

RESEARCH

Open Access

Admixture-enabled selection for rapid adaptive evolution in the Americas



Emily T. Norris^{1,2,3}, Lavanya Rishishwar^{1,2,3}, Aroon T. Chande^{1,2,3}, Andrew B. Conley^{2,3}, Kaixiong Ye^{4,5}, Augusto Valderrama-Aguirre^{3,6,7} and I. King Jordan^{1,2,3*} 

Abstract

Background: Admixture occurs when previously isolated populations come together and exchange genetic material. We hypothesize that admixture can enable rapid adaptive evolution in human populations by introducing novel genetic variants (haplotypes) at intermediate frequencies, and we test this hypothesis through the analysis of whole genome sequences sampled from admixed Latin American populations in Colombia, Mexico, Peru, and Puerto Rico.

Results: Our screen for admixture-enabled selection relies on the identification of loci that contain more or less ancestry from a given source population than would be expected given the genome-wide ancestry frequencies. We employ a combined evidence approach to evaluate levels of ancestry enrichment at single loci across multiple populations and multiple loci that function together to encode polygenic traits. We find cross-population signals of African ancestry enrichment at the major histocompatibility locus on chromosome 6, consistent with admixture-enabled selection for enhanced adaptive immune response. Several of the human leukocyte antigen genes at this locus, such as *HLA-A*, *HLA-DRB51*, and *HLA-DRB5*, show independent evidence of positive selection prior to admixture, based on extended haplotype homozygosity in African populations. A number of traits related to inflammation, blood metabolites, and both the innate and adaptive immune system show evidence of admixture-enabled polygenic selection in Latin American populations.

Conclusions: The results reported here, considered together with the ubiquity of admixture in human evolution, suggest that admixture serves as a fundamental mechanism that drives rapid adaptive evolution in human populations.

Keywords: Rapid adaptive evolution, Positive selection, Genetic ancestry, Admixture, Population genomics, Polygenic traits

Background

Admixture is increasingly recognized as a ubiquitous feature of human evolution [1]. Recent studies on ancient DNA have underscored the extent to which human evolution has been characterized by recurrent episodes of population isolation and divergence followed by convergence and admixture. In this study, we considered the implications of admixture for human adaptive evolution [2]. We hypothesized that admixture is a critical mechanism that enables rapid adaptive evolution in human populations, and we tested this hypothesis via the analysis of admixed genome sequences from four Latin American populations:

Colombia, Mexico, Peru, and Puerto Rico. We refer to the process whereby the presence of distinct ancestry-specific haplotypes on a shared population genomic background facilitates adaptive evolution as “admixture-enabled selection.”

The conquest and colonization of the Americas represents a major upheaval in the global migration of our species and is one of the most abrupt and massive admixture events known to have occurred in human evolution [3, 4]. The ancestral source populations—from Africa, Europe, and the Americas—that admixed to form modern Latin American populations evolved separately for tens of thousands of years before coming together over the last 500 years. This 500-year time frame, corresponding to approximately 20 generations, amounts to less than 1% of the time that has elapsed since modern humans first emerged from Africa [5, 6]. Considered together, these facts point to admixed Latin American

* Correspondence: king.jordan@biology.gatech.edu

¹School of Biological Sciences, Georgia Institute of Technology, 950 Atlantic Drive, Atlanta, GA 30332, USA

²IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, GA, USA

Full list of author information is available at the end of the article



populations as an ideal system to study the effects of admixture on rapid adaptive evolution in humans [7].

A number of previous studies have considered the possibility of admixture-enabled selection in the Americas, yielding conflicting results. On the one hand, independent studies have turned up evidence for admixture-enabled selection at the major histocompatibility complex (MHC) locus in Puerto Rico [8], Colombia [9], and Mexico [10], and another study found evidence for admixture-enabled selection on immune system signaling in African-Americans, particularly as it relates to influenza and malaria response [11]. Together, these studies highlighted the importance of the immune system as a target for admixture-enabled selection among a diverse group of admixed American populations. However, a follow-up study on a different cohort of African-Americans found no evidence for admixture-enabled selection in the Americas [12]. The latter study concluded that the observed differences in local ancestry reported by previous studies, which were taken as evidence for selection, could have occurred by chance alone given the large number of hypotheses that were tested (i.e., the number of loci analyzed across the genome). This work underscored the importance of controlling for multiple hypothesis testing when investigating the possibility of admixture-enabled selection in the Americas.

We attempted to resolve this conundrum by performing integrated analyses that combine information from (1) single loci across multiple populations and (2) multiple loci that encode polygenic traits. We also used admixture simulation, along with additional lines of evidence from haplotype-based selection scans, to increase the stringency of, and confidence in, our screen for admixture-enabled selection. This combined evidence approach has proven to be effective for the discovery of admixture-enabled selection among diverse African populations [13, 14]. We found evidence for admixture-enabled selection at the MHC locus across multiple Latin American populations, consistent with previous results, and our polygenic screen uncovered novel evidence for adaptive evolution on a number of inflammation, blood, and immune-related traits.

Results

Genetic ancestry and admixture in Latin America

We inferred patterns of genetic ancestry and admixture for four Latin American (LA) populations characterized as part of the 1000 Genomes Project: Colombia ($n = 94$), Mexico ($n = 64$), Peru ($n = 85$), and Puerto Rico ($n = 104$) (Fig. 1). Genome-wide continental ancestry fractions were inferred using the program ADMIXTURE [15], and local (haplotype-specific) ancestry was inferred using the program RFMix [16]. The results from both programs are highly concordant, and local ancestry assignments are robust to the use of distinct reference populations or variable recombination parameters (Additional file 1: Figures S1-S3). As

expected [17–20], the four LA populations show genetic ancestry contributions from African, European, and Native American source populations, and they are distinguished by the relative proportions of each ancestry. Overall, these populations show primarily European ancestry followed by Native American and African components. Puerto Rico has the highest European ancestry, whereas Peru shows the highest Native American ancestry. Mexico shows relatively even levels of Native American and European ancestry, while Colombia shows the highest levels of three-way admixture. Individual genomes vary greatly with respect to the genome-wide patterns of local ancestry, i.e., the chromosomal locations of ancestry-specific haplotypes (Additional file 1: Figure S4). If the process of admixture is largely neutral, then we expect ancestry-specific haplotypes to be randomly distributed throughout the genome in proportions corresponding to the genome-wide ancestry fractions.

Ancestry enrichment and admixture-enabled selection

For each of the four LA populations, local ancestry patterns were used to search for specific loci that show contributions from one of the three ancestral source populations which are greater than can be expected based on the genome-wide ancestry proportions for the entire population (Additional file 1: Figure S5). The ancestry enrichment metric that we use for this screen (z_{anc}) is expressed as the number of standard deviations above or below the genome-wide ancestry fraction. Previous studies have used this general approach to look for evidence of admixture-enabled selection at individual genes within specific populations, yielding mixed results [8–12]. For this study, we have added two new dimensions to this general approach in an effort to simultaneously increase the confidence for admixture-enabled selection inferences and to broaden the functional scope of previous studies. To achieve these ends, we searched for (1) concordant signals of ancestry enrichment for single genes (loci) across multiple populations, and (2) concordant signals of ancestry enrichment across multiple genes that function together to encode polygenic phenotypes. The first approach can be considered to increase specificity, whereas the second approach increases sensitivity. Loci that showed evidence for ancestry enrichment using this combined approach were interrogated for signals of positive selection using the integrated haplotype score (iHS) [21] to further narrow the list of potential targets of admixture-enabled selection.

