

Genomic Analyses of Human European Diversity at the Southwestern Edge: Isolation, African Influence and Disease Associations in the Canary Islands

Beatriz Guillen-Guio,¹ Jose M. Lorenzo-Salazar,² Rafaela González-Montelongo,² Ana Díaz-de Usera,² Itahisa Marcelino-Rodríguez,¹ Almudena Corrales,^{1,3} Antonio Cabrera de León,¹ Santos Alonso,⁴ and Carlos Flores^{*,1,2,3}

¹Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain

²Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain

³CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain

⁴Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country UPV/EHU, Leioa, Bizkaia, Spain

*Corresponding author: E-mail: cflores@ull.edu.es.

Associate editor: Connie Mulligan

Abstract

Despite the genetic resemblance of Canary Islanders to other southern European populations, their geographical isolation and the historical admixture of aborigines (from North Africa) with sub-Saharan Africans and Europeans have shaped a distinctive genetic makeup that likely affects disease susceptibility and health disparities. Based on single nucleotide polymorphism array data and whole genome sequencing (30×), we inferred that the last African admixture took place ~14 generations ago and estimated that up to 34% of the Canary Islander genome is of recent African descent. The length of regions in homozygosity and the ancestry-related mosaic organization of the Canary Islander genome support the view that isolation has been strongest on the two smallest islands. Furthermore, several genomic regions showed significant and large deviations in African or European ancestry and were significantly enriched in genes involved in prevalent diseases in this community, such as diabetes, asthma, and allergy. The most prominent of these regions were located near *LCT* and the HLA, two well-known targets of selection, at which 40–50% of the Canarian genome is of recent African descent according to our estimates. Putative selective signals were also identified in these regions near the *SLC6A11-SLC6A1*, *KCNMB2*, and *PCDH20-PCDH9* genes. Taken together, our findings provide solid evidence of a significant recent African admixture, population isolation, and adaptation in this part of Europe, with the favoring of African alleles in some chromosome regions. These findings may have medical implications for populations of recent African ancestry.

Key words: consanguinity, ROH, local ancestry, MHC, natural selection.

Introduction

The Canarian Archipelago consists of seven main islands located in the Atlantic Ocean ~100 km from the northwest African coast. By no later than 2,500 years B.P. (Onrubia Pintado 1987) and until the XVth century, when the Spanish conquest began, the Canary Islands were inhabited by the indigenous Guanche population (Crosby 1999). Many anthropological, archaeological, and cultural traits indicate that the most likely origin of Guanche aborigines was the Berber population from North Africa (Hooton 1970), and supporting evidence indicates more than one aboriginal settlement from the Maghreb and the Sahara likely occurred (Arco and Navarro 1987). The Spanish conquest can be divided into two stages: 1) occupation of the less populated islands, which was concluded rapidly and peacefully, and

2) a subsequent slower and violent invasion of the more populated islands, which ended in 1496 (de Abreu Galindo and Cioranescu 1977). Despite the devastating effect of the conquest, many aborigines remained in the territory, either freed or enslaved. The strategic location of the Canary Islands (located between the Americas and Africa) stimulated a continuous immigration between the XVIth and XIXth centuries from Europe and sub-Saharan Africa, with immigration from the latter occurring as a result of the slave trade (Lobo-Cabrera 1993). By the XVIth century, the chronicles estimated the population size as ~35,000 inhabitants, of which nearly 11,000 were potentially of aboriginal or sub-Saharan African origin (Wölfel 1930). Physical anthropology studies of the inhabitants provided evidence of the continuity of indigenous traits during the XXth century (Falkenburger 1942; Ara 1959; Berthelot 1978).

© The Author(s) 2018. 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.

Open Access

The aboriginal, historical, and contemporary populations inhabiting the Canary Islands have been the subjects of many population-genetics studies. Among these studies, those focusing on uniparental inheritance markers in samples from current inhabitants (Rando et al. 1999; Flores et al. 2003) and ancient DNA studies of aboriginal (Maca-Meyer et al. 2004; Fregel et al. 2009; Rodríguez-Varela et al. 2017) and historical remains (Maca-Meyer et al. 2005) have yielded consistent findings. Overall, the genetic evidence strongly supports a nearby North African origin of Guanches. Although this ancestry is still evident in present-day inhabitants of the archipelago, other genetic influences from Europe and Africa are also apparent. These other genetic influences have been explained by the shifts in ethnic mixtures after the Spanish conquest. However, because of the sexual asymmetry in the genetic contributions of the ancestral populations (Flores et al. 2001), autosomal markers offer more unbiased estimates than uniparental markers of the recent admixture of European (EUR), North African (NAF) and sub-Saharan African (SSA) ancestry among present-day Canary Islanders. Classical studies analyzing a few blood groups, red-blood-cell enzymes, or approximately a dozen polymorphic *Alu* insertions have shown that this three-way admixture model encompasses a prominent EUR ancestry (62–78%) and as much as 20–38% NAF and 3–10% SSA ancestries (Flores et al. 2001; Maca-Meyer et al. 2004). Studies using single nucleotide polymorphisms (SNPs), although limited by number of genetic markers (Pino-Yanes et al. 2011) or sample size (Botigué et al. 2013), have shown agreement in placing the population ancestries at ~75–83% EUR, <2% SSA, and as much as 17–23% NAF. Therefore, although NAF ancestry is widespread in Southern Europe (particularly in southwestern populations), these estimates suggest that NAF ancestry in Canary Islands populations reaches the highest levels so far described for Europe (Botigué et al. 2013).

We recently showed that the genetic diversity related to NAF ancestry has important biomedical implications for EUR populations (Botigué et al. 2013). Because of its unique genetic admixture and/or environmental exposures (likely involving evolutionary adaptations), the population inhabiting the Canarian Archipelago suffers from a disproportionate burden of prevalent chronic conditions and associated complications. For example, the prevalence of asthma and allergic diseases in children in the Canary Islands is markedly higher than that in mainland Spain (Sánchez-Lerma et al. 2009). Diabetes, obesity, and hypertension are also more prevalent among Canary Islanders than in other Spanish populations in all age groups (Marcelino-Rodríguez et al. 2016). Moreover, despite the Canarian Archipelago and mainland Spain share the healthcare system, diabetes-related mortality in the Canary Islands remains the highest in the country, and the incidence of diabetes-related morbidities, such as end-stage renal disease and lower limb amputation, differ between the two regions (Aragón-Sánchez et al. 2009; Lorenzo et al. 2010).

Here, we aimed to characterize in detail the genomes of Canary Islanders at an unprecedented scale using SNP arrays and whole genome sequencing (WGS) of samples from all seven islands. We also assessed these data to evidence loci

showing large deviations in ancestry with respect to the average of the genome, putative targets of selection and disease associations. In addition, because the population offers a uniquely challenging admixture scenario (i.e., a three-way admixture combined with the admixture of parental populations exhibiting small to moderate degrees of differentiation), this study secondarily evaluated the performance of two of the fastest and most accurate methods of local ancestry estimation for multiway admixtures (Chimusa et al. 2014; Guan 2014).

Results

Time since Admixture and Genomic Ancestry Proportions

After quality control procedures, the intersection with reference data sets, and the exclusion of regions in high-linkage disequilibrium (LD), a total of 100,175 SNPs (r^2 threshold of 0.5) from 416 individuals (34 from El Hierro, 35 from La Palma, 78 from La Gomera, 64 from Tenerife, 117 from Gran Canaria, 32 from Fuerteventura, and 56 from Lanzarote) were used for the analyses. The leading principal components (PCs) (explaining 62.54% of total variation) from the principal component analysis (PCA) including reference populations evidenced the intermediate position of Canary Islanders between the NAF and EUR populations (separated by PC2) and their more distant relationships with SSA, separated by PC1 (fig. 1). In addition, despite the relative homogeneity of ancestry supported by the tight clustering of populations from the different islands, individuals from La Gomera and El Hierro clustered closer to NAF, whereas those from Tenerife and Gran Canaria clustered closer to EUR. These results suggest that there is no clear relationship between the geographic and genetic distances from Africa in the Canary Islands.

We next examined whether Canary Islanders as a whole can be considered an admixed population based on formal tests. To do so, for each pair of reference populations considered as surrogates for the true ancestral populations, we used ALDER to calculate a weighted two-locus admixture LD statistic based on the decay (supplementary fig. 1, Supplementary Material online) and assess whether this statistic supports an admixture of the ancestral populations. In all of the pairs considering one EUR population and one SSA population as proxies of the ancestral populations, there was consistent and significant evidence of admixture ($P < 3.1 \times 10^{-69}$). Using ALDER, we also estimated the number of generations since the last admixture event. The results were consistent across comparisons and indicated that the admixture event took place $\sim 13.6 \pm 0.7$ generations ago (429–495 years BP), which is within the timescale of the historical conquest of the archipelago in the XVth century.

To complement these results, we examined the ancestry proportions based on a clustering analysis using ADMIXTURE with K varying from 2 to 7 (supplementary fig. 2, Supplementary Material online). This revealed a clear ancestry component separating SSA from the other populations from $K = 2$ through $K = 7$ and being absent in EUR individuals. At

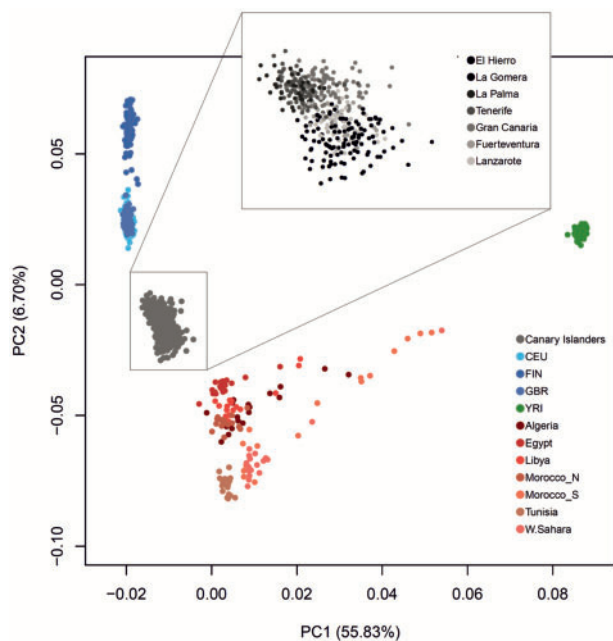


Fig. 1. Plot of the first two principal components (explaining 62.5% of variability) from PCA of Canary Islanders and samples from reference populations from Europe, North Africa, and sub-Saharan Africa. The inset depicts a detailed view per island. Results are based on a subsample of 100,175 SNPs excluding those in high LD (pairwise r^2 threshold = 0.5).

$K \geq 3$, a new ancestry cluster became apparent, which reached its maximum frequency in EUR and Canary Islanders. At $K = 4$ and above, ~ 40 – 60% of the ancestry clusters detected in NAF individuals were assigned to a new cluster with its maximum in these populations ($>95\%$ on an average in Tunisians). This NAF-related ancestry was also evident in Canary Islanders. Other ancestry clusters emerged for $K > 4$, which generally reflected both further ancestry sharing or unresolved clustering in NAF and Canary Islanders and additional ancestry subdivisions among NAF populations compatible with the mixed pattern of ancestry components evidenced elsewhere (Henn et al. 2012; Arauna et al. 2017). Cross-validation error was lowest for $K = 4$ (fig. 2). At this value, the clusters roughly corresponded to the ancestries, reaching their maximum frequencies in SSA and NAF plus two other ancestry clusters defined for EUR. The results for EUR are largely compatible with previous observations (Seldin et al. 2006). To provide further support to the ancestry clusters evidenced by the ADMIXTURE analysis, we evaluated the fitting of the admixture model with the optimal measures of haplotype sharing between groups using badMIXTURE. For $K = 4$, we observed small residuals with no systematic patterns on the Canary Islanders, whereas larger residuals were concentrated in one of the ancestral components (NAF) (supplementary fig. 3, Supplementary Material online). Therefore, regardless of the genetic heterogeneity resulting from the merging of different North African populations into a single group, the structure of the residuals indicate that the admixture proportions provided by the ADMIXTURE analysis constitute robust inferences of the ancestries in Canary Islanders.

