The genomic signatures of natural selection in admixed human populations

Authors

Sebastian Cuadros-Espinoza, Guillaume Laval,
Lluis Quintana-Murci, Etienne Patin

Correspondence

epatin@pasteur.fr (E.P.),
quintana@pasteur.fr (L.Q.-M.)

CellPress

ARTICLE

The genomic signatures of natural selection in admixed human populations

Sebastian Cuadros-Espinoza,[1,2] Guillaume Laval,[1] Lluis Quintana-Murci,[1,3,*] and Etienne Patin[1,*]

Summary

Admixture has been a pervasive phenomenon in human history, extensively shaping the patterns of population genetic diversity. There is increasing evidence to suggest that admixture can also facilitate genetic adaptation to local environments, i.e., admixed populations acquire beneficial mutations from source populations, a process that we refer to as "adaptive admixture." However, the role of adaptive admixture in human evolution and the power to detect it remain poorly characterized. Here, we use extensive computer simulations to evaluate the power of several neutrality statistics to detect natural selection in the admixed population, assuming multiple admixture scenarios. We show that statistics based on admixture proportions, F_{adm} and LAD, show high power to detect mutations that are beneficial in the admixed population, whereas other statistics, including iHS and F_{ST}, falsely detect neutral mutations that have been selected in the source populations only. By combining F_{adm} and LAD into a single, powerful statistic, we scanned the genomes of 15 worldwide, admixed populations for signatures of adaptive admixture. We confirm that lactase persistence and resistance to malaria have been under adaptive admixture in West Africans and in Malagasy, North Africans, and South Asians, respectively. Our approach also uncovers other cases of adaptive admixture, including *APOL1* in Fulani nomads and *PKN2* in East Indonesians, involved in resistance to infection and metabolism, respectively. Collectively, our study provides evidence that adaptive admixture has occurred in human populations whose genetic history is characterized by periods of isolation and spatial expansions resulting in increased gene flow.

Introduction

Over the last two decades, the search for molecular signatures of natural selection in the human genome has played an integral part in understanding human evolution and population differences in disease risk.[1–6] Genome scans for local adaptation have shed light on the environmental pressures that populations have faced for the last 100,000 years, including reduced exposure to sunlight, altitude-related hypoxia, new nutritional resources, or exposure to local pathogens. Candidate genes for local genetic adaptation have been identified on the basis of expected signatures of positive selection, such as extended haplotype homozygosity or strong differences in allele frequencies between geographically diverse populations. In doing so, selection studies have implicitly assumed that advantageous variation occurred in a single population that has remained isolated from other populations since their separation. Yet, ancient and modern genomics studies have clearly demonstrated that the last millennia of human history have been characterized by large-scale spatial expansions followed by extensive gene flow.[1,7,8] These findings indicate that most human populations descend from admixture between formerly isolated groups, highlighting the need for detailed studies of the expected genomic signatures of natural selection in admixed populations.

Several studies have searched for evidence of genetic adaptation in admixed populations as a means to detect genes under positive selection in their ancestral sources prior to admixture.[9–15] These studies showed that admixture can obscure signals of selective sweeps in the source populations and proposed approaches to alleviate this problem, such as local ancestry masking. Conversely, few studies have yet explored the patterns of diversity expected under admixture with selection as a means to detect genes under positive selection in the admixed population since admixture.[16,17] Studying the genomic signatures of "adaptive admixture," that is, positive selection in the admixed population of an allele that was beneficial in one of its ancestral sources, could shed light on the role of gene flow in spreading beneficial alleles among populations[18] and the prevalence of recent, ongoing selection in humans.

While an increasing number of studies have revealed how introgression from ancient hominins, such as Neanderthals or Denisovans, facilitated genetic adaptation in modern humans,[19] the occurrence of adaptive admixture among modern humans remains largely unexplored. Nonetheless, several empirical studies have reported candidate loci for positive selection in admixed populations.[16,20–38] A striking example is the Duffy null $FY*B^{ES}$ allele, which confers protection against *Plasmodium vivax* malaria.[39,40] Selection signals have been detected at the locus in diverse African-descent admixed populations from Madagascar, Cabo Verde, Sudan, and Pakistan,[22,29,30,32,34] suggesting strong, ongoing selection owing to *vivax* malaria in these regions. A variety of methods has been used to detect the signatures of adaptive admixture, relying on classic neutrality statistics, such as iHS or F_{ST}, and deviations from allele frequencies[22,41] or admixture proportions[21,24,27,30,31,34–38] expected under admixture and neutrality.[42] However, little

