

1 **Cryptic Native American ancestry recapitulates population-specific migration and settlement of the**  
2 **continental United States**

3 I. King Jordan<sup>1,2,3</sup>, Lavanya Rishishwar<sup>1,2,3</sup>, and Andrew B. Conley<sup>1,2,3</sup>

4

5 <sup>1</sup>School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia USA

6 <sup>2</sup>IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, Georgia, USA

7 <sup>3</sup>PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia

8

9 \*Correspondence

10 Andrew B. Conley

11 IHRC-Georgia Tech Applied Bioinformatics Laboratory

12 Atlanta, Georgia, USA

13 678-613-7921

14 [andrew.conley@biosci.gatech.edu](mailto:andrew.conley@biosci.gatech.edu)

15

16

17

18

19

20

## 21 **Abstract**

22 European and African descendants settled the continental US during the 17<sup>th</sup>-19<sup>th</sup> centuries, coming into contact  
23 with established Native American populations. The resulting admixture among these groups yielded a significant  
24 reservoir of cryptic Native American ancestry in the modern US population. We analyzed the patterns of Native  
25 American admixture seen for the three largest genetic ancestry groups in the US population: African American,  
26 European American, and Hispanic/Latino. The three groups show distinct Native American ancestry profiles,  
27 which are indicative of their historical patterns of migration and settlement across the country. Native American  
28 ancestry in the modern African American population does not coincide with local geography, instead forming a  
29 monophyletic group with origins in the southeastern US, consistent with the Great Migration of the early 20<sup>th</sup>  
30 century. European Americans show Native American ancestry that tracks their geographic origins across the US,  
31 indicative of ongoing contact during westward expansion, and Native American ancestry can resolve  
32 Hispanic/Latino individuals into distinct local groups formed by more recent migration from Mexico and Puerto  
33 Rico. We found an anomalous pattern of Native American ancestry from the US southwest, which most likely  
34 corresponds to the *Nuevomexicano* descendants of early Spanish settlers to the region. We addressed a number  
35 of controversies surrounding this population, including the extent of Sephardic Jewish ancestry. *Nuevomexicanos*  
36 are less admixed than nearby Mexican-American individuals, with more European and less Native American and  
37 African ancestry, and while they do show demonstrable Sephardic Jewish ancestry, the fraction is no greater than  
38 seen for other Hispanic/Latino populations.

39

## 40 Introduction

41 Native Americans inhabited the area that now makes up the continental US for thousands of years prior to the  
42 arrival of the first European settlers. The ancestors of modern Native Americans are thought have arrived in the  
43 Americas from Asia, by way of the Bering Strait, in several successive waves of migration<sup>1</sup>. The current model,  
44 based on archaeology and comparative genomic studies, holds that the earliest ancestors of Native Americans  
45 arrived in the Americas ~23,000 years ago<sup>2</sup>. The earliest evidence for Native Americans in the continental US dates  
46 to ~14,000 years ago<sup>3</sup>. The much later arrival of Europeans in the Americas, followed shortly thereafter by Africans  
47 who were brought by force via the trans-Atlantic slave trade, had a drastic effect on the demographic makeup of  
48 the region. Native American population numbers declined rapidly in the face of continuous immigration,  
49 settlement, and conflict, and as a result the modern US population is made up mainly of descendants of European  
50 and African immigrants.

51 Europeans arrived in the Americas more than 20,000 years after the first Native Americans. The first European  
52 settlers to reach the continental US were Spaniards led by the conquistador Ponce de León, who claimed Florida  
53 for the Spanish crown in 1513<sup>4</sup>. British settlers arrived more than 70 years later, initially establishing the ill-fated  
54 colony of Roanoke in 1585 and later the permanent settlement of Jamestown in 1607<sup>5</sup>. An estimated 400,000  
55 British had migrated to the US by the end of the 17<sup>th</sup> century. The first Africans were brought to Jamestown in  
56 1619 by Dutch pirates who traded them to the British settlers as indentured servants<sup>6</sup>. The social status of Africans  
57 in the US changed quickly, with slavery first legally sanctioned by 1640. The trans-Atlantic slave trade would  
58 eventually bring ~400,000 enslaved Africans to the continental US<sup>7</sup>.

59 The arrival of Europeans and Africans in the Americas, and the conflict that followed, would prove to be  
60 catastrophic for the indigenous population. It has been estimated that 10-100 million Native Americans may have  
61 died in the first 150 years after Columbus' arrival in the New World, amounting to a 95% reduction in the  
62 population<sup>8</sup>. This massive Native American population decline is mainly attributed to the introduction of European  
63 and African endemic infectious diseases – *e.g.* malaria, measles, and smallpox – for which the indigenous  
64 population had little or no immune defense.

65 The story of conflict between Native Americans and European and African settlers, along with the devastating  
66 consequences for the indigenous population, is by now well-known. However, there is another, perhaps less  
67 appreciated, aspect of the encounter between these population groups that has also had profound consequences  
68 for the genetic demography of the Americas. Here, we are referring to the process of genetic admixture, whereby  
69 individuals from previously isolated population groups reproduce, resulting in the combination of ancestry-

70 specific haplotypes within individual genomes. Admixture has been a fundamental feature of human evolution  
71 and migration<sup>9</sup>. Whenever previously isolated human populations meet, no matter what the circumstances, they  
72 mix and give rise to individuals with a mosaic of different genetic ancestries.

73 As European and African descendants settled the continental US, they inevitably came into contact with  
74 established Native American populations resulting in admixture and the introgression of Native American genomic  
75 sequence into the expanding US population. Accordingly, the genomes of European and African descendants in  
76 the US are expected to contain some fraction of Native American ancestry. In other words, a significant reservoir  
77 of Native American ancestry currently exists outside of recognized indigenous communities. We refer to this  
78 ancestry component as ‘cryptic’ Native American ancestry given the fact that its low levels may often lead it to go  
79 unrecognized. In this study, we ask how the historical processes of migration and settlement affected the  
80 distribution of cryptic Native American ancestry across the continental US. We address this question for the three  
81 largest genetic ancestry groups in the modern US population: African American, European American, and  
82 Hispanic/Latino.

83

## 84 **Material and Methods.**

85 **Genotype datasets.** Whole genome genotype data of US individuals from the Health and Retirement Study (HRS)  
86 dataset ( $n=15,620$ ) were merged with whole genome sequence variant data from the 1000 Genomes Project  
87 (1KGP)<sup>10, 11</sup> ( $n=1,718$ ) and whole genome genotype data from the Human Genome Diversity Project (HGDP)<sup>12-14</sup>  
88 ( $n=230$ ) (Table S1). HRS genotype data were accessed via the NCBI dbGaP database and the study was conducted  
89 with Institutional Review Board approval from the Georgia Institute of Technology (protocol number H17029).  
90 Individual HRS genotypes are provided along with geographical origin data for sample donors from the nine census  
91 regions in the continental US. A collection of Native American genotypes from 21 populations across the Americas  
92 was taken from a comprehensive study on Native American population history<sup>2</sup> ( $n=314$ ). These Native American  
93 genotype data were accessed according to the terms of a data use agreement from the Universidad de Antioquia.  
94 Whole genome genotype data from 5 populations of Sephardic Jewish individuals ( $n=40$ ) were also included as  
95 reference populations<sup>15</sup>. The genotypes from HRS individuals were merged with the comparative genomic data  
96 sources using PLINK version 1.9<sup>16</sup>, keeping only those sites common to all datasets and correcting SNP strand  
97 orientations for consistency as needed. The final merged dataset includes 228,190 SNPs across 17,882 individuals.  
98 The merged genotype dataset was phased using ShapeIT version 2.r837<sup>17</sup>. SNPs that interfered with the ShapeIT  
99 phasing process were excluded from subsequent analyses. ShapeIT was run without reference haplotypes, and

100 all individuals were phased at the same time. Individual chromosomes were phased separately, and the X  
101 chromosome was phased with the additional '-X' flag.

