Received: 25 July 2022 | Revised: 18 October 2022 | Accepted: 25 October 2022

DOI: 10.1002/ajpa.24656

RESEARCH ARTICLE

AMERICAN JOURNAL OF BIOLOGICAL ANTHROPOLOGY

WILEY

Genetic adaptations to potato starch digestion in the Peruvian Andes

Kelsey Jorgensen[1,2] | Obed A. Garcia[3] | Melisa Kiyamu[4] | Tom D. Brutsaert[5] | Abigail W. Bigham[1]

[1]Department of Anthropology, University of California, Los Angeles, California, USA

[2]Department of Anthropology, Wayne State University, Detroit, Michigan, USA

[3]Department of Biomedical Data Science, Stanford University, Stanford, California, USA

[4]Departamento de Ciencias Biológicas y Fisiológicas, Universidad Peruana Cayetano Heredia, Lima, Peru

[5]Department of Exercise Science, Syracuse University, Syracuse, New York, USA

Correspondence

Kelsey Jorgensen, Department of Anthropology, University of California, Los Angeles, CA, USA.

Email: kcjorgensen@ucla.edu

Funding information

National Science Foundation, Grant/Award Number: BCS-1132310

Abstract

Objectives: Potatoes are an important staple crop across the world and particularly in the Andes, where they were cultivated as early as 10,000 years ago. Ancient Andean populations that relied upon this high-starch food to survive could possess genetic adaptation(s) to digest potato starch more efficiently. Here, we analyzed genomic data to identify whether this putative adaptation is still present in their modern-day descendants, namely Peruvians of Indigenous American ancestry.

Materials and methods: We applied several tests to detect signatures of natural selection in genes associated with starch-digestion, *AMY1*, *AMY2*, *SI*, and *MGAM* in Peruvians. These were compared to two populations who only recently incorporated potatoes into their diets, Han Chinese and West Africans.

Results: Overlapping statistical results identified a regional haplotype in *MGAM* that is unique to Peruvians. The age of this haplotype was estimated to be around 9547 years old.

Discussion: The *MGAM* haplotype in Peruvians lies within a region of high transcriptional activity associated with the REST protein. The timing of this haplotype suggests that it arose in response to increased potato cultivation and attendant consumption. For Peruvian populations that relied upon the high-starch potato as a major source of nutrition, natural selection likely favored these *MGAM* variant(s) that led to more efficient digestion and increased glucose production. This research provides further support that detecting subtle shifts in human diet can be a major driver of human evolutionary change, as these results indicate that there is global variation in human ability to better digest high-starch foods.

KEYWORDS

Andes, diet, natural selection, potato, starch digestion

INTRODUCTION

Human evolutionary change has been shaped by dietary shifts including the introduction of cooking, plant and animal domestication, as well as other cultural events (Luca et al., 2010). Identifying genetic targets of diet-related selective pressures can improve our understanding of human evolutionary history, the nature of evolutionary transitions, and can also shed light on the genetic basis of modern metabolic diseases that affect human populations today. One example of an extreme dietary shift was the increase of starch-based foods in the diets of post-agricultural populations. While archaeological evidence suggests that high-starch foods constituted an important part of past hunter-gatherer diets, they comprised a smaller proportion than the estimated 50%–55% of daily caloric intake in agricultural societies today (Robertson et al., 2018). However, in especially arid or challenging environments, modern hunter-gatherers often consume higher quantities of starch-based foods when other resources are scarce (Hardy et al., 2009; Yasuoka, 2006). Of the world's major staple crops, the potato is a calorically rich food source that thrives in harsh conditions, making it historically and currently one of the most consumed starch-based crops in the world (Devaux et al., 2014; Singh et al., 2020). The potato's origin was within the Peruvian Andes, and it is likely that early human occupation of the Andean Altiplano was facilitated by this plentiful tuber (Li et al 2018; Rumold & Aldenderfer, 2016; Spooner et al., 2005).

Molecular and archaeological data indicate that prehistoric populations began intensely consuming and cultivating the potato relatively soon after initial human habitation of the Andean highlands. Archaeological evidence for the oldest confirmed preserved domesticated potato dates to 6950 BP from the Tres Ventanas Cave in the central Peruvian Andes (Pearsall, 2008). Available groundstone and dental data from the Lake Titicaca Basin indicates tubers were processed by hunter-gatherers for consumption between 6700 and 8000 BP (Haas & Llave, 2015). Later sedentary populations based in the Lake Titicaca Basin, such as the Tiwanaku culture (2150 BP–950 BP) or "potato civilization," practiced agronomy methods of raised field agriculture combined with potato freeze-drying (chuño) that led to the potato's dominance as a critical dietary staple of the region (Peñarrieta et al., 2011). Genomic data indicates an even earlier start date of potato domestication around 10,000 BP, with a single domestication origin in the southern Peruvian Andes (Hardigan et al., 2017; Li et al., 2018; Spooner et al., 2005). Today, Andean populations rely heavily upon the potato as a nutritional staple, with the potato making up around 54% of their daily caloric intake (Berti et al., 2010). Andean populations who have relied upon the high-starch potato as a major source of nutrition may have undergone natural selection in genes that encode for starch hydrolysis, leading to more efficient digestion and increased glucose production.

Human starch digestion is facilitated by a suite of salivary, pancreatic, and intestinal enzymes that each have their own role in the starch digestion pathway. In the first stage of starch digestion, salivary α-amylase enzymes encoded by the genes *AMY1A*, *AMY1B*, and *AMY1C* hydrolyze starch polysaccharides into smaller disaccharides such as maltose, sucrose, and other dextrins (Butterworth et al., 2011). Once these starch fragments pass through the esophagus to the stomach, the acidic environment of the stomach halts all action of the salivary α-amylase enzyme. The majority of starch digestion then begins after the chyme from the stomach enters the small intestine. There, α-amylase digestive enzymes, encoded by the genes *AMY2A* and *AMY2B*, are secreted from the pancreas and released into the upper small intestine, or duodenum, to aid with starch digestion. These pancreatic amylase enzymes continue to hydrolyze starch polysaccharides into smaller disaccharides such as maltose. The final suite of maltase-glucoamylase and sucrase-isomaltase enzymes, encoded by *MGAM* and *SI*, respectively, are produced by the brush border membrane of the small intestine. The process of starch digestion is completed when glucosidase enzymes break down these disaccharides to monosaccharides, including glucose (Nichols et al., 2009).

Both salivary α-amylase and pancreatic α-amylase enzymes only produce small amounts of glucose from starches, and the majority of glucose production occurs during the last few steps of digestion within the small intestine (Fernández & Wiley, 2017; Nichols et al., 2009). Since glucose is the most important source of energy in all organisms, this final digestion step in the small intestine is the most critical stage of this process and a more probable target of natural selection related to starch digestion. This is supported by two lines of evidence. First, previous studies found conflicting evidence of selection for other genetic regions related to starch digestion, and second, a strong signal of selection has been detected in association with the intestinal enzyme encoded by *MGAM* for Andeans (Lindo et al., 2018). Higher copy-number variation (CNV) of the *AMY1* gene was previously hypothesized to be related to increased starch in the diets of post-agricultural populations (Perry et al., 2007). However, physiological evidence suggests that salivary α-amylase plays a limited role in glucose production from starch digestion (Nichols et al., 2009). Therefore, an alternative hypothesis is that high CNV of *AMY1* could be related to reproductive factors such as sperm fertility or production of uterine fluid (Esterhuizen et al., 1995; Heitlinger et al., 1983; O'Mahony et al., 2013; Singh, 1995). Later analyses found that archaic hominins possessed multiple copies of the *AMY1* gene, suggesting that increased CNV occurred prior to the post-agricultural upsurge of dietary starch (Inchley et al., 2016; Perry et al., 2015). These findings indicate that multiple physiological processes could be causative for high CNV in *AMY1*, particularly because of the limited role that salivary α-amylase enzymes play in facilitating glucose production from starch digestion. In addition, a previous study of ancient Andean individuals (dated to 1400 BP to 6800 BP) found evidence for natural selection within *MGAM*, but not higher CNV for *AMY1*, lending further strength to the hypothesis that selection favored the final step of starch digestion in the small intestine (Lindo et al., 2018).

