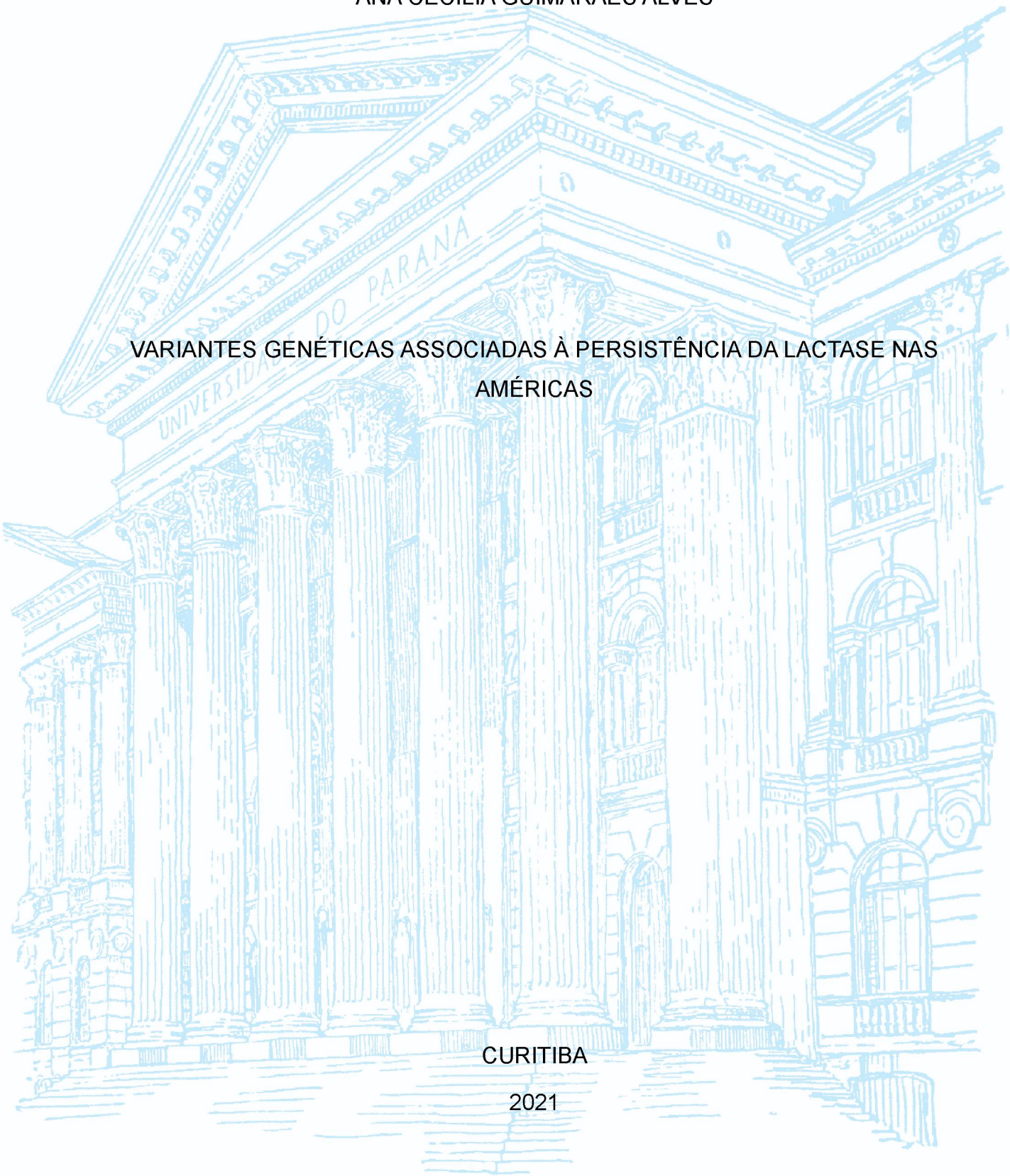
UNIVERSIDADE FEDERAL DO PARANÁ

ANA CECÍLIA GUIMARÃES ALVES



VARIANTES GENÉTICAS ASSOCIADAS À PERSISTÊNCIA DA LACTASE NAS AMÉRICAS

CURITIBA

2021

ANA CECÍLIA GUIMARÃES ALVES

VARIANTES GENÉTICAS ASSOCIADAS À PERSISTÊNCIA DA LACTASE NAS AMÉRICAS

Dissertação apresentada ao Curso de Pós-Graduação em Genética, Setor de Ciências Biológicas, da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Genética.

Orientadora: Profa. Dra. Marcia Holsbach Beltrame
Coorientadora: Profa. Dra. Maria Luiza Petzl-Erler

CURITIBA

2021

Universidade Federal do Paraná
Sistema de Bibliotecas
(Giana Mara Seniski Silva – CRB/9 1406)

Alves, Ana Cecília Guimarães
Variantes genéticas associadas à persistência da lactase nas américas. / Ana Cecília Guimarães Alves. – Curitiba, 2021.
167 p.: il.

Orientadora: Marcia Holsbach Beltrame.
Coorientador: Maria Luiza Petzl-Erler.

Dissertação (mestrado) - Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Lactase. 2. Genética de populações. 3. América Latina. I. Título. II. Beltrame, Márcia Holsbach, 1984-. III. Petzl-Erler, Maria Luiza, 1953-. IV. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

CDD (22. ed.) 576.58

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **ANA CECÍLIA GUIMARÃES ALVES** intitulada: **VARIANTES GENÉTICAS ASSOCIADAS À PERSISTÊNCIA DA LACTASE NAS AMÉRICAS**, sob orientação da Profa. Dra. MARCIA HOLSBACH BELTRAME, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 31 de Março de 2021.

Assinatura Eletrônica
03/04/2021 11:51:59.0
MARCIA HOLSBACH BELTRAME
Presidente da Banca Examinadora

Assinatura Eletrônica
06/04/2021 10:28:50.0
LIANA ALVES DE OLIVEIRA
Avaliador Externo (FACULDADES INTEGRADAS DO BRASIL)

Assinatura Eletrônica
06/04/2021 15:28:11.0
MARCIA REGINA PINCERATI
Avaliador Externo (UNIVERSIDADE POSITIVO)

Assinatura Eletrônica
07/04/2021 11:13:35.0
KELLY NUNES
Avaliador Externo (UNIVERSIDADE DE SãO PAULO - USP)

Dedico este trabalho a todos que, de alguma forma, contribuíram com ele.

Em especial, aos voluntários participantes deste estudo.

AGRADECIMENTOS

Agradeço à minha mãe, Rosana Zumach Guimarães, que me incentivou a seguir o caminho da academia, pelo apoio e amor incondicional de sempre.

Ao meu namorado, João Victor Archegas, pelo companheirismo, por ter compartilhado comigo as alegrias, angústias e frustrações da vida acadêmica e pela compreensão nos momentos de ausência.

À minha orientadora, Marcia Holsbach Beltrame, e à minha coorientadora, Maria Luiza Petzl-Erler, por terem confiado em mim, por compartilharem suas experiências e pelos comentários que moldaram este trabalho. Foi uma honra ser orientada por cientistas como vocês!

À minha estagiária, Natalie Mary Sukow, por ter me acompanhado e auxiliado durante o desenvolvimento deste trabalho. Fico muito feliz por ter compartilhado o meu projeto com uma aluna esforçada e interessada como você.

Aos meus amigos, Victor Vardanega, Jessica Magno, Giuliana Biagini, Bruna Rachwal e Daniela Hartmann, que me acolheram durante o TCC e desde então compartilham comigo lágrimas e risadas.

Aos demais colegas do LGMH, especialmente à Iriel Luque e Hellen Issler, pelo companheirismo e convivência durante esses anos.

Aos técnicos, Priscila dos Santos e Valter de Baura, pelo suporte prestado e conselhos dados que guiaram principalmente o sequenciamento realizado neste trabalho.

Aos membros da banca de acompanhamento, Angelica Boldt e Gabriel Cipolla, pelos comentários e correções sempre muito pertinentes.

Cada um de vocês foi imprescindível para a conclusão deste trabalho, com certeza a jornada tornou-se muito mais leve com vocês ao meu lado. Muito obrigada!

Aqueles que passam por nós, não vão sós, não nos deixam sós.
Deixam um pouco de si, levam um pouco de nós.

(Antoine de Saint-Exupéry)

RESUMO

Em humanos, a persistência da lactase (LP) é a capacidade de digerir a lactose durante a idade adulta. O padrão de distribuição do fenótipo LP está associado à cultura de pastoralismo e ordenha, sendo um exemplo de coevolução gene-cultura e evolução convergente. Além disso, o sinal de seleção observado para essa região genômica faz com que o fenótipo LP seja um dos exemplos de seleção natural em populações humanas mais conhecido. O gene *LCT* codifica a enzima lactase e sua expressão está diretamente relacionada ao fenótipo LP. Polimorfismos de nucleotídeo único (SNPs) localizados no íntron 13 do gene *MCM6*, região sobreposta ao acentuador do gene *LCT*, estão associados ao fenótipo LP. Os SNPs mais conhecidos são: *–13910C>T* (rs4988235, amplamente distribuído pela Europa), *–13907G>C* (rs41525747), *–13915T>G* (rs41380347) e *–14010G>C* (rs145946881) — comuns em populações pastoralistas da África Oriental e Oriente Médio. A população pan-americana possui, principalmente, componentes de ancestralidade genômica nativo americano, africano e europeu, em proporções que variam entre e dentro dos países. Poucos estudos sobre variantes associadas ao fenótipo LP em populações miscigenadas pan-americanas foram desenvolvidos, havendo informações limitadas à variação de origem europeia (*–13910C>T*). Dessa forma, este trabalho tem como objetivo (i) abordar a distribuição geográfica da variação *–13910C>T* ao longo das Américas; (ii) analisar este *locus* para evidências de seleção positiva em diferentes populações miscigenadas pan-americanas; e (iii) descrever em nível populacional as principais variantes associadas ao fenótipo LP em populações de ancestralidade africana da região de Curitiba (PR) e da comunidade quilombola de Sertão do Valongo, região rural de Porto Belo (SC). Avaliando 14.978 indivíduos de 13 populações pan-americanas distintas, observamos que a frequência do alelo *–13910*T* está diretamente correlacionada à porcentagem de ancestralidade europeia média (*rho* = 0,866; $p = 1{,}526 \times 10^{-10}$); que a frequência do alelo *–13910*T* tende para a frequência alélica observada nas populações europeias ancestrais; e que não há super-representação de haplótipos europeus na região flanqueadora da variação *–13910C>T*. Esses resultados sugerem que não houve seleção natural nas Américas após eventos de miscigenação que deram origem às populações pan-americanas modernas. Ainda, foram sequenciados 259 indivíduos para o íntron 13 do gene *MCM6*, as sequências geradas foram analisadas no *software Mutation Surveyor*®, versão 3.30 (*SoftGenetics*). Para a população afro-brasileira de Curitiba, o fenótipo LP está associado à variação *–13910C>T* em 98,78% dos casos; entretanto, não foram identificadas variações associadas ao fenótipo LP na população quilombola de Sertão do Valongo. Para avaliar os contextos haplotípicos nas populações afro-brasileiras de Curitiba e em Sertão do Valongo, foram genotipados três marcadores haplotípicos localizados no gene *LCT*, são eles: *–678A>G* (rs56211644), *+666G>A* (rs3754689) e *+5579T>C* (rs2278544). Em afro-brasileiros de Curitiba foram observados 20 haplótipos distintos, dos quais apenas 4 foram observados na população de Sertão do Valongo — em ambos os grupos o haplótipo mais frequente contém os alelos ancestrais. Os resultados obtidos mostraram que afro-brasileiros da região de Curitiba possuem ancestralidade europeia importante para a região genômica avaliada, além disso os

resultados são compatíveis com a diversidade genética relatada para a população pan-americana.

Palavras-chave: Gene *MCM6* 1. SNP 2. Hipolactasia primária 3. América Latina 4. Quilombo 5.

ABSTRACT

In humans, lactase persistence (LP) is the ability to digest lactose during adulthood. The distribution pattern and the prevalence of LP are associated with pastoralism and milking culture; lactase persistence is also among the strongest selected phenotypes in human populations, being one of the best examples of convergent evolution and gene-culture co-evolution. The *LCT* gene encodes the lactase enzyme, and its expression is directly related to the LP phenotype. Single nucleotide polymorphisms (SNPs) located at intron 13 of the *MCM6* gene, a region overlapped to the *LCT* gene's enhancer, are associated with the LP phenotype. The most well-known SNPs are: *–13910C>T* (rs4988235, widely distributed in Europe), *–13907G>C* (rs41525747), *–13915T>G* (rs41380347), and *–14010G>C* (rs145946881) — common in herder populations in East Africa and the Middle East. The Pan-American admixed populations have primarily Native American, African, and European ancestries, which proportions vary among and within the countries. Few studies about variants associated with the LP phenotype in Pan-American populations have been developed, with information limited to the European origin variation (*–13910C>T*). Thus, this work aims to (i) address the geographical distribution of the *–13910C>T* SNP along the Americas; (ii) analyze this locus for evidence of positive selection in different Pan-American admixed populations; and (iii) describe at the population level the main SNPs associated with the LP phenotype in populations of African ancestry from the urban region of Curitiba (PR) and the quilombola community of Sertão do Valongo, rural area of Porto Belo (SC). Evaluating 14.978 individuals from thirteen different Pan-American admixed populations, we observed a positive correlation between *–13910*T* allele frequencies and the average proportion of European ancestry (*rho* = 0.866; $p = 1.526 \times 10^{-10}$); the *–13910*T* allele frequency tends to the European sources' allele frequency; and no overrepresentation of European haplotypes in the *–13910C>T* flanking region was observed. These findings suggest no selective pressure after admixture events that gave rise to modern Pan-American populations. Moreover, we sequenced two hundred fifty-nine samples for the *MCM6* gene (intron 13), and the generated sequences were analyzed using the Mutation Surveyor® software, version 3.30 (SoftGenetics). While the LP phenotype is associated with the *–13910C>T* SNP in 98.78% of cases in the Afro-Brazilian from Curitiba, in the quilombola community from Sertão do Valongo, there was no allele associated with the LP phenotype. The haplotypic backgrounds in the Afro-Brazilian from Curitiba and Sertão do Valongo were evaluated by genotyping three markers located in the *LCT gene*: *–678A>G* (rs56211644), *+666G>A* (rs3754689), and *+5579T>C* (rs2278544). We observed twenty distinct haplotypes in the Afro-Brazilian from Curitiba, of which only four haplotypes were also observed in the quilombola community from Sertão do Valongo. In both populations, the most frequent haplotype contained ancestral alleles. Our results showed that Afro-Brazilians from Curitiba have an important European ancestry for the *MCM6* genomic region, agreeing with the genetic diversity reported for Pan-American populations.

Keywords: *MCM6* gene 1. SNP 2. Adult-type hypolactasia 3. Latin America 4. Quilombo 5.

LISTA DE FIGURAS

FIGURA 1: A DIGESTÃO DA LACTOSE E OS FENÓTIPOS LACTASE PERSISTENTE (LP) E LACTASE NÃO PERSISTENTE (LNP).....20
FIGURA 2: LOCALIZAÇÃO E ESTRUTURA DOS GENES *LCT* E *MCM6*.....23
FIGURA 3: MARCADORES HAPLOTÍPICOS DO GENE *LCT*.....30
FIGURA 4: REPRESENTAÇÃO GRÁFICA DOS HAPLÓTIPOS DO GENE LCT.....33
FIGURA 5: MECANISMOS GENÉTICOS ASSOCIADOS À NÃO PERSISTÊNCIA (LNP) E PERSISTÊNCIA DA LACTASE (LP) EM HUMANOS.....36
FIGURA 6: DISTRIBUIÇÃO FENOTÍPICA DA PERSISTÊNCIA DA LACTASE NA EUROPA.....37
FIGURA 7: FREQUÊNCIA E DISTRIBUIÇÃO FENOTÍPICA MUNDIAL DA PERSISTÊNCIA DA LACTASE.....38
FIGURA 8: VARIAÇÃO NA FREQUÊNCIA E DISTRIBUIÇÃO DOS ALELOS – *14010*C*, *–13910*T*, *–13915*G* E *–13907*G* ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE EM POPULAÇÕES AFRICANAS, DA PENÍNSULA ARÁBICA E DO ORIENTE MÉDIO.....42
FIGURA 9: DISTRIBUIÇÃO DE IMC ENTRE OS SEXOS FEMININO E MASCULINO DE ACORDO COM O FENÓTIPO LACTASE PERSISTENTE (LP) OU LACTASE NÃO PERSISTENTE (LNP).....128

LISTA DE TABELAS

TABELA 1: SNPs ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE LOCALIZADOS NO GENE *MCM6*.....29
TABELA 2: NÚMERO DE HAPLÓTIPOS E FREQUÊNCIA HAPLOTÍPICA PARA DIFERENTES GRUPOS POPULACIONAIS.....31
TABELA 3: ALELOS ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE E SEUS RESPECTIVOS HAPLÓTIPOS EM POPULAÇÕES DISTINTAS.....32
TABELA 4: ÍNDICES DE MASSA CORPORAL (IMC) E CLASSES DE PESO PARA INDIVÍDUOS ADULTOS ACIMA DE 20 ANOS.....61
TABELA 5: PROTOCOLOS 1 E 2 DA PCR.....71
TABELA 6: CICLOS DA PCR.....71
TABELA 7: POLIMORFISMOS SELECIONADOS PARA GENOTIPAGEM, IDENTIFICAÇÃO DOS SEUS RESPECTIVOS INICIADORES, SEQUÊNCIA E TAMANHO DO FRAGMENTO AMPLIFICADO.....73
TABELA 8: PROTOCOLO DA PCR ALELO ESPECÍFICA *MULTIPLEX*.....73
TABELA 9: CICLOS DA PCR ALELO ESPECÍFICA *MULTIPLEX*.....74
TABELA 10: FREQUÊNCIAS ALÉLICAS, GENOTÍPICAS E FENOTÍPICAS EM AFRO-BRASILEIROS DE CURITIBA E REGIÃO METROPOLITANA (PR) E DA COMUNIDADE QUILOMBOLA DE SERTÃO DO VALONGO (SC).....113
TABELA 11: FREQUÊNCIA HAPLOTÍPICA E HAPLÓTIPOS IDENTIFICADOS EM AFRO-BRASILEIROS DE CURITIBA E REGIÃO METROPOLITANA (PR).....122
TABELA 12: FREQUÊNCIA HAPLOTÍPICA E HAPLÓTIPOS IDENTIFICADOS EM AFRO-BRASILEIROS DA COMUNIDADE QUILOMBOLA DE SERTÃO DO VALONGO (SC).....123
TABELA 13: PARÂMETROS DE DESEQUILÍBRIO DE LIGAÇÃO D', r^2, χ^2 E p PARA AFRO-BRASILEIROS DE CURITIBA E REGIÃO METROPOLITANA (PR) E DA COMUNIDADE QUILOMBOLA DE SERTÃO DO VALONGO (SC).....126
TABELA 14: FREQUÊNCIAS ALÉLICAS E GENOTÍPICAS NA SUBAMOSTRA DE AFRO-BRASILEIROS DE CURITIBA E REGIÃO METROPOLITANA (PR).....129
TABELA 15: NÚMERO DE INDIVÍDUOS DISTRIBUÍDOS DE ACORDO COM A CLASSE DE PESO (NORMAL OU ACIMA DO PESO E OBESIDADE) E FENÓTIPO

LACTASE PERSISTENTE (LP) OU LACTASE NÃO PERSISTENTE (LNP) E DE ACORDO COM O SEXO (MASCULINO OU FEMININO).....130
TABELA 16: TESTES REALIZADOS PARA A ANÁLISE DOS DADOS, SEUS RESPECTIVOS OBJETIVOS, NÚMERO DE OBSERVAÇÕES, PARÂMETROS UTILIZADOS E CONCLUSÕES.....132

LISTA DE ABREVIATURAS OU SIGLAS

aC	antes de Cristo
AIMs	*Ancestry informative markers*
CAAE	Certificado de Apresentação de Apreciação Ética
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CONEP	Comissão Nacional de Ética em Pesquisa
CQ	Comunidade Quilombola
dC	depois de Cristo
DL	Desequilíbrio de ligação
EHH	*Extended haplotype homozygosity*
HGNC	*HUGO Gene Nomenclature Committee*
IBGE	Instituto Brasileiro de Geografia e Estatística
iHS	*Integrated haplotype score*
IMC	Índice de massa corporal
LGMH	Laboratório de Genética Molecular Humana
LNP	Lactase não persistente
LP	Lactase persistente
LPL	Laboratório de Polimorfismos e Ligação
meQTLs	*Methylation quantitative trait loci*
NCBI	*National Center for Biotechnology Information*
OMIM	*Online Mendelian Inheritance in Man*
OR	*Odds ratio*
PCR	*Polymerase chain reaction*
ppm	Partes por milhão
PR	Paraná
SC	Santa Catarina
SNP	*Single nucleotide polymorphism*
UFPR	Universidade Federal do Paraná
UTR	*Untranslated region*

SUMÁRIO

1 INTRODUÇÃO.....19
2 REVISÃO BIBLIOGRÁFICA.....23
2.1 GENES E ALELOS ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE.....23
2.2 HAPLÓTIPOS ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE...29
2.3 MECANISMOS REGULATÓRIOS PÓS-TRANSCRICIONAIS ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE.....34
2.4 GENÉTICA DE POPULAÇÕES E ASPECTOS EVOLUTIVOS DA PERSISTÊNCIA DA LACTASE.....36
2.5 COMPOSIÇÃO GENÉTICA DA POPULAÇÃO PAN-AMERICANA.....45
2.5.1 A comunidade quilombola de Sertão do Valongo.....48
2.6 O FENÓTIPO LACTASE PERSISTENTE EM POPULAÇÕES PAN-AMERICANAS.....50
2.6.1 O fenótipo lactase persistente em populações brasileiras.....53
2.7 O FENÓTIPO LACTASE PERSISTENTE E AS CLASSES DE PESO.....60
3 JUSTIFICATIVA.....65
4 HIPÓTESE.....66
5 OBJETIVOS.....67
5.1 OBJETIVOS GERAIS.....67
5.2 OBJETIVOS ESPECÍFICOS.....67
6 METODOLOGIA.....68
6.1 AMOSTRAS.....68
6.2 CONJUNTO DE DADOS.....69
6.3 SEQUENCIAMENTO DO ÍNTRON 13 DO GENE *MCM6*.....70
6.3.1 PCR.....70
6.3.2 Reações de sequenciamento.....72
6.4 GENOTIPAGEM DE MARCADORES HAPLOTÍPICOS.....72
6.5 ANÁLISE DE DADOS.....74
6.5.1 Sequenciamento do íntron 13 do gene *MCM6*.....74
6.5.2 Frequências alélicas e genotípicas.....75

6.5.3 Contribuição genética e miscigenação.....75
6.5.4 Inferência de fenótipos lactase persistente e lactase não persistente.....76
6.5.5 Inferência de haplótipos e desequilíbrio de ligação.....77
6.5.6 Avaliação piloto: associação entre o fenótipo lactase persistente e as classes de peso.....77
7 RESULTADOS E DISCUSSÕES.....79
7.1 TRACING THE DISTRIBUTION OF EUROPEAN LACTASE PERSISTENCE GENOTYPES ALONG THE AMERICAS.....79
7.2 POPULAÇÕES DE ANCESTRALIDADE AFRICANA DO SUL DO BRASIL 112
7.2.1 Frequências alélicas e genotípicas.....112
7.2.2 Inferência de fenótipos lactase persistente e lactase não persistente.....119
7.2.3 Inferência de haplótipos e desequilíbrio de ligação.....121
7.2.4 Avaliação piloto: associação entre o fenótipo lactase persistente e as classes de peso.....127
8 CONCLUSÕES.....134
REFERÊNCIAS.....137
APÊNDICE 1.....152
ANEXO 1.....160
ANEXO 2.....162
ANEXO 3.....164
ANEXO 4.....166
ANEXO 5.....167

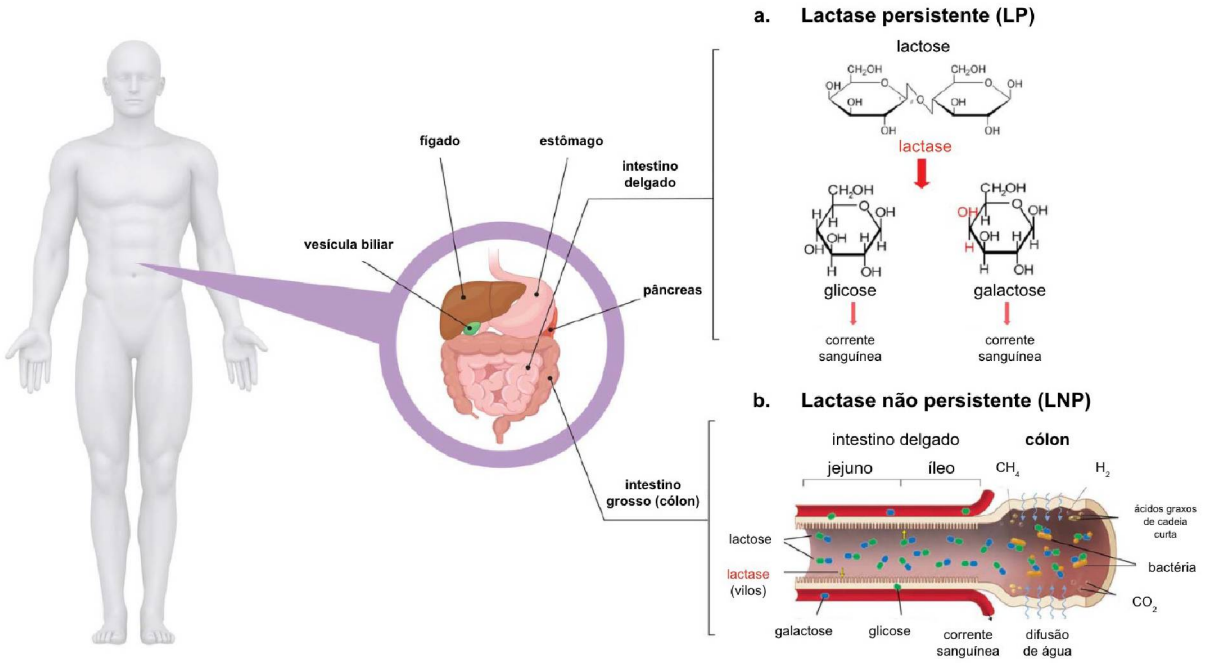
1 INTRODUÇÃO

A lactase florizina hidrolase (E.C. 3.2.1.23-62) — mais conhecida como lactase — é uma enzima expressa pelas células dos vilos do intestino delgado responsável pela hidrólise da lactose, dissacarídeo não absorvível, em glicose e galactose, açúcares simples absorvíveis pela mucosa intestinal (SAHI, 1994). A lactose é abundante no leite de mamíferos, que é essencial para o desenvolvimento de recém-nascidos. Por isso, todos os mamíferos possuem atividade de lactase intestinal máxima durante o período perinatal (WANG *et al.*, 1994). Após o desmame, a dieta passa a ter diferentes fontes além do leite e, em mamíferos adultos, a expressão gênica da lactase é reduzida naturalmente, permanecendo baixa ao decorrer do desenvolvimento (SWAGERTY, WALLING e KLEIN, 2002).

Em humanos, a expressão gênica da lactase e, consequentemente, a atividade enzimática dela, pode ou não sofrer alterações originando dois grupos distintos de indivíduos: indivíduos de fenótipo lactase persistente (LP) — observado em cerca de 35% da população mundial adulta, porém com heterogeneidade pronunciada entre populações —, aqueles com atividade de lactase similar à observada em fases de desenvolvimento anteriores, possuem alta capacidade de digestão da lactose (FIGURA 1a); e indivíduos de fenótipo lactase não persistente (LNP), aqueles com baixa atividade de lactase e baixa capacidade de digestão da lactose quando adultos (FIGURA 1b) (SWAGERTY, WALLING e KLEIN, 2002; TROELSEN, 2005; RASINSPERÄ *et al.*, 2005).

A redução na expressão da lactase ou ausência dela em adultos foi descrita pela primeira vez em 1963 por Auricchio e colaboradores. Geralmente assintomática, o quadro clínico está relacionado à indigestão e não absorção ou absorção reduzida da lactose presente no leite, podendo evoluir para hipolactasia, na qual o indivíduo desenvolve intolerância à lactose (OMIM #223100) associada aos sintomas de indigestão, inchaço, náusea, vômito, cólica abdominal, distensão abdominal, flatulência, gordura nas fezes e diarreia (HOLZEL *et al.*, 1959; AURICCHIO *et al.*, 1963).

FIGURA 1: A DIGESTÃO DA LACTOSE E OS FENÓTIPOS LACTASE PERSISTENTE (LP) E LACTASE NÃO PERSISTENTE (LNP)



a. Em indivíduos lactase persistentes, fenótipo LP, a lactose é hidrolisada no intestino delgado pela enzima lactase em dois monossacarídeos, glicose e galactose, que são rapidamente absorvidos pela corrente sanguínea; b. Em indivíduos lactase não persistentes, fenótipo LNP, a lactose não é hidrolisada no intestino delgado e chega íntegra ao cólon (intestino grosso). As bactérias presentes no intestino grosso realizam a fermentação da lactose. A fermentação bacteriana gera ácidos graxos de cadeia curta e gases (CO_2, H_2 e CH_4) como metabólitos finais, que contribuem para o desenvolvimento dos sintomas gastrointestinais associados à intolerância à lactose. LP: lactase persistente; LNP: lactase não persistente. FONTE: Adaptado de Campbell e Ranciaro, 2021.

A hipolactasia pode acometer qualquer faixa etária, sendo classificada em três tipos: hipolactasia congênita (OMIM #223000), uma característica autossômica recessiva extremamente rara (desde 1966 foi diagnosticada em 42 pacientes de 35 famílias distintas na Finlândia, onde estima-se que a incidência seja de 1 a cada 60.000 indivíduos), na qual o indivíduo portador possui atividade enzimática mínima ao nascer (NEVANLINNA, 1972; NORIO *et al.*, 1973; ROBAYO-TORRES e NICHOLS, 2007); hipolactasia primária (OMIM #223100), caracterizada pela perda irreversível da atividade enzimática após o desmame (SWAGERTY, WALLING e KLEIN, 2002; TROELSEN, 2005); e hipolactasia secundária, caracterizada pela perda reversível da atividade enzimática devido a condições pré-existentes que causam danos à mucosa intestinal como enteroparasitoses, doença celíaca e doença de Crohn (SWAGERTY, WALLING e KLEIN, 2002; ROBAYO-TORRES e

NICHOLS, 2007). Neste trabalho, serão abordadas especificamente as variantes genéticas associadas à hipolactasia primária (OMIM #223100).

A hipolactasia primária é considerada binária, bem como a determinação do fenótipo LP, que segue o padrão de herança autossômica dominante. Entretanto, a atividade enzimática da lactase é uma característica que obedece ao padrão de herança de dominância incompleta. Nesse caso, é observada atividade enzimática distinta entre homozigotos e heterozigotos para o fenótipo LP, ambos digerem a lactose, porém a capacidade de digestão é alterada em certas condições, sendo possível observar digestão intermediária em heterozigotos (SWALLOW, 2003).

Por muito tempo acreditava-se que a expressão do fenótipo LP na fase adulta fosse determinada por modificações na sequência do gene *LCT* (HGNC: 6530; *HUGO Gene Nomenclature Committee*), codificador da enzima lactase. Entretanto, atualmente, sabe-se que a expressão do fenótipo LP está associada a diversos polimorfismos de nucleotídeo único (SNP, do inglês, *single nucleotide polymorphism*) presentes no gene *MCM6* (HGNC: 6949), principalmente no íntron 13, numa região que atua como intensificadora da expressão gênica de *LCT* (TROELSEN *et al.*, 2003; LEWINSKY *et al.*, 2005; TISHKOFF *et al.*, 2007; LIEBERT *et al.*, 2016). O gene *MCM6* (do inglês, *minichromosome maintenance complex component 6*), por sua vez, codifica uma das proteínas que compõem o complexo proteico de manutenção do minicromossomo. Essas proteínas são altamente conservadas e desempenham papel essencial para a iniciação da replicação em genomas eucarióticos (NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION, 2021).

Estudos recentes demonstraram que outros fatores influenciam a expressão do fenótipo LNP e o desenvolvimento dos sintomas associados à intolerância à lactose, são eles: composição da microbiota intestinal de cada indivíduo (ZHONG *et al.*, 2004; HE *et al.*, 2008); modificações epigenéticas (LABRIE *et al.*, 2016); e a concentração de lactose em alimentos distintos (BUTRRISS, 1997). O genótipo sozinho (para LNP) não necessariamente determina a manifestação dos sintomas mencionados anteriormente (ANGUITA-RUIZ, AGUILERA e GIL, 2020). Assim, indivíduos de fenótipo LNP não necessariamente desenvolvem a hipolactasia primária associada a intolerância à lactose.

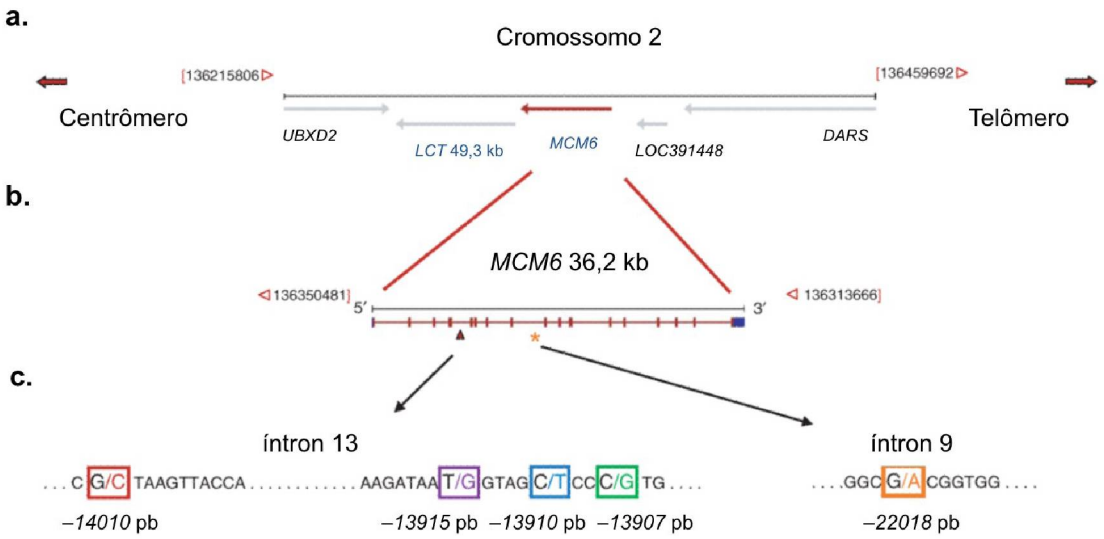
A avaliação do fenótipo LP, no geral, é interessante, pois a característica é um dos exemplos de seleção natural em populações humanas melhor conhecido, dado sinal de seleção observado para a região genômica (íntron 13, gene *MCM6*) (BERSAGLIERI *et al.*, 2004). Além disso, a característica é prevalente em populações com histórico pastoralista, sendo um dos melhores exemplos de evolução convergente e coevolução gene-cultura (BEJA-PEREIRA *et al.*, 2003; TISHKOFF *et al.*, 2007; INGRAM *et al.*, 2009; RANCIARO *et al.*, 2014). Embora muitos estudos tenham avaliado a distribuição da persistência da lactase e os alelos associados à característica em populações europeias, as informações para populações miscigenadas pan-americanas (populações da América Central, do Sul e do Norte) são restritas à variação identificada em europeus, não havendo amplas informações com relação às variações de origem não europeia. Portanto, do ponto de vista da genética de populações e, até mesmo, da genética epidemiológica, estudos que abordam a distribuição e prevalência do fenótipo LP em populações miscigenadas pan-americanas se tornam necessários.

2 REVISÃO BIBLIOGRÁFICA

2.1 GENES E ALELOS ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE

O gene *LCT*, identificado em 1988 (KRUSE *et al.*, 1988), está localizado no braço longo do cromossomo 2 (2q21.3) e possui 17 éxons que dão origem a um transcrito de 6 kb. Boll e colaboradores (1991) investigaram as sequências do transcrito maduro (RNA mensageiro) do gene *LCT* em indivíduos de fenótipo LP e LNP. As sequências avaliadas eram idênticas, sugerindo que a determinação dos fenótipos se dá por mecanismos regulatórios transcricionais. Posteriormente, foram identificados SNPs numa região localizada 13,9 kb a montante do gene *LCT* sobreposta ao gene *MCM6*, que atua como intensificadora do gene *LCT* (FIGURA 2).

FIGURA 2: LOCALIZAÇÃO E ESTRUTURA DOS GENES *LCT* E *MCM6*



a. Localização dos genes *LCT* e *MCM6* no cromossomo 2; b. Estrutura do gene *MCM6*; c. Localização dos SNPs associados à persistência da lactase nos íntrons 9 e 13 do gene *MCM6* em populações africanas (*–14010G>C*, *–13915T>G* e *–13907C>G*) e europeias (*–13910C>T* e *–22018G>A*). FONTE: adaptado de Tishkoff *et al.* (2007).

Enattah e colaboradores (2002) identificaram SNPs associados ao fenótipo LP em 196 indivíduos finlandeses não aparentados — população na qual o traço é observado em frequências superiores a 80% (JUSSILA, 1969; JUSSILA, ISOKOSKI

e LAUNIALA, 1970; SAHI, 1974). Foram identificadas substituições *C>T* e *G>A* nos íntrons 13 (*–13910C>T*, rs4988235) e 9 (*–22018G>A*, rs182549), respectivamente, do gene *MCM6*. Nas populações finlandesa e não finlandesas (sul-coreana, italiana e alemã) os alelos *–13910*T* e *–22018*A* (derivados) determinam, mesmo em heterozigose, o fenótipo LP, ao passo que os alelos *–13910*C* e *–22018*G* (ancestrais) determinam, apenas em homozigose, o fenótipo LNP. Avaliando as frequências alélica e fenotípica observadas, foi possível constatar que os alelos *–13910*T* e *–22018*A* estão, respectivamente, 100% e 97% associados estatisticamente a determinação do fenótipo LP na idade adulta na população finlandesa estudada, sendo o alelo *–13910*T* melhor preditor de fenótipo (ENATTAH *et al.*, 2002). Nas populações finlandesa e não finlandesas investigadas (sul-coreana, italiana e alemã), a frequência do alelo *–13910*T* é compatível com a frequência fenotípica para LP — relação também veridicada posteriormente em outras populações europeias (ITAN *et al.*, 2010). Já o alelo *–22018*A* possui frequência compatível com a frequência fenotípica para LP na população finlandesa em questão, entretanto, indivíduos não finlandeses de fenótipo LNP apresentaram o alelo *–22018*A*, sendo incompatível com o genótipo esperado para o traço. Além disso, foi realizada inferência de haplótipos com base em 33 alelos associados ao fenótipo LP na população finlandesa avaliada no estudo. Desses, 20 alelos estão presentes num haplótipo longo de 1 Mb (ENATTAH *et al.*, 2002) — denominado de haplótipo *A* (tópico 2.2 deste trabalho) — também reportado em outras populações europeias (HARVEY *et al.*, 1998; HOLLOX *et al.*, 2001), enquanto os alelos restantes estão presentes em outros haplótipos observados em menor frequência (ENATTAH *et al.*, 2002).

Troelsen e colaboradores (2003) descreveram uma diferença funcional entre os alelos da posição *–13910* que corrobora a ideia de que o fenótipo LP é devido à substituição *C>T* neste sítio. O efeito regulador das posições *–13910* e *–22018* em indivíduos de fenótipo LP e LNP foi avaliado em ensaios de luciferase, a partir da transfecção de gene repórter em linhagens de células intestinais Caco-2. Foi possível demonstrar que a região na qual ocorre a variação *–13910C>T* atua como acentuadora da transcrição do gene *LCT*, na presença da variante *–13910*T*, a transcrição do gene repórter é aumentada em até quatro vezes quando comparada com o mesmo acentuador contendo o alelo *–13910*C*. Todavia, os alelos *–22018*G*

*e –22018*A* não apresentaram efeito regulador diferencial, o que sugere que a variação *–22018G>A* não desempenha função acentuadora da transcrição do gene *LCT* (como observado para a variação *–13910C>T*) e, consequentemente, na determinação do fenótipo LP (TROELSEN *et al.*, 2003). Ainda, os alelos *–13910*T* e *–22018*A* estão em desequilíbrio de ligação completo (r^2 = 1,0) na população finlandesa avaliada (ENATTAH *et al.*, 2002) e em forte desequilíbrio (r^2 > 0,8) em populações registradas no banco de dados *1000 Genomes Project* fase 3 (THE 1000 GENOMES PROJECT CONSORTIUM, 2015), o que sugere que o SNP *–13910C>T* possui efeito biológico, é causal, enquanto o SNP *–22018G>A* é apenas um marcador.

Posteriormente, Lewinsky e colaboradores (2005) analisaram a mesma região e identificaram que o fator de transcrição OCT-1 liga-se num segmento que inclui a variação *–13910C>T*. Os resultados obtidos indicam que a persistência da lactase em seres humanos é consequência do aumento da expressão gênica devido à atividade acentuadora mediada pelo fator de transcrição OCT-1. Este, por possuir maior afinidade pela variante *–13910*T*, determina a manutenção da expressão do gene *LCT*, mesmo após o desmame (LEWINSKY *et al.*, 2005), corroborando os estudos anteriores.

Outros estudos buscaram caracterizar populações não europeias para a posição *–13910*. Mulcare e colaboradores (2004) investigaram a presença do alelo *–13910*T* em 1.671 indivíduos de 20 populações africanas distintas, dentre elas, populações pastoralistas e não pastoralistas da mesma região. Foi relatado que a baixa frequência do alelo *–13910*T*, mesmo presente em algumas populações urbanas das etnias Fulani e Hausa de Camarões, não era compatível com a alta frequência do fenótipo LP relatada na literatura — em contraste com a compatibilidade descrita no estudo de Enattah e colaboradores (2002) e observada em amostras irlandesas utilizadas para fins comparativos. Desse modo, a baixa frequência, e até ausência, da variante *–13910*T* nas populações africanas em questão sugerem que o alelo *T* não é preditor confiável do fenótipo LP, e que, para as populações dessa região, a substituição *–13910C>T* não é causal ou não é a única causa para o fenótipo (MULCARE *et al.*, 2004).

Ingram e colaboradores (2007) investigaram polimorfismos associados ao fenótipo LP nas populações Fulani e Shuwa (árabes) de Camarões, Afar e Somali da

Etiópia, pastoralistas Beni Amir (grupo Beja) do Sudão e beduínos de Israel (área da Autoridade Palestina/PAA), da Jordânia e da Arábia Saudita. A partir do sequenciamento da região de 13,9 kb a montante do gene *LCT*, foi possível identificar variantes não descritas em populações europeias anteriormente avaliadas. São elas: *–13907*G* (*–13907C>G*, rs41525747), observada nas populações das regiões do Sudão e Etiópia, sendo mais frequente na população Afar (Etiópia); *–13913*C* (*–13913T>C*, rs41456145), a mais rara delas, observada com maior frequência nas populações de etnia Somali (Etiópia) e Fulani (Camarões), que também possui o alelo *–13910*T* em alta frequência devido a eventos de migração e fluxo gênico europeu no norte da África (VICENTE *et al.*, 2019); e *–13915*G* (*–13915T>G*, rs41380347), observada nas populações da África Oriental e do Oriente Médio. Essa última variante é comum entre beduínos da Arábia Saudita, estando associada ao consumo de leite, de acordo com o relatado pelos participantes do estudo. As variantes *–13907*G*, *–13913*C* e *–13915*G* estão localizadas muito próximas da variante *–13910*T* (FIGURA 2). Além disso, *–13913*C* e *–13915*G* se sobrepõem ao sítio de ligação de OCT-1 e *–13907*G* está apenas três nucleotídeos fora dele (INGRAM *et al.*, 2007).

Corroborando o apresentado anteriormente, Tishkoff e colaboradores (2007) observaram as variantes *–13907*G* e *–13915*G* e descreveram uma nova substituição *G>C* na posição *–14010* (*–14010G>C*, rs145946881) (FIGURA 2). O estudo foi desenvolvido em 470 indivíduos africanos de 43 grupos étnicos distintos originários da Tanzânia, Quênia e Sudão. O alelo *–14010*C*, associado ao fenótipo LP, é frequente nas populações afro-asiáticas (46%) e nilo-saarianas (39%) da Tanzânia e nilo-saarianas (32%) do Quênia, porém é ausente no Sudão. Os alelos africanos *–13907*G*, *–13915*G* e *–14010*C*, para as populações em questão, estão presentes nos haplótipos *B*, *C* e *F*, respectivamente — distintos do haplótipo europeu (denominado haplótipo *A*) no qual é observado o alelo *–13910*T* (TISHKOFF *et al.*, 2007; tópico 2.2 deste trabalho).

Além desses, o alelo *–14009*G* (*–14009T>G*, rs869051967), também associado ao fenótipo LP, foi descrito em populações de etnia Somali da Etiópia e em indivíduos etíopes de diversos grupos étnicos (INGRAM *et al.*, 2009; JONES *et al.*, 2013). Ingram e colaboradores (2009) demonstraram que, para as populações de etnia Somali da Etiópia, os alelos *–13907*G* e *–13915*G* também estão

associados ao fenótipo LP. Ainda, foi possível inferir os haplótipos nos quais os alelos estariam inseridos: os alelos *–13915*G*, *–14009*C* e *–14010*C* estão presentes, respectivamente, nos haplótipos *C*, *X* e *B*, para as populações de etnia Somali avaliadas (INGRAM *et al.*, 2009; tópico 2.2 deste trabalho).

Para avaliar o efeito das variações *–13907C>G*, *–13915T>G* e *–14010G>C* sobre a expressão gênica e como elas interagem com o promotor do gene *LCT*, Tishkoff e colaboradores (2007) realizaram ensaios de expressão de luciferase a partir da transfecção de gene repórter em linhagens de células intestinais Caco-2, de forma semelhante ao realizado em 2003 por Troelsen e colaboradores. Foram testados vetores contendo a região promotora do gene *LCT* combinada aos haplótipos do gene *MCM6* originando cinco vetores: haplótipo ancestral (alelos *–13907*C*, *–13915*T* e *–14010*G*) ligado a posição *–13495* contendo alelo *C* (controle interno do experimento); haplótipo ancestral ligado a posição *–13495* contendo alelo *T* (utilizado para controlar o efeito da variante *–13495*C*); haplótipo derivado que difere do haplótipo ancestral apenas em *–14010*C*; haplótipo derivado que difere do haplótipo ancestral em *–13907*G* e *–13495*T*; e haplótipo derivado que difere do haplótipo ancestral em *–13915*G*. Os vetores contendo haplótipos derivados para os alelos *–14010*C*, *–13915*G* e *–13907*G* potencializaram a transcrição do gene repórter em 18% a 30% quando comparados ao vetor contendo o haplótipo ancestral, assim, os alelos derivados promovem *in vitro* maior expressão gênica de *LCT*. Ainda, não foi observada diferença estatisticamente significante entre os vetores contendo os haplótipos ancestrais com e sem o alelo *–13495*T*, e entre os vetores contendo os haplótipos derivados (alelos *–13907*G*, *–13915*G* e *–14010*C*) (TISHKOFF *et al.*, 2007).

Liebert e colaboradores (2016) desenvolveram análises funcionais semelhantes às descritas por TROELSEN e colaboradores (2003) e TISHKOFF e colaboradores (2007). A partir delas, foi possível demonstrar que os alelos *–14009*G*, *–14010*C* e *–14011*T* (*–14011C>T*, rs4988233) potencializaram a transcrição do gene repórter quando comparados aos alelos ancestrais (*T*, *G* e *C*, respectivamente), dessa forma, os alelos derivados promovem *in vitro* maior expressão gênica de *LCT*. Além disso, foi possível concluir que as variantes em questão atuam de formas distintas: *–14010*C* e *–14011*T* ligam-se com o fator de transcrição OCT-1, de forma similar ao descrito para o alelo *–13910*T* (LEWINSKY

et al., 2005); enquanto *–14009*T* liga-se com um fator de transcrição da família ETS, provavelmente, C-Ets-1 (LIEBERT *et al.*, 2016). Também foram avaliadas as variações *–13779G>C* (rs527991977) e *–14028T>C* (rs759157971). A presença do alelo *–13779*C* potencializa a transcrição do gene repórter, porém não foram observadas diferenças significativas na interação com fatores de transcrição. Já o alelo *–14028*C* apresentou interação clara com fatores de transcrição, entretanto não apresentou efeito potencializador nos ensaios de transfecção, quando comparado ao efeito do alelo ancestral *–14028*T* (LIEBERT *et al.*, 2016).