Single gene admixture-enabled selection

Gene-specific ancestry enrichment values (z_{anc}) were computed for each of the three continental ancestry components within each of the four admixed LA populations analyzed here. We then integrated gene-specific z_{anc} values across the four LA populations using a Fisher combined

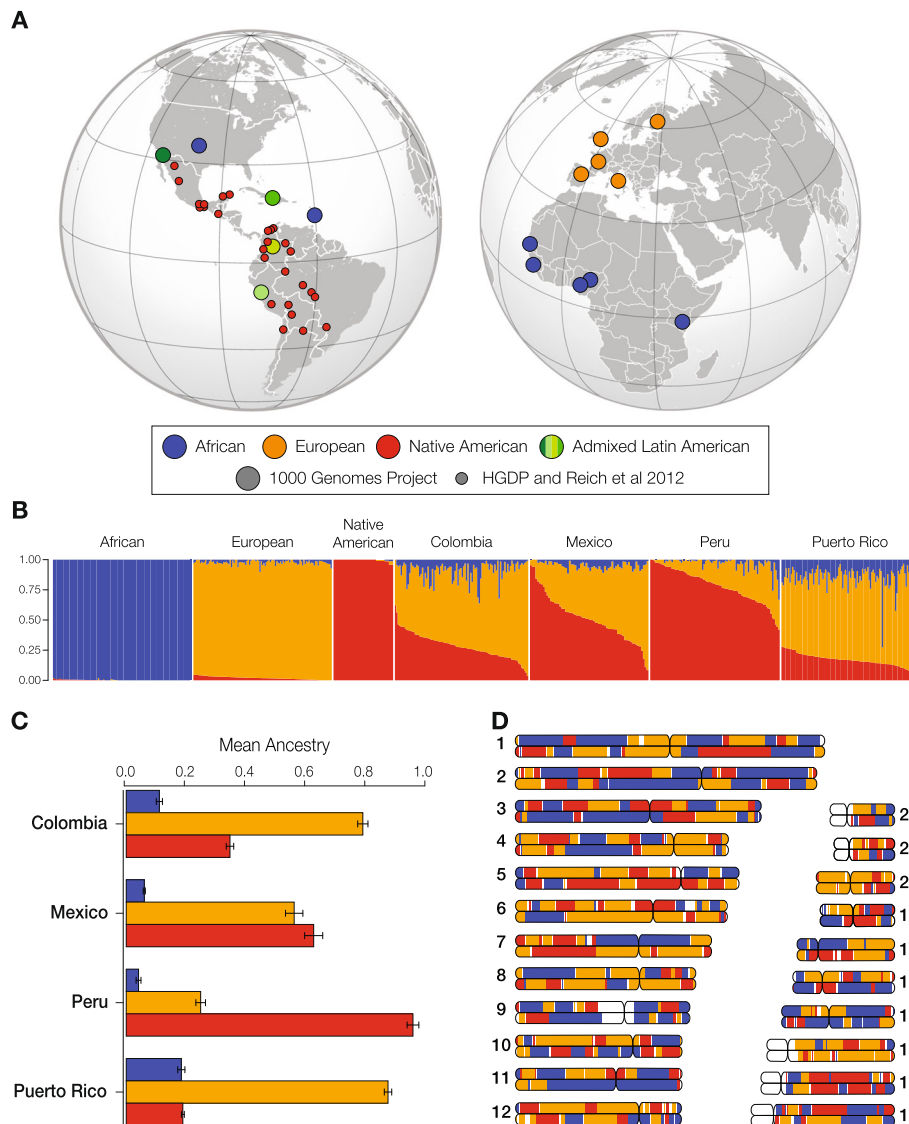


Fig. 1 Genetic ancestry and admixture in Latin America. **a** The global locations of the four LA populations analyzed here (green) are shown along with the locations of the African (blue), European (orange), and Native American (red) reference populations. The sources of the genomic data are indicated in the key. **b** ADMIXTURE plot showing the three-way continental ancestry components for individuals from the four LA populations—Colombia, Mexico, Peru, and Puerto Rico—compared to global reference populations. **c** The mean (\pm se) continental ancestry fractions for the four LA populations. **d** Chromosome painting showing the genomic locations of ancestry-specific haplotypes for an admixed LA genome.

score (F_{CS}). The strongest signals of single gene ancestry enrichment were seen for African ancestry at the major histocompatibility complex (MHC) locus on the short arm of chromosome 6 (Fig. 2a). Three out of the four LA populations show relatively high and constant African ancestry enrichment across this locus, with the highest levels of enrichment seen for Mexico and Colombia (Fig. 2b). This signal is robust to control for multiple statistical tests using the Benjamini–Hochberg false discovery rate (FDR).

We used two independent approaches to simulate random admixture across the four LA populations in an effort to further assess the probability that this signal

could be generated by chance alone (i.e., by genetic drift). The first genome-wide simulation was parameterized by populations’ ancestry proportions; the second simulation focused on chromosome 6 and included additional known demographic features of LA populations. The demographic features taken from the literature on these LA populations, starting effective population size ($n = 100$) and generations since admixture ($g = 10$), were chosen to simulate a recent bottleneck that could be expected to yield a high variance in local ancestry fractions by chance alone [22, 23]. Based on the first genome-wide simulation, the observed levels of cross-population

