Genomic Analyses of Pre-European Conquest Human Remains from the Canary Islands Reveal Close Affinity to Modern North Africans

Highlights
- The first genome-wide data from the Guanches confirm a North African origin
- The Guanches were genetically most similar to modern North African Berbers
- Modern inhabitants of Gran Canaria carry an estimated 16%–31% Guanche autosomal ancestry

Authors
Ricardo Rodríguez-Varela, Torsten Günther, Maja Krzewińska, ..., Mattias Jakobsson, Anders Götherström, Linus Girdland-Flink

Correspondence
ricardo.rodriguez.varela@arklab.su.se (R.R.-V.), linusflink@hotmail.com (L.G.-F.)

In Brief
Rodríguez-Varela et al. report the first genome-wide data from the aboriginals of the Canary Islands, the Guanches, confirming the long-held hypothesis that the Guanches originated from a North African Berber-like population and showing that modern inhabitants of Gran Canaria carry an estimated 16%–31% Guanche autosomal ancestry.
Genomic Analyses of Pre-European Conquest Human Remains from the Canary Islands Reveal Close Affinity to Modern North Africans

Ricardo Rodríguez-Varela,1,2,11,* Torsten Günter,3 Maja Krzewinska,1 Jan Stora˚,1 Thomas H. Gillingwater,4 Malcolm MacCallum,4 Juan Luis Arsuaga,2,5 Keith Dobney,6,7,8 Cristina Valdiosera,2,9 Mattias Jakobsson,3 Anders Götherström,1 and Linus Girdland-Flink1,10,*

1Department of Archaeology and Classical Studies, Stockholm University, 10691 Stockholm, Sweden
2Centro Mixto, Universidad Complutense de Madrid–Instituto de Salud Carlos III de Evolución y Comportamiento Humano, 28029 Madrid, Spain
3Department of Organismal Biology, Uppsala University, 75236 Uppsala, Sweden
4Anatomy, Edinburgh Medical School: Biomedical Sciences, University of Edinburgh, Edinburgh EH4 9AG, UK
5Departamento de Paleontología, Facultad de Ciencias Geológicas, Universidad Complutense de Madrid, 28040 Madrid, Spain
6Department of Archaeology, School of Geosciences, University of Aberdeen, St. Mary’s, Aberdeen AB24 3UF, UK
7Department of Archaeology, Classics and Egyptology, University of Liverpool, Liverpool L69 7WZ, UK
8Department of Archaeology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
9Department of Archaeology and History, La Trobe University, Melbourne, VIC 3086, Australia
10Research Centre in Evolutionary Anthropology and Palaeoecology, School of Natural Sciences and Psychology, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK
11Lead Contact
*Correspondence: ricardo.rodriguez.varela@arklab.su.se (R.R.-V.), linusflink@hotmail.com (L.G.-F.)

https://doi.org/10.1016/j.cub.2017.09.059

SUMMARY

The origins and genetic affinity of the aboriginal inhabitants of the Canary Islands, commonly known as Guanches, are poorly understood. Though radiocarbon dates on archaeological remains such as charcoal, seeds, and domestic animal bones suggest that people have inhabited the islands since the 5th century BCE [1–3], it remains unclear how many times, and by whom, the islands were first settled [4, 5]. Previously published ancient DNA analyses of uniparental genetic markers have shown that the Guanches carried common North African Y chromosome markers (E-M81, E-M78, and J-M267) and mitochondrial lineages such as U6b, in addition to common Eurasian haplogroups [6–8]. These results are in agreement with some linguistic, archaeological, and anthropological data indicating an origin from a North African Berber-like population [1, 4, 9]. However, to date there are no published Guanche autosomal genomes to help elucidate and directly test this hypothesis. To resolve this, we generated the first genome-wide sequence data and mitochondrial genomes from eleven archaeological Guanche individuals originating from Gran Canaria and Tenerife. Five of the individuals (directly radiocarbon dated to a time transect spanning the 7th–11th centuries CE) yielded sufficient autosomal genome coverage (0.21 to 3.93) for population genomic analysis. Our results show that the Guanches were genetically similar over time and that they display the greatest genetic affinity to extant Northwest Africans, strongly supporting the hypothesis of a Berber-like origin. We also estimate that the Guanches have contributed 16%–31% autosomal ancestry to modern Canary Islanders, here represented by two individuals from Gran Canaria.

RESULTS AND DISCUSSION

We processed twelve samples from twelve different individuals, of which eleven yielded genome-wide sequence data (Table 1). We sampled human teeth only from intact skulls. The ancient individuals, collected from caves on Tenerife and Gran Canaria (the two largest islands of the Canary archipelago) (Figure 1), were donated to the skull collection housed in the AMEU (Anatomical Museum Edinburgh University) in the 19th century (see Experimental Model and Subject Details).

