



# Postadmixture Selection on Chileans Targets Haplotype Involved in Pigmentation, Thermogenesis and Immune Defense against Pathogens

Lucas Vicuña<sup>1</sup>, Olga Klimenkova<sup>2</sup>, Tomás Norambuena<sup>1</sup>, Felipe I. Martinez <sup>3</sup>, Mario I. Fernandez <sup>4,5</sup>, Vladimir Shchur<sup>2</sup>, and Susana Eyheramendy<sup>1,6,\*</sup>

<sup>1</sup>Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Peñalolén, Santiago, Chile

<sup>2</sup>National Research University Higher School of Economics, Russian Federation, Moscow, Russia

<sup>3</sup>Center for Intercultural and Indigenous Research, School of Anthropology, Faculty of Social Sciences, Pontificia Universidad Católica de Chile, Santiago, Chile

<sup>4</sup>Department of Urology, Clínica Alemana, Santiago, Chile

<sup>5</sup>Center for Genetics and Genomics, Faculty of Medicine, Clínica Alemana, Universidad del Desarrollo, Santiago, Chile

<sup>6</sup>Instituto Milenio de Investigación sobre los Fundamentos de los Datos (IMFD)

\*Corresponding author: E-mail: susana.eyheramendy@uai.cl.

Accepted: 24 June 2020

## Abstract

Detection of positive selection signatures in populations around the world is helping to uncover recent human evolutionary history as well as the genetic basis of diseases. Most human evolutionary genomic studies have been performed in European, African, and Asian populations. However, populations with Native American ancestry have been largely underrepresented. Here, we used a genome-wide local ancestry enrichment approach complemented with neutral simulations to identify postadmixture adaptations underwent by admixed Chileans through gene flow from Europeans into local Native Americans. The top significant hits ( $P=2.4\times 10^{-7}$ ) are variants in a region on chromosome 12 comprising multiple regulatory elements. This region includes *rs12821256*, which regulates the expression of *KITLG*, a well-known gene involved in lighter hair and skin pigmentation in Europeans as well as in thermogenesis. Another variant from that region is associated with the long noncoding RNA *RP11-13A1.1*, which has been specifically involved in the innate immune response against infectious pathogens. Our results suggest that these genes were relevant for adaptation in Chileans following the Columbian exchange.

**Key words:** adaptation, genetic ancestry, admixture, gene flow, genetic drift.

## Introduction

Among other evolutionary forces such as positive selection and drift, genetic admixture has been shown to play an important role in shaping human diversity (Nielsen et al. 2017). The majority of human groups studied to date experienced ancient (<7,000 years) to recent (<500 years) admixture, whereas a limited number of populations have remained in isolation for longer periods (Pickrell and Reich 2014). Selective pressures exerted long enough over populations as well as genetic drift generate regional differences in allele frequencies (Coop et al. 2010), whereas admixture tends to neutralize these differences through gene flow (Pickrell and Reich

2014). However, standing genetic variation inherited by the admixed individuals can sweep into high frequency over generations if this variation aids them to adapt to endemic or foreign selective pressures. This *post-admixture* selection (hereafter referred to as “PAS”) is similar to adaptive introgression at the intraspecies level (Bhatia et al. 2014; Jeong et al. 2014).

Some studies have shown that PAS underlies adaptation in diverse human populations. For instance, modern high-altitude Tibetans, who resulted from the admixture of populations ancestral to the high-altitude Sherpa and low-altitude Han Chinese, underwent PAS in the *EGLN1* and *EPAS1* genes.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

These genes are important components of the hypoxia-inducible pathway, which is involved in changes in O<sub>2</sub> supply and is thus relevant for breathing at high altitudes (Jeong et al. 2014). In addition, the African-specific *Duffy-null* blood group—which confers resistance against *Plasmodium vivax* infection—underwent PAS in the Makranis of Pakistan after it was introduced by sub-Saharan African slaves into Pakistan (Laso-Jadart et al. 2017). Among admixed Mexicans, PAS acted on the major histocompatibility complex region, as revealed by significant departures of African ancestry in that genomic region (Guan 2014). Similarly, a study used microsatellite data from a set of pulled Latin American populations with Native American, European, and African ancestry to detect PAS. The top signals were located in gene regions with important roles in immune defense, including the HLA region (excess of African ancestry) and the immunoglobulin heavy chain gene complex region (excess of European ancestry). This suggests that the corresponding genes were relevant for adaptation to infectious diseases brought by nonnative immigrants (Deng et al. 2016). PAS acting on Native American genetic variation in lipid-related genes was associated with a higher incidence of obesity and dyslipidemias in admixed Mexicans (Ko et al. 2014). Similarly, a study on admixed Brazilians found PAS acting on a Native American haplotype encompassing the *PPP1R3B* gene, which is involved in glycogen synthesis. Such variants in Mexicans and Brazilians may have been advantageous due to increased fat and glucose storage under a restrictive diet environment (Secolin et al. 2019; Ko et al. 2014). Further, a recent study found that PAS acting on Native American genetic variation may have aided a population with Native American and European ancestry from the Atacama Desert of Northern Chile to adapt to the pathophysiological effects of chronic exposure to high arsenic levels in water (Vicuna et al. 2019).

The Chilean population is a suitable model for analyzing how PAS has acted through gene flow from Europeans into Native Americans. Chileans have similar proportions of European and Native American ancestry (0.52 and 0.45, respectively) and a small African component (0.03) (Eyheramendy et al. 2015). These proportions resulted from an admixture process that began in the 16<sup>th</sup> century with European settlers (mostly Spaniards) who admixed with Native Americans from the lowlands of Central-Southern Chile (the Mapuche) and the Andes highlands of Northern Chile (the Aymara). The admixture process continued with several minor migratory events at later times (mainly from Europe), including a small African component brought in during the 17<sup>th</sup> century (Eyheramendy et al. 2015).

The aim of this study was to evaluate whether Europeans provided adaptive genetic variation to admixed Chileans by acting through PAS following the Columbian exchange. We analyzed this selective force in a population from the Atacama Desert of Northern Chile (Vicuna et al. 2019). We identified significant PAS signatures at a haplotype rich in regulatory

elements. This haplotype included *rs12821256*, an enhancer variant that regulates the expression on *KITLG*. Although *rs12821256* is a causative variant of lighter hair color in northern Europeans (Guenther et al. 2014), *KITLG* regulates the amount of melanin pigments in hair follicles as well as in the skin and is also involved in thermogenesis (Yang et al. 2018). *KITLG* variants, including *rs12821256*, have undergone strong positive selection in Europeans (Grossman et al. 2010; Yang et al. 2018). The selected haplotype also includes a variant associated with the long noncoding RNA (lncRNA) *RP11-13A1.1*, which has been specifically implicated in the immune response against fungal infections (Riege et al. 2017). Our results suggest that these regulatory elements were relevant for adaptation among admixed Chileans following the Columbian Exchange.

## Materials and Methods

### Genomic Data

We used whole-genomes from 11 Huilliche/Pehuenche individuals from Vidal et al. (2019), which belong to the macro-ethnic Mapuche population. Whole-genome data were taken from 108 Yoruba in Ibadan, Nigeria (YRI); 107 Iberian Population in Spain (IBS); 107 Toscani in Italy (TSI); 99 Utah Residents with Northern and Western European Ancestry (CEU); 91 British in England and Scotland (GBR); and 99 Finnish in Finland (FIN) individuals from the 1000 G Project (1000 Genomes Project Consortium et al. 2015). We used SNP array data of 25 Aymara, 24 Quechua, 25 Maya, and 14 Nahua Native Americans from Bigham et al. (2010). We used an SNP array data set from a previous study (Vicuna et al. 2019) consisting of 185 admixed subjects (121 males and 64 females) with mostly European and Native American ancestry from the Atacama Region of Northern Chile. We performed the same quality controls as in Vicuna et al. (2019). Briefly, using Plink 1.9 (Purcell et al. 2007), we excluded genetic variants with minor allele frequency <0.01, SNP calling rate <90%, individual calling rate <90%, and Hardy–Weinberg Equilibrium *P* value <0.00001. We obtained a final set of 895 individuals (including the 185 Chileans) and 772,277 SNPs.