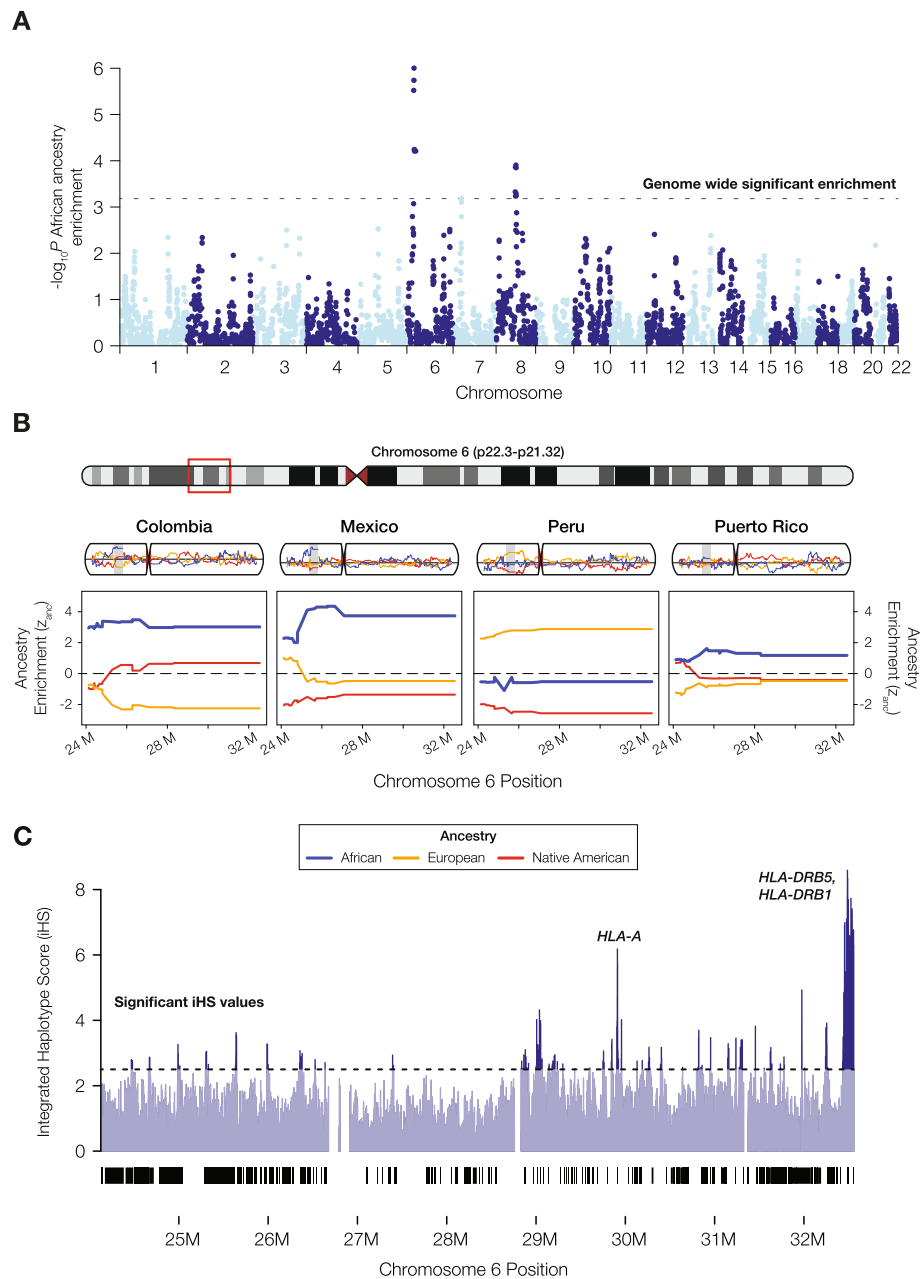


Fig. 2 African ancestry enrichment at the major histocompatibility complex (MHC) locus. **a** Manhattan plot showing the statistical significance of African ancestry enrichment across the genome. **b** Haplotype on chromosome 6 with significant African ancestry enrichment for three of the four LA populations: Colombia, Mexico, and Puerto Rico. This region corresponds to the largest peak of African ancestry enrichment on chromosome 6 seen in **a**. Population-specific African (blue), European (orange), and Native American (red) ancestry enrichment values (Z_{anc}) are shown for chromosome 6 and the MHC locus. **c** Integrated haplotype score (iHS) values for African continental population from the 1KGP are shown for the MHC locus; peaks correspond to putative positively selected human leukocyte antigen (*HLA*) genes.

African ancestry enrichment at the MHC locus are highly unlikely to have occurred by chance ($P < 5 \times 10^{-5}$), whereas the observed patterns of European and Native American ancestry enrichment are consistent with the range of expected levels generated by the random admixture simulation (Additional file 1: Figure S6). Results of the admixture simulation analysis were also

used to demonstrate that the cross-population approach to single locus ancestry-enrichment is sufficiently powered to detect selection at the population sizes analyzed here (Additional file 1: Figures S7 and S8). The demographic simulation of admixture on chromosome 6 also confirmed that African ancestry enrichment at the MHC locus could not have occurred by chance alone, whereas

the observed patterns of European and Native American ancestry enrichment are consistent with the range of expected levels considering the populations demographic features and time since admixture (Additional file 1: Figure S9). The statistical power of the ancestry enrichment approach used in this study rests on the cross-population comparisons, as the probability of observing the same ancestry enrichment at the same locus across multiple LA populations is diminishingly low.

The MHC locus of chromosome 6 also shows a number of peaks for the iHS metric of positive selection from the African continental population (Fig. 2c). These peaks rise well above the value of 2.5, which is taken as a threshold for putative evidence of positive selection [21]. The iHS threshold of 2.5 corresponds to the top ~ 1.4% of values in the data analyzed here. The highest African iHS scores are seen for the human leukocyte antigen (HLA) encoding genes *HLA-A*, *HLA-DRB5*, and *HLA-DRB1* (Fig. 3a, b). These HLA protein encoding genes make up part of the MHC class I (*HLA-A*) and MHC class II (*HLA-DRB5* and *HLA-DRB1*) antigen presenting pathways of the adaptive immune system (Fig. 3c), consistent with shared selective pressures on immune response in admixed LA populations.

We modeled the magnitude of selection pressure that would be needed to generate the observed levels of cross-population African ancestry enrichment at the MHC locus, using a tri-allelic recursive population genetics model that treats ancestry haplotype fractions as allele frequencies (Fig. 4). The average selection coefficient value for African MHC haplotypes is $s = 0.05$ (Additional file 1: Figure S10), indicating strong selection at this locus over the last several hundred years since the admixed LA populations were formed, consistent with previous work [10]. It should be noted that this is an upper bound selection coefficient since ancestry-specific haplotype frequencies are modeled here, and there could be multiple specific haplotypes (alleles) for any given ancestral haplotype.

Polygenic admixture-enabled selection

For each of the three continental ancestry components, we combined gene-specific ancestry enrichment values (z_{anc}), for genes that function together to encode polygenic phenotypes, via the polygenic ancestry enrichment score (PAE) (Fig. 5a). Observed PAE values were compared to expected values generated by randomly permuting size-matched gene sets to search for functions (traits) that show evidence of admixture-enabled selection (Additional file 1: Figure S11). As with the single locus approach, we narrowed our list of targets to traits that showed evidence of polygenic admixture enrichment across multiple LA populations. This approach yielded evidence of statistically significant enrichment and depletion, across multiple ancestries, for a number of inflammation, blood, and immune-related traits (Fig. 5b). Inflammation-related phenotypes that show

polygenic ancestry enrichment include a variety of skin conditions and rheumatoid arthritis. A number of different blood metabolite pathways show evidence for primarily European and Native American ancestry enrichment, while both the adaptive and innate components of the immune system show evidence of admixture-enabled selection.

Several interconnected pathways of the innate immune system—the RIG-I-like receptor signaling pathway, the Toll-like receptor signaling pathway, and the cytosolic DNA-sensing pathway—all show evidence of Native American ancestry enrichment (Fig. 6). All three of these pathways are involved in rapid, first-line immune response to a variety of RNA and DNA viruses as well as bacterial pathogens. Genes from these pathways that show evidence of Native American ancestry enrichment encode a number of distinct interferon, interleukin, and cytokine proteins.

Discussion

Rapid adaptive evolution in humans

Human adaptive evolution is often considered to be a slow process, which is limited by relatively low effective population sizes and long generation times [24–26]. The rate of human adaptive evolution is further constrained by the introduction of new mutations [27]. Initially, positive selection acts very slowly to gradually increase the frequency of newly introduced beneficial mutations, which by definition are found at low population frequencies. The process of admixture, whereby previously diverged populations converge, brings together haplotypes that have not previously existed on the same population genomic background [28]. In so doing, it can provide raw material for rapid adaptive evolution in the form of novel variants that are introduced at intermediate frequencies, many of which may have evolved adaptive utility over thousands of years based on local selection pressures faced by ancestral source populations [7].

