptadas. O efeito da seleção natural no *pool* gênico de uma população depende dos valores de *fitness* dos genótipos na população. *Fitness* é definido como o sucesso reprodutivo de um genótipo em comparação com o sucesso reprodutivo de outros genótipos na população. Devido à seleção natural, os valores de *Fitness* diferenciais entre os genótipos, levam a mudanças nas frequências genotípicas da população ao longo do tempo que, por sua vez, levam a mudanças nas frequências alélicas que as compõem (GILLESPIE, 1998).

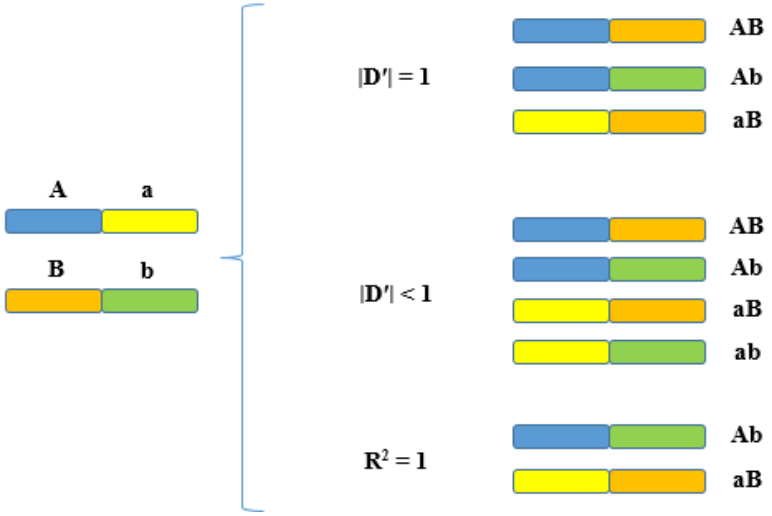
O processo final que traz mudanças nas frequências alélicas é a força evolutiva conhecida como **migração ou fluxo gênico**. Referindo-se novamente aos pressupostos da lei de HW, esses nos dizem que um grupo populacional permanece fechado para outros grupos populacionais, mas muitas populações naturais experimentam migração de outras populações. De fato, os seres humanos vivem em milhares de populações ao redor do mundo, todas interconectadas por diferentes níveis de fluxo gênico no presente e no passado.

A migração causa mudanças nas frequências alélicas de uma população ao introduzir alelos de outras populações, diminuindo as diferenças genéticas entre as populações e aumentando a variação genética dentro delas. A magnitude da mudança, devido à migração, vai depender da extensão da migração, da diferença nas frequências alélicas entre a população fonte e as populações receptoras e dos efeitos de deriva genética (PIERCE, 2012). Como foi exposto anteriormente, a deriva genética aumenta as diferenças genéticas entre as populações. Por ser um processo aleatório, com o tempo, diferentes alelos se fixarão em diferentes populações, aumentando assim a variação genética entre elas. A migração, ao contrário, neutraliza esse efeito da deriva genética, diminuindo as diferenças genéticas entre as populações.

Embora um modelo simples de *locus* único com dois alelos tenha sido usado, o equilíbrio de HW pode ser facilmente estendido para mais *loci*. Ao analisar dois *loci* com dois alelos, devemos nos referir como haplótipo às possíveis combinações aleatórias que podem ocorrer entre *loci*. Tal como acontece com o equilíbrio de HW para um único *locus*, a ação das forças evolutivas pode levar, neste caso, a um desvio na expectativa de combinações aleatórias entre os *loci*. Quando isso ocorre, os *loci* são considerados em um estado de desequilíbrio de ligação (também conhecido como desequilíbrio gamético). Ou seja, alguns haplótipos ocorrem com maior e outros com menor frequência do que o esperado (RELETHFOR, 2011). O desequilíbrio de ligação (LD) é a associação não aleatória de alelos de dois ou mais *loci* em uma população, não necessariamente próximos. Várias medidas têm sido propostas para avaliar a força do LD. Entre as medidas mais importantes que capturam a força da associação entre

pares de *loci* bialélicos estão: R^2 e |D′| (HEDRICK, 1987). Ambas as medidas variam de 0 (sem desequilíbrio) a 1 (desequilíbrio completo), mas a interpretação é ligeiramente diferente. A medida |D′| é igual a 1 se apenas dois ou três dos possíveis haplótipos estão presentes, e é <1 se os quatro haplótipos possíveis estão presentes. A medida R^2 representa a correlação estatística entre pares de *loci* e assume o valor 1 se apenas dois haplótipos estiverem presentes (Figura **2**) (WALL; PRITCHARD, 2003).

Figura 2 - Interpretação das medidas de LD entre pares de *loci* bialélicos



Legenda: Possíveis combinações haplotípicas com seus respectivos valores de |D′| e R^2
Fonte: O autor, 2022

Como foi descrito anteriormente, se uma população não está inicialmente em equilíbrio de HW para um único *locus*, na ausência de forças evolutivas, um único evento de acasalamento aleatório irá produzir as frequências de equilíbrio. No entanto, esse não é o caso do LD. Se as forças evolutivas que levaram ao LD não continuarem a agir e houver acasalamento aleatório, então a recombinação irá eventualmente alterar as frequências dos haplótipos até alcançar o equilíbrio. Ao contrário do caso de *locus* único, levará algum tempo para isso acontecer. A quantidade de tempo dependerá da taxa de recombinação e do número de gerações que passaram desde o estado inicial de LD (RELETHFOR, 2011).

Marcadores genéticos

"Marcador clássico" é um termo genérico para denotar qualquer marcador genético baseado na variação nos produtos dos genes, em oposição à variação ao nível do ácido desoxirribonucleico (DNA). Os primeiros marcadores desse tipo a serem usados em estudos de variação humana, foram os grupos sanguíneos ABO. Esses grupos sanguíneos são um dos sistemas mais conhecidos, todos baseados em vários antígenos encontrados na superfície dos glóbulos vermelhos. Os grupos sanguíneos em geral são marcadores imunogênicos, que têm em comum a característica de se basearem no reconhecimento de antígenos por anticorpos (YAMAMOTO et al., 1990).

Existem também outros tipos de variação nos produtos proteicos de genes que têm sido usados em estudos de genética humana. Esses são os polimorfismos bioquímicos, que refletem a variação na sequência de aminoácidos das proteínas (isoenzimas), que por sua vez refletem a variação da sequência de DNA subjacente (ROBERTS; PAPIHA; SSEBABI, 1977).

Um evento essencial, que possibilitou a manipulação e o estudo do DNA, foi a descoberta em 1970 de enzimas de restrição em bactérias. As enzimas de restrição são proteínas que reconhecem uma sequência específica de nucleotídeos dentro de uma molécula de DNA e cortam o DNA naquele ponto específico. Portanto, o interesse nestas enzimas decorre de seu uso na detecção de polimorfismos do DNA. Esses polimorfismos são chamados "Polimorfismos de comprimento de fragmentos de restrição" ou RFLPs. Essas melhorias tecnológicas na detecção de polimorfismos permitiram a análise de sequências não codificantes e a detecção de mudanças silenciosas em uma sequência codificadora de proteínas.

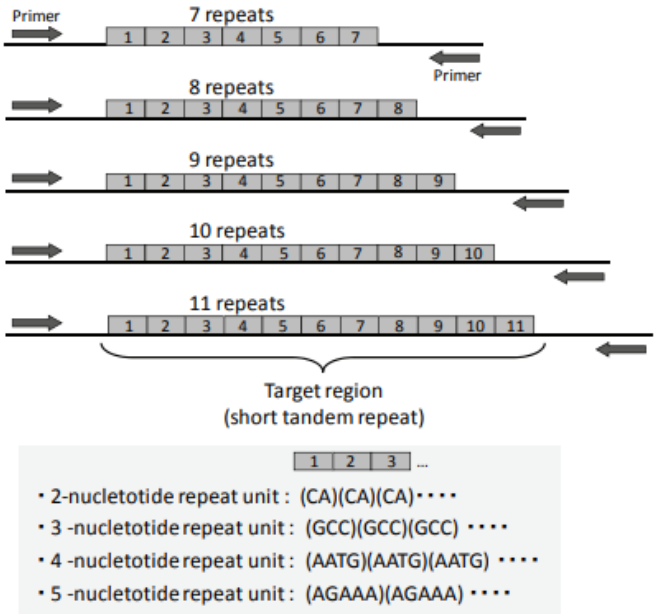
A primeira variação de repetições em *tandem* a ser descoberta em humanos, foi graças ao trabalho de Allec Jeffreys em 1980 (JEFFREYS; WILSON; THEIN, 1985). Essa variação foi denominada como minissatélite ou VNTR (Repetição em *tandem* de número variável). Apesar de representar uma classe de marcador conceitualmente diferente, a primeira etapa da análise também envolveu a digestão do DNA genômico com enzimas de restrição, semelhante à análise por RFLPs. No entanto, o polimorfismo extremamente alto dos minissatélites revolucionou a identificação genética dos indivíduos. Essa técnica recebeu o nome de impressão digital de DNA.

O principal avanço dos marcadores moleculares baseados em DNA foi impulsionado pela invenção da Reação em cadeia da polimerase (PCR) (TAUTZ, 1989). Pela primeira vez, qualquer região genômica poderia ser amplificada e analisada em muitos indivíduos sem a

necessidade de clonar ou isolar grandes quantidades de DNA genômico. Os primeiros marcadores a tirar o máximo proveito da tecnologia de PCR foram os microssatélites ou STRs (Repetições curtas em *tandem*). Como os minissatélites, os microssatélites também são sequências repetidas em *tandem*, mas seus motivos de repetição envolvem unidades de 1–6 pares de bases (pb) (WEBER; MAY, 1989).

O polimorfismo dos STRs é, em sua maioria, originado a partir de variações no número de unidades dos motivos de repetição. Essas unidades de repetição que formam o STR são nomeadas em função do número de nucleotídeo que as compõem (LAREU, 2013). Portanto, uma unidade de repetição é denominada como di-, tri-, tetra- ou pentanucleotídica se for formada por dois, três, quatro ou cinco nucleotídeos, respectivamente (Figura **3**). É importante ressaltar que nem todos os alelos contêm unidades de repetição completas, existem também alelos formados por unidades de repetição incompletas, denominados microvariantes (BUTLER, 2007).

Figura 3 - Estrutura da repetição curta em *tandem* (STR)

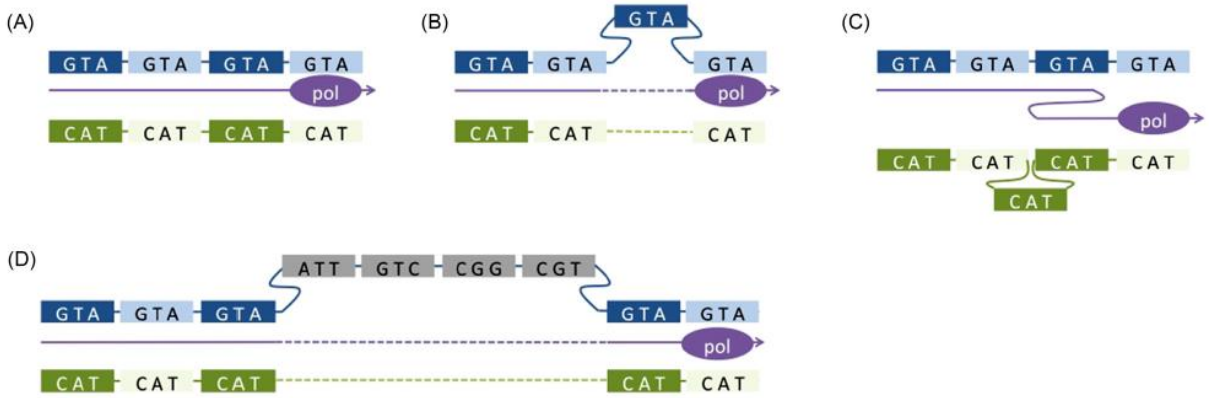


Fonte: Adaptado de HASHIYADA (2011)

O mecanismo mutacional dos loci STR é dado pelo deslizamento da DNA polimerase durante a replicação do DNA (slippage); a DNA polimerase, basicamente, perde o controle de quantas repetições existem. O deslizamento afeta classicamente as regiões do DNA com repetições de 1-4 bases e resulta na inserção ou deleção de apenas uma ou duas repetições adicionais (Figura 4). No entanto, também pode ocorrer em repetições não contínuas, resultando

em inserções ou deleções de sequências mais longas, flanqueadas pelas repetições em tandem (SEHN, 2015). Devido a estas características, os STR têm uma taxa de mutação muito maior do que outros tipos de marcadores (com taxas médias em torno de 2,1x10-3) (ELLEGREN, 2000), se mostrando potencialmente úteis em estudos de identificação humana e em análises de diversidade genética (BUTLER, 2012).

Figura 4 - Mecanismo mutacional causado pelo *slippage*



Legenda: Deslizamento da DNA polimerase. (A) Replicação normal em um trato de repetição de trinucleótidos, onde a linha superior (azul) representa a sequência de DNA molde, a linha roxa representa o caminho seguido pela polimerase (pol) e a linha inferior (verde) mostra o DNA complementar que é sintetizado pela polimerase. Quatro cópias da repetição estão presentes na sequência molde e na fita de DNA recém-sintetizada. Quando a polimerase desliza para a frente (B), ela pula uma das repetições da sequência molde GTA, com apenas três repetições copiadas para a fita de DNA recém-sintetizada (isto é, deleção de uma repetição). Correspondentemente, quando a polimerase desliza para trás (C), ela cópia uma repetição extra na fita de DNA recém-sintetizada (isto é, inserção de uma repetição). A polimerase também pode "deslizar" para uma repetição que não é imediatamente adjacente à sequência molde que estava se replicando, resultando na deleção ou inserção de DNA não repetitivo (D).
Fonte: Adaptado de SEHN (2015).

A variação de sequência de base única entre indivíduos em um ponto específico do genoma é referida como **polimorfismo de nucleotídeo único**, ou SNP. Os SNPs são uma das principais fontes de variação do genoma e representam 90% de toda a variação genética em humanos. Os SNPs se encontram tanto em regiões não codificantes quanto em regiões codificantes do genoma (BUTLER, 2012).

Outra forma de variação de sequência é a inserção-deleção (Indel) de um ou mais nucleotídeos (Figura 5). Indel é um termo geral que pode se referir à inserção, deleção ou inserção e deleção de nucleotídeos no DNA genômico. Por definição, os Indels variam de 1 nucleotídeo até 1000 nucleotídeos de comprimento (SEHN, 2015). Depois dos SNPs, os Indels são os polimorfismos de DNA mais abundantes. Quase 20% de todas as variações genéticas humanas no genoma são Indels, dos quais 8% são bialélicos (WEBER et al., 2002). As taxas

de mutação esperadas nos Indels são da ordem de magnitude de 10^{-9}, muito mais baixa do que a observada para STRs (MARTINEZ et al., 2019).

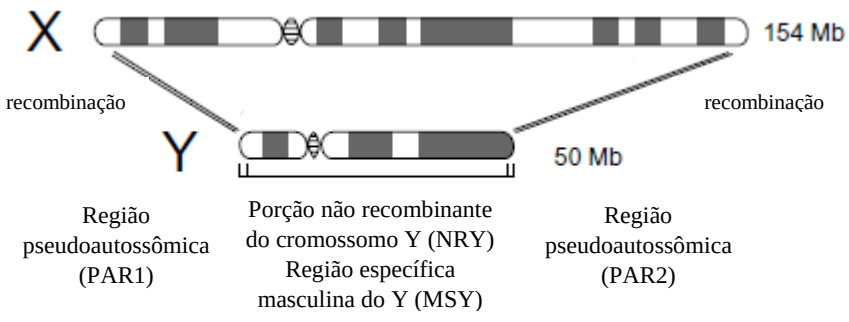
Figura 5 - Variação de sequência do tipo inserção-deleção

Legenda: Variações no nível d**e** nucleotídeo
Fonte: Adaptado de CARDOSO *et al.* (2015)

Cromossomo X

Os cromossomos X e Y são responsáveis pela determinação do sexo em seres humanos. Ohno em 1967 propôs que esses dois cromossomos evoluíram de um par de autossomos homólogos ancestrais (OHNO, 1967). Ao longo do tempo, ambos cromossomos sofreram mudanças estruturais e de tamanho. Essa diferenciação resultou em um par de cromossomos diferentes com a perda considerável de material genético do cromossomo Y, enquanto o cromossomo X adquiriu um mecanismo epigenético único de regulação da expressão do gene, a inativação, com a consequência de igualar a dosagem de genes do cromossomo X entre mulheres homogaméticas e homens heterogaméticos (LYON, 1961). O par de cromossomos sexuais é caracterizado pela ausência de recombinação ao longo da maior parte de seu comprimento, exceto nas regiões teloméricas pseudoautossômicas (PAR 1 e PAR 2) (Figura **6**).

Figura 6 - Esquema dos cromossomos sexuais X e Y



Legenda: As duas regiões pseudoautossômicas conhecidas como (PAR1) e (PAR2) recombinam entre os dois cromossomos sexuais. Os 95% restantes do cromossomo Y são referidos como a porção não recombinante (NRY).
Fonte: Adaptado de STONEKING (2017).

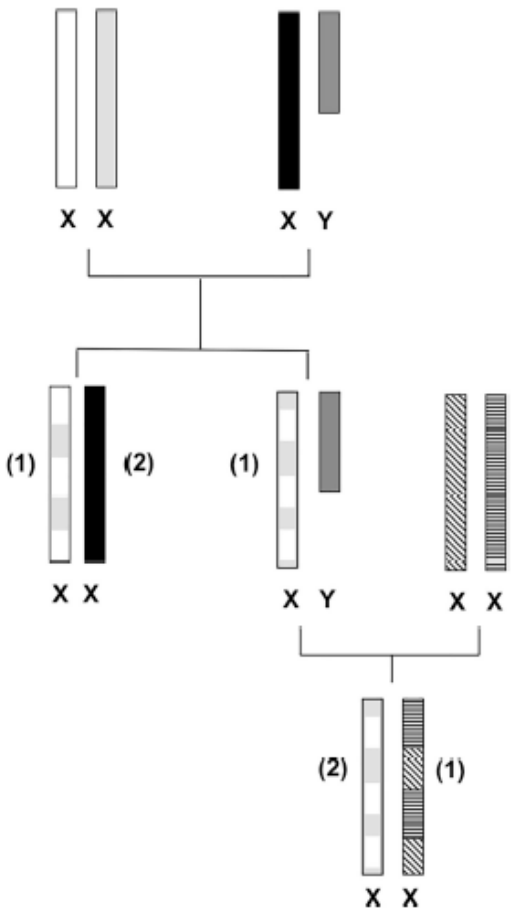
O cromossomo X tem cerca de 160 milhões de pares de bases (Mb) de comprimento e representa cerca de 5% do conteúdo genético do genoma humano (CHAN; HANNEMA; ACHERMANN, 2020). O baixo número de genes funcionais detectados (~1100) confere ao cromossomo X uma das menores densidades gênicas entre os cromossomos. Além de apresentar baixa concentração de genes, apenas 1,7% da sequência cromossômica dos genes funcionais é composta por exons, responsáveis pela transcrição de 33% do cromossomo X (ROSS et al., 2005).

O cromossomo X na genética populacional

As características distintivas do cromossomo X derivam em grande parte de como é herdado (Figura **7**). Os homens têm apenas uma cópia do cromossomo X, o que implica que cada cromossomo X existente numa população passou dois terços de sua história em mulheres. Consequentemente, as mutações ocorrem com menos frequência no cromossomo X em comparação com os autossomos, porque a taxa de mutação nas mulheres é cerca de 6 vezes menor do que nos homens (SCIENCE STAFF, 1997). Isto faz com que a diversidade genética seja menor para marcadores do cromossomo X, em comparação com marcadores autossômicos (SCHAFFNER, 2004).

A presença de uma única cópia do cromossomo X nos homens, que só é transmitida às filhas, reduz o seu tamanho efetivo na população em relação aos autossomos. Portanto, as forças evolutivas, como seleção natural e deriva genética, são mais pronunciadas e mantidas por mais tempo no cromossomo X (SCHAFFNER, 2004). Isso também é observado em padrões de desequilíbrio de ligação. O LD pode ser criado quando duas populações com frequências alélicas muito diferentes trocam informações genéticas. Durante um determinado período, os alelos das populações parentais irão se associar de forma não aleatória. Somente após algumas gerações, o LD será quebrado, e esses efeitos desaparecerão gradualmente. Como a recombinação ocorre apenas em mulheres, para *loci* que estão à mesma distância genética, demorará mais para quebrar o LD no cromossomo X em comparação com os autossomos.

Figura 7 - Herança do cromossomo X



Legenda: Descendentes femininos e masculinos herdam um cromossomo X materno recombinado (1) que resultou da meiose feminina. Descendentes do sexo feminino herdam um cromossomo X paterno inalterado (2), devido à falta de recombinação (com exceção das regiões pseudoautossômicas)
Fonte: Adaptado de GOMES *et al.* (2020).

Essas propriedades específicas do cromossomo X, duas cópias recombinantes em mulheres e uma única cópia não recombinante em homens (exceto para as regiões PAR), tornam os marcadores do cromossomo X uma ferramenta útil na genética populacional, bem como em outras áreas de investigação, como estudos evolutivos humanos, genética forense e genética médica (SZIBOR, 2007).

Marcadores moleculares de cromossomo X

Atualmente, existe um número considerável de marcadores polimórficos característicos do cromossomo X empregados em estudos populacionais e forenses. Em genética populacional,

os marcadores do cromossomo X são utilizados, por exemplo, na análise de subestrutura e diversidade genética e na caracterização de movimentos migratórios. Em estudos forenses, estes marcadores são particularmente uteis na análise de parentescos, tanto em investigações criminais como na identificação de vítimas de acidentes em massa. Esses marcadores apresentam diferenciais na sequência de nucleotídeos (polimorfismos de sequência; como é o caso dos SNPs) ou no comprimento de segmentos de DNA (polimorfismos de comprimento; como é o caso dos STRs e Indels) entre indivíduos.

As primeiras publicações sobre polimorfismos de cromossomo X do tipo STR (X-STRs) datam dos anos 90 (EDWARDS et al., 1992; SLEDDENS et al., 1992). Desde então, vários conjuntos de marcadores X-STR continuaram a ser descritos com o objetivo de se obter um maior poder de diferenciação entre indivíduos (ver Tabela **1**).

Em 2006, o sistema multiplex Argus X-UL da Biotype (Dresden, Alemanha) foi o primeiro kit comercial a ser desenvolvido para fins forenses, composto por quatro marcadores STR, localizados em posições distantes ao longo do cromossomo X para evitar a ligação entre marcadores (Tabela **1**). Em 2008, esse kit foi expandido com quatro X-STRs adicionais (Argus X-8). Ao contrário do kit anterior, este seria composto por quatro pares de X-STRs ligados entre si (Tabela **1**). Mais tarde, o multiplex Argus X-8 expandiu o número de marcadores incluindo quatro X-STRs adicionais, recebendo o nome de Argus X-12 (Qiagen, Hilden, Alemanha) (Tabela **1**). Desta forma, o kit composto por doze X-STRs é a versão mais recente e mais usada do kit Argus. Atualmente, está disponível a versão otimizada do kit, Argus X-12 QS, que apresenta os mesmos marcadores, organizados em quatro grupos de ligação (Figura **8**).

Em relação aos kits não comerciais, o sistema multiplex composto por dez marcadores X-STR (Decaplex) foi desenvolvido, como resultado de um trabalho colaborativo realizado pelo Grupo de Trabalho de Línguas Espanhola e Portuguesa da Sociedade Internacional de Genética Forense (GHEP-ISFG) (GUSMÃO et al., 2009).