### Local Ancestry Inference

We estimated local ancestry in Chileans at these 772,277 SNPs using LAMP-LD 1.1. We used reference panels for Native American (*n* = 88), European (*n* = 911), and African (*n* = 229) populations, as described in Vicuna et al. (2019). LAMP-LD uses a Hidden Markov Model (HMM) that integrates out all possible phase switches in the admixed genotypes when estimating the total genotype probability. To speed up computation, only phase switches that change ancestry (i.e., breakpoints) are constrained to occur at boundaries of windows of *L* SNPs; these breakpoints are subsequently further resolved in a final step using a simplified HMM that

allows a single ancestry switch for local region around that breakpoint (Baran et al. 2012). We used windows of size  $L = 10$  and an HMM with number of hidden states  $S = 15$ . We used phased samples from the parental populations and unphased data from the Chilean population. Phasing was performed using Beagle (Browning and Browning 2007). As readout, LAMP-LD produces individual chromosome files where each SNP allele is coded as 0, 1, or 2, depending on whether the haplotype block harboring the SNP has a higher probability of being of European, Native American, or African ancestry, respectively. After filtering only autosomal SNPs with known rs IDs, we obtained a final data set of 633,744 SNPs.

### Global Ancestry Inference

We selected autosomal SNPs that intersect between all the population data sets and then merged the 12 populations using Plink. We further excluded SNPs with SNP calling rate  $< 95\%$  to the merged file, obtaining a final set of 420,688 SNPs and 895 individuals. We used the *Ohana* software suite (Cheng et al. 2017) to infer the individual ancestry components of our Chilean population. For this, the clean data set was converted to .dgm format with *Ohana* using `convert ped2dgm`. Then, the converted data set was down-sampled to 5% of the number of SNPs (sampled randomly) using the `sample-sites.py` script provided in the *Ohana* GitHub repository <https://github.com/jade-cheng/ohana> (Cheng et al. 2017). In this way, we obtained a high-quality subset of  $\sim 100,000$  unlinked markers to use for the structure analysis, as recommended in that repository. Then, `qpas -e 0.0001 -k i -qo` was used to create Q matrices with the likelihood of the ancestral component proportions, where  $i$  represents the number of ancestry components and ranges from 3 to 10. Admixture plots were plotted using `pong` (Behr et al. 2016). ADMIXTURE was used to compute the cross-validation error among different  $K$ s (Alexander et al. 2009). For each value of  $K$ , ten replicates were performed, and the  $K$  with the lowest mean cross-validation error was assumed to have the best predictive accuracy.

### Principal Component Analysis

Principal component (PC) analysis on the clean data set of 420,688 SNPs and 796 individuals (FIN samples were excluded from the original data set) was carried out with Plink, using the following filters to obtain an LD-pruned subset of variants. We only included variants with minor allele frequency  $> 0.01$ , SNP calling rate  $> 99\%$ , and we pruned SNPs using an independent pairwise approach with window size of 50 kb, a step size of five SNPs, and a  $r^2$  cutoff threshold of 0.15. Eigenvectors were produced with Plink. PC plots were generated with R. The PCs from figure 1B and C were obtained using the same filters, but excluding all Native American and European samples, respectively.

### Deviations from Mean Genome-Wide Local Ancestry

Following local ancestry estimations with LAMP-LD (see Local Ancestry Inference), we calculated the mean local ancestry for every SNP using an in-house R script (supplementary file 1, Supplementary Material online). We used one-tailed  $t$ -tests to identify SNPs with deviations in the mean European ancestry. At each SNP, we compared its European ancestry proportion with the genome-wide ancestry mean  $p_0 = 0.52$ . We performed a statistical hypothesis test  $H_0: p_i = p_0$  versus  $H_1: p_i > p_0$  at each SNP, where  $p_i$  represents the European ancestry at the  $i$ th SNP. We assumed that  $X_{ij}$  is a random variable following a Binomial distribution with parameters  $n = 2$  and  $p_i$ , where  $X_{ij}$  takes values in  $\{0, 1, 2\}$  representing the number of European alleles at SNP  $i$  and individual  $j$ .  $p_i$  is the proportion of European ancestry at SNP  $i$ . The asymptotic results of the maximum likelihood estimator for  $p_i$  were used to assume a normal distribution on this estimator and to design our hypothesis test. Variants reaching a significance threshold of  $P < 10^{-5}$  were considered to be under PAS (Bhatia et al. 2014).

### Estimation of the Effective Population Size of Chileans

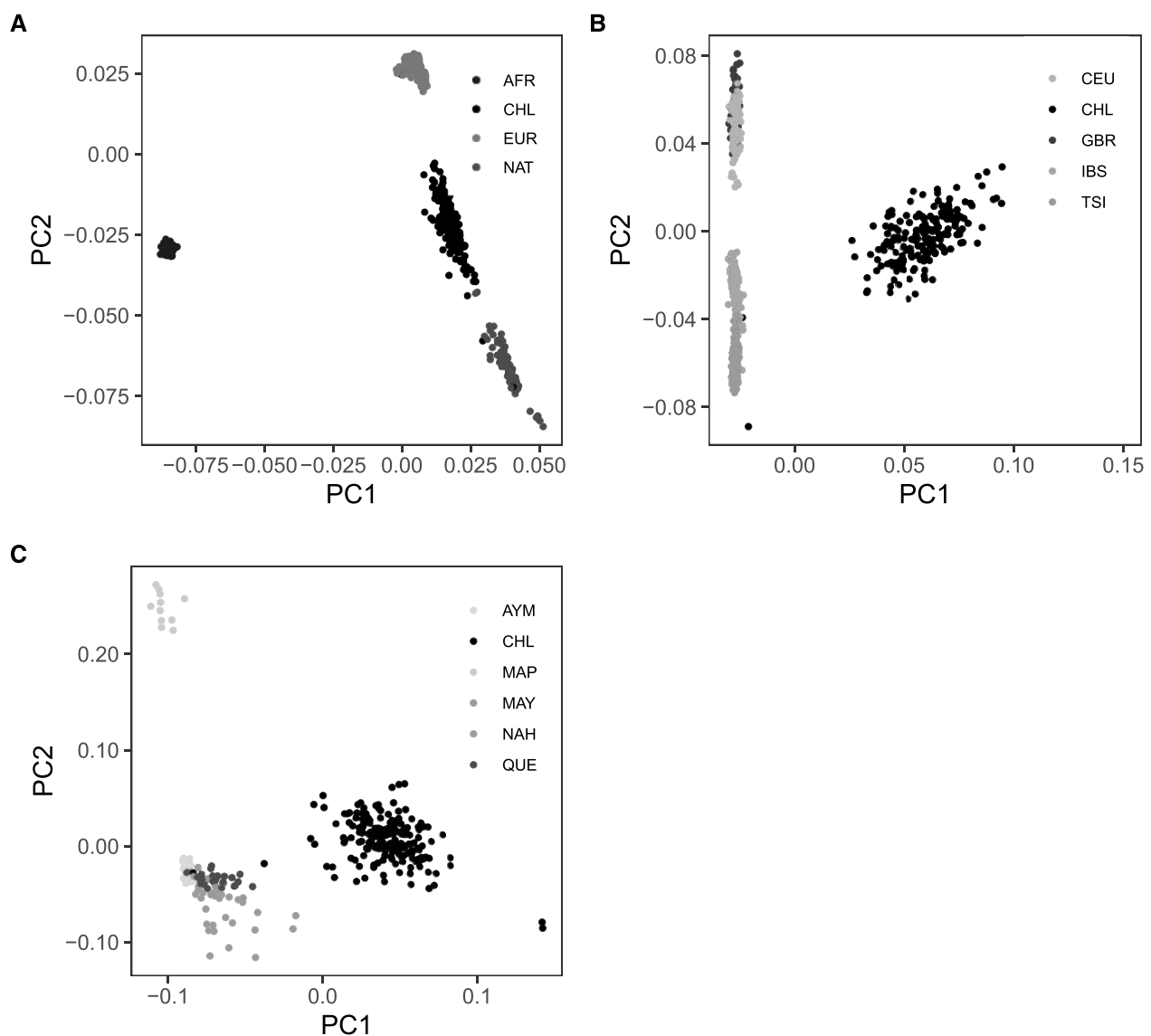
We estimated the effective population size ( $N_e$ ) of our Chilean sample as follows. The initial ancestry proportion  $p_0$  of European component changes over time due to genetic drift. At the sampling time, the European proportion  $p_1$  is a random variable with the variance approximately equal to (Tataru et al. 2017):