Here, we test for evidence of genetic adaptation in all of the main genes related to starch digestion, *AMY1*, *AMY2*, *SI*, and *MGAM*, among modern-day Peruvians. We hypothesize that natural selection in response to high potato consumption shaped variation in these genes to facilitate improved starch digestion. We further hypothesize that the genes associated with the last and most important step of starch digestion and glucose production, *MGAM* and *SI*, show greater evidence of recent positive selection. Our findings provide valuable insight into genetic adaptations to human dietary shifts and help clarify the timeline for when potatoes began to be intensely consumed in the Andes.

MATERIALS AND METHODS

Populations

Publicly available genomes from the 1000 Genomes Project (1KG) were used to identify evidence of recent positive selection among a modern-day Andean descendant population, Peruvians. For selection tests, Peruvians included 85 individuals from Lima, Peru (PEL). Control populations included Han Chinese in Beijing (CHB, $n = 103$), Southern Han Chinese (CHS, $n = 105$), the Mende in Sierra Leone (MSL, $n = 85$), and Gambian Mandinkas (GWD, $n = 113$) (Consortium, 2015). Each control population has a unique history of starch consumption. Rice was domesticated as early as 9000 BP in China (Liu et al., 2007), and in West Africa, the indigenous yam was domesticated around 6950 BP (Andres et al., 2017). Potatoes were introduced to both populations less than 500 years ago, with West Africans among the lowest consumers of potatoes worldwide today (Wijesinha-Bettoni & Mouillé, 2019). Therefore, adaptation to starch digestion may have occurred independently among Han Chinese and West African populations.

Genotype data

Genotype data for the tandem genes *AMY1-AMY2*, as well as *SI* and *MGAM* were extracted from 1KG Phase 3 variant calls for each population using the Ensembl Data Slicer tool (Consortium, 2015). Ancestral alleles were queried from the 1KG Project using BCFtools and recoded using PLINK 2.0 to preserve genome phasing (Chang et al., 2015; Li et al., 2009; Purcell et al., 2007). In total 11,280 SNPs were extracted from the four genes. To place our starch digestion genes in a genome-wide context, we extracted 502,214 autosomal SNPs that were included on the Affymetrix Human Origins Array (Patterson et al., 2012). This dataset was merged with the 11,280 SNPs from our candidate genes to form our working dataset of 513,494 SNPs.

Principal component analysis (PCA), ADMIXTURE, and RFMix

To determine if the 1KG admixed PEL were an appropriate proxy for Andean populations, we performed a principal components analysis (PCA) including 575 Quechua-speaking individuals recruited in Cerro de Pasco, Peru (4200 m) (Brutsaert et al., 2019). Data for these Quechua-speaking individuals were derived from the Affymetrix (Santa Clara, CA) Axiom Biobanking array, but had low SNP density (107 SNPs) for the genes related to starch digestion. Quechua data files were phased using Beagle 5.1, processed with PLINK 2.0, and manipulated with VCFtools and BCFtools (Browning et al., 2018; Browning & Browning, 2007; Chang et al., 2015; Danecek et al., 2011; Li et al., 2009; Purcell et al., 2007). Additional control populations were from the 1KG project and included East Asians (EAS) with (CHB, CHS, JPT), Africans (AFR) with (GWD, MSL, YRI), Europeans (EUR) with (IBS), and Americans (AMR) with (CLM and MXL). The 1 KG Iberian population from Spain (IBS) was used to represent the European component for all ancestral tests given the legacy of Spanish colonization in modern day Peru. We extracted 610,000 genetic markers from the 1KG populations that were included on the Affymetrix Axiom Biobanking Array. QC was performed by removing markers that had greater than 10% missing data. After pruning for LD ($R^2 > 0.2$) and QC, we used 141,412 SNPs for the PCA.

We performed an unsupervised model in ADMIXTURE to determine individual Peruvian (PEL) ancestral proportions ($n = 85$), using 65,940 SNPs after pruning for LD ($R^2 > 0.1$) (Alexander & Lange, 2011). 1 KG individuals were clustered into the following super-populations of European (IBS), East Asian (CHB and JPT), African (YRI), and Indigenous American (PEL) with $K = 4$. Next, we assigned locus-specific ancestry to the entire dataset using RFMix 2, and analyzed chromosomal segments in windows for each of the candidate genes (Maples et al., 2013). The number of generations since contact was set to 20 to correspond to the arrival of Spanish colonizers in Peru approximately 500 years ago. Based on ADMIXTURE results, 19 out of 85 PEL had negligible ancestral proportions from the other super-populations (<1%) and were used as Indigenous American haplotype references in this analysis. We selected 19 random 1KG individuals from IBS, YRI, and CHB as the additional ancestral references.

Selection tests

Four statistical tests were applied to identify signatures of natural selection in our dataset: (1) Locus specific branch length (LSBL) (Shriver et al., 2004), (2) Tajima's *D* (Tajima, 1989), (3) the single-population haplotype-based statistic: Number of Segregating Sites by Length (nSL) (Ferrer-Admetlla et al., 2014), and (4) its extension based on two different populations (XP-nSL) (Szpiech et al., 2021). An unavoidable limitation of available tests for positive selection is that population histories are complex and difficult to model. This can lead to demographic events possibly producing false signatures of positive selection in statistics that compare allele frequencies (Bamshad & Wooding, 2003). To account for this, we tested for positive selection using multiple orthogonal methods that are sensitive to different characteristics of the data: allele frequency distribution (Tajima's *D*), genetic differentiation (LSBL), and haplotype decay (XP-nSL, nSL) methods. To additionally account for underlying demography, we used a genome-wide empirical distribution to test for statistical significance, as natural selection causes locus-specific deviations from neutrality.

For LSBL, we compared Peruvians (PEL) against Han Chinese (CHB/CHS) and West African (MSL/GWD) populations. Pairwise F_{st} values were computed for each SNP using Weir-Cockerham's equation (Shriver et al., 2004; Weir & Cockerham, 1984). A statistical significance threshold of $\alpha = 0.01$ was determined using a genome-wide empirical distribution generated using 513,494 SNPs (Akey et al., 2002). Tajima's *D* was calculated using non-overlapping 10,000 base pair windows across the entire genome-wide dataset in VCFtools (Danecek et al., 2011). *D* values were standardized using the *Z*-score approach, and *p*-values were assigned based on the genome-wide empirical distribution (Tajima, 1989). nSL scores were calculated for all SNPs with a MAF > 0.05. PEL were compared with CHB/CHS for

