Tracing the Distribution of European Lactase Persistence Genotypes Along the Americas

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

[1] Laboratório de Genética Molecular Humana, Departamento de Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, Brazil, [2] Programa de Pós-Graduação em Genética, Departamento de Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, Brazil, [3] Laboratório de Diversidade Genética Humana, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, 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, Peru, [7] Emerging Diseases and Climate Change Research Unit, School of Public Health and Administration, Universidad Peruana Cayetano Heredia, Lima, Peru, [8] Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, United States, [9] Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD, United States, [10] Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, United States, [11] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, United States, [12] Department of Epidemiology, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [13] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States, [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, Brazil, [16] Fundação Oswaldo Cruz, Instituto René Rachou, Belo Horizonte, Brazil, [17] Universidade Federal de Minas Gerais, Programa de Pós-Graduação em Saúde Pública, Belo Horizonte, Brazil, [18] Facultad de Ciencias de la Salud, Universidad de Huánuco, Huánuco, Peru, [19] Carrera de Medicina Humana, Facultad de Ciencias de la Salud, Universidad Científica del Sur, Lima, Peru, [20] Veterans Affairs Puget Sound Health Care System, Seattle, WA, United States, [21] Department of Neurology, University of Washington, Seattle, WA, United States, [22] Lerner Research Institute, Genomic Medicine, Cleveland Clinic, Cleveland, OH, United States, [23] Neurology Institute, Universidad de la República, Montevideo, Uruguay, [24] Department of Genetics, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay, [25] Ribeirão Preto Medical School, Universidade de São Paulo, Ribeirão Preto, Brazil, [26] Movement Disorders Unit, Department of Neurology and Neurosurgery, Universidade Federal de São Paulo, São Paulo, Brazil, [27] Departamento de Neurologia, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil, [28] Serviço de Neurologia, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil, [29] Departamento de Farmacologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [30] Instituto de Ciências da Saúde, Universidade Federal do Pará, Belém, Brazil, [31] CETRAM, Facultad de Ciencias Médicas, Universidad de Santiago de Chile, Santiago, Chile, [32] Neuroscience and Cell Death Research Groups, Medical School and Genetic Institute, Universidad Nacional de Colombia, Bogotá, Colombia, [33] Program for Personalized and Genomic Medicine, School of Medicine, University of Maryland, Baltimore, Baltimore, MD, United States, [34] Department of Medicine, School of Medicine, University of Maryland, Baltimore, Baltimore, MD, United States

In adulthood, the ability to digest lactose, the main sugar present in milk of mammals, is a phenotype (lactase persistence) observed in historically herder populations, mainly Northern Europeans, Eastern Africans, and Middle Eastern nomads. As the *−13910*T*

allele in the *MCM6* gene is the most well-characterized allele responsible for the lactase persistence phenotype, the *−13910C > T* (rs4988235) polymorphism is commonly evaluated in lactase persistence studies. Lactase non-persistent adults may develop symptoms of lactose intolerance 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 complex ancestry of Pan-American admixed populations, we studied the distribution of *−13910C > T* lactase persistence genotypes and its flanking haplotypes of European origin in 7,428 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 milk of mammals, 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 −14 kb upstream of the *LCT* gene, in its enhancer region, are believed to be responsible for the LP phenotype. The *−13910*T* allele (*−13910C > T*, rs4988235, intron 13, *MCM6* gene) is widely distributed in the European population; the allele frequencies vary from 8.9% in Tuscany (Italy) to 72.0% among British (England and Scotland), with an average of 50.8% in Europe (1000 Genomes Project Consortium, Auton 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 the LP phenotype. Among these, the most well-known are *−14010G > C* (rs145946881), *−14009T > G* (rs869051967), *−13915T > G* (rs41380347), and *−13907C > G* (rs41525747) (Tishkoff et al., 2007; Ranciaro et al., 2014; Liebert et al., 2017).

In the Americas, archaeozoological evidence supports the occurrence of mammal 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 fifteenth century (Gade, 1999) – cattle were not introduced in the Americas until 1493 (Primo, 1992). Not surprisingly, 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 most people cannot digest lactose naturally. Countries such as 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 frequency of $-14010*C$ ($-14010G > C$, rs145946881) in Afro-Brazilians from the South is 0.27% (Friedrich et al., 2012b); the frequencies of $-14011*T$ ($-14011C > T$, rs4988233) in admixed Brazilians from the North and the Northeast are 0.25 and 0.58%, respectively (Friedrich et al., 2012b); and the frequencies of $-13913*C$ ($-13913T > C$, 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) substitution 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).

Because of 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 historical records of the subcontinental origin of African peoples, previous genetic studies confirmed that West-Central African ancestry is the most prevalent in the Americas (Gouveia et al., 2020). Further on, between the nineteenth and twentieth 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 across 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

Publicly Available Datasets

We analyzed genomic information of 7,428 unrelated individuals from North, Central, and South Americas (**Supplementary Table 1**). This includes whole-genome sequencing (WGS) data from admixed individuals of the 1000 Genomes Project (1000 Genomes Project Consortium, Auton et al., 2015) [African Americans from Southwest United States (ASW); African Caribbeans from Barbados (ACB); individuals of Mexican ancestry from Los Angeles, United States (MXL); Puerto Ricans from Puerto Rico (PUR); and Colombians from Medellin (CLM)], as well as WGS data from individuals of 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) and the Peruvian Genome Project (Harris et al., 2018). We also included genotype array data from the LARGE-PD Project (Loesch et al., 2020) (individuals from Brazil, Colombia, Chile, and Uruguay), two Brazilian EPIGEN population-based cohorts (Kehdy et al., 2015) (individuals from Salvador and Bambui), and from admixed and Native American individuals of the Peruvian Genome Project (Harris et al., 2018; Borda et al., 2020). All these datasets were generated from different sources (**Supplementary Table 1**) but include the $-13910C > T$ SNP (rs4988235). In order to keep a higher SNP density for haplotype-based analyses, we organized the genomic information in two datasets: (i) LARGE-PD only and (ii) a merged dataset (including all other datasets). During the quality control of these two datasets, we excluded SNPs with significant missing data ($>10\%$), loci with 100% of heterozygosity, non-chromosomal information, A/T–C/G genotypes, and SNPs with minor allele count = 1 using PLINK 1.9 (Chang et al., 2015). For the merging process, we used the flags –bmerge and –flip when necessary. Finally, we kept biallelic SNPs and removed singletons and monomorphic positions as they are not informative for population genetic analyses. We ended with 1,010,078 and 1,528,206 SNPs for the LARGE-PD Project and the merged dataset, respectively.