is known about how these neutrality statistics behave under scenarios of admixture with selection and, therefore, about the power of these statistics to detect adaptive admixture. More worrying, it has been suggested that artifactual signals of adaptive admixture can be observed because of errors in local ancestry inference (LAI) in complex genomic regions[43,44] and/or when the populations used as ancestral sources are poor proxies of the true source populations.[16,31] Lastly, reported signals of adaptive admixture are still limited to few populations relative to the large number of admixture events reported in humans.[1,7,8]

In this study, we compared the power of various neutrality statistics to detect adaptive admixture through computer simulations under different admixture with selection scenarios. We then used a combination of the most powerful statistics to scan the genomes of 15 different admixed human populations from around the world and detect candidate loci for adaptive admixture. In doing so, we confirm several, iconic signals of ongoing positive selection since admixture and identify cases that highlight pathogens as key drivers of recent genetic adaptation in humans.

Material and methods

General simulation settings

All the simulations were computed with the SLiM 3.2 engine[45] under the Wright-Fisher model. Each simulation consisted of a 2-Mb long locus characterized by varying recombination and mutation rates. For each simulation, we sampled the physical coordinates of a random 2-Mb genomic window in the human genome, excluding telomeric and centromeric regions, and assigned recombination rates based on the 1000 Genomes phase 3 genetic map[46] and mutation rates based on Francioli et al. mutation map.[47] To account for background selection, which is thought to be prevalent in the human genome and could affect the power of neutrality tests,[48] we simulated exon-like genetic elements positioned according to the position of exons in the sampled 2-Mb genomic window. Each simulated exon is made of positions under negative selection or under neutrality, mimicking non-synonymous and synonymous positions, respectively. We set deleterious mutations to occur three times more frequently than neutral mutations to account for codon degeneracy. The fitness effects of deleterious mutations were sampled from the gamma distribution inferred in Europeans by Boyko and colleagues.[49] For simulations that include positive selection, the beneficial mutation was set to appear in the middle of the 2-Mb simulated locus and assumed to be semi dominant. Because we used computationally intensive forward-in-time simulations, we rescaled population sizes and times according to N/λ and T/λ, with $\lambda = 10$, and used rescaled mutation, recombination, and selection parameters, $\lambda\mu$, λr, and λs.[45] Of note, we found that simulating background selection has little impact on the power to detect alleles under strong positive selection in the admixed population ($s \geq 0.05$; data not shown).

