# Population structure in Argentina

Marina Muzzio[1,2]*, Josefina M. B. Motti[3], Paula B. Paz Sepulveda[1], Muh-ching Yee[4], Thomas Cooke[4], María R. Santos[1,2], Virginia Ramallo[5], Emma L. Alfaro[6], Jose E. Dipierri[6], Graciela Bailliet[1], Claudio M. Bravi[1,2], Carlos D. Bustamante[4], Eimear E. Kenny[7]

1 Instituto Multidisciplinario de Biología Celular (IMBICE) CCT-La Plata CONICET-CICPBA, La Plata, Buenos Aires, Argentina, **2** Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina, **3** Universidad Nacional del Centro de la Provincia de Buenos Aires, FACSO, NEIPHPA, Quequén, Buenos Aires, Argentina, **4** Stanford University, Stanford, California, United States of America, **5** CENPAT CONICET, Puerto Madryn, Chubut, Argentina, **6** INECOA (Instituto de Ecorregiones Andinas) UNJu-CONICET, Instituto de Biología de la Altura, Universidad Nacional de Jujuy, San Salvador de Jujuy, Jujuy, Argentina, **7** Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, United States

* marinamuzzio@gmail.com
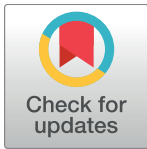
Abstract

We analyzed 391 samples from 12 Argentinian populations from the Center-West, East and North-West regions with the Illumina Human Exome Beadchip v1.0 (HumanExome-12v1-A). We did Principal Components analysis to infer patterns of populational divergence and migrations. We identified proportions and patterns of European, African and Native American ancestry and found a correlation between distance to Buenos Aires and proportion of Native American ancestry, where the highest proportion corresponds to the Northernmost populations, which is also the furthest from the Argentinian capital. Most of the European sources are from a South European origin, matching historical records, and we see two different Native American components, one that spreads all over Argentina and another specifically Andean. The highest percentages of African ancestry were in the Center West of Argentina, where the old trade routes took the slaves from Buenos Aires to Chile and Peru. Subcontinentaly, sources of this African component are represented by both West Africa and groups influenced by the Bantu expansion, the second slightly higher than the first, unlike North America and the Caribbean, where the main source is West Africa. This is reasonable, considering that a large proportion of the ships arriving at the Southern Hemisphere came from Mozambique, Loango and Angola.

Introduction

One of the most important applications of admixture research is to reduce bias of association studies, since it has long been known that the underlying genetic structure can produce a high percentage of false positive results due to differences in the genetic composition of cases and controls [1–6]. This bias occurs when the frequency of the case disease varies between populations, so that the probability of selecting affected individuals from specific subpopulations grows; thus, any allele with a higher frequency in the over represented population will show an association with the phenotype [7–9].

DNA studies allow researchers to analyze population differences as well as the ancestry of a given individual, whether following a single line of ancestry (uniparental markers) or focusing on the autosomal regions and the X chromosome, which reveal different ancestral components. There is great interest in identifying admixture to show continental ancestry in genetically divergent populations [10–14]. These methods have been applied to African-American populations [15–20](Zhu et al. 2005, Smith et al. 2004, Tian et al. 2008, Reich et al. 2005, Deo et al. 2007, Reich et al. 2007), and have been powerful in identifying ancestry in Latin American populations [21–24].

Populations in the Americas are the result of admixture between Native Americans, West Eurasians and Africans; however, we are just beginning to understand the finer details of this process and its genetic correlates [25–27]. Thanks to large consortia, new studies analyzing full genomes are being published [28]. It is of biomedical and anthropological interest to survey admixture in these populations, which would help to elucidate the unique Latin American genetic landscape.

Early evidence of people in South America is seen at sites such as Monte Verde in Chile [29], with people arriving before 12,000 years ago (ya) [30], and a rapid settlement of the continent is supported by DNA studies [31](Bodner et al. 2012), mostly explained by a coastal route [29, 31–34].

Argentina is a country of a large spectrum of climates and geographical features: from subtropical rainforests in the North East to the glaciers in the Andes and the South, and from the highest peak outside Asia (the Aconcagua) to the flats of Chaco and Pampas. Evidence that hunter-gatherer groups were well settled in the North West and Center-West regions by 10,000 ya is shown by the presence of older sites such as Inca Cueva 4 (10,620 ya), Pintoscayoc (10,200 ya), Gruta del Indio (10,3500) and Agua de la Cueva (10,950) [35], and archaeological industries such as La Fortuna (8500 ya), Intihuasi (8000 ya) and Quebrada Seca 3 (9.790 ya) [36, 37]. The first signs of domestication in the North West date to near 4000 ya, seen in the sites of Inca Cueva, Huachichocana and Puente del Diablo. By 2500 ya there was the development of villages based on herding and agriculture as well as ceramic and metallurgic technology. Early sedentary villages are spread all over the North West, and between 1400 and 1100 ya there was the Aguada style expansion (simultaneous with Tiahuanaco-Wari in Central Andes, sharing characteristics such as the hallucinogenic complex and symbolism), throughout the North West except on the Puna where there was the interaction circuit Yavi-Isla that involved the North of Chile and the South West of Bolivia. In 1476 the Tawantisuyo expanded along the Andes as far South as to the present day Mendoza province in the Center-West.