We also considered 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 the 1000 Genomes Project (**Supplementary Table 1**).

Newly Generated Datasets

Moreover, we generated sequencing data for the *LCT* enhancer (*MCM6* gene, intron 13) for 259 Afro-Brazilians 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 classified 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 settlement history and isolation of the community 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

The Afro-Brazilians sample (n = 259) was sequenced for a fragment of 594 bp in the *MCM6* gene, including the *–13910C* > *T* SNP (rs4988235). The fragment was amplified by polymerase chain reaction (PCR) on a Mastercycler EP Gradient S® (Eppendorf, Germany). The forward and reverse primers used were as follows: 5'-GGCAGGGGTTTGGAACTTTC-3' and 5'-CTGTTGAATGCTCATACGACCA-3', respectively. Other reagents and the protocol used are described in **Supplementary Tables 2, 3**, respectively.

The PCR products were purified using exonuclease I (Fermentas, United States) and alkaline phosphatase (Thermo Fisher Scientific, United States) on a Mastercycler EP Gradient S® (Eppendorf) at 37°C for 1 h and 80°C for 15 min. The sequencing was performed using the same primers used in the PCR and BigDye® Terminator Cycle Sequencing Standard v3.1 (Life Technologies, United States), according to the instructions of the manufacturer. Sequencing reactions consisted of a first step at 95°C for 1 min, followed by 25 cycles of 95°C for 10 s, 50°C for 5 s, and 60°C for 4 min. The sequencing products were purified using ethanol (Merck, Germany), resuspended in Hi-Di Formamide (Life Technologies), and, finally, submitted to capillary electrophoresis in a 3500xl Genetic Analyzer Sequencer (Life Technologies).

We analyzed the obtained sequences using the Mutation Surveyor® 3.30 software (SoftGenetics, United States), which aligns the forward and reverse sequences to a human genome reference sequence (GRCh38) available in NCBI (National Center for Biotechnology Information) resources, enabling the evaluation of amplified sequences and the genotype annotation for the *–13910C* > *T* SNP (**Supplementary Table 4**). The sequences were deposited in the GenBank at NCBI website under the accession numbers MZ362598 to MZ362856.

Allelic and Genotypic Frequencies 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 dataset. For both analyses, we merged each dataset with a reference panel of 848 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 228 unadmixed Native American individuals from the Peruvian Genome Project (Borda et al., 2020). As the genetic information available for Afro-Brazilians from Curitiba and Sertão do Valongo corresponds only 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 were available for Afro-Brazilians from Curitiba.

For global ancestry analysis, we removed linked variants (r^2 > 0.1) using the PLINK flag –indep-pairwise with the following parameters: 50 10 0.1, and applied a minor allele frequency (MAF) filter of 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 (ρ) 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 analyzed the complete chromosome 2, which includes the *MCM6* gene, for each dataset (LARGE-PD Project dataset and the merged dataset) without removing linked variants. First, for each dataset, all chromosomes were 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, which perform 10 burn-in iterations, followed by four 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. After RFMix completion, we used the RFMix msp output, which includes the most likely assignment of ancestry per conditional random field point to obtain the size of the haplotypes per ancestry. The size was determined by considering the uninterrupted length (in base pairs) of the genomic region flanking the *–13910C* > *T* SNP until a switch in ancestry along the haplotype. 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.[1]

---

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

Natural Selection Analysis

In order to identify evidence of natural selection acting over the *MCM6* region, 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 inferred by RFMix for each population. Moreover, we analyzed the length distribution of uninterrupted European haplotypes around the *−13910C > T* SNP to explore the genomic region dynamics. 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 populations of European source is also observed in the admixed Pan-American populations. For this purpose, we used the rehh package (Gautier et al., 2017). Then, we estimated the integrated haplotype score (iHS), which indicates if a locus is under recent positive selection. This score compares the levels of linkage disequilibrium surrounding a positively selected allele with the ancestral allele background at the same position (Voight et al., 2006). For such iHS inference, we used the selscan software (Szpiech and Hernandez, 2014) with default parameters for phased information for each population. Finally, we calculated the genome-wide iHS Z-scores value by a normalization using the norm package, provided with selscan, in derived allele frequency bins with the −bin option set as 20. We calculated a two-tailed p-value of the SNPs based on the normalized iHS Z-scores by dividing the proportional rank of the statistic by the total number of values in the distribution.

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 25 Pan-American populations of 12 countries. The highest frequency of the *−13910*T* allele in the Americas occurs in the Uruguayan population (35%, **Figure 1** and **Supplementary Table 5**). Conversely, the lowest frequencies were observed for Peruvian and African descendant populations. A positive correlation was observed between *−13910*T* allele frequencies and the average proportion of European ancestry (ρ = 0.843, p = 4.414e-7). 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**). Moreover, in Afro-Brazilians from Curitiba, we identified the *−14011*T* (*−14011C > T*, rs145946881) and *−13915*G* (*−13915T > G*, rs41380347) LP-associated alleles in low frequencies (0.2%).