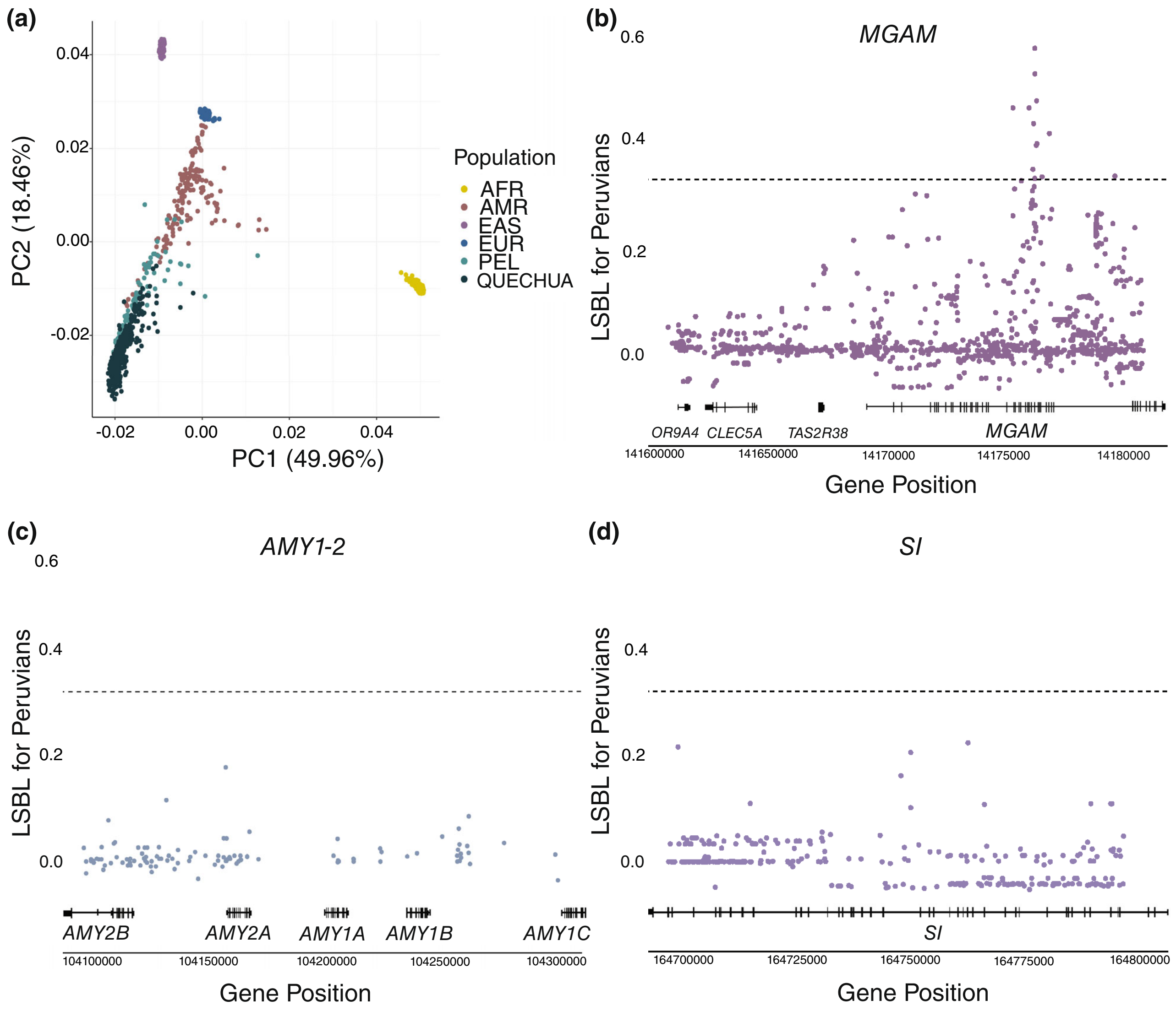

FIGURE 1 (a) Individual genetic structure using PCA with 141,412 SNPs. PC1 explains 49.96% of the variance and PC2 explains 18.46% of the variance. Populations include Quechua-speaking individuals and Peruvians (PEL) from the 1KG project. Control populations are Africans (MSL/GWD/YRI), Europeans (IBS), Americans (CLM/MXL), and East Asians (CHB/CHS/JPT) from the 1KG project. Peruvian LSBL values for *MGAM* (b), *AMY1-2* (c), and *SI* (d). Peruvian branch lengths were calculated using Han Chinese (CHB/CHS) and West Africans (MSL/GWD) as outgroups. Dotted lines indicate the 1% significance level (LSBL $= 0.320$) based on the genome-wide empirical distribution generated for 513,494 SNPs. Chromosome location is depicted along the *x*-axis and LSBL is on the *y*-axis. Genes are indicated including their intron and exon structure for each test statistic.

XP-nSL. Both nSL and XP-nSL were normalized using allele frequency bins in Selscan (Szpiech & Hernandez, 2014; Voight et al., 2006).

To estimate the age of any resulting haplotype(s) we used extended haplotype homozygosity (EHH) scores (Sabeti et al., 2002; Voight et al., 2006). EHH assumes a starlike phylogeny to assess the age of the haplotype by the decay of associated alleles, based on their distance from the core SNP. Core SNPs are determined through their comparatively higher population frequency, and extreme statistic scores from the haplotype test used across a region. In this case, extreme XP-nSL values were used to determine the core SNP, and EHH statistic scores were calculated using Selscan v1.3.0. Given that EHH $\approx P$(Homozygosity), or the probability of homozygosity, the following equation was used: $P(\text{Homozygosity}) = e^{-2rg}$, $r =$ Haplotype length in M (morgans), $g =$ Generation time (Sabeti et al., 2002; Voight et al., 2006). The natural log of EHH decay was plotted with a line of best fit ($-\ln(\text{EHH}) = g \times 2r$) using a linear regression model through the origin. The regression coefficient was used to determine generation time (25 years) with a 95% confidence interval (CI) in *R*.

To provide concordance for resulting estimates of haplotype age, we used a hidden Markov model (HMM) that performs under Markov chain Monte Carlo (MCMC) to estimate the time frame for when a beneficial allele began increasing in frequency (Smith et al., 2018). Under the assumption that the variant with the most extreme XP-nSL value was the target of positive selection, the interval range estimate was generated for time to the most recent common ancestor (T_{MRCA}) with a 95% CI. The MCMCs were run following the mathematical model

TABLE 1 Significant results for *MGAM* in Peruvians

LSBL				
rsID	**LSBL**	**PE = 0.01**	**Gene position**	**Function**
rs2961071[a,b]	0.572	0.0001	141760830	Intron
rs2961070[a,b]	0.523	0.0004	141760842	Intron
rs7793854[a]	0.471	0.001	141761561	Intron
rs2960745	0.458	0.001	141751710	Intron
rs2960757	0.458	0.001	141758405	Intron
rs2961072[a]	0.427	0.002	141758405	Intron
rs4283959[a]	0.408	0.003	141766892	Intron
rs11762139[a,b]	0.389	0.004	141761754	Intron
rs7810984[a,b]	0.386	0.004	141761438	Intron
rs4492290[a]	0.340	0.008	141760008	Intron
rs3087322	0.327	0.009	141794263	Synonymous Variant
rs7797360[a]	0.325	0.009	141763866	Intron
rs6963697[a,b]	0.322	0.010	141760798	Intron

Tajima's *D*			
Position start	**Position end**	**Tajima's *D***	**P_E for 10 kb non-overlapping windows**
141670000	141680000	−2.364	0.001
141680000	141690000	−2.404	0.0006
141720000	141730000	−2.111	0.006
141740000	141750000	−2.071	0.007
141760000	141770000[b]	−2.015	0.009
141770000	141780000[b]	−2.716	4.69E−05
141780000	141790000[b]	−2.215	0.003

[a]Indicates significance with Tajima's *D*.
[b]Indicates significance with XP-nSL.

outlined in Smith et al. (2018), with 5000 iterations and a burn-in of the first 500 iterations using a mutation rate of $1.61 \pm 0.13 \times 10^{-8}$ (Lipson et al., 2015). Han Chinese (CHB) were used as the outgroup for a diverged reference panel in cases where the final allele frequency is close to fixation in PEL, and using a generation time of 25 years.