Admixture with selection models

We performed simulations of a population that originates from admixture between two source populations, referred to as P_1 and P_2 (Figure S1). We assumed that P_1 and P_2 contributed α_1 and α_2 admixture proportions to the admixed population, with $\alpha_1 + \alpha_2 = 1$. We also assumed that P_1 and P_2 diverged T_{div} generations ago and the single-pulse admixture event occurred T_{adm} generations ago. We simulated three scenarios of admixture with selection (Figures 1 and S2). For scenarios 1 and 2, a beneficial mutation was set to appear in the P_1 source population and is transmitted to the admixed population with either the same selection coefficient (scenario 1) or a selection coefficient set to 0 (scenario 2). For scenario 3, we adapted a combination of recipes 9.6.2 and 14.7 from the SLiM manual,[50] introducing a set of "ancestry marker" neutral mutations in the P_1 source population, and randomly choosing one of them to become beneficial by setting its selection coefficient to $s > 0$ in the admixed population only. We computed 500 simulations for each admixture with selection scenario, as well as 500 simulations for the null scenario (i.e., no positive selection). Because the goal of these simulations was to compare the power of neutrality statistics to detect positive selection, only the selection coefficient of the beneficial mutation s was given different values, ranging from $s = 0.01$ to $s = 0.05$. All the other parameters were given fixed values: population sizes of source and admixed populations $N = 10,000$; divergence time between source populations $T_{div} = 2,000$ generations; admixture proportions $\alpha_1 = 0.35$ and $\alpha_2 = 0.65$; time of the single pulse admixture event $T_{adm} = 70$ generations; time when the beneficial mutation appears $T_{mut} = 350$ generations ago.
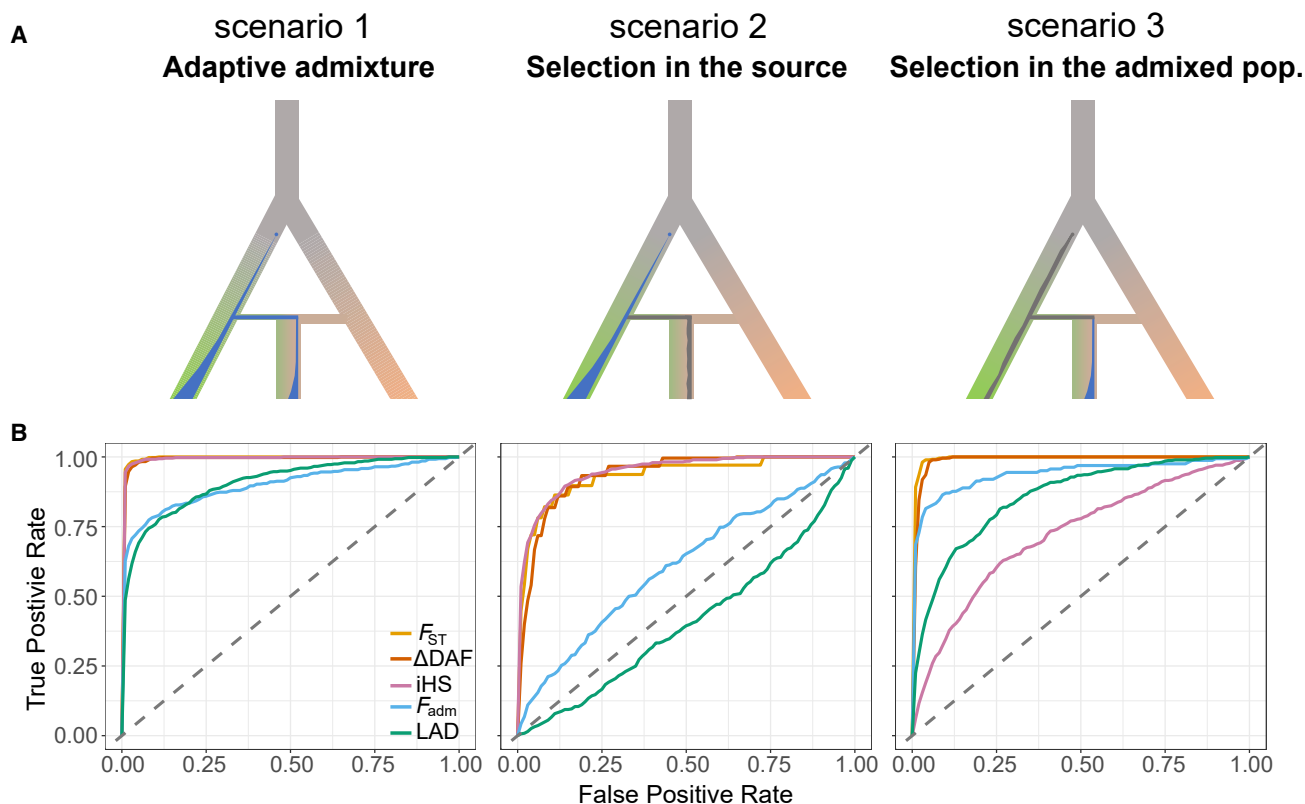
Power of explored neutrality statistics

Neutrality statistics were computed for all genetic variants within the 2-Mb simulated loci under no positive selection (H0) and only for the selected mutation for simulated 2-Mb loci under positive selection (H1). We estimated detection power (i.e., the true positive rate [TPR]) for each statistic as the proportion of values under H1 that are above a varying threshold value under H0, corresponding to a given false positive rate (FPR). We computed F_{ST}, ΔDAF, and iHS by using selink.[51] We computed F_{ST} and ΔDAF between the admixed population and the source population that does not experience positive selection. For iHS, we used a 200-kb window and normalized the values by bins of similar derived allele frequency (DAF).

