Indigenous Ancestry and Admixture in the Uruguayan Population

Lucía Spangenberg[1], María Inés Fariello[2], Darío Arce[3], Gabriel Illanes[4], Gonzalo Greif[5], Jong-Yeon Shin[6], Seong-Keun Yoo[6], Jeong-Sun Seo[6,7], Carlos Robello[5,8], Changhoon Kim[9], John Novembre[10,11]*, Mónica Sans[12]* and Hugo Naya[1,13]*

[1]Bioinformatics Unit, Institut Pasteur de Montevideo, Montevideo, Uruguay, [2]Instituto de Matemática y Estadística Rafael Laguardia, Facultad de Ingeniería, Universidad de la República (UDELAR), Montevideo, Uruguay, [3]Agregado de Cooperación Lingüística y Cultural de la Embajada de Francia en Uruguay, Montevideo, Uruguay, [4]Centro de Matemáticas, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay, [5]Molecular Biology Unit, Institut Pasteur de Montevideo, Montevideo, Uruguay, [6]Precision Medicine Institute, Macrogen Inc., Seoul, South Korea, [7]Asian Genome Institute, Seoul National University Bundang Hospital, Gyeonggi-do, South Korea, [8]Departamento de Bioquímica, Facultad de Medicina, Universidad de la República (UDELAR), Montevideo, Uruguay, [9]Bioinfomatics Institute, Macrogen Inc., Seoul, South Korea, [10]Department of Human Genetics, University of Chicago, Chicago, IL, United States, [11]Department of Ecology and Evolution, University of Chicago, Chicago, IL, United States, [12]Departamento de Antropología Biológica, Facultad de Humanidades y Ciencias de la Educación, Universidad de la República (UDELAR), Montevideo, Uruguay, [13]Departamento de Producción Animal y Pasturas, Facultad de Agronomía, Universidad de la República (UDELAR), Montevideo, Uruguay

The Amerindian group known as the Charrúas inhabited Uruguay at the timing of European colonial contact. Even though they were extinguished as an ethnic group as a result of a genocide, Charrúan heritage is part of the Uruguayan identity both culturally and genetically. While mitochondrial DNA studies have shown evidence of Amerindian ancestry in living Uruguayans, here we undertake whole-genome sequencing of 10 Uruguayan individuals with self-declared Charruan heritage. We detect chromosomal segments of Amerindian ancestry supporting the presence of indigenous genetic ancestry in living descendants. Specific haplotypes were found to be enriched in "Charrúas" and rare in the rest of the Amerindian groups studied. Some of these we interpret as the result of positive selection, as we identified selection signatures and they were located mostly within genes related to the infectivity of specific viruses. Historical records describe contacts of the Charrúas with other Amerindians, such as Guaraní, and patterns of genomic similarity observed here concur with genomic similarity between these groups. Less expected, we found a high genomic similarity of the Charrúas to Diaguita from Argentinian and Chile, which could be explained by geographically proximity. Finally, by fitting admixture models of Amerindian and European ancestry for the Uruguayan population, we were able to estimate the timing of the first pulse of admixture between European and Uruguayan indigenous peoples in approximately 1658 and the second migration pulse in 1683. Both dates roughly concurring with the Franciscan missions in 1662 and the foundation of the city of Colonia in 1680 by the Spanish.

Keywords: population genomics, human genomics, indigenous ancestry, admixed population, bioinformatics, South America

INTRODUCTION

At the time of European contact with the peoples of present-day Uruguay (estimated in 1516, date of the landing of first Spanish conqueror Juan Díaz de Solís), the land was populated by several different indigenous groups for which there is varying and scarce data (Arce, 2015). Historians have progressively reduced the number of described ethnic groups from three or four. The main one, the macro-ethnic Charrúa group (expression first proposed by Vidart (1973) to describe the former Pampeans), was composed by groups known as the Charrúa and Guenoa and other minor groups (eg., Bohan and Yaro), while the Guaraní, an indigenous group more abundant in regions farther north from Uruguay and present-day Paraguay, arrived at the territory probably few centuries before Spanish conquerors (Cabrera, 1989). Charrúas, even though they participated in the independence war against Spain, were continuously persecuted, including by the Uruguayan government after independence (1830). The "Salsipuedes" massacre, in 1831, has been recorded as the final episode of the persecution of the Charrúa by the Uruguayan army, where the government of the country deliberately killed the majority of them. Only few indigenous men were able to escape, while women and children were taken as prisoners and distributed as servants across the country. The genocide, or extermination, of this ethnic group, has been an object of different historic and ethnohistoric studies (Acosta Y Lara, 1985a; Acosta Y Lara, 1985b).

Originating further north, in present-day Paraguay, the migration of Guaraní to Uruguay has been considered the first migration wave into Uruguay in historical times (González Risotto and Rodríguez, 1989). The last recorded arrival was in 1829 when General Fructuoso Rivera (first president of the Uruguayan Republic) brought several thousands of natives from the Jesuits Missions (probably mostly Guaraní) to found two cities, Santa Rosa del Cuareim and lastly, San Borja del Yi. In 1767 the Jesuits Missions were expelled from the territory, leaving thousands of individuals living in those regions (Curbelo, 2009).

Despite Guaraní's presence, Uruguayan national identity has been related to the disappearance of the Charrúas and for many years, it was believed that it was a "native-free" country.

After the genocide, the indigenous contribution was overwhelmingly underestimated: in the first National Census (1852) indigenous peoples were not even mentioned (categories were only "Whites", "Mulattoes", "Blacks", "foreigners"); The most recent National Census (2011) introduced questions about ancestry and ethnicity, revealing that only 4.9% of the population self-report as having at least one indigenous ancestor, though contemporary studies on mitochondrial genomes have revealed indigenous haplogroups in frequencies of 31–37%, and the nuclear genome provides estimates of 10–14% indigenous ancestry (Sans et al., 1997; Bonilla et al., 2004). At present, Uruguay is experiencing a reemergence of the Charrúas cultural identity (Magalhaes and Michelena, 2017), even though it is not necessarily linked to ancestry.

While admixture has been sufficient and not denied by any present association or self-recognized ethnic group, full Uruguayan indigenous genomes are lost. However, as shown by studies of other Amerindian ancestries in Latin America (Schroeder et al., 2018) the reconstruction of indigenous gene pools using admixed gene pools opens a new tool to understand the origins of the ancient inhabitants of America. Moreover, they can be used to identify genomic segments that could belong to different ethnic groups, interesting to history and individual and national identity. As Bodner et al. states, data about indigenous or mixed South American populations will allow a more detailed panorama of migrations inside the continent (Bodner et al., 2012). Admixed or urban populations have proven that they are useful predictors of indigenous diversity, at least considering mtDNA (Tavares et al., 2019).

Studies of ancient DNA have also contributed to the understanding of past populations of the continent through the recovering of extinct genomes, as well as to prove continuity from past to present times. At present, several complete genomes from prehistoric indigenous groups have been sequenced, eg., four individuals from the Northwestern coast (Lindo et al., 2017), as well as 15 from different parts of America (Moreno-Mayar et al., 2018). More recently, a database of over a thousand European ancient genomes was released (Olalde et al., 2018). In Uruguay, three ancient mitogenomes have been published. One of these belonged to the mtDNA haplogroup C1d3, which has been found only in Uruguayans and established continuity from an individual dated to 1610 ± 46 years before present with one Charrúa Indian from the 18–19th centuries, and present population individuals (Sans et al., 2015), while the other two lineages, belonging to haplogroup C1b, seem to have no continuity in present times (Figueiro et al., 2016). Moreover, the lack of prehistoric remains that can be identified with ethnic groups living in historical times, as well as the lack of historical remains which belong to indigenous groups, will be an obstacle to use them as the unique approximation to the knowledge of Amerindian genetic characteristics. Hence, the possibility of identifying indigenous segments of DNA from present population will be a useful proxy to reconstruct indigenous genomes and, when possible, to identify different indigenous ethnic origins.

In the present work we report whole genome sequences of 10 individuals with known (self-declared) Charrúas ancestry to gain insight to the indigenous ancestry of Uruguayans. Genome wide ancestry proportions, differences in self-declared and genomic estimation, similarity to other indigenous tribes in the region and selection signatures within indigenous tracts were determined. Additionally, the timing of early admixture with Europeans and Africans was estimated from the genomic data, and concurs with historical events.

MATERIAL AND METHODS

Sample Selection

Participants were selected after an interview with a social anthropologist. Interviews allowed to determine a high percentage of (expected) Amerindian ancestry and that participants were not related to each other. In general,

participants declared to have at least a great grandparent or great great grandparent of Amerindian ancestry. Genealogies were constructed and several metadata information were gathered (ie., birth and death dates and places, occupation/craft, etc.) for these individuals.

15 participants enrolled in the study and 10 were randomly selected for whole genome sequencing, so that anonymity could be assured. All 15 participants (with or without WGS) were given the results of ancestry proportion inference (based on a custom SNP array of ancestry information markers), as feedback for having participated in the study. Of the 10 selected for sequencing, six individuals declared to have "Charrúa" ancestry, two declared a mixed (Guaraní and Charrúa) ancestry, one declared Guenoa and one unknown indigenous group. This information was blinded from the genetic analysis to assist in protecting anonymity of the participants. These procedures were approved by Institut Pasteur de Montevideo ethics committee with reference number IP011-17/CEI/LC/MB and with informed consent signed by each participant. Genomic data generated in this study is available at http://urugenomes.org/lovd/variants.

From here on, we will address these indigenous descendants in this study as "Charrúa," between quotes, since there is some uncertainty regarding the exact ancestry of the individuals. At present, Charrúas are considered the main group, but there were also other additional groups in the territory, that might also have left descendants in the region.

Sequencing and Post Processing of Data

Ten whole genomes were sequenced in an Illumina HiSeq X ten with 30X coverage.

Data was mapped with BWA (Li and Durbin, 2009) with default parameters, variant calling was performed according to the best practices with GATK version 3.4 (Mckenna et al., 2010) using following steps: Indel Realigner, base Recalibrator, and joint Haplotype Caller (with -stand_call_conf 30, -stand_emit_conf 10 and -genotyping_mode DISCOVERY). Annotation was done with ANNOVAR (Wang et al., 2010). A total of 10776269 high quality variants (SNPs and INDELS) were found among the 10 individuals.