The North West of Argentina (NW) was one of the most densely populated areas of the country in pre-Columbian times [38]. Its contact with Europeans dates back to the founding of the first European city in Argentinian territory, Santiago del Estero, in 1533 [39]. The Chaco area is a vast plain of which there is little archaeological data of its first settlers except that it has been inhabited for about 4,000 or 5,000 years, before which it was a swamp [40]. Native Americans from the Argentinian Chaco successfully repelled colonization, including the pre-Hispanic Tawantisuyo, until the military expeditions sent between the years 1884 and 1911 [41], with the exploitation of its forests for exotic wood the initial economic attractant for the region [42] and, in the 1920s, oil exploitation in its West. The Center-West in colonial times was part of the Capitanía General de Chile and was an important part of the route that brought goods from the port of Buenos Aires into Chile and then Peru. Until the creation of the Virreinato del Río de La Plata in 1776, the colonial control nuclei were Santiago de Chile (for the Center-West) and Lima (for the rest of present-day Argentina). Founded in 1580 and currently the capital of Argentina, Buenos Aires was an entry port for people and goods, smuggled at first and later legally transported once it became the capital of the Virreinato. Slave

ships introduced Africans until 1812 [43] and there was massive immigration from Europe from the 1870s until the 1950s [44]. This was a large transoceanic migration considering the relative size of the source and receiving populations [45, 46]. As mentioned by Motti et al. [47], in the 1914 national census, 30% of Argentina's inhabitants had been born abroad but were distributed unevenly across the country: 50% of the people in the city of Buenos Aires, 30% in the provinces of Buenos Aires and Mendoza and only 2% in the provinces of Catamarca and La Rioja (North West region) were foreign-born.

Historical and genetic evidence demonstrate that the geographical distribution of the massive European wave of immigrants to Argentina was concentrated in the Pampas region of the Center-East, plus the North-East of Argentina. These patterns have been studied through uniparental lineages, both in the Creole population and the Native American communities [48–59] and in the differing ancestry proportion contribution from the Americas, Europe and Africa [23, 60–62], but their resolution was only of a continental level. The aim of this paper is to describe Argentina's genetic composition and population structure, taking into account its geographical and historical diversity. We are especially interested in possible regional differences in the Native American component of these populations, as well as those in both the sources and extent of foreign migration.

Material and methods

Sampling and generation of genotype data

We collected blood samples from over 3,300 participants from 17 Argentinian cities and small towns over multiple sampling efforts between 2006 and 2013. Genomic DNA was extracted from whole blood and de-identified for subsequent genotyping, sequencing and analysis. Written informed consent in Spanish was obtained from all participants under ethics approvals and permits granted by the Comité de Ética del IMBICE, and included local governmental approval from each of the 11 provinces included in the study. Donors answered a short genealogical survey to identify recent migrants from other locations or countries. Given the historical and cultural differences between regions of Argentina, these locations were grouped as shown in S1 Table. We genotyped a subset of 391 samples from 12 populations on Illumina Human Exome Beadchip v1.0 (HumanExome-12v1_A) at the Hussman Institute for Human Genomics, University of Miami. Sampling locations and summary data for the populations included in the study are detailed in S1 Table and shown on Fig 1. Using PLINK v1.07 software, we applied the following filters to the genotype data for all downstream analysis. We removed individuals with a missingness rate of > 0.1 (excluding n = 2 individuals), and SNPs with a call rate < 0.9 and missingness > 0.1. We excluded sites that were monomorphic in our sample, sites with a minor allele frequency (MAF) < 0.01, removed SNPs in high linkage disequilibrium (removing one of a pair of SNPs with r2 >0.8). Finally, we randomly removed one of a pair of individuals (excluding n = 8 individuals) demonstrating direct or distant relatedness, as determined as having pairwise pi_hat > 0.5, given that algorithms for estimating identity by descent in non-homogenous populations give a systematically inflated degree of relatedness among individuals of the same group [63]. After applying our filters, we were left with 29,145 of the 247,870 SNPs segregating in 381 individuals.

We used public data from the 1000 Genomes Project as reference populations [64], specifically the Luhya in Webuye, Kenya (LWK), Yoruba in Nigeria (YRI), Afro-Caribbean (ACB), African Americans from the US South West (ASW), Finnish (FIN), British (GBR), individuals of European Ancestry from the South West US (CEU), Tuscans (TSI), Iberians (IBS), Puerto Rican (PUR), Colombians from Medellín (CLM), Mexicans from Los Ángeles (MXL), Peruvians from Lima (PEL). Since these panels are trio data, we excluded the children, applied the
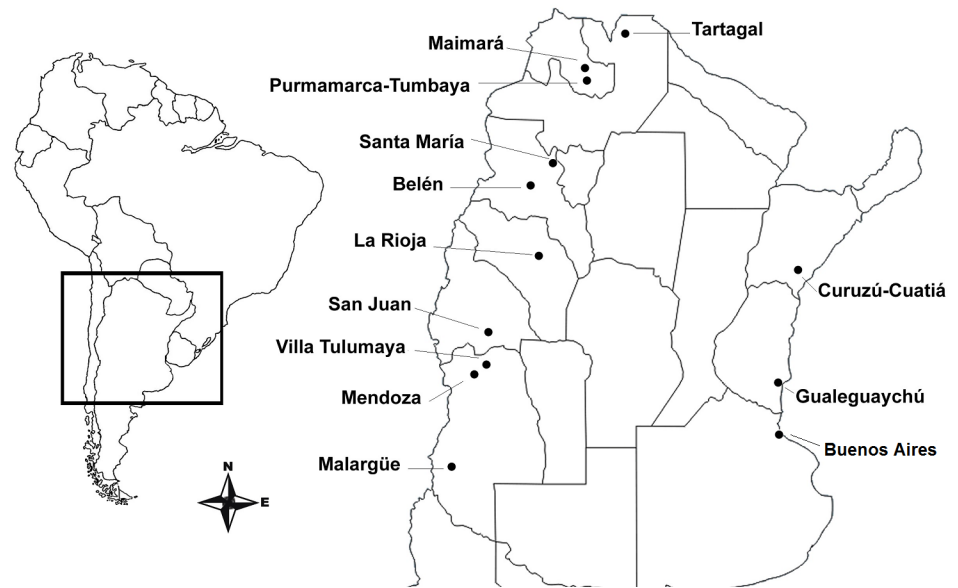
Fig 1. Distribution of the sampling locations in Argentina. Dots show the sampling locations included in this study.

https://doi.org/10.1371/journal.pone.0196325.g001

same QC filters of our dataset and looked up strand flips before merging. After merging, there were 21,847 SNPs.