The average read length (DNA fragmentation) and patterns of cytosine deamination, clustering at a significantly higher frequency at fragment termini, are consistent with expectations for ancient DNA (Figure S1A) [10]. Mitochondrial contamination estimates [11] for the five individuals included in further autosomal genomic analysis range from point estimates of 0.53% to 5.97%, showing that contamination is largely negligible (Table 1). The five individuals with the highest autosomal genome coverage (0.21 to 3.93) were directly radiocarbon dated at the Svedberg Laboratory (Uppsala University) (Figure S1B) and span approximately 400 calendar years from the 7th to the 11th centuries CE (Table 1). All individuals predate the European colonization of the Canary Islands (15th century) and one predated the...
Muslim conquest of the Maghreb (7th–8th centuries), events that have had significant impact on the gene pool of the Canary Islands and North Africa, respectively (Figure S1 B) [6, 12]. Hence, our sample is a good representative of the pre-conquest aboriginals.

Analysis of Uniparental Genetic Markers

The mitochondrial genome coverage for the eleven Guanche individuals ranges from 3.4× to 931×. The number of SNPs that support each haplotype varies between 38 and 58 and are reported as deviations from the Reconstructed Sapiens Reference Sequence (RSRS) [13] (Table S1). All positions supporting the predicted haplogroup calls and mutations missing for the absolute haplogroup assignment are listed in Table S1. We found six different mitochondrial haplogroups in eleven individuals (Tables 1 and S1). The Guanches analyzed here carried mitochondrial lineages such as J1c3, H2a, U6b, L3b1a, and T2c1d2 that are common across West Eurasia and/or North Africa [14] (Tables 1 and S1) and are consistent with previous studies on ancient Guanche mitochondrial DNA [6, 7]. Two individuals from Tenerife (gun001 and gun012) and one from Gran Canaria (gun013) carried the U6b1a haplotype, which is hypothesized to be endemic to and a founder lineage of the Canary Islands (Tables 1 and S1) [15]. We also found the H1cf haplotype in one individual from Tenerife (gun002), which is defined by a mutation at position 16260T (Table S1). H1-16260 is also considered a founder lineage of the Canary Islands, for two reasons: (1) it is present in all modern Canary populations [15] and rare outside the Canary Islands (found previously in only a single Algerian individual) [7, 16]; and (2) it was found in pre-Hispanic populations from Tenerife [6], La Palma [7], El Hierro [15], and La Gomera [17]. The three males from whom haplogroup-defining Y chromosome SNPs were retrieved carried the E1b1b1b1a1 (E-M183) haplotype (a major sub-clone of the

Figure 1. Map of the Canary Islands

Map template is modified from Google Earth (https://www.google.com/earth).
haplogroup E1b1b1b1, defined by the derived M81 marker), which is again consistent with previous analyses of ancient Guanches [5]. This haplogroup is ubiquitous across modern North African populations and particularly common in Berber-speaking populations of North Africa [8, 18]. The derived mutations supporting the Y chromosome haplogroups are listed in Data S1, sheet 1.

Population Genomic Analysis of Autosomal DNA

A principal component analysis (PCA) of the five samples with the highest autosomal genome coverage, performed using genome-wide autosomal SNPs overlapping with Human Origins (HO) data [19, 20], reveals close affinity to modern Northwest African populations such as Tunisians and Algerians, but with a tendency (especially for individuals from Gran Canaria) to occupy a space outside modern Northwest African variation, closer to Europeans (Figures 2 and S2). However, outgroup f3 statistics [19] suggest that the Guanches share more genetic drift with non-African test populations than with African test populations, including Northwest African populations of Berber origin (Data S1, sheet 2). This observation is inconsistent with the PCA and the uniparental genetic marker data, indicating that the outgroup f3 statistic may be misleading, possibly due to the complex history of recent sub-Saharan admixture events in North African populations [12, 21] and the sensitivity of the f3 estimator to such patterns. This issue seems to extend to other statistics based on allele frequency correlations such as the D statistic [19] since D(Outgroup, Guanches; North African, Sardinian/Anatolian farmer) consistently produces highly significant positive values of D (Z > 4), which would imply a closer relationship between Guanches and Sardinians and Anatolian farmers than between Guanches and North African populations (Data S1, sheet 3).

