The Genomic Prehistory of the Indigenous People of Uruguay

John Lindo1*, Rosseirys De La Rosa1, Andre Luiz Campelo dos Santos2,3, Mónica Sans4, Michael DeGiorgio3, Gonzalo Figueiro4*

1Department of Anthropology, Emory University, Atlanta, GA, 30322, USA
2Department of Archeology, Federal University of Pernambuco, Recife, Brazil
3Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA
4Departamento de Antropología Biológica, Facultad de Humanidades y Ciencias de la Educación, Universidad de la República, Montevideo, Uruguay

*Corresponding Author

Abstract

The prehistory of the people of Uruguay is greatly complicated by the dramatic and severe effects of European contact, as with most of the Americas. After the series of military campaigns that exterminated the last remnants of nomadic peoples, Uruguayan official history masked and diluted the former indigenous ethnic diversity into the narrative of a singular people that all but died out. Here we present the first whole genome sequences of the Indigenous people of the region before the arrival of Europeans, from an archaeological site in eastern Uruguay that dates from 2,000 years before present. We find a surprising connection to ancient individuals from Panama and eastern Brazil, but not to modern Amazonians. This result may be indicative of a distinct migration route into South America that may have occurred along the Atlantic coast. We also find a distinct ancestry previously undetected in South America. Though this work begins to piece together some of the demographic nuance of the region, the sequencing of ancient individuals from across Uruguay is needed to better understand the ancient prehistory and genetic diversity that existed before European contact, thereby helping to rebuild the history of the indigenous population of what is now Uruguay.

Introduction

Historically, Uruguayan identity has been marked by the extermination of the indigenous populations found in the territory at the time of European contact in the 16th century and up until the 19th century. The extermination was carried out through a series of military campaigns of which the massacre at Salsipuedes creek in 1831 is considered the culminating event. (1) The target of the Salsipuedes campaign was the ethnic group known as the Charrúa, which at the time was the term employed for the remnants of various hunter-gatherer groups in the recently independent territory of Uruguay. Subsequently, it was held that in sharp contrast to all other South American countries, Uruguay lacked indigenous populations, an idea still widely accepted.
The research presented here aims to elucidate the genomic prehistory of the Indigenous people of Uruguay by presenting low-coverage whole genomes from the CH2D01-A archeological site in Rocha, Uruguay, which date from ~1,450 to ~668 years before present (BP; Table 1). This represents the first ancient genomic DNA from the region and presents a starting point to examine the evolutionary history of the Indigenous people of Uruguay and their diversity from a genomic perspective.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Age*</th>
<th>mtDNA Haplogroup</th>
<th>Contamination†</th>
<th>Endogenous DNA</th>
<th>Average Sequencing Depth</th>
<th>Genome Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH-13</td>
<td>668 ±22 BP</td>
<td>C1d3</td>
<td>1%, 95% CI[0-2%]</td>
<td>3.4%</td>
<td>10.7</td>
<td>0.137</td>
</tr>
<tr>
<td>CH-19B</td>
<td>1450±70 BP</td>
<td>C1c</td>
<td>5%, 95% CI[4-6%]</td>
<td>5.1%</td>
<td>7.14</td>
<td>0.344</td>
</tr>
</tbody>
</table>

Table 1. Ancient samples sequenced in this study from the archaeological site CH2D01-A in Rocha, Uruguay.

*Radiocarbon dating was conducted at the University of Arizona AMS Laboratory. †Contamination was estimated from mtDNA alignments using Schmutzi(2).

Results

To assess the relationship of the ancient Uruguay samples with global and regional populations, both ancient and modern, we merged the dataset with samples from the Simons Genomes Diversity Project (3) and ancient whole genomes from the Americas (4–8). To prevent a batch effect from the method used to call the ancient Uruguay genotypes, the ancient reference BAM files were also called with the ancient DNA caller ARIADNA(9) (see Extended Methods). To maximize the number of overlapping sites for the various population genetic analyses, only whole genomes were chosen for comparison (Fig 1A).

Mitochondrial DNA

The mitochondrial genome of CH-13 carries all diagnostic mutations of haplogroup C1d, namely 194T and 16051G, including the additional mutations 12378T, 16140C and 16288C which place it within subhaplogroup C1d3. This subhaplogroup had its origin approximately 9,000 years ago and apparently evolved entirely in what is now Uruguay, and is found also in CH-20, an individual found in the same site but between 700 and 1000 years older (10). The mtDNA of CH-19B carries mutations 1888A and 15930A, diagnostic of haplogroup C1c, but lacks further mutations associated with registered subhaplogroups (C1c1 through C1c8). Furthermore, CH-19B carries 606G and a deletion in position 7471, neither of which have been found in published mitochondrial sequences. The lineage represented by CH-19B might very well be extinct.