Based on this evidence, we interpreted the EUR-, NAF-, and SSA-related fractions indicated by the ADMIXTURE analysis as the admixture proportions in Canary Islanders. For $K = 4$, the overall ancestry proportions among Canary Islanders were, on an average, 22% NAF and 3% SSA (table 1). These estimates are highly concordant with our previous results obtained under standard settings (Botigué et al. 2013) despite the latter being based on a data set containing SNPs with moderate levels of LD (r^2 threshold set at 0.5). However, in this study, which comprised a larger and more diverse sample of Canary Islanders, we found a wider NAF-related ancestry interval of 14.9–29.9%, and a slightly higher SSA-related ancestry of as much as 9.2% (table 1 and fig. 2). The use of alternative African data sets from The 1000 Genomes Project (1KGP) as SSA representatives in the ADMIXTURE analysis did not qualitatively change the results as the ancestry estimates were highly correlated (for $K = 4$, $R^2 \geq 0.995$; $P < 2.2 \times 10^{-16}$ in all comparisons). Similarly, balancing the number of individuals from the reference populations ($n = 75$ for each) resulted in similar ancestry estimates (averages of 26% NAF and 1% SSA). Overall, these results demonstrate that the Canary Islanders are closely related to EUR but show substantial influences from NAF and distant relatedness with SSA.

There was no difference in the NAF-related ancestry between the eastern (Fuerteventura, Lanzarote, and Gran Canaria) and western islands (Tenerife, La Gomera, La Palma, and El Hierro) (22.4% vs. 21.8%, respectively; Wilcoxon test $P = 0.160$), although the SSA-related ancestry differed significantly between the eastern and western islands (3.1% vs. 2.8%, respectively; Wilcoxon test $P = 6.4 \times 10^{-5}$). However, the largest differences were observed between the islands that were conquered first and more peacefully by the Spanish (Fuerteventura, Lanzarote, La Gomera, and El Hierro) and those in which the conquest took longer (Tenerife, Gran Canaria, and La Palma) (24.0% vs. 20.3%, respectively, for the NAF-related fraction, and 3.5% vs. 2.4%, respectively, for the SSA-related fraction; Wilcoxon test $P < 3.0 \times 10^{-5}$ for both comparisons). This observation agrees with findings based on mitochondrial DNA (mtDNA) lineages that suggest increased African affinities on the former islands (Rando et al. 1999). However, they differ from those based on the non-recombining portion of the Y chromosome (NRY) that indicate limited NAF affinities of paternal genetic markers on all of the Canary Islands but particularly the populations from the westernmost islands (Flores et al. 2003). Such different distributions of NAF and SSA ancestries, previously evidenced by only uniparental markers, are expected given the sexual asymmetry of parental contributions detected in the current and historical populations of the archipelago (Flores et al. 2001; Fregel et al. 2009).

Population Isolation

Recent studies have indicated that runs of homozygosity (ROHs) are common to all world populations, are longer than expected, and have profiles that can indicate distinctive demographic histories of the population and of inbreeding (Kirin et al. 2010; Pemberton et al. 2012). Here, we have

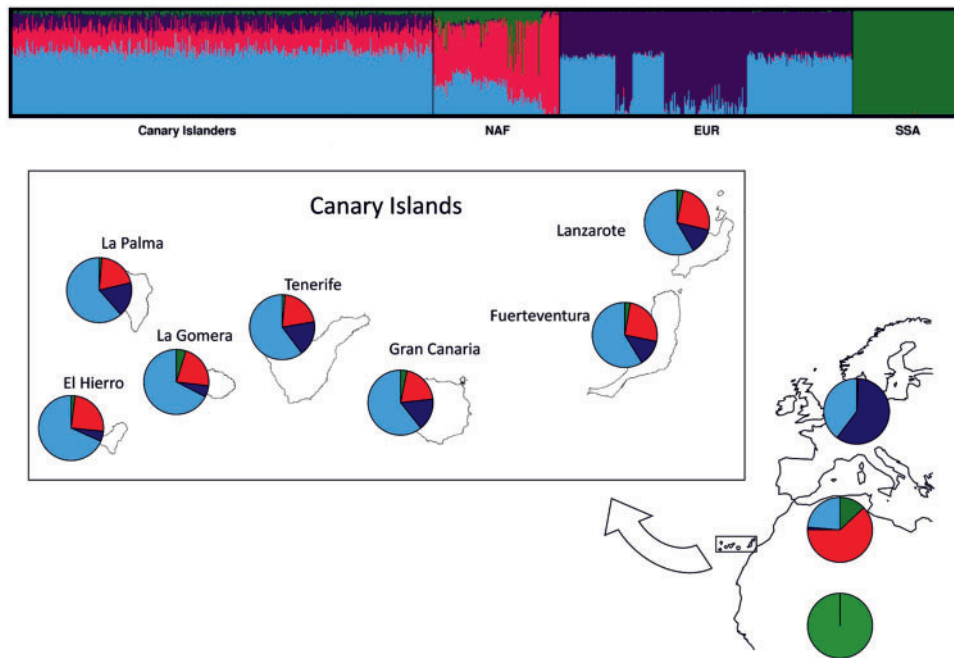


Fig. 2. ADMIXTURE estimates for $K = 4$ for Canary Islanders and samples from reference populations from Europe, North Africa, and sub-Saharan Africa.

Table 1. Genomic Ancestry Proportions (from ADMIXTURE, $K = 4$) in Canary Islanders.

	North African			Sub-Saharan African		
	Min.	Mean	Max.	Min.	Mean	Max.
Fuerteventura	0.218	0.255	0.296	0.011	0.027	0.046
Lanzarote	0.214	0.254	0.296	0.014	0.032	0.057
Gran Canaria	0.155	0.200	0.264	0.005	0.032	0.082
Tenerife	0.149	0.208	0.255	0.002	0.015	0.057
La Gomera	0.160	0.221	0.289	0.013	0.048	0.092
La Palma	0.170	0.200	0.245	0.000	0.013	0.032
El Hierro	0.192	0.246	0.299	0.005	0.020	0.032

analyzed, for the first time, genome-wide ROH patterns in Canary Islanders to reveal the level of population isolation. Focusing first on the average total sum of ROHs over all subjects sampled, we found that <10 Mb of the Canary Islanders genome on an average was in ROHs of >2 Mb in length. For smaller ROHs (≤ 1.6 Mb), which have been associated with geographic distances from East Africa (Pemberton et al. 2012), the profiles were very similar across the island populations (fig. 3). However, among all ROHs >1.6 Mb, the samples from La Gomera and El Hierro showed consistently higher average total ROHs than did the samples from the remaining islands, suggestive of increased recent inbreeding in these two islands.

An exploration of average total ROHs by island again indicated El Hierro and La Gomera as population outliers. The genomes from these two islands showed the largest number of fragments in ROHs and the longest average total ROH length (fig. 4), the latter demonstrating significant differences from all the other islands (Wilcoxon test $P < 1.0 \times 10^{-3}$ for all pairwise comparisons) except La Palma. Interestingly, La Gomera and El Hierro showed average total ROH lengths of

114 and 134 Mb. The conclusions remained unchanged when considering only ROHs >1.6 Mb for the average total ROHs estimates (Wilcoxon test $P < 1.0 \times 10^{-3}$ for all pairwise comparisons), whereas the differences disappeared when using the average total of ROHs ≤ 1.6 Mb (lowest $P = 4.5 \times 10^{-3}$). Given that variant ascertainment of the array is less of a problem for ROHs >1 Mb (Ceballos et al. 2018), we interpret the ROH patterns observed for La Gomera and El Hierro as signatures of genetic isolation, reduced population size and consequently, endogamy within the archipelago. In this respect, although the ROH patterns suggested a tendency toward larger average total ROH length in the western and smaller islands of the archipelago, the mean number of genomic regions in ROHs significantly increased with geographical longitude (Spearman's rank correlation $\rho = 0.93$, $P = 6.7 \times 10^{-3}$).

Local Ancestry: Assessing Estimators and the Average Size of Blocks

In cases of recent admixture, the genomes of admixed individuals become a mosaic of chromosomal stretches originating from the ancestral populations, which can be detected by examining locus-specific ancestry (usually termed local ancestry). Many local ancestry methods perform well in two-way admixture scenarios with large genetic differences between the ancestral populations (e.g., in African–Americans). However, the difficulty increases with the number of ancestral populations (e.g., three-way admixtures), particularly if a small to moderate degree of differentiation exists between some of the ancestral populations (e.g., between NAF and EUR). Given that this study constitutes the first time that the patterns of ancestry blocks in Canary Islander genomes have been assessed, we first evaluated two of the fastest and most

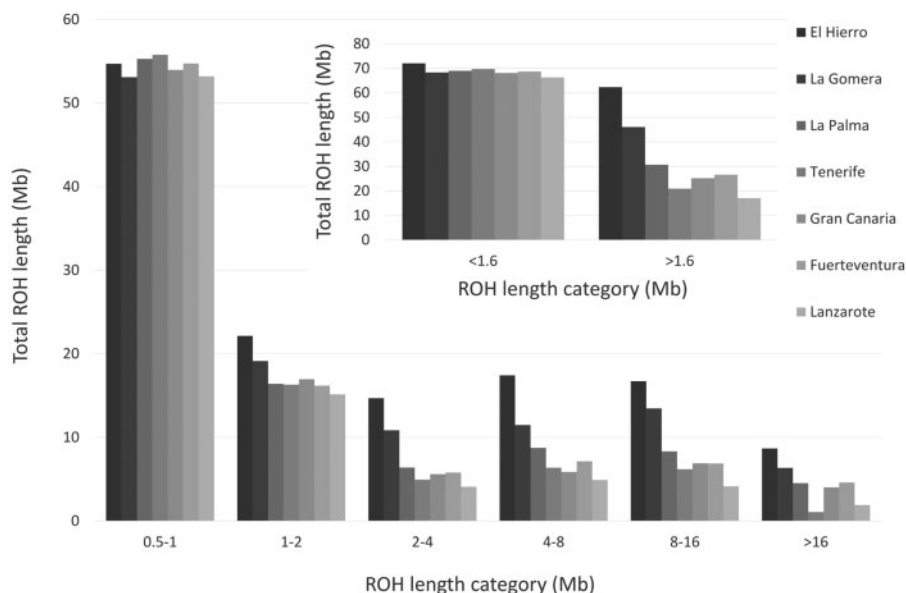


Fig. 3. Average length (Mb) of ROHs using two classifications into categories in the populations from the Canary Islands.

accurate methods of local ancestry estimation for multiway admixtures, ELAI and LAMP-LD, both of which have been fruitfully utilized with Hispanic populations (Baran et al. 2012; Zhou et al. 2016). We compared their results of these methods with those obtained previously by ADMIXTURE, pooling the two EUR-related ancestry fractions for the comparisons. The Pearson correlations between the estimates of ADMIXTURE and either ELAI or LAMP-LD were strongly significant ($P < 5.0 \times 10^{-10}$ in all cases) (table 2). However, ELAI offered better fitting estimates than LAMP-LD based on the overall estimates of NAF- and SSA-related ancestries obtained. Whereas the ELAI estimates were similar to those of ADMIXTURE (estimates for NAF and SSA of 23.6% and 2.3%, respectively), LAMP-LD provided an inflated estimate of NAF-related ancestry (estimates for NAF and SSA of 32.7 and 1.4%, respectively) (fig. 5). A least squares estimator indicated that individual fractions were between 2 (SSA) and 4 times (EUR and NAF) as different between ADMIXTURE and LAMP-LD than between ADMIXTURE and ELAI (not shown). In addition, whereas the lengths of the ancestry blocks provided by ELAI and LAMP-LD were correlated for the three ancestries (fig. 6), ELAI provided greater sensitivity for the detection of smaller stretches of ancestry. Based on these results, all further local ancestry analyses were conducted with ELAI estimates.