Principal component and population diversity by distance analysis

The SNPRelate R package [65] was used to perform principal components analysis (PCA). Color schemes for plots were chosen from RColorBrewer. Pairwise Fst were calculated with the Weir and Cockerham formula [66] with the SNPRelate package [65]. Correlation between pairwise Fst and linear population distances was assessed with Pearson's correlation coefficient.

Structure analysis

We used block relaxation algorithm as implemented in ADMIXTURE [67] to estimate ancestry proportions for each individual given K ancestral populations, where K was incremented from 3 to 10. We used the default setting (folds = 5) to perform ADMIXTURE's cross-validation procedure for evaluating fit of different values of K (S1 Fig), where K = 9 showed the lowest error estimates (CV = 0.51).

Integration of ancestry proportions and geographic coordinates

We used Surfer® software to geographically visualize the spatial distribution of average proportions of ancestry and we estimated the correlation between distance to Buenos Aires and percentage of Native American ancestry with Pearson's correlation coefficient.

From the clustering patterns observed in the ADMIXTURE analysis, a clear pattern could be distinguished between clusters and the location of sampling of different groups.

TreeMix analysis

We used TreeMix v1.12 to infer patterns of population divergence and migrations [68]. First we ran it with all our samples, pooled by region and chose the CEU and GBR as the root position using the -root flag. We accounted for linkage disequilibrium with the -k flag using block

of 20 SNPs and performed a round of the hill-climbing to optimize migration edges with the -climb flag. We included information about known migration events using the -cor_mig flag [69]. The percentage was estimated from the Admixture results and the putative sources were selected from historical information. Those events were defined as a 0.5 flow from TSI to East, 0.4 from IBS to the Center-West, and 0.2 IBS to the North East and to the North West.

Results and discussion

On Fig 2 we represent the Principal Component Analysis (PCA) results (Fig 2). First, we used the YRI and LWK for African references, the IBS and TSI for European references and the PEL as a Latin American reference that has samples with high Native American ancestry (Fig 2A). The x axis, PC1, separates African samples from the rest while the y axis, PC2 shows Europe and the samples with highest Native American ancestry in each extreme, with the remaining admixed samples distributed along. Then, we removed the African, the TSI and all admixed individuals that had over 3% African ancestry and added the MXL and CLM panels that fit this criterion, so that we could have a better understanding of differences among the Native American component of these populations (Fig 2B). PC1 (x-axis) shows us the spectrum of admixture between full European ancestry near -0.6 while near 0.8 we have the Native American extreme. PC2 (y axis) shows that the samples with a high Native American ancestry from the North West segregate from MXL and overlap with PEL while those from the North East do not group with PEL and overlap with the MXL. Argentinians and Colombians overlapping with Mexicans is due to the low resolution of our SNP set; CLM and MXL segregate from each other in PC space when denser sets are assayed, as shown in by Gravel et al. [28].

As expected, there was a significant correlation between linear distance and Fsts (r = 0.46, p < 0.001). S2 Fig shows this correlation and S2 Table shows the pairwise Fsts and distances between populations.
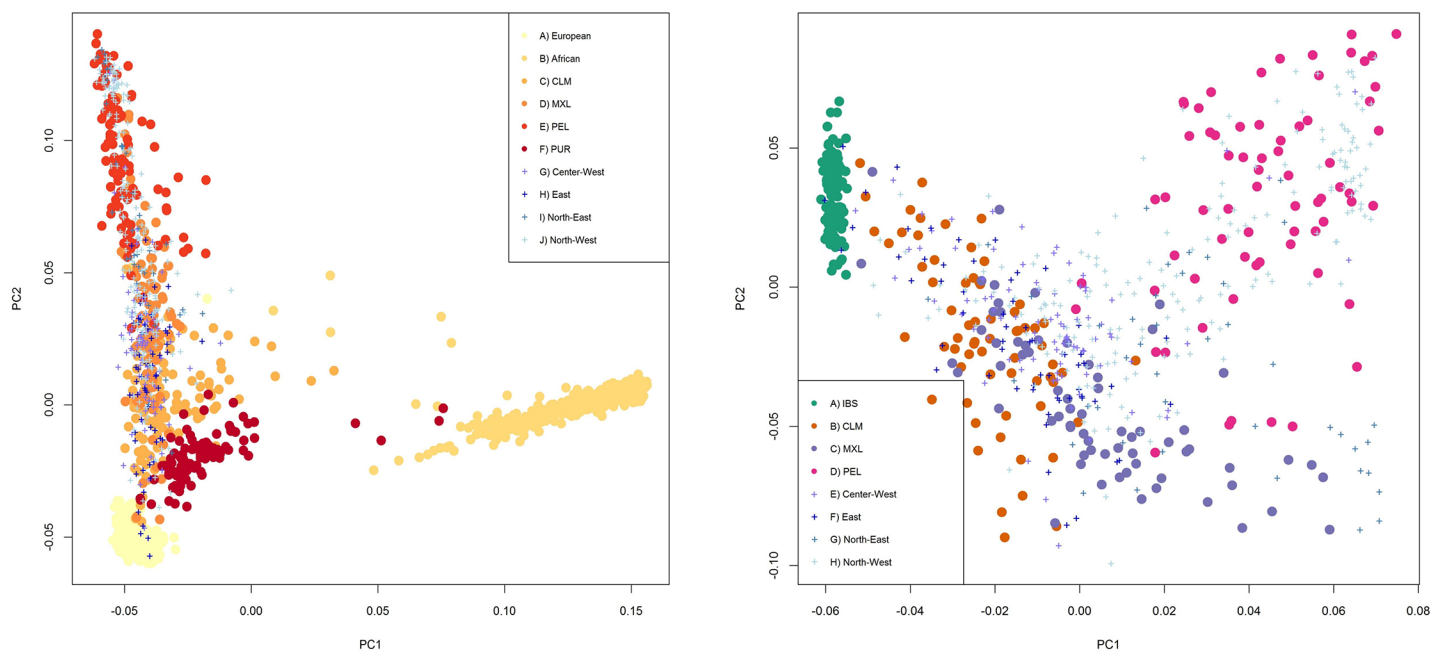