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 wi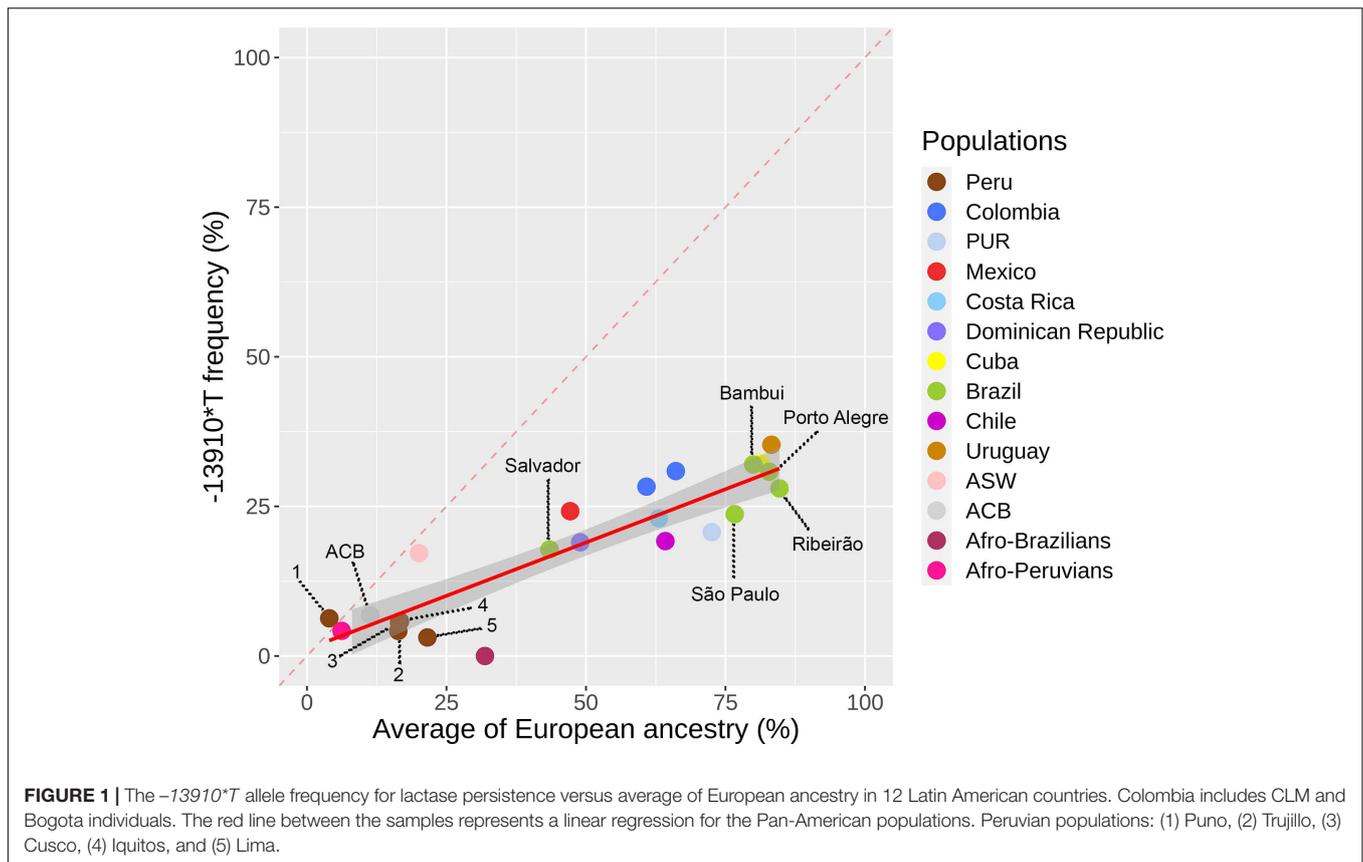th this phenotype. The observed genotype frequencies did not differ from those expected under Hardy–Weinberg equilibrium in all Pan-American admixed populations (**Supplementary Table 5**). Higher proportions of *T* allele carriers were observed in Uruguay (61%), Southeast–South Brazil (∼54%, including Bambui and Porto Alegre), and Cuba (53%), which also have higher proportions of European ancestry (**Figure 2** and **Supplementary Table 5**), suggesting that more than half of the individuals of these samples are LP. Conversely, this inferred phenotype is absent in ancient and some present-day Peruvian Native Americans (**Figure 2** and **Supplementary Table 5**). Furthermore, admixed Peruvians and African 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 nor other LP-associated alleles.

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 haplotypes (with or without the *−13910*T* allele) in the *MCM6* region, we estimated the proportion of uninterrupted European haplotypes and compared it to the genome-wide proportion of European ancestry for each population. Our results (**Figure 3B**) revealed a high correlation between European haplotype frequencies and the percentage of genome-wide European ancestry (ρ = 0.9435, p = 4.042e-6). 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 explored the admixture dynamics of this region by analyzing the length distribution of uninterrupted 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 African 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 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 2**). Peruvians from Cusco and the populations of Bogota, Chile, and Porto Alegre have a linkage disequilibrium (LD) block of 1 Mb with 100% EHH that includes the derived allele (**Figure 4** and **Supplementary Figure 2**). Observing

FIGURE 1 | The *–13910*T* allele frequency for lactase persistence versus average of European ancestry in 12 Latin American countries. Colombia includes CLM and Bogota individuals. The red line between the samples represents a linear regression for the Pan-American populations. Peruvian populations: (1) Puno, (2) Trujillo, (3) Cusco, (4) Iquitos, and (5) Lima.

the length of the EHH blocks (**Figure 4** and **Supplementary Figure 3**), it is remarkable that African American and Porto Alegre populations have more extended and more homogeneous 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 African Americans, some haplotypes reach 5 Mb. Moreover, several Pan-American populations presented iHS values greater than 2 (**Supplementary Table 5**), whereas African Americans presented the highest value (iHS = 4.2, **Figure 5**).

Interestingly, when we calculated the derived haplotype frequency, which includes the *–13910*T* allele, in European haplotypes for each Pan-American population, the African Americans from Southwest United States (ASW) had the highest value (80%) (**Supplementary Table 5**). This frequency is quite similar to the frequency of the derived haplotype in Northern European populations (**Supplementary Figure 4**). Moreover, we observed that populations that were Spaniard or Portuguese colonies (Peru, Colombia, Mexico, Chile, Uruguay, and Brazil) had derived haplotype frequencies more similar to those of Iberians from the 1000 Genomes Project (**Supplementary Figure 4**). 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

LP 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 (Duncan and Scott, 1972; Caskey et al., 1977). 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. Furthermore, the identification of the *–14011*T* and *–13915*G* alleles in low frequencies in Afro-Brazilians from Curitiba agrees with the frequencies previously reported (Friedrich et al., 2012b; Paz-Y-Miño et al., 2016), corroborating the suggestion that the phenotypic variation observed in this study is mainly due to the *–13910C > T* variation of European origin.