$$\text{var} = p_0(1 - p_0)\left(1 - e^{-\frac{T}{N_e}}\right).$$

The expectation of  $p_1$  equals the admixture proportion  $p_0$ , and the distribution of  $p_1$  can be approximated by a Beta distribution with the mean and variance equal to these values. Next, we took into account that the European ancestry observed in our empirical data depends on the sampling. At each SNP, the observed proportion  $k$  of European ancestry was sampled from the binomial distribution with sample size  $K$  (in our case  $K = 370$ ) and probability  $p_1$ . So, the observed European ancestry is distributed as follows:

$$k \sim \int_0^1 P(\text{Binom}(p_1, K) = k) d(p_1).$$

We adjusted  $N_e$  so that the variance of  $k$  matches the empirical variance of European ancestry across all the SNPs in our Chilean sample of size  $N = 370$  haplotypes. For different times since admixture, we got the following values for the effective population sizes:  $N_e = 4,500$  at  $T = 10$ ,  $N_e = 6,000$  at  $T = 12$ , and  $N_e = 7,000$  at  $T = 15$ .



**FIG. 1.**—Principal component analysis. Principal component clustering of admixed Chileans together with populations from Europe, the United States, and Africa. (A) PC1 versus PC2 of Chilean (CHL), European (EUR), Native American (NAT), and African (AFR) populations. Populations included are: CEU, GBR, IBS, TSI [Europe]; Aymara (AYM), Mapuche (MAP), Maya (MAY), Nahua (NAH), Quecha (QUE) [the United States]; and YRI [Africa]. (B) PC1 versus PC2 of Chileans and CEU, GBR, IBS, and TSI. (C) PC1 versus PC2 of Chileans and the five Native American populations.

### Simulations of Local Ancestry

We performed forward-time local ancestry simulations using SELAM (Corbett-Detig and Jones 2016). We assumed a scenario with a single pulse of admixture of three populations  $T$  generations ago (setting  $T$  to three different values of 10, 12, and 15 generations). The admixture proportions were set to 0.521 (European), 0.442 (Native American), and 0.037 (African), as those estimated by LAMP-LD in our Chilean sample. We assumed the  $N_e$  of the admixed population to be constant. We explored three different diploid (haploid) sample sizes:  $N = 93$  (186), 185 (370), and 370 (740). For the case

of 185 individuals, we simulated sexual populations (64 females and 121 males), whereas for the other sample sizes, we simulated hermaphroditic populations. We simulated 189 chromosomes each of length = 20 Morgans, so that the total length of the simulated genomes is equal to the length of our data. Then we mapped the physical positions of the SNPs in our data on the simulated chromosomes, which gives the same correlation structure in the SNP ancestry as in the real data. We calculated  $t$ -test  $P$  values for the SNP with maximal European ancestry deviations for different combinations of  $N$ ,

$N_e$ , and  $T$  (Table 1), similarly as indicated in supplementary file 1, Supplementary Material online.

### Variant and Gene Annotations

All variant annotations used in this study corresponded to the GRCh37 (hg19) assembly. Variant annotations including the Sequence Ontology (SO) consequence type, associated alleles, Gencode biotypes, and combined annotation dependent depletion (CADD) scores were retrieved using the web tool Variant Effect Predictor (VEP) from Ensembl (McLaren et al. 2016). Downstream variants were defined as those located within 10 kb downstream of a particular gene. Expression quantitative trait loci (eQTL) associations were retrieved with the HaploReg web tool (Ward and Kellis 2012). We used the GWAS Catalog (Welter et al. 2014) to identify phenotypes that have been associated through GWAS with our selected variants and genes.

## Results

### Identification of the Main Ancestral Components of Admixed Chileans

We first estimated genetic affinities between the admixed Chilean population from the Atacama Desert and their proxy parental populations from Europe, the United States, and Africa by PCs clustering (fig. 1A). From Europe, we included Iberians (IBS), Italians (TSI), Utah residents with European ancestry (CEU), and British (GBR) populations (1000 Genomes Project Consortium et al. 2015). From Native America, we used genotype data from Mapuche (Vidal et al. 2019), Aymara, Quechua, Nahua, and Maya (Bigham et al. 2010). From Africa, we used the Yoruba (YRI) population (1000 Genomes Project Consortium et al. 2015) as a proxy population for the small (~3%) African component of Chileans (Eyheramendy et al. 2015). The final merged data set consists in 796 individuals. (See Materials and Methods for the details of the populations included in the analysis.) As expected, Chileans are scattered between the European and Native American samples, with Africans farther away (fig. 1A). Because sample size differences may distort the PCA (McVean 2009), we analyzed an equal number of random Chilean, Native American, European (IBS), and African (YRI) individuals ( $n = 99 \times 4 = 396$ ; supplementary fig. 1, Supplementary Material online). We find that the PCA results obtained using unequal and equal sample sizes are consistent between each other (fig. 1A and supplementary fig. 1, Supplementary Material online).

To test for affinities between Chileans and Europeans, we performed a second PCA considering only Chileans and the aforementioned European populations (total sample size = 476). We find that IBS is the most closely related European populations to Chileans, as it can be seen along the PC2 axis

in figure 1B. This is in agreement with previous studies (Eyheramendy et al. 2015).

To evaluate genetic relationships between Chileans and Native Americans, we performed a third PCA comparing Chileans with the aforementioned Native American populations (total sample size = 284). We find that the Aymara and Quechua cluster closest to Chileans, as revealed by the PC2 axis of figure 1C. This finding is in line with the historical geographic range of the Aymara and Quechua in The Andes of Northern Chile, Southern Perú, and Southeastern Bolivia (Browman 1984). The Mapuche are the major Native American population of Chile. Unexpectedly, we found that besides clustering farther away from Chileans than the Aymara and Quechua along the second PC, the Mapuche cluster farther than Maya and Nahua Native Mesoamerican populations as well. This may be a distortion caused by a very low genetic variation of the Mapuche sample that resulted from at least two strong population bottlenecks experienced by them in the last centuries (see Discussion).