Fig 2. Principal component analysis. On the x axis is PC 1 while PC2 is the y axis. Plus symbols represent Argentinian samples and circles are for reference panels. Fig 2a (left) Argentinians with YRI and LWK for African references ("African"), IBS and TSI for European references ("European") and the PEL, MXL, PUR and CLM as a Latin American references. Fig 2b (right) samples from Argentina with IBS, MXL, CLM and PEL.
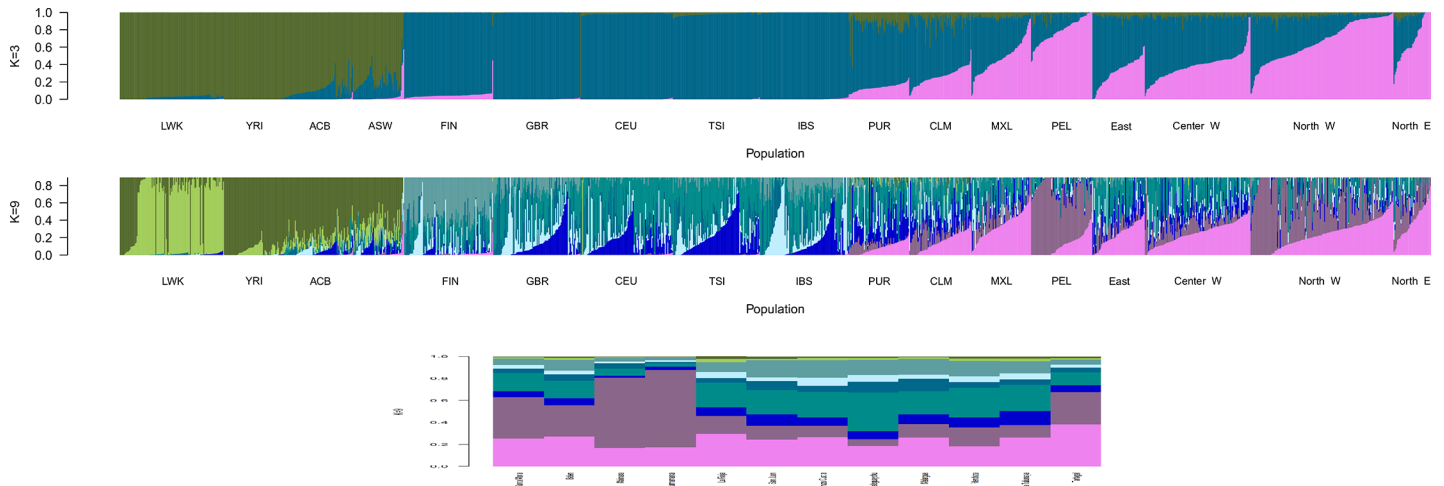
https://doi.org/10.1371/journal.pone.0196325.g002

Fig 3. Admixture results. With 3 (K = 3) (Fig 3a, top), 9 groups (K = 9) (Fig 3b, middle) and ancestry components averaged across individuals in each per population (Fig 3c, bottom). In K = 9 there is an Andean component among the Peruvians (PEL) and the Argentinians from the North West (ANW).

https://doi.org/10.1371/journal.pone.0196325.g003

Through Admixture analysis [67] we can see the admixture at both a populational and individual level (Fig 3). In Argentina, Native American ancestry is highly anti-correlated to distance to Buenos Aires (r = 0.99; p < 0.0001), where the highest percentages are the farthest from the capital. The spatial distribution of this component is shown on S3A and S3B Fig. We show the spatial distribution of each of the remaining components on S3C to S3I Fig. As mentioned in the Introduction, Buenos Aires was the entry port for the large immigration that arrived between the last decades of the 19th century and a few years after the World War II, many of whom were rural workers and merchants, who settled in the Capital and areas that were suitable for agriculture and stock breeding, mainly the Pampas and the East. This has been previously reported by surname studies [70].

The highest African percentages are on the Center West of the country, mainly Mendoza and the southernmost populations of the North West, San Juan and La Rioja. These locations were along the route that took slaves from Buenos Aires into Chile and Peru, and they had a much lower density of Native population at the time of the Conquistadors, so it was expected that they would be the ones with the highest African ancestry. This component is represented by both West African (represented by Nigerian Yorubas) and Bantu (represented by the Kenyan Luhyas) ancestries, the second slightly larger than the first. The presence of a Bantu component is explained by the slave routes that reached the La Plata basin, about half of them directly from Africa -mostly ships from Mozambique, Loango and Angola, and some from Bight of Biafra- and the other half from Brazil, where the majority came from Luanda and Benguela with a minority from Bight of Benin [43]. Also, matches to Mozambican mitDNA sequences have been found in Brazil and Santo Domingo [71–73], in concordance to our Bantu findings since these areas were affected by the Bantu expansion [74–76].