102 **Local ancestry inference.** The RFMix algorithm<sup>18</sup> is able to accurately characterize the local ancestry of admixed  
103 individuals but is prohibitively slow when run on a dataset of the size used here. To reduce the runtime, we  
104 modified RFMix version 1.5.4 so that the expectation-maximization (EM) procedure samples from, and creates a  
105 forest for, the entire set of individuals rather than each individual. This modified RFMix was run in the PopPhased  
106 mode with a minimum node size of five, using 12 generations and the "--use-reference-panels-in-EM" for two  
107 rounds of EM, generating local ancestry inference for both the reference and admixed populations. Continental  
108 African, European, and Native American populations were used as reference populations. Contiguous regions of  
109 ancestral assignment, "ancestry tracts," were created where RFMix ancestral certainty was at least 95%. Genome-  
110 wide ancestry estimates from the modified RFMix algorithm closely correlate with those from ADMIXTURE (Figure  
111 S1).

112 The extent of Sephardic Jewish (*Converso*) ancestry in individuals from the Hispanic/Latino group in HRS (as  
113 defined in the genome-wide ancestry section below), and Latin American populations from 1KGP, was inferred via  
114 ancestry-specific haplotype comparisons with Sephardic Jewish reference populations using the program  
115 ChromoPainter<sup>29</sup> (kindly provided by Garrett Hellenthal). First, African and Native American haplotypes were  
116 masked from the RFMix output. Then, the remaining European haplotypes were compared against genomes from  
117 the European reference populations together with the Sephardic Jewish populations. The extent of Jewish  
118 ancestry for any individual genome is defined as the 'copying fraction' from the Sephardic Jewish populations,  
119 where the copying fraction is taken as the fraction of sites with best matches to the Sephardic Jewish reference  
120 genomes. It should be noted that this procedure results in a relative fraction of Sephardic Jewish ancestry for all  
121 individuals under consideration, which is directly comparable among individuals but likely to be an overestimate  
122 of the total ancestry derived from a single source population.

123 **Genome-wide ancestry inference.** ADMIXTURE<sup>19</sup> version 1.3.0 was used with  $K=4$  to infer continental ancestry  
124 fractions for individuals in the dataset via comparison with reference populations from Africa, Europe, the  
125 Americas, and East Asia. Sub-continental ancestry was inferred independently for each of the three major  
126 continental ancestry components – African, European, and Native American – using an ancestry-specific masking  
127 procedure that we developed as previously described<sup>20</sup>. This procedure relies on the local continental ancestry  
128 assignments, along with the re-phased genotypes, generated by RFMix as described above. Sub-continental  
129 ancestry was characterized by first masking out two of the three continental ancestries (African, European, and/or  
130 Native American) at a time and then analyzing the genomic regions (haplotypes) corresponding to the remaining

131 continental ancestry. For sub-continental ancestry analysis of any given continental ancestry component, only  
132 those individuals with at least 1.5% genome-wide ancestry for that same continental group were used.

133 We developed a novel machine learning based approach to distinguish Spanish from other (primarily Western)  
134 European descendants in the HRS dataset via analysis of European-specific haplotypes. First, ADMIXTURE was run  
135 with  $K=5$  on the RFMix characterized European haplotypes for the HRS individuals to stratify sub-continental  
136 European ancestries based on comparison with Northern (Finnish and Russian), Western (French and British),  
137 Spanish, and Southern (Italian and Sardinian) European reference populations from the 1KGP and HGDP datasets.  
138 A Support Vector Machine (SVM) classifier<sup>21</sup> was then trained using the resulting ADMIXTURE ancestry vectors for  
139 the European reference populations from the four sub-continental groups: Northern, Western, Spanish, and  
140 Southern. The European-specific ADMIXTURE ancestry vectors for the HRS individuals were then classified into  
141 one of the four European sub-continental groups defined by the SVM classifier. A confidence threshold of 0.8 was  
142 used for sub-continental group assignments. For the purpose of analysis here, we consider two major groups of  
143 European descendants in the HRS data set: Spanish descendants and all others. We refer to Spanish descendants  
144 as Hispanic/Latino (HL). Non-Spanish HRS individuals with <5% African ancestry are defined as European  
145 American, whereas non-Spanish HRS individuals with at least 20% African ancestry were defined as African  
146 American (see Supplementary Methods for additional details).

147 **Sex-biased ancestry inference.** Sex-biased ancestry contributions were inferred by comparing the RFMix  
148 characterized fractions of each continental ancestry component on the X chromosomes versus the autosomes as  
149 previously described<sup>22; 23</sup>. For each individual genome, and each ancestry component, the normalized difference  
150 between the X chromosome ancestry fraction and the autosomal ancestry fraction ( $\Delta Admix$ ) is defined as:

$$151 \quad \Delta Admix = F_{anc,total} \times (F_{anc,X} - F_{anc,auto}) / (F_{anc,X} + F_{anc,auto})$$

152 where  $F_{anc,total}$ ,  $F_{anc,X}$ , and  $F_{anc,auto}$  are the genome-wide, X chromosome, and autosome ancestry fractions,  
153 respectively.

154

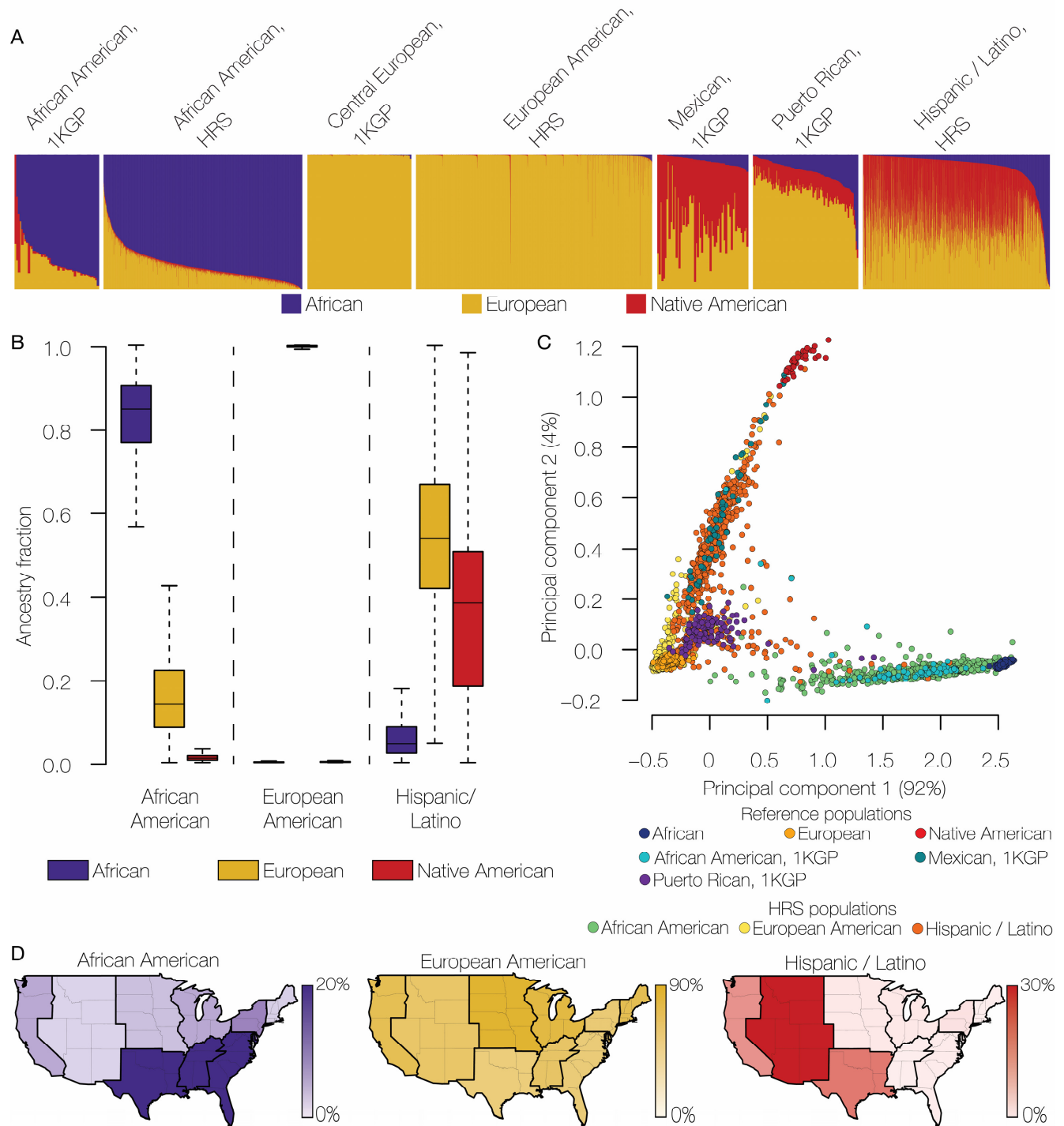
155 **Phylogenetic inference.** We used the RFMix defined Native American haplotypes for individuals from the HRS  
156 and reference populations to infer the phylogenetic relationships between populations. For each pair of  
157 populations with at least 25 Native American sites covered by 5 individuals, we computed the weighted  $F_{ST}$   
158 between all pairs of populations using PLINK. The resulting  $F_{ST}$  distance matrix was used to create a neighbor-  
159 joining tree<sup>24</sup> with the program MEGA6<sup>25</sup>.