Data was phased using Shape-it (Delaneau et al., 2013) using as reference panel the 1000 Genomes data and with input data format in BED/BIM/FAM. We used a two-step approach, first we used the parameter -check to identify the SNPs to exclude from further analysis (missing or flipped in reference panel). Then we used shapeit with -exclude-snp to exclude those SNPs. We processed every chromosome separately in parallel.

MDS [using plink (Purcell et al., 2007)] was performed on a thinned (plink with parameter thin 0.1) merged data set of the Uruguayans, 1000G version 3 (The 1000 Genomes Project, 2015) and Simons Foundations (Mallick et al., 2016) Amerindian individuals resulting in a 39909 SNPs after LD-pruning (using parameter -indep-pairwise 10 5 0.1).

Global and Local Ancestry Estimations

Global ancestry estimations were done with ADMIXTURE (Alexander et al., 2009), using as reference panel 1000G populations and 10 native genomes obtained from the Simon's Foundation Project (Mallick et al., 2016). We used the same thinned data set as for the MDS analysis. In addition, in order to compare Uruguayan data set to a larger set of different indigenous groups, we used the SNP microarray data of (Reich et al., 2012). In this study 52 Native Americans were genotyped at 364,470 single nucleotide polymorphisms, together with European and Africans, among others. The Uruguayan data set was assessed at those positions and was integrated to this data set obtaining a total of 364,398 with a genotypic rate of 0.9988. ADMIXTURE was run using this combined data set, using as reference unadmixed individuals as done in (Reich et al., 2012).

In both cases (thinned data set with 1000G and chip data set) ADMIXTURE was run with K ranging from 3 to 19. For each K, an ad hoc assignment of the clusters labels was done. For K = 3, the labels are most likely "European", "African" and "Native." This was used to calculate global ancestry proportions, in order to compare with RFMix results. For higher K's we evaluated the cross-validation errors and the log-likelihood values in order to choose the right K for our data set (**Supplementary Figure S1**). Even though Ks around the value of 9 could be considered a significant value (**Supplementary Figure S1**), we considered for our analysis higher Ks (eg., K = 19) since for our population several European clusters were dominant even at K = 9 level, and the interest was the fine-structure of the native clusters. Therefore, we focused on the higher values of K, which have high likelihoods. The cluster assignment of the labels for higher K's is less straight forward, but can be made by analyzing the individuals present in each group.

Local ancestry estimation was done with RFMix, using a reference panel that included 21 whole genomes of natives (Simon's Foundation Project), 20 Africans and 23 Europeans from the 1000 Genomes Project (other set of parameters were evaluated, see Supplementary data S1).

Detecting Signatures of Selection

On the whole genomes of masked Uruguayan natives, we applied iHS calculations (Gautier and Vitalis, 2012) for detecting intra population selection signatures. Each individual is considered as two independent haplotypes and we considered for the calculations only those haplotypes that had more than 15% of SNPs genotyped (min_perc_geno.hap = 15) and SNPs genotyped on more than 40% percent of the haplotypes were considered (min_perc_geno.snp = 40).

With these settings we were able to calculate an iHS statistic for 728935 sites. Considering an adjusted p-value of $4.12e^{-8}$, which corresponds to a family wise error of 0.03 with a Bonferroni correction 0.03/728935. The −log value is −7.4 (significance line in **Figure 2A**).

On both most significant peaks (chromosome 1 and 4) a visual inspection of the haplotype was made. Starting from the focal SNPs that iHS identifies, the haplotype is defined to the right and left. When one base differs from the extending haplotype, the extension stops and the haplotype breaks. After the definition of the haplotype, its presence in other natives was evaluated. Conserved haplotypes can be seen as long horizontal bars including many individuals (eg., blue haplotype in **Figure 2B**,

left) and haplotypes that are conserved in a population are seen as long bars in a subpopulation (eg., blue haplotypes at the bottom of **Figure 2B**, right), and in other populations they appear as small centered bars.

The different colors in the figure correspond to different haplotypes that were extended from the focal SNPs.

Determining the Generation at Which an Individual has a Complete Native Ancestor With High Probability

The objective of this hypothesis test is, fixed a number T of generations, and given the length of the maximum Charrúa haplotype for every non-sexual chromosome, assess if it is possible that one of the individual's great$^{\{T-2\}}$ grandparent is a complete Charrúa ancestor.

The test considers the family tree of the individual up to T generations in the past, regards the chromatids as intervals (measured in Morgans) rather than sequences of base pairs, and assumes the following meiosis model to generate a daughter chromatid from two chromatids for every pairing of the family tree:

1) Simulate the recombination points using a Poisson process with parameter L_i (length of the chromosome in Morgans). After adding the borders of the interval $[0,L_i]$, we obtain $\{x_0 = 0, x_1, \ldots ,x_n,x_{n+1} = L_i\}$. As long as we measure the intervals in Morgans, using a Poisson process to simulate the recombination points is not a strong assumption for the model.
2) Select a chromatid at random, and consider the segment $tr_1 = [0,x_1]$ in the selected chromatid. This will be the first segment of the daughter chromatid.
3) Switch to the other chromatid, and concatenate the segment $tr\_2 = [x_1,x_2]$.
4) Iterate the last step until the length of the daughter chromatid is L_i.

This is a model which is closer to the biological model than other widely used models, such as the Wright-Fisher model. However, this model does not fulfill the Markov property, so we need to simulate all the pairings of the family tree if we want to obtain an approximation of the distribution function of the statistic under the null hypothesis. This can only be achieved if we restrict the null hypothesis to a simpler scenario:

H_0: a^0 has exactly one pure Charrúa ancestor, T generations ago.

The other ancestors are pure from some other ancestral population.

H_1: a^0 has no pure Charrúa ancestors, T generations ago.

If we consider a smart choice for the test statistic, H_0 will become a borderline case of H_0, meaning that the statistic will be stochastically smaller under H_0. The test statistic requires a scaled score for every chromosome, and a way to combine them in a final test statistic. In this work, we choose to use the (scaled) length of the maximum tract for every chromosome, and consider the maximum of the scores as the test statistic.

To note is that the hypothesis test satisfies that, if we reject under the base null hypothesis H_0 (that is, one unadmixed

Amerindian ancestor and 2^{t-1} unadmixed non-Amerindian ancestors), it would also reject any other scenario where the null hypothesis is true (that is, one unadmixed Amerindian ancestor and the other 2^{t-1} ancestors admixed in any way). This works as many other composite hypothesis tests: the idea is controlling the "worst case scenario" where H_0 is true, and use it as a bound for any other scenario where H_0 is true.

More details and choices for a test statistic can be found in Illanes et al. (manuscript submitted).

Migration Models

The results of local ancestry estimations with RFMix (using as reference panel whole native genomes, and 1000G Africans and Europeans) were used as input for the TRACTS tool (Gravel, 2012). Here, the length distribution of African, European, and Native American ancestry tracts along each chromosome is determined and an extended Markov model is applied to compare the observed data with predictions from different demographic models considering various migration scenarios.

As stated, two different models were tested, using predefined models provided by TRACTS and examples in https://github.com/ armartin/ancestry_pipeline. We chose models including the three main ancestries that made sense for the historical process. First, Europeans came in contact with Amerindians and posteriorly Africans were most likely brought to the continent. Probably there were several pulses of Europeans after that initial contact. Hence, following models would be appropriate for our population: 1) EUR, NAT + AFR and, 2) EUR, NAT + AFR + EUR. For each, log likelihood was obtained, and bootstrapping was performed as follows: first 100 iterations were done, and if results was promising, further 100 were run, and so on until a maximum of 400. Model 1) was run with 200 iterations and 2) with 400.

We corroborated the estimations with the MALDER tool (Loh et al., 2013) that also estimates admixture times, but based on the weighted linkage disequilibrium (LD) statistic and gives for the estimated admixture times also a confidence interval.

Analysis of Heterozygosity

Most of the individuals had the same amount of heterozygosity (around 0.355) with the exception of two individuals. To understand from which population it comes from, a linear model with heterozygosities as the dependent variable and the ancestry proportion (EUR, NAT, AFR) from the "Charrua" individuals as the covariates was fitted.

HET = a*EUR + b*NAT + c*AFR.

With the result of the linear model it is possible to write each individual's heterozygosity as a linear combination of the expected heterozygosities of each ancestral population. We then compared the heterozygosity of each ancestral population in the Uruguayan one, with the heterozygosities of non-admixed populations, with different ancestries.

Fst, SmartPCA, Adxmiture Graphs and TreeMix

Uruguayan native genomes were assessed at chip positions. RFMix was run on chip positions including 525 individuals +10

"Charrúa," considering the Europeans, Africans and Natives from the chip as reference population. Native tracts were considered for these further analyses.

Fst was calculated only for South American tribes: Toba, Wichi and Chane (Gran Chaco region), Diaguita, Quechua and Aymara (Andean region), Arara and Kaingang (Brazilian Planalto), Guaraní, Surui and Karitiana (Amazonian region), Guahibo and Piapoco (Orinoco Plains), Chilote (Patagonian region) and URU (Pampa region). Only variants with high coverage of native tracts were considered, since Fst calculations are sensitive to extreme frequency values. We tried different coverage values, and decided to have at least 6 native haplotypes. Four haplotypes gave similar results regarding Fst/f3 statistics but produced worse p-values (**Supplementary Figure S8**). Arara was excluded from the analysis since this group had very few individuals and allele frequencies were biased.

For SmartPCA and admixture graphs (qpgraph), both from admixtools, we generated pseudo-haploids. For the positions in the genome with half-missing ancestry calls (ie., one SNP native, the other one EUR/AFR hence set to missing data, which affects several positions in the genome) we "copied" that indigenous SNP to that homologous missing position (it can be the genomic reference or alternative). For the positions with both missing alleles, we left the missing data character. We obtained an average of ~60% of missing data among the "Charrua" individuals. The missing data positions range from 332,410 to 152,707 covering different genomic regions (from 363,578 chip SNPs).

Those pseudo-haploids were used to generate the PCA and the admixture graph to compare only native tracts of Uruguayans together with other natives from the chip.