Torniainen e colaboradores (2009) buscaram por variantes associadas ao fenótipo LP na população de etnia Xhosa (n = 109) e na população de ancestralidade miscigenada (n = 62) da África do Sul, que consomem laticínios, e compararam os resultados obtidos com variantes observadas na população de agricultores de etnia Akan (n = 196) de Gana, que não possuem o mesmo hábito de consumo. No total, foram observados seis alelos, dos quais *–13910*T*, *–14010*C* e *–13937*A (–13937G>A,* rs4988234) já eram conhecidos, e *–14091*T*, *–14107*A* e *–14176*C* foram observados pela primeira vez. O alelo *–13910*T* foi observado apenas na população de ancestralidade miscigenada (21,8%), enquanto o alelo *–14010*C* foi observado em 13,3% na população de etnia Xhosa, em 6,5% na população de ancestralidade miscigenada e é ausente na população de etnia Akan. A presença de ambos os alelos na população de ancestralidade miscigenada reflete a origem étnica diversa, composta por, principalmente, africanos, europeus do oeste, povos de etnia Khoisan, indonésios e malaios. O alelo *–13937*A* foi observado em 1,4% na população de etnia Xhosa e é ausente nas outras populações. Os alelos *–14091*T*, *–14107*A* e *–14176*C* estão presentes em 1,4%, 0,5% e 0,9%, respectivamente, na população de etnia Xhosa, estão ausentes na população de ancestralidade miscigenada e, na população de etnia Akan, apenas *–14107*A* foi observado na frequência de 0,3% (TORNIAINEN *et al.*, 2009). Segundo os autores, a ausência de variações associadas ao fenótipo LP na população de etnia Akan sugere que não houve pressão seletiva suficiente para o aumento de frequência dos genótipos associados à LP, dado que a população em questão não consome laticínios (TORNIAINEN *et al.*, 2009) — ainda é possível que a ação da seleção sobre essa população não tenha sido identificada por ser muito recente.

Na TABELA 1, pode-se observar os SNPs associadas ao fenótipo LP apresentados nesse tópico e suas respectivas identificação, posição referente ao início da transcrição do gene *LCT*, substituição, validação funcional, posição genômica (referência GRCh 38.p12) e informações complementares.

TABELA 1: SNPs ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE LOCALIZADOS NO GENE *MCM6*

Identificação (rs)	Posição*	Substituição	Validação funcional	Posição genômica**	Informação complementar
rs41525747	*–13907*	*C>G*	Possui associação funcional	chr2:135851073	Variação rara (q < 5,0%)
rs9488235	*–13910*	*C>T*	Possui associação funcional	chr2:135851076	Amplamente estudado
rs41380347	*–13915*	*T>G*	Possui associação funcional	chr2:135851081	Amplamente estudado
rs869051967	*–14009*	*T>G*	Possui associação funcional	chr2:135851175	Amplamente estudado
rs145946881	*–14010*	*G>C*	Possui associação funcional	chr2:135851176	Amplamente estudado
rs9488233	*–14011*	*C>T*	Possui associação funcional	chr2:135851177	Variação rara (q < 5,0%)
rs182549	*–22018*	*G>A*	Não possui associação funcional	chr2:135859184	Em DL absoluto com *–13910C>T* (r2 = 1,0)***
rs527991977	*–13779*	*G>C*	Possui evidência de função	chr2:135850945	Variação rara (q < 5,0%)
rs41456145	*–13913*	*T>C*	Possui evidência de função	chr2:135851079	Variação rara (q < 5,0%)
rs9488234	*–13937*	*G>A*	Possui evidência de função	chr2:135851103	Variação rara (q < 5,0%)

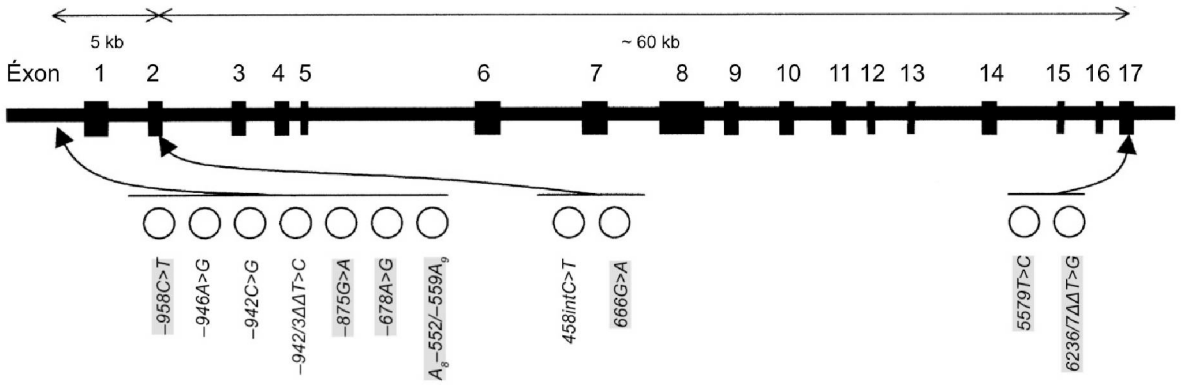
* Posição referente ao início da transcrição do gene *LCT*; ** Genoma referência GRCh 38.p12; *** Populações europeias CEU, FIN, GBR e IBS, americanas CLM e PUR, e africanas ACB, ASW e GWD do banco de dados *1000 Genomes Project* fase 3 (2015); DL: desequilíbrio de ligação; q: frequência alélica. FONTE: adaptado de Anguita-Ruiz, Aguilera e Gil (2020).

2.2 HAPLÓTIPOS ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE

Como descrito no tópico 2.1 deste trabalho, o primeiro SNP (gene *MCM6*, íntron 13) associado ao fenótipo LP foi identificado em populações finlandesas apenas em 2002 (ENATTAH *et al.*, 2002). Anteriormente, os esforços para identificação de variantes associadas ao fenótipo LP estavam focados no sequenciamento do gene *LCT*, codificador da enzima lactase (BOLL *et al.*, 1991; LLOYD *et al.*; 1992). Entretanto, os estudos de sequenciamento desenvolvidos possibilitaram a identificação de diversos polimorfismos na região de aproximadamente 70 kb que abrange o gene *LCT* (BOLL *et al.*, 1991; HARVEY *et al.*, 1995). Assim, os polimorfismos identificados passaram a ser utilizados para a definição dos contextos haplotípicos da região que contém o gene *LCT* (HARVEY *et al.*, 1998; HOLLOX *et al.*, 2001). Atualmente, em estudos que avaliam o fenótipo LP, esses polimorfismos são amplamente empregados para definição dos contextos haplotípicos da região que contém o gene *MCM6*, sendo denominados de marcadores haplotípicos.

Em 1995, um estudo preliminar de Harvey e colaboradores em sete famílias de origem europeia do CEPH (Fundação Jean Dausset) revelou três haplótipos comuns do gene *LCT*. Ao avaliarem famílias de origem francesa e de Utah, nos Estados Unidos — região povoada, principalmente, por indivíduos provenientes do Reino Unido, Dinamarca, Suécia, Noruega e Finlândia (MCLELLAN *et al.*, 1984) — separadamente, foi possível observar frequência haplotípica distinta entre os grupos (HARVEY *et al.*, 1995). Para avaliar se a diferença de frequência haplotípica observada refletia a diferença de distribuição do fenótipo LP observada na Europa (tópico 2.4 deste trabalho), Harvey e colaboradores (1998) determinaram a frequência alélica de sete marcadores haplotípicos do gene *LCT* (FIGURA 3, destacados em cinza) em indivíduos europeus e indianos de fenótipo (LP ou LNP) inicialmente desconhecido.

FIGURA 3: MARCADORES HAPLOTÍPICOS DO GENE *LCT*



Os éxons estão representados como barras pretas. A série de círculos abaixo da identificação dos marcadores haplotípicos representam os sítios polimórficos que compõem o haplótipo. Destacados em cinza estão os marcadores haplotípicos avaliados por Harvey e colaboradores (1998). FONTE: adaptado de Hollox *et al.* (2001).

Os indivíduos avaliados foram agrupados de acordo com a região de origem ou etnia, gerando três classes distintas: Norte da Europa (residentes em Londres, Reino Unido), Sul da Europa (residentes em Londres de origem mediterrânea e naturais de Nápoles, Itália) e indianos (indivíduos do subcontinente indiano — Índia, Bangladesh e Paquistão — residentes no Reino Unido) (HARVEY *et al.*, 1998). Foram identificados quatro haplótipos (*A*, *B*, *C* e *E*) em comum entre os grupos do Norte da Europa, Sul da Europa e indianos e dois haplótipos relativamente comuns

entre europeus (*D*) e indianos (*G*) (TABELA 2). Ainda, em ambos os grupos europeus o haplótipo mais frequente foi o haplótipo *A*, com os haplótipos *B* e *C* mais frequentes na população do sul do que na população do norte. Comparando os grupos europeus avaliados com a população francesa (CEPH; HARVEY *et al.*, 1995), foi possível observar um gradiente decrescente de distribuição no sentido norte-sul para o haplótipo *A*. Para o grupo indiano, os haplótipos *A* e *C* foram igualmente frequentes, enquanto o haplótipo *B* foi o terceiro mais frequente (HARVEY *et al.*, 1998). Por último, os indivíduos foram testados para o fenótipo LP. O haplótipo *A* foi observado em maior frequência em populações que apresentam alta frequência do fenótipo LP. Entretanto, o haplótipo *A* também foi observado em indivíduos de fenótipo LNP e esses indivíduos apresentaram maior diversidade haplotípica (HARVEY *et al.*, 1998).

TABELA 2: NÚMERO DE HAPLÓTIPOS E FREQUÊNCIA HAPLOTÍPICA PARA DIFERENTES GRUPOS POPULACIONAIS

Haplótipo	europeus (norte)		europeus (sul)		indianos	
	N	%	N	%	N	%
A	90	86,54	40	37,74	17	34,00
B	6	5,77	34	32,08	9	18,00
C	3	2,88	13	12,26	18	36,00
D	3	2,88	6	5,66	-	-
E	2	1,92	7	6,60	2	4,00
G	-	-	-	-	4	8,00
I	-	-	2	1,89	-	-
J	-	-	2	1,89	-	-
L	-	-	1	0,94	-	-
M	-	-	1	0,94	-	-
TOTAL	104	100	106	100	50	100

Europeus do norte residentes em Londres, Reino Unido. Europeus do sul residentes em Londres de origem mediterrânea e naturais de Nápoles, Itália. Indianos, indivíduos do subcontinente indiano (Índia, Bangladesh e Paquistão) residentes no Reino Unido. N: número de haplótipos observados; %: frequência haplotípica. FONTE: adaptado de Harvey *et al.* (1998).

Em 2001, Hollox e colaboradores ampliaram as análises descritas por Harvey e colaboradores (1995; 1998) com o objetivo de avaliar a distribuição mundial dos haplótipos do gene *LCT*. Para isso, foram avaliados onze marcadores haplotípicos do gene *LCT* (FIGURA 3) em 669 indivíduos de onze populações distintas (África: Namíbia e África do Sul; Ásia: Índia, Sri Lanka, China, Malásia e Japão; Oceania: Papua-Nova Guiné; e Europa: Reino Unido, Itália, Rússia e

Eslováquia). Foram identificados 42 haplótipos distintos (FIGURA 4), dos quais apenas quatro apresentaram frequência mundial igual ou superior a 5,0% (*A*, *B*, *C* e *U*). A maior diversidade haplotípica registrada foi observada para as populações africanas, que apresentaram haplótipos exclusivos e valores mais altos de heterozigosidade (0,91 no grupo de etnia San, Namíbia; e 0,87 em populações falantes da língua Bantu, África do Sul). Ainda, segundo os autores, a maior diversidade haplotípica e os altos valores de heterozigosidade corroboram o modelo de povoamento dos continentes *Out of Africa* (HOLLOX *et al.*, 2001).

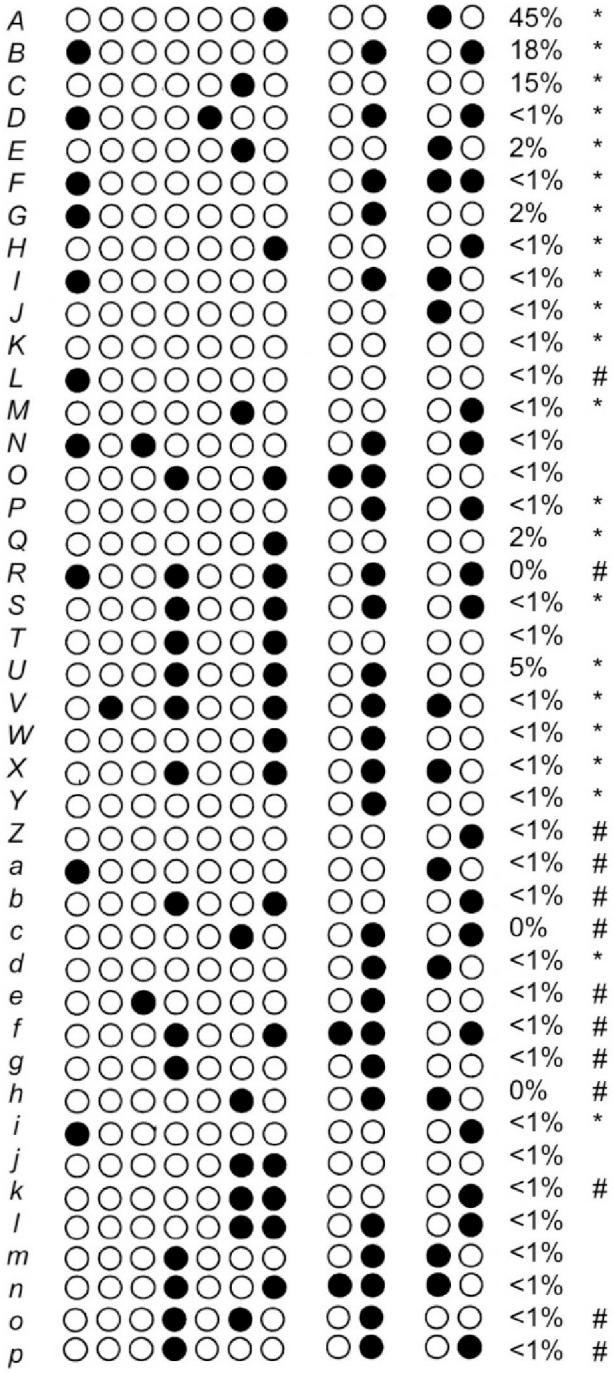
Assim, a avaliação do contexto haplotípico é interessante, uma vez que o mesmo alelo pode estar inserido em haplótipos distintos, de acordo com a população de origem. O alelo *–14010*C*, por exemplo, em populações da Etiópia está inserido no haplótipo *B* (INGRAM *et al.*, 2009), enquanto em populações da Tanzânia, Quênia e Sudão está inserido no haplótipo *F* (TISHKOFF *et al.*, 2007). Como a inserção de um determinado alelo em um contexto haplotípico não é concordante entre as populações, as diferenças observadas podem informar o local e a forma de surgimento do alelo. Na TABELA 3, pode-se observar os haplótipos que contém as variantes associadas ao fenótipo LP apresentadas no tópico 2.1 deste trabalho.

TABELA 3: ALELOS ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE E SEUS RESPECTIVOS HAPLÓTIPOS EM POPULAÇÕES DISTINTAS

Alelo	Haplótipo	Referência
*–13907*G*	***A***, *B, G, K, F*	Tishkoff *et al.* (2007); Enattah *et al.* (2008); Ingram *et al.* (2009); Liebert *et al.* (2017)
*–13910*T*	***A***, *B, C, G, J, K, P, W*	Poulter *et al.* (2003); Tishkoff et al. (2007); Friedrich *et al.* (2012a, 2012b); Ingram *et al.* (2007, 2009); Liebert *et al.* (2017)
*–13913*C*	***B***, *F, G, L*	Ingram *et al.* (2007, 2009); Jones *et al.* (2013); Liebert *et al.* (2017)
*–13915*G*	***C***, *E, M, K, c*	Tishkoff *et al.* (2007); Ingram *et al.* (2007, 2009); Jones *et al.* (2013); Liebert *et al.* (2017)
*–14009*G*	*H, O, S, T U,* ***X***	Ingram *et al.* (2009); Jones *et al.* (2013); Liebert *et al.* (2017)
*–14010*C*	*B, F,* ***P***	Tishkoff *et al.* (2007); Ingram *et al.* (2009); Friedrich *et al.* (2012a); Jones *et al.* (2013); Liebert *et al.* (2017)
*–14011*T*	***A, B***, *E, S*	Friedrich *et al.* (2012a); Liebert *et al.* (2017)

Destacados em negrito os haplótipos reportados em maior frequência contendo o alelo derivado. FONTE: a autora (2021).

FIGURA 4: REPRESENTAÇÃO GRÁFICA DOS HAPLÓTIPOS DO GENE *LCT*



A série de círculos representam os sítios polimórficos que compõem os haplótipos, na mesma ordem apresentada na FIGURA 3. Na coluna ao lado esquerdo da figura pode-se observar a identificação dos haplótipos. Na coluna ao lado direito da figura pode-se observar a frequência haplotípica. A coloração em preto indica se os alelos são derivados, enquanto a ausência de coloração indica alelos ancestrais. * Haplótipos definidos por homozigosidade observada em um indivíduo, heterozigosidade em apenas um sítio, ou deduzidos a partir de estudos anteriores; # Frequências observadas em todas as populações avaliadas. FONTE: adaptado de Hollox *et al.* (2001).

2.3 MECANISMOS REGULATÓRIOS PÓS-TRANSCRICIONAIS ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE

Apesar da identificação de diversos alelos associados ao fenótipo LP (descritos no tópico 2.1 deste trabalho), os mecanismos de determinação dos fenótipos (tanto LP quanto LNP) não foram completamente elucidados. Estudos recentes demonstraram que mecanismos regulatórios pós-transcricionais também desempenham papel importante na regulação negativa da expressão do gene *LCT* após o desmame (LABRIE *et al.*, 2016; OH *et al.*, 2017; LESEVA *et al.*, 2018).

Em 2016, Labrie e colaboradores desenvolveram estudo que avaliou em conjunto as variações genéticas já conhecidas para o fenótipo LP, mapas de cromatina e testes funcionais em camundongos mutantes e em linhagens de células intestinais Caco-2. A partir dos resultados obtidos foi possível observar que modificações epigenéticas em elementos regulatórios influenciam a expressão diferencial do gene *LCT* em nível tecidual e de espécie, além de influenciar a variação na expressão gênica observada entre indivíduos (LABRIE *et al.*, 2016). Os autores demonstraram que o haplótipo que contém o alelo ancestral *–13910*C* (*–13910C>T*, rs4988235, gene *MCM6*) acumula metilações que silenciam a transcrição do gene *LCT* em mamíferos na fase adulta. Em contraste com o haplótipo ancestral, o haplótipo derivado (que contém o alelo *–13910*T*) exibe alterações relacionadas à idade que, por sua vez, contribuem para a manutenção da expressão gênica de *LCT* mesmo após o desmame (LABRIE *et al.*, 2016). Assim, a variação genética na posição *–13910* atua de forma a mediar o acúmulo (alelo *C*) ou perda (alelo *T*) de modificações epigenéticas em regiões potencializadoras da transcrição do gene *LCT* localizadas no gene *MCM6* (LABRIE *et al.*, 2016).

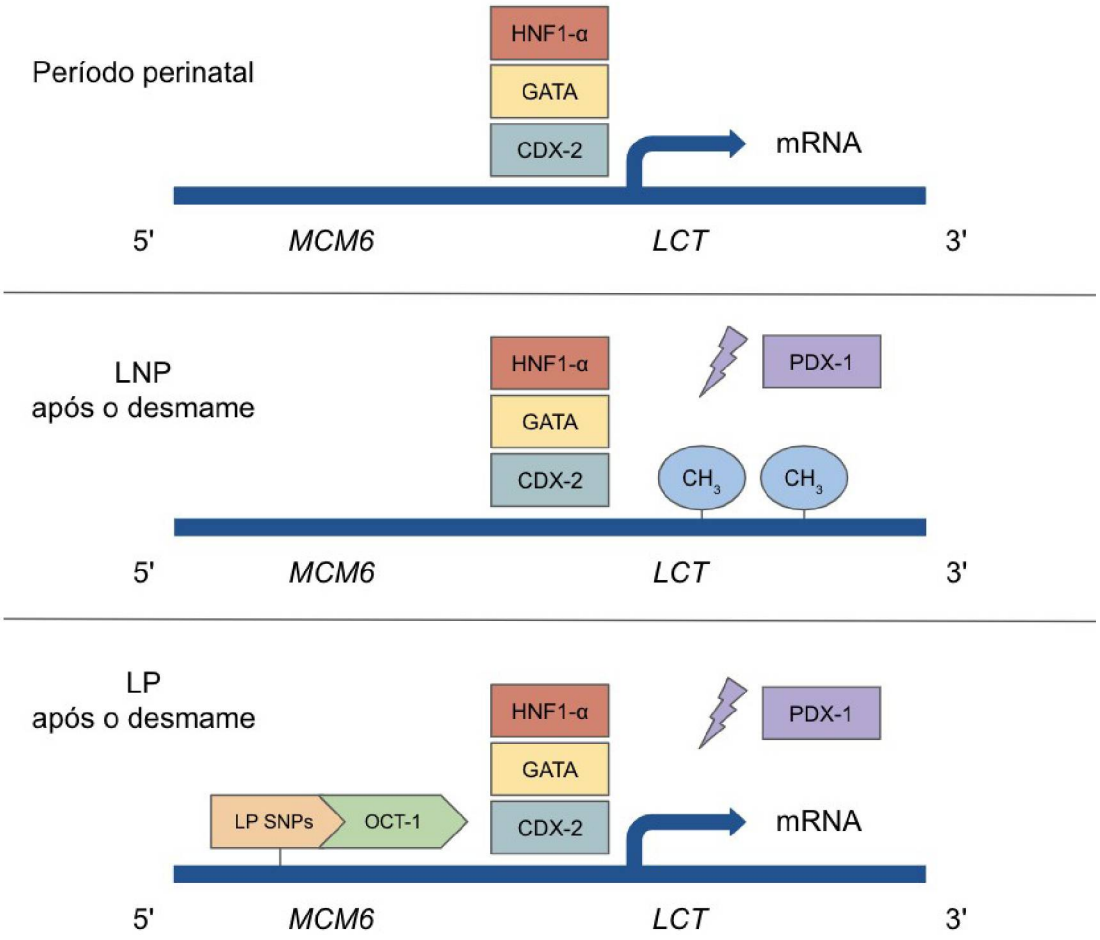
Oh e colaboradores (2017) demonstraram que assinaturas epigenéticas no gene *LCT* estão associadas à diferenças na transcrição do gene em enterócitos de camundongos. Os autores avaliaram perfis epigenéticos associados ao gradiente de transcrição do gene *LCT* que ocorre ao longo do eixo proximal-distal no intestino delgado e avaliaram a influência da idade e da exposição ambiental em assinaturas epigenéticas. Indicadores ambientais específicos para idade e fenótipo (como a exposição à lactose após o desmame) induzem modificações epigenéticas na ligação de CTCF (proteína que facilita interações entre sequências reguladoras

transcricionais ao alterar a estrutura da cromatina) em regiões que estão diretamente relacionadas à transcrição diferencial de mRNA no intestino. Dessa forma, a epigenética também pode contribuir para expressão diferencial do gene *LCT* de acordo com subtipos celulares e regiões intestinais distintas (OH *et al.*, 2017).

Corroborando os resultados de Labrie e colaboradores (2016) e Oh e colaboradores (2017), Leseva e colaboradores (2018), identificaram meQTLs (do inglês, *methylation quantitative trait loci*) no gene *LCT* em indivíduos de fenótipo LNP e LP e descreveram metilação diferencial na posição *–13910* do gene *MCM6.* Segundo os autores, o nível de metilação na região potenciadora (gene *MCM6*) da transcrição do gene *LCT* e em sítios CpG na região promotora do gene *LCT* apresentou maior poder de predição para os fenótipos LNP e LP do que os alelos descritos anteriormente no gene *MCM6*. Estes alelos promovem (fenótipo LP) ou reprimem (fenótipo LNP) a ligação de fatores de transcrição intestinais, como CDX-2, POU2F1, GATA4, GATA6 e HNF1-α (FIGURA 5), na região promotora do gene *LCT*. Entretanto, para a amostra avaliada, a variação na expressão desses fatores de transcrição não foi correlacionada com a determinação do fenótipo (LESEVA *et al.*, 2018).

Como visto no estudo anterior (LESEVA *et al.*, 2018), em alguns casos a presença da variação *–13910C>T* (rs4988235) não prediz estritamente o fenótipo LP. Devido a essa observação, Oliveira e colaboradores (2019) avaliaram a associação entre o polimorfismo rs140433552**CA>del*, localizado na região 3'UTR do gene *LCT*, e a determinação do fenótipo LNP, que pode levar ao desenvolvimento da intolerância à lactose. Foram avaliados 122 indivíduos de ancestralidade predominantemente europeia da região sul do Brasil, os SNPs *–13910C>T* (rs4988235) e rs140433552**CA>del* foram genotipados pela técnica de PCR-SSP (OLIVEIRA *et al.*, 2019). Nos indivíduos avaliados, o rs140433552**CA>del* não apresentou efeito claro na atribuição de fenótipo, dado que o genótipo *CA* foi observados em indivíduos de ambos os fenótipos (LNP e LP). Dessa forma, não foi observada associação entre a presença do polimorfismo e a determinação do fenótipo LNP (OLIVEIRA *et al.*, 2019).

FIGURA 5: MECANISMOS GENÉTICOS ASSOCIADOS À NÃO PERSISTÊNCIA (LNP) E PERSISTÊNCIA DA LACTASE (LP) EM HUMANOS



Os retângulos representam fatores de transcrição conhecidos por interagirem com o promotor do gene *LCT*. Os fatores de transcrição CDX-2, HNF1-α, GATA e OCT-1 atuam promovendo a transcrição de *LCT*, enquanto o fator PDX-1 atua reprimindo a expressão gênica. Os ovais identificados como "CH_3" representam metilações na região promotora do gene *LCT*, que também atuam reprimindo a expressão gênica. Os polimorfismos associados ao fenótipo LP, localizados no gene *MCM6*, estão representados em alaranjado. LNP: lactase não persistente; LP: lactase persistente; CDX-2: homeobox 2 do tipo caudal; HNF1-α: fator nuclear 1 dos hepatócitos; OCT-1: proteína 1 de ligação ao octâmero; PDX-1: homeobox 1 pancreático e duodenal. FONTE: adaptado de Anguita-Ruiz, Aguilera e Gil (2020).

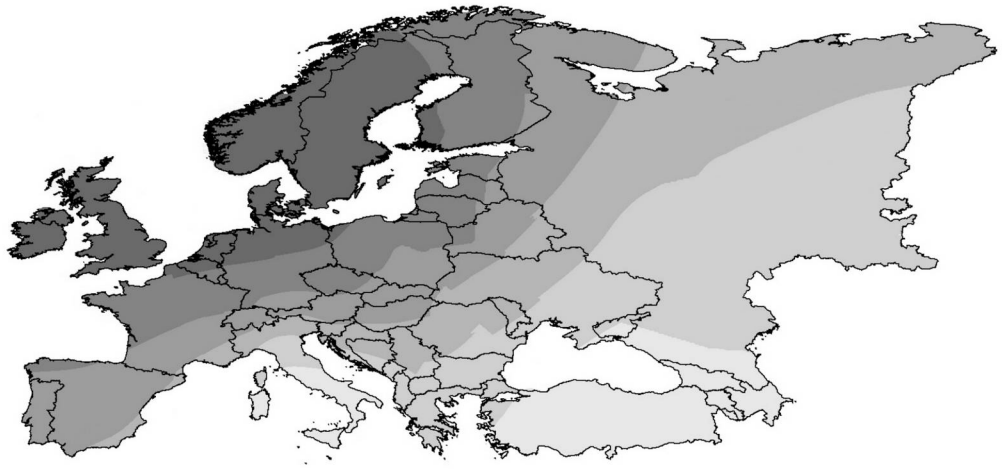
2.4 GENÉTICA DE POPULAÇÕES E ASPECTOS EVOLUTIVOS DA PERSISTÊNCIA DA LACTASE

O fenótipo LP é comum na Europa e em algumas populações da África e Oriente Médio. O padrão de distribuição dele, bem como a prevalência, estão

associados à cultura de pastoralismo e ordenha (SIMOONS, 1970; MCCRACKEN, 1971), que tem como característica principal o consumo de leite fresco e derivados.

Na Europa, em média 80% da população possui o fenótipo LP, sendo esse mais frequente entre os grupos pastoralistas da Finlândia, Irlanda e Grã-Bretanha (ITAN *et al.*, 2010). A frequência da característica diminui gradualmente do norte ao sul e do oeste ao leste do continente (FIGURA 6), sendo observada em 15% da população do sul da Itália (ITAN *et al.*, 2010). Já em Portugal, a frequência estimada do fenótipo LP é de 60%, condizente com a localização do país no continente (COELHO *et al.*, 2005).

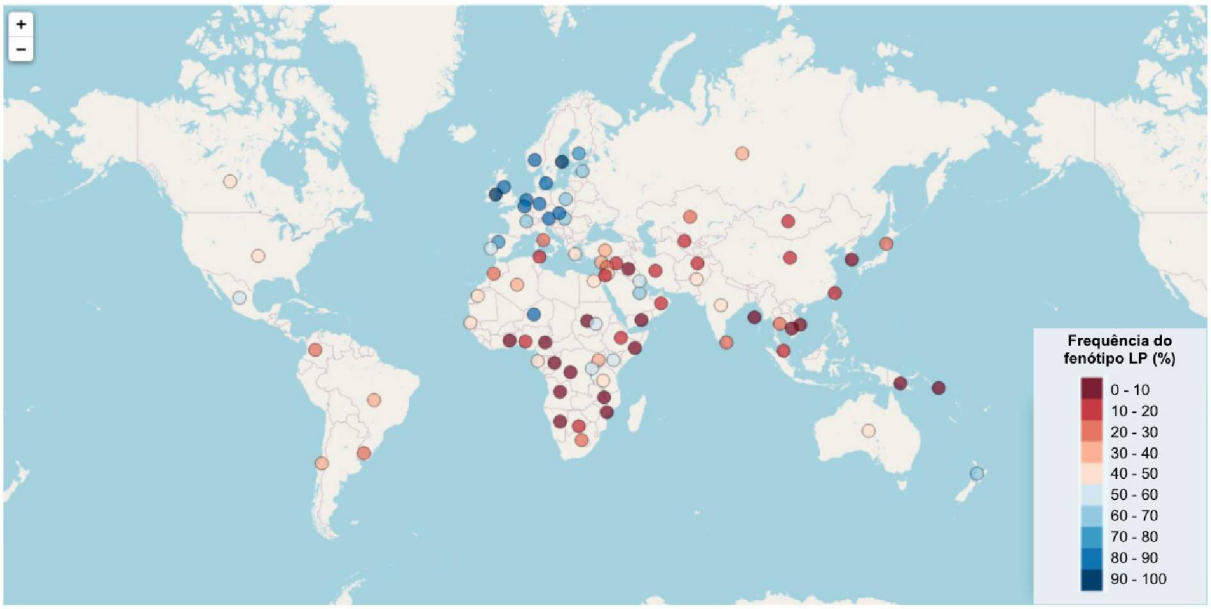
FIGURA 6: DISTRIBUIÇÃO FENOTÍPICA DA PERSISTÊNCIA DA LACTASE NA EUROPA



O mapa de interpolação foi gerado pelo programa ArcGIS 9.2 utilizando as frequências fenotípicas disponíveis no *Global Lactase Persistence Association Database*, publicado originalmente por Itan e colaboradores (2010). Quanto mais escuro o tom de cinza, maior a frequência do fenótipo lactase persistente na população. FONTE: Rocha (2012).

Na África e no Oriente Médio, o fenótipo LP é distribuído irregularmente (FIGURA 7), sendo muito comum nos grupos pastoralistas Fulani de Camarões, Beja do Sudão e beduínos da Arábia Saudita, em contraste com populações vizinhas não pastoralistas (INGRAM *et al.*, 2009; ITAN *et al.*, 2010). No ANEXO 1 (p. 160), ANEXO 2 (p. 162), ANEXO 3 (p. 164) e ANEXO 4 (p. 166) pode-se observar as frequências para o fenótipo LP em diversas populações da África, Europa, Oriente Médio, Ásia e Australásia, respectivamente.

FIGURA 7: FREQUÊNCIA E DISTRIBUIÇÃO FENOTÍPICA MUNDIAL DA PERSISTÊNCIA DA LACTASE



Mapa contendo as frequências fenotípicas médias da persistência da lactase em populações modernas. Inclui dados disponíveis no *Global Lactase Persistence Association Database*, publicado originalmente por Itan e colaboradores (2010), e dados reportados em outros estudos até 20 de julho de 2020. FONTE: adaptado de Anguita-Ruiz, Aguilera e Gil (2020).

Para explicar a variação no padrão de distribuição do fenótipo LP, foram propostas diversas hipóteses. A hipótese cultural-histórica (do inglês, *cultural-historical*), melhor aceita atualmente, dita que o aumento na frequência de alelos determinantes do fenótipo LP se deve ao fato de que a população pré-moderna, no período neolítico, teria alterado o seu estilo de vida, incorporando o pastoralismo e ordenha como métodos de subsistência, e consequentemente, adicionando o leite à sua dieta. Dessa forma, indivíduos LP (fenótipo derivado) estariam aptos a desfrutar dos benefícios do consumo do leite, enquanto indivíduos LNP (fenótipo ancestral) ao consumirem laticínios poderiam desenvolver os sintomas característicos da intolerância à lactose (SIMOONS, 1970; McCRACKEN, 1971). Em síntese, os alelos descritos anteriormente (tópico 2.1 deste trabalho) teriam aumentado de frequência devido à seleção natural em populações que praticavam o pastoralismo e ordenha, sendo um exemplo de coevolução gene-cultura (BEJA-PEREIRA *et al.*, 2003). Nesse cenário, a evolução do fenótipo LP pode ser considerada um exemplo de construção de nicho humano, no qual o ambiente seletivo leva ao aumento de frequência alélica, genotípica e fenotípica influenciado por uma prática cultural (LALAND *et al.*,

2010; GERBAULT *et al.*, 2011). Na ausência da cultura de pastoralismo e ordenha associada ao consumo de leite fresco, não seriam observadas diferenças de valor adaptativo entre indivíduos de fenótipo LP e LNP (DURHAM, 1991).

Complementando a hipótese cultural-histórica, outras hipóteses sobre a dependência humana do leite foram propostas: pressões seletivas diversas teriam atuado sobre populações habitantes de regiões de clima árido, nas quais o leite seria uma fonte limpa de água (COOK e ALTORKI, 1975); e em populações habitantes de latitudes altas, na qual a ingestão de leite por indivíduos de fenótipo LP seria facilitadora da absorção de cálcio, que, na deficiência de vitamina D devido à baixa radiação solar, reduziria os riscos de raquitismo e osteomalácia (FLATZ e ROTTHAUWE, 1973).

Em contrapartida, a hipótese causa reversa (do inglês, *reverse-cause*) sugere que as populações pré-modernas já possuíam os alelos determinantes do fenótipo LP em alta frequência e isso fez com que fosse possível a prática do pastoralismo, ordenha e consumo de leite fresco. Assim, a variação de frequência do fenótipo LP entre populações humanas teria surgido por deriva genética anteriormente às alterações nos padrões de subsistência ocorridas durante o período neolítico (MCCRACKEN, 1971). Contudo, estudos envolvendo a avaliação de DNA ancestral suportam o proposto pela hipótese cultural-histórica.

Burger e colaboradores (2007) avaliaram amostras de DNA de esqueletos provenientes de sítios arqueológicos dos períodos meso e neolítico, distribuídos entre a Europa Central, Nordeste e Sudeste. Foi investigada a presença das variações *–13910C>T* e *–22018G>A* em uma amostra do período mesolítico (2.276 ± 116 aC), oito amostras do período neolítico (entre 5.800 e 5.000 aC) e uma amostra do período medieval (400 - 600 anos depois de Cristo), as quais foram datadas por radiocarbono. Foram identificados nove indivíduos homozigotos para o alelo *–13910*C*, sendo referentes aos períodos meso e neolítico, e o indivíduo do período medieval heterozigoto (*CT*) para a posição. Para a variação *–22018G>A*, todas as amostras, com exceção da medieval (*AA*), são homozigotas para o alelo *–22018*G*. A partir disso, foi estimada a frequência do alelo *–13910*T*. Tratando os oito indivíduos do período neolítico como membros de uma única população, a frequência máxima estimada seria de 17% (confiança de 95%) e 25% (confiança de 99%), sendo que esses valores podem ter sido muito inferiores. Assumindo o

equilíbrio de Hardy-Weinberg e que o alelo *–13910*T* seria a única causa do fenótipo LP em populações europeias pré-modernas, as frequências fenotípicas seriam de 31% e 44% para intervalos de confiança de 95% e 99%, respectivamente (BURGER *et al.*, 2007). Ainda, em outro estudo, foram avaliadas amostras de DNA coletadas em um cemitério medieval na cidade de Dalheim, Alemanha (KRÜTTLI *et al.*, 2014). Foi realizada a genotipagem da variante *–13910C>T* em 18 amostras de DNA extraídas da dentina. Desses indivíduos, 13 (72%) apresentariam o fenótipo LP, sendo 44% de genótipo *CT* e 28% de genótipo *TT*. Assim, foi demonstrado que a frequência do fenótipo LP na população medieval alemã (72%) é similar à observada na população moderna alemã e austríaca — 71% e 80%, respectivamente —, sugerindo que o alelo *–13910*T* atingiu sua atual frequência de acordo com a hipótese cultural-histórica (KRÜTTLI *et al.*, 2014).

Embora evidências arqueológicas de criação de gado, ovelhas e cabras para ordenha nas populações europeias pré-modernas sejam escassas, dados zooarqueológicos e biomoleculares sugerem que o início da criação de gado para a produção de leite tenha sido concomitante ao início do período neolítico, cerca de 5.000 anos antes de Cristo (aC) (BEJA-PEREIRA *et al.*, 2003; VIGNE e HELMER, 2007; EVERSHED *et al.*, 2008). Nas Américas, evidências zooarqueológicas sugerem a ocorrência da domesticação de mamíferos há cerca de 7.000 - 6.000 anos, esses mamíferos seriam os ancestrais de lhamas, alpacas, vicunhas e guanacos modernos (WHEELER, 1995; KADWELL et al., 2001). Apesar da criação destes animais, não há evidências arqueológicas ou culturais que sugiram o consumo de leite fresco e derivados por povos nativos americanos até a chegada de povos europeus no final do século XV (GADE, 1999) — o gado, por exemplo, foi introduzido nas Américas apenas em 1493 (PRIMO, 1992).

Bersaglieri e colaboradores (2004) avaliaram as frequências populacionais do haplótipo no qual o alelo *–13910*T*, de origem europeia, está inserido. Os resultados obtidos sustentam evidências para a hipótese cultural-histórica, indicando duas assinaturas de seleção recente: frequências alélicas discrepantes entre as populações — diferentes pressões de seleção causaram o aumento da frequência dos alelos em algumas, mas não em todas as populações; e a ocorrência de um haplótipo longo comum a diversas populações — uma vez que houve o aumento rápido de frequência alélica, não se passou tempo suficiente para que a

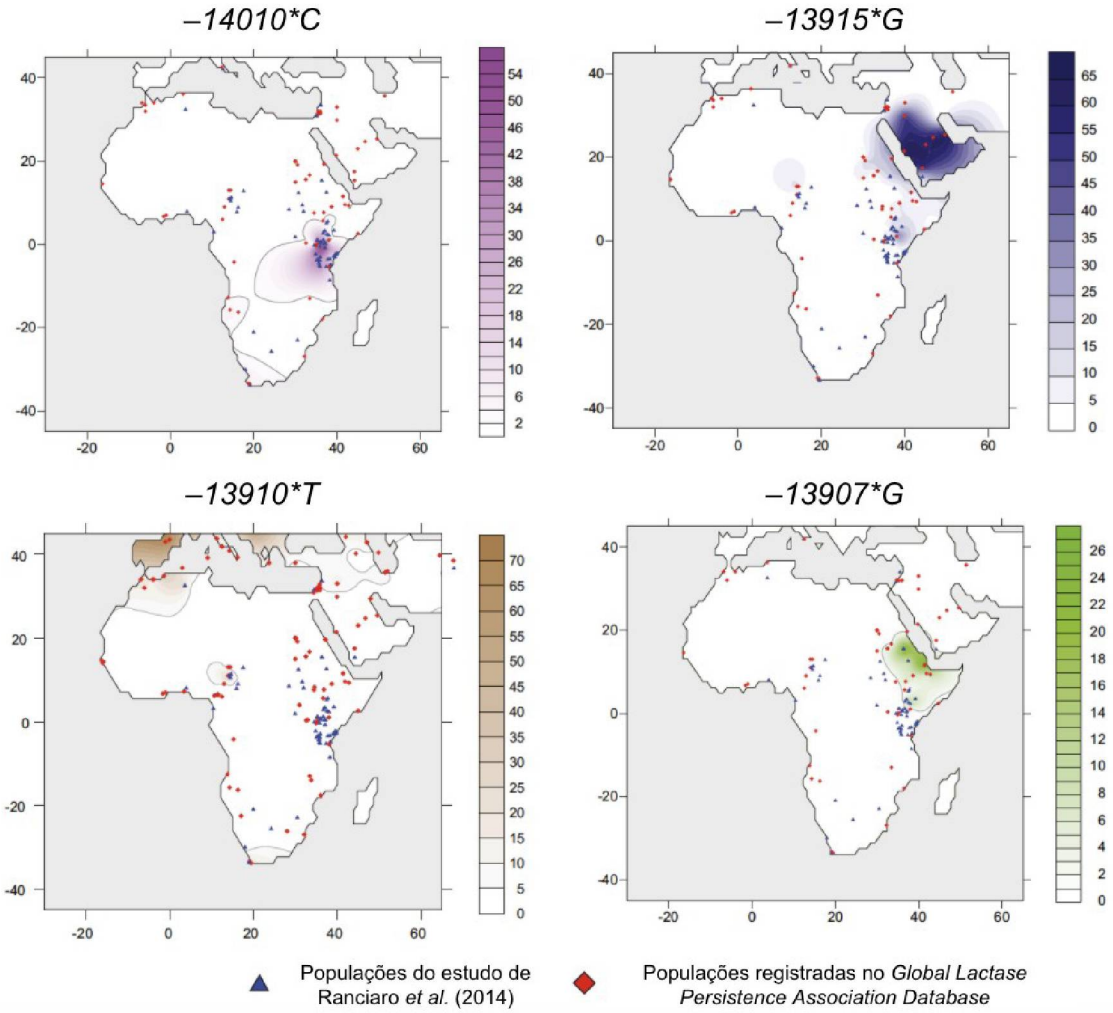
recombinação diminuísse o desequilíbrio de ligação entre os alelos desse haplótipo. Ainda, foi possível inferir que o alelo *–13910*T* aumentou de frequência entre 2.188 e 20.650 anos atrás (BERSAGLIERI *et al.*, 2004). Estudos de variação em microssatélites, reduziram a amplitude do intervalo, sugerindo que o alelo surgiu entre 7.450 e 12.300 anos atrás (COELHO *et al.*, 2005), consistente com o início do período neolítico (cerca de 5.000 anos aC) no qual foi introduzida a cultura de pastoralismo e ordenha, bem como a criação de gado, na Europa (BEJA-PEREIRA *et al.*, 2003).

Quanto à dispersão do alelo *–13910*T* em populações europeias, Itan e colaboradores (2009) empregaram método baseado no cálculo bayesiano aproximado, que permitiu estimar parâmetros evolutivos e demográficos envolvidos na dispersão do fenótipo LP pela Europa. As simulações realizadas indicaram que a disseminação do alelo *–13910*T* foi moldada pela expansão, entre 6.256 e 8.683 anos atrás, de populações que praticavam o pastoralismo e ordenha, vindas da região entre a península Balcânica e a Europa Central (ITAN *et al.*, 2009). Os autores propuseram que, após o início do período neolítico no sudeste da Europa e a crescente importância da criação de gado e uso de laticínios, a seleção natural começou a atuar sobre os alelos associados ao fenótipo LP em populações ao norte dos Balcãs. Após o lento aumento inicial da frequência do fenótipo LP nessas populações e o início da cultura de pastoralismo e ordenha na Europa Central (há cerca de 7.500 anos), a frequência do alelo *–13910*T* e do fenótipo LP teriam aumentado rapidamente em um processo coevolutivo gene-cultura, frente à expansão demográfica, que levou ao estabelecimento de economias leiteiras durante o período neolítico na Europa Central (há cerca de 6.500 anos) — favorecendo a hipótese cultural-histórica. A combinação de fatores demográficos com a seleção teria levado à dispersão da variação associada ao fenótipo LP na Europa, sendo o alelo *–13910*T*, aparentemente, o único responsável pela determinação do fenótipo em populações europeias (ITAN *et al.*, 2009).

Em contraste com a Europa, na África são observados diversos alelos (descritos no tópico 2.1 deste trabalho) associados ao fenótipo LP em diferentes populações (FIGURA 8). O maior número de alelos reflete a maior diversidade genética das populações africanas, sob as quais a seleção pôde atuar de formas distintas de acordo com o surgimento e disseminação da cultura de pastoralismo e

ordenha, como sugerido pela hipótese cultural-histórica apresentada anteriormente (TISHKOFF *et al.*, 2007; RANCIARO *et al.*, 2014).

FIGURA 8: VARIAÇÃO NA FREQUÊNCIA E DISTRIBUIÇÃO DOS ALELOS *–14010*C*, *–13910*T*, *–13915*G* E *–13907*G* ASSOCIADOS AO FENÓTIPO LACTASE PERSISTENTE EM POPULAÇÕES AFRICANAS, DA PENÍNSULA ARÁBICA E DO ORIENTE MÉDIO



Mapas de contorno para a visualização de frequências e distribuição dos alelos *–14010*C*, *–13910*T*, *–13915*G* e *–13907*G* associados ao fenótipo lactase persistente. As barras ao lado direito de cada quadro mostram as frequências alélicas correspondentes à escala de cor. Os triângulos azuis representam a localização das populações amostradas no estudo de Ranciaro e colaboradores (2014). Os losangos vermelhos representam a localização das populações disponíveis no *Global Lactase Persistence Association Database*, publicado originalmente por Itan e colaboradores (2010). FONTE: Adaptado de Ranciaro *et al.* (2014).

Ranciaro e colaboradores (2014) sequenciaram os íntrons 9 (1.297 bp) e 13 (3.218 bp) do gene *MCM6* e aproximadamente 2 kb da região promotora do gene *LCT* em 819 amostras de 63 populações africanas e 154 amostras não africanas de

9 populações distintas. Foram identificados 68 SNPs nessas regiões e a maioria das variações ocorreu em frequência inferior a 5%. Os dados gerados mostraram distribuição geográfica restrita dos alelos já conhecidos *–14010*C*, *–13910*T*, *–13915*G* e *–13907*G* (FIGURA 8) (RANCIARO *et al.*, 2014).

O alelo *–13910*T* foi observado em populações pastoralistas da África Central e do Norte, porém é ausente em outras populações da África sul-saariana (FIGURA 8). A distribuição do alelo e a similaridade entre os contextos haplotípicos em populações africanas e não africanas sugere que a variante *–13910*T* tenha sido inserida através de fluxo gênico entre populações de origem europeia e populações da África Central e do Norte (BERSAGLIERI, *et al.*, 2004; TISHKOFF *et al.*, 2007; TISHKOFF *et al.*, 2009; RANCIARO *et al.*, 2014; VICENTE *et al.*, 2019). As principais ondas migratórias humanas (dentro e de dentro para fora da África) percorreram a porção oriental do continente, o que explica a presença de três alelos distintos nessa região (CAMPBELL e TISHKOFF, 2008; TISHKOFF *et al.*, 2009). O alelo *–13915*G* é mais comum na península arábica e no Oriente Médio, dessa forma, provavelmente, a variante foi originada nessa região (FIGURA 8). O alelo *–13915*G* foi observado em populações pastoralistas africanas de ancestralidade árabe de contexto haplotípico comum ao observado nas populações da península arábica e do Oriente Médio, o que reforça o surgimento do alelo nessa região e posterior inserção dele em populações africanas (ENATTAH *et al.*, 2007; RANCIARO *et al.*, 2014).

O alelo *–13907*G* foi identificado em populações falantes de línguas afro-asiáticas do ramo cuchítico da Etiópia e das regiões norte do Sudão e do Quênia (FIGURA 8). Ausente em outras regiões avaliadas, provavelmente, o alelo *–13907*G* foi originado em populações falantes de línguas do ramo cuchítico na região leste da Etiópia que migraram para as regiões norte do Quênia e Sudão nos últimos 5.000 anos (INGRAM *et al.*, 2007; ENATTAH *et al.*, 2008; TISHKOFF *et al.*, 2007; RANCIARO *et al.*, 2014).