The highest Native American percentages are on the Chaco and the extreme North West. This concurs with what is known from the archaeological records -the area with the highest populational density in prehispanic times was the North West- and the fact that the Chaco was the last territory to be effectively colonized by the Argentinian government, with few Argentinean migrants to it after the oil industry established wells in the mid-1920s [77]. Hunter-gatherer Native American groups live in the Chaco and many of them even today do not speak Spanish [78].
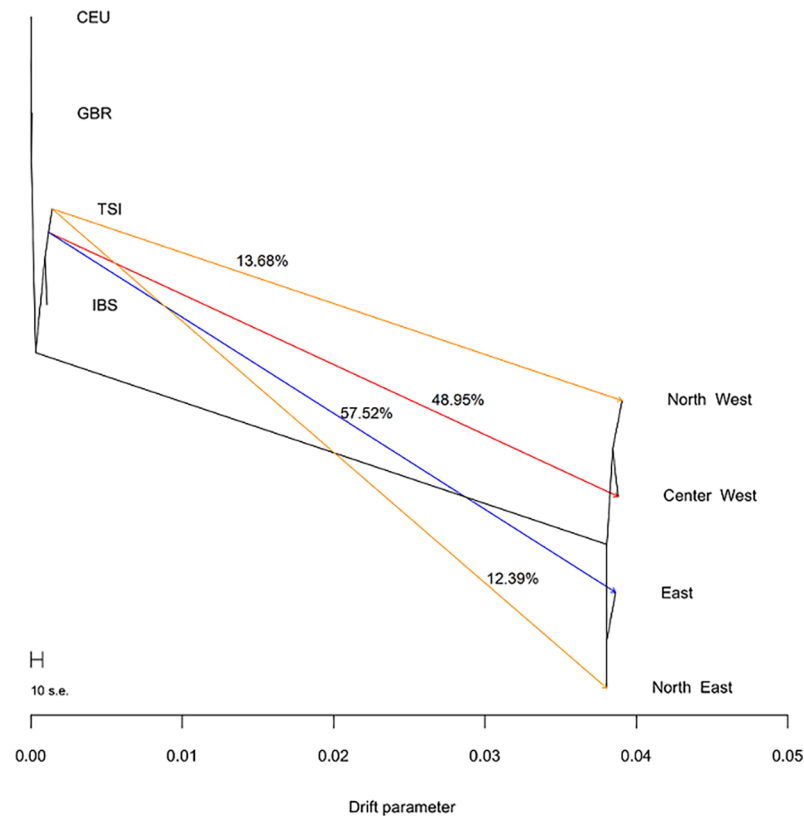
Fig 4. Treemix results. Fig 4 (left) Argentinian populations pooled by region and the GBR, CEU, TSI and IBS reference panels from the 1000 Genomes Project. The arrows reflect gene flow from the European source towards each Argentinian group with its percentage written.

https://doi.org/10.1371/journal.pone.0196325.g004

In concordance with previous studies [79] we see different European components, with mainly a North-European and Southern-European distinction. In the Argentinian data we find a mostly a Southern-European component, which is concordant with the main immigration sources.

The North West, is part of the broad Central Andes cultural area in South America. It was part of the Kollasuyo during the Tawantisuyo and took part in other pre-hispanic cultural phenomena such as Tiawanako-Wari, the archaeological complex Yavi-Isla and Aguada 1400 ya [80], it is interesting to point out that the Native American component in these samples is Andean, different from the one found in the Chaco and other areas of the country. The Inka were not able to conquer the Chaco, they were repelled by the Guarani, Wichi, Chorote, Qom and other groups that managed to keep their independence until the 20th century [78].

On Fig 4 we show Treemix [68] results for the Exome Array set. We analyzed all samples sorted by region, with some European populations of the 1000 Genomes Project (Fig 4). Matching with our Admixture results, we can see how the largest European gene flow starts from the Spanish and Italian references, being the lowest for the Northern Argentinean populations, while on the East and Center-West it is high, and also coming from slightly different origins: Italian for the East and Center-West and with a Spanish influence for the North East and North West.Residual plot is available as S4 Fig.

These findings are similar to what Avena et al. [62] found, specifically regarding regional differences in ancestry components, and Campana et al. [81] found in the Center-West region

and Cordoba (a province not sampled on this study). Our work allows us to have a finer resolution so that we can differentiate between different Native American ancestries, which has not been previously reported for Argentina with autosomal markers, and an African component that points to a Bantu origin instead of Western African, as seems to be the case for African ancestry in North America, the Caribbean, Colombia and Peru but not Brazil [82,83]. It is interesting to note that Brazil seems very similar regarding the sources of all three major components (except perhaps the Andean sub-component within the Native American component), with a clear difference between the proportions of African and Native American ancestries. In Brazil, the African component is much higher than in Argentina (50.8–14.7% for Brazil, 1.21–5.46% in Argentina), whereas the Native American component is much higher in Argentina (6.4–8% for Brazil, 24.84–87.49% for Argentina), which shows that even though both countries have similar migration histories in the late 19th and during the 20th centuries, the impact of the African slave trade and the interactions with the local Native Americans was much different.

As expected when considering the population history, ancestry in Argentina varies highly depending on which area is sampled, from high percentages of European ancestry to almost exclusively Native American. These results will enable future region-specific medical studies that can take advantage of Argentina's highly diverse populations, understanding the differences within each continental source regarding each region will allow better research designs that capture this complex matrix.