### Global Ancestry Estimation of Admixed Chileans

We estimated global ancestry proportions of our Chilean sample. For this, we used the program *qpas* from the *Ohana* suite software (Cheng et al. 2017), which uses an unsupervised algorithm to estimate global ancestry components based on the allele frequencies of proxy populations for individual ancestry components. Using  $K=3$ , *qpas* clearly separates Chileans into European, Native American, and African ancestral components (fig. 2).  $K=4$  clusters the European populations into two main ancestral components; one more enriched in Southern European populations (IBS and TSI; fig. 2, in blue), and another more enriched in Northern European populations, particularly in FIN (fig. 2, in purple). Chileans have mostly the Southern European component, in line with our PCA results. The main Native American components of Chileans are Mapuche and Aymara, with an average Chilean having 40% Mapuche and 8% Aymara ancestry, respectively (Lorenzo Bermejo et al. 2017). To get a better estimate of these proportions in our cohort from Northern Chile, we run *qpas* using  $K=5$ , which showed the best predictive accuracy across  $K=3-12$ , given its lowest cross-validation error (CV error = 0.457) inferred with ADMIXTURE (Alexander et al. 2009). We could identify Aymara and Mapuche as the main Native American subancestry components of admixed Chileans from Atacama, as revealed when contrasted with the reference native Mapuche and Aymara groups (fig. 2, in orange and green, respectively). *qpas* ( $K=5$ ) estimates mean ancestry proportions of 20.7% Mapuche and 25.9% Aymara, 43.3% Southern European, 6.5% Northern European, and 3.5% African among Chileans. When summing up the contribution of each of these subancestry components, we obtain global European, Native American, and African proportions of 49.8%, 46.6%, and 3.5%, respectively. These estimates are

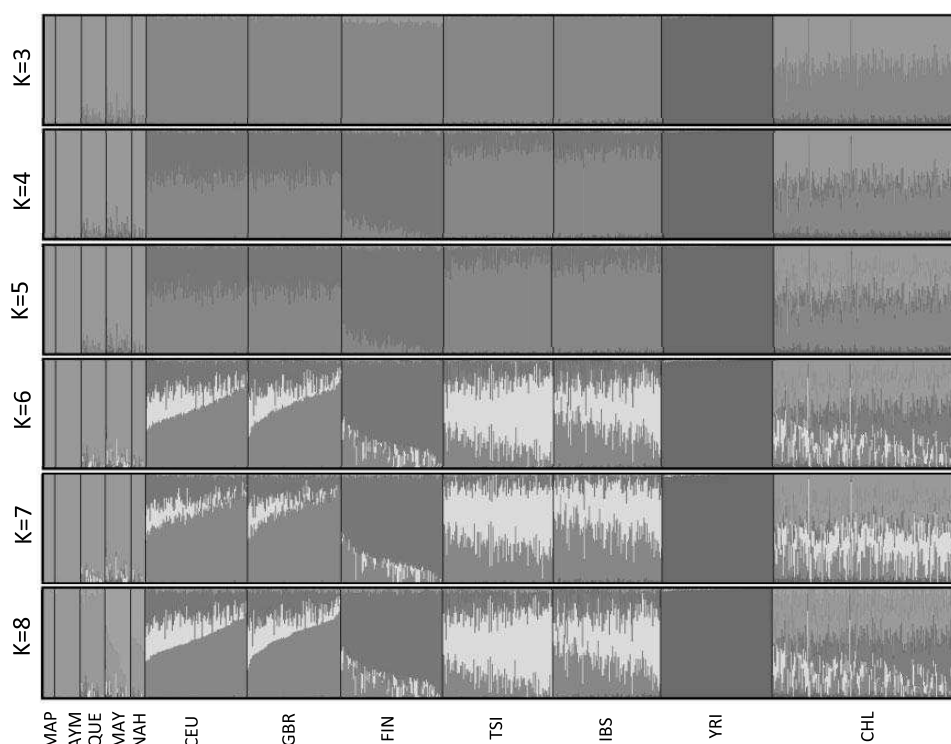


Fig. 2.—Global ancestry proportions of Chileans, European, Native American, and African populations. Cluster analysis generated by *qpas* for Chileans as well as a set of Native American, European, African, and Chilean populations. The number of displayed clusters is  $K = 3$ –8.

slightly different to the genome-wide local ancestry means estimated by LAMP-LD (52.1%, 44.2%, and 3.7%, respectively), probably reflecting differences between the underlying algorithms. Of note, most populations included in this study have slight traces of other ancestry types. For instance, IBS and TSI have 1.5–1.7% African proportions, whereas FIN shows 6.3% of an ancestral component related with Native American populations (fig. 2). This latter observation possibly reflects ancient Siberian gene flow into the Finnish population (Lamnidis et al. 2018) and Native Americans (Nielsen et al. 2017).  $K=6$ –7 identifies separate clusters within the Southern and Northern European components (fig. 2, in yellow and brown, respectively).  $K=8$  differentiates between Mesoamerican (Maya-Nahua) and Andean (Aymara-Quechua) subancestry components (fig. 2, in pink, green, and orange, respectively).

#### Postadmixture Adaptation Mediated by Gene Flow in Chileans

We hypothesized that admixed Chileans underwent PAS acting on European standing genetic variation (PAS acting on Native American variation in this population was analysed in Vicuña et al. (2019)). PAS can be identified by detecting SNPs with SDs in the mean local ancestry from the genome-wide mean (Bhatia et al. 2014; Jeong et al. 2014). At each SNP, the

mean European ancestry was compared with the genome-wide ancestry mean 0.52. Using  $t$ -tests, we evaluated the hypothesis  $H_0: \mu_{\text{EUR}, i} = 0.52$  versus  $H_1: \mu_{\text{EUR}, i} > 0.52$ , at each variant  $i$ . We found 85 SNPs reaching the statistical significance threshold of  $P < 10^{-5}$  recommended for recently admixed populations (Bhatia et al. 2014). Figure 3A shows the mean European ancestry at each SNP along autosomal chromosomes, highlighting selected genes. Figure 3B shows the corresponding  $t$ -test  $P$  values. Supplementary table 1, Supplementary Material online, contains a list of variants showing an excess of European ancestry as well as the corresponding association  $P$  values. The 85 SNPs map a peak of European ancestry in chromosome 12 associated with several regulatory regions, including two lncRNAs (*RP11-13A1.1* and *RP11-13A1.3*) and one processed pseudogene (*RP11-13A1.2*; supplementary table 1, Supplementary Material online).

To evaluate whether these extreme deviations in local ancestry are due to genetic drift, we performed simulations of local ancestry under a model of neutral evolution using the SELAM software (Corbett-Detig and Jones 2016). We modeled a single-pulse of admixture in a Chilean population with global European, Native American, and African ancestry proportions of 0.521, 0.441, and 0.037, respectively, according to our LAMP-LD estimates. Admixture took place  $T = 10, 12$ , or 15 generations ago. These values of  $T$  were chosen because admixture among Chileans began in the 16<sup>th</sup> century