Admixture and rapid adaptive evolution

Our results suggest that admixture can enable extremely rapid adaptive evolution in human populations. In the case of the LA populations studied here, we found evidence of adaptive evolution within the last 500 years (or ~ 20 generations) since the conquest and colonization of the Americas began [3, 4]. We propose that, given the ubiquity of admixture among previously diverged populations [1, 2], it should be considered as a fundamental mechanism for the acceleration of human evolution.

The haplotypes that show evidence of ancestry enrichment in our study evolved separately for tens-of-thousands of years in the ancestral source populations—African, European, and Native American—that mixed to form modern, cosmopolitan LA populations. Many of these haplotypes are likely to contain variants, or combinations of variants, that provided a selective advantage in their ancestral

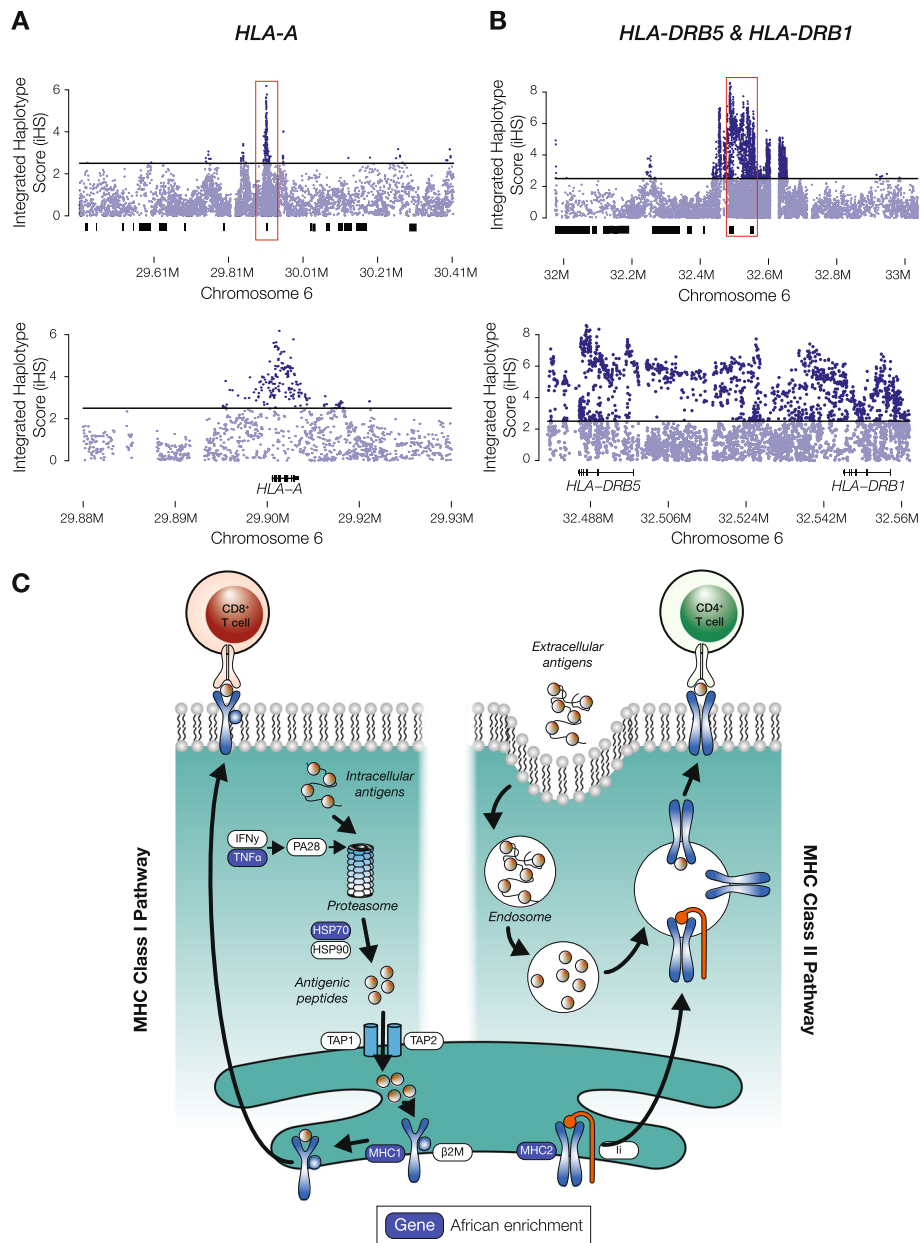


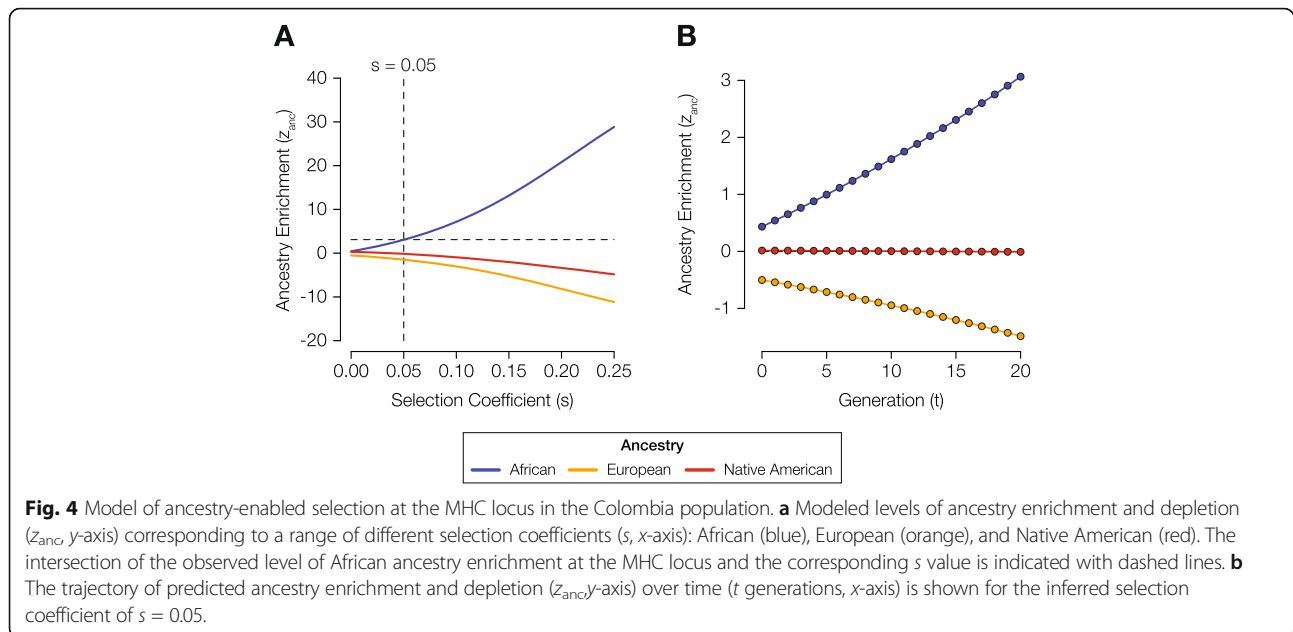
Fig. 3 Admixture-enabled selection at human leukocyte antigen (*HLA*) genes. Integrated haplotype score (iHS) peaks for the African continental population from the 1KGP are shown for **a** the MHC Class I gene *HLA-A* and **b** the MHC Class II genes *HLA-DRB5* and *HLA-DRB1*. **c** Illustration of the MHC Class I and MHC Class II antigen presenting pathways, with African enriched genes shown in blue.

environments [29]. These adaptive variants would have increased in frequency over long periods of time and then later provided source material for rapid adaptation of admixed populations, depending on their utility in the New World environment. Variants that reached high frequency in ancestral source populations via genetic drift could also serve as targets for positive selection in light of the distinct environments and selection pressures faced by modern admixed populations. In either case, admixture-enabled selection can be taken as a special case of selection on

standing variation, or soft selective sweeps, underscoring its ability to support rapid adaptation in the face of novel selective pressures [30, 31].