In order to resolve this, we used different statistics to measure genetic similarity between Guanches and modern populations, as well as between different modern North African populations and other populations (Table S2 and Quantification and Statistical Analysis). Both outgroup f3 statistics and average pairwise differences identify Sardinians as the population sharing the most drift with both modern North Africans and Guanches (Table S2). This is in stark contrast to our expectation that North Africans would cluster with other North Africans. However, it replicates recent findings [21] showing that both ancient and modern Egyptians share more genetic drift with European populations than they do with other North African groups, while still showing evidence of strong genetic affinity to one another in other types of analyses. This behavior might be due to varying degrees of admixture from highly divergent sources (e.g., sub-Saharan populations) into different populations, which can affect the interpretation of f3 statistics [22]. Therefore, we conclude that these statistics are not suitable to identify the modern population most similar to Guanches in the sense that we intend it for this study (Quantification and Statistical Analysis). In contrast, our extended analyses using FST (fixation index) [23, 24] and f2 [25] identify different North African and Near Eastern populations as most similar to modern North African populations and Guanches, results that are consistent with the PCA and the uniparental genetic analysis (Table S2; Figures 2 and S2). This finding is in agreement with anthropological, archaeological, linguistic [1, 4, 9, 26], and uniparental genetic data [6–8] and adds to a growing body of new evidence suggesting that some ancient domesticated plants [27] and animals [28] from the Canary Islands originated from North Africa.

The Guanches’ Berber-like affinity is further supported by ADMIXTURE [29] analysis (Figures 3 and S3), where Guanches largely behave like modern Berbers across all values of K. At K = 10, a Northwest African-specific ancestry component makes up the greatest amount of autosomal ancestry in the Guanche and Berber populations in the HO dataset, such as the Mozabite and Saharawi. It is also ubiquitous across other Northwest African populations with Berber ancestry, such as Algerians and Tunisians, consistent with the PCA results. This ancestry component is also represented in present-day Canary Islanders and at a low proportion in some South European populations (Figures 3 and S3). Interestingly, it is also shared by Middle Eastern populations, including some Natufians (Figure 3). Y chromosome E1b1b1b haplotypes (though not M183 variants) were also common in Natufians (circa 11,000 BCE) and pre-pottery Neolithic male individuals from the Levant (circa 7,000 BCE), suggesting some affinity to North Africans [30].

The results of the ADMIXTURE analysis furthermore show that the Guanches carried early European farmer (EEF)-like ancestry; this ancestry component is widespread (though at varying proportions) in present-day North Africans and Middle Easterners but rare or largely absent in some Berber populations (Figure 3). The EEF component is strongly associated with early Neolithic farmers from Anatolia and Europe (as well as present-day Sardinians), hinting at a possible link between present-day North Africans and the expansion of Neolithic culture through the
cultures prior to the European colonization in the 15th century were in at least sporadic contact with other peoples and finds of Phoenician-Punic amphora in Buena Vista (Lanzarote), indicating that the islands (and local islanders) and Roman amphora fragments retrieved from El Bebedero, North Africans than Northern Europeans [34]. However, other European source populations provided intermediate values, such as a Basque population from Spain (27.2%, SE 4.92%) (Table S3). These results are smaller than the estimates of genetic contribution obtained using mitochondrial DNA (42%–73%) [6] but higher than estimates based on Y chromosome data (17%) [8], suggesting that the male Guanche contribution to modern-day Canary Islanders is lower than the female Guanche contribution. This may have been due to the violent process of colonization by the Europeans, which led to the death of a proportionally greater number of Guanche males than females [8, 35]. However, we caution that Canary Islanders in the HO dataset may not be representative of all Canary Islands, not only because of the small sample size (n = 2) but also because previous studies have shown significant differences in the relative proportion of mitochondrial ancestry between the different islands [17].

Inferring Phenotypes

Lastly, we obtained phenotypically informative SNPs from the five individuals with the highest genome coverage; however, only individual gun011 yielded high enough coverage to infer genotypes (Table S4). We relied on the HirisPlex and 8-plex prediction systems, which are based on 24 and 8 SNPs respectively, for skin, hair, and eye color prediction [36, 37], as well as 3 SNPs involved in lactose tolerance [38–40] (Table S4). The results reveal that this individual likely was lactose intolerant and had brown eyes, dark hair, and light or medium skin color. These results are similar for the other individuals where SNP information is available, albeit with lower coverage, suggesting that—at least for this sample of Guanches—the dominating phenotype was lactose intolerant, dark hair, light or medium skin color, and brown eyes (Table S4).

In summary, by generating the first genome-wide sequence data from several individuals of the aboriginal population of the Canary Islands, the Guanches, we confirm the long-held hypothesis that they were genetically most similar to modern Berber populations from Northwest Africa [1, 4, 9]. Importantly, our data represent a genomic time transect spanning from the early 7th century to the early 11th century CE, allowing us to explore temporal structure and to test the extent to which the European conquest of the Canary Islands (15th century CE) replaced Guanche ancestry. Although we find no clear indication that the Muslim conquest of the Maghreb (mid to late 7th century) significantly impacted the genetic variation and ancestry of the Guanches, we caution that this does not reject the possibility of limited contribution to the Guanche gene pool from either non-African or even African populations throughout the studied time period. However, overall our data suggest that the Guanches on Tenerife and Gran Canaria were genetically similar during the 7th–11th centuries CE. On the other hand, we show that the European conquest led to a decline in the overall degree
of Guanche autosomal ancestry and provide an estimate that modern Canary Islanders (as represented by two individuals on the HO panel) carry between 16% and 31% autosomal ancestry derived from the Guanches (Table S3).

