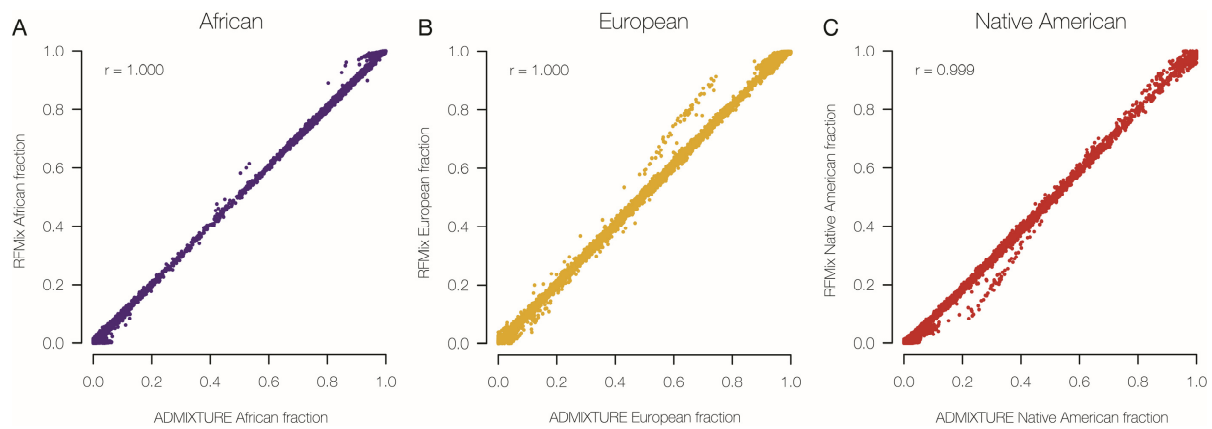


Supplementary Information for:

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

I. King Jordan, Lavanya Rishishwar, and Andrew B. Conley



Supplementary Figure S1. **Comparison of genome-wide ancestry estimates for HRS individuals inferred using our modified version of RFMix (y-axis) versus ADMIXTURE (x-axis).** African (blue), European (yellow), and Native American (red) ancestry fractions for HRS individuals inferred using the two methods are regressed, and the Pearson correlation coefficient values (r -values) are shown.

Delineation of genetic ancestry groups in the US. We compared whole genome genotypes from 15,620 HRS individuals to genotype data from African, European, Native American, and East Asian global reference populations in an effort to define the major genetic ancestry groups. The individual HRS donors who contributed genotype data were placed into two broad groups by the study according to their self-identified race/ethnicity: African Americans and all others. We initially ran ADMIXTURE, comparing these two HRS-defined groups to global reference populations, in order to visualize the continental ancestry fractions – African, European, Native American, and East Asian – for individuals in each group (Supplementary Figure 2).

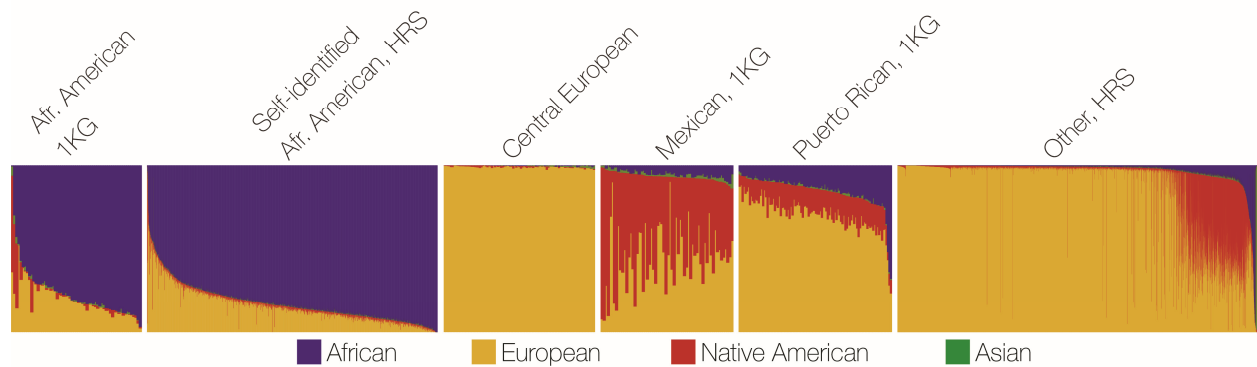
Self-identified African Americans from the HRS show a wide range of African ancestry (0 – 100%) with an average African ancestry of 81.7%. Interestingly, this group includes a number of individuals with anomalously low levels African genetic ancestry, including 11 individuals with little to no discernable African ancestry. The second HRS-defined group, containing all of the remaining genotype donors, shows a mix of individuals with primarily European ancestry along with a second large group that contains individuals with a mix of primarily European and Native American ancestry. That second group, which most likely corresponds to Hispanic/Latino individuals, also contains lower levels of African ancestry. There is a third smaller group made up of individuals with almost entirely East Asian ancestry.