Tabela 1 - Descrição dos conjuntos de marcadores X-STRs usados em genética populacional e forense

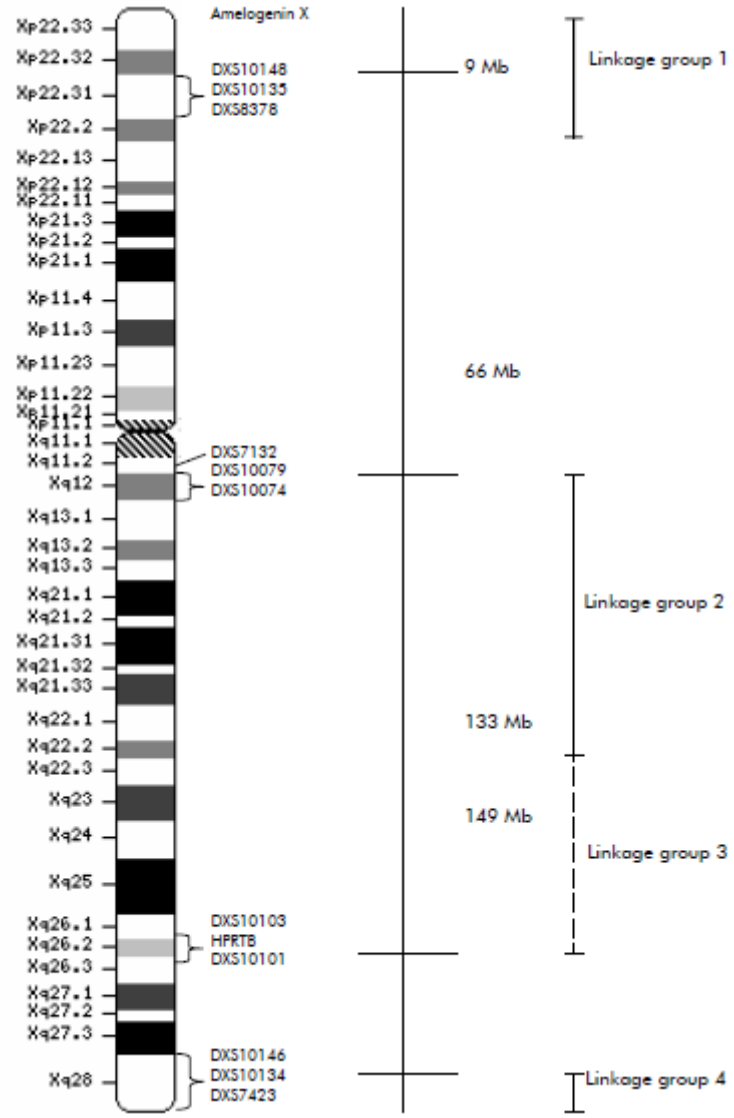
Conjunto de marcadores	**Número de *loci* X-STRs**	***Loci* amplificados**
Argus X-UL kit	4	DXS8378, DXS7132, HPRTB, DXS7423
Investigator Argus X-8 (Qiagen) kit	8	DXS8378, DXS10135, DXS7132, DXS10074, DXS10101, DXS7423, DXS10134, HPRTB
Investigator Argus X-12 QS (Qiagen) kit	12	DXS7132, DXS7423, DXS8378, DXS10074, DXS10079, DXS10101, DXS10103, DXS10134, DXS10135, DXS10146, DXS10148, HPRTB
GHEP-ISFG declaplex	10	DXS8378, DXS9898, DXS7133, GATA31E08, GATA172D05, DXS7423, DXS6809, DXS7132, DXS9902, DXS6789

Fonte: Adaptado de GOMES *et al.* (2020)

A outra classe de marcadores que tem sido utilizada na análise do cromossomo X são os marcadores bialélicos, nomeadamente os SNPs e os Indels. Geralmente, esses marcadores apresentam apenas dois alelos e, portanto, oferecem menos informação a nível intrapopulacional em comparação com os marcadores X-STR (CASTO et al., 2010; EDELMANN et al., 2009; SZIBOR, 2007). No entanto, os X-Indels possuem uma menor taxa de mutação em relação aos X-STRs, portanto, a probabilidade de um determinado Indel sofrer mutação independente em dois genomas é muito pequena. Assim, os Indels são capazes de detectar baixos níveis de LD, permitindo inferir o nível de subestrutura genética em uma determinada população.

A caracterização genética de uma população com ambos marcadores, simultaneamente, facilita a análise eficiente de eventos relacionados com as migrações masculinas e femininas ao longo da história das populações humanas ou com diferenças nos padrões de mutação e recombinação entre os sexos.

Figura 8 - Distribuição dos doze *loci* STR analisados no *kit* Investigator Argus X-12 QS



Legenda: Descrição da localização física dos *loci* STR e grupos de ligação que podem ser analisados usando o Investigator Argus X-12 QS Kit. Os doze *loci* encontram-se agrupados em quatro grupos de ligação (três marcadores por grupo) e, portanto, cada conjunto de três marcadores é tratado como um haplótipo para análise

Fonte: Adaptado de QIAGEN®(2013)

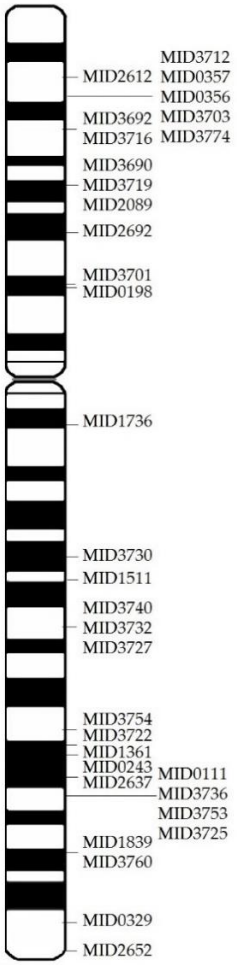
Pelas razões acima mencionadas, vários conjuntos de marcadores X-Indel tem sido descritos nos últimos 20 anos (Tabela **2**). Particularmente, PEREIRA *et al.* (2012a), descreveram um conjunto de 32 marcadores de inserção-deleção no cromossomo X (Figura **9**), que mostrou ser altamente polimórfico em grupos populacionais da África Subsaariana, Europa e do leste Asiático, além de apresentar eficientes parâmetros estatísticos forenses.

Tabela 2 - Descrição dos conjuntos de marcadores X-Indels descritos na literatura forense

Número de *loci*	**Sistema de genotipagem**	**Referência**
32	Multiplex simples (EC)	(PEREIRA *et al.*, 2012a)
33	Multiplex simples (EC)	(FREITAS et al., 2010)
16 X-Indels (de um total de 45 marcadores)	Multiplex simples (EC)	(TAO et al., 2019)
17 X-Indels (de um total de 60 marcadores, incluindo STRs)	Sequenciamento paralelo massivo	(ZHANG et al., 2017)
21	Multiplex simples (EC)	(EDELMANN et al., 2016)

Legenda: EC, eletroforese capilar
Fonte: Adaptado de GOMES *et al.* (2020)

Figura 9 - Distribuição dos 32 marcadores de tipo X-Indel analisados



Legenda: Descrição da localização física dos 32 *loci* Indels (números MID) analisados no conjunto de marcadores descrito por PEREIRA *et al.* (2012a)
Fonte: Adaptado de PEREIRA *et al.* (2012a)

Diversidade genética entre populações

A determinação da estrutura genética das populações constitui uma parte fundamental da genética populacional. Em 1950, Sewall Wright introduziu o valor estatístico F como uma ferramenta para descrever a partição da diversidade genética dentro e entre as populações. Ele mostrou que a quantidade de diferenciação genética entre as populações tem uma relação previsível com as taxas de processos evolutivos como migração, mutação e deriva. Wright descreveu o índice de fixação F_{ST} como um dos três parâmetros interrelacionados para descrever a estrutura genética de populações diploides (WRIGHT, 1950). Esses parâmetros são: F_{IT}, a correlação entre gametas dentro de um indivíduo em relação a toda a população; F_{IS}, a correlação entre gametas dentro de um indivíduo em relação à subpopulação à qual esse indivíduo pertence; e F_{ST}, a correlação entre gametas escolhidos aleatoriamente dentro da mesma subpopulação em relação a toda a população. Este último parâmetro é usado para medir o grau de diferenciação genética de subpopulações. (Ver equação 2) (NEI, 1973).

$$F = 1 - F_{IT} = (1 - F_{IS})(1 - F_{ST}) \tag{2}$$

Anos mais tarde, em 1973, Masatoshi Nei introduziu uma abordagem um tanto diferente para o estudo da diferenciação genética entre populações. Ele mostrou que, quando a diversidade genética da população é definida como a heterozigosidade esperada (frequência de heterozigotos) no equilíbrio de HW, a diversidade total pode ser dividida em componentes, ou seja, diversidade genética intra e interpopulacional. Essa diversidade genética é definida pelo uso das frequências genéticas na geração atual. Esta análise de diversidade foi projetada principalmente para se aplicar à diversidade genética média de um grande número de *loci*. No entanto, também pode ser aplicado a um único *locus* (NEI, 1973). Posteriormente, Weir e Cockerham (1984) utilizaram uma análise de variância (ANOVA) para estimar os componentes de variabilidade dentro e entre populações, que são usados para estimar os valores de F_{ST}.

O valor de F_{ST} não é uma distância genética em sentido estrito, é mais um valor representativo da consanguinidade relativa de subamostras em comparação com a consanguinidade na amostra total. Portanto, torna-se uma medida direta de consanguinidade devido à subdivisão e, portanto, é frequentemente considerada como uma medida de diferenciação entre subamostras. Uma das virtudes dos valores de F_{ST} é que são interpretados

como o componente da variância genética total entre populações e podem ser facilmente estendidos para níveis hierárquicos adicionais. No início da década de 1990, Laurent Excoffier e colegas forneceram uma estrutura estatística útil para analisar e decompor a variância genética total análoga à ANOVA, que é uma ferramenta padrão em estatística que eles chamaram de AMOVA (análise de variância molecular) (EXCOFFIER; SMOUSE; QUATTRO, 1992). Em AMOVA, a significância dos F_{ST} é testada usando uma abordagem de permutação não paramétrica descrita em EXCOFFIER *et al.* (1992), consistindo na permutação de haplótipos, indivíduos ou populações, entre indivíduos, populações ou grupos de populações, respectivamente. O valor de *p* obtido do teste é a proporção de permutações que levam a um valor de F_{ST} maior ou igual ao observado nas amostras populacionais analisadas (REYNOLDS; WEIR; CLARK, 1983; SLATKIN, 1995).

De forma geral, os resultados das análises genéticas entre populações são representados através de uma matriz de distâncias genéticas. Dependendo do número de populações analisadas, essas tabelas podem facilmente atingir proporções gigantescas. Portanto, são necessárias ferramentas computacionais que permitam simplificar e interpretar as relações entre um conjunto de populações, ao mesmo tempo minimizando a perda de informação que tal simplificação necessariamente acarreta. Existem vários métodos comumente usados para trabalhar com dados genéticos, entre eles, podemos mencionar o escalonamento multidimensional (MDS) e a análise de componentes principais (PCA). Eles são chamados também de métodos de “plotagem”, porque produzem um gráfico bidimensional ou tridimensional das relações entre as populações (STONEKING, 2017).

O escalonamento multidimensional MDS mostra as populações em um espaço dimensional, enquanto mantém as relações de distância entre as populações o mais próximas possíveis. Normalmente, a análise é feita para duas dimensões, então o resultado do MDS é um gráfico bidimensional em que as populações que estão próximas são interpretadas como sendo mais semelhantes geneticamente do que as populações que estão distantes no gráfico. Um aspecto importante desse método é a avaliação de quão bem a plotagem resultante retém a estrutura dos dados. Essa avaliação é fornecida pelo valor de estresse, que compara os valores de distância observados entre cada par de populações com aqueles obtidos no gráfico. Valores de estresse baixos indicam um bom ajuste entre as distâncias estimadas e as observadas no gráfico de MDS, enquanto valores de estresse altos indicam um mau ajuste à matriz de distâncias utilizada. Infelizmente, não há nenhum teste de significância que possa ser avaliado para indicar se o valor de estresse é aceitável ou não, então surgiu a convenção de que os valores de estresse abaixo de 0,15 são considerados aceitáveis, enquanto os valores de estresse acima

disso são inaceitáveis (HOUT; PAPESH; GOLDINGER, 2013). Portanto, se um gráfico MDS bidimensional resultar em um valor de estresse inaceitável, pode-se aumentar as dimensões para três, quatro ou mais, até se obter um valor abaixo do acima indicado.

A análise de componentes principais (PCA) é um algoritmo matemático que reduz a dimensionalidade dos dados, enquanto retém a maior parte da variação no conjunto de dados. O algoritmo realiza essa redução identificando direções, chamadas de componentes principais, ao longo das quais a variação nos dados é máxima. Através de alguns componentes, cada amostra pode ser representada por relativamente poucos números em vez de valores para milhares de variáveis, tornando possível avaliar visualmente as semelhanças e diferenças entre as amostras e determinar se as amostras podem ser agrupadas (STEFFENS, 2008).

Embora os dois métodos permitam visualizar diferenças ou semelhanças entre conjuntos de dados, a diferença mais importante entre PCA e MDS é que, a PCA é uma ferramenta de análise exploratória de dados, enquanto o MDS é uma representação gráfica de uma análise em si. Outra característica importante é que na PCA o primeiro eixo captura a maior parte das informações nos dados, e os eixos sucessivos capturam cada vez menos informações. Enquanto isso, no MDS os eixos não têm nenhum significado inerente. A PCA não produz um valor de estresse, em vez disso, a quantidade da variância total explicada por cada componente é fornecida pela análise, que então fornece uma ideia aproximada de quanto da informação nos dados é capturada pelos vários componentes. Por definição, o primeiro componente representará a maior proporção da variância e cada componente sucessivo representará uma parte menor da variância. Matematicamente, pode ser mostrado que a PCA captura mais informações nos dados do que outras análises e, portanto, caracteriza-se por ser mais útil porque fornece evidências objetivas da existência de grupos específicos de populações (STONEKING, 2017).

Equador

Na época pré-colombiana, os territórios que atualmente constituem o Equador eram ocupados por grupos nativos americanos diversos, tanto na cultura quanto no idioma. No entanto, durante o século XV, esses grupos nativos foram dominados pelo maior império da América pré-colombiana, o império Inca. Dessa forma, novas culturas, tradições e dialetos foram impostos. A civilização Inca se originou no Peru e se espalhou rumo ao norte através do

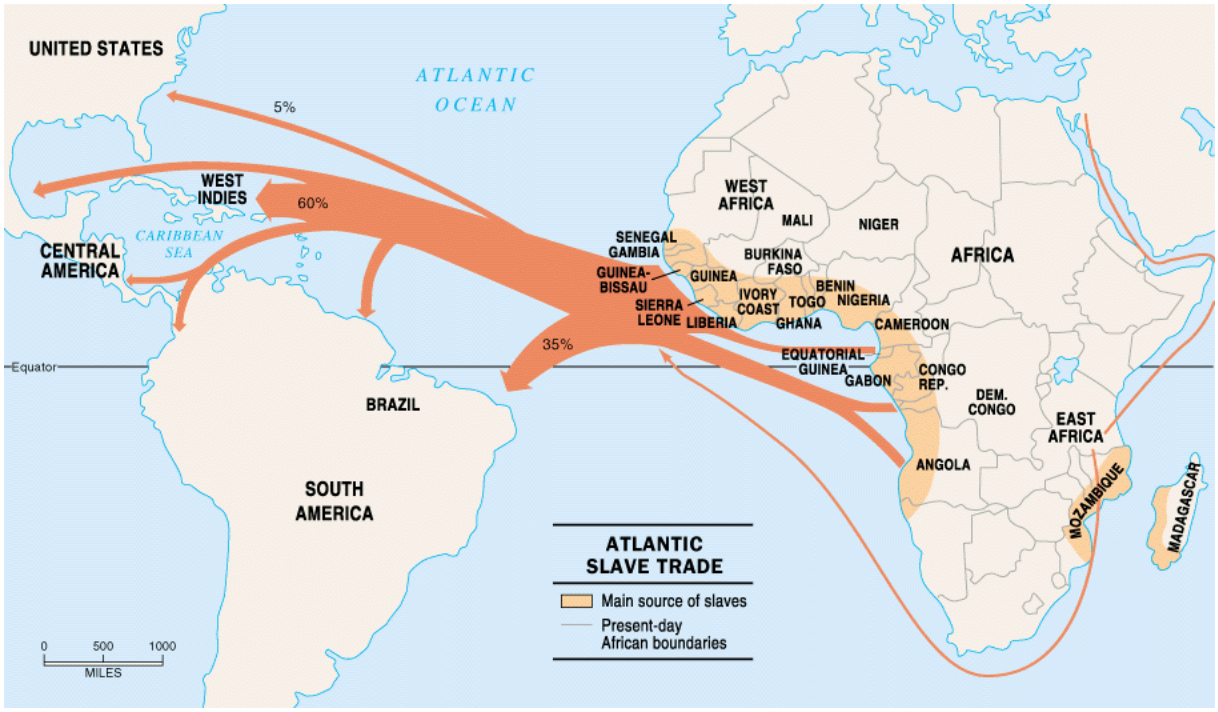
oeste do Equador até o sudoeste da Colômbia. Ao sul, para o oeste e centro-sul da Bolívia, noroeste da Argentina e uma grande parte do que hoje é o Chile (Figura **10**) (AYALA, 2002).

Figura 10 - Expansão do império inca

Legenda: Território que constituiu o império Inca (1437 d.C. a 1526 d.C.)
Fonte: O autor, 2022

Um século depois, os grupos nativos americanos tiveram a primeira interação com os europeus, durante as expedições espanholas ao longo da costa sul-americana. As contínuas expedições e os confrontos constantes entre os colonizadores e as populações nativas levaram à conquista e estabelecimento da era colonial no Equador. Nas décadas seguintes, escravos africanos foram trazidos desde Cartagena (um dos principais portos de chegada dos escravos africanos à América) para o Equador como mão de obra para as plantações de cana de açúcar (Figura **11**). Diferentes regiões da costa africana, entre Senegâmbia e Angola, foram citadas como origem de alguns grupos africanos no Equador (SALAZAR, 2020). Com o fim da escravidão no Equador, as populações afrodescendentes passaram a se localizar principalmente ao norte do país (AYALA, 2002).

Figura 11 - Mapa representando as principais regiões de onde os escravos africanos foram trazidos para a América



Legenda: Principais rotas de fluxo de escravos africanos para a América do Sul. A maioria dos escravos africanos que chegaram ao Equador era provenientes das regiões entre Senegâmbia e Angola.
Fonte: Adaptado de GAMBÉLA (2009)

O Equador é um país situado no noroeste da América do Sul, limitado ao norte pela Colômbia, ao sul e leste pelo Peru e a oeste pelo oceano Pacífico (Figura **12**). É caracterizado por ser um país étnica e demograficamente diverso. Em termos demográficos, o Equador está dividido em quatro regiões naturais: Costa do Pacífico, Andes, Amazônia e Galápagos (INEC, 2010). Atualmente é habitado por populações miscigenadas, afrodescendentes e nativas americanas. Essas populações são o resultado de diferentes graus de miscigenação entre povos nativos, europeus e africanos, bem como de imigrantes recentes, como chineses e árabes que chegaram ao Equador no início do século XX (AYALA, 2008).

Em suas três regiões continentais (Costa do Pacífico, Andes e Amazônia), quinze nacionalidades indígenas coexistem com diversas tradições e própria cultura. As nacionalidades indígenas localizadas na Amazônia são: *Huaorani*, *Achuar*, *Shuar*, *Cofán*, *Siona-Secoya*, *Shiwiar* e *Záparo*. Os *Tagaeri*, parentes dos *Huaorani*, constituem outra nacionalidade da região, mas este povo foi declarado "intangível" pelo Estado equatoriano, em respeito à sua vontade de viver longe da civilização. Nos Andes estão os *Quichuas*, formados por povos como *Otavalos*, *Salasacas*, *Cañaris* e *Saraguros*. Na parte norte dos Andes está a nacionalidade *Awa*,

mas expandiram seus territórios ao norte da Costa do Pacífico. Atualmente, os *Awás*, fazem parte da Costa do Pacífico junto com os *Chachis*, *Tsáchilas* e *Huancavilcas* (AYALA, 2008).

Figura 12 - Localização do Equador

Legenda: Divisão geográfica nas 4 regiões naturais: Costa do Pacífico, Andes, Amazônia e Galápagos
Fonte: Adaptado de http://earth.google.com/, 2009

De acordo com o último censo nacional (INEC, 2010), os indivíduos da população se auto identificam como mestiços (72,3%), afrodescendentes (7,2%), montubios (7,4%), nativos americanos (7%) ou descendentes de europeus (6,1%). Tanto mestiços quanto montubios são considerados uma miscigenação de nativos, europeus e africanos. No entanto, a população miscigenada da Costa do Pacífico é identificada como montubio e não como mestiça (AYALA,

2002; INEC, 2010). Atualmente, os mestiços/montubios representam a maior fração da população e estão amplamente distribuídos nas quatro regiões do Equador.

Para avaliar a composição genética e subestruturação da população equatoriana, vários estudos têm sido realizados usando diferentes tipos de marcadores, tanto de ancestralidade como forenses, em diferentes populações (BAETA et al., 2009, 2012; BAEZA et al., 2007; BURGOS et al., 2019a, 2019b; CARDOSO et al., 2012; GONZÁLES-ANDRADE; SÁNCHEZ; MARTÍNEZ-JARRETA, 2006; GONZÁLEZ-ANDRADE et al., 2008, 2009; SANTANGELO et al., 2017; TOSCANINI et al., 2018; ZAMBRANO et al., 2017, 2019a).

Estudos para estimar a contribuição genética das três principais populações parentais (nativa americana, europeia e africana), usando marcadores informativos de ancestralidade (AIMs) localizados nos cromossomos autossômicos, mostraram que grupos populacionais miscigenados conservam uma ancestralidade predominantemente nativa americana, seguida por contribuições europeias e africanas (POULSEN et al., 2011; SANTANGELO et al., 2017; ZAMBRANO et al., 2017, 2019a). Uma ancestralidade diferenciada entre os grupos é claramente evidenciada, destacando a necessidade de analisar uma possível subestrutura genética presente na população. Estudos do cromossomo Y mostraram uma herança paterna europeia predominante em todos os grupos miscigenados (BURGOS et al., 2019a; TOSCANINI et al., 2018). Em contraste, as linhagens de mtDNA são majoritariamente de ascendência nativa americana (BAETA et al., 2012; BURGOS et al., 2019b).

Poucos estudos foram realizados usando marcadores do cromossomo X em populações equatorianas. Os conjuntos de dados publicados incluem dez ou doze STRs do cromossomo X (X-STRs) selecionados para aplicações forenses. Com base nesses marcadores, nenhum sinal de subestrutura populacional foi encontrado para a população miscigenada (BAETA et al., 2013; CARDOSO et al., 2012; GAVIRIA et al., 2013; PINTO et al., 2020; ZAMBRANO et al., 2015, 2019b).

1 OBJETIVOS

1.1 Objetivo geral

O presente trabalho teve como objetivo estudar a diversidade genética das populações miscigenadas do Equador, através da análise de marcadores moleculares localizados no cromossomo X.

1.2 Objetivos específicos

Para cumprir o objetivo geral proposto, foram determinados os seguintes objetivos específicos:

a) Caracterizar geneticamente as populações miscigenadas do Equador, através da análise de marcadores moleculares do tipo STR (X-STRs) e inserção/deleção (X-Indels);
b) Analisar a subestrutura genética presente nas populações miscigenadas do Equador, por meio do estudo dos desvios ao equilíbrio de HW e dos padrões de desequilíbrio de ligação;
c) Comparar os dados gerados com os já publicados para outras populações da América do Sul, através da análise de distâncias genéticas;
d) Interpretar os resultados genéticos obtidos, levando em consideração a geografia e história demográfica do Equador.

2 MATERIAIS E MÉTODOS

Este trabalho é parte de uma cooperação entre dois grupos de pesquisa em genética, do Laboratório de Diagnóstico por DNA (LDD) da Universidade do Estado do Rio de Janeiro (UERJ), e do Laboratório de Pesquisa da Universidade das Américas no Equador (UDLA).

2.1 Seleção das amostras

Neste estudo, um total de 795 haplótipos de população miscigenada do Equador (377 masculinos e 418 haplótipos femininos) foram analisados. Um subconjunto de 149 amostras femininas foi incluído a partir de um estudo anterior sobre as taxas de mutação de X-STR (PINTO et al., 2020). Nesse estudo, ao selecionar as filhas dos trios estudados, foi possível determinar a fase gamética do genótipo e criar um banco de dados composto por 298 haplótipos não relacionados. Amostras de duplas pai / filha foram selecionadas de casos de paternidade. As amostras foram coletadas mediante consentimento informado, aprovado pelo comité de ética da Universidade das Américas (CEISH-UDLA 2017-0301).

O material genético dos indivíduos foi extraído a partir de manchas de sangue depositadas em papel Whatman® FTA® (Sigma-Aldrich Co.), utilizando a resina quelante Chelex® 100 (Laboratórios Bio-Rad, CA, USA) a 10%, e a partir de esfregaço bucal, utilizando o método orgânico de extração, com o uso de fenol-clorofórmio/álcool isoamílico (BUTLER, 2012).

2.2 Análise de marcadores moleculares do cromossomo X

A análise de DNA foi realizada mediante o emprego de dois *sets* de marcadores moleculares específicos do cromossomo X: (a) doze marcadores STR incluídos no kit

Investigator® Argus X-12 QS (Qiagen, Hilden, Alemanha); e (b) 32 marcadores de tipo Indel descritos por (PEREIRA et al., 2012a).

Para a análise dos doze marcadores STR, foi utilizado o *kit* comercial *Investigator®* Argus X-12 QS, seguindo a metodologia indicada pelo fabricante. O conjunto de 32 marcadores de inserção-deleção foi amplificado em uma única reação de PCR, seguindo a metodologia e especificações descritas por (PEREIRA et al., 2012a).

Nas Tabelas 3 e 4, encontra-se o total de marcadores do cromossomo X incluídos em cada um dos *sets* estudados.