## 160 **Results**

161 **Genetic ancestry groups in the US.** The first aim of our study was to objectively define the major genetic ancestry  
162 groups for the continental US based on observable patterns of ancestry and admixture seen for the 15,620 HRS  
163 genotypes analyzed here. Having defined the US genetic ancestry groups, we then considered the distribution of  
164 Native American ancestry within and between ancestry groups and among geographic regions. We provide a  
165 detailed description, along with supporting results (Supplementary Figures S2-S5), of how we defined the three  
166 main US ancestry groups – African American, European American, and Hispanic/Latino – in the Supplementary  
167 Material.

168 The distribution of HRS individuals among the three major US genetic ancestry groups is shown in Figure 1. Visual  
169 inspection of the continental ancestry fractions seen for members of the three groups supports our objective  
170 approach to genetic ancestry-based classification (Figure 1A). For example, the majority of Hispanic/Latino  
171 individuals show substantial levels of Native American ancestry compared to individuals from the European  
172 American ancestry group (Figure 1A); the median Native American ancestry for the Hispanic/Latino group is 38%  
173 compared to 0.1% for the European American group (Figure 1B). In addition, individuals from the Hispanic/Latino  
174 group cluster tightly with the Mexican reference population from the 1KGP, along the second axis between the  
175 European and Native American populations in the principal components analysis (PCA) plot of the pairwise  
176 genome distances (Figure 1C). It is important to note that we did not use Native American ancestry for the  
177 purposes of classification. Rather, European ancestry alone was sufficient to recapitulate known levels of Native  
178 American ancestry for Hispanic/Latino individuals.

179



180

181 **Figure 1. Genetic ancestry groups in the modern US population.** (A) ADMIXTURE plot ( $K=3$ ) showing the African  
 182 (blue), European (gold), and Native American (red) ancestry components for individuals from different US  
 183 population groups. Data are from the 1000 Genomes Project (1KGP) and the Health and Retirement Study (HRS).  
 184 (B) Distributions of African, European, and Native American ancestry fractions for the three main US genetic  
 185 ancestry groups defined here: African American, European American, and Hispanic/Latino. (C) Principal  
 186 components analysis (PCA) plot showing the relationships among individuals from reference populations and



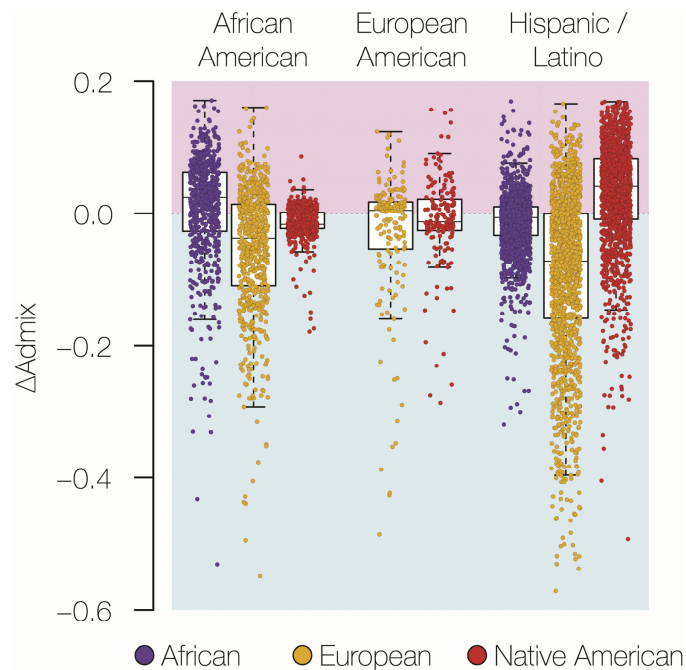
187 individuals from the HRS dataset corresponding to the three US genetic ancestry groups. (D) Percentages of  
188 individuals from each of the three US genetic ancestry groups are shown for the nine census regions in continental  
189 US.

190

191 Individuals from the African American ancestry group show medians of 85% African ancestry, 14% European  
192 ancestry, and 1% Native American ancestry (Figure 1B). Most of these individuals group along the first PCA axis  
193 separating the African and European reference populations. In contrast to the admixed Hispanic/Latino and  
194 African American ancestry groups, individuals from the European American ancestry group show extremely low  
195 levels of admixture with non-European populations, with a median value of 99.8% European ancestry. Given their  
196 relatively low numbers (see Supplementary Figure S2), as well as their relatively late historical arrival in the  
197 continental US, we did not consider Asian Americans further in this study.

198 Individuals assigned to the three main genetic ancestry groups show distinct geographic distributions across the  
199 continental US, which are largely consistent with demographic data for the country. African ancestry is highest in  
200 the three southern census regions, European ancestry is highest in the two north central regions, and  
201 Hispanic/Latino ancestry is highest in the Mountain census region, which includes Arizona and New Mexico (Figure  
202 1D).

203 **Sex-biased admixture in US genetic ancestry groups.** We compared the patterns and extent of sex-biased  
204 admixture among the three US genetic ancestry groups by comparing the continental ancestry fractions – African,  
205 European, and Native American – seen for the X chromosomes versus the autosomes. For any given ancestry  
206 component, a relative excess of X chromosome ancestry is indicative of female-biased admixture, whereas an  
207 excess of autosomal ancestry reflects male-biased admixture<sup>26</sup>. This was only done for admixed individuals that  
208 had two or more continental ancestry fractions at >1.5% of the overall ancestry. Almost all individuals from the  
209 African American and Hispanic/Latino groups met this criterion, but only a small minority of European American  
210 individuals with Native American admixture did. African American and Hispanic/Latino ancestry groups showed  
211 marked patterns of sex-biased admixture, whereas the European Americans did not show any appreciable  
212 evidence of sex-biased admixture (Figure 2). The strongest pattern of sex-biased admixture was seen for  
213 Hispanic/Latino individuals, with female-biased Native American admixture and male-biased European admixture.  
214 African Americans show female-biased African ancestry and male-biased European ancestry.