Supporting information

S1 Table. Sampled populations. List of populations included with their coordinates in decimal degrees format, distance to Buenos Aires, region assigned and N.
(XLS)

S2 Table. Paired distances and Fsts.
(XLS)

S1 Fig. CV values for K 2 to 10. The lowest CV is for K = 9.
(TIF)

S2 Fig. Scatterplot between paired Fsts (y axis) and paired distance between populations (x axis). There is a correlation between paired linear distances and paired Fsts (p < 0.001).
(TIF)

S3 Fig. Geographic distribution of ancestry components. a: Native American (whole) b: Andean c: Non-Andean, d: LWK, e: YRI, f: European 1, g: European 2, h: European 3, i: European 4 j: European 5
(TIF)

S4 Fig. Residual plot for the Treemix results. S3 Fig shows Argentinian populations pooled by region and the GBR, CEU, TSI and IBS reference panels from the 1000 Genomes Project.
(TIF)

Acknowledgments

Author Contributions

References

1. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. Am. J. Hum. Genet. 2000; 67, 170. https://doi.org/10.1086/302959 PMID: 10827107

2. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, et al. Assessing the impact of population stratification on genetic association studies. Nat Genet. 2004; 36: 388–93. https://doi.org/10.1038/ng1333 PMID: 15052270

3. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. Nat Genet. 2004; 36: 512–17. https://doi.org/10.1038/ng1337 PMID: 15052271

4. Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, et al. Demonstrating stratification in a European American population. Nat Genet. 2005; 37: 868–72. https://doi.org/10.1038/ng1607 PMID: 16041375

5. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet. 2005; 37: 1243–46. https://doi.org/10.1038/ng1653 PMID: 16228001

6. Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefansson K. An Icelandic example of the impact of population structure on association studies. Nat Genet. 2005; 37: 90–5. https://doi.org/10.1038/ng1492 PMID: 15608637

7. Landeer ES and Schork NJ. Genetic Dissection of Complex Traits. Science. 1994; 265(5181):2037–48. PMID: 8091226

8. Ewens WJ, Spielman RS. The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet. 1995; 57:455–64. PMID: 7668272

9. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet. 1999; 65:220–28. https://doi.org/10.1086/302449 PMID: 10364535

10. Zhu X, Cooper RS, Elston RC. Linkage analysis of a complex disease through use of admixed populations. Am J Hum Genet. 2004; 74:1136–53. https://doi.org/10.1086/421329 PMID: 15131754

11. McKeigue PM. Prospects for admixture mapping of complex traits. Am J Hum Genet. 2005; 76: 1–7. https://doi.org/10.1086/426949 PMID: 15540159

12. Smith MW, O'Brien SJ. Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. Nat Rev Genet. 2005; 6: 623–32. https://doi.org/10.1038/nrg1657 PMID: 16012528

13. Tang H, Coram M, Wang P, Zhu X, Rish N. Reconstruction genetic ancestry blocks in admixed individuals. Am J Hum Genet. 2006; 79: 1–12. https://doi.org/10.1086/504302 PMID: 16773560

14. Seklin MF. Admixture mapping as a tool in gene discovery. Curr Opin Genet Dev. 2007; 17:177–81. https://doi.org/10.1016/j.gde.2007.03.002 PMID: 17466511

15. Zhu X, Luke A, Cooper RS, Quertermous T, Hanis C, Mosley T, et al. Admixture mapping for hypertension loci with genome-scan markers. Nat Genet. 2005; 37: 177–81. https://doi.org/10.1038/ng1510 PMID: 15665825

16. Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, et al. A high-density admixture map for disease gene discovery in African Americans. Am J Hum Genet. 2004; 74: 1001–13. https://doi.org/10.1086/420856 PMID: 15088270

17. Deo RC, Patterson N, Tandon A, McDonald GJ, Haiman CA, Ardlie K, et al. A high-density admixture scan in 1,679 African Americans with Hypertension. PLoS Genet. 2007; 3: e196. https://doi.org/10.1371/journal.pgen.0030196 PMID: 18020707

18. Reich D, Patterson N, Ramesh V, De Jager PL, McDonald GJ, Tandon A, et al. Admixture mapping of an allele affecting interleukin 6 soluble receptor and interleukin 6 levels. Am J Hum Genet. 2007; 80: 716–26. https://doi.org/10.1086/513206 PMID: 17357077

19. Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, Selmi C, et al. Analysis and application of European genetic substructure using 300 K SNP information. PLoS Genet. 2008; 4(1): e4. https://doi.org/10.1371/journal.pgen.0040004 PMID: 18208329

20. Reich D, Patterson N, De Jager PL, McDonald GJ, Waliszewska A, Tandon A, et al. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. Nat Genet. 2005; 37: 1113–18. https://doi.org/10.1038/ng1646 PMID: 16186815

21. González BE, Borrell LN, Choudhry S, Naqvi M, Tsai HJ, Rodriguez-Santana JR, et al. Latino populations: a unique opportunity for the study of race, genetics and social environment in epidemiological research. Am J Public Health. 2005; 95: 2161–68. https://doi.org/10.2105/AJPH.2005.068668 PMID: 16257940

22. Silva-Zolezzi I, Hidalgo-Miranda A, Estrada-Gil A, Fernández-López JC, Uribe-Figueroa L, Contreras A, et al. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. PNAS. 2009; 106 (21) 8611–16. https://doi.org/10.1073/pnas.0903045106 PMID: 19433783

23. Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, et al. Geographic patterns of genome Admixture in Latin American Mestizos PLoS Genet. 2008; 4(3): e1000037. https://doi.org/10.1371/journal.pgen.1000037 PMID: 18369456

24. Bryc K, Christopher Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Autona A, et al. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. PNAS. 2010; 107 (2): 8954–61.