Single locus versus polygenic selection

Our initial analysis of individual LA populations turned up numerous instances of apparent ancestry enrichment instances genome-wide, including enrichment for all three ancestry components in each of the four populations studied here (Additional file 2: Table S1). However, when ancestry



enrichment signals were combined across all four populations, only a handful of significant results remained after correcting for multiple tests. Finally, when random admixture was simulated, only two peaks of African ancestry enrichment were found to be shared among populations at levels greater than expected by chance (Fig. 2 and Additional file 1: Figure S6). These findings support the conservative nature of our combined evidence approach to using cross-population ancestry enrichment as a criterion for inferring admixture-enabled selection, and also reflect the fact that selection needs to be extremely strong to be detected at single loci. This is especially true given the relatively short period of time that has elapsed since modern LA populations were formed via admixture of ancestral source populations. The results of our population genetic model support this notion, showing an average selection coefficient value of $s = 0.05$ for African haplotypes at the MHC locus.

A number of recent studies have underscored the ubiquity of polygenic selection on complex traits that are encoded by multiple genes, emphasizing the fact that weaker selection dispersed across multiple loci may be a more common mode of adaptive evolution than strong single locus selection [32–35]. The results of our polygenic ancestry enrichment analysis are consistent with these findings, as the polygenic approach yielded signals of admixture-enabled selection for numerous traits across different ancestry components and populations. Thus, the polygenic ancestry enrichment that we employed to infer admixture-enabled selection is both more biologically realistic and better powered compared to the single locus approach.

Admixture-enabled selection and the immune system

Both the single locus and polygenic selection tests turned up multiple cases of admixture-enabled selection on the immune system, including genes and pathways of the both innate and adaptive immune response (Figs. 2, 3, 5, and 6). These results are not surprising when you consider that (1) the immune system represents the interface between humans and their environment and is widely known to be a target of selection [36], and (2) the demographic collapse of Native American populations in the New World is attributed primarily to the introduction of novel pathogens from Africa and Europe, for which they had no natural immune defense [4]. However, the latter point does not seem to be consistent with our finding that three innate immune pathways—the RIG-I-like receptor signaling pathway, the Toll-like receptor signaling pathway, and the cytosolic DNA-sensing pathway—actually show evidence of Native American ancestry enrichment (Fig. 6). This result suggests the possibility of distinct selection pressures acting on innate versus the adaptive immune response in the New World environment.

The innate immune system provides a rapid first-line defense against invading pathogens, whereas the adaptive immune system provides a slower secondary defense. It could be that the Native American innate immune system provided an adequate defense against pathogens that are endemic to the New World, while the corresponding adaptive immune system was not tuned to defend against non-native pathogens introduced from African and Europe. A relatively weak adaptive Native American immune system could also be related to the paucity of domesticated animals, which are the source of many zoonotic diseases, in the New World prior to the Columbian

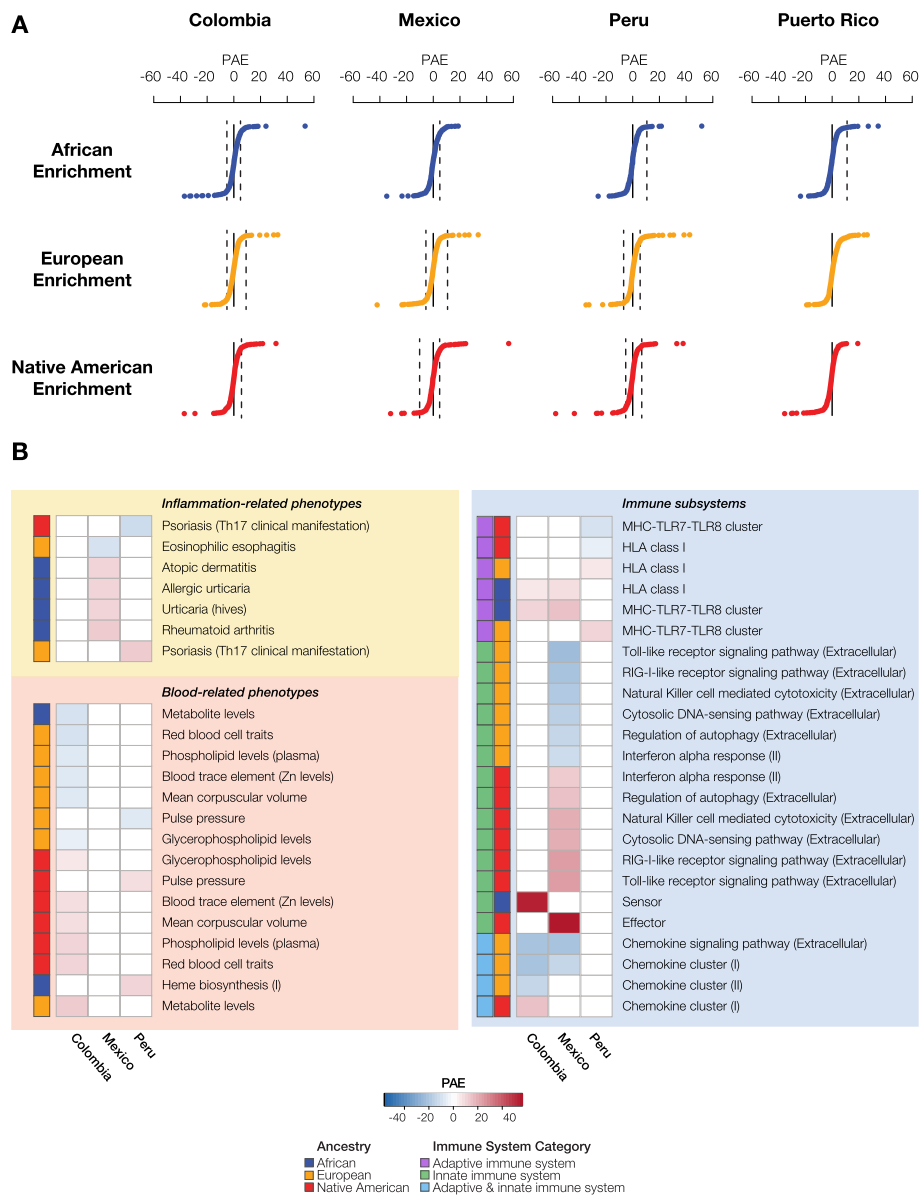


Fig. 5 Polygenic ancestry enrichment (PAE) and admixture-enabled selection. **a** Distributions of the PAE test statistic are shown for each of the three ancestry components—African (blue), European (orange), and Native American (red)—across the four LA populations. Points beyond the dashed lines correspond to polygenic traits with statistically significant PAE values, after correction for multiple tests. **b** Polygenic traits that show evidence of PAE in multiple LA populations. PAE values are color coded as shown in the key, and the ancestry components are indicated for each trait. Immune system traits are divided into adaptive (purple), innate (green), or both (blue).