For the admixture graph we proceeded as follows. First a subset of 9 Southern Natives were considered to build a reasonable topology of the graph (Quechua, Diaguita, Wichi, Toba, Guarani, Kaingang, Chane, Surui, Karitina) according to historical records, Those topologies had 2 outliers with the same Z-score of 3. Next, we tried to add the "Charrúa" branch, first without admixture and then with. The best fitting topology is the one presented in **Figure 5C**. An equivalent one, with the same Z-score is in **Supplementary Figure S9**.

To evaluate the effect of the pseudohaploids in the results of the admixturegraph we masked in turn each population included in the graph as follows: 50% was set to half-missing calls and then pseudohaploids were constructed. This was done randomly 20 times for each population. Then, admixture graphs were constructed on the same topology but with less "accurate" genomic information (more pseudohaploids). We reported the Z-score for every iteration.

We found that pseudohaploids are robust for the construction of the admixture graph for populations with sufficient number of individuals (more than 2) and that are not very distant different from the rest of the populations under consideration (eg., very long edges). The populations that we expected to work well on simulations were Guaraní, Quechua, Kaingang, Diaguita and Chane. Their distributions of Z-scores are close to the original value (**Supplementary Figure S10**). The exception is Chane that has a wide distribution but has only two individuals.

However, Wichi, Toba, Surui and Kaingang, are very different populations, some even highly endogamous (eg., Surui). Masking 50% of those individuals and doing pseudohaploids results in more differences, hence we expected to have high Z-score in the admixture graphs (**Supplementary Figure S10**).

TreeMix (Pickrell and Pritchard, 2012) was run using the data set in 2.3 for the ADMIXTURE with the microarray data (364,398 SNPs) with Aleutian as root and the amount of admixture edges ranging from 1 to 3 (m parameter).

RESULTS

High Indigenous Signal is Observed in Uruguayans With Self-Declared Indigenous Ancestry

Ten individuals with self-declared known Charrúas ancestry were included in the study. Family information and historical records were validated to include each sample. Samples had a very high quality as seen by the high throughput of each run (mean of 801542015 reads), low number of duplicated reads, high percentage of mapped reads (mean of 94.81% on the de-duplicated reads) and high percentage of reads with QC > 30 (mean 83.95%).

Principal component analysis (PCA) of the samples was carried out using as a reference set, individuals from the 1000 Genomes Project (The 1000 Genomes Project, 2015), Simons Diversity Project (Mallick et al., 2016).

In the PC1/PC2 space, Charrúas descendants were placed on the European-Asian/Amerindian axis (blue triangles), consistent with a high degree of admixture degree (**Figure 1A**). In higher PCs (**Supplementary Figure S1**), the "Charrúas" samples locate more closely towards other Amerindian reference samples than Asian samples. Hence, it is shown that the sampled individuals have admixture with Amerindian, proving that there is still indigenous signal in this population. **Figure 1C** shows global ancestry estimations using ADMIXTURE (K = 4) of the 10 individuals, which are consistent with the PCA, showing a significant indigenous signal on the Uruguayan genomes (light blue color bars; higher K values are shown in **Supplementary Figure S2**; comparisons with different methods are detailed in **Supplementary material S3**).

We next carried out local ancestry inference using a three-way model of admixture using RfMix. **Figure 1B** shows local ancestry estimations for two example individuals (one with the most Amerindian ancestry and the other with the most African ancestry; full results, **Supplementary Figure S3**). One of the individuals shows a considerably high proportion of indigenous ancestry (top, in green), while the other has a more even contribution of all three ancestries (bottom, AFR, EUR and NAT).

Additionally, we determined mitochondrial and Y-chromosomes haplogroups. Six of the mitochondrial haplogroups found are typical of living Amerindian individuals: B2e, B2 (n = 2), B2b, C1d and C1b; three European: J1c1f, V, J1c and one African: L1b1a. None of the
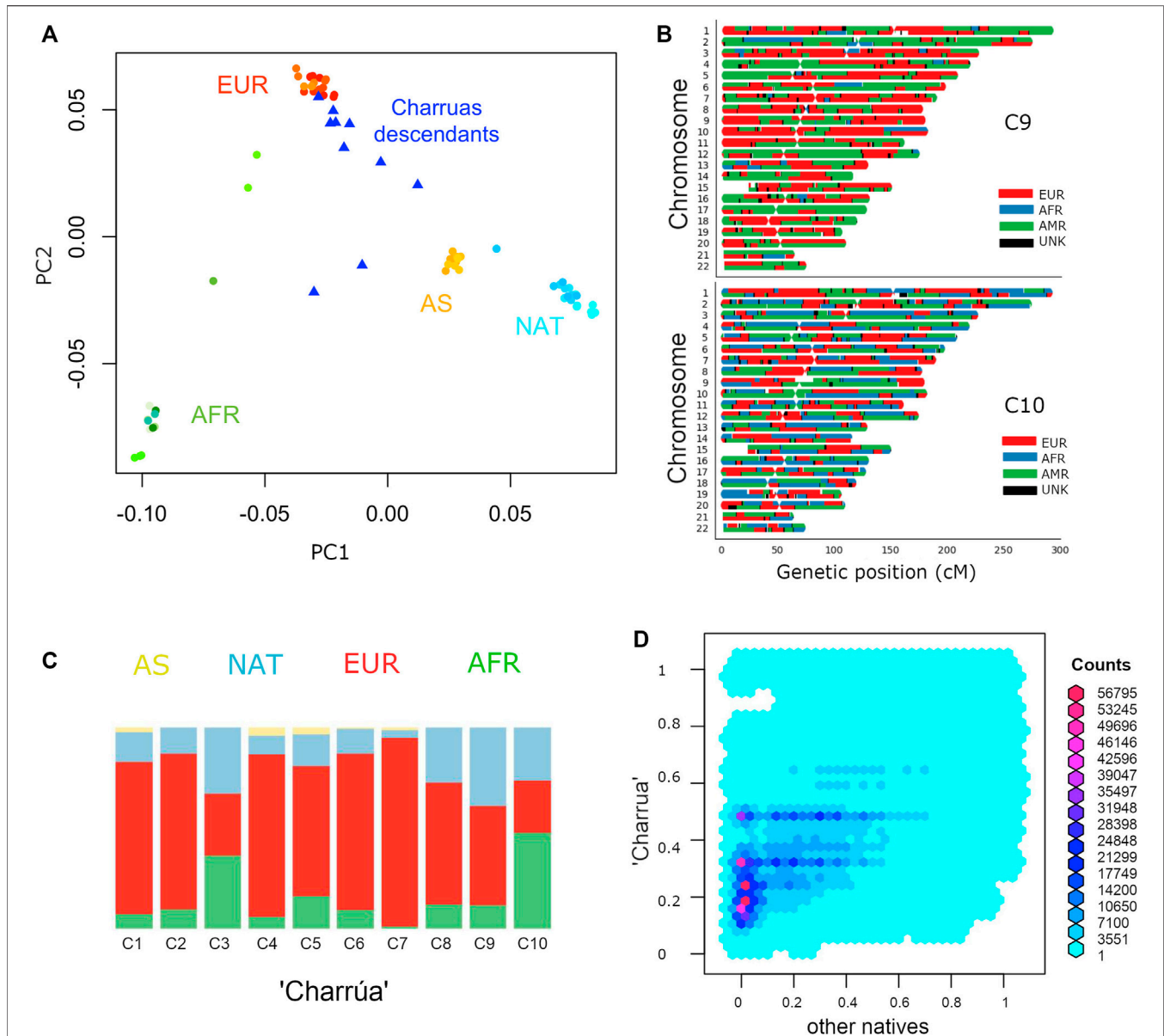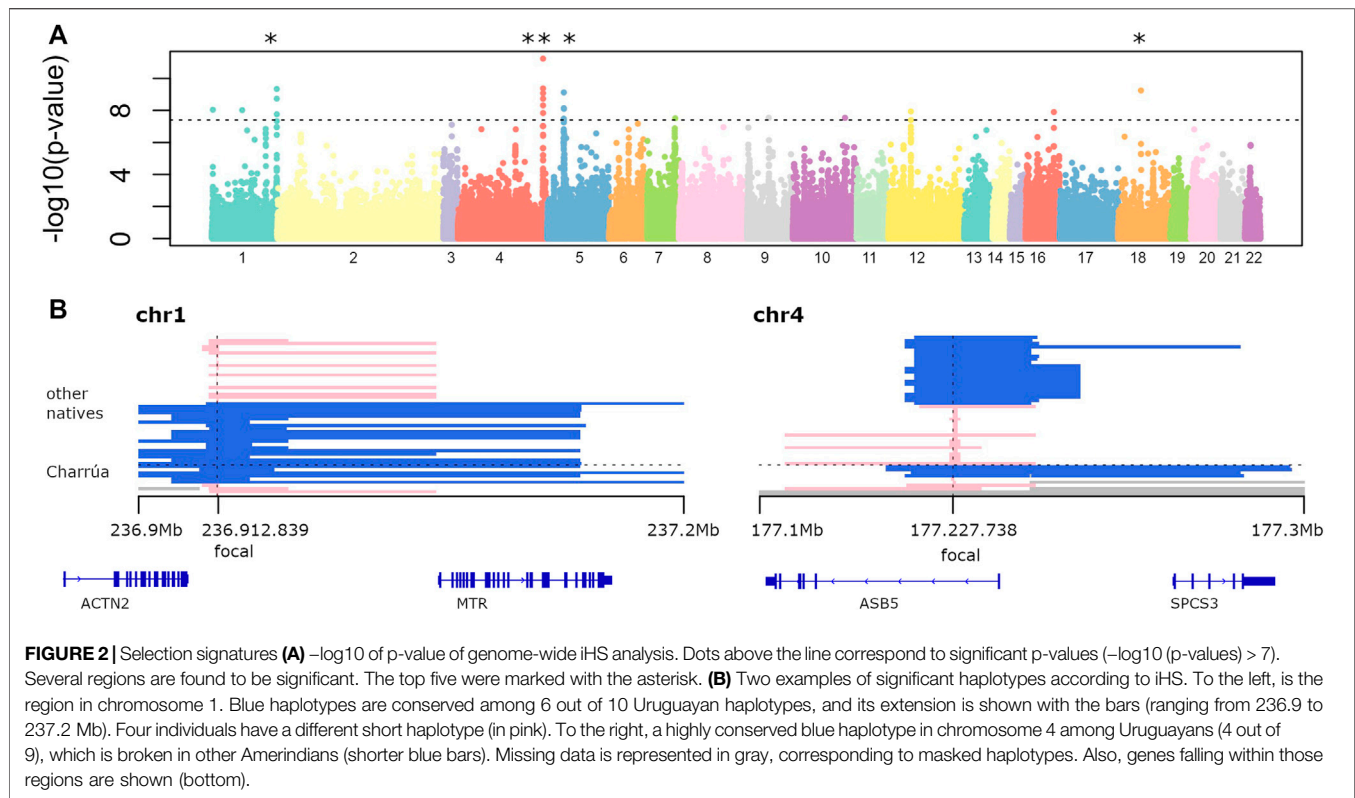
FIGURE 1 | Admixture of Uruguayan samples. **(A)** PC1 against PC2 of a PCA with complete genomes from different populations: Amerindian (cyan), Europeans (red), Asians (yellow), Africans (green) and "Charrúa" descendants (blue triangles). **(B)** Karyograms for two sample individuals showing results from local ancestry inference using a three-way model of admixture. Colors indicate the most probable local ancestry assignment (red: European, blue: African, green: Amerindian). **(C)** Admixture plot corresponding to k = 4. Reference populations were the same as in A: Africans (green), Asians (yellow), European (red) and Amerindian (light blue). 10 bars corresponding to the "Charrúas" descendants. **(D)** Frequency comparison between haplotypes coming from Uruguayan Amerindian and the rest of the indigenous groups. Reddish hexagons, correspond to a high number of haplotypes corresponding to that pair of frequencies. Cyan and blue hexagons correspond to less haplotypes for that pair of frequencies.