Block sizes followed lognormal distributions (supplementary fig. 4, Supplementary Material online), with the largest blocks on an average corresponding to the EUR-related component (13.05 Mb) followed by the NAF (8.46 Mb) and SSA (7.48 Mb) components, and with all pairwise comparisons yielding significant differences (Wilcoxon test $P < 3.0 \times 10^{-8}$). Based on these block length estimates and the ADMIXTURE proportions calculated for all Canary Islanders, we would expect averages of ~ 181 EUR-, 82 NAF-, and 13 SSA-related blocks per haploid genome (i.e., 276 ancestry blocks in total), which are lower than but within

the range of the estimates previously suggested for African-Americans (Shriner et al. 2011).

Deviations in Local Ancestry and Selection

Based on the previous results, we then used the ELAI estimates to assess the existence of regional deviations in any of the three ancestries in Canary Islanders. We detected eight peaks with large ancestry deviations: two enriched in EUR ancestry, and six associated with higher proportions of African ancestries (fig. 7, table 3, and supplementary fig. 5, Supplementary Material online). Notably, the EUR- and NAF-related peak positions largely overlapped. The EUR-related peaks were located on chr2 (with the lead SNP showing an average local ancestry of 55% and flanked by the *CXCR4* and *THSD7B* genes) and chr6 (with all lead SNPs located within *HLA-B*, with an average local ancestry of 49%). The two peaks of NAF-related ancestry enrichment were located on chr2 (with the lead SNP showing an average local ancestry of 43.2% and flanked by the same genes flanking the EUR-related peaks) and chr6 (with the lead SNP showing an average local ancestry of 46.3% and flanked by the *NFKB1* and *LTA* genes). Peaks associated with SSA-related ancestry were situated on chr3 (two hits: one with the lead SNP located within *SLC6A11*, the other with the lead SNP located within *KCNMB2*, both presenting an average local ancestry of $\sim 4.0\%$), chr6 (with the lead SNP within *ZNRD1*, showing an average local ancestry of 5.9%), and chr13 (with the lead SNP near *PCDH9* and flanked by two long intergenic noncoding RNA genes: *LINC00355* and *LINC01052*, showing an average local ancestry of 4.4%). These regions span several Mbs (range of 1.2–12.3 Mb), meaning that, while the observed ancestry deviations may indicate adaptive processes, any adaptation may be unrelated to the gene harboring the lead SNP. However, it is likely that the *LCT* and *HLA* loci caused the peaks on chr2 and chr6. One regulatory variant of *LCT* is the major determinant for lactase persistence in EUR populations

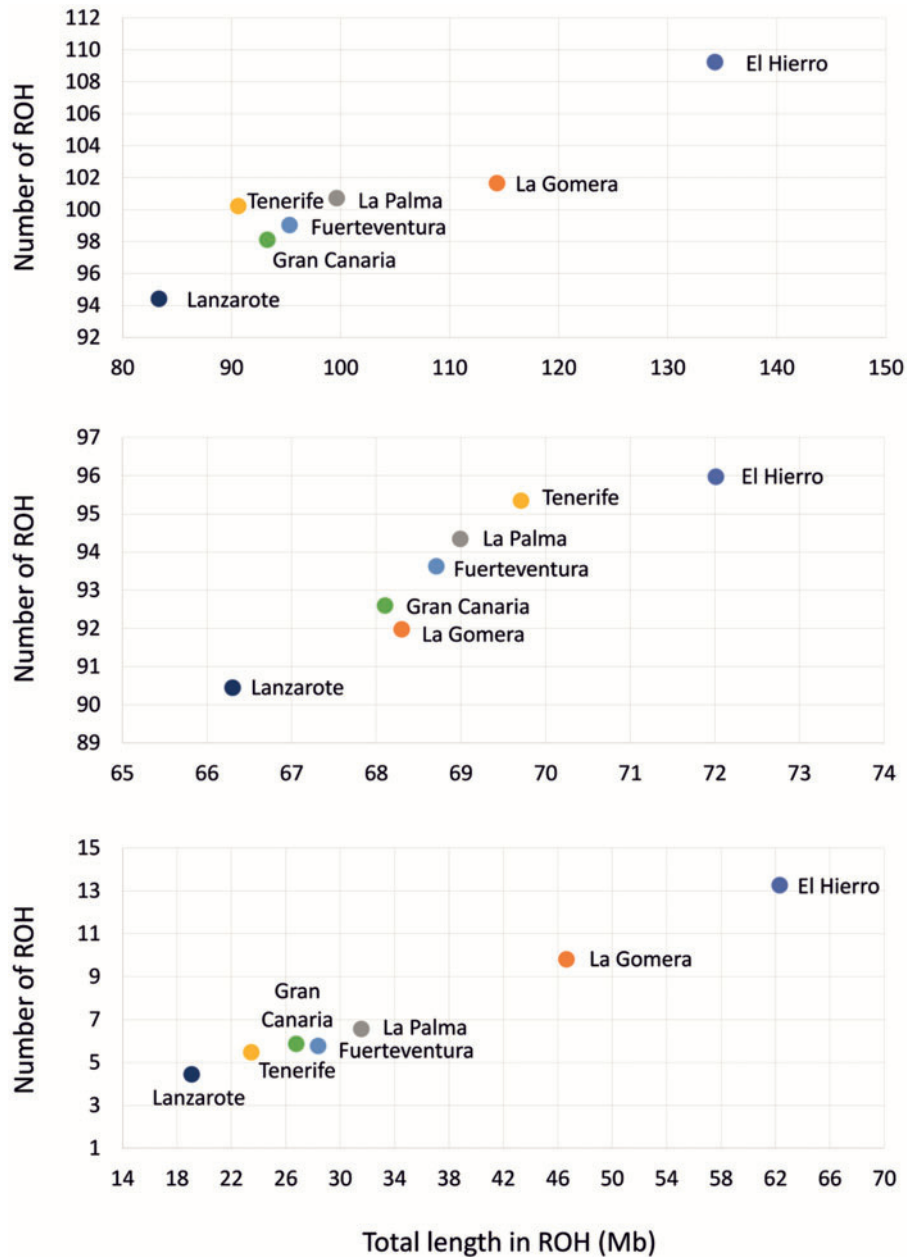


Fig. 4. Average number of genome regions in ROHs with respect to the average total (top), ≤ 1.6 Mb (middle), and > 1.6 Mb (bottom) lengths of ROHs per island.

Table 2. Pearson Correlation between ADMIXTURE ($K = 4$) and ELAI or LAMP-LD Estimates of Ancestry in Canary Islanders.

Ancestry	ELAI		LAMP-LD	
	Coef.	95% CI	Coef.	95% CI
European	0.95	0.94–0.96	0.83	0.80–0.86
North African	0.88	0.85–0.90	0.74	0.68–0.78
Sub-Saharan African	0.93	0.92–0.94	0.89	0.87–0.91

and underlies the strongest signal of selection in the human genome identified to date (Mathieson et al. 2015). Similarly, this region has been shown to be under positive selection in Bantu-speaking populations (Patin et al. 2017). As for HLA, the gene-rich human leukocyte antigen region is also a well-known target of selection in EUR and African

populations, likely involving balancing selection (Gurdasani et al. 2015; Mathieson et al. 2015; Patin et al. 2017) as well as directional selective processes, such as that described for HLA-B alleles and malaria protection in the Sahel (Sanchez-Mazas et al. 2017). Based on this evidence, we reasoned that all or some of the remaining regions with ancestry deviations detected (two on chr3 and one on chr13) could also have resulted from selective processes. One of the hallmarks of selection is increased homozygosity and reduced diversity in the surrounding regions (Pickrell et al. 2009). Therefore, we then assessed the mean heterozygosity and the number of samples with ROHs containing SNPs from the ancestry-enriched regions. We performed this assessment for all peaks associated with African ancestry (including those in chr2 and chr6) to obtain references for comparisons (table 4) and

found that one of the regions on chr3 and the region on chr13 showed signals of reduced diversity among Canary Islanders. To formally test for the existence of selective signals in chr3 and chr13 loci, we assessed the Tajima's D and the Population Branch Statistic (PBS) in a subset of individuals for which WGS data was available. Deviations from neutrality at three locations were revealed by at least one test for the three

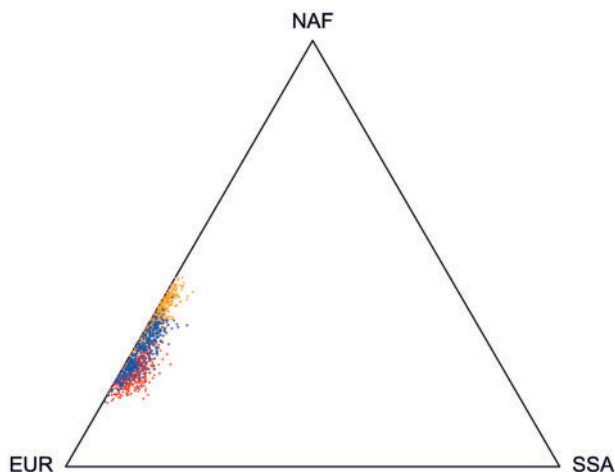


Fig. 5. Triangle plot of individual genomic admixture proportions in Canary Islanders as estimated by ADMIXTURE with $K = 4$ (red), ELAI (blue), and LAMP-LD (orange).

loci, with frequent support for more than one hit obtained for some regions after accounting for the number of comparisons ($P < 0.017$ and $P < 0.01$ for Tajima's D and PBS, respectively, after Bonferroni correction). These three locations were as follows: in the intergenic region between the *SLC6A11* and *SLC6A1* genes on chr3 (Tajima's $D = -1.733$, $P < 0.061$; PBS = 0.108, $P < 2.0 \times 10^{-4}$), near *KCNMB2* on the same chromosome (Tajima's $D = -1.511$, $P < 0.130$; PBS = 0.064, $P < 2.0 \times 10^{-4}$), and in the intergenic region between *PCDH20* and *PCDH9* on chr13 (Tajima's $D = -2.505$, PBS = 0.136, $P < 2.0 \times 10^{-4}$ in both comparisons) (table 5). Interestingly, common genetic variants previously found to be associated with height (He et al. 2015), bone traits (Kiel et al. 2007), and asthma (Ferreira et al. 2011) reside in two of these regions (table 5). This observation is in agreement with the observation that variants for inflammatory diseases located via genome-wide association studies are significantly enriched in signatures of positive selection in European populations (Raj et al. 2013).