For the admixture-specific statistics, we introduced an allele frequency-based statistic, F_{adm}, that measures the difference between x_i, the observed frequency of allele i in the admixed population, and y_i, the expected allele frequency under admixture and neutrality. It was shown that $y_i = \sum_p \alpha_p x_{i,p}$, which is the average of allele frequencies $x_{i,p}$ observed in the source populations p weighted by estimated admixture proportions α_p, where $\sum_p \alpha_p = 1$ (Bernstein[52]). Under neutrality, the squared difference between x_i and y_i, $(x_i - y_i)^2$, is the variance of allele frequencies in the admixed population due to genetic drift.[42] Thus, $(x_i - y_i)^2$ can be interpreted as the genetic distance between the current admixed population and its ancestral population at the time of admixture. Analogously to F_{ST}, this genetic distance can be used to detect natural selection, as the change in frequency of a beneficial allele in time depends on its selection coefficient.[53] F_{adm} is thus defined as follows:

$$F_{adm} = \frac{\sum_i (x_i - y_i)^2}{2(1 - \sum_i y_i^2)},$$

where $1 - \sum_i y_i^2$ is the expected heterozygosity in the admixed population, used here to allow comparisons among SNPs.

Figure 1. Performance of neutrality statistics under different scenarios of admixture with selection

(A) Explored scenarios of admixture with selection (from left to right): adaptive admixture, positive selection in the source population only, and positive selection in the admixed population only. The blue and gray points indicate the appearance of a new beneficial or neutral mutation, respectively. The blue and gray areas indicate changes in frequency of the beneficial and neutral mutation, respectively.

(B) Receiver operating characteristic (ROC) curves comparing the performance of classic neutrality statistics F_{ST}, iHS, and ΔDAF and the admixture-specific statistics F_{adm} and LAD across the three explored scenarios. The selection coefficient was fixed to $s = 0.05$ to highlight the differences between statistics and between models (see Figure S2 for lower s values). False positive rate (FPR) is the fraction of simulated neutral sites that are incorrectly detected as under selection, and true positive rate (TPR) is the fraction of simulated adaptive mutations that are correctly detected as under selection.

When calculating F_{adm} in the simulated and observed data, the allele frequencies $x_{i,p}$ at the time of admixture were estimated by the allele frequencies in the current generation, which is accurate when genetic drift in source populations is weak or when admixture is recent. We used as admixture proportions the simulated proportions α_{sim}, for the simulated data, and the estimated proportions $\overline{\alpha}$, for the observed data, obtained by running ADMIXTURE v.1.23 (Alexander et al.;[54] see empirical detection of adaptive admixture). We verified with simulations that errors in the estimation of admixture proportions do not affect F_{adm} detection power (Figure S3A) by computing F_{adm} with α sampled from a normal distribution $\mathcal{N}(\mu = \alpha_{sim}, \sigma^2 = 0.026^2)$; 0.026 is the highest root-mean-square deviation of the ADMIXTURE estimation.[54] Additionally, we excluded sites where the observed allele frequency in the admixed population x_i is higher (or lower) than the maximum (or minimum) of the frequencies $x_{i,p}$ in the source populations. Although this can reduce the detection power in scenario 3, this filter increases power for the adaptive admixture scenario (Figure S3B), which is the focus of this study.

We also computed an LAI-based neutrality statistic, LAD, which measures the local ancestry deviation from the average genome-wide ancestry, defined as follows:

$$LAD_{w,p} = \alpha_{w,p} - \overline{\alpha_p},$$

where $\alpha_{w,p}$ is the admixture proportion from source population p for a given window w and $\overline{\alpha_p}$ is the estimated genome-wide admixture proportion. Natural selection has been proposed to bias the estimation of admixture proportions since the first estimates of this parameter were obtained.[55–57] The rationale is that, when a beneficial allele is transmitted from a source population to the admixed population, estimated admixture proportions from this source population are expected to increase at the locus relative to neutral loci. As single-marker estimates of admixture proportions are sensitive to errors in the estimation of allele frequencies, more powerful haplotype-based methods have preferentially been used to detect natural selection since admixture.[38]

We used RFMix v1.5.4 to estimate local ancestry,[58] with default parameter values (except for –G, which was replaced with the simulated T_{adm} value) and using the forward-backward option with three expectation maximization steps. Because LAD is sensitive to phasing errors,[58] we incorporated potential phasing errors in our simulations by phasing, with SHAPEIT v.4.2.1 (Delaneau et al.[59]), unphased diploid individuals obtained from the combination of two simulated haploid individuals. Admixture proportions $\overline{\alpha_p}$