215

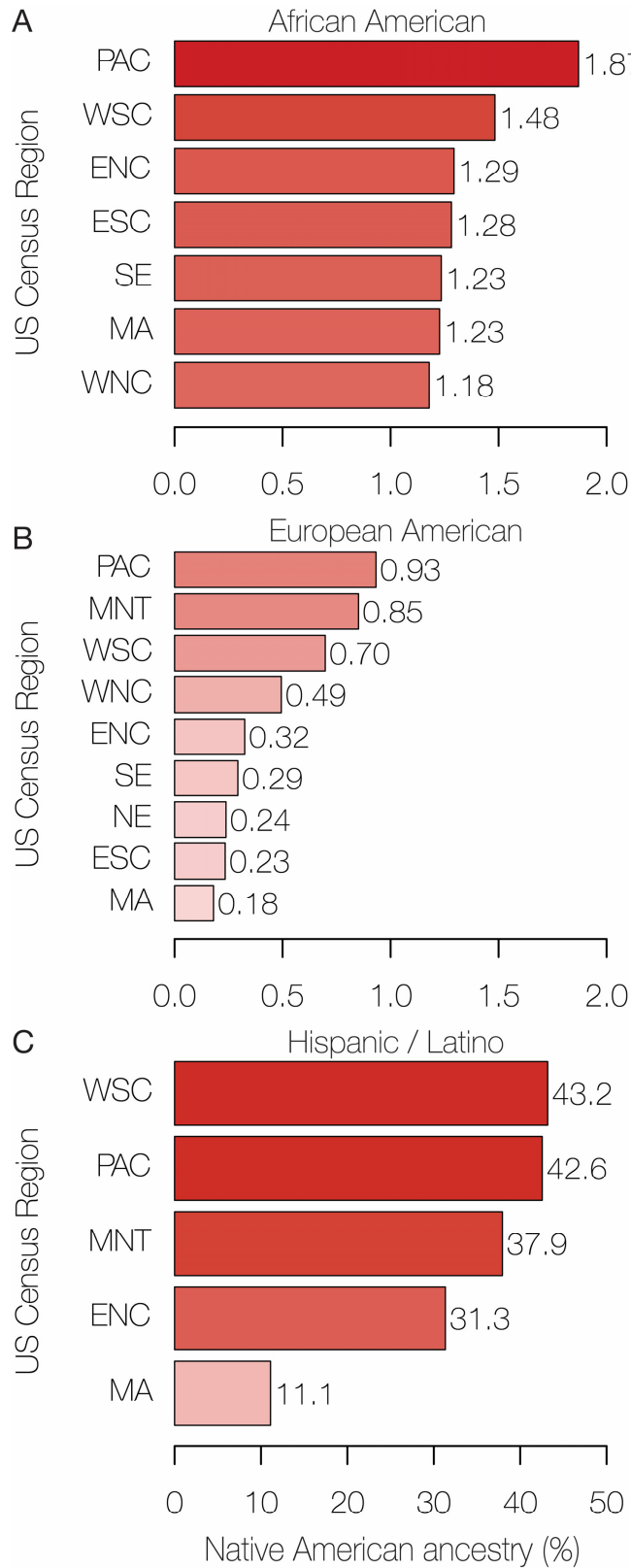
216 **Figure 2. Sex-biased admixture in US genetic ancestry groups.** Normalized differences between X chromosome  
217 ancestry fractions and autosomal ancestry fractions ( $\Delta Admix$ ) are shown on the y-axis.  $\Delta Admix$  values are  
218 shown for each ancestry component – African (blue), European (gold), and Native American (red) – in each  
219 individual genome.  $\Delta Admix$  values above zero (pink) indicate female-biased admixture, and values below zero  
220 (blue) indicate male-biased admixture.

221

222 **Native American ancestry distribution across the US.** For each US genetic ancestry group, we considered three  
223 distinct characteristics of Native American ancestry across the continental US: (1) the relative levels of Native  
224 American ancestry genome-wide, (2) the patterns of Native American allele frequencies, and (3) the phylogenetic  
225 relationships among US populations based on their Native American ancestry.

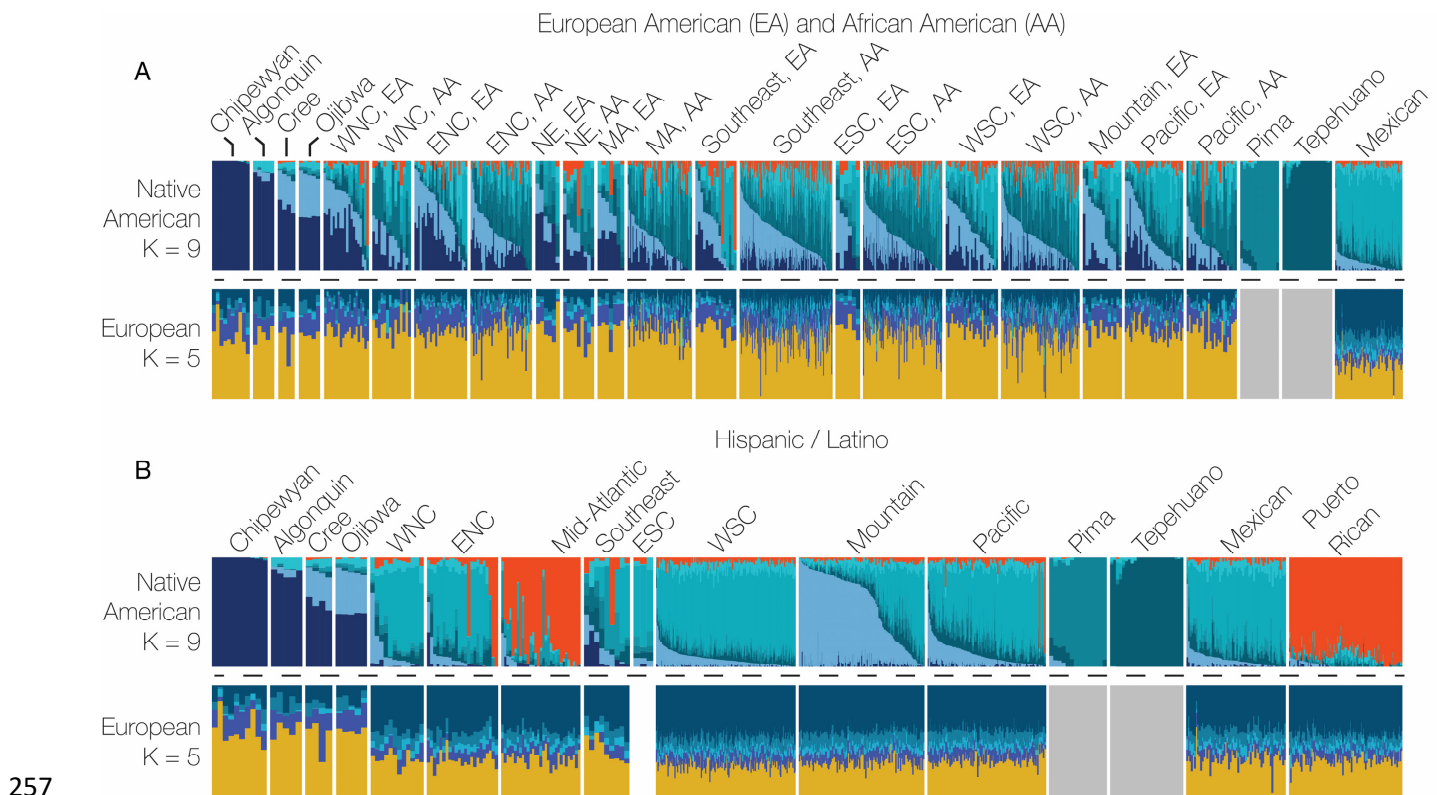
226 As we showed previously, overall Native American ancestry is highest for the Hispanic/Latino group (median 38%),  
227 followed by the African American (1%) and European American groups (0.1%) (Figure 1B). Among all three  
228 ancestry groups, the highest levels of Native American ancestry are seen for the West-South-Central (WSC;  
229 including Texas), Pacific (PAC; including California), and Mountain (MNT; including Arizona and New Mexico)  
230 census regions (Figure 3). Native American ancestry levels show the highest variability among regions for the  
231 Hispanic/Latino group (coefficient of variation [c.v.]=1.08), followed by the European American (c.v.=0.65) and  
232 then African American (c.v.=0.60) groups.

233



**Figure 3. Native American ancestry percentages in the modern US population.** The average percentages of Native American ancestry are shown for the three US genetic ancestry groups across the nine geographic census regions (see Supplementary Figure S3). Data for census regions with less than five individuals for any ancestry group are considered unreliable and are not shown.

244 We measured the patterns of Native American allele frequencies across the continental US using ADMIXTURE  
 245 analysis of Native American haplotypes for individuals from the three ancestry groups. Visualization of the  
 246 ancestry vectors produced by ADMIXTURE shows that the African American and European American groups have  
 247 patterns that are similar to each other (Figure 4A; top panel) and distinct from the patterns seen for the  
 248 Hispanic/Latino group (Figure 4B; top panel). Furthermore, the African American and European American groups  
 249 show ancestry patterns that are intermediate to the Canadian (Chipewyan, Algonquin, Cree, and Ojibwa) and  
 250 Northern Mexican (Pima and Tepehuano) Native American reference populations, whereas the Hispanic/Latino  
 251 group shows Native American ancestry patterns that are more similar to either the admixed Mexican and Native  
 252 American reference populations or the admixed Puerto Rican population. There is substantially more regional  
 253 variation in Native American ancestry seen for the Hispanic/Latino group, with characteristically Mexican patterns  
 254 seen in the Pacific (PAC) and West South-Central (WSC) regions and a strongly Puerto Rican pattern in the Mid-  
 255 Atlantic (MA) region. The Mountain region (MNT) shows a distinct and highly variable pattern of Native American  
 256 ancestry, which we explore in more detail in the following section.