canonical Amerindian haplogroups for the Y chromosome were observed.

Frequent "Charrúas" Variants and Haplotypes are Rare in Other Populations

A total of 4524317 genome-wide variants (according to reference genome GRCh37) were found within "Charrúas" tracts (at least present in one individual). Of those variants (within indigenous tracts), 809497 were highly frequent in observed "Charrúas" tracts (frequency>0.8), and of those, 105 were very rare in the rest of the world (frequency <0.01 in 1000Genomes and gnomAD). **Supplementary Table S1** compares the frequency of all "Charrúa" variants with the rest of the world (1000G). The 105 globally rare variants fall into the following functional categories: 62 intergenic, 32 intronic, 6 ncRNA intronic, two downstream (variant overlaps 1-kb region downstream of transcription end site), one lncRNA, one 3′UTR and one 5′UTR.

FIGURE 2 | Selection signatures (A) –log10 of p-value of genome-wide iHS analysis. Dots above the line correspond to significant p-values (–log10 (p-values) > 7). Several regions are found to be significant. The top five were marked with the asterisk. (B) Two examples of significant haplotypes according to iHS. To the left, is the region in chromosome 1. Blue haplotypes are conserved among 6 out of 10 Uruguayan haplotypes, and its extension is shown with the bars (ranging from 236.9 to 237.2 Mb). Four individuals have a different short haplotype (in pink). To the right, a highly conserved blue haplotype in chromosome 4 among Uruguayans (4 out of 9), which is broken in other Amerindians (shorter blue bars). Missing data is represented in gray, corresponding to masked haplotypes. Also, genes falling within those regions are shown (bottom).

Of the original 4524317 variants only 26628 were exonic and the rest non-coding, including splicing, UTRs, intronic, intergenic, etc. Of the total, only one was novel (not reported in dbSNP version 151) in chromosome 14 and genomic position 48780303, an intergenic variant.

We also assessed patterns of haplotypic variation. Haplotypes were defined on the basis of non-overlapping windows of 100 variants and their frequency were determined for the "Charrúa" population and the rest of the indigenous groups.

Figure 1D compares the frequency of such haplotypes in "Charrúa" versus all other indigenous groups. Enrichment of reddish hexagons are seen with low frequencies in other indigenous groups and in higher frequencies in "Charrúa." This could imply that "Charrúa" might have characteristic, private haplotypes, which might not be present in the other indigenous groups.

In order to see if some of these haplotypes may be the result of recent positive selection, we analyzed signatures of selective sweeps within the "Charrúa" individuals. Taking only the indigenous tracts into account, iHS analysis was performed (Voight et al., 2006).

Figure 2A shows the log transformed p-values of the iHS analysis for each "indigenous" position in the genome. The locations of the top 30 iHS values are reported in **Supplementary Table S2** together with the genes located within 1 Mb distance, and we further investigate the top five of these regions (marked with asterisks in **Figure 2A**).

One is found on chromosome 1 (0.3 Mb long), which is present in 6 out of 10 Amerindian haplotypes. It is also found in the rest of the indigenous groups: Karitiana (4), Quechua (1), Zapotec (1), Mixe (3), Mayan (1), Piapoco (1), Chane (1).

Genes falling in that region are ACTN2 (only partially, the 3′ end) and the complete MTR gene. ACTN2 gene encodes a muscle-specific, alpha actinin isoform that is expressed in both skeletal and cardiac muscles. It is related to RET signaling and signaling by GPCR pathways (Rebhan et al., 1997). The MTR gene encodes the 5-methyltetrahydrofolate-homocysteine methyltransferase. This enzyme catalyzes the final step in methionine biosynthesis and is involved also in folate metabolism (Rebhan et al., 1997).

The second large iHS value region (with two significant focal SNPs) was found on chromosome 4 with a 0.2 Mb extended haplotype (**Figure 2B**, right panel), that is found in 4 out of 9 observed 'Charrúa' haplotypes (see *Detecting Signatures of Selection*).

The extended haplotype is also found in Karitiana haplotypes (5), Quechua (2), Zapotec (3), Mixtec (1), Mixe (2), Mayan (3), Pima (1), Piapoco (2), Surui (2) and Chane (1).

The other 4 Uruguayan individuals have a smaller haplotype, and it is conserved among a smaller fraction of indigenous groups: Quechua (1), Maya (1), Chane 1). Genes falling within that region are ASB5 and SPCS3. ABS5 is a protein coding gene, related to the pathways of the Innate Immune System and Class I MHC mediated antigen processing and presentation. SPCS3 is also a protein coding gene, related to viral mRNA translation and gene expression (Rebhan et al., 1997). A further region on chromosome 5 had a focal significant SNP in position 33,694,366, which covers the complete ADAMTS-12 gene,
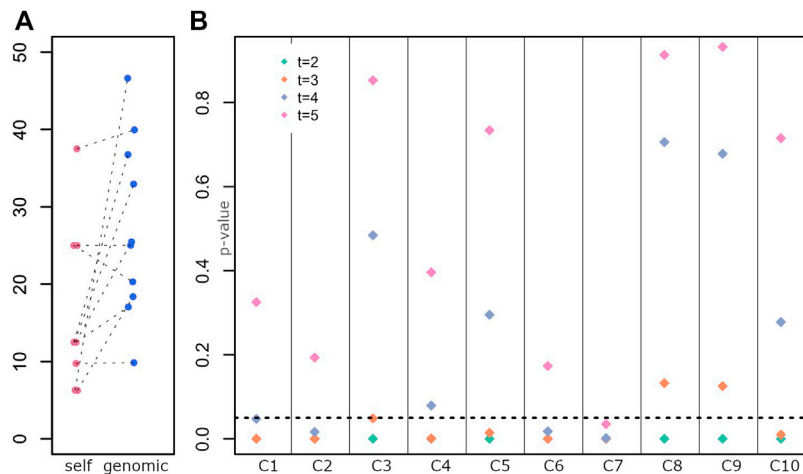
FIGURE 3 | Determination of the generation of the last full indigenous ancestor **(A)** Distribution of self-declared Amerindian ancestry (self) and genomic estimates (genomic). Each dot is an individual. **(B)** Test of a model with a full indigenous ancestor in each generation for each individual. *p*-values for rejecting the null hypothesis are on the y-axis, and individuals are displayed along the x-axis.

one of the 19 members of the ADAMTS family of secreted zinc metalloproteinases (Kelwick et al., 2015). The 1 Mb region surrounding intergenic focal SNP in position 7,127,444 on chromosome 18 includes LAMA1, LINC00668, ARHGAP28, LOC112577592 genes.

To investigate whether the low coverage of indigenous tracts per SNP in the calculations of the iHS values might generate artificial extreme values, we compared the fraction of observed Amerindian haplotypes per locus with the calculated iHS values (**Supplementary Figure S4**).

Even though we observe a range compression in the highest frequencies (from 0.7 to 0.8), there is no general tendency of range compression from the lower coverage frequencies towards the intermediate frequencies. Hence, there is no obvious bias in the results due to the uneven/low Amerindian coverage. A similar approach was applied in (Gautier and Vitalis, 2012), by using the proportion of a specific ancestry to help in the detection of selection signatures.

Correlation Between Self-Declared Ancestry and Genomics Estimations

The average self-declared indigenous ancestry in the sample was 15%. When compared to the genomic estimation of ancestry, the self-declared ancestry is on average lower (**Figure 3A**, paired t-tests p-value < 0.03).

Five individuals have a relatively low self-declared Amerindian ancestry and a high genomic estimation, with a mean difference between both ancestry estimates of 22%. Those individuals potentially have indigenous ancestry from both matrilineal and patrilineal branches. **Figure 1B** shows the karyogram of two individuals, where homozygous tracts of indigenous ancestry along the whole genome can be observed (similar homozyous tracts are found for other individuals, **Supplementary Figure S2**). Additionally, two individuals

who declared European maternal ancestry have mitochondrial DNA from Amerindian origin.

Given the difficulty to identify the last ancestor with full indigenous ancestry for each individual from historical records, we developed a test (see 5.5) that tests, for each generation, whether the individual has a full indigenous ancestor at that generation or earlier in time. For each individual and each generation, we tested whether a full indigenous ancestor exists (null hypothesis). If p-values are smaller than 0.05, the null hypothesis is rejected, hence there is a low probability of the observed data if a full indigenous ancestor was present at that generation (for that individual).