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - DNA extraction
  - Library construction
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Sequence processing and alignment
  - SNP genotype data
  - Principal Component Analysis
  - ADMIXTURE analysis
  - Affinity to modern-day North African populations
  - Mitochondrial DNA analysis
  - Y chromosome analysis
  - Damage patterns
  - Mitochondrial DNA authentication
  - Outgroup f3 statistics
  - D statistics
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, four tables, and the extended dataset and can be found with this article online at https://doi.org/10.1016/j.cub.2017.09.059.

AUTHOR CONTRIBUTIONS


ACKNOWLEDGMENTS

This work has been funded by Swedish Research Council grant 2013-1905. R.R.-V. was funded by the Knut and Alice Wallenberg Foundation (1000 Ancient Genome Project) and a Fundación Ramón Areces and Fundación Atapuerca postdoctoral scholarship. L.G.-F. was supported by a LJMU Collaborative Research Fellowship. Computations were performed using UPPMAX (Uppsala Multidisciplinary Centre for Advanced Computational Science) research grants. Computations were performed using UPPMAX (Uppsala Multidisciplinary Centre for Advanced Computational Science) research grants. R.R.-V. was funded by the Knut and Alice Wallenberg Foundation (1000 Ancient Genome Project) and a Fundación Ramón Areces and Fundación Atapuerca postdoctoral scholarship. L.G.-F. was supported by a LJMU Collaborative Research Fellowship. Computations were performed using UPPMAX (Uppsala Multidisciplinary Centre for Advanced Computational Science) research grants.

Received: July 28, 2017
Revised: September 7, 2017
Accepted: September 27, 2017
Published: October 26, 2017; corrected online: May 1, 2018

REFERENCES

34. Botigu


### STAR METHODS

#### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biological Samples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gun001</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 52</td>
</tr>
<tr>
<td>gun011</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 57</td>
</tr>
<tr>
<td>gun005</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 47</td>
</tr>
<tr>
<td>gun012</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 62</td>
</tr>
<tr>
<td>gun008</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 46</td>
</tr>
<tr>
<td>gun014</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 61</td>
</tr>
<tr>
<td>gun006</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 48</td>
</tr>
<tr>
<td>gun002</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 49</td>
</tr>
<tr>
<td>gun007</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 45</td>
</tr>
<tr>
<td>gun013</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 60</td>
</tr>
<tr>
<td>gun004</td>
<td>Anatomical Museum Edinburgh University (AMEU)</td>
<td>Cat#XXVI A 53</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteinase K</td>
<td>VWR Sweden</td>
<td>Cat#1.24568.0100</td>
</tr>
<tr>
<td>EDTA buffer solution pH 8.0 (0.5 mol/l) for biotechnology, sterile</td>
<td>VWR Sweden</td>
<td>Car#E522-100ML</td>
</tr>
<tr>
<td>Sodium acetate</td>
<td>Sigma Aldrich Sweden</td>
<td>Cat#S2889</td>
</tr>
<tr>
<td>Guanidine hydrochloride</td>
<td>Sigma Aldrich Sweden</td>
<td>Cat#50933</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>Sigma-Aldrich Sweden</td>
<td>Cat#67-63-0</td>
</tr>
<tr>
<td>Tween-20</td>
<td>Sigma-Aldrich Sweden</td>
<td>Cat#9005-64-5</td>
</tr>
<tr>
<td>Sodium phosphate PH 6.0</td>
<td>VWR Sweden</td>
<td>Cat#101447-426</td>
</tr>
<tr>
<td>ATP</td>
<td>Fermentas/Thermo Scientific</td>
<td>Cat#R0441</td>
</tr>
<tr>
<td>T4 Polynucleotide Kinase</td>
<td>Thermo Scientific</td>
<td>Cat#EK0032</td>
</tr>
<tr>
<td>T4 DNA Polymerase</td>
<td>Fermentas/Thermo Scientific</td>
<td>Cat#EP0062</td>
</tr>
<tr>
<td>Bst polymerase (supplied with 10X ThermoPol reaction buffer)</td>
<td>NEB/BioNordika</td>
<td>Cat#M0275S</td>
</tr>
<tr>
<td>AmpliTaq Gold</td>
<td>Invitrogen/life technologies</td>
<td>Cat#4311816</td>
</tr>
<tr>
<td>ATP</td>
<td>Fermentas/Thermo Scientific</td>
<td>Cat#R0441</td>
</tr>
<tr>
<td>10X Tango Buffer</td>
<td>Fermentas/Thermo Scientific</td>
<td>Cat#BY5</td>
</tr>
<tr>
<td>High Pure Viral Nucleic Acid Large Volume Kit</td>
<td>Roche</td>
<td>Cat#5114403001</td>
</tr>
<tr>
<td>T4 DNA Ligase</td>
<td>Fermentas/Thermo Scientific</td>
<td>Cat#EL0011</td>
</tr>
<tr>
<td>Min Elute PCR Purification Kit</td>
<td>QIAGEN</td>
<td>Cat#28006</td>
</tr>
<tr>
<td>PEG-4000</td>
<td>Sigma</td>
<td>Cat#1546569</td>
</tr>
<tr>
<td>Agencourt AMPure XP beads (60ml)</td>
<td>Beckman Coulter</td>
<td>Cat#A63881</td>
</tr>
</tbody>
</table>