Tabela 3 - Conjunto de marcadores Indel estudados

32 X-Indels		
Locus	*Primer Forward* 5'— 3'	*Primer Reverse* 5'— 3'
MID2612	gACCCACGGTGTTGAATTCAG	NED-CACAGCACCAGGAAAATAGC
MID3712	gtttAGTCTTGCTGCAATGTACCC	VIC-TTCAAGGGCAATGATGTTTG
MID357	gTTTTATAGACTGTGGCCCCC	PET-GTTAGTGGTTGGATTGCTCG
MID356	gtttCCAACTCCACGTGAGAAATG	PET-AGTCTGATGCAGTGGCAAAC
MID3703	VIC-AGCTTCCAAGTAGTTCTGCC	GTTTGGCTTACTTCCTCCTCC
MID3774	gAAGACGGGAATTGAGTCACC	NED-TTTTTGTGCACAGGCACTCC
MID3692	6FAM-ACATAAAAGCAAGCTTTGGC	gtttcttCCCGGTGTGTGAACTTTTTC
MID3716	6FAM-AAAGGGAGCATCTACTCCAG	gtttctAGGGCAATCCAGAATTGGAC
MID3690	GGGCACCATATTAGGCATGT	VIC-CCCACCATCTAACCCATTTC
MID3719	gTTCTTTCTCATCTGGCACCC	VIC-CTATGAAGCCTATAGATTGG
MID2089	VIC-AATCCATTCTGGAATAAGATGTCA	gtttcttTCCACTCTCAGGGATTCCTTT
MID2692	gtttcttCAAGTTCATATGGTGTCTTGG	PET-TGCATTACACAGAGCAACTC
MID3701	gtttctAGTTGGAGATGCAATGAAGC	NED-AGAGACAGGTGAATTGAGGC
MID198	6FAM-CAGGCACAGGAGAGGAAGAG	gTCCACCCCTAGTTAAACAGC
MID1736	VIC-GTGAAAGGTGAGCTTGTCTG	gtttctAGGCCTTTTTGGTTAACTGG
MID3730	6FAM-AGGATCCTGACTAAGATAGC	GAATCTCTGGAAACACTTGG
MID1511	gCTGCCTGGGATTTTTCCTTT	PET-CAGGGGAGAACACCCACTAA
MID3740	gtttctACTTGCTTTGCTTTTCCCTC	6FAM-GTACAACTGCAAGGAACRAG

Legenda: Lista dos marcadores moleculares X-Indel analisados neste estudo, e as sequências dos *primers forward* e *reverse* no sentido 5'→ 3'

Fonte: Adaptado de PEREIRA. *et al.*(2012a)

Tabela 4 - Conjunto de marcadores STR estudados

12 X-STRs	
Locus	Grupo de ligação
DXS10148	1
DXS10135	1
DXS8378	1
DXS7132	2
DXS10079	2
DXS10074	2
DXS10103	3
HPRTB	3
DXS10101	3
DXS10146	4
DXS10134	4
DXS7423	4

Legenda: Lista dos marcadores moleculares X-STR analisados neste estudo, e os grupos de ligamento conformados no *kit* Argus X-12 QS.
Fonte: Adaptado de QIAGEN® (2013).

2.2.1 Reação de amplificação com o kit Investigator® Argus X-12 QS

A reação de amplificação dos doze X-STRs foi baseada no protocolo recomendado pelo fabricante (Qiagen, Hilden, Alemanha), no qual foram utilizados 5 ηg de DNA, 1,5 µL de *Fast Reaction Mix 2.0* (Qiagen), 0,5 µL de *Primer Mix* Argus X-12 QS e água miliQ até completar 5 µL de volume final de reação. Como controle positivo de amplificação foi utilizado o DNA 9947A (Thermo Fisher Scientific Inc.). Como controle negativo foi preparada uma reação adicional, substituindo a quantidade adicionada de DNA por água miliQ.

As reações de amplificação foram realizadas em um termociclador modelo 9600 (Thermo Fisher Scientific Inc), seguindo as condições de amplificação apresentadas na Tabela 5.

Tabela 5 - Condições de amplificação para o *kit* Argus X12 QS

Temperatura (°C)	Tempo	Ciclos
98	60 s	3
61	100 s	
72	5 s	
96	10 s	27
61	100 s	
72	5 s	
68	2 min	1
10	∞	-

Legenda: Temperaturas, tempo e número de ciclos utilizados na reação de amplificação

2.2.2 Reação de amplificação com o sistema *Multiplex* X-Indels

A reação de amplificação dos 32 X-Indels foi baseada no protocolo descrito por Pereira e colaboradores (2012a), no qual foram utilizados 5 ηg de DNA, 2,5 µL de QIAGEN *Multiplex* PCR *Master Mix* (Qiagen), 0,5 µL de *Primer Mix* e água miliQ até completar 5 µL de volume final de reação. Como controle positivo de amplificação foi utilizado o DNA 9947A (Thermo Fisher Scientific Inc.). Como controle negativo foi preparada uma reação adicional, substituindo a quantidade adicionada de DNA por água miliQ.

As reações de amplificação foram realizadas em um termociclador modelo 9600 (Thermo Fisher Scientific Inc), seguindo as condições de amplificação apresentadas na Tabela 6.

Tabela 6 - Condições de amplificação para o sistema *Multiplex* X-Indels

Temperatura (°C)	Tempo	Ciclos
95	15 min	1
94	30 s	30
60	90 s	
72	45 s	
72	60 min	1
10	∞	-

Legenda: Temperaturas, tempo e número de ciclos utilizados na reação de amplificação

2.2.3 Eletroforese capilar e detecção dos produtos de amplificação para o *kit* Argus X-12 QS

A separação e detecção dos fragmentos foram realizadas por eletroforese capilar no aparelho ABI 3130 (Thermo Fisher Scientific Inc.). O volume de 1 µL de cada uma das amostras amplificadas foi misturado com 9,0 µL de Hi-Di Formamide (Thermo Fisher Scientific Inc.) e 0,3 µL de Liz-600 (Thermo Fisher Scientific Inc.), e carregado no aparelho ABI 3130. As análises dos fragmentos obtidos foram feitas no *software* GeneMapper versão 3.5 (Thermo Fisher Scientific Inc).

2.2.4 Eletroforese capilar e detecção dos produtos de amplificação para o sistema *Multiplex* X-Indels

A separação e a detecção dos fragmentos foram realizadas por eletroforese capilar no aparelho ABI 3130 (Thermo Fisher Scientific Inc.). O volume de 1 µL de cada uma das amostras amplificadas foi misturado com 9,0 µL de Hi-Di Formamide (Thermo Fisher Scientific Inc.) e 0,3 µL de Liz-500 (Thermo Fisher Scientific Inc.), e carregado no aparelho ABI 3130. As análises dos fragmentos obtidos foram feitas no *software* GeneMapper versão 3.5 (Thermo Fisher Scientific Inc).

2.3 **Análise estatística**

O número de amostras usadas na análise de dados variou dependendo do teste realizado, e apenas haplótipos não relacionados foram selecionados em cada análise (Tabela **7**).

Por meio do s*oftware* Arlequin v.3.5.2.2 (EXCOFFIER; LISCHER, 2015) foram calculadas tanto as frequências haplotípicas, estimativas de diversidade genética e valores estatísticos R^2, como o equilíbrio de HW e a análise de LD. Da mesma forma, os cálculos de distancias genéticas (F_{ST}), análises de variância (AMOVA) entre grupos populacionais, bem como a probabilidade de não-diferenciação (valores de *p*) foram avaliados através do mesmo *software*. Para testes múltiplos, o nível de significância de 0,05 foi ajustado pela aplicação da correção de *Bonferroni*.

Por outro lado, os cálculos de distâncias genéticas de Nei (NEI, 1973) foram geradas usando o *software* Phylip 3.69 (https://evolution.genetics.washington.edu/phylip.html).

Os dados obtidos das distâncias genéticas (F_{ST}) foram representados através da análise do escalonamento multidimensional (MDS) usando o *software* STATISTICA 13.2 Trial (StatSoft Inc., http://www.statsoft.com.br). A análise de componentes principais (PCA) foi realizada usando o *software* XLSTAT versão 2020.5.1 (https://www.xlstat.com).

Tabela 7 - Número total e características das amostras/haplótipos incluídos em cada uma das análises realizadas durante este estudo

	Total	Homens tipificados para X-STRs e X-Indels [este estudo]	Homens tipificados para X-Indels [este estudo]	Homens tipificados para X-STRs [este estudo]	Mulheres tipificadas para X-STRs [Pinto et al. (2020)]	Mulheres tipificadas para X-STRs [este estudo]	
Número de amostras‡	586	318	54	5	149	37	23
Número de haplótipos	795	318	54	5	298	74	46
Fase gamética§	-	-	-	-	conhecida	conhecida	desconhecida
Teste de equilíbrio de Hardy-Weinberg (nº de amostras)	209	-	-	-	149	37	23
X-STRs: LG-frequências haplotípicas, análises de LD, e comparações em pares com outras populações (nº de haplótipos)	558	181	-	5	298	74	-
X-STRs: Comparações em pares com outras populações (nº de haplótipos)¶	527	304	-	5	207	11	-
X-Indels: Análises de LD (nº de amostras/haplótipos)	372	318	54	-	-	-	-
X-Indels: PCA e comparações em pares com outras populações (nº de amostras/haplótipos)¶	355	304	51	-	-	-	-
Conjunto total: Análises de LD (nº de amostras/haplótipos)	318	318	-	-	-	-	-
Conjunto total: FSTs e AMOVA entre as regiões equatorianas (nº de amostras/haplótipos)¶	304	304	-	-	-	-	-

Legenda: ‡Um total de 137 homens apresentam uma relação de pai/filha com mulheres incluídas no conjunto de amostras, incluídas nas colunas "Mulheres tipificadas para X-STRs, com fase gamética conhecida" §A fase gamética foi determinada pela genotipagem do pai ¶Amostras sem informações sobre a origem geográfica do doador foram excluídas dessas análises

3 RESULTADOS E DISCUSSÃO

Os dados obtidos nesta dissertação foram publicados em formato de artigo científico na revista internacional *American Journal of Physical Anthropology* (APÊNDICE A).

3.1 Análise de equilíbrio de Hardy-Weinberg

Ao comparar as frequências genotípicas observadas com as esperadas de acordo ao equilíbrio de HW, nenhum desvio estatisticamente significativo foi detectado na população equatoriana para os doze X-STRs (Tabela **8**). No entanto, todos, exceto um X-STR (DXS10146), mostraram um excesso de homozigotos em comparação com o esperado assumindo o equilíbrio de HW (Tabela **8**). O valor médio dos heterozigotos observados na amostra (0,77) foi inferior ao esperado para uma população em equilíbrio de HW (0,80).

Tabela 8 - Valores observados e esperados dos heterozigotos e o valor de *p* para o teste de equilíbrio de Hardy-Weinberg, para os doze X-STRs

Loci	Equador (N=209)		
	Heterozigosidade. Observada.	Heterozigosidade. Esperada.	valor de p
DXS10148	0,84689	0,84933	0,09901
DXS10135	0,92344	0,93977	0,05899
DXS8378	0,55024	0,61298	0,12603
DXS7132	0,75598	0,78776	0,69660
DXS10079	0,83654	0,86398	0,06555
DXS10074	0,77990	0,81391	0,34557
DXS10103	0,65072	0,68334	0,43699
HPRTB	0,70335	0,75685	0,22089
DXS10101	0,85646	0,86669	0,21134
DXS10146	0,85646	0,85165	0,16948
DXS10134	0,81818	0,86388	0,73654
DXS7423	0,66986	0,68286	0,75524
Média	0,77067	0,79775	-

Embora um excesso de homozigotos seja sugestivo de subestrutura populacional ou algum grau de endogamia, os desvios não foram estatisticamente significativos. O mesmo resultado foi obtido anteriormente em uma amostra menor de 150 mulheres (PINTO et al., 2020).

O fato de não se encontrarem desvios significativos na proporção de genótipos observados em relação aos esperados numa população em HW não prova que cruzamentos ao acaso estejam ocorrendo ou que a população esteja livre de seleção natural, mutação, migração e/ou deriva genética. Além disso, pequenos desvios ao equilíbrio de HW apenas podem ser detectados em amostras populacionais de efetivo elevado. Para confirmar se a população está realmente em equilíbrio de HW, um tamanho de amostra maior ainda é necessário.

3.2 **Análise de desequilíbrio de ligação entre X-STRs**

Os doze X-STRs incluídos no *kit* Argus estão organizados em quatro grupos de ligação (LGs), cada um compreendendo três X-STRs. Os marcadores no mesmo LG têm baixas taxas de recombinação e estão localizados próximos uns dos outros (DG ≤ 0,46 cM). Portanto, eles são mais suscetíveis a mostrar LD mesmo várias gerações após a ocorrência dos eventos que os geraram. A análise de LD para os doze X-STRs foi realizada para uma amostra total de 558 haplótipos (Tabela **7**-9).

Os resultados revelaram associações estatisticamente significativas entre os três *loci* no LG 3 e entre dois *loci* no LG 1 e LG 4. Não foram obtidos valores significativos para os *loci* no LG 2 (Tabela **9**). Valores de LD significativo também foram observados entre *loci* distantes: HPRTB e DXS10148 (129,750 cM), HPRTB e DXS10135 (129,600 cM) e DXS10074 e DXS10101 (58,410 cM). A detecção de desvios no equilíbrio de ligação pode ser difícil em amostras com um número efetivo baixo, especialmente quando muitas combinações de haplótipos estão presentes na população. Para os X-STRs altamente variáveis que analisamos, KLING e colaboradores (2015) mostraram que são necessárias amostras com um efetivo elevado para se detectar LD com base em testes exatos. Esses autores realizaram simulações usando amostras de 100 a 600 haplótipos de uma população sueca exibindo LD para marcadores no mesmo LG. No estudo citado, valores de LD estatisticamente significativos foram encontrados entre os marcadores do LG 3 para todas as iterações de N = 200, o que só foi

possível para N = 600 no LG 1 e LG 4. No LG 2, mesmo para N = 600, valores de LD significativos foram obtidos, apenas, em 83% das interações. Embora os níveis de LD sejam específicos da população, é interessante notar que nas populações sueca e equatoriana, o LG 3 apresentou os menores valores de *p* (associação entre os *loci*), seguido por LG 1 e LG 4, e os maiores valores foram obtidos para LG 2 (Tabela **9**). Estudos anteriores também mostraram LD significativo entre marcadores no LG 3 em populações americanas, europeias, africanas e do leste asiático (FERRAGUT et al., 2017; GOMES et al., 2017; KLING; DELL; TILLMAR, 2015; PINTO et al., 2015, 2020; TOMAS; PEREIRA; MORLING, 2012).

Tabela 9 - Resultados do teste exato de desequilíbrio de ligação realizado em uma amostra de 558 haplótipos

	LG1			LG2			LG3			LG4		
	DXS10148	DXS10135	DXS8378	DXS7132	DXS10079	DXS10074	DXS10103	HPRTB	DXS10101	DXS10146	DXS10134	DXS7423
DXS10148	*											
DXS10135	0,3671	*										
DXS8378	0,0317	0,0000	*									
DXS7132	0,3316	0,5468	0,2724	*								
DXS10079	0,6144	0,4521	0,0262	0,9835	*							
DXS10074	0,5738	0,0496	0,0109	0,8408	0,0341	*						
DXS10103	0,1700	0,7732	0,2811	0,6671	0,6202	0,0648	*					
HPRTB	0,0000	0,0000	0,0174	0,5519	0,2683	0,0318	0,0000	*				
DXS10101	0,1040	0,1855	1,0000	0,7146	0,1154	0,0000	0,0000	0,0000	*			
DXS10146	0,0321	0,8236	0,2311	0,4600	0,1354	0,1692	0,0752	0,0870	0,7245	*		
DXS10134	0,0878	0,3203	0,1072	0,8535	0,4890	0,3147	0,2223	0,2771	0,7736	0,4781	*	
DXS7423	0,6574	0,9354	0,3187	0,7217	0,2586	0,5722	0,2806	0,6879	0,8339	0,0000	0,1470	*

Legenda: Em vermelho, estão os valores de *p* significativos (nível de significância = 0,00076, após a aplicação da correção de *Bonferroni*) para 100.172 permutações (s.e. <0,0015).

Na Tabela 10, calculamos os parâmetros de diversidade, bem como as distâncias genéticas que separam os marcadores mais distantes em cada grupo de ligação. LG 3 apresentou as menores distâncias genéticas entre os loci, quando comparado aos outros grupos de ligação. No entanto, a distância genética por si só não explica os resultados obtidos, e uma maior taxa de mutação e de diversidade de haplótipos pode levar a uma menor sensibilidade de amostras de baixo efetivo para se detectar LD. De fato, podemos ver que o LG 3 apresenta a menor taxa de mutação e de diversidade de haplótipos na população miscigenada equatoriana (Tabela 10). Apoiando a necessidade de um maior número de amostras para detectar LD, o LG 1 apresentou a maior diversidade de haplótipos, o LG 2 teve a maior taxa de mutação e de diversidade média por locus e o LG 4 apresentou a maior distância genética entre loci.

Tabela 10 - Taxa média de mutação e diversidade genética por *locus*, número de haplótipos diferentes, diversidade de haplótipos e distância genética entre os marcadores mais distantes em cada grupo de ligação

	Taxa média de mutação¶	**Diversidade média por *locus*§**	**Número de haplótipos§**	**Diversidade haplotípica§**	**Distância genética (cM)**
LG 1	5,29E-03	0,793	178	0,995	0,37
LG 2	5,79E-03	0,821	158	0,993	0,38
LG 3	2,17E-03	0,766	102	0,971	0,36
LG 4	2,89E-03	0,804	158	0,994	0,46

Legenda: ¶As taxas de mutação foram obtidas em PINTO e colaboradores (2020); §Estes valores foram calculados na amostra do Equador (presente estudo)

3.3 Análise de desequilíbrio de ligação entre o conjunto completo de *loci*

Para aumentar a possibilidade de detecção de baixos níveis de LD, uma nova análise de LD foi realizada considerando os doze X-STRs e os 32 X-Indels simultaneamente, usando um subconjunto de 318 amostras masculinas (aquelas genotipadas para todos os *loci*). Além das associações já discutidas para X-STRs nos LGs, valores significativos para LD foram encontrados entre pares de X-Indels a distâncias de 0,204 cM e menores (Tabela 11). Para esses pares de marcadores, valores de LD significativos foram previamente observados em outras populações de diferentes continentes (CAPUTO et al., 2017; EDELMANN et al., 2016; FERRAGUT et al., 2017; FREITAS et al., 2010; IBARRA et al., 2014; MARTINEZ et al.,

2019; PEREIRA et al., 2012b, 2012c; PINTO et al., 2015) (Tabela 11). Uma associação significativa também foi observada entre os indels MID3754 e MID3722, os quais estão separados por 1,433 cM, não detectada em outras populações previamente estudadas. Além disso, os X-STRs do LG 3 apresentaram uma associação significativa com os indels contíguos MID1839 e MID3736 (Tabela 11), mostrando que este LG possivelmente compreende seis *loci* que se estendem até um bloco de 4,396 cM.

No total, 20 das 946 comparações entre pares de *loci* revelaram valores de LD estatisticamente significativos. Em 13 casos (descritos acima), os marcadores estavam separados por menos de 2,8 cM. Uma vez que a população mestiça resultou de miscigenação recente, esses resultados podem ser explicados pela baixa recombinação entre os *loci*, preservando o LD na população por várias gerações. No entanto, associações significativas também foram observadas em sete comparações envolvendo pares de marcadores em diferentes grupos de ligação, que estão separados por distâncias variando entre 10,891 e 163,53 cM (Tabela 11). Junto com os resultados do teste de equilíbrio de HW (excesso de homozigosidade), as associações significativas observadas entre marcadores de diferentes grupos de ligação apontam para a presença de subestrutura dentro da população miscigenada equatoriana.

Para determinar se os níveis de LD observados na população miscigenada geral do Equador podem ser devido a diferenças entre as regiões, repetimos a análise considerando três subamostras, divididas de acordo com a geografia e história demográfica do país (conforme descrito na Seção 1): Amazônia, Andes e Costa do Pacífico. O padrão de LD na região dos Andes foi muito semelhante ao observado na população geral. Em contraste, as outras duas regiões apresentaram menos valores de *p* estatisticamente significativos, especialmente a Amazônia. No entanto, os valores de *p* dependem fortemente do tamanho da amostra e não podem ser usados para comparar LD entre amostras com tamanhos muito diferentes (KLING; DELL; TILLMAR, 2015). Por esse motivo, para avaliar a força das associações de LD encontradas, calculamos o valor R^2. O valor médio do R^2 foi menor nos Andes do que na Costa do Pacífico e na Amazônia (Figura 13).

No total, 20 das 946 comparações entre pares de *loci* revelaram valores de LD estatisticamente significativos. Em 13 casos (descritos acima), os marcadores estavam separados por menos de 2,8 cM. Uma vez que a população mestiça resultou de miscigenação recente, esses resultados podem ser explicados pela baixa recombinação entre os *loci*, preservando o LD na população por várias gerações. No entanto, associações significativas também foram observadas em sete comparações envolvendo pares de marcadores em diferentes

grupos de ligação, que estão separados por distâncias variando entre 10,891 e 163,53 cM (Tabela 11). Junto com os resultados do teste de equilíbrio de HW (excesso de homozigosidade), as associações significativas observadas entre marcadores de diferentes grupos de ligação apontam para a presença de subestrutura dentro da população miscigenada equatoriana.

Para determinar se os níveis de LD observados na população miscigenada geral do Equador podem ser devido a diferenças entre as regiões, repetimos a análise considerando três subamostras, divididas de acordo com a geografia e história demográfica do país (conforme descrito na Seção 1): Amazônia, Andes e Costa do Pacífico. O padrão de LD na região dos Andes foi muito semelhante ao observado na população geral. Em contraste, as outras duas regiões apresentaram menos valores de *p* estatisticamente significativos, especialmente a Amazônia. No entanto, os valores de *p* dependem fortemente do tamanho da amostra e não podem ser usados para comparar LD entre amostras com tamanhos muito diferentes (KLING; DELL; TILLMAR, 2015). Por esse motivo, para avaliar a força das associações de LD encontradas, calculamos o valor R^2. O valor médio do R^2 foi menor nos Andes do que na Costa do Pacífico e na Amazônia (Figura 13).