RESULTS

Peruvians from Lima are an appropriate proxy for Andean populations

Results from the PCA showed that the Peruvian (PEL) individuals exhibited a high degree of overlap with the Quechua, and formed a population cluster distinct from 1KG East Asian, European, and African populations (Figure 1a). As expected, some individuals from this PEL and Quechua population cluster showed slight overlap with 1KG American populations (CLM/MXL). PC1 explained 49.96% of the variance and PC2 explained 18.46%. ADMIXTURE was used to quantify ancestry estimates for PEL individuals originating from the Americas, Europe, East Asia, and Africa ($K = 4$). Nineteen PEL individuals had greater than 99% Indigenous American ancestry, and 66 were of admixed ancestry. Of those admixed individuals, European ancestry constituted the greatest non-Indigenous component and ranged from less than 0.0001% to 51%. The majority of PEL had at least 75% Indigenous American ancestry, with four individuals possessing lower ranges of 33% to 48% Indigenous American ancestry (Figure S1). Given the presence of European, East Asian, and/or African ancestry in 66 PEL individuals, locus-specific ancestry was calculated using RFMix 2. We extracted the chromosomal window segments overlapping with the starch gene regions at *AMY1-AMY2*, *SI*, and *MGAM* to determine if these loci contained sufficient Indigenous American ancestry for subsequent selection scan tests. *SI* contained the greatest amount of Indigenous American ancestry with 82.9%, followed by *AMY1-2* with 80%, and *MGAM* with 78.2%. Collectively, these results indicate that at the loci for these starch candidate genes the admixed Peruvian (PEL) individuals are largely of Indigenous American ancestry and a suitable representative population for Andeans.

MGAM shows evidence of recent positive selection among Peruvians

We tested for statistical evidence of recent positive selection in four starch digestion genes, *AMY1*, *AMY2*, *SI*, and *MGAM*, among Peruvians

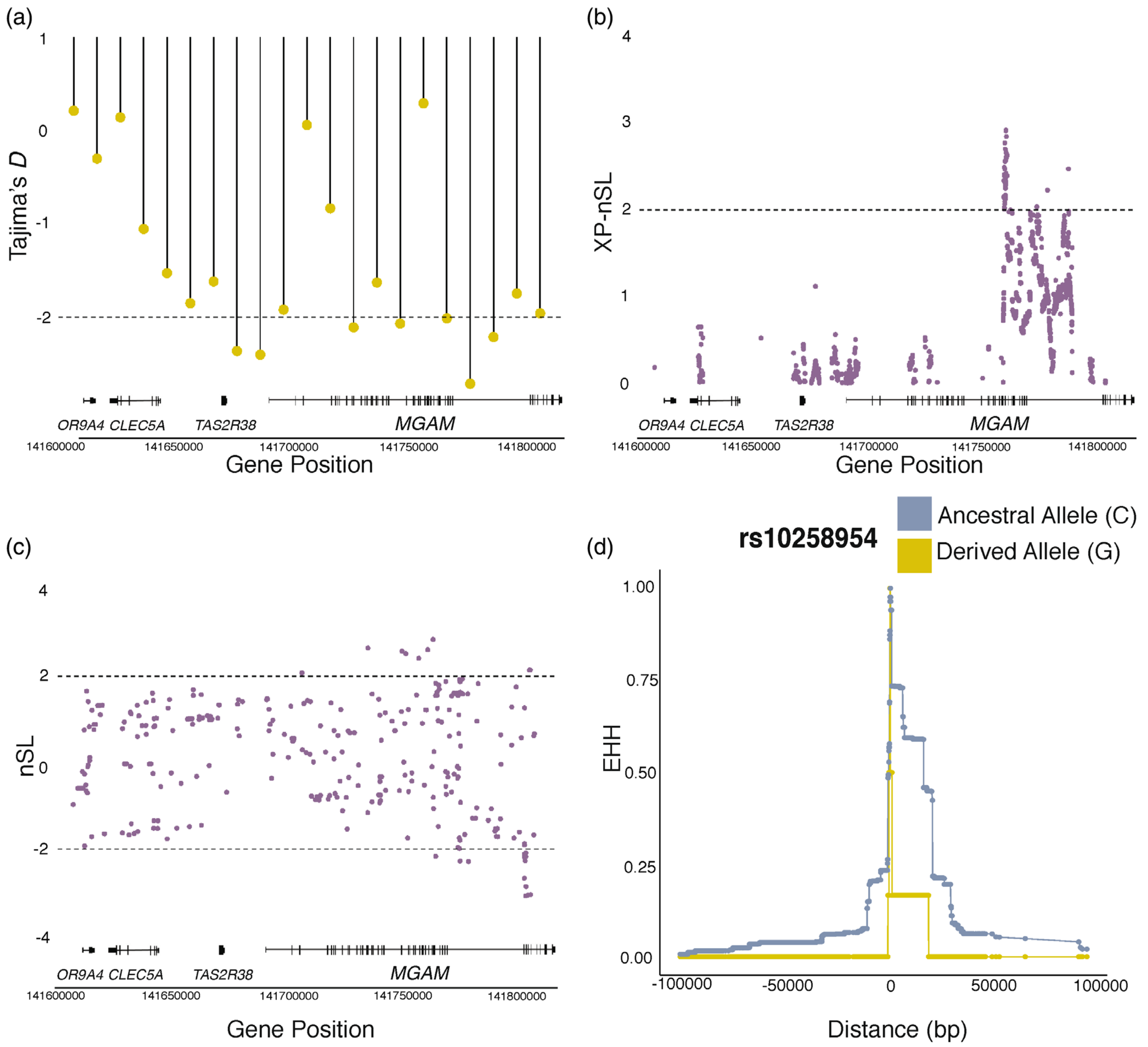

FIGURE 2 (a) Tajima's *D* for *MGAM* in Peruvians calculated using non-overlapping 10,000 base pair windows, with *D* values standardized using the Z-score approach. Dotted lines indicate the 1% significance level (Tajima's $D = -1.83$) based on the genome-wide empirical distribution generated for 513,494 SNPs. Chromosome location is depicted along the *x*-axis and the value of the statistic is shown on the *y*-axis. (b) *MGAM* XP-nSL calculated for Peruvians and Han Chinese. Gene location is depicted on the *x*-axis and XP-nSL value is represented on the *y*-axis. Values above 0 indicate directional selection in the Peruvian population, with values above 2 indicating significant extreme values. (c) Distribution of *MGAM* normalized nSL values. Values above or below the dotted lines (nSL = 2, nSL = −2) indicate extreme values for positive selection based on the genome-wide empirical distribution. Gene location is depicted along the *x*-axis and the test statistic is shown on the *y*-axis. (d) EHH for *MGAM* for Peruvians from the putative core SNP (rs10258954) at varying physical distances. The yellow line represents the distribution of derived EHH values, whereas the blue line represents the distribution of ancestral EHH values. Distance from the core SNP in base pairs is depicted along the *x*-axis and the value of the statistic is shown on the *y*-axis.