In the present study, we conducted the most geographically extensive analysis of the *–13910*T* allele spanning several Pan-American admixed populations (North American African 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 the current Latin American
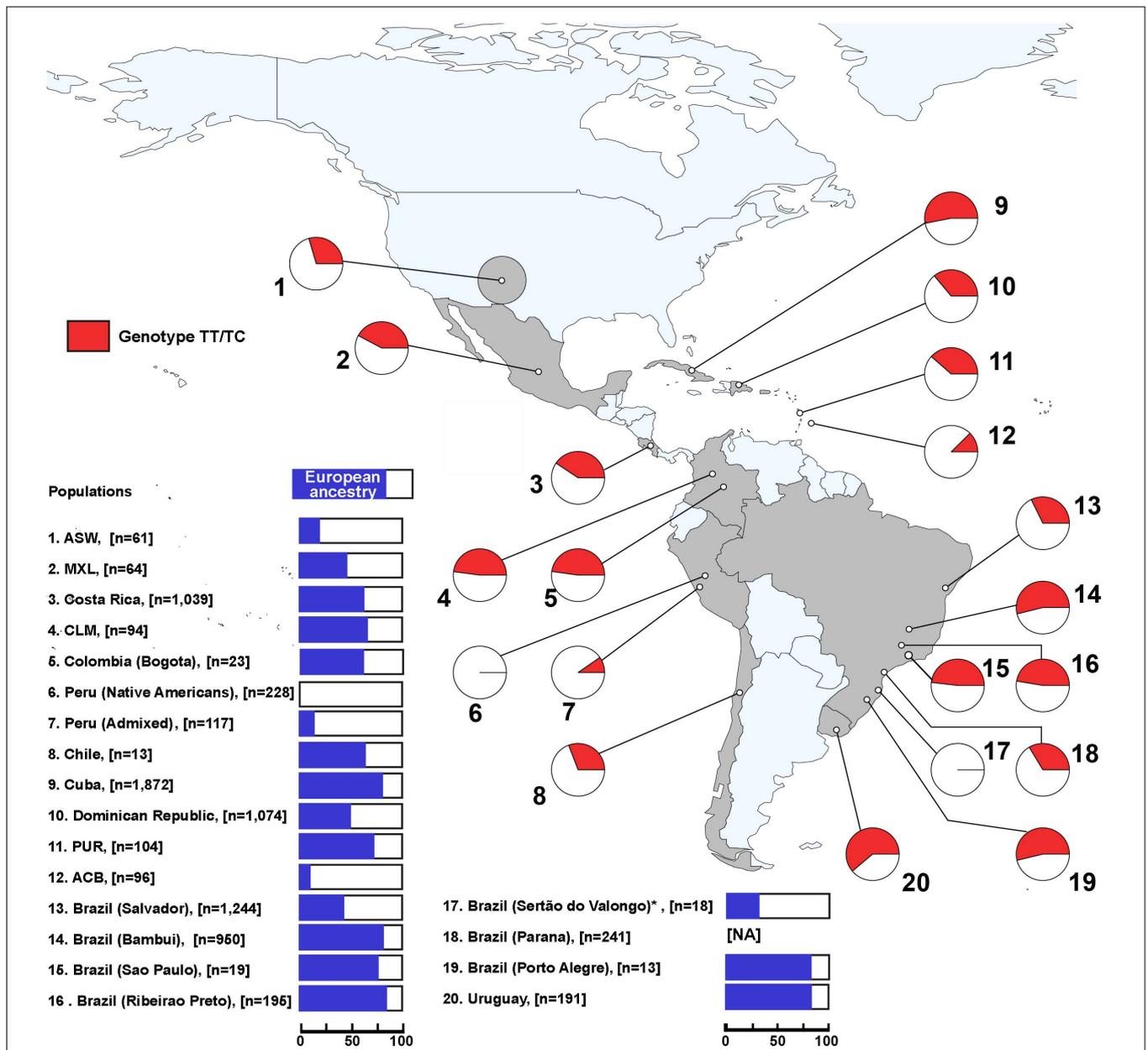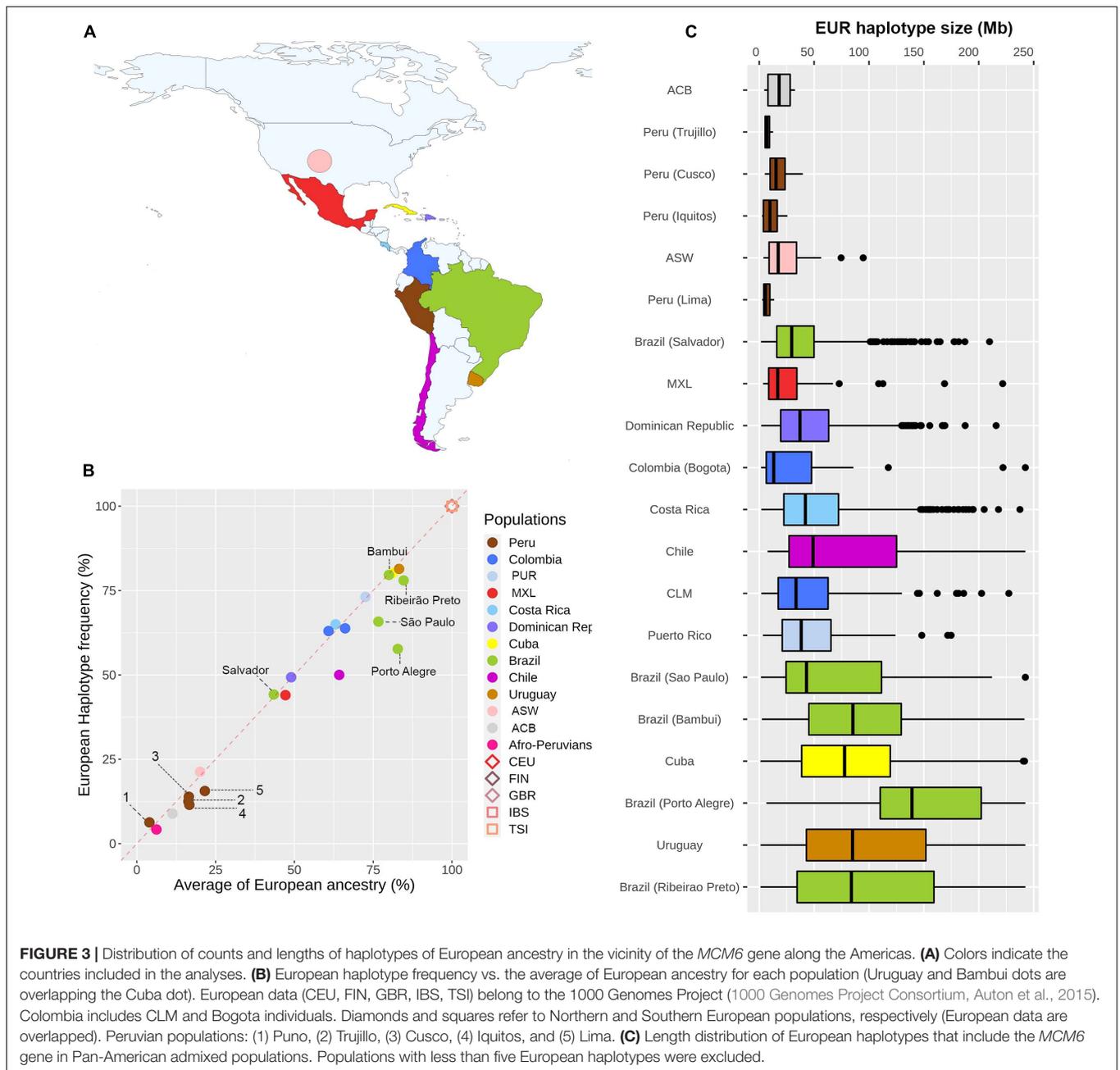
FIGURE 2 | Geographical distribution of *TT* and *TC* genotypes of −13910C > T along the Americas. The red color in the pie charts represents the combined frequencies of *TT* and *TC* genotypes in each population. On the left, the blue color in the bar plots indicates the average proportion of European ancestry for each population. NA, no available data; ASW, African Americans from Southwest United States; MXL, Mexican ancestry from Los Angeles, United States; ACB, African 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).