One striking observation relates to the chr2 peak, which indicates a higher proportion of NAF-related alleles in Canary Islanders in this region. Given that at least three other *LCT* variants have been linked with lactase persistence in other populations, we then explored whether any of those variants existed in this population and could offer a potential explanation for this peak. By accessing the WGS data available for a subset of 14 subjects, we found that the only detected variant

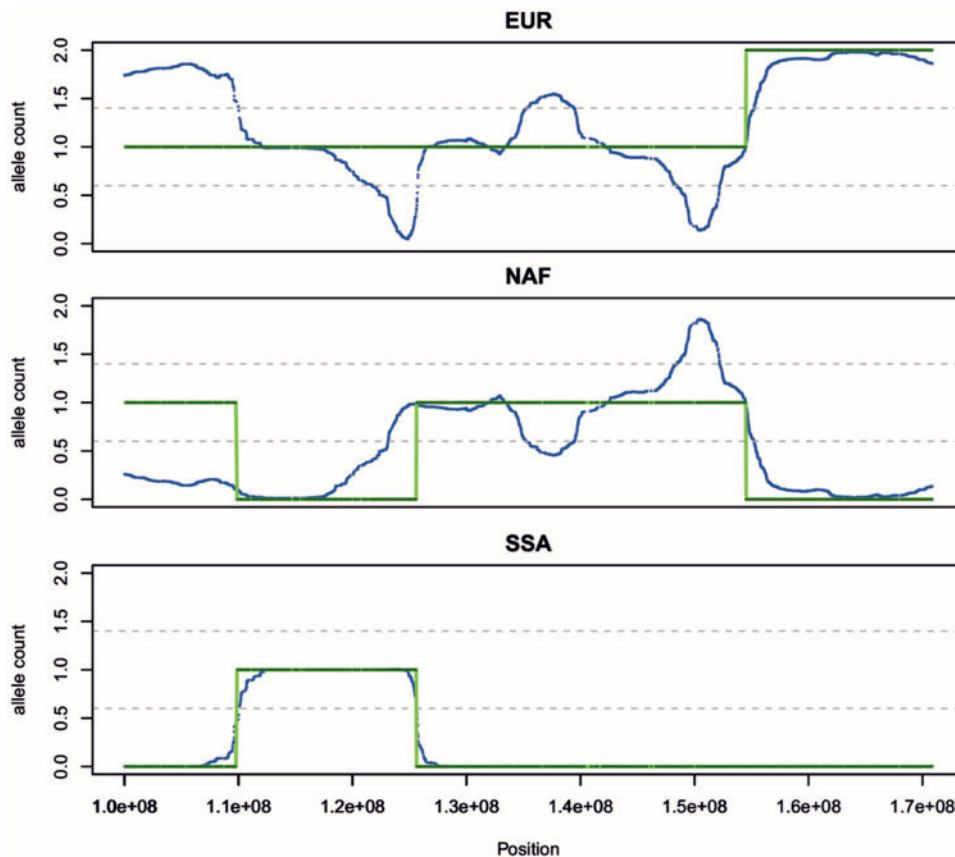


Fig. 6. Inference of local ancestry by ELAI and LAMP-LD. The plot shows an example of inference in a chromosome region (one panel of each parental population) comparing ELAI (blue) with LAMP-LD (green) allele dosages.

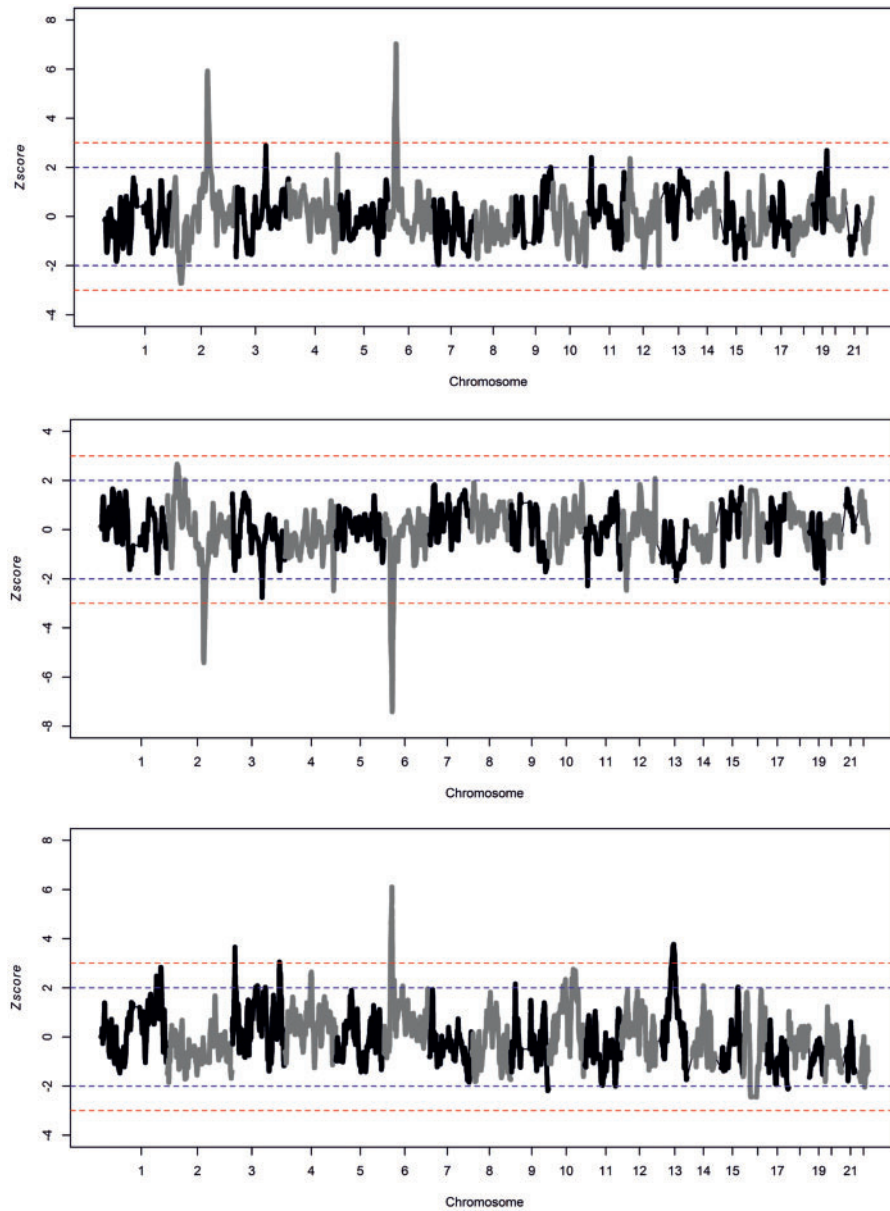


Fig. 7. Genome-wide Z-score scan of ELAI local admixtures in Canary Islanders for NAF (top), EUR (middle), and SSA (bottom). Horizontal broken lines (blue $> |2|$; red $> |3|$) indicate score thresholds.

Table 3. Genomic Regions with Supported Deviations in Ancestry among Canary Islanders.

Ancestry	Region	Lead SNP	Z Score	Mean Ancestry
North African	chr2: 133,952,040-144,266,489	rs10177911	5.92	43.2
	chr6: 24,703,442-36,288,651	rs2844484	7.03	46.3
European	chr2: 134,088,150-142,882,593	rs4954402	-5.41	55.0
	chr6: 24,120,456-36,653,597	(*)	-7.41	49.0
Sub-Saharan African	chr3: 10,539,482-11,710,471	rs17033567	3.65	4.3
	chr3: 177,443,968-178,679,751	rs13061192	3.02	3.9
	chr6: 24,790,462-32,192,083	rs16896944	6.10	5.9
	chr13: 57,962,413-70,091,195	rs9540226	3.77	4.4

(*) rs2524095, rs16899203, rs16899205, rs16899207, rs2524089, rs2394967, rs2524066, rs9366778.

position associated with lactase persistence was rs4988235 (a.k.a. -13,910), which is the major determinant for lactase persistence in Europe. The frequency of this lactase persistence allele in Canary Islanders (-13,910*T, 40%) is

intermediate between the frequency reported in other central and northern EUR populations (60–80%) and that in NAF populations (24%) (Bersaglieri et al. 2004; Ben Halima et al. 2017) (<http://www.ucl.ac.uk/mace-lab/resources/glad>).

Table 4. Diversity Estimates in Regions with Large Deviations in African Ancestry in Canary Islanders.

Region	Heterozygosity		SNPs in ROHs	
	Mean	P value ^a	Mean	P value ^a
Genome	0.300	–	12.508	–
chr2: 133,952,040..144,266,489	0.298	0.452	11.934	1.38 × 10 ^{-7b}
chr3: 10,539,482..11,710,471	0.278	0.010	9.155	0.037
chr3: 177,443,968..178,679,751	0.309	0.418	14.172	2.20 × 10 ^{-16b}
chr6: 24,703,442..36,288,651	0.296	0.015	25.934	2.20 × 10 ^{-16b}
chr6: 24,790,462..32,192,083	0.285	5.28 × 10 ^{-7b}	29.364	2.20 × 10 ^{-16b}
chr13: 57,962,413..70,091,195	0.287	2.55 × 10 ^{-4b}	22.050	2.20 × 10 ^{-16b}

^aWilcoxon rank sum test.^bStatistically significant after adjusting for multiple tests ($P < 8 \times 10^{-3}$).**Table 5.** Tajima's *D* and PBS Test Results Calculated for chr3 and chr13 Regions with a *Z* score >|3|.

Chr.	Tajima's <i>D</i>			PBS		
	Region (× 1,000)	Score (probability)	Gene	Region (× 1,000)	Score ($P < 2.0 \times 10^{-4}$)	Gene (traits)
3p25.3	11,010–11,020	–1.733 (0.061)	SLC6A11-SLC6A1	11,025–11,035	0.108	SLC6A1
				11,635–11,645	0.087	VGLL4 (height ^a)
3q26.32	178,120–178,130	–1.510 (0.130)	KCNMB2	177,465–177,475	0.177	LINC00578
				178,573–178,583	0.064	KCNMB2
13q21.32	66,060–66,070	–2.505 (<2.0 × 10 ⁻⁴)	PCDH20-PCDH9	63,600–63,650	0.136	PCDH20-PCDH9 (bone traits ^b , asthma ^c)

^aHe et al. (2015).^bKiel et al. (2007).^cFerreira et al. (2011).

Therefore, we suggest that the prevalence of lactase nonpersistence alleles in Canary Islanders and NAF populations likely explains the chr2 peak in Canary Islanders. In fact, a recent ancient DNA study of pre-Hispanic human teeth from a small sample of five Guanche people from Tenerife and Gran Canaria also suggested that the dominant phenotype was lactose intolerance (Rodríguez-Varela et al. 2017). Taken together, this evidence reduces the possibility that the known African or Arabian *LCT* variants (Ingram et al. 2009; Ranciuro et al. 2014) are responsible for the chr2 peak, although we cannot rule out the possibility that there may be other rare variants associated with lactase persistence in this population that remain undiscovered.

Links between Ancestry, Diseases, and Biological Pathways

To determine whether the genomic regions with large deviations in ancestry are linked with human diseases and biological pathways, we applied enrichment analysis to the 341 unique genes mapping to the regions with significant evidence of EUR-, NAF-, or SSA-related ancestry deviations (fig. 8 and supplementary table 1, Supplementary Material online). The top annotations were dominated by skin, vascular, renal, autoimmune, and neuropsychiatric diseases as well as by DNA metabolism, amyloids, meiosis, and transcription pathways. In addition, many prevalent diseases, such as diabetes, asthma, and allergy, and infectious diseases as well as some severe conditions, such as oncologic and severe acute respiratory syndrome, were also significantly enriched (q value < 0.05). Regulation of inflammatory response, the

complement pathway, telomere maintenance, and antigen processing and presentation were among the pathways significantly enriched (q value < 0.05) but ranked lower in the results (supplementary tables 2 and 3, Supplementary Material online).