Tabela 11 - Resultados do teste exato de desequilíbrio de ligação para todos os pares de marcadores que apresentaram valores de *p* significativos (em vermelho) ou menores que 0,009 (em verde), em uma amostra de 318 haplótipos

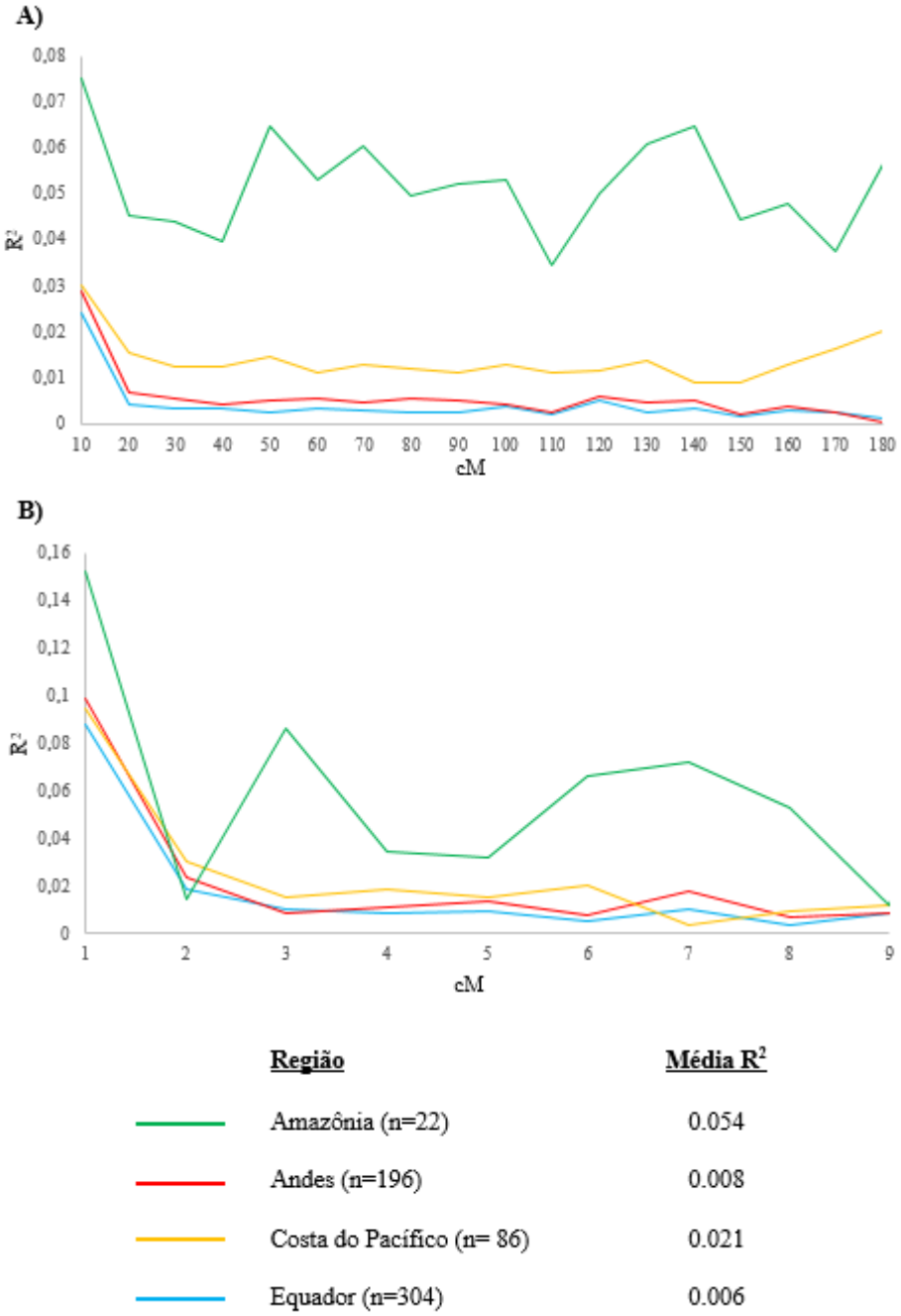
Loci		Valores de p	Distância cM	LD em outras populações
357	356	0,000000	0,007	Europa, África, Ásia leste e América [1][2][3][4][5][6][7][8][9]
3690	3719	0,000000	0,051	Europa, África, Ásia leste e América [3][5][6][8]
HPRTB	DXS10101	0,000000	0,060	Europa, África, Ásia leste e América [3][5][9][10][11][12]
3719	2089	0,000000	0,127	Europa, África, Ásia leste e América [3][5][6][8]
3703	3774	0,000000	0,204	América [5][6]
DXS10135	DXS8378	0,000000	0,220	Europa, África, Ásia leste e América [3][5][9][10][11][12]
DXS10103	HPRTB	0,000000	0,300	Europa, África, Ásia leste e América [3][5][9][10][11][12]
DXS10103	DXS10101	0,000000	0,360	Europa, África, Ásia leste e América [3][5][9][10][11][12]
DXS10146	DXS7423	0,000000	0,460	Europa, África, Ásia leste e América [3][5][9][10][11][12]
3740	3727	0,008515	1,237	Não detectado em populações previamente estudadas
3754	3722	0,000000	1,433	Não detectado em populações previamente estudadas
3736	DXS10103	0,000000	1,639	Não testado antes
3736	DXS10101	0,006733	1,999	Não testado antes
DXS10101	1839	0,000000	2,397	Não testado antes
DXS10103	1839	0,000000	2,757	Não testado antes
356	3774	0,004158	3,197	Não detectado em populações previamente estudadas
357	3774	0,002475	3,204	Não detectado em populações previamente estudadas

3712	3774	0,001782	4,383	Não detectado em populações previamente estudadas
111	DXS10103	0,005644	4,583	Não testado antes
3774	3692	0,000693	6,508	Não detectado em populações previamente estudadas
2637	3753	0,000792	6,513	Não detectado em populações previamente estudadas
3722	243	0,004950	7,488	Não detectado em populações previamente estudadas
243	3736	0,006139	10,130	Não detectado em populações previamente estudadas
2637	1839	0,006733	10,700	Não detectado em populações previamente estudadas
3712	3692	0,000000	10,891	Não detectado em populações previamente estudadas
DXS10148	3692	0,003564	16,849	Não testado antes
2692	DXS10079	0,001980	24,784	Não testado antes
3760	329	0,004257	26,211	Não detectado em populações previamente estudadas
DXS10101	2652	0,000000	45,969	Não detectado em populações previamente estudadas
HPRTB	2652	0,004554	46,029	Não testado antes
3692	DXS10074	0,001683	54,551	Não testado antes
DXS10074	HPRTB	0,000000	58,350	Não detectado em populações previamente estudadas
DXS10074	DXS10101	0,000000	58,410	Não detectado em populações previamente estudadas
3719	3732	0,006634	61,955	Não detectado em populações previamente estudadas
DXS8378	DXS10079	0,000000	71,000	Não detectado em populações previamente estudadas
DXS10135	3740	0,001089	91,138	Não testado antes
DXS10079	DXS10146	0,000000	92,530	Não detectado em populações previamente estudadas
DXS10148	HPRTB	0,007525	129,750	Não detectado em populações previamente estudadas
3716	DXS10146	0,005050	136,874	Não testado antes
DXS8378	DXS10146	0,000000	163,530	Não detectado em populações previamente estudadas

Legenda: Em vermelho, estão os valores de *p* significativos (nível de significância = 0,000053, após a aplicação da correção de Bonferroni), para 100.172 permutações (s.e. <0,0015). Em verde, estão os valores de *p* menores que 0,009. [1] (CAPUTO et al., 2017); [2] (EDELMANN et al., 2016); [3] (FERRAGUT et al., 2017); [4] (FREITAS et al., 2010); [5] (GOMES et al., 2017); [6] (IBARRA et al., 2014); [7] (MARTINEZ et al., 2019); [8] (PEREIRA et al., 2012a); [9] (PINTO et al., 2015); [10] (KLING; DELL; TILLMAR, 2015); [11] (PINTO et al., 2020); [12] (TOMAS; PEREIRA; MORLING, 2012)

Ao calcular os valores médios de R^2 para marcadores agrupados por distâncias genéticas, podemos observar que o R^2 diminui com a distância entre os pares de marcadores (Figura 13). Há uma grande diminuição quando a distância aumenta de 1 para 2 cM e uma leve diminuição de 2 para 3 cM. Acima desta distância, os valores tornam-se estáveis, exceto para a Amazônia, que mostra LD basal entre marcadores independentemente da distância genética (Figura 13b). Pelo que se sabe sobre a Amazônia, a ação sincrônica de miscigenação recente e deriva genética acentuada por efeito gargalo (BARBIERI et al., 2019) provavelmente agiu para criar o nível observado de LD nesta região.

Figura 13 - Valores médios de R^2 medidos na amostra total e nas três regiões equatorianas. (A) Valores médios de R^2 (eixo Y) calculados para pares de marcadores agrupados em intervalos de 10 cM (eixo X). (B) Valores médios de R^2 (eixo Y) calculados para pares de marcadores agrupados em intervalos de 1 cM (eixo X) para distâncias abaixo de 9 cM



3.4 Parâmetros de diversidade, distâncias genéticas e AMOVA

Os valores médios de diversidade genética entre o total de *loci* mostraram uma heterozigosidade esperada ligeiramente menor na amostra da Amazônia (Tabela 12), o que pode ser devido ao pequeno tamanho da amostra. Para aumentar o tamanho da amostra, as diversidades genéticas foram estimadas novamente para cada marcador, de forma independente (Tabela 12).

Tabela 12 - Valores médios de heterozigosidade esperada calculados para o conjunto completo de marcadores do cromossomo X, doze X-STRs e 32 X-Indels, nas três regiões equatorianas

	Amazonia	Andes	Costa do Pacífico
	32 X-Indels (N=355)		
N	27	222	106
Média Heterozigosidade.	0,331	0,346	0,347
s.e.	0,184	0,136	0,142
	12 X-STRs (N=526)		
N	36	344	147
Média Heterozigosidade.	0,792	0,793	0,811
s.e.	0,095	0,096	0,089
	Conjunto completo (N=304)		
N	22	196	86
Média Heterozigosidade.	0,451	0,468	0,473
s.e.	0,265	0,240	0,242

Legenda: N: número de amostras; s.e: erro padrão

Para avaliar a estrutura genética da população miscigenada equatoriana, foram realizadas análises de AMOVA e F_{ST} considerando as três regiões estudadas e o conjunto completo de *loci* do cromossomo X. A análise AMOVA mostrou que 99,29% da variação está dentro das populações (F_{ST} = 0,0071; p < 0,00000).

Os resultados de AMOVA para os X-STRs e para os X-Indels, calculados de forma separada, mostraram, novamente, valores baixos de F_{ST} (0,0038 e 0,0091, respectivamente).

Valores baixos de F_{ST} também foram obtidos entre as três regiões do Equador (Tabela **13**). Um teste de não diferenciação mostrou um valor de F_{ST} estatisticamente não significativo entre as regiões da Amazônia e dos Andes. Esses resultados podem estar relacionados ao fato de que os Andes e a Amazônia foram amplamente habitados por nativos americanos com o mesmo dialeto (Quichua), mas esses resultados também podem ser devido a uma proporção semelhante de ancestralidade nativa americana nas duas regiões. O teste de não diferenciação realizado entre a Costa do Pacífico e as outras duas regiões equatorianas mostrou valores de *p* estatisticamente significativos (Tabela **13**). Portanto, nas análises subsequentes, dividimos os dados por região.

Tabela 13 - Distâncias genéticas (F_{ST}) entre as três regiões do Equador (diagonal inferior) e os correspondentes valores de não diferenciação *p* (diagonal superior), para todos os marcadores do cromossomo X

	N	**Amazônia**	**Andes**	**Costa do Pacífico**
Amazônia	22	*	0,01802	0,00000
Andes	196	0,0097	*	0,00000
Costa do Pacífico	86	0,0211	0,0048	*

Legenda: Em vermelho, estão os valores de *p* significativos (nível de significância = 0,01667, após a aplicação da correção de *Bonferroni*), após 10.100 permutações (s.e. <0,005)

3.5 Comparação com outras populações

As distâncias genéticas foram avaliadas de forma separada para os dois tipos de marcadores, uma vez que nenhuma outra população foi estudada com o conjunto completo de 44 *loci* do cromossomo X incluídos neste estudo (Tabela **14** e 15).

Para os marcadores estudados, foram detectadas diferenças significativas entre as populações equatorianas e as populações miscigenadas do Brasil ($0.069 < F_{ST} < 0.085$, para X-Indels; $0.013 < F_{ST} < 0.032$, para X-STRs), Colômbia ($0.020 < F_{ST} < 0.074$ para X-Indels), e Argentina ($0.005 < F_{ST} < 0.021$, para X-STRs). No entanto, não foram encontrados valores de F_{ST} estatisticamente significativos em comparações com populações do sudoeste dos Andes, México, Peru e Rio Negro (Tabela **14** e 15). Isso está de acordo com a ancestralidade nativa americana relativamente alta encontrada para essas populações (CORACH et al., 2010; NORRIS et al., 2018; OSSA et al., 2016). Junto com a região sudoeste dos Andes da Colômbia,

o Equador tem uma maior ancestralidade nativa americana do que outros países da América do Sul, incluindo Peru e México.

Na representação gráfica (MDS) dos valores de F_{ST} (Figura 14), os dois conjuntos de marcadores do cromossomo X separam claramente as populações europeias, africanas e nativas americanas (Figura 14). Esses três grupos de referência estão distantemente posicionados, com as populações miscigenadas da América do Sul localizadas entre eles.

Tabela 14 - Distâncias genéticas (F_{ST}) entre as populações estudadas (diagonal inferior) e os valores correspondentes de probabilidade de não diferenciação (diagonal superior), para os 23 X-indels

População¶	N	Referência	África	Europa	Nativos Americanos (Pastos)	Amazônia	Andes	Costa do Pacífico	Lima	México DF	Centro-Leste Andino	Centro-Oeste Andino	Orinoquia	Sudoeste Andino	Região do Caribe	Costa do Pacífico Norte	São Paulo
África	47	[1]	*	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0180	0,0000
Europa	191	[1]	0,1299	*	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Colômbia, Nativos Americanos (Pastos)	15	[2]	0,1729	0,1739	*	0,0451	0,0180	0,0811	0,0090	0,0000	0,0000	0,0000	0,0000	0,6216	0,0000	0,0000	0,0000
Equador, Amazônia	27	Este estudo	0,1422	0,1343	0,0332	*	0,3604	0,0360	0,1982	0,2613	0,0000	0,0000	0,0000	0,2162	0,0000	0,0000	0,0000
Equador, Andes	222	Este estudo	0,1204	0,1158	0,0252	0,0016	*	0,0991	0,1081	0,1982	0,0000	0,0000	0,0000	0,0901	0,0000	0,0000	0,0000
Equador, Costa do Pacífico	106	Este estudo	0,1039	0,1247	0,0192	0,0214	0,0031	*	0,0721	0,0180	0,0000	0,0000	0,0000	0,0631	0,0000	0,0000	0,0000
Peru, Lima	41	[1]	0,1090	0,1053	0,0437	0,0078	0,0075	0,0080	*	0,5946	0,0090	0,0000	0,0000	0,0090	0,0000	0,0000	0,0000
México, Mexico DF	32	[1]	0,1167	0,0719	0,0586	0,0057	0,0068	0,0191	0,0000	*	0,2613	0,0000	0,0090	0,0451	0,0000	0,0000	0,0000
Colômbia, Centro-Leste Andino	164	[1][2]	0,0857	0,0391	0,0660	0,0306	0,0197	0,0267	0,0203	0,0031	*	0,1351	0,0360	0,0090	0,0000	0,0000	0,0000
Colômbia, Centro-Oeste Andino	47	[2]	0,0931	0,0185	0,1129	0,0678	0,0518	0,0626	0,0475	0,0349	0,0055	*	0,0000	0,0000	0,0090	0,0000	0,0721
Colômbia, Orinoquia	59	[2]	0,0639	0,0409	0,0896	0,0658	0,0498	0,0411	0,0366	0,0200	0,0082	0,0253	*	0,0000	0,0090	0,0000	0,0000
Colômbia, Sudoeste Andino	26	[2]	0,1115	0,1235	0,0000	0,0075	0,0099	0,0148	0,0268	0,0238	0,0284	0,0707	0,0500	*	0,0090	0,0090	0,0000
Colômbia, Região do Caribe	52	[2]	0,0387	0,0635	0,0763	0,0425	0,0470	0,0467	0,0373	0,0281	0,0159	0,0287	0,0175	0,0290	*	0,1982	0,0000
Colômbia, Costa do Pacífico Norte	34	[2]	0,0225	0,0977	0,1174	0,0729	0,0735	0,0740	0,0742	0,0709	0,0467	0,0665	0,0395	0,0521	0,0042	*	0,0000
Brasil, São Paulo	218	[3]	0,0798	0,0091	0,1237	0,0847	0,0691	0,0752	0,0677	0,0417	0,0129	0,0051	0,0158	0,0777	0,0269	0,0486	*

Legenda: Em vermelho, estão os valores de *p* significativos (nível de significância = 0,000641, após a aplicação da correção de *Bonferroni*), após 10.100 permutações (s.e. <0,005). ¶Apenas perfis masculinos foram incluídos. [1] (THE 1000 GENOMES PROJECT CONSORTIUM, 2015); [2] (IBARRA et al., 2014); [3] (MARTINEZ et al., 2019)

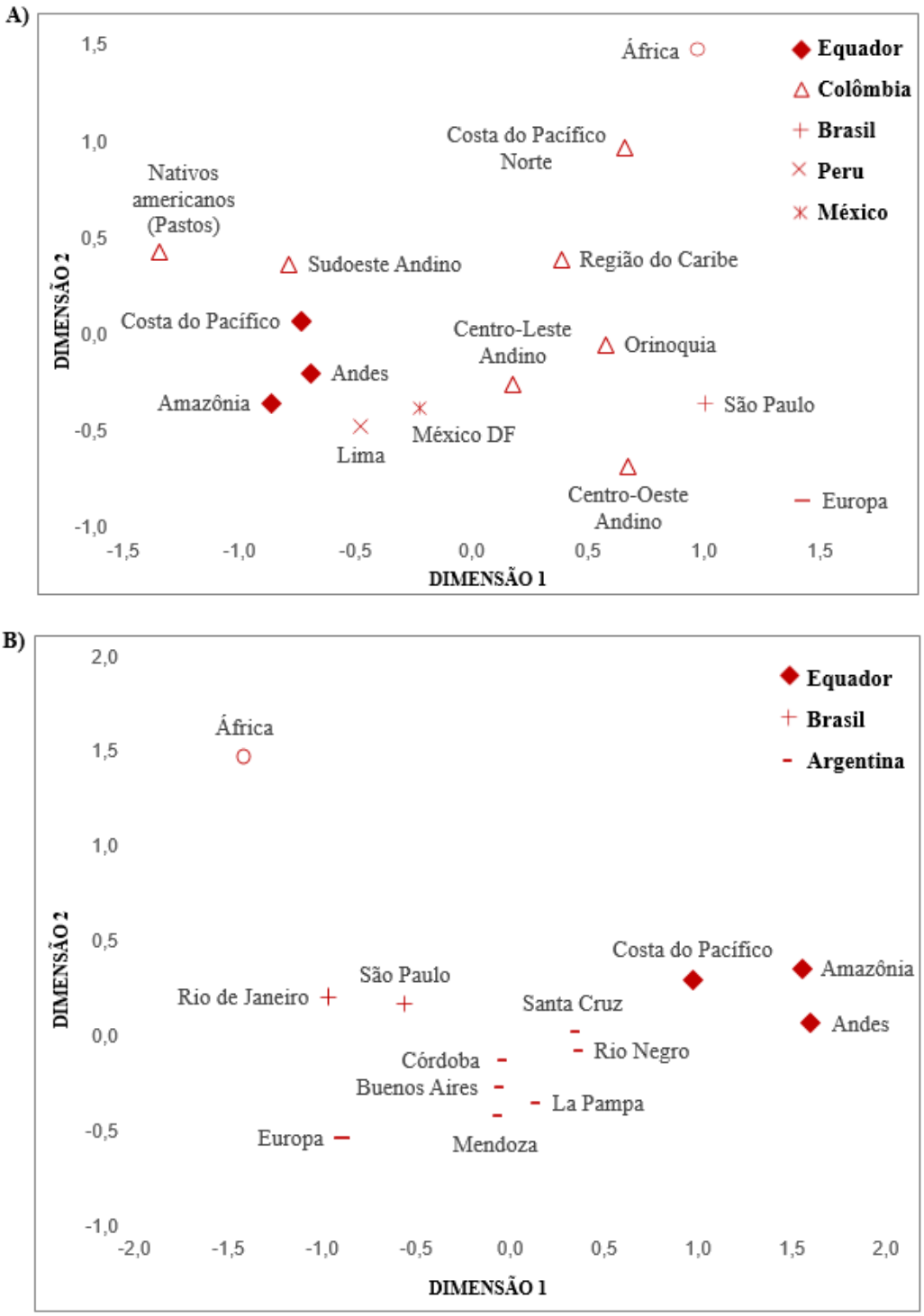
Tabela 15 - Distâncias genéticas (F_{ST}) entre as populações estudadas (diagonal inferior) e os valores correspondentes de probabilidade de não diferenciação, *p* (diagonal superior), para os doze X-STRs

População¶	N	Referência	Amazônia	Andes	Costa do Pacífico	La Pampa	Córdoba	Buenos Aires	Rio Negro	Santa Cruz	Mendoza	Rio de Janeiro	São Paulo	África	Europa
Equador, Amazônia	36	Este estudo	*	0,8378	0,4144	0,0000	0,0000	0,0000	0,0000	0,0180	0,0000	0,0000	0,0000	0,0000	0,0000
Equador, Andes	344	Este estudo	0,0027	*	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Equador, Costa do Pacífico	147	Este estudo	0,0008	0,0042	*	0,0000	0,0000	0,0000	0,0991	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Argentina, La Pampa	136	[1]	0,0146	0,0119	0,0049	*	0,1351	0,5856	0,3063	0,1171	0,4505	0,0000	0,0000	0,0000	0,0000
Argentina, Córdoba	286	[1]	0,0147	0,0141	0,0084	0,0013	*	0,6577	0,0901	0,1081	0,8559	0,0000	0,0000	0,0000	0,0000
Argentina, Buenos Aires	140	[1]	0,0164	0,0147	0,0066	0,0004	0,0006	*	0,3243	0,2072	0,7568	0,0000	0,0090	0,0000	0,0000
Argentina, Rio Negro	142	[1]	0,0095	0,0089	0,0019	0,0008	0,0012	0,0006	*	0,6036	0,0901	0,0000	0,0000	0,0000	0,0000
Argentina, Santa Cruz	250	[1]	0,0070	0,0092	0,0045	0,0014	0,0010	0,0008	0,0003	*	0,0000	0,0000	0,0000	0,0000	0,0000
Argentina, Mendoza	140	[1]	0,0208	0,0145	0,0099	0,0001	0,0009	0,0007	0,0017	0,0023	*	0,0000	0,0000	0,0000	0,0000
Brasil, Rio de Janeiro	266	[1]	0,0315	0,0319	0,0208	0,0077	0,0062	0,0055	0,0094	0,0097	0,0068	*	0,2973	0,0000	0,0000
Brasil, São Paulo	282	[1]	0,0227	0,0236	0,0131	0,0037	0,0033	0,0024	0,0049	0,0054	0,0038	0,0003	*	0,0000	0,0000
África	332	[2]	0,0462	0,0493	0,0344	0,0263	0,0242	0,0267	0,0259	0,0263	0,0261	0,0108	0,0118	*	0,0000
Europa	572	[1]	0,0369	0,0325	0,0214	0,0064	0,0058	0,0036	0,0087	0,0112	0,0047	0,0038	0,0043	0,0229	*

Legenda: Em vermelho, estão os valores de *p* significativos (nível de significância = 0,000641, após a aplicação da correção de *Bonferroni*), após 10.100 permutações (s.e. <0,005). ¶Apenas perfis masculinos foram incluídos. [1] (PINTO et al., 2020); [2] (GOMES et al., 2017)

Para os X-Indels (Figura 14 a), a Costa do Pacífico equatoriana apresentou menor distância genética com os *Pastos* (nativos americanos) do que as populações da Amazônia e dos Andes. A etnia *Pastos* se estabeleceu nos territórios anteriormente conhecidos como "*Franja Andina*", que se estendem desde *Nariño* (departamento administrativo no sul da Colômbia) até a província de Imbabura (região dos Andes, norte do Equador) (URIBE, 1986). Atualmente, os *Pastos* estão apenas no sul da Colômbia. No entanto, existe um grupo étnico chamado Awá que compartilha o mesmo grupo linguístico dos *Pastos*. Os *Awás* também se estabeleceram na "*Franja Andina*", mas expandiram seus territórios ao norte da costa do Pacífico do Equador. Esses resultados sugerem que a menor distância entre a Costa do Pacífico e as populações dos *Pastos* pode ser devido a interações entre grupos étnicos ancestrais (Figura 14 a). No entanto, algumas populações no MDS estão deslocadas em uma direção diferente daquela dos *Pastos*, sugerindo diferentes grupos ancestrais de nativos americanos entre as populações analisadas.

Figura 14 - Análise de escala multidimensional (MDS) com base nas distâncias genéticas F_{ST} calculadas para (A) 32 X-Indels (*Stress* = 0.0678) e (B) 12 X-STRs (*Stress* = 0.0308)



Legenda: (A) Os dados usados foram retirados do CONSÓRCIO PROJETO 1000 GENOMAS (2015) (África, Colômbia, Peru e México), IBARRA *et al.* (2014) (Colômbia) e MARTINEZ *et al.* (2019) (Brasil). Apenas um subconjunto de 23 X-Indels foi usado nesta análise, uma vez que não há dados disponíveis para os 9 X-Indels restantes nas amostras do Projeto 1000 Genomas. (B) Os dados usados foram recuperados de PINTO *et al.* (2020) (Europa, Argentina e Brasil) e GOMES *et al.* (2017) (África).

3.6 Diferenciação populacional para indels de cromossomo X e autossômicos

Uma vez que as mulheres contribuem com dois terços dos cromossomos X em uma população (assumindo uma proporção igual de descendentes masculinos e femininos), as diferenças de ancestralidade para marcadores do cromossomo X e autossômicos refletirão acasalamentos enviesados. Portanto, foram comparadas as distâncias genéticas das três amostras equatorianas relativamente a populações europeias, africanas e nativas americanas, obtidas por PCA para indels do cromossomo X e autossômicos. Nesta análise, foram usados os 32 X-Indels e 46 indels informativos de ancestralidade (distribuídos em 18 dos 22 autossomos), relatados anteriormente nos mesmos grupos populacionais (PEREIRA et al., 2012b, 2012c; ZAMBRANO et al., 2019a).

Os resultados dos dois primeiros PCAs revelaram uma maior proximidade das populações equatorianas com os nativos americanos para o cromossomo X (Figura 15a) do que para Indels autossômicos (Figura 15b). Embora uma contribuição europeia de 26% tenha sido relatada por Zambrano e colaboradores (2019) para marcadores autossômicos, nenhum sinal de miscigenação europeia foi detectado na população amazônica para os Indels de cromossomo X. Estes resultados indicaram uma contribuição materna europeia inferior à paterna, o que é consistente com dados publicados de genomas não recombinantes (BAETA et al., 2012; BURGOS et al., 2019b, 2019a; TOSCANINI et al., 2018). Estudos anteriores comparando dados do mtDNA e do cromossomo Y, além das diferenças entre as populações miscigenadas do Equador, também revelaram processos de acasalamento enviesado por sexo, envolvendo predominantemente descendentes de homens europeus e mulheres nativas americanas. Em relação à herança materna, os estudos disponíveis mostram que a maioria das populações miscigenadas do Equador apresentam um conjunto de linhagens maternas que são principalmente de ascendência nativa americana (acima de 96%) (BAETA et al., 2012; BURGOS et al., 2019b). Em contraste, as linhagens de cromossomo Y são principalmente europeias (acima de 55%), embora as populações conservem uma parte significativa de sua ancestralidade paterna nativa americana (entre 26% na Costa do Pacífico e 44% na Amazônia) (BURGOS et al., 2019a; TOSCANINI et al., 2018).