We wanted to more objectively classify HRS individuals from these two broad groups based on their observable patterns of genetic ancestry, as opposed to using self-identified race/ethnicity, which can be an unreliable indicator of ancestry. Individuals with primarily European and/or Native American ancestry were classified into the European American or Hispanic/Latino ancestry groups based largely on their sub-continental European ancestry. To do so, we considered only the European haplotypes for these individuals and compared them to haplotypes from (1) Spanish versus (2) Finnish, British, French, and Italian European reference populations. As described in the Materials and Methods section, a support vector machine (SVM) classifier was trained on ADMIXTURE vectors for these two groups of European populations and then the SVM was used to classify the HRS individuals as having primarily Spanish-European ancestry or other European ancestry. This was done for HRS individuals grouped by their geographic origins among nine census regions in the continental US (Supplementary Figure S3).

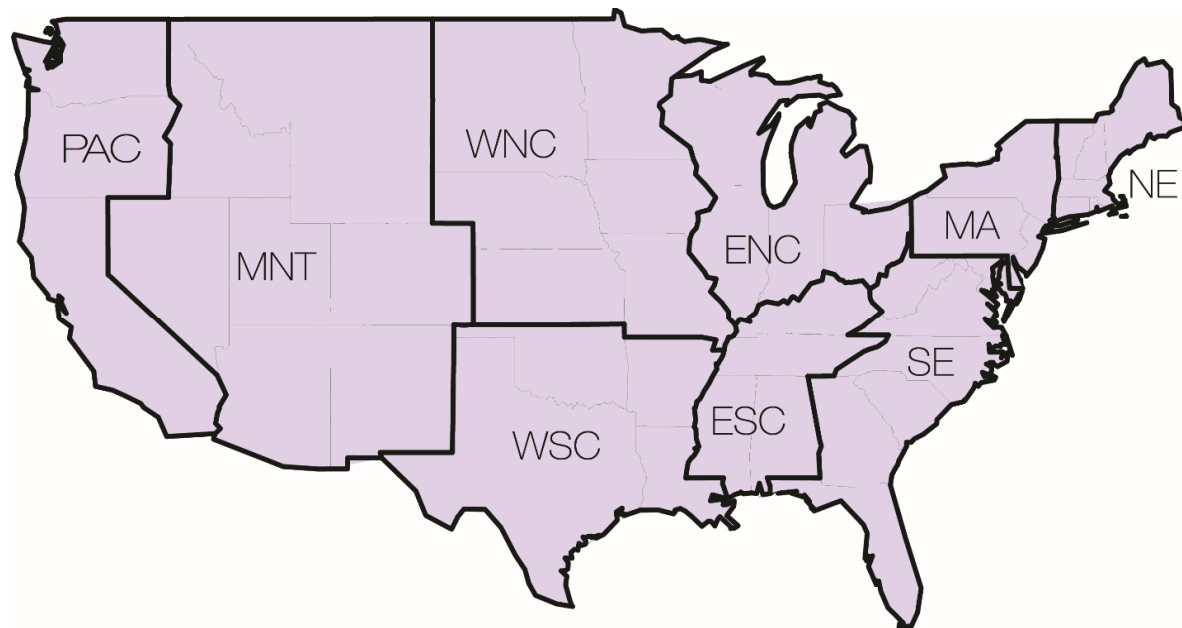
Individuals classified by the ADMIXTURE-based SVM as having Spanish-European ancestry were considered as members of the Hispanic/Latino ancestry group, whereas individuals with other European ancestry are considered as belonging to the European American ancestry group. The ADMIXTURE ancestry fractions (vectors) used for this classification can be seen in Supplementary Figures S4 (European American) and S5 (Hispanic/Latino). European American individuals show European ancestry patterns that are similar to the British and French European reference populations (Supplementary Figure S4), whereas the Hispanic/Latino individuals show European ancestry patterns that more closely resemble both the Spanish European reference population and the Mexican reference population (Supplementary Figure S5).