Figure 3B shows the p-value for rejecting the null hypothesis for each generation (t = 2, 3, 4 and 5) and each individual.

While for few individuals we reject models where the full ancestor was 4 or 5 generations ago (C2, C6, C7), for most individuals, we reject models where the full ancestor was 2 or 3 generations ago (C4, C5, C1, C10), suggesting an ancestor around ~1830 or earlier, e.g., a great-great grandparent. In some cases, we could not reject an ancestor 3 generators ago (C3, C8, C9), consistent with a more recent ancestor (e.g., towards the end of 19th century, a great grandparent).

These results agree with the genomic estimations of Amerindian contributions (C3, C8, C9 have the highest indigenous ancestry, **Figure 1C**).

Inference of European Contact and Subsequent Migrations

Since recombination events in each generation break down chunks of genome coming from the same parental chromosome, the length of resulting tracts assigned to distinct ancestries in admixed genomes might be informative of the migration modes and times (Pool and Nielsen, 2009). To

TABLE 1 | Two migration models were tested. 1) EUR, NAT + AFR is a simple model were NAT has first contact with EUR in time T_1, with a subsequent pulse of AFR migration in T_2; 2) EUR, NAT + AFR + EUR is model with a first NAT and EUR contact at time T_1, followed by an AFR pulse at T_2 and with a subsequent EUR pulse at T_3. Log-likelihood of each model is given. TRACTS estimates the generations (in GA units) of the first NAT-EUR contact, which is given under GA.

	EUR, NAT + AFR	EUR, NAT + AFR + EUR
log-likelihood	−171.56	−174.26
αNAT_prop	0.29	0.29
τtime_(EUR,NAT)	8.3	8.5
τtime_(AFR)	6.7	7.8
τtime_(EUR2)	NA	0.36
GA	11	11

FIGURE 4 | Representation of best fitting migration model. Top panel shows the number of tracts vs the tract length. Observed distribution of ancestry tracts is represented by scatter points, while the solid-colored lines represent the distribution of the model, with shaded areas indicating 68.3% confidence intervals. Models implemented in TRACTS were used to find the best model that fits the observed data. Admixture time estimates (in number of generations ago), migration events, volume of migrants, and ancestry proportions over time are given for each population under the best-fitting model.

explore the timing of European and African contact with the Amerindians in the region, the length distribution of continuous ancestry specific tracts of the ten individuals was considered. According to historical records, the Europeans (Spanish and Portuguese) came first to the present-day Uruguay, followed by probably more Europeans along with Africans, who were brought as slaves. Hence, in this study, two different migrations scenarios were evaluated: 1) EUR, NAT + AFR, meaning a first contact of Europeans and Amerindians, followed by a pulse of Africans, and 2) EUR, NAT + AFR + EUR, meaning European and Amerindian contact, followed by a pulse of Africans, followed by a subsequent pulse of European migrants. Model 2) has two additional parameters, corresponding to time and proportion of the subsequent European migration.

Table 1 shows the log-likelihood for both fitted model, with estimated parameters including the number of generations ago for admixture (GA) for both models. We performed a likelihood ratio test to compare the nested models, 2 * (174.26–171.56) = 5.4, which under a chi-square distribution (df = 2) has a p-value of 0.033. Hence, we selected the model 2) for further consideration (Figure 4).

The migration matrix is represented in the bottom panel of Figure 4 and reveals that the first European contact with the Uruguayan "Charrúa" was 11 generations ago, which corresponds to around the mid-end of XVII century (1658 with formula 1958 - (GA + 1) * 25 as in (Fortes-Lima et al., 2017), being 1958 in this case the median of all participant's birth dates. GA: generations ago). The second European migration contributed 30% to the population and was 10 GA ago (1683). The first African pulse was 10 GA (1683), with an ancestry contribution of 7%. The second one was 9 GA with an ancestry contribution of 1%. Of note, the results obtained with MALDER are similar to the TRACTs approach for the first Indigenous-European and Indigenous-African contact: 8.26353±2.07157 (Z = 3.98902) for both.

To further investigate the contribution of different migrants to the Uruguayan Amerindians we determined the expected heterozygosity according to the ancestral contributions. For this we did a linear model with the ancestral proportions as the independent variables and heterozygosity as the observed
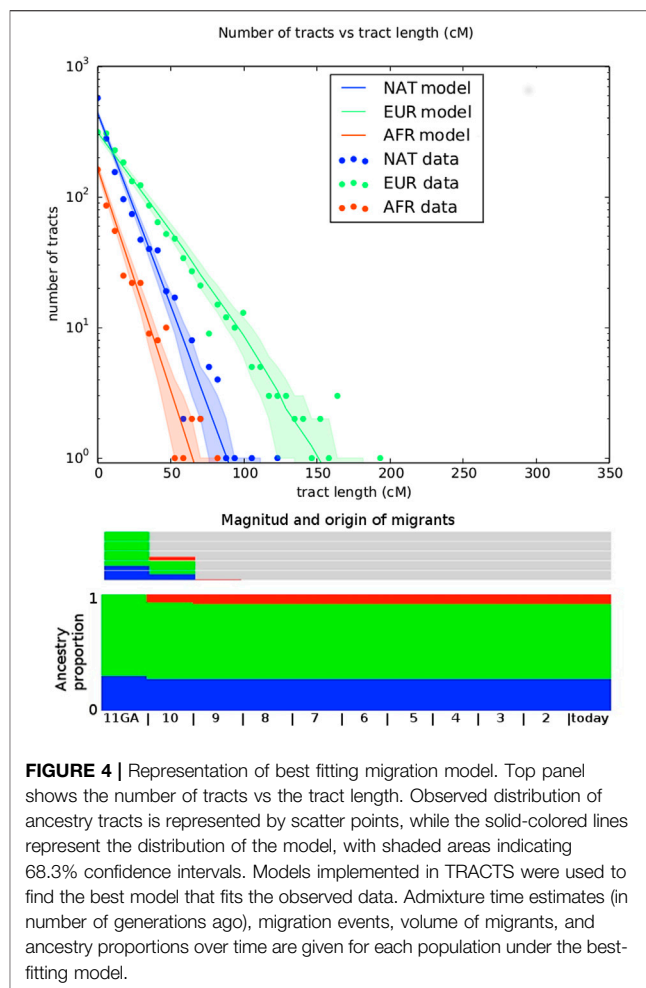
variable (see Analysis of Heterozygosity). The calculated coefficients for the Amerindian, European and African, are 0.29, 0.31 and 0.46, respectively (Supplementary Figure S5B, first left panel). Hence, the African ancestry component has the larger effect on the heterozygosity of the Uruguayan Amerindians individuals studied here. Also, we compared the heterozygosity of individuals with full European, Amerindian and African ancestry, with the determined heterozygosities by the linear model (for each ancestry component) of Uruguayan indigenous descendants. We observed that those European and Amerindian heterozygosities in Uruguayans are comparable with those of individuals of that ancestries. Sardinians have almost the exact same heterozygosity as the European component of Uruguayans and, the "Amerindian" heterozygosity of Uruguayans lies between Mayan and Karitiana. However, African heterozygosity in Uruguayan individuals is larger than African individuals such as the YRI (Yoruba), LWK (Luhya in Webuye, Kenya) samples (Supplementary Figure S5A). This means, on the one hand, that the high heterozygosity in Uruguayan individuals is mostly due to the African ancestry and on the other, that African component in Uruguayan sample in highly diverse.

Genomic Similarity Between "Charrúa," Kaingang, Diaguita and Guaraní

The similarity of "Charrúas" to other South American tribes and the history of admixture for "Charrúas" are unknown. The only available information comes from relatively scarce historical records and oral histories. To assess similarity between "Charrúas" and other South American tribes we used the data collected here along with that of (Reich et al., 2012). The latter is a merged dataset describing 525 individuals from different populations (214 European, 109 African, 202 Amerindian), including 169 individuals from 14 South American populations, with a final genotype set covering 364K SNPs. Guaraní (from the Paraguay-Brazil and Paraguay-Argentina border), Aymara, Quechua, Chilote, Diaguita, Kaingang, Wichi, Toba, Guahibo, Piapoco, Karitiana and Surui were some of the represented populations.

To analyze the ancestry components in the "Charrúas" descendant genomes collected here, the ADMIXTURE method was used, with K ranging from 3 to 19. The higher the number of clusters K, the finer the structure that can be analyzed. With K = 19 all "Charrúas" had a high contribution of three mostly European clusters (predominately represented by TSI, Sardinian and Russian individuals) and two mostly African clusters (predominantly, YRI and LWK, **Supplementary Figure S2**). The 14 remaining clusters correspond mostly to indigenous groups. **Figure 5A** shows the classical admixture plot (with K = 19) but without the European/African contribution, to allow a closer inspection of the Amerindian ancestry of our sample. A shared cluster is observed in almost each "Charrúas" individual (indicated in blue in the figure). That cluster is mainly shared by Wichi, Ticuna, Toba, Guahibo and additional groups with more complex ancestry (Piapoco, Chane, Guaraní, Diaguita and Yaghan). A cluster (light orange) shared with Andean groups (Aymara and Quechua), is present in some "Charrúas" descendants (C1, C5 and C8) but in a smaller proportion than the blue one. Only three Charrua descendants have a non-negligible contribution of a cluster (green), which is shared mostly with Mexican indigenous individuals (Zapotec, Mixe, Mixetec, and Maya) though present also in small amounts in indigenous groups geographically closer to Uruguay (e.g., Diaguita). Several of the individuals share some ancestry with two clusters (dark orange and peach), that is shared with Amazonian groups (Karitiana and Surui). Finally, a small proportion of a cluster (purple) shared with Pima individuals is found in several of the Charrúa descendant individuals. These results show that this sample is generally homogeneous with respect to its indigenous genetic background (One individual, C7, is a visual outlier, though we note the ancestry proportions may be more difficult to infer as this individual was inferred to have ~9% indigenous ancestry).