(PEL), a modern-day Andean descendant population. We used Han Chinese and West Africans as comparative populations as they have had more recent adoption of potato consumption. For LSBL, 13 significant SNPs were identified within *MGAM* ranging from LSBL = 0.57 to LSBL = 0.32 (Table 1, Figure 1b). The two top SNPs were intronic variants rs2961071 (LSBL = 0.57, $P_E = 0.0001$) and rs2961070 (LSBL = 0.52, $P_E = 0.0004$) that lie within a region associated with the repressor element-1 silencing transcription factor (REST) protein. Seven proximate *MGAM* SNP windows were significant for Tajima's *D* ($\alpha = 0.01$) (Table 1, Figure 2a). Ten of the 13 SNPs that were significant for LSBL also fell within these significant Tajima's *D* windows (Table 1). No significant LSBL values (Figure 1c,d) or Tajima's *D* windows were identified within the *AMY1-2* gene cluster or within *SI*.

We additionally tested for selection using two haplotype-based tests, nSL and XP-nSL, calculated per SNP using a genome-wide distribution (Figures S2 and S3). All of the significant results for nSL and XP-nSL were located within *MGAM*, with nSL values ranging from -3.088 to 2.848, and XP-nSL values ranging from 2.924 to 2.001 (Tables S1 and S2). Negative nSL values signify selection on a derived allele whereas positive values indicate selection on an ancestral allele. XP-nSL values above two indicate a hard or "classic" selective sweep within Peruvians compared to Han Chinese. Both nSL and XP-nSL results were split into significant windows of 10 kb, with three significant windows for nSL in *MGAM* located within: 7:141,750,000–141,760,000, 7:141,760,000–141,770,000, and 7:141,780,000–141,790,000. There were also three significant windows for XP-nSL in *MGAM* located within: 7:141,760,000–141,770,000, 7:141,770,000–141,780,000, and 7:141,780,000–141,790,000. For XP-nSL, 69 out of the 73 significant SNPs clustered within one region (7:141,760,000–141,761,500) (Figure 2b,c). The most extreme XP-nSL value (rs10258954, XP-nSL = 2.92) occurred within this clustered region of SNPs, and was also significant for Tajima's *D* (7:141,760,000–141,761,500) (Table S2). Within this same *MGAM* region there was one significant SNP for nSL (rs11772559, nSL = −2.28). However, the most extreme nSL value (rs73524539, nSL = −3.09) was located within a downstream cluster of 15 significant SNPs (7:141,780,000–141,790,000) (Table S1).

In comparison, significant results for Han Chinese did not occur within overlapping regions of *MGAM*. There were no significant Tajima's *D* windows, and significant results for LSBL, XP-nSL, and nSL all occurred within different regions of *MGAM* (Table S3, Figure S4). Further, only three nSL SNPs fell within the significant *MGAM* region identified in Peruvians (7:141,760,000–141,761,500). For West Africans, significant results for nSL and XP-nSL overlapped with significant Tajima's *D* windows in a downstream region of *MGAM*. This region did not have any significant LSBL results. Significant LSBL, XP-nSL, and nSL values occurred in different regions of *MGAM*, and overlapping signals across the three independent statistical tests did not converge in the same *MGAM* region identified among Peruvians. However, three nSL SNPs for West Africans fell within the significant *MGAM* region identified in Peruvians (7:141,760,000–141,761,500) (Table S4, Figure S4).

There were no significant results identified for Han Chinese in *AMY1-2*. For *SI*, only four significant SNPs using nSL overlapped with significant Tajima's *D* windows, and there were no significant results for either LSBL or XP-nSL (Table S3). For West Africans, significant results did not occur within the same region in *AMY1-2* for LSBL and Tajima's *D*, and there were no significant results for XP-nSL or nSL. For *SI*, there were no significant results for XP-nSL. A significant Tajima's *D* window contained a single significant LSBL SNP and three significant nSL SNPs for *SI* (Table S4). Altogether, the lack of overlapping results across three independent tests (LSBL, Tajima's *D*, and XP-nSL or nSL) indicates that there is not a strong signal of selection in either Han Chinese or West Africans across all starch digestion genes (*AMY1-2*, *SI*, *MGAM*).

Peruvian *MGAM* selection pre-dates archaeological evidence for potato domestication

Strikingly, the top three SNPs in Peruvians with the highest LSBL values (rs2961071, rs2961070, rs7793854), and 63 out of 73 significant SNPs for XP-nSL, were both located within a clustered region in *MGAM* that was highly significant for Tajima's *D* ($D = -2.01$) (7:141,760,000–141,761,500) (Tables 1 and S2). Taken together, the overlapping results of statistical significance using multiple tests indicate a regional haplotype in Peruvians. The age of this regional haplotype in *MGAM* within Peruvians was estimated using the decay of EHH. Calculations were based on the SNP with the most extreme XP-nSL value (rs10258954, XP-nSL = 2.924). We estimated that this haplotype dated to about 398 generations or about 9938 years ago, (95% CI: 9795–10,081 years ago) (Figure 2d). This estimate predates the earliest archaeological evidence of domesticated potatoes by over 2000 years, but is within the time period that molecular analyses estimate for potato domestication. To test for concordance of this estimated haplotype age, we applied a hidden Markov model (HMM) under MCMC to this same SNP with the most extreme XP-nSL value (rs10258954) to generate an interval for when the beneficial allele began increasing in frequency. The ancestral allele is found at high-frequency among Peruvians (C: 97.6%, G: 2.4%) and at intermittent frequency among the outgroup reference Han Chinese (C: 52.4%, G: 42.7%). Our estimate found that the allele rose in frequency about 366 generations or about 9156 years ago, (95% CI: 5393–12,918 years ago) (Figure S5).

DISCUSSION

The transition from hunting and gathering to an agriculturally based subsistence strategy was a global cultural shift that affected the economy as well as the health and diets of human populations. Most post-agricultural societies today have a heavily starch-based diet as a result of this transition, a dramatic distinction from when prehistoric foraging populations primarily relied upon and consumed greater amounts of high-starch foods only in periods of food scarcity (Hardy et al., 2009; Robertson et al., 2018; Yasuoka, 2006). For prehistoric populations living year-round in the Andes, the high-starch potato was a critical resource and continues to be a prominent food source there today. The intense cultivation and consumption of the potato could have led to a genetic adaptation for better starch digestion in Andean populations, with improved nutrition and survival when other resources were scarce. Our results suggest that one gene, *MGAM*, involved in the starch digestion pathway, underwent positive directional selection among Peruvian populations.

Overlapping significant results using Tajima's *D*, LSBL, and the haplotype test, XP-nSL, show a highly significant signal of selection in *MGAM* only identified among Peruvians (7:141,760,000–141,761,500). Only one significant SNP using nSL was identified within this region, as the majority were instead located downstream (7:141,786,200–141,788,800). nSL may have failed to detect a significant haplotype in this region as it is designed to optimize and detect partial sweeps, and it likely removed markers that are close to fixation (Ferrer-Admetlla et al., 2014; Voight et al., 2006). Together, the overlap of XP-nSL and LSBL significant results, and the lack of significant nSL results within this same region, suggests that this *MGAM* haplotype may have experienced a hard selective sweep among Peruvians.