Relativamente à contribuição africana detectada por meio de marcadores autossômicos na população da Costa do Pacífico 16%, segundo Zambrano e colaboradores (2019), para os marcadores de cromossomo X houve um aumento da distância para a população africana, bem como maior proximidade com o eixo que liga as populações nativas americanas e africanas

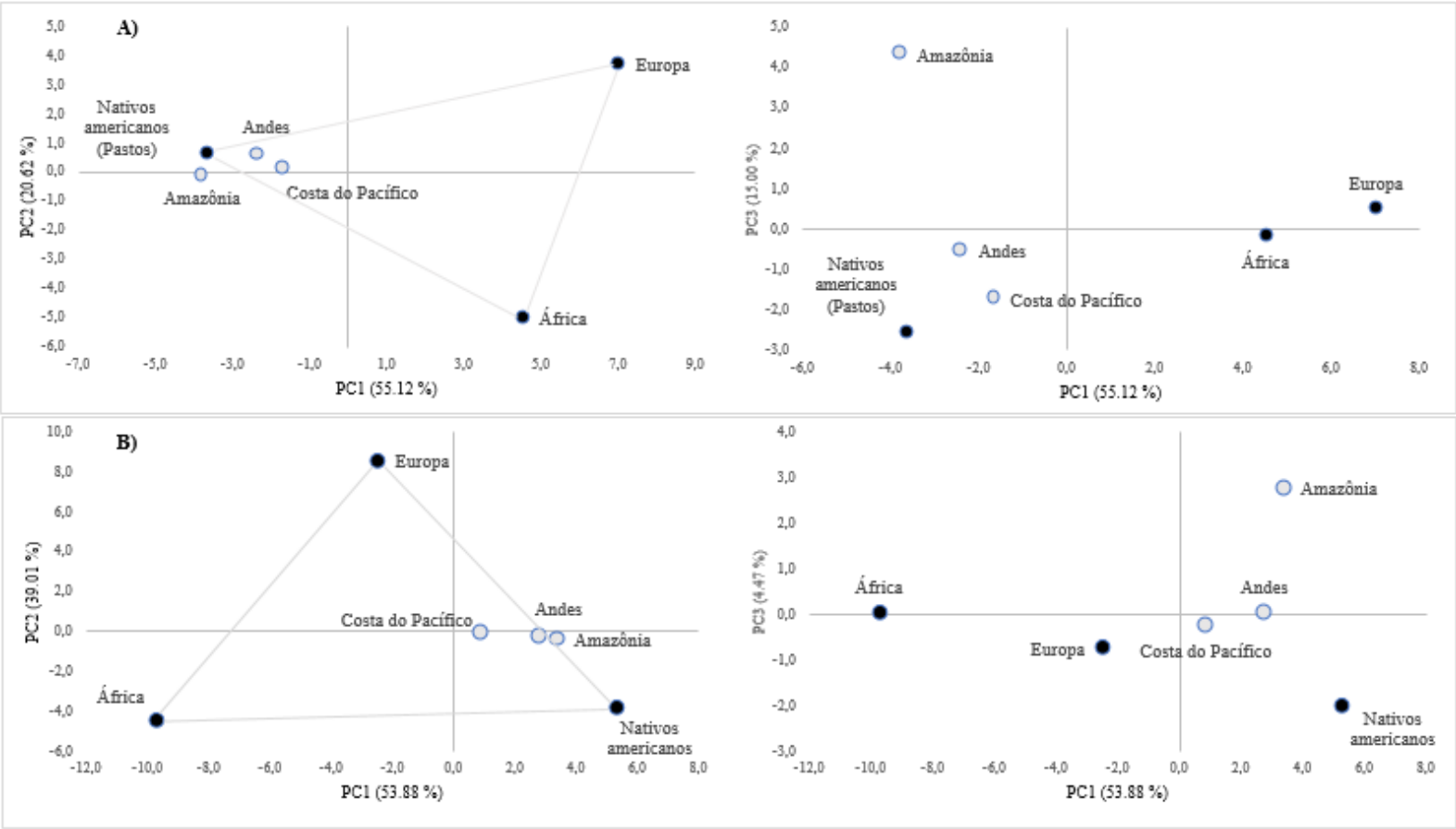
(Figura 15). Esses resultados podem ser explicados pelo acasalamento enviesado por sexo, não apenas entre homens europeus e mulheres africanas, mas também entre homens africanos e mulheres nativas americanas, o que não foi relatado anteriormente para esta população.

Quando o terceiro componente principal é investigado, podemos ver uma separação clara entre a Amazônia e as outras regiões equatorianas, tanto nas análises do cromossomo X quanto nas análises autossômicas (Figura 15a). Os X-Indels mostram uma ancestralidade nativa americana na Amazônia que é diferente daquela dos Andes, Costa do Pacífico e na população nativa americana dos *Pastos*. Os marcadores autossômicos mostram diferenças menores na ancestralidade nativa americana dessas populações, o que é consistente com uma maior mobilidade masculina do que feminina entre a Amazônia e as outras regiões.

Dada a discrepância no tamanho da amostra entre as três regiões estudadas, uma nova PCA foi realizada após subdividir a amostra total dos Andes em seis subgrupos de 37 indivíduos e a amostra da Costa do Pacífico em três subamostras de 35, 35 e 36 indivíduos cada. Os resultados obtidos são consistentes com aqueles obtidos ao usar as amostras totais (Figura 16). As subamostras dos Andes se agruparam perto do grupo de nativos, e as três amostras da Costa do Pacífico se agruparam próximo ao eixo entre as amostras de nativos e africanos.

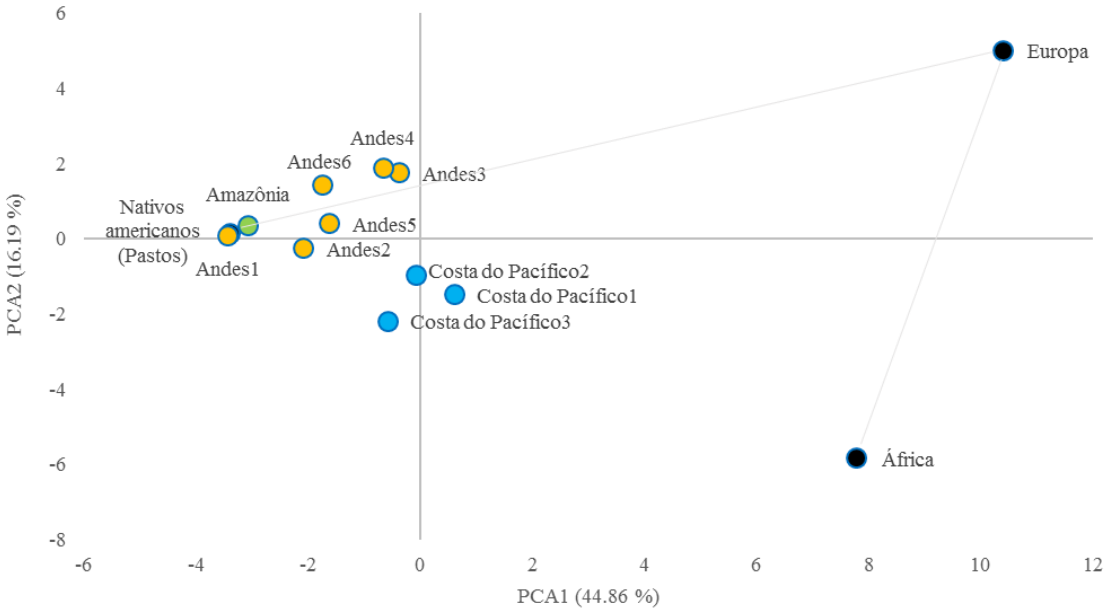
Além disso, a consistência dos resultados foi verificada através da análise das distâncias genéticas de Nei, medidas para os X-Indels (LIBIGER; NIEVERGELT; SCHORK, 2009). Da mesma forma, a amostra dos Andes se agrupou perto do grupo de nativos americanos, e a amostra da Costa do Pacífico se situou perto do eixo entre as amostras usadas como referência de nativos americanos e africanos (Figura **17**). Por outro lado, a terceira dimensão apresentou uma separação semelhante à encontrada na análise do terceiro componente principal (Figura **15**) entre a Amazônia e as outras regiões equatorianas (Figura **17**).

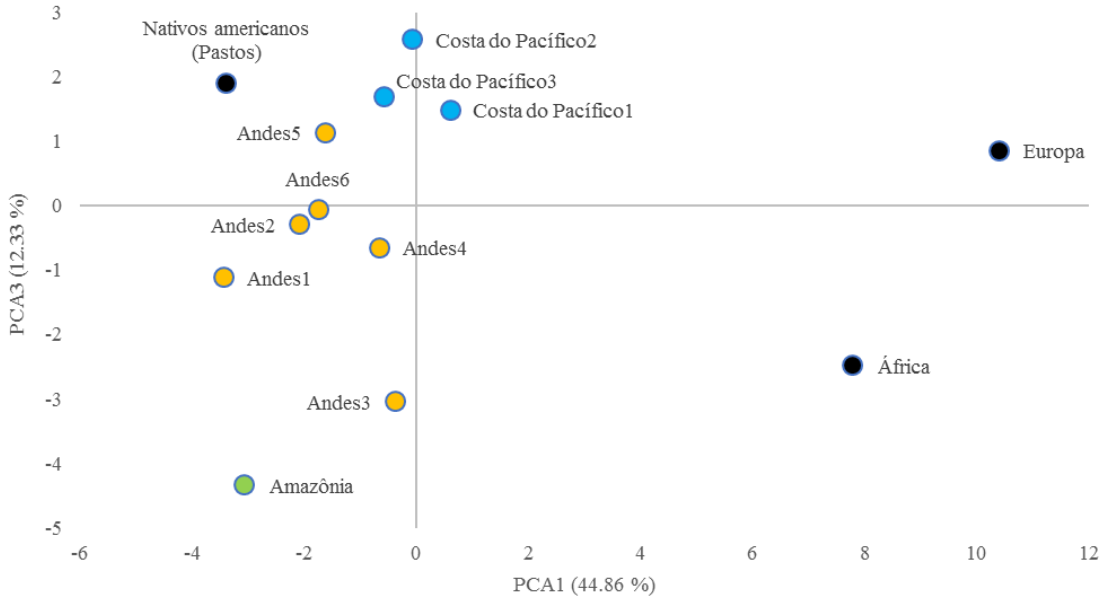
Figura 15 - Análise de componentes principais (PCA)



Legenda: Análise de componentes principais entre (A) Equador (este estudo) e populações de referência nativas americanas (IBARRA *et al.*, 2014), africanas e europeias (PROJETO 1000 GENOMAS, 2015) medidas por marcadores do cromossomo X; (B) Equador (ZAMBRANO *et al.*, 2019) e populações de referência nativas americanas, africanas e europeias (PEREIRA, PEREIRA, *et al.*, 2012; PEREIRA, PILLIPS, *et al.*, 2012) medidas por marcadores informativos de ancestralidade autossômica (AIMs).

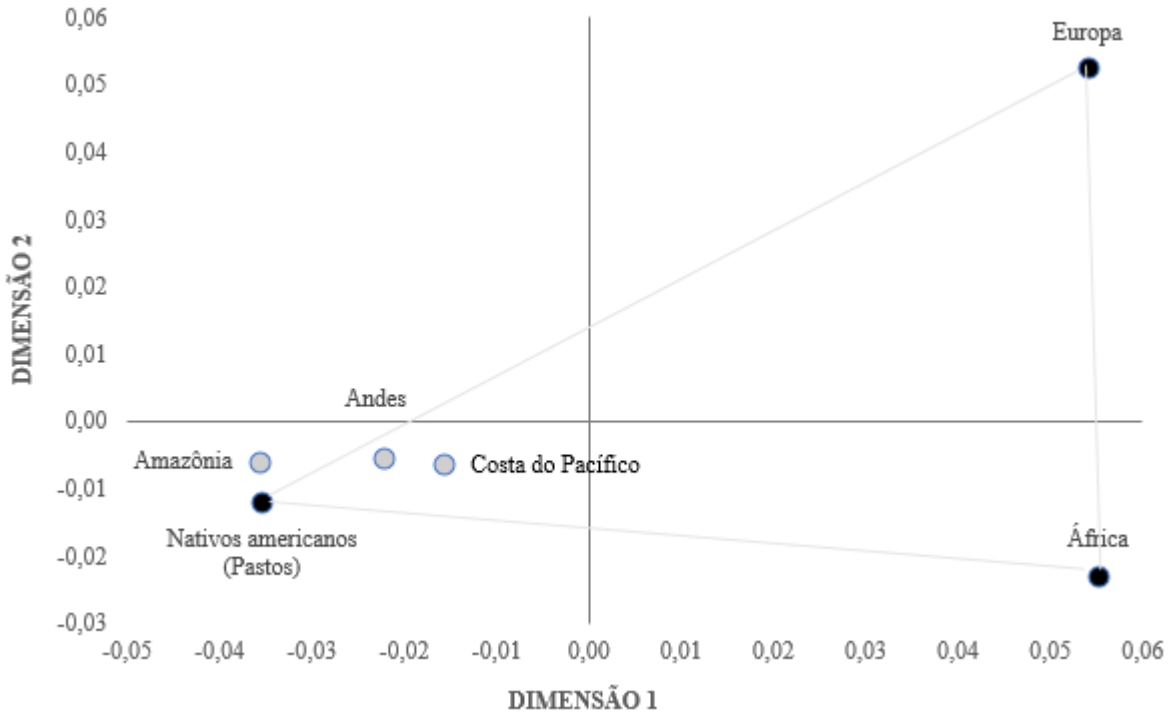
Figura 16 - Análise de componentes principais entre: populações de referência equatorianas (este estudo) e nativas americanas (IBARRA *et al.*, 2014), africanas e europeias (PROJETO 1000 GENOMAS, 2015) medidas por X-Indels

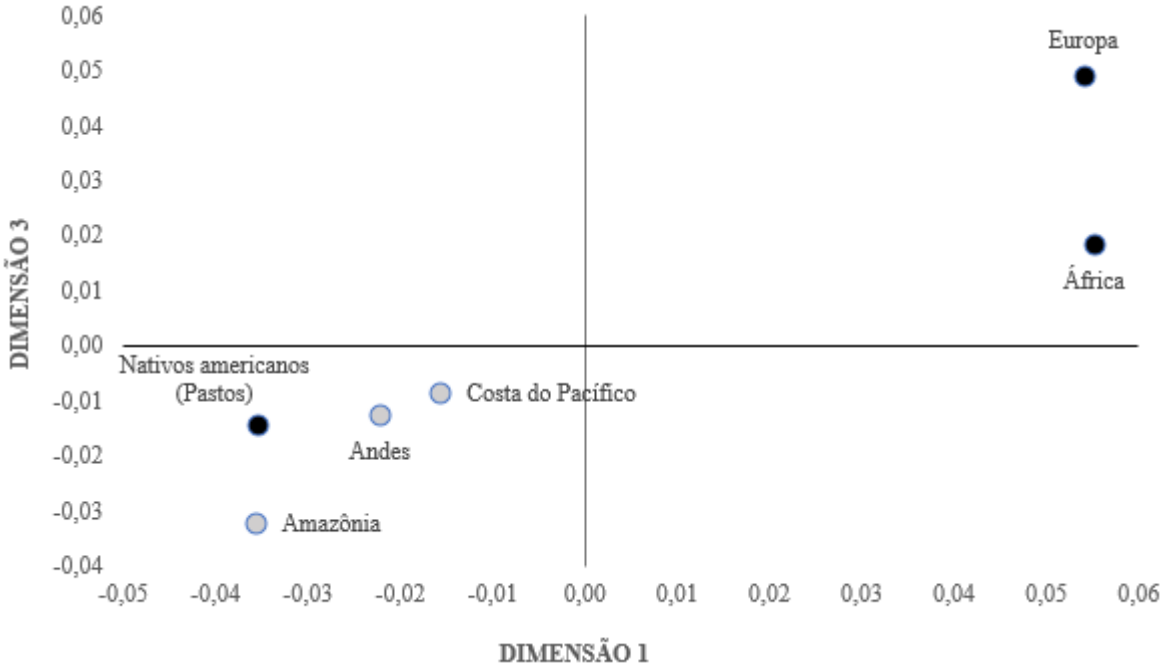




Legenda: Para obter tamanhos aproximados de cada uma das subamostras, as amostras totais dos Andes e da Costa do Pacífico foram subdivididas em seis subamostras de 37 perfis e três subamostras de 35/35/36 perfis, respectivamente.

Figura 17 - Análise de escala multidimensional (MDS) com base nas distâncias genéticas de Nei calculadas para os 32 X-Indels





Legenda: Análise de escala multidimensional (MDS) entre: populações de referência equatorianas (este estudo) e nativas americanas (IBARRA *et al.*, 2014), africanas e europeias (PROJETO 1000 GENOMAS, 2015) com base nas distâncias genéticas de Nei medidas para 32 X-Indels.

As populações da América do Sul têm histórias de colonização similares, caracterizadas pela chegada de europeus e, posteriormente, de africanos, os quais foram trazidos como escravos e espalhados pelo continente. No entanto, os processos de miscigenação ocorreram de forma diferenciada em todo o continente, dando origem a diversos grupos populacionais com proporções de ancestralidade nativa americana, europeia e africana muito variáveis. A diversidade que os grupos populacionais atuais apresentam é imensamente grande, a qual consegue-se observar na composição genética obtida pelo estudo de marcadores moleculares do cromossomo X.

A população do Equador, apesar de ter sofrido influência europeia e africana durante o período colonial, é caracterizada por preservar uma ancestralidade nativa americana significativa, quando comparada com outras populações da América do Sul. Estudos anteriores comparando dados do mtDNA e do cromossomo Y, revelaram que a composição genética atual da população equatoriana resultou de eventos de acasalamento enviesado por sexo, envolvendo predominantemente descendentes de homens europeus e mulheres nativas americanas. No entanto, este estudo, permitiu determinar que além do acasalamento entre homens europeus e mulheres nativas, existe um acasalamento enviesado por sexo entre homens africanos e mulheres nativas. Por outro lado, foi possível observar a presença de diferentes componentes de ancestralidade nativa americana, insinuando que este componente pode ter várias origens.

Nossos resultados indicam que a população miscigenada do Equador não é um grupo homogêneo, provavelmente refletindo diferenças na ancestralidade nativa e nos níveis de miscigenação continental nas três regiões analisadas. A região da Amazônia apresentou altos níveis de LD, o que pode ser devido a um isolamento geográfico prolongado, eventos recentes de miscigenação e/ou efeito da deriva genética devido a um baixo efetivo populacional. É importante ressaltar que apesar da divisão por regiões da amostra analisada, ainda há indícios de subestrutura populacional (altos níveis de LD). Portanto, para se obter uma visão mais abrangente da história genética e demográfica do Equador, é necessário coletar uma amostra mais ampla e representativa dessas regiões e avaliar uma divisão alternativa para as regiões.

A metodologia aplicada neste estudo permitiu comparar informações genéticas entre várias populações miscigenadas da America do Sul, e aprofundar o conhecimento existente sobre marcadores do cromossomo X, destacando a sua influência na composição genética das populações. Por outra parte, foram gerados dados que contribuirão para a formação de bases de dados robustas sobre a população miscigenadas do Equador, que poderão ser utilizados para fins de identificação forense ou como base para estudos de genética médica.

Até o momento, vários estudos genéticos têm sido realizados com o objetivo de analisar as populações atuais para obter informações das populações ancestrais, como as origens e rotas migratórias. Apesar disso, existem ainda várias questões não resolvidas, sendo interessante continuar a explorar através da análise integrada de diferentes tipos de marcadores, que oferecem informações únicas para complementar o entendimento histórico desde um ponto de vista genético e demográfico das populações da América do Sul. A importância de se integrar e estudar em conjunto as informações das diferentes regiões do genoma é refletida nos resultados obtidos nesta dissertação.

CONCLUSÕES

a) O estudo da população miscigenada do Equador mostrou uma alta diversidade genética no cromossomo X, e que a ancestralidade medida através de marcadores do cromossomo X é majoritariamente de origem nativa americana;

b) Os resultados das análises de cromossomo X mostraram um excesso de homozigosidade e associações significativas, envolvendo pares de marcadores em diferentes grupos de ligação, apontando para a presença de subestrutura genética dentro da população miscigenada equatoriana;

c) A população miscigenada do Equador é geneticamente mais próxima das populações do Rio Negro (Argentina), Lima (Peru), México e sudoeste andino Colombiano, de acordo com a ancestralidade nativa americana relativamente alta encontrada para essas populações. Por outro lado, observar-se a presença de diferentes componentes de ancestralidade nativa, sugerindo diferentes grupos ancestrais de nativos entre as populações analisadas;

d) Os resultados das comparações entre indels de cromossomo X e autossômicos, revelaram uma maior proximidade das populações equatorianas com os nativos americanos para o cromossomo X do que para indels autossômicos. Por outro lado, para o cromossomo X, foi detectada uma maior proximidade com o eixo que liga as populações nativas americanas e africanas, mostrando um acasalamento enviesado por sexo, entre homens africanos e mulheres nativas americanas, fato que não foi relatado anteriormente para esta população.

REFERÊNCIAS

AYALA, E. **Ecuador: Patria de todos. La nación ecuatoriana, unidad en la diversidad.** Quito.

AYALA, E. **Resumen de Historia del Ecuador** (Corporación Editora Nacional, Ed.). Quito.

BAETA, M.; NÚÑEZ, C.; AZNAR, J. M.; SOSA, C.; CASALOD, Y.; BOLEA, M.; GONZÁLEZ-ANDRADE, F.; DE PANCORBO, M. M.; MARTÍNEZ-JARRETA, B. Analysis of 10 X-STRs in three population groups from Ecuador. **Forensic Science International: Genetics**, [s. l.], v. 7, n. 1, p. 19–20, 2013.

BAETA, M.; NÚÑEZ, C.; GONZÁLEZ-ANDRADE, F.; SOSA, C.; CASALOD, Y.; BOLEA, M.; ZWEYNERT, S.; VACAS CRUZ, O. W.; GONZÁLEZ-SOLORZANO, J.; GEPPERT, M.; ROEWER, L.; MARTÍNEZ-JARRETA, B. Mitochondrial analysis revealed high homogeneity in the Waorani population-The last nomadic group of hunter-gatherers from Ecuador. **Forensic Science International: Genetics Supplement Series**, [s. l.], v. 2, n. 1, p. 313–314, 2009.

BAETA, M.; NÚÑEZ, C.; SOSA, C.; BOLEA, M.; CASALOD, Y.; GONZÁLES-ANDRADE, F.; ROEWER, L.; MARTÍNEZ-JARRETA, B. Mitochondrial diversity in Amerindian Kichwa and Mestizo populations from Ecuador. **Int J Legal Med**, v. 126, p. 299–302, 2012.

BAEZA, C.; GUZMÁN, R.; TIRADO, M.; LÓPEZ-PARRA, A. M.; RODRÍGUEZ, T.; MESA, M. S.; FERNÁNDEZ, E.; ARROYO-PARDO, E. Population data for 15 Y-chromosome STRs in a population sample from Quito (Ecuador). **Forensic Science International**, v. 173, p. 214–219, 2007.

BARBIERI, C.; BARQUERA, R.; ARIAS, L.; SANDOVAL, J. R.; ACOSTA, O.; ZURITA, C.; AGUILAR-CAMPOS, A.; TITO-ÁLVAREZ, A. M.; SERRANO-OSUNA, R.; GRAY, R. D.; MAFESSONI, F.; HEGGARTY, P.; SHIMIZU, K. K.; FUJITA, R.; STONEKING, M.; PUGACH, I.; FEHREN-SCHMITZ, L. The Current Genomic Landscape of Western South America: Andes, Amazonia, and Pacific Coast. **Molecular Biology and Evolution**, v. 36, n. 12, p. 2698–2713, 2019.

BURGOS, G.; GARZÓN-SALAZAR, A.; MALDONADO-UQUILLAS, K.; ÁVILA, C.; TOSCANO, M.; PAZ-CRUZ, E. A look of paternal ancestry in a sample of Ecuadorian "MESTIZO" population analyzed through PowerPlex Y23. **Forensic Science International: Genetics Supplement Series**, v. 7, n. 1, p. 534–536, 2019. a.

BURGOS, G.; SIMÃO, F.; FLORES-ESPINOZA, R.; YEPEZ-SANTOS, J. I.; GARZÓN-SALAZAR, A.; PAZ-CRUZ, E.; FREIRE-PASPUEL, B.; CARVALHO, E. F.; GUSMÃO, L. An approach to maternal ancestry in a sample of Ecuadorian "mestizo" population by sequencing the control region of mtDNA. **Forensic Science International: Genetics Supplement Series**, v. 7, n. 1, p. 537–538, 2019. b.

BUTLER, J. M. Short tandem repeat typing technologies used in human identity testing. **BioTechniques**, v. 43, n. 4, 2007.

BUTLER, J. M. **Advanced Topics in Forensic DNA Typing: Methodology**. California: Elsevier Inc., 2012.

CAPUTO, M.; AMADOR, M. A.; SANTOS, S.; CORACH, D. Potential forensic use of a 33 X-InDel panel in the Argentinean population. **International Journal of Legal Medicine**, v. 131, n. 1, p. 107–112, 2017.

CARDOSO, J. G. R.; ANDERSEN, M. R.; HERRGÅRD, M. J.; SONNENSCHEIN, N. Analysis of genetic variation and potential applications in genome-scale metabolic modeling. **Frontiers in Bioengineering and Biotechnology**, v. 3, n. 13, 2015.