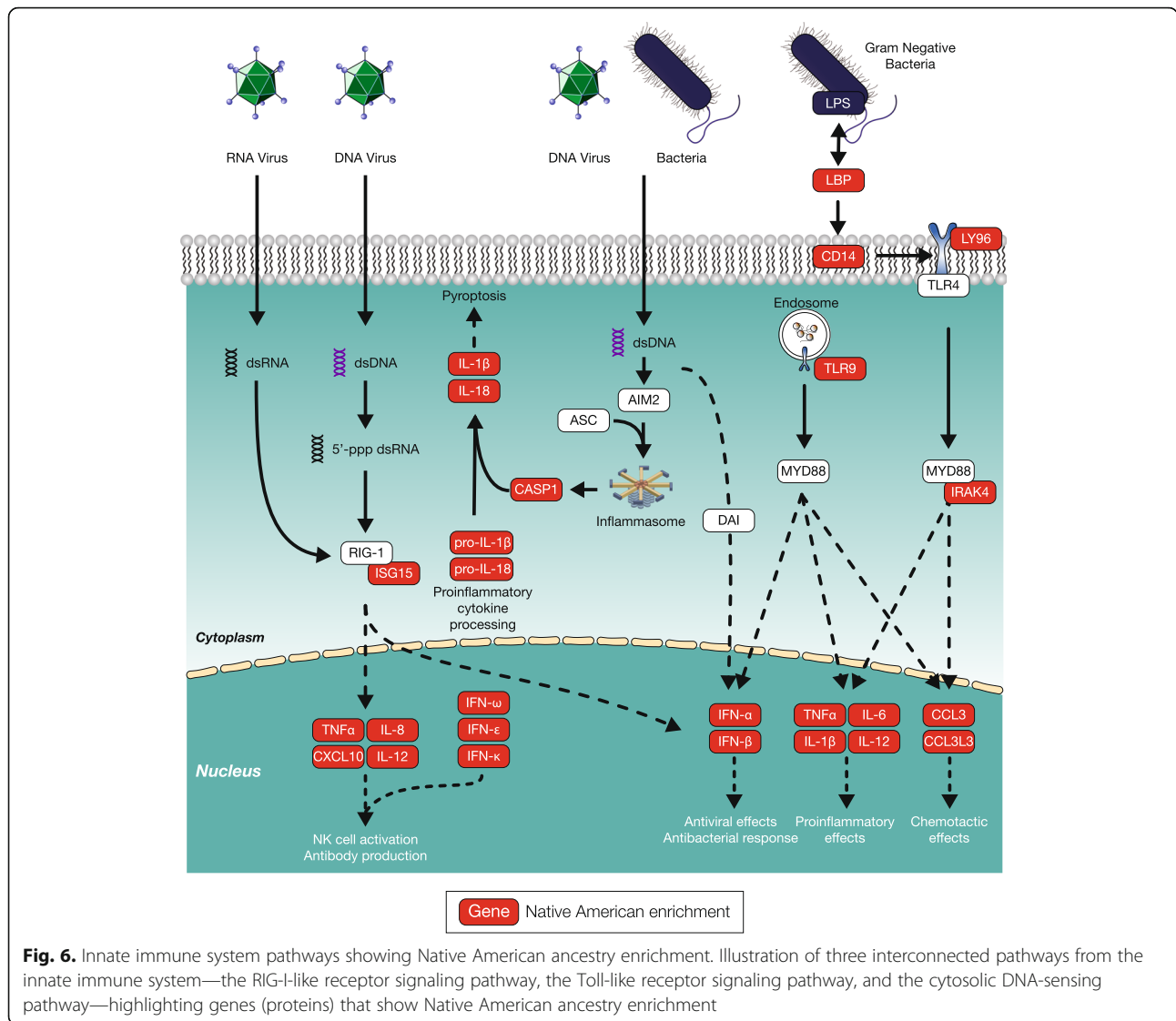
Exchange. Thus, it could be that admixture-enabled selection facilitated the emergence of hybrid immune systems made up of ancestral components best suited to combat both endemic and non-native pathogens.

Conclusions

We report abundant evidence for admixture-enabled selection within and between Latin American populations that were formed by admixture among diverse African, European, and Native American source populations within the

last 500 years. The MHC locus shows evidence of particularly strong admixture-enabled selection for several *HLA* genes, all of which appear to contain pre-adapted variants that were selected prior to admixture in the Americas. In addition, a number of related immune system, inflammation, and blood metabolite traits were found to evolve via polygenic admixture-enabled selection.

Over the last several years, it has become increasingly apparent that admixture is a ubiquitous feature of human evolution. Considering the results of our study



together with the prevalence of admixture leads us to conclude that admixture-enabled selection has been a fundamental mechanism for driving rapid adaptive evolution in human populations.

Methods

Genomic data

Whole genome sequence data for four admixed LA populations—Colombia, Mexico, Peru, and Puerto Rico—were taken from the Phase 3 data release of the 1000 Genomes Project (1KGP) [37, 38]. Whole genome sequence data and whole genome genotypes for proxy ancestral reference populations from Africa, Asia, Europe, and the Americas were taken from multiple sources, including the 1KGP, the Human Genome Diversity Project (HGDP) [39] and a previous study on Native American genetic ancestry [40] (Additional file 1: Table S2). Whole genome sequence and whole genome genotype data

were harmonized using the program PLINK [41], keeping only those sites common to all datasets and correcting SNP strand orientations as needed. A genotyping filter of 95% calls was applied to all populations.

Global and local ancestry inference

Global continental ancestry estimates for each individual from the four LA populations were inferred using the program ADMIXTURE [15]. The harmonized SNP set was pruned using PLINK [41] with window size of 50 bp, a step size of 10 bp, and a linkage disequilibrium (LD) threshold of $r^2 > 0.1$, and ADMIXTURE was run with $K = 4$ corresponding to African, European, Asian, and Native American ancestry components. Local continental ancestry estimates for each individual from the four LA populations were inferred using a modified version of the program RFMix [16] as previously described [42]. The complete harmonized SNP set was phased

using the program SHAPEIT, and RFMix was run to assign African, European, or Native American ancestry to individual haplotypes from the LA populations. Haplotype ancestry assignments were made with a conservative RFMix confidence threshold ≥ 0.98 . The chromosomal locations of ancestry-specific haplotypes were visualized with the program Tagore (<https://github.com/jordanlab/tagore>).