Discussion

The recent history of the Canary Islands has involved heterogeneous genetic influences from Europe and Africa since their aboriginal settlement from nearby North Africa during the first millennium BC. Here, we report new high-density genotyping and WGS data that allowed an exhaustive exploration of the time since the last admixture event, genetic diversity and isolation, disease links, and putative selective processes in these populations. By using the largest and most diverse Canary Islands sample analyzed to date at a genomic scale, we recognized a wider range of interindividual African ancestry assignments (as much as 29.9% NAF and 9.2% SSA), the largest so far identified in southwestern EUR populations (Botigué et al. 2013). Furthermore, a between-islands pattern of variation in African ancestries was observed and interpreted according to the impact of the Spanish conquest and settlement of the territory during and after the XVth century. For the first time, we have found genomic signals of inbreeding in the population, suggesting that isolation has been especially drastic in the two smallest islands, El Hierro and La Gomera, the latter of which is associated with the highest frequency reported to date of the Northwest African U6 mtDNA lineage (>36%) in a non-African

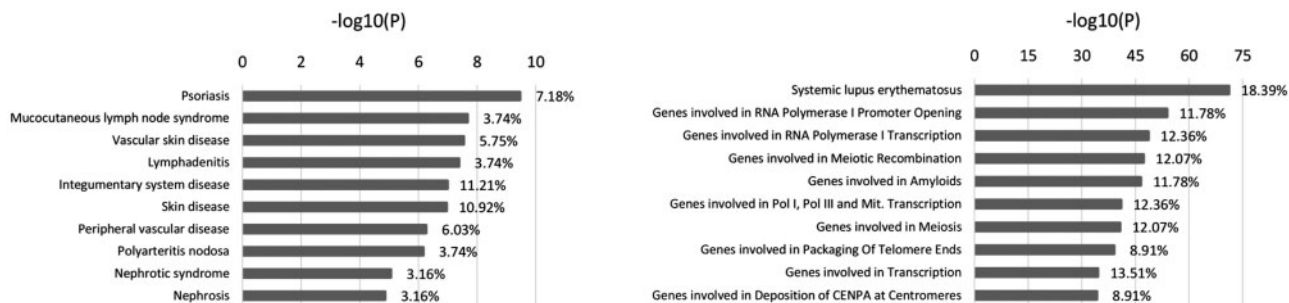


Fig. 8. Enrichment analysis on regions with large deviations for any ancestry. Top ten significantly enriched human diseases (left) and MSigDB pathways (right).

population (Rando et al. 1999). In addition, we uncovered the mosaic nature of Canary Islander genomes, finding that a typical haploid genome would consist of fewer than 300 ancestry blocks and the number of EUR-related blocks would be approximately twice and 10 times that of the NAF- and SSA-related blocks, respectively. Finally, we used this information to focus on particular chromosome regions that showed an overrepresentation of one of the recognized ancestries. These regions included the *LCT* gene, the HLA, and other genes that appear to have undergone selective processes. Most importantly, chromosome regions with large deviations in the ancestries were enriched in genes underlying important diseases that disproportionately affect the archipelago, including respiratory diseases and diabetes, among others.

Despite the profound impact of the European immigration starting in the XVth century, our results indicate that significant genetic footprints of African ancestry persist in the current inhabitants of the Canary Islands. Our data also imply that gene flow between the island populations must have been high to maintain the relative homogeneity observed for the African ancestries, despite the significant discontinuity favoring African affinities in the populations of El Hierro, La Gomera, Fuerteventura, and Lanzarote. Strikingly, such a pattern was also revealed in mtDNA studies of independent samples (Rando et al. 1999), suggesting that African ancestry has been better preserved in the populations that were conquered at the beginning of the historic process, where the literature attests to a more peaceful occupation of those territories (Suárez et al. 1988). In contrast to our previous study, which was limited by the number of autosomal markers utilized at the time (Pino-Yanes et al. 2011), in the present study, we were able to measure the existence of a minor but substantial SSA-related influence in the populations of the seven islands. SSA migrations into NAF populations are well known (Rando et al. 1998; Wright 2007), suggesting that continuous gene flow started before the aboriginal settlement of the islands. In addition, our results evidenced significant differences in the mean ancestry block lengths interpreted as of recent SSA and NAF origin in Canary Islanders. While this result may be indicative of admixture events involving different African components, as has been supported by many previous studies (Arco and Navarro 1987; de Castro 1987; Navarro Mederos 1997; Flores et al. 2003), independent analyses should aim to establish the time periods of such events.

For example, in a recent genomic study of ancient DNA, a very low proportion of SSA ancestry was observed in pre-Hispanic human remains despite their genetic resemblance to other modern NAF populations (Rodríguez-Varela et al. 2017). This finding suggests that the proportion of SSA ancestry we observed in Canary Islanders likely originated in the postconquest importation of enslaved African people.

Previous genetic studies of pre-Hispanic remains, burial remains from the XVIIth century, and samples from current inhabitants indicate declining frequencies of the NRY lineages ascribed to aboriginal people (Fregel et al. 2009). However, a continuity of aboriginal mtDNA lineages, possibly evolving locally into Canarian-specific U6b subhaplogroups (Rando et al. 1999; Maca-Meyer et al. 2004), has been evidenced in the aboriginal remains and current inhabitants, suggesting an important sexual asymmetry involved in the demographic process of the population (Flores et al. 2001). In this study, we found that five other regions in the autosomal genome, including the HLA and *LCT*, paralleling this scenario, where genetic affinities with African populations remain regionally preserved today. Independent studies and our own data support that putative selective processes are the cause of such autosomal affinities. However, given that the bulk of human adaptation occurs via selection on standing variation (Schridder and Kern 2017), it is possible that these signals are not due to private alleles but to preexisting variations in any of the parental populations involved in the admixture. Indeed, extensive evidence indicates pervasive selection processes in the HLA and *LCT* regions in different populations; these processes have been explained by dietary and cultural changes occurring with the Neolithic dispersion, including population density increases and the establishment of permanent settlements, resulting in novel pathogen exposures (Marciniak and Perry 2017).

Additionally, the *SLC6A11* and *KCNMB2* genes have been suggested to be under selective pressures in African and American populations (Hu 2012; Watkins et al. 2012), and literature evidence supports *DIAPH3*, not *PCDH9*, as the closest candidate target for selection in the chr13 region (Pickrell et al. 2009). One important consequence of this observation is that those regions contain genes in which mutations are linked to rare disorders that are usually screened in families with genetic risks, such as *PCDH9*, *SLC6A11*, *SLC6A1*, and *COL11A2*. Therefore, as local ancestry will act as a confounder

in the context of genetic tests, disease genes linked to those regions could be candidates for interpretative problems in diagnosis (Manrai et al. 2016). However, beyond monogenic conditions, such problems can arise in the context of complex disease mapping and in the evaluation of known risk factors. The HLA region, a central locus in the predisposition for or protection from many diseases (Hill et al. 1991; Oksenberg et al. 2004; Pelak et al. 2010), contains genetic markers that have been previously identified as related to complex traits such as asthma susceptibility (*MUC22*) (Galanter et al. 2014) and HIV-1 infection control (*HLA-C*) (Thomas et al. 2009), and their attributable risks would have passed unnoticed if local ancestry was not appropriately accounted for in the studies. Furthermore, the regions exhibiting local ancestry deviations have significant associations with distinct prevalent diseases in this population. Therefore, it can be speculated that the prevalence of some of these diseases in the Canary Islands population might be influenced by the distinctive genetic makeup of this population. This hypothesis is in agreement with epidemiological studies of cardiovascular risk factors (Cabrera de León et al. 2006; Bueno et al. 2008; Marcelino-Rodríguez et al. 2016), asthma and allergies (Sánchez-Lerma et al. 2009; Juliá-Serdá et al. 2011).

To the best of our knowledge, this is the first study to evaluate the proportion of the genome in ROHs in this population. Our results support distinct degrees of isolation and consanguinity between the seven islands, which are most extreme in the two smallest islands, La Gomera and El Hierro. Because of isolation, the inhabitants of these islands would be enriched in low-frequency functional variants (Xue et al. 2017) that can lead to novel discoveries of disease genes (Moltke et al. 2014; Jakobsdottir et al. 2016). As a consequence, there is an increased number of recessive variants that can confer risk for complex diseases (Rudan et al. 2003; Campbell et al. 2007; Lencz et al. 2007; Ghani et al. 2015; Thomsen et al. 2016). In addition, founder monogenic mutations are expected, as observed in distinct Canarian populations for type 1 primary hyperoxaluria (Santana et al. 2003; Lorenzo et al. 2006, 2014), sickle-cell anemia (Castella et al. 2011), Wilson's disease (García-Villarreal et al. 2000), and cardiovascular traits (Rodríguez-Esparragón et al. 2017), highlighting the singular genetic characteristics of Canary Islanders.

It is important to declare some limitations of our study. First, scarce genomic data are available for NAF populations in public databases (Henn et al. 2012). As a consequence, there was limited overlap in SNP contents with the SNP array utilized, leaving us with as few as 114,567 SNPs for some components of the study. This circumstance forced us to use 1KGP population references to maximize SNP density in the comparisons. Along with the number of generations since admixture and the ascertained nature of the contents of the array, these conditions likely had direct impacts on the local ancestry estimates, the average lengths and the regions identified by the admixture scan. However, further studies will aim to improve this overlap by further WGS and/or SNP array genotyping of novel NAF samples to be able to refine future scans. The regions identified by the admixture scan are

relatively wide, on the order of several Mb, which limits the precise allocation of ancestry peaks. Therefore, the genes highlighted by proximity to the detected peaks should be considered with caution. In this respect, the final data set had low or no coverage of centromeres, offering no basis for local ancestry inference in those regions. Although this situation would have affected all ancestries equally, we excluded those regions from the analysis to avoid an upward bias in the average block length measures. Furthermore, the choice of reference populations for ancestry inference is a common concern in these studies. In our case, we balanced choosing a reference data set with sufficient genetic resolution to retain the minimal number of SNPs required for the analyses (>100K) with the use of EUR or SSA populations where NAF or Near East influences were absent or minimal, as both factors hamper local NAF ancestry inferences. In any case, a ubiquitous Berber substrate was recently found in both pre-Hispanic remains from the aboriginal Guanche people and samples from modern NAF populations, supporting a close genetic affinity between these populations (Rodríguez-Varela et al. 2017). Finally, even in the ideal scenario of having access to population data sets from all over the world, most natural populations are expected to represent heterogeneous groups as well. This heterogeneity was the justification for considering all available NAF data sets as a single population source in the model. However, we admit that the reference population sources utilized in the study constitute model simplifications.

In conclusion, here we have provided a genetic dissection of population ancestries and isolation in Canary Islanders at an unprecedented level. We estimate that up to 34% of Canarian genomes are of recent African descent and that the geographical distribution of ancestries still reflects historical events. Local ancestry estimates enabled the identification of Mb-size chromosome regions with higher-than-expected African affinities, most likely involving putative adaptive signals. Our results suggest that these observations may have implications for the major health disparities affecting the population. In addition, genetic testing and genetic mapping studies of diseases in Canary Islanders should take local ancestry into account. Finally, because the adaptive signals were previously described in populations from Africa and America, our conclusions could also have repercussions for the identification of disease loci in other recently admixed populations.