CARDOSO, S.; ALFONSO-SÁNCHEZ, M.; VALVERDE, L.; SÁNCHEZ, D.; ZARRABEITIA, M. T.; ODRIOZOLA, A.; MARTÍNEZ-JARRETA, B.; DE PANCORBO, M. M. Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon. **Heredity**, v. 108, n. 6, p. 609–615, 2012.

CASTO, A. M.; LI, J. Z.; ABSHER, D.; MYERS, R.; RAMACHANDRAN, S.; FELDMAN, M. W. Characterization of X-Linked SNP genotypic variation in globally-distributed human populations. **Genome Biology**, v. 11, n. 1, 2010.

CHAN, Y.-M.; HANNEMA, S. E.; ACHERMANN, J. C. Disorders of Sex Development. In: **Williams Textbook of Endocrinology**. 14. ed. . p. 867–936.

CORACH, D.; LAO, O.; BOBILLO, C.; GAAG, K. Van Der; ZUNIGA, S.; VERMEULEN, M.; DUIJN, K. Van; GOEDBLOED, M.; VALLONE, P. M.; PARSON, W.; KNIJFF, P. De; KAYSER, M. Inferring Continental Ancestry of Argentineans from Autosomal , Y-Chromosomal and Mitochondrial DNA. **Annals of Human genetics**, v. 74, p. 65–76, 2010.

CROW, J. F. The beanbag lives on. **Nature**, v. 409, n. 6822, p. 771, 2001.

EDELMANN, J.; HERING, S.; AUGUSTIN, C.; SZIBOR, R. Indel polymorphisms-An additional set of markers on the X-chromosome. **Forensic Science International: Genetics Supplement Series**, v. 2, n. 1, p. 510–512, 2009.

EDELMANN, J.; KOHL, M.; DRESSLER, J.; HOFFMANN, A. X-chromosomal 21-indel marker panel in German and Baltic populations. **International Journal of Legal Medicine**, v. 130, n. 2, p. 357–360, 2016.

EDWARDS, A.; HAMMOND, H. A.; JIN, L.; CASKEY, C. T.; CHAKRABORTY, R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. **Genomics**, v. 12, n. 2, p. 241–253, 1992.

ELLEGREN, H. Heterogeneous mutation processes in human microsatellite DNA sequences. **Nature Genetics**, v. 24, n. 4, p. 400–402, 2000.

EXCOFFIER, L.; LISCHER, H. **An Integrated Software Package for Population Genetics Data Analsis**, 2015. Disponível em: <http://cmpg.unibe.ch/software/arlequin35>

EXCOFFIER, L.; SMOUSE, P.; QUATTRO, J. Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. **Genetics**, v. 131, n. 2, p. 479–491, 1992.

FERRAGUT, J. F.; BENTAYEBI, K.; PEREIRA, R.; CASTRO, J. A.; AMORIM, A.; RAMON, C.; PICORNELL, A. Genetic portrait of Jewish populations based on three sets of X-chromosome markers: Indels, Alu insertions and STRs. **Forensic Science International: Genetics**, v. 31, p. 5–11, 2017.

FREITAS, S. C. N.; RESQUE, R. L.; RIBEIRO-RODRIGUES, E. M.; GUERREIRO, J. F.; SANTOS, N. P. C.; RIBEIRO-DOS-SANTOS, Â.; SANTOS, S. X-linked insertion/deletion

polymorphisms: Forensic applications of a 33-markers panel. **International Journal of Legal Medicine**, v. 124, n. 6, p. 589–593, 2010.

GAMBÉLA, B. **Afri-Asiatic Anthropology Blog**. 2009. Disponível em: <https://billygambelaafroasiaticanthropology.wordpress.com/>.

GAVIRIA, A.; ZAMBRANO, A. K.; MOREJON, G.; AGUIRRE, V.; VELA, M.; BUILES, J. J.; BURGOS, G. Genetic characterization of an X-STR decaplex on the population of Pichincha (Ecuador). **Forensic science international. Genetics Supplement Series**, v. 4, n. 1, p. 144–145, 2013.

GILLESPIE, J. **Population Genetics**. United States: The Johns Hopkings University Press, 1998. v. 54

GOMES, I.; PEREIRA, P. J. P.; HARMS, S.; OLIVEIRA, A. M.; SCHNEIDER, P. M.; BREHM, A. Genetic characterization of Guinea-Bissau using a 12 X-chromosomal STR system: Inferences from a multiethnic population. **Forensic Science International: Genetics**, v. 31, p. 89–94, 2017.

GOMES, I.; PINTO, N.; ANTÃO-SOUSA, S.; GOMES, V.; GUSMÃO, L.; AMORIM, A. Twenty Years Later: A Comprehensive Review of the X Chromosome Use in Forensic Genetics. **Frontiers in Genetics**, v. 11, n. September, p. 1–17, 2020.

GONZÁLES-ANDRADE, F.; SÁNCHEZ, D.; MARTÍNEZ-JARRETA, B. EL MESTIZAJE GENÉTICO EN ECUADOR Y SU APLICACIÓN MÉDICO FORENSE. **Ciencia Forense**, v. 8, p. 133–154, 2006.

GONZÁLEZ-ANDRADE, F.; ROEWER, L.; WILLUWEIT, S.; SÁNCHEZ, D.; MARTÍNEZ-JARRETA, B. Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis. **Forensic Science International: Genetics**, v. 3, n. 3, p. 83–91, 2009.

GONZÁLEZ-ANDRADE, F.; SÁNCHEZ, D.; MARTÍNEZ-JARRETA, B.; BUDOWLE, B. Y-Chromosome STR Haplotypes in Three Different Population Groups From Ecuador (South America). **Forensic Science International**, v. 53, n. 2, p. 512–514, 2008.

GUSMÃO, L.; SÁNCHEZ-DIZ, P.; ALVES, C.; ABOVICH, M.; ATMETLLA, I.; BOBILLO, C.; CALVO, R.; CARVALHO, E.; CARVALHO, M.; ESPINOZA, M.; GARCÍA, Ó.; MALAGHINI, M.; MARTINS, J.; PINHEIRO, F.; PORTO, M. J.; RAIMONDI, E.; RIANCHO, J. A.; RODRÍGUEZ, A.; RODRÍGUEZ, A. A GEP-ISFG collaborative study on the optimization of an X-STR decaplex : data on 15 Iberian and Latin American populations. **Int J Legal Med**, v. 123, p. 227–234, 2009.

HARDY, G. H. Mendelian Proportions in a Mixed Population Published. **Advancement Of Science**, v. 28, n. 706, p. 49–50, 1908.

HASHIYADA, M. DNA biometrics. **Biometrics**, n. June, 2011.

HEDRICK, P. W. Gametic disequilibrium measures: proceed with caution. **Genetics**, v. 117, n. 2, p. 331–341, 1987.

HOUT, M. C.; PAPESH, M. H.; GOLDINGER, S. D. Multidimensional scaling. **Wiley Interdisciplinary Reviews: Cognitive Science**, v. 4, n. 1, p. 93–103, 2013.

IBARRA, A.; RESTREPO, T.; ROJAS, W.; CASTILLO, A.; AMORIM, A.; MARTÍNEZ, B.; BURGOS, G.; OSTOS, H.; ÁLVAREZ, K.; CAMACHO, M.; SUAREZ, Z.; PEREIRA,

R.; GUSMÃO, L. Evaluating the X chromosome-specific diversity of Colombian populations using insertion/deletion polymorphisms. **PLoS ONE**, v. 9, n. 1, p. 1–10, 2014.

INEC. **Instituto Nacional de Estadisticas y Censos. Proyecciones Poblacionales**. 2010. Disponível em: <https://www.ecuadorencifras.gob.ec/proyecciones-poblacionales/>.

JEFFREYS, A.; WILSON, V.; THEIN, S. Hypervariable “minisatellite” regions in human DNA. **Nature**, v. 314, n. 7, p. 67–73, 1985.

KLING, D.; DELL, B.; TILLMAR, A. FamLinkX – implementation of a general model for likelihood computations for X-chromosomal marker data. **Forensic Science International: Genetics**, v. 17, p. 1–7, 2015.

LAREU, M. **Short Tandem Repeats**. 2. ed. [s.l.] : Elsevier Ltd., 2013. Disponível em: <http://dx.doi.org/10.1016/B978-0-12-382165-2.00040-4>

LIBIGER, O.; NIEVERGELT, C. M.; SCHORK, N. J. Comparison of genetic distance measures using human SNP genotype data. **Human Biology**, v. 81, n. 4, p. 389–406, 2009.

LYON, M. . Gene Action in the X-Chromosome of the Mouse (Mus musculus L.). **Nature**, v. 190, p. 372–373, 1961.

MARTINEZ, J.; POLVERARI, F. S.; SILVA, F. A. de J.; BRAGANHOLI, D. F.; FERRAZ, J. A. M. L.; GUSMÃO, L.; CICARELLI, R. M. B. Genetic characterization of 32 X-InDels in a population sample from São Paulo State (Brazil). **International Journal of Legal Medicine**, v. 133, n. 5, p. 1385–1388, 2019.

NEI, M. F-statistics and gene diversity in subdivided populations. **Ann. Hum. Gent.**, v. 41, n. 225, p. 225–233, 1973.

NORRIS, E. T.; WANG, L.; CONLEY, A. B.; RISHISHWAR, L.; MARIÑO-RAMÍREZ, L.; VALDERRAMA-AGUIRRE, A.; JORDAN, I. K. Genetic ancestry, admixture and health determinants in Latin America. **BMC Genomics**, v. 19, n. 861, p. 76–99, 2018.

OHNO, S. Sex Chromosomes and Sex-Linked Genes. **Teratology**, Berlin, v. 4, n. 1, 1967.

OSSA, H.; AQUINO, J.; PEREIRA, R.; IBARRA, A.; OSSA, R. H.; GRANDA, J. D.; LATTIG, M. C.; GROOT, H. Outlining the Ancestry Landscape of Colombian Admixed Populations. **PLoS ONE**, v. 11, n. 10, p. 1–15, 2016.

PEREIRA, R.; PEREIRA, V.; GOMES, I.; TOMAS, C.; MORLING, N.; AMORIM, A.; PRATA, M.; CARRACEDO, A.; GUSMÃO, L. A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR. **Int J Legal Med**, v. 126, p. 97–105, 2012. a.

PEREIRA, R.; PILLIPS, C.; PINTO, N.; SANTOS, C.; EMANUEL BATISTA DOS SNTOS, S.; AMORIM, A.; CARRACEDO, A.; GUSMÃO, L. Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing. **PLoS ONE**, v. 7, n. 1, p. 1–10, 2012. b.

PEREIRA, V.; GUSMÃO, L.; VALENTE, C.; PEREIRA, R.; CARNEIRO, J.; GOMES, I.; MORLING, N.; AMORIM, A.; JOÃO PRATA, M. Refining the genetic portrait of Portuguese Roma through X-chromosomal markers. **American Journal of Physical Anthropology**, v. 148, n. 3, p. 389–394, 2012. c.

PIERCE, B. A. **Genetics**. 4. ed. New York: W. H. Freeman and Company, 2012.

PINTO, J. C.; PEREIRA, V.; MARQUES, S. L.; AMORIM, A.; ALVAREZ, L.; PRATA, M. J. Mirandese language and genetic differentiation in Iberia: A study using X chromosome markers. **Annals of Human Biology**, v. 42, n. 1, p. 20–25, 2015.

PINTO, N.; PEREIRA, V.; TOMAS, C.; LOIOLA, S.; CARVALHO, E. F.; MODESTI, N.; MAXZUD, M.; MARCUCCI, V.; CANO, H.; CICARELLI, R.; JANUARIO, B.; BENTO, A.; BRITO, P.; BURGOS, G.; PAZ-CRUZ, E.; DÍEZ-JUÁREZ, L.; VANNELLI, S.; LURDES, M. De; BERARDI, G.; FURFURO, S.; FERNANDEZ, A.; SUMITA, D.; BOBILLO, C.; GABRIELA, M.; GUSMÃO, L. Genetics Paternal and maternal mutations in X-STRs : A GHEP-ISFG collaborative study. **Forensic Science International: Genetics**, v. 46, 2020.

POULSEN, L.; BØRSTING, C.; TOMAS, C.; GONZÁLES-ANDRADE, F.; MORLING, N.; LOPEZ-PULLES, R.; GONZÁLES-SOLÓRZANO, J. Genetics Typing of Amerindian Kichwas and Mestizos from Ecuador with the SNP for ID multiplex. **Forensic Science International: Genetics**, v. 5, p. 105–107, 2011.

QIAGEN®. **Investigator Argus X-12 QS Handbook**, 2013. Disponível em: <https://www.qiagen.com/us/resources/download.aspx?id=d3ab4ec3-33ec-4be3-8a88-6b66ebebdcf0&lang=en>

RELETHFOR, J. **Human Population Genetics**. New Yersey: WILEY-BLACKWELL, 2011.

REYNOLDS, J.; WEIR, B.; CLARK, C. Estimation of the coancestry coefficient: basis for a short-term genetic distance. **Genetics**, v. 105, p. 767–779, 1983.

ROBERTS, D. F.; PAPIHA, S. S.; SSEBABI, C. T. Red Cell Isoenzymes in East Africa. **Human Biology**, v. 49, n. 3, p. 301–308, 1977.

ROSS, M. T. et al. The DNA sequence of the human X chromosome. **Nature**, v. 434, n. 7031, p. 325–337, 2005.

SALAZAR, J. **Cimarronaje en el Pacífico Sur**. Abya-Yala ed. Quito.

SANTANGELO, R.; GONZÁLEZ-ANDRADE, F.; BØRSTING, C.; TORRONI, A.; PEREIRA, V.; MORLING, N. Forensic Science International : Genetics Analysis of ancestry informative markers in three main ethnic groups from Ecuador supports a trihybrid origin of Ecuadorians. **Forensic Science International: Genetics**, v. 31, p. 29–33, 2017.

SCHAFFNER, S. F. The X chromosome in population genetics. **Nature reviews. Genetics**, England, v. 5, n. 1, p. 43–51, 2004.

SCIENCE STAFF. Males Mutate more than Females. **Science**, 1997.

SEHN, J. K. Insertions and Deletions (Indels). In: **Clinical Genomics**. . p. 129–150.

SLATKIN, M. A measure of population subdivision based on microsatellite allele frequencies. **Genetics**, v. 139, n. 1, p. 457–462, 1995.

SLEDDENS, H.; OOSTRA, B.; BRINKMANN, A.; TRAPMAN, J. Trinucleotide repeat polymorphism in the androgen receptor gene (AR). **Nucleic Acids Research**, v. 20, n. 6, p. 1427, 1992.

STEFFENS, F. E. What is principal components analysis? **Nature Biotechnology**, v. 26, n. 3, p. 303–304, 2008.

STONEKING, M. **An Introduction to Molecular Anthropology**. New Yersey: John Wiley

& Sons, Inc., 2017.

SZIBOR, R. X-chromosomal markers: Past, present and future. **Forensic Science International: Genetics**, v. 1, n. 2, p. 93–99, 2007.

TAO, R.; ZHANG, J.; SHENG, X.; ZHANG, J.; YANG, Z.; CHEN, C.; BIAN, Y.; LIU, X.; ZHANG, S.; LI, C. Development and validation of a multiplex insertion/deletion marker panel, SifaInDel 45plex system. **Forensic Science International: Genetics**, v. 41, p. 128–136, 2019.

TAUTZ, D. Nucleic Acids Research. **Nucleic Acids Research**, v. 17, n. 16, p. 6463–6471, 1989.

THE 1000 GENOMES PROJECT CONSORTIUM. A global reference for human genetic variation. **Nature**, v. 526, p. 68–74, 2015.

TOMAS, C.; PEREIRA, V.; MORLING, N. Analysis of 12 X-STRs in greenlanders, danes and somalis using argus X-12. **International Journal of Legal Medicine**, v. 126, n. 1, p. 121–128, 2012.

TOSCANINI, U.; GAVIRIA, A.; MOSCOSO, F.; VELA, M.; COBOS, S.; LUPERO, A.; ZAMBRANO, A.; MARTINÓN-TORRES, F.; CARABAJO-MARCILLO, A. Genetics The geographic mosaic of Ecuadorian Y-chromosome ancestry. **Forensic Science International: Genetics**, v. 33, p. 59–65, 2018.

URIBE, M. V. **Etnohistoría de las comunidades indígenas prehispánicas del sur de ColombiaAnuario Colombiano de Historia Social y de la Cultura**. .

WAHLUND, V. S. Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus Betrachtet. **Heredity**, v. 11, p. 1–42, 1928.

WALL, J.; PRITCHARD, J. Haplotype block and Linkage disequilibrium in the Human genome. **Nature Reviews Genetics**, v. 4, p. 587–597, 2003.

WEBER, J. L.; DAVID, D.; HEIL, J.; FAN, Y.; ZHAO, C.; MARTH, G. Human diallelic insertion/deletion polymorphisms. **American Journal of Human Genetics**, v. 71, n. 4, p. 854–862, 2002.

WEBER, J.; MAY, P. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. **American Journal of Human Genetics**, v. 44, n. 3, p. 388–396, 1989.

WRIGHT, S. Genetical Structure of Populations. **Nature**, v. 4215, p. 247–249, 1950.

YAMAMOTO, F. I.; CLAUSEN, H.; WHITE, T.; MARKEN, J.; HAKOMORI, S. I. Molecular genetic basis of the histo-blood group ABO system. **Nature**, v. 345, n. 6272, p. 229–233, 1990.

ZAMBRANO, A.; GAVIRIA, A.; COBOS-NAVARRETE, S.; GRUEZO, C.; RODRÍGUEZ-POLLIT, C.; ARMENDÁRIZ-CASTILLO, I.; GARCÍA-CÁRDENAS, J. M.; GUERRERO, S.; LÓPEZ-CORTÉS, A.; LEONE, P. E.; PÉREZ-VILLA, A.; GUEVARA-RAMÍREZ, P.; YUMICEBA, V.; FIALLOS, G.; VELA, M.; PAZ-Y-MIÑO, C. The three-hybrid genetic composition of an Ecuadorian population using AIMs-InDels compared with autosomes, mitochondrial DNA and Y chromosome data. **Scientific Reports**, v. 9, n. 9247, p. 1–8, 2019. a.

ZAMBRANO, A.; GAVIRIA, A.; VELA, M.; COBOS, S.; LEONE, P. E.; GRUEZO, C.;

FIALLO, G.; GARCÍA-CÁRDENAS, J. M.; LÓPEZ-CORTÉS, A.; CABRERA-ANDRADE, A.; PAZ-Y-MIÑO, C. Ancestry characterization of Ecuador's Highland mestizo population using autosomal AIM-INDELs. **Forensic Science International: Genetics Supplement Series**, v. 6, p. 477–478, 2017.

ZAMBRANO, A. K.; MOREJÓN, G.; GALARZA, J.; PASTAS, V.; CASTILLO, G.; VELA, M.; GAVIRIA, A. Genetic data of an X-STR decaplex in Ecuadorian population (P). **Forensic science international. Genetics Supplement Series**, v. 5, p. 91–92, 2015.

ZAMBRANO, A. K.; VACA-PÓLIT, M.; BOADA, L.; VELA, M.; RODRÍGUEZ-PÓLIT, C.; FIALLOS, G.; GRUEZO, C.; PAZ-Y-MIÑO, C.; GAVIRIA, A. X-STR decaplex study in the population of Imbabura-Ecuador. **Forensic Science International: Genetics Supplement Series**, v. 7, n. 1, p. 288–290, 2019. b.

ZHANG, S.; LIN, Y.; BIAN, Y.; LI, C. Parallel sequencing of 60 X-chromosome genetic markers including STRs, SNPs and InDels. **Forensic Science International: Genetics Supplement Series**, v. 6, p. e317–e319, 2017.

APÊNDICE - Artigo "*Investigating genetic diversity in admixed populations from Ecuador*"

Received: 3 February 2021 | Revised: 21 April 2021 | Accepted: 23 May 2021
DOI: 10.1002/ajpa.24341

RESEARCH ARTICLE

American Journal of PHYSICAL ANTHROPOLOGY WILEY

Investigating genetic diversity in admixed populations from Ecuador

Rodrigo Flores-Espinoza[1,2] | Elius Paz-Cruz[3] | Viviana A. Ruiz-Pozo[3] | Marcelo Lopez-Carrera[3] | Alejandro Cabrera-Andrade[4,5] | Leonor Gusmão[1] | German Burgos[6,7]

[1]Laboratório de Diagnóstico por DNA (LDD), Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
[2]Laboratorios de Investigación, Universidad de Las Américas (UDLA), Quito, Ecuador
[3]Laboratorio de ADN, Fiscalía General del Estado, Quito, Ecuador
[4]Grupo de Bio-Quimioinformática, Universidad de Las Américas (UDLA), Quito, Ecuador
[5]Carrera de Enfermería, Facultad de Ciencias de la Salud, Universidad de Las Américas (UDLA), Quito, Ecuador
[6]Escuela de Medicina, Facultad de Ciencias de la Salud, Universidad de Las Américas (UDLA), Quito, Ecuador
[7]Grupo de Medicina Xenómica, Universidad de Santiago de Compostela, Santiago de Compostela, Spain

Correspondence
German Burgos, Escuela de Medicina, Facultad de Ciencias de la Salud, Universidad de Las Américas (UDLA), Quito, Ecuador.
Email: hgermanburgosf@yahoo.com

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Numbers: 131709/2020-8, 306342/2019-7; Dirección General de Investigación, Universidad de Las Américas, Quito, Ecuador, Grant/Award Number: MED.GB.18.06; Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Grant/Award Number: CNE-2018

Abstract

Objectives: According to demographic history, Ecuador has experienced shifts in its Native American populations caused by European colonization and the African slave trade. The continuous admixture events among Europeans, Native Americans, and Africans occurred differently in each region of the country, producing a stratified population. Thus, the aim of this study was to investigate the level of genetic substructure in the Ecuadorian Mestizo population.

Materials and methods: A total of 377 male and 209 female samples were genotyped for two sets of X-chromosomal markers (32 X-Indels and 12 X-STRs). Population analyses performed included Hardy–Weinberg equilibrium tests, LD analysis, PCA, pairwise F_{ST}s, and AMOVA.

Results: Significant levels of LD were observed between markers separated by distances of less than 1 cM, as well as between markers separated by distances varying from 10.891 to 163.53 cM. Among Ecuadorian regions, Amazonia showed the highest average R^2 value.

Discussion: When X-chromosomal and autosomal differentiation values were compared, a sex-biased admixture between European men and Native American and African women was revealed, as well as between African men and Native American women. Moreover, a distinct Native American ancestry was discernible in the Amazonian population, in addition to sex-biased gene flow between Amazonia and the Andes and Pacific coast regions. Overall, these results underline the importance of integrating X chromosome information to achieve a more comprehensive view of the genetic and demographic histories of South American admixed populations.

KEYWORDS
Indels, linkage disequilibrium, Mestizo, STRs, X chromosome

1 | INTRODUCTION

The current territory of Ecuador was once home to very diverse Native American groups that were subjugated by the Inca Empire during the 15th century. These native groups first interacted with Europeans a century later, during the Spanish expeditions along the South American Pacific coast. The period of European conquest and colonization in Ecuador started from the coast and was concentrated in the Andes region. Amazonia, however, was not directly reached by the Spanish (Ayala, 2008). The period of European colonization was also characterized by the arrival of

Leonor Gusmão and German Burgos should be considered joint senior author.

 © 2021 Wiley Periodicals LLC.

African slaves entering America through Caribbean ports. This contingent of slave labor reached a final destination of the Pacific coast and the warm valleys of the Andes (Ayala, 2002).

Ecuador is currently an ethnically and demographically diverse country. According to the latest national census (INEC, 2010), Ecuadorians self-identify as Mestizos (72.3%), Afro-descendants (7.2%), Montubios (7.4%), Native Americans (7%), or European descendants (6.1%). Both Mestizos and Montubios harbor Native American, European and African ancestry. However, the admixed population from the Pacific coast identifies itself as Montubio instead of Mestizo. In geographic terms, the country is divided into four main regions: the Pacific coast, Andes, Amazonia, and Galápagos. The Mestizos/Montubios represent the largest fraction of the population and are distributed across all regions. Most Native American populations are dispersed throughout the Andean region. However, there are also some native groups living in Amazonia that are characterized by strong geographic and sociocultural isolation. Afro-Ecuadorian populations are mainly found in the northern portion of the country, namely in Esmeraldas (Pacific coast region) and the Valle del Chota (Andes region). It is known that Esmeraldas was the final destination of fugitive slaves who sought freedom and took refuge in this region. There are also several records describing fully loaded slave ships sinking off the coast of Esmeraldas. The survivors entered the mainland and lived free, coexisting with the native population (Ayala, 2002; Salazar & Catherine, 2020).