Single locus ancestry enrichment

Single gene (locus) ancestry enrichment (z_{anc}) values were calculated for all three continental ancestry components (African, European, and Native American) across all four LA populations. Genomic locations of NCBI RefSeq gene models were taken from the UCSC Genome Browser (hg19 build) [43], and gene locations were mapped to the ancestry-specific haplotypes characterized using RFMix for each individual genome. For each gene, population-specific three-way ancestry fractions (f_{anc}) were computed as the number of ancestry-specific haplotypes (h_{anc}), divided by the total number of ancestry-assigned haplotypes for that gene (h_{tot}): $f_{\text{anc}} = h_{\text{anc}}/h_{\text{tot}}$. Ancestry enrichment analysis was limited to genes that had h_{tot} values within one standard deviation of the genome-wide average for any population. Distributions of gene-specific ancestry fractions (f_{anc}) for each population were used to calculate population-specific genome-wide average (μ_{anc}) and standard deviation (σ_{anc}) ancestry fractions. Then, for any given gene in any given population, ancestry enrichment (z_{anc}) was calculated as the number of standard deviations above (or below) the genome-wide ancestry average: $z_{\text{anc}} = (f_{\text{anc}} - \mu_{\text{anc}})/\sigma_{\text{anc}}$, with gene-specific ancestry enrichment P values computed using the z distribution. A Fisher's combined score (F_{CS}) was used to combine gene-specific ancestry enrichment P values across the four LA populations as: $F_{\text{CS}} = -2 \sum_{i=1}^4 \ln(P_i)$. The statistical significance of F_{CS} was computed using the χ^2 distribution with 8 ($2k$) degrees of freedom. Correction for multiple F_{CS} tests was performed using the Benjamini-Hochberg false discovery rate (FDR), with a significance threshold of $q < 0.05$ [44].

Admixture simulation

Three-way admixed individuals were randomly simulated for each LA population—Colombia, Mexico, Peru, and Puerto Rico—and used to calculate expected levels of ancestry enrichment z_{anc} as described in the previous section. Expected levels of z_{anc} were combined across the four LA populations to yield expected Fisher's combined scores (F_{CS}) and their associated P values as described in the previous section. Two independent approaches to admixture simulation were used here. For the first approach, admixed populations were simulated

as collections of genes (i.e., ancestry-specific haplotypes) randomly drawn from the genome-wide ancestry distributions for each LA population. Sized matched admixed populations were simulated for each LA population and combined to generate expected (F_{CS}) and their associated P values, and admixture simulation was also conducted across a range of population sizes ($n = 10$ to 10,000) to evaluate the power of the combined evidence cross-population approach used to detect ancestry-enabled selection. This approach was applied across the entire genome for all four LA populations.

For the second approach, admixed populations were simulated using the “Admixture simulation tool,” which can be found at <https://github.com/slowkoni/admixture-simulation>. Each admixed LA population was simulated using a Wright-Fisher forward simulation over 10 generations with an effective population size of $n = 100$ individuals. These parameters represent lower bound estimates for generations since admixture and founding population sizes in the populations studied here [22, 23]. For each population, the initial population was a collection of single-ancestry individuals—proxy African, European, and Native American reference populations—with the proportion of individuals with each ancestry corresponding to the genome-wide average for the given population. In each generation, a portion of the previous generation of admixed individuals was chosen to mate and produce the subsequent generation. Chromosomal recombination rates were accounted for within the software, using HapMap-inferred recombination rates. As with the previous simulation, size-matched admixed populations were created and combined to generate expected (F_{CS}) and their associated P values for chromosome 6.

Polygenic ancestry enrichment

Polygenic ancestry enrichment values (PAE) were computed by combining single locus ancestry enrichment values (z_{anc}) across genes that function together to encode polygenic traits. Gene sets for polygenic traits were curated from a number of literature and database sources to represent a wide array of phenotypes (Additional file 1: Table S3). All gene sets were LD pruned with a threshold of $r^2 > 0.1$ using PLINK. Additional details on the curation of polygenic trait gene sets can be found in Additional file 1 (page 14). For any trait-specific gene set, in any population, PAE was calculated by summing the gene-specific z_{anc} values for all of the genes in the trait set: $\text{PAE} = \sum_1^n z_{\text{anc}}$, where n is the number of genes in the set. Since z_{anc} values can be positive or negative, depending on over- or under-represented ancestry, values of PAE are expected to be randomly distributed around 0. The statistical significance levels of observed PAE values were calculated via comparison against distributions of expected PAE values calculated

from 10,000 random permutations of gene sets, each of which is made up of the same number of genes as the trait-specific gene set being compared (Additional file 1: Figure S11). Observed values (PAE_{obs}) were compared against the mean (μ_{PAE}) and standard deviation (σ_{PAE}) of the expected PAE values to compute the statistical significance for each trait: $z_{PAE} = (PAE_{obs} - \mu_{PAE}) / \sigma_{PAE}$, with P values computed using the z distribution. Correction for multiple tests was performed using the Benjamini-Hochberg false discovery rate (FDR), with a significance threshold of $q < 0.05$.

Integrated haplotype scores (iHS)

Integrated haplotype scores (iHS) [21] were calculated for European and African continental populations from the 1KGP using the software selscan (version 1.1.0a) [45]. |iHS| scores were overlaid on genes with evidence of ancestry enrichment to scan for concurrent signals of selection.

Modeling admixture-enabled selection

Admixture-enabled selection was modeled for the African enriched chromosome 6 MHC haplotype using a standard recursive population genetics model for positive selection [46]. Three allelic states were used for the selection model, each of which corresponds to a specific ancestry component: African, European, or Native American. Population-specific models were initialized with allele (ancestry) frequencies based on the genome-wide background ancestry fractions and run across a range of selection coefficient (s) values to determine the values of s that correspond to the observed African ancestry enrichment levels. This allowed us to compute a positive selection coefficient corresponding to the strength of African ancestry selection at the MHC locus for each population. Additional details of this model can be found in Additional file 1 (page 11–12 and Figure S10).

Review history The review history is available as Additional file 3.

Peer review information Barbara Cheifet was the primary editor on this article and managed its peer review and editorial process in collaboration with the rest of the editorial team.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13059-020-1946-2>.

Additional file 1: Figures S1–S11, Tables S2–S3, and Supplementary Methods.

Additional file 2: Table S1. Ancestry enrichment test statistic values.

Additional file 3. Review history.

Authors' contributions

ETN conducted the ancestry enrichment analyses. LR performed the simulation analyses. ABC performed the genetic ancestry analyses. KY

calculated the integrated haplotype scores. ETN, ATC, and AV-A curated the polygenic phenotype gene sets. ATC performed the polygenic phenotype significance testing. ETN and LR generated the manuscript figures. IKJ conceived of, designed, and supervised the project. ETN, LR, and IKJ wrote the manuscript. All authors read and approved the final manuscript.

Funding

ETN, LR, ATC, ABC, and IKJ were supported by the IHRC-Georgia Tech Applied Bioinformatics Laboratory. KY was supported by the University of Georgia Research Foundation. AV-A was supported by Fulbright Colombia.

Availability of data and materials

1000 Genomes Project (1KGP) data are available from <http://www.internationalgenome.org/data/>.

Human Genome Diversity Project (HGDP) data are available from <http://www.hagsc.org/hgdp/>.

Previously published Native American genotype data can be accessed from a data use agreement governed by the University of Antioquia as previously described [40].