The most extreme SNP result from XP-nSL (rs10258954, XP-nSL = 2.924), represents the most likely target of selection for the Peruvian *MGAM* haplotype core. To estimate the age for this haplotype, EHH was calculated based on the decay around this putative core SNP. We dated the age of this *MGAM* haplotype to be about 398 generations, or 9938 years ago, (95% CI: 9795–10,081 years ago). We further tested this haplotype age by applying an alternative model to generate a timeframe for when this SNP (rs10258954) began increasing in frequency. Results using an HMM model that leveraged an MCMC to generate a posterior distribution for the T_{MRCA} found the age of the *MGAM* haplotype to be about 366 generations or about 9156 years ago, (95% CI: 5393–12,918 years ago). The average estimated age of this haplotype from these two tests of about 9547 years better supports the earlier potato domestication date of about 10,000 BP put forward by molecular analyses (Hardigan et al., 2017; Li et al., 2018; Spooner et al., 2005), than the later domestication date of 6700–8000 BP suggested by available archaeological evidence (Haas & Llave, 2015; Pearsall, 2008). However, a lack of ancient preserved organic material in the archaeological record has led to gaps in crop domestication timelines. For example, molecular studies of the genetic diversity of rice and maize posited earlier domestication dates than those suggested by archaeologists (Liu et al., 2007; Matsuoka et al., 2002). Only with recent methodological developments in archaeology have domestication dates for maize and rice been pushed to earlier time periods that better align with molecular data (Larson et al., 2014). Future studies integrating archaeological and molecular data could refine the timeline for the beginnings of potato domestication in the Peruvian Andes.

When further considering available support for the prominence of potatoes in ancient Peruvian diets, some of the earliest evidence comes from stable carbon and nitrogen isotopic analyses of human bones. Isotopic analysis of individuals from Soro Mik'aya Patjxa in the Lake Titicaca Basin dating to 7000 BP indicated inhabitants at this site consumed a diet predominately consisting of "meat and potatoes" (Chen et al., 2019). Further isotopic analysis of individuals from the nearby Wilamaya Patjxa site dating to 9000 BP indicated a diet of mixed terrestrial C_3 plants, such as potatoes, and camelid animals (Haas et al., 2020). Available groundstone and dental data from sites in the Lake Titicaca Basin indicate intensive exploitation of tubers as a resource from 3550 to 8000 BP (Haas & Llave, 2015; Rumold & Aldenderfer, 2016). While it is likely that potato cultivation efforts were in conjunction with increased consumption, it is not possible to determine whether positive selection in association with starch digestion began in response to consumption of domesticated or wild tubers. However, previous research indicates year-round occupation of the Andean Altiplano began by 12,000 BP (Rademaker et al., 2014), making it likely that wild tuber consumption occurred in tandem with a slow cultivation process prior to the shift to sedentary agricultural domestication. Ethnographic data lends support to this idea, as research shows that within challenging environments hunter-gatherer populations consumer high quantities of starch-based foods when other resources are scarce (Hardy et al., 2009; Yasuoka, 2006). Taken together, selection for increased starch digestion could have occurred in response to attendant potato consumption as a form of optimizing nutrient extraction.

The *MGAM* haplotype (7:141,760,000–141,761,500) in Peruvians lies within a region of high transcriptional activity, specifically Intron 34 (7:141,760,174–141,762,367). Intron 34 is associated with the repressor element-1 silencing transcription factor (REST) protein. As a major influencer of the cellular epigenome, REST is highly dependent on contextual environmental clues. One of its major roles is to undergo context-dependent alternative splicing resulting in multiple protein isoforms being encoded by a single gene (Chen & Miller, 2018). Distinct isoforms of *MGAM* could mediate differential expression patterns, which could lead to enhanced post-transcriptional functional activity and improved gene regulation (Li et al., 2014). In this case, the polymorphisms identified from the selection scan results could have increased or decreased REST ability to suppress transcription of *MGAM*. It is possible that alterations in gene expression or regulation might have been a potential route of genetic adaptation for enhanced potato-starch digestion in Andeans. Exact predictions for the transcription factor function of *MGAM* intron 34 are presently unknown, making this region an ideal candidate for future studies of transcriptional gene regulation related to starch digestion.

It is especially striking that this Peruvian *MGAM* region (7:141,760,000–141,761,500) did not show the same overlapping evidence of recent positive selection across tests among Han Chinese or West Africans. Both are populations with a deep history of starch consumption, including rice and yams, respectively, but very recent introduction of dietary potato starch. This could be explained in two ways. First, the mutation(s) under selection in Peruvians may be specific to potato-starch digestion rather than generalized starch digestion. It is plausible that the maltase-glucoamylase enzymes encoded by *MGAM* are better suited to break down potato-based polysaccharides, whereas other enzymes associated with starch-digestion such as *AMY1*, *AMY2*, or *SI* would be better suited to break down other starch-based foods. There is evidence from microscopy analyses that varying surfaces of starch granules are differentially affected by enzyme activity (Dona et al., 2010). For instance, the hexagonal crystalline structure of potato starch granules are digested more slowly by amylase enzymes than the orthogonal crystalline structure of cereal starch granules (maize, wheat, rice). However, starch digestion is complex and can be affected by many factors including enzyme concentration, starch structure, individual microbiomes, and even cooking. When potatoes are heated through cooking, their crystalline structures are disrupted and more easily digested by amylase enzymes (Butterworth et al., 2011; Dona et al., 2010). Further research into the biochemical properties of starch-based foods and the associated mechanisms of enzymatic hydrolysis is needed to better isolate potential genotype-phenotype relationships of digestion adaptations.

Second, independent genetic adaptations to high starch foods could have evolved in each of the three populations. This would be similar to the pattern of convergent adaptation to high-altitude hypoxia among Tibetans and Andeans (Bigham et al., 2009; Scheinfeldt & Tishkoff, 2010), or lactase persistence among Europeans and Africans (Tishkoff et al., 2007). We did detect evidence of selection in *MGAM* for both Han Chinese and West Africans, but in a region distinct from that detected in Peruvians, lending more support to this second explanation. We did not observe an overlap in selection signals across tests in Han Chinese, but did observe overlap for the haplotype tests in West Africans although not the allele frequency test. However, the lack of overlap in selection signals across three orthogonal tests (LSBL, Tajima's *D*, and XP-nSL or nSL) for West Africans and Han Chinese indicates that evidence for selection in Han Chinese and West Africans for *MGAM* is less compelling than that observed among Peruvians. When considering other instances of strong selective pressure related to diet, a similar case study is found in Inuit and Yu'pik populations. The L479 variant of *CPT1a* experienced a strong selective sweep in only the last 6000 to 23,000 years that was likely related to their high protein consumption leading to chronic ketosis (Hale, 2020). Given the complexity of both starch digestion and evolutionary forces, it is quite possible that both explanations occurred; where selection for *MGAM* could have been specifically in response to potato consumption in Andeans, and West Africans, Han Chinese, and Peruvians all experienced convergent adaptation to starch more generally.

Due to the low SNP density for the genes related to starch digestion in our Quechua-speaking cohort, we chose Peruvians (PEL) from the 1000 Genomes Project as a proxy Andean population given the absence of other sufficient publicly available whole genome sequencing data. Since not all of these Peruvian individuals possessed fully Indigenous Andean ancestry, we demonstrated through our admixture analyses that these Peruvian individuals suffice as an appropriate proxy for Andean populations. Nonetheless, this represents a limitation of our analysis. Similarly, a related population to Peruvians without a long history of starch consumption would be a more ideal outgroup for testing signatures of positive selection. Unfortunately, we do not presently know of or have access to such a genomic database with this type of population. Publicly available genome data for other American populations with sufficient sample sizes all possess comparably long histories of starch consumption, namely maize. While this is a limitation of the present study, it is an active consideration for future research related to testing for positive selection related to starch digestion.