nourishment policies, our results represent an important aspect to be considered for the understanding of the relationship between the consumption of dairy products and well-being across the continent.

We can address two limitations of our study. The first was the uneven representation of the population sampled within the countries: sample sizes varied from 13 individuals (Chile) to more than 1,000 individuals (Costa Rica, *n* = 1,039). However, if we restrict our analyses to countries with more

than 30 sampled individuals, one may still observe the patterns described. Second, to maintain a balance between individuals and SNP density, we used two separate datasets for the analyses (LARGE-PD and a merged dataset). However, as the study was focused on the −13910C > T SNP and its flanking haplotypes of European origin, the genomic ancestry would still be the same if inferred with a different set of SNPs, as long as the inference is performed with the same SNP density.
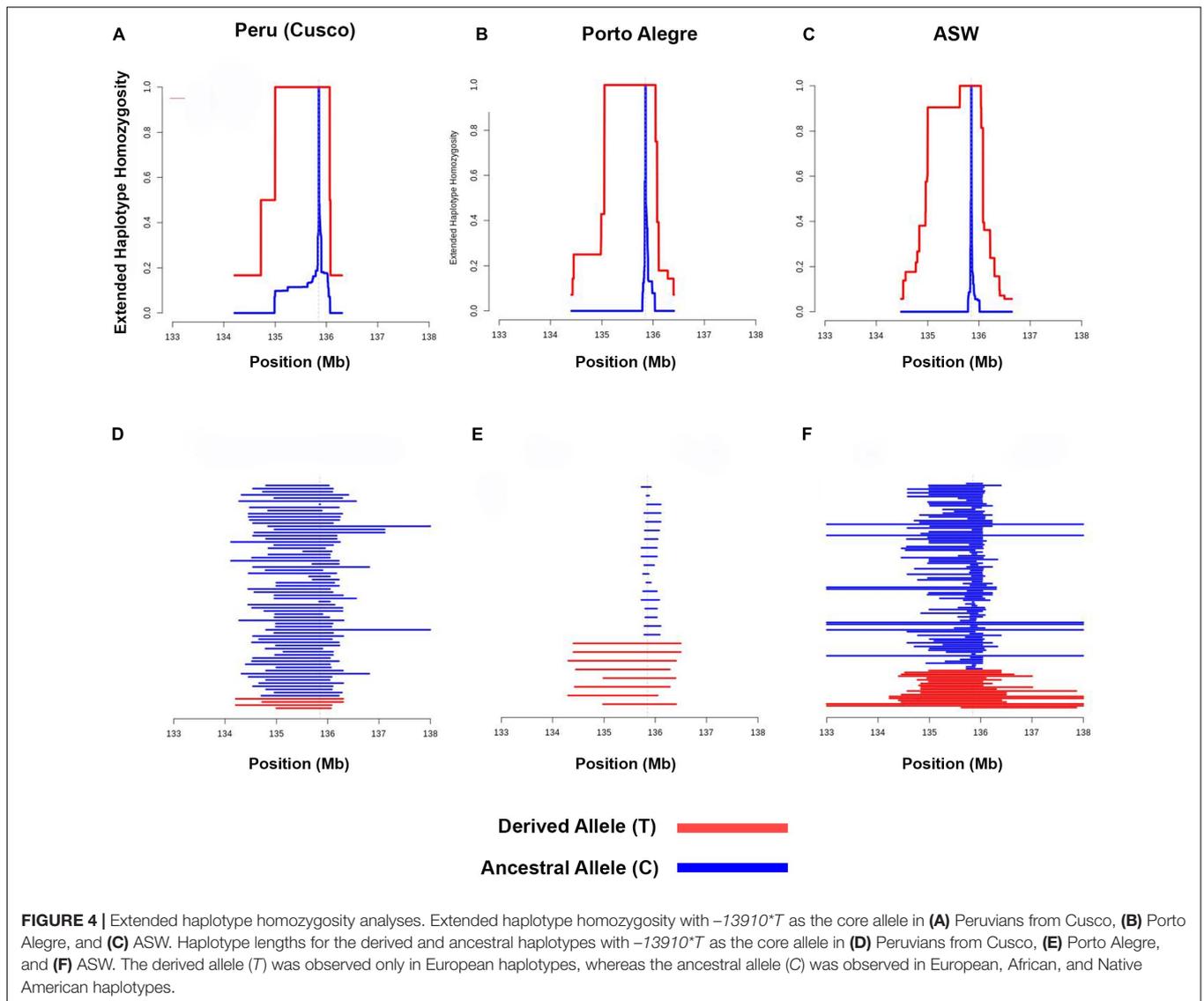
FIGURE 3 | Distribution of counts and lengths of haplotypes of European ancestry in the vicinity of the *MCM6* gene along the Americas. **(A)** Colors indicate the countries included in the analyses. **(B)** European haplotype frequency vs. the average of European ancestry for each population (Uruguay and Bambui dots are overlapping the Cuba dot). European data (CEU, FIN, GBR, IBS, TSI) belong to the 1000 Genomes Project (1000 Genomes Project Consortium, Auton et al., 2015). Colombia includes CLM and Bogota individuals. Diamonds and squares refer to Northern and Southern European populations, respectively (European data are overlapped). Peruvian populations: (1) Puno, (2) Trujillo, (3) Cusco, (4) Iquitos, and (5) Lima. **(C)** Length distribution of European haplotypes that include the *MCM6* gene in Pan-American admixed populations. Populations with less than five European haplotypes were excluded.