Genomic analyses

We performed a principal component analysis to better understand the relationship of the ancient Uruguay individuals with other ancient individuals from the Americas. C/T and G/A transitions were removed from the dataset to guard against the most common forms of postmortem DNA damage (11) and to prevent false affinities among the ancient samples. Since the Uruguay ancient individuals are low coverage compared to the modern and ancient high coverage samples in the reference panel (3–6, 8, 12), we projected the ancient genomes onto top two principal components identified from the modern samples using SmartPCA (13). Interestingly, the contemporary ancient samples from Uruguay (CH19B, ~1400 BP) and Panama (PAPV173(8), ~1400 BP) show a strong affinity on the first two principal components (Fig. 1B), with the younger Uruguay sample (CH13,
~600 BP) showing a more distant relationship. To further elucidate the relationship among the ancient Uruguay individuals and the Americas, we performed an ADMIXTURE-based cluster analysis, with \( K=4 \) clusters showing the best cross-validation value (Fig. 1C). Transitions were also removed for this analysis. The Uruguay ancient samples exhibit a green ancestry cluster which is shared with USR1(5) (Alaska, ~11,500 BP) and Anzick-1(4) (Montana, ~12,500 BP). In relation to the South America, the green cluster is shared with Sumidouro5(6) (Brazil, ~10,000 BP) and PAP173(8) (Panama, ~1400 BP), which shows the largest shared fraction. With regard to the modern samples in the reference panel, the green cluster is apparent in a Mayan individual but is not observed in other populations.

To further assess the relationship of the ancient individuals from Uruguay with global populations, we utilized outgroup \( f^3 \) statistics to assess the shared genetic ancestry with the modern individuals from Simons(3). Outgroup \( f^3 \) statistics of the worldwide dataset demonstrate that both ancient Uruguay individuals display greater affinity with Indigenous groups from South America than with other populations (Figs. 2A and 2C). Ranked outgroup \( f^3 \) statistics suggest that both ancient individuals from the two time periods (CH19B: ~1450 BP and CH13: ~668 BP) tend to share the greatest affinity with Brazilian living groups, the Surui and Karitiana (Fig. 2B). We also note that we do not see an Austronesian signal with either individual, which may suggest a more complicated ancestry despite the apparent affinity.
To further explore the relationships between the ancient Uruguay individuals and the Americas, we utilized maximum likelihood trees inferred with TreeMix(14). Transitions were also removed from this analysis to prevent false affinities among ancient samples. Though migration events were not well supported by our bootstrapping validation, we do have support for the structure of a tree that shows a nuanced relationship between the ancient Uruguay individuals and South America. The ~10,000 BP ancient sample from Brazil, Sumidouro5, falls basal to both ancient Panama and ancient Uruguay (Fig. 3A). It should be noted that Sumidouro5 is from an archaeological site on the by the eastern coast of Brazil, as opposed to the Amazon populations, the Surui and Karaitiana, which are located in the western part of the country (Fig. 1A). Though we are careful to claim definitive relationships among the ancient and modern samples, the tree does seemingly correctly position the ancient individuals from the highlands of Peru (Rio Uncanalle(7)) with modern-day individuals from the same region, the Quechua(3). Furthermore, the ~11,500-year-old USR1(5), a terminal individual from Alaska, is placed as an outgroup to all populations from the Americas, which lends additional validation of the tree structure.