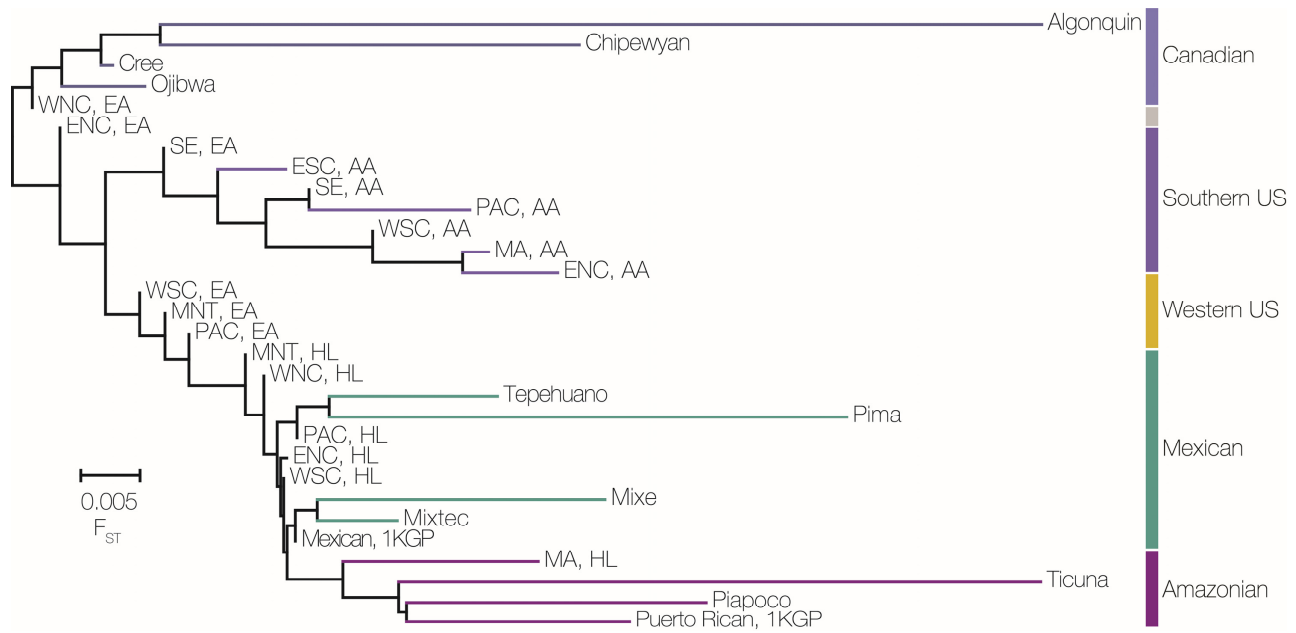


257  
 258 **Figure 4. Native American and European ancestry profiles for US ancestry groups.** Native American ( $K=9$ ) and  
 259 European ( $K=5$ ) ancestry-specific ADMIXTURE plots are shown for the European American (EA) and African  
 260 American (AA) groups combined (A) and for the Hispanic/Latino group (B). The individual panels shown  
 261 correspond to Native American reference populations (Chipewyan, Algonquin, Cree, Ojibwa, Pima, Tepehuano),

262 1000 Genomes Project reference populations (Mexican and Puerto Rican) and the HRS data from the different US  
263 census regions (see Supplementary Figure S3).

264

265 The phylogenetic relationships among genetic ancestry groups across the US were inferred by calculating the  
266 fixation index ( $F_{ST}$ ) between pairs of populations based on Native American haplotypes (Figure 5). The Canadian  
267 and Amazonian Native American reference populations occupy the most distant clades on the phylogeny with the  
268 admixed Mexican and Mexican Native American reference populations adjacent to the Amazonian group. African  
269 Americans from all of the census regions form a single monophyletic clade, with the European Americans from  
270 the Southeast region (SE) as the closest sister taxon. European Americans from the West North-Central (WNC)  
271 and East North-Central (ENC) regions group most closely with the Canadian Native American reference  
272 populations, and the European Americans from this region form a distinct group adjacent to a Mexican group of  
273 populations. Members of the Hispanic/Latino ancestry group from most of the census regions group closely with  
274 Mexican populations, with the exception of the Mid-Atlantic region (MA) which groups most closely with the  
275 admixed Puerto Rican and Amazonian reference populations.

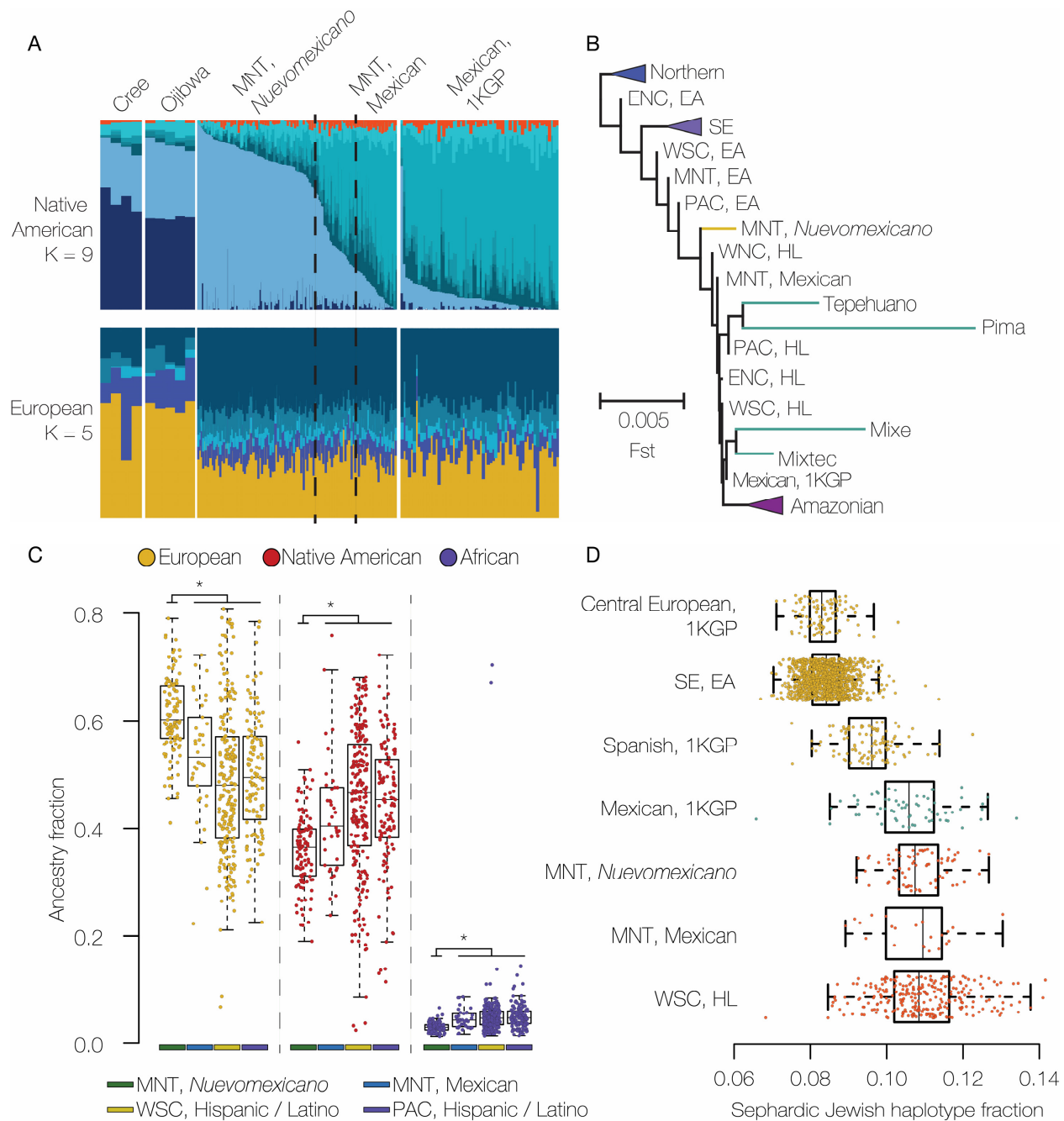


276

277 **Figure 5. Native American ancestry phylogeny.** Phylogenetic relationships are shown for the Native American  
278 ancestry-specific components of Native American reference populations (Algonquin, Chipewyan, Cree, Ojibwa,  
279 Tepehuano, Pima, Mixe, Mixtec, Ticuna, Piapoco), 1000 Genomes Project reference populations (Mexican and  
280 Puerto Rican) and HRS groups. The HRS groups are labeled according to their US census region origins (see  
281 Supplementary Figure S3) and genetic ancestry group: African American (AA), European American (EA), and  
282 Hispanic/Latino (HL). Broad geographic and genetic groupings are indicated by the bars on the right side. The  
283 scale bar corresponds to the pairwise  $F_{ST}$  values used to generate the phylogeny.