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

Author details

¹School of Biological Sciences, Georgia Institute of Technology, 950 Atlantic Drive, Atlanta, GA 30332, USA. ²IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, GA, USA. ³PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia. ⁴Department of Genetics, University of Georgia, Athens, GA, USA. ⁵Institute of Bioinformatics, University of Georgia, Athens, GA, USA. ⁶Biomedical Research Institute (COL0082529), Cali, Colombia. ⁷Universidad Santiago de Cali, Cali, Colombia.

Received: 25 September 2019 Accepted: 24 January 2020

Published online: 07 February 2020

References

- Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. A genetic atlas of human admixture history. *Science*. 2014;343:747–51.
- Reich D. Who we are and how we got here: Ancient DNA and the new science of the human past. Oxford: Oxford University Press; 2018.
- Crosby AW. The Columbian exchange: biological and cultural consequences of 1492. Westport: Greenwood Publishing Group; 2003.
- Mann CC. 1493: Uncovering the new world Columbus created. New York: Vintage; 2011.
- Schraiber JG, Akey JM. Methods and models for unravelling human evolutionary history. *Nat Rev Genet*. 2015;16:727–40.
- Veeramah KR, Hammer MF. The impact of whole-genome sequencing on the reconstruction of human population history. *Nat Rev Genet*. 2014;15: 149–62.
- Jordan IK. The Columbian Exchange as a source of adaptive introgression in human populations. *Biol Direct*. 2016;11:17.
- Tang H, Choudhry S, Mei R, Morgan M, Rodriguez-Cintron W, Burchard EG, Risch NJ. Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am J Hum Genet*. 2007;81:626–33.
- Rishishwar L, Conley AB, Wigington CH, Wang L, Valderrama-Aguirre A, Jordan IK. Ancestry, admixture and fitness in Colombian genomes. *Sci Rep*. 2015;5:12376.
- Zhou Q, Zhao L, Guan Y. Strong selection at MHC in Mexicans since admixture. *PLoS Genet*. 2016;12:e1005847.
- Jin W, Xu S, Wang H, Yu Y, Shen Y, Wu B, Jin L. Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Res*. 2012;22:519–27.
- Bhatia G, Tandon A, Patterson N, Aldrich MC, Ambrosone CB, Amos C, Bandera EV, Berndt SJ, Bernstein L, Blot WJ, et al. Genome-wide scan of

- 29,141 African Americans finds no evidence of directional selection since admixture. *Am J Hum Genet.* 2014;95:437–44.
13. Lopez M, Choin J, Sikora M, Siddle K, Harmant C, Costa HA, Silvert M, Mouguiama-Daouda P, Hombert JM, Froment A, et al. Genomic evidence for local adaptation of hunter-gatherers to the African rainforest. *Curr Biol.* 2019;29:2926–35 e2924.
 14. Patin E, Lopez M, Grollemund R, Verdu P, Harmant C, Quach H, Laval G, Perry GH, Barreiro LB, Froment A, et al. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science.* 2017;356:543–6.
 15. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
 16. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet.* 2013;93:278–88.
 17. Conley AB, Rishishwar L, Norris ET, Valderrama-Aguirre A, Marino-Ramirez L, Medina-Rivas MA, Jordan IK. A comparative analysis of genetic ancestry and admixture in the Colombian populations of Choco and Medellin. *G3 (Bethesda).* 2017;7:3435–47.
 18. Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, Ortiz-Tello PA, Martinez RJ, Hedges DJ, Morris RW, et al. Reconstructing the population genetic history of the Caribbean. *PLoS Genet.* 2013;9:e1003925.
 19. Homburger JR, Moreno-Estrada A, Gignoux CR, Nelson D, Sanchez E, Ortiz-Tello P, Pons-Estel BA, Acevedo-Vasquez E, Miranda P, Langefeld CD, et al. Genomic insights into the ancestry and demographic history of South America. *PLoS Genet.* 2015;11:e1005602.
 20. Ruiz-Linares A, Adhikari K, Acuna-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, Fuentes M, Pizarro M, Everardo P, de Avila F, et al. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* 2014;10:e1004572.
 21. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006;4:e72.
 22. Browning SR, Browning BL, Daviglus ML, Durazo-Arvizu RA, Schneiderman N, Kaplan RC, Laurie CC. Ancestry-specific recent effective population size in the Americas. *PLoS Genet.* 2018;14:e1007385.
 23. Norris ET, Rishishwar L, Wang L, Conley AB, Chande AT, Dabrowski AM, Valderrama-Aguirre A, Jordan IK. Assortative mating on ancestry-variant traits in admixed Latin American populations. *Front Genet.* 2019;10:359.
 24. Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sinsky JJ, Sunyaev SR, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* 2008;4:e1000083.
 25. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varily P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. Positive natural selection in the human lineage. *Science.* 2006;312:1614–20.
 26. Fay JC, Wyckoff GJ, Wu CI. Positive and negative selection on the human genome. *Genetics.* 2001;158:1227–34.
 27. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics.* 1998;148:1667–86.
 28. Norris ET, Wang L, Conley AB, Rishishwar L, Marino-Ramirez L, Valderrama-Aguirre A, Jordan IK. Genetic ancestry, admixture and health determinants in Latin America. *BMC Genomics.* 2018;19:861.
 29. Fan S, Hansen ME, Lo Y, Tishkoff SA. Going global by adapting local: a review of recent human adaptation. *Science.* 2016;354:54–9.
 30. Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol.* 2013;28:659–69.
 31. Peter BM, Huerta-Sanchez E, Nielsen R. Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genet.* 2012;8:e1003011.
 32. Racimo F, Berg JJ, Pickrel JK. Detecting polygenic adaptation in admixture graphs. *Genetics.* 2018;208:1565–84.
 33. Berg JJ, Coop G. A population genetic signal of polygenic adaptation. *PLoS Genet.* 2014;10:e1004412.
 34. Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, Genetic Investigation of ATC, Hirschhorn JN. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat Genet.* 2012;44:1015–9.
 35. Beiter ER, Khramtsova EA, Van Der Merwe C, Chimusa ER, Simonti C, Stein J, Thompson P, Fisher SE, Stein DJ, Capra JA, et al. Polygenic selection underlies evolution of human brain structure and behavioral traits. *bioRxiv.* 2017:164707. <https://www.biorxiv.org/content/10.1101/164707v1>. <https://doi.org/10.1101/164707>.
 36. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. *Nat Rev Genet.* 2014;15:379–93.
 37. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature.* 2015;526:68–74.
 38. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MH, et al. An integrated map of structural variation in 2,504 human genomes. *Nature.* 2015;526:75–81.
 39. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. Worldwide human relationships inferred from genome-wide patterns of variation. *Science.* 2008;319:1100–4.
 40. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, et al. Reconstructing Native American population history. *Nature.* 2012;488:370–4.
 41. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
 42. Jordan IK, Rishishwar L, Conley AB. Native American admixture recapitulates population-specific migration and settlement of the continental United States. *PLoS Genet.* 2019;15:e1008225.
 43. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods.* 2013;10:5–6.
 44. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* 2019;47:D853–8.
 45. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Series B (Methodological).* 1995;57:289–300.
 46. Szpiech ZA, Hernandez RD. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol.* 2014;31:2824–7.
 47. Hartl D, Clark A. Principles of population genetics. Sunderland: Sinauer Associates; 1989.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