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 fifteenth century, European migration to the Americas was intensified during the late nineteenth 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 low percentage of European ancestry correlated with a low −13910*T allele frequency and a low European haplotype frequency. Moreover, a short European haplotype was observed, possibly introduced earlier in the Salvador population than in other Brazilian populations. A similar pattern was observed in

FIGURE 4 | Extended haplotype homozygosity analyses. Extended haplotype homozygosity with –13910*T as the core allele in **(A)** Peruvians from Cusco, **(B)** Porto Alegre, and **(C)** ASW. Haplotype lengths for the derived and ancestral haplotypes with –13910*T as the core allele in **(D)** Peruvians from Cusco, **(E)** Porto Alegre, and **(F)** ASW. The derived allele (T) was observed only in European haplotypes, whereas the ancestral allele (C) was observed in European, African, and Native American haplotypes.

African descendant populations, in which the lowest average of European ancestry was detected. In African Americans from Southwest United States, 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 United States 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 African Americans. The patterns observed in our work are consistent with one-way admixture events that have been 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 low –13910*T allele frequency, a low European haplotype frequency, and a short European haplotype – similarly to those observed in African 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 the independence of Peru 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 short European haplotypes observed in populations from 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 marked European migration to this country in the early twentieth century (Salzano and Bortolini, 2005).
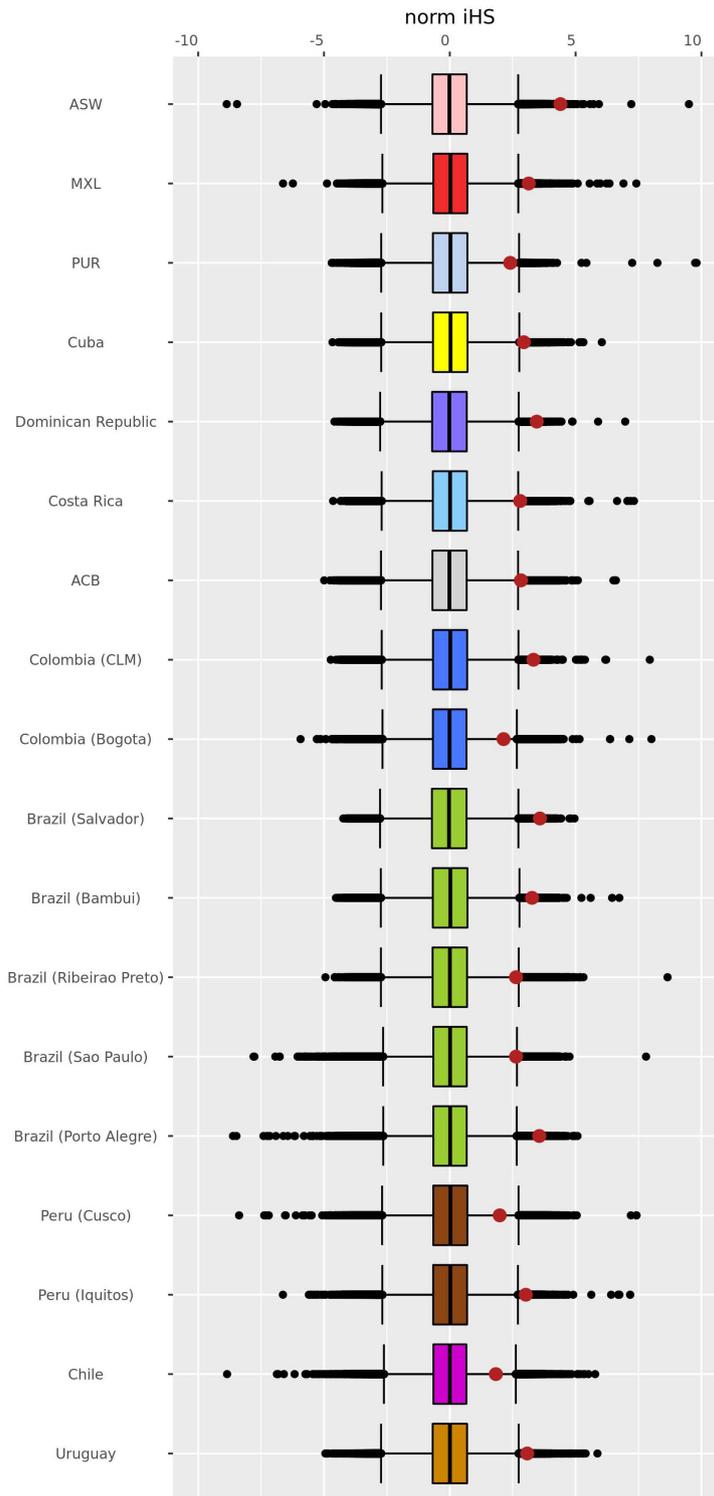
FIGURE 5 | Genome-wide distribution of normalized iHS Z-scores. The scores were estimated using the genome-wide distribution of the unstandardized iHS in Pan-American admixed populations with –13910*T allelic frequencies > 5% and more than 10 individuals. After normalization, the mean and median are 0, and the standard deviation is 1. The whisker ends of the box plot indicate a value of ± 2.69, delimiting 99.3% of the distribution. The middle line inside the box plot corresponds to the median of the normalized iHS Z-score distribution. The red dots refer to normalized iHS Z-scores for the –13910*T allele. ASW, African Americans from Southwest United States; MXL, Mexican ancestry from Los Angeles, United States; PUR, Puerto Ricans from Puerto Rico; ACB, African Caribbeans from Barbados; CLM, Colombians from Medellin. Colors represent the countries just as the map in **Figure 3A**. Normalized iHS Z-scores and empirical p-values for the estimated –13910*T allele are described in **Supplementary Table 5**.

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

LP in an Afro-Brazilian Quilombo Community

Between the sixteenth and mid-nineteenth 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 a European one of 31.9%, therefore with no reported Native American 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 agree with the settlement history of the community 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 (de Almeida, 2010). Hence, as 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 Postadmixture Selective Pressure for LP in the Pan-American Admixed Populations

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 greater than 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, African 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.