284

285 **Native American ancestry of the *Nuevomexicanos*.** The ADMIXTURE results for the Hispanic/Latino group in the  
286 Mountain region (MNT) point to the presence of two distinct sub-populations, one of which is clearly Mexican in  
287 origin, whereas the second group has a very distinct pattern compared to any other Hispanic/Latino group  
288 analyzed here (Figure 4B and Figure 6A). If these two apparent Mountain Hispanic/Latino sub-populations are  
289 considered separately, they form distinct phylogenetic groups (Figure 6B). One group clearly falls into the clade  
290 with the other Mexican origin populations (see MNT, Mexican), whereas the distinct group is basal to the Mexican  
291 clade and intermediate between the Western US and Mexican clades (see MNT, *Nuevomexicano*). The results of  
292 the ADMIXTURE and phylogenetic analyses are consistent with historical records indicating the presence of a  
293 unique group of Spanish descendants in the American Southwest, known as the ‘Hispanos of New Mexico’ or  
294 *Nuevomexicanos*. This population is descended from very early Spanish settlers to the Four Corners region of the  
295 US, primarily New Mexico and southern Colorado, and distinct from Mexican-American immigrants who arrived  
296 later<sup>27</sup>.



297

298 **Figure 6. Genetic ancestry of the Nuevomexicanos.** (A) Native American ( $K=9$ ) and European ( $K=5$ ) ancestry-  
 299 specific ADMIXTURE plots comparing the Mountain census region (MNT) in the middle panel to Native American  
 300 Cree, Ojibwa (left) and admixed Mexican (right) reference populations. Native American ancestry profiles for the  
 301 Mountain region can be divided into *Nuevomexicano* (left) and Mexican-American (right) components. (B) Native  
 302 American ancestry phylogeny (as shown in Figure 5) with the Mountain census region (MNT) broken down into  
 303 *Nuevomexicano* and Mexican-American components. (C) Distributions of European, Native American, and African  
 304 ancestry fractions are shown for the Mountain (MNT) *Nuevomexicano*, Mountain (MNT) Mexican, West South  
 305 Central (WSC) Hispanic/Latino, and Pacific (PAC) Hispanic/Latino groups. The \* indicates significant differences in



306 median ancestry fractions between the *Nuevomexicano* and other groups ( $P < 0.01$  Wilcoxon Rank-Sum test). (D)  
307 Distributions of the Sephardic Jewish haplotype copying fractions are shown for European reference populations  
308 from the 1000 Genomes Project (Central European and Spanish), HRS European Americans from the Southeast  
309 census region (SE, EA), Mountain (MNT) *Nuevomexicano*, Mountain (MNT) Mexican, and West South Central  
310 Hispanic/Latino (WSC, HL) groups.

311

312 Members of the *Nuevomexicano* population have maintained a distinct cultural identity for centuries, and the  
313 ability to isolate individuals from this group based on analysis of their genotypes allowed us to address open  
314 questions related to their ancestry. In addition to characterizing their distinct pattern of Native American  
315 ancestry, we also compared the levels of Native American admixture between *Nuevomexicanos* and the other  
316 nearby Hispanic/Latino groups, which show a more Mexican pattern of Native American ancestry. Consistent with  
317 previous results<sup>28</sup>, we show that *Nuevomexicanos* have significantly more European ancestry and less Native  
318 American ancestry than other Hispanic/Latino groups from the Western Census regions (Figure 6C).  
319 *Nuevomexicanos* also show significantly lower levels of African ancestry compared to the other Hispanic/Latino  
320 groups.

321 *Nuevomexicano* cultural and historical traditions suggest that many of the early Spanish settlers in the region were  
322 *Conversos*, or crypto-Jewish individuals, who ostensibly converted to Catholicism in an effort to avoid religious  
323 persecution and pogroms, while secretly maintaining Jewish identity and traditions<sup>29</sup>. We interrogated this idea  
324 by comparing the extent of Sephardic Jewish admixture found among individuals with the *Nuevomexicano*  
325 ancestry pattern compared to other Hispanic/Latino populations. Sephardic Jewish admixture was measured by  
326 comparing European haplotypes from Hispanic/Latino individuals to a reference panel including both European  
327 and Sephardic Jewish populations. *Nuevomexicanos* show elevated levels of matching to Jewish haplotypes  
328 compared to Spanish and other European populations, consistent with substantial *Converso* ancestry among New  
329 World Hispanic/Latino populations<sup>30</sup> (Figure 6d). However, *Nuevomexicanos* do not show a higher level of  
330 *Converso* ancestry compared to the other New World Hispanic/Latino populations.

331



## 332 Discussion

333 **Native American admixture patterns for distinct US ancestry groups.** We were able to delineate three  
334 predominant genetic ancestry groups – African American, European American, and Hispanic/Latino – using  
335 comparative analysis of whole genome genotypes from >15,000 individuals from across the continental US. Each  
336 of these different groups of people experienced distinct historical trajectories in the US, which we found to be  
337 manifested as group-specific patterns of Native American ancestry.

338 Individuals from the African American ancestry group show low (Figure 1B) and relatively invariant (Figure 3A)  
339 levels of Native American ancestry across the continental US. The patterns of Native American ancestry seen for  
340 the African American group are also more constant among US census regions compared to individuals from the  
341 other two ancestry groups (Figure 4A). With respect to the Native American component of their ancestry, African  
342 Americans from all US census groups form a single monophyletic clade, for which the Southeast European  
343 American (SE, EA) group is basal (Figure 5). Considered together, these results point to a most likely scenario  
344 whereby African descendants admixed with local Native American groups in the antebellum South. Early  
345 admixture with Native Americans in the South was followed by subsequent dispersal across the US during the  
346 Great Migration in the early to mid-twentieth century<sup>31</sup>. The genetic legacy of the Great Migration has previously  
347 been explored based on overall patterns of African American genetic diversity<sup>32</sup>. Here, we were able to uncover  
348 traces of this same history based solely on the relatively low Native American ancestry component found in the  
349 genomes of African Americans.

350 The European American group shows the lowest levels of Native American ancestry for the three US ancestry  
351 groups analyzed here (Figure 1B), consistent with a large and fairly constant influx of European immigrants to the  
352 US along with social and legal prohibitions against miscegenation<sup>33</sup>. Compared to African Americans, individuals  
353 from the European American ancestry group show more variant levels of Native American ancestry among US  
354 census regions (Figure 3B) along with substantially more region-specific patterns of Native American ancestry  
355 (Figure 4A). Their region-specific patterns of Native American ancestry are also reflected in the Native American  
356 ancestry-based phylogeny, whereby the European American groups are related according to their geographic  
357 distribution across the country (Figure 5). These results point to a historical pattern of continuous, albeit  
358 infrequent, admixture between local Native American groups and European settlers as they moved westward  
359 across the continental US.

360 As can be expected, the Hispanic/Latino group shows by far the highest (Figure 1B) and most variable (Figure 3C)  
361 levels of Native American ancestry across the US. In particular, individuals from the Hispanic/Latino group show

362 highly regional-specific patterns of Native American ancestry (Figure 4B), which are consistent with known  
363 demographic trends. For example, analysis of the Native American component of Hispanic/Latino ancestry is  
364 sufficient to distinguish Puerto Rican immigrants from the Mid-Atlantic census region from Mexican Americans  
365 who predominate in the western census regions. Perhaps most striking, the patterns of Native American ancestry  
366 seen for the Mountain census regions were alone sufficient to distinguish descendants of very early Spanish  
367 settlers to the region, the group known as *Hispanos* or *Nuevomexicanos*, from subsequent waves of  
368 Hispanic/Latino immigrants who arrived later from Mexico.