Materials and Methods

Samples, Genotyping, and Reference Population Data Sets

The sample of Canary Islanders consisted of 429 unrelated subjects who self-declared as having two generations of ancestors born on the same island of the Archipelago. Samples were selected from a large cohort study entitled "CDC of the Canary Islands" (Cabrera de León et al. 2004), which included ~7,000 randomly selected representatives of the general Canarian population aged between 18 and 75 years and unbiased for gender. Informed consent and an extensive health survey were obtained from all participants

through personal interviews. Genotyping of 587,352 variants was conducted using the Axiom Genome-Wide Human CEU 1 Array (Affymetrix, Santa Clara, CA) with the support of the National Genotyping Center (CeGen), Universidad de Santiago de Compostela Node. Genotyping quality control was performed with R 3.2.2 and PLINK v1.07 (Purcell et al. 2007). Thus, samples with genotype call rates <95%, discordant sex and family relationships (PIHAT > 0.2) were removed from the study, leaving a total of 416 individuals, 205 men and 211 women, for further analyses (34 from El Hierro, 35 from La Palma, 78 from La Gomera, 64 from Tenerife, 117 from Gran Canaria, 32 from Fuerteventura, and 56 from Lanzarote). Additionally, those SNPs with a genotyping rate <95%, a minor allele frequency (MAF) <0.01, or deviating from Hardy–Weinberg expectations ($P < 1.0 \times 10^{-6}$) were excluded, leaving a total of 516,348 SNPs. To maximize SNP density in the downstream analyses, we relied on the 1KGP Phase 3 data to extract the data sets serving as EUR and SSA sources (Sudmant et al. 2015). According to recent genome-wide ancestry estimates (Botigué et al. 2013), the presence of Near East or NAF influences in EUR and SSA populations is minimal. In addition, NAF ancestry in EUR populations is clearly distinguishable from Near Eastern influences (Botigué et al. 2013). This information motivated the selection of British (GBR) and Finnish (FIN) people and Utah residents with NW EUR ancestry (CEU) (overall $n = 289$) as well as Yoruba Nigerian (YRI) people ($n = 108$) as the representatives for EUR and SSA sources, respectively. To perform sensitivity analyses of the results, random subsamples of 75 individuals from each set of EUR, NAF, and SSA samples were alternatively used as well as other data sets from 1KGP, namely the Gambian Mandenka (GWD; $n = 113$) and Sierra Leone Mende (MSL; $n = 85$) data sets. The NAF representative grouping ($n = 125$) was gathered from samples with origins in North and South Morocco, Occidental Sahara, Algeria, Tunisia, Egypt, and Libya that had previously been genotyped with the Genome-Wide Human SNP Array 6.0 (Affymetrix) (Henn et al. 2012). Using PLINK, we ensured that all samples in the reference data had a genotyping call rate >95% and excluded SNPs with a > 5% missing rate or with a Hardy–Weinberg equilibrium $P < 1 \times 10^{-6}$ in at least one population. The intersection of data sets and postfiltering (of SNPs located in mtDNA or the sex chromosomes) left a total of 114,567 SNPs for downstream analyses (data available in http://www.iter.es/wp-content/uploads/2018/09/AffyCEU1_data_from_Canary_Islanders_MBE-Guillen-Guio-et-al.2018.zip).

Admixture Inference and Population Analyses

Principal Component Analysis (PCA) was performed using PLINK v1.9 (Chang et al. 2015). ALDER v1.03 (Loh et al. 2013) was used to calculate the two-locus decay of admixture LD to test the existence of admixture and to infer the time of the most recent admixture event in Canary Islanders (assuming 33 years per generation). ALDER was first used to pretest all reference populations to determine the best pair of populations to be considered as ancestral for the estimate, avoiding the presence of long-range LD correlations with the

admixed population. Then, we were able to test for admixture and date the admixture event using FIN or CEU as surrogates of EUR and YRI, GWD, or MSL as surrogates of SSA. All of the other populations failed in the pretest.

We used ADMIXTURE v1.22 (Alexander et al. 2009), which uses a maximum likelihood estimation of individual ancestries averaged across the genome, to compute the ancestry clusters of each individual and to serve as a reference for assessing the performance of the local ancestry estimators. The ADMIXTURE calculations assumed 2 to 7 ancestral populations (K) and used a 10-times cross-validation with random seeds to estimate the best predictive K . A subsample of 100,175 SNPs from EUR, NAF, SSA, and Canary Islanders was used for this assessment. This subset resulted from excluding with PLINK those SNPs in LD (window size = 50, step = 10, pairwise r^2 threshold = 0.5) or located in regions of long-range LD according to hg19 positions (chr5: 43,964,243–51,464,244; chr6: 24,892,021–33,392,022; chr8: 7,962,590–11,962,591; and chr11: 45,043,424–57,243,424). The ADMIXTURE results were also assessed for the effects of downsampling the number of samples from the reference populations and the use of alternative SSA surrogates other than YRI (GWD and MSL). To provide further support to the ADMIXTURE results, we assessed the goodness of fit of the ADMIXTURE model to the underlying genomic data based on the patterns of haplotype sharing between individuals using badMIXTURE v0.0.0.9000 (<https://github.com/danjlawson/badMIXTURE>), whose residuals provide information on the ancestral relationships between the population groups. Statistical differences in ADMIXTURE ancestry estimates between islands and regions were assessed by Wilcoxon test.

To further explore the population history and isolation of Canary Islanders, ROHs were calculated with PLINK based on a previously described sliding window approach (Kirin et al. 2010) considering regions of 5,000 kb (ensuring a minimum density of 50 kb/SNP per window to reduce biases due to differences in local SNP densities), allowing for one heterozygous variant and up to five missing calls per window, and counting those ROHs with a minimum length of 500 kb. As a measure of the average total extent of homozygosity per population, we then calculated the average lengths of the ROHs in six categories (0.5–1, 1–2, 2–4, 4–8, 8–16, and >16 Mb). Additionally, to provide further support to the findings, we classified ROHs according to the size limits suggested by Pemberton et al. (2012) but simply stratified into ROHs ≤ 1.6 Mb and > 1.6 Mb. ROH length patterns in Canary Islanders were also explored by assessing the average number of genome regions in ROHs and by the average total length of ROHs per island. Differences in the average total ROH lengths were assessed by Wilcoxon test adjusting for the number of comparisons via Bonferroni correction ($P < 2.4 \times 10^{-3}$ considered significant).

Local Ancestry Assessments and Block Sizes

Local ancestry blocks across autosomes were inferred by using LAMP-LD v1.0 (Baran et al. 2012) and ELAI v1.0 (Guan 2014) assuming three admixing populations (EUR, NAF, and SSA). These two methods do not require SNPs to be independent and perform well with recent multiway admixtures. However,

while ELAI uses multilocus genotype data, overcoming the inherent uncertainty of the phasing step, LAMP-LD requires haplotype data for the parental populations. For that step, we used SHAPEIT v2.727 (Delaneau et al. 2014) for haplotype reconstruction under the default settings. We assumed 15 generations since admixture, which is consistent with our results as well as with assumptions made on studies in populations with similar historic scenarios (Price et al. 2007). We then compared the LAMP-LD and ELAI estimates to those provided by ADMIXTURE by Pearson correlation coefficients and by a least squares estimator of the differences in individual ancestry estimates. Ancestry block sizes were calculated excluding the centromere regions, as they are not adequately covered by genotyping arrays. In addition, given that ELAI provides ancestry dosages in a continuum (0–2), ancestry block size estimates in this case were derived from dosage approximations to the next nonnegative integer estimates (thresholds set at 0.6 and 1.4) (supplementary fig. 6, Supplementary Material online).

Local Ancestry and Diversity Deviations and Enrichment Analyses

We used the method proposed by Zhu (ZHu 2012) to estimate a Z-score statistic at each SNP position as a measure of the deviation in the local ancestries with respect to the average ancestry in the genome. This scan was conducted for the three ancestries separately, and an excess of locus-specific ancestry in a segment was considered significant if a large deviation was detected (i.e., $Z \text{ score} > |3|$, $P < 2.7 \times 10^{-3}$). In those regions showing large ancestry deviations, genetic diversity was evaluated on the basis of the mean SNP heterozygosity estimates and the mean number of subjects with SNPs from those regions that were contained in ROH stretches. The statistical significance of the differences in these regions compared with data from the whole genome was assessed by Wilcoxon rank sum tests. Enrichment analyses were conducted for all regions with ancestry deviations together based on the peak regions defined by a Z score > 3 on hg19 using the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al. 2010). A hypergeometric test was used to estimate the significance of ontology term enrichment in the unique genes extracted in those regions with respect to the set of all genes in the genome. This analysis was evaluated for annotations in two particular ontologies (human diseases and MSigDB pathways), and the significance was corrected for multiple comparisons with a false discovery rate (FDR q value).

Whole Genome Sequencing Data

Given that deviations of local ancestry often contain well-known targets of selection (i.e., *LCT* on chr2 and the HLA genes on chr6), we accessed data from deep WGS from a subset of 14 individuals (two per island) to further explore if the other regions showing deviations also harbored putative signals of selection. Briefly, DNA samples were processed with a Nextera DNA Prep kit with dual indexes following the manufacturer's recommendations (Illumina Inc., San Diego, CA). Library sizes were checked on a TapeStation 4200

(Agilent Technologies, Santa Clara, CA). The concentration of each library was determined by a Qubit dsDNA HS Assay (Thermo Fisher, Waltham, MA). As a control, a PhiX DNA sample (1%) was also sequenced with the samples. Samples were sequenced to an average depth of $36.5 \times$ (range $24\text{--}45 \times$) with paired-end 150-base reads on a HiSeq 4000 instrument (Illumina). Reads were preprocessed with bcl2fastq v2.18 and aligned to hg19 with BWA-MEM 0.7.15-r1140 (Li and Durbin 2010), and the BAMs were processed with Qualimap v2.2.1 (Okonechnikov et al. 2016), SAMtools v1.3 (Li et al. 2009), and Picard v2.1.1 (<http://broadinstitute.github.io/picard>). Variant calls were obtained with HaplotypeCaller in GATK (v3.7) (DePristo et al. 2011) following best practices workflow recommendations. These analyses were conducted in the Teide-HPC Supercomputing facility (<http://teidehpc.iter.es/en>). Sequencing data are available in http://www.iter.es/wp-content/uploads/2018/09/chr3_data_from_Canary_Islanders_MBE-Guillen-Guio-et-al.2018.vcf.tar.gz for chromosome 3 and in http://www.iter.es/wp-content/uploads/2018/09/chr13_data_from_Canary_Islanders_MBE-Guillen-Guio-et-al.2018.vcf.tar.gz for chromosome 13.

Detection of Natural Selection

WGS data from Canary Islanders were used to extract the genotypes for the four described SNPs located upstream of *LCT* that are involved in lactase persistence as reported previously in the literature: rs4988235, rs41525747, rs41380347, and rs145946881 (Ingram et al. 2009). Furthermore, biallelic SNPs identified from WGS from the chr3 and chr13 regions defined by a Z score $> |3|$ were used to ascertain the existence of candidate signals of evolutionary adaptation. To do so, we assessed Tajima's *D* and PBS (Yi et al. 2010) in 10-kb windows, registering the observed local minima for Tajima's *D* and the largest PBS scores for each chromosome region. Observed Tajima's *D* and PBS statistics were compared against a null distribution generated from 5,000 neutral simulations under a simplified demographic model (details below). Statistical significance was calculated using Hudson's "sample_stats" software (<http://home.uchicago.edu/rhudson1/source/mksamples.html>).