In contrast, by evaluating goat herders from Coquimbo (Chile), Montalva et al. (2019) reported a significant enrichment for European ancestry in the *LCT* gene when compared to genome-wide European ancestry, suggesting recent positive selection after admixture. However, the enrichment for European ancestry in the *LCT* gene was absent in urban non-pastoralist Latin American populations evaluated [e.g., MXL, CLM, and PEL (Peruvians in Lima, Peru) from the 1000 Genomes Project].

Thus, the lack of post-admixture selective pressure observed in our study – which datasets included MXL, CLM (1000 Genomes Project Consortium, Auton et al., 2015), and Peruvians (Harris et al., 2018; Borda et al., 2020) – agrees with Montalva et al. (2019), supporting their hypothesis of specific adaptation to milking agropastoralism in the Coquimbo population.

It should be noted that admixture processes could hinder the detection of selection signals due to changes in the allelic frequencies and linkage disequilibrium patterns (Lohmueller et al., 2011; Hamid et al., 2021). Furthermore, a recent study shows that a strong selective pressure of an adaptive phenotype could lead to genome-wide changes, modifying chromosome regions not directly involved in the phenotype, which can potentially bias the ancestry inference (Lohmueller et al., 2011; Hamid et al., 2021). Therefore, although it is unlikely that the LP phenotype has promoted differential survival in the American continent in the last centuries, we cannot ignore that the aspects early mentioned may have influenced our results.

Impact of Our Results on Public Health Policies

Dairy plays an essential role in human nutrition because of its supply of protein, lipids, and micronutrients, such as calcium, magnesium, and vitamins B_5 and B_{12}, 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. Whereas some studies have linked its consumption to the development of cardiovascular diseases and obesity – possibly because of the fat content of milk (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 Mdel 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, 1990; Porrata-Maury, 2004; Comisión Intersectorial de Guías Alimentarias Para Costa Rica Ministerio de Salud, 2010; Secretaría 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 US program 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 US 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 complex 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 who 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 is also dependent on the lactose concentration of the food, 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, only a few populations from Cuba, Brazil (Bambui and Porto Alegre), and Uruguay have an elevated proportion (greater than 50%) of individuals who 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 is rare (Kiple and Ornelas, 2001). However, it should be noted that European immigrants (and their descendants) have established such practices in culinary in America, resulting in the recent development of typically 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, as consumption of dairy products is known to be lower in developing countries than in developed ones, despite the changes over the last decades due to 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 African descendant populations (African Americans, African Caribbeans, Afro-Brazilians from Curitiba,

and the quilombola community of Sertão do Valongo). According to the data 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 fully understood, alternative milk products and plant-based dairy substitutes would possibly be a better option for the dietary guidelines of America.

DATA AVAILABILITY STATEMENT

Newly generated sequences were deposited in the GenBank at NCBI website under the accession numbers MZ362598–MZ362856.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Brazilian CONEP (Comissão Nacional de Ética em Pesquisa). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MHB, CS, and ViB designed the study. AG and NS performed the Sanger sequencing and the analysis of Afro-Brazilian populations. MHB, MP-E, RL, and IR contributed with samples and reagents. CS, MD, J-YM, RK, KN, SW, MLB, ML-C, HG, OC, CP, ET-S, IM, ED, VR, AL, VT, VaB, HF, CR, AS-S, BS-L, PC-C, WF, GA, HA, and CA-B generated the datasets and were responsible by them. ViB, MHB, AG, NS, GA, MMe, TL, MS, CS, DL, MMa, and TO'C, analyzed the data. AG, NS, MHB, and ViB drafted the manuscript. GAC, MMe, TL, MP-E, TO'C, OC, MD, and ET-S critically edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.671079/full#supplementary-material

REFERENCES

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., and Kang, H. M. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393

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. doi: 10.1016/j.gde.2016.09.003

Albala, K. (2011). *Food Cultures of the World Encyclopedia*. Santa Barbara, CA: 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. doi: 10.1101/gr.094052.109

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. doi: 10.3390/nu12092689

Auricchio, S., Rubino, A., Landolt, M., Semenza, G., and Prader, A. (1963). Isolated intestinal lactase deficiency in the adult. *Lancet* 2, 324–326. doi: 10.1016/s0140-6736(63)92991-x

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. doi: 10.1001/archpedi.159.6.543

Bermudez, O. I., and Tucker, K. L. (2003). Trends in dietary patterns of Latin American populations. *Cad. Saude Publica* 19, S87–S99.

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. doi: 10.1086/421051

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, 889–902.

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.* 117, 32557–32565. doi: 10.1073/pnas.2013773117

Buttriss, J. (1997). Nutritional properties of fermented milk products. *Int. J. Dairy Technol.* 50, 21–27. doi: 10.1111/j.1471-0307.1997.tb01731.x

Casey, J. (2005). "Holocene occupations of the forest and savanna," in *African Archaeology*, ed. A. B. Stahl (Oxford: Blackwell Publishing).

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. doi: 10.1007/bf01072952

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*. San José: Comisión Intersectorial de Guias Alimentarias Para Costa Rica.

de Almeida, A. W. B. (2010). *Cadernos de Debates Nova Cartografia Social: Territórios Quilombolas e Conflitos*. Norwich: UEA.

de Mello e Souza, M. (2007). *África e Brasil Africano*. Brzil: Editora Ática.

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.

Dillehay, T. D. (2009). Probing deeper into first American studies. *Proc. Natl. Acad. Sci. U. S. A.* 106, 971–978. doi: 10.1073/pnas.0808424106

Dirección General de la Salud Ministerio de Salud de Uruguay (2019). *Guía Alimentaria Para la Población Uruguaya: Para una Alimentación Saludable, Compartida y Placentera*. Montevideo: Ministerio de Salud de Uruguay.

Duncan, I. W., and Scott, E. M. (1972). Lactose intolerance in Alaskan Indians and Eskimos. *Am. J. Clin. Nutr.* 25, 867–868. doi: 10.1093/ajcn/25.9.867

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–734S.