O alelo *–14010*C* foi identificado em populações da Tanzânia, Quênia e na região sul do continente africano (populações de falantes da língua Bantu da região sudoeste da Angola e em grupos de etnia Xhosa da África do Sul) (FIGURA 8). A ausência do alelo na região sul do Sudão e sua idade — possivelmente o alelo surgiu entre 2.700 e 6.800 anos atrás — sugerem que a variante *–14010*C* tenha

sido originada em populações falantes de línguas afro-asiáticas e, posteriormente, foi introduzida em populações falantes de línguas nilo-saarianas (TISHKOFF *et al.*, 2007; COELHO *et al.*, 2009; TORNIAINEN *et al.*, 2009; RANCIARO *et al.*, 2014). O alelo *–14010*C* foi observado em maior frequência na Tanzânia, provavelmente, foi inserido na região através de populações falantes de línguas afro-asiáticas do ramo cuchítico originárias do sul do país (TISHKOFF *et al.*, 2007; RANCIARO *et al.*, 2014).

Na região sul do continente africano, o alelo *–14010*C* foi identificado na população caçadora-coletora San e na população pastoralista de etnia Xhosa falante da língua Bantu, que possuem ancestralidade San (FIGURA 8). A população Xhosa possui o alelo *–14010*C* inserido em contexto haplotípico comum ao observado nas populações da Tanzânia e do Quênia, o que sugere que o alelo tenha sido inserido mais recentemente, através de fluxo gênico (TISHKOFF *et al.*, 2009; RANCIARO *et al.*, 2014). Outra possibilidade para a introdução da variante *–14010*C* em populações da região sul diz respeito ao movimento de expansão Bantu (RANCIARO *et al.*, 2014). A população Bantu, original da região sudoeste da Nigéria e noroeste de Camarões, iniciou o movimento de expansão territorial, linguística e cultural em direção ao sul do continente, entre 10.000 e 4.000 anos atrás, dividindo-se em duas principais rotas: ocidental, ao longo da costa atlântica; e oriental, em direção às florestas equatoriais tropicais e aos Grandes Lagos africanos (PHILLIPSON, 1993; NEWMAN, 1995). Podem ter ocorrido convergências entre as rotas em períodos distintos, porém a extensão dos encontros é desconhecida. Assim, a população Bantu gradualmente ocupou a África Central, do Sul e porção da África Oriental, estendendo-se por diversas regiões (PHILLIPSON, 1993). A variante *–14010*C* pode ter sido inserida em populações sul africanas há cerca de 2.400 anos por fluxo gênico. Num primeiro momento, o alelo foi introduzido na população Bantu a partir do contato com populações pastoralistas, então, o alelo *–14010*C* foi disseminado devido ao movimento de expansão, no qual a população migrou em sentido sul (RANCIARO *et al.*, 2014).

Os diferentes haplótipos observados entre populações africanas (TISHKOFF *et al.*, 2007; RANCIARO *et al.*, 2014) e europeias (ENATTAH *et al.*, 2002; MULCARE *et al.*, 2004), os diversos alelos associados ao fenótipo LP descritos em populações africanas (TISHKOFF *et al.*, 2007; INGRAM *et al.*, 2007) e o surgimento independente deles — em diferentes ambientes e em tempos distintos (RANCIARO

et al., 2014) — caracterizam evolução convergente. Assim, a persistência da lactase em seres humanos, além de ser um exemplo de coevolução gene-cultura (de acordo com a hipótese cultural-histórica), é um exemplo de evolução convergente (TISHKOFF *et al.*, 2007; RANCIARO *et al.*, 2014).

2.5 COMPOSIÇÃO GENÉTICA DA POPULAÇÃO PAN-AMERICANA

O processo de formação da população miscigenada pan-americana pode ser segmentado em três principais ondas migratórias — povoamento, colonização e miscigenação —, que refletem a ancestralidade genômica dessas populações (SANTOS, 2002; SALZANO e BORTOLINI, 2005; PENA *et al.*, 2011). Neste trabalho, utilizamos o termo "populações pan-americanas" para nos referirmos às populações das Américas (América Central, do Sul e do Norte).

O povoamento das Américas ocorreu há cerca de 20 mil anos, quando povos de origem asiática, vindos das Montanhas Altai (região sul da Sibéria), teriam acessado o continente através do Estreito de Bering, que liga a região nordeste da Sibéria ao Alasca (GOEBEL, WATERS e O'ROURKE, 2008; DILLEHAY, 2009; REICH *et al.*, 2012; MENDES *et al.*, 2020). Os povos nativos americanos são descendentes, principalmente, de um único grupo ancestral asiático que povoou o continente americano. As divisões populacionais sequenciais e pouco fluxo gênico após divergência deram origem aos povos nativos da América Central e do Sul — com grupos populacionais altamente diferenciados, como mesoamericano, andino e amazônico (SALZANO, 2011; REICH *et al.*, 2012; MENDES *et al.*, 2020) — e da América do Norte, como os povos falantes das línguas Esquimó-Aleuta no Ártico e Na-Dene (grupos Chipewyan) no Canadá (GREENBERG, TURNER e ZEGURA, 1986; REICH *et al.*, 2012). Ainda, estudos que avaliaram a morfologia e a ancestralidade de populações nativas da América Central e do Sul sugerem que ocorreu mais de uma onda migratória que deu origem a essas populações. Eventos de miscigenação entre populações nativas e australasianas ocorreram, provavelmente, antes da população que povoou a América Central e do Sul chegar à Amazônia, assim, populações ancestrais já possuíam porção de ancestralidade australasiana (GONZÁLEZ-JOSÉ *et al.*, 2005; SKOGLUND *et al.*, 2015).

Quando os europeus chegaram às Américas, no final do século XV, a população nativa americana era estimada em 43 milhões de pessoas (BETHELL, 1997; VERANO e UBELAKER, 1992; SALZANO e BORTOLINI, 2005). Enquanto os portugueses deram início à invasão e exploração da América do Sul pela costa do Oceano Atlântico, os espanhóis acessaram as Américas pelas ilhas da América Central; no continente, exploraram a América Central, a América do Sul (pela costa do Oceano Pacífico) e a bacia amazônica. No processo, europeus exterminaram impérios nativos americanos pré-estabelecidos, como o Asteca e o Inca, além de grupos menores (BETHELL, 1997; WEHLING e WEHLING, 1994; SALZANO e BORTOLINI, 2005).

Mais tarde, entre os séculos XVI e meados do século XIX, os europeus trouxeram cerca de 9 milhões de africanos escravizados para as Américas — dos quais estima-se que 4 milhões foram para o Brasil (IBGE, 2000) — através do tráfico negreiro transatlântico (READER, 1998; SOUZA, 2008). Embora não haja registros históricos oficiais acerca da origem subcontinental dos povos africanos, estudos genéticos anteriores demonstraram que a ancestralidade centro-oeste africana é a mais prevalente nas Américas (GOUVEIA *et al.*, 2020). Ainda, é possível observar prevalência das ancestralidades centro-oeste e oeste africanas em latitudes ao norte (72% em afro-caribenhos de Barbados; 57% nas populações do nordeste do Brasil; e 56% em afro-peruanos e afro-americanos dos Estados Unidos), enquanto em latitudes ao sul as ancestralidades sul-leste africanas são mais prevalentes (44% nas populações do sul e 54% do sudeste do Brasil) (GOUVEIA *et al.*, 2020). Além da colonização espanhola e portuguesa, britânicos, holandeses e franceses — que também empregavam trabalho escravo — ocuparam e exploraram o Caribe, a região norte da América do Sul e as ilhas ao sul do continente entre os séculos XVI e XIX. Entretanto, somente no século XVII, a América do Norte foi colonizada principalmente por britânicos (BETHELL, 1997; SALZANO e BORTOLINI, 2005).

A imigração europeia teve escala e influência variadas na América do Sul e Central. Os países de língua espanhola também receberam italianos, portugueses, alemães, franceses e ingleses. Enquanto o fluxo de imigrantes europeus foi mais intenso no sul da América do Sul, os países caribenhos receberam mais imigrantes africanos do que europeus (SALZANO e BORTOLINI, 2005). No Brasil, a migração europeia foi incentivada pelo governo principalmente após a abolição da escravatura

(1888), em um processo conhecido como "branqueamento" da população brasileira (SANTOS, 2002). A partir dos anos 1890, o governo brasileiro amparou a vinda de imigrantes, majoritariamente italianos. Dessa forma, no final do século XIX, o Brasil recebeu cerca de 6 milhões de imigrantes italianos, portugueses, espanhóis e, em menor escala, alemães, poloneses e japoneses (IBGE, 2000; SANTOS, 2002). Além disso, no século XX, povos do Leste Asiático (principalmente chineses e japoneses), indianos e paquistaneses migraram para a América Central, região norte da América do Sul e Brasil (SALZANO e BORTOLINI, 2005).

A miscigenação resultante alterou drasticamente a composição genética da população pan-americana, que em sua maioria possui origem tri-híbrida: nativa americana, africana e europeia, com proporções de ancestralidade variando entre os países e dentro deles de acordo com a região (PENA *et al.*, 2011; RUIZ-LINARES *et al.*, 2014). O Brasil registra a maior ancestralidade europeia (78,6%), seguido pela Colômbia, com proporções de 61,2%; entretanto, a ancestralidade nativa americana é observada em 12,1% e 29,2%, respectivamente. No Chile, a ancestralidade europeia (46,2%) e nativa americana (49,3%) são relativamente similares. O oposto é observado no Peru e México que possuem ancestralidade nativa americana predominante, com proporções de 64,8% e 57,5%, respectivamente; enquanto a ancestralidade europeia é observada em 30,6% e 37,7%, respectivamente. Já a ancestralidade africana é observada em proporções menores, 9,6% e 9,3% na Colômbia e no Brasil (em comunidades quilombolas, por exemplo, a proporção pode variar), respectivamente, e em menos de 5% no México, Peru e Chile (RUIZ-LINARES *et al.*, 2014; ADHIKARI *et al.*, 2016). Para os Estados Unidos, estimativas de ancestralidade levando em conta o genoma completo (do inglês, *genome-wide*) indicam maior ancestralidade europeia na população euro-americana (98,6%), seguida pelas populações latina e afro-americana com proporções de 65,1% e 24%, respectivamente. A ancestralidade nativa americana é predominante na população latina (18,0%), com proporções reduzidas em afro-americanos (0,80%) e euro-americanos (0,18%). Já a ancestralidade africana é observada nas proporções de 73,2%, 6,2% e 0,19% nas populações afro-americana, latina e euro-americana, respectivamente (BRYC *et al.*, 2015).

2.5.1 A comunidade quilombola de Sertão do Valongo

No período anterior à abolição da escravatura (1888), era comum que africanos escravizados e seus descendentes formassem comunidades — como maneira de resistência ao regime escravocrata — denominadas Quilombos, termo de origem Bantu que significa acampamento ou fortaleza. Atualmente, no Brasil, os grupos que residem em territórios quilombolas são chamados de comunidades quilombolas (CQ) (IBGE, 2000; RAGGIO *et al.*, 2008; LEITE, 2008).

A comunidade quilombola de Sertão do Valongo, na região rural de Porto Belo em Santa Catarina, foi fundada em 1880 por quatro casais (geração I), sendo sete indivíduos de origem africana e apenas um indivíduo de origem inglesa (SOUZA e CULPI, 1992). A partir de dados coletados por Souza e Culpi entre os anos de 1984 e 1985, foi possível gerar um heredograma (ANEXO 5, p. 167) que contém informações de 151 indivíduos (43 já eram falecidos) dos quais 81 viviam na comunidade (36 homens e 45 mulheres; correspondente a 75% dos indivíduos vivos) (SOUZA e CULPI, 1992) — dados atuais não estão disponíveis.

Dos 20 casamentos avaliados na época, 17 (85%) eram consanguíneos. O coeficiente de endocruzamento (F) foi calculado para cada núcleo do heredograma que possuísse pelo menos um ancestral em comum; ponderando os valores de F para o número total de indivíduos avaliados, foi obtido o coeficiente médio de endocruzamento ($F_{MÉDIO}$). Dessa forma, $F_{MÉDIO}$ = 0,06906 (SOUZA e CULPI, 1992), o valor é superior ao estimado para o estado de Santa Catarina e para o Brasil, 0,00032 (DE BASSI, 1983) e 0,00088 (FREIRE-MAIA, 1990), respectivamente. Também foi obtido o coeficiente médio de endocruzamento para cada uma das gerações retratadas no heredograma, são eles: $F_{MÉDIO}$ = 0, para as gerações I e II; $F_{MÉDIO}$ = 0,01171, para a geração III; $F_{MÉDIO}$ = 0,05450, para a geração IV; e $F_{MÉDIO}$ = 0,06882, para a geração V (SOUZA e CULPI, 1992). Segundo a literatura, valores altos de F são esperados em populações isoladas, enquanto em populações panmíticas os valores observados costumam ser menores. O valor de F em populações isoladas tende a ser maior em gerações mais recentes devido à ocorrência de consanguinidade múltipla, já em populações não isoladas o valor de F costuma reduzir ao decorrer das gerações, justamente por receber indivíduos de outros grupos — migração e fluxo gênico entre populações distintas. Na CQ Sertão

do Valongo, mesmo com a inserção de dois imigrantes na geração IV (IV-1 e IV-21), não foi observada redução no valor de $F_{MÉDIO}$ (SOUZA e CULPI, 1992). Além desses, outros parâmetros genéticos foram obtidos: tamanho efetivo da população, N_e = 15 indivíduos (18,5%); tamanho da população reprodutora, N_r = 28 indivíduos (34,6%); e migração da população reprodutora, m_e = 2 indivíduos (IV-1 e IV-21), correspondente a 2,5% da população total, tendo ambos contribuído para a geração V, representando 7,1% da população reprodutora (SOUZA e CULPI, 1992).

A endogamia observada por Souza e Culpi (1992) através do alto coeficiente médio de endocruzamento possui diversas causas. A principal delas diz respeito às práticas religiosas e segregação cultural: a comunidade quilombola de Sertão do Valongo pratica casamentos entre seus membros, pois foi convertida à Igreja Adventista no início do século XX, na qual casamentos entre indivíduos do mesmo grupo são estimulados. Ainda, as populações vizinhas são compostas tradicionalmente por católicos de origem alemã, italiana e portuguesa, o que auxilia tanto a segregação cultural quanto racial da comunidade (BORBA, 2016; SOUZA e CULPI, 1992). Beiguelman (1994, p. 301) define um isolado populacional como "um conjunto humano que está separado dos outros por alguma barreira, seja ela geográfica, política, sócio-econômica, religiosa ou cultural, que impede ou dificulta a troca de genes com outro conjunto". Assim, a combinação dos fatores expostos anteriormente com a dificuldade de troca de genes caracteriza a CQ Sertão do Valongo como uma população isolada — desconsiderando-se o grau de isolamento —, com alto impacto da deriva genética, devida ao efeito fundador, sobre a população, o que pode levar a fixação ou eliminação de alelos (SOUZA e CULPI, 1992).

Em 2005, Souza e Culpi avaliaram 14 *loci* (*ABO*, *FY*, *P*, *RHD-RHCE*, *GPA-GPB*, *KEL*, *HBB*, *HP*, *TF*, *ALB*, *ESD*, *CA2*, *BCHE* e *CHE2*) em 49 indivíduos da CQ Sertão do Valongo (SOUZA e CULPI, 2005). Posteriormente, aplicaram um modelo de miscigenação no qual compararam as frequências alélicas obtidas com as frequências correspondentes em africanos da região sul-saariana e em europeus. A população em questão teve ancestralidade africana estimada em 97,33% (com desvio calculado de ± 10,41%, podendo variar entre 86,92% a 100%) e europeia de 2,67% (com desvio calculado de ± 10,41%, podendo variar entre 0% a 13,08%), com o erro quadrático médio de 27,07%, que representa a proporção de variação na

frequência alélica não explicada pelo modelo (SOUZA e CULPI, 2005). Posteriormente, Luizon (2007) desenvolveu um estudo sobre marcadores informativos de ancestralidade (AIMs, do inglês, *ancestry informative markers*) — polimorfismos de nucleotídeo único que possuem alelos significativamente mais comuns em determinadas populações do que em outras (JOBLING *et al.*, 2013) — no qual avaliou os marcadores *FY*, *RB*, *LPL*, *AT3*, *Sb19.3*, *APO*, *PV92* e *CYP1A1*2C* em populações de comunidades quilombolas da Bahia e em 25 indivíduos da população de Sertão do Valongo. Uma vez que está presente em maior frequência nas populações nativas brasileiras, o alelo *CYP1A1*2C* é um excelente AIM para distinguir populações nativas de europeias e asiáticas. Assim, a ausência do alelo na população da CQ Sertão do Valongo está de acordo com o histórico de fundação da população descrito por Souza e Culpi (1992), não havendo contribuição nativa americana ao *pool* gênico dessa população (LUIZON, 2007). A partir da avaliação dos marcadores citados acima, foi possível estimar que a população da CQ Sertão do Valongo possui 68,1% ± 0,4% e 31,9% ± 0,4% de ancestralidade africana e europeia, respectivamente (LUIZON, 2007).

A avaliação genética de populações isoladas, como a comunidade quilombola de Sertão do Valongo, submetidas ao efeito fundador e ao baixo fluxo gênico recente, pode ampliar significativamente o conhecimento sobre a persistência da lactase em populações afro-brasileiras, dado que essas condições geram divergências fenotípicas e genotípicas, não observadas em populações não isoladas. O tamanho reduzido da população, a baixa taxa de migração e a alta taxa de endocruzamento descritos por Souza e Culpi (1992; 2005) são fatores que levam a ocorrência de fenômenos genéticos interessantes de serem discutidos, como a baixa diversidade genética, alto desequilíbrio de ligação e aumento da frequência de homozigotos. Não há estudos disponíveis relacionados ao fenótipo LP na população da CQ Sertão do Valongo, assim como, não há estudos sobre o tema nas demais populações quilombolas brasileiras.

2.6 O FENÓTIPO LACTASE PERSISTENTE EM POPULAÇÕES PAN-AMERICANAS

Como descrito no tópico 2.1 deste trabalho, já foram identificados diversos SNPs associados ao fenótipo LP, dos quais *–13910C>T* (rs4988235), *–13915T>G* (rs41380347), *–14009T>G* (rs869051967), *–14010G>C* (rs145946881) e *–14011C>T* (rs4988233) possuem validação funcional e tiveram o impacto funcional determinado. Entretanto, existem poucos estudos sobre variantes associadas ao fenótipo LP em populações pan-americanas. Os estudos existentes possuem limitações, e — apesar das populações em questão possuírem componente genético nativo americano, africano e europeu — analisam, principalmente, a variação de origem europeia *–13910C>T*.

Em 1972, Duncan e Scott avaliaram indivíduos esquimós de ancestralidade nativa americana (n = 36) e indivíduos de ancestralidade europeia (n = 16) do Alasca, Estados Unidos. Ambos os grupos tiveram sangue coletado antes e após a ingestão de 50 g de lactose dissolvida em 350 ml de água. A tolerância à lactose foi testada medindo a concentração de glicose no sangue antes e após a ingestão da solução. Dos indivíduos nativos americanos, 92% apresentaram aumento menor que 20 mg de glicose por 100 ml de sangue, limite inferior sugerido para tolerância à lactose. Em contraste com o observado para indivíduos de ancestralidade europeia, dos quais 25% apresentaram valores inferiores ao critério citado acima. Ainda, de 36 indivíduos nativos americanos, 30 reportaram a ocorrência de diarréia, enquanto apenas um indivíduo do outro grupo (n = 16) apresentou o sintoma. O diagnóstico de intolerância à lactose foi dado para indivíduos que apresentaram sintomas associados ao exame negativo para tolerância à lactose. Assim, 83,3% e 6,25% dos indivíduos de ancestralidade nativa americana e europeia, respectivamente, foram diagnosticados com intolerância à lactose (DUNCAN E SCOTT, 1972). Em outro estudo, Caskey e colaboradores (1977) aplicaram testes de respiração de hidrogênio para o diagnóstico da má absorção de lactose em nativos americanos de Oklahoma (n = 60), também nos Estados Unidos. Os indivíduos foram classificados em seis faixas etárias: idade de 3 a 5 anos; 6 a 12; 13 a 19; 20 a 30; 31 a 44; e 45 a 64. Apenas três indivíduos de idade inferior a 13 anos (n = 20) foram diagnosticados com má absorção da lactose. Enquanto dos indivíduos de idade superior a 13 anos, 82,5% apresentaram má absorção da lactose e após os 31 anos a porcentagem aumentou para 90%. Esses resultados evidenciam que a má absorção da lactose

em nativos americanos de Oklahoma é pouco frequente durante a infância, entretanto torna-se frequente a partir da adolescência (CASKEY *et al.*, 1977).

Estudos averiguando a frequência da variação *–13910C>T* (rs4988235) em populações pan-americanas revelaram que o alelo *–13910*T* está presente em baixa frequência em nativos americanos: 0,5% na população Mapuche (Chile); 1,2% na população Nahuan (México); 7,4% na população nativa do Equador; e entre 0,5% e 7,6% nas populações nativas brasileiras, como será descrito no tópico 2.6.1 deste trabalho (FERNÁNDEZ e FLORES, 2014; OJEDA-GRANADOS *et al.*, 2016; PAZ-Y-MIÑO *et al.*, 2016; FRIEDRICH *et al.*, 2012b). Utilizando a frequência alélica obtida a partir da genotipagem da variação *–13910C>T* foi possível inferir a frequência do fenótipo LP em populações nativas Huicholes e Nahuas (México), nas quais está ausente (0%) e presente em 2,3% dos indivíduos, respectivamente; em contraste com o observado para indivíduos mexicanos *mestizos*, nos quais o fenótipo LP está presente entre 31,4% e 56,3% (OJEDA-GRANADOS *et al.*, 2016). Da mesma forma, em populações nativas do Equador e do Chile (Mapuche), o fenótipo LP está presente nas frequências de 12,5% e 10,0%, respectivamente (PAZ-Y-MIÑO *et al.*, 2016; FERNÁNDEZ *et al.*, 2016). Assim, as frequências alélicas e fenotípicas obtidas demonstram que populações nativas americanas apresentam baixa prevalência do fenótipo LP (< 20%), o que sugere que essas populações não digerem lactose naturalmente.

Além da variação *–13910C>T*, outros SNPs associados ao fenótipo LP foram reportados em baixas frequências nas populações pan-americanas, são eles: *–14010G>C* (rs145946881) em afro-brasileiros do sul do Brasil, na frequência de 0,27% (FRIEDRICH *et al.*, 2012a); *–14011C>T* (rs4988233) em populações brasileiras miscigenadas das regiões norte e nordeste do Brasil, nas frequências de 0,25% e 0,58%, respectivamente (FRIEDRICH *et al.*, 2012a); *–13913T>C* (rs4145614) e *–13915T>G* (rs41380347) em *mestizos* do Equador, nas frequências de 0,20% e 0,50%, respectivamente (PAZ-Y-MIÑO *et al.*, 2016).

Países como Peru, México e Chile apresentam alta proporção de ancestralidade nativa americana — 80%, 57,5% e 49,3%, respectivamente (RUIZ-LINARES *et al.*, 2014; ADHIKARI *et al.*, 2016) —, o que poderia explicar a baixa frequência do fenótipo LP. Embora não haja frequências para o alelo *–13910*T* relatadas em povos nativos americanos ancestrais, as baixas frequências

observadas atualmente em nativos americanos provavelmente ocorrem devido a ancestralidade desses povos, que é comum à ancestralidade de povos asiáticos (FLATZ, 1987; SALZANO e BORTOLINI, 2005). Apesar da ocorrência de populações pastoralistas na Ásia, a frequência do fenótipo LP é inferior à relatada na Europa e na África (SÉGUREL e BON, 2017) e as frequências alélicas de *–13910*T* são baixas, raramente excedendo 10% (ITAN *et al.*, 2010). Assim, os eventos de miscigenação que levaram à formação das populações pan-americanas modernas (mencionados no tópico 2.5 deste trabalho) contribuíram para a introdução de alelos associados ao fenótipo LP em grupos nativos americanos (FRIEDRICH *et al.*, 2012b; MENDOZA TORRES *et al.*, 2012; LATORRE *et al.*, 2014; FERNÁNDEZ *et al.*, 2016; OJEDA-GRANADOS *et al.*, 2016; PAZ-Y-MIÑO *et al.*, 2016; VALENCIA *et al.*, 2017; MONTALVA *et al.*, 2019). Consequentemente, frequências mais altas do fenótipo LP são observadas em populações miscigenadas brasileiras (FRIEDRICH *et al.*, 2012a) e *mestizos* do Chile (FERNÁNDEZ *et al.*, 2016), Colômbia (MENDOZA TORRES *et al.*, 2012), Equador (PAZ-Y-MIÑO *et al.*, 2016) e México (OJEDA-GRANADOS *et al.*, 2016).

2.6.1 O fenótipo lactase persistente em populações brasileiras

Similar ao observado para populações miscigenadas pan-americanas, existem poucos estudos que abordem variantes associadas ao fenótipo LP além do SNP de origem europeia *–13910C>T* (rs4988235) em populações miscigenadas brasileiras. Assim, não estão disponíveis amplas informações quanto às variantes africanas associadas ao fenótipo LP nessas populações.

Bulhões e colaboradores (2007) avaliaram a presença das variações *–22018G>A* e *–13910C>T* em vinte indivíduos de Porto Alegre, Rio Grande do Sul, sendo 19 euro-brasileiros e um descendente de asiáticos. Desses, dez indivíduos acreditavam possuir intolerância à lactose (associado ao fenótipo LNP) e o restante acreditava não possuí-la (fenótipo LP). Foram genotipadas as variações *–22018G>A* e *–13910C>T*, e foram aplicados testes de respiração de hidrogênio (H_2) para o diagnóstico da má absorção de lactose.

O teste de respiração de hidrogênio consiste em medir a concentração do gás na respiração do paciente após a ingestão de quantidade pré-determinada de

lactose; nesse caso, 50 g de lactose diluída em 250 ml de água (BULHÕES *et al.*, 2007). O hidrogênio é produzido exclusivamente por bactérias intestinais a partir do metabolismo de carboidratos. Dessa forma, pacientes de fenótipo LNP não metabolizam a lactose e, consequentemente, apresentam concentrações mais altas de hidrogênio na respiração, produzido pelo metabolismo das bactérias, do que indivíduos de fenótipo LP.

A excreção de hidrogênio foi medida em partes por milhão (ppm) a cada 60 minutos, durante o intervalo de 180 minutos após a administração da solução de lactose. Indivíduos com excreção superior a 20 ppm em relação à concentração de hidrogênio anterior a ingestão da solução foram diagnosticados com má absorção da lactose, possuindo o fenótipo LNP (BULHÕES *et al.*, 2007).

Dos dez indivíduos que acreditavam não possuir intolerância à lactose, todos apresentaram teste de respiração de hidrogênio negativo e nenhum sintoma associado à intolerância à lactose. Os genótipos observados para as variações *–22018G>A* e *–13910C>T* foram, respectivamente: *GA* e *CT*, em seis indivíduos; *AA* e *CT*, em um indivíduo; e *AA* e *TT*, em três indivíduos (BULHÕES *et al.*, 2007). Dos dez indivíduos que acreditavam possuir intolerância à lactose, nove apresentaram teste de respiração de hidrogênio positivo e um apresentou teste negativo. Porém, todos os indivíduos apresentaram sintomas associados à intolerância à lactose (distensão abdominal, náusea, flatulência e diarreia) nas 24 horas após a realização do teste (BULHÕES *et al.*, 2007). Os níveis de excreção de hidrogênio observados no grupo diagnosticado com má absorção de lactose foi estatisticamente maior do que no outro grupo ($p < 0,001$). A máxima excreção de hidrogênio observada nesses pacientes ocorreu em 120 minutos após a administração de lactose. Os genótipos *GG* e *CC* (para as variações *–22018G>A* e *–13910C>T*, respectivamente) foram observados nos nove indivíduos que apresentaram teste de respiração de hidrogênio positivo. O indivíduo que apresentou teste de respiração de hidrogênio negativo e sintomas associados à intolerância à lactose, possui genótipo *GA* e *CT* para as variações *–22018G>A* e *–13910C>T*, respectivamente. Para esse indivíduo, a presença de sintomas após a realização do teste de respiração de hidrogênio está relacionada, possivelmente, a causas secundárias de intolerância à lactose (BULHÕES *et al.*, 2007). Esses resultados sugerem que, para a população brasileira avaliada (predominantemente euro-brasileira), o teste de respiração de hidrogênio

possui alta correlação ($p < 0,001$) com a presença do alelo *–13910*T*, que possui efeito funcional validado na determinação do fenótipo LP (TROELSEN *et al.*, 2003; BULHÕES *et al.*, 2007).

Mattar e colaboradores (2009) avaliaram 567 brasileiros para a substituição *–13910C>T*. As amostras foram agrupadas em euro-brasileiros (n = 399), afro-brasileiros (n = 50), miscigenados (descendentes de euro e afro-brasileiros, n = 65) e nipo-brasileiros (n = 53). A prevalência dos genótipos *CC*, *CT* e *TT* variou entre os grupos, sendo o genótipo *CC* observado em maior frequência em indivíduos nipo-brasileiros, 100%; seguido de afro-brasileiros, 80%; miscigenados e euro-brasileiros, 57% em ambos os grupos. O genótipo *CT* foi observado nos grupos de miscigenados, euro-brasileiros e afro-brasileiros nas frequências de 40%, 37,3% e 16%, respectivamente. Enquanto o genótipo *TT*, mais raro, foi observado nos grupos de miscigenados e euro-brasileiros nas frequências de 3,1% e 6%, respectivamente. Considerando que os genótipos *CT* e *TT* estão associados ao fenótipo LP, a frequência fenotípica observada para os grupos de miscigenados, euro-brasileiros e afro-brasileiros foi de 43,1%, 43,4% e 20%, respectivamente (MATTAR *et al.*, 2009).

Como a variação *–13910C>T* não foi observada no grupo de nipo-brasileiros previamente investigados, Mattar e colaboradores (2010) avaliaram a substituição *–22018G>A* em 56 nipo-brasileiros. Das 56 amostras previamente genotipadas para a variação *–13910C>T* (genótipo *CC* observado em 100%) (MATTAR *et al.*, 2009), 5,4% possuem genótipo *GA*, associado ao fenótipo LP, e 94,6% possuem genótipo *GG*, associado ao fenótipo LNP, para a variação *–22018G>A* (MATTAR *et al.*, 2010). Anteriormente, foi observado que testes de respiração de hidrogênio realizados em nipo-brasileiros diagnosticaram a má absorção de lactose em 100% dos casos (SEVÁ-PEREIRA *et al.*, 1982). Esse resultado está em concordância com a frequência genotípica obtida a partir da genotipagem de *–22018G>A* (MATTAR *et al.*, 2010). Extrapolando os resultado obtidos para a população nipo-brasileira no geral, não é possível excluir a presença da variação *–13910C>T*, dado que os alelos *A* e *T* estão em forte desequilíbrio de ligação ($r^2 > 0,82$), atingindo $r^2 = 1,0$ em populações do sul asiático (THE 1000 GENOMES PROJECT CONSORTIUM, 2015). Assim, o trabalho propõe que as variações *–22018G>A* e *–13910C>T* devem ser avaliadas em conjunto para um diagnóstico adequado (MATTAR *et al.*, 2010).

Friedrich e colaboradores (2012a) investigaram alelos associados à LP em quatro populações brasileiras: em afro (n = 182) e euro-brasileiros (n = 337) da região sul (Porto Alegre, Rio Grande do Sul), em populações do norte (n = 200; Belém, Pará) e do nordeste (n = 262; Recife, Pernambuco). O alelo *–13910*T* foi observado nas três regiões estudadas como provável contribuição europeia, mesmo que em diferentes proporções, ao pool gênico brasileiro. A frequência alélica mais alta foi de 29,5%, observada em euro-brasileiros de Porto Alegre, seguida por 20,4%, 18,4% e 17,5% nas populações de Recife, nos afro-brasileiros de Porto Alegre e em Belém, respectivamente (FRIEDRICH *et al.*, 2012a).

Além do alelo europeu *–13910*T*, foram observados, em baixa frequência, os alelos *–13779*C*, *–13937*A*, *–14010*C* e *–14011*T*, também associados ao fenótipo LP (FRIEDRICH *et al.*, 2012a). O alelo *–13779*C*, identificado em dois indivíduos de Recife, foi descrito em um indivíduo de fenótipo LNP de etnia Somali (INGRAM *et al.*, 2009), porém também é observado em populações indianas pastoralistas na frequência de 2,4% (ROMERO *et al.*, 2012), em 0,5% na população SAS (Sul Asiático) do banco de dados *1000 Genomes Project* fase 3 (THE 1000 GENOMES PROJECT CONSORTIUM, 2015) e na frequência de 0,01% em populações europeias não finlandesas e do sul asiáticas presentes no banco de dados *gnomAD* versão 3 (KARCZEWSKI e MARTIN, 2020). O alelo *–13937*A*, identificado em um indivíduo afrodescendente de Porto Alegre e em um indivíduo de Recife, é observado em populações de etnia Xhosa da África do Sul, que possuem o hábito de consumir leite e produtos fermentados, na frequência de 1,4% (TORNIAINEN *et al.*, 2009), em 0,8% na população ASW (afro-americanos do sudoeste dos Estado Unidos) do banco de dados *1000 Genomes Project* fase 3 (THE 1000 GENOMES PROJECT CONSORTIUM, 2015) e na frequência de 0,04% no banco de dados *gnomAD* versão 3 (KARCZEWSKI e MARTIN, 2020). O alelo *–14010*C*, identificado em um indivíduo afro-brasileiro de Porto Alegre, é observado nas frequências de 46%, 39% e 32% nas populações da região afro-asiática e nilo-saariana da Tanzânia e nilo-saariana do Quênia, respectivamente (TISHKOFF *et al.*, 2007). Além dessas, o alelo *–14010*C* também é observado em populações de etnia Somali da Etiópia (INGRAM *et al.*, 2009) e de etnia Xhosa da África do Sul (TORNIAINEN *et al.*, 2009). O alelo *–14011*T*, identificado em quatro indivíduos (um de Belém e três de Recife), possui distribuição global, sendo observado em

frequência inferior a 5% em populações da Europa, África, Ásia e Oriente Médio (INGRAM *et al.*, 2009; JONES *et al.*, 2013; LIEBERT *et al.*, 2016).

Boschmann e colaboradores (2016) investigaram a frequência da variação *–13910C>T* em 292 indivíduos euro-brasileiros da região de Curitiba (PR) e em 151 indivíduos Menonitas (PR), grupo de ancestralidade alemã e holandesa. A população Menonita passou por diversos gargalos de garrafa desde o século XVI, até que, em 1930, uma parcela chegou ao Brasil (LOPES *et al.*, 2016). As frequências para o alelo *–13910*T* e para o fenótipo LP inferido para as populações euro-brasileiras de Curitiba e Menonita foram de 33,0% e 55,5%, e, 65,0% e 88,0%, respectivamente. Ainda, a frequência alélica obtida foi comparada com a frequência reportada no estudo de Friedrich *et al.* (2012a), citado anteriormente. Assim, as frequências alélicas das populações euro-brasileiras de Curitiba (PR) e de Porto Alegre (RS) foram semelhantes (p = 0,2) (BOSCHMANN *et al.*, 2016).

Poucos estudos envolvendo o fenótipo LP foram realizados em populações nativas brasileiras. Porém, sabe-se que as frequências alélicas e genotípicas associadas ao fenótipo LP variam de acordo com a taxa de miscigenação, que difere bastante entre as populações nativas e é ausente ou muito baixa em populações isoladas (SALZANO e CALLEGARI-JACQUES, 1988). Friedrich e colaboradores (2012b) investigaram 316 indivíduos pertencentes a quatro populações nativas brasileiras (região central e sul) de grupos linguísticos distintos, sendo eles: Jê (populações Kaingang, n = 71, e Xavante, n = 101) e Tupi (populações Guarani–Kaiowá, n = 84, e Guarani–Nãndeva, n = 59). Foram identificados 12 polimorfismos, dos quais *–22018G>A* e *–13910C>T* foram selecionados para a inferência dos haplótipos dos indivíduos. As variantes selecionadas estão em desequilíbrio de ligação nas quatro populações avaliadas. A combinação entre os alelos *–22018*A* e *–13910*T* foi observada em Guarani e Kaingang, estando inseridos no haplótipo *A*, e em Xavante, estando inseridos no haplótipo *C* — ambos os haplótipos são associados ao fenótipo LP. Outras combinações observadas foram: *–22018*G* e *–13910*T* em Guarani–Ñandeva (0,8%) — em um indivíduo heterozigoto para os haplótipos *B* e *C* —, e *–22018*A* e *–13910*C* em Kaingang (0,7%). A frequência do alelo *–13910*T* foi de 0,5%, 0,6%, 4,9% e 7,6% nas populações Xavante, Guarani–Kaiowá, Kaingang e Guarani–Nãndeva, respectivamente (FRIEDRICH *et al.*, 2012b).

Provavelmente as variações *–22018G>A* e *–13910C>T* foram originadas da seguinte forma: *GC → AC → AT* (POULTER *et al.*, 2003). Portanto, a combinação entre os alelos *–22018*G* e *–13910*T* seria possível apenas se tivesse ocorrido recombinação. Estudos avaliando microssatélites e a evolução da região genômica em questão (61,4 kb ao redor de *–22018G>A* e *–13910C>T*) verificaram que o surgimento e aumento da frequência alélica das variantes foram dirigidos pela ação da seleção natural, não havendo recombinação na região genômica avaliada (COELHO *et al.*, 2005). Como a variação *–22018G>A* surgiu antes da variação *–13910C>T*, um único evento de mutação seria necessário para originar o haplótipo *A* que contém a combinação *–22018*A* e *–13910*T* na população nativa brasileira atual, se o alelo *–22018*A* estivesse presente na população ancestral que povoou as Américas (FRIEDRICH *et al.*, 2012b). Somado a isso, a ocorrência de *–22018*G* e *–13910*T* num mesmo haplótipo, na ausência de recombinação, sugere que a presença do alelo *–13910*T* é devida a mutação *de novo* (FRIEDRICH *et al.*, 2012b). Nas populações Guarani e Kaingang, a frequência do alelo *–13910*T* e a taxa estimada de miscigenação com não nativos (0% em Guarani–Kaiowá, 14,3% em Guarani–Nãndeva e 7,2% em Kaingang (TSUNETO *et al.*, 2003)) se correlacionam. Desse modo, a combinação entre os alelos *–22018*A* e *–13910*T* (haplótipo *A*), provavelmente, provêm de linhagens europeias, sendo a variante *–13910*T* observada de origem não nativa (FRIEDRICH *et al.*, 2012b). Para a população Xavante, devido a baixa miscigenação com não nativos, provavelmente, a combinação entre os alelos *–22018*A* e *–13910*T* (haplótipo *C*) surgiu a partir de mutação *de novo — apesar do alelo –13910*T* já ter sido observado em haplótipo *C* anteriormente (HARVEY *et al.*, 1998). Assim, a variante *–13910*T* observada na população Xavante, idêntica a variante europeia, possui origem nativa (FRIEDRICH *et al.*, 2012b).

Quanto aos haplótipos observados em populações brasileiras, Friedrich e colaboradores (2012a) observaram 26 haplótipos distintos, sendo o haplótipo *A* (europeu, FIGURA 4) o mais comum, identificado em 46,0%, 38,0%, 37,8% e 34,3% dos euro-brasileiros da região sul, de indivíduos do norte, do nordeste e em afro-brasileiros da região sul, respectivamente (FRIEDRICH *et al.*, 2012a). Em populações brasileiras, o alelo *–13910*T* foi observado nos haplótipos *A*, *B*, *J* e *K*. O haplótipo *A* é característico de populações europeias, e, provavelmente, os

haplótipos *B*, *J* e *K* foram originados a partir de eventos de recombinação envolvendo o haplótipo *A*, dado a origem miscigenada da população brasileira (tópico 2.5 deste trabalho). O alelo *–13779*C* foi observado em dois indivíduos heterozigotos para os haplótipos *C* e *U* e *C* e *B* (ambos de Recife). O alelo *–13937*A* foi observado em um indivíduo heterozigoto para os haplótipos *A* e *U* (afro-brasileiro de Porto Alegre) e em um indivíduo heterozigoto para os haplótipos *A* e *g* (de Recife). O alelo *–14011*T* foi observado em dois indivíduos homozigotos para o haplótipo *A* (um de Belém e o outro de Recife), e em dois indivíduos heterozigotos para os haplótipos *A* e *E* e *A* e *S* (ambos de Recife). Por fim, o alelo *–14010*C* foi observado em um indivíduo heterozigoto para os haplótipos *A* e *P*. Anteriormente o alelo *–14010*C* foi descrito em haplótipo *F* (TISHKOFF *et al.*, 2007) e em haplótipo *B* (INGRAM *et al.*, 2009); entretanto, no estudo de FRIEDRICH e colaboradores (2012a), a variante *–958C>T* (rs56064699) que diferencia os haplótipos *B* e *P* não foi avaliada (FIGURA 4). Assim, o haplótipo *P* descrito em populações brasileiras, provavelmente, é correspondente ao haplótipo *B*, descrito por Ingram e colaboradores (2009) em populações de etnia Somali da Etiópia (FRIEDRICH *et al.*, 2012a).

Nas populações nativas brasileiras avaliadas por Friedrich e colaboradores (2012b), a distribuição de haplótipos é heterogênea, porém dos 15 haplótipos identificados apenas os haplótipos *A*, *B*, *C* e *U* (mais comuns; HOLLOX *et al.*, 2001; FIGURA 4) foram observados em frequência superior ou igual a 5,0%. Avaliando as populações Kaingang, Xavante, Guarani–Kaiowá e Guarani–Nãndeva em conjunto, o haplótipo *C* foi o mais frequente (41%), em contraste com as populações asiática e africana, nas quais é observado em média em 20% da população, e com a população europeia, na qual é observado em 9% da população. Além disso, foi observado o haplótipo *E* (pouco frequente em outros grupos), na frequência de 10% (também avaliando as populações em conjunto), sendo esse haplótipo uma provável recombinação entre os haplótipos *A* e *C* (FRIEDRICH *et al.*, 2012b). Ainda, aplicando o teste de neutralidade de Ewens-Watterson, foi possível concluir que os haplótipos em questão não estão sob pressão de seleção nas populações nativas brasileiras avaliadas. Provavelmente esses haplótipos estavam presentes na população ancestral que povoou as Américas, dessa forma, a diversidade

haplotípica atual deve-se à deriva genética, como indicado pelo teste de neutralidade (FRIEDRICH *et al.*, 2012b).

2.7 O FENÓTIPO LACTASE PERSISTENTE E AS CLASSES DE PESO

Estudos que relacionam o fenótipo LP e as classes de peso, definidas pelo índice de massa corporal (IMC), levam em conta que indivíduos de fenótipo LNP possuem dieta restrita quando comparada à dieta de indivíduos de fenótipo LP, como dietas distintas possuem balanços energéticos diferentes, pode-se observar impacto direto na massa corporal (KETTUNEN *et al.*, 2010). Os laticínios possuem alto teor de energia, assim, o consumo diário desses produtos faz com que indivíduos de fenótipo LP apresentem maior ingestão de calorias diárias frente à ingestão calórica diária por indivíduos de fenótipo LNP, impactando, novamente, o balanço energético (KETTUNEN *et al.*, 2010). Adicionalmente, o consumo de laticínios por indivíduos de fenótipo LNP pode levar ao desenvolvimento de sintomas intestinais, que, por sua vez, podem impedir a absorção correta de nutrientes, influenciando o ganho de peso. Além disso, o leite também é rico em cálcio, mineral essencial para o desenvolvimento dos ossos. Portanto, indivíduos de fenótipo LP que consomem leite e não apresentam sintomas associados à intolerância à lactose podem apresentar maior estatura quando comparados a indivíduos que ao consumirem leite apresentam tais sintomas (KETTUNEN *et al.*, 2010). Assim, a depender do fenótipo, indivíduos LNP e LP podem apresentar diferenças no peso e altura, o que leva a discrepâncias no valor do IMC entre os grupos.

Em 2016, mais de 1,9 bilhão de adultos (a partir de 18 anos) estavam acima do peso, desses, mais de 650 milhões eram obesos, o que corresponde a 39% e 13% da população mundial, respectivamente. Sobrepeso e obesidade são definidos como o acúmulo excessivo ou anormal de gordura corporal (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2020). O IMC é uma medida de massa corporal por altura comumente utilizada para classificar sobrepeso e obesidade em adultos e é definida pela razão massa corporal (kg) pelo quadrado da altura (m) (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2020; CENTRO DE CONTROLE E PREVENÇÃO DE DOENÇAS, 2020).

A Organização Mundial da Saúde define que, para indivíduos de idade igual ou superior a 20 anos, o IMC ≥ 25 kg/m^2 caracteriza sobrepeso e o IMC ≥ 30 kg/m^2 caracteriza obesidade. No Brasil, o Ministério da Saúde recomenda que, para adultos entre 20 e 59 anos, a avaliação nutricional seja realizada de acordo com o IMC e circunferência da cintura, como sugerido pela Organização Mundial da Saúde (MINISTÉRIO DA SAÚDE, 2017). As classes de peso e os valores de IMC correspondentes podem ser observados na TABELA 4 (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2020).

TABELA 4: ÍNDICES DE MASSA CORPORAL (IMC) E CLASSES DE PESO PARA INDIVÍDUOS ADULTOS ACIMA DE 20 ANOS

IMC	Classes de peso
Abaixo de 18,5	Abaixo do peso
Entre 18,5 e 24,9	Peso normal
Entre 25 e 29,9	Acima do peso
Entre 30 e 34,9	Obesidade I
Entre 35 e 39,9	Obesidade II (severa)
Acima de 40	Obesidade III (mórbida)

FONTE: Organização Mundial da Saúde (2020).

Para crianças e adolescentes, entre 0 e 19 anos, o IMC é calculado da mesma forma apresentada anteriormente. Entretanto, durante a infância e adolescência, a variação entre o peso e a altura ocorre de acordo com a idade e sexo, por isso é necessário que o IMC seja associado a outros parâmetros. Desse modo, a avaliação do IMC é recomendada, desde que os parâmetros escores-z e percentis (os quais são definidos em duas escalas: para idades entre 0 e 5 anos e entre 5 e 19 anos) sejam levados em consideração para a categorização em classes de peso (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2020; CENTRO DE CONTROLE E PREVENÇÃO DE DOENÇAS, 2020).

O IMC fornece um parâmetro muito útil para ambos os sexos e todas as idades, entretanto deve ser avaliado individualmente, pois pode não corresponder a mesma composição corporal em diferentes indivíduos, de diferentes populações (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2020). Em 1998, Michels, Greenland e Rosner avaliaram o uso do IMC em estudos epidemiológicos. O uso de medidas

antropométricas como covariável nesses estudos, por muitas vezes, é substituído apenas pelo uso do IMC. Por exemplo, estudos de fatores de risco para pressão alta, utilizam a obesidade como um preditor poderoso para a pressão arterial. Entretanto, modelos de regressão linear relacionando o IMC a outras medidas antropométricas demonstraram que o uso do IMC sozinho não é suficiente para capturar a variação sistemática na pressão arterial. Dessa forma, é necessária a adição de medidas antropométricas que levem em conta o tamanho corporal, como a massa corporal ou a altura, além da proporção entre essas duas medidas (MICHELS, GREENLAND e ROSNER, 1998). Os autores concluíram que, embora o IMC seja uma medida antropométrica útil para densidade tecidual, é insuficiente para análises epidemiológicas amplas. Ainda, o uso exclusivo do IMC como relação de tamanho corporal pode ser uma aproximação errônea a depender do estudo, portanto, o uso combinado com medidas antropométricas adicionais torna a aproximação mais precisa (MICHELS, GREENLAND e ROSNER, 1998).

O uso do IMC também possui limitações relacionadas à detecção da obesidade e composição corporal. Segundo a Organização Mundial da Saúde, o IMC ≥ 30 kg/m^2 caracteriza obesidade, porém existem outros parâmetros que podem ser utilizados para a categorização de indivíduos de acordo com classes de peso. Uma das definições alternativas ao IMC é a porcentagem de gordura corporal, na qual valores iguais ou superiores a 25% e 30% da massa corporal em mulheres e homens, respectivamente, caracterizam obesidade (BROZEK *et al.*, 1963). Em estudo avaliando 88 mulheres e 53 homens (em sua maioria de ancestralidade europeia) foi demonstrado que o emprego da porcentagem de gordura corporal é mais seguro do que o uso do IMC sozinho para a identificação de casos de obesidade (FRANKENFIELD *et al.*, 2001). Das mulheres e homens avaliados, 38% e 40% apresentaram IMC ≥ 30 kg/m^2, respectivamente, sendo classificados como obesos. Nenhum desses indivíduos apresentou porcentagem de gordura corporal abaixo de 25% (para mulheres) ou 30% (para homens) (SEGAL *et al.*, 1988; FRANKENFIELD *et al.*, 2001). Entretanto, 30% dos homens e 46% das mulheres que apresentaram IMC ≤ 30 kg/m^2 (classificados como não obesos) exibiram porcentagem de gordura corporal acima dos valores de corte, suficiente para classificá-los como obesos de acordo com os critérios de bioimpedância elétrica (FRANKENFIELD *et al.*, 2001). Assim, a utilização do IMC sozinho para a

classificação de indivíduos em classe de peso não é eficiente, uma vez que a obesidade detectada a partir da porcentagem de gordura corporal ocorre em valores de IMC inferiores a 30 kg/m^2 (FRANKENFIELD *et al.*, 2001).