Additionally, we did a PCA on the Amerindian ancestry tracts using PCAadmix (Brisbin et al., 2012) with local ancestry called by RFMix (**Figure 5B**). The PCA shows "Charrúas" descendants located between the distribution of Guaraní/Kaingang (red/light orange diamonds) and Diaguita (green triangles) individuals, suggesting closest similarity with those groups. Wichi/Toba

and Guahibo/Piapoco groups remain in distant positions in the PCA space with respect to the "Charrúas" descendant individuals. Moreover, the group Kaingang, Guaraní and Diaguita had the lowest F_{ST} values with the "Charrúas" descendants (**Supplementary Figure S6**).

Using the same data, admixture graphs were estimated with qpgraph (Patterson et al., 2012). **Figure 5C** shows a plausible graph for the current data set [Z score = −3.197, corresponding to the following D-statistic outlier: (Surui, Kaingang), (Wichi, Guaraní)]. Admixture branches did not improve the score. In the graph the "Charrúas"are positioned among current Brazilian Amerindian groups; specifically, the "Charrúas" are an outgroup to a subtree with Karitiana and Surui, with Kaingang as the most immediate outgroup to the Charrúa, Surui, Karitiana clade. The branch lengths (measured in units of genetic drift) suggest a star-like topology where many of the ancestral nodes are close to the root.
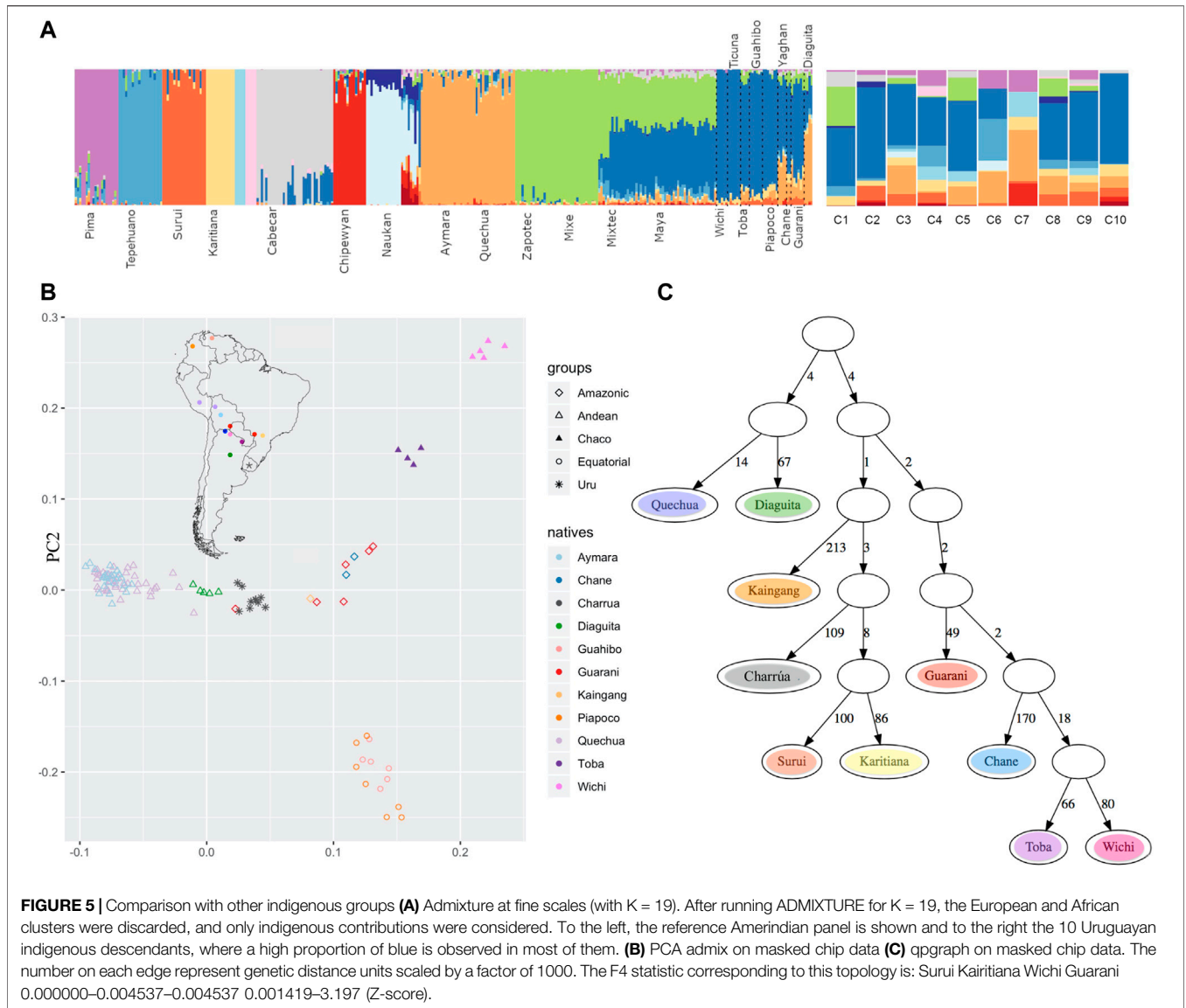
DISCUSSION

We found a substantial indigenous genomic contribution in the "Charrúas" descendants that participated in this study, despite there are no isolated indigenous communities living in Uruguay in the present.

These individuals have twice the estimated indigenous ancestry as the expected for the general Uruguayan population and is in most cases also higher than the estimations for Montevideo (around 15%), where the individuals were sampled (Bonilla et al., 2004) (ancestry estimations range from 9% to over 40% with a mean of 27%).

In general, the overall level of Amerindian ancestry was higher than expected from self-report in this sample. Particularly, in most cases, individual genomic data was consistent with indigenous ancestry on both parental lines, which was remarkable as each participant self-reported indigenous ancestry only from a single paternal side. This is consistent with the underestimation of indigenous ancestry in the general Uruguayan population (Sans, 2011), and is perhaps notable that this underestimation occurs also in this set of individuals that self-identify with Charrúan heritage. Two individuals also had a high proportion of African ancestry (0.21 and 0.30) though African ancestry was generally not reported by the participants, probably meaning that their sense of identity was more related to their Charrúan ancestry, or the admixture event was sufficiently in the past to be unknown to them.

As previously shown, in the post-colonial period matings that occurred in Latin America were mostly between Amerindian women and European men (Ongaro et al., 2019). In addition, indigenous genocide in Uruguay ("Salsipuede's") exterminated the majority of "Charrúas" men. Thus, we found indigenous mtDNA haplotypes in 6 out of the 10 individuals, while none of the five men had an indigenous Y-chromosome (R-DF27*, R-S127*,R-Z49, I-Z140*, R-DF27*).

The indigenous mitochondrial haplogroups found in this sample were B and C, which are typically found in Uruguay (Bonilla et al., 2004). Figueiro and Sans (The 1000 Genomes

FIGURE 5 | Comparison with other indigenous groups **(A)** Admixture at fine scales (with K = 19). After running ADMIXTURE for K = 19, the European and African clusters were discarded, and only indigenous contributions were considered. To the left, the reference Amerindian panel is shown and to the right the 10 Uruguayan indigenous descendants, where a high proportion of blue is observed in most of them. **(B)** PCA admix on masked chip data **(C)** qpgraph on masked chip data. The number on each edge represent genetic distance units scaled by a factor of 1000. The F4 statistic corresponding to this topology is: Surui Kairitiana Wichi Guarani 0.000000–0.004537–0.004537 0.001419–3.197 (Z-score).

Project, 2015) discussed the difference of origins between the haplogroups A, B and C, with A being more commonly found in Amazonian peoples (particularly Guaraní), and B and C more associated with Uruguayan prehistoric sites and possibly Argentinean Chaco or Pampa (in spite of having a broad distribution).

A scan for recent positive selection (using iHS) showed several regions that might be under selection.

A selection signal was observed in chromosome 4 spanning the ASB5 and the SPCS3 genes. The SPCS3 (Signal peptidase complex subunit 3) gene plays an important role in virion production of flaviviruses such as West Nile virus, Japanese enchephalitis virus, Yellow Fever virus, Zika and Dengue virus type 2 (Zhang et al., 2016). The latter two, most likely present in the Uruguayan region before colonization. A subset of endoplasmic reticulum-associated signal peptidase complex proteins was necessary for proper cleavage of the flavivirus

structural proteins (prM and E) and secretion of viral particles. Hence, those proteins (among them SPCS3) are required for flavivirus infectivity, and these were associated with endoplasmic reticulum functions including translocation, protein degradation, and N-linked glycosylation. The ASB5 gene is a member of the ankyrin repeat and SOCS box-containing (ASB) family of proteins. The SOCS box serves to couple suppressor of cytokine signalling (SOCS) proteins and their binding partners with the elongin B and C complex, possibly targeting them for degradation. It is related to Class I MHC mediated antigen processing and it is involved in the innate immune system (Rebhan et al., 1997).

The gene region within the selected haplotype, which is present in 4 out of 9 Uruguayan haplotypes, presents almost no variants, meaning that in our sample mostly the ancestral allele is conserved. Only one variant is found at position 177253006, which has no frequency reported in the gnomAD

database (Ongaro et al., 2019). If this variant could have implications in immune system evolution and adaptation for Uruguayan indigenous groups cannot be stated without further studies.

A region in chromosome 1 was also found, that is highly frequent in Uruguayan Amerindians and other indigenous groups considered in this study. The genes included in this region are ACTN2 and MTR. The MTR gene is an enzyme that is involved in methionine biosynthesis. Methionine synthase helps convert the amino acid homocysteine to methionine (Oohora and Hayashi, 2016). The ACTN2 gene encodes alpha-actinin-2, which is highly expressed in the sarcomeric Z-disk or Z-line in cardiac and skeletal muscle. It plays an important structural and functional role in the sarcomere and contractile apparatus (Lornage et al., 2019). The region found in chromosome 5 includes gene.

ADAMST-12, that performs essential roles in modulation and recovery from inflammatory processes such as asthma, colitis, endotoxic sepsis and pancreatitis (Mohamedi et al., 2021). ADAMTS-12 has also been involved in cancer development acting either as a tumor suppressor or as a pro-tumoral agent (Rabadan et al., 2020). One of the genes included in chromosome 18 region is LAMA1, which encodes one of the alpha 1 subunits of laminin. The laminins make up a major component of the basement membrane and have been implicated in a wide variety of biological processes including cell adhesion, differentiation, migration, signaling, neurite outgrowth and metastasis (Rebhan et al., 1997).