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. doi: 10.1136/jech.2006. 053157

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. doi: 10.1038/ng826

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. doi: 10.1002/ajhb.22775

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. doi: 10.1007/bf02238168

Food and Agriculture Organization of the United Nations (2018). *FAOSTAT Database. New Food Balances. FAOSTAT Database*. Available Online 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. doi: 10.1002/ajpa.22010

Friedrich, D. C., Santos, S., Ribeiro-dos-Santos, ÂK. 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. doi: 10.1371/journal.pone.0046520

Gade, D. W. (1999). *Nature and Culture in the Andes*. Madison: University 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. doi: 10.1111/1755-0998.12634

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. doi: 10.1126/science.1153569

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. doi: 10.1093/molbev/msaa033

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. doi: 10.1086/203472

Hamid, I., Korunes, K. L., Beleza, S., and Goldberg, A. (2021). Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. *Elife* 10:e63177.

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.* 115, E6526–E6535.

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. *Eur. J. Clin. Invest.* 38, 541–547. doi: 10.1111/j.1365-2362.2008.01966.x

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. doi: 10.1093/ajcn/64.2.232

Hespanhol, R. A. 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. doi: 10.1590/s1982-45132013000300003

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. doi: 10.1007/s00439-008-0593-6

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

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*. Rome: FAO.

Instituto Nacional de Nutrición (1990). *Guías de Alimentación Para Venezuela*. Mexico: 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. doi: 10.1186/1471-2148-10-36

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. doi: 10.1093/nutrit/nuz093

John, R. (1997). *Africa: a Biography of the Continent*. London: Penguin Adult.

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. doi: 10.1098/rspb.2001.1774

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. doi: 10.1073/pnas.1504447112

Kiple, K. F., and Ornelas, K. C. (2001). *The Cambridge World History of Food*. Cambridge: 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. doi: 10.1038/nsmb.3227

Lahiri, D. K., and Nurnberger, J. I. (1991). A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res.* 19:5444. doi: 10.1093/nar/19.19.5444

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. doi: 10.1159/000363229

Liebert, A., López, S., Jones, B. L., Montalva, N., Gerbault, P., Lau, W., et al. (2017). World-wide distributions of lactase persistence alleles and the complex effects of recombination and selection. *Hum. Genet.* 136, 1445–1453. doi: 10.1007/s00439-017-1847-y

Loesch, D., Andrea, R. V., Heilbron, K., Sarihan, E. I., Inca-Martinez, M., and Mason, E. (2020). Characterizing the genetic architecture of Parkinson's disease in Latinos. *Ann. Neurol.* doi: 10.1002/ana.26153

Lohmueller, K. E., Bustamante, C. D., and Clark, A. G. (2011). Detecting directional selection in the presence of recent admixture in African-Americans. *Genetics* 187, 823–835. doi: 10.1534/genetics.110.122739

Luizon, M. R. (2007). *Dinâmica da Mistura Étnica em Comunidades Remanescentes de Quilombos Brasileiras. Tese de Doutorado, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto*. Available online at: https://www.teses.usp.br/teses/disponiveis/17/17135/tde-31052011-092816/pt-br.php

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. doi: 10.1002/j.1460-2075.1988.tb03124.x

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. doi: 10.1016/j.ajhg.2013.06.020

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. doi: 10.1016/j.gde.2020.06.007

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. doi: 10.1590/s0004-28032012000100002

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. doi: 10.1111/ahg.12277

Morales-Ruán Mdel, 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, S199–S205.

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*. Maceioì, AL: EdUFAL.

Muehlhoff, E., Bennett, A., and McMahon, D. (2013). *Milk and Dairy Products in Human Nutrition*. Rome: 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. 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. doi: 10.1159/000446241

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. doi: 10.1002/ajhb.22865

Pena, S. D. J., Di Pietro, G., Fuchshuber-Moraes, M., Genro, J. P., Hutz, M. H., Kehdy, F. 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. doi: 10.1371/journal.pone.0017063

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. doi: 10.1002/ajmg.c.31853

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. Ministerio de Salud publica (Havana: Ministerio de Salud Pública).

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 Histórica Sobre a População Negra no Estado do Paraná*. Paraná: 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. doi: 10.1016/j.ajhg.2014.02.009

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. doi: 10.1371/journal.pgen.1004572

Salzano, F. M. (2011). The prehistoric colonization of the Americas: evidence and models. *Evol. Educ. Outreach* 4, 199–204. doi: 10.1007/s12052-011-0330-9

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

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor: 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. doi: 10.1111/ehr.12787

Secretaria de Salud de México (2010). *Guía de Alimentos Para la Población Mexicana. {Secretaria de Salud de Mexico}*. Available Online at: https://www.yumpu.com/es/document/read/63974919/guia-de-alimentos-para-la-poblacion-mexicana

Segall, J. J. (1994). Dietary lactose as a possible risk factor for ischaemic heart disease: review of epidemiology. *Int. J. Cardiol.* 46, 197–207. doi: 10.1016/0167-5273(94)90242-9

Ségurel, L., and Bon, C. (2017). On the evolution of lactase persistence in humans. *Annu. Rev. Genomics Hum. Genet.* 18, 297–319. doi: 10.1146/annurev-genom-091416-035340

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. doi: 10.1371/journal.pbio.3000742

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. doi: 10.1038/ejhg.2008.156

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.* 6, 208–223. doi: 10.1007/s40142-018-0151-z

Souza, I. R., and Culpi, L. (1992). Valongo, an isolated braziliann black community. i. structure of the population. *Rev. Bras. Genét.* 15, 439–447.

Souza, I. R., and Culpi, L. (2005). Valongo, genetic studies on an isolated Afro-Brazilian community. *Genet. Mol. Biol.* 28, 402–406. doi: 10.1590/s1415-47572005000300012

Stifel, D., and Alderman, H. (2006). The "glass of milk" subsidy program and malnutrition in peru. *World Bank Econ. Rev.* 20, 421–448. doi: 10.1093/wber/lhl002

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. doi: 10.1093/molbev/msu211

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. doi: 10.1038/ng1946

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. doi: 10.1080/00365513.2017.1318445

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. doi: 10.1371/journal.pbio.0040072

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. doi: 10.1203/00006450-199410000-00018

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

Wheeler, J. C. (1995). Evolution and present situation of the South American camelidae. *Biol. J. Linn. Soc. Lond.* 54, 271–295. doi: 10.1016/0024-4066(95)90021-7

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., and He, T. (2004). The role of colonic microbiota in lactose intolerance. *Dig. Dis. Sci.* 49, 78–83. doi: 10.1023/b:ddas.0000011606.96795.40