Além das limitações citadas acima, a ancestralidade também pode influenciar o uso do IMC. Deurenberg, Yap e Van Staveren (1998) avaliaram a relação entre a porcentagem de gordura corporal, o IMC e os pontos de corte para a classificação de obesidade pelo critério de IMC em diferentes populações (afro-americanos, europeus do Reino Unido e Holanda, chineses, etíopes, indonésios, polinésios e tailandeses). Considerando a mesma porcentagem de gordura corporal, idade e sexo foi possível identificar diferenças com relação ao IMC: afro-americanos, chineses, etíopes, indonésios, polinésios e tailandeses possuem IMC 1,3 kg/m^2, 1,9 kg/m^2, 4,6 kg/m^2, 3,2 kg/m^2, 4,5 kg/m^2 e 2,9kg/m^2, respectivamente, mais baixo quando comparados com europeus (do Reino Unido e Holanda) avaliados (DEURENBERG, YAP e VAN STAVEREN, 1998). A discrepância observada entre as populações pode ser devida a diferenças no balanço energético (de acordo com a dieta) e na constituição corporal característica de cada um dos grupos. Dessa forma, mostra-se necessária a definição de linhas de corte distintas e população-específicas para cada uma das classes de peso mencionadas anteriormente (DEURENBERG, YAP e VAN STAVEREN, 1998).

Considerando a hipótese exposta no início deste tópico, Bergholdt e colaboradores (2015) desenvolveram estudo de randomização mendeliana no qual investigaram a associação de consumo de leite com diabetes do tipo 2 e sobrepeso-obesidade (IMC ≥ 25 kg/m^2) em 97.811 indivíduos dinamarqueses. A determinação do fenótipo LP foi baseada na identificação genética da variação *–13910C>T* (rs4988235), genótipos *CT* e *TT* (BERGHOLDT *et al.*, 2015). A média de consumo de leite foi de, aproximadamente, 1.000 e 600 ml/semana para indivíduos de fenótipo LP e LNP, respectivamente. O *odds ratio* (OR, razão entre a chance de um evento ocorrer em um grupo, LP, e a chance de ocorrer em outro grupo, LNP) observado para diabetes do tipo 2 foi de 0,96 (intervalo de confiança de 95%: 0,86, 1,08; p = 0,50) e para sobrepeso-obesidade foi de 1,06 (intervalo de confiança de 95%: 1,00, 1,12; p = 0,04). Em análise estratificada para diabetes do tipo 2, o OR é de 0,88 (intervalo de confiança de 95%: 0,76, 1,03; p = 0,11) e 1,35 (intervalo de confiança de 95%: 1,07, 1,70; p = 0,01) (interação de p = 0,002), considerando o

consumo de leite e desconsiderando-o, respectivamente (BERGHOLDT *et al.*, 2015). Portanto, não há associação significativa entre consumo de leite e diabetes do tipo 2 e sobrepeso-obesidade, na população dinamarquesa (BERGHOLDT *et al.*, 2015).

Hartwig e colaboradores (2016) investigaram a associação de consumo de leite com sobrepeso-obesidade (IMC ≥ 25 kg/m^2) e pressão arterial a partir do fenótipo LP, também baseados na identificação genética da variante *–13910C>T* (rs4988235), genótipos *CT* e *TT*, em uma coorte de 2.843 indivíduos da região de Pelotas (RS). Os resultados obtidos pelo estudo de randomização mendeliana foram combinados com dados de outros estudos de coorte em uma avaliação sistemática de meta-análise. Para a coorte de Pelotas, o consumo de leite foi 42 ml/dia (intervalo de confiança de 95%: 18, 67) superior em indivíduos de fenótipo LP (HARTWIG *et al.*, 2016). Análises posteriores mostraram que o aumento de 100 ml/dia no consumo de leite está associado a redução de 0,26 kg/m^2 (intervalo de confiança de 95%: -0,33, -0,19) no IMC, ao aumento de 0,31 mmHg (intervalo de confiança de 95%: 0,46, 0,16) na pressão sistólica e a redução de 0,35 mmHg (intervalo de confiança de 95%: -0,46; -0,23) na pressão diastólica. Em contrapartida, a meta-análise indicou que indivíduos de fenótipo LP apresentaram IMC 0,17 kg/m^2 superior (intervalo de confiança de 95%: 0,07; 0,27) e maior tendência ao sobrepeso e obesidade (1,09, intervalo de confiança de 95%: 1,02; 1,17), sem associação confiável com a pressão sanguínea. Assim, foi possível concluir que o fenótipo LP está positivamente associado com o sobrepeso e obesidade (correção para sexo e ancestralidade genômica) (HARTWIG *et al.*, 2016). Ainda, os autores listaram possíveis limitações do estudo, entre elas: a estratificação da amostra a partir do consumo de leite — alguns indivíduos de fenótipo LNP (genótipo *CC*) podem ter parado de consumir laticínios devido ao desenvolvimento de sintomas associados à intolerância à lactose; a subestrutura populacional, para reduzir o viés da análise os autores adicionaram a ancestralidade genômica como covariável; e o uso limitado de apenas uma variante associada ao fenótipo LP frente ao número de variantes já identificadas em populações de ancestralidades distintas (HARTWIG *et al.*, 2016).

3 JUSTIFICATIVA

Como demonstrado pelos estudos expostos no tópico 2.1 deste trabalho, o teste genético para diagnóstico da hipolactasia primária baseado apenas na variação *–13910C>T* (rs4988235), apesar de ser adequado para populações europeias e eurodescendentes, não é adequado para as demais. Em outras populações, principalmente nas africanas, já foram reportadas variações distintas, localizadas no mesmo gene (*MCM6*), associadas ao fenótipo de persistência da lactase, são elas: *–13907G>C* (rs41525747), *–13913T>C* (rs41456145), *–13915T>G* (rs41380347), *–13937G>C* (rs4988234), *–14009T>G* (rs869051967), *–14010G>C* (rs145946881), entre outras.

Em populações brasileiras, bem como nas demais populações pan-americanas, existem lacunas nos poucos estudos sobre variantes associadas ao fenótipo LP, uma vez que esses estudos abordam, principalmente, as variações de origem europeia *–22018G>A* (rs182549) e *–13910C>T* (rs4988235). Sabemos que estas populações possuem componentes genéticos nativo americano, africano e europeu, assim, mostra-se necessária a avaliação de variantes de origem não europeias.

Nesse panorama, é interessante acessar populações pan-americanas miscigenadas, incluindo populações brasileiras de diferentes regiões, para avaliar se o fenótipo LP nas Américas passou por seleção natural, bem como traçar se há relação entre ancestralidade e prevalência desse fenótipo nas populações analisadas. Avaliar populações afro-brasileiras não isoladas e isoladas — como por exemplo a CQ Sertão do Valongo, pela primeira vez estudada para o traço em questão — quanto às variações associadas ao fenótipo LP, descrever as variantes identificadas, inferir fenótipos e os possíveis haplótipos também pode nos fornecer informações quanto à ancestralidade e evolução da região que compreende o gene *MCM6*. Embora seja apenas uma região cromossômica, a região localizada 13,9 kb a montante do gene *LCT* está submetida a pressões de seleção e possui variantes população-específicas. As variantes africanas, em particular, podem nos fornecer informações sobre ancestralidade subcontinental, dado que são observadas variações exclusivas em populações africanas distintas.

4 HIPÓTESE

Sabemos que a população pan-americana foi originada principalmente a partir da miscigenação de populações nativas americanas, africanas e europeias. De acordo com a região do país, localização dele no continente, ou com a população específica, a contribuição relativa desses ancestrais difere, da mesma forma que as populações ancestrais específicas. Para nativos americanos, por exemplo, as populações mesoamericanas, andinas e amazônicas são grupos populacionais altamente diferenciados.

Avaliando a distribuição do fenótipo LP nas Américas, esperamos observar oscilações nas frequências alélicas da variação *–13910C>T* (rs4988235) relacionadas, principalmente, à origem das populações ancestrais, as proporções de contribuição de populações ancestrais para a formação das populações miscigenadas pan-americanas e efeitos demográficos decorrentes do processo de formação destas. Já quanto à seleção natural do fenótipo LP, esperamos observar sinal de seleção reduzido ou ausente nas populações analisadas (quando comparado ao sinal observado em populações ancestrais europeia e africana). A avaliação da região que compreende o gene *MCM6* é necessária para averiguar se houve seleção recente nas populações pan-americanas, além da seleção já reportada nas populações ancestrais europeia e africana.

Para as populações afro-brasileiras, esperamos observar variantes associadas ao fenótipo LP de origem africana, bem como a variante europeia — em menor frequência quando comparada à frequência alélica observada em populações euro-brasileiras. Provavelmente, as variantes africanas identificadas serão correspondentes àquelas observadas nas populações africanas ancestrais, em frequências distintas de acordo com a região do Brasil, devido ao processo de formação da população brasileira, marcado por sucessivas ondas migratórias e fluxo gênico. Para os haplótipos, esperamos observar tanto os haplótipos que contêm as variações africanas quanto o haplótipo que contém a variação europeia. Ainda, podemos observar haplótipos recombinantes que contenham mais de uma variação africana, ou variações africana e europeia em conjunto.

5 OBJETIVOS

5.1 OBJETIVOS GERAIS

Avaliar o fenótipo LP em populações miscigenadas pan-americanas sob o aspecto da genética de populações e evolução.

5.2 OBJETIVOS ESPECÍFICOS

- Determinar a distribuição geográfica da variação *–13910C>T* (rs4988235) ao longo do continente americano;
- Avaliar a região flanqueadora da variação *–13910C>T* (rs4988235) para sinais de seleção positiva em populações miscigenadas pan-americanas distintas;
- Identificar variantes associadas ao fenótipo LP, em populações afro-brasileiras da região de Curitiba (PR) e da comunidade quilombola de Sertão do Valongo da região de Porto Belo (SC);
- Avaliar se as populações afro-brasileiras da região de Curitiba (PR) e da CQ Sertão do Valongo da região de Porto Belo (SC) diferem entre si com relação às frequências alélicas e genotípicas;
- Avaliar se as populações afro-brasileiras investigadas diferem de outras populações (nativa americana, africana e europeia) descritas na literatura com relação às frequências alélicas e genotípicas;
- Avaliar a população afro-brasileira da região de Curitiba (PR) e da CQ Sertão do Valongo da região de Porto Belo (SC) com relação aos haplótipos identificados;
- Determinar se há associação entre o fenótipo LP e as classes de peso numa subamostra da população afro-brasileira da região de Curitiba (PR).

6 METODOLOGIA

6.1 AMOSTRAS

Foram incluídas no estudo 338 amostras de DNA, previamente coletadas, provenientes de afro-brasileiros de Curitiba e Região Metropolitana (PR), n = 319, e da comunidade quilombola de Sertão do Valongo da região rural de Porto Belo (SC), n = 19. Não foram coletados dados fenotípicos para a persistência da lactase. Das amostras de afro-brasileiros da região de Curitiba, 59 foram cedidas pela Profª. Drª. Maria Luiza Petzl-Erler, tiveram a coleta e uso aprovados pela Comissão Nacional de Ética em Pesquisa (CONEP) sob o nº 2.970.200 (CAAE: 02727412.4.0000.0096), e estão armazenadas no painel do Laboratório de Genética Molecular Humana da Universidade Federal do Paraná (LGMH-UFPR); e 260 foram cedidas pelo Prof. Dr. Ricardo Lehtonen Rodrigues de Souza, tiveram a coleta e uso também aprovados pela CONEP sob o nº 180/2001, e estão armazenadas no Laboratório de Polimorfismos e Ligação da mesma instituição (LPL-UFPR). As amostras referentes à CQ Sertão do Valongo foram cedidas pela Profª. Drª. Ilíada Rainha de Souza, foram coletadas de acordo com as diretrizes e normas vigentes na época e estão armazenadas no Laboratório de Genética Molecular Humana da Universidade Federal do Paraná (LGMH-UFPR). Todos os participantes informaram consentimento oralmente ou por escrito. A extração do DNA foi realizada previamente pelo método fenol-clorofórmio-álcool isoamílico (SAMBROOK *et al.*, 1989) ou pelo método de *salting-out* (LAHIRI e NURNBERGER, 1991).

Os indivíduos da região de Curitiba foram categorizados como afro-brasileiros através de classificação (observação da cor da pele e de características fenotípicas) realizada pelos pesquisadores responsáveis pela coleta das amostras de DNA. Estudos anteriores desenvolvidos pelo grupo do LGMH-UFPR para genes *HLA* indicaram que estes indivíduos possuem componentes genéticos africano, europeu e nativo americano nas proporções de 49,5%, 41,8% e 8,7%, respectivamente (BRAUN-PRADO *et al.*, 2000; PROBST *et al.*, 2000). Como não possuímos dados de ancestralidade genômica para essas amostras, assumimos, neste estudo, que o termo afro-brasileiro refere-se a indivíduos miscigenados, que possuem os componentes genéticos nativo americano em menor proporção,

africano e europeu em proporções similares, apesar da porcentagem exata variar entre os indivíduos. Estamos cientes que as características empregadas para a classificação dos participantes de Curitiba não necessariamente estão biologicamente correlacionadas com a ancestralidade genômica real dos indivíduos. Para os indivíduos da comunidade quilombola de Sertão do Valongo, não possuímos dados fenotípicos. Nesse caso, a partir do histórico de fundação e composição da população e da investigação de AIMs (ancestralidade africana de 68,1% e europeia de 31,9%; tópico 2.5.1 deste trabalho) os indivíduos foram classificados como afro-brasileiros, pois são descendentes de indivíduos de ancestralidade predominantemente africana (SOUZA e CULPI, 2005; LUIZON, 2007).

6.2 CONJUNTO DE DADOS

Para a avaliação da distribuição geográfica da variação *–13910C>T* (rs4988235) ao longo do continente americano foram analisadas informação genômica de 14.978 indivíduos não aparentados das Américas do Norte, Central e do Sul. Foram incluídas informações depositadas no banco de dados *1000 Genomes Project* fase 3 (THE 1000 GENOMES PROJECT CONSORTIUM, 2015) para afro-americanos do sudoeste dos Estado Unidos (ASW), afro-caribenhos de Barbados (ACB), indivíduos de ancestralidade mexicana de Los Angeles nos Estados Unidos (MXL), porto-riquenhos de Porto Rico (PUR) e colombianos de Medellín (CLM). Além de informações referentes a indivíduos da República Dominicana (coorte HCHS/SOL), de Cuba (coorte HCHS/SOL) e da Costa Rica (coorte CRA) depositadas no banco de dados *TOPMED Project* (TALIUN *et al.*, 2021); indivíduos do Brasil, Colômbia, Chile e Uruguai registrados no *LARGE-PD Project* (LOESCH *et al.*, 2020); indivíduos registrados no banco de dados *EPIGEN* das coortes brasileiras de Salvador e Bambuí (KEHDY *et al.*, 2015); e indivíduos peruanos miscigenados e nativos registrados no banco de dados *Peruvian Genome Project* (HARRIS *et al.*, 2018; BORDA *et al.*, 2020). Também foram incluídas nas análises informações (frequência alélica) para populações ancestrais nativas americanas (POSTH *et al.*, 2018; NAKATSUKA *et al.*, 2020) e para as populações residentes de Utah (CEPH) com ancestralidade da Europa do Norte e Ocidental (CEU), britânicos da Inglaterra e Escócia (GBR), finlandeses da Finlândia (FIN),

ibéricos da Espanha (IBS) e indivíduos da Toscana na Itália (TSI) registrados no banco de dados *1000 Genomes Project* fase 3 (THE 1000 GENOMES PROJECT CONSORTIUM, 2015).

6.3 SEQUENCIAMENTO DO ÍNTRON 13 DO GENE *MCM6*

6.3.1 PCR

Utilizando os iniciadores direto *5' GGCAGGGGTTTGGAACTTTC 3'* e reverso *5' CTGTTGAATGCTCATACGACCA 3'*, foram realizadas PCRs (do inglês, *polymerase chain reaction*) do íntron 13 do gene *MCM6*, com o objetivo de amplificar um fragmento de 594 pares de bases que compreende as variações identificadas nas posições *–13779G>C* (rs527991977), *–13907C>G* (rs41525747), *–13910C>T* (rs4988235), *–13913T>C* (rs41456145), *–13915T>G* (rs41380347), *–13937G>A* (rs4988234) *–14009T>G* (rs869051967), *–14010G>C* (rs145946881) e *–14011C>T* (rs4988233). Em todos os experimentos foram incluídas reações de controle negativo e de amostra controle. Para as reações de controle negativo, a amostra de DNA foi substituída por água e para as reações de amostra controle, foram utilizadas duas amostras do painel do LGMH-UFPR de genótipo previamente conhecido (*CC* e *CT*) para a variação *–13910C>T*.

Para as amostras provenientes de afro-brasileiros da região de Curitiba, armazenadas no painel do LGMH-UFPR, as reações foram preparadas de acordo com o protocolo 1; já para as amostras armazenadas no LPL-UFPR e para as amostras provenientes de afro-brasileiros da CQ Sertão do Valongo as reações foram preparadas de acordo com o protocolo 2, ambos explicitados na TABELA 5. Após o preparo, as reações foram submetidas ao termociclador *Mastercycler EP Gradient S® (Eppendorf, Alemanha)* de acordo com o protocolo descrito na TABELA 6.

As amplificações resultantes foram observadas em gel de agarose 1%, corado com *Unisafe Dye®* (*Uniscience*) fluorescente sob luz ultravioleta, após corrida eletroforética a 80 volts, durante 45 minutos. As reações que possuíam fragmentos amplificados foram armazenadas em freezer à temperatura de -20°C, enquanto as reações que não geraram amplificações foram refeitas. Ao final do

processo, foram obtidas 276 reações de amplificação — das quais 258 e 18 são referentes aos afro-brasileiros da região de Curitiba (PR), amostra inicial de 319 indivíduos, e da CQ Sertão do Valongo (SC), amostra inicial de 19 indivíduos, respectivamente. A redução do número amostral é devida a falhas nas amplificações mesmo após repetição e a baixa qualidade ou degradação das amostras de DNA.

TABELA 5: PROTOCOLOS 1 E 2 DA PCR

PROTOCOLO 1			
Reagentes	[Inicial]	[Final]	Volume (µL)
Tampão (x)	10,00	1,50	1,50
MgCl2 (mM)	50,00	1,50	0,30
dNTP (mM)	2,00	0,20	1,00
Iniciador direto (10 pmol/µL)	10,00	0,25	0,25
Iniciador reverso (10 pmol/µL)	10,00	0,25	0,25
Taq Polimerase Convencional (U/µL)	5,00	0,02	0,04
DNA (ng/uL)	20,00	3,00	1,50
Água (MiliQ)	-	-	5,16
TOTAL	-	-	10,00
PROTOCOLO 2			
Reagentes	[Inicial]	[Final]	Volume (µL)
Taq Polimerase Mix*	-	-	3,33
Iniciador direto (10 pmol/µL)	10,00	0,20	0,20
Iniciador reverso (10 pmol/µL)	10,00	0,20	0,20
DNA (ng/uL)	20,00	3,00	1,50
Água (MiliQ)	-	-	4,77
TOTAL	-	-	10,00

* Kit molecular doado ao LGMH-UFPR em 2019. FONTE: a autora (2020).

TABELA 6: CICLOS DA PCR

Etapa	Nº de ciclos	Temperatura (°C)	Tempo
Desnaturação inicial	1	94	2 min
Desnaturação	30	94	15 s
Anelamento		55	30 s
Extensão		72	1 min
Extensão final	1	72	1 min
Espera	1	12	5 min

FONTE: a autora (2020).

6.3.2 Reações de sequenciamento

Os produtos de PCR obtidos foram purificados utilizando as enzimas exonuclease I (*Fermentas*, EUA) e fosfatase alcalina (*Thermo Fisher Scientific*, EUA) no termociclador *Mastercycler EP Gradient S*® (*Eppendorf*, Alemanha) a 37°C por 1 hora e 80°C por 15 minutos. O sequenciamento foi realizado utilizando os iniciadores direto e reverso descritos anteriormente (tópico 6.3.1) e o reagente *BigDye*® *Terminator Cycle Sequencing Standard v3.1* (*Life Technologies*, EUA), de acordo com as instruções do fabricante. As reações de sequenciamento consistiram em uma primeira etapa a 95°C por 1 minuto, seguida por 25 ciclos de 95°C por 10 segundos, 50°C por 5 segundos e 60°C por 4 minutos, também no termociclador *Mastercycler EP Gradient S*® (*Eppendorf*, Alemanha). Os produtos das reações de sequenciamento foram purificados com etanol (*Merck*, Alemanha), ressuspendidos em Hi-Di Formamida (*Life Technologies*, EUA) e, por fim, submetidos à eletroforese capilar no equipamento *3500xl Genetic Analyzer Sequencer* (*Life Technologies*, USA). O protocolo utilizado foi otimizado anteriormente pelo LGMH-UFPR e o equipamento de eletroforese capilar empregado está disponível no Departamento de Bioquímica da Universidade Federal do Paraná.

6.4 GENOTIPAGEM DE MARCADORES HAPLOTÍPICOS

Para a genotipagem dos marcadores haplotípicos foram selecionados os marcadores *–678A>G* (rs56211644), *+666G>A* (rs3754689) e *+5579T>C* (rs2278544) que definem os haplótipos *A*, *B* e *C* (FIGURA 3 e FIGURA 4). Para isso foram desenhados os iniciadores direto e reverso descritos na TABELA 7. Então, foram realizadas PCRs alelo específicas *multiplex* para os marcadores haplotípicos selecionados. Em todos os experimentos foram incluídas reações de controle negativo, nas quais a amostra de DNA foi substituída por água, e de controle endógeno com a coamplificação de um fragmento de 434 pb do gene que codifica o hormônio de crescimento humano, *HGH* (do inglês, *human growth hormone*). Foram realizadas duas reações *multiplex* de acordo com o protocolo explicitado na TABELA 8. A primeira *multiplex* contém os iniciadores diretos para os alelos ancestrais (destacados em azul na TABELA 7), seus respectivos iniciadores reversos e os

iniciadores direto e reverso para o controle endógeno das reações (*HGH*). A segunda *multiplex* contém os iniciadores diretos para os alelos derivados (destacados em vermelho na TABELA 7), seus respectivos iniciadores reversos e os iniciadores direto e reverso para o controle endógeno das reações (*HGH*).

TABELA 7: POLIMORFISMOS SELECIONADOS PARA GENOTIPAGEM, IDENTIFICAÇÃO DOS SEUS RESPECTIVOS INICIADORES, SEQUÊNCIA E TAMANHO DO FRAGMENTO AMPLIFICADO

X	Identificação	Iniciadores direto (*F*) e reverso (*R*)	Tamanho do fragmento amplificado (pb)
Controle endógeno *HGH*	*HGH_F* *HGH_R*	*TGCCTTCCCAACCATTCCCTTA* *CCACTCACGGATTTCTGTTGTGTTTC*	434
Polimorfismo *–678A>G* (rs56211644)	*LCT-678A_F* *LCT-678G_F* *LCT-678_R*	*CACACCTATGGTCCCACCTA* *CACACCTATGGTCCCACCTG* *CCTATACAAGTTAAGCAGAGGAACC*	210
Polimorfismo *+666G>A* (rs3754689)	*LCT666G_F* *LCT666A_F* *LCT666_R*	*TTCAGGCGGAAAACTCTCTG* *TTCAGGCGGAAAACTCTCTA* *CAGGTATATTTTTGGGCTGCTGT*	111
Polimorfismo *+5579T>C* (rs2278544)	*LCT5579T_F* *LCT5579C_F* *LCT5579_R*	*TGTGTCACACTCTCCTAGATGCT* *TGTGTCACACTCTCCTAGATGCC* *GACCCTAAGGTGTTTGGTGGC*	317

Identificados em azul e vermelho os alelos ancestrais e derivados, respectivamente. A sequência dos indicadores apresentados está no sentido *5'-.3'* FONTE: a autora (2021).

TABELA 8: PROTOCOLO DA PCR ALELO ESPECÍFICA *MULTIPLEX*

Reagentes	[Inicial]	[Final]	Volume (µL)
Tampão (x)	10,00	1,50	1,50
MgCl2 (mM)	50,00	1,50	0,30
dNTP (mM)	2,00	0,20	1,00
Iniciadores direto (10 pmol/µL)	10,00	0,25*	0,75**
Iniciadores reverso (10 pmol/µL)	10,00	0,25*	0,75**
HGH_F (10 pmol/uL)	10,00	0,25	0,25
HGH_R (10 pmol/uL)	10,00	0,25	0,25
Taq Polimerase Convencional (U/µL)	5,00	0,02	0,04
DNA (ng/uL)	20,00	3,00	1,50
Água (MiliQ)	-	-	3,66
TOTAL	-	-	10,00

* Concentração final para cada um dos iniciadores individualmente; ** Volume total considerando os três iniciadores. Iniciadores utilizados na primeira *multiplex*: *LCT-678A_F*, *LCT666G_F*, *LCT5579T_F*, seus respectivos iniciadores reversos, *HGH_F* e *HGH_R*. Iniciadores utilizados na segunda multiplex: *LCT-678G_F*, *LCT666A_F*, *LCT5579C_F*, seus respectivos iniciadores reversos, *HGH_F* e *HGH_R.* FONTE: a autora (2021).

Após o preparo, as reações foram submetidas ao termociclador *Mastercycler EP Gradient S® (Eppendorf, Alemanha)* de acordo com o protocolo descrito na TABELA 9 abaixo.

TABELA 9: CICLOS DA PCR ALELO ESPECÍFICA *MULTIPLEX*

Etapa	Nº de ciclos	Temperatura (°C)	Tempo
Desnaturação inicial	1	94	2 min
Desnaturação		94	15 s
Anelamento	10	63,5	30 s
Extensão		72	1 min
Desnaturação		94	15 s
Anelamento	10	59,3	30 s
Extensão		72	1 min
Desnaturação		94	15 s
Anelamento	10	55	30 s
Extensão		72	1 min
Extensão final	1	72	1 min
Espera	1	12	∞

FONTE: a autora (2021).

As amplificações resultantes foram observadas em gel de agarose 2,5%, corado com *Unisafe Dye®* (*Uniscience*, Brasil) fluorescente sob luz ultravioleta, após corrida eletroforética a 80 volts, durante 60 minutos. Todos os géis de eletroforese foram corridos com a aplicação de *ladder* de 100 pb (*QuatroG*, Brasil). Ao total foram genotipados para os marcadores haplotípicos 196 indivíduos da região de Curitiba (PR) e 5 indivíduos da CQ Sertão do Valongo (SC), a redução do número amostral é devida a falhas nas amplificações realizadas em *multiplex* e a impossibilidade de repetição em decorrência da pandemia de COVID-19.

6.5 ANÁLISE DE DADOS

6.5.1 Sequenciamento do íntron 13 do gene *MCM6*

As sequências obtidas a partir das reações de sequenciamento foram analisadas no *software Mutation Surveyor®*, versão 3.30 (*SoftGenetics*). O *software* atua alinhando as sequências direta e reversa obtidas com uma sequência consenso depositada no NCBI (referência GRCh 38.p12), possibilitando a avaliação das

regiões de interesse, bem como a anotação do genótipo dos indivíduos para cada uma das posições contendo os SNPs.

6.5.2 Frequências alélicas e genotípicas

As frequências alélicas e genotípicas foram calculadas por contagem direta utilizando a linguagem de programação R (R DEVELOPMENT CORE TEAM, 2020). Para o cálculo do equilíbrio de Hardy-Weinberg foi utilizado o pacote "genetics 1.3.8" (Genética de Populações), função "HWE.chisq" na linguagem de programação R (R DEVELOPMENT CORE TEAM, 2020).

Para a comparação entre frequências em subgrupos de número amostral superior a 20 foram aplicados testes de χ^2, pacote "stats", função "chisq.test"; já em subgrupos de número amostral inferior a 20 foram aplicados testes exatos de Fisher, pacote "stats", função "fisher.test", ambos na linguagem de programação R (R DEVELOPMENT CORE TEAM, 2020).

6.5.3 Contribuição genética e miscigenação

Para avaliar a proporção de contribuição de populações originais para a formação da população miscigenada brasileira utilizamos a seguinte fórmula (BEIGUELMAN, 1994):

$$x = \frac{\bar{a} - a_2}{a_1 - a_2}$$

Conhecendo os valores ā (frequência alélica na população miscigenada), a_1 (frequência alélica na população original 1) e a_2 (frequência alélica na população original 2) é possível calcular x e (1 - x), que representam a contribuição genética das populações originais para a formação da população miscigenada (BEIGUELMAN, 1994). Optamos por não avaliar a população afro-brasileira de Sertão do Valongo, pois o coeficiente de endocruzamento médio para esse grupo é de 0,06906 (SOUZA e CULPI, 1992), superior ao estimado para o estado de Santa Catarina e para o Brasil, 0,00032 (DE BASSI, 1983) e 0,00088 (FREIRE-MAIA,

1990), respectivamente. Assim, como a fórmula aplicada não considera endocruzamentos, avaliamos apenas a população afro-brasileira de Curitiba.

Para isso, foram utilizados dados de frequência alélica da variante *–13730*G* (rs4954492) — 2,71% para a população em questão (ā); 3,49% (a_1) para a população africana (África Oriental; RANCIARO *et al.*, 2014); e 0% (a_2) para a população europeia, pois a variação é exclusiva de populações africanas e afrodescendentes (THE 1000 GENOMES PROJECT CONSORTIUM, 2015) — e da variante *–13910*T* (rs4988235) — 18,67% para a população em questão (ā); 1,0% (a_1) para a população africana (Moçambique; COELHO *et al.*, 2005); e 60,0% (a_2) para a população europeia (Portugal; COELHO *et al.*, 2005).

Todavia a estimativa de contribuição genética das populações originais para a formação da população miscigenada apresenta algumas limitações: as frequências reportadas na população moderna, e empregadas no cálculo em questão, podem não ser correspondentes àquelas observadas nas populações ancestrais que deram origem à população brasileira devido a ação de fatores evolutivos; a estimativa não leva em conta as diferenças subcontinentais com relação às populações originais que deram origem à população brasileira (por exemplo, mais de uma população de origem europeia pode ter contribuído para a formação da população miscigenada); as frequências alélicas observadas nos grupos que chegaram ao Brasil podem não corresponder às frequências alélicas observadas na população original devido ao efeito fundador, ainda, no Brasil pode ter ocorrido deriva genética, alterando as frequências alélicas observadas; e, por último, existe heterogeneidade nas frequências alélicas ao longo dos continentes europeu (ITAN *et al.*, 2010) e africano (INGRAM *et al.*, 2009; ITAN *et al.*, 2010) e dentro de seus países, assim, os valores utilizados para a_2 tratam-se de aproximações de acordo com o disponível na literatura.

6.5.4 Inferência de fenótipos lactase persistente e lactase não persistente

A inferência do fenótipo de cada indivíduo e o cálculo das frequências fenotípicas foram realizados utilizando a linguagem de programação R (R DEVELOPMENT CORE TEAM, 2020). A inferência de fenótipos ocorreu a partir dos genótipos anotados levando em conta as variações de origem africana (ou comum

em populações africanas) *–13907C>G* (rs41525747), *–13913T>C* (rs41456145), *–13915T>G* (rs41380347), *–13937G>A* (rs4988234), *–14009T>G* (rs869051967) e *–14010G>C* (rs145946881); as variações de origem não determinada *–13779G>C* (rs527991977, reportada em populações do sul asiático e de Recife, Brasil) e *–14011C>T* (rs4988233, reportada em populações da Europa, África e Oriente Médio); e a variação de origem europeia *–13910C>T* (rs4988235), todas associadas ao fenótipo LP e localizadas no íntron 13 do gene *MCM6*.

6.5.5 Inferência de haplótipos e desequilíbrio de ligação

A inferência de haplótipos e o cálculo das frequências haplotípicas foram realizados utilizando o algoritmo *PHASE* (STEPHENS, SMITH e DONNELLY, 2001) implementado no *software DnaSP*, versão 6.12.03 (ROZAS *et al.*, 2017). Para isso, foram consideradas as posições *–13730*, *–13732*, *–13752*, *–13779*, *–13907*, *–13910*, *–13913*, *–13915*, *–13937*, *–14009*, *–14010* e *–14011*. Além das posições *–678*, *+666* e *+5579*, referentes aos marcadores haplotípicos empregados neste estudo.

Para o cálculo do desequilíbrio de ligação, para as regiões nas quais foram observadas variações e para os SNPs *–678A>G* (rs56211644), *+666G>A* (rs3754689) e *+5579T>C* (rs2278544), foi utilizado o pacote "genetics 1.3.8" (Genética de Populações), função "LD", parâmetros r^2 e D', na linguagem de programação R (R DEVELOPMENT CORE TEAM, 2020).

6.5.6 Avaliação piloto: associação entre o fenótipo lactase persistente e as classes de peso

Para as análises de associação entre o fenótipo LP e a classe de peso foram utilizados dados referentes a subamostra dos indivíduos afro-brasileiros da região de Curitiba. A classe de peso foi determinada pelo índice de massa corporal (IMC), calculada a partir da razão massa corporal (kg) pelo quadrado da altura (m). Os dados necessários para o cálculo do IMC estavam disponíveis para uma subamostra de 68 indivíduos, sendo 52 do sexo masculino e 16 do sexo feminino. Portanto, as análises foram aplicadas nessa subamostra (n = 68 indivíduos). Esses indivíduos foram classificados fenotipicamente a partir dos genótipos *CC* (LNP), *CT* e *TT* (LP)

determinados pela variação *–13910C>T* (rs4988235), uma vez que foi a única variação observada na subamostra.

Foram realizados testes de comparação de frequências, testes de comparação de médias aritméticas, testes de poder estatístico (devido ao tamanho reduzido da subamostra avaliada) e regressão logística multivariada para o fenótipo LP, classes de peso (IMC) e sexo utilizando, principalmente, os pacotes “stats” e “pwr” na linguagem de programação R (R DEVELOPMENT CORE TEAM, 2020).

7 RESULTADOS E DISCUSSÕES

7.1 TRACING THE DISTRIBUTION OF EUROPEAN LACTASE PERSISTENCE GENOTYPES ALONG THE AMERICAS

O manuscrito a seguir foi submetido (22/02/2021) para publicação na revista *Frontiers in Genetics*, seção *Evolutionary and Population Genetics*. No momento, o manuscrito está sob revisão (31/03/2021). O material suplementar está disponível no APÊNDICE 1, p. 152.

Tracing the distribution of European lactase persistence genotypes along the Americas

Ana Cecília Guimarães Alves 1,3, Natalie Mary Sukow 1, Gabriel Adelman Cipolla 1, Marla Mendes 2, Thiago P. Leal 2, Maria Luiza Petzl-Erler 1,3, Ricardo Lehtonen Rodrigues de Souza 3,4, Ilíada Rainha de Souza 1,5, Cesar Sanchez 6, Meddly Santolalla 7, Douglas Loesch 8, Michael Dean 9, Moara Machado 2, Jee-Young Moon 10, Robert Kaplan 10,11, Kari North 12, Scott Weiss 13, Mauricio L. Barreto 14,15, M. Fernanda Lima-Costa 16,17, Bernardo Horta 18, Heinner Guio 6,19, Omar Cáceres 6,20, Carlos Padilla 6, Eduardo Tarazona-Santos 2, Ignacio F. Mata 21,22,23, Elena Dieguez 24, Víctor Raggio 25, Andres Lescano 24, Vitor Tumas 26, Vanderci Borges 27, Henrique B. Ferraz 27, Carlos R. Rieder 28, Artur Schumacher-Schuh 29,30, Bruno L. Santos-Lobato 31, Pedro Chana-Cuevas 32, William Fernandez 33, Gonzalo Arboleda 33, Humberto Arboleda 33, Carlos E. Arboleda-Bustos 33, Timothy D. O'Connor 8,34,35, Marcia Holsbach Beltrame 1,3†, and Victor Borda 8†

†These authors have contributed equally to this work and share last authorship.

Affiliations

1. Laboratório de Genética Molecular Humana, Departamento de Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, Brazil.

2. Laboratório de Diversidade Genética Humana, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

3. Programa de Pós-Graduação em Genética, Departamento de Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, Brazil.

4. Laboratório de Polimorfismos e Ligação, Departamento de Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, Brazil.

5. Laboratório de Polimorfismos Genéticos, Departamento de Biologia Celular, Embriologia e Genética, Centro de Ciências Biológicas, Universidade Federal de Santa Catarina, Florianópolis, Brazil.

6. Laboratorio de Biotecnología y Biología Molecular, Instituto Nacional de Salud, Lima, Perú.

7. Universidad Peruana Cayetano Heredia, Lima, Perú.

8. Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD.

9. Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD, 20892, USA.

10. Department of Epidemiology and Population Health, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY, 10461, USA.

11. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

12. Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA.

13. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

14.Universidade Federal da Bahia, Instituto de Saúde Coletiva, Salvador, Brazil.

15.Fundação Oswaldo Cruz, Centro de Integração de Dados e Conhecimentos para Saúde(Cidacs), Salvador-Bahia.

16.Fundação Oswaldo Cruz, Instituto de Pesquisas Rene Rachou, Belo Horizonte, Brazil

17.Universidade Federal de Minas Gerais, Programa de Pós Graduação em Saúde Pública, Belo Horizonte, Brazil.

18.Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, 464, 96001-970 Pelotas, Rio Grande do Sul, Brazil

19.Universidad de Huánuco, Huánuco, Perú

20.Carrera de Medicina Humana, Facultad de Ciencias de la Salud, Universidad Científica del Sur, Lima, 150142, Perú

21.Veterans Affairs Puget Sound Health Care System, Seattle, WA, USA

22.Department of Neurology, University of Washington, Seattle, WA, USA

23.Lerner Research Institute, Genomic Medicine, Cleveland Clinic, Cleveland, OH, USA

24.Neurology Institute, Universidad de la República, Montevideo, Uruguay

25.Department of Genetics, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay

26.Ribeirão Preto Medical School, Universidade de São Paulo, Ribeirão Preto, Brazil

27.Movement Disorders Unit, Department of Neurology and Neurosurgery, Universidade Federal de São Paulo, São Paulo, Brazil

28.Departamento de Neurologia, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil

29.Serviço de Neurologia, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

30.Departamento de Farmacologia, Universidade Federal do Rio Grande do Sul, Brazil

31.Instituto de Ciências da Saúde, Universidade Federal do Pará, Belém, Brazil

32. CETRAM, Facultad de Ciencias Médicas, Universidad de Santiago de Chile, Chile

33. Neuroscience and Cell Death Research Groups, Medical School and Genetic Institute, Universidad Nacional de Colombia, Bogotá, Colombia

34. Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD

35. Department of Medicine, University of Maryland School of Medicine, Baltimore, MD

Corresponding Author

Marcia Holsbach Beltrame marcia.hbeltrame@gmail.com

Victor Borda VBorda@som.umaryland.edu

Keywords: *–13910C>T*; *MCM6* gene; Lactose intolerance; Dairy consumption; Nutrition policies; Latin America.

All Perl and R scripts used for this study are freely available in: https://github.com/vicbp1/Lactase-persistance-in-the-Americas.git

Abstract

In adulthood the ability to digest lactose, the main sugar present in mammals' milk, is a phenotype (lactase persistence) observed in historically herder populations, mainly Northern Europeans, Eastern Africans, and Middle Eastern nomads. Since the *–13910*T* allele in the *MCM6* gene is the most well-known allele associated with the lactase persistence phenotype, the *–13910C>T* (rs4988235) polymorphism is commonly evaluated in lactase persistence studies. Lactase non-persistent adults may develop lactose intolerance events when consuming dairy products. In the Americas, there is no evidence of the consumption of these products until the arrival of Europeans. However, several American countries' dietary guidelines recommend consuming dairy for adequate human nutrition and health promotion. Considering the extensive use of dairy and the tri-hybrid ancestry of Pan-American admixed

populations, we studied the distribution of *–13910C>T* lactase persistence genotypes and its flanking haplotypes of European origin in 14,978 individuals from several Pan-American admixed populations. We found that the *–13910*T* allele frequency in Pan-American admixed populations is directly correlated with allele frequency of the European sources. Moreover, we did not observe any overrepresentation of European haplotypes in the *–13910C>T* flanking region, suggesting no selective pressure after admixture in the Americas. Finally, considering the dominant effect of the *–13910*T* allele, our results indicate that Pan-American admixed populations are likely to have higher frequency of lactose intolerance, suggesting that general dietary guidelines deserve further evaluation across the continent.

Introduction

Lactase phlorizin hydrolase, popularly known as lactase, is an enzyme expressed by enterocytes from the small intestine. Lactase is responsible for the hydrolysis of lactose, a non-absorbable sugar present in mammals' milk, into glucose and galactose, which in turn are simple sugars absorbable by the intestinal mucosa (Mantei et al., 1988). In mammals, lactase is highly expressed by newborns; after weaning, however, its expression is naturally downregulated (Wang et al., 1994), leading to a phenotype known as lactase non-persistence (LNP). This trait, also known as adult-type hypolactasia, is related to lactose indigestion and malabsorption. It may progress to lactose intolerance associated with indigestion, bloating, abdominal pain, nausea, vomiting, flatulence, and diarrhea (Auricchio et al., 1963; Boll et al., 1991). In contrast, a high prevalence of lactase persistence (LP) occurs in populations with a long history of pastoralism and dairy consumption, mainly in Northern Europe and among nomads in Africa and the Middle East (Ingram et al., 2009). Single nucleotide polymorphisms (SNPs) in the *MCM6* gene, located in the 14 kb enhancer region upstream of the *LCT* gene, are strongly associated with the LP phenotype. The *–13910*T* allele (*–13910C>T*, rs4988235, intron 13) is widely distributed in the European population; the allele frequencies vary from 8.9% in Tuscany (Italy) to 72.0% among Britishers (England and Scotland), with an average of 50.8% in Europe (1000 Genomes Project Consortium et al., 2015). The *T* allele enhances *LCT* expression at the mRNA level, causing LP in adulthood, even among

heterozygotes, thus determining a dominant trait (*TT* and *TC* genotypes) (Enattah et al., 2002). In Asia, despite the occurrence of pastoralist populations, the LP phenotype frequency is lower than that reported in Europe and Africa (Ségurel and Bon, 2017), and the *–13910*T* allele frequencies rarely exceed 10% (Itan et al., 2010). Other SNPs in African and Middle Eastern populations are observed in the same *MCM6* gene region (intron 13) and are also associated with LP phenotype. Among these, the most well-known are *–14010G>C* (rs145946881), *–13915T>G* (rs41380347), and *–13907C>G* (rs41525747) (Tishkoff et al., 2007; Ranciaro et al., 2014).

In the Americas, archaeozoological evidence supports the occurrence of mammals domestication around 7,000 - 6,000 years ago and included the ancestors of modern llamas, alpacas, vicuñas, and guanacos (Wheeler, 1995; Kadwell et al., 2001). Despite breeding these animals, there is no archeological or cultural evidence of dairy consumption by Native American populations until Europeans arrived in the late 15th century (Gade, 1999) — cattle were not introduced in the Americas until 1493 (Primo, 1992). Hence, studies regarding living Native Americans from Brazil (Friedrich et al., 2012a), Chile (Fernández et al., 2016), Ecuador (Paz-Y-Miño et al., 2016), Mexico (Ojeda-Granados et al., 2016), Peru (Figueroa et al., 1971), and the United States (Duncan and Scott, 1972; Casey, 2005) identified lower LP phenotype frequencies in these populations (20% on average), suggesting that they cannot digest lactose naturally. Countries like Peru, Mexico, and Chile have a high proportion of Native American ancestry — 80%, 57.5%, and 49.3%, respectively (Ruiz-Linares et al., 2014; Adhikari et al., 2016; Harris et al., 2018) —, which could explain the low frequency of the LP phenotype. Although there are no *–13910*T* allele frequencies reported in ancient Native American samples, the low frequencies observed in Asian peoples and present-day Native Americans suggest that the *–13910*T* allele was introduced in the Americas after the arrival of Europeans. Moreover, African and Middle Eastern SNPs have been reported at very low frequencies in the Americas. The *–14010*C* (*–14010G>C*, rs145946881) in Afro-Brazilians from the South is 0.27% (Friedrich et al., 2012b); the *–14011*T* (*–14011C>T*, rs4988233) in admixed Brazilians from the North and the Northeast are 0.25% and 0.58%, respectively (Friedrich et al., 2012b); and the *–13913*C* (*–13913T>C*, rs4145614) and *–13915*G* (*–13915T>G*, rs41380347) in Mestizos from

Ecuador are 0.20% and 0.50%, respectively (Paz-Y-Miño et al., 2016). Therefore, the phenotypic variation observed is mainly due to the *–13910C>T* (rs4988235) variation of European origin. Thus, the admixture events that led to the formation of current Latin American populations contributed to introducing LP-associated alleles in these groups (Friedrich et al., 2012a; Mendoza Torres et al., 2012; Latorre et al., 2014; Fernández et al., 2016; Ojeda-Granados et al., 2016; Paz-Y-Miño et al., 2016; Valencia et al., 2017; Montalva et al., 2019). Consequently, higher frequencies of the LP phenotype are observed in admixed populations from Brazil (Friedrich et al., 2012b), and Mestizos from Chile (Fernández et al., 2016), Colombia (Mendoza Torres et al., 2012), Ecuador (Paz-Y-Miño et al., 2016), and Mexico (Ojeda-Granados et al., 2016).

Due to the Latin American population formation process, genetic analyses have revealed the heterogeneous pattern of ancestries across the continent (Kehdy et al., 2015; Chacón-Duque et al., 2018; Harris et al., 2018; Soares-Souza et al., 2018). The peopling of the Americas occurred about 20,000 years ago, when an East Asian-derived group accessed the continent through the Bering Strait (Goebel et al., 2008; Dillehay, 2009; Reich et al., 2012; Mendes et al., 2020). Sequential population divisions and little gene flow from other continents after divergence gave rise to distinct Native American populations, with highly differentiated population groups such as the Mesoamericans, Andeans, and Amazonians, distributed across the continent (Greenberg et al., 1986; Salzano, 2011; Reich et al., 2012; Mendes et al., 2020). Later on, during the Colonial period, Europeans brought about 9 million Africans through the Transatlantic Slave Trade to the Americas (Wehling et al., 1994; John, 1997; de Mello e Souza, 2007). Although there are no official records of the sub-continental origin of African peoples, previous studies demonstrated that West-Central African ancestry is the most prevalent in the Americas (Gouveia et al., 2020). Further on, between the 19th and 20th centuries, European immigration was intensified mostly in South America. The resulting admixture drastically modified the genetic makeup of the continent.

The complex demographic processes involved in the formation of the Americas gave rise to mosaic populations in which ancestry proportions vary among and within the countries, affecting the distribution of Mendelian and complex phenotypes (Pena et al., 2011, 2020; Ruiz-Linares et al., 2014; Adhikari et al., 2016).

Considering this, in the present study we (i) addressed the geographical distribution of the *–13910C>T* SNP along the Americas, as well as (ii) analyzed this locus for evidence of positive selection in Pan-American admixed populations (American continent populations, i.e., North, Central, and South Americas). Moreover, we discussed our results in light of the use of dairy products in public health policies for nourishment in Latin American countries.

Materials and Methods

Datasets

We analyzed genomic information of 14,978 unrelated individuals from North, Central, and South Americas (Supplementary Table 1). This includes admixed individuals from the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015) [Afro-Americans from Southwest USA (ASW), Afro-Caribbeans from Barbados (ACB), individuals of Mexican Ancestry from Los Angeles USA (MXL), Puerto Ricans from Puerto Rico (PUR), and Colombians from Medellin (CLM)], as well as individuals from the TOPMED Project (Taliun et al., 2021) [Individuals from the Dominican Republic and Cuba from HCHS/SOL cohort, and individuals from Costa Rica from CRA cohort], the LARGE-PD Project (Loesch et al., 2020) [Individuals from Brazil, Colombia, Chile, and Uruguay], and admixed and Native American individuals from the Peruvian Genome Project (Harris et al., 2018; Borda et al., 2020). We also included allele frequency information of ancient Native Americans from Posth et al. (2018) and Nakatsuka et al. (2020), and of European populations [Utah residents (CEPH) with Northern and Western European ancestry (CEU), British from England and Scotland (GBR), Finnish from Finland (FIN), Iberian from Spain (IBS), and individuals from Toscani in Italy (TSI)] from 1000 Genomes Project (Supplementary Table 1). We disconsidered SNPs with significant missing data (> 10%), loci with 100% of heterozygosity, non chromosomal information and A/T-C/G genotypes using PLINK 1.9 (Chang et al., 2015).