Of note is that only 20 chromosomes were available for such analyses and only adaptation events such as pre-admixture hard-sweeps might be detected. Other types of signals are very difficult to assess using this approach.

According to our estimated time frame of admixture, the first European-Amerindian was around 1660 and the second around 1680. In 1516 the Spanish Juan Díaz de Solís, reached the (now) Uruguayan coast for the first time. During the 16th century, forts and villages were founded by the Spanish, however, none of them survived more than 5 years. Thus, no admixture persisting to today was expected during that period and this was confirmed by the genomic data. Interestingly, the first estimated admixture event concurs with the establishment of the Franciscan mission, Santo Domingo de Soriano, founded in 1662–1664. According to historical records, allegedly mostly Chaná (a different indigenous group with uncertain relationship to the Charrúas) lived in the Santo Domingo de Soriano Franciscan mission (Acosta Y Lara, 1985a). This could imply that, either "Charrúas" were also living together with Chaná and Franciscan individuals, or simply that we cannot distinguish between these indigenous sub-populations with our samples yet, or even that we cannot determine the exact time frame of the contact (a change in one generation and/or using a different assumed generation length can move the estimated date by more than 2 decades). In the same line, the second migration pulse in ~1680, tentatively concurs with the foundation of the Nova Colônia do Santíssimo Sacramento, that brought together according to historical records, more than 400 soldiers, around 60 African slaves, and some Amerindians (Azarola, 1940).

Using the TRACTS model, the European genomic contribution in the first generation of admixture was estimated as 70% with 30% of indigenous contribution. This over-simplifies the probably real historical process, where Amerindian were the vast majority in the region (at least until 17th and probably also 18th century, but for the latter there is scarce data regarding population numbers), only few admixed with Europeans and the rest of the unadmixed Amerindian probably left no footprint in the genome of living descendants, most likely due to the massive genocide (Salsipuedes). Further, the TRACTS model fit a second European wave 10 generations ago (approximately 1683) that contributed 30% to the population already living in the territory. Additionally, the first inferred African pulse was also ten generations ago. These data may suggest that the first pulse of African slaves came here together with the Iberian soldiers. Even though historical reports show several thousand of slaves came after the foundation of Colonia, their genomic contribution is inferred genetically to be low, around 7%. This could reflex the social behavior during slavery times. A second African pulse was also estimated in 1708, contributing only 1% to the gene pool of the historical population. This may correspond to a further pulse of African slaves coming from further Spanish ships. In 1726 was the foundation of the city of Montevideo and even though, according to historical records, only Spanish participated in this process, a couple of years later due to the high activity of a port city, slaves were brought as workforce to keep up with the labor that the city required (Bracco et al., 2012). However, the process of integration of American populations in general and in Uruguayan ones in particular was very complex, and after the first episode of admixture of each group, admixture took place not only between the main groups but also between the admixed people themselves.

Uruguayan indigenous group is placed between Andean and Amazonic natives in the PCA. The closest populations from those groups are Diaguita (Andean) and Guaraní and Kaingang (Amazonic). This genomic similarity between these groups can be explained based on admixture events or shared ancestors. In particular, a Guaraní signal was expected, since different sources have shown that Guaraní arrived in Uruguay around the 13th century, and they continue coming during historical times. Most of them came from the Jesuitic Missions, as mentioned before. Nevertheless, the signal was lower than expected, probably due to the heterogeneity (two different locations) and the small size of the Guaraní sample. Both facts could weaken admixture signals.

Kaingang, despite belonging to different linguistic families and having some degree of genetic differences (Petzl-Erler et al., 1993), lived close in southern Brazil, a region also inhabited by Charrúa. Also, kidnaps and exchanges, especially of women, were common practices in different Amerindian groups during the historical period (Bracco, 1998). These facts, together with close geographic distances, probably favored admixture events between Guaraní, Kaingang and "Charrúas."

In contrast, the Diaguita similarity with Uruguayan indigenous descendants (and also to Guaraní and Kaingang), was at first, unexpected. However, the admixture between "Charrúas" and Diaguita, especially from the Argentinean Northwest and Chilean "Norte Chico" (groups from the

Chaco) is supported by several ethnohistorical records: Based on linguistics, (Rona (1964) pointed out that the "Charrúas" languages (as well as Minuán and Chaná) are related to Chaquean languages, specially the Vilela language group.

Additionally, Alcides D'Orbigny already in the 19th century, associated Charrúa together with Tobas and Mbocobíes for their customs, features and language (D'orbigny, 1839).

More recently, we found several records that support the Charrúa presence in the Chaco (Sanchez, 2007). Between the Chaco and today's Uruguayan territory, the "Argentinian Mesopotamia", there is evidence of Charrúa presence in the 18th century in the Cayastá's reduction (Bracco, 2016). Some studies even indicate that the expression "Band of the Charrúas" was used to describe the indigenous groups of the Paraná River (in Argentina) and also, since the end of the 20th century descendants of Charrúa have been establishing in the province of Entre Ríos (Argentina). Moreover, inter-ethnic admixture in the Jesuit reductions has been repeatedly described (Levinton, 2009; Wilde, 2009) and it could have nucleated different ethnic groups with an extensive geographical dispersion.

The admixture graph supports the similarity between "Charrúas" and Kaingang, nevertheless Guaraní and Diaguita appear in different branches of the graph. However, those branch lengths (that are proportional to genetic drift) are small, meaning that a star-like phylogenetic structure could still be possible (only binary topologies are allowed in this approach). Admixture branches did not improve the score of the tree. Also, Diaguita-Guarani and Diaguita-Charrua have low F_{ST} values (**Supplementary Figure S6**), which could also support early day differentiation with subsequent high amount of admixture events, or a later and recent differentiation of these populations. Additionally, Treemix results (**Supplementary Figure S7**) support the relation between Charrúa and the ancestors of Diaguita (Andean group).

Of note, the construction of pseudohaploids to generate those results (qpgraph, Treemix, PCA) might add noise to the derivations since at the positions were there is no SNP information a variant is sampled. Those sources of errors should be taken into account to properly interpret the conclusion.

Not all tribes in South America are currently sampled. There is an under sampling of groups belonging to adjacent regions from the Charrúas, such as Chaná, Minuán and other groups. Hence, similarity and relations to those groups cannot be assessed. Additionally, the groups that were sampled are sometimes small (Kaingang, Chané) and sometimes sampled from very different locations, which can affect heterogeneity of the group, eg., Guaraní. In addition, a caveat is our small sample sizes of Charruan ancestry individuals and the amount of Amerindian information per individual.

The African ancestry component has the strongest contribution to the expected heterozygosity, and higher than expected for contemporary west African samples. Even though a high heterozygosity is expected in African ancestries, due to the evolution of populations within Africa, the calculated heterozygosity in Uruguayan sample is higher than African

populations (**Supplementary Figure S5A**). This could represent the diverse origin of African slaves that came to the territory. Conversely, the founders of European ancestry were most likely less diverse, since Spanish/Portuguese were the populations that invaded the territory. Even though African ancestry is in general low in this sample (0.08) some individuals have high contributions (0.30 and 0.21). For those, the more African ancestry, the highest the heterozygosity (**Supplementary Figure S5B**, right and left panels). No correlation was observed for the other ancestries.

Regarding technical aspects, the relatively small indigenous signal of the genomes causes long stretches of missing data, where European and African ancestry were masked out. Also, this causes a frequency bias, since positions that are covered by few individuals have extreme frequencies, but this is due to sampling and not as a result of evolution. Either this can cause a bias in results (such as F_{ST}) or a loss of information due to different strategies, such as filtering (only work with haplotypes that have at least a 'reasonable' number of haplotypes) or construction of pseudo-haploids, which were both used here.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found online under http://urugenomes.org/lovd/variants.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institut Pasteur de Montevideo ethics committee with reference number IP011-17/CEI/LC/MB. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LS: Raw data curation and bioinformatics analysis, visualization of results, Writing – Original Draft Preparation. MF: Visualization of results, statistical/biomathematical analysis, methodology, Writing – Original Draft Preparation. DA: Interviews with participants, anthropological investigations, Writing – Review and Editing. GI: Statistical methodology, Writing – Review and Editing. GG: DNA extraction, Writing – Review and Editing. J-YS: Data Curation, bioinformatic analysis, Writing – Review and Editing. S-KY: Data Curation, bioinformatic analysis, Writing – Review and Editing. J-SS: Coordination, Resources, Writing – Review and Editing. CR: Supervisor of molecular biology procedures, Writing – Review and Editing. CK: Supervisor of bioinformatic analysis, Writing – Review and Editing. JN: Supervisor of statistical analysis, visualization of results, Writing – Original Draft Preparation. MS: Conceptualization, Project Supervision, Coordination of anthropologist team. HN: Conceptualization, Funding Acquisition, Project supervision, Coordination of the complete project.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.733195/full#supplementary-material

REFERENCES

Acosta Y Lara, E. F. (1985). Salsipuedes 1831 (Los Lugares). *Revista de la Facultad de Humanidades y Ciencias, Ciencias Antropológicas* 1, 65–68.

Acosta Y Lara, E. F. (1985). Salsipuedes 1831 (Los Protagonistas). *Revista Del. Instituto Histórico y Geográfico Del. Uruguay* 26, 73–104.

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/gr.094052.109

Arce, D. (2015). Etnónimos indígenas en la historiografía uruguaya: Desensamblando piezas de diferentes puzzles. *Anuario de Antropología Soc. y Cult. en Uruguay* 13, 23–34.

Azarola, G. (1940). *Historia de Colonia del Sacramento*. Montevideo: Barreiro y Ramos.

Bodner, M., Perego, U. A., Huber, G., Fendt, L., Röck, A. W., Zimmermann, B., et al. (2012). Rapid Coastal Spread of First Americans: Novel Insights from South America's Southern Cone Mitochondrial Genomes. *Genome Res.* 22, 811–820. doi:10.1101/gr.131722.111

Bonilla, C., Bertoni, B., González, S., Cardoso, H., Brum-Zorrilla, N., and Sans, M. (2004). Substantial Native American Female Contribution to the Population of Tacuarembó, Uruguay, Reveals Past Episodes of Sex-Biased Gene Flow. *Am. J. Hum. Biol.* 16, 289–297. doi:10.1002/ajhb.20025

Bracco, D. (2016). Charrúas y aculturación : la primera década en Concepción de Cayastá (1750-1760). *TEFROS* 14, 5–6.