369 The three main US ancestry groups are also distinguished by their patterns of sex-biased ancestry in a way that  
370 reflects the unique history of each group (Figure 2). European Americans show very little evidence for sex-biased  
371 ancestry, along with very low levels of overall admixture, compared to the African American and Hispanic/Latino  
372 groups. The strongest pattern of sex-biased ancestry was seen for the Hispanic/Latino group followed by African  
373 Americans. Sex-bias for Hispanic/Latinos is characterized by a strong female-bias for Native American ancestry  
374 coupled with European male-biased ancestry. This pattern has been observed in a number of previous studies  
375 and is consistent with the history of male-biased migration to the region dating back to the era of the  
376 conquistadors<sup>22;34</sup>. The African American group shows female-biased African ancestry and male-biased European  
377 ancestry, a pattern which has also been documented previously and tied to the legacy of slavery and racial  
378 oppression in the US<sup>35;36</sup>. It has not been previously possible to directly compare the extent of sex-biased  
379 admixture among the three largest ancestry groups in the US as we have done here. As such, it is interesting to  
380 note that the history of the Spanish colonization in Latin America had a stronger impact on sex-biased ancestry  
381 than the legacy of slavery in the US.

382

383 **Implications of genetic ancestry for the historical and cultural traditions of *Nuevomexicanos*.** Our ability to  
384 distinguish *Nuevomexicanos* from the HRS dataset, using their patterns of Native American ancestry, allowed us  
385 to address a number of open questions and controversies regarding the history and culture of this interesting  
386 population. *Nuevomexicanos* from the American southwest are historically defined as the descendants of early  
387 Spanish settlers, those who arrived in the period from 1598 to 1848, as opposed to immigrants from Mexico who  
388 arrived the region considerably later. The two distinct patterns of Native American ancestry seen for  
389 Hispanic/Latino individuals from the Mountain census region are very much consistent with this historical  
390 definition. The *Nuevomexicanos* show a pattern of Native American ancestry that is intermediate to the Canadian  
391 and Mesoamerican reference populations analyzed here, whereas the Mexican American individuals from the  
392 same region are more closely related to Mesoamerican reference populations. This is consistent with early

393 admixture with local Native American groups in the US southwest, for the *Nuevomexicanos*, versus admixture with  
394 Mesoamerican groups in Mexico for the later Mexican immigrants. A more precise characterization of  
395 *Nuevomexicanos'* Native American ancestry would require access to genomic data from US Native American  
396 reference populations, which are not readily available owing to cultural resistance to genetic testing for ancestry  
397 among these groups<sup>37</sup>.

398 Historically, *Nuevomexicanos* have identified strongly with their European (Spanish) ancestry, while downplaying  
399 ancestral ties to Native Americans<sup>38</sup>. This tradition of exclusive European identity is rooted in the colonial era  
400 when Spanish descendants in the region were preoccupied with the notion of maintaining so-called pure blood,  
401 and the local aristocracy identified as Castilian. Mexicans, on the other hand, have long identified as *Mestizo* with  
402 an explicit recognition of their Native American heritage<sup>39</sup>. Our comparative analysis of genetic ancestry for  
403 *Nuevomexicanos* and Mexican ancestry groups yielded results that are partly consistent with this historical  
404 narrative. On the one hand, *Nuevomexicanos* do have a substantial amount of Native American ancestry, with a  
405 median of just under 40% (Figure 6C), which is far more than seen for the African American and European  
406 American groups analyzed here, and also more than seen for an number of other Latin American populations in  
407 the Caribbean and South America<sup>40</sup>. Nevertheless, the *Nuevomexicanos* have significantly less Native American  
408 ancestry, and more European ancestry, than nearby Mexican descendant populations (Figure 6C). Our results are  
409 consistent with a recent study that used microsatellite-based ancestry analysis on a much smaller sample of self-  
410 identified *Nuevomexicanos*, who were also found to have higher European ancestry and lower Native American  
411 ancestry compared to Mexican Americans<sup>28</sup>. Interestingly, we found that the *Nuevomexicanos* also have  
412 significantly less African ancestry than Mexican descendant populations, which likely reflects higher levels of early  
413 African admixture in Mexico<sup>41</sup>.

414 Perhaps the most controversial aspect of *Nuevomexicano* history relates to the influence of *Conversos*, or crypto-  
415 Jewish individuals, on the culture and traditions of the local community. *Conversos* are Jewish people who  
416 converted to Catholicism under intense pressure from religious persecution in Spain, and elsewhere in Europe,  
417 and many Spanish *Conversos* immigrated to the New World<sup>42</sup>. Despite their forced conversion to Catholicism,  
418 some New World *Conversos* apparently maintained Jewish religious traditions over the centuries since their  
419 immigration from Spain. For example, the persistence of rituals and symbols related to Jewish traditions in New  
420 Mexico has been taken as evidence for an influential presence of *Conversos* among the *Nuevomexicanos*, a  
421 position championed by the historian Stanley Hordes<sup>29</sup>. On the other hand, the folklorist Judith Neulander and  
422 others have been fiercely critical of this narrative based on what they perceive to be misunderstandings of the  
423 origins of many of the cultural traditions tied to Jewish rituals and even deliberate misrepresentations of facts<sup>43</sup>.

424 Neulander's interpretation relates the notion of *Converso* identity among *Nuevomexicanos* back to the colonial  
425 assertions of pure Spanish ancestry given that the Sephardim are Spanish and would presumably be loath to marry  
426 outside of their religious group<sup>44</sup>.

427 We evaluated the extent of Sephardic Jewish ancestry among *Nuevomexicanos*, via comparative analysis of their  
428 European haplotypes to both European and Sephardic Jewish reference populations, in attempt to assess the  
429 genetic evidence in support of the *Converso* narrative. While we did find more Sephardic Jewish ancestry among  
430 *Nuevomexicanos* compared to Spaniards or other Europeans, they did not show any more Sephardic Jewish  
431 ancestry than Mexican descendants from nearby regions (Figure 6D). Our results are consistent with a recent  
432 study that used haplotype-based ancestry methods to uncover widespread *Converso* ancestry in Latin American  
433 populations<sup>30</sup>. Taken together, we interpret these results to indicate that, while *Nuevomexicanos* do in fact have  
434 a demonstrable amount of Jewish ancestry, they are no more, or less, Jewish than other New World Latin  
435 American populations. Of course, we cannot weigh in on the strength of evidence for or against the persistence  
436 of Jewish cultural traditions among *Nuevomexicanos* based on our genetic evidence alone. Nevertheless, there  
437 does not seem to be anything particularly unique, at least from the genetic perspective, with respect to the extent  
438 of Sephardic Jewish heritage among *Nuevomexicanos*.

439 **Conclusion.** Much of the genetic legacy of the original inhabitants of the area that is now the continental US can  
440 be found in the genomes of the descendants of European and African immigrants to the region. In this study, we  
441 analyzed signals of cryptic Native American genetic ancestry that can be gleaned from comparative analysis of  
442 genomes from three distinct US ancestry groups: African American, European American, and Hispanic/Latino. Our  
443 study was enabled by the use of haplotype-based methods for genetic ancestry inference and leveraged a large  
444 dataset of whole genome genotypes. This approach allowed for detailed analysis of Native American ancestry  
445 patterns even when the per-genome levels of Native American ancestry were quite low, *i.e.* cryptic genetic  
446 ancestry. Each of the three genetic ancestry groups analyzed here shows distinct profiles of Native American  
447 ancestry, which reflect population-specific historical patterns of migration and settlement across the US. Analysis  
448 of the Native American ancestry component for members of these groups allowed for the delineation of region-  
449 specific subpopulations, such as the *Nuevomexicanos* from the American southwest, and facilitated the  
450 interrogation of specific historical scenarios.