Andeans have been consuming potatoes for thousands of years (Haas & Llave, 2015; Pearsall, 2008; Peñarrieta et al., 2011; Rumold & Aldenderfer, 2016). We hypothesized that Andeans may have developed genetic adaptations to better digest potato starch more efficiently, and that these adaptations may still be present in modern-day Peruvians. A strong signal of selection for a regional haplotype was identified within Peruvians in *MGAM*, and was not under selection in either Han Chinese or West African control populations. This *MGAM* haplotype (7:141,760,000-141,761,500) is located within Intron 34, which could potentially be an important region for gene expression affecting starch digestion. Calculations of EHH and an HMM model that implements an MCMC averaged a haplotype age of about 382 generations, or about 9547 years ago, which best corresponds to early molecular estimates of potato domestication in the southern Peruvian Andes (Hardigan et al., 2017; Li et al., 2018; Spooner et al., 2005). Our research demonstrates that there is dynamic variation in the human ability to better digest high-starch foods, provides a foundational basis for future studies of nutritional biochemistry in the Andes, and contributes to our understanding of independently evolved dietary genetic adaptations among global populations.

AUTHOR CONTRIBUTIONS

Kelsey Jorgensen: Conceptualization (lead); data curation (lead); formal analysis (equal); investigation (lead); methodology (equal); project administration (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Obed Aram Garcia:** Formal analysis (equal); methodology (equal); resources (equal); validation (supporting); visualization (supporting). **Melisa Kiyamu:** Resources (equal). **Tom D. Brutsaert:** Resources (equal). **Abigail W Bigham:** Conceptualization (supporting); data curation (equal); funding acquisition (lead); investigation (supporting); methodology (supporting); project administration (supporting); resources (supporting); supervision (lead); validation (supporting); visualization (supporting); writing – review and editing (supporting).

ACKNOWLEDGMENTS

First and foremost, we would like to thank the study participants of Cerro de Pasco, Peru and all of the individuals that participated in the 1000 Genomes Project. We are also thankful to Dr. Maria Rivera-Chira, Dr. Fabiola Leon-Velarde, Dr. Sudipta Ghosh, and Dr. Gianpietro Elias Revolledo for assistance with participant recruitment in Cerro de Pasco, Peru. Data for the Quechua-speaking individuals used for the PCA was funded by the National Science Foundation grant BCS-1132310 (T.D.B., F.L., and A.W.B.). We would also like to thank Dr. Julie Lesnik, Dr. Tamara Bray, Dr. Chuanzhu Fan, and Dr. Dennis O'Rourke for their early feedback on this project while serving on Dr. Jorgensen's PhD committee.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in GitHub at: https://github.com/KelseyJorgensen/Genetic-adaptations-to-potato-starch-digestion-in-the-Peruvian-Andes.git.

ORCID

Kelsey Jorgensen https://orcid.org/0000-0002-8681-5248
Obed A. Garcia https://orcid.org/0000-0001-6150-6118

REFERENCES

Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, *12*(12), 1805–1814.

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 1–6.

Andres, C., AdeOluwa, O. O., & Bhullar, G. S. (2017). Yam (*Dioscorea* spp.): A rich staple crop neglected by research. In *Encyclopedia of applied plant sciences* (Vol. 2, pp. 435–441). Academic Press.

Bamshad, M., & Wooding, S. P. (2003). Signatures of natural selection in the human genome. *Nature Reviews Genetics*, *4*(2), 99–110.

Berti, P. R., Jones, A. D., Cruz, Y., Larrea, S., Borja, R., & Sherwood, S. (2010). Assessment and characterization of the diet of an isolated population in the Bolivian Andes. *American Journal of Human Biology*, *22*(6), 741–749.

Bigham, A. W., Mao, X., Mei, R., Brutsaert, T., Wilson, M. J., Julian, C. G., Parra, E. J., Akey, J. M., Moore, L. G., & Shriver, M. D. (2009). Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. *Human Genomics*, *4*(2), 79.

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, *103*(3), 338–348.

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, *81*(5), 1084–1097.

Brutsaert, T. D., Kiyamu, M., Elias Revollendo, G., Isherwood, J. L., Lee, F. S., Rivera-Ch, M., Leon-Velarde, F., Ghosh, S., & Bigham, A. W. (2019). Association of EGLN1 gene with high aerobic capacity of Peruvian Quechua at high altitude. *Proceedings of the National Academy of Sciences*, *116*(48), 24006–24011.

Butterworth, P. J., Warren, F. J., & Ellis, P. R. (2011). Human α-amylase and starch digestion: An interesting marriage. *Starch-Stärke*, *63*(7), 395–405.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), s13742-13015-10047-13748.

Chen, G.-L., & Miller, G. M. (2018). Alternative REST splicing underappreciated. *eNeuro*, *5*(5), ENEURO.0034-18.2018.

Chen, J., Haas, R., Eerkens, J., & Hull, B. (2019). Meat and potatoes: A mixed 7,000-year-old-diet. Paper presented at the The 84th Annual Meeting of the Society for American Archaeology, Albuquerque, NM.

Consortium, G. P. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.

Devaux, A., Kromann, P., & Ortiz, O. (2014). Potatoes for sustainable global food security. *Potato Research*, *57*(3-4), 185–199.

Dona, A. C., Pages, G., Gilbert, R. G., & Kuchel, P. W. (2010). Digestion of starch: In vivo and in vitro kinetic models used to characterise oligosaccharide or glucose release. *Carbohydrate Polymers*, *80*(3), 599–617.

Esterhuizen, A., Lourens, J., Lindeque, H., Groenewald, C., Giesteira, M., & Labuschagne, G. (1995). The effect of α-amylase on the acrosomal membrane of human sperm. *Journal of Assisted Reproduction and Genetics*, *12*(4), 283–287.

Fernández, C. I., & Wiley, A. S. (2017). *Rethinking the starch digestion hypothesis for AMY1 copy number variation in humans*. Wiley Online Library.

Ferrer-Admetlla, A., Liang, M., Korneliussen, T., & Nielsen, R. (2014). On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular Biology and Evolution*, *31*(5), 1275–1291.

Haas, R., Watson, J., Buonasera, T., Southon, J., Chen, J. C., Noe, S., Smith, K., Llave, C. V., Eerkens, J., & Parker, G. (2020). Female hunters of the early Americas. *Science Advances*, *6*(45), eabd0310.

Haas, W. R., & Llave, C. V. (2015). Hunter-gatherers on the eve of agriculture: investigations at Soro Mik'aya Patjxa, Lake Titicaca Basin, Peru, 8000–6700 BP. *Antiquity*, *89*(348), 1297–1312.

Hale, N. (2020). Inuit metabolism revisited: what drove the selective sweep of CPT1a L479? *Molecular Genetics and Metabolism*, *129*(4), 255–271.

Hardigan, M. A., Laimbeer, F. P. E., Newton, L., Crisovan, E., Hamilton, J. P., Vaillancourt, B., Wiegert-Rininger, K., Wood, J. C., Douches, D. S., Farré, E. M., Veilleux, R. E., & Buell, C. R. (2017). Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proceedings of the National Academy of Sciences*, *114*(46), E9999–E10008 https://www.pnas.org/content/pnas/114/46/E9999.full.pdf

Hardy, K., Blakeney, T., Copeland, L., Kirkham, J., Wrangham, R., & Collins, M. (2009). Starch granules, dental calculus and new perspectives on ancient diet. *Journal of Archaeological Science*, *36*(2), 248–255.