We also tested the relationship of the 1,400-year-old ancient Uruguay individual with modern South American populations using qpGraph(15), which extends the $f_3$ statistic to numerous populations by incorporating the topology of an admixture graph. We found that that the model fit best with two migration events, incorporating the individuals from key populations of the Simons...
Genome Diversity Project (3) (Fig. 3B). The topology of the graph also suggests that ancestry of the ancient Uruguay sample is deriving from two sources: a deep ancestral source and a source that led to the Karaitiana and Surui of Brazil. Both the maximum likelihood tree and the qpGraph show a more complicated picture than what was shown with the outgroup f3 statistic, whereby the Amazonian populations share a more distant connection to the ancient Uruguay individuals. This connection may relate to a more general ancestry signal from South America, rather than a direct one, and may be a consequence of different migrations upon entry into the continent. In contrast, we see a connection with ancient Brazil, Panama, and Uruguay on the maximum likelihood tree, where they form their own branch (Fig. 3A). The qpGraph shows a connection between the ancient Panama sample and the oldest Uruguay individual, demonstrating a migration event between the two (Fig. 4B). Taken together, it’s possible that the connections are reflective of migration routes that occurred along the Atlantic coast of South America. We also note a deep ancestral component contributing to the ancient sample form Uruguay (Fig. 3B), which combined with the ancestry cluster results (Fig. 1C), may represent a previously undetected ancestry in South America.

Figure 3. Maximum likelihood Tree and qpGraph. A) Maximum likelihood trees generated by TreeMix (14) using whole-genome sequencing data from the Simons Genome Diversity Project (3). The tree shows a connection between ancient samples from eastern Brazil, Panama, and Uruguay. B) The qpGraph best fitting model with two migration events. The populations included derive from the Simons (3) and shows a deep ancestral event in the direction of the ancient Uruguay sample (CH19B), along with a migration signal associated with ancient Panama.
Discussion

Here we start to elucidate the origins of the Indigenous people of Uruguay. We find that the ancestral population of the ancient individuals may have derived from a migration that stemmed closer to the Atlantic coast and may be different than the migrations that led to modern Amazonian indigenous populations from Brazil. We also find a close relationship to an ancient Panama sample, which could correlate with the source of the migration, Sumidouro Cave (part of the Lagoa Santa archaeological area), in Brazil, some 5,000 km away (Fig. 4).

That being said, besides indicating a past migration to Uruguay from Lagoa Santa through the Atlantic coast, this finding also suggests the existence of a putative south-to-north ancient migration route linking Lagoa Santa and Panama. Moreover, these movements seem to have occurred independently of the migratory waves that occurred near the Pacific coast in South American populations. Altogether, this presents genomic evidence for ancient migration events along, or at least closer to, South America’s Atlantic coast, which is supported by paleoenvironmental and chronological(16).

While we begin to unravel the relationship of the Indigenous people of Uruguay on a continental level, in addition to the potential discovery of a distinct ancestral component in South America, it also important to point out the need for ancient DNA from other archaeological sites from across Uruguay, especially those close to the time of European contact. Such samples will help to better capture the genetic diversity of the Indigenous that existed upon the Spanish in the 16th Century, and by extension, better understand the diversity of Indigenous groups that existed. In doing so, future DNA studies could assist living people in Uruguay to potentially identify indigenous ancestry that is not limited to the “Charrúa.”(17).

Supplemental Information

Extended Methods

Archaeology and Samples

Site CH2D01 is a group of two mounds (A and B) located on the edge of the San Miguel wetlands, in the department of Rocha, in eastern Uruguay. The radiocarbon dating place the occupation of mound A between 2000 BP and the period of European contact. The mound is approximately 1.20 m high with a diameter of 35 m and is presumed to be an area of differentiated activity within a larger site of about 20,000 square meters. The archaeological materials recovered from mound A do not show clear spatial arrangements and were interpreted as "displaced primary contexts": materials that were carried along with the sediments that make up the mound from the places where the activity was carried out. Implicit in this interpretation is the intentional character of the mound construction. Although there is ongoing debate about the exact mechanism of the formation of the "Cerritos," there have been no subsequent interpretations about site CH2D01. In excavation IA, a 25 square meter excavation carried out in the center of mound A, several bone assemblages representing the primary and secondary burials of at least 21 individuals were recovered.
**Ancient DNA and Sequencing**

The ancient tooth samples were extracted, and sequencing libraries were constructed at the Lindo Ancient DNA laboratory at Emory University using the Dabney protocol\(^{(18)}\). Libraries were prepared with the NEB Ultra II DNA Library Prep for Illumina, with modifications for ancient DNA, which including quartering the reagents, the utilization of 1:20 adaptor dilution, and 1.5 ul of premixed NEB indexes. Samples were preliminarily screened for endogenous DNA on the Illumina iSeq 100 at the Lindo Lab, with libraries that were not treated with the USER enzymes. Samples that were selected for deep sequencing on the NovaSeq 6000 at Dante Labs (L’Aquila, Italy), included libraries treated with the USER enzyme to help compensate for DNA damage.