Moreover, we generated sequencing data for the *LCT* enhancer (*MCM6* gene, intron 13) for 259 Afro-Brazilian individuals from the south region of Brazil. The sample included 241 individuals from Curitiba and its metropolitan region (Paraná

state) and 18 individuals from the quilombola community of Sertão do Valongo (Santa Catarina state) (Supplementary Figure 1). For Curitiba individuals, first-degree relatives were excluded based on self declared information. Blood samples had been previously collected as authorized under the Brazilian CONEP (*Comissão Nacional de Ética em Pesquisa*) registry numbers 180/2001 and 2.970.200 (CAAE: 02727412.4.0000.0096). These participants were heteroclassified as Afro-Brazilians. For the quilombola community, blood samples were collected according to the ethical guidelines in effect at that time and individuals were classified as Afro-Brazilians considering the community's settlement history and isolation described elsewhere (Souza and Culpi, 1992, 2005). All participants gave their written informed consent. DNA was extracted either by the phenol-chloroform-isoamyl alcohol method (Sambrook et al., 1989) or by the salting-out method (Lahiri and Nurnberger, 1991).

Sequencing

We sequenced a fragment of 594 bp in the –14 kb region of the *LCT* enhancer (*MCM6* gene). The fragment was amplified by polymerase chain reaction (PCR), using forward and reverse primers by Ranciaro et al. (2014) (5' GGCAGGGGTTTGGAACTTTC 3' and 5' CTGTTGAATGCTCATACGACCA 3', respectively), on a Mastercycler EP Gradient S® (Eppendorf, Germany). The reagents and protocol used are described in Supplementary Tables 2 and 3, respectively.

The PCR products were purified using Exonuclease I (Fermentas, USA) and Alkaline Phosphatase (ThermoFisher Scientific, USA) on a Mastercycler EP Gradient S® (Eppendorf, Germany) at 37°C for 1 hour and 80°C for 15 minutes. The sequencing was performed using BigDye® Terminator Cycle Sequencing Standard v3.1 (Life Technologies, USA), according to the manufacturer's instructions. Sequencing reactions consisted of a first step at 95°C for 1 minute, followed by 25 cycles of 95°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. The sequencing products were purified, resuspended in Hi-Di Formamide (Life Technologies, USA), and, finally, submitted to capillary electrophoresis in a 3500xl Genetic Analyzer Sequencer (Life Technologies, USA).

We analyzed the obtained sequences using the Mutation Surveyor® 3.30 software (SoftGenetics), which aligns the forward and reverse sequences to a human genome reference sequence (GRCh 38) available in NCBI (National Center for Biotechnology Information) resources, enabling the evaluation of amplified sequences and the genotype annotation.

Allelic and genotypic frequency for the *–13910*T* variant

We calculated allelic frequencies for the *–13910*T* allele (rs4988235) in each population using the --extract and --freq flags of PLINK 1.9. For genotype frequencies, we used the --hardy flag, which also calculates whether the population is in Hardy-Weinberg equilibrium for the given locus.

Ancestry analyses

Global and local ancestry inferences were performed for each population. For both analyses, we merged each population with a reference panel of 842 individuals from four continental ancestries: 206 European (IBS and CEU), 207 African (LWK and YRI), and 207 East Asian (CDX and JPT) individuals from the 1000 Genomes Project, and 222 unadmixed Native American individuals from the Peruvian Genome Project (Borda et al., 2020). Since the genetic information available for Afro-Brazilians from Curitiba and Sertão do Valongo corresponds just to the *MCM6* region, it was not possible to perform global or local ancestry analyses for these samples. Instead, we extracted the global ancestry information for Sertão do Valongo individuals from Luizon (2007) who performed this inference using eight ancestry-informative markers in the same population. No data that could enable this analysis was available for Afro-Brazilians from Curitiba.

For global ancestry analysis, we removed highly linked variants ($r^2 < 0.1$) using the PLINK flag --indep-pairwise with the following parameters: 50 10 0.1. We inferred global ancestry proportions using ADMIXTURE (Alexander et al., 2009) on supervised mode for four ancestry clusters for each population merged with the references. After ADMIXTURE runs, we calculated the average proportion of European ancestry for each population. We used R to estimate the Spearman

correlation (rho) between the average proportion of European ancestry and allelic and genotypic frequencies for all populations.

To infer the ancestry of the genomic region flanking the *–13910C>T* SNP, we extracted the chromosome 2 sequence, which includes the *MCM6* gene, for each merged dataset (admixed and ancestral references). First, each of these merged datasets was phased with shapeit4 (Delaneau et al., 2019) using the GRCh38 genetic map and the MCMC parameters: --mcmc-iterations 10b,1p,1b,1p,1b,1p,1b,1p,10m, that perform 10 burn-in iterations, followed by 4 paired runs of pruning and burn-in, and, finally, 10 main iterations of sampling. Then, we ran RFMix ver2 in order to infer the local ancestry (Maples et al., 2013) using the phased dataset with two Expectation-Maximization runs. We used in-house perl and R scripts to determine the ancestry of the flanking region of *–13910*T* allele and the length distribution of European haplotypes of the flanking region of the *MCM6* gene.

Natural Selection analysis

In order to identify evidences of natural selection, we applied two approaches. First, to identify a potential overrepresentation of European haplotypes due to recent, post-admixture natural selection acting over the *MCM6* region, we compared the average global European ancestry with the proportion of European haplotypes for each population. Moreover, we analyzed the length distribution of European haplotypes to explore the genomic region dynamic. Second, we inferred the pattern of extended haplotype homozygosity (EHH) and the length distribution of derived haplotypes around the *–13910C>T* SNP in each population in order to identify whether the selection signal observed in the European source populations is also observed in the admixed Pan-American populations. For this purpose we used the rehh package (Gautier et al., 2017). Finally, we estimated the integrated haplotype score (iHS), which indicates if a loci is under recent positive selection. This score compares the levels of linkage disequilibrium surrounding a positively selected allele with the background ancestral allele at the same position (Voight et al., 2006). For the iHS inference, we used the selscan software (Szpiech and Hernandez, 2014) with default parameters for phased information for each population.

Results

The –13910*T allele distribution is correlated with European ancestry in the Americas

We explored the geographical distribution of the *–13910*T* allele in 13 Pan-American and five European populations. The highest frequency of the *–13910*T* allele in the Americas occurs in the Uruguayan population (35%, Figure 1 and Supplementary Table 4). A positive correlation was observed between *–13910*T* allele frequencies and the average proportion of European ancestry (rho = 0.866, p-value = 1.526e-10). Additionally, when we fit a line to the Pan-American data there is a tendency towards the IBS allele frequency for the *–13910*T* allele. Consequently, we did not observe an overrepresentation of the *–13910*T* allele compared with the average proportion of European ancestry in any population (Figure 1).

Considering that the LP phenotype is a dominant trait, we explored the geographical distribution of *–13910*T* allele carriers (*TT* and *TC* genotypes) as a proxy for the proportion of individuals with this phenotype. The genotype frequencies did not differ from those expected under Hardy-Weinberg equilibrium in all Pan-American admixed populations (Supplementary Table 4). Higher proportions of *T* allele carriers were observed in Uruguay (61%), Southeast-South Brazil (~54%, including Bambui, Porto Alegre, and Pelotas), and Cuba (53%) — which also have higher proportions of European ancestry (Figure 2 and Supplementary Table 4) —, suggesting that more than half of the individuals of these samples are LP. Conversely, this inferred phenotype is absent in ancient and some Peruvian present-day Native Americans (Figure 2 and Supplementary Table 4). Furthermore, admixed populations with higher Native American ancestry (Peruvian and Guatemalan) and Afro-descendants (ASW, ACB, Afro-Peruvians, and Afro-Brazilians) have lower LP frequencies; especially, Afro-Brazilians from Sertão do Valongo did not carry the *–13910*T* allele.

Moreover, our dataset includes several Peruvian and Brazilian populations corresponding to different biogeographical regions (Figure 2, Supplementary Table 1 and Supplementary Figure 2). Admixed Peruvians have the lowest *–13910*T* allele frequencies in South America (< 13%). Interestingly, most admixed Peruvian

populations from the coast have twice the frequency of *–13910*T* than Andean and Amazonian populations (Supplementary Table 4 and Supplementary Figure 2). Analyzing Northeastern, Southeastern, and Southern Brazilian populations, we observed a North-South cline in the frequency of genotypes correlated with the proportion of European ancestry; an exception to this cline is the quilombola community of Sertão do Valongo from Southern Brazil.

The haplotype distribution of the *MCM6* gene reflects the history along the Americas

To infer which forces are playing a significant role in the evolution of the *MCM6* gene in the Americas, we explored the distribution of counts and lengths of European haplotypes that include this gene. To address this goal, we performed a local ancestry inference for this locus and confirmed that the *–13910*T* allele is observed only in European haplotypes in all Pan-American admixed populations. To determine if there is an overrepresentation of European alleles (with or without the *–13910*T* allele) in the *MCM6* region, we estimated the proportion of European haplotypes and compared it to the genome-wide proportion of European ancestry for each population. Our results (Figure 3B) revealed a high correspondence between European haplotype frequencies and the percentage of genome-wide European ancestry (rho = 0.962, p-value = 2.2e-16). This pattern suggests the possibility of neutral evolution of this genomic region after the admixture processes that gave rise to these Pan-American admixed populations.

We deconvoluted the admixture dynamics of this region by exploring the length distribution of European haplotypes that include the *MCM6* gene. We observed a high density of small European segments in populations with a lower proportion of European ancestry (Figure 3C), such as Peruvians and Afro-descendants, suggesting an early gene flow from Europeans. Furthermore, populations with higher European ancestry have a wide length distribution of European segments, indicating a continuous or multidate gene flow from Europeans. Both patterns are in agreement with historical records and population genetics studies.

We performed extended haplotype homozygosity (EHH) analysis for each population to evaluate the haplotypes and to compare their backgrounds for the *–13910*T* allele with the ancestral *–13910*C* allele. Most of the assessed populations presented a decay pattern of haplotype homozygosity around the *–13910*T* allele (Supplementary Figure 3). Afro-Peruvians and the populations of Bogota, Chile, and Porto Alegre have a LD block of 1 Mb with 100% EHH that includes the derived allele (Figure 4 and Supplementary Figure 3). Observing the length of the EHH blocks (Figure 4 and Supplementary Figure 4), it is remarkable that Afro-American and Porto Alegre populations have more extended haplotypes for the derived allele in opposition to the haplotypes containing the *–13910*C* allele (Figure 4). In both populations, all derived haplotypes are longer than 1 Mb. Specifically, in Afro-Americans, some haplotypes reach 5 Mb. Moreover, several Pan-American populations presented iHS (integrated haplotype score) values above 2 (Supplementary Table 4), while Afro-Americans presented the highest value (iHS = 4.2).

Interestingly, when we calculated the derived haplotype frequency, that includes the *–13910*T* allele, in European haplotypes for each Pan-American population, the Afro-Americans from Southwest USA (ASW) had the highest value (80%) (Supplementary Table 4). This frequency is quite similar if compared with the frequency of the derived haplotype in Northern European populations (Supplementary Figure 5). Moreover, we observed that populations that were Spaniard and Portuguese colonies (Peru, Colombia, Mexico, Chile, Uruguay, and Brazil) had derived haplotype frequencies more similar to those of Iberians from 1000 Genomes Project (Supplementary Figure 5). All these observations suggest that in Pan-American admixed populations, for the genomic segment that includes the *MCM6* gene, demography plays a more significant role than natural selection after the admixture processes.

Discussion

Lactase persistence is among the most strongly selected phenotypes in human populations; for LP, positive selection occurred during the last 5,000 - 10,000 years (Bersaglieri et al., 2004). In the Americas, there is no evidence of dairy

consumption by native groups until the arrival of Europeans (Gade, 1999). Moreover, the digestion test performed in Native American people evidenced the highest incidence (> 80%) of lactose intolerance in the Americas (Figueroa et al., 1971; Duncan and Scott, 1972; Caskey et al., 1977; Friedrich et al., 2012b; Fernández et al., 2016; Ojeda-Granados et al., 2016; Paz-Y-Miño et al., 2016). This is consistent with the absence of the *–13910*T* allele in ancient and present-day unadmixed Native Americans, and with the hypothesis that this allele was probably introduced in the Americas by Europeans since the Colonial period. In the present study, we conducted the most geographically extensive analysis of the *–13910*T* allele spanning several Pan-American admixed populations (North American Afro-descendants and Latin Americans), demonstrating high positive correlation between the distribution of the derived allele and European ancestries, with no evidence of post-admixture positive selection. Furthermore, in light of Latin American nourishment policies, our results represent an important aspect to be considered for the understanding of the relationship between the consumption of dairy products and health issues across the continent.

Impact of European colonization in the distribution of LP alleles

European immigration had varied scales and influences across Central and South American regions. Despite occurring since the 15th century, European migration to the Americas was intensified during the late 19th century. Between the 1870s and the 1930s, approximately 13 million Europeans arrived in Latin America, of which more than 90% migrated to Argentina, Brazil, Cuba, and Uruguay (Sánchez-Alonso, 2019). In agreement, higher proportions of European ancestry have been observed in Brazilian populations (except for Salvador), Cubans, and Uruguayans. Also, the length distribution of European haplotypes is wider in these populations, which is consistent with demographic data of a more continuous European gene flow.

Previous studies and historical records reported earlier European colonization in Northeast Brazil and higher proportions of African ancestry in this region (Instituto Brasileiro de Geografia e Estatística, 2000; Pena et al., 2011; Kehdy et al., 2015). We observed in the Salvador sample (Northeast Brazil) a lower

percentage of European ancestry correlated with a lower *–13910*T* allele frequency and a lower European haplotype frequency. Moreover, a shorter European haplotype was observed, possibly introduced earlier in the Salvador population than in other Brazilian populations. A similar pattern was observed in Afro-descendant populations, in which the lowest average of European ancestry was detected. In Afro-Americans from Southwest USA, the European ancestry corresponds to a minor proportion, which explains the low frequency of the *–13910*T* allele in this population. Interestingly, the largest part of the European immigrants who came to the USA originated from countries where the frequencies of the *–13910*T* allele are typically high (Bersaglieri et al., 2004; Smith et al., 2009; Itan et al., 2010), resulting in the presence of this variant in most of the European haplotypes observed in Afro-Americans. The patterns observed in our work are consistent with one-way admixture events that were inferred for these populations in previous studies (Harris et al., 2018; Ongaro et al., 2019; Gouveia et al., 2020).

In the admixed Peruvian populations, we also observed a lower *–13910*T* allele frequency, a lower European haplotype frequency, and a shorter European haplotype — similarly to those observed in Afro-descendant populations, including Afro-Peruvians. This pattern is concordant with the hypothesis that the admixture between European colonizers (mainly Spanish) and Native Americans occurred mostly after Peru's independence in 1824, about 300 years after the European settlement, resulting in the low European ancestry reported in Peruvian and Afro-Peruvian populations, which range between 14% and 22% (Harris et al., 2018).

Genetic studies indicated that the admixture between the Caribbeans and Europeans occurred shortly after the arrival of these populations in the American continent (Moreno-Estrada et al., 2013), resulting in the shorter European haplotypes observed in populations from Guatemala, Costa Rica, Dominican Republic, Barbados, and Puerto Rico. In contrast to these Central American countries, the wider European haplotype observed in Cubans was possibly due to the later European migration to this country in the early 20th century (Salzano and Bortolini, 2005).

Altogether, our results show that the distribution of lengths and frequencies related to the *–13910*T* allele reflects the demographic history of the Pan-American admixed populations.

LP in an Afro-Brazilian Quilombo community

Between the 16th and mid 19th centuries, Europeans brought to Brazil around 4 million enslaved Africans (Instituto Brasileiro de Geografia e Estatística, 2000). Africans and their descendants formed communities named Quilombos, a form of resistance to slavery and a stronghold for African culture (Moura, 2001; Raggio et al., 2018). The quilombola community of Sertão do Valongo (Santa Catarina State, Southern Brazil) has a reported African genomic ancestry of 68.1% and an European one of 31.9%, therefore with no reported Amerindian ancestry (Luizon, 2007). The absence of the *–13910*T* allele in Afro-Brazilians from Sertão do Valongo, and, consequently, the absence of the inferred LP phenotype, agrees with the community's settlement history marked by the small and early European contribution. Quilombola communities, in general, are not genetically isolated from other populations, although the degree of interaction with neighboring groups may vary (Almeida, 2010). Hence, since each community has a distinct population structure, settlement history, and cultural practices, the evaluation of other quilombola communities is essential to the understanding of the *–13910*T* allele distribution in these groups.

Lack of post-admixture selective pressure for LP in the Americas

The high correlation between European haplotype proportion and European genome-wide proportion suggests no selective pressure over the *MCM6* gene after admixture. Nonetheless, our EHH analysis showed that several populations have longer LD blocks containing the derived allele. Moreover, the iHS values for these populations were above 2, which indicates some level of positive selection (Voight et al., 2006). Considering that there is no overrepresentation of European haplotypes in the *MCM6* gene and that the length distribution of these haplotypes reflects demographic history, we hypothesized that the detected iHS signals result from positive selection in the source population. Specifically, Afro-Americans, which had the highest iHS value, have Northern Europe as the main source of their admixture (Ongaro et al., 2019; Gouveia et al., 2020). Other populations with higher European

ancestry levels (i.e., Brazilian populations) could have moderate iHS values due to the highly diverse European sources introduced into Brazil during the last century.

Impact of our results on public health policies

Dairy plays an essential role in human nutrition due to its supply of protein, lipids, and micronutrients, such as calcium, magnesium, vitamins B5 and B12, among others (Muehlhoff et al., 2013). However, it is necessary to emphasize that, in recent years, the consumption benefits of dairy in adulthood have been questioned. While some studies have linked its consumption to the development of cardiovascular diseases and obesity — possibly because of milk's fat content (Segall, 1994; Berkey et al., 2005; Muehlhoff et al., 2013) —, other studies have suggested that milk and dairy intake could reduce the risk of metabolic syndrome, diabetes, and cancer development (Elwood et al., 2007, 2008).

Dairy consumption is frequently encouraged — and even financed — by governmental initiatives to fight famine in Latin American countries due to these products' unarguable nutrient content. The "*Programa de Abasto Social de Leche LICONSA*" in Mexico (created in 1944), the "*Programa Nacional de Alimentación Complementaria*" or PNAC in Chile (created in 1952), the "*Programa del Vaso de Leche*" or PVL in Peru (created in 1984), and the "*Programa de Aquisição de Alimentos*" in Brazil (created in 2003), are examples of governmental initiatives (Uauy et al., 2001; Stifel and Alderman, 2006; Hespanhol, 2013; Morales-Ruán et al., 2013). Moreover, in Cuba, Costa Rica, México, Venezuela, Colombia, and Uruguay, dietary guidelines also recommend the consumption of dairy products (Instituto Nacional de Nutrición (Venezuela), 1990; Porrata-Maury, 2004; Comisión Intersectorial de Guías Alimentarias Para Costa Rica Ministerio de Salud, 2010; Secretaria de Salud de México, 2010; Instituto Colombiano de Bienestar Familiar and FAO Colombia, 2015; Dirección General de la Salud Ministerio de Salud de Uruguay, 2019). These initiatives follow the United States programs' pattern, in which the dietary guidelines include milk as an essential nourishment element since its first publication (Bertron et al., 1999; Jacobs et al., 2020). However, the United States programs relied on early nutrition scientific studies developed mostly in Northern European populations (Jacobs et al., 2020), where a high prevalence of LP is observed (Ingram et al.,

2009). Considering the prevalence of the inferred LNP phenotype in the Americas and the tri-hybrid ancestry of Pan-American populations, dairy intake could contribute to the generation of health issues due to the development of lactose intolerance-associated events.

It is important to highlight that the gut microbiota composition and epigenetics modifications affect the manifestation of lactose intolerance-associated events, attenuating or eliminating them in individuals that typically cannot digest the disaccharide (Zhong et al., 2004; He et al., 2008; Labrie et al., 2016; Anguita-Ruiz et al., 2020). The manifestation of these events are also dependent on the food's lactose concentration, which tends to be reduced in fermented dairy foods, such as cheese and yogurt (Buttriss, 1997). According to this, cultural and colonic adaptation mechanisms could explain the consumption of milk and dairy products without causing gastrointestinal events in non-carriers of genetic variants associated with LP (Hertzler and Savaiano, 1996; Segurel et al., 2020; Bleasdale et al., 2021). Thus, the absence of the *–13910*T* allele alone may not be completely predictive of lactose intolerance, especially considering the existence of other alleles of non-European origin associated with the LP phenotype (Tishkoff et al., 2007; Ranciaro et al., 2014) — although the low frequencies of such alleles in Pan-American admixed populations (Friedrich et al., 2012b; Paz-Y-Miño et al., 2016) reinforce the effectiveness of the *–13910*T* allele for the genetic diagnosis of LP.

Our results revealed that, in the Americas, just a few populations from Cuba, Brazil (Bambui, Pelotas, and Porto Alegre), and Uruguay have an elevated proportion (above 50%) of individuals that are likely to be lactase persistent — agreeing with the fact that these countries (except for Cuba) are among the largest consumers of dairy products (excluding butter) per year in the continent: 175.3 and 141.8 kg per capita for Uruguayans and Brazilians, respectively (Food and Agriculture Organization of the United Nations, 2018). A lower consumption in other Latin American countries (Food and Agriculture Organization of the United Nations, 2018) is associated with the prevalence of the LNP phenotype, which is consistent with the recent introduction of cattle and, consequently, with the consumption of dairy products in the continent (Gade, 1999). Therefore, traditional consumption of these products by Native American populations are rare (Kiple and Ornelas, 2001). However, it should be noted that European immigrants (and their descendants) have established such

practices in America's culinary, resulting in the recent development of typical milk-based dishes in several countries of this continent (Albala, 2011; Wiley, 2015). The socioeconomic status of Latin American countries also contributes to such mentioned lower levels, since consumption of dairy products is known to be lower in developing countries than in developed ones — despite this scenario is being modified in the last decades by increasing dairy consumption (Bermudez and Tucker, 2003; Muehlhoff et al., 2013).

The reduced frequency of the *–13910*T* allele may be a hint that dairy products are not the best option for dietary guidelines in populations with lower proportions of European ancestry, such as Peruvian and Afro-descendant populations (Afro-Americans, Afro-Caribbeans, Afro-Brazilians from Curitiba and the quilombola community of Sertão do Valongo). According to the discussion above, European-biased policies that include dairy products should be rediscussed and reconsidered. Furthermore, until the role of dairy intake in the pathogenesis of complex diseases is not fully understood, alternative milk products and plant-based dairy substitutes would possibly be a better option for America's dietary guidelines.

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

MHB and VB designed the study. ACGA and NMS performed the sequencing and the analysis of Afro-Brazilian populations. VB, MHB, ACGA, NMS, GAC and TDO analyzed the data. MHB, MLP-E, RLRS, and IRS contributed with samples and/or reagents. ACGA, NMS, MHB, and VB drafted the manuscript. GAC, MM, TPL, MLP-E, TDO, and MHB critically edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by scholarships from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/PROAP – Finance Code 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação Araucária, Brazil, provided to ACGA and NMS.

Acknowledgments

We would like to thank all the volunteers who participated in this study, Priscila Ianzen dos Santos and Valter Antonio de Baura for the technical support, and colleagues for the helpful discussions, comments and support.

References

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.

Adhikari, K., Mendoza-Revilla, J., Chacón-Duque, J. C., Fuentes-Guajardo, M., and Ruiz-Linares, A. (2016). Admixture in Latin America. *Curr. Opin. Genet. Dev.* 41, 106–114.

Albala, K. (2011). *Food Cultures of the World Encyclopedia*. ABC-CLIO.

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664.

Almeida, A. W. B. de (2010). *Cadernos de debates Nova Cartografia Social: Territórios quilombolas e conflitos*. UEA Edições.

Anguita-Ruiz, A., Aguilera, C. M., and Gil, Á. (2020). Genetics of lactose intolerance: An updated review and online interactive world maps of phenotype and genotype frequencies. *Nutrients* 12, 2689.

Auricchio, S., Rubino, A., Landolt, M., Semenza, G., and Prader, A. (1963). ISOLATED INTESTINAL LACTASE DEFICIENCY IN THE ADULT. *Lancet* 2, 324–326.

Berkey, C. S., Rockett, H. R. H., Willett, W. C., and Colditz, G. A. (2005). Milk, dairy fat, dietary calcium, and weight gain: a longitudinal study of adolescents. *Arch. Pediatr. Adolesc. Med.* 159, 543–550.

Bermudez, O. I., and Tucker, K. L. (2003). Trends in dietary patterns of Latin American populations. *Cad. Saude Publica* 19 Suppl 1, S87–99.

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., et al. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* 74, 1111–1120.

Bertron, P., Barnard, N. D., and Mills, M. (1999). Racial bias in federal nutrition policy, Part I: The public health implications of variations in lactase persistence. *J. Natl. Med. Assoc.* 91, 151–157.

Bleasdale, M., Richter, K. K., Janzen, A., Brown, S., Scott, A., Zech, J., et al. (2021). Ancient proteins provide evidence of dairy consumption in eastern Africa. *Nat. Commun.* 12, 632.

Boll, W., Wagner, P., and Mantei, N. (1991). Structure of the chromosomal gene and cDNAs coding for lactase-phlorizin hydrolase in humans with adult-type hypolactasia or persistence of lactase. *Am. J. Hum. Genet.* 48, 889.

Borda, V., Alvim, I., Mendes, M., Silva-Carvalho, C., Soares-Souza, G. B., Leal, T. P., et al. (2020). The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. *Proc. Natl. Acad. Sci. U. S. A.* doi:10.1073/pnas.2013773117.

Buttriss, J. (1997). Nutritional properties of fermented milk products. *Int. J. Dairy Technol.* Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-0307.1997.tb01731.x.

Casey, J. (2005). Holocene occupations of the forest and savanna. *African archaeology*.

Caskey, D. A., Payne-Bose, D., Welsh, J. D., Gearhart, H. L., Nance, M. K., and Morrison, R. D. (1977). Effects of age on lactose malabsorption in Oklahoma Native Americans as determined by breath H2 analysis. *Am. J. Dig. Dis.* 22, 113–116.

Chacón-Duque, J.-C., Adhikari, K., Fuentes-Guajardo, M., Mendoza-Revilla, J., Acuña-Alonzo, V., Barquera, R., et al. (2018). Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. *Nat. Commun.* 9, 5388.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.

Comisión Intersectorial de Guías Alimentarias Para Costa Rica Ministerio de Salud (2010). Guías alimentarias para Costa Rica. Comisión Intersectorial de Guias Alimentarias para Costa Rica San José.

Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L., and Dermitzakis, E. T. (2019). Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* 10, 5436.

de Mello e Souza, M. (2007). *África e Brasil africano*. Editora Ática.

Dillehay, T. D. (2009). Probing deeper into first American studies. *Proc. Natl. Acad. Sci. U. S. A.* 106, 971–978.

Dirección General de la Salud Ministerio de Salud de Uruguay (2019). Guía Alimentaria para la Población Uruguaya: para una alimentación saludable, compartida y placentera. Ministerio de Salud de Uruguay. Available at: http://www.scielo.edu.uy/scielo.php?pid=S2301-15482019000100092&script=sci_arttext&tlng=pt.

Duncan, I. W., and Scott, E. M. (1972). Lactose intolerance in Alaskan Indians and Eskimos. *Am. J. Clin. Nutr.* 25, 867–868.

Elwood, P. C., Givens, D. I., Beswick, A. D., Fehily, A. M., Pickering, J. E., and Gallacher, J. (2008). The survival advantage of milk and dairy consumption: an overview of evidence from cohort studies of vascular diseases, diabetes and cancer. *J. Am. Coll. Nutr.* 27, 723S–34S.

Elwood, P. C., Pickering, J. E., and Fehily, A. M. (2007). Milk and dairy consumption, diabetes and the metabolic syndrome: the Caerphilly prospective study. *J. Epidemiol. Community Health* 61, 695–698.

Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L., and Järvelä, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 30, 233–237.

Fernández, C. I., Montalva, N., Arias, M., Hevia, M., Moraga, M. L., and Flores, S. V. (2016). Lactase non-persistence and general patterns of dairy intake in indigenous and mestizo chilean populations. *Am. J. Hum. Biol.* 28, 213–219.

Figueroa, R. B., Melgar, E., Jó, N., and García, O. L. (1971). Intestinal lactase deficiency in an apparently normal Peruvian population. *Am. J. Dig. Dis.* 16, 881–889.

Food and Agriculture Organization of the United Nations (2018). FAOSTAT Database. New Food Balances. *FAOSTAT Database.* Available at: http://www.fao.org/faostat/en/#data/FBS [Accessed February 02, 2021].

Friedrich, D. C., Callegari-Jacques, S. M., Petzl-Erler, M. L., Tsuneto, L., Salzano, F. M., and Hutz, M. H. (2012a). Stability or variation? Patterns of lactase gene and its enhancer region distributions in Brazilian Amerindians. *Am. J. Phys. Anthropol.* 147, 427–432.

Friedrich, D. C., Santos, S. E. B., Ribeiro-dos-Santos, Â. K. C., and Hutz, M. H. (2012b). Several different lactase persistence associated alleles and high diversity of the lactase gene in the admixed Brazilian population. *PLoS One* 7, e46520.

Gade, D. W. (1999). *Nature and Culture in the Andes*. Univ of Wisconsin Press.

Gautier, M., Klassmann, A., and Vitalis, R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* 17, 78–90.

Goebel, T., Waters, M. R., and O'Rourke, D. H. (2008). The late Pleistocene dispersal of modern humans in the Americas. *Science* 319, 1497–1502.

Gouveia, M. H., Borda, V., Leal, T. P., Moreira, R. G., Bergen, A. W., Kehdy, F. S. G., et al. (2020). Origins, Admixture Dynamics, and Homogenization of the African Gene Pool in the Americas. *Mol. Biol. Evol.* 37, 1647–1656.

Greenberg, J. H., Turner, C. G., Zegura, S. L., Campbell, L., Fox, J. A., Laughlin, W. S., et al. (1986). The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence [and Comments and Reply]. *Curr. Anthropol.* 27, 477–497.

Harris, D. N., Song, W., Shetty, A. C., Levano, K. S., Cáceres, O., Padilla, C., et al. (2018). Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. *Proc. Natl. Acad. Sci. U. S. A.*, 201720798.

Hertzler, S. R., and Savaiano, D. A. (1996). Colonic adaptation to daily lactose feeding in lactose maldigesters reduces lactose intolerance. *Am. J. Clin. Nutr.* 64, 232–236.

Hespanhol, R. A. de M. (2013). Programa de Aquisição de Alimentos: limites e potencialidades de políticas de segurança alimentar para a agricultura familiar. *Soc. Nat. Resour.* 25, 469–483.

He, T., Venema, K., Priebe, M. G., Welling, G. W., Brummer, R.-J. M., and Vonk, R. J. (2008). The role of colonic metabolism in lactose intolerance. *European Journal of Clinical Investigation* 38, 541–547. doi:10.1111/j.1365-2362.2008.01966.x.

Ingram, C. J. E., Mulcare, C. A., Itan, Y., Thomas, M. G., and Swallow, D. M. (2009). Lactose digestion and the evolutionary genetics of lactase persistence. *Hum. Genet.* 124, 579–591.

Instituto Brasileiro de Geografia e Estatística (2000). *Brasil: 500 anos de povoamento*. IBGE Rio de Janeiro.

Instituto Colombiano de Bienestar Familiar and FAO Colombia (2015). *Documento Técnico de las Guías Alimentarias Basadas en Alimentos para la población colombiana mayor de 2 años*. Food & Agriculture Org.

Instituto Nacional de Nutrición (Venezuela) (1990). *Guías de alimentación para Venezuela*. , ed. A. A. A. C. A. J. M. B. P. D. E. E. Z. M. L. de G. C. R. M. C. R. de Ornés Instituto Nacional de Nutrición.

Itan, Y., Jones, B. L., Ingram, C. J. E., Swallow, D. M., and Thomas, M. G. (2010). A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol. Biol.* 10, 36.

Jacobs, E. T., Foote, J. A., Kohler, L. N., Skiba, M. B., and Thomson, C. A. (2020). Re-examination of dairy as a single commodity in US dietary guidance. *Nutr. Rev.* 78, 225–234.

John, R. (1997). Africa: A Biography of the Continent.

Kadwell, M., Fernandez, M., Stanley, H. F., Baldi, R., Wheeler, J. C., Rosadio, R., et al. (2001). Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proc. Biol. Sci.* 268, 2575–2584.

Kehdy, F. S. G., Gouveia, M. H., Machado, M., Magalhães, W. C. S., Horimoto, A. R., Horta, B. L., et al. (2015). Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc. Natl. Acad. Sci. U. S. A.* 112, 8696–8701.

Kiple, K. F., and Ornelas, K. C. (2001). *The Cambridge World History of Food*. Cambridge University Press.

Labrie, V., Buske, O. J., Oh, E., Jeremian, R., Ptak, C., Gasiūnas, G., et al. (2016). Lactase nonpersistence is directed by DNA-variation-dependent epigenetic aging. *Nat. Struct. Mol. Biol.* 23, 566–573.

Lahiri, D. K., and Nurnberger, J. I., Jr (1991). A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res.* 19, 5444.

Latorre, G., Besa, P., Parodi, C. G., Ferrer, V., Azocar, L., Quirola, M., et al. (2014). Prevalence of lactose intolerance in Chile: a double-blind placebo study. *Digestion* 90, 18–26.

Loesch, D., Andrea R V, Heilbron, K., Sarihan, E. I., Inca-Martinez, M., Mason, E., et al. (2020). Characterizing the genetic architecture of Parkinson's disease in Latinos. *medRxiv*, 2020.11.09.20227124.

Luizon, M. R. (2007). Dinâmica da mistura étnica em comunidades remanescentes de quilombo brasileiras. 2007.

Mantei, N., Villa, M., Enzler, T., Wacker, H., Boll, W., James, P., et al. (1988). Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. *EMBO J.* 7, 2705–2713.

Maples, B. K., Gravel, S., Kenny, E. E., and Bustamante, C. D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* 93, 278–288.

Mendes, M., Alvim, I., Borda, V., and Tarazona-Santos, E. (2020). The history behind the mosaic of the Americas. *Curr. Opin. Genet. Dev.* 62, 72–77.

Mendoza Torres, E., Varela Prieto, L. L., Villarreal Camacho, J. L., and Villanueva Torregroza, D. A. (2012). Diagnosis of adult-type hypolactasia/lactase persistence: genotyping of single nucleotide polymorphism (SNP C/T-13910) is not consistent with breath test in Colombian Caribbean population. *Arq. Gastroenterol.* 49, 5–8.

Montalva, N., Adhikari, K., Liebert, A., Mendoza-Revilla, J., Flores, S. V., Mace, R., et al. (2019). Adaptation to milking agropastoralism in Chilean goat herders and nutritional benefit of lactase persistence. *Ann. Hum. Genet.* 83, 11–22.

Morales-Ruán, M. del C., Shamah-Levy, T., Mundo-Rosas, V., Cuevas-Nasu, L., Romero-Martínez, M., Villalpando, S., et al. (2013). [Food assistance programs in Mexico, coverage and targeting]. *Salud Publica Mex.* 55 Suppl 2, S199–205.

Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J. L., Byrnes, J. K., Gignoux, C. R., et al. (2013). Reconstructing the population genetic history of the Caribbean. *PLoS Genet.* 9, e1003925.

Moura, C. (2001). *Os quilombos na dinâmica social do Brasil*. EdUFAL.

Muehlhoff, E., Bennett, A., McMahon, D., and Others (2013). *Milk and dairy products in human nutrition*. Food and Agriculture Organization of the United Nations (FAO).

Nakatsuka, N., Lazaridis, I., Barbieri, C., Skoglund, P., Rohland, N., Mallick, S., et al. (2020). A Paleogenomic Reconstruction of the Deep Population History of the Andes. *Cell* 181, 1131–1145.e21.

Ojeda-Granados, C., Panduro, A., Rebello Pinho, J. R., Ramos-Lopez, O., Gleyzer, K., Malta, F. de M., et al. (2016). Association of Lactase Persistence Genotypes with High Intake of Dairy Saturated Fat and High Prevalence of Lactase Non-Persistence among the Mexican Population. *J. Nutrigenet. Nutrigenomics* 9, 83–94.

Ongaro, L., Scliar, M. O., Flores, R., Raveane, A., Marnetto, D., Sarno, S., et al. (2019). The Genomic Impact of European Colonization of the Americas. *Curr. Biol.* 29, 3974–3986.e4.

Paz-Y-Miño, C., Burgos, G., López-Cortés, A., Herrera, C., Gaviria, A., Tejera, E., et al. (2016). A study of the molecular variants associated with lactase persistence in different Ecuadorian ethnic groups. *Am. J. Hum. Biol.* 28, 774–781.

Pena, S. D. J., Di Pietro, G., Fuchshuber-Moraes, M., Genro, J. P., Hutz, M. H., Kehdy, F. de S. G., et al. (2011). The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* 6, e17063.

Pena, S. D. J., Santos, F. R., and Tarazona-Santos, E. (2020). Genetic admixture in Brazil. *Am. J. Med. Genet. C Semin. Med. Genet.* 184, 928–938.

Porrata-Maury, C. (2004). "Guías Alimentarias para la Población Cubana mayor de dos años de edad," in *Educación Alimentaria y Nutricional e Higiene de los Alimentos. Manual de Capacitación*, ed. Instituto de Nutrición e Higiene de los Alimentos. Ministerio de Salud Pública, Cuba (WFP, PAHO, National Center of Health Promotion and Education, Cuba).

Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T. C., et al. (2018). Reconstructing the Deep Population History of Central and South America. *Cell* 175, 1185–1197.e22.

Primo, A. T. (1992). El ganado bovino ibérico en las Américas: 500 años después. *Arch. Zootec.* 41, 13.

Raggio, A. N. A. Z., Bley, R. B., and Trauczynski, S. C. eds. (2018). *Abordagem historica sobre a populacao negra no Estado do Parana*. Curitiba, Parana: SEJU,.

Ranciaro, A., Campbell, M. C., Hirbo, J. B., Ko, W.-Y., Froment, A., Anagnostou, P., et al. (2014). Genetic origins of lactase persistence and the spread of pastoralism in Africa. *Am. J. Hum. Genet.* 94, 496–510.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., et al. (2012). Reconstructing Native American population history. *Nature* 488, 370–374.

Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., et al. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* 10, e1004572.

Salzano, F. M. (2011). The Prehistoric Colonization of the Americas: Evidence and Models. *Evolution: Education and Outreach* 4, 199–204.

Salzano, F. M., and Bortolini, M. C. (2005). *The Evolution and Genetics of Latin American Populations*. Cambridge University Press.

Sambrook, J., Fritsch, E. F., Maniatis, T., and Others (1989). *Molecular cloning: a laboratory manual*. Cold spring harbor laboratory press.

Sánchez-Alonso, B. (2019). The age of mass migration in Latin America: THE AGE OF MASS MIGRATION IN LATIN AMERICA. *Econ. Hist. Rev.* 72, 3–31.

Secretaria de Salud de México (2010). Guía de Alimentos para la Población Mexicana. {Secretaria de Salud de Mexico} Available at: http://www.imss.gob.mx/sites/all/statics/salud/guia-alimentos.pdf.

Segall, J. J. (1994). Dietary lactose as a possible risk factor for ischaemic heart disease: review of epidemiology. *Int. J. Cardiol.* 46, 197–207.

Ségurel, L., and Bon, C. (2017). On the Evolution of Lactase Persistence in Humans. *Annu. Rev. Genomics Hum. Genet.* 18, 297–319.

Segurel, L., Guarino-Vignon, P., Marchi, N., Lafosse, S., Laurent, R., Bon, C., et al. (2020). Why and when was lactase persistence selected for? Insights from Central Asian herders and ancient DNA. *PLoS Biol.* 18, e3000742.

Smith, G. D., Lawlor, D. A., Timpson, N. J., Baban, J., Kiessling, M., Day, I. N. M., et al. (2009). Lactase persistence-related genetic variant: population substructure and health outcomes. *Eur. J. Hum. Genet.* 17, 357–367.

Soares-Souza, G., Borda, V., Kehdy, F., and Tarazona-Santos, E. (2018). Admixture, Genetics and Complex Diseases in Latin Americans and US Hispanics. *Curr. Genet. Med. Rep.* doi:10.1007/s40142-018-0151-z.

Souza, I. R. de, and Culpi, L. (1992). Valongo, an isolated braziliann black community. I. Structure of the population. *Rev. bras. genét*, 439–447.

Souza, I. R. de, and Culpi, L. (2005). Valongo, genetic studies on an isolated Afro-Brazilian community. *Genet. Mol. Biol.* 28, 402–406.

Stifel, D., and Alderman, H. (2006). The “Glass of Milk” Subsidy Program and Malnutrition in Peru. *World Bank Econ. Rev.* 20, 421–448.

Szpiech, Z. A., and Hernandez, R. D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* 31, 2824–2827.

Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., et al. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 39, 31–40.

Uauy, R., Albala, C., and Kain, J. (2001). Obesity trends in Latin America: transiting from under- to overweight. *J. Nutr.* 131, 893S–899S.

Valencia, L., Randazzo, A., Engfeldt, P., Olsson, L. A., Chávez, A., Buckland, R. J., et al. (2017). Identification of novel genetic variants in the mutational hotspot region 14 kb upstream of the LCT gene in a Mexican population. *Scand. J. Clin. Lab. Invest.* 77, 311–314.

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4, e72.

Wang, Y., Harvey, C., Rousset, M., and Swallow, D. M. (1994). Expression of human intestinal mRNA transcripts during development: analysis by a semiquantitative RNA polymerase chain reaction method. *Pediatr. Res.* 36, 514–521.

Wehling, A., de Macedo Wehling, M. J. C., and da Silva, J. L. W. (1994). *Formação do Brasil colonial*. Nova Fronteira Rio de Janeiro.

Wheeler, J. C. (1995). Evolution and present situation of the South American camelidae. *Biol. J. Linn. Soc. Lond.* 54, 271–295.

Wiley, A. S. (2015). *Cultures of Milk: The Biology and Meaning of Dairy Products in the United States and India*. Cambridge: Harvard University Press.

Zhong, Y., Priebe, M. G., Vonk, R. J., Huang, C.-Y., Antoine, J.-M., He, T., et al. (2004). The role of colonic microbiota in lactose intolerance. *Dig. Dis. Sci.* 49, 78–83.

Figures

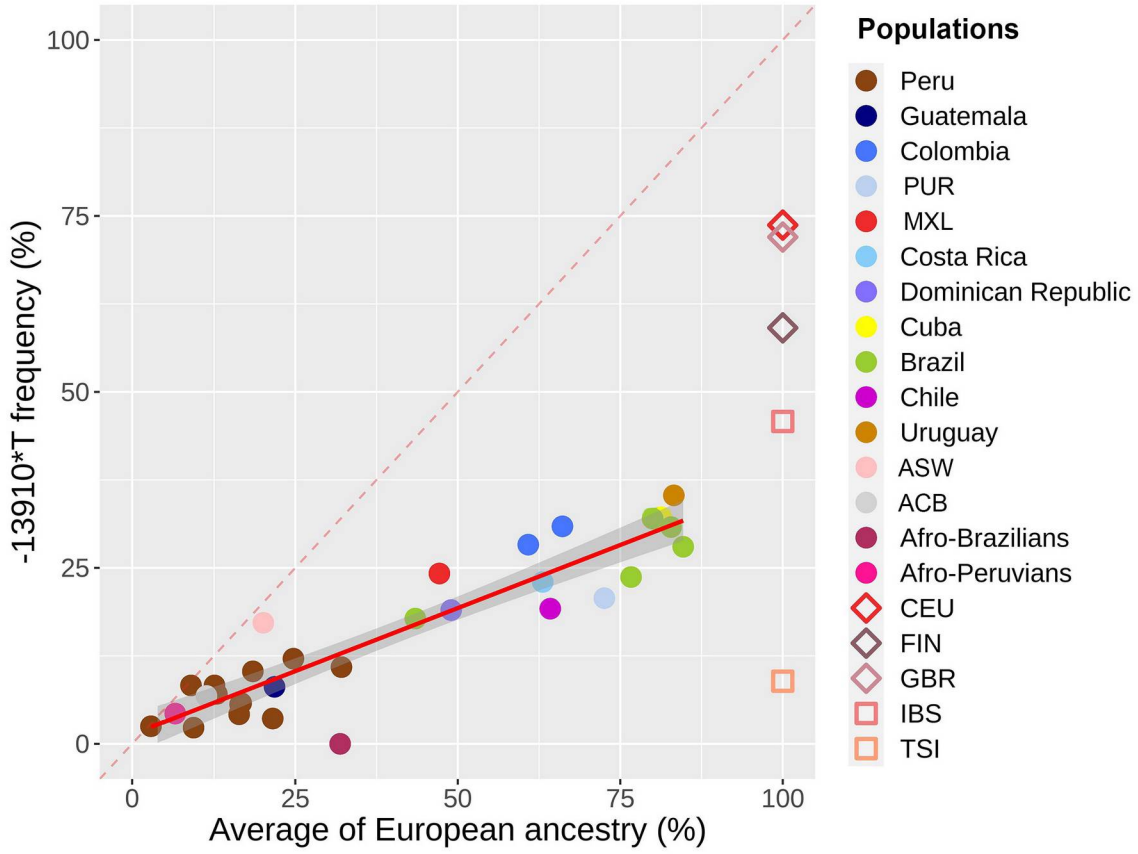


Figure 1. **The *–13910*T* allele frequency for lactase persistence vs. average of European ancestry in 13 Latin American countries and five European populations**. European data (CEU, FIN, GBR, IBS, TSI) belong to the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015). Colombia includes CLM and Bogota individuals. Diamonds and squares refer to Northern and Southern European populations, respectively. The red line between the samples represents a linear regression for the Pan-American populations.

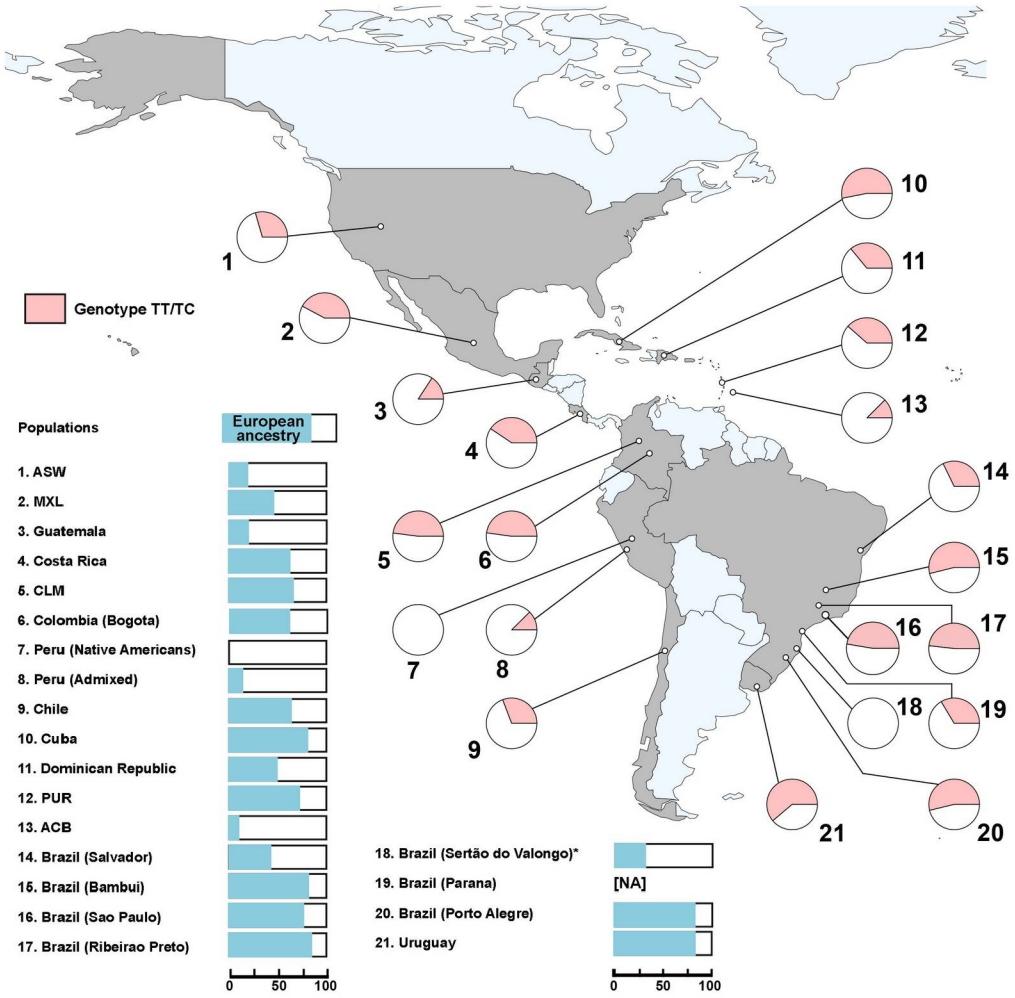


Figure 2. Geographical distribution of *TT* and *TC* genotypes of *–13910C>T* along the Americas. The red color in the pie charts represents the combined frequencies of *TT* and *TC* genotypes in each population. On the left, the blue color in the bar plots indicates the average proportion of European ancestry for each population. NA: No available data. ASW: Afro-Americans from Southwest USA. MXL: Mexican Ancestry from Los Angeles, USA. ACB: Afro-Caribbeans from Barbados. PUR: Puerto Ricans from Puerto Rico. CLM: Colombians from Medellin. *European ancestry proportion for Afro-Brazilians from Sertão do Valongo was estimated using different samples from the same population from Luizon (2007).