Heitlinger, L. A., Lee, P. C., Dillon, W. P., & Lebenthal, E. (1983). Mammary amylase: a possible alternate pathway of carbohydrate digestion in infancy. *Pediatric Research*, *17*(1), 15–18.

Inchley, C. E., Larbey, C. D., Shwan, N. A., Pagani, L., Saag, L., Antão, T., Jacobs, G., Hudjashov, G., Metspalu, E., Mitt, M., Eichstaedt, C. A., Malyarchuk, B., Derenko, M., Wee, J., Abdullah, S., Ricaut, F. X., Mormina, M., Mägi, R., Villems, R., Metspalu, M., … Kivisild, T. (2016). Selective sweep on human amylase genes postdates the split with Neanderthals. *Scientific Reports*, *6*(1), 1–10.

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., Barton, L., Climer Vigueira, C., Denham, T., Dobney, K., Doust, A. N., Gepts, P., Gilbert, M. T., Gremillion, K. J., Lucas, L., Lukens, L., Marshall, F. B., Olsen, K. M., Pires, J. C., Richerson, P. J., … Fuller, D. Q. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences*, *111*(17), 6139–6146.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, H.-D., Menon, R., Omenn, G. S., & Guan, Y. (2014). The emerging era of genomic data integration for analyzing splice isoform function. *Trends in Genetics*, *30*(8), 340–347.

Li, Y., Colleoni, C., Zhang, J., Liang, Q., Hu, Y., Ruess, H., Simon, R., Liu, Y., Liu, H., Yu, G., Schmitt, E., Ponitzki, C., Liu, G., Huang, H., Zhan, F., Chen, L., Huang, Y., Spooner, D., & Huang, B. (2018). Genomic analyses yield markers for identifying agronomically important genes in potato. *Molecular Plant*, *11*(3), 473–484.

Lindo, J., Haas, R., Hofman, C., Apata, M., Moraga, M., Verdugo, R. A., Watson, J. T., Viviano Llave, C., Witonsky, D., Beall, C., Warinner, C., Novembre, J., Aldenderfer, M., & Di Rienzo, A. (2018). The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Science Advances*, *4*(11), eaau4921 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6224175/pdf/aau4921.pdf

Lipson, M., Loh, P.-R., Sankararaman, S., Patterson, N., Berger, B., & Reich, D. (2015). Calibrating the human mutation rate via ancestral recombination density in diploid genomes. *PLoS Genetics*, *11*(11), e1005550.

Liu, L., Lee, G.-A., Jiang, L., & Zhang, J. (2007). Evidence for the early beginning (c. 9000 cal. BP) of rice domestication in China: a response. *The Holocene*, *17*(8), 1059–1068.

Luca, F., Perry, G., & Di Rienzo, A. (2010). Evolutionary adaptations to dietary changes. *Annual Review of Nutrition*, *30*, 291.

Maples, B. K., Gravel, S., Kenny, E. E., & Bustamante, C. D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *The American Journal of Human Genetics*, *93*(2), 278–288.

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, J., Buckler, E., & Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences*, *99*(9), 6080–6084.

Nichols, B. L., Quezada-Calvillo, R., Robayo-Torres, C. C., Ao, Z., Hamaker, B. R., Butte, N. F., Marini, J., Jahoor, F., & Sterchi, E. E. (2009). Mucosal maltase-glucoamylase plays a crucial role in starch digestion and prandial glucose homeostasis of mice. *The Journal of Nutrition*, *139*(4), 684–690.

O'Mahony, J., Fox, P., & Kelly, A. (2013). Indigenous enzymes of milk. In *Advanced dairy chemistry* (pp. 337–385). Springer.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Pearsall, D. M. (2008). Plant domestication and the shift to agriculture in the Andes. In *The handbook of South American archaeology* (pp. 105–120). Springer.

Peñarrieta, J., Alvarado, J. A., Bravo, J., & Bergenståhl, B. (2011). Chuño and Tunta; the traditional andean sun-dried potatoes. In *Potatoes: Production, Consumption and Health Benefits*, Nova Science Publishers, 1–12.

Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., Werner, J., Villanea, F. A., Mountain, J. L., Misra, R., Carter, N. P., Lee, C., & Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*, *39*(10), 1256–1260.

Perry, G. H., Kistler, L., Kelaita, M. A., & Sams, A. J. (2015). Insights into hominin phenotypic and dietary evolution from ancient DNA sequence data. *Journal of Human Evolution*, *79*, 55–63.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575.

Rademaker, K., Hodgins, G., Moore, K., Zarrillo, S., Miller, C., Bromley, G. R., Leach, P., Reid, D. A., Álvarez, W. Y., & Sandweiss, D. H. (2014). Paleoindian settlement of the high-altitude Peruvian Andes. *Science*, *346*(6208), 466–469.

Robertson, T. M., Alzaabi, A. Z., Robertson, M. D., & Fielding, B. A. (2018). Starchy carbohydrates in a healthy diet: the role of the humble potato. *Nutrients*, *10*(11), 1764.

Rumold, C. U., & Aldenderfer, M. S. (2016). Late Archaic–Early Formative period microbotanical evidence for potato at Jiskairumoko in the Titicaca Basin of southern Peru. *Proceedings of the National Academy of Sciences*, *113*(48), 13672–13677.

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, *419*(6909), 832–837.

Scheinfeldt, L. B., & Tishkoff, S. A. (2010). Living the high life: High-altitude adaptation. *Genome Biology*, *11*(9), 1–3.

Shriver, M. D., Kennedy, G. C., Parra, E. J., Lawson, H. A., Sonpar, V., Huang, J., Akey, J. M., & Jones, K. W. (2004). The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Human Genomics*, *1*(4), 1–13.

Singh, B., Raigond, P., Dutt, S., & Kumar, M. (2020). Potatoes for food and nutritional security. In *Potato* (pp. 1–12). Springer.

Singh, V. N. (1995). Human uterine amylase in relation to infertility. *Hormone and Metabolic Research*, *27*(1), 35–36.

Smith, J., Coop, G., Stephens, M., & Novembre, J. (2018). Estimating time to the common ancestor for a beneficial allele. *Molecular Biology and Evolution*, *35*(4), 1003–1017.

Spooner, D. M., McLean, K., Ramsay, G., Waugh, R., & Bryan, G. J. (2005). A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proceedings of the National Academy of Sciences*, *102*(41), 14694–14699.

Szpiech, Z. A., & Hernandez, R. D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, *31*(10), 2824–2827.

Szpiech, Z. A., Novak, T. E., Bailey, N. P., & Stevison, L. S. (2021). Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. *Evolution letters*, *5*(4), 408–421. https://doi.org/10.1101/2020.05.19.104380

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B., Ghori, J., Bumpstead, S., Pritchard, J. K., Wray, G. A., & Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, *39*(1), 31–40.

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, *4*(3), e72.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Wijesinha-Bettoni, R., & Mouillé, B. (2019). The contribution of potatoes to global food security, nutrition and healthy diets. *American Journal of Potato Research*, *96*(2), 139–149.

Yasuoka, H. (2006). Long-term foraging expeditions (Molongo) among the Baka hunter-gatherers in the Northwestern Congo Basin, with special reference to the "wild yam question". *Human Ecology*, *34*(2), 275–296.

SUPPORTING INFORMATION

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

How to cite this article: Jorgensen, K., Garcia, O. A., Kiyamu, M., Brutsaert, T. D., & Bigham, A. W. (2023). Genetic adaptations to potato starch digestion in the Peruvian Andes. *American Journal of Biological Anthropology*, *180*(1), 162–172. https://doi.org/10.1002/ajpa.24656