To capture the genetic diversity present in current populations of Ecuador, several studies have been developed focusing on different population groups and different regions of the genome (Baeta et al., 2009, 2012; Baeza et al., 2007; Burgos, Garzón-Salazar, et al., 2019; Burgos, Simão, et al., 2019; Cardoso et al., 2012; González-Andrade et al., 2006; González-Andrade et al., 2008; González-Andrade et al., 2009; Santangelo et al., 2017; Toscanini et al., 2018; Zambrano et al., 2017, 2019). Studies using autosomal ancestry informative markers (AIMs) showed that admixed populations in Ecuador (Mestizos) preserve a significant Native American ancestry, followed by European and African contributions (Poulsen et al., 2011; Santangelo et al., 2017; Zambrano et al., 2017, 2019). As expected, the Native American genetic background is higher in communities that preserved their Native American identity, and lower in Afro-descendent groups, which harbor a higher African ancestry (Santangelo et al., 2017; Zambrano et al., 2019). The same trend was found for the Y chromosome variation. Native American ancestry is higher in Amazonia, followed by the Andes and has the lowest value on the Pacific coast (Toscanini et al., 2018). However, paternal European inheritance is predominant in all Mestizo groups (Burgos, Garzón-Salazar, et al., 2019; Burgos, Simão, et al., 2019; Toscanini et al., 2018). In contrast, the mtDNA lineages are mainly of Native American ancestry (Baeta et al., 2012; Burgos, Garzón-Salazar, et al., 2019; Burgos, Simão, et al., 2019).

The X-chromosomal markers combine the interesting genetic features of both autosomal and uniparental markers. Unlike mtDNA and the Y chromosome, due to recombination, the X chromosome presents diverse informative regions that reveal different histories. The presence of a single copy of this chromosome in males that is only transmitted to the daughters reduces the effective population size in relation to the autosomes. Therefore, we expect linkage disequilibrium (LD) to be greater and the population structure more pronounced for X-chromosomal than for autosomal markers (Schaffner, 2004). LD patterns are largely affected by population history (Tishkoff et al., 1996). Consequently, the analysis of genetic markers located on the X chromosome can be useful to investigate admixture, substructure, isolation, selection, and drift events in human populations (Pereira et al., 2011, 2015).

Few studies have been performed on Ecuadorian populations using X-chromosomal markers. The published data sets include 10 or 12 X-chromosomal STRs (X-STRs) selected for forensic applications. Based on these markers, no statistically significant stratification was found for the Mestizo population (Baeta et al., 2013; Cardoso et al., 2012; Gaviria et al., 2013; Pinto et al., 2020; Zambrano et al., 2015). However, due to a high mutation rate, X-STRs are highly diverse, and large databases are mandatory to detect LD between markers (Kling et al., 2015). An excess of homozygosity and significant LD among X-STRs located in distant positions on the chromosome were reported by Pinto et al. (2020), suggesting the presence of population substructure in the Ecuadorian Mestizo sample analyzed. Nevertheless, the authors conclude that a larger sample will be necessary to investigate population stratification.

To assess levels of genetic substructure within the Ecuadorian Mestizo population, in the present study we enlarged the sample group previously studied for 12 X-STRs (Pinto et al., 2020). To increase the possibility of detecting low levels of LD, we have also studied 32 insertion/deletion polymorphisms along the X chromosome (X-Indels), which have lower mutation rates than STRs.

2 | MATERIALS AND METHODS

2.1 | Sample collection and extraction

In this study, a total of 795 haplotypes from Ecuadorian Mestizos (377 male and 418 female haplotypes) were analyzed. A subset of 149 female samples were included in a previous study on X-STR mutation rates (Pinto et al., 2020). Samples from father/daughter duos were selected from paternity cases. Individuals from different paternity investigations, not sharing first-generation ancestors, were assumed to be unrelated. The sample set was divided into three groups/regions according to the birthplace of the individuals, namely, the Pacific coast, Andes, and Amazonia.

Blood samples and buccal swabs were collected under written informed consent. The project was approved by the Fiscalía General del Estado and by CEISH-UDLA 2017-0301 (Ethical Committee in Human Research from the Universidad de las Américas).

Genomic DNA was extracted using 10% Chelex 100 resin (Bio-Rad Laboratories, CA, USA) or a standard phenol-chloroform method.

2.2 | Genotyping of X-chromosomal markers

DNA samples were genotyped for two marker sets. The 12 X-STRs included in the Investigator® Argus X-12 Kit (Qiagen, Hilden, Germany) were genotyped following the manufacturer's

recommendations. This kit can be particularly useful for evaluating LD, since it comprises 12 X-STRs organized into four linkage groups. A set of 32 X-chromosomal insertion–deletion polymorphisms (X-Indels) were amplified in a single multiplex PCR, followed by fragment size separation by capillary electrophoresis. Primer sequences and PCR conditions were as previously described by Pereira, Pereira, et al. (2012) and Pereira, Pillips, et al. (2012). For both marker sets, separation and detection of the PCR products were carried out in an ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, California, USA). Genotypes were determined using the GeneMapper software V3.2 (Applied Biosystems, Foster City, California, USA).

2.3 | Statistical analysis

The number of samples used in the data analysis varied depending on the test performed, and only unrelated haplotypes were selected in each analysis (Table 1).

Haplotype frequencies, gene diversity, and values for the R^2 statistic were estimated using the Arlequin software v.3.5.2.2 (Excoffier & Lischer, 2010). The same software was used for Hardy–Weinberg equilibrium (HWE) and pairwise LD analyses. To estimate R^2 in the STR loci, we converted the allele frequencies to binary variables, where the most common allele was coded as 0, and all alternative alleles were pooled as 1 (Willems et al., 2014). Pairwise genetic distances (F_{ST}) and analysis of

TABLE 1 Total number and characteristics of the samples/haplotypes included in each analyses performed during this study

	Total	Males typed for both X-STRs and X-Indels [this study]	Males typed for X-Indels [this study]	Males typed for X-STRs [this study]	Females typed for X-STRs [Pinto et al., 2020]	Females typed for X-STRs [this study]	Females typed for X-STRs [this study]
Number of samples[a]	586	318	54	5	149	37	23
Number of haplotypes	795	318	54	5	298	74	46
Gametic phase[b]	–	–	–	–	Known	Known	Unknown
Hardy–Weinberg equilibrium test (n° of samples)	209	–	–	–	149	37	23
X-STRs: LG-haplotype frequencies, LD analyses, and pairwise comparison with other populations (n° of haplotypes)	558	181	–	5	298	74	–
X-STRs: Pairwise comparison with other populations (n° of haplotypes)[c]	527	304	–	5	207	11	–
X-Indels: LD analyses (n° of samples/haplotypes)	372	318	54	–	–	–	–
X-Indels: PCA and Pairwise comparison with other populations (n° of samples/haplotypes)[c]	355	304	51	–	–	–	–
Full set: LD analyses (n° of samples/haplotypes)	318	318	–	–	–	–	–
Full set: Pairwise FSTs and AMOVA among Ecuadorian regions (n° of samples/haplotypes)[c]	304	304	–	–	–	–	–

[a]A total of 137 males are related to females included in the sample set of "Female samples typed for X-STRs, with known gametic phase", as father/daughter duos.
[b]Gametic phase was determined by genotyping the father.
[c]Samples without information about the donor's geographic origin were excluded from these analyzes.

molecular variance (AMOVA), as well as the probability of non-differentiation (*p*-values) were also assessed using Arlequin software. For multiple tests, the significance level of 0.05 was adjusted by applying Bonferroni's correction. The pairwise F_{ST} matrix was represented in a multidimensional scaling (MDS) plot using the software STATISTICA 13.2 Trial (StatSoft Inc., http://www.statsoft.com.br). Principal component analysis (PCA) was performed using XLSTAT version 2020.5.1 (https://www.xlstat.com).

3 | RESULTS AND DISCUSSION

The full female and male profiles of the Ecuadorian samples are reported in Tables S1 and S2, respectively. Haplotype frequencies for the 12 X-STRs in a sample of unrelated individuals are reported in Table S3.

3.1 | Hardy–Weinberg equilibrium analysis

No statistically significant deviations from the HWE expectations were detected in the Ecuadorian population for the 12 X-STRs (Table S4). Nevertheless, all but one X-STR (DXS10146) showed an excess of homozygotes compared with the expected HWE (Table S4). The average value of observed heterozygotes in the sample (0.77) was lower than that expected for a population in HWE (0.80).

Although an excess of homozygotes is suggestive of population substructure or some degree of inbreeding, deviations were not statistically significant. The same result was previously obtained in a smaller sample set of 150 females (Pinto et al., 2020). To confirm whether the population is indeed in HWE, a larger sample size is still needed.

3.2 | Linkage disequilibrium analysis among X-STRs

The 12 X-STRs included in the Argus kit are organized into four linkage groups (LGs), each comprising three X-STRs. Markers on the same LG have low recombination rates and are located close together (GD ≤ 0.46 cM). Therefore, they are more susceptible to show LD even several generations after the occurrence of the events that generated it. Analysis of pairwise LD for the 12 X-STRs was performed for a total sample of 558 haplotypes (Table 1). The results revealed statistically significant pairwise associations between the three loci on LG 3 and between two loci on LG 1 and LG 4. No significant values were obtained for loci on LG 2 (Table S5). Significant LD was also observed between distant loci: HPRTB and DXS10148 (129.750 cM), HPRTB and DXS10135 (129.600 cM), and DXS10074 and DXS10101 (58.410 cM).

The detection of deviations from linkage equilibrium can be difficult in samples with a low effective number, especially when a high number of haplotype combinations is present in the population. For the highly variable X-STRs that we analyzed, Kling et al. (2015) showed that large sample sizes are necessary to detect LD based on exact tests. These authors performed simulations using samples from 100 to 600 haplotypes from a Swedish population displaying LD for markers on the same LG. In the cited study, statistically significant LD was found between markers on LG 3 for all iterations of $N = 200$, which was only possible for $N = 600$ on LG 1 and LG 4. On LG 2, for $N = 600$, significant LD was only obtained for 83% of the interactions. Although levels of LD are population specific, it is interesting to note that in both Swedish and Ecuadorian populations, LG 3 showed the lowest *p*-values (no association between loci), followed by LG 1 and LG 4, and the highest values were obtained for LG 2 (Table S5). Previous studies also showed significant LD between markers on LG 3 in American, European, African, and East Asian populations (Ferragut et al., 2017; Gomes et al., 2017; Kling et al., 2015; Pinto et al., 2015; Pinto et al., 2020; Tomas et al., 2012).

In Table 2, we calculated the diversity parameters, as well as the genetic distances separating the most distant markers in each linkage group. LG 3 showed the smallest genetic distances between loci when compared to the other linkage groups. However, genetic distance alone does not explain the results obtained, and a higher mutation rate and haplotype diversity may lead to a lower sensitivity of small samples to detect LD. In fact, we can see that LG 3 presents the lowest mutation rate and haplotype diversity in the Ecuadorian Mestizo population. Supporting the need for larger samples to detect LD, LG 1 showed the highest haplotype diversity, LG 2 had the highest mutation rate and average single locus diversity, and LG 4 had the largest genetic distance between loci.

3.3 | Linkage disequilibrium analysis among the full set of loci

To increase the possibility of detecting low levels of LD, a new pairwise LD analysis was performed considering the 12 X-STRs and

TABLE 2 Average single locus mutation rate and genetic diversity, number of different haplotypes and haplotype diversities, and genetic distance between the most distant markers in each linkage group

	Average mutation rate[a]	Average single locus diversity[b]	Number of haplotypes[b]	Haplotype diversity[b]	Genetic distance (cM)
LG 1	5.29E-03	0.793	178	0.995	0.37
LG 2	5.79E-03	0.821	158	0.993	0.38
LG 3	2.17E-03	0.766	102	0.971	0.36
LG 4	2.89E-03	0.804	158	0.994	0.46

[a]Mutation rates were retrieved from Pinto et al. (2020);
[b]These values were calculated in the sample from Ecuador (present study).

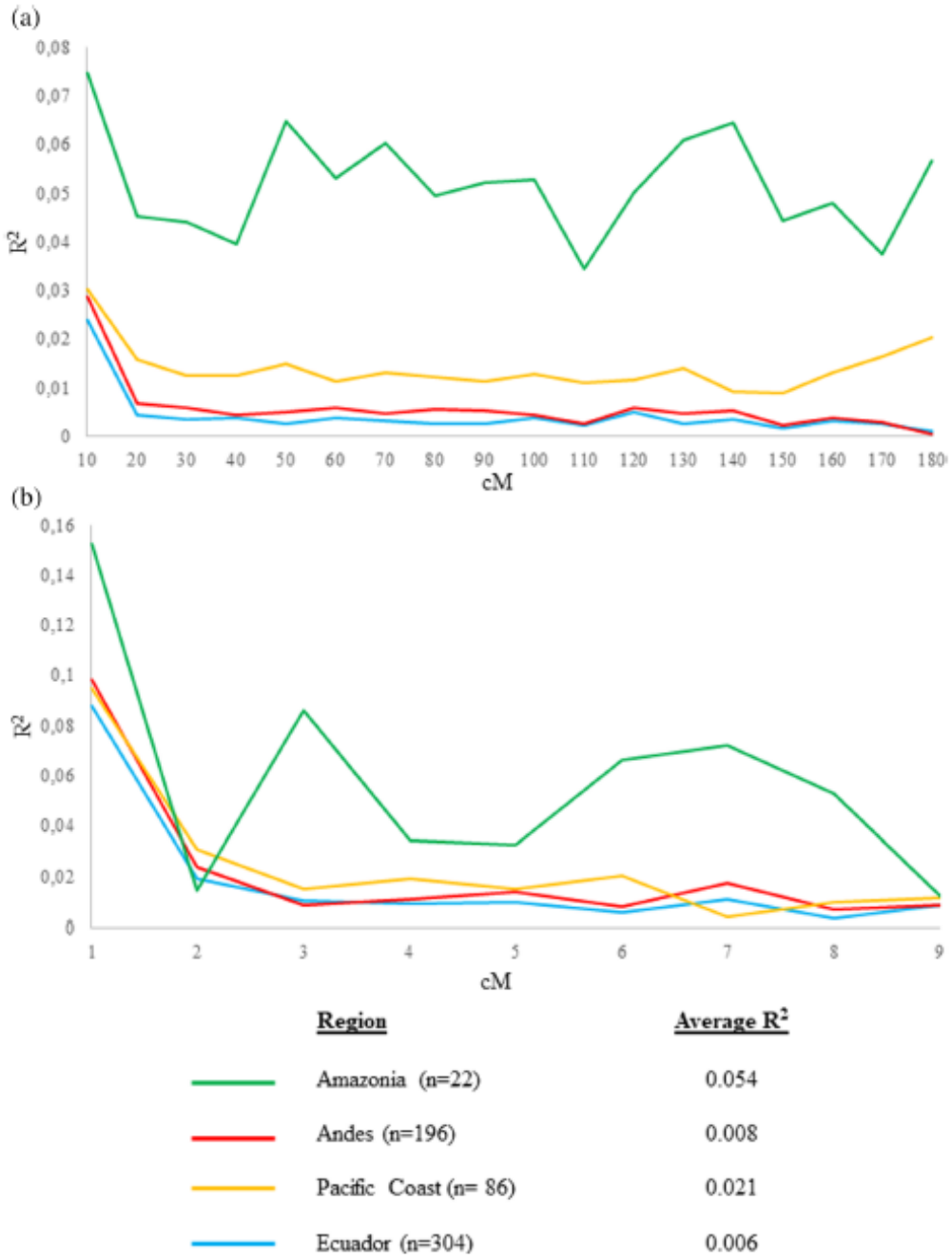


FIGURE 1 Average R^2 values in the total sample from Ecuador and in the three Ecuadorian regions. (a) Average R^2 values (Y-axis) calculated for marker pairs grouped in 10 cM intervals (X-axis). (b) Average R^2 values (Y-axis) calculated for marker pairs grouped in 1 cM intervals (X-axis) for distances below 9 cM

the 32 X-Indels together, using a subset of 318 male samples (those genotyped for all loci). Apart from the associations already discussed for X-STRs on the LGs, significant values for LD were found between pairs of X-Indels at distances of 0.204 cM and smaller (Table S6). For these marker pairs, significant *p*-values for LD were previously observed in other populations from different continents (Caputo et al., 2017; Edelmann et al., 2016; Ferragut et al., 2017; Freitas et al., 2010; Ibarra et al., 2014; Martinez et al., 2019; Pereira, Pereira,

TABLE 3 Average values of expected heterozygosity calculated for the full set of X-markers, the 12 X-STRs, and the 32 X-Indels, in the three Ecuadorian regions

	Amazonia	Andes	Pacific coast
	32 X-Indels ($N = 355$)		
N	27	222	106
Av Het.	0.331	0.346	0.347
SE	0.184	0.136	0.142
	12 X-STRs ($N = 526$)		
N	36	344	147
Av Het.	0.792	0.793	0.811
SE	0.095	0.096	0.089
	Full X-set ($N = 304$)		
N	22	196	86
Av Het.	0.451	0.468	0.473
SE	0.265	0.240	0.242

et al., 2012; Pereira, Pillips, et al., 2012; Pinto et al., 2015; see Table S6). A significant association was also observed between indels MID3754 and MID3722, which were separated by 1.433 cM and not detected in previously studied populations. Furthermore, X-STRs from LG 3 presented a significant association with the contiguous indels MID1839 and MID3736 (Table S6), showing that this LG possibly comprises six loci extending to a block of 4.396 cM.

In total, 20 out of the 946 pairwise comparisons revealed statistically significant LD. In 13 cases (described above), the markers were separated by less than 2.8 cM. Since the Mestizo population resulted from recent admixture, these results can be explained by the low recombination between loci, preserving LD in the population for many generations. However, significant associations were also observed in seven pairwise comparisons involving markers from different linkage groups, which were separated by distances varying between 10.891 and 163.53 cM (Table S6). Along with the results from the HWE test (excess homozygosity), the significant associations observed between markers from different linkage groups point to the presence of substructure inside the Ecuadorian Mestizo population.

To determine if the levels of LD observed in the general Mestizo population of Ecuador could be due to differences among regions, we repeated the analysis considering three subsamples, divided according to the geographic and demographic history of the country (as described in Section 1): Amazonia, Andes, and the Pacific coast. The LD pattern in the Andes region was very similar to that observed in the general population. In contrast, the other regions showed less statistically significant p-values, especially Amazonia. Nonetheless, *p*-values depend strongly on sample size and cannot be used to compare LD between samples with very different sizes (Kling et al., 2015). For this reason, we calculated R^2 as a standardization of the gametic determinant *D*. The average R^2 value was lower in the Andes than in the Pacific coast and Amazonia (Figure 1).

When calculating the average R^2 values for markers grouped by genetic distances, we can see that R^2 decreases with the distance between pairs of markers (Figure 1). There is a large decrease when the distance increases from 1 to 2 cM and a slight decrease from 2 to 3 cM; above this distance, the values become stable, except for Amazonia, which shows background LD between markers independent of the genetic distance (Figure 1(b)). From what is known about Amazonia, the synchronous action of recent admixture and genetic drift accentuated by recent bottlenecks (Barbieri et al., 2019) probably worked toward creating the observed level of background LD in this region.

3.4 | Pairwise genetic distances and AMOVA

To evaluate the genetic structure of the Ecuadorian Mestizo population, AMOVA, and pairwise F_{ST} analyses were performed considering the three studied regions, and the full set of X-chromosomal loci. AMOVA showed that 99.29% of the variation is within populations ($F_{ST} = 0.0071$; $p < 0.00000$). The average values of genetic diversity among all loci showed a slightly lower expected heterozygosity in the Amazonian sample (Table 3), which may be due to a small sample size. To increase the sample sizes, genetic diversities were estimated again for each marker set independently (Table 3). The AMOVA results for the X-STRs and the X-Indels, separately, again showed low F_{ST} values (0.0038 and 0.0091, respectively).

Low F_{ST} values were also obtained in pairwise comparisons between the three Ecuadorian regions (Table S7). A nondifferentiation test showed a nonstatistically significant F_{ST} between the Amazonian and Andes regions. These results may be related to the fact that the Andes and Amazonia have been largely inhabited by Native American people with the same dialect (Quichua), but these results may also be due to a similar proportion of Native American ancestry. The nondifferentiation test performed between the Pacific coast and the other two Ecuadorian regions showed statistically significant *p*-values (Table S7). Therefore, in subsequent analyses, we divided the data by region.

3.5 | Comparison with other populations

Genetic distances were assessed separately for the two types of markers since no other populations were studied with the full set of 44 X-chromosomal loci (Tables S8 and S9). In the MDS plots of the pairwise F_{ST} values (Figure 2), the two sets of X-markers can clearly separate European, African, and Native American populations (Figure 2). These three reference groups are distantly positioned, with the South American admixed populations located between them.

Significant differences in the studied markers were detected between the Ecuadorian and most admixed populations from Brazil ($0.069 < F_{ST} < 0.085$ for X-Indels, $0.013 < F_{ST} < 0.032$ for X-STRs), Colombia ($0.020 < F_{ST} < 0.074$ for X-Indels), and Argentina ($0.005 < F_{ST} < 0.021$ for X-STRs). However, nonstatistically significant F_{ST}s were found in comparisons with populations from the southwest Andes, Mexico, Peru, and Rio Negro (Tables S8 and S9). This is in accordance with the relatively high Native American ancestry

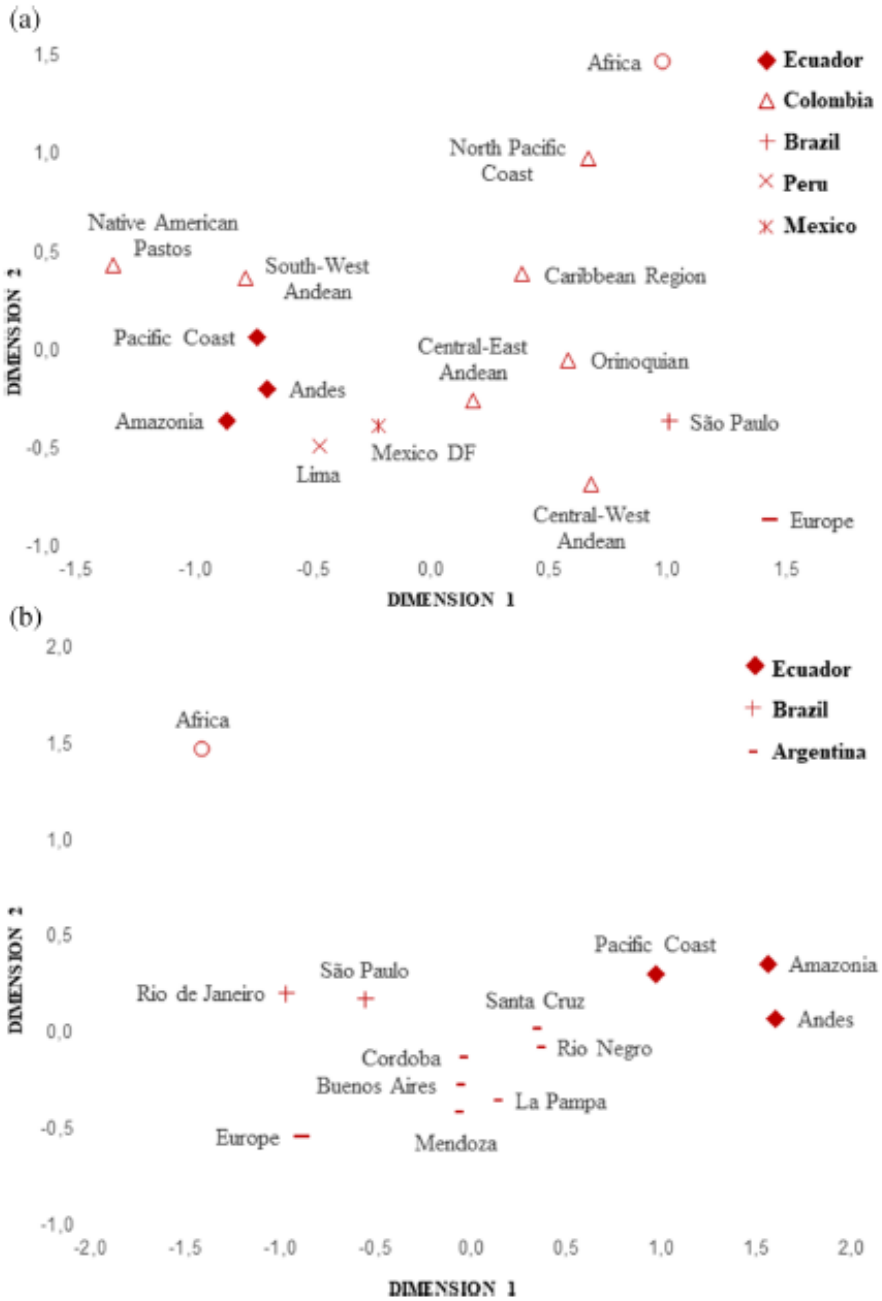


FIGURE 2 Multidimensional scaling analysis (MDS) based on F_{ST} genetic distances calculated for (a) 32 X-Indels (Stress = 0.0678) and (b) 12 X-STRs (Stress = 0.0308). (a) Data used were retrieved from The 1000 Genomes Project Consortium (2015) (Africa, Colombia, Peru, and Mexico), Ibarra et al. (2014) (Colombia), and Martinez et al. (2019) (Brazil). Only a subset of 23 X-Indels was used in this analysis, since no data are available for the remaining 9 X-Indels in the samples from the 1000 Genomes Project. (b) Data used were retrieved from Pinto et al. (2020) (Europe, Argentina, and Brazil), Gomes et al. (2017) (Africa)

reported for these populations (Corach et al., 2010; Norris et al., 2018; Ossa et al., 2016). Together with the southwest Andean region of Colombia, Ecuador has a higher Native American ancestry than other South American countries, including Peru and Mexico.