(Continued on next page)
Continued

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Sensitivity DNA (chips + reagents) (Bioanalyzer 2100)</td>
<td>Agilent Technologies</td>
<td>Cat#5067-4626</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European Nucleotide Archive</td>
<td><a href="https://www.ebi.ac.uk/ena/data/view/PRJEB86458">https://www.ebi.ac.uk/ena/data/view/PRJEB86458</a></td>
<td>ENA: PRJEB86458</td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IS1_adapter.P5: 5'-A<em>C</em>A<em>C</em>TCTTTCCCTACAGACGCTTCCG<em>A</em>T<em>C'T-3' (</em> indicates a PTO bond)</td>
<td>[41] Biomers</td>
<td></td>
</tr>
<tr>
<td>IS2_adapter.P7: 5'-G<em>T</em>G<em>A</em>CTGGAGTTCAAGCTACACTCTTCCG<em>A</em>T<em>C'T-3' (</em> indicates a PTO bond)</td>
<td>[41] Biomers</td>
<td></td>
</tr>
<tr>
<td>IS4: (5'-AATGATACGGCGACCACCGAGTCTACACTCTTTCCCTACAGACGCTTCC)</td>
<td>[41] Biomers</td>
<td></td>
</tr>
<tr>
<td>P7 indexing: (5'-CAACAGAAAGACGCGATACGAGATxxxxxxxGTGACTGGAGTTCAGACGTGT 3') where x is one of 228 different 7 bp indexes provided in [42]</td>
<td>[41] Biomers</td>
<td></td>
</tr>
<tr>
<td><strong>Software and Algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MergeReadsFastq_cc.py</td>
<td>[43]</td>
<td><a href="https://bioinf.eva.mpg.de/fastqProcessing/">https://bioinf.eva.mpg.de/fastqProcessing/</a></td>
</tr>
<tr>
<td>FilterUniqSAMCons_cc.py</td>
<td>[43]</td>
<td><a href="https://bioinf.eva.mpg.de/fastqProcessing/">https://bioinf.eva.mpg.de/fastqProcessing/</a></td>
</tr>
<tr>
<td>Samtools-0.1.19</td>
<td>[45]</td>
<td><a href="https://sourceforge.net/projects/samtools/files/samtools/0.1.19/">https://sourceforge.net/projects/samtools/files/samtools/0.1.19/</a></td>
</tr>
<tr>
<td>mpileup from the Samtools package</td>
<td>[45]</td>
<td><a href="https://sourceforge.net/projects/samtools/files/samtools/0.1.19/">https://sourceforge.net/projects/samtools/files/samtools/0.1.19/</a></td>
</tr>
<tr>
<td>vcfutils.pl from the Samtools package</td>
<td>[45]</td>
<td><a href="https://sourceforge.net/projects/samtools/files/samtools/0.1.19/">https://sourceforge.net/projects/samtools/files/samtools/0.1.19/</a></td>
</tr>
<tr>
<td>POPSTATS</td>
<td>[46]</td>
<td><a href="https://github.com/pontussk/popstats">https://github.com/pontussk/popstats</a></td>
</tr>
<tr>
<td>Haplofind</td>
<td>[47]</td>
<td>https:// haplofind.unibo.it/new/</td>
</tr>
<tr>
<td>PhyloTree Build 17 (18th February 2016)</td>
<td>[48, 49]</td>
<td><a href="http://www.phylotree.org/">http://www.phylotree.org/</a></td>
</tr>
<tr>
<td>PMDtools</td>
<td>[50]</td>
<td><a href="https://code.google.com/archive/p/pmdtools/">https://code.google.com/archive/p/pmdtools/</a></td>
</tr>
<tr>
<td>EIGENSOFT v.6.0.1</td>
<td>[51]</td>
<td><a href="https://github.com/DRechLab/EIG">https://github.com/DRechLab/EIG</a></td>
</tr>
<tr>
<td>qp3Pop v. 204. ADMIXTOOLS package</td>
<td>[19]</td>
<td><a href="https://github.com/DRechLab/AdmixTools">https://github.com/DRechLab/AdmixTools</a></td>
</tr>
<tr>
<td>qpDist v. 450. ADMIXTOOLS package</td>
<td>[19]</td>
<td><a href="https://github.com/DRechLab/AdmixTools">https://github.com/DRechLab/AdmixTools</a></td>
</tr>
<tr>
<td>ADMIXTURE</td>
<td>[29]</td>
<td><a href="https://www.genetics.ucla.edu/software/admixture/">https://www.genetics.ucla.edu/software/admixture/</a></td>
</tr>
<tr>
<td>Pong</td>
<td>[52]</td>
<td><a href="https://github.com/ramachandran-lab/pong">https://github.com/ramachandran-lab/pong</a></td>
</tr>
<tr>
<td>Integrative Genomics Viewer (IGV)</td>
<td>[53]</td>
<td><a href="http://software.broadinstitute.org/software/igv/download">http://software.broadinstitute.org/software/igv/download</a></td>
</tr>
<tr>
<td>Plink v1.90</td>
<td>[54, 55]</td>
<td><a href="https://www.cog-genomics.org/plink/2.0/">https://www.cog-genomics.org/plink/2.0/</a></td>
</tr>
</tbody>
</table>
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ricardo Rodríguez Varela (ricardo.rodriguez.varela@arklab.su.se).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All activities relating to this project complied with the Human Tissue (Scotland) Act 2006 [56], the University of Edinburgh’s Research Ethics Framework, the Guidelines for the Care of Human Remains in Scottish Museum Collections [57], and Historic Scotland’s guide for The Treatment of Human Remains in Archaeology [58]. All sampling of teeth was carried out by L.G.-F. at the Anatomical Museum of Edinburgh under the supervision of T.H.C. and M.M. All the sampled individuals were adults; the gender is provided in Table 1.