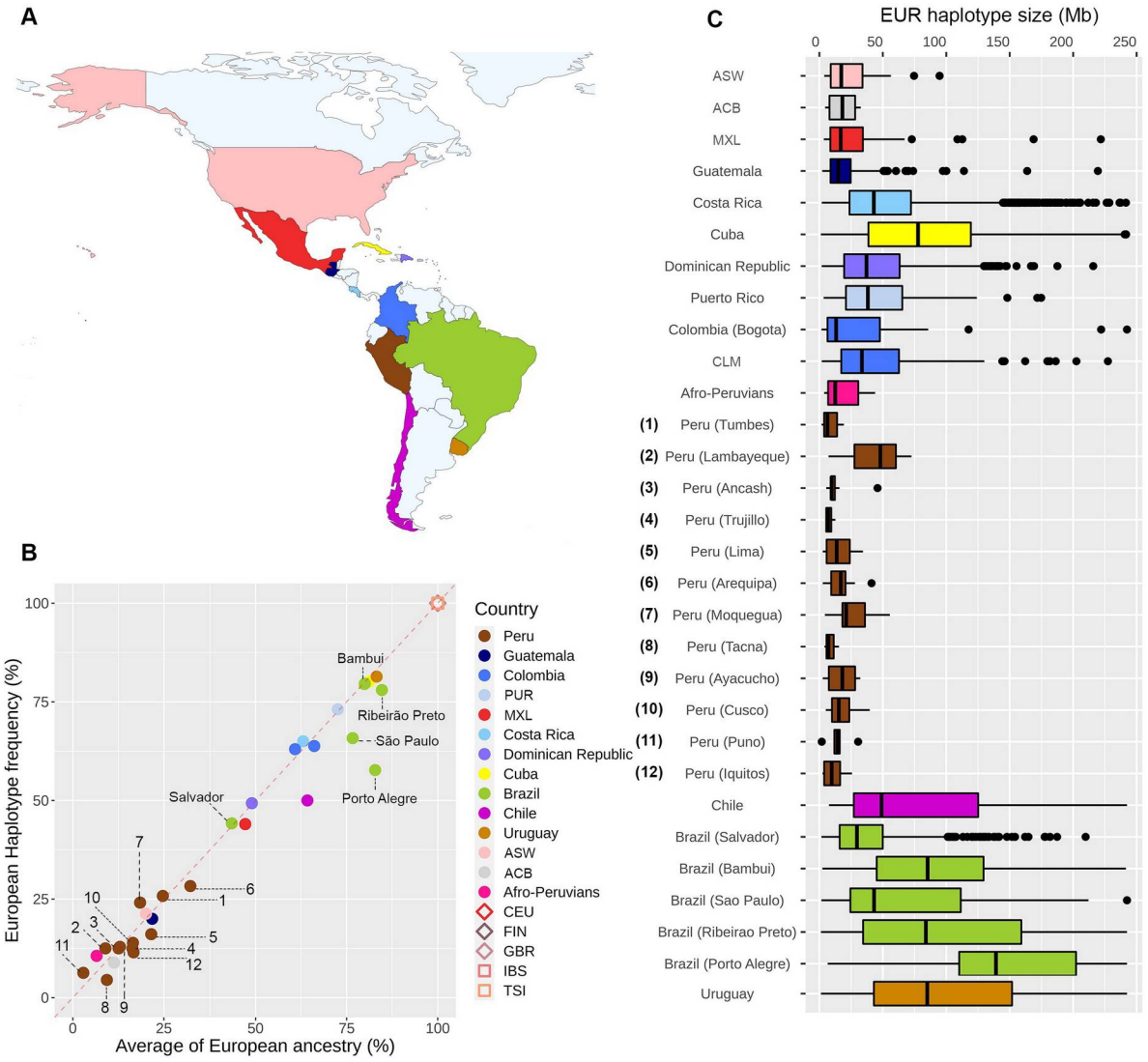


Figure 3. Distribution of counts and lengths of haplotypes of European ancestry in the vicinity of the *MCM6* gene along the Americas. A) Colors indicate the countries included in the analyses. B) European haplotype frequency vs. the average of European ancestry for each population (Uruguay and Bambui dots are overlapping Cuba dot). European data (CEU, FIN, GBR, IBS, TSI) belong to the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015). Colombia includes CLM and Bogota individuals. Diamonds and squares refer to Northern and Southern European populations, respectively (European data are overlapped). C) Length distribution of European haplotypes that include the *MCM6* gene in Pan-American admixed populations.

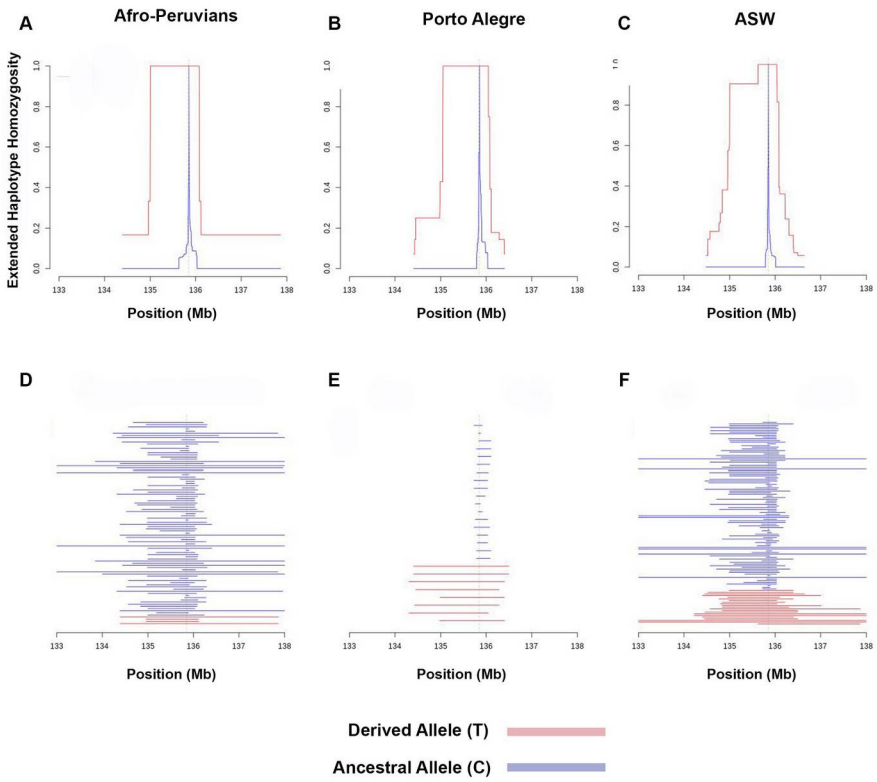


Figure 4. Extended Haplotype Homozygosity analyses. Haplotype lengths for the derived and ancestral haplotypes with *–13910*T* as the core allele in A) Afro-Peruvians, B) Porto Alegre, and C) ASW. Extended Haplotype Homozygosity with *–13910*T* as the core allele in D) Afro-Peruvians, E) Porto Alegre, and C) Afro-Americans.

7.2 POPULAÇÕES DE ANCESTRALIDADE AFRICANA DO SUL DO BRASIL

7.2.1 Frequências alélicas e genotípicas

Para o cálculo das frequências alélicas e genotípicas, foram considerados 259 indivíduos dos 276 sequenciados. A redução do número de indivíduos deve-se a exclusão daqueles que possuem parentes consanguíneos até 1° grau (conforme parentesco declarado pelos participantes) — com exceção dos indivíduos da CQ Sertão do Valongo, que são consanguíneos devido à estrutura da população — e daqueles que tiveram o DNA utilizado como controles das reações de

sequenciamento (genótipo já conhecido). As análises foram realizadas de acordo com os grupos afro-brasileiros de Curitiba e da Região Metropolitana (n = 241) e da comunidade quilombola de Sertão do Valongo, Porto Belo (n = 18). Na TABELA 10 pode-se observar as frequências alélicas, genotípicas e fenotípicas das variantes identificadas nas populações de Curitiba e da Região Metropolitana e da comunidade quilombola de Sertão do Valongo.

TABELA 10: FREQUÊNCIAS ALÉLICAS, GENOTÍPICAS E FENOTÍPICAS EM AFRO-BRASILEIROS DE CURITIBA E REGIÃO METROPOLITANA (PR) E DA COMUNIDADE QUILOMBOLA DE SERTÃO DO VALONGO (SC)

Curitiba e Região Metropolitana (PR)											
Variação	Alelos				Genótipos						Equilíbrio de Hardy-Weinberg
	N	%	N	%	N	%	N	%	N	%	
–13730T>G (rs4954492)*	*T* 467	97,29	*G* 13	2,71	*TT* 228	95,00	*TG* 11	4,58	*GG* 1	0,42	$p = 0,138$
–13732C>T (rs182609517)	*C* 481	99,79	*T* 1	0,21	*CC* 240	99,59	*CT* 1	0,41	*TT* 0	-	$p = 0,974$
–13752C>A (rs893578023)	*C* 481	99,79	*A* 1	0,21	*CC* 240	99,59	*CA* 1	0,41	*AA* 0	-	$p = 1$
–13910C>T (rs4988235)	*C* 392	81,33	***T*** 90	18,67	*CC* 160	66,39	***CT*** 72	29,88	***TT*** 9	3,73	$p = 0,829$
–13915T>G (rs41380347)	*T* 481	99,79	***G*** 1	0,21	*TT* 240	99,59	***TG*** 1	0,41	***GG*** 0	-	$p = 0,974$
–14011C>T (rs4988233)	*C* 481	99,79	***T*** 1	0,21	*CC* 240	99,59	***CT*** 1	0,41	***TT*** 0	-	$p = 1$
Frequência fenotípica					LNP 159	65,98%	LP 82	34,02%			-
Comunidade quilombola de Sertão do Valongo (SC)											
Variação	Alelos				Genótipos						Equilíbrio de Hardy-Weinberg
	N	%	N	%	N	%	N	%	N	%	
–13730T>G (rs4954492)	*T* 30	83,33	*G* 6	16,67	*TT* 12	66,67	*TG* 6	33,33	*GG* 0	-	$p = 1$
Frequência fenotípica					LNP 18	100%	LP 0	0%			-

* A anotação do genótipo para a posição em um indivíduo não foi possível, assim, foram contabilizados 480 alelos referentes a 240 indivíduos. Destacados em negrito os alelos e genótipos associados ao fenótipo lactase persistente. N: número de alelos observados; %: frequência alélica e genotípica. FONTE: a autora (2020).

As rotas portuguesas utilizadas durante o tráfico negreiro transatlântico partiam principalmente de Costa da Mina (atualmente corresponde a região de Gana, Togo, Benin e Nigéria) e Costa da Angola (atualmente corresponde a região Norte da Angola) (SOUZA, 2008). Ainda, a partir do século XVIII porção considerável de africanos escravizados foram trazidos da África Oriental, principalmente de

Moçambique (SOUZA, 2008). De forma geral, os povos traficados através dos portos de Costa da Mina e Angola eram do grupo Bantu de etnias diversas, enquanto os povos traficados através dos portos de Moçambique eram majoritariamente do grupo Iorubá, falantes da língua nigero-congolesa, principalmente (SOUZA, 2008).

Assim, esperávamos observar variantes africanas da costa central-sul e leste (SOUZA, 2008; GOUVEIA *et al.*, 2020). O alelo *–14010*C* (*–14010G>C*, rs145946881) está presente em populações da Tanzânia, Quênia e na região sul do continente africano — populações falantes da língua Bantu da região sudoeste da Angola e em grupos de etnia Xhosa da África do Sul (FIGURA 8), enquanto os alelos *–13915*G* (*–13915T>G*, rs41380347) e *–13907*G* (*–13907C>G*, rs41525747) estão restritos à porção sudeste do continente africano, também conhecida como "Chifre da África" (FIGURA 8). Ainda, o alelo *–13915*G* possui alta frequência na península árabe e Oriente Médio (FIGURA 8) (INGRAM *et al.*, 2007; ENATTAH *et al.*, 2007; TISHKOFF *et al.*, 2007; COELHO *et al.*, 2009; TORNIAINEN *et al.*, 2009; RANCIARO *et al.*, 2014).

Para os afro-brasileiros da região de Curitiba, foram identificadas as variações *–13730T>G* (rs4954492), *–13910C>T* (rs4988235), *–13915T>G* (rs41380347), *–13732C>T* (rs182609517), *–13752C>A* (rs893578023) e *–14011C>T* (rs4988233). O alelo *–13730*G*, exclusivo de populações africanas e afrodescendentes, não possui associação confirmada com o fenótipo LP, bem como os alelos *–13732*T* e *–13752*A*, identificados em populações de origem africana. O alelo *–13910*T*, de origem europeia, foi o alelo derivado observado em maior frequência na população avaliada. Os alelos *–13915*G*, presente em populações do Oriente Médio e da África Oriental, e *–14011*T*, de origem não determinada (reportado em baixas frequências em populações europeias, africanas e do Oriente Médio), foram observados em indivíduos heterozigotos distintos (genótipos *TG* e *CT*, respectivamente); em ambos os casos as reações de PCR e sequenciamento foram refeitas de modo a confirmar a presença dos alelos *–13915*G* e *–14011*T*. Para as demais posições, todos os indivíduos apresentaram o genótipo homozigoto para os alelos ancestrais, que está associado ao fenótipo LNP (TABELA 10).

Já para afro-brasileiros da comunidade quilombola de Sertão do Valongo — pela primeira vez avaliada para o gene *MCM6* —, foram observadas variações apenas em *–13730T>G* (rs4954492). Não foram observadas variações nas demais

posições, portanto, todos os indivíduos apresentaram o genótipo homozigoto para os alelos ancestrais, que estão associados ao fenótipo LNP (TABELA 10). Tanto para a população afro-brasileira de Curitiba, quanto para a população quilombola de Sertão do Valongo, as frequências genotípicas dos SNPs identificados encontram-se em equilíbrio de Hardy-Weinberg (TABELA 10).

O único polimorfismo em comum entre as populações avaliadas foi *–13730T>G* (rs4954492). O alelo *–13730*G*, que não possui associação confirmada com o fenótipo LP (ENATTAH *et al.*, 2008; JONES *et al.*, 2013), foi observado nas frequências de 2,71% e 16,67% em afro-brasileiros da região de Curitiba e da CQ Sertão do Valongo, respectivamente (TABELA 10). Segundo o *1000 Genomes Project* fase 3, o alelo ancestral *–13730*T* possui frequência de 100% em populações europeias e asiáticas e o alelo derivado *–13730*G* pode ser observado em frequência de 8% nas populações africanas e de ancestralidade africana nas Américas e em menos de 1% nas populações americanas (THE 1000 GENOMES PROJECT CONSORTIUM, 2015). Ranciaro e colaboradores (2014) observaram o alelo *–13730*G* nas frequências de 11,79%, 9,62%, 7,02% e 3,49% na África Central, Ocidental, do Norte e Oriental, respectivamente (RANCIARO *et al.*, 2014). Assim, a variação *–13730T>G* é exclusiva de populações africanas e afrodescendentes e está amplamente distribuída pelo continente africano.

Embora frequências similares sejam esperadas em indivíduos não miscigenados e não necessariamente indicam que as populações comparadas possuem origem geográfica em comum (também deve-se considerar fatores evolutivos atuando de formas distintas entre as populações), os resultados obtidos refletem os processos de miscigenação que deram origem à população brasileira. Comparando a frequência alélica (*–13730*G*) observada na população de Curitiba (2,71%; TABELA 10) com a frequência alélica reportada em populações africanas por Ranciaro e colaboradores (2014), temos frequências estatisticamente diferentes para a África Central, Ocidental, do Norte (p = 4,006 x 10^{-8}, p = 0,0235, p = 0,0423, respectivamente) e similar (p = 0,5216) à frequência observada nas populações da África Oriental. Dessa forma, a presença da variação africana (*–13730T>G*) indica que a população afro-brasileira da região de Curitiba possui porção de ancestralidade africana na região genômica investigada. A partir da aplicação da fórmula explicitada no tópico 6.5.3 deste trabalho, as populações africana e europeia

originais contribuíram para a formação da população brasileira em questão, com relação à posição *–13730*, em 78% e 22%, respectivamente.

Na população afro-brasileira da CQ Sertão do Valongo, o único SNP observado foi *–13730T>G*, que não possui associação confirmada com o fenótipo LP. O alelo *–13730*G*, na população em questão, possui frequência de 16,67%, a mais alta já registrada, devido provavelmente ao grau de parentesco entre os indivíduos avaliados e aos fatores evolutivos atuando sobre a população. Quando comparamos as frequências alélicas observada em Sertão do Valongo com a reportada para a África Central — de 11,79%, a mais alta até então registrada (RANCIARO *et al.*, 2014) —, temos frequências estatisticamente semelhantes (p = 0,423). Entretanto, comparando a frequência alélica de 16,67% com as frequências registradas para as demais populações africanas (RANCIARO *et al.*, 2014), também observamos frequências estatisticamente semelhantes entre a África Ocidental e do Norte (p = 0,3461 e p = 0,1019, respectivamente) e diferente entre a África Oriental (p = 0,0022). Novamente, frequências similares seriam esperadas em indivíduos não miscigenados e não necessariamente indicam que as populações comparadas possuem origem geográfica em comum.

Esses resultados são compatíveis com a sugestão de que a população da CQ Sertão do Valongo passou por efeito fundador no passado, com forte efeito de deriva genética e possível aumento da frequência alélica observada. Possivelmente a população fundadora possuía *pool* gênico no qual não estavam presentes alelos associados ao fenótipo LP e o baixo fluxo gênico (desde a fundação em 1880, até a data da coleta entre 1984 e 1985) com outras populações, o que pode ter impossibilitado a introdução desses alelos na população atual da CQ Sertão do Valongo. Dessa forma, não é possível descartar a hipótese de que os alelos associados ao fenótipo LP estivessem presentes na população fundadora em baixa frequência, ou ainda, é possível que esses alelos estejam presentes na população atual em frequência não detectável pela amostragem realizada, igual ou inferior a 2,78% (caracterizando o alelo derivado como polimorfismo).

Para afro-brasileiros da região de Curitiba, alelo *–13910*T*, de origem europeia, possui frequência de 18,67% (TABELA 10), similar (p = 0,9216) a frequência de 18,4% relatada por Friedrich e colaboradores (2012a) em populações afro-brasileiras de Porto Alegre (RS), também do sul do Brasil. Contudo, quando

comparada às frequências registradas para euro-brasileiros de Pelotas (RS), Porto Alegre (RS) e Curitiba (PR), de 38,7%, 29,5% e 33,0% (HARTWIG *et al.*, 2016; FRIEDRICH *et al.*, 2012a; BOSCHMANN *et al.*, 2016), respectivamente, temos frequências estatisticamente diferentes ($p \leq 0,001$ em todos os casos). A diferença estatística observada evidencia contraste com relação à prevalência do fenótipo LP em populações de ancestralidades distintas, nesse caso, proporções distintas de ancestralidade africana e europeia. A partir da aplicação da fórmula explicitada no tópico 6.5.3 deste trabalho, as populações africana e europeia originais contribuíram para a formação da população brasileira em questão, com relação à posição *–13910*, em 70% e 30%, respectivamente.

A variação *–13915T>G*, possui origem na península arábica e Oriente Médio, porém é observada em populações africanas da região leste (ENATTAH *et al.*, 2007; RANCIARO *et al.*, 2014). Pela primeira vez identificamos o alelo *–13915*G* na população brasileira, especificamente em um único indivíduo heterozigoto de genótipo *TG*, na frequência alélica de 0,21% (TABELA 10). O alelo *–13915*G* pode ter sido introduzido na população brasileira pela miscigenação com povos africanos a partir da diáspora africana. Entretanto, deve-se notar a presença de imigrantes árabes no Brasil, a partir da segunda metade do século XIX a migração de povos sírios e libaneses foi intensificada (ANBA, 2020). Atualmente, segundo dados da Câmara de Comércio Árabe-Brasileira (ANBA, 2020), 11,61 milhões de brasileiros são descendentes de árabes (auto-declaração), desses, cerca de 27% possuem ascendência libanesa e 13% possuem ascendência síria. Dessa forma, o fluxo gênico entre povos árabes também pode ter contribuído para a introdução do alelo *–13915*G* na população brasileira.

Em frequências similares à observada para a variação *–13915T>G*, identificamos as substituições *–13732C>T*, *–13752C>A* e *–14011C>T*, cada uma em um indivíduo (0,21%, TABELA 10). As três substituições já foram descritas na literatura, porém, apenas o alelo *–14011*T* possui associação com o fenótipo LP confirmada *in vitro* (LIEBERT *et al.*, 2016). Segundo o *1000 Genomes Project* fase 3, a variação *–13732C>T* possui frequência de 0,8%, 0,5% e 1,2% em afro-americanos do sudoeste dos Estado Unidos, indivíduos de etnia Luhya em Webuye no Quênia, e indivíduos de etnia Mende em Serra Leoa, respectivamente (THE 1000 GENOMES PROJECT CONSORTIUM, 2015). Já a variação *–13752C>A*, possui frequência de

0,3% em africanos e afro-americanos presentes no banco de dados *gnomAD* versão 3 (KARCZEWSKI e MARTIN, 2020). Quanto à variação *–14011C>T*, o alelo *T* foi reportado em maior frequência, 2,5%, na população de Kgalagadi na Botsuana (MACHOLDT *et al.*, 2014). No Brasil, o alelo *–14011*T* já foi observado em Recife (PE) e em Belém (PA), nas frequências de 0,8% e 0,5%, respectivamente (FRIEDRICH *et al.*, 2012a), porém, a origem do alelo não foi identificada, dado que já foi reportado em populações da Europa, África e Oriente Médio (LIEBERT *et al.*, 2016). Quando comparamos a frequência observada de *–14011*T* em afro-brasileiros da região de Curitiba (0,21%) com as frequências reportadas em Recife e Belém, não temos diferença significativa (p = 0,210 e p = 0,458, respectivamente). Assim, a frequência alélica observada na população de Curitiba concorda com a baixa frequência reportada em outras regiões brasileiras.

O alelo *–14010*C* é observado em baixas frequências em populações angolanas falantes das línguas Ovimbundu (1%), Nyaneka-Nkhumbi (3%) e Kuvale (6%), todas da região sudoeste, e está ausente em indivíduos de etnias distintas avaliados em Moçambique (COELHO *et al.*, 2009). Entretanto, é observado nas frequências de 46% e 39% nas populações da região afro-asiática e nilo-saariana da Tanzânia, respectivamente (TISHKOFF *et al.*, 2007). O alelo *–14010*C* foi observado por Friedrich e colaboradores (2012a) em um indivíduo afro-brasileiro de Porto Alegre (RS) heterozigoto para a posição (*GC*), porém não foi observado nas populações avaliadas neste trabalho. A baixa frequência relatada em Porto Alegre e a ausência do alelo *–14010*C* em Curitiba pode ser explicada pela ocorrência de deriva genética: as populações ancestrais poderiam possuir a variação *–14010G>C* em baixa frequência e, com o passar do tempo, o alelo derivado pode ter sido perdido por deriva genética (na população atual) ou estar presente em frequência não detectável pela amostragem realizada neste trabalho, inferior a 0,21% (alelo raro na população, diferente do observado em Sertão do Valongo). A ausência de outras variantes africanas associadas ao fenótipo LP em ambas as populações avaliadas neste trabalho pode ser atribuída aos mesmos fatores elencados para justificar a baixa frequência do alelo *–14010*C* em Porto Alegre e ausência dele em Curitiba.

7.2.2 Inferência de fenótipos lactase persistente e lactase não persistente

Para a inferência de fenótipos, foram considerados os alelos *–13779*C*, *–13907*G*, *–13910*T*, *–13913*C*, *–13915*G*, *–13937*A*, *–14009*G*, *–14010*C* e *–14011*T* que estão associados ao fenótipo LP (validados funcionalmente ou com evidências de função, TABELA 1).

Para os afro-brasileiros da região de Curitiba, dos 241 indivíduos avaliados, 81 possuem o alelo europeu *–13910*T* (72 de genótipo *CT* e 9 de genótipo *TT*), um possui o alelo africano *–13915*G* (genótipo *TG*) e um possui o alelo *–14011*T* (genótipo *CT*) de origem não determinada (reportada em populações da Europa, África e Oriente Médio). Como o indivíduo que possui o alelo *–14011*T* também possui o alelo *–13910*T*, no total, 82 indivíduos possuem o fenótipo LP, sendo a frequência fenotípica observada de 34,02% (TABELA 10). Enquanto 159 indivíduos (65,98%, TABELA 10) possuem o fenótipo LNP e podem desenvolver sintomas associados à intolerância à lactose ao consumirem leite fresco e derivados. A frequência para o fenótipo LP em indivíduos miscigenados (ancestralidade europeia e africana) e afro-brasileiros, descrita por Mattar e colaboradores (2009), foi de 43,1% e 20%, respectivamente. Quando comparadas com a frequência fenotípica observada na população de afro-brasileiros da região de Curitiba, temos frequências estatisticamente semelhantes (p = 0,1771 e p = 0,0654).

O consumo de leite no Brasil, em 2018, foi em média de 170 litros *per capita* (RENTERO *et al.*, 2019), valor inferior ao registrado em países da União Europeia, nos quais o consumo *per capita* durante o mesmo período foi entre 250 e 300 litros (IFCN, 2018; GDP, 2019), e superior ao registrado em alguns países africanos (não especificados) nos quais o consumo *per capita* é, em média, de 30 litros de leite ao ano (SIQUEIRA, 2019). Entretanto, em populações pastoralistas africanas o consumo de leite pode representar até 90% da ingestão calórica total, com porcentagens variando entre populações e de acordo com a época do ano (LITTLE, GRAY e CAMPBELL, 2001). Com algumas exceções, o consumo de leite está diretamente relacionado ao poder aquisitivo da população e à frequência do fenótipo LP. À medida que a população possui maior renda *per capita* o consumo de leite aumenta (SIQUEIRA, 2019), com maior consumo de leite observa-se maior frequência do fenótipo LP. Assim, a frequência fenotípica observada na população

afro-brasileira da região de Curitiba (34,02%) reflete a ancestralidade miscigenada da população, sendo intermediária quando comparada às frequências fenotípicas registradas em populações africanas e europeias, nas quais o consumo de leite *per capita* é inferior e superior, respectivamente.

Para afro-brasileiros da comunidade quilombola de Sertão do Valongo, nenhum dos 18 indivíduos avaliados é portador de alelos associados ao fenótipo LP — frequência fenotípica inferida de 0% (TABELA 10). A frequência para o fenótipo LP em indivíduos miscigenados (ancestralidade europeia e africana) e afro-brasileiros, descrita por Mattar e colaboradores (2009), foi de 43,1% e 20%, respectivamente. Quando comparadas com a frequência fenotípica observada na população de afro-brasileiros da CQ Sertão do Valongo, temos frequências estatisticamente diferente (p = 0,0004) e semelhante (p = 0,5237), respectivamente. Ainda, comparando as frequências fenotípicas observadas em afro-brasileiros da região de Curitiba (34,02%) e da CQ Sertão do Valongo (0%), também temos frequências estatisticamente diferentes (p = 0,0011). Os resultados obtidos são devidos, principalmente, ao histórico de fundação (deriva genética) e isolamento (baixo fluxo gênico), citados anteriormente, para a população da CQ Sertão do Valongo. Contudo, pôde-se observar que a população em questão possui semelhanças e diferenças com as demais populações de ancestralidade africana comparadas.

Em contraste com a população da região de Curitiba, a CQ Sertão do Valongo subsiste da agricultura familiar, principalmente, da plantação de cana-de-açúcar e mandioca — cultura agriculturalista (WANDERLEY, 2001; CASTELLS, 2008). Logo, o consumo de leite fresco e derivados não é recorrente e a digestão da lactose não é essencial para indivíduos adultos. Ainda, a correlação positiva observada entre a frequência do alelo *–13910*T* e a ancestralidade europeia média, os padrões de decaimento da homozigose dos haplótipos observados em populações pan-americanas (tópico 7.1 deste trabalho) e o histórico de formação delas (tópicos 2.5 deste trabalho) sugerem que fatores demográficos possuem papel mais importante na distribuição do fenótipo LP nas Américas do que a seleção natural. Assim, é possível que os alelos associados ao fenótipo LP estivessem presentes, mesmo que em baixa frequência, na população fundadora da CQ Sertão do Valongo, porém foram perdidos por deriva genética.

7.2.3 Inferência de haplótipos e desequilíbrio de ligação

Os haplótipos foram avaliados considerando os marcadores haplotípicos *–678A>G*, *+666G>A* e *+5579T>C* descritos por Hollox e colaboradores (2001) e as variações *–13730T>G*, *–13732C>T*, *–13752C>A*, *–13779G>C*, *–13907C>G*, *–13910C>T*, *–13913T>C*, *–13915T>G*, *–13937G>A*, *–14009T>G*, *–14010G>C* e *–14011C>T*.

Atualmente são empregados onze marcadores haplotípicos localizados no gene *LCT* para a definição dos contextos haplotípicos do gene *MCM6* (HARVEY *et al.*, 1995; HARVEY *et al.*, 1998; HOLLOX *et al.*, 2001). A FIGURA 3 apresenta os marcadores haplotípicos em questão e a FIGURA 4 apresenta os hapótipos definidos por esses marcadores e suas possíveis combinações. Para esse trabalho foram selecionados os marcadores *–678A>G*, *+666G>A* e +5579T>C, pois eles diferenciam individualmente os haplótipos *A*, *B* e *C*, respectivamente. Esses haplótipos são os mais frequentes mundialmente (HOLLOX *et al.*, 2001), com frequências de 45% (*A*), 18% (*B*) e 15% (*C*). Ademais, os marcadores *–678A>G* e *+5579T>C* determinam o haplótipo *E* (frequência mundial de 2%), quando observados em conjunto (HOLLOX *et al.*, 2001). Devido a diversidade haplotípica já descrita e ao número reduzido de marcadores selecionados, não será possível confirmar a nomenclatura de todos os haplótipos identificados e, provavelmente, os haplótipos observados em menor frequência serão agrupados. Ainda, é necessário notar que o haplótipo *U*, que possui frequência mundial de 5% (logo após os haplótipos *A*, *B* e *C*; HOLLOX *et al.*, 2001), não pôde ser diferenciado pelos marcadores selecionados.

Em afro-brasileiros da região de Curitiba foram identificados vinte haplótipos distintos em 392 cromossomos avaliados, correspondentes a 196 indivíduos — a redução do número amostral é devida a falhas no protocolo empregado para a genotipagem dos marcadores (tópico 6.4 deste trabalho) e a impossibilidade de repetição em decorrência da pandemia de COVID-19. Na TABELA 11 pode-se observar os haplótipos identificados, a frequência haplotípica para cada um deles, o número de cromossomos nos quais o haplótipo foi observado e a sua provável nomenclatura.

TABELA 11: FREQUÊNCIA HAPLOTÍPICA E HAPLÓTIPOS IDENTIFICADOS EM AFRO-BRASILEIROS DE CURITIBA E REGIÃO METROPOLITANA (PR)

Afro-brasileiros de Curitiba (n = 196)																		
Identificação	Nomenclatura*	Haplótipos															Frequência (%)	Número de cromossomos
		-678	666	5579	-13730	-13732	-13752	-13779	-13907	-13910	-13913	-13915	-13937	-14009	-14010	-14011		
Hap_1	*B*, ***D*****, *U*	*A*	*A*	*T*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	23,724	93
Hap_2	***A***, *J*	*A*	*G*	*C*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	18,112	71
Hap_3	***C***, *M*	*G*	*G*	*T*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	17,602	69
Hap_4	***A***, *J*	*A*	*G*	*C*	*T*	*C*	*C*	*G*	*C*	*T*	*T*	*T*	*G*	*T*	*G*	*C*	16,837	66
Hap_5	*H, K, L, Q, T, Z*	*A*	*G*	*T*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	6,633	26
Hap_6	-	*G*	*A*	*T*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	4,847	19
Hap_7	*F, I, V, X*	*A*	*A*	*C*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	2,806	11
Hap_8	*H, K, L, Q, T, Z*	*A*	*G*	*T*	*G*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	2,296	9
Hap_9	***E****	*G*	*G*	*C*	*T*	*C*	*C*	*G*	*C*	*T*	*T*	*T*	*G*	*T*	*G*	*C*	1,786	7
Hap_10	-	*G*	*A*	*C*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	1,531	6
Hap_11	***E****	*G*	*G*	*C*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	0,765	3
Hap_12	*H, K, L, Q, T, Z*	*A*	*G*	*T*	*T*	*C*	*C*	*G*	*C*	*T*	*T*	*T*	*G*	*T*	*G*	*C*	0,765	3
Hap_13	*F, I, V, X*	*A*	*A*	*C*	*T*	*C*	*C*	*G*	*C*	*T*	*T*	*T*	*G*	*T*	*G*	*C*	0,510	2
Hap_14	*B*, ***D*****, *U*	*A*	*A*	*T*	*G*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	0,255	1
Hap_15	***C***, *M*	*G*	*G*	*T*	*T*	*T*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	0,255	1
Hap_16	*H, K, L, Q, T, Z*	*A*	*G*	*T*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*T*	0,255	1
Hap_17	***A***, *J*	*A*	*G*	*C*	*G*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	0,255	1
Hap_18	***C***, *M*	*G*	*G*	*T*	*T*	*C*	*A*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	0,255	1
Hap_19	***C***, *M*	*G*	*G*	*T*	*G*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	0,255	1
Hap_20	-	*G*	*A*	*T*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*G*	*G*	*T*	*G*	*C*	0,255	1

Sublinhados estão os alelos derivados para cada uma das variantes. Em vermelho, os alelos associados ao fenótipo lactase persistente. Em negrito, os haplótipos mais prováveis. * Nomenclatura de acordo com Hollox e colaboradores (2001). ** Haplótipos ancestrais de acordo com Tishkoff e colaboradores (2007). FONTE: a autora (2021).

Em afro-brasileiros da região de Curitiba, os haplótipos 1, 2, 3 e 4 foram observados em maior frequência (TABELA 10): 23,72%, 18,11%, 17,60% e 16,84%, respectivamente. Provavelmente, esses haplótipos são correspondentes aos haplótipos *D*, *A*, *C* e *A*, respectivamente. Entretanto, se agruparmos as classes de acordo com a nomenclatura sugerida, temos: haplótipo *A* na frequência de 35,20%; haplótipo *D*, 23,98%; haplótipo *C*, 18,37%; haplótipo *E*, 2,55%; e demais haplótipos (classes *H*, *K*, *L*, *Q*, *T*, *Z* ou *F*, *I*, *V*, *X*), 19,90%. Os haplótipos *D* e *E*, já foram reportados em populações africanas, europeias e asiáticas, entretanto, em populações africanas são considerados ancestrais, por não conterem as variantes associadas ao fenótipo LP (HARVEY *et al.*, 1998; HOLLOX *et al.*, 2001; TISHKOFF *et al.*, 2007). Dessa forma, presumimos que os haplótipos 3 e 8 são correspondentes

ao haplótipo *D* (que não pôde ser diferenciado de *B*), e os haplótipos 9 e 13, ao haplótipo *E* (TABELA 11).

Em afro-brasileiros da comunidade quilombola de Sertão do Valongo foram avaliados 10 cromossomos (5 indivíduos), nos quais observamos quatro haplótipos distintos (TABELA 12), idênticos aos haplótipos 5, 3, 8 e 2, respectivamente, observados em afro-brasileiros da região de Curitiba (TABELA 11). Novamente, o número amostral reduzido é devido a falhas no protocolo empregado para a genotipagem dos marcadores (tópico 6.4 deste trabalho) e a impossibilidade de repetição em decorrência da pandemia de COVID-19.

TABELA 12: FREQUÊNCIA HAPLOTÍPICA E HAPLÓTIPOS IDENTIFICADOS EM AFRO-BRASILEIROS DA COMUNIDADE QUILOMBOLA DE SERTÃO DO VALONGO (SC)

Afro-brasileiros da comunidade quilombola de Sertão do Valongo (n = 5)																		
Identificação	Nomenclatura*	Haplótipos															Frequência (%)	Número de cromossomos
		-678	*666*	*5579*	*-13730*	*-13732*	*-13752*	*-13779*	*-13907*	*-13910*	*-13913*	*-13915*	*-13937*	*-14009*	*-14010*	*-14011*		
Hap_5	*H, K, L, Q, T, Z*	*A*	*G*	*T*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	40,00	4
Hap_3	***C**, M*	*G*	*G*	*T*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	20,00	2
Hap_8	*H, K, L, Q, T, Z*	*A*	*G*	*T*	*G*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	20,00	2
Hap_2	***A**, J*	*A*	*G*	*C*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	20,00	2

Sublinhados estão os alelos derivados para cada uma das variantes. Destacados em negrito, os haplótipos mais prováveis. * Nomenclatura de acordo com Hollox e colaboradores (2001). FONTE: a autora (2021).

Curiosamente, os quatro haplótipos identificados na população de Sertão do Valongo (TABELA 12) contém os alelos ancestrais inseridos em contextos distintos de *D* e *E*. Os haplótipos mais frequentes, presentes em 60% dos cromossomos avaliados, não puderam ser distinguidos (classe *H*, *K*, *L*, *Q*, *T, Z*). Como os marcadores haplotípicos utilizados não permitem diferenciar o haplótipo *D* de *B* e *U*, nem os haplótipos *H*, *K*, *L*, *Q*, *T* e *Z*, seria necessária a genotipagem dos demais marcadores disponíveis para possibilitar a diferenciação completa dos haplótipos observados em Sertão do Valongo e em Curitiba. Para ambas as populações não foram observados haplótipos contendo mais de uma variação africana, ou variação africana e europeia em conjunto.

Na população afro-brasileira de Curitiba, o alelo *–13910*T* foi observado em quatro contextos haplotípicos distintos: 4 (frequência de 16,84%) e 9 (1,79%), provavelmente correspondentes aos haplótipos *A* e *E*, respectivamente; 12

(0,77%) e 13 (0,51%), que não puderam ser diferenciados pelos marcadores avaliados (TABELA 11). O haplótipo *A* trata-se de um haplótipo longo de 1 Mb (ENATTAH *et al.*, 2002; POULTER *et al.*, 2003) reportado em populações europeias em frequências variando de 86,5% a 37,7% entre as regiões norte e sul (HARVEY *et al.*, 1998; HOLLOX *et al.*, 2001); e o haplótipo *E*, frequência mundial de 2%, não contém variantes associadas ao fenótipo LP (HARVEY *et al.*, 1998; HOLLOX *et al.*, 2001; TISHKOFF *et al.*, 2007). Na população de Curitiba, a variante *–13910*T* (alelo derivado) foi observada no contexto haplotípico *E* (ancestral) em baixa frequência, 1,79%, o que sugere que o haplótipo pode ter sofrido recombinação, porém, seriam necessárias avaliações futuras, bem como a genotipagem dos demais marcadores haplotípicos, para confirmar essa hipótese.

Liebert e colaboradores (2017) descreveram haplótipo único que contém as variantes *–13910*T* e *–14011*T*. Os autores concluíram que os alelos teriam sido originados em contextos evolutivos distintos, dessa forma, provavelmente, *–13910*T* e *–14011*T* surgiram mais de uma vez em diferentes populações. Portanto, o haplótipo descrito em Liebert e colaboradores (2017) não se trata de um haplótipo recombinante. O método de inferência haplotípica utilizado neste estudo definiu haplótipos distintos para as variantes *–13910*T* e *–14011*T*, identificados na TABELA 11 pelos números 12 (frequência de 0,765%) e 16 (0,255%), respectivamente. Entretanto, as variantes foram observadas em haplótipos de contexto idêntico — observado a partir da avaliação dos marcadores *–678A>G*, *+666G>A* e *+5579T>C* —, o que indica que os alelos *–13910*T* e *–14011*T* podem estar presentes num mesmo haplótipo, como descrito anteriormente. Nesse caso, a avaliação dos demais marcadores haplotípicos seria interessante para confirmar a identidade dos haplótipos contendo as variantes em questão na população afro-brasileira de Curitiba.

O alelo *–13915*G* é bastante comum na península arábica, Oriente Médio e em populações do Quênia, Sudão e Tanzânia na África, nessas populações o contexto haplotípico é semelhante, com o haplótipo *C* contendo o alelo *–13915*G* (ENATTAH *et al.*, 2007; TISHKOFF *et al.*, 2007; RANCIARO *et al.*, 2014). Este alelo foi observado num indivíduo heterozigoto de Curitiba, entretanto, não foi possível identificar o haplótipo no qual o alelo está inserido (TABELA 11, Hap_20). Assim, a genotipagem dos demais marcadores haplotípicos citados anteriormente confirmaria

se o haplótipo identificado neste estudo é similar ao descrito para essas regiões. Ainda, se semelhante, não teríamos como distinguir a origem do haplótipo 20.

O haplótipo *C* é mais comum em populações da região sul da Europa do que em populações da região norte, reportado nas frequências de 12,3% e 2,3%, respectivamente (HARVEY *et al.*, 1998). Observamos o haplótipo de contexto *C* que não contém o alelo *–13915*G* em ambas as populações avaliadas (TABELA 11 e TABELA 12; Hap_3), nas frequências de 20,0% e 17,6% na população de Sertão do Valongo e de Curitiba, respectivamente. Dessa forma os haplótipos 3 (TABELA 11 e TABELA 12), 15, 18, 19 e 20 (TABELA 11), provavelmente, possuem origens distintas, dado que apenas o haplótipo 20 contém o alelo *–13915*G* associado ao fenótipo LP.

Ainda na população afro-brasileira de Curitiba, os haplótipos 15 e 18 (TABELA 11), que contém as variantes *–13732*T* e *–13752*A*, respectivamente, foram observados na frequência de 0,26%. Não há informações disponíveis na literatura com relação ao contexto haplotípico no qual essas variantes estariam inseridas, contudo, ambos os alelos foram identificados no contexto haplotípico *C* ou *M*. Em populações Bantu os haplótipo *C* e *M* são observados nas frequências de 31% e 5%, respectivamente. Já em populações San, o haplótipo *C* é observado na frequência de 3% e o haplótipo *M* está ausente (HOLLOX *et al.*, 2001). Como as variantes foram identificadas em populações africanas ou afrodescendentes (THE 1000 GENOMES PROJECT CONSORTIUM, 2015; KARCZEWSKI e MARTIN, 2020), é provável que os haplótipos em questão tenham sido introduzidos na população brasileira a partir da miscigenação com povos africanos.

Tendo em vista os eventos de miscigenação que deram origem à população brasileira (tópico 2.5 deste trabalho), a avaliação realizada e a nomenclatura haplotípica considerada, concorda com os demais resultados apresentados nos tópicos 7.1 e 7.2.1 deste trabalho.

Quanto ao desequilíbrio de ligação, os cálculos foram realizados par a par — de acordo com a TABELA 13 — para os marcadores haplotípicos selecionados (*–678A>G*, *+666G>A* e *+5579T>C*) e para as variações *–13730T>G* e *–13910C>T*, caracterizadas como polimorfismos nas população de Curitiba e Sertão do Valongo. Os parâmetros *D'* e r^2 são medidas complementares de desequilíbrio de ligação. O *D'* representa o valor de *D* como uma fração do seu máximo (ou mínimo) valor

possível, e o r^2 representa o coeficiente de correlação entre alelos de um mesmo gameta (HARTL e CLARK, 2007). O r^2 nos informa sobre a ocorrência das mutações nos haplótipos — quando (idade dos alelos nos diferentes sítios) e onde (distância genética entre os diferentes sítios) (HARTL e CLARK, 2007).

TABELA 13: PARÂMETROS DE DESEQUILÍBRIO DE LIGAÇÃO *D'*, r^2, χ^2 E *p* PARA AFRO-BRASILEIROS DE CURITIBA E REGIÃO METROPOLITANA (PR) E DA COMUNIDADE QUILOMBOLA DE SERTÃO DO VALONGO (SC)

Curitiba e Região Metropolitana (PR)					
X	Parâmetros	rs3754689 (*+666G>A*)	rs2278544 (*+5579T>C*)	rs4954492 (*–13730T>G*)	rs4988235 (*–13910C>T*)
rs56211644 (*–678A>G*)	*D'*	0,3438	0,5257	0,6306	0,4971
	r2	0,0231	0,0780	0,0048	0,0233
	χ2	9,0493	30,5797	1,8717	9,1512
	p	0,0026	0,0032	0,1713	0,0025
rs3754689 (*+666G>A*)	*D'*		0,5838	0,4623	0,9041
	r2		0,1299	0,0035	0,1043
	χ2		50,9169	1,3588	40,8703
	p		9,64E-07	0,2437	1,63E-04
rs2278544 (*+5579T>C*)	*D'*			0,7153	0,9140
	r2			0,0120	0,2796
	χ2			4,7013	109,5953
	p			0,0301	< 2,2204e-16
rs4954492 (*–13730T>G*)	*D'*				0,9911
	r2				0,0077
	χ2				3,0204
	p				0,0822
Comunidade quilombola de Sertão do Valongo (SC)					
X	Parâmetros	-	*rs2278544* (*+5579T>C*)	rs4954492 (*–13730T>G*)	-
rs56211644 (*–678A>G*)	*D'*	-	0,9988	0,9988	-
	r2	-	0,0623	0,0623	-
	χ2	-	0,6234	0,6234	-
	p	-	0,4298	0,4298	-
rs2278544 (*+5579T>C*)	*D'*	-		0,2381	-
	r2	-		0,0567	-
	χ2	-		0,5667	-
	p	-		4,52E-01	-

Em vermelho estão destacados os valores de *p* significativos (< 0,05), nesses casos as variantes estão em desequilíbrio de ligação. FONTE: a autora (2021).

Para a população afro-brasileira de Curitiba, as variações *–13910C>T* e *–13730T>G* não estão em desequilíbrio de ligação: p = 0,0822, r^2 = 0,0077, *D'* =

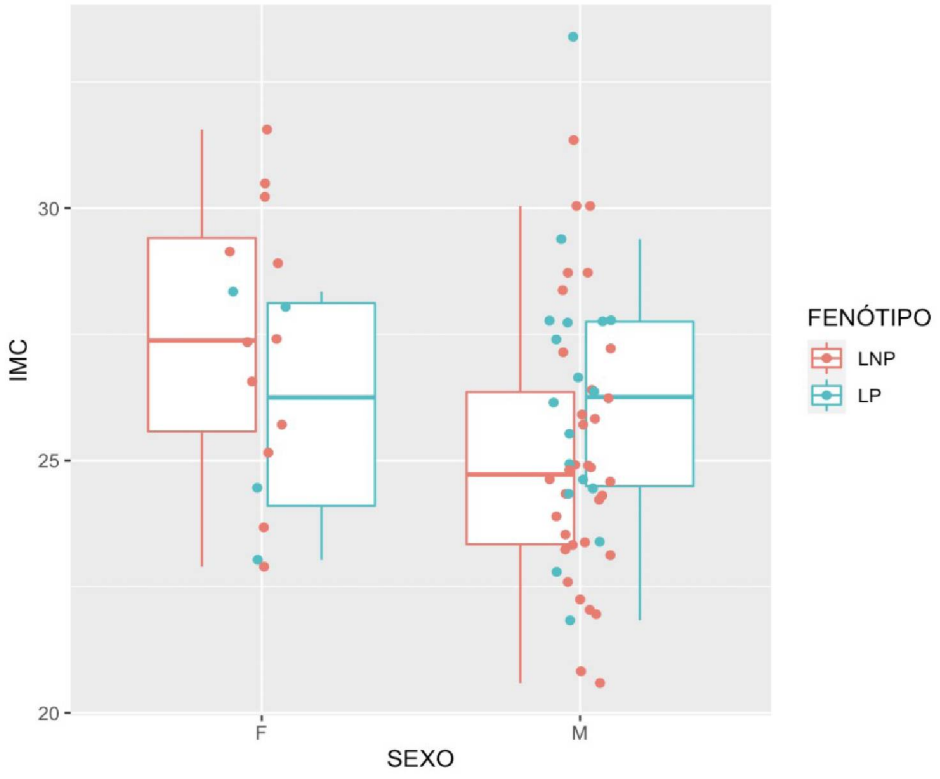
0,9911. O valor obtido para r^2 nos indica que os alelos derivados nas posições *–13910C>T* e *–13730T>G*, provavelmente, não serão observados num mesmo haplótipo, e surgiram em contextos haplotípicos distintos. Em concordância com a inferência haplotípica apresentada na TABELA 11, na qual os haplótipos que contém o alelo *–13910*T* (derivado) também contém o alelo *–13730*T* (ancestral) e vice-versa. O alto valor de *D'* e o baixo valor de r^2 são devidos, possivelmente, à diferença de frequência entre os alelos *–13910*T* e *–13730*G* (18,67% e 2,71%, respectivamente). Ainda, vale ressaltar que as variações *–13910C>T* e *–13730T>G* surgiram em populações distintas (europeia e africana, respectivamente) e são características dessas populações. Dessa forma, as frequências haplotípicas observadas são similares às esperadas pela combinação aleatória das variações *–13910C>T* e *–13730T>G*.