For the X-Indels (Figure 2(a)), the Ecuadorian Pacific coast presented lower genetic distance with the Pastos (Native Americans) than the Amazonia and Andes populations. The Pasto ethnic group settled in the territories previously known as the "*Franja Andina*," which extends from Nariño (an administrative department in southern Colombia) to the province of Imbabura (Andes region, in northern Ecuador) (Uribe, 1986). Currently, the Pastos are only in southern Colombia. However, there is an ethnic group called Awá that shares the same linguistic group as the Pastos. The Awás also settled in the "*Franja Andina*" but expanded their territories to the north of the Pacific coast of Ecuador. These results suggest that the lower distance between the Pacific coast and Pastos populations may be due to interactions between ancestral ethnic groups (Figure 2(a)). Nevertheless, some populations in the MDS are displaced in a different

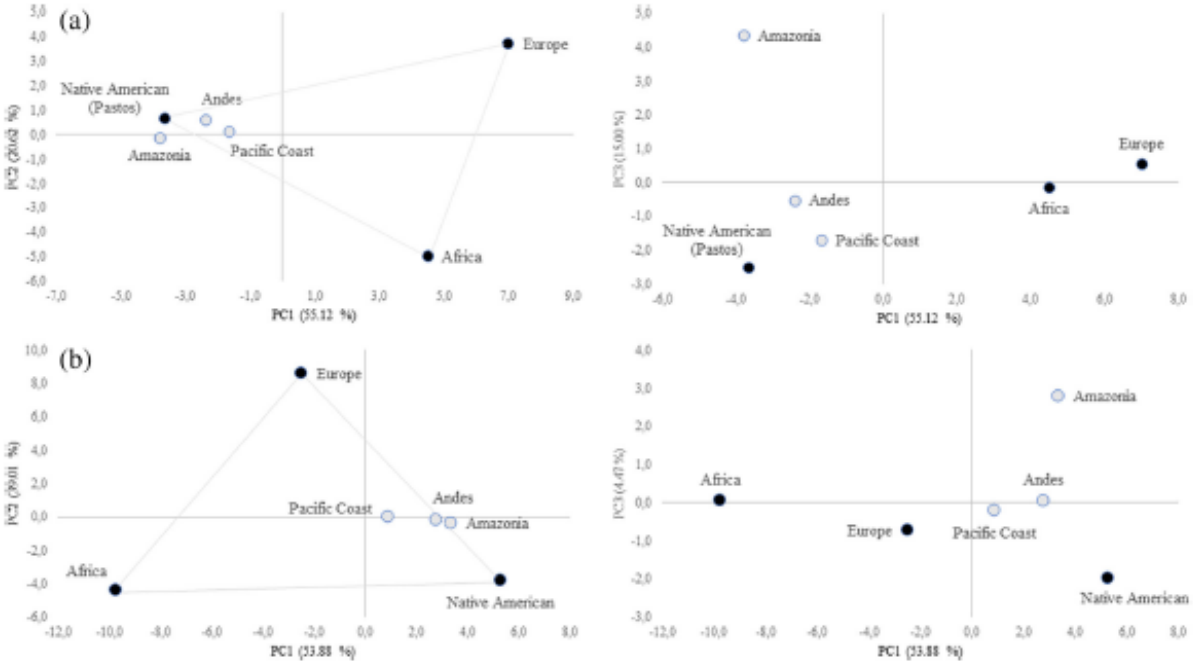


FIGURE 3 Principal component analysis between (a) Ecuadorian (this study) and Native American (Ibarra et al., 2014), African, and European (The 1000 Genomes Project Consortium, 2015) reference populations measured by X chromosome markers; (b) Ecuadorian (Zambrano et al., 2019) and Native American, African, and European reference populations (Pereira, Pereira, et al., 2012; Pereira, Pillips, et al., 2012) measured by autosomal ancestry informative markers (AIMs)

direction from that of the Pastos, suggesting different Native American ancestries among the analyzed populations.

3.6 | Population differentiation for X-chromosomal versus autosomal indels

Since women contribute two-thirds of the X chromosomes in a population (assuming an equal proportion of male and female descendants), differences in X-chromosomal and autosomal ancestries will reflect sex-biased matings. Therefore, the separation of the three Ecuadorian samples from European, African (The 1000 Genomes Project Consortium, 2015) and Native American populations was investigated by PCA using both X-Indels and 46 ancestry informative indels (distributed over 18 of the 22 autosomes) previously reported in the same population groups (Pereira, Pereira, et al., 2012; Pereira, Pillips, et al., 2012; Zambrano et al., 2019).

The results of the two first principal components revealed a closer proximity of the Ecuadorian populations to Native Americans for X-chromosomal (Figure 3(a)) than for autosomal indels (Figure 3 (b)). Although a 26% European contribution was reported by (Zambrano et al., 2019) for autosomal markers, no signs of European admixture were detected in the Amazonian population for the X-chromosomal indels. These results indicated a lower maternal than paternal European contribution, which is consistent with published data from nonrecombining genomes (Baeta et al., 2012; Burgos, Garzón-Salazar, et al., 2019; Burgos, Simão, et al., 2019; Toscanini et al., 2018). Previous studies comparing mtDNA and Y chromosome data, apart from differences among Ecuadorian Mestizo populations, also revealed clear sex-biased mating processes, involving predominantly descendants from European men and Native American women. Regarding maternal inheritance, the available studies show that most Ecuadorian Mestizo populations present a set of mtDNA lineages that are mainly of Native American ancestry (above 96%) (Baeta et al., 2012; Burgos, Garzón-Salazar, et al., 2019; Burgos, Simão, et al., 2019). In contrast, Y-chromosomal lineages are primarily European (above 55%), although populations still retain a significant part of their ancient paternal ancestry (between 26% in the Pacific coast and 44% in Amazonia) (Burgos, Garzón-Salazar, et al., 2019, Burgos, Simão, et al., 2019, Toscanini et al., 2018).

Concerning the African contribution detected using autosomal markers in the population from the Pacific coast (16% according to Zambrano et al., 2019), for the X-chromosomal markers, there was an increase in the distance to the African population as well as closer proximity to the axis that links Native American and African populations (Figure 3). These results can be explained by sex-biased mating not only between European men and African women but also between African men and Native American women, which was not previously reported for this population.

When the third principal component is investigated, we can see a clear separation between Amazonia and the other Ecuadorian regions, in both X-chromosomal and autosomal analyses (Figure 3(a)). The

X-Indels shows a Native American ancestry in Amazonia that is different from that of the Andes, Pacific coast, and Native American (Pastos) populations. The autosomal markers show smaller differences in the Native American ancestry of these populations, which is consistent with a higher male than female mobility between Amazonia and the other regions.

Given the discrepancy in sample size between the three studied regions, a new PCA was performed after subdividing the total sample from the Andes into six subgroups of 37 individuals and the sample from the Pacific coast into three subsamples of 35, 35 and 36 individuals each. The results obtained are consistent with those obtained when using the full-size samples (Figure S1). The subsamples from the Andes cluster near the Native American group, and the three samples from the Pacific coast cluster next to the axis between the Native American and African samples.

4 | CONCLUSIONS

In summary, our results indicate that the Ecuadorian Mestizo population is not a homogeneous group, probably reflecting differences in Native American ancestry and in levels of continental admixture in the three regions analyzed. The most distinctive region is Amazonia. This region showed high levels of LD, which could be due to its prolonged geographic isolation, recent admixture events, the effect of genetic drift on a small effective population size, and population bottlenecks. It is important to point out that despite the division of the analyzed sample by region, there are still signs of population substructure. Therefore, to obtain a more comprehensive view of the genetic and demographic history of Ecuador, it is necessary collect a broader and more representative sample of these regions.

This study also highlights the importance of an integrated analysis of different types of markers for a more comprehensive view of the genetic and demographic history of the admixed populations from South America. The comparison of autosomal and X-chromosomal information allowed us to discern sex-biased admixture and migration events among Ecuadorian populations, which was not previously detected through the analysis of uniparental genomes.

ACKNOWLEDGMENTS

We thank the support of Dirección Nacional de Investigaciones de la Fiscalía General del Estado (Quito-Ecuador) in authorizing access to its collection of samples. We would also like to thank Alejandra Garzón, Eduardo Tejera, and Victor Calderón for their support in the processing and analysis of the sample data; to Iva Gomes for kindly provide the full X-STR genotypes of the Africa samples included in Gomes et al. (2017); to Alberto Fernandez from Labgenetics for kindly provide information on the country region origin of their Ecuadorian samples included in the article from Pinto et al. (2020). RF and LG were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq (ref. 131709/2020-8 and ref. 306342/2019-7), Brazil. LG was supported by Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro—FAPERJ (CNE-2018), Brazil. GB was supported by MED.GB.18.06 grant from Dirección General de Investigación, Universidad de Las Américas, Quito, Ecuador.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

Rodrigo Flores-Espinoza: Conceptualization; data curation; formal analysis; investigation; methodology; software; supervision; validation; writing-original draft; writing-review & editing. **Ellus Paz-Cruz:** Resources; writing-review & editing. **Viviana Ruiz-Pozo:** Resources; writing-review & editing. **Marcelo Lopez-Carrera:** Resources; writing-review & editing. **Alejandro Cabrera-Andrade:** Resources; writing-review & editing. **Leonor Gusmao:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; software; supervision; validation; writing-review & editing. **German Burgos:** Conceptualization; data curation; funding acquisition; investigation; methodology; project administration; resources; supervision; validation; writing-review & editing.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Rodrigo Flores-Espinoza https://orcid.org/0000-0002-7861-0808
German Burgos https://orcid.org/0000-0002-8400-2753

REFERENCES

Ayala, E. (2002). *Ecuador: Patria de todos. La nación ecuatoriana, unidad en la diversidad*. Quito.

Ayala, E. (2008). *Resumen de Historia del Ecuador* (3rd ed.). Corporación Editora Nacional, Ed.

Baeta, M., Núñez, C., Aznar, J. M., Sosa, C., Casalod, Y., Bolea, M., González-Andrade, F., de Pancorbo, M. M., & Martínez-Jarreta, B. (2013). Analysis of 10 X-STRs in three population groups from Ecuador. *Forensic Science International: Genetics*, *7*(1), 19–20. https://doi.org/10.1016/j.fsigen.2012.08.004

Baeta, M., Núñez, C., González-Andrade, F., Sosa, C., Casalod, Y., Bolea, M., Zweynert, S., Cruz, O. W. V., González-Solorzano, J., Geppert, M., Roewer, L., & Martínez-Jarreta, B. (2009). Mitochondrial analysis revealed high homogeneity in the Waorani population-the last nomadic group of hunter-gatherers from Ecuador. *Forensic Science International: Genetics Supplement Series*, *2*(1), 313–314. https://doi.org/10.1016/j.fsigss.2009.08.025

Baeta, M., Núñez, C., Sosa, C., Bolea, M., Casalod, Y., Gonzáles-Andrade, F., Roewer, L., & Martínez-Jarreta, B. (2012). Mitochondrial diversity in Amerindian Kichwa and Mestizo populations from Ecuador. *International Journal of Legal Medicine*, *126*, 299–302. https://doi.org/10.1007/s00414-011-0656-4

Baeza, C., Guzmán, R., Tirado, M., López-Parra, A. M., Rodríguez, T., Mesa, M. S., Fernández, E., & Arroyo-Pardo, E. (2007). Population data for 15 Y-chromosome STRs in a population sample from Quito (Ecuador). *Forensic Science International*, *173*, 214–219. https://doi.org/10.1016/j.forsciint.2006.09.011

Barbieri, C., Barquera, R., Arias, L., Sandoval, J. R., Acosta, O., Zurita, C., Aguilar-Campos, A., Tito-Álvarez, A. M., Serrano-Osuna, R., Gray, R. D., Mafessoni, F., Heggarty, P., Shimizu, K. K., Fujita, R., Stoneking, M., Pugach, I., & Fehren-Schmitz, L. (2019). The current genomic landscape of Western South America: Andes, Amazonia, and Pacific Coast. *Molecular Biology and Evolution*, *36*(12), 2698–2713. https://doi.org/10.1093/molbev/msz174

Burgos, G., Garzón-Salazar, A., Maldonado-Uquillas, K., Ávila, C., Toscano, M., & Paz-Cruz, E. (2019). A look of paternal ancestry in a sample of Ecuadorian "MESTIZO" population analyzed through PowerPlex Y23. *Forensic Science International: Genetics Supplement Series*, *7*(1), 534–536. https://doi.org/10.1016/j.fsigss.2019.10.080

Burgos, G., Simão, F., Flores-Espinoza, R., Yepez-Santos, J. L, Garzón-Salazar, A., Paz-Cruz, E., Freire-Paspuel, B., Carvalho, E. F., & Gusmão, L. (2019). An approach to maternal ancestry in a sample of Ecuadorian "mestizo" population by sequencing the control region of mtDNA. *Forensic Science International: Genetics Supplement Series*, *7*(1), 537–538. https://doi.org/10.1016/j.fsigss.2019.10.081

Caputo, M., Amador, M. A., Santos, S., & Corach, D. (2017). Potential forensic use of a 33 X-InDel panel in the Argentinean population. *International Journal of Legal Medicine*, *131*(1), 107–112. https://doi.org/10.1007/s00414-016-1399-z

Cardoso, S., Alfonso-Sánchez, M., Valverde, L., Sánchez, D., Zarrabeitia, M. T., Odriozola, A., Martínez-Jarreta, B., & De Pancorbo, M. M. (2012). Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon. *Heredity*, *108*(6), 609–615. https://doi.org/10.5061/dryad.f6hs62hq

Corach, D., Lao, O., Bobillo, C., Van Der Gaag, K., Zuniga, S., Vermeulen, M., Van Duijn, K., Goedbloed, M., Vallone, P. M., Parson, W., De Knijff, P., & Kayser, M. (2010). Inferring continental ancestry of Argentineans from autosomal, Y-chromosomal and mitochondrial DNA. *Annals of Human Genetics*, *74*, 65–76. https://doi.org/10.1111/j.1469-1809.2009.00556.x

Edelmann, J., Kohl, M., Dressler, J., & Hoffmann, A. (2016). X-chromosomal 21-indel marker panel in German and Baltic populations. *International Journal of Legal Medicine*, *130*(2), 357–360. https://doi.org/10.1007/s00414-015-1221-3

Excoffier, L. & Lischer, H. E. L. (2010). Arlequin 3.5.2 manual. Molecular ecology resources (Vol. 10). Switzerland. Retrieved from http://cmpg.unibe.ch/software/arlequin35

Ferragut, J. F., Bentayebi, K., Pereira, R., Castro, J. A., Amorim, A., Ramon, C., & Picornell, A. (2017). Genetic portrait of Jewish populations based on three sets of X-chromosome markers: Indels, Alu insertions and STRs. *Forensic Science International: Genetics*, *31*, 5–11. https://doi.org/10.1016/j.fsigen.2017.09.008

Freitas, S. C. N., Resque, R. L., Ribeiro-Rodrigues, E. M., Guerreiro, J. F., Santos, N. P. C., Ribeiro-Dos-Santos, Â., & Santos, S. (2010). X-linked insertion/deletion polymorphisms: Forensic applications of a 33-markers panel. *International Journal of Legal Medicine*, *124*(6), 589–593. https://doi.org/10.1007/s00414-010-0441-9

Gaviria, A., Zambrano, A. K., Morejon, G., Aguirre, V., Vela, M., Builes, J. J., & Burgos, G. (2013). Genetic characterization of an X-STR decaplex on the population of Pichincha (Ecuador). *Forensic Science International. Genetics Supplement Series*, *4*(1), 144–145. https://doi.org/10.1016/j.fsigss.2013.10.075

Gomes, I., Pereira, P. J. P., Harms, S., Oliveira, A. M., Schneider, P. M., & Brehm, A. (2017). Genetic characterization of Guinea-Bissau using a 12 X-chromosomal STR system: Inferences from a multiethnic population. *Forensic Science International: Genetics*, *31*, 89–94. https://doi.org/10.1016/j.fsigen.2017.08.016

González-Andrade, F., Roewer, L., Willuweit, S., Sánchez, D., & Martínez-Jarreta, B. (2009). Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis. *Forensic Science International: Genetics*, *3*(3), 83–91. https://doi.org/10.1016/j.fsigen.2008.08.003

González-Andrade, F., Sánchez, D., & Martínez-Jarreta, B. (2006). Announcement of population data genetic analysis of the Amerindian Kichwas and Afroamerican descendents populations from Ecuador characterised by 15 STR-PCR polymorphisms. *Forensic Science International*, *160*, 231–235. https://doi.org/10.1016/j.forsciint.2005.08.002

González-Andrade, F., Sánchez, D., Martínez-Jarreta, B., & Budowle, B. (2008). Y-chromosome STR haplotypes in three different population groups from Ecuador (South America). *Forensic Science International*, *53*(2), 512–514. https://doi.org/10.1111/j.1556-4029.2008.00692.x

Ibarra, A., Restrepo, T., Rojas, W., Castillo, A., Amorim, A., Martínez, B., Burgos, G., Ostos, H., Álvarez, K., Camacho, M., Suarez, Z., Pereira, R., & Gusmão, L. (2014). Evaluating the X chromosome-specific diversity of Colombian populations using insertion/deletion polymorphisms. *PLoS One*, *9*(1), 1–10. https://doi.org/10.1371/journal.pone.0087202

INEC. (2010). Retrieved from https://www.ecuadorencifras.gob.ec/resultados/

Kling, D., Dell, B., & Tillmar, A. (2015). FamLinkX—Implementation of a general model for likelihood computations for X-chromosomal marker data. *Forensic Science International: Genetics*, *17*, 1–7. https://doi.org/10.1016/j.fsigen.2015.02.007

Martínez, J., Polverari, F. S., Silva, F. A. d. J., Braganholi, D. F., Ferraz, J. A. M. L., Gusmão, L., & Cicarelli, R. M. B. (2019). Genetic characterization of 32 X-InDels in a population sample from São Paulo state (Brazil). *International Journal of Legal Medicine*, *133*(5), 1385–1388. https://doi.org/10.1007/s00414-018-01988-w

Norris, E. T., Wang, L., Conley, A. B., Rishishwar, L., Mariño-ramírez, L., Valderrama-aguirre, A., & Jordan, I. K. (2018). Genetic ancestry, admixture and health determinants in Latin America. *BMC Genomics*, *19*(861), 76–99. https://doi.org/10.1186/s12864-018-5195-7

Ossa, H., Aquino, J., Pereira, R., Ibarra, A., Ossa, R. H., Granda, J. D., Lattig, M. C., Groot, H., de Carvalho, E. F., & Gusmão, L. (2016). Outlining the ancestry landscape of Colombian admixed populations. *PLoS One*, *11*(10), 1–15. https://doi.org/10.1371/journal.pone.0164414

Pereira, R., Pereira, V., Gomes, I., Tomas, C., Morling, N., Amorim, A., Prata, M. J., Carracedo, Á., & Gusmão, L. (2012). A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR. *International Journal of Legal Medicine*, *126*, 97–105. https://doi.org/10.1007/s00414-011-0593-2

Pereira, R., Phillips, C., Pinto, N., Santos, C., Emanuel Batista dos Santos, S., Amorim, A., Carracedo, Á., & Gusmão, L. (2012). Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One*, *7*(1), 1–10. https://doi.org/10.1371/journal.pone.0029684

Pereira, V., Moncada, E., Diez, I. E., Tomas, C., Amorim, A., Morling, N., Gusmão, L., & Prata, M. J. (2011). Genetic characterization of Somali and Iraqi populations using a set of 33 X-chromosome Indels. *Forensic Science International: Genetics Supplement Series*, *3*(1), 137–138. https://doi.org/10.1016/j.fsigss.2011.08.069

Pereira, V., Tomas, C., Sanchez, J. J., Syndercombe-Court, D., Amorim, A., Gusmão, L., Prata, M. J., & Morling, N. (2015). The peopling of Greenland: Further insights from the analysis of genetic diversity using autosomal and X-chromosomal markers. *European Journal of Human Genetics*, *23*(2), 245–251. https://doi.org/10.1038/ejhg.2014.90

Pinto, J. C., Pereira, V., Marques, S. L., Amorim, A., Alvarez, L., & Prata, M. J. (2015). Mirandese language and genetic differentiation in Iberia: A study using X chromosome markers. *Annals of Human Biology*, *42*(1), 20–25. https://doi.org/10.3109/03014460.2014.944215

Pinto, N., Pereira, V., Tomas, C., Loiola, S., Carvalho, E. F., Modesti, N., Maxzud, M., Marcucci, V., Cano, H., Cicarelli, R., Januario, B., Bento, A., Brito, P., Burgos, G., Paz-Cruz, E., Díez-Juárez, L., Vannelli, S., de Lurdes Pontes, M., Berardi, G., ... Gusmão, L. (2020). Genetics paternal and maternal mutations in X-STRs : A GHEP-ISFG collaborative study.

Forensic Science International: Genetics, *46*, 102258. https://doi.org/10.1016/j.fsigen.2020.102258

Poulsen, L., Børsting, C., Tomas, C., Gonzáles-Andrade, F., Morling, N., Lopez-Pulles, R., & Gonzáles-Solórzano, J. (2011). Genetics typing of Amerindian Kichwas and Mestizos from Ecuador with the SNP for ID multiplex. *Forensic Science International: Genetics*, *5*, 105–107. https://doi.org/10.1016/j.fsigen.2011.03.006

Salazar, J., & Catherine, W. (Eds.). (2020). *Cimarronaje en el Pacífico Sur*. Abya-Yala.

Santangelo, R., González-Andrade, F., Børsting, C., Torroni, A., Pereira, V., & Morling, N. (2017). Analysis of ancestry informative markers in three main ethnic groups from Ecuador supports a trihybrid origin of Ecuadorians. *Forensic Science International: Genetics*, *31*, 29–33. https://doi.org/10.1016/j.fsigen.2017.08.012

Schaffner, S. F. (2004). The X chromosome in population genetics. *Nature Reviews Genetics*, *5*(1), 43–51. https://doi.org/10.1038/nrg1247

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*, 68–74. https://doi.org/10.1038/nature15393

Tishkoff, S., Dietzsch, E., Speed, W., Pakstis, A. J., Kidd, J. R., Cheung, K., Bonné-Tamir, B., Santachiara-Benerecetti, A. S., Moral, P., Krings, M., Pääbo, S., Watson, E., Risch, N., Jenkins, T., & Kidd, K. K. (1996). Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science*, *271*, 1380–1387.

Tomas, C., Pereira, V., & Morling, N. (2012). Analysis of 12 X-STRs in greenlanders, danes and somalis using argus X-12. *International Journal of Legal Medicine*, *126*(1), 121–128. https://doi.org/10.1007/s00414-011-0609-y

Toscanini, U., Gaviria, A., Moscoso, F., Vela, M., Cobos, S., Lupero, A., Zambrano, A. K., Martinón-Torres, F., Carabajo-Marcillo, A., Yunga-León, R., Ugalde-Noritz, N., Ordoñez-Ugalde, A., & Salas, A. (2018). The geographic mosaic of Ecuadorian Y-chromosome ancestry. *Forensic Science International: Genetics*, *33*, 59–65. https://doi.org/10.1016/j.fsigen.2017.11.011

Uribe, M. V. (1986). Etnohistoria de las comunidades indígenas prehispánicas del sur de Colombia. Anuario Colombiano de Historia Social y de la Cultura

Willems, T., Gymrek, M., Highnam, G., Mittelman, D., & Erlich, Y. (2014). The landscape of human STR variation. *Genome Research*, *24*(11), 1894–1904. https://doi.org/10.1101/gr.177774.114

Zambrano, A., Gaviria, A., Vela, M., Cobos, S., Leone, P. E., Gruezo, C., Fiallo, G., García-Cárdenas, J. M., López-Cortés, A., Cabrera-Andrade, A., & Paz-y-Miño, C. (2017). Ancestry characterization of Ecuador's Highland mestizo population using autosomal AIM-INDELs. *Forensic Science International: Genetics Supplement Series*, *6*, 477–478. https://doi.org/10.1016/j.fsigss.2017.09.191

Zambrano, A. K., Morejón, G., Galarza, J., Pastas, V., Castillo, G., Vela, M., & Gaviria, A. (2015). Genetic data of an X-STR decaplex in Ecuadorian population (P). *Forensic Science International: Genetics Supplement Series*, *5*, 91–92. https://doi.org/10.1016/j.fsigss.2015.09.037

Zambrano, A., Gaviria, A., Cobos-Navarrete, S., Gruezo, C., Rodríguez-Pollit, C., Armendáriz-Castillo, I., García-Cárdenas, J. M., Guerrero, S., López-Cortés, A., Leone, P. E., Pérez-Villa, A., Guevara-Ramírez, P., Yumiceba, V., Fiallos, G., Vela, M., & Paz-y-Miño, C. (2019). The three-hybrid genetic composition of an Ecuadorian population using AIMs-InDels compared with autosomes, mitochondrial DNA and Y chromosome data. *Scientific Reports*, *9*(9247), 1–8. https://doi.org/10.1038/s41598-019-45723-w

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Flores-Espinoza, R., Paz-Cruz, E., Ruiz-Pozo, V. A., Lopez-Carrera, M., Cabrera-Andrade, A., Gusmão, L., & Burgos, G. (2021). Investigating genetic diversity in admixed populations from Ecuador. *American Journal of Physical Anthropology*, 1–11. https://doi.org/10.1002/ajpa.24341