The Guanches who survived the Spanish conquest in the 15th century assimilated into Spanish culture and there is currently no ethnic minority claiming aboriginal status or groups of living descendants claiming ancestral ownership of the Guanche human remains held at the Anatomical Museum of Edinburgh. However, the history of the Guanche human remains analyzed for this study is in itself interesting and important from an ethical perspective: the materials were donated to the Anatomical Museum of Edinburgh in the late 19th century by various collectors and anthropologists. Western physicians, medical doctors, and antiquaries (as well as private collectors) amassed numerous collections of human remains during the Victorian or Edwardian eras (1837-1914). Extensive trading and poor, partial or often missing, documentation, in addition to growing ethical concerns regarding both their acquisition, ownership, and public display, has rendered these, often disparate, collections increasingly marginalized in terms of new research. Thus, potentially unique human remains may be largely forgotten by modern scientific as well as culture-historical enquiry [59]. As such, the current project forms part of the Anatomical Museum of Edinburgh’s ongoing detailed inventory of the collections that comprise some 12,000 objects reflecting 300 years of teaching Anatomy at the University of Edinburgh [60].

METHOD DETAILS

DNA extraction

Samples were prepared in dedicated ancient DNA (aDNA) facilities at the Archaeological Research Laboratory, University of Stockholm, Sweden. We targeted the cementum-rich root tip of the teeth as it has been shown to preserve DNA better than most other types of bone [61–63]. We first wiped the teeth with 1% sodium hypochlorite and ddH2O, and obtained ca. 75 mg tooth powder/sample using a multitool drill at the lowest possible rpm (Dremel). DNA extractions were carried out in batches of six plus one extraction blank. The tooth powder was incubated (washed) three times for 15 min under constant rotation in 500 µL 1M sodium phosphate buffer, pH 6, in order to remove microbial contaminants and enrich for endogenous DNA [42]. Buffers were replaced after each incubation step and pellets were washed with 1 mL water after the incubation. DNA was extracted by incubating the bone powder for 24 to 48 hr at 37°C in 1 mL of digestion buffer (0.45 M EDTA pH 8.0 and 0.25 mg/ml of proteinase K). After incubation, samples were centrifuged and 1 mL of supernatant was transferred to an Eppendorf Lo-Bind 1.5 mL tube. DNA was extracted and purified using a silica-binding method optimized for short DNA fragments [64, 65]. The 1 mL of supernatant extract was added to a 13 mL of a binding buffer containing 5 M guanidine hydrochloride, 40% (vol/vol) isopropanol, 0.05% Tween-20 and 90 mM sodium acetate (pH 5.2) in a 50 mL Falcon tube. The 14 mL solution containing the binding buffer and the extraction supernatant was transferred into a 50 mL silica column (Roche, High Pure Viral Nucleic Acid Large Volume Kit). We centrifuged the 50 mL Roche tube for 6 min at 4,000 rpm and remove the Roche falcon tube placing the inside silica column into a new 2 mL collection tube. After centrifuging the silica column tube for 1 min at 6,000 rpm we purified the DNA by adding 750 µL PE buffer (QIAGEN) to the silica column followed by centrifugation at 6,000 rpm. The flow-through was discarded and we repeated this step one time. After these two wash steps, we centrifuged the dry column tube for 1 min at max speed (16,000 rpm) and placed the silica column in a new 1.5 mL collection tube. We pipetted 22 µL of TE buffer (EB QIAGEN buffer plus 0.05% Tween) onto the silica membrane for elution, and after 5-min incubation at 37°C, the DNA eluate was collected by centrifugation for 30 s at maximum rpm. This step was repeated once for a total of 44 µL of DNA extract.