For the simulations, we used msHOT software (Hellenthal and Stephens 2007). As parameters, we used those from the reference model of Gutenkunst et al. (2009) for the general African (simulated population #1), European (#2) and Asian (#3) populations. These parameters were slightly modified to include a fourth population sample representing the Canary Islanders (#4). The output only included a set of 28 haplotypes representing the Canarian sample, equivalent in number to those assessed by WGS. To model the Canary Islanders' demography, we assumed a simplified model that attempted to reflect, in a broad sense, their history according to historical records and genetically supported evidence of their effective population size (N_e) in pre-European times. No attempt was made to infer the most likely parameters for the Canary Islanders. In general terms, we assumed a small constant-sized isolated population of African origin, starting from 10,000 years BP, which suffered a strong decline in population size after European colonization of the Archipelago

(simulated as an instant admixture occurring 500 years BP). The specific command line was as follows:

```
./msHOT 28 5000 -t tbs -r tbs 10000 -l 4 0 0 28 -n 1 1.68202
-n 2 3.73683 -n 3 7.29205 -n 4 3.73683 -g 2 116.010723 -g 3
160.246047 -ma x 0.881098 0.561966 0 0.881098 x 2.79746 10
0.561966 2.79746 x 0 0 1 0 x -es 0.0005 4 0.1 -ej 0.0005 5 2 -en
0.0005 4 0.068 -ej 0.0135 4 1 -ej 0.028985 3 2 -en 0.028985 2
0.287184 -eM 0.028985 28 -ej 0.197963 2 1 -en 0.303501 1 1 <
random_thetas-rhos.txt > output_file.ms
```

The initial values of theta were inferred from the comparison of human and chimp orthologous sequences for each 10-kb region using a mutation rate based on the average number of substitutions per site (with the Jukes and Cantor correction) as assessed by DnaSP v5.1 (Librado and Rozas 2009). The rho parameter was estimated from the deCODE Genetics sex-averaged recombination-rate track in the UCSC Genome Browser (<https://genome-euro.ucsc.edu>). Uncertainty in the estimates of theta and rho was considered by generating a random and normally distributed set of values for these parameters, with mean equal to the estimated value and variance equal to the mean, from which we sampled a pair of values in each simulation.

To estimate the Ne of Canary Island aborigines in pre-European times, we accessed the publicly available Guanche genome (gun011; accession number ENA: PRJEB86458) with the largest depth-of-coverage (3.9×) from Rodríguez-Varela et al. (2017). These data correspond to a male found in Tenerife for which radiocarbon dating supports an age of $1,216 \pm 27$ years BP (the oldest one analyzed with the lowest contamination levels), which predates the European colonization of the Canary Islands. We used a total of 187.7 million reads mapped to the GRCh37 human reference genome as single-end reads. SAMtools (v1.3) and BCFtools (v1.3.1) were used to generate whole-genome consensus sequences at loci with a minimum mapping quality of 30 and a minimum and maximum read-depth of 3 and 20, respectively. Finally, the Pairwise Sequentially Markovian Coalescent (PSMC) method (Li and Durbin 2011) was used on the reads with base qualities >20 to estimate the Ne of the ancestral aboriginal population assuming 33 years per generation, a mutation rate of 2.5×10^{-8} , and a range of background false negative rates (FNRs) of 0.0–0.3 for variant discovery. Considering the broad range of FNRs assumed, this analysis yielded a uniform Ne in the range of 470–560 for the Guanche population.

Supplementary Material

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

Acknowledgments

This work was supported by Instituto de Salud Carlos III (grant numbers PI11/00623, PI14/00844, PI17/00610) cofinanced by the European Regional Development Funds, “A way of making Europe” from the European Union; by Ministerio de Ciencia, Innovación y Universidades (RTC-2017-6471-1; MINECO/AEI/FEDER, UE); and by the agreement OA17/008 with Instituto Tecnológico y de Energías

Renovables (ITER) to strengthen scientific and technological education, training, research, development and innovation in Genomics, Personalized Medicine and Biotechnology. S.A. was funded by the Spanish MINECO CGL2014-58526-P and the Basque Government—Research Groups, IT1138-16. B.G.G. was supported by a fellowship from the Canarian Agency for Research, Innovation and Information Society (ACIISI, grant number TESIS2015010057) cofunded by the European Social Fund (ESF). A.D.U. was supported by a fellowship from the Spanish Ministry of Education, Culture and Sports (grant number FPU16/01435). Teide-HPC thanks to INP-2011-0063-PCT-430000-ACT (INNPLANTA program) from the Spanish Ministry of Economy and Competitiveness. The authors thank Servicio de Apoyo Informático a la Investigación (SAII) from Universidad de La Laguna for the HPC support. The content of this publication is solely the responsibility of the authors and does not necessarily reflect the views or policies of funding agencies. Finally, the authors would like to thank the anonymous reviewers for their advices to improve this manuscript.

References

- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664.
- Ara MF. 1959. Algunas observaciones acerca de la Antropología de las poblaciones prehistórica y actual de Gran Canaria. Las Palmas de Gran Canaria: *El Museo Canario*.
- Aragón-Sánchez J, García-Rojas A, Lázaro-Martínez JL, Quintana-Marrero Y, Maynar-Moliner M, Rabellino M, Hernández-Herrero MJ, Cabrera-Galván JJ. 2009. Epidemiology of diabetes-related lower extremity amputations in Gran Canaria, Canary Islands (Spain). *Diabetes Res Clin Pract.* 86(1):e6–e8.
- Arauna LR, Mendoza-Revilla J, Mas-Sandoval A, Izaabel H, Bekada A, Benhamamouch S, Fadhlaoui-Zid K, Zalloua P, Hellenthal G, Comas D. 2017. Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa. *Mol Biol Evol.* 34(2):318–329.
- Arco MC, Navarro JF. 1987. Historia popular de Canarias. Vol. 1. Los aborígenes. S/C de Tenerife: Centro de la Cultura Popular Canaria.
- Baran Y, Pasiński B, Sankararaman S, Torgerson DG, Gignoux C, Eng C, Rodríguez-Cintron W, Chapela R, Ford JG, Avila PC, et al. 2012. Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* 28(10):1359–1367.
- Ben Halima Y, Kefi R, Sazzini M, Giuliani C, De Fanti S, Nouali C, Nagara M, Mengozzi G, Elouej S, Abid A, et al. 2017. Lactase persistence in Tunisia as a result of admixture with other Mediterranean populations. *Genes Nutr.* 12:20.
- Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. 2004. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet.* 74(6):1111–1120.
- Berthelot S. 1978. Etnografía y anales de la conquista de las islas Canarias. Santa Cruz de Tenerife: Goya Ediciones.
- Botigué LR, Henn BM, Gravel S, Maples BK, Gignoux CR, Corona E, Atzmon G, Burns E, Ostrer H, Flores C, et al. 2013. Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc Natl Acad Sci U S A.* 110(29):11791–11796.
- Bueno H, Hernández R, Hernández AV. 2008. Type 2 Diabetes mellitus and cardiovascular disease in Spain: a narrative review. *Rev Esp Cardiol Supl.* 8:50C–58C.
- Cabrera de León A, González DA, Méndez LIP, Aguirre-Jaime A, del Cristo Rodríguez Pérez M, Coello SD, Trujillo IC. 2004. Leptin and altitude in the cardiovascular diseases. *Obes Res.* 12(9):1492–1498.

- Cabrera de León A, Rodríguez-Pérez M, del C, del Castillo-Rodríguez JC, Brito-Díaz B, Pérez-Méndez LI, Muros de Fuentes M, Almeida-González D, Batista-Medina M, Aguirre-Jaime A. 2006. [Coronary risk in the population of the Canary Islands, Spain, using the Framingham function]. *Med Clin*. 126(14):521–526.
- Campbell H, Carothers AD, Rudan I, Hayward C, Biloglav Z, Barac L, Pericic M, Janicijevic B, Smolej-Narancic N, Polasek O, et al. 2007. Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. *Hum Mol Genet*. 16(2):233–241.
- Castella M, Pujol R, Callén E, Trujillo JP, Casado JA, Gille H, Lach FP, Auerbach AD, Schindler D, Benítez J, et al. 2011. Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood* 117(14):3759–3769.
- Ceballos FC, Hazelhurst S, Ramsay M. 2018. Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics* 19(1):106.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.
- Chimusa ER, Zaitlen N, Daya M, Möller M, van Helden PD, Mulder NJ, Price AL, Hoal EG. 2014. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. *Hum Mol Genet*. 23(3):796–809.
- Crosby AW. 1999. Imperialismo ecológico. La expansión biológica de Europa, 900-1900. Barcelona: Crítica.
- de Abreu Galindo J, Cioranescu A. 1977. Historia de la conquista de las Siete Islas de Canaria. Santa Cruz de Tenerife: Goya Ediciones.
- de Castro JB. 1987. Quantitative analysis of the molar-size sequence in human prehistoric populations of the Canary Isles. *Arch Oral Biol*. 32(2):81–86.
- Delaneau O, Marchini J; 1000 Genomes Project Consortium. 2014. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun*. 5:3934.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 43(5):491–498.
- Falkenburger F. 1942. Ensayo de una nueva clasificación craneológica de los antiguos habitantes de Canarias. *Actas y Memorias de la Sociedad Española de Antropología, Etnografía y Prehistoria*. XVII: 5–52.
- Ferreira MAR, Matheson MC, Duffy DL, Marks GB, Hui J, Le Souëf P, Danoy P, Baltic S, Nyholt DR, Jenkins M, et al. 2011. Identification of *IL6R* and chromosome 11q13.5 as risk loci for asthma. *Lancet* 378(9795):1006–1014.
- Flores C, Larruga JM, González AM, Hernández M, Pinto FM, Cabrera VM. 2001. The origin of the Canary Island Aborigines and their contribution to the modern population: a molecular genetics perspective. *Curr Anthropol*. 42(5):749.
- Flores C, Maca-Meyer N, Pérez JA, González AM, Larruga JM, Cabrera VM. 2003. A predominant European ancestry of paternal lineages from Canary Islanders. *Ann Hum Genet*. 67(Pt 2):138–152.
- Fregel R, Gomes V, Gusmão L, González AM, Cabrera VM, Amorim A, Larruga JM. 2009. Demographic history of Canary Islands male gene-pool: replacement of native lineages by European. *BMC Evol Biol*. 9:181.
- Galanter JM, Gignoux CR, Torgerson DG, Roth LA, Eng C, Oh SS, Nguyen EA, Drake KA, Huntsman S, Hu D, et al. 2014. Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: the Genes-environments & Admixture in Latino Americans study. *J Allergy Clin Immunol*. 134(2):295–305.
- García-Villarreal L, Daniels S, Shaw SH, Cotton D, Galvin M, Geskes J, Bauer P, Sierra-Hernández A, Buckler A, Tugores A. 2000. High prevalence of the very rare Wilson disease gene mutation Leu708Pro in the Island of Gran Canaria (Canary Islands, Spain): a genetic and clinical study. *Hepatology* 32(6):1329–1336.
- Ghani M, Reitz C, Cheng R, Vardarajan BN, Jun G, Sato C, Naj A, Rajbhandary R, Wang L-S, Valladares O, et al. 2015. Association of long runs of homozygosity with Alzheimer disease among African American individuals. *JAMA Neurol*. 72(11):1313–1323.
- Guan Y. 2014. Detecting structure of haplotypes and local ancestry. *Genetics* 196(3):625–642.
- Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, Karthikeyan S, Iles L, Pollard MO, Choudhury A, et al. 2015. The African Genome Variation Project shapes medical genetics in Africa. *Nature* 517(7534):327–332.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*. 5(10):e1000695.
- He M, Xu M, Zhang B, Liang J, Chen P, Lee J-Y, Johnson TA, Li H, Yang X, Dai J, et al. 2015. Meta-analysis of genome-wide association studies of adult height in East Asians identifies 17 novel loci. *Hum Mol Genet*. 24(6):1791–1800.
- Hellenthal G, Stephens M. 2007. msHOT: modifying Hudson's ms simulator to incorporate crossover and gene conversion hotspots. *Bioinformatics* 23(4):520–521.
- Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, et al. 2012. Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet*. 8(1):e1002397.
- Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM. 1991. Common west African HLA antigens are associated with protection from severe malaria. *Nature* 352(6336):595–600.
- Hooton EA. 1970. The ancient inhabitants of the Canary Islands. New York: Kraus Reprint.
- Hu M. 2012. Positive natural selection in the human genome. University of Cambridge.
- Ingram CJE, Raga TO, Tarekegn A, Browning SL, Elamin MF, Bekele E, Thomas MG, Weale ME, Bradman N, Swallow DM. 2009. Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. *J Mol Evol*. 69(6):579–588.
- Jakobsdottir J, van der Lee SJ, Bis JC, Chouraki V, Li-Kroeger D, Yamamoto S, Grove ML, Naj A, Vronska M, Salazar JL, et al. 2016. Rare functional variant in *TM2D3* is associated with late-onset Alzheimer's disease. *PLoS Genet*. 12(10):e1006327.
- Juliá-Serdá G, Cabrera-Navarro P, Acosta-Fernández O, Martín-Pérez P, Losada-Cabrera P, García-Bello MA, Carrillo-Díaz T, Antó-Boqué J. 2011. High prevalence of asthma and atopy in the Canary Islands, Spain. *Int J Tuberc Lung Dis*. 15(4):536–541.
- Kiel DP, Demissie S, Dupuis J, Lunetta KL, Murabito JM, Karasik D. 2007. Genome-wide association with bone mass and geometry in the Framingham Heart Study. *BMC Med Genet*. 8(Suppl 1):S14.
- Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. 2010. Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5(11):e13996.
- Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, Kucherlapati R, Malhotra AK. 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci U S A*. 104(50):19942–19947.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.
- 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.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452.
- Lobo-Cabrera M. 1993. La esclavitud en Fuerteventura en los Siglos XVI y XVII. *V Jornadas de estudios sobre Fuerteventura y Lanzarote* 1:13–40.
- Loh P-R, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. 2013. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193(4):1233–1254.