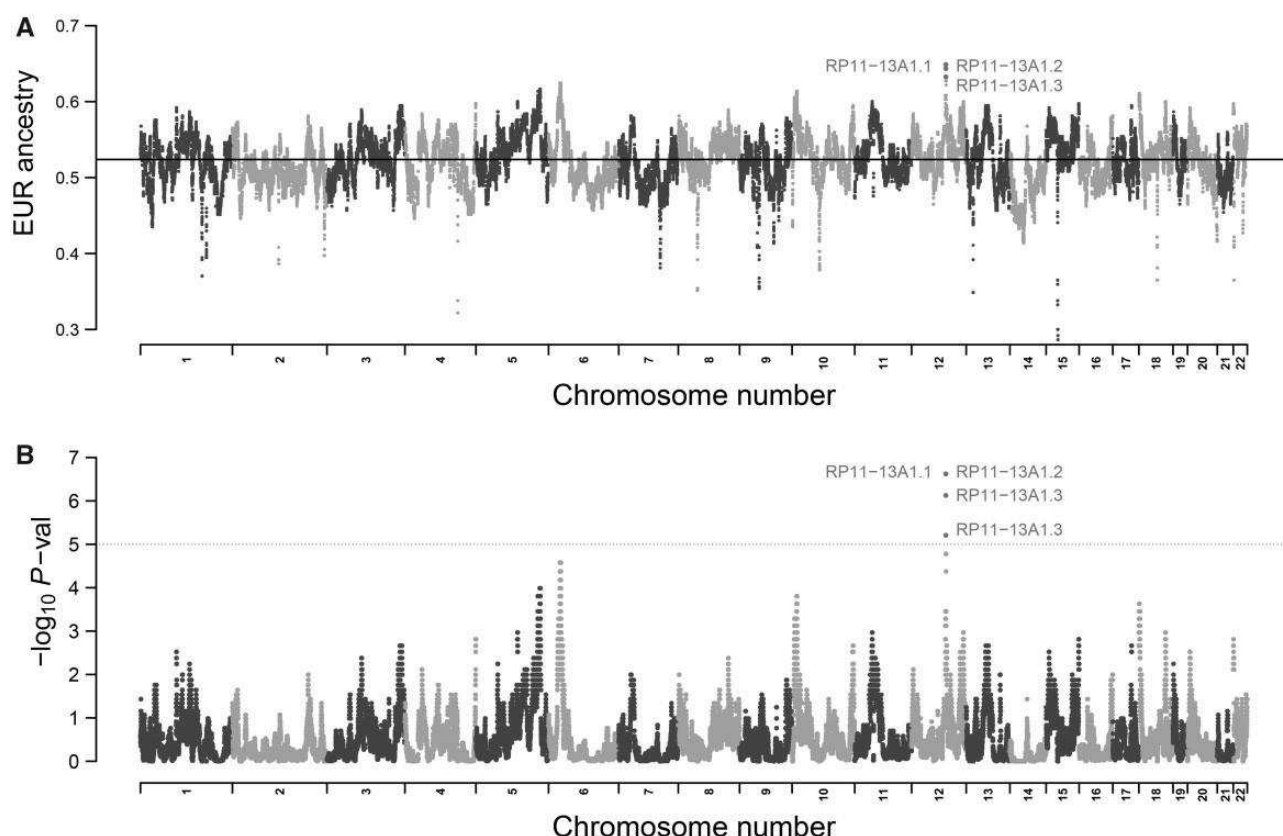
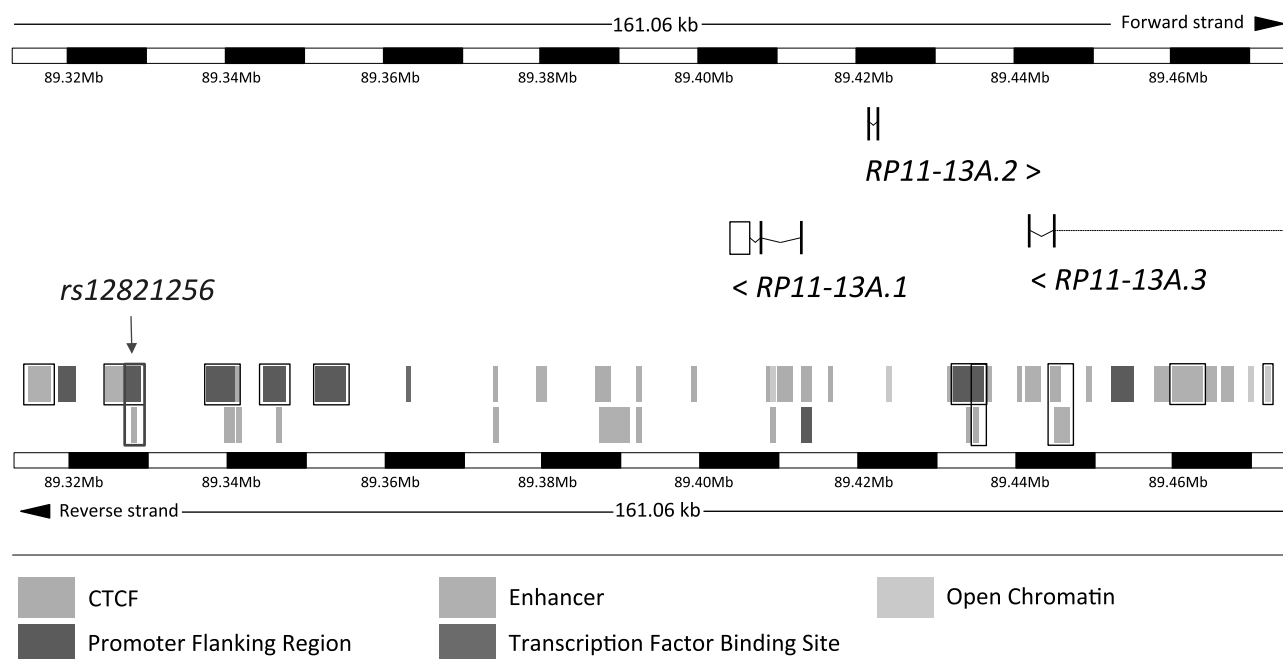


FIG. 3.—Genome-wide deviations in the local mean European ancestry among Chileans. (A) Deviations in the mean local European ancestry over the genome-wide mean along autosomes. The black horizontal line represents the genome-wide local European ancestry mean (0.521) (B).  $t$ -Test  $P$  values of the deviations expressed as  $-\log_{10} P$  value. The dotted black line represents the cutoff threshold of  $P < 10^{-5}$  for SDs (Bhatia et al. 2014). Highlighted are genes associated with variants significantly enriched in European ancestry.

(about 14–15 generations ago) (Encina 1993), but the average time of admixture among Chileans was previously estimated to be 10 generations ago (Eyheramendy et al. 2015). We explored samples of size  $N=93$ , 185, or 370 diploid individuals. For the case of 185 individuals (the size of the real sample), we simulated sexual populations (64 females and 121 males; matching the sexes of the real individuals), whereas for the other sample sizes, we simulated only hermaphroditic populations. We numerically estimated three effective population sizes ( $N_e$ ) based on the values of  $T$ , obtaining  $N_e = 4,500$  at  $T=10$ ,  $N_e = 6,000$  at  $T=12$ , and  $N_e = 7,000$  at  $T=15$  (see Materials and Methods). We simulated scenarios of neutral evolution under different combinations of  $N$ ,  $T$ , and  $N_e$ . For each scenario, we calculated  $t$ -test  $P$  values for the SNP with maximal European ancestry deviations (Table 1). No  $P$  value achieved the significant threshold of  $P < 10^{-5}$  (Bhatia et al. 2014). Hence, our results indicate that the extreme deviations in European local ancestry observed in our Chilean population are not due to genetic drift and are consistent with an effect of PAS.