451

## 452 **References**

- 453 1. Reich, D. (2018). *Who We Are and How We Got Here: Ancient DNA and the new science of the human*  
454 *past.*(Oxford University Press).
- 455 2. Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa,  
456 N., et al. (2012). Reconstructing Native American population history. *Nature* 488, 370-374.
- 457 3. Jenkins, D.L., Davis, L.G., Stafford, T.W., Jr., Campos, P.F., Hockett, B., Jones, G.T., Cummings, L.S., Yost, C.,  
458 Connolly, T.J., Yohe, R.M., 2nd, et al. (2012). Clovis age Western Stemmed projectile points and human coprolites  
459 at the Paisley Caves. *Science* 337, 223-228.
- 460 4. Fuson, R.H. (2000). *Juan Ponce de León and the Spanish Discovery of Puerto Rico and Florida.*(McDonald and  
461 Woodward Publishing Company).
- 462 5. Price, D.A. (2007). *Love and hate in Jamestown: John Smith, Pocahontas, and the start of a new nation.*(Vintage).
- 463 6. Hashaw, T. (2007). *The birth of Black America: The first African Americans and the pursuit of freedom at*  
464 *Jamestown.*(Basic Books).
- 465 7. Eltis, D., and Richardson, D. (2010). *Atlas of the Transatlantic Slave Trade.*(Yale University Press).
- 466 8. Mann, C.C. (2005). *1491: New revelations of the Americas before Columbus.*(Alfred a Knopf Incorporated).
- 467 9. Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. (2014). A genetic atlas  
468 of human admixture history. *Science* 343, 747-751.
- 469 10. Genomes Project, C., Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles,  
470 M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467,  
471 1061-1073.
- 472 11. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini,  
473 J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68-74.
- 474 12. Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-  
475 Tamir, B., Cambon-Thomsen, A., et al. (2002). A human genome diversity cell line panel. *Science* 296, 261-262.
- 476 13. Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech, Z.A., Degnan, J.H., Wang,  
477 K., Guerreiro, R., et al. (2008). Genotype, haplotype and copy-number variation in worldwide human populations.  
478 *Nature* 451, 998-1003.
- 479 14. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S.,  
480 Feldman, M., Cavalli-Sforza, L.L., et al. (2008). Worldwide human relationships inferred from genome-wide  
481 patterns of variation. *Science* 319, 1100-1104.
- 482 15. Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev,  
483 I., Yudkovsky, G., et al. (2010). The genome-wide structure of the Jewish people. *Nature* 466, 238-242.

- 484 16. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK:  
485 rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
- 486 17. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and  
487 population genetic studies. *Nature methods* 10, 5-6.
- 488 18. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach  
489 for rapid and robust local-ancestry inference. *Am J Hum Genet* 93, 278-288.
- 490 19. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated  
491 individuals. *Genome Res* 19, 1655-1664.
- 492 20. Conley, A.B., Rishishwar, L., Norris, E.T., Valderrama-Aguirre, A., Marino-Ramirez, L., Medina-Rivas, M.A., and  
493 Jordan, I.K. (2017). A Comparative Analysis of Genetic Ancestry and Admixture in the Colombian Populations of  
494 Choco and Medellin. *G3 (Bethesda)* 7, 3435-3447.
- 495 21. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2015). e1071: Misc Functions of the  
496 Department of Statistics (e1071), TU Wien, 2014. R package version, 1.6-4.
- 497 22. Rishishwar, L., Conley, A.B., Wigington, C.H., Wang, L., Valderrama-Aguirre, A., and Jordan, I.K. (2015).  
498 Ancestry, admixture and fitness in Colombian genomes. *Sci Rep* 5, 12376.
- 499 23. Fortes-Lima, C., Gessain, A., Ruiz-Linares, A., Bortolini, M.C., Migot-Nabias, F., Bellis, G., Moreno-Mayar, J.V.,  
500 Restrepo, B.N., Rojas, W., Avendano-Tamayo, E., et al. (2017). Genome-wide Ancestry and Demographic History  
501 of African-Descendant Maroon Communities from French Guiana and Suriname. *Am J Hum Genet* 101, 725-736.
- 502 24. Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic  
503 trees. *Mol Biol Evol* 4, 406-425.
- 504 25. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary  
505 Genetics Analysis version 6.0. *Mol Biol Evol* 30, 2725-2729.
- 506 26. Goldberg, A., and Rosenberg, N.A. (2015). Beyond 2/3 and 1/3: The Complex Signatures of Sex-Biased  
507 Admixture on the X Chromosome. *Genetics* 201, 263-279.
- 508 27. Gonzales, P.B. (2006). Whither the Nuevomexicanos: The career of a southwestern intellectual discourse,  
509 1907-2004. *Soc Sci J* 43, 273-286.
- 510 28. Healy, M.E., Hill, D., Berwick, M., Edgar, H., Gross, J., and Hunley, K. (2017). Social-group identity and  
511 population substructure in admixed populations in New Mexico and Latin America. *PLoS One* 12, e0185503.
- 512 29. Hordes, S.M. (2008). *To the end of the earth: a history of the crypto-Jews of New Mexico.* (Columbia University  
513 Press).



- 514 30. Chacon-Duque, J.C., Adhikari, K., Fuentes-Guajardo, M., Mendoza-Revilla, J., Acuna-Alonzo, V., Lozano, R.B.,  
515 Quinto-Sanchez, M., Gomez-Valdes, J., Martinez, P.E., and Villamil-Ramirez, H. (2018). Latin Americans show wide-  
516 spread Converso ancestry and the imprint of local Native ancestry on physical appearance. *bioRxiv*, 252155.
- 517 31. Wilkerson, I. (2011). *The warmth of other suns: The epic story of America's great migration.*(Vintage).
- 518 32. Baharian, S., Barakatt, M., Gignoux, C.R., Shringarpure, S., Errington, J., Blot, W.J., Bustamante, C.D., Kenny,  
519 E.E., Williams, S.M., Aldrich, M.C., et al. (2016). The Great Migration and African-American Genomic Diversity.  
520 *PLoS Genet* 12, e1006059.
- 521 33. Pascoe, P. (2009). *What comes naturally: Miscegenation law and the making of race in America.*(Oxford  
522 University Press on Demand).
- 523 34. Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., Hammer, M., Bustamante, C.D., and  
524 Ostrer, H. (2010). Colloquium paper: genome-wide patterns of population structure and admixture among  
525 Hispanic/Latino populations. *Proc Natl Acad Sci U S A* 107 Suppl 2, 8954-8961.
- 526 35. Bryc, K., Auton, A., Nelson, M.R., Oksenberg, J.R., Hauser, S.L., Williams, S., Froment, A., Bodo, J.M., Wambebe,  
527 C., Tishkoff, S.A., et al. (2010). Genome-wide patterns of population structure and admixture in West Africans and  
528 African Americans. *Proc Natl Acad Sci U S A* 107, 786-791.
- 529 36. Lind, J.M., Hutcheson-Dilks, H.B., Williams, S.M., Moore, J.H., Essex, M., Ruiz-Pesini, E., Wallace, D.C., Tishkoff,  
530 S.A., O'Brien, S.J., and Smith, M.W. (2007). Elevated male European and female African contributions to the  
531 genomes of African American individuals. *Hum Genet* 120, 713-722.
- 532 37. TallBear, K. (2013). *Native American DNA: Tribal belonging and the false promise of genetic science.*(University  
533 of Minnesota Press).
- 534 38. Nieto-Phillips, J.M. (2008). *The language of blood: The making of Spanish-American identity in New Mexico,  
535 1880s-1930s.*(UNM Press).
- 536 39. Vasconcelos, J. (1925). *La raza cósmica: misión de la raza iberoamericana.*(Agencia Mundial de Librería).
- 537 40. Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J.L., Byrnes, J.K., Gignoux, C.R., Ortiz-Tello, P.A.,  
538 Martinez, R.J., Hedges, D.J., Morris, R.W., et al. (2013). Reconstructing the population genetic history of the  
539 Caribbean. *PLoS Genet* 9, e1003925.
- 540 41. Moreno-Estrada, A., Gignoux, C.R., Fernandez-Lopez, J.C., Zakharia, F., Sikora, M., Contreras, A.V., Acuna-  
541 Alonzo, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., et al. (2014). Human genetics. The genetics of Mexico  
542 recapitulates Native American substructure and affects biomedical traits. *Science* 344, 1280-1285.
- 543 42. Roth, N. (2002). *Conversos, Inquisition, and the Expulsion of the Jews from Spain.*(Univ of Wisconsin Press).
- 544 43. Ferry, B., and Nathan, D. (2000). Mistaken Identity? The Case of New Mexico's 'Hidden Jews,'. *Atlantic Monthly*  
545 283, 85-96.

- 546 44. Weitz, J.S. (2014). Let my people go (home) to Spain: a genealogical model of Jewish identities since 1492.  
547 PLoS One 9, e85673.  
548