25. Price AL, Patterson N, Yu F, Cox DR, Waliszewska A, McDonald GJ, et al. A genomewide admixture map for Latino populations. Am J Hum Genet. 2007; 80: 1024–36. https://doi.org/10.1086/518313 PMID: 17503322

26. Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, et al. Genotype, haplotype and copy-number variation in world-wide human populations. Nature. 2008; 451: 998–1003. https://doi.org/10.1038/nature06742 PMID: 18288195

27. Li JZ, Absher DM, Tang H, Southwick AM, Castro AM, Ramachandran S, et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 2008; 319: 1100–04. https://doi.org/10.1126/science.1153717 PMID: 18292342

28. Gravel S, Zakharia F, Moreno-Estrada A, Byrnes JK, Muzzio M, Rodriguez-Flores JL, et al. Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. PLoS Genet. 2013; 9(12): e1004023. https://doi.org/10.1371/journal.pgen.1004023 PMID: 24385924

29. Dillehay TD, Ramrez C, Pino M, Collins MB, Rossen J, Pino-Navarro JD. Monte Verde: seaweed, food, medicine, and the peopling of South America. Science. 2008; 320: 784–786. https://doi.org/10.1126/science.1156533 PMID: 18467586

30. Steele J, Politis G. AMS 14D dating of early human occupation of southern South America. Journal of Archaeological Science. 2009; 36(2):419–29.

31. Bodner M, Perego UA, Huber G, Fendt L, Rock AW, Zimmermann B, et al. Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. Genome Res. 2012; 22: 811–820. https://doi.org/10.1101/gr.131722.111 PMID: 22333566

32. Dillehay TD. Probing deeper into first American studies. Proc Natl Acad Sci USA. 2009 106: 971–978. https://doi.org/10.1073/pnas.0808424106 PMID: 19164556

33. Hurst CT. A Folsom site in a mountain valley of Colorado. American Antiquity. 1943; 8: 250–253. https://doi.org/10.2307/275905

34. Meltzer DJ. First peoples in a New World. Colonizing Ice Age America. Univ of California Press. 2009.

35. Rivero D, Berberian E. El poblamiento de la región central del territorio argentino durante la transicón Pleistoceno-Holoceno (12,000–9,000AP). Revista española de Antropología. 2008; 38:17–37.

36. Gonzalez AR. La estratigrafia de la gruta de Intihuasi (Prov. de San Luis, R.A.) y sus relaciones con otros sitios preceramicos de Sudamerica. Revista del Instituto de Antropologia. 1960; 1:5–302.

37. Rivero D, Roldan F. Initial peopling of the Cordoba mountains, Argentina: first evidence from El Alto 3. Current Research in the Pleistocene. 2005; 22: 33–5.

38. Martínez Sarasola C. Nuestros paisanos los indios. Vida, historia y destino de las comunidades indígenas de la Argentina. EMECE ed. 1992.

39. Binayan Carmona N. Historia Genealogica Argentina. Emece (de). Buenos Aires, Argentina; 1999.

40. Demarchi DA and Garcia Ministro A. Genetic Structure of native populations from the Gran Chaco Region, South America. Int J Hum Genet. 2008; 8(1–2): 131–141

41. Wright PG. Colonizacion del espacio, la palabra y el cuerpo en el Chaco argentino. Horiz. Antropol.2003; 9 (19). https://doi.org/10.1590/S0104-71832003000100006

42. Zarrilli A. Transformacion ecologica y precariedad economica en una economia marginal. El Gran Chaco argentino, 1890–1950. Theomai 1. 2000.

43. Borucki A. The slave trade to the rio de La Plata, 1777–1812: Trans-imperial networks and Atlantic warfare. Colonial Latin American Review. 2011; 20 (1): 81–107.

44. Avena S. Mezcla génica en una muestra poblacional de la ciudad de Buenos Aires. Medicina (B. Aires). 2006; 66 (2):113–118.

45. Lattes A. Migraciones hacia América Latina y Caribe desde principios del siglo XX. Cuadernos del CENEP. 1985; 35.

46. Esteban F. Dinámica migratoria argentina: inmigración y exilios. Am Latina Hoy. 2003; 15–34.

47. Motti JMB, Muzzio M, Ramallo V, Rodenak Kladniew B, Alfaro EL, Dipierri JE, et al. Origen y distribucion espacial de linajes maternos nativos en el noroeste y centro oeste argentinos. Rev Arg Antrop Biol. 2013; 15(1):03–14.

48. Castilla EE and Adams JP. Migration and genetic structure in an isolated population in Argentina. Aicuña. In Adams JP (ed.) Proceedings of Convergent Questions in Genetics and Demography. Oxford. Oxford University Press 45–62. 1990

49. Bianchi NO, Catanesi CI, Bailliet G, Martínez-Marignac VL, Bravi CM, Vidal-Rioja LB, et al. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am J Hum Genet. 1998; 63: 1862–71. https://doi.org/10.1086/302141 PMID: 9837838

50. Dipierri JE, Alfaro EL, Martínez-Marignac VL, Bailliet G, Bravi CM, Cejas S, et al. Paternal directional mating in two Amerindian subpopulations from the northwest of Argentina. Hum Biol. 1998; 70 (6):1001–10. PMID: 9825592

51. Bailliet G, Castilla EE, Adams JP, Orioli IM, Martínez-Marignac VL, Richard SM, et al. Correlation between molecular and conventional genealogies in Aicuña: A rural population from Northwestern Argentina. Hum Hered. 2001; 51: 150–59. https://doi.org/10.1159/000053336 PMID: 11173966

52. Demarchi DA and Mitchell RJ. Genetic structure and gene flow in Gran Chaco populations of Argentina: evidence from Y-chromosome markers. Hum Biol. 2004; 76(3): 413–29. PMID: 15481676

53. Fejerman L, Carnese FR, Goicochea AS, Avena SA, Dejean CB, Ward RH. African ancestry of the population of Buenois Aires. Am J Phys Anthropol. 2005; 128(1): 164–70. https://doi.org/10.1002/ajpa.20083 PMID: 15714513

54. Salas A, Jaime JC, Álvarez-Iglesias V, Carracedo A. Gender bias in the multiethnic genetic composition of central Argentina J Hum Gen. 2008; 53: 662–74.

55. Bailliet G, Ramallo V, Muzzio M, García A, Santos MR, Alfaro EL, et al. Restricted geographic distribution for Y-Q* haplogroup in South America. Am J Phys Ant. 2009; 140(3): 578–82.