HRS individuals were classified into the African American (AA) ancestry group largely based on their observed African genetic ancestry; this group includes all individuals that have >20% African ancestry and non-Spanish European ancestry. Individuals who had non-Spanish European ancestry and showed <5% African ancestry, including a number of self-identified African Americans, were placed into the European American ancestry group. Individuals from the African American genetic ancestry group show European

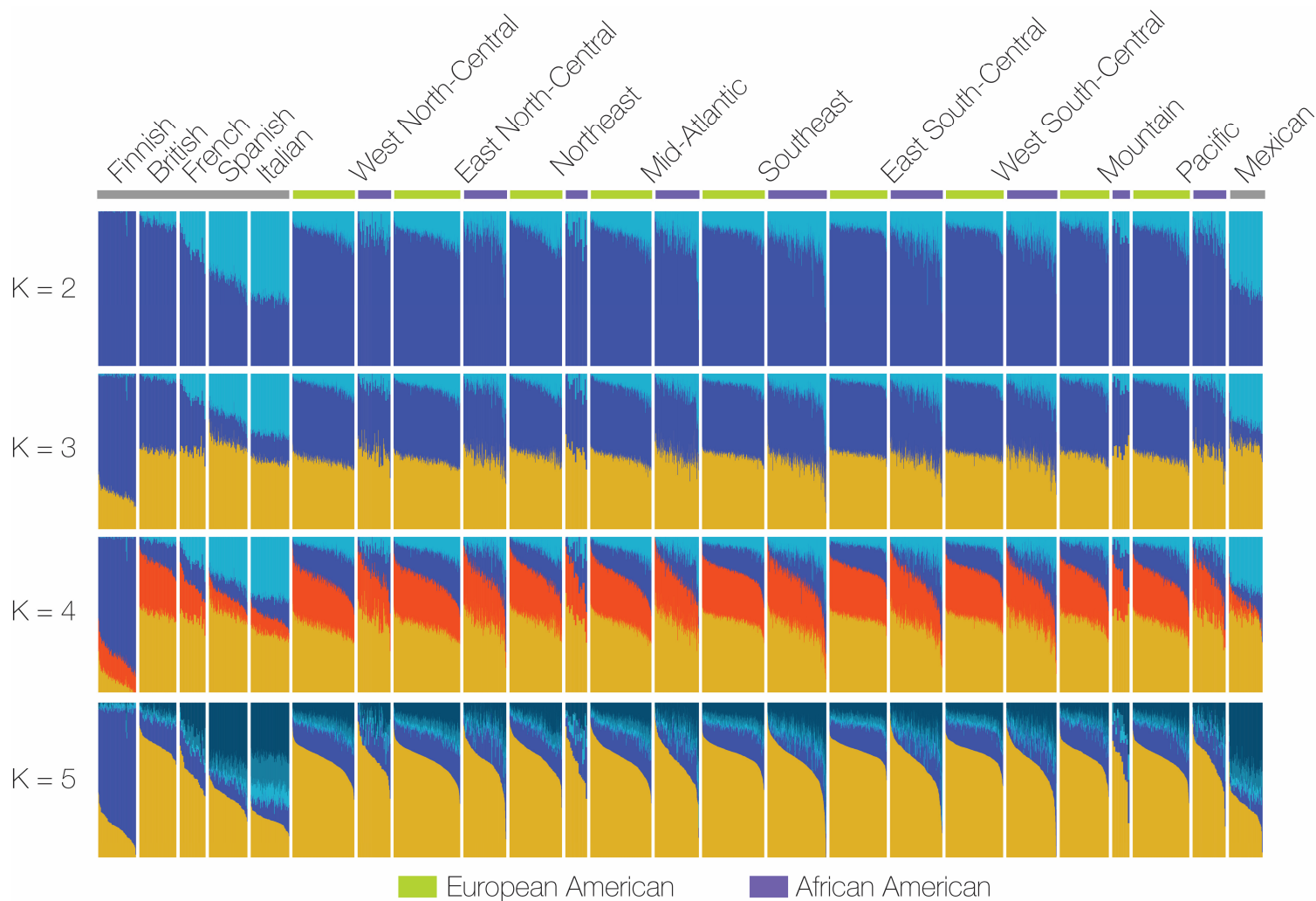
ancestry patterns that are similar to those seen for the European American ancestry group and distinct from those of the Hispanic/Latino ancestry group.