- Lorenzo V, Alvarez A, Torres A, Torregrosa V, Hernández D, Salido E. 2006. Presentation and role of transplantation in adult patients with type 1 primary hyperoxaluria and the I244T AGXT mutation: single-center experience. *Kidney Int.* 70(6):1115–1119.
- Lorenzo V, Boronat M, Saavedra P, Rufino M, Maceira B, Novoa FJ, Torres A. 2010. Disproportionately high incidence of diabetes-related end-stage renal disease in the Canary Islands. An analysis based on estimated population at risk. *Nephrol Dial Transplant.* 25(7):2283–2288.
- Lorenzo V, Torres A, Salido E. 2014. Primary hyperoxaluria. *Nefrología* 34(3):398–412.
- Maca-Meyer N, Arnay M, Rando JC, Flores C, González AM, Cabrera VM, Larruga JM. 2004. Ancient mtDNA analysis and the origin of the Guanches. *Eur J Hum Genet.* 12(2):155–162.
- Maca-Meyer N, Cabrera VM, Arnay M, Flores C, Fregel R, González AM, Larruga JM. 2005. Mitochondrial DNA diversity in 17th–18th century remains from Tenerife (Canary Islands). *Am J Phys Anthropol.* 127(4):418–426.
- Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, Margulies DM, Loscalzo J, Kohane IS. 2016. Genetic misdiagnoses and the potential for health disparities. *N Engl J Med.* 375(7):655–665.
- Marcelino-Rodríguez I, Elosua R, del Cristo Rodríguez Pérez M, Fernández-Bergés D, Guembe MJ, Alonso TV, Félix FJ, González DA, Ortiz-Marrón H, Rigo F, et al. 2016. On the problem of type 2 diabetes-related mortality in the Canary Islands, Spain. The DARIOS Study. *Diabetes Res Clin Pract.* 111:74–82.
- Marciniak S, Perry GH. 2017. Harnessing ancient genomes to study the history of human adaptation. *Nat Rev Genet.* 18(11):659–674.
- Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528(7583):499–503.
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 28(5):495–501.
- Moltke I, Grarup N, Jørgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, Korneliusen TS, Andersen MA, Nielsen TS, Krarup NT, et al. 2014. A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes. *Nature* 512(7513):190–193.
- Navarro Mederos JF. 1997. *Arqueología de las Islas Canarias. Espacio, tiempo y forma, Serie I, Prehistoria y arqueología* 10:447–478.
- Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32(2):292–294.
- Oksenberg JR, Barcellos LF, Cree BAC, Baranzini SE, Bugawan TL, Khan O, Lincoln RR, Swerdlin A, Mignot E, Lin L, et al. 2004. Mapping multiple sclerosis susceptibility to the *HLA-DR* locus in African Americans. *Am J Hum Genet.* 74(1):160–167.
- Onrubia Pintado J. 1987. Les cultures préhistoriques des Îles Canaries, état de la question. *Anthropologie* 91:653–678.
- Patin E, Lopez M, Grollemund R, Verdu P, Harmant C, Quach H, Laval G, Perry GH, Barreiro LB, Froment A, et al. 2017. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science* 356(6337):543–546.
- Pelak K, Goldstein DB, Walley NM, Fellay J, Ge D, Shianna KV, Gumbs C, Gao X, Maia JM, Cronin KD, et al. 2010. Host determinants of HIV-1 control in African Americans. *J Infect Dis.* 201(8):1141–1149.
- Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. 2012. Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet.* 91(2):275–292.
- Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, Srinivasan BS, Barsh GS, Myers RM, Feldman MW, et al. 2009. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* 19(5):826–837.
- Pino-Yanes M, Corrales A, Basaldúa S, Hernández A, Guerra L, Villar J, Flores C. 2011. North African influences and potential bias in case-control association studies in the Spanish population. *PLoS One* 6(3):e18389.
- Price AL, Patterson N, Yu F, Cox DR, Waliszewska A, McDonald GJ, Tandon A, Schirmer C, Neubauer J, Bedoya G, et al. 2007. A genomewide admixture map for Latino populations. *Am J Hum Genet.* 80(6):1024–1036.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.
- Raj T, Kuchroo M, Replogle JM, Raychaudhuri S, Stranger BE, De Jager PL. 2013. Common risk alleles for inflammatory diseases are targets of recent positive selection. *Am J Hum Genet.* 92(4):517–529.
- Ranciaro A, Campbell MC, Hirbo JB, Ko W-Y, Froment A, Anagnostou P, Kotze MJ, Ibrahim M, Nyambo T, Omar SA, et al. 2014. Genetic origins of lactase persistence and the spread of pastoralism in Africa. *Am J Hum Genet.* 94(4):496–510.
- Rando JC, Cabrera VM, Larruga JM, Hernández M, González AM, Pinto F, Bandelt HJ. 1999. Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet.* 63(Pt 5):413–428.
- Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt HJ. 1998. Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet.* 62(Pt 6):531–550.
- Rodríguez-Esparragón F, López-Fernández JC, Buset-Ríos N, García-Bello MA, Hernández-Velázquez E, Cappiello L, Rodríguez-Pérez JC. 2017. Paraoxonase 1 and 2 gene variants and the ischemic stroke risk in Gran Canaria population: an association study and meta-analysis. *Int J Neurosci.* 127(3):191–198.
- Rodríguez-Varela R, Günther T, Krzewińska M, Storå J, Gillingwater TH, MacCallum M, Arsuaga JL, Dobney K, Valdiosera C, Jakobsson M, et al. 2017. Genomic analyses of pre-european conquest human remains from the canary islands reveal close affinity to modern North Africans. *Curr Biol.* 27(21):3396–3402.e5.
- Rudan I, Smolej-Narancic N, Campbell H, Carothers A, Wright A, Janicijevic B, Rudan P. 2003. Inbreeding and the genetic complexity of human hypertension. *Genetics* 163(3):1011–1021.
- Sánchez-Lerma B, Morales-Chirivella FJ, Peñuelas I, Blanco Guerra C, Mesa Lugo F, Aguinaga-Ontoso I, Guillén-Grima F. 2009. High prevalence of asthma and allergic diseases in children aged 6 to [corrected] 7 years from the Canary Islands. [corrected]. *J Investig Allergol Clin Immunol.* 19:383–390.
- Sanchez-Mazas A, Černý V, Di D, Buhler S, Podgorná E, Chevallier E, Brunet L, Weber S, Kervaire B, Testi M, et al. 2017. The *HLA-B* landscape of Africa: signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection. *Mol Ecol.* 26(22):6238–6252.
- Santana A, Salido E, Torres A, Shapiro LJ. 2003. Primary hyperoxaluria type 1 in the Canary Islands: a conformational disease due to I244T mutation in the P11L-containing alanine: glyoxylate aminotransferase. *Proc Natl Acad Sci U S A.* 100(12):7277–7282.
- Schrider DR, Kern AD. 2017. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol Biol Evol.* 34(8):1863–1877.
- Seldin MF, Shigeta R, Villoslada P, Selmi C, Tuomilehto J, Silva G, Belmont JW, Klareskog L, Gregersen PK. 2006. European population substructure: clustering of northern and southern populations. *PLoS Genet.* 2(9):e143.
- Shriner D, Adeyemo A, Rotimi CN. 2011. Joint ancestry and association testing in admixed individuals. *PLoS Comput Biol.* 7(12):e1002325.
- Suárez JJ, Rodríguez F, Quintero CL. 1988. Historia popular de Canarias. Vol. 2. Conquista y colonización. Santa Cruz de Tenerife: Gobierno de Canarias. Centro de Cultura Popular Canaria.
- Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH-Y, et al. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature* 526(7571):75–81.
- Thomas R, Apps R, Qi Y, Gao X, Male V, O’Hugin C, O’Connor G, Ge D, Fellay J, Martin JN, et al. 2009. HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of *HLA-C*. *Nat Genet.* 41(12):1290–1294.

- Thomsen H, Chen B, Figlioli G, Elisei R, Romei C, Cipollini M, Cristaudo A, Bambi F, Hoffmann P, Herms S, et al. 2016. Runs of homozygosity and inbreeding in thyroid cancer. *BMC Cancer* 16:227.
- Watkins WS, Xing J, Huff C, Witherspoon DJ, Zhang Y, Perego UA, Woodward SR, Jorde LB. 2012. Genetic analysis of ancestry, admixture and selection in Bolivian and Totonac populations of the New World. *BMC Genet.* 13:39.
- Wölfel DJ. 1930. Sind die Ureinwohner der Kanaren ausgestorben? Eine siedlungsgeschichtliche Untersuchung, ausgeführt mit Hilfe der Notgemeinschaft der Deutschen Wissenschaft. *Zeitschrift Für Ethnol.* 62:282–302.
- Wright J. 2007. The trans-Saharan slave trade. London: Routledge.
- Xue Y, Mezzavilla M, Haber M, McCarthy S, Chen Y, Narasimhan V, Gilly A, Ayub Q, Colonna V, Southam L, et al. 2017. Enrichment of low-frequency functional variants revealed by whole-genome sequencing of multiple isolated European populations. *Nat Commun.* 8:15927.
- Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliusson TS, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329:75–78.
- Zhou Q, Zhao L, Guan Y. 2016. Strong selection at MHC in Mexicans since admixture. *PLoS Genet.* 12(2):e1005847.
- Zhu X. 2012. The analysis of ethnic mixtures. *Methods Mol Biol.* 850:465–481.