Library construction

One Illumina double stranded library was built from 20 µl of extract using the blunt end ligation protocol described in [41] with modifications as in [66]. Due to the fragmented state of ancient DNA the initial nebulization step was omitted. Indexing PCR’s were set up in a total volume of 25 µl with a final concentration of 1X Gold Buffer (Invitrogen/life technologies), 2.5 mM Magnesium Chloride (Invitrogen/life technologies), 250 µM dNTP (each), 3 µL of DNA library, 0.2 µM IS4 PCR primer (5’-AATGATACGGCGACCACCGAGATCTCAGACGGGATCC TCACTCTTTCCCTACACGACGCTTCTT 3’) and 0.2 µM indexing primer (5’-CAAGCAGAAGACGGCATACGAGATXXXXXXGT GAAGTCGAGATCCGACGTGTT 3’) (x is one of 228 different 7 bp indexes provided in [41]) and 0.1 U/µl of AmpliTaq Gold (Invitrogen/life technologies). Cycling conditions were performed as follows: a 2 min activation step at 94°C, followed by 8-15 cycles of 30 s at 94°C, 30 s at 60°C, 45 s at 72°C, with a final extension of 10 min at 72°C. qPCR was used to assess the optimal number of PCR cycles for amplification that varied between 14 and 21 [41]. Each library was amplified in quintuplicates together with two PCR blanks with an index in the P7 primer. PCR products from the same library were pooled and purified with Agencourt AMPure XP beads (Beckman Coulter). The concentration and size profiles of the purified libraries were identified on a Bioanalyzer 2100 using
the High Sensitivity DNA chip (Agilent) for DNA. Purified libraries were pooled in equimolar concentration and sequenced on an Illumina HiSeq X at the SNP and SEQ technology platform at the SciLife Sequencing Centre in Stockholm.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sequence processing and alignment
Paired-end reads were merged requiring an overlap of at least 11 bp and remaining adaptor sequences were trimmed using MergeReadsFastq_cc.py [43]. The fragments were then mapped to the human reference genome as single-end reads using BWA aln version 0.7.8 [44] with the non-default parameters -i 16500 -n 0.01 -o 2. BAM files of different sequencing runs were merged per library using Samtools [45]. PCR duplicates were then removed using a modified version of FilterUniqSAMCons_cc.py [43]. Fragments of at least 35 bp length, with less than 10% mismatches to the reference genome and minimum mapping quality of 30 were kept for downstream analysis. Biological sex was estimated based on the number of reads mapping to the sex chromosomes and compared to a reference panel [67]. We restricted the analysis to sequence alignments with mapping quality of at least 30.

SNP genotype data
We built pseudo-haploid genomes by randomly choosing one read with minimum mapping quality 30 and a base quality of 30 or higher covering the site of interest [66]. Sites showing indels and transition were excluded to avoid potential post-mortem damage and only transversion sites were used in the analysis.

Principal Component Analysis
The principal component analysis of the Guanches was performed using EIGENSOFT v.6.0.1 [51] with a selection of European, Middle East and North African populations from the Human Origins dataset [19, 20]. The PCAs for the modern populations were made using all SNPs and each of the ancient individuals using the merged SNP data. The number of overlapping SNPs between ancient individuals and the Human Origins dataset ranged from 74,642 (gun002) to 370,595 (gun011). We use Procrustes analysis [68, 69] to transform coordinates of each individual and to plot them together with each individual. The individual PCAs for each ancient individual are shown in the Figure S2.

ADMIXTURE analysis
ADMIXTURE [29] was run on all African and western Eurasian populations of the Human Origins dataset [19, 20] together with the Guanches and 239 relevant published ancient genomes [20, 30–33, 66, 70–75]. To avoid an excess of drift in ancient samples due to post-mortem damages, they were coded missing at all transition sites. The data were pruned for linkage disequilibrium between markers using Plink v1.90 [54, 55] and the parameters–indep-pairwise 200 25 0.4. ADMIXTURE was then run for 2 to 15 clusters (K) with 20 independent runs (different random seeds) per K. Pong [52] was then used in greedy mode to identify common modes between markers using Plink v1.90 and the parameters–indep-pairwise 200 25 0.4. ADMIXTURE was then run for 2 to 15 clusters (K) with 20 independent runs (different random seeds) per K. Pong [52] was then used in greedy mode to identify common modes among the different ADMIXTURE runs per K and to align clusters between different numbers of clusters. We display K = 10 in the main text as it represents the lowest K with a separate North African cluster and the mode shown is common among 15 out of the 20 different runs. All Ks are shown in Figure S3.