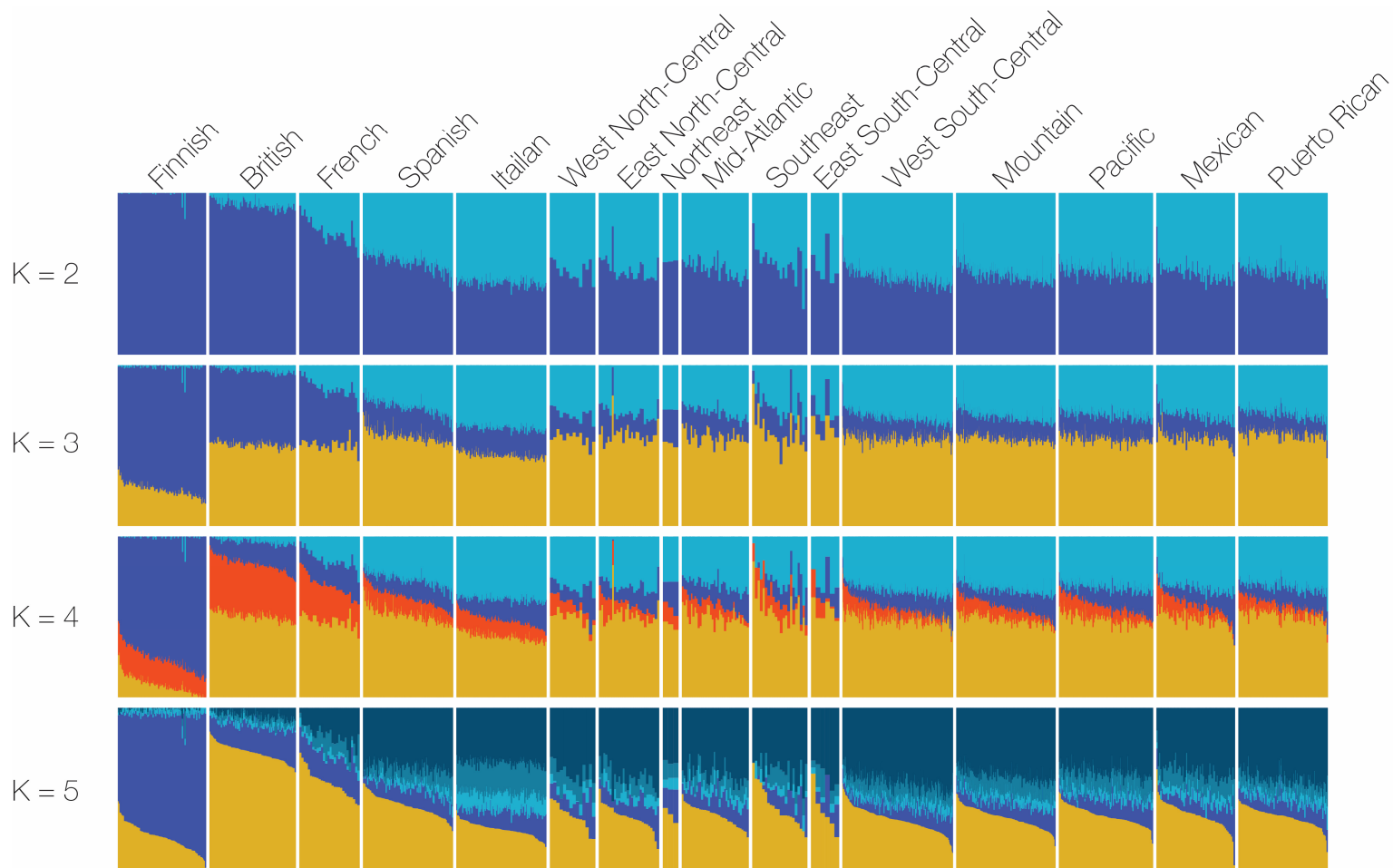
Supplementary Figure S2. **ADMIXTURE plot showing HRS individuals' continental ancestry fractions.** African (blue), European (yellow), Native American (red), and East Asian (green) ancestry components are shown. The HRS individuals were divided into two groups based on their self-identified status as African-Americans and all others. HRS individuals shown in comparison to African (African American – ASW), European (Central European – CEU), Latin American (Mexican – MXL and Puerto Rican – PUR) 1KGP reference populations.



Supplementary Figure S3. **Map showing the nine US census regions used to assign the geographic origins of the HRS individuals analyzed here.** The census regions, in semi-clockwise order, are: West North-Central (WNC), East North-Central (ENC), Northeast (NE), Mid-Atlantic (MA), Southeast (SE), East South-Central (ESC), West South-Central (WSC), Mountain (MNT), and Pacific (PAC).



Supplementary Figure S4. **ADMIXTURE analysis of European haplotypes for HRS individuals classified by genetic ancestry as African American (AA) or European American (EA).** Individuals were placed into genetic ancestry groups using the SVM classifier as described in the Materials and Methods. ADMIXTURE was run on HRS individuals from the two genetic ancestry groups (AA-purple and EA-lime) grouped according to their geographic region of origin along with individuals from European (Finnish, British, French, Spanish, and Italian) and Mexican 1KGP reference populations. ADMIXTURE was run using $K=2, 3, 4,$ and 5 populations.



Supplementary Figure S5. **ADMIXTURE analysis of European haplotypes for HRS individuals classified by genetic ancestry as Hispanic/Latino (HL).** Individuals were placed into genetic ancestry groups using the SVM classifier as described in the Materials and Methods. ADMIXTURE was run on HRS individuals from the Hispanic/Latino genetic ancestry groups grouped according to their geographic region of origin along with individuals from European (Finnish, British, French, Spanish, and Italian), Mexican, and Puerto Rican 1KGP reference populations. ADMIXTURE was run using $K=2, 3, 4,$ and 5 populations.