The ancient raw sequences were trimmed for Illumina adapters using AdapterRemoval2\(^{(19)}\) and aligned to the hg19 human reference sequence using the BWA mem algorithm\(^{(20)}\), which has been shown to increase accuracy with ancient DNA mapping over the aln algorithm\(^{(21)}\). The alignments from the shotgun sequences that were not treated with the USER enzyme were used to validate their ancient authenticity with MapDamage2\(^{(22)}\). Both ancient individuals showed deamination patterns consistent with ancient DNA (Supp. Fig. 1).

Genotypes were called from both ancient individuals using the ancient DNA caller ARIADNA, which uses a machine learning method to overcome issue with DNA damage and contamination\(^{(9)}\). The resulting VCF was further filtered to remove genotype calls with allele counts below 3. Since CH-19B showed a relatively high contamination rate, we further filtered the associated VCF using RFMix \(^{(23)}\) to identify and remove sites that showed a high probability of deriving from Europeans. The VCFs were then merged with modern and ancient samples from the Americas with bcftools.

**SmartPCA**

We conducted Principal Component Analysis using the ‘smartpca’ program from the EIGENSOFT v7.2.1 package. Principal Components (PCs) were estimated with the ‘poplistname’ option and using representative individuals from present-day Native American and indigenous South African and Papuan populations from Simons Genome Diversity Panel\(^{(3)}\). The ancient individuals were then projected onto the computed PCs with the ‘lsqproject: YES’ option. No outliers were excluded, and the analysis was based on 4,978,400 loci.

**Assessment of population structure using ADMIXTURE**

We started with the identical filtered dataset of called genotypes described above. We further pruned the dataset by removing sites in strong linkage disequilibrium \((r^2 > 0.1)\) using PLINK\(^{(24)}\). The program ADMIXTURE was used to assess global ancestry of the ancient and present-day samples from this study. We computed cluster membership for \(K=2\) through \(K=15\) and found the lowest cross-validation value to be at \(K=4\). The PONG program was used to visualize the admixture plots.

**Outgroup \(f_3\) analysis**

We applied the \(qp3Pop\) module of ADMIXTOOLS\(^{(15)}\) to compute \(f_3\) statistics with the target population as the African Yoruban population and the two reference populations set as one of the ancient Uruguay samples (CH13 or CH19B) and the other as one of the non-African populations.
from the Simons Genome Diversity Project(3). For this analysis we retained C/T and G/A transitions, as the CH13 and CH19B ancient samples have been treated with uracil-DNA glycosylase to guard against this form of DNA damage.

TreeMix analysis
We started with the identical filtered dataset of called genotypes described above. TreeMix was applied the dataset to generate maximum likelihood trees and admixture graphs from allele frequency data. The Mbuti from the Simons dataset was used to root the tree (with the –root option). We accounted for linkage disequilibrium by grouping M adjacent sites (with the –k option), and we chose M such that a dataset with L sites will have approximately L/M ≈ 20,000 independent sites. At the end of the analysis (i.e., number of migrations) we performed a global rearrangement (with the –global option). We considered admixture scenarios with m = 0 and m = 3 migration events. Each migration scenario was run with 500 bootstrapping replicates, and the replicates were used to determine the confidence of each node.

qpGraph analysis
We employed the ADMIXTOOLS2 (https://uqrmaiel.github.io/admixtools/index.html, ADMIXTOOLS2 is currently under preparation) R package to perform qpGraph(15) estimation. We extracted f_2 statistics between population pairs using a two megabase SNP block size while retaining C/T and G/A transitions, as the CH19B ancient sample has been treated with uracil-DNA glycosylase to guard against this form of DNA damage. We considered scenarios with M ∈ {0,1,2,3,4,5,6} migration events, with graph searches initiated by a random graph and Mbuti population set as the outgroup, stopping the search after 100 generations. If the best graph with M events did not have a better score than those with fewer events, then the graph search was rerun. The best-fit model of two migration events was chosen by assessing statistical differences between model score distributions computed from 500 bootstrap replicates of f_2 blocks.

Acknowledgements. This work was supported by National Science Foundation grants BCS-1926075, BCS-1945046, and BCS-2001063, and by Fundação de Amparo à Ciência e Tecnologia de Pernambuco grant BFP-0191-7.04/20.

Supplementary Figure 1. DNA Damage Patterns.
References