### Genomic Context and Functional Annotations of Selected Variants

We queried functional annotations of the 85 selected variants using the VEP web tool from Ensembl (McLaren et al. 2016). The strongest PAS hits ( $P = 2.4 \times 10^{-7}$ ) were 65 variants associated with the lncRNA *RP11-13A1.1*, the processed pseudogene *RP11-13A1.2*, the lncRNA *RP11-13A1.3* as well as several regulatory regions, including enhancers, promoter-flanking regions (PFR), CTCF-binding sites (CTCF-BS)—regions that binds CTCF, the insulator protein that demarcates open and closed chromatin (McLaren et al. 2016)—and an open chromatin region (supplementary table 1, Supplementary Material online). Interestingly, *RP11-13A1.1* has been implicated in host immune defense against infectious fungal pathogens (Riege et al. 2017) (see Discussion). One variant, *rs12821256*, maps a transcription factor binding site (TFBS) that overlaps a PFR as well as a CTCF-BS (supplementary table 1, Supplementary Material online). Figure 4 shows the detailed genomic context of *rs12821256* and *RP11-13A1.1*, highlighting associated regulatory elements undergoing



**FIG. 4.**—Genomic context of region in chromosome 12 affected by PAS. The image includes chromosome bands intercalated between black and white blocks with 10 kb of size, as well as several—sometimes overlapping—regulatory elements. The blue box encloses the regulatory region embracing a TFBS, a PFR, and a CTCF-BS that harbors *rs12821256*. Black boxes enclose enhancers, promoter-flanking regions, CTCF-BSs, and open chromatin regions associated by SNPs affected by PAS. Shown are the structures of *RP11-13A1.1*, *RP11-13A1.2*, and *RP11-13A1.3*, as well as the orientation of their coding sequences on the reference genome (“>” and “<” represent forward [5′–3′] and reverse [3′–5′] strands, respectively). Due to space limitations, we show the subregion with physical coordinates 12:89313336–89474394 instead of the whole region with coordinates 12:89193164–89556437. The image was adapted from an image retrieved from Ensembl.

PAS. Supplementary figure 1, Supplementary Material online, shows an expanded image of the whole chromosomal region.

To evaluate the deleteriousness potential of selected variants, we retrieved CADD scores from VEP. CADD scores integrate multiple annotations into a single metrics. Variants with CADD >15 are within the top 5% variants with the highest deleteriousness potential across the genome (Rentzsch et al. 2019). We found that some variants located in PFRs, CTCF-BSs, a TFBS, and one inter-PFR intergenic region have unusually high CADD scores (14.7–18.9), including the A, C, and G alleles of *rs12821256* (supplementary table 1, Supplementary Material online).

We queried the HaploReg tool (Ward and Kellis 2012) for eQTL associations of the regulatory variants with high CADD scores as well as for other epigenetic marks. Interestingly, *rs12821256* is an eQTL associated with the expression of *SUGT1* ( $P = 2.6 \times 10^{-6}$ ) and *SLC38A2* ( $P = 2 \times 10^{-6}$ ) in blood. In addition, *rs12821256* has promoter histone marks in skin cells (epidermal kerinocyte primary cells) and enhancer histone marks in 14 diverse tissues. We next queried the GWAS Catalog (Welter et al. 2014) to identify associations between selected genes and phenotypes. We found that significant GWAS associations have been reported between *rs12821256* and hair color (*rs12821256-T*;  $P = 1 \times 10^{-100}$ ) (Hysi et al. 2018) as well as with light versus dark hair color

(*rs12821256-C*;  $P = 2 \times 10^{-308}$ ) (Morgan et al. 2018). We did not find known associations for *RP11-13A1.2* or *RP11-13A1.3*. Moreover, variants in *SUGT1* have been GWAS-associated with asthma ( $P = 7 \times 10^{-12}$ ) and respiratory system disease ( $P = 2 \times 10^{-14}$ ). Further, we found associations between *SLC38A2* variants and hair color ( $P = 2 \times 10^{-11}$ ) as well as lean body mass ( $P = 1 \times 10^{-15}$ ). Remarkably, *rs12821256* has been reported under strong selection in Europeans and East Asians (see Discussion).

## Discussion

The emergence of vast human population genetic data and the development of advanced analytical methods have enabled the detection of genetic variants involved in adaptation and complex diseases across diverse worldwide populations (Vitti et al. 2013; Lowe and Reddy 2015; Nielsen et al. 2017). Most of these studies have focused on African, European, and Asian populations. However, populations with Native American ancestry have been largely underrepresented (Cheng et al. 2019). This is due in part to: 1) a lack of publicly available genomes from several Native American populations and their admixed Latino descendants; 2) the technical difficulty in detecting ancestry-specific genetic factors in admixed populations with different continental ancestries and complex



demographic histories, such as Latinos. In contrast to Peruvians, Colombians, Mexicans, and Puerto Ricans, for which there are publicly available whole-genome sequencing data (1000 Genomes Project Consortium et al. 2015), this is not the case for Chileans.

Chileans trace their genetic ancestry to Spaniards who admixed with Mapuche and Aymara Native Americans during the Columbian Exchange. On an average, Chileans have European, Native American, and African proportions of 0.52, 0.45, and 0.03, respectively (Eyheramendy et al. 2015). In addition, on an average, Chileans have 0.32 Mapuche and 0.11 Andean proportions (Andean proportion considered as the sum of the related Aymara, Quechua, Colla, and Southern Peru Andean proportions) (Chacon-Duque et al. 2018). However, ancestry and subancestry proportions vary considerably along Chile's north–south geographic axis as well as among socio-economic groups. Although in the north Chileans have higher proportions of Aymara-related ancestry, in the south the main Native American subancestry is the Mapuche (Verdugo et al. 2020). In the present study, we estimated genetic relationships between admixed Chileans and several European and Native American populations. We analyzed a population from the Atacama Desert of Northern Chile (Vicuna et al. 2019). We found that this population has higher Aymara ancestry and lower Mapuche ancestry (25.9% vs. 20.7%, respectively;  $K = 5$ ) than other more geographically heterogeneous Chilean samples with predominant Mapuche ancestry (Chacon-Duque et al. 2018). Therefore, our estimates from the admixed Atacama population are in agreement with a scenario of higher gene flow between Europeans and Andean populations in Northern Chile but lower admixture with the Mapuche from Central-Southern Chile. Of note, our PCs and global ancestry results show that the native Mapuche form a differentiated cluster when compared with the other Native American populations ( $K = 5–8$ ). This is probably due to at least two effects resulting in a very low genetic diversity: 1) The Mapuche suffered a ~95% decrease in their effective population size following the European colonization, based on an analysis that used imputed SNP array data from Mapuche genomes (Lindo et al. 2018); 2) The Mapuche individuals used in the present study were sampled in the small and relatively isolated Huapi Island and are expected to have extremely high-genetic relatedness. Indeed, 2 of the 11 original samples show second-degree relationships ( $IBD = 0.23–0.25$ ) (Vidal et al. 2019); Unfortunately, these are the only available whole-genome sequences from modern Mapuche.

In the present study, we analyzed how gene flow from Europeans into Native Americans contributed adaptive genetic variation through PAS to admixed Chileans. We found that the strongest signals are associated with genes previously involved in pigmentation, thermogenesis, and immune defense against pathogens (Guenther et al. 2014; Yang et al. 2018), all of them phenotypes under strong selection in

diverse human populations (Karlsson et al. 2014; Amorim et al. 2015; Yang et al. 2018). In relation to pigmentation, populations with dark skin tend to be protected against the lower folic acid levels induced by UV light, whereas light skin is an adaptation to maintain proper vitamin D levels (Deng and Xu 2018). Adaptation acting on thermogenesis enabled the Inuit as well as ancient Native American populations to adapt to extremely cold climates (Amorim et al. 2015, 2017; Fumagalli et al. 2015). Adaptation to dangerous infectious pathogens has occurred in response to the high mortality produced by them (Karlsson et al. 2014).

After the Columbian exchange, European colonization of the Americas resulted in the exposure of immigrants and natives to novel selective pressures. For example, many infectious diseases brought in by Europeans decimated native populations. These include smallpox, measles, influenza, and several others (Cook 1999; Waldman and Braun 2009). In contrast, there is little evidence supporting a role of endemic American diseases in the mortality of European settlers. Perhaps the strongest claim is syphilis, which caused a devastating pandemic in Europe that began in 1495, right after Columbus arrived to the United States. However, whether or not syphilis is endemic of the Americas is still subject of debate (de Melo et al. 2010). Arguably, to cope with the stronger selective pressures exerted by European infectious diseases, it is expected that the Latino admixed descendants of Native Americans and Europeans underwent PAS due to adaptive European genetic variation related with the immune response against pathogens brought by Europeans.