Com relação aos marcadores haplotípicos, *–678A>G* está em desequilíbrio de ligação com *+666G>A* e *+5579T>C*; e *+666G>A* e *+5579T>C* estão em desequilíbrio de ligação entre si. Os valores obtidos para r^2 nos indicam que os alelos derivados para as posições em questão, provavelmente, serão observados num mesmo haplótipo (haplótipos *A* ou *J* para *–678A>G* e *+666G>A*; haplótipo *E* para *–678A>G* e *+5579T>C*; e haplótipos *F*, *I*, *V* ou *X* para *+666G>A* e *+5579T>C*), e surgiram em contextos haplotípicos semelhantes. Os marcadores *+666G>A* e *+5579T>C* estão em desequilíbrio de ligação com *–13910C>T*, entretanto possuem alto valor de *D'* (0,9041 e 0,9140, respectivamente) e baixo valor de r^2, refletindo a diferença de frequência entre os alelos derivados *–13910*T* (18,67%) e *+666*A* (33,93%) e *–13910*T* (18,67%) e *+5579*C* (42,60%). Além disso, *+5579T>C* está em desequilíbrio de ligação com *–13730T>G* em ambas as populações avaliadas. O desequilíbrio de ligação observado entre os marcadores haplotípicos selecionados (*–678A>G*, *+666G>A* e *+5579T>C*) indica que a escolha deles em conjunto foi assertiva, entretanto não exclui a necessidade da genotipagem dos demais marcadores haplotípicos descritos por Hollox e colaboradores (2001).

7.2.4 Avaliação piloto: associação entre o fenótipo lactase persistente e as classes de peso

Dos 258 indivíduos afro-brasileiros da região de Curitiba, 80 possuíam registro de peso e altura necessários para o cálculo do índice de massa corporal (IMC). Desses, 12 tiveram o peso e altura anotados antes do participante completar 20 anos. Como explicado anteriormente (tópico 2.7 deste trabalho), a avaliação do IMC para menores de 20 anos deve ser associada a outros parâmetros, por isso, optamos por excluir esses participantes das análises. Portanto, foram avaliados 68 indivíduos, sendo 52 do sexo masculino com média de idade de 32,42 ± 9,32 anos e IMC médio de 24,9 ± 2,7 kg/m^2 (na data da coleta) (FIGURA 9); e 16 do sexo feminino, com média de idade de 36,25 ± 10,29 anos e IMC médio de 27,1 ± 2,7 kg/m^2 (na data da coleta) (FIGURA 9).

FIGURA 9: DISTRIBUIÇÃO DE IMC ENTRE OS SEXOS FEMININO E MASCULINO DE ACORDO COM O FENÓTIPO LACTASE PERSISTENTE (LP) OU LACTASE NÃO PERSISTENTE (LNP)



O traço no interior de cada caixa representa a mediana dos valores de IMC. As duas primeiras caixas (rosa e verde) são referentes a indivíduos do sexo feminino (F) e as caixas subsequentes a indivíduos do sexo masculino (M). Os pontos representam cada indivíduo avaliado. FONTE: a autora (2020).

Dentre os polimorfismos avaliados neste estudo, o SNP *–13910C>T* (rs4988235), de origem europeia, foi selecionado para a inferência dos fenótipos

LNP e LP, pois foi o único com dados disponíveis na subamostra (68 indivíduos) pré-definida. Desses indivíduos, apenas um é portador de variante associada ao fenótipo LP alternativa à variante *–13910*T* (*–14011*T*). Entretanto, o indivíduo também é portador de *–13910*T*, de forma que se tivéssemos adicionado a subamostra avaliada todos os indivíduos que possuíssem dados de peso e altura (de idade ≥ 20 anos) e que fossem portadores de qualquer variante associada ao fenótipo LP, a subamostra selecionada seria idêntica.

Dos 68 indivíduos avaliados, 46 possuem genótipo *CC*, associado ao fenótipo ancestral LNP, enquanto 21 possuem genótipo *CT* e um possui genótipo *TT*, ambos associados ao fenótipo LP (TABELA 14). Para a subamostra em questão, as frequências genotípicas observadas para a variação *–13910C>T* (rs4988235) encontram-se em equilíbrio de Hardy Weinberg.

TABELA 14: FREQUÊNCIAS ALÉLICAS E GENOTÍPICAS NA SUBAMOSTRA DE AFRO-BRASILEIROS DE CURITIBA E REGIÃO METROPOLITANA (PR)

Variação	Alelos				Genótipos						Equilíbrio de Hardy-Weinberg
	N	%	N	%	N	%	N	%	N	%	
–13910C>T (rs4988235)	*C*		***T***		*CC*		***CT***		***TT***		$p = 0,363$
	113	83,00	23	17,00	46	68,00	21	31,00	1	1,00	

Destacados em negrito os alelos e genótipos associados ao fenótipo lactase persistente. N: número de alelos observados; %: frequência alélica e genotípica. FONTE: a autora (2020).

Para comparar as proporções de indivíduos distribuídos de acordo com as classes de peso entre os fenótipos LNP e LP (FIGURA 9), foram realizados três testes exatos de Fisher compreendendo o total da amostra, indivíduos do sexo masculino e indivíduos do sexo feminino (TABELA 15). Foram consideradas duas classes de peso: peso normal (IMC entre 18,5 e 24,9 kg/m^2), acima do peso e obesidade (IMC ≥ 25 kg/m^2), sendo que as duas últimas foram avaliadas em conjunto. Os valores de *p* estão apresentados na TABELA 15. Como em todos os casos $p > 0,05$, para a população afro-brasileira da região de Curitiba avaliada, as proporções de indivíduos de fenótipo LP e LNP categorizados de acordo com as classes de peso (IMC) não diferem de forma estatisticamente significativa.

TABELA 15: NÚMERO DE INDIVÍDUOS DISTRIBUÍDOS DE ACORDO COM A CLASSE DE PESO (NORMAL OU ACIMA DO PESO E OBESIDADE) E FENÓTIPO LACTASE PERSISTENTE (LP) OU LACTASE NÃO PERSISTENTE (LNP) E DE ACORDO COM O SEXO (MASCULINO OU FEMININO)

Total de indivíduos (n = 68)			
Classes de peso	LNP	LP	Valor de *p*
Peso normal	23	9	0,6053
Acima do peso e obesidade	23	13	
TOTAL	46 (67,6%)	22 (32,4%)	-
Sexo masculino (n = 52)			
Classes de peso	LNP	LP	Parâmetros
Peso normal	21	7	0,1492
Acima do peso e obesidade	13	11	
TOTAL	34 (65,4%)	18 (34,6%)	-
Sexo feminino (n = 16)			
Classes de peso	LNP	LP	Parâmetros
Peso normal	2	2	0,2445
Acima do peso e obesidade	10	2	
TOTAL	12 (75,0%)	4 (25,0%)	-

FONTE: a autora (2020).

A partir da aplicação de uma regressão logística multivariada para classes de peso (IMC) e o fenótipo LP, com correção para sexo, foi possível observar que, para os indivíduos avaliados, não há correlação significativa (0,084, *p* = 0,098) entre as variáveis. Para essa análise, a mais influente das limitações seria a ancestralidade local (gene *MCM6*), dado que determinados alelos associados ao fenótipo LP são mais comuns em populações de ancestralidades distintas. Por exemplo, o alelo *–13910*T* é mais comum em populações de ancestralidade predominantemente europeia, enquanto os alelos *–13907*C*, *–13915*G* e *–14010*C* são mais comuns em populações pastoralistas da África Oriental e do Oriente Médio. Além disso, utilizando apenas a variação *–13910C>T* para a classificação fenotípica, há a possibilidade de segregar indivíduos de ancestralidade europeia predominante para a região (íntron 13, gene *MCM6*) no grupo LP e indivíduos de ancestralidades locais distintas no grupo LNP.

Embora os indivíduos analisados tenham sido classificados como afro-brasileiros (tópico 6.2 deste trabalho), não possuímos dados de ancestralidade genômica para eles, não sendo possível realizar a correção do parâmetro. Assim,

consideramos que para a região em questão (íntron 13, gene *MCM6*), os indivíduos avaliados possuem ancestralidade local europeia expressiva, dado que a frequência do alelo *–13910*T* é superior a frequência observada para as variantes *–13915*G* e *–14011*T*; e a frequência de *–13910*T* está diretamente relacionada à porcentagem de ancestralidade europeia em populações pan-americanas, de acordo com o exposto no tópico 7.1 deste trabalho.

Ainda foram comparadas as médias aritméticas de IMC entre os fenótipos LNP e LP, as médias aritméticas de IMC entre os sexos masculino e feminino de fenótipo LNP e LP e a influência da idade (avaliação de indivíduos de idade igual ou superior a 20 anos) na determinação do fenótipo LNP e LP de acordo com o sexo. Em indivíduos de fenótipo LNP as médias aritméticas de IMC entre os sexos masculino (25,1 ± 2,6 kg/m^2) e feminino (27,4 ± 2,7 kg/m^2) são diferentes. Curiosamente, estudos anteriores do grupo LGMH-UFPR (dados não publicados) avaliando IMC e outras medidas antropométricas em indivíduos de ambos os sexos, com idade superior a 21 anos, de diferentes populações africanas revelaram IMC superior em indivíduos do sexo masculino. A divergência observada pode ser devida a diferenças entre o número de indivíduos avaliados: no primeiro estudo foram incluídos 535 homens e 836 mulheres (no total), enquanto nesse estudo foram incluídos 34 homens e 12 mulheres. Na TABELA 16 pode-se observar um resumo das análises realizadas contendo o objetivo do teste aplicado, a análise estatística empregada, o número de observações, o parâmetro avaliado e as conclusões de cada uma das análises.

Os resultados obtidos podem ter sido influenciados pelo tamanho amostral (n = 68), especialmente nas análises em que os indivíduos foram separados de acordo com o sexo — a subamostra do sexo feminino representa aproximadamente um quarto do total (TABELA 15). A análise de poder estatístico nos mostrou que para o tamanho amostral disponível, significância estatística de 5% e efeito do tamanho amostral de 0,5 (efeito médio), o poder estatístico é de 31%. Em contrapartida, para poder estatístico de 80%, significância estatística de 5% e efeito do tamanho amostral de 0,5, seria necessária amostragem de no mínimo 64 indivíduos por grupo (sexo masculino de fenótipos LNP e LP e sexo feminino de fenótipos LNP e LP).

TABELA 16: TESTES REALIZADOS PARA A ANÁLISE DOS DADOS, SEUS RESPECTIVOS OBJETIVOS, NÚMERO DE OBSERVAÇÕES, PARÂMETROS UTILIZADOS E CONCLUSÕES

Objetivo	Análise empregada	Nº de observações	Valor de *p**	Conclusão
Comparação das médias aritméticas de IMC entre os fenótipos LNP e LP	Teste de Normalidade de Shapiro-Wilk	68	0,3877	O IMC possui distribuição normal
	Teste de homogeneidade das variâncias de Levene	52 (M) 16 (F)	0,9996 0,9734	As variâncias de IMC são iguais entre os grupos LNP e LP
	Teste t de Student (não pareado)	52 (M) 16 (F)	0,1548 0,1548	As médias aritméticas de IMC são iguais entre os grupos LNP e LP
Comparação das médias aritméticas de IMC entre os sexos masculino e feminino de fenótipos LNP e LP	Teste de Normalidade de Shapiro-Wilk	68	0,3877	O IMC possui distribuição normal
	Teste de homogeneidade das variâncias de Levene	46 (LNP)** 22 (LP)***	0,7420 0,8126	As variâncias de IMC são iguais entre os sexos masculino e feminino para indivíduos de fenótipo LNP e LP
	Teste t de Student (não pareado)	46 (LNP)**	0,0135	Para o fenótipo LNP, as médias aritméticas de IMC entre os sexos masculino e feminino são diferentes
		22 (LP)***	0,8585	Para o fenótipo LP, as médias aritméticas de IMC entre os sexos masculino e feminino são iguais
Influência da idade (≥ 20 anos) na determinação do fenótipo LNP e LP de acordo com o sexo	Teste de Normalidade de Shapiro-Wilk	68	0,0139	A idade possui distribuição não normal
	Teste de Mann-Whitney	46 (LNP)** 22 (LP)***	0,3163 0,4422	As medianas de idade entre os sexos masculino e feminino são iguais, tanto para indivíduos de fenótipo LNP, quanto para indivíduos de fenótipo LP

* Valores de $p > 0,05$ foram considerados significativos; ** Dos quais 34 são do sexo masculino e 12 do sexo feminino; *** Dos quais 18 são do sexo masculino e 4 do sexo feminino. M: sexo masculino; F: sexo feminino; LNP: fenótipo lactase não persistente; LP: fenótipo lactase persistente. FONTE: A autora (2021).

Ao todo, os resultados obtidos nos indicam que, para a população afro-brasileira da região de Curitiba investigada não há associação entre as classes de peso, determinadas pelo IMC, e o fenótipo LP, assim como no estudo de Bergholdt e colaboradores (2015) que avaliou maior número de indivíduos (n = 97.811). É possível que a divergência entre os resultados dos estudos apresentados anteriormente com relação à associação entre as classes de peso (IMC) e o fenótipo LP (observada em indivíduos pelotenses (HARTWIG *et al.*, 2016) e não observada em indivíduos dinamarqueses (BERGHOLDT *et al.*, 2015)) tenha ocorrido devido à subestrutura das populações, mesmo que a ancestralidade genômica tenha sido levada em conta como parâmetro de correção. De forma similar, a subestrutura da população avaliada pode ter influenciado os resultados obtidos no presente estudo, principalmente tendo em vista que a classificação dos indivíduos em afro-brasileiros foi baseada em características fenotípicas (tópico 6.2 deste trabalho). Apesar de termos utilizado a categorização em classes de peso de acordo com o IMC sugerida pela Organização Mundial da Saúde, a categorização dos participantes também é limitante quando nos referimos à ancestralidade dos indivíduos, já que linhas de corte população-específica são necessárias (DEURENBERG, YAP e VAN

STAVEREN, 1998). Ainda, como demonstrado por Frankenfield e colaboradores (2001), o emprego da porcentagem de gordura corporal é mais seguro do que o uso do IMC sozinho para a identificação de casos de obesidade. Portanto, a utilização da porcentagem de gordura corporal associada ao IMC poderia tornar as análises realizadas mais robustas, porém, infelizmente não possuímos esse dado para os indivíduos avaliados.

Por fim, a alimentação e consumo energético em geral limita a hipótese desenvolvida por Kettunen e colaboradores (2010). Segundo os autores, diferenças no IMC refletem diferenças na ingestão calórica entre indivíduos de fenótipo LP e LNP, que está diretamente relacionada ao consumo de laticínios. Entretanto, os autores não levaram em consideração os hábitos alimentares de cada indivíduo, independente do fenótipo. Logo, não é possível estabelecer que dietas distintas entre indivíduos de fenótipo LP e LNP estão, de fato, relacionadas às diferenças na ingestão de calorias diárias devido, exclusivamente, ao consumo de laticínios. Da mesma forma, a ausência de investigação dos hábitos alimentares dos participantes do presente estudo é limitante para os resultados obtidos.

8 CONCLUSÕES

Em populações miscigenadas pan-americanas observamos correlação positiva entre a frequência do alelo *–13910*T* e a porcentagem de ancestralidade europeia média no cromossomo 2, no qual está localizado o gene *MCM6.* Além disso, a frequência do alelo *–13910*T* está diretamente relacionada à frequência alélica observada nas populações europeias ancestrais, não havendo super-representação de haplótipos europeus na região flanqueadora da variação *–13910C>T*. Esses resultados sugerem que não houve pressão de seleção (ou a pressão de seleção é muito recente para ser detectada) após eventos de miscigenação que contribuíram para a formação das populações em questão. Considerando o padrão de herança da variante *–13910*T*, os resultados obtidos indicam que as populações miscigenadas pan-americanas, possivelmente, apresentam intolerância à lactose em alta frequência. Dessa forma, nas Américas as políticas públicas de alimentação que incluem o uso de laticínios devem ser reavaliadas e reconsideradas, principalmente nas populações com ancestralidade europeia reduzida.

Reunindo as variantes identificadas no íntron 13 do gene *MCM6*, observamos alelos de origem africana (*–13730*G*, *–13915*G*), de origem europeia (*–13910*T*) e de origem não identificada (*–13732*T*, *–13752*A* e *–14011*T*). Portanto, as frequências alélicas observadas são compatíveis com o histórico de miscigenação da população brasileira, dado que a população afro-brasileira de Curitiba possui tanto ancestralidade europeia quanto africana. Para a população da comunidade quilombola de Sertão do Valongo, observamos apenas a variante de origem africana *–13730*G*. Sabe-se que essa população está sujeita ao fenômeno de deriva genética em maior intensidade (devido ao tamanho diminuto, n = 81) o que pode resultar em fixação ou eliminação de alelos e a redução de heterozigosidade. Assim, a inexistência de variação nas demais posições, em contraste com o observado para a população afro-brasileira de Curitiba, é compatível com o contexto histórico no qual os indivíduos da comunidade quilombola de Sertão do Valongo estão inseridos.

Para a população afro-brasileira da região de Curitiba, o exame genético para a posição *–13910C>T*, de origem europeia, pode ser utilizado como preditor

relativamente confiável do fenótipo LP, pois de todas as variações identificadas foi a única na qual o alelo derivado foi observado em frequência expressiva (18,67%). Além disso, foram observados 82 indivíduos de fenótipo LP, desses, apenas um (heterozigoto para a variação *–13915T>G*) não possui o alelo *–13910*T*. Dessa forma, a variação fenotípica inferida a partir dos genótipos observados na população afro-brasileira da região de Curitiba deve-se à variante de origem europeia em 98,78% dos casos, similar ao observado em populações europeias, nas quais o alelo *–13910*T* está associado ao fenótipo LP em 100% dos casos. O fato da variante ser característica de populações europeias nos mostra que indivíduos afro-brasileiros possuem ancestralidade europeia importante para a região genômica em questão (ao menos nas populações urbanas miscigenadas avaliadas) — de acordo com as estimativas de ancestralidade local realizadas e com a relação positiva entre a frequência do alelo *–13910*T* e a porcentagem de ancestralidade europeia média para o cromossomo 2.

Em contraste, nenhum alelo associado ao fenótipo LP foi identificado na população afro-brasileira da comunidade quilombola de Sertão do Valongo. A partir da inferência fenotípica realizada, o fenótipo LP não é observado na população em questão, assim, a princípio, não haveria necessidade da realização do exame genético para predição do fenótipo nesses indivíduos. Entretanto, seria necessária a avaliação genética de outros membros da CQ Sertão do Valongo para confirmar a ausência de variantes associadas ao fenótipo LP, bem como a avaliação fenotípica deles (através de testes de respiração de hidrogênio, por exemplo). Ainda, as frequências alélicas e genotípicas observadas na CQ Sertão do Valongo não podem ser extrapoladas para outras comunidades quilombolas. Como o fenótipo LP nunca foi avaliado em outras CQs, cada comunidade quilombola deve ser avaliada individualmente, dado que cada uma possui estrutura populacional, histórico de fundação e práticas culturais distintas.

A variação *–13730T>G*, exclusiva de populações africanas, não está associada à determinação do fenótipo LP, porém foi observada em indivíduos de ambas as populações investigadas. O alelo *–13730*G* está amplamente distribuído pelo continente africano, dessa forma, não é possível determinar uma ancestralidade subcontinental específica. Consideramos que o alelo foi herdado devido à miscigenação com populações africanas da região central-sul e leste, da qual há

maior probabilidade dos povos africanos no Brasil terem sido originados. A alta frequência observada para o alelo *–13730*G* na população afro-brasileira da CQ Sertão do Valongo (mais alta já registrada mundialmente) deve-se, provavelmente, ao efeito fundador.

Quanto às classes de peso (IMC) e o fenótipo LP, para os indivíduos avaliados (n = 68), não foi possível observar associação entre os dois parâmetros. Contudo, o baixo poder estatístico observado (31%) para a subamostra avaliada pode ter interferido nos resultados obtidos. Logo, seria interessante ampliar a avaliação utilizando número amostral maior e inserindo outros parâmetros na análise (ancestralidade genômica e porcentagem de gordura corporal, por exemplo), não disponíveis para os participantes deste trabalho.

Os resultados e discussões expostos nesse estudo contribuirão para o melhor entendimento da distribuição de alelos associados ao fenótipo LP (não apenas da variante *–13910C>T*) em populações pan-americanas, de forma a agregar ao panorama mundial de estudos relacionados à persistência da lactase em populações até então pouco estudadas. Além disso, a avaliação de populações afrodescendentes residentes no continente americano é de grande valia, principalmente para o mapeamento das variantes *–13907C>G*, *–13913T>C*, *–13915T>G*, *–14009T>G* e *–14010G>C*, pouco estudadas em populações fora da África.

REFERÊNCIAS

ADHIKARI, K.; MENDOZA-REVILLA, J.; CHACÓN-DUQUE, J. C.; FUENTES-GUAJARDO, M.; RUIZ-LINARES, A. Admixture in Latin America. **Curr. Opin. Genet. Dev.**, v. 41, p. 106–114, 2016.

Agência de Notícias Brasil-Árabe (ANBA). Comunidade árabe é 6% da população brasileira, diz pesquisa. 22 de julho de 2020. Disponível em: < https://anba.com.br/comunidade-arabe-e-6-da-populacao-brasileira-diz-pesquisa/ > Acesso em: 18 de maio de 2021.

Agência de Notícias Brasil-Árabe (ANBA). De 1500 a 2020, árabes encontram no Brasil seu novo lar. 16 de julho de 2020. Disponível em: < https://anba.com.br/de-1500-a-2020-arabes-encontram-no-brasil-seu-novo-lar/ > Acesso em: 18 de maio de 2021.

ANGUITA-RUIZ, A.; AGUILERA, C. M.; GIL, A. Genetics of Lactose Intolerance: An Updated Review and Online InteractiveWorld Maps of Phenotype and Genotype Frequencies. **Nutrients**, v. 12, 2689, 2020. doi:10.3390/nu12092689.

AURICCHIO, S.; RUBINO, A.; LANDOLT, M.; SEMENZA, G.; PRADER, A. Isolated intestinal lactase deficiency in the adult. **Lancet**, v. 2, p. 324–326, 1963.

AURICCHIO, S.; RUBINO, A.; TOSI, R.; SEMENZA, G.; LANDOLT, M.; KISTLER, H. J.; PRADER, A. The quantitative disaccharidase activity of the human small intestine and acquired lactase deficiency in adults. **Helv Med Acta**, v. 30, n. 4, p. 690–692, 1963.

BEIGUELMAN, B. **Dinâmica dos Genes nas Famílias e nas Populações**. 2a ed. Ribeirão Preto: Sociedade Brasileira de Genética, 1994.

BEJA-PEREIRA, A.; LUIKART, G.; ENGLAND, P. R.; BRADLEY, D. G.; JANN, O. C.; BERTORELLE, G.; CHAMBERLAIN, A. T.; NUNES, T. P.; METODIEV, S.; FERRAND, N.; ERHARDT, G. Gene-culture coevolution between cattle milk protein genes and human lactase genes. **Nat Genet**, v. 35, p. 311–313, 2003.

BERGHOLDT, H. K. M.; NORDESTGAARD, B. G; ELLERVIK, C. Milk intake is not associated with low risk of diabetes or overweight-obesity: a Mendelian randomization study in 97,811 Danish individuals. **American Journal of Clinical Nutrition**, v. 102, p. 487–96, 2015.

BERSAGLIERI, T.; SABETI, P. C.; PATTERSON, N.; VANDERPLOEG, T.; SCHAFFNER, S. F.; DRAKE, J. A.; RHODES, M.; REICH, D. E.; HIRSCHHORN, J. N. Genetic Signatures of Strong Recent Positive Selection at the Lactase Gene. **The American Journal of Human Genetics**, v. 74, p. 1111–1120, 2004.

BETHELL, L. **História da América Latina: América Latina Colonial**. vol. 1. São Paulo: Editora da Universidade de São Paulo, 1997.

BOLL, W.; WAGNER, P.; MANTEI N. Structure of the chromosomal gene and cDNAs coding for lactase-phlorizin hydrolase in humans with adult-type hypolactasia or persistence of lactase. **Am J Hum Genet**, v. 48, n. 5, p. 889–902, 1991.

BORBA, F. M. Usos do passado e apropriações patrimonais em territórios quilombolas de Santa Catarina. **Revista Confluências Culturais**, v. 5, n. 2, p. 138, 2016.

BORDA, V.; ALVIM, I.; MENDES, M.; SILVA-CARVALHO, C.; SOARES-SOUZA, G. B.; LEAL, T. P. *et al.* The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. **Proc. Natl. Acad. Sci. U. S. A.**, v. 117, n. 51, p. 32557–32565, 2020.

BOSCHMANN, S. P.; BOLDT, A. B.; SOUZA, I. R.; PETZL-ERLER, M. L.; MESSIAS-REASON, I. J. The Frequency of the LCT * –13910C>T Polymorphism Associated with Lactase Persistence Diverges among Euro-Descendant Groups from Brazil. **Med Princ Pract**, v. 25, p. 18–20, 2016.

BRAUN-PRADO, K.; VIEIRA MION, A. L.; PEREIRA, N. F.; CULPI, L.; PETZL-ERLER, M. L. HLA class I polymorphism, as characterised by PCR-SSOP, in a Brazilian exogamic population. **Tissue Antigens**, v. 56, p. 417–427, 2000.

BRYC, K.; DURAND, E. Y.; MACPHERSON, M.; REICH, D.; MOUNTAIN, J. L. The Genetic Ancestry of African Americans, Latinos, and European Americans across the United States. **The American Journal of Human Genetics**, v. 96, p. 37–53, 2015.

BROZEK, J.; KIHLBERG, J. K.; TAYLOR, H. L.; KEYS, A. Skinfold distributions in middle-aged American men: A contribution to norms of leanness-fatness. **Ann N Y Acad Sci.**, v. 110, p. 492-502, 1963. doi: 10.1111/j.1749-6632.1963.tb15776.x.

BULHÕES, A. C.; GOLDANI, H. A. S.; OLIVEIRA, F. S.; MATTE, U. S.; MAZZUCA, R. B.; SILVEIRA, T. R. Correlation between lactose absorption and the C/T-13910 and G/A-22018 mutations of the lactase-phlorizin hydrolase (LCT) gene in adult-type hypolactasia. **Brazilian Journal of Medical and Biological Research**, v. 40, p. 1441–1446, 2007.

BURGER, J.; KIRCHNER, M.; BRAMANTI, B.; HAAK, W.; THOMAS, M. G. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. **Proc Natl Acad Sci**, v. 104, p. 3736–3741, 2007.

BUTTRISS, J. Nutritional properties of fermented milk products. **International Journal of Dairy Technology**, v. 50, n. 1, p. 21–27, 1997.

CAMPBELL, M. C.; RANCIARO, A. Human adaptation, demography and cattle domestication: an overview of the complexity of lactase persistence in Africa. **Human Molecular Genetics**, v. 30, R1, p. R98–R109, 2021.

CAMPBELL, M. C.; TISHKOFF, S. A. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. **Annu. Rev. Genomics Hum. Genet.**, v. 9, p. 403–433, 2008.

CASKEY, D. A.; PAYNE-BOSE, D.; WELSH, J. D.; GEARHART, H. L.; NANCE, M. K.; MORRISON, R. D. Effects of age on lactose malabsorption in Oklahoma Native Americans as determined by breath H2 analysis. **The American Journal of Digestive Diseases**, v. 22, n. 2, p. 113–116, 1977.

CASTELLS, A. N. G. A cor do lugar: o Sertão do Valongo como patrimônio cultural. **Ilha Revista de Antropologia**, v. 8, p. 417-439, 2008.

CASTRO E SILVA, M. A.; NUNES, K.; LEMES, R. B.; MAS-SANDOVAL, À.; GUERRA AMORIM, C. E.; KRIEGER, J. E.; MILL, J. G.; SALZANO, F. M.; BORTOLINI, M. C.; PEREIRA, A. DA C.; COMAS, D.; HÜNEMEIER, T. Genomic insight into the origins and dispersal of the Brazilian coastal natives. **Proc Natl Acad Sci**, v. 117, p. 2372–2377, 2020.

Centro de Controle e Prevenção de Doenças. Healthy Weight. 10 de abril de 2020. Disponível em: < https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html > Acesso em: 15 de abril de 2020.

COELHO, M.; LUISELLI, D.; BERTORELLE, G.; LOPES, A. I.; SEIXAS, S.; DESTRO-BISOL, G.; ROCHA, J. Microsatellite variation and evolution of lactase persistence. **Human Genetics**, v. 117, n. 4, p. 329–339, 2005.

COELHO, M.; SEQUEIRA, F.; LUISELLI, D.; BELEZA, S.; ROCHA, J. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. **BMC Evol. Biol.**, v. 9, n. 80, 2009.

COOK, G. C.; ALTORKI, M. T. High Intestinal Lactase Concentrations in Adult Arabs in Saudi-Arabia. **British Medical Journal**, v. 3, p. 135–136, 1975.

CUNHA, M. C. **História dos índios no Brasil**. São Paulo: Companhia das Letras, 1992.

DE BASSI, R. A. **Casamentos consangüíneos em populações brasileiras e alguns parâmetros migracionais e etários, associados a casamentos, em Curitiba**. 1983. 85 f. Dissertação (Mestrado em Genética) Faculdade de Ciências Biológicas, Universidade Federal do Paraná, Curitiba. 1983.

DEURENBERG, P.; YAP, M.; VAN STAVEREN, W. A. Body mass index and percent body fat: a metaanalysis among different ethnic groups. **International Journal of Obesity**, v. 22, p. 1164–1171, 1998.

DILLEHAY, T. D. Probing deeper into first American studies. **Proc. Natl. Acad. Sci. U. S. A.**, v. 106, p. 971–978, 2009.

DUNCAN, I. W.; SCOTT, E. M. Lactose intolerance in Alaskan Indians and Eskimos. **American Journal of Clinical Nutrition**, v. 25, n. 9, p. 867–868, 1972.

DURHAM, W. H. **Coevolution: genes, culture and human diversity**. Stanford: Stanford University Press, 1991.

ENATTAH, N. S.; JENSEN, T. G.; NIELSEN, M.; LEWINSKI, R.; KUOKKANEN, M.; RASINPERA, H.; ET AL. Independent introduction of two lactase persistence alleles into human populations reflects different history of adaptation to milk culture. **Am J Hum Genet**, v. 82, n. 1, p. 57–72, 2008.

ENATTAH, N. S.; SAHI, T.; SAVILAHTI, E.; TERWILLIGER, J. D.; PELTONEN, L.; JÄRVELÄ, I. Identification of a variant associated with adult-type hypolactasia. **Nat Genet**, v. 30, n. 2, p. 233–237, 2002.

ENATTAH, N. S.; TRUDEAU, A.; PIMENOFF, V.; MAIURI, L.; AURICCHIO, S.; GRECO L.; et al. Evidence of still-ongoing convergence evolution of the lactase persistence T-13910 alleles in humans. **Am J Hum Genet**, v. 81, n. 3, p. 615–25, 2007.

EVERSHED, R. P.; PAYNE, S.; SHERRATT, A. G. *et al.* Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. **Nature**, v. 455, p. 528–531, 2008.

FERNÁNDEZ, C. I.; MONTALVA, N.; ARIAS, M.; HEVIA, M.; MORAGA, M. L.; FLORES, S. V. Lactase non-persistence and general patterns of dairy intake in indigenous and mestizo chilean populations. **American Journal of Human Biology**, v. 28, n. 2, p. 213–219, 2016.

FERNÁNDEZ, C. I.; FLORES, S. V. Brief communication: Lactase persistence and dairy intake in Mapuche and Mestizo populations from Southern Chile. **American Journal of Physical Anthropology**, v. 155, n. 3, p. 482–487, 2014.

FLATZ, G. Genetics of lactose digestion in humans. **Advances in Human Genetics**, v. 16, p. 1–77, 1987.

FLATZ, G.; ROTTHAUWE, H. W. Lactose nutrition and natural selection. **Lancet**, v. 2, p. 76–77, 1973.

FRANKENFIELD, D. C.; ROWE, W. A.; COONEY, R. N.; SMITH, S.; BECKER, D. Limits of Body Mass Index to Detect Obesity and Predict Body Composition. **Nutrition**, v. 17, p. 26–30, 2001.

FREIRE-MAIA, N. Genetic effects in Brazilian populations due to consanguineous marriages. **Am. J. Med. Gen.**, v. 35, p. 115-117, 1990.

FRIEDRICH, D. C.; SANTOS, S. E.; RIBEIRO-DOS-SANTOS, Â. K.; HUTZ, M. H. Several different lactase persistence associated alleles and high diversity of the lactase gene in the admixed Brazilian population. **PLoS One**, v. 7, n. 9, p. 1–8, 2012a.

FRIEDRICH, D.C.; CALLEGARI-JACQUES, S. M.; PETZL-ERLER, M. L.; TSUNETO, L.; SALZANO, F. M.; HUTZ, M. H. Stability or variation? Patterns of

lactase gene and its enhancer region distributions in Brazilian Amerindians. **Am. J. Phys. Anthropol**, v. 147, p. 427–432, 2012b.

GERBAULT, P.; LIEBERT, A.; ITAN, Y.; POWELL, A.; CURRAT, M.; BURGER, J.; SWALLOW, D. M.; THOMAS, M. Evolution of lactase persistence: an example of human niche construction. **Philos Trans R Soc Lond B Biol Sci**, v. 366, n. 1566, p. 863–877, 2011.

GERBAULT, P.; MORET, C.; CURRAT, M.; SANCHEZ-MAZAS. A. Impact of selection and demography on the diffusion of lactase persistence. **PLoS One**, v. 4, n. 7, e6369, 2009.

Global Dairy Platform. 2018 Annual Review. Disponível em: < https://www.globaldairyplatform.com/wp-content/uploads/2019/02/gdp-annual-report-2018-compressed.pdf > Acesso em: 05 de agosto de 2020.

Global Lactase Persistence Association Database. 2018. Disponível em: < https://www.ucl.ac.uk/biosciences/departments/genetics-evolution-and-environment/research/molecular-and-cultural-evolution-lab/glad#:~:text=Global%20Lactase%20persistence%20Association%20Database,%2C%20medical%2C%20and%20evolutionary%20research > Acesso em: 17 de julho de 2020.

GOEBEL, T.; WATERS, M. R.; O'ROURKE, D. H. The late Pleistocene dispersal of modern humans in the Americas. **Science**, v. 319, p. 1497–1502, 2008.

GONZÁLEZ-JOSÉ, R.; NEVES, W.; LAHR, M. M.; GONZÁLEZ, S.; PUCCIARELLI, H.; MARTÍNEZ, M. H.; CORREL, G. Late Pleistocene/Holocene craniofacial morphology in Mesoamerican Paleoindians: implications for the peopling of the New World. **Am. J. Phys. Anthropol.**, v. 128, n. 4, p. 772–780, 2005.

GOUVEIA, M. H.; BORDA, V.; LEAL, T. P. *et al.* Origins, admixture dynamics and homogenization of the African gene pool in the Americas. **Molecular Biology and Evolution**, msaa033, 2020.

GREENBERG, J. H.; TURNER II, C. G.; ZEGURA, S. L. The Settlement of the Americas: A Comparison of the Linguistic, Dental, and Genetic Evidence. **Current Anthropology**, v. 27, n. 5, p. 477–497, 1986.

HARRIS, D. N.; SONG, W.; SHETTY, A. C.; LEVANO, K. S.; CÁCERES, O.; PADILHA, C. *et al.* Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. **Proc. Natl. Acad. Sci. U. S. A**., v. 115, n. 28, E6526–E6535, 2018.

HARTL, D. L.; CLARK, A. G. **Principles of population genetics**. 4a ed. Sunderland, 2007.

HARTWIG, F. P.; HORTA, B. L.; SMITH, G. D.; DE MOLA, C. L.; VICTORA, C. G. Association of lactase persistence genotype with milk consumption, obesity and blood pressure: a Mendelian randomization study in the 1982 Pelotas (Brazil) Birth

Cohort, with a systematic review and meta-analysis. **International Journal of Epidemiology**, vol. 45, n. 5, p. 1573–1587, 2016.

HARVEY, C. B. et al. DNA polymorphisms in the lactase gene: Linkage Disequilibrium across the 70-kb region. **European Journal of Human Genetics**, v. 3, n. 1, p. 27–41, 1995.

HARVEY, C. B.; HOLLOX, E. J.; POULTER, M.; WANG, Y.; ROSSI, M.; AURICCHIO, S.; IQBAL, T. H.; COOPER, B. T.; BARTON, R.; SARNER, M.; KORPELA, R.; SWALLOW, D. M. Lactase haplotype frequencies in Caucasians: association with the lactase persistence/non-persistence polymorphism. **Ann Human Genet**, v. 62, p. 215–223, 1998.

HE, T.; VENEMA, K.; PRIEBE, M. G.; WELLING, G. W.; BRUMMER, R. J. M.; VONK, R. J. The role of colonic metabolism in lactose intolerance. **European Journal of Clinical Investigation,** v. 38, n. 8, p. 541–547, 2008.

HOLLOX, E. J.; POULTER, M.; ZVARIK, M.; FERAK, V.; KRAUSE, A.; JENKINS, T.; SAHA, N.; KOZLOV, A. I.; SWALLOW, D. M. Lactase haplotype diversity in the old world. **American Journal of Human Genetics**, v. 68, n. 1, p. 160–172, 2001.

HOLZEL, A.; SCHWARZ, V.; SUTCLIFFE, K. W. Defective lactose absorption causing malnutrition in infancy. **Lancet,** v.1, p.1126–1128, 1959.

IBGE. **Brasil: 500 anos de povoamento**. Rio de Janeiro: IBGE, 2000.

International Farm Comparison Network. Dairy Report 2019. Disponível em: < https://ifcndairy.org/wp-content/uploads/2019/10/Dairy-Report_2019_extraction_for-marketing.pdf > Acesso em: 05 de agosto de 2020.

INGRAM, C. J. E.; ELAMIN, M. F.; MULCARE, C. A.; WEALE, M. E.; TAREKEGN, A.; RAGA, T. O.; BEKELE, E.; ELAMIN, F. M.; THOMAS, M. G.; BRADMAN, N.; SWALLOW, D. M. A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? **Hum Genet**, v. 120, n. 6, p. 779–788, 2007.

INGRAM, C. J. E.; MULCARE, C. A.; ITAN, Y.; THOMAS, M. G.; SWALLOW, D. M. Lactose digestion and the evolutionary genetics of lactase persistence. ***Human Genetics***, v. 124, n. 6, p. 579–591, 2009.

INGRAM, C. J. E.; RAGA, T. O.; TAREKEGN, A.; BROWNING, S. L.; ELAMIN, M. F.; BEKELE, E.; THOMAS, M. G.; WEALE, M. E.; BRADMAN, N.; SWALLOW, D. M. Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. **J Mol Evol**, v. 69, n. 6, p. 579–588, 2009.

ITAN, Y.; JONES, B.; INGRAM, C.; SWALLOW, D.; THOMAS, M. G. A worldwide correlation of lactase persistence phenotype and genotypes. **BMC Evol Biol**, v. 10, p. 36, 2010.

ITAN, Y.; POWELL, A.; BEAUMONT, M. A.; BURGER, J.; THOMAS, M. G. The Origins of Lactase Persistence in Europe. **PLoS Comput Biol**, v. 5, n. 8, e1000491, 2009.

JOBLING, M.; HOLLOX, E.; HURLES, M.; KIVISILD, T.; TYLER-SMITH, C. **Human Evolutionary Genetics**. 2a ed. Nova Iorque: Garland Science, 2013.

JONES, B. L.; RAGA, T. O.; LIEBERT, A.; ZMARZ, P.; BEKELE, E.; DANIELSEN, E. T.; OLSEN, A. K.; BRADMAN, N.; TROELSEN, J. T; SWALLOW, D. M. Diversity of Lactase Persistence Alleles in Ethiopia: Signature of a Soft Selective Sweep. **Am J Hum Genet**, v. 93, n. 3, p. 538–544, 2013.

JUSSILA, J. Milk intolerance and lactose malabsorption in hospital patients and young servicemen in Finland. **Annals of Clinical Research**, v.1, p. 199–207, 1969.
JUSSILA, J.; ISOKOSKI, M.; LAUNIALA, KARI. Prevalence of Lactose Malabsorption in a Finnish Rural Population. **Scandinavian Journal of Gastroenterology**, v. 5, n. 1, p. 49–56, 1970.

KARCZEWSKI, K. J.; MARTIN, A. R. Analytic and Translational Genetics. Annual Review of Biomedical Data Science, v. 18, n. 4, p. 217–241, 2020.

KEHDY, F. S. G.; GOUVEIA, M. H.; MACHADO, M.; MAGALHÃES, W. C. S.; HORIMOTO, A. R. V. R.; HORTA, B. L. *et al.* Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. **Proc. Natl. Acad. Sci. U. S. A.**, v. 112, n. 28, p. 8696–8701, 2015.

KETTUNEN, J.; SILANDER, K.; SAARELA, O.; *et al.* European lactase persistence genotype shows evidence of association with increase in body mass index. **Human Molecular Genetics**, v. 19, n. 6, p. 1129–1136, 2010. doi:10.1093/hmg/ddp561

KRUSE, T. A.; BOLUND, L.; GRZESCHIK, K-H.; ROPERS, H. H.; SJOSTROM, H.; NOREN, O.; MANTEI, N.; SEMENZA, G. The human lactase-phlorizin hydrolase gene is located on chromosome 2. **FEBS Lett**, v. 240, p. 123–126, 1988.

LABRIE, V; BUSKE, O. J.; OH, E.; JEREMIAN, R.; PTAK, C.; GASIUNAS, G.; MALECKAS, A.; PETEREIT, R.; ŽVIRBLIENE, A.; ADAMONIS, K. *et al.* Lactase nonpersistence is directed by DNA-variation-dependent epigenetic aging. **Nature Structural & Molecular Biology**, v. 23, p. 566–573, 2016.

LAHIRI, D. K.; NURNBERGER, J. I. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. **Nucleic Acids Res.**, v. 19, 5444, 1991.

LALAND, K. N.; ODLING-SMEE, J.; MYLES, S. How culture shaped the human genome: bringing genetics and the human sciences together. **Nature Reviews Genetics**, v. 11, p. 137–148, 2010.

LATORRE, G.; BESA, P.; PARODI, C. G.; FERRER, V.; AZOCAR, L.; QUIROLA, M., *et al*. Prevalence of lactose intolerance in Chile: a double-blind placebo study. **Digestion**, v. 90, p. 18–26, 2014.

LEITE, I. B. O projeto político quilombola: Desafios, conquistas e impasses atuais. **Revista Estudos Feministas**, v. 16, n. 3, p. 965–977, 2008.

LEONARDI, M.; GERBAULT, P.; THOMAS, M. G.; BURGER, J. The evolution of lactase persistence in Europe. A synthesis of archaeological and genetic evidence. **International Dairy Journal**, v. 22, p. 88–97, 2012.

LESEVA, M. N.; GRAND, R. J.; KLETT, H.; BOERRIES, M.; BUSCH, H.; BINDER, A. M.; MICHELS, K. B. Differences in DNA Methylation and Functional Expression in Lactase Persistent and Non-persistent Individuals. **Scientific Reports**, v. 8, 5649, 2018.

LEWINSKI, R. H. T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. **Hum Mol Genet**, v. 14, n. 24, p. 3945–3953, 2005.

LIEBERT, A.; JONES, B. L.; DANIELSEN, E. T.; OLSEN, A. K.; SWALLOW, D. M.; TROELSEN, J. T. In Vitro Functional Analyses of Infrequent Nucleotide Variants in the Lactase Enhancer Reveal Different Molecular Routes to Increased Lactase Promoter Activity and Lactase Persistence. **Annals of Human Genetics**, v. 80, p. 307–318, 2016.

LIEBERT, A.; LÓPEZ, S.; JONES, B. L.; MONTALVA, N.; GERBAULT, P.; LAU, W.; THOMAS, M. G.; BRADMAN, N.; MANIATIS, N.; SWALLOW, D. M. World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection. **Hum. Genet.**, v. 136, p. 1445–1453, 2017.

LITTLE, M. A.; GRAY, S. J.; CAMPBELL, B. C. Milk consumption in African pastoral peoples. *In:* de GARINE, I; de GARINE, V. **Drinking: Anthropological Approaches**. Nova Iorque: Berghahn Books, 2001, p. 66 – 86.

LLOYD, M.; MEVISSEN, G.; FISCHER, M.; OLSEN, W.; GOODSOEED, D.; GENINI, M. *et al.* Regulation of intestinal lactase in adult hypolactasia. **The Journal of Clinical Investigation**, v. 89, n. 2, p. 524–529, 1992.

LOESCH, D.; HORIMOTO, A. R. V. R.; HEILBRON, K.; SARIHAN, E. I.; INCA-MARTINEZ, M.; MANSON, E. *et al.* Characterizing the genetic architecture of Parkinson's disease in Latinos. **medRxiv**, 2020. doi: 11.09.20227124.

LOPES, F. L.; HOU, L.; BOLDT, A. B. W.; KASSEM, L.; ALVES, V. M.; NARDI, A. E.; MCMAHON, F. J. Finding Rare, Disease-Associated Variants in Isolated Groups: Potential Advantages of Mennonite Populations. **Human Biology**, v. 88, n. 2, p. 109–120, 2016.

LUIZON, M. R. **Dinâmica da mistura étnica em comunidades remanescentes de quilombo brasileiras**. 2007. 115 f. Tese (Doutorado em Genética) Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2007.

LUIZON, M. R.; MENDES-JUNIOR, C. T.; DE OLIVEIRA, S. B.; SIMÕES, A. L. Ancestry Informative Markers in Amerindians from Brazilian Amazon. **American Journal of Human Biology**, v. 20, p. 86–90, 2008.

MACHOLDT, E.; LEDE, V.; BARBIERI, C.; MPOLOKA, S. W.; CHEN, H.; SLATKIN, M.; PAKENDORF, B.; STONEKING, M. Tracing pastoralist migrations to southern Africa with lactase persistence alleles. **Current Biology**, v. 24, n. 8, p. 875–879, 2014.

MATTAR, R.; MAZO, D. F. C.; CARRILHO, F. J. Lactose intolerance: diagnosis, genetic, and clinical factors. **Clinical and Experimental Gastroenterology**, v. 5, n. 1, p. 113–121, 2012.

MATTAR, R.; MONTEIRO, M. S.; SILVA, J. M. K.; CARRILHO, F. J. LCT-22018G>A single nucleotide polymorphism is a better predictor of adult-type hypolactasia/lactase persistence in Japanese-Brazilians than LCT-13910C>T. **Clinics**, v. 65, n. 12, p. 1399–1400, 2010.

MATTAR, R.; MONTEIRO, M. S.; VILLARES, C. A.; SANTOS, A. F.; SILVA, J. M. K.; CARRILHO, F. J. Frequency of LCT -13910C>T single nucleotide polymorphism associated with adult-type hypolactasia/lactase persistence among Brazilians of different ethnic groups. **Nutrition Journal**, v. 8, n. 46, 2009.

MCCRACKEN, R. D. Lactase deficiency: an example of dietary evolution. **Current Anthropology**, v. 12, n. 4/5, p. 479–517, 1971.

MCLELLAN, T.; JORDE, L.B.; SKOLNICK, M. H. Genetic distance between the Utah Mormons and related populations. **American Journal of Human Genetics,** v. 36, p. 836–857, 1984.

MENDES, M.; ALVIM, I.; BORDA, V.; TARAZONA-SANTOS, E. The history behind the mosaic of the Americas. **Curr. Opin. Genet. Dev.**, v. 62, p. 72–77, 2020.

MENDOZA TORRES, E.; VARELA PRIETO, L. L.; VILLARREAL CAMACHO, J. L.; VILLANUEVA TORREGROZA, D. A. Diagnosis of adult-type hypolactasia/lactase persistence: genotyping of single nucleotide polymorphism (SNP C/T-13910) is not consistent with breath test in Colombian Caribbean population. **Arq. Gastroenterol.**, v. 49, p. 5–8, 2012.

MONTALVA, N.; ADHIKARI, K.; LIEBERT, A.; MENDOZA-REVILLA, J.; FLORES, S. V.; MACE, R., *et al*. Adaptation to milking agropastoralism in Chilean goat herders and nutritional benefit of lactase persistence. **Ann. Hum. Genet.**, v. 83, p. 11–22, 2019.

MICHELS, K. B.; GREENLAND, S.; ROSNER, B. A. Does Body Mass Index Adequately Capture the Relation of Body Composition and Body Size to Health Outcomes? **Am J Epidemiol**, v. 147, p. 167-172, 1998.
Ministério da Saúde. Avaliação do peso em adultos (20 a 59 anos). 30 de maio de 2017. Disponível em: <

http://www.saude.gov.br/component/content/article/804-imc/40509-imc-em-adultos > Acesso em: 07 de abril de 2020.

MULCARE, C. A.; WEALE, M. E.; JONES, A. L.; CONNELL, B.; ZEITLYN, D.; TAREKEGN, A.; SWALLOW, D. M.; BRADMAN, N.; THOMAS, M. G. The T Allele of a Single-Nucleotide Polymorphism 13.9 kb upstream of the lactase gene (*LCT*) (*C513.9kbT*) does not predict or cause the lactase-persistence phenotype in africans. **Am J Hum Genet**, v. 74, n. 6, p. 1102–1110, 2004.