Bracco, D. (1998). *Guenoas*. Montevideo: Ministerio de Educación y Cultura.

Bracco, R., Lopez Mazz, J., Orrego Rojas, B., Batalla, N., and Bongiovanni, R. (2012). *Esclavitud y afrodescendientes en Uruguay Una mirada desde la antropología*. Montevideo: Facultad de Humanidades y Ciencias de la Educación, Universidad de la República. UDELAR, 1–40.

Brisbin, A., Bryc, K., Byrnes, J., Zakharia, F., Omberg, L., Degenhardt, J., et al. (2012). PCAdmix: Principal Components-Based Assignment of Ancestry along Each Chromosome in Individuals with Admixed Ancestry from Two or More Populations. *Hum. Biol.* 84, 343–364. doi:10.3378/027.084.0401

Cabrera, L. (1989). Los 'indios infieles' de la Banda Oriental y su participación en la guerra guaranítica. *Estudos Ibero-Americanos* 15, 215–227.

Curbelo, C. (2009). *Lengua y poder en San Borja del Yi (1833 - 1862). Ensayo para un análisis sociolingüístico*. Linardi y Risso. Montevideo: Primeras Jornadas de Genealogía Indígena del Mercosur, 117–135.

D'orbigny, A. L. (1839). *homme américain (de l'Amérique méridionale) considéré sous ses rapports physiologiques et moraux*. Paris: Pitois-Levrault.

Delaneau, O., Howie, B., Cox, A. J., Zagury, J. F., and Marchini, J. (2013). Haplotype Estimation Using Sequencing Reads. *Am. J. Hum. Genet.* 93, 687–696. doi:10.1016/j.ajhg.2013.09.002

Figueiro, G., Cabrera Pérez, L., Lindo, J., Mallott, E. K., Owings, A., Malhi, R. S., et al. (2016). Análisis Del Genoma Mitocondrial De Dos Individuos Inhumados En El Sitio Arqueológico Cg14E01 "Isla Larga" (Rocha, Uruguay). *Rev. Arg Antrop Biol.* 19, 17. doi:10.17139/raab.19.1.17

Fortes-Lima, C., Gessain, A., Ruiz-Linares, A., Bortolini, M.-C., Migot-Nabias, F., Bellis, G., et al. (2017). Genome-wide Ancestry and Demographic History of African-Descendant Maroon Communities from French Guiana and Suriname. *Am. J. Hum. Genet.* 101, 725–736. doi:10.1016/j.ajhg.2017.09.021

Gautier, M., and Vitalis, R. (2012). Rehh: An R Package to Detect Footprints of Selection in Genome-wide SNP Data from Haplotype Structure. *Bioinformatics* 28, 1176–1177. doi:10.1093/bioinformatics/bts115

González Risotto, R., and Rodríguez, V. (1989). La importancia de las Misiones Jesuíticas en la formación de la sociedad uruguaya. *Estudos Ibero-Americanos (PUCRS)* 15, 191–214. doi:10.15448/1980-864X.1989.1.30514

Gravel, S. (2012). Population Genetics Models of Local Ancestry. *Genetics* 191, 607–619. doi:10.1534/genetics.112.139808

Kelwick, R., Desanlis, I., Wheeler, G. N., and Edwards, D. R. (2015). The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin Motifs) Family. *Genome Biol.* 16, 113. doi:10.1186/s13059-015-0676-3

Levinton, N. (2009). *Guaraníes Y Charrúas : Una Frontera Exclusivista-Inclusivista Espacio Territorio*. Contratiempo, 398.

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Lindo, J., Achilli, A., Perego, U. A., Archer, D., Valdiosera, C., Petzelt, B., et al. (2017). Ancient Individuals from the North American Northwest Coast Reveal 10,000 Years of Regional Genetic Continuity. *Proc. Natl. Acad. Sci. U.S.A.* 114, 4093–4098. doi:10.1073/pnas.1620410114

Loh, P-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., et al. (2013). Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. *Genetics* 193 (4), 1233–1254.

Lornage, X., Romero, N. B., Grosgogeat, C. A., Malfatti, E., Donkervoort, S., Marchetti, M. M., et al. (2019). ACTN2 Mutations Cause 'multiple Structured Core Disease' (MsCD). *Acta Neuropath* 137, 501–519. doi:10.1007/s00401-019-01963-8

Magalhaes, A., and Michelena, M. (2017). Refexiones sobre los escencialismos en la Antropología uruguaya. Una etnografía Invertida. *Conversaciones del Cono Sur* 3 (1).

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., et al. (2016). The Simons Genome Diversity Project: 300 Genomes from 142 Diverse Populations. *Nature* 538, 201–206. doi:10.1038/nature18964

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

Mohamedi, Y., Fontanil, T., Cal, S., Cobo, T., and Obaya Álvaro, J. (2021). ADAMTS-12: Functions and Challenges for a Complex Metalloprotease. *Front. Mol. Biosciences* 8, 378. doi:10.3389/fmolb.2021.686763

Moreno-Mayar, J. V., Vinner, L., de Barros Damgaard, P., de la Fuente, C., Chan, J., Spence, J. P., et al. (2018). Early Human Dispersals within the Americas. *Science* 362, 362. doi:10.1126/science.aav2621

Olalde, I., Brace, S., Allentoft, M. E., Armit, I., Kristiansen, K., Booth, T., et al. (2018). The Beaker Phenomenon and the Genomic Transformation of Northwest Europe. *Nature* 555, 190–196. doi:10.1038/nature25738

Ongaro, L., Scliar, M. O., Flores, R., Raveane, A., Marnetto, D., Sarno, S., et al. (2019). The Genomic Impact of European Colonization of the Americas. *Curr. Biol.* 29, 3974–3986. doi:10.1016/j.cub.2019.09.076

Oohora, K., and Hayashi, T. (2016). Peptide, Protein and Enzyme Design. *Methods Enzymol.* 580, 439–454.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient Admixture in Human History. *Genetics* 192, 1065–1093. doi:10.1534/genetics.112.145037

Petzl-Erler, M., Luz, R., and Sotomaior, V. (1993). The HLA Polymorphism of Two Distinctive South-American Indian Tribes: the Kaingang and the Guarani. *Tissue Antigens* 41, 227–237. doi:10.1111/j.1399-0039.1993.tb02011.x

Pickrell, J. K., and Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-wide Allele Frequency Data. *Plos Genet.* 8 (11), e1002967. doi:10.1371/journal.pgen.1002967

Pool, J. E., and Nielsen, R. (2009). Inference of Historical Changes in Migration Rate from the Lengths of Migrant Tracts. *Genetics* 181, 711–719. doi:10.1534/genetics.108.098095

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795

Rabadan, R., Mohamedi, Y., Rubin, U., Chu, T., Alghalith, A. N., Elliott, O., et al. (2020). Identification of Relevant Genetic Alterations in Cancer Using Topological Data Analysis. *Nat. Commun.* 11 (1), 3808. doi:10.1038/s41467-020-17659-7

Rebhan, M., Chalifa-Caspi, V., Prilusky, J., and Lancet, D. (1997). GeneCards: Integrating Information about Genes, Proteins and Diseases. *Trends Genet.* 13, 163. doi:10.1016/s0168-9525(97)01103-7

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., et al. (2012). Reconstructing Native American Population History. *Nature* 488, 370–374. doi:10.1038/nature11258

Rona, J. (1964). *Nuevos elementos acerca de la lengua charrúa*. Montevideo: Departamento de Lingüística, Facultad de Humanidades y Ciencias, Universidad de la República.

Sanchez, O. (2007). *Historias de los aborígenes tobas del Gran Chaco contadas por sus ancianos*. Roque Sáenz Peña: Acción Apostólica Común Instituto Universitario ISEDET Sociedad Bíblica Argentina.

Sans, M., Figueiro, G., Hughes, C. E., Lindo, J., Hidalgo, P. C., and Malhi, R. S. (2015). A South American Prehistoric Mitogenome: Context, Continuity, and the Origin of Haplogroup C1d. *PLOS ONE* 10, e0141808–14. doi:10.1371/journal.pone.0141808

Sans, M. (2011). National Identity, Census Data, and Genetics in Uruguay. *Racial Identities, Genet. Ancestry, Health South America* 84, 196. doi:10.1057/9781137001702_10

Sans, M., Salzano, F. M., and Chakraborty, R. (1997). Historical Genetics in Uruguay: Estimates of Biological Origins and Their Problems. *Hum. Biol.* 69, 161–170.

Schroeder, H., Sikora, M., Gopalakrishnan, S., Cassidy, L. M., Maisano Delser, P., Sandoval Velasco, M., et al. (2018). Origins and Genetic Legacies of the Caribbean Taino. *Proc. Natl. Acad. Sci. USA* 115, 2341–2346. doi:10.1073/pnas.1716839115

Tavares, G., Reales, G., Bortolini, M., and Fagundes, N. (2019). Measuring the Impact of European Colonization on Native American Populations in Southern Brazil and Uruguay: Evidence from mtDNA. *Am. J. Hum. Biol.* 31, e23243. doi:10.1002/ajhb.23243

The 1000 Genomes Project Consortium (2015). A Global Reference for Human Genetic Variation. *Nature* 526, 68–74. doi:10.1038/nature15393

Vidart, D. (1973). *Diez mil años de prehistoria uruguaya*. Montevideo: Fundación editorial unión del magisterio.

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A Map of Recent Positive Selection in the Human Genome. *Plos Biol.* 4, e72. doi:10.1371/journal.pbio.0040072

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Res.* 38, e164. doi:10.1093/nar/gkq603

Wilde, G. (2009). Territorio y etnogenesis misional en el Paraguay del siglo XVIII. *Fronteiras (Dourados)* 11, 86–106.

Zhang, R., Miner, J., Gorman, M., Rausch, K., Ramage, H., White, J. P., et al. (2016). A CRISPR Screen Defines a Signal Peptide Processing Pathway Required by Flaviviruses. *Nature* 535, 164–168. doi:10.1038/nature18625