Our results show that one target undergoing PAS among Chileans is the *RP11-13A1.1* lncRNA. lncRNAs are an important class of noncoding RNAs that regulate gene expression through diverse mechanisms (Marchese et al. 2017). Interestingly, a transcriptome-wide study on lncRNAs following pathogen infection showed that *RP11-13A1.1* was strongly upregulated in monocytes upon infection of the fungal pathogens *Candida albicans* and *Aspergillus fumigatus*, but not upon infection of the bacteria *Escherichia coli*. Hence, the authors suggested that this lncRNA is a marker specific to fungal—but not bacterial—infections (Riege et al. 2017). Yet, a recent study showed that *RP11-13A1.1* was significantly downregulated in monocytes isolated from patients with Q fever, an infection caused by the zoonotic bacterium *Coxiella burnetii* (Raijmakers et al. 2019), suggesting that this lncRNA may affect the immune response against nonfungal infections as well. In addition, *RP11-13A1.1* was significantly and differentially upregulated in CD4<sup>+</sup> T-cells of obese children with asthma compared with normal-weight children. The authors hypothesized that a possible role of T-cells during this condition is to increase neutrophilic airway inflammation among obese asthmatic children (Rastogi et al. 2018). Neutrophils are important players in the innate and adaptive immune systems, are the primary defense line

against infection and one of the key cell types involved in initiation of the inflammatory response (Rosales et al. 2017).

Another relevant hit undergoing PAS is *rs12821256*, which maps an enhancer that regulates *KITLG* gene expression. *KITLG* encodes a ligand for the KIT receptor tyrosine kinase and regulates skin as well as hair pigmentation through the production of melanin pigments (Guenther et al. 2014; Yang et al. 2018). The molecular function of *rs12821256* in relation with hair pigmentation has been characterized in detail (Guenther et al. 2014). Transgenic mice carrying two *rs12821256-G* allele copies are significantly lighter than mice carrying two copies of the *rs12821256-A* ancestral allele, and also show significantly decreased *KITLG* expression in skin cells. Among humans, the blond-associated *rs12821256-A > G* substitution is prevalent in northern Europeans, but it is almost absent among Africans and Asians (Guenther et al. 2014). In addition, *rs12821256* has a significant effect on skin color in African-Americans (Miller et al. 2007). Further, *rs12821256* has been targeted by positive selection in Europeans (Grossman et al. 2010) and East Asians (Sturm and Duffy 2012; Yang et al. 2018). Moreover, other *KITLG* variants have been associated with skin color in the Han Chinese. Indeed, *KITLG* has undergone recurrent selection in European and East Asian populations (Williamson et al. 2007), possibly due to its adaptive effects on pigmentation and low temperatures at high latitude regions (Yang et al. 2018).

Besides its role in pigmentation and thermogenesis, *rs12821256* is also an eQTL for *SUGT1* expression in blood. *SUGT1* is a highly conserved gene among humans and mammals—including humans—and is indispensable for the activity of inflammasomes (Mayor et al. 2007). These are a specialized group of intracellular sensors that are essential components of the host innate immune system against microbial and cellular insults, including those that occur in autoinflammatory diseases (Tartey and Kanneganti 2020). *SUGT1* variants have been significantly GWAS-associated with respiratory diseases with a strong inflammatory component, such as asthma. Noteworthy, higher proportions of Mapuche ancestry among admixed Chileans have been significantly associated with increase mortality rates due to asthma (Lorenzo Bermejo et al. 2017). Thus, *rs12821256* could also have contributed to PAS due to its regulatory effects over *SUGT1*. This might have been relevant in the context of the inflammatory response that occurred during asthma or other chronic inflammatory diseases in Chilean individuals under genetic risk during the Columbian Exchange.

Due to the pleiotropic associations between *rs12821256* and *KITLG* with pigmentation, thermogenesis, and expression levels of immune genes, it is difficult to identify the selective pressure that drove *rs12821256* into PAS. However, because the role of *KITLG* in adaptation related to skin pigmentation (Williamson et al. 2007; Grossman et al. 2010; Yang et al. 2018) and cold climate (Yang et al. 2018) has already been characterized in Eurasians, the most parsimonious hypothesis

seems to be that *KITLG* variants may have been advantageous for Chileans due to similar reasons. Some observations support this hypothesis. 1) UV radiation ranges from extremely high in Northern Chile to extremely low in Southern Chile, and the highest levels of surface UV irradiance in the world have been measured in the Atacama Desert (Cordero et al. 2018). 2) There is high temperature variation along the 4,270 km north-south axis of Chile, and a big fraction of this geographical range—including the Patagonia region, where the Mapuche lived for centuries—shows very low temperatures (Miller 1976). It seems unlikely that *KITLG* variants were selected through sexual selection due to mating preferences for lighter hair color, since at least in Europeans, selection on hair color is mostly a pleiotropic effect of *KITLG* being selected for tolerance to climate and UV radiation (Stern et al. 2020).

Regarding *RP11-13A1.1*, we speculate that it may have contributed to immune defense against European infectious pathogens during the Columbian Exchange. This is because of two reasons. 1) The regulation of *RP11-13A1.1* has been well characterized in the context of immune defense against pathogens. 2) European infectious pathogens were an extreme selective pressure for Native American populations. However, it is also possible that *RP11-13A1.1* variants were hitchhiked into high frequency by *KITLG*-associated variants affected by PAS.

In conclusion, European variants of the haplotype in chromosome 12 underwent PAS most likely due to pleiotropic advantages that were relevant for Chileans following the Columbian Exchange.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

This work was supported by the Fondo Nacional de Desarrollo Científico y Tecnológico “FONDECYT” (3170038 to L.V., 1120987 to M.I.F., and 1200146 and 1160833 to S.E.). S.E. was additionally supported by the Instituto Milenio de Investigación sobre los Fundamentos de los Datos (IMFD) (Iniciativa Científica Milenio). O.K. and V.S. did this research within the framework of the HSE University Basic Research Program. F.I.M. acknowledges funds from Conicyt/Fondap 15110006 and VRI, P. Universidad Católica de Chile (Proyecto Investigación Interdisciplinaria 5/2015). L.V. thanks Rasmus Nielsen for hosting him from August 2018 to January 2019 at the Center for Theoretical Evolutionary Genomics, UC Berkeley. We thank the two reviewers for their constructive feedback.

## Author Contributions

L.V. and S.E. conceived the study. O.K. performed the simulations. V.S. designed and supervised the simulations, with

input from L.V. L.V. analyzed the data, with contributions from T.N. F.I.M. provided intellectual input. M.I.F. provided genotype data. L.V. wrote the article, with contributions from F.I.M and V.S.