Admixture f3 statistics and f4 ratios [19] were calculated using popstats (https://github.com/pontussk/popstats [46]) and the Human Origins dataset. Transition polymorphisms were excluded and standard errors were calculated using a weighted block-jackknife with block sizes of 5Mb. The proportion of Guanche ancestry in modern Canary Islanders α was calculated using the following f4 ratio

\[
\alpha = \frac{1 - f_4(\text{FIN}, \text{Chimp}; \text{CanaryIslanders}, \text{Guanches})}{f_4(\text{FIN}, \text{Chimp}; X, \text{Guanches})}
\]

where X is a modern western European population and FIN the Finish population from the 1000 genome project [76].

Affinity to modern-day North African populations
We used different statistics to measure genetic similarity between Guanches and modern populations. We calculate f2 statistics [25], Hudson’s FST [24], Weir and Cockerham’s FST [23] as well as the commonly used outgroup f2 statistics [19, 71] and the related measurement of average pairwise differences (as proposed by [77]) (Table S2). These statistics were calculated with custom scripts or popstats [46]; https://github.com/pontussk/popstats. The estimators differ in their sensibility to drift, rare alleles, and recent admixture – factors important to take into account for the analysis of North African populations and ancient DNA due to the complex history of admixture events [12, 78] and data quality of low coverage ancient DNA. As a sanity check, we calculated these statistics for modern North Africans as a ‘good’ estimator for our purposes should identify another North African population as closest relative.

Mitochondrial DNA analysis
Consensus sequences for the mitochondrial genomes were called using mpileup and vcfutils.pl (vcf2fq) from the Samtools package [45]. Only reads with a minimum mapping score of 30 and a minimum base quality of 30 were used to call confident bases for the final consensus sequences. However, haplogroup diagnostic positions were checked manually with the program (IGV) [53]. Mitochondrial haplogroups were identified using HAPLOFIND [47] and PhyloTree Build 17 (18th February 2016) [48] (Table S1).
**Y chromosome analysis**
The Y chromosome sequences were filtered out using mpileup from the Samtools package [45]. The pileup file was then merged with the PyloTree Y haplogroup definitions [49]. Y chromosome haplotypes were called using the nomenclature of the International Society of Genetic Genealogy ISOGG database (https://www.isogg.org) v. 11.349 (accessed 04, 2016) (Data S1, sheet 1). We excluded all non-SNP sites, transition sites (to avoid deamination damage), and A/T and G/C SNPs (to avoid strand misidentification).

**Damage patterns**
The presence of 3’ and 5’ cytosine deamination patterns characteristic of ancient DNA [10, 79–82] was estimated using PMDtools [50]. All individuals presented deamination patterns consistent with presence of ancient DNA templates (Figure S1A).

**Mitochondrial DNA authentication**
Contamination levels were estimated based on the analyses of polymorphic site distribution in mitochondrial sequences [11]. The contamination point estimates ranged between 0.53% up to 13.79% (0.53% to 5.97% for the five individuals included in further autosomal genomic analysis) rendering the obtained genomic data reliable (Table 1 and S1).

**Outgroup \( f_3 \) statistics**
In order to check the genetic affinities between the Guanches and the modern populations we performed outgroup \( f_3 \) statistics using qp3Pop v. 204. of the ADMIXTOOLS package [19]. The outgroup \( f_3 \) statistics using of the form (O; A, B) [19, 71] represent the amount of genetic affinity shared between test populations, A and B. The analyses were performed with the Human Origins dataset (using between 21,864 and 108,190 SNPs). The outgroup (O) was Ju’hoansi population, the tested individual was A, and any of the populations from the reference panel was the test population B. The analyses were performed with up to 705 jackknife blocks. The results for each of the individuals are listed in Data S1, sheet 2.

**D statistics**
Deviations from tree-like population history between the analyzed individuals were tested using \( D \) statistics included in ADMIXTOOLS package qpDstat v. 450 [19]. We exclude transition sites in all analysis and used pseudo-haploid genomes [66].

The \( D \) statistics of the form \( D(O, X; A, B) \) [19] represent the amount of genetic affinity shared between test individual X, A and B, where O is the outgroup the Ju’hoansi population; X Guanche individual; A North African population; B Sardinians or AnatoliaEN (Early Neolithic)

The analyses were performed with the Human Origins dataset (using between 20,227 and 108,190 SNPs). The standard errors were estimated by performing block jackknife over blocks of 705 SNPs. The results are summarized in Data S1, sheet 3.

**DATA AND SOFTWARE AVAILABILITY**
The newly generated genome data have been deposited at the European Nucleotide Archive with the accession number ENA: PRJEB86458 (https://www.ebi.ac.uk/ena/data/view/PRJEB86458).