56. Ramallo V, Mucci JM, García A, Muzzio M, Motti JMB, Santos MR, et al. Comparison of Y-chromosome haplogroup frequencies in eight Provinces of Argentina. For Sci Int: Gen Supp Series. 2009; 431–432.

57. Ramallo V, Santos MR, Muzzio M, Motti JMB, Salceda S, Bailliet G. Linajes masculinos y su diversidad en comunidades Wichí de Formosa. Revista del Museo de Antropología (Córdoba, Argentina). 2009; 2: 67–74.

58. Motti JMB, Rodenak B, Muzzio M, Ramallo V, Santos MR, Castro C, et al. The genetic composition of Argentina prior to the massive immigration era: insights from matrilineages of extant criollos in central-western Argentina. For Sci Int: Gen Supp Series. 2009; 342–43.

59. Muzzio M, Ramallo V, Motti JMB, Santos MR, López Camelo JS, Bailliet G. Software Haplogroup Predictions: a word of caution. Int J Leg Med. 2010; https://doi.org/10.1007/s00414-009-0404-1 PMID: 20082090

60. Martínez-Marignac VL, Bertoni B, Parra EJ, Bianchi NO. Characterization of Admixture in an Urban Sample from Buenos Aires, Argentina, using uniparentally and biparentally inherited genetic markers. Hum Biol. 2004; 4:543–57.

61. Corach D, Lao O, Bobillo C, Van Der Gaag K, Zuniga S, Vermeulen M, et al. Inferring Continental Ancestry of Argentineans from Autosomal, Y-Chromosomal and Mitochondrial DNA. Ann Hum Gen. 2010; 74: 65–76.

62. Avena S, Via M, Ziv E, Pérez-Stable EJ, Gignoux CR, Dejean C, et al. Heterogeneity in Genetic Admixture across Different Regions of Argentina. PLoS ONE. 2012; 7(4): e34695. https://doi.org/10.1371/journal.pone.0034695 PMID: 22506044

63. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies.Bioinformatics. 2010; 26 (22): 2867–2873. https://doi.org/10.1093/bioinformatics/btq559 PMID: 20926424

64. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature. 2010; 467: 1061–1073. https://doi.org/10.1038/nature09534 PMID: 20981092

65. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. Bioinformatics. 2012; https://doi.org/10.1093/bioinformatics/bts606 PMID: 23060615

66. Weir BS and Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 38(6):1358–1370 https://doi.org/10.1111/j.1558-5646.1984.tb05657.x PMID: 28563791

67. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research. 2009; 19:1655–1664. https://doi.org/10.1101/gr.094052.109 PMID: 19648217

68. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet. 2012; 8(11): e1002967. https://doi.org/10.1371/journal.pgen.1002967 PMID: 23166502

69. Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, Guldemann T, et al. The genetic prehistory of southern Africa. Nat Commun. 2012; 3:1143. https://doi.org/10.1038/ncomms2140 PMID: 23072811

70. Dipierri JE, Alfaro EL, Scapoli C, Mamolini E, Rodriguez-Larralde A, Barrai I. Surnames in Argentina: A Population Study Through Isonymy. Am J Phys Anthropol. 2005; https://doi.org/10.1002/ajpa.20027 PMID: 15761858

71. Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bntu expansions and the slave trade. Ann. Hum. Genet. 2001; 65: 439–458. https://doi.org/doi:10.1017/S0003480001008855 PMID: 11806853

72. Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, et al. The African diaspora: mithocondrial DNA and the Atlantic slave trade. Am. J. Hum. Genet. 2004; 74.

73. Hunemeier T, Carvalho C, Marrero AR, Salzano FM, Junho Pena SD, Bortolini MC. Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. Am J Phys Ant. 2007; 133. https://doi.org/10.1002/ajpa.20604 PMID: 17427922

74. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, et al. The Genetic Structure and History of Africans and African Americans. Science. 2009; 324(5930): 1035–1044. https://doi.org/10.1126/science.1172257 PMID: 19407144

75. Patin E, Lopez M, Grollemund R, Verdu P, Harmant C, Quach H, et al. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. 2017; 356, 543–546 https://doi.org/10.1126/science.aal1988 PMID: 28473590

76. Currie TE, Meade A, Guillon M, Mace R. Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. Proc R Soc B. 2013; 280: 20130695. https://doi.org/10.1098/rspb.2013.0695 PMID: 23658203

77. Benclowicz JD. Contributions to the History of Northern Salta. Formation and development of the towns of Tartagal and General Mosconi in the first half of the Twentieth century. Andes. 2009; 22(1)

78. Miller ES. Peoples of the Gran Chaco. Praeger. 2001.

79. The 1000 Genomes Project. A global reference for human genetic variation. Nature. 2015; 526, 68–74 https://doi.org/10.1038/nature15393 PMID: 26432245

80. Mandrini R. La Argentina aborigen. De los primeros pobladores a 1910. Buenos Aires. Siglo XXI. 2008.

81. Campana H, Pawluk MS, López Camelo JS and Grupo de Estudio del ECLAMC. Prevalencia al nacimiento de 27 anomalías congénias seleccionadas, en 7 regiones geográficas de la Argentina. Arch.

argent. Pediatr. 2010. 108(5): 409–417. https://doi.org/10.1590/S0325-00752010000500006 PMID: 21132229

82. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. AmJHumGenet. 2017; 100 (5) 635–649 http://dx.doi.org/10.1016/j.ajhg.2017.03.004

83. Kehdy FSG, Gouveia MH, Machado M, Magalhaes WC, Horimoto AR, Horta BL, et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. PNAS. 2015; 112 (28): 8696–8701. https://doi.org/10.1073/pnas.1504447112 PMID: 26124090