MULCARE, C. A. **The evolution of the lactase persistence phenotype**. London: University of London, 2006.

NAKATSUKA, N.; LAZARIDIS, I.; BARBIERI, C.; SKOGLUND, P.; ROHLAND, N.; MALLICK, S. *et al.* A Paleogenomic Reconstruction of the Deep Population History of the Andes. **Cell**, v. 181, n. 5, p. 1131–1145.e21, 2020.

National Center for Biotechnology Information (NCBI). MCM6 minichromosome maintenance complex component 6 [*Homo sapiens* (human)]. Atualizado em 03 de janeiro de 2021. Disponível em: < https://www.ncbi.nlm.nih.gov/gene/4175#gene-expression > Acesso em: 26 de janeiro de 2021.

NEVANLINNA, H. R. The Finnish population structure. A genetic and genealogical study. **Hereditas**, v. 71, p. 195–236, 1972.

NEWMAN, J. L. **The peopling of Africa**. New Haven: Yale University Press, 1995.

NORIO, R.; NEVANLINNA, H. R.; PERHEENTUPA, J. Hereditary diseases in Finland; rare flora in rare soul. **Ann Clin Res**, v. 5, p. 109–141, 1973.

OH, E.; JEREMIAN, R.; OH, G.; GROOT, D.; SUSIC, M.; LEE, K.; FOY, K.; LAIRD, P. W.; PETRONIS, A.; LABRIE, V. Transcriptional heterogeneity in the lactase gene within cell-type is linked to the epigenome. **Scientific Reports**, n. 7, 41843, 2017.

OJEDA-GRANADOS, C.; PANDURO, A.; REBELLO PINHO, J. R. *et al.* Association of lactase persistence genotypes with high intake of dairy saturated fat and high prevalence of lactase non-persistence among the Mexican population. **Journal of Nutrigenetics and Nutrigenomics**, v. 9, n. 2–4, p. 83–94, 2016.

OLDS, L. C.; SIBLEY, E. Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. **Human Molecular Genetics**, v. 12, n. 18, p. 2333–2340, 2003.

OLIVEIRA, L. C.; BARROS, A. L. D.; BOLDT, A. B. W.; CIPOLLA, G. A. Lack of Association between *LCT_rs140433552*CA>del* Indel Polymorphism and Lactose Intolerance in a Southern Brazilian Population. **Lifestyle Genomics**, v. 13, n. 4, 2020.

Organização Mundial da Saúde. Body mass index - BMI. 2020. Disponível em: < http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi > Acesso em: 07 de abril de 2020.

Organização Mundial da Saúde. Obesity and overweight. 1o de abril de 2020. Disponível em: < https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight > Acesso em: 07 de abril de 2020.

PAZ-Y-MIÑO, C.; BURGOS, G.; LOPEZ-CORTES, A.; HERRERA, C.; GAVIRIA, A.; TEJERA, E; CABRERA-ANDRADE, A. A study of the molecular variants associated with lactase persistence in different Ecuadorian ethnic groups. **American Journal of Human Biology**, v. 28, n. 6, p. 774–781, 2016.

PENA, S. D.; DI PIETRO, G.; FUCHSHUBER-MORAES, M. *et al*. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. **PLoS One**, v. 6, n. 2, p. 1–9, 2011.
PHILLIPSON, D. W. **African archaeology**. Cambridge: Cambridge University Press, 1993.

POSTH, C.; NAKATSUKA, N.; LAZARIDIS, I.; SKOGLUND, P.; MALLICK, S.; LAMNIDIS, T. C. *et al.* Reconstructing the Deep Population History of Central and South America. **Cell**, v. 175, n. 5, p. 1185–1197.e22, 2018.

POULTER, M.; HOLLOX, E.; HARVEY, C. B.; MULCARE, C.; PEUHKURI, K.; KAJANDER, K.; SARNER, M.; KORPELA, R.; SWALLOW, D. M. The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. **Ann Hum Genet**, v. 67, p. 298–311, 2003.

PROBST, C. M.; BOMPEIXE, E. P.; PEREIRA, N. F.; DALALIO, M. M. O.; VISENTAINER, J. E.; TSUNETO, L. T.; PETZL-ERLER, M. L. HLA polymorphism and evaluation of European, African, and Amerindian contribution to the white and mulatto populations from Paraná, Brazil. **Hum Biol**, v. 72, n. 4, p. 597–617, 2000.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I. W.; DALY, M. J.; SHAM, P. C. PLINK: a toolset for whole-genome association and population-based linkage analysis. **American Journal of Human Genetics**, v. 81, p. 559–75, 2007.

R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, 2020. Disponível em: < http://www.R-project.org >. Acesso em: 03 de abril de 2020.

RAGGIO, A. N. A. Z.; BLEY, R. B.; TRAUCZYNSKI, S. C. **Abordagem Histórica sobre a População Negra no Estado de Paraná**. Primeira ed. Curitiba: SEPPIR/MDH e SEJU, 2008.

RANCIARO, A.; CAMPBELL, M. C.; HIRBO, J. B.; KO, W-Y.; FROMENT, A.; ANAGNOSTOU, P.; KOTZE, M. J.; IBRAHIM, M.; NYAMBO, T.; OMAR, S. A.; TISHKOFF, S. A. Genetic Origins of Lactase Persistence and the Spread of Pastoralism in Africa. **The American Journal of Human Genetics**, v. 94, p. 496–510, 2014.

RASINPERÄ, H.; KUOKKANEN, M.; KOLHO, K. L.; LINDAHL, H.; ENATTAH, N. S.; SAVILAHTI, E. Transcriptional down regulation of the lactase (*LCT*) gene during childhood. **Gut: journal of the British Society of Gastroenterology**, v. 54, p.1660–1661, 2005.

READER, J. **Africa: A Biography of the Continent**. New York: Knopf, 1998.

REICH, D.; PATTERSON, N.; CAMPBELL, D. *et al.* Reconstructing Native American population history. **Nature**, v. 488, p. 370–388, 2012.

RENTERO, N. *et al.* **ANUÁRIO LEITE 2019**. São Paulo: Texto Comunicação Corporativa elaborado por concessão da Embrapa Gado de Leite, 2019.

ROBAYO-TORRES, C. C.; NICHOLS, B. L. Molecular Differentiation of Congenital Lactase Deficiency from Adult Type Hypolactasia. **Nutrition Reviews**, v. 65, n. 2, p. 95–98, 2007.

ROCHA, J. The evolution of lactase persistence. **Antropologia Portuguesa**, v. 29, p. 121-137, 2012. doi: 10.14195/2182-7982_29_9.

ROMERO, I. G.; MALLICK, C. B.; LIEBERT, A. *et al.* Herders of Indian and European Cattle Share Their Predominant Allele for Lactase Persistence. **Mol. Biol. Evol.**, v. 29, n. 1, p. 249–260, 2012.

ROZAS, J.; FERRER-MATA, A.; SÁNCHEZ-DELBARRIO, J. C.; GUIRAO-RICO, S.; LIBRADO, P.; RAMOS-ONSINS, S. E.; SÁNCHEZ-GRACIA, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. **Molecular Biology and Evolution**, v. 34, n. 12, p. 3299–3302, 2017.

RUIZ-LINARES, A.; ADHIKARI, K.; ACUÑA-ALONZO, V.; QUINTO-SANCHEZ, M.; JARAMILLO, C.; ARIAS, W. *et al.* Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. **PLoS Genet.**, v. 10, e1004572, 2014.

SAHI, T. Lactose malabsorption in Finnish-speaking and Swedish-speaking populations in Finland. **Scandinavian Journal of Gastroenterology**, v. 9, n. 3, p. 303–308, 1974.

SAHI, T. Hypolactasia and Lactase Persistence. **Scandinavian Journal of Gastroenterology**, v. 29, n. 202, p. 1–6, 1994.

SAHI, T.; ISOKOSKI, M.; JUSSILA, J.; LAUNIALA, K.; PYÖRÄLÄ, K. Recessive inheritance of adult-type lactose malabsorption. **The Lancet**, v. 2, n. 7833, 1973.

SALZANO, F. M. The prehistoric colonization of the Americas: evidence and models. **Evol Educ Outreacht**, v. 4, p. 199–204, 2011.

SALZANO, F. M.; BORTOLINI, M. C. **The Evolution and Genetics of Latin American Populations**. Cambridge: Cambridge University Press, 2005.

SALZANO, F. M.; CALLEGARI-JACQUES, S. **South American Indians: a case study in evolution**. Oxford: Clarendon Press, 1988.

SAMBROOK, J. ; FRITSCH, E. F.; MANIATIS, T. **Molecular cloning: a laboratory manual**. New York: Cold spring harbor laboratory press, 1989.

SANTOS, S. A. **A formação do mercado de trabalho livre em São Paulo: tensões raciais e marginalização social**. Tese de Mestrado, Departamento de Sociologia, Universidade de Brasília, 1997.

SANTOS, S. A. Historical roots of the "whitening" of Brazil. **Latin American Perspectives**, v. 29, n. 1, p. 61–82, 2002.

SEGAL, K. R.; VAN LOAN, M.; FITZGERALD, P. I.; HODGDON, J. A.; VAN ITALIE, T. B. Lean body mass estimation by bioelectrical impedance analysis: a four-site cross-validation study. **Am J Clin Nutr**, v. 47, n.1, p. 7–14, 1988. doi: 10.1093/ajcn/47.1.7.

SÉGUREL, L.; BON, C. On the Evolution of Lactase Persistence in Humans. **Annu. Rev. Genom. Hum. Genet.**, v. 18, p. 297–319, 2017.

SEVÁ-PEREIRA, A.; BEIGUELMAN, B. Primary lactose malabsorption in healthy Brazilian adult caucasoid, negroid and mongoloid subjects. **Arq Gastroenterol.**, v. 19, p. 133–138, 1982.

SIMOONS, F. J. Primary adult lactose intolerance and the milking habit: a problem in biological and cultural interrelations. II. A culture historical hypothesis. **American Journal of Digestive Diseases**, v. 15, n. 8, p. 695–710, 1970.

SIQUEIRA, K. B. **O Mercado Consumidor de Leite e Derivados**. Minas Gerais: Embrapa Gado de Leite, 2019. 17 p. (Embrapa Gado de Leite. Circular Técnica, 120).

SKOGLUND, P.; MALLICK, S.; BORTOLINI, M. C.; CHENNAGIRI, N.; HÜNEMEIER, T.; PETZL-ERLER, M. L.; SALZANO, F. M.; PATTERSON, N.; REICH, D. Genetic evidence for two founding populations of the Americas. **Nature**, v. 525, p. 104–108, 2015.

SOUZA, I. R.; CULPI, L. Valongo, an isolated Brazilian Black community. **Brazilian Journal of Genetics**, v. 15, p. 439–447, 1992.

SOUZA, I. R.; CULPI, L. Valongo, genetic studies on an isolated Afro-Brazilian community. **Genetics and Molecular Biology**, v. 28, n. 3, p. 402–406, 2005.

SOUZA, M. M. **África e Brasil Africano**. São Paulo: Ática, 2008.
STEPHENS, M.; SMITH, N. J.; DONNELLY, P. A new statistical method for haplotype reconstruction from population data. **American Journal of Human Genetics**, v. 68, n. 4, p. 978–989, 2001.

SWAGERTY, L. D.; WALLING, D. A.; KLEIN, M. R. Lactose Intolerance. **American Family Physician**, v. 65, n. 9, p. 1845–1850, 2002.

SWALLOW, D. M. Genetics of lactase persistence and lactose intolerance. **Annu Rev Genet**, v. 37, p. 197–219, 2003.

TALIUN, D.; HARRIS, D. N.; KESSLER, M. D.; CARLSON, J.; SZPIECH, Z. A.; TORRES, R. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. **Nature**, v. 590, 290–299, 2021.

THE 1000 GENOMES PROJECT CONSORTIUM. A global reference for human genetic variation. **Nature**, v. 526, n. 7571, p. 68–74, 2015.

TISHKOFF, S.A.; REED, F. A.; FRIEDLAENDER, F. R. *et al.* The genetic structure and history of Africans and African Americans. **Science**, v. 324, p. 1035–1044, 2009.

TISHKOFF, S. A.; REED, F. A.; RANCIARO, A. *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. **Nat Genet**, v. 39, n. 1, p. 31–40, 2007.

TORNIAINEN, S.; PARKER, M. I.; HOLMBERG, V.; LAHTELA, E.; DANDARA, C.; JARVELA, I. Screening of variants for lactase persistence/non-persistence in populations from South Africa and Ghana. **BMC Genetics**, v. 10, n. 31, 2009.

TROELSEN, J. T.; OLSEN, J.; MØLLER, J.; SJÖSTRÖM, H. An upstream polymorphism associated with lactase persistence has increased enhancer activity. **Gastroenterology**, v. 125, p. 1686–1694, 2003.

TROELSEN, J. T. Adult-type hypolactasia and regulation of lactase expression. **Biochim Biophys Acta**, v. 1723, n. 1–3, p. 19–32, 2005.

TSUNETO, L. T.; PROBST, C. M.; HUTZ, M. H. *et al.* HLA class II diversity in seven Amerindian populations. Clues about the origins of the Aché. **Tissue Antigens**, v. 62, n. 6, p. 512–526, 2003.

VALENCIA, L.; RANDAZZO, A.; ENGFELDT, P.; OLSSON, L. A.; CHÁVEZ, A.; BUCKLAND, R. J., *et al.* Identification of novel genetic variants in the mutational hotspot region 14 kb upstream of the LCT gene in a Mexican population. **Scand. J. Clin. Lab. Invest.**, v. 77, p. 311–314, 2017.

VERANO, J. W., UBELAKER, D. H. **Disease and Demography in the Americas**. Washington: Smithsonian Institution Press, 1992.

VICENTE, M.; PRIEHODOVÁ, E.; DIALLO, I.; PODGORNÁ, E.; POLONI, E. S.; ČERNÝ, V.; SCHLEBUSCH, C. M. Population history and genetic adaptation of the Fulani nomads: inferences from genome-wide data and the lactase persistence trait. **BMC Genomics**, v. 20, n. 915, 2019.

VIGNE, J-D.; HELMER, D. Was milk a "secondary product" in the Old World Neolithisation process? Its role in the domestication of cattle, sheep and goats. **Anthropozoologica**, v. 42, p. 9–40, 2007.

WANDERLEY, M. N. Raízes históricas do campesinato brasileiro. In: TEDESCO, J. C. Agricultura familiar: realidades e perspectivas. 3.ed. Passo Fundo: UPF, 2001.

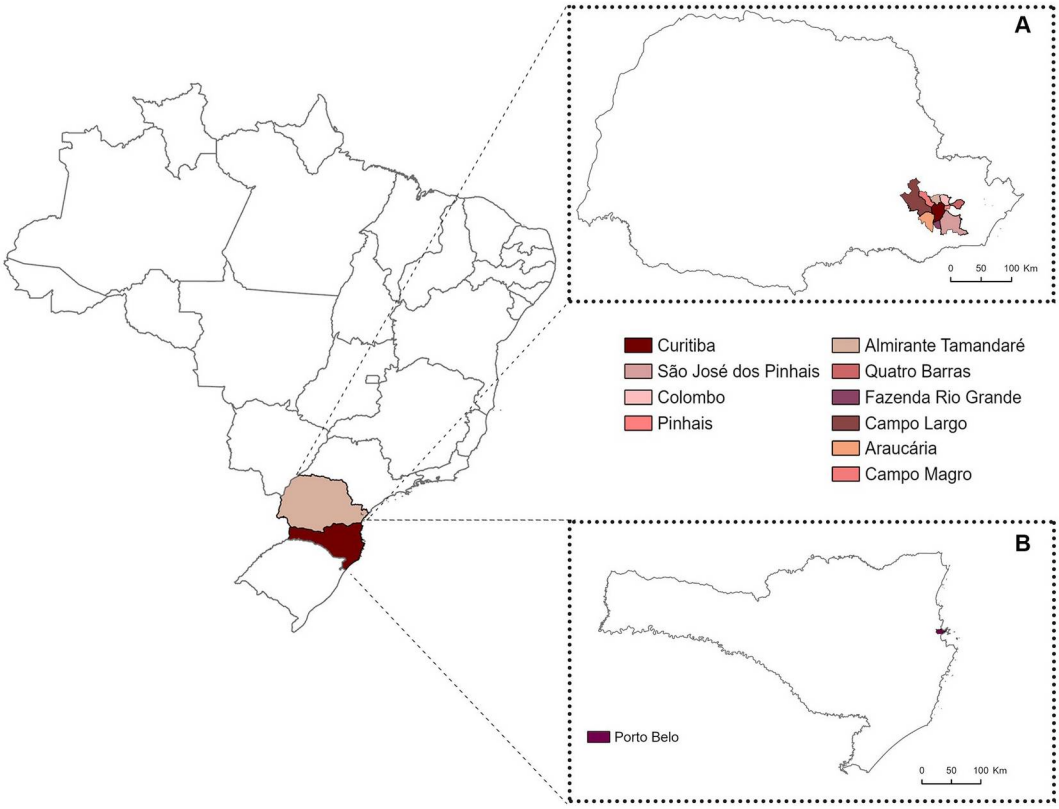
WANG, Y.; HARVEY, C.; ROUSSET, M.; SWALLOW, D. M. Expression of human intestinal mRNA transcripts during development: analysis by a semiquantitative RNA polymerase chain reaction method. **Pediatric Research**, v. 36, p. 514–521, 1994.

WEHLING, A.; WEHLING, M. J. C. **Formação do Brasil Colonial**. Rio de Janeiro: Editora Nova Fronteira, 1994.

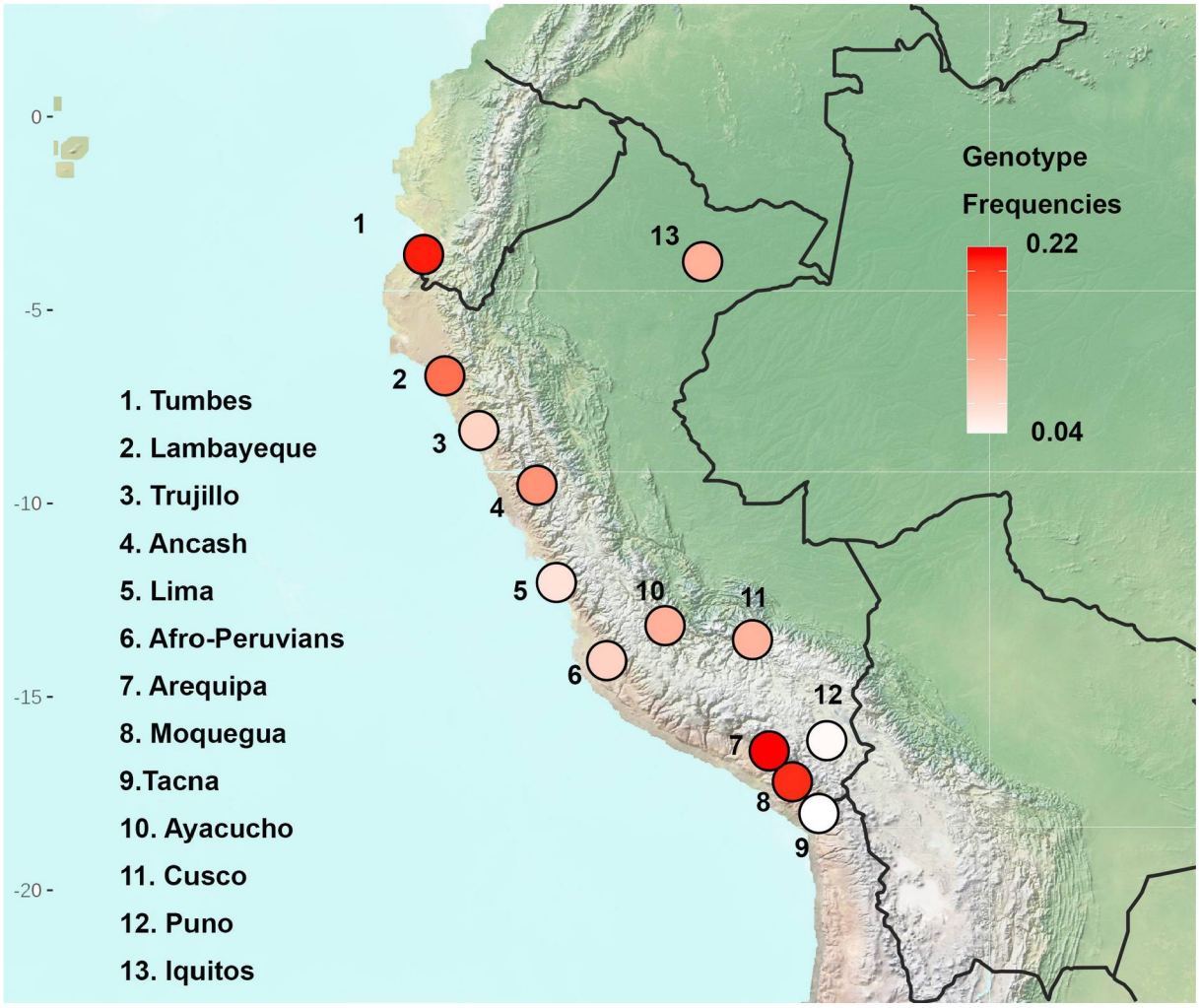
ZHONG, Y.; PRIEBE, M. G.; VONK, J.; HUANG, C.; ANTOINE, J.; HE, T.; HARMSEN, H. J. M.; WELLING, G. W. The Role of Colonic Microbiota in Lactose Intolerance. **Digestive Diseases and Sciences**, v. 49, n. 1, p. 78–8, 2004.

APÊNDICE 1

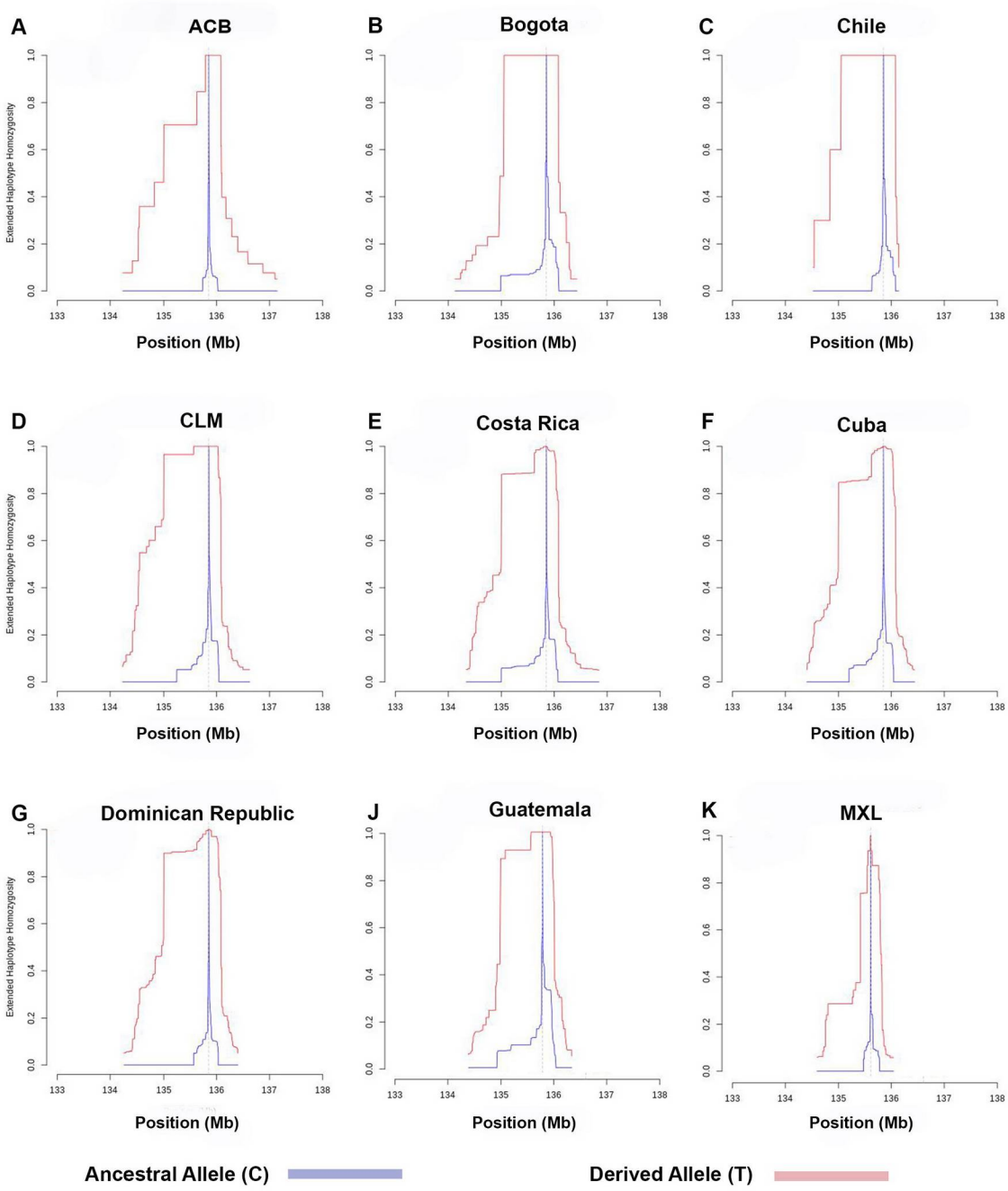
SUPPLEMENTARY MATERIAL



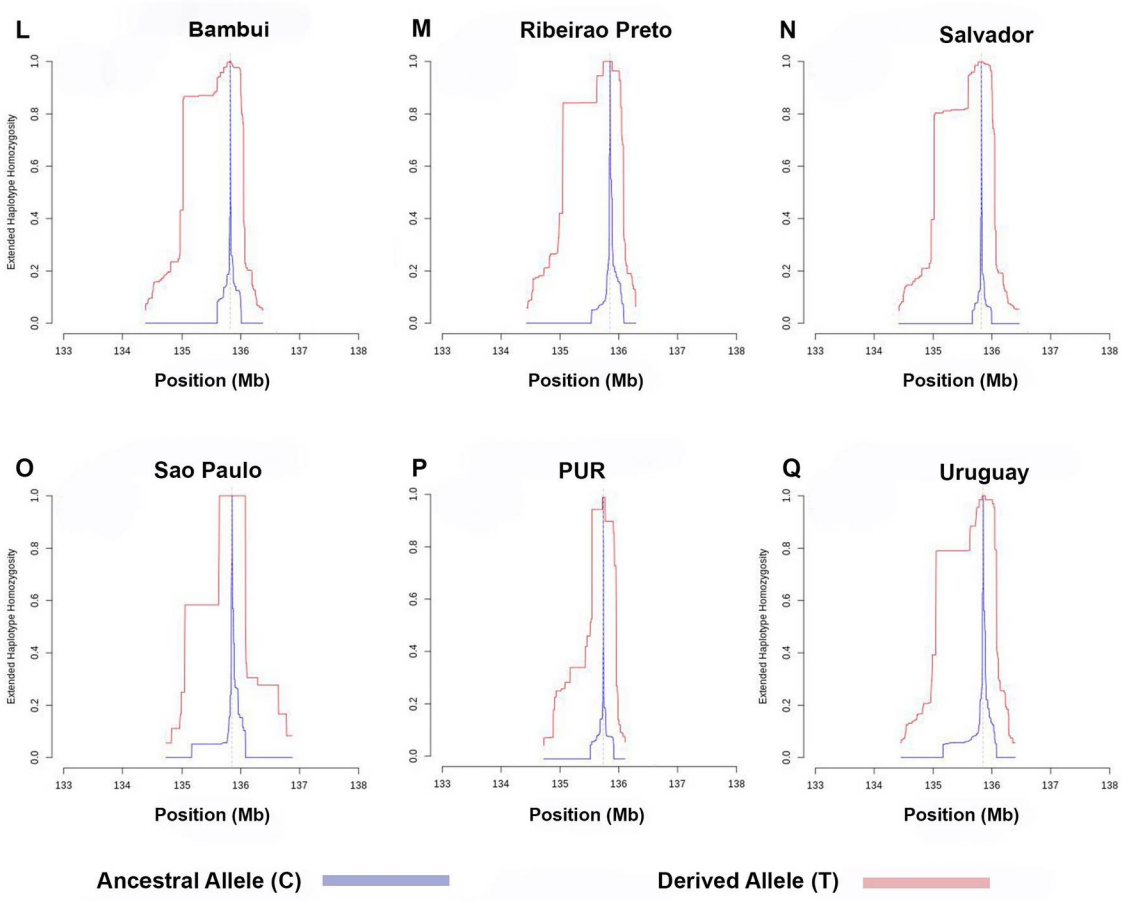
Supplementary Figure 1. Location of sampled Afro-Brazilians for the sequencing. Individuals were sampled from A) Curitiba (Paraná state) and its metropolitan region and B) Sertão do Valongo (Santa Catarina state).



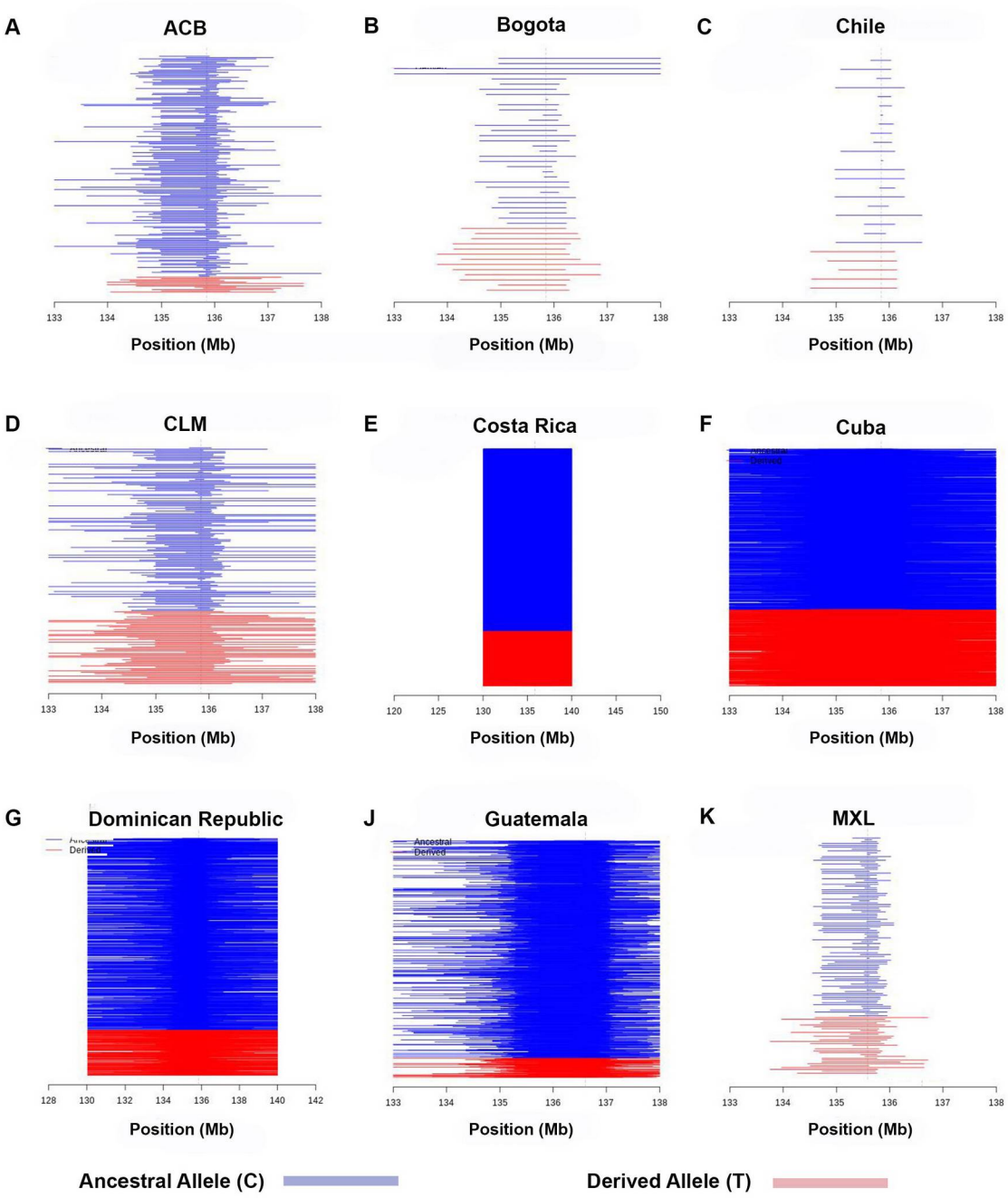
Supplementary Figure 2. Geographical distribution of the genotypes *TT/TC* for the *–13910C>T* polymorphism in admixed populations from Peru.



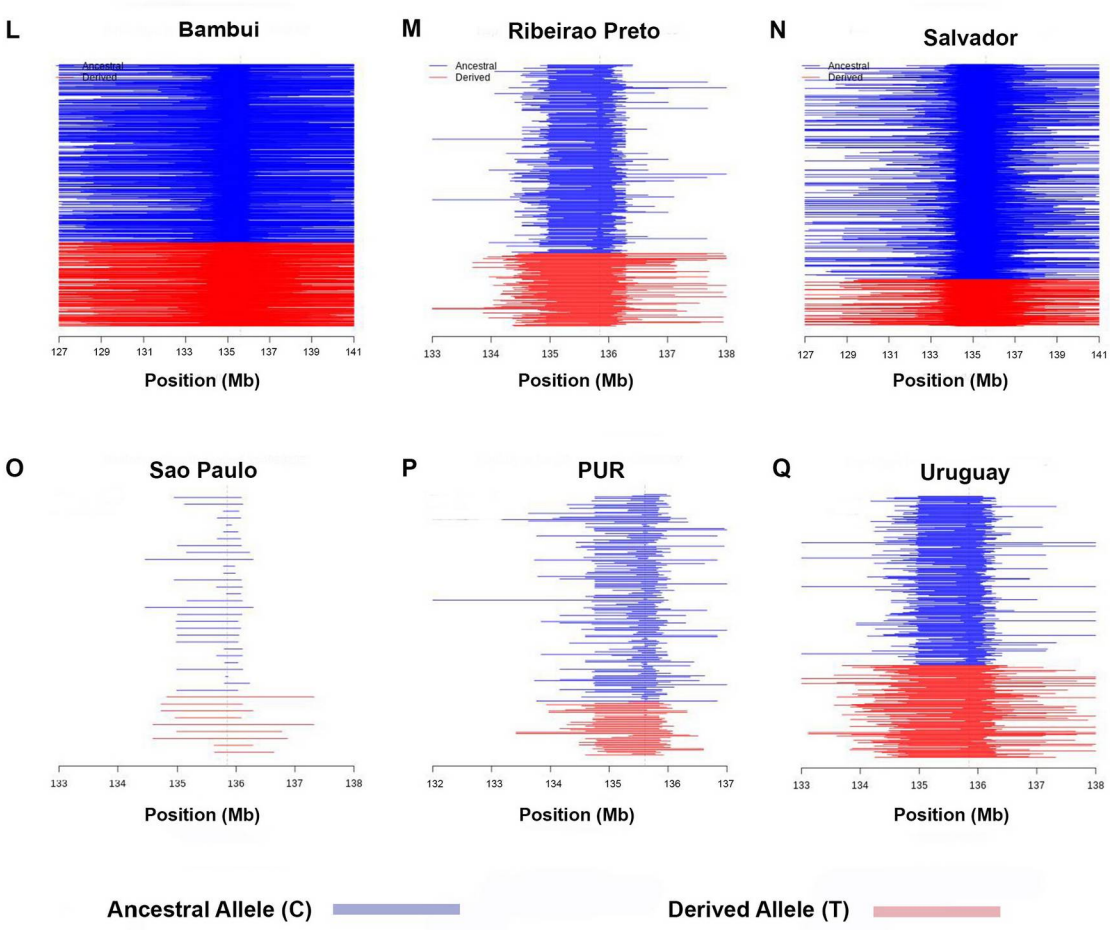
Supplementary Figure 3. Pattern of Extended Haplotype Homozygosity in Pan-American populations. The core allele corresponds to the *–13910*T* allele in the *MCM6* gene.



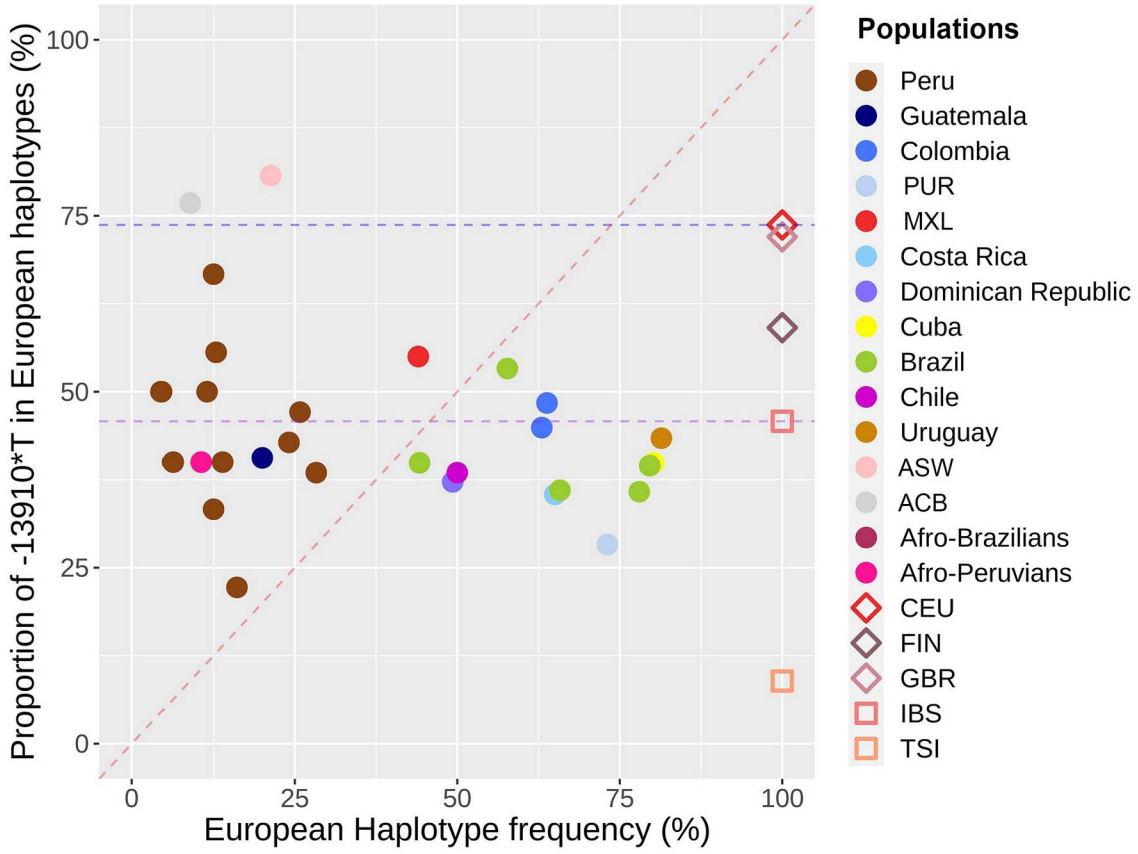
Supplementary Figure 3 (Continuation). Pattern of Extended Haplotype Homozygosity in Pan-American populations. The core allele corresponds to the –*13910*T* allele in the *MCM6* gene.



Supplementary Figure 4. Lengths of Extended Haplotype Homozygosity in Pan-American populations. The core allele corresponds to the *–13910*T* allele in the *MCM6* gene.



Supplementary Figure 4 (Continuation). Lengths of Extended Haplotype Homozygosity in Pan-American populations. The core allele corresponds to the –*13910*T* allele in the *MCM6* gene.



Supplementary Figure 5. Proportion of the *–13910*T* allele in European haplotypes vs Proportion of European haplotypes in the *MCM6* gene in each population. Blue and purple dashed lines correspond to the frequency of the derived allele in CEU (Northern European ancestry) and IBS (Southern European ancestry), respectively. Colombia includes CLM and Bogota individuals. Also, Diamonds and squares belong to Northern and Southern European populations, respectively.

Supplementary Tables captions

Supplementary Table 1. Detailed information of the population included in this study.

Supplementary Table 2. Concentration and volume of reagents used for the PCR reaction per sample.

Supplementary Table 3. Protocol used for PCR reaction.

Supplementary Table 4. Frequencies (allelic, genotypic and haplotypic), European ancestry and iHS values calculated for each population.

ANEXO 1

FREQUÊNCIA OBSERVADA PARA O FENÓTIPO LP EM POPULAÇÕES AFRICANAS

As frequências foram obtidas pelo *Global Lactase Persistence Association Database*, publicado originalmente por Itan e colaboradores (2010) e atualizado em 2013. Foram mantidas as populações que possuíam no mínimo dez indivíduos.

País	Etnia / Região	Número amostral	Frequência do fenótipo LP
África do Sul	Zulu	61	0,100
Egito	Cairo e Giza	67	0,330
	Delta do rio Nilo	291	0,270
	Canal de Suez	16	0,310
	Alto Egito, região norte	111	0,150
	Alto Egito, região sul	85	0,400
Etiópia	Somali	90	0,240
Gabão	Bantu	20	0,400
Namíbia	!Kung, da região noroeste do deserto do Kalahari	40	0,025
Níger	Tuareg	118	0,873
Nigéria	Hausa/Fulani	15	0,400
	Ibo	11	0,180
	Yoruba	48	0,160
Quênia	Maasai	26	0,880
	Ogiek	11	0,550
	Pokot	10	0,600
	Sengwer	12	0,170
	Tugen	11	0,730
	Yaaku	11	0,730
Ruanda (região do lago Kivu)	Hutu	36	0,417
	Tutsi	27	0,926
	Hutu-Tutsi	11	0,456
	Twa	22	0,179
	Shi	28	0,036

País	Etnia / Região	Número amostral	Frequência do fenótipo LP
Sudão	Misseri, povos Baggara	20	0,400
	Habbani, povos Baggara	19	0,474
	Nuba, povos aborígenes	58	0,207
	Bedja	9	0,889
	Gomoeia	31	0,677
	Kahli	21	0,619
	Nubians	21	0,333
	Shaygi	42	0,381
	Dongolawi	16	0,188
	Nilotic	18	0,333
	Jaali	113	0,531
	Jaali	94	0,479
	Amarar	82	0,870
	Artega	22	0,820
	Beni Amir	40	0,880
	Bisharin	22	0,860
	Dinka	208	0,250
	Haddendoa	137	0,800
	Nuer	23	0,220
	Shilluk	8	0,380
Tanzânia	Akie	11	0,550
	Burunge	16	0,380
	Hadza	15	0,600
	Iraqw	19	0,950
	Maasai	15	0,670
	Mbugu	23	0,430
	Rangi	26	0,650
	Sandawe	23	0,350
Tunísia	Tunisianos	43	0,170
Zambia	Bantu	26	0,039

ANEXO 2

FREQUÊNCIA OBSERVADA PARA O FENÓTIPO LP EM POPULAÇÕES EUROPEIAS

As frequências foram obtidas pelo *Global Lactase Persistence Association Database*, publicado originalmente por Itan e colaboradores (2010) e atualizado em 2013. Foram mantidas as populações que possuíam no mínimo dez indivíduos.

País	Etnia / Região	Número amostral	Frequência do fenótipo LP
Alemanha	Alemães, região Baden-Wurttemberg*	136	0,765
	Alemães, região Bayern*	221	0,864
	Alemães, região leste*	246	0,776
	Alemães, região noroeste*	341	0,912
	Alemães, região Rheinland e Pfalz*	182	0,863
	Alemães, região Schleswig-Holstein*	100	0,940
Áustria	Austríacos, região leste	181	0,773
	Austríacos, região oeste	166	0,855
Dinamarca	dinamarqueses	91	0,960
Espanha	Galegos (Galízia), região noroeste	850	0,675
Finlândia	finlandeses	638	0,830
França	franceses, região oeste	102	0,765
Grécia	Gregos, região continental	600	0,553
	Cretenses	50	0,440
	Cipriotas gregos (Chipre)	50	0,340

* Probandos divididos por região de acordo com o local de nascimento de seus avós.

País	Etnia / Região	Número amostral	Frequência do fenótipo LP
Hungria	Húngaros, região leste*	70	0,714
	Húngaros, região oeste*	100	0,720
	Húngaros, região nordeste*	103	0,583
	Húngaros miscigenados**	262	0,590
	Matyo, grupo étnico distinto desde o século 18	172	0,634
	Romai, origem Sinti-Romani (ciganos)	113	0,443
Irlanda	Irlandeses de Dublin***	50	0,960
Itália	Italianos de Nápoles	44	0,227
Itália	Italianos da Sicília	100	0,290
Itália	Italianos, região norte*	89	0,480
	Italianos, região central*	65	0,810
	Italianos, região sul*	51	0,590
Polônia	Poloneses	275	0,625
Polônia	Poloneses	21	0,714
República Tcheca	Tchecos	17	0,824
Reino Unido	Britânicos	150	0,953

* Probandos divididos por região de acordo com o local de nascimento de seus avós; **Probandos com avós nascidos em diferentes regiões; ***Probandos nascidos na Irlanda, com pais e pelo menos três avós irlandeses. Para aqueles com três avós irlandeses, o quarto deve ser britânico ou escocês ou galês.

ANEXO 3

FREQUÊNCIA OBSERVADA PARA O FENÓTIPO LP EM POPULAÇÕES DO ORIENTE MÉDIO

As frequências foram obtidas pelo *Global Lactase Persistence Association Database*, publicado originalmente por Itan e colaboradores (2010) e atualizado em 2013. Foram mantidas as populações que possuíam no mínimo dez indivíduos.

País	Etnia / Região	Número amostral	Frequência do fenótipo LP
Afeganistão	Hazara, áreas montanhosas da região central	10	0,200
	Miscigenados, indivíduos falantes das línguas Deri e Pashtu de etnia mista ou desconhecida	34	0,235
	Pasha-I, indivíduos assentados ao sul das montanhas Indocuche, nordeste de Kabul	60	0,133
	Pashtun, região central e sudeste	71	0,211
	Tajik, região central e oeste	79	0,177
	Uzbek, região norte	16	0
Irã	Iranianos	21	0,140
Israel	Árabes israelenses, bacia Mediterrânea	67	0,194
Jordânia	Indivíduos da zona urbana/agrícola ao oeste da Jordânia e Palestina	148	0,250
	Beduínos jordanianos	162	0,759
	Árabes jordanianos, origem Mediterrânea	56	0,232

País	Etnia / Região	Número amostral	Frequência do fenótipo LP
Kuwait	Árabes kuwaitianos	70	0,529
	kuwaitianos asiáticos	79	0,418
Líbano	Libaneses	75	0,220
Arábia Saudita	Árabes sauditas não beduínos	109	0,430
	Beduínos Árabes	21	0,810
	Yemeni	17	0,530
Turquia	Indivíduos da região da Anatólia central	104	0,290
	Indivíduos da região da Anatólia oriental	122	0,260
	Indivíduos da região da Anatólia ocidental e Turquia europeia	126	0,300
	Indivíduos da costa norte da Turquia	64	0,310
	Indivíduos da costa sul da Turquia	54	0,280

ANEXO 4

FREQUÊNCIA OBSERVADA PARA O FENÓTIPO LP EM POPULAÇÕES ÁSIA E DA AUSTRALÁSIA

As frequências foram obtidas pelo *Global Lactase Persistence Association Database*, publicado originalmente por Itan e colaboradores (2010) e atualizado em 2013. Foram mantidas as populações que possuíam no mínimo dez indivíduos.

Continente	País	Etnia / Região	Número amostral	Frequência do fenótipo LP
Ásia	China	Kazakh, região norte	195	0,236
		Mongols, região norte	198	0,121
		Han, região norte	248	0,077
	Índia	Indianos, região oeste	100	0,360
	Índia	Indianos, região norte	-	0,726
		Indianos, região sul	-	0,333
	Japão	Japoneses	40	0,275
	Paquistão	Baluchistani	32	0,380
		Kashmiri	27	0,300
		Punjabi	322	0,410
		Sindhi	33	0,420
	Sri Lanka	Ceilonenses	200	0,275
	Tailândia	Tailandeses	140	0,030
	Uzbequistão	Tajiko-Uzbek	100	0,110
		Kazakh	83	0,250
Ásia / Europa	Rússia	Khanty, região norte	115	0,280
		Komi-Izhems	56	0,370
		Mansi, região norte	81	0,290
		Nenets, Sibéria ocidental	9	0,220
		Kildin Saami	50	0,520
		Komi-Permiaks, falantes da língua fino-permiana	112	0,500
		Udmurtians, falantes da língua fino-permiana	75	0,410
Australásia	Austrália	Aborígines australianos	45	0,160
		Australianos não Aborígines	37	0,800
	Papua Nova Guiné	Indivíduos residentes das províncias de Sepik oriental e ocidental	35	0,229
	Papua Nova Guiné	Indivíduos de descendência Huli, Mendi e Dunai, região sul	30	0,100

ANEXO 5

Heredograma referente a Comunidade Quilombola de Sertão do Valongo (SC). Retirado de Souza e Culpi, 1992.