## Literature Cited

- 1000 Genomes Project Consortium, et al. 2015. A global reference for human genetic variation. *Nature* 526:68–74.
- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.
- Amorim CE, Daub JT, Salzano FM, Foll M, Excoffier L. 2015. Detection of convergent genome-wide signals of adaptation to tropical forests in humans. *PLoS One* 10(4):e0121557.
- Amorim CE, et al. 2017. Genetic signature of natural selection in first Americans. *Proc Natl Acad Sci U S A.* 114(9):2195–2199.
- Baran Y, et al. 2012. Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* 28(10):1359–1367.
- Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. 2016. pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* 32(18):2817–2823.
- Bhatia G, et al. 2014. Genome-wide scan of 29,141 African Americans finds no evidence of directional selection since admixture. *Am J Hum Genet.* 95(4):437–444.
- Bigham A, et al. 2010. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* 6(9):e1001116.
- Browman DL. 1984. Prehispanic Aymara expansion, the Southern Altiplano and San Pedro de Atacama. *Estud Atacameños: Editorial Lautaro Núñez Atencio.* (7):179–190.
- Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet.* 81(5):1084–1097.
- Chacon-Duque JC, et al. 2018. Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. *Nat Commun.* 9(1):5388.
- Cheng JY, Mailund T, Nielsen R. 2017. Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics* 33(14):2148–2155.
- Cheng JY, Racimo F, Nielsen R. 2019. Ohana: detecting selection in multiple populations by modelling ancestral admixture components. *BioRxiv.*
- Cook DN. 1999. *Born to die: disease and New World Conquest, 1492–1650.* Cambridge University Press.
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185(4):1411–1423.
- Corbett-Detig R, Jones M. 2016. SELAM: simulation of epistasis and local adaptation during admixture with mate choice. *Bioinformatics* 32(19):3035–3037.
- Cordero RR, et al. 2018. Ultraviolet radiation in the Atacama Desert. *Antonie Van Leeuwenhoek.* 111(8):1301–1313.
- de Melo FL, de Mello JC, Fraga AM, Nunes K, Eggers S. 2010. Syphilis at the crossroad of phylogenetics and paleopathology. *PLoS Negl Trop Dis.* 4(1):e575.
- Deng L, Ruiz-Linares A, Xu S, Wang S. 2016. Ancestry variation and footprints of natural selection along the genome in Latin American populations. *Sci Rep.* 6(1):21766.
- Deng L, Xu S. 2018. Adaptation of human skin color in various populations. *Hereditas* 155(1).
- Encina F. 1993. *Historia de Chile.* Santiago de Chile: Sociedad Editora Revista Ercilla.
- Eyheramendy S, Martinez FI, Manevy F, Vial C, Repetto GM. 2015. Genetic structure characterization of Chileans reflects historical immigration patterns. *Nat Commun.* 6(1):6472.
- Fumagalli M, et al. 2015. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* 349(6254):1343–1347.
- Grossman SR, et al. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 327(5967):883–886.
- Guan Y. 2014. Detecting structure of haplotypes and local ancestry. *Genetics* 196(3):625–642.
- Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM. 2014. A molecular basis for classic blond hair color in Europeans. *Nat Genet.* 46(7):748–752.
- Hysi PG, et al. 2018. Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat Genet.* 50(5):652–656.
- Jeong C, et al. 2014. Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat Commun.* 5(1):3281.
- Karlsson EK, Kwiatkowski DP, Sabeti PC. 2014. Natural selection and infectious disease in human populations. *Nat Rev Genet.* 15(6):379–393.
- Ko A, et al. 2014. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. *Nat Commun.* 5(1):3983.
- Lamnidis TC, et al. 2018. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nat Commun.* 9(1):5018.
- Laso-Jadart R, et al. 2017. The genetic legacy of the Indian Ocean slave trade: recent admixture and post-admixture selection in the Makranis of Pakistan. *Am J Hum Genet.* 101(6):977–984.
- Lindo J, et al. 2018. The genetic prehistory of the Andean highlands 7000 years BP though. *Sci Adv.* 4(11):eaau4921.
- Lorenzo Bermejo J, et al. 2017. Subtypes of Native American ancestry and leading causes of death: mapuche ancestry-specific associations with gallbladder cancer risk in Chile. *PLoS Genet.* 13(5):e1006756.
- Lowe WL Jr, Reddy TE. 2015. Genomic approaches for understanding the genetics of complex disease. *Genome Res.* 25(10):1432–1441.
- Marchese FP, Raimondi I, Huarte M. 2017. The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* 18(1):206.
- Mayor A, Martinon F, De Smedt T, Pétrilli V, Tschopp J. 2007. A crucial function of SGT1 and HSP90 in inflammasome activity links mammalian and plant innate immune responses. *Nat Immunol.* 8(5):497–503.
- McLaren W, et al. 2016. The Ensembl Variant Effect Predictor. *Genome Biol.* 17(1):122.
- McVean G. 2009. A genealogical interpretation of principal components analysis. *PLoS Genet.* 5(10):e1000686.
- Miller A. 1976. The climate of Chile. In: Schwerdtfeger W, editor. *Climates of Central and South America. World survey of climatology.* Vol. 12. Amsterdam: Elsevier Co. p. 113–145.
- Miller CT, et al. 2007. cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* 131(6):1179–1189.
- Morgan MD, et al. 2018. Genome-wide study of hair colour in UK Biobank explains most of the SNP heritability. *Nat Commun.* 9(1):5271.
- Nielsen R, et al. 2017. Tracing the peopling of the world through genomics. *Nature* 541(7637):302–310.
- Pickrell JK, Reich D. 2014. Toward a new history and geography of human genes informed by ancient DNA. *Trends Genet.* 30(9):377–389.
- Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.
- Rajmakers RP, et al. 2019. Long-lasting transcriptional changes in circulating monocytes of acute Q fever patients. *Open Forum Infect Dis.* 6(7):ofz296.
- Rastogi D, et al. 2018. CDC42-related genes are upregulated in helper T cells from obese asthmatic children. *J Allergy Clin Immunol.* 141(2):539–548.e537.

- Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. 2019. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 47(D1):D886–D894.
- Riege K, et al. 2017. Massive effect on lncRNAs in human monocytes during fungal and bacterial infections and in response to vitamins A and D. *Sci Rep.* 7(1):40598.
- Rosales C, Lowell CA, Schnoor M, Uribe-Querol E. 2017. Neutrophils: their role in innate and adaptive immunity 2017. *J Immunol Res.* 2017:1–2.
- Secolin R, et al. 2019. Distribution of local ancestry and evidence of adaptation in admixed populations. *Sci Rep.* 9(1):13900.
- Stern AJ, Spiedel L, Zaitlen NA, Nielsen R. 2020. Disentangling selection on genetically correlated polygenic traits using whole-genome genealogies. *BioRxiv.*
- Sturm RA, Duffy DL. 2012. Human pigmentation genes under environmental selection. *Genome Biol.* 13(9):248.
- Tartey S, Kanneganti TD. 2020. Inflammasomes in the pathophysiology of autoinflammatory syndromes. *J Leukoc Biol.* 107(3):379–391.
- Tataru P, Simonsen M, Bataillon T, Hobolth A. 2017. Statistical inference in the Wright-Fisher model using allele frequency data. *Syst Biol.* 66(1):e30–e46.
- Verdugo RA, et al. 2020. Development of a small panel of SNPs to infer ancestry in Chileans that distinguishes Aymara and Mapuche components. *Biol Res.* 53(1):15.
- Vicuña L, et al. 2019. Adaptation to extreme environments in an admixed human population from the Atacama Desert. *Genome Biol Evol.* 11(9):2468–2479.
- Vidal EA, et al. 2019. Whole genome sequence, variant discovery and annotation in Mapuche-Huilliche native South Americans. *Sci Rep.* 9(1):2132.
- Vitti JJ, Grossman SR, Sabeti PC. 2013. Detecting natural selection in genomic data. *Annu Rev Genet.* 47(1):97–120.
- Waldman C, Braun M. 2009. *Atlas of the North American Indian*. New York: Facts On File.
- Ward LD, Kellis M. 2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40(Database issue):D930–D934.
- Welter D, et al. 2014. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42(D1):D1001–D1006.
- Williamson SH, et al. 2007. Localizing recent adaptive evolution in the human genome. *PLoS Genet.* 3(6):e90.
- Yang Z, et al. 2018. Darwinian positive selection on the pleiotropic effects of KITLG explain skin pigmentation and winter temperature adaptation in Eurasians. *Mol Biol Evol.* 35(9):2272–2283.

Associate editor: Soojin Yi