1 **Fine-scale genomic analyses of admixed individuals reveal**

2 **unrecognized genetic ancestry components in Argentina**

3 **Native American, African and European genetic ancestries in Argentina**

4 Pierre Luisi*[1], Angelina García[1,2,3], Juan Manuel Berros[4], Josefina Motti[5], Darío
5 Demarchi[1,2,3], Emma Alfaro[6,7], Eliana Aquilano[8], Carina Argüelles[9,10], Sergio
6 Avena[11,12], Graciela Bailliet[8], Julieta Beltramo[8,13], Claudio M. Bravi[8], Mariela Cuello[8],
7 Cristina Dejean[11,12], José Edgardo Dipierri[7], Laura S. Jurado Medina[8], José Luis
8 Lanata[14], Marina Muzzio[8], María Laura Parolin[15], Maia Pauro[1,2,3], Paula B. Paz
9 Sepúlveda[8], Daniela Rodríguez Golpe[8], María Rita Santos[8], Marisol Schwab[8], Natalia
10 Silvero[8], Jeremias Zubrzycki[16], Virginia Ramallo[17], Hernán Dopazo*[18,19]
11
12
13 **1.** Universidad Nacional de Córdoba, Facultad de Filosofía y Humanidades,
14 Departamento de Antropología, Argentina
15
16 **2.** Instituto de Antropología de Córdoba (IDACOR), Consejo Nacional de
17 Investigaciones Científicas y Técnicas – Universidad Nacional de Córdoba, Córdoba,
18 Argentina
19
20 **3.** Universidad Nacional de Córdoba, Facultad de Filosofía y Humanidades, Museo de
21 Antropología, Córdoba, Argentina.
22
23 **4.** Consejo Nacional de Investigaciones Científicas y Técnicas – Laboratorio de
24 Análisis de Datos, Biocódices S.A, Buenos Aires, Argentina
25
26 **5.** Núcleo de Estudios Interdisciplinarios de Poblaciones Humanas de Patagonia
27 Austral (NEIPHA), Consejo Nacional de Investigaciones Científicas y Técnicas –
28 Universidad Nacional del Centro de la Provincia de Buenos Aires, Quequén,
29 Argentina.
30
31 **6.** Instituto de Ecorregiones Andinas (INECOA), Consejo Nacional de Investigaciones
32 Científicas y Técnicas – Universidad Nacional de Jujuy, Argentina
33
34 **7.** Instituto de Biología de la Altura, Universidad Nacional de Jujuy, Argentina
35
36 **8.** Instituto Multidisciplinario de Biología Celular (IMBICE), Consejo Nacional de
37 Investigaciones Científicas y Técnicas – Comisión de Investigaciones Científicas –
38 Universidad Nacional de La Plata, Argentina
39
40 **9.** Grupo de Investigación en Genética Aplicada (GIGA), Facultad de Ciencias Exactas,
41 Químicas y Naturales, Departamento de Genética, Instituto de Biología Subtropical

(IBS) - Nodo Posadas, Universidad Nacional de Misiones (UNaM) – Consejo Nacional de Investigaciones Científicas y Técnicas, Posadas, Argentina.

10. Cátedra de Biología Molecular. Carrera de Medicina. Facultad de Ciencias de la Salud. Universidad Católica de las Misiones (UCAMI), Posadas, Argentina.

11. Instituto de Ciencias Antropológicas (ICA), Facultad de Filosofía y Letras, Universidad de Buenos Aires, Argentina

12. Centro de Estudios Biomédicos, Biotecnológicos, Ambientales y Diagnóstico (CEBBAD), Universidad Maimónides, Buenos Aires, Argentina,

13. Laboratorio de Análisis Comparativo de ADN, Corte Suprema de Justicia de la Provincia de Buenos Aires, La Plata, Argentina

14. Instituto de Investigaciones en Diversidad Cultural y Procesos de Cambio (IIDyPCa), Consejo Nacional de Investigaciones Científicas y Técnicas – Universidad Nacional de Río Negro, San Carlos de Bariloche, Argentina

15. Instituto de Diversidad y Evolución Austral (IDEAus), Consejo Nacional de Investigaciones Científicas y Técnicas – Centro Nacional Patagónico, Puerto Madryn, Argentina

16. Laboratorio de Genómica. Biocódices S.A., Buenos Aires, Argentina

17. Instituto Patagónico de Ciencias Sociales y Humanas (IPCSH) – Consejo Nacional de Investigaciones Científicas y Técnicas - Centro Nacional Patagónico, Puerto Madryn, Argentina

18. Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina

19. Consejo Nacional de Investigaciones Científicas y Técnicas – Biocodices S.A., Buenos Aires, Argentina

* Corresponding Authors: pierrespc@gmail.com, dopazoh@gmail.com

79 ## 1 Abstract

80   We are at the dawn of the efforts to describe and understand the origins of genetic

81   diversity in Argentina from high-throughput data. This knowledge is a primary step

82   in the intent of deciphering the specific genetic bases of diseases and drug response

83   in the country. Similarly to other populations across the Americas, genetic ancestry

84   in Argentinean populations traces back into African, European and Native American

85   ancestors, reflecting a complex demographic history with multiple migration and

86   admixture events in pre- and post-colonial times. However, little is known about the

87   sub-continental origins of these three main ancestries. We present new high-

88   throughput genotyping data for 87 admixed individuals across Argentina. This data

89   was combined to previously published data for admixed individuals in the region

90   and then compared to different reference panels specifically built to run population

91   structure analyses at a sub-continental level. Concerning the European and African

92   ancestries, we confirmed previous results about their main origins, and we provide

93   new insights into the presence of other origins that reflect historical records. As for

94   the Native American ancestry, leveraging genotype data for archaeological samples in the

95   region in order to gain temporal depth in our analyses, we could identify four Native

96   American components segregating in modern Argentinean populations. Three of them

97   are also found in modern South American populations and are specifically

98   represented in Central Chile/Patagonia, Lowlands and Central Andes geographic

99   areas. The fourth one may be specific to the Central Western region of Argentina.

100   Identifying such component has not been straightforward since it is not well

101   represented in any genomic data from the literature. Altogether, we provide useful

102   insights into the multiple population groups from different continents that have

103   contributed to present-days genetic diversity in Argentina. We encourage the

104   generation of massive genotype data locally to further describe the genetic structure

105   in Argentina.

2 Author Summary

106

107   The human genetic diversity in Argentina reflects demographic mechanisms during

108   which the European colonists invaded a territory where Native American

109   populations were settled. During colonial period, the slave trade also prompted

110   many African people to move to Argentina. Little is known about the origins of the

111   Native American and African components in Argentinean populations nowadays.

112   Genotyping data for 87 admixed individuals throughout Argentina was generated

113   and data from the literature was re-analyzed to shed light on this question. We

114   confirmed that most of the European genetic ancestry comes from the South,

115   although several individuals are related to Northern Europeans. We found that

116   African origins in Argentina trace back from different regions. As for the Native

117   American ancestry, we identified that it can be divided into four main components

118   that correspond to Central Chile/Patagonia, Lowlands, Central Andes and Central

119   Western region of Argentina. In order to understand the specificity of the genetic

120   diversity in Argentina, we should not rely on knowledge generated in other

121   populations. Instead, more effort is required to generate specific massive genomic

122   knowledge at the local level.

3   Introduction

123

124   Natural selection, genetic drift, migrations and admixture shape the genetic

125   composition of human populations. In order to decipher the demographic history of

126   populations and get insights into their specific genetic bases of diseases and drug

127   response, it is primordial to first characterize their geographical genetic structure

128   [1]. Much effort has been made at broad scales in South America. In this article, we

129   present analyses at much finer scales for Argentina.

130   The first systematic investigation of human genetic variation in the country focused

131   on a limited number markers either uniparental (mtDNA, Y-STRs, Y-SNP; [2–11] or

132   autosomal (Short Tandem Repeats, Ancestry Informative Markers, Alu sequences,

133   indels, and blood groups [12–17]). Studies based on autosomal markers identified

134   an important inter-individual heterogeneity for the African, Native American and

135   European genetic ancestry proportions [12–17]. Accordingly, most of the studies

136   based on uniparental markers showed large differences in the genetic composition

137   of Argentinean populations, accounting for the different demographic histories

138   within the country [2,3,18–21,4–10,17]. A recently published book compiles most of

139   these studies and others to gives a general view about the admixture in Argentinean

140   populations [22]. Although the idea of a 'white' country with most of inhabitant's

141   descendants from European immigrants has now been rejected by these studies, the

142   Argentine founding myth of a white and European nation remains perceptible today

143   [23].

144   There is a wealth of information about the European ancestors, both from familial

145   stories and from historical records. On the contrary, little is known about the

146   African and Native American populations that contributed to the admixture events

147   between the continental components. During the great wave of European

148   immigration between the mid-19th to mid-20th centuries, people from Southern

149   Europe (mainly Italy and Spain) settled in Argentina, along with migrants from

150   Western Europe (mainly French), Central Europe (mainly German) and

151   Northeastern Europe (mainly Poles, Ukrainians, Russians) [24]. Historical records

152   attest that immigration waves from Southern Europe were predominant from this

153   continent [24].

154   The presence of human groups of African origins in Argentina traces back to the

155   slave trade that took place during colonial era, most of the time illegally, and is

156   therefore poorly documented in historical records [25]. It is even very difficult to

157   estimate how many African slaves were brought into current Argentinean territory.

158   Moreover, most of the European people at that time were unaware of African

159   societies with which they traded. Thus, the slaves were ethnically identified

160   according to the last ship they boarded before travelling to the Americas [25,26].

161   However, this is a poor proxy to infer their actual origins since captains often had to

162   dock in different harbors along the African coasts in order to complete the human

163   load. In addition, many slaves first arrived in Brazil and then were taken to the Río

164   del Plata harbor to be finally distributed across the current Argentinean territory

165   [25]. The latest Brazilian harbor is often the only source of origin information

166   available. In addition, extrapolate the historical records from the entrance point to

167   the whole Argentinean territory is a hazardous task since the slaves origin and

168   destination have been changing during colonial era according to varying economical

169   interests and political situations [26,27]. Even then, historical records about the

170   arrival flows of African slaves to Río de la Plata show that Luanda, in current Angola

171   territory in Central Western region of Africa, was the main departure harbor,

172   followed by harbors located in the Gulf of Guinean and on the coast of the present-

173   day Senegalese, Gambian and Sierra Leone territories. The coasts of Indian Ocean, in

174   the current territory of Mozambique, were also an important departure location

175   [27–29].

176   As for the Native American component, it is difficult to study its origin focusing on

177   present-day communities since their organization has changed drastically from the

178   16th century, and the arrival of the first conquerors, until the beginning of the 20th

179   century [23]. During the period of conquest and colonization, wars, diseases and

180   forced labor decimated the Native populations. The system of colonial exploitation

181   also often meant the relocation of individuals, families, and communities [30]. Then,

182   the expansion of the nation-state by the late 19th Century can be described as a

183   territorial annexation process and subjugation of the indigenous peoples

184   perpetrated by the Argentinean national armed forces between 1876 and 1917 [31].

185   For example, the miscalled Conquest of the Desert in the 1870s to integrate

186   Patagonia into the Argentinean state caused that the indigenous populations were

187   subjected to practices of extermination, deportations and social fragmentation [32].

188   Moreover, the arbitrariness of several academic, historical, economic and political

189   factors have greatly influenced the classification of the Native American groups

190   settled in the Argentinean territory

191   Recently, two articles presented high-throughput genotyping data for hundreds of

192   thousands of autosomal markers through the genome for modern Argentine

193   individuals, and provided the first inferences of which populations contributed to

194   the admixture events among the three main continental components [33,34].

195   Homburger *et al.* studied 437 admixed individuals from urban centers in 5 South

196   American countries: Argentina, Chile, Colombia, Ecuador and Peru. On the other

197   hand, Muzzio *et al.* obtained genotype data from nearly 400 individuals from 12

198   populations in Northern and Central Argentina using Illumina exome-based array

199   technology [34]. In both studies, it was found that the European ancestry in

200   Argentina is mainly explained by Italian and Iberian ancestry components according

201   to the historical records. Homburger *et al.* observed a strong gradient in Native

202   American ancestry of South American Latinos between Andean and other South

203   American Native American populations. This gradient is associated with the country

204   of origin and geography of local indigenous populations [33]. In addition, Muzzio *et*

205   *al.* found that African ancestry is explained by a Western and a Bantu-influenced

206   components and described a Native American component not represented in the

207   1000 Genomes project (1KGP) [34].

208   Despite the progress in studying the genetic diversity of the Argentinean

209   populations, the challenge of achieving a fine-scale knowledge has inevitably been

210   scarce. In this context, studies of a wider region, namely South America, can provide

211   important insights into the Native American specific genetic diversity in Argentina

212 and its origins. Ancient DNA analyses of the peopling of the Americas suggest that

213 Native American populations of South America descend from two streams from

214 Northern America: one mainly present in the Andes and another one present

215 elsewhere [35]. These streams have replaced the first people settled in the region,

216 whose ancestry was related to the Clovis culture [35]. Genetic continuity for the

217 Native American component appears to have prevailed in the region ever since

218 these replacements [35–37]. However, little is known about the legacy in modern

219 populations of the different ancestries associated to the several waves of population

220 arrivals in South America. Modern Native American populations in the Southern

221 Cone of South America seems to be divided between a Lowlands component and

222 another component that includes Andean, Central Chile and Patagonian populations

223 [38]. A recent genomic study of ancient and modern populations from Central Chile

224 and Western Patagonia further identified that they are differentiated from the

225 Andean and Lowlands populations, supporting a differentiation process throughout

226 South America [37]. In another recent genomic study of modern samples, in which

227 the Southern Cone is only represented by Gran Chaco populations, it has been

228 described a likely common lineage for non-Andean South Americans populations

229 which do probably not share with the Andeans any common ancestor from Central

230 America [39].This supports the hypothesis of many migrations back to Central

231 America different from the Andeans ancestors [39]. It has also been described that

232 the genetic interactions between the peopling routes on both sides of the Andes

233 were limited [39,40].

234   The Native American specific genetic diversity in South America has been studied

235   focusing almost exclusively on indigenous communities. In many regions of

236   Argentina, particularly were the Spanish conquerors settled (in the Northwest,

237   Central and Central Western regions), indigenous communities did not maintain

238   their traditional lifestyle. They have rather been incorporated as a workforce in the

239   colonial society. The sedentary life of these communities at the moment of contact

240   partially contributed to this phenomena [23]. The early assimilation to the colonial

241   society made gradually indigenous identity invisible since the communities were

242   then referred to as peasants [41,42]. It implied a cultural annihilation and several

243   Argentinean regions were considered "Indian free" in the mid-20st century [43].

244   However, the cultural incorporation did not necessarily imply a biological

245   extinction. We argue that focusing only on indigenous communities by

246   "implementing accurate sampling strategies and of selecting representative

247   populations based on historical/linguistic and anthropological information", as

248   advocated by Gnecchi-Ruscone *et al.* [39] for example, is not sufficient to decipher

249   the pre-Columbian history of Native Americans. Although studies based on such

250   sampling strategies [37–39] provide decisive information to understand the

251   evolutionary history of Native American ancestry [22], it implies to disregard many

252   regions of the continent. This is particularly the case of Argentinean territory where

253   only the extreme North and South, separated by ~ 3,700 kilometers, have been

254   studied under this scope. Therefore, alternative strategies must be considered to fill

255   this gap in the intent of describing the Native American ancestry.

256   Here, we carry out a fine-scale population genomic study to get insight into the

257   genetic structure and the complex origins of the African, European and Native

258   American ancestries in Argentina populations. Using Affymetrix, Axiom-Lat1 array

259   technology, we genotyped 87 admixed individuals nationwide for more than

260   800,000 SNPs covering coding and non-coding regions of the genome. Additionally,

261   a dataset with close to 500 individuals from modern and ancient populations

262   throughout South America was constructed. We also pulled together genotype data

263   for over 1,600 individuals from some ~50 Sub-Saharan African populations from

264   the literature. The new data generated in the present study, compared to those data

265   sets, allowed to broaden the knowledge of the sub-continental origins of the three

266   main genetic ancestries of Argentinean populations.

267   ## 4   Results and Discussion

268   High-throughput genotyping data was generated for 87 admixed individuals

269   throughout Argentina (**Fig 1** and **S1 Table**). For clarity in the visualization of the

270   results, we used Administrative Regions to classify the admixed individuals

271   analyzed as shown in **Fig 1**. This classification has not been used for any statistical

272   analyses. The generated data was compared to different data sets to understand the

273   origins of the genetic diversity in the country. These data sets are called **DS<n>** and

274   are described in **S2 Table**.
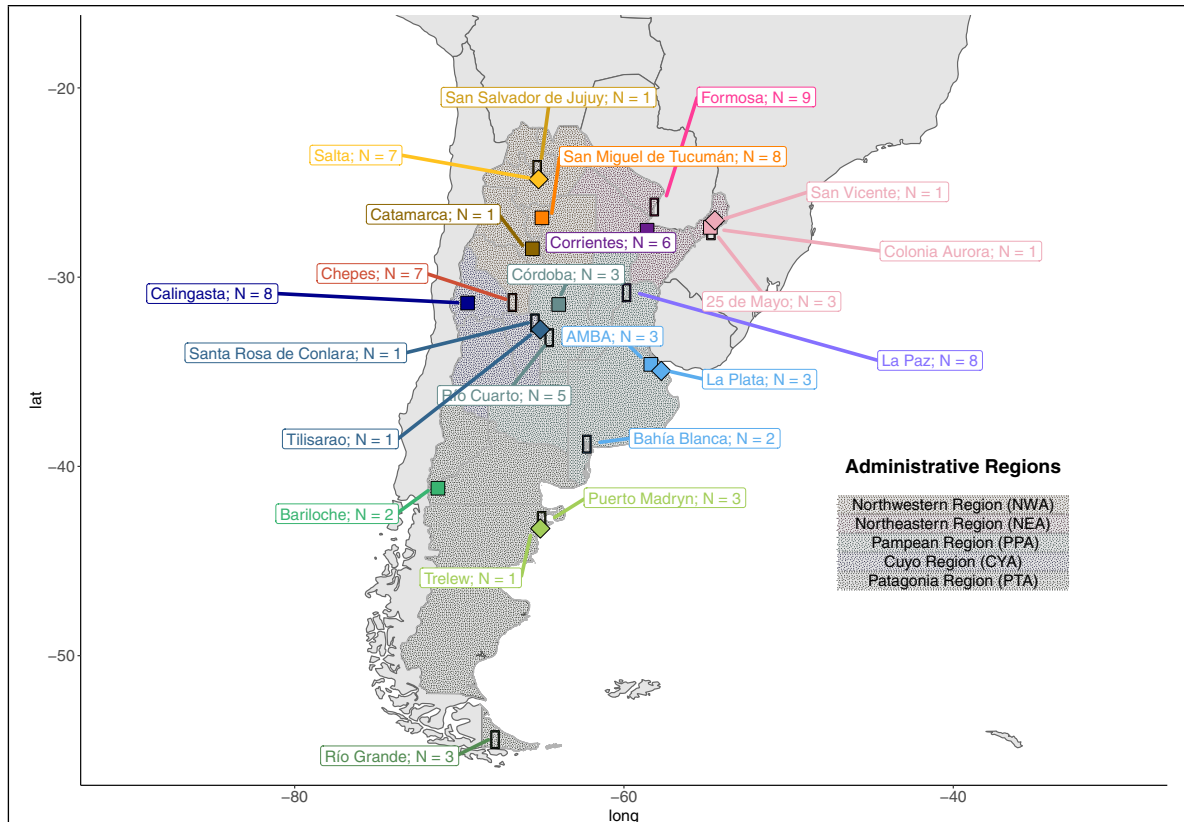
Fig 1. Sample locations from the present study.
The samples are divided according to administrative regions for visualization only.
N: The number of samples that passed genotyping Quality Controls.

4.1 Ancestry in a worldwide context

In order to characterize the genetic diversity observed in Argentina within a

worldwide genetic context, Principal Component Analysis (PCA [44]; **S1 Fig**) was

applied to the dataset **DS1,** which contains genotype data for the samples from

Argentina generated in the present study, as well as admixed and Native American

individuals from South America characterized in previous studies [33,37,38] and

individuals from Europe and Africa from the 1000 Genomes Project (1KGP; [45]).

The PCA shows, as expected, that Argentinean individuals have different

proportions of European, Native American and, much less represented, African

12

288   ancestry (**S1A Fig**). This pattern, which has already been documented in other

289   admixed populations in South America [33,45], including in Argentina [14–17,34]

290   supports the heterogeneous genetic origins throughout the country. The third and

291   the fifth principal components discriminate three different main Native American

292   ancestries that are also represented in the Argentinean samples. PC4 discriminates

293   between Southern and Northern European individuals (**S1B Fig**) while PC6 divides

294   Luhya population, a Bantu-speaking population in Kenya, to Western African

295   populations (**S1C Fig**). The Argentinean samples do not exhibit any striking gradient

296   along these two PCs. The ancestry composition of the Argentinean individuals will

297   be discussed below.

298   We ran unsupervised clustering models with Admixture software [46] on **DS1**. We

299   used from 3 to 12 putative ancestral populations (K ranging from 3 to 12).

300   Comparing the cross-validation scores obtained for each run, we estimated that the

301   genotype data analyzed was best explained with a model with $K=8$ ancestral

302   populations (**S2A Fig**). At $K=3$ (**S2B Fig**), the algorithm allows estimating the

303   proportions of European, Native American and African ancestry. For $K=4$ to $K=7$,

304   (**S2C-F Figs**), the model detects sub-continental ancestries, i.e. it divides the main

305   continental ancestry into different components, that are finally all observed in the

306   model with $K=8$ (**S2G Fig**)*. That is, the European ancestry is divided into Northern

307   and Southern components, while African ancestry is composed of Westernmost

308   African, Gulf of Guinea and Bantu-influenced components. Moreover, three

309   components are observed for Native American ancestry: one represented in Central

310   Chile/Patagonia populations (hereafter referred to as **CCP**), another present in

311   Central Andes populations (hereafter referred to as **CAN**) and a third one in

312   populations mainly from Lowlands (here after referred to as **LWL**).

313   **S3 Table** lists, for individuals included in **DS1**, the proportion of continental

314   ancestry estimated from the model with K=3, its proportion of sub-continental

315   component estimated from the model with K=8, as well as the proportion of

316   continental ancestry estimated from K=8 (obtained from the sum of the sub-

317   continental component proportions). We confirmed that the estimates of the

318   continental ancestry proportion obtained from a model with K=3 and K=8 are highly

319   consistent (**S3 Fig**).

320   From the Admixture results for K=8 (**S2G Fig**), we observed that Peruvian and

321   Chilean admixed samples [33] exhibit very low northern European ancestry

322   proportions. The Native American ancestry for Peruvian admixed samples is mainly

323   represented by CAN, although the Lowlands component is present in smaller

324   proportions. Moreover, the Native American ancestry for the Chilean samples is

325   mainly characterized by CCP. As for the Argentinean samples from both the present

326   study and from [33], their European ancestry is divided by Southern and Northern

327   components, the former being the most abundant. Arguably surprisingly, all three

328   components of Native American ancestry are present in most Argentinean samples

329   (**S2G Fig**). The low proportions of African ancestry in Argentinean samples makes

330   difficult to interpret its sub-continental origins from analyses within a global

331   context.

14

332   The Native American ancestry proportions are higher in our samples than in the

333   Argentinean sample studied by [33] (**S4C Fig**; Wilcoxon test P = 2.57x10^{-8}). This

334   could demonstrate that the selection of individuals with previously characterized

335   uniparental Native American lineages can improve the overall representation of

336   Native American ancestry in autosomal genomic regions, which would refute a

337   previous study in Argentina [12]. However, this hypothesis can not be formally

338   proven since our sample is almost exclusively composed of individuals with Native

339   American uniparental lineages, and the genotyping data from [33] does not allow

340   the inference of the uniparental lineages. Alternatively, this significant difference

341   could also underscore that the exclusively urban samples from [33] do not allow the

342   study of the total extent of Native American ancestry in the country. Moreover, the

343   proportions of African ancestry are slightly higher in our samples than in the

344   Argentinean sample studied by [33] (**S4A Fig** Wilcoxon test; P = 2.40x10^{-2}).

345   Altogether, our sample is suitable for further studying the origins of Native

346   American and African ancestry in Argentina.

347   **4.2   Sub-continental ancestry components in Argentina**

348   **4.2.1   Local Ancestry Estimates**

349   To explore the origins of the three main continental ancestries at a finer

350   geographical level, we first estimated local ancestry patterns in phased **DS2** and

351   **DS3** (**DS2p** and **DS3p**) separately. Across phased chromosomes, we assigned

352   whether a genomic region was of Native American, European or African ancestry

353   using RFMix v2 [47]. A sanity check confirmed that the ancestry proportions across

15

354   the genome estimated by local and global ancestry approaches are concordant (**S5**

355   **Fig**). It seems that local ancestry, as compared to global ancestry, tends to

356   overestimate the proportion of the genome with European ancestry while

357   underestimating the proportion of African and Native American ancestry. The

358   present article main conclusions concern the origins of African and Native American

359   ancestries, and we argue that our approach of assigning regions to one of those two

360   ancestries is conservative.

361   Next, we masked the data in order to keep, for each individual, only the regions with

362   ditypes of a given ancestry, i.e. regions that exhibit the ancestry of interest on both

363   chromosomes (**S6 Fig**; for details see Material and Methods).

364   To decipher the sub-continental origin(s) of the European, African and Native

365   American ancestries segregating in Argentina, we applied principal component and

366   unsupervised clustering analyses on the masked data compiled with reference data

367   sets describing the genetic diversity within each continent (**S2 Table**).

4.2.2   European ancestry components in Argentina

369   To explore the European origins of the Argentinean, we used **DS4**, a combination of

370   the masked genotype data for admixed individuals with a set of individuals carefully

371   selected to be representative of the genetic diversity in their sampling area [48,49].

372   We estimated that the genotype data analyzed was best explained with a model with

373   K=2 ancestral populations for the unsupervised clustering model (**S7A Fig**).

374   Although it is rather difficult to assign a clear label to each of these two ancestral

375   populations, it seems that the algorithm discriminates between Northern and

16

376   Southern Europe components (**S7B Fig**). For *K*=3, we observed that Iberians

377   individuals exhibited a higher proportion of a new component highly represented in

378   Southern American samples (**S7C Fig**). This pattern most likely reflects the legacy of

379   first Iberian migrations in South America during colonial period [12].

380   From the PCA applied on the same dataset (**DS4)**, we observed, as previously

381   described [33,34], that most of the European ancestry of admixed Argentinean

382   individuals clusters with Iberians and Italians (**S8 Fig**). However, some individuals

383   trace their European ancestry back to Central or Northern Europe. Both PCA and

384   Admixture showed that the individuals with higher proportion of Central / Northern

385   Europe ancestry are from Misiones province (Northeastern region) (**S7** and **S8**

386   **Figs**), consistent with the historical record of settlement of Polish, German, Danish

387   and Swedish colonies promoted by governmental or private enterprises in the

388   province [50]. This demonstrates that to identify the genetic legacy from secondary

389   migration streams and thus better reflect the evolutionary history of populations in

390   Argentina, it is primordial to increase the number of studied individuals in the

391   country.

392   ### 4.2.3   African ancestry components in Argentina

393   To investigate the sub-continental components that explains African ancestry in

394   Argentina, we used **DS5** which combines masked genotype data from admixed

395   individuals with a published data set of African individuals [45,51–53] (**Fig 2A**). The

396   African reference populations used here can be divided into five main groups:

17

397   Bantu-influenced, Hunter-Gatherers, Western Africa, San and Eastern Africa

398   populations [51].
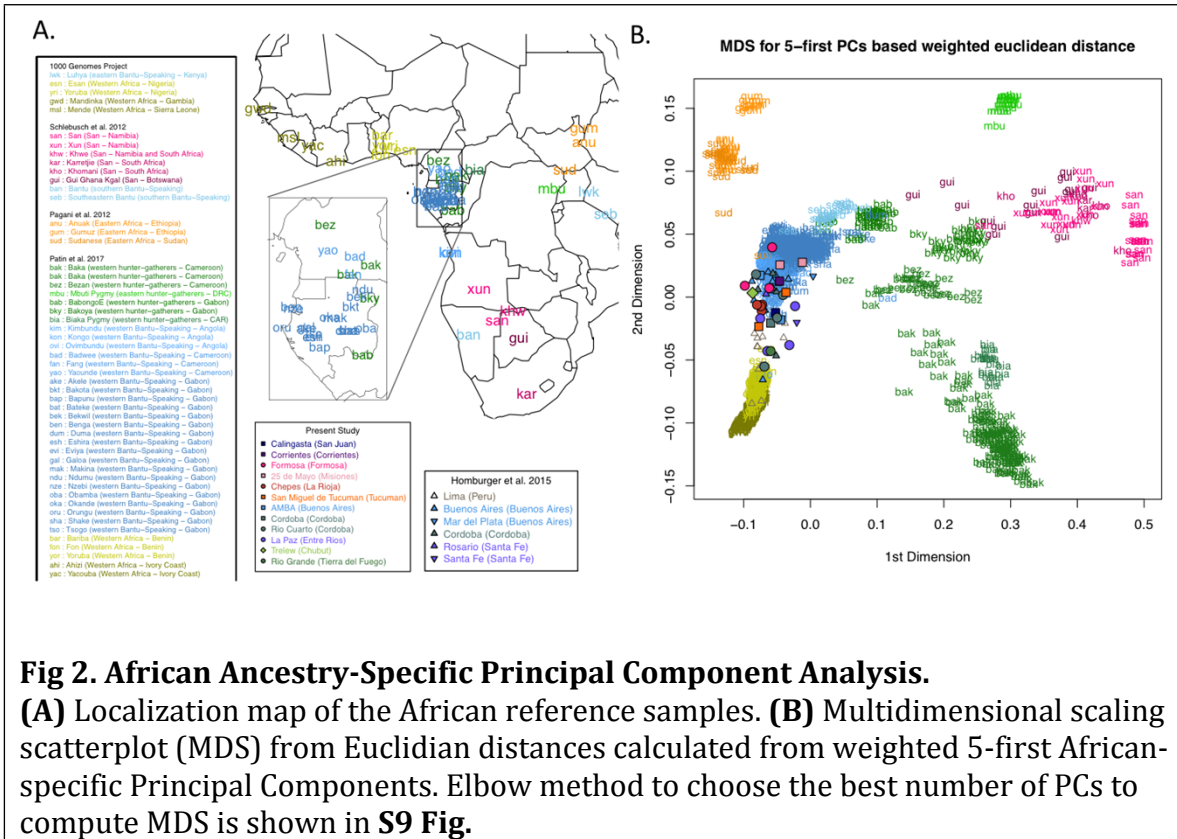
399

Fig 2. African Ancestry-Specific Principal Component Analysis.
(A) Localization map of the African reference samples. **(B)** Multidimensional scaling scatterplot (MDS) from Euclidian distances calculated from weighted 5-first African-specific Principal Components. Elbow method to choose the best number of PCs to compute MDS is shown in **S9 Fig.**

405   The PCA applied to this data allowed the distinction among those groups (**Fig 2B**).

406   The individuals from Lima (Peru) tend to group with the Western Africa

407   populations, consistent with historic records that account to a Western African

408   origin of the slaves deported to Peru through the Caribbean harbors [54]. Most of

409   the Argentinean individuals are within the Bantu-influenced population cluster

410   while others individuals are clustering with Western Africa populations. Several

411   other Argentinean individuals do not group within any African cluster and rather

412   are found between the Western Africa and Bantu-influenced clusters (**Fig 2B**).

413   These patterns were confirmed by applying Admixture algorithm. The Cross-

18

414   Validation procedure points to a best fit of the data with K=5 (**S10A Fig**). With five

415   ancestral populations, African individuals cluster in a similar fashion as observed in

416   the PCA. The Bantu-influenced and Western Africa components are the most

417   represented in Argentinean individuals. Moreover, this analysis also showed that

418   some individuals exhibit smaller proportions of Eastern Africa component,

419   particularly in Northern Argentina, as well as Western Hunter-Gatherer ancestry

420   (**Fig 3**).

Fig 3. African Ancestry-Specific Admixture Analysis.
Admixture plots for K=5
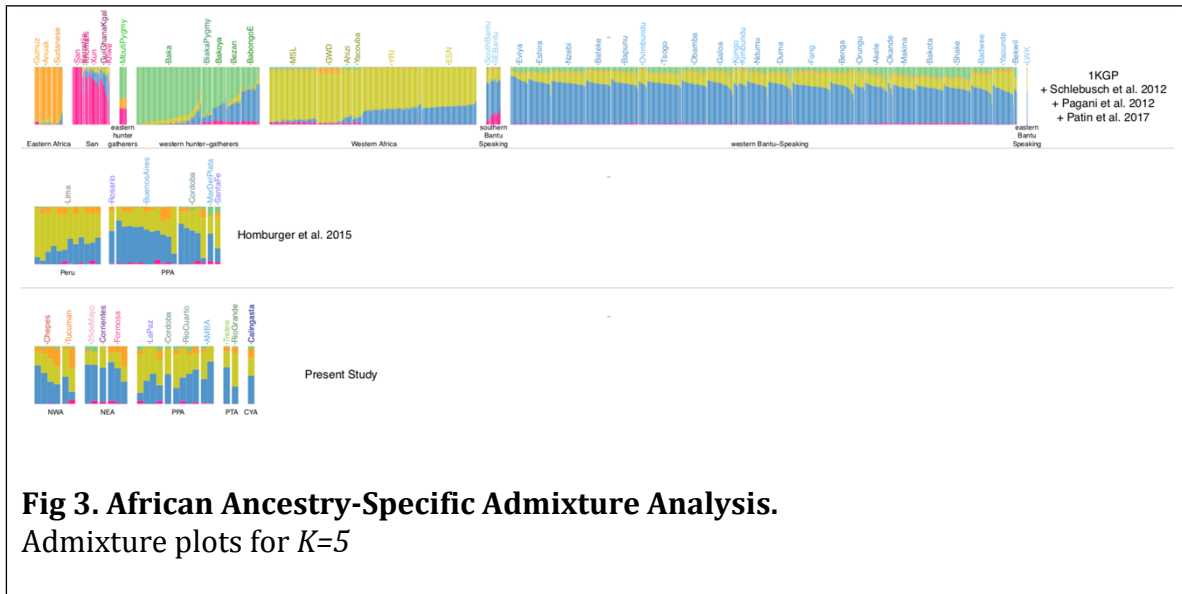
424   The legacy of Western Africa on the African genetic diversity in the Americas has

425   been preeminent [55–57], along with an impact of Bantu-influenced populations

426   from Central Western Africa, particularly in Brazil South and the Caribbean [55,56].

427   These two African ancestries have also been previously documented in Argentina

428   from studies of autosomal [34] and maternal markers [58,59]. Maternal lineages

429   specific to populations from Central Western Africa – particularly from Angola - are

430   the most common African lineages in Argentina according to studies in the Central

431   region [58] and in four urban centers (Puerto Madryn, Rosario, Resistencia and

19

432   Salta) [59]. These results are concordant with the predominance of the Bantu-

433   influenced origin identified in the present study from autosomal markers. In

434   addition, the presence of Southeastern Africa maternal lineages in Argentina [58,59]

435   is consistent with African ancestry of this origin identified here. Additional analyses

436   with larger number of individuals and higher SNP density for autosomal are

437   required to follow-up these results.

438   ### 4.2.4 Native American ancestry components in Argentina

439   We compared masked genotype data for admixed individuals with masked genotype

440   data of Native American reference populations from South America [37,38] and a set

441   of ancient DNA data for samples in the region [35–37,60,61] (**DS6**; **Fig 4A**). The PCA

442   using modern individuals, and projecting the ancient samples, confirmed the three

443   main South Native American clusters described by de la Fuente *et al.* (2018) (**Fig

444   4B**). Namely, we can distinguish between the three clusters already mentioned in

445   the analyses at the global level: CAN, CCP and LWL. The admixed individuals from

446   Lima are mainly located within CAN, while those from Santiago de Chile are found

447   next to CCP. As for the Argentinean admixed individuals, some are located within

448   one of the three clusters while many others are located in-between.
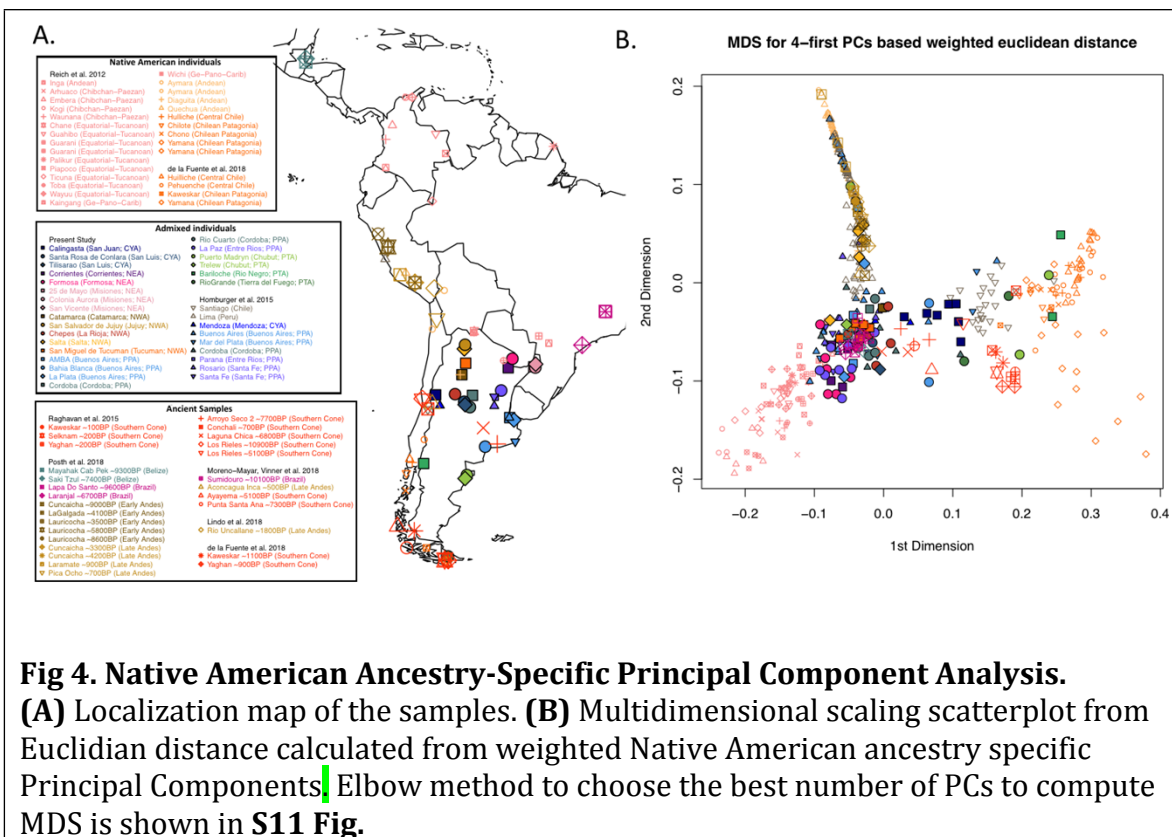
449

Fig 4. Native American Ancestry-Specific Principal Component Analysis.
(A) Localization map of the samples. **(B)** Multidimensional scaling scatterplot from
Euclidian distance calculated from weighted Native American ancestry specific
Principal Components Elbow method to choose the best number of PCs to compute
MDS is shown in **S11 Fig.**

Similar results were also observed when applying Admixture algorithm. The Cross-

Validation procedure points to a best fit of the data with *K*=3 (**S12A Fig**). The

proportion estimates for the three main Native American ancestries observed for

the South America Native American individuals and for the ancient samples are

consistent with CAN, CCP and LWL labels that we attributed for the interpretation of

the PCA (**Fig 5**). Concerning the admixed individuals, CCP ancestry is the most

represented in Santiago de Chile, while CAN ancestry prevails in Lima, Peru (**Fig 5**).

In Argentina, the three Native American ancestries are observed in almost all the

admixed individuals studied. In Northeastern Argentina (NEA), LWL ancestry is the

most frequent. In Northwestern Argentina (NWA), CAN ancestry is not clearly

dominating since LWL is also observed in important proportions. In the South, CCP

is observed in greater proportions. However, two individuals sampled in Trelew and

467 Puerto Madryn, next to the Atlantic coast of Patagonia, exhibit greater proportions

468 of LWL and CAN, respectively (**Fig 5**). The reduced genealogical information for the

469 individual from Trelew does not inform about his Native American ancestors. On the

470 other hand genealogical information of the individual from Puerto Madryn attests to

471 a recent migration to this locality since he was born in Jujuy province (NWA), as well

472 as her both parents and paternal grandparents while maternal grandparents were

473 born in Bolivia. This particular case is a perfect example of recent migrations to the

474 Patagonia region [15], and more generally within or to Argentina [62].

475

476 **Fig 5. Native American Ancestry-Specific Admixture Analysis.**
477 Admixture plots for *K=3*
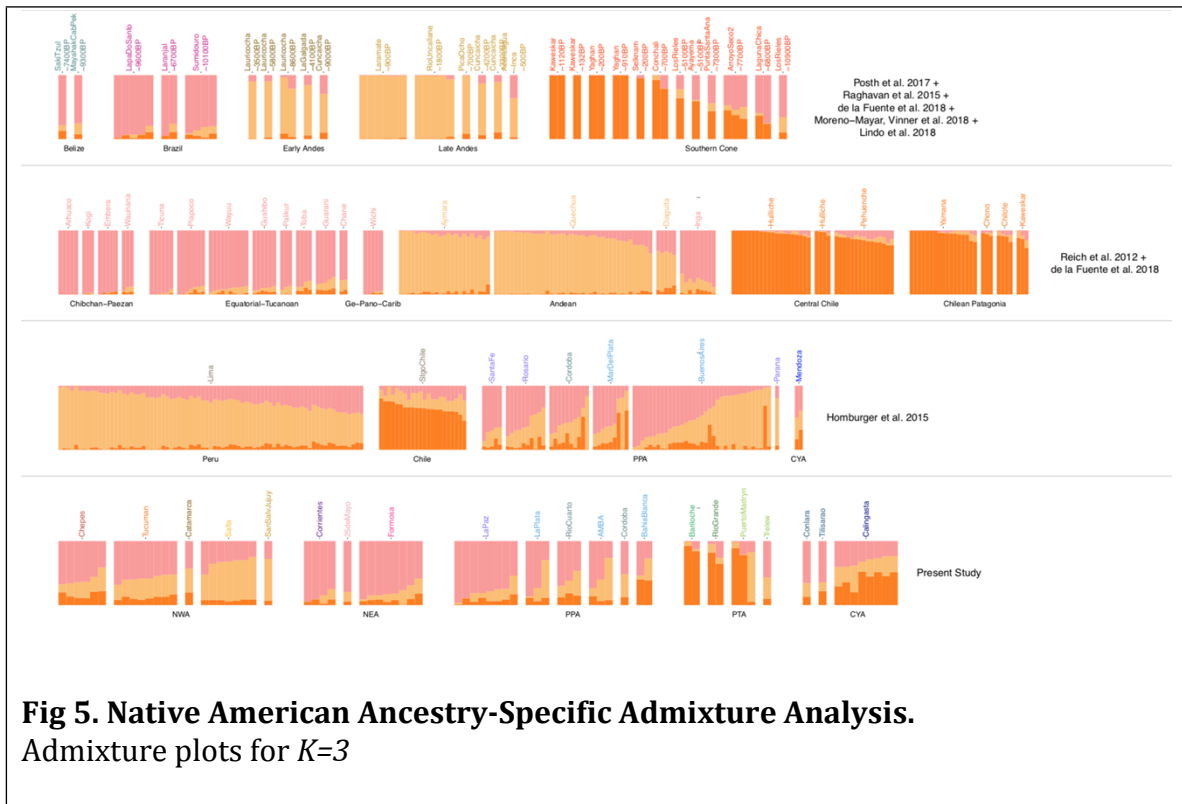
478 Although migration events tend to reduce the genetic distance among these

479 components, we still identified correlations between geographical coordinates and

480 the proportions of each Native American ancestry component. Indeed, the

481 proportion of CAN and LWL is higher further North (**S13A-B Fig**) while the

22

482   proportion of CCP is higher further South (**S13C Fig**). In addition, the proportion of

483   CAN is higher further West (**S13D Fig**) while the proportion of LWL is higher

484   further East (**S13E Fig**). This association between Native American ancestry

485   proportions and geography is expected under a model of ancient divergence and

486   genetic drift.

487   Note that a model with *K=3* at the South American scale with masked data is

488   consistent with the results from Admixture at a global scale with unmasked data for

489   which the best model contains 3 putative ancestral populations specific to South

490   American (**S2 Fig**). Moreover, the proportions for each of the three Native American

491   components estimated from masked data highly correlate with estimates from

492   unmasked data (**S14 Fig**).

493   Many individuals from the Cuyo and Pampean regions of Argentina (San Juan and

494   Córdoba provinces as well as Southern of Buenos Aires province) exhibit a complex

495   Native American ancestry in both the Principal Component (**Fig 4**) and Admixture

496   analyses (**Fig 5**). This pattern of intermediate position in the PCA graph and with

497   mid proportions of different ancestries in Admixture analyses can be interpreted as

498   the result of a mixture between the ancestries depicted. Alternatively, it can reflect

499   genetic diversity underrepresented by these ancestries because of relative limited

500   shared history with them. Increasing the number of ancestral populations to 4 in the

501   Admixture analyses, we observed that CCP component is actually structured and

502   that modern individuals from Argentinean Patagonia exhibit mainly a Central Chile

503   ancestry instead of a Patagonian ancestry (*K*=4; **S12C Fig**). With an additional

23

504 ancestral population in the model, we identified a component specific to the Middle

505 Holocene samples from the Southern Cone and Brazil (*K*=5; **S12D Fig**).

506 In order to analyze further the evolutionary history of the Native American ancestry,

507 we assigned the 452 modern South American individuals (the 53 ancient samples

508 included in **DS6** were not analyzed here) to a Native American ancestry cluster

509 using an objective quantitative approach based on *K*-means clustering (**S15 Fig**; see

510 Material and Methods for details). Thus, we assigned 163, 161 and 70 individuals to

511 the clusters representing CAN, LWL and CCP, respectively. 32 individuals were

512 assigned to a fourth cluster. The remaining 26 individuals were removed for further

513 analyses because their group assignation was not consistent across the three

514 clustering approaches. **S4 Table** describes the cluster assignation for the studied

515 individuals. We acknowledge that these groups are culturally, ethnically and

516 linguistically heterogeneous. However, we argue that analyzing such groups built

517 from genetic similarities may provide interesting insights at a broad scale into

518 evolutionary mechanisms that shaped the Native American ancestry in South

519 America.

520   **4.2.5 Relationship among the four identified Native American components**

521 We ran f_3*(Target; S1, S2*) to test whether *Target*, *S1* and *S2* are related in a form of a

522 tree (positive f_3 expected) or whether *Target* is the result of admixture between *S1*

523 and *S2* (negative f_3 expected). Significant f_3 positive scores were obtained for all the

524 possible comparisons among the four Native American clusters identified (**Fig 6A**;

525 **S5A Table**). In addition to confirming the differentiation among the three

526   components labeled geographically, this analysis showed that the fourth group most

527   likely represents a Native American component on its own never described in any

528   previous study of autosomal genetic markers. Actually, the distribution of 1 –*f3(YRI;*

529   *Ind1, Ind2)* between pairs of individuals from different groups (**S16 Fig**) does not

530   suggest a mixture among the 3 geographically labeled components to explain why

531   individuals from the fourth group are located at intermediate positions in the PCA

532   and exhibit mid Admixture proportion estimates (as shown in **Fig 4** and **Fig 5**).

533   We estimated $f_4(YRI, Target; S1, S2)$ to test whether a component (*Target*) shares

534   more affinity with any of the other two components (*S1* and *S2*). We observed that

535   (i) CAN has no particular genetic affinity with any component relative to the others;

536   (ii) LWL is closer to CAN as compared to CCP and the fourth component; and (iii) the

537   fourth component and CCP exhibit higher genetic affinity between them than with

538   CAN or LWL (**Fig 6B**; **S5B Table**). However, a neighbor-joining analysis [63] from

539   distances of the form $1/f_3(YRI; X, Y)$ suggests that CAN is more closely related to CCP

540   and the fourth component than to LWL (**Fig 6C**; **S5C Table**).

25

541
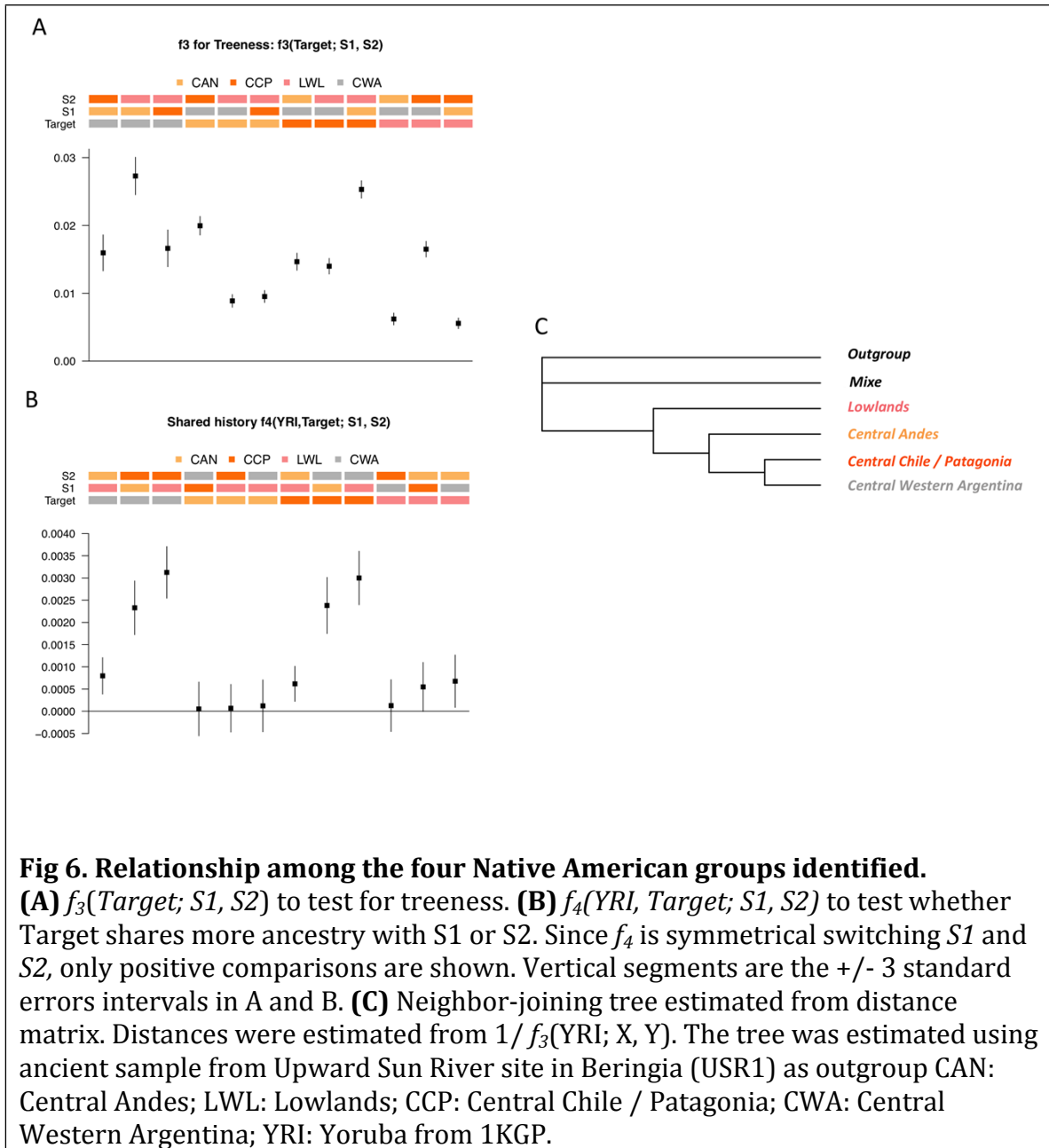
Fig 6. Relationship among the four Native American groups identified.
(A) f_3(*Target; S1, S2*) to test for treeness. **(B)** f_4*(YRI, Target; S1, S2)* to test whether Target shares more ancestry with S1 or S2. Since f_4 is symmetrical switching *S1* and *S2,* only positive comparisons are shown. Vertical segments are the +/- 3 standard errors intervals in A and B. **(C)** Neighbor-joining tree estimated from distance matrix. Distances were estimated from 1/ f_3(YRI; X, Y). The tree was estimated using ancient sample from Upward Sun River site in Beringia (USR1) as outgroup CAN: Central Andes; LWL: Lowlands; CCP: Central Chile / Patagonia; CWA: Central Western Argentina; YRI: Yoruba from 1KGP.

551   Seven out of eight individuals from Calingasta, a locality in the San Juan province

552   and located in the Northwest Monte and Thistle of the Prepuna ecoregion, were

553   assigned to the fourth cluster (the only remaining individual from this locality is not

554   consistently assigned to a cluster and has been removed). Moreover, out of 18

555   individuals from Santiago de Chile assigned to any given cluster, 16 have also been

556   assigned to the fourth component. The genealogical record for the studied

26

557 Calingasta individuals attests to a local origin of their direct ancestors up to two

558 generations ago and their region-specific maternal lineages (**S1 Table**; [64]).

559 Moreover, all the Huilliche and Pehuenche individuals from Central Chile [37] have

560 been consistently assigned to CCP. Altogether, we argue that this fourth component

561 may represent a Native American ancestry that diverged from CCP and established

562 on the Eastern side of the Andes in Cuyo region. It may have extended its influence

563 to the Pampas and Argentinean Patagonia region as suggested by the assignation to

564 this cluster of individuals from the provinces of Mendoza, Córdoba Capital, Buenos

565 Aires and Tierra del Fuego (**S15 Fig**; **S4 Table**).

566 The existence of a specific differentiated component in the Central Western and

567 Central regions of Argentina has been previously suggested from mitochondrial

568 analyses accounting to the genetic relationship between these two regions [11] and

569 the presence of specific clades underrepresented elsewhere [11,58,65]. These

570 studies support the hypothesis of a common origin and/or important gene flow

571 [66], and the authors referred to a meta-population with great temporal depth and

572 differentiated from other regions in Argentina. Moreover, the ethnographic

573 description of the populations that were settled at the moment of contact with the

574 Spanish colonies accounts to a potential relationship between Huarpes in the Cuyo

575 region and Comechingones in present-day Córdoba province ([67], cited by [68]).

576 Thus, we suggest labeling as Central Western Argentina (**CWA**) this fourth Native

577 American component never described before from autosomal markers.

578 Moreover, different, perhaps complementary, historical facts may explain the

579 representation of this component in Santiago de Chile and its absence in Huilliche

27

580   and Pehuenche populations. First, in pre-Colombian times, Central Andean

581   civilizations may never have reached Southern latitudes than the Biobío River.

582   Indeed, the troops of the Inca Empire, the Central Andean civilization that reached

583   the highest latitudes, apparently never crossed that river [69,70], leaving Huilliche

584   and Pehuenche territories, among others, outreach the expansion of any Central

585   Andean civilization [69]. Second, during colonial times, the Spanish colonies could

586   not settle South of the Biobío River neither, and they organized massive deportation

587   of Huarpes individuals from Cuyo region to palliate the lack of indigenous workforce

588   [71]. For example, in Santiago de Chile, in 1614, 37% of the indigenous people that

589   lived in the suburbs were Huarpes according to the chronicler Vázquez de Espinosa

590   [72].

591   The present study, in which we analyzed individuals that do not belong to any

592   indigenous community, made possible the identification of a Native American

593   component never described previously from autosomal markers. This demonstrates

594   the importance of broadening studies beyond the indigenous communities in order

595   to reduce the underrepresentation of many regions in Argentina, and thus, to reach

596   a better characterization of the Native American specific genetic diversity.

597   **4.2.6   Genetic affinity of the four Native American ancestry components with**

598   **ancient populations**

599   Pseudo-haploid genotype data were used for ancient samples published in the

600   literature [35–37,60,61,73,74], and their genetic affinity to the four Native American

601   ancestry components identified in Argentina was evaluated. The ancient samples

28

602   were classified according to their spatiotemporal distribution following Posth *et al.*

603   2018 [35].

604   Graphical summaries of pairwise distances are available as the neighbor-joining

605   analysis [63] from 1/f_3*(YRI; X, Y)* distances (**S17A Fig**) and a Multidimensional

606   scaling analyses from 1-f_3*(YRI; X, Y)* distances (**S17B Fig**).

607   When comparing the genetic affinity of a given component with the different ancient

608   groups using either the *f3-outgroup* or the *f4* statistics (**S18 Fig; S5D,E Tables**), we

609   identified that CAN tends to exhibit greater genetic affinity with ancient Andean

610   populations than with other ancient groups (**S18A,E Figs**). Strikingly, the genetic

611   affinity of this component with both Late Andes and Early Andes ancient groups

612   would point to a genetic continuity across the whole temporal transect that the

613   archaeological samples provide. This would suppose that the replacement of an

614   early population arrival by a later stream of gene flow in Central Andes identified

615   previously [35] does not fully explain the current-day genetic diversity for CAN. For

616   CCP, we observed higher genetic affinity with ancient Southern Cone groups

617   (**S18C,G Figs**). Overall CWA appears to be closer to ancient Southern Cone groups,

618   particularly with Conchalí ~ 700BP, although its genetic affinity with some Andean

619   groups (La Galgada ~4100 BP, Pica Ocho ~700 BP and Aconcagua ~500 BP) is

620   relatively important (**S18D,H Figs**). As for LWL, we observed intermediate genetic

621   affinity with ancient samples from both the Andes and the Southern Cone (**S18B,F**

622   **Figs**). Indeed, the three ancient samples with the greatest genetic affinity with

623   Lowlands components are from the Andes (La Galgada ~4100 BP, Aconcagua ~500

624   BP, and Pica Ocho ~700 BP), followed by three ancient groups in the Southern Cone

29

625   (Conchalí ~700 BP, Yaghan ~ 200 BP, and Selk'nam ~200 BP). The fact that there is

626   no ancient sample group exhibiting outstanding genetic affinity with LWL points to

627   the underrepresentation of this component in ancient samples. Indeed, the temporal

628   period covered by the Brazilian ancient samples is much older than for other

629   regions, and the geographical range restricted to Brazil does not cover the entire

630   area where LWL currently prevails. Moreover, LWL is a heterogeneous group that

631   includes individuals from the Gran Chaco in Argentina, the Amazonas and Northern

632   Andes. Even so, Admixture analyses demonstrated that LWL shares more ancestry

633   with ancient samples in Brazil (**Fig 5**).

634   To study further the history of the four modern components of Native American

635   ancestry, we evaluated the relationship of the time depth of the ancient samples

636   from either the Andes or the Southern Cone, with their genetic affinity to the

637   modern components of Native American ancestry (**Fig 7**). For that purpose, we

638   performed a linear regression between $f3($YRI, X, Anc$)$ – where X is one the four

639   Native American component – and the age of the ancient population tested (*Ancient*

640   in the f_3 formula*). We observed a significant relationship between the age of the

641   ancient Southern Cone samples and their genetic affinity with CCP ($P = 3.15 \times 10^{-4}$)

642   and CWA ($P = 3.94 \times 10^{-2}$) while no significant relationship was identified for LWL

643   and CAN ($P = 0.523$ and $P = 0.596$, respectively; **Fig 7A**). These patterns could be due

644   to a relationship between geography and the age of the ancient samples since the

645   most recent samples are concentrated in the Southern tip of the subcontinent (**Fig**

646   **4A**). To correct this putative confounding effect, we repeated the same analyses but

647   using the residuals of the linear regression between the age of the ancient samples

30

648   and their geographic coordinates. This correction intensified the relationship

649   described for CCP and CWA (P = 4.78x10^{-5} and P = 2.21x10^{-6}, respectively; **Fig 7C**). It

650   also allowed to actually identifying significant relationships for LWL and CAN (P =

651   8.07x10^{-4} and P = 1.56x10^{-3}, respectively). On the other hand, CAN is the only

652   modern Native American component that exhibits a significant relationship between

653   its genetic affinity with ancient Andean samples and their age (**Fig 7B**). This pattern

654   holds the correction for geography (**Fig 7D**). Repeating the same analyses using f_4

655   statistics, instead of f_3-Outgroup, we reached the same conclusions regarding the

656   relationships between the age of the ancient Andes samples or the Southern Cone

657   ancient samples and their genetic affinity with any of the four modern Native

658   American components (**S19 Fig**).
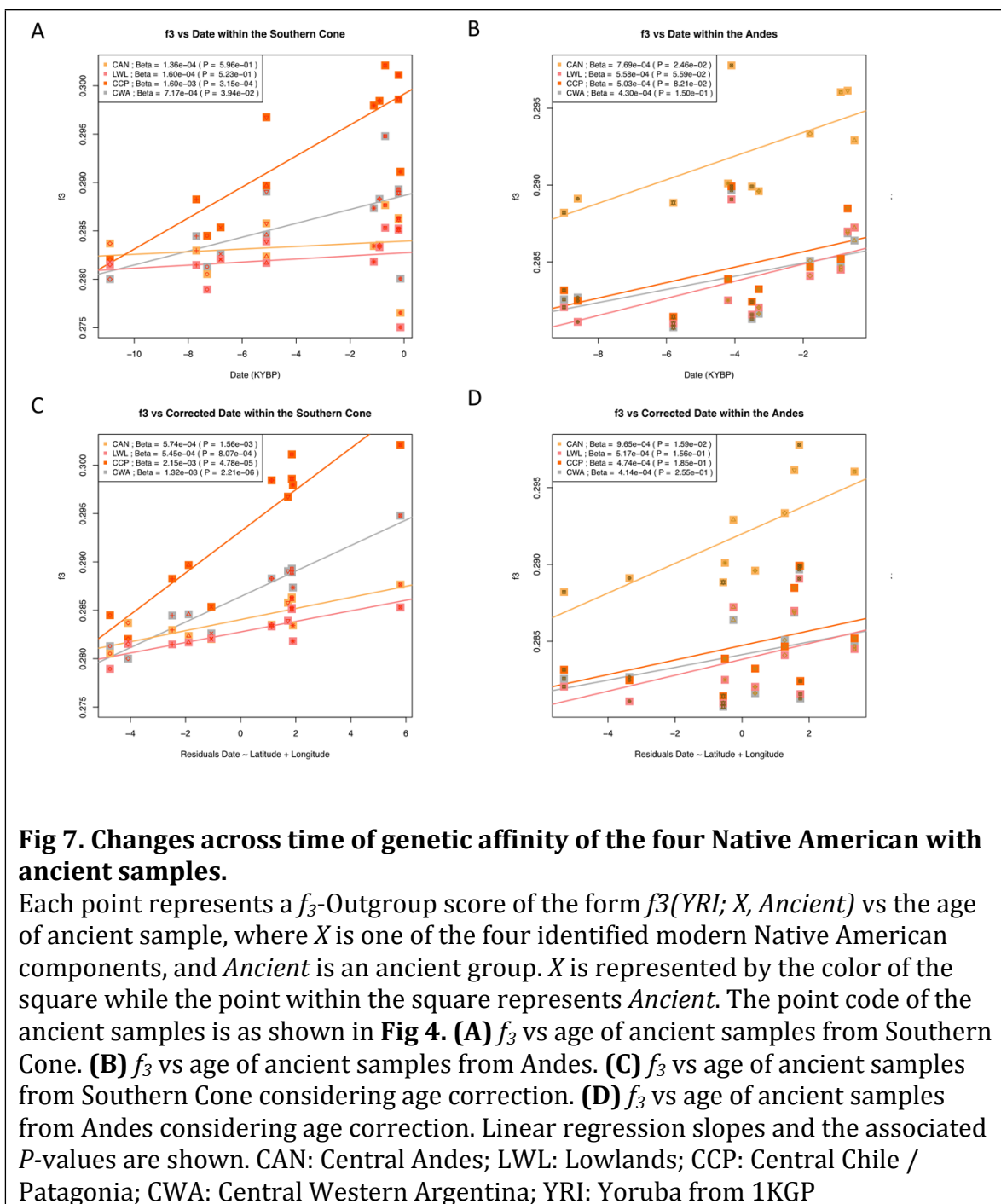
Fig 7. Changes across time of genetic affinity of the four Native American with ancient samples.
Each point represents a f_3-Outgroup score of the form *f3(YRI; X, Ancient)* vs the age of ancient sample, where *X* is one of the four identified modern Native American components, and *Ancient* is an ancient group. *X* is represented by the color of the square while the point within the square represents *Ancient*. The point code of the ancient samples is as shown in **Fig 4**. **(A)** f_3 vs age of ancient samples from Southern Cone. **(B)** f_3 vs age of ancient samples from Andes. **(C)** f_3 vs age of ancient samples from Southern Cone considering age correction. **(D)** f_3 vs age of ancient samples from Andes considering age correction. Linear regression slopes and the associated *P*-values are shown. CAN: Central Andes; LWL: Lowlands; CCP: Central Chile / Patagonia; CWA: Central Western Argentina; YRI: Yoruba from 1KGP

4.2.7 Comparison of the shared history of two components with ancient DNA
samples

The analysis of archaeological samples gives insights into the temporal depth of the

main demographic events that occurred during the evolutionary history of the

32

676   analyzed populations. Concerning the times of divergence of the four components

677   discussed in this study, although it is not possible to estimate them exactly, we can

678   establish lower and upper bounds. To explore this issue, we used another setting of

679   the *f4* statistics (**S5F Table**; **S20 Fig**). In this analysis, of the form *f4(YRI, Ancient; X,*

680   *Y),* a negative result is expected if *X* shares more history with an *Ancient* relative to

681   *Y,* while a positive result is expected if *Y* shares history with *Ancient* relative to *X.* We

682   observed that all Southern Cone ancient samples – except Los Rieles ~10900 BP

683   which has been described as derived from an ancestry shared with the Clovis

684   culture that has been replaced later [35]) – are more similar to CCP than to CWA

685   (**S20F Fig**), pointing to divergence time between those two components older than

686   ~7700 BP. CWA is not more attracted by any Brazilian and Andean ancient sample

687   as compared to CCP, reinforcing the idea that CWA is not a mix of CCP, CAN and LWL

688   related ancestries. Moreover, we observed that all the ancient groups of the

689   Southern Cone (with the exception of Los Rieles ~10900 BP) have greater genetic

690   affinity with CCP than with CAN and LWL (**S20B,D Figs**). With the exceptions of the

691   two Middle Holocene ancient samples from Patagonia (Ayayema ~5100 BP and

692   Punta Santa Ana ~ 7300BP) and Los Rieles ~10900 BP, ancient groups of the

693   Southern Cone exhibit higher genetic affinity with CWA than with LWL (**S20C Fig**).

694   In addition, the comparisons including LWL and CWA also point to a closer genetic

695   relation between ancient groups from the Southern Cone with CWA, although

696   significance is reached only for Late Holocene Patagonia samples.

697   All the ancient samples from the Andes, indifferently if there have been previously

698   assigned to the Early or Late population stream [35], are genetically more similar to

33

699 CAN than to any of the three other components (**S20A,D,E Figs**). These results

700 support the hypothesis that the ancestry associated with an early Andes population

701 arrival remains present to some extent in modern Central Andes populations and

702 has not been totally replaced by a later stream of gene flow. Although, ancient

703 samples from Brazil does not exhibit striking differences in their genetic affinity

704 with a modern Native American component as compared to another, it seems that

705 they are closer to LWL than to CWA and CCP but not than to CAN (**S20A-C Figs**).

706 Moreover, the archaeological samples in Brazil have higher genetic affinity with CAN

707 than with CWA and CCP (not significant for the latter comparisons). This points that

708 divergence of LWL with CCP and CWA could have occurred before than the

709 divergence of LWL with CAN (also observed in **S17A Fig**).

710 Altogether, the trends depicted by these analyzes are expected under a model of

711 early divergence among the four components. Our results also advocate to genetic

712 continuity over long period of time for the different components after they

713 differentiated.

714 **4.2.8 Past influence of a component on another since their divergence**

715 In order to get insights into the past genetic influence between two components

716 since their divergence, we applied a last *f4*-statistics analysis (**S5G Table**; **S21 Fig**).

717 In this analysis of the form *f4(Ancient, X; Y, YRI),* a negative result is expected if *Y*

718 shares history with *X* relative to *Ancient,* while a positive result is expected if *Y*

719 shares history with *Ancient* relative to *X.* We observed that the influence of LWL into

720 the three other components was relatively limited (**S21A,C,E Figs**). However, this

34

721   result suffers poor resolution due the temporal and geographical gaps in the ancient

722   samples representing this component. The influence of CAN appears to have been

723   more pronounced into LWL component than into CWA and CCP components

724   (**S21B,G,I Figs**). Strikingly, Pica Ocho ~700 BP, La Galgada ~4100 BP and

725   Aconcagua ~500 BP attract any of the four components relative to any other

726   component (not always reaching |Z| >3), with the exception of comparisons

727   including simultaneously CCP and CWA components. In other words, the relations of

728   a given modern component with Pica Ocho ~700 BP, La Galgada ~4100 BP and

729   Aconcagua ~500 BP are stronger than with any other modern component, except

730   when including both CWA and CCP in the comparison. The genetic influence of these

731   three Andes ancient groups outside the Central Andes may be related to events

732   geographical expansions of some Andean populations at given times. Discussing this

733   pattern further is beyond the scope of the present study.

734   The influence of CCP into LWL and CAN also appears to have been reduced, although

735   the negative *f4* with Kaweskar ~100 BP suggests gene flow between CCP and these

736   two components in very recent times (**S21D,H Figs**).

737   The mutual influence of CCP and CWA components seems to have been important in

738   remote times (**S21K,L Figs**). However, the available ancient samples provide much

739   more precise information on the influence of CCP into CWA component than for the

740   other way around. Indeed, CWA is not attracted by any ancient DNA group relative

741   to CCP (**S21K Fig**). Although Conchalí ~700 BP exhibit higher genetic affinity with

742   CWA among all the analyzed ancient groups (**S18D,H Figs**), it clearly attracts CCP

743   relative to CWA (**S21L Fig**) while it does not attract CWA relative to CCP. The

35

744   comparisons including CCP and CWA demonstrate that the archaeological record for

745   which genetic data has been generated does not include ancient samples

746   representing CWA since its early divergence with CCP neither ancient samples

747   related to the common ancestors specific to these two components.

748   **4.3   Conclusions**

749   We studied genetic ancestry at the sub-continental scale in Argentinean populations.

750   First, we confirmed that European ancestry is mostly explained by Southern

751   European origins, although we identified several individuals with Northern

752   European ancestry. Second, we described three main components of African

753   ancestry in Argentina. Although the Bantu-influenced and Western components are

754   clearly the most represented, we also found that an Eastern origin can explain some

755   of African ancestry in Argentina and represents up to ~30% of the African ancestry

756   for some Argentinean individuals. Studying more admixed individuals in Argentina

757   would help to get a better representation of the complex origins of African and

758   European ancestries in the country.

759   Concerning Native American ancestry, we concluded that Argentinian populations

760   share, in varying proportions, at least three distinct components: from the Central

761   Andes, from Central Chile/Patagonia and from the Lowlands. These components

762   correlate with geographic coordinates. Moreover, we present here some arguments

763   that the Native American ancestry in the Central Western region of Argentina may

764   actually derive from a fourth component that diverged early from the other Native

765   American components, and it maintained a tight link with the Central

36

766   Chile/Patagonia component. This relation is not explained by a putative

767   contribution of admixed individuals from Santiago de Chile (**S22 Fig**). Having

768   identified this component from admixed individuals demonstrates that to

769   characterize the Native American genetic diversity focusing only on indigenous

770   communities is insufficient, at least in Argentina.

771   It is important to acknowledge that our results should be taken cautiously since

772   studying admixed individuals make the analyses more complex. Leveraging a pure

773   statistical approach, we grouped individuals from rather culturally, ethnically and

774   linguistically heterogeneous groups to represent the four Native American

775   components discussed here. These groups exhibit within population structure, and

776   gene flow are most likely to have occurred among them after divergence. In

777   addition, temporal and geographical gaps in the archaeological samples analyzed,

778   particularly for the Central Western Argentina and Lowlands components, limits the

779   interpretative resolution of our analyses. Yet, the present study provides useful

780   insights into the routes followed by the main populations arrivals in the Southern

781   Cone (**S23 Fig**).

782   Further efforts are needed to better characterize the Native American ancestry

783   component identified in the Central Western region of Argentina. Particularly we

784   encourage future studies to confirm the tentative geographical label that we suggest

785   here, and to estimate its influence in the region. Besides these specific questions,

786   many other general questions remain to be answered to better understand the pre-

787   Colombian population dynamics in the Southern Cone such as the time and place of

788   the splits among the components described here, and the extent of genetic

37

789   exchanges among them. More genotype data for ancient samples, modern

790   indigenous communities and admixed individuals, particularly in Central,

791   Northeastern and Patagonia regions of Argentina, would help to decipher these

792   issues.

793   The genomic characterization of populations is an unavoidable practice for many

794   issues ranging from the understanding of our biological heritage, the rational use of

795   biobanks, the definition of an adequate reference genome, the estimation of

796   polygenic risk scores, the study and treatment of simple and complex diseases, and

797   the design of a national program of genomic medicine in our country. This study is a

798   common effort of exclusively Argentinean laboratories and totally financed by the

799   national scientific system. It is the first step of the *Consorcio Poblar*, a national

800   consortium for the creation of a public reference biobank to support biomedical

801   genomic research in the Republic of Argentina [75]. Genomic knowledge of local

802   populations should be a priority of developing countries for unbiased

803   representation of public databases and the scientific development of our countries.

5   Material and Methods

804

805   We genotyped 94 individuals with the Axiom LAT1 array (Affymetrix) from 24

806   localities and 17 provinces across Argentina (**Fig 1**). These samples were selected

807   among 240 collected by different population genetics groups (*Consorcio PoblAR*) in

808   the country. Besides ensuring an extended geographical range, inclusion criteria

809   were: (i) sufficient DNA concentration; (ii) Native American maternal lineage; (iii)

38

810   Native American paternal lineage (for males when available). All saliva and blood

811   samples have been collected with informed consent for research use, and we

812   dispose of the necessary approvals from different ethical committees in Argentina.

813   87 samples and 791,543 autosomal variants passed the standard Affymetrix

814   genotyping Quality Controls (**S1 Table**).

815   Most of the genotype data processing was performed using in-house scripts in R

816   [76] and perl [77], leveraging *plink2* [78], *vcftools v1.13* [79], and *bedTools v2* [80].

817   We compiled the genotype data for the 87 Argentinean samples with different

818   genotype data available in the literature. We focused our study on biallelic SNPs

819   (removing indels and SNPs with more than 2 alleles). Any putative inconsistent

820   strand had been fixed processing to the relevant flip, filtering out any SNP with

821   ambiguous genotype (A/T, G/C).

822   Cryptic relatedness among samples were assessed using King software [81]. To

823   avoid any 1st degree relationship, we filtered out individuals, minimizing the total

824   number of removals. No admixed individuals had been removed at this step.

825   In order to search for the fine-scale genetic structure of Argentinean we used

826   different data through this work. **S2 Table** depicts the main dataset arrangements

827   (named **DS<n>**).

828   ### 5.1   Argentinean genetic diversity within a worldwide context

829   To analyze genetic diversity in Argentina within a worldwide context, we built the

830   Dataset1 (**DS1**). This dataset contains 87 Argentinean samples, 654 African, 503

831   European and 347 American samples from 1KGP [45], 54 modern unrelated Chilean

832   samples [37], and 161 Native-American individuals from South America [38].

833   Moreover, **DS1** included genotype data from [33] which consists in 82 individuals

834   from Lima (Peru), 27 from Santiago de Chile, and 161 from Argentinean urban

835   centers.

836   We filtered out any variant and individual exhibiting in the compiled data set more

837   than 2% and 5% of missing genotypes (--geno 0.02 and --mind 0.05 flags in plink

838   1.9), respectively, as well as Minor Allele Frequency below 1% (--maf 0.01 flag in

839   plink 1.9). The filtered data was then pruned for Linkage-Disequilibrium (--indep-

840   pairwise 50 5 0.5 option in plink 1.9). The combined data set have a total

841   intersection of 59,237 SNPs and 2,076 individuals. With this curated data we

842   performed Principal Component Analyses [44] and Admixture [46] (**S1 and S2**

843   **Figs**).

5.2   Local Ancestry

845   Local ancestry analyses rely on haplotype reconstruction (phasing) and require high

846   SNP density. Since, admixed individuals of interest were genotyped on different

847   microarray platforms, we decided to perform two separate local ancestry analyses

848   on two different data sets (**DS2** and **DS3**). **DS2** consists of 87 Argentinean

849   individuals from the present study, and 54 Chilean individuals [37], all of them

850   genotyped with the Axiom LAT1 microarray. **DS3** includes 175 Argentinean, 27

851   Chilean, and 119 Peruvian genotyped with the Illumina OMNI1 microarray [33].

852   Both **DS2**, **DS3** were merged with 1KGP data consisting in 503 phased reference

40

853   samples for each of the African and European genetic ancestry, and 347 Latin

854   American individuals.

855   The 87 Argentinean, and the 54 Chilean samples were phased with shapeIT2

856   [82,83]. The genetic map, and 5,008 haplotypes panel provided by 1KGP were

857   downloaded from http://mathgen.stats.ox.ac.uk/impute/1000GP_Phase3/.

858   Algorithm, and model parameters were used by default, filtering out monomorphic

859   SNPs, and those with more than 2% of missing genotype. We obtained phased

860   genotype data for 608,501 autosomal SNPs. This data was then merged with phased

861   1KGP genotype data for African, American and European samples described before.

862   Since this data set derived from DS2, we call it **DS2p** (phased DS2).

863   The 175 Argentinean, 27 Chilean and 119 Peruvian samples from [33] were phased

864   separately using the same procedure with shapeIT2. After filtering for missing

865   genotypes and merging with phased 1KGP genotype data for African, American and

866   European samples we obtained phased data for 694,626 autosomal SNPs. Since this

867   data set derived from DS3, we call it **DS3p** (phased DS3).

868   In **DS2** and **DS3** we ran Admixture using K parameter minimizing the Cross-

869   Validation score. We used individuals with more than 99% of Native American

870   ancestry as references for local ancestry estimation. For **DS2** we used $K = 7$, and

871   Native American ancestry was defined as the sum of the two American specific

872   components observed (**S24 Fig**). According to this criterion, 48 individuals were

873   assigned as Native American reference: 20 PEL, 2 MXL, 23 Chilean from [37] (1

874   Huilliche, 9 Kaweskar, and 13 Pehuenche), and 3 Argentinean – 1 from the

41

875   Metropolitan Area of Buenos Aires (AMBA), 1 from San Salvador de Jujuy, and 1

876   from Salta). All the other American samples were defined as Admixed.

877   For **DS3**, we used K = 5, and Native American ancestry was defined as the single

878   American specific component observed (**S25 Fig**). According to this criterion, 19

879   individuals were assigned as Native American reference: 10 PEL, 5 from Lima, 1

880   from Santiago de Chile, and 3 Argentineans from Buenos Aires. All the other

881   American samples were defined as Admixed.

882   RFMix v2 ([47] downloaded at https://github.com/slowkoni/rfmix on 15th August

883   2018) was run on **DS2P** and **DS3P** separately using parameter settings similar to

884   [84]. The reference panels consist of the African and European samples from 1KGP,

885   as well as Native American individuals identified through Admixture procedure

886   described before. We used 1 Expectation-Maximization iteration (-e 1) using the

887   reference panels in this process (--reanalyze-reference). We used CRF spacing size

888   and random forest window size of 0.2 cM (-c 0.2 and -r 0.2). We use a node size of 5

889   (-n 5). We set the number of generations since admixture to 11 (-G 11) considering

890   the estimates from [33]. The forward-backward output was then interpreted to

891   assign allele ancestry to the one exhibiting major posterior probability, conditioning

892   that it was greater than 0.9. Otherwise, the allele ancestry was assigned to Unknown

893   (UKN). The global ancestry proportions estimated by this RFMix analysis were

894   compared with those obtained with Admixture software. The global ancestry

895   proportion estimates obtained by both procedures matched: spearman's correlation

896   greater than 0.9 in American samples for any of the 3 continental ancestries (**S5**

42

897   **Fig**).

5.3   Ancestry Specific Population Structure

899   In order to analyze the ancestry-specific population structure we masked the data,

900   i.e. for each individual, we assigned missing genotype for any position for which at

901   least one of the two alleles was not assigned to the relevant ancestry. In other

902   words, to study ancestry A, we kept for each individual, regions exhibiting ancestry

903   A on both haplotypes (ditypes) as illustrated in **S6 Fig**.

5.3.1   European ancestry specific population structure

905   To study European ancestry specific population structure, we analyzed together

906   masked data for this ancestry for individuals from **DS2P** and **DS3P** excluding

907   individuals from Chilean Native American communities [37]. This data was merged

908   with a set of reference individuals with European ancestry [49], which is a subset of

909   the POPRES dataset [48]. We call this data set as **DS4**. We removed individuals with

910   less than 30% SNPs with the ancestry ditypes (--mind 0.7 with *plink* 1.9). We also

911   removed SNPs with more than 50% of missing genotypes (--geno 0.5 with *plink* 1.9).

912   Thus, **DS4** contains 147 modern Argentinean individuals (71 from the present study

913   and 76 from [33]), 4 individuals from Santiago de Chile, 1 from Lima. **DS4**

914   encompasses 60,968 SNPs of which 54,073 remained after LD-pruning (--indep-

915   pairwise 50 5 0.5 flag in *plink2*).

916   *Smartpca* from Eigensoft v7.2.0 run on **DS4** [44] with the *lsqproject* option ON. We

917   report the PCA results summarized into a 2-dimensional space by applying

43

918   Multidimensional Scaling on weighted Euclidian distance matrix for the first N PCs.

919   We weighted each PC by the proportion of variance it explains. We selected the N

920   most informative PCs according to the Elbow method on the proportion of explained

921   variance.

922   Admixture [46] run with K ranging from 2 to 10 with cross-validation procedure.

923   5.3.2   African ancestry specific population structure

924   To study African ancestry specific population structure, we analyzed together

925   masked data for this ancestry for individuals from **DS2P** and **DS3P**. This data was

926   merged with a compilation of reference individuals with African ancestry from

927   [45,51–53]. We removed African individuals with less than 99% of African ancestry

928   when comparing them to the 2504 individuals from 1KGP (with K=7 minimizing

929   cross-validation score; results not shown). We thus reduced the African reference to

930   1685 individuals. We call as **DS5** the data set containing both the masked data for

931   admixed South American individuals and African reference individuals.

932   We removed SNPs with more than 10% of missing genotypes (--geno 0.1 with *plink*

933   1.9), and individuals with less than 5% of the ancestry ditypes (--mind 0.95 with

934   *plink* 1.9). Thus, **DS5** contains, 26 modern Argentinean individuals (all from the

935   present study), and 12 individuals from Lima (*9*). **DS5** consisted in 137,136 SNPs, of

936   which 128,086 remained after LD-pruning (--indep-pairwise 50 5 0.5 flag in *plink2*).

937   PCA and Admixture following as for European ancestry specific population structure

938   analyses (described before).

5.3.3   Native American ancestry specific population structure

940   To study Native American ancestry specific population structure, we analyzed

941   together masked data for this ancestry for individuals from **DS2P** and **DS3P**. This

942   data was merged with pseudo-haploid data for ancient samples within South and

943   Central America [35–37,60,61], as well as with masked data for Native American

944   individuals from [38]. The pseudo-haploid data for ancient samples was

945   downloaded from the Reich Laboratory webpage

946   (https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-

947   dna-data-compiled-published-papers) the 15th of May 2019. For each sample, we

948   used annotations (geographical coordinates and approximate date) from the

949   metafile provided at the same url.

950   We call this data set **DS6.** We removed individuals with less than 30% SNPs with the

951   ancestry ditypes (--mind 0.7 with *plink* 1.9). We also removed SNPs with less than

952   50% individuals with the ancestry ditypes (--geno 0.5 with *plink* 1.9). **DS6** contains

953   146 modern Argentinean individuals (74 from the present study and 72 from [33]),

954   22 individuals from Santiago and 77 from Lima, along with 207 Native South

955   American individuals from [37,38] and 53 ancient samples. **DS6** encompasses

956   47,003 SNPs, of which 39,423 remained after LD-pruning (--indep-pairwise 50 5 0.5

957   flag in *plink2*).

958   PCA and Admixture following as for European ancestry specific population structure

959   analyses (described before), with the difference that the *poplistname* option was set

45

960   for *smartpca* in order to estimate the PC using only modern samples and project the

961   ancient samples.

5.3.4   Statistical assignation of modern South American individuals to Native American components

964   Given a distance matrix among modern individuals from **DS6**, we performed *K-*

965   *means* clustering with the number of clusters ranging from 2 to 20 and selected the

966   *K-means* output minimizing the Bayesian Information Criterion (BIC). This

967   procedure was applied to three different distance matrix: (i) Weighted Euclidean

968   distance based on the two first PCs from Principal Component Analysis (ii)

969   Euclidean distance based on the ancestry proportions estimated from Admixture

970   (K=3) and (iii) $1 - f_3(Ind1,Ind2,Yoruba)$, where *Ind1* and *Ind2* are two individuals. In

971   the three cases, the K-means procedure minimized the BIC when considering 4

972   clusters, and the clusters tend to correspond to the three Native American

973   Component discussed in this paper (Central Chile/Patagonia, Lowlands, Andes)

974   along with one laying in-between, and which we finally attributed to Central

975   Western Argentina. However, individual assignment to one of these four clusters

976   was not totally consistent according to the distance matrix we used. To obtain a

977   robust assignation, an individual was assigned to a given cluster when it

978   consistently belonged to it across the three K-means procedures (**S15 Fig**),

979   otherwise it had been removed for following analyses. The cluster assignation for

980   each individual is given in **S4 Table**.

5.3.5 *F*-statistics to infer relationship between the four Native American components in Argentina and their genetic affinity with ancient populations

Starting from genotype data for the individuals in **DS6**, we also included genotype data from Yoruba (YRI) population from [45], Mixe population from [38] and, pseudo-haploid data for Anzick individual from the Clovis culture [74] and USR1 individual from Upward Sun River in Beringia [73]. The resulting data set is called **DS7**. Within **DS7,** we grouped modern individuals according to their assigned Native American group and we removed those with inconsistent assignation (**S15 Fig, S4 Table**). Thus, **DS7** contains 426 modern individuals, 55 ancient samples, 108 Yoruba and 17 Mixe individuals. **DS7** encompasses 88,564 SNPs. Note that we did not apply any SNP filtering overall **DS7** in order to maximize the number of SNPs included in each group comparison considered.

Using modern individuals from **DS7**, and considering any possible combination of the four groups identified (LWL, CCP, CAN and CWA), we computed the *f3* statistics [85] in the form of *f3(Target; S1, S2).* This allowed to contrast whether *Target, S1* and *S2* could be organized in the form of a phylogenetic tree *(*positive *f3)* or whether the *Target* group is the result of an admixture event between *S1* and *S2* groups (negative *f3*). We also computed *f4(YRI, Target; S1, S2)* to test whether group *Target* shares more evolutionary history with group *S1* (negative *f4)* or group *S2* (positive *f4*). Moreover, we reconstructed the Neighbour-joining tree from the matrix of distances with Phylip v3.2 [63] and *USR1* ancient sample from the Upward Sun

47

1003   River Site in Beringia [73] as outgroup. The distances were estimated as $1/f3$-

1004   *outgroup(YRI; X, Y).*

1005   We computed f_3-Outgroup and f_4 statistics in order to estimate the genetic affinity of

1006   the four Native American components with ancient populations from South and

1007   Central America. We computed the *f3-outgroup* and *f4* statistics in the form of

1008   *f3(Yoruba; X, Ancient)* and *f4(Yoruba, X; Ancient Beringia, Ancient)*, where *X* the

1009   represent the cluster containing all individuals assigned to a given Native American

1010   ancestry component, Ancient Beringia is the pseudo-haploid individual USR1 from

1011   the Upward Sun River site in Alaska [73].

1012   We also computed f_4 of the form *f4(YRI, Ancient; X, Y)*. This *f4* setting allowed testing

1013   which of *X* or *Y*, each referring to one of the four Native American components.

1014   shares more ancestry with a given *Ancient* group. *X* and *Y*.

1015   We finally computed f_4 of the form *f4(Ancient, X; Y, YRI)* to test whether a given

1016   modern Native American component *Y* shares exhibit closer genetic affinity with a

1017   given *Ancient* group (negative f_4) or with another modern Native American

1018   component *X* (positive f_4).

1019   All the results based on *F-statistics* are listed in **S5 Table**. We assessed significance

1020   of a comparison considering 3 standard errors ($|Z| > 3$).

1021

6 Abbreviations

1022

1023   **1KGP:** 1000 Genomes Project

1024   **AMBA:** Metropolitan Area of Buenos Aires

1025   **CCP:** Central Chile / Patagonia

1026   **CWA:** Central Western Region in Argentina

1027   **CYA:** Cuyo Region in Argentina

1028   **DS:** Data set

1029   **LD:** Linkage Disequilibrium

1030   **LWL:** Lowlands

1031   **NEA:** Northeastern Region in Argentina

1032   **NWA:** Northwestern Region in Argentina

1033   **PCA:** Principal Component Analysis

1034   **PPA:** Pampean Region in Argentina

1035   **PTA:** Patagonia Region in Argentina

1036   **CAN:** Central Andes

1037   **SNP:** Single Nucleotide Polymorphism

1038   **YRI:** Yoruba individuals in Ibadan, Nigeria

1039

7   Acknowledgments

8   Author Contributions

Conceptualization: Pierre Luisi, Angelina García, Darío Demarchi and Hernán Dopazo
Data acquisition: Emma Alfaro, Eliana Aquilano, Carina Argüelles, Sergio Avena, Graciela Bailliet, Julieta Beltramo, Claudio M. Bravi, Raul Carnese, Mariela Cuello, Cristina Dejean, Darío Demarchi, José Edgardo Dipierri, Angelina García, José Luis Lanata, Laura Jurado Medina, Josefina Motti, Maria Laura Parolin, Maia Pauro, Paula Paz, Virginia Ramallo, Daniela Rodríguez Golpe, María Rita Santos, Marisol Schwab, Natalia Silvero
Data curation: Pierre Luisi, Juan Manuel Berros and Jeremias Zubrzycki
Formal analysis: Pierre Luisi
Funding acquisition: Pierre Luisi, Virginia Ramallo and Hernán Dopazo
Investigation:
Methodology: Pierre Luisi, Juan Manuel Berros, Darío Demarchi and Hernán Dopazo

50

1067 **Project administration:** Pierre Luisi and Hernán Dopazo

1068 **Resources:** Juan Manuel Berros and Hernán Dopazo

1069 **Supervision:** Hernán Dopazo

1070 **Visualization:** Pierre Luisi.

1071 **Writing – original draft:** Pierre Luisi, Angelina García and Hernán Dopazo.

1072 **Writing – review & editing:** Carina Argüelles, Sergio Avena, Darío Demarchi,

1073 Josefina Motti, Maria Laura Parolin, Maia Pauro.

9 Disclosure Declaration

1074

1075 PL provides consulting services to myDNAmap S.A. JMB is employed by Biocódices

1076 S.A. HD is the scientific director of Biocódices S.A.

10 Data Availability Statement:

1077

1078 The data analyzed here comprises both newly generated and previously reported

1079 data sets. Access to publicly available data sets should be requested through the

1080 distribution channels indicated in each published study. For newly genotyped

1081 samples, individual genotype data will be made available at

1082 http://biocodices.hopto.org/Affymetrix_Pop_Arg/ at time of publication.

1083

51

11 Bibliography

1. Tishkoff SA, Verrelli BC. Patterns og Human Genetic Diveristy: Implications for Human Evolutionary History and Disease. Annu Rev of Genomics Hum Genet. 2003;4: 293–340.

2. Bravi CM, Sans M, Bailliet G, Martínez-Marignac V, Portas M, Barreto I, et al. Characterization of mitochondrial DNA and Y-chromosome haplotypes in a Uruguayan population of African ancestry. Hum Biol. 1997;69(5): 641–652.

3. Sala A, Penacino G, Corach D. Comparison of allele frequencies of eight STR loci from Argentinean Amerindian and European populations. Hum Biol. 1998;70(5): 937–947.

4. Goicoechea AS, Carnese FR, Dejean CB, Avena SA, Weimer TA, Franco MH, et al. Genetic Relationships Between Amerindian Populations of Argentina. Am J Phys Anthropol. 2001;115(2): 133–143.

5. Dejean CB, Crouau-Roy B, Goicoechea AS, Avena SA, Carnese FR. Genetic variability in Amerindian populations of northern Argentina. Genet Mol Biol. 2004;27(4): 489–495.

6. Fejerman L, Carnese FR, Goicoechea AS, Avena SA, Dejean CB, Ward RH. African ancestry of the population of Buenos Aires. Am J Phys Anthropol. 2005;128(1): 164–70.

7. 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(1): e1000037.

8. Bobillo MC, Zimmermann B, Sala A, Huber G, Röck A, Bandelt HJ, et al. Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database. Int J Legal Med. 2010;124(4): 263–8.

9. Pauro M, Garcia A, Bravi CM, Demarchi DA. Distribucion de haplogrupos mitocondriales aloctonos en poblaciones rurales de Cordoba y San Luis. Rev Argentina Antropol Biol. 2010;12(1): 47–55.

10. Pauro M, García A, Nores R, Demarchi DA. Analysis of uniparental lineages in two villages of Santiago Del Estero, Argentina, seat of Pueblos de Indios in colonial times. Hum Biol. 2013;85(5): 699–720.

11. García A, Pauro M, Bailliet G, Bravi CM, Demarchi DA. Genetic variation in populations from central Argentina based on mitochondrial and Y chromosome DNA evidence. J Hum Genet. 2018;63(4):493-507

12. 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;76(4): 543–57.

1123 13. Seldin MF, Tian C, Shigeta R, Scherbarth HR, Silva G, Belmont JW, et al.
1124   Argentine Population Genetic Structure: Large Variance in Amerindian
1125   Contribution. Am J Phys Anthropol. 2007;132(3): 455–62.

1126 14. Corach D, Lao O, Bobillo C, van Der Gaag K, Zuniga S, Vermeulen M, et al.
1127   Inferring continental ancestry of argentineans from autosomal, Y-
1128   chromosomal and mitochondrial DNA. Ann Hum Genet. 2010;74(1): 65–76.

1129 15. Parolin ML, Toscanini UF, Velázquez IF, Berardi GL, Holley A, Tamburrini C, et
1130   al. Genetic admixture patterns in Argentinian Patagonia. PLoS One.
1131   2019;14(6): e0214830.

1132 16. Avena SA, Via M, Ziv E, Pérez-Stable EJ, Gignoux CR, Dejean CB, et al.
1133   Heterogeneity in genetic admixture across different regions of Argentina.
1134   PLoS One. 2012;7(4): e34695.

1135 17. García A, Dermarchi DA, Tovo-Rodrigues L, Pauro M, Callegari-Jacques SM,
1136   Salzano FM, et al. High interpopulation homogeneity in Central Argentina as
1137   assessed by ancestry informative markers (AIMs). Genet Mol Biol.
1138   2015;38(3): 324–31.

1139 18. Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, et al. The
1140   African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum
1141   Genet. 2004;74(3): 454–65.

1142 19. Corach D, Figueira Risso L, Marino M, Penacino G, Sala A. Routine Y-STR
1143   typing in forensic casework. Forensic Sci Int. 2001;118(2-3): 131–5.

1144 20. Salas A, Jaime JC, Álvarez-Iglesias V, Carracedo A. Gender bias in the
1145   multiethnic genetic composition of central Argentina. J Hum Genet.
1146   2008;53(7): 662–74.

1147 21. García A, Pauro M, Nores R, Bravi CM, Demarchi DA. Phylogeography of
1148   mitochondrial haplogroup D1: An early spread of subhaplogroup D1j from
1149   Central Argentina. Am J Phys Anthropol. 2012;149(4): 583–90.

1150 22. Carnese FR. El mestizaje en la Argentina: indígenas, europeos y africanos: una
1151   mirada desde la antropología biológica. 1st ed . Ciudad Autónoma de Buenos
1152   Aires: Editorial de la Facultad de Filosofía y Letras Universidad de Buenos
1153   Aires; 2019.

1154 23. Martínez Sarasola C. Nuestros paisanos los indios. 9th ed. Buenos Aires: Del
1155   Nuevo Extremo; 2011.

1156 24. Devoto F. Historia de la inmigración en la Argentina. 1st ed. Buenos Aires:
1157   Editorial Sudamericana; 2003.

1158 25. Rosal MA (UNESCO). Africanos Esclavizados Llegados al Plata Durante el lapso
1159   tardocolonial. Afrodescendencia Aproximaciones contemporáneas de América
1160   Latina y el Caribe.; 2011. pp. 5–10.

1161 26. Guzmán F. Africanos en la Argentina. Una reflexión desprevenida. Andes.
1162   2006;17.

1163   27.   Borucki A. The Slave Trade to the Río de la Plata, 1777–1812: Trans-Imperial
1164   Networks and Atlantic Warfare. Colon Latin Am Rev. 2011;20(1):81–107.

1165   28.   Studer E. La trata de negros en el Río de la Plata durante el siglo XVIII. Buenos
1166   Aires; Instituto de Historia de Argentina "Doctor Emilio Ravignani"; 1958.

1167   29.   Rodríguez Molas R. Los sometidos de la conquista. Buenos Aires: Centro
1168   Editor de America Latina; 1985.

1169   30.   Mandrini R. La Argentina aborigen. De los primeros pobladores a 1910.
1170   Buenos Aires: Siglo XXI; 2008.

1171   31.   Lenton DI, Delrio WM, Pérez PMV, Papazian AER, Nagy MA, Musante M.
1172   Argentina's Constituent Genocide: Challenging the Hegemonic National
1173   Narrative and Laying the Foundation for Reparations to Indigenous Peoples.
1174   Armen Rev. 2012;53(1-4): 63–84.

1175   32.   de Jong IL. Guerra, Genocidio y Resistencia: Apuntes para discutir el fin de las
1176   fronteras en Pampa y Norpatagonia, Siglo XIX. Habitus. 2018;16: 229–254.

1177   33.   Homburger JR, Moreno-Estrada A, Gignoux CR, Nelson D, Sanchez E, Ortiz-
1178   Tello P, et al. Genomic Insights into the Ancestry and Demographic History of
1179   South America. PLOS Genet. 2015;11(12): e1005602.

1180   34.   Muzzio M, Motti JMB, Sepulveda PB, Yee MC, Santos MR, Ramallo V, et al.
1181   Population structure in Argentina. PLoS One. 2018;13(5): e0196325.

1182   35.   Posth C, Nakatsuka N, Lazaridis I, Fehren-schmitz L, Krause J, Reich D, et al.
1183   Reconstructing the Deep Population History of Central and South America.
1184   Cell. 2018;175(5): 1185–1197.e22

1185   36.   Moreno-Mayar JV, Vinner L, de Barros Damgaard P, de la Fuente C, Chan J, et
1186   al. Early human dispersals within the Americas. Science. 2018;362(6419): pii:
1187   eaav2621.

1188   37.   de la Fuente C, Ávila-arcos MC, Galimany J, Carpenter ML, Homburger JR,
1189   Blanco A, et al. Genomic insights into the origin and diversification of late
1190   maritime hunter-gatherers from the Chilean Patagonia. Proc Natl Acad Sci U S
1191   A. 2018;115(17): E4006–E4012.

1192   38.   Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, et al.
1193   Reconstructing Native American population history. Nature. 2012;488(7411):
1194   370–4.

1195   39.   Gnecchi-ruscone GA, Sarno S, De Fanti S, Gianvincenzo L, Giuliani C, Boattini A,
1196   et al. Dissecting the Pre-Columbian Genomic Ancestry of Native Americans
1197   along the Andes – Amazonia Divide. Mol Biol Evol. 2019;36(6): 1254–1269.

1198   40.   Gomez-Carballa A, Pardo-Seco J, Brandini S, Achilli A, Perego UA, Coble MD, et
1199   al. The peopling of South America and the trans-Andean gene flow of the first
1200   settlers. Genome Res. 2018;28(6):767-779

1201   41.   Michieli CT. La disolución de la categoría Jurídico-Social de "indio" en el siglo
1202   XVIII: El caso de San Juan (Región de Cuyo). San Juan: Instituto de

1203   Investigaciones Arqueoloógicas y Museo UNSJ; 2000.

1204 42. Katzer L. El mestizaje como dispositivo biopolítico. Biblos. Pueblos indígenas
1205   Interculturalidad, colonialidad, política. La Plata: Biblos; 2009. pp. 59–7.

1206 43. Briones C. Construcciones de aboriginalidad en Argentina. Société suisse des
1207   Américanistes / Schweizerische Amerikanisten-Gesellschaft. 2004;Bulletin
1208   68: 73–90.

1209 44. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS
1210   Genet. 2006;2(12): e190

1211 45. The 1000 Genomes Project Consortium. A global reference for human genetic
1212   variation. Nature. 2015;526(7571): 68–74.

1213 46. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry
1214   in unrelated individuals. Genome Res. 2009;19(9): 1655–64.

1215 47. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: A discriminative
1216   modeling approach for rapid and robust local-ancestry inference. Am J Hum
1217   Genet. 2013;93(2): 278–288.

1218 48. Nelson MR, Bryc K, King KS, Indap A, Boyko AR, Novembre J, et al. The
1219   Population Reference Sample, POPRES: A Resource for Population, Disease,
1220   and Pharmacological Genetics Research. Am J Hum Genet. 2008;83(3): 347–
1221   58.

1222 49. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, et al. Genes
1223   mirror geography within Europe. Nature. 2008;456(7218): 98–101.

1224 50. Gallero MC, Krautstofl EM. Proceso de poblamiento y migraciones en la
1225   Provincia de Misiones , Argentina (1881-1970). Posadas: Avá Rev Antropol.
1226   2010;16: 1881–1970.

1227 51. Patin E, Lopez M, Grollemund R, Verdu P, Harmant C, Quach H, et al.
1228   Dispersals and genetic adaptation of Bantu-speaking populations in Africa and
1229   North America. Science. 2017;356(6337): 543–546.

1230 52. Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Romero IG, et al.
1231   Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex
1232   Influences on the Ethiopian Gene Pool. Am J Hum Genet. 2012;91(1): 83–96.

1233 53. Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, Jay F, et al.
1234   Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and
1235   Complex African History. Science. 2012;338(6105): 374–9.

1236 54. Judica C. Análisis antropogenético de la población afrodescendiente de la
1237   Región de Noryungas, Bolivia. PhD Thesis. Universidad de Buenos Aires. 2017.

1238 55. Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR,
1239   et al. Reconstructing the Population Genetic History of the Caribbean. PLoS
1240   Genet. 2013;9(11): e1003925

1241 56. Fortes-Lima C, Gessain A, Ruiz-Linares A, Bortolini MC, Migot-Nabias F, Bellis

G, et al. Genome-wide Ancestry and Demographic History of African-Descendant Maroon Communities from French Guiana and Suriname. Am J Hum Genet. 2017;101(5): 725–736.

57. Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Williams S, et al. Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc Natl Acad Sci U S A. 2010;107(2): 786–91.

58. Pauro M. Análisis molecular de linajes uniparentales en poblaciones humanas del centro de Argentina. PhD Thesis. Universidad de Córdoba. 2015.

59. Di Fabio Rocca F. La presencia subsahariana en el acervo génico de poblaciones cosmopolitas de la Argentina. PhD Thesis. Universidad de Buenos Aires. 2016.

60. Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science. 2015;349(6250): aab3884.

61. Lindo J, Haas R, Hofman C, Apata M, Moraga M, Verdugo RA, et al. The genetic prehistory of the Andean highlands 7000 years BP though European contact. Sci Adv. 2018;4(11): eauu4921.

62. Modolo VE. Análisis histórico-demográfico de la inmigración en la Argentina del Centenario al Bicentenario. Papeles de Población. 2016;22(89): 201–222.

63. Felsenstein J. PHYLIP: Phylogeny Inference Package (Ver. 3.2). Cladistics. 1989;5: 164–166.

64. Motti JMB, Muzzio M, Ramallo V, Kladniew BR, Alfaro EL, Dipierri JE, et al. Origen Y Distribución Espacial De Linajes Maternos Nativos En El Noroeste Y Centro Oeste Argentinos. Rev Argentina Antropol Biol. 2013;15(1): 3–14.

65. Motti JMB, Schwab ME, Beltramo J, Jurado-medina LS, Muzzio M, Ramallo V, et al. Regional differentiation of native American populations from the analysis of maternal lineages. Intersecc en Antropol. 2017;18(3): 271–282.

66. Motti J. Caracterización de Linajes Maternos en la Población Actual del Noroeste y Centro-Oeste Argentinos. PhD Thesis. Universidad Nacional de La Plata. 2012.

67. Canals Frau S. Etnología de los Huarpes. Una síntesis. Anales del Instituto de Etnología Americana. Mendoza; Anales del Instituto de Etnología Americana, Tomo VII; 1946. pp. 9–147.

68. Escolar D. Los dones étnicos de la Nación. Identidades huarpe y modos de producción de soberanía en Argentina. Buenos Aires; Prometeo; 2007.

69. Bengoa J. Historia de los antiguos mapuches del sur. Desde antes de la llegada de los españoles hasta las paces de Quilín. Santiago: Catalonia; 2003.

70. Galdames S. ¿Detuvo la batalla del Maule la expansión Inca hacia el sur de Chile? Cuadernos Hist. 1989;3: 7–25.

1282 71. Michieli CT. Antigua historia de Cuyo. San Juan: Ansilta; 1994.

1283 72. Gascón M. Cuyo en el espacio imperial. La fase de configuración: 1580-1680.
1284   Rev Tefros. 2011;9(1-2): 6.

1285 73. Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, Terhorst
1286   J, et al. Terminal Pleistocene Alaskan genome reveals first founding
1287   population of Native Americans. Nature. 2018;553(7687): 203–207.

1288 74. Rasmussen M, Anzick SL, Waters MR, Skoglund P, Degiorgio M, Stafford TW Jr,
1289   et al. The genome of a Late Pleistocene human from a Clovis burial site in
1290   western Montana. Nature. 2014;506(7487): 225–229.

1291 75. Dopazo H, Llera AS, Berenstein M, Gonzáles-José R. Genomas , enfermedades y
1292   medicina de precisión : un Proyecto Nacional. Ciencia, Tecnol y Política.
1293   2019;2(2): 1–10.

1294 76. Team RC. R: A language and environment for statistical computing. R
1295   Foundation for Statistical Computing, Vienna, Austria. 2019.

1296 77. Wall L, Christiansen T, Orwant J. Programming perl. O'Reilly M. 2000.

1297 78. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-
1298   generation PLINK: Rising to the challenge of larger and richer datasets.
1299   Gigascience. 2015;4:7.

1300 79. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA, et al. The
1301   variant call format and VCFtools. Bioinformatics. 2011;27(15): 2156–8.

1302 80. Quinlan AR, Hall IM. BEDTools : a flexible suite of utilities for comparing
1303   genomic features. Bioinformatics. 2010;26(6): 841–842.
1304   doi:10.1093/bioinformatics/btq033

1305 81. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust
1306   relationship inference in genome-wide association studies. Bioinformatics.
1307   2010;26(22): 2867–73.

1308 82. Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for
1309   thousands of genomes. Nat Methods. 2011;9(2): 179–181.

1310 83. Delaneau O, Zagury J, Marchini J. Improved whole chromosome phasing for
1311   disease and population genetic studies. Nat Methods. 2013;10(1): 5-6.

1312 84. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al.
1313   Human Demographic History Impacts Genetic Risk Prediction across Diverse
1314   Populations. Am J Hum Genet. 2017;100(4): 635–649.

1315 85. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient
1316   admixture in human history. Genetics. 2012;192(3): 1065–93.

1317

12 Supplementary Figure Captions

S1 Fig. Principal component analysis in a Worldwide Context.
The percentage of variance explained by each principal component (PC) is given.
Each point represents an individual following the color and point codes given in
legend. **A:** PC2 vs PC1; **B:** PC4 vs PC3; **C:** PC6 vs PC5.

S2 Fig. Admixture Analyses in a Worldwide Context.
(A) Cross-Validation score for Admixture runs on the global meta dataset with K
from 3 to 12. **(B-G)** Admixture results with K=3 to K=8. 1KGP: 1000 Genomes
Project; CYA: Cuyo Region in Argentina; NEA: Northeastern Region in Argentina,
NWA: Northwestern Region in Argentina; PPA: Pampean Region in Argentina; PTA:
Patagonia Region in Argentina.

S3 Fig. Comparison of Different Admixture Continental Ancestry Proportion
Estimates in a Worldwide Context.
Comparison of the African, European and Native American ancestry proportion
estimates obtained with Admixture with K=3 and K=8. Individuals are represented
with points according to the legend of Figure 2. **(A)** African ancestry proportions for
K=3 are as observed in green in Supplementary S2B Fig, while for K=8 they are
estimated as the sum of the three greenish colors observed in S1D-G Figs. **(B)**
European ancestry proportions for K=3 are as observed in blue in Supplementary
S2B Fig, while for K=8 they are estimated as the sum of the two bluish colors
observed in S1D-G Figs. **(C)** Native American ancestry proportions for K=3 are as
observed in orange in S2B Fig, while for K=8 they are estimated as the sum of the
three reddish colors observed in S2D-G Figs.

S4 Fig. Comparison of Different Admixture Continental Ancestry Proportions
Among Different Samples.
Comparison of ancestry proportion estimates with admixture (K=8) among samples
from the present study, and Argentinean, Chilean and Peruvian samples from [33].
(A) African Ancestry, **(B)** European Ancestry, **(C)** Native American Ancestry. P-value
of the Wilcoxon test for each pairwise comparison is shown.

S5 Fig. Comparison of Admixture and RFMix Ancestry Proportion Estimates in
DS2p and DS3p.
(A-F) For DS2p: Argentinean samples from the present study with reference panel
that consists in 1KGP individuals from Africa, Europe and America [45] and Chilean
individuals from [37]. **(A-C)** Native American, European and African ancestry
proportions estimates with RFMix vs with Admixture with K=3. **(D-F)** Native
American, European and African ancestry proportions estimates with RFMix vs with
Admixture (K=7). **(G-L) For DS3p:** Argentinean samples from [33] with reference
panel that consists in 1KGP individuals from Africa, Europe and America [45]. **(G-I)**
Native American, European and African ancestry proportions estimates with RFMIx

58

1357 vs with Admixture with *K=3*. **(J-L)** Native American, European and African ancestry
1358 proportions estimates with RFMIx vs with Admixture with *K=5*.

S6 Fig. Example of a Local Ancestry Output.
1360 **(A)** RFMIX output for a given admixed individual. **(B)** Masked genotype showing
1361 ditypes of Native American (red), European (blue) and African (green) ancestry.
1362 Gaps are represented in grey and regions with unassigned ancestry (Unknown) are
1363 in black.

S7 Fig. European Ancestry Specific Admixture Analysis.
1365 **(A)** Cross-Validation scores for *K* from 2 to 10. **(B)** Admixture plots for *K=2*. **(C)**
1366 Admixture plots for *K=3*. CYA: Cuyo Region in Argentina; NEA: Northeastern Region
1367 in Argentina, NWA: Northwestern Region in Argentina; PPA: Pampean Region in
1368 Argentina; PTA: Patagonia Region in Argentina.

S8 Fig. European Ancestry Specific Principal Component Analysis.
1370 **(A)** Multidimensional scaling based on the weighted Euclidean distance for PCs 1 to
1371 3. Individuals from the European reference panel are colored according to main
1372 European geographic regions as shown in **(B)** while South American admixed
1373 individuals are represented as shown in the legend**.** Elbow method to choose the
1374 number of relevant PCs is shown as inset within (A).

S9 Fig. Choose The Number Of Principal Components From African Ancestry-
1376 **Specific Principal Component Analysis.**
1377 Elbow method to determine which PC minimizes the angle of the curve from the
1378 chat "Percentage of variance explained *versus* Number of PCs"

S10 Fig. African Ancestry-Specific Admixture Analysis
1380 **(A)** Cross-validation scores *K* from 2 to 10. **(B)** Admixture plots for *K=2*. **(C)**
1381 Admixture plots for *K=3*. **(D)** Admixture plots for *K=4*. CYA: Cuyo Region in
1382 Argentina; NEA: Northeastern Region in Argentina, NWA: Northwestern Region in
1383 Argentina; PPA: Pampean Region in Argentina; PTA: Patagonia Region in Argentina

S11 Fig. Choose The Number Of Principal Components From Native American
1385 **Ancestry-Specific Principal Component Analysis.**
1386 Elbow method to determine which PC minimizes the angle of the curve from the
1387 chart "Percentage of variance explained *versus* Number of PCs"

S12 Fig. Native American Ancestry-Specific Admixture Analysis
1389 **(A)** Cross-validation scores *K* from 2 to 10. **(B)** Admixture plots for *K=2*. **(C)**
1390 Admixture plots for *K=4*. **(D)** Admixture plots for *K=5*. CYA: Cuyo Region in
1391 Argentina; NEA: Northeastern Region in Argentina, NWA: Northwestern Region in
1392 Argentina; PPA: Pampean Region in Argentina; PTA: Patagonia Region in Argentina

S13 Fig. Correlation of Native American Ancestry Proportions and
1394 **Geographical Coordinates in Argentina**
1395 **(A)** Central Andes ancestry proportions vs Latitude. **(B)** Central Andes ancestry
1396 proportions vs Longitude. **(C)** Lowlands ancestry proportions vs Latitude. **(D)**

1397 Lowlands ancestry proportions vs Longitude. **(E)** Central Chile/Patagonia ancestry
1398 proportions vs Latitude. **(F)** Central Chile/Patagonia ancestry proportions vs
1399 Longitude. Linear regression slopes and the associated *P*-values are shown.

1400 **S14 Fig. Comparison of Native American Ancestry Proportion Estimates**
1401 **Obtained with Admixture Run on Unmasked Data (*K=8)* and Masked data**
1402 **(*K=3*).**
1403 Spearman correlation coefficients and associated *P*-values are shown. CAN: Central
1404 Andes; LWL: Lowlands; CCP: Central Chile / Patagonia.

1405 **S15 Fig. Individual Assignation to a Native American Ancestry Cluster.**
1406 Consensus clustering South American individuals based on three K-means
1407 procedures run with different pairwise distances among individuals. **(Top)** K-means
1408 results using Ancestry-Specific PCA and Admixture (ASPCA and AS-Admixture), and
1409 f_3 results to compute pairwise distances. Individuals are represented as in **Main Fig**
1410 **4**. Insets: BIC score for number of clusters set to *K-means* ranging from 2 to 20. In all
1411 the three cases, *K*-means BIC was minimized when considering 4 clusters. **(Bottom)**
1412 Same as top with point colors corresponding to the assigned cluster.

1413 **S16 Fig. Pairwise Genetic Affinity Among Individuals Assigned to Different**
1414 **Groups.**
1415 Boxplots of the *1-f3(YRI; Ind1, Ind2),* where *Ind1* and *Ind2* are two individuals
1416 belonging to Group 1 and Group 2, respectively. The groups are either the fourth
1417 Native American components identified or ancient Southern Cone groups from
1418 Middle Holocene. For clarity, boxplot outliers are not shown. YRI: Yoruba from
1419 1KGP.

1420 **S17 Fig. Graphical Visualization of Pairwise Genetic Distances Among Groups**
1421 **in South America**
1422 **(A)** Neighbor-joining tree from distances of the form $1/f_3(YRI; X, Y)$. USR1 from
1423 Ancient Beringia was used as *outgroup* **(B)** Multidimensional-scaling from distances
1424 of the form $1-f_3(YRI; X, Y)$. Each group is represented as appearing in the leaf of (A).
1425 USR1 and Anzick-1 were not considered in (B). YRI: Yoruba from 1KGP

1426 **S18 Fig. Genetic Affinity of the Four Native American Components with Ancient**
1427 **Groups.**
1428 **(A-D)** *f3(YRI; X; Ancient)*. **(E-H)** *f4(YRI, X; Ancient Beringia, Ancient).* **(A)** and **(E)**:
1429 with Central Andes (CAN) as *X.* **(B)** and **(F)**: With Lowlands (LWL) as *X.* **(C)** and **(G)**:
1430 With Central Chile / Patagonia (CCP) as *X.* **(D)** and **(H)**: With Central Western
1431 Argentina (CWA) as *X.* YRI: Yoruba from 1KGP; *Ancient Beringia*: USR1 individual
1432 from [73]; *X*: Native American component in Argentina (one plot per *X). Ancient*:
1433 ancient group labeled on the *x*-axis and represented with a point/color scheme as in
1434 **Main Figure 4**. Vertical segments are the +/- 3 standard errors intervals.

1435 **S19 Fig. Changes Across Time of Genetic Affinity of the Four Native American**
1436 **Components with Ancient Groups.**

Each point represents a *f4* score of the form *f4(YRI, X; Ancient Beringia, Ancient)* vs the age of ancient sample, where *X* is one of the four identified Native American components, and *Ancient* is an ancient group. *X* is represented by the color of the square while *Ancient* is represented by the point within the square. The point code of the ancient samples is shown in Fig.4. *Ancient Beringia*: USR1 individual from [73]. **(A)** f_4 vs age of ancient samples from Southern Cone. **(B)** f_4 vs age of ancient samples from Andes. **(C)** f_4 vs age of ancient samples from Southern Cone considering age correction. **(D)** f_4 vs age of ancient samples from Andes considering age correction. Linear regression slopes and the associated *P*-values are shown. CAN: Central Andes; LWL: Lowlands; CCP: Central Chile / Patagonia; CWA: Central Western Argentina; YRI: Yoruba from 1KGP

S20 Fig. Comparison of Genetic Affinity of an Ancient Group to a Native American Component Relative to Another.
f_4*(YRI, Ancient, X, Y)* where *X* and *Y* are two of the four identified Native American components (one plot per *X-Y* combination), and *Ancient* is ancient group labeled on the *x*-axis and represented with a point/color scheme as in Main Figure 4. Vertical segments are the +/- 3 standard errors intervals. Note this setting for f_4 statistics is symmetrical when switching *X* and *Y.*

S21 Fig. Comparison of Genetic Affinity of a Native American Component to Another Relative to an Ancient Group.
f_4*(Ancient, X; Y, YRI)* where *X* and *Y* are two of the four identified Native American components (one plot per *X-Y* combination), and *Ancient* is ancient group labeled on the *x*-axis and represented with a point/color scheme as in Main Figure 4. Vertical segments are the +/- 3 standard errors intervals. CAN: Central Andes; LWL: Lowlands; CCP: Central Chile / Patagonia; CWA: Central Western Argentina; YRI: Yoruba from 1KGP

S22 Fig. Removing Admixed Santiago de Chile Individuals from Analyses Does Not Affect the Results.
Admixed individuals from Santiago de Chile were removed to perform the analyses presented in this figure. **(A)** f_3(*Target; S1, S2*) to test for treeness; **(B)** f_4*(YRI, Target; S1, S2)* to test whether Target shares more ancestry with S1 or S2; **(C)** *f3(YRI; CWA; Ancient)*; **(D)** *f4(YRI, CWA; Ancient Beringia, Ancient)*; **(E)** f_4*(Ancient, CCP; CWA, YRI)*; **(F)** *f4(Ancient, CWA; CCP YRI)*; **(G)** *f4(YRI, Ancient; CWA, CCP).* CCP: Central Chile / Patagonia.

S23 Fig. Schematic Routes for the Main Population Arrivals in the Southern Cone.
Each arrow represents one of the four components discussed throughout the article. Neither disentangling the time and place of the splits among these components nor the extent of potential gene flow have been addressed in this study.

S24 Fig. Admixture Analyses in DS2 to Define European, African and Native American Reference Individuals for Local Ancestry Analyses
(A) Cross-validation scores from *K*=3 to *K*=10. **(B)** Admixture plot for *K=7*

1479 **S25 Fig. Admixture Analyses in DS3 to Define European, African and Native**
1480 **American Reference Individuals for Local Ancestry Analyses**
1481 **(A)** Cross-validation scores from *K*=3 to *K*=10. **(B)** Admixture plot for *K=5*

13 Supplementary Table Captions
1482

1483 **S1 Table. Sample information.**
1484 Sampling location, Gender, uniparental lineages, Affymetrix QC metrics, Color and
1485 Point coding used for plots throughout the article.

1486 **S2 Table. Data sets (DS) used throughout the article.**

1487 **S3 Table. Ancestry proportion estimates in a worldwide context.**
1488 Ancestry proportion estimates from Admixture analyses with *K=3* and *K*=8 at the
1489 worldwide level. The column names describe the labels attributed to each ancestry
1490 detecting for both Admixture analyses, as well as the hexadecimal code for the color
1491 used to represent it in the corresponding admixture plot. The columns "Point",
1492 "Color" and "cex" list the graphic parameters used to represent each individual in
1493 the different plots throughout the article.

1494 **S4 Table. Native American cluster assignation.**
1495 Individual Native American cluster assignation is given for each of the *K*-means
1496 procedures and for the consensus call (columns "F3", "PCA", "Admixture" and
1497 "Consensus"). The ancestry proportion estimates from Admixture analyses with *K=3*
1498 on the masked data for Native American ancestry are also provided. The column
1499 names explicit the labels attributed to each ancestry detecting for both Admixture
1500 analyses as well as the hexadecimal code for the color. The column "Argentinean
1501 describes whether an individual was used for Argentinean territory specific
1502 analyses or not. For admixed individuals (from the present study and [33]), the
1503 "Population" and "Region" columns list the locality and province, respectively, while
1504 for Native American population (from [37,38])) the "Population" and "Region"
1505 columns list the ethnic and main ethnic groups, respectively.

1506 **S5 Table. *F*-statistics results**
1507 **(A)** *f3(Target; S1, S2)* only for comparisons including Native American components
1508 **(B)** *f4(YRI, Target; S1, S2)* only for comparisons including Native American
1509 components **(C)** *f3(YRI; X, Y)* only for comparisons including Ancient Beringia, Mixe
1510 and Native American components **(D)** *f3(YRI; X, Y)* where *X* and *Y* can be either an
1511 ancient group or one of the four Native American components. **(E)** *f4(YRI, X; Ancient*
1512 *Beringia, Ancient)* only for comparisons including between a Native American
1513 components (*X*) and an ancient group (*Ancient*). **(F)** *f4(YRI, Ancient, X, Y)* only for
1514 comparisons including two Native American components (*X, Y*) and an ancient
1515 group (*Ancient*). **(G)** *f4(Ancient, X; Y, YRI)* only for comparisons including two Native
1516 American components (*X, Y*) and an ancient group (*Ancient*). YRI: Yoruba from 1KGP
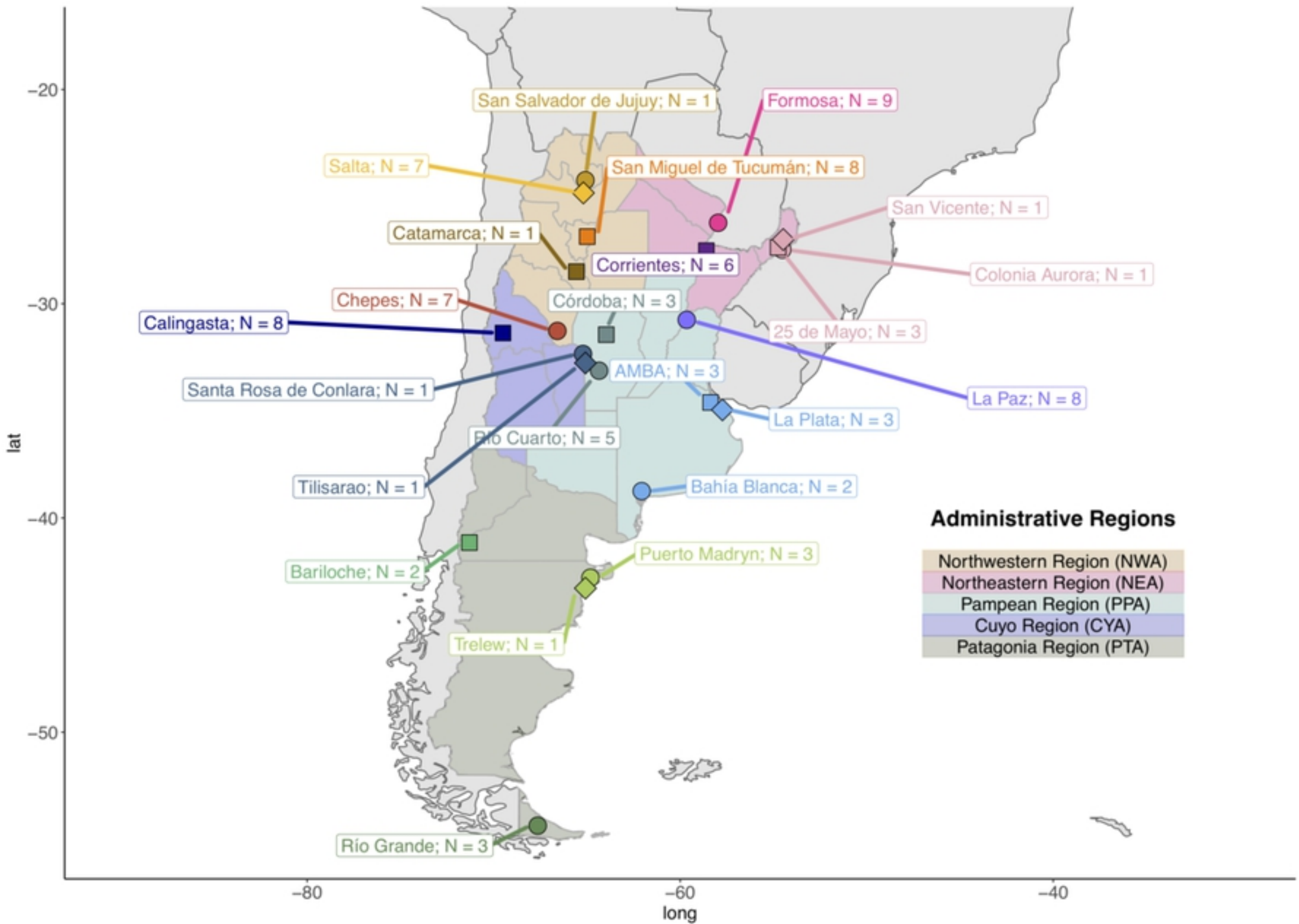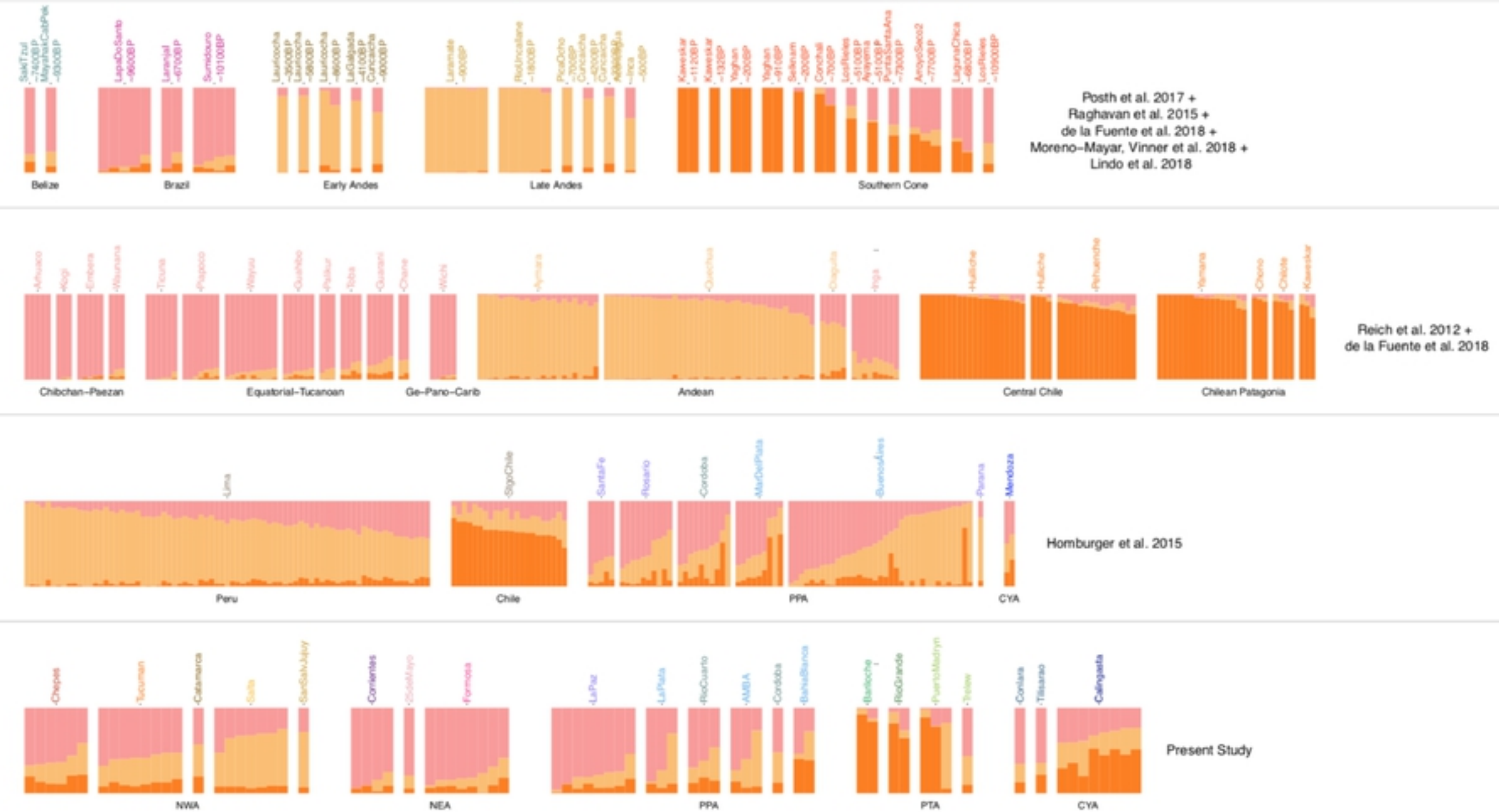
Fig 1. Sample locations from the present study

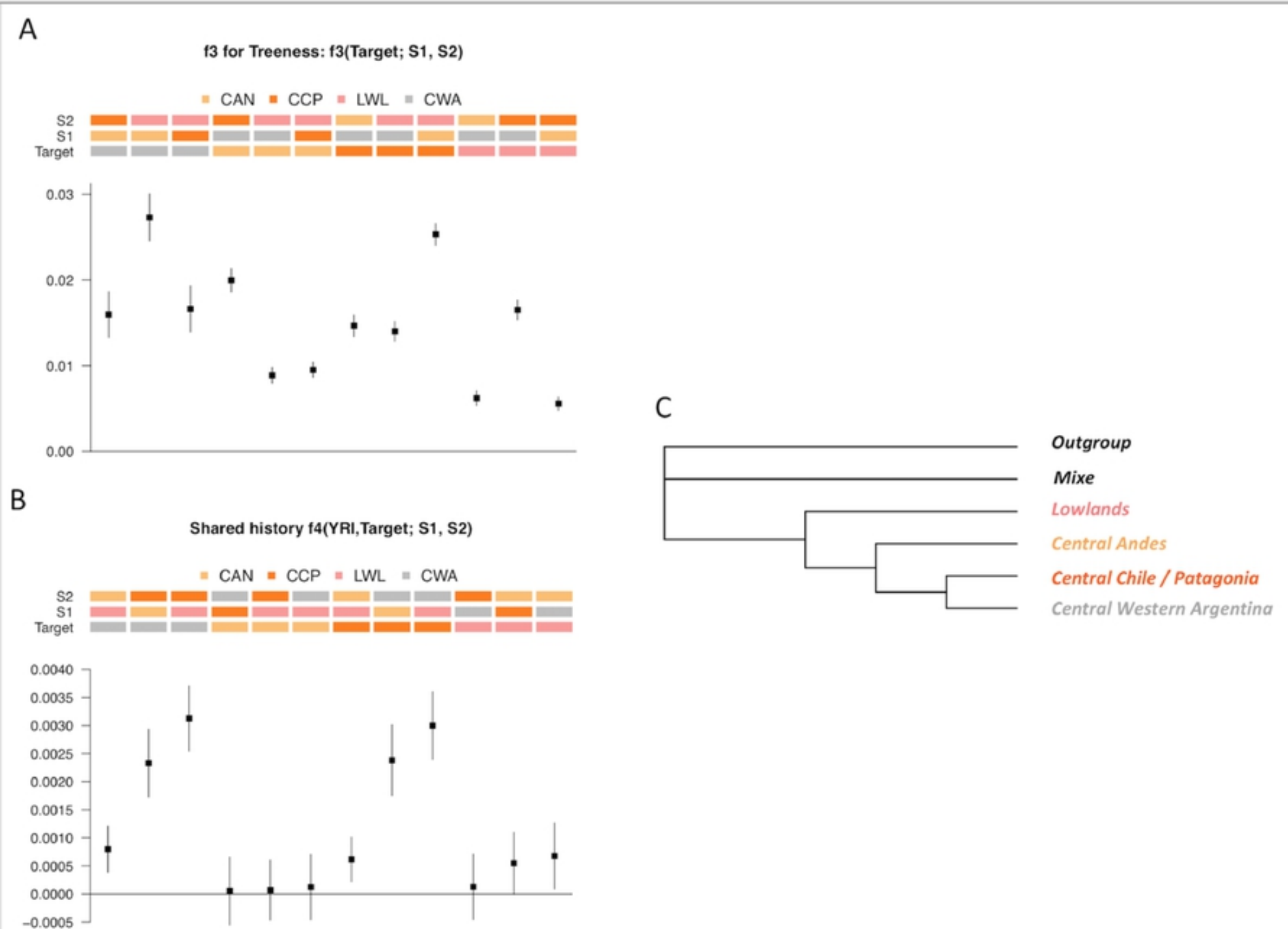Fig 5. Native American Ancestry-Specific Admixture Analysis.

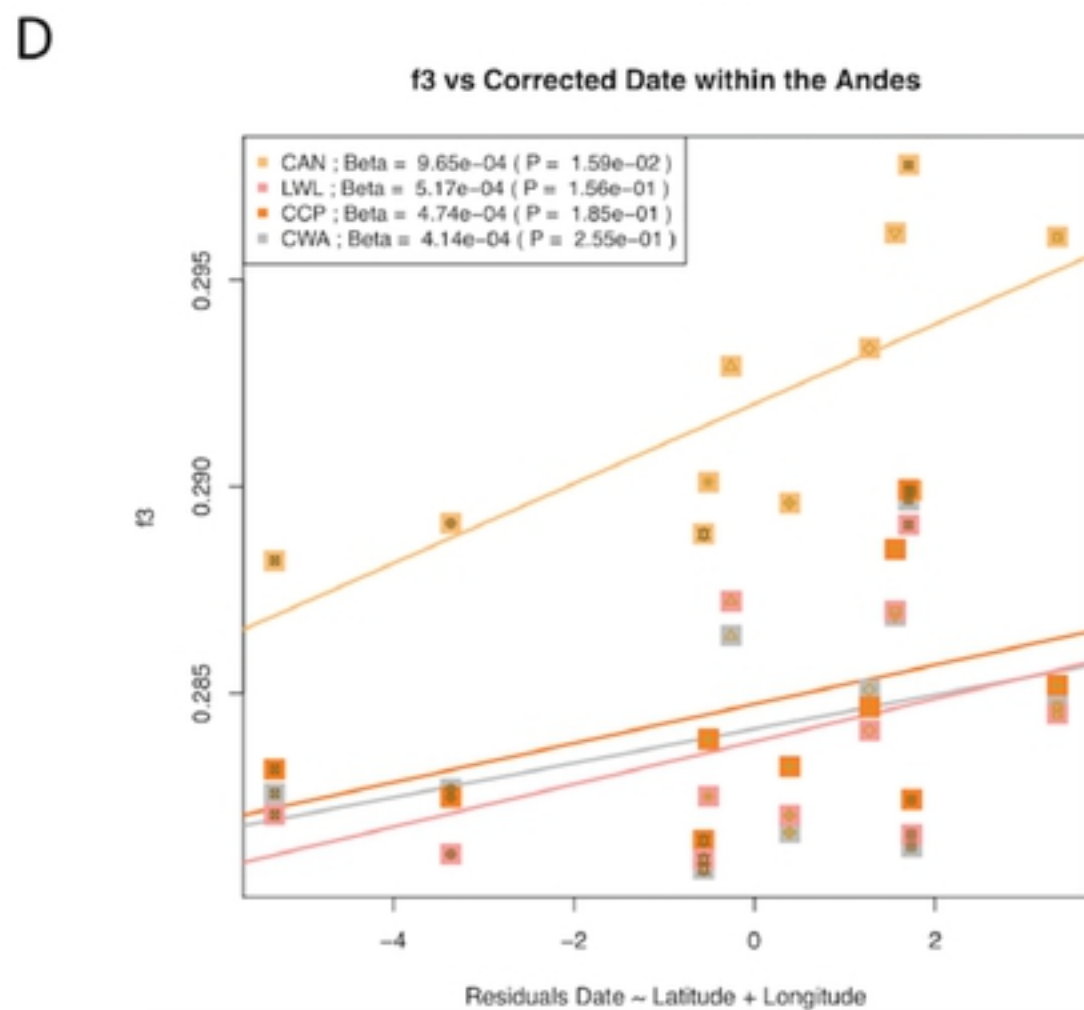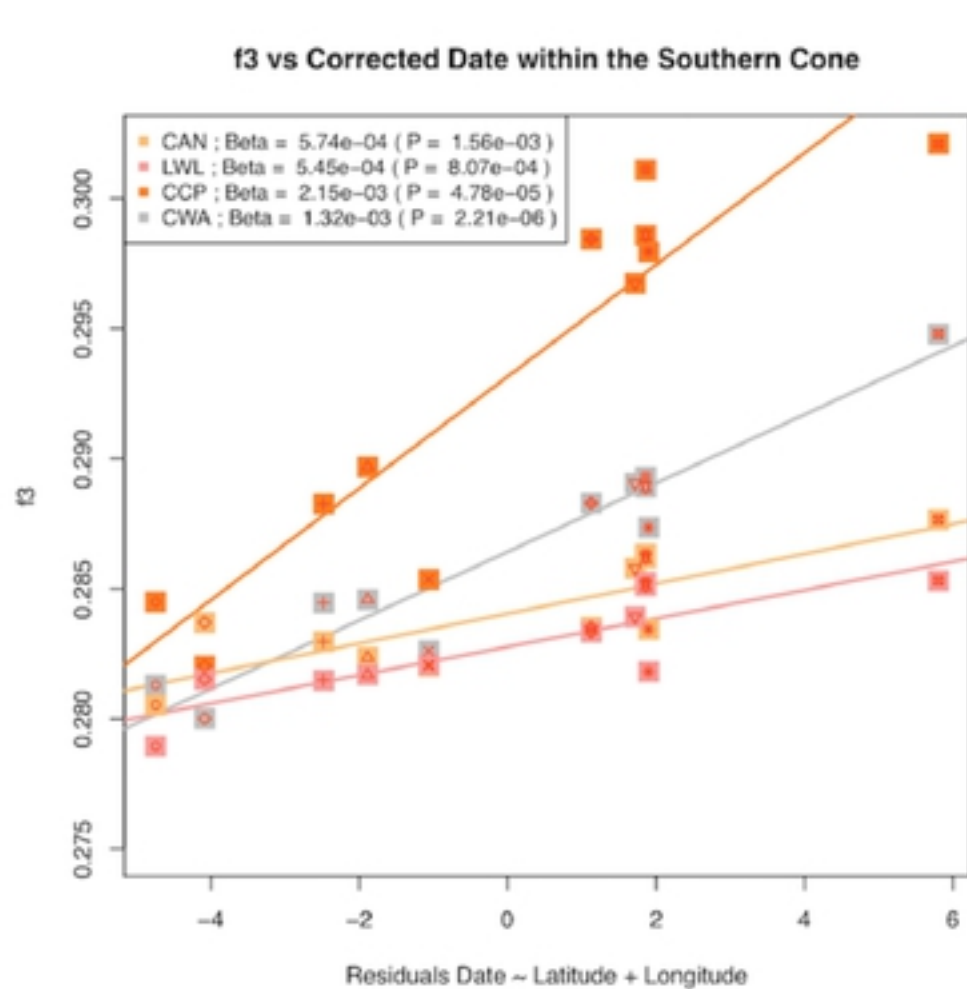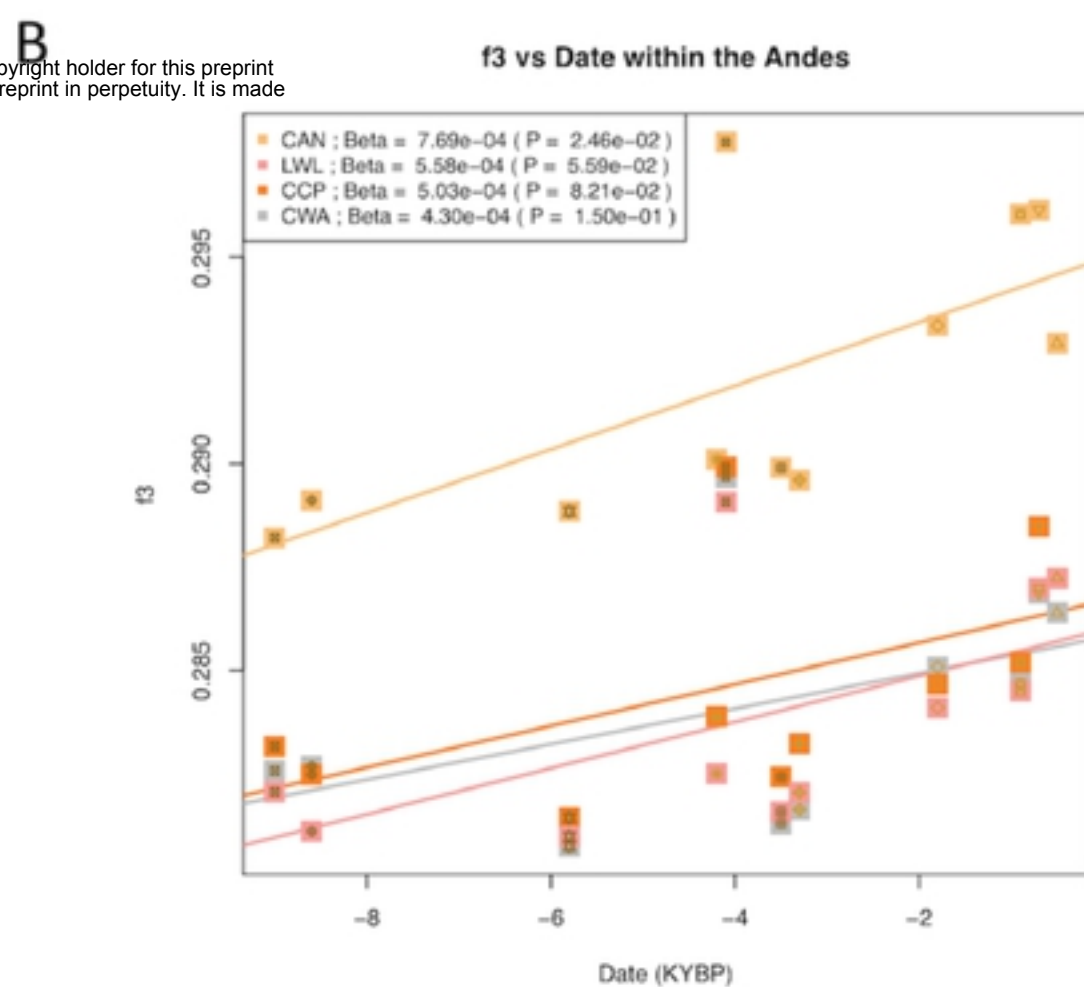Fig 6. Relationship among the four Native American groups iden

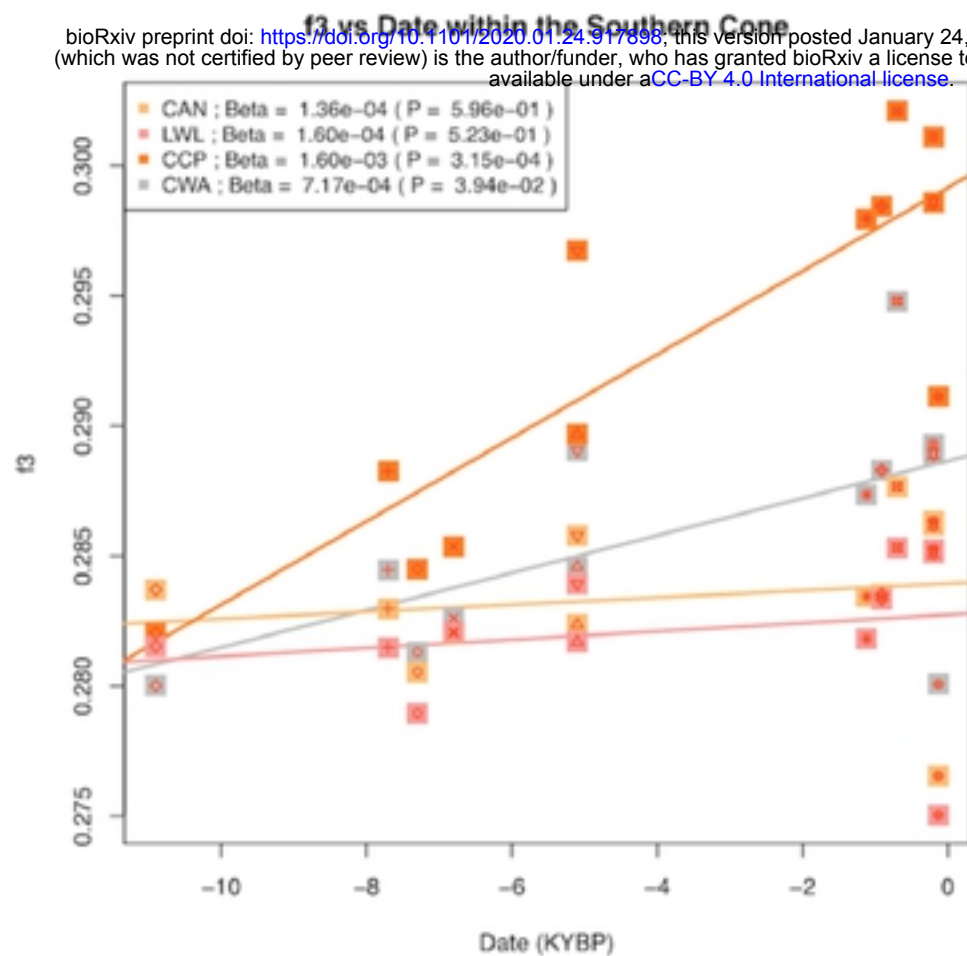Fig 7. Changes across time of genetic affinity of the four Native A

A.

Native American individuals

Reich et al. 2012

- Inga (Andean)
- × Arhuaco (Chibchan-Paezan)
- Embera (Chibchan-Paezan)
- Kogi (Chibchan-Paezan)
- + Waunana (Chibchan-Paezan)
- Chane (Equatorial-Tucanoan)
- ▽ Guahibo (Equatorial-Tucanoan)
- Guarani (Equatorial-Tucanoan)
- Guarani (Equatorial-Tucanoan)
- Palikur (Equatorial-Tucanoan)
- Piapoco (Equatorial-Tucanoan)
- Ticuna (Equatorial-Tucanoan)
- Toba (Equatorial-Tucanoan)
- Wayuu (Equatorial-Tucanoan)
- Kaingang (Ge-Pano-Carib)

- Wichi (Ge-Pano-Carib)
- Aymara (Andean)
- Aymara (Andean)
- + Diaguita (Andean)
- + Quechua (Andean)
- + Huilliche (Central Chile)
- ▽ Chilote (Chilean Patagonia)
- × Chono (Chilean Patagonia)
- Yamana (Chilean Patagonia)
- Yamana (Chilean Patagonia)

de la Fuente et al. 2018

- △ Huilliche (Central Chile)
- △ Pehuenche (Central Chile)
- Kaweskar (Chilean Patagonia)
- Yamana (Chilean Patagonia)

Admixed individuals

Present Study

- Calingasta (San Juan; CYA)
- Santa Rosa de Conlara (San Luis; CYA)
- Tilisarao (San Luis; CYA)
- Corrientes (Corrientes; NEA)
- Formosa (Formosa; NEA)
- 25 de Mayo (Misiones; NEA)
- Colonia Aurora (Misiones; NEA)
- San Vicente (Misiones; NEA)
- Catamarca (Catamarca; NWA)
- San Salvador de Jujuy (Jujuy; NWA)
- Chepes (La Rioja; NWA)
- Salta (Salta; NWA)
- San Miguel de Tucuman (Tucuman; NWA)
- AMBA (Buenos Aires; PPA)
- Bahia Blanca (Buenos Aires; PPA)
- La Plata (Buenos Aires; PPA)
- Cordoba (Cordoba; PPA)

- Rio Cuarto (Cordoba; PPA)
- La Paz (Entre Rios; PPA)
- Puerto Madryn (Chubut; PTA)
- Trelew (Chubut; PTA)
- Bariloche (Rio Negro; PTA)
- RioGrande (Tierra del Fuego; PTA)

Homburger et al. 2015

- ▽ Santiago (Chile)
- △ Lima (Peru)
- ▲ Mendoza (Mendoza; CYA)
- ▲ Buenos Aires (Buenos Aires; PPA)
- ▽ Mar del Plata (Buenos Aires; PPA)
- Cordoba (Cordoba; PPA)
- Parana (Entre Rios; PPA)
- Rosario (Santa Fe; PPA)
- ▽ Santa Fe (Santa Fe; PPA)

Ancient Samples

Raghavan et al. 2015

- Kaweskar ~100BP (Southern Cone)
- Selknam ~200BP (Southern Cone)
- Yaghan ~200BP (Southern Cone)

Posth et al. 2018

- Mayahak Cab Pek ~9300BP (Belize)
- Saki Tzul ~7400BP (Belize)
- Lapa Do Santo ~9600BP (Brazil)
- Laranjal ~6700BP (Brazil)
- Cuncaicha ~9000BP (Early Andes)
- LaGalgada ~4100BP (Early Andes)
- Lauricocha ~3500BP (Early Andes)
- Lauricocha ~5800BP (Early Andes)
- Lauricocha ~8600BP (Early Andes)
- Cuncaicha ~3300BP (Late Andes)
- Cuncaicha ~4200BP (Late Andes)
- Laramate ~900BP (Late Andes)
- ▽ Pica Ocho ~700BP (Late Andes)

- Arroyo Seco 2 ~7700BP (Southern Cone)
- Conchali ~700BP (Southern Cone)
- × Laguna Chica ~6800BP (Southern Cone)
- Los Rieles ~10900BP (Southern Cone)
- ▽ Los Rieles ~5100BP (Southern Cone)

Moreno-Mayar, Vinner et al. 2018

- Sumidouro ~10100BP (Brazil)
- △ Aconcagua Inca ~500BP (Late Andes)
- △ Ayayema ~5100BP (Southern Cone)
- Punta Santa Ana ~7300BP (Southern Cone)

Lindo et al. 2018

- Rio Uncallane ~1800BP (Late Andes)

de la Fuente et al. 2018

- Kaweskar ~1100BP (Southern Cone)
- Yaghan ~900BP (Southern Cone)

B.

MDS for 4-first PCs based weighted euclidean distance

2nd Dimension

1st Dimension

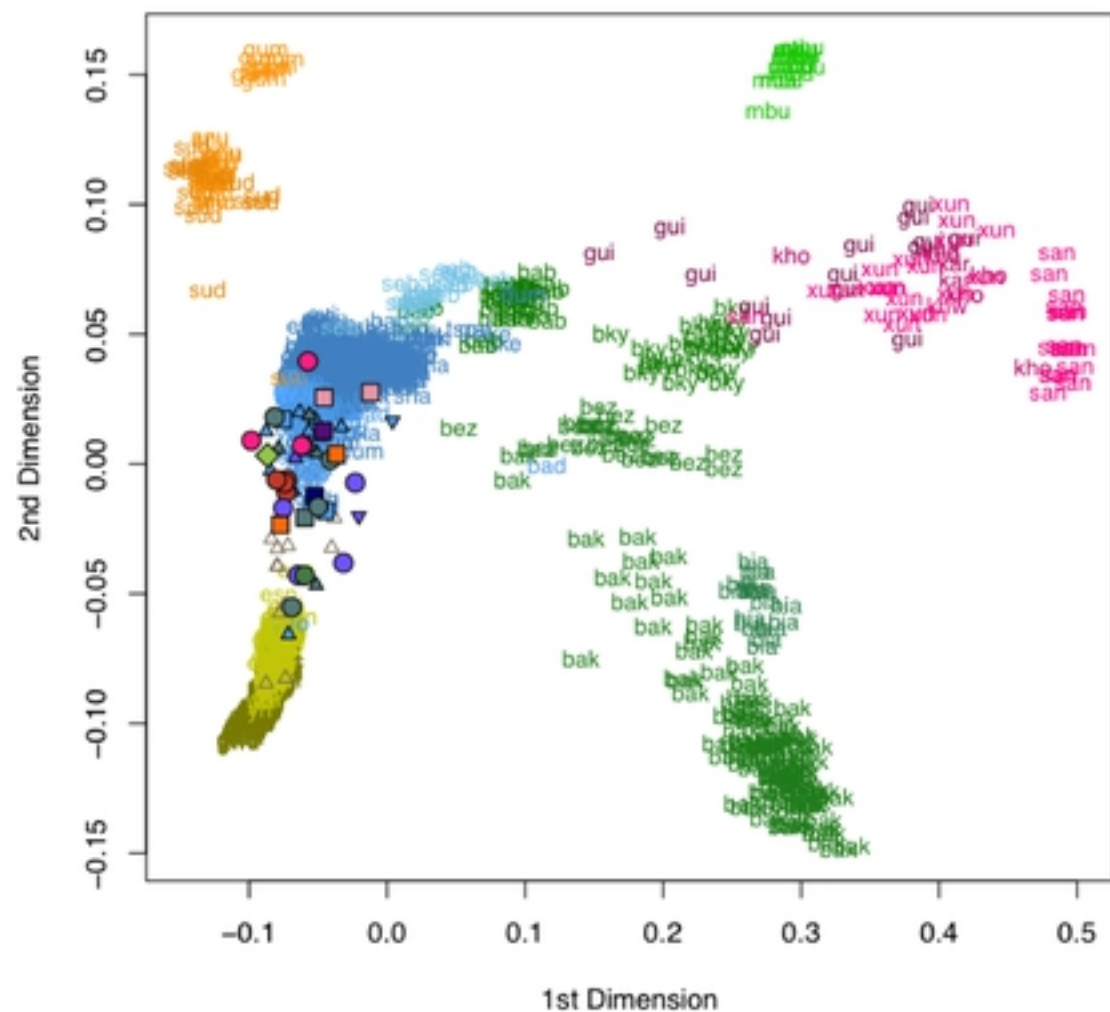Fig 4. Native American Ancestry-Specific Principal Component An
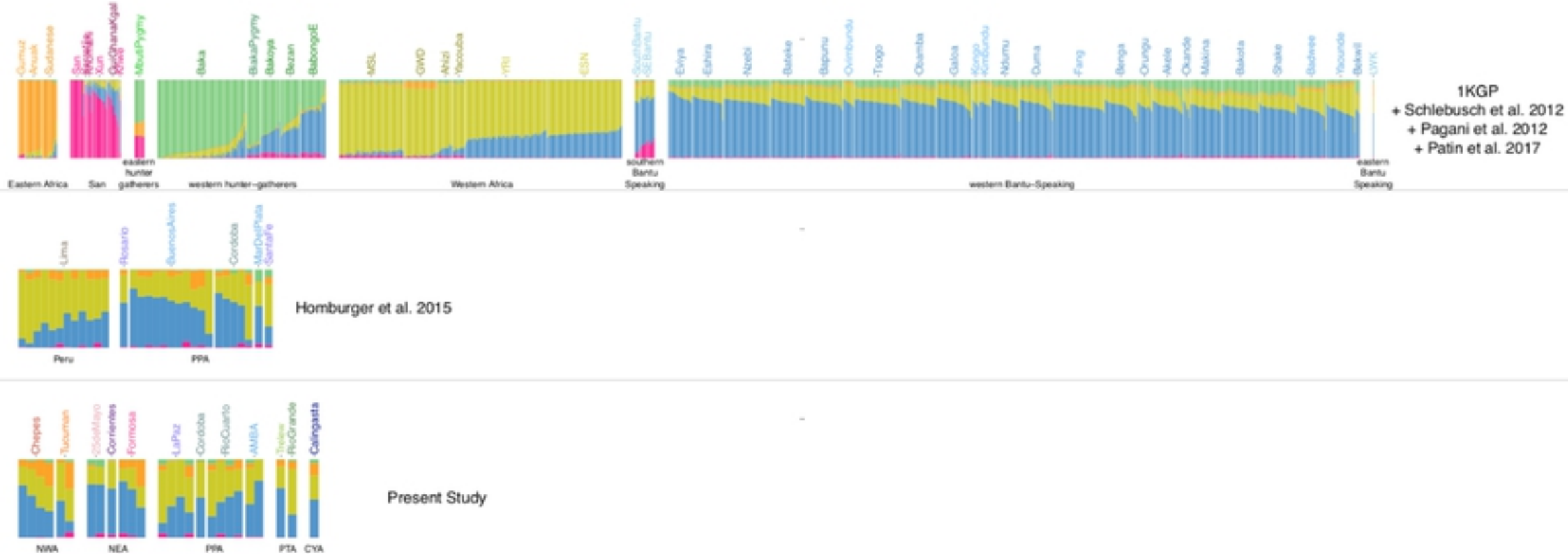
Fig 2. African Ancestry-Specific Principal Component Analysis.

Fig 3. African Ancestry-Specific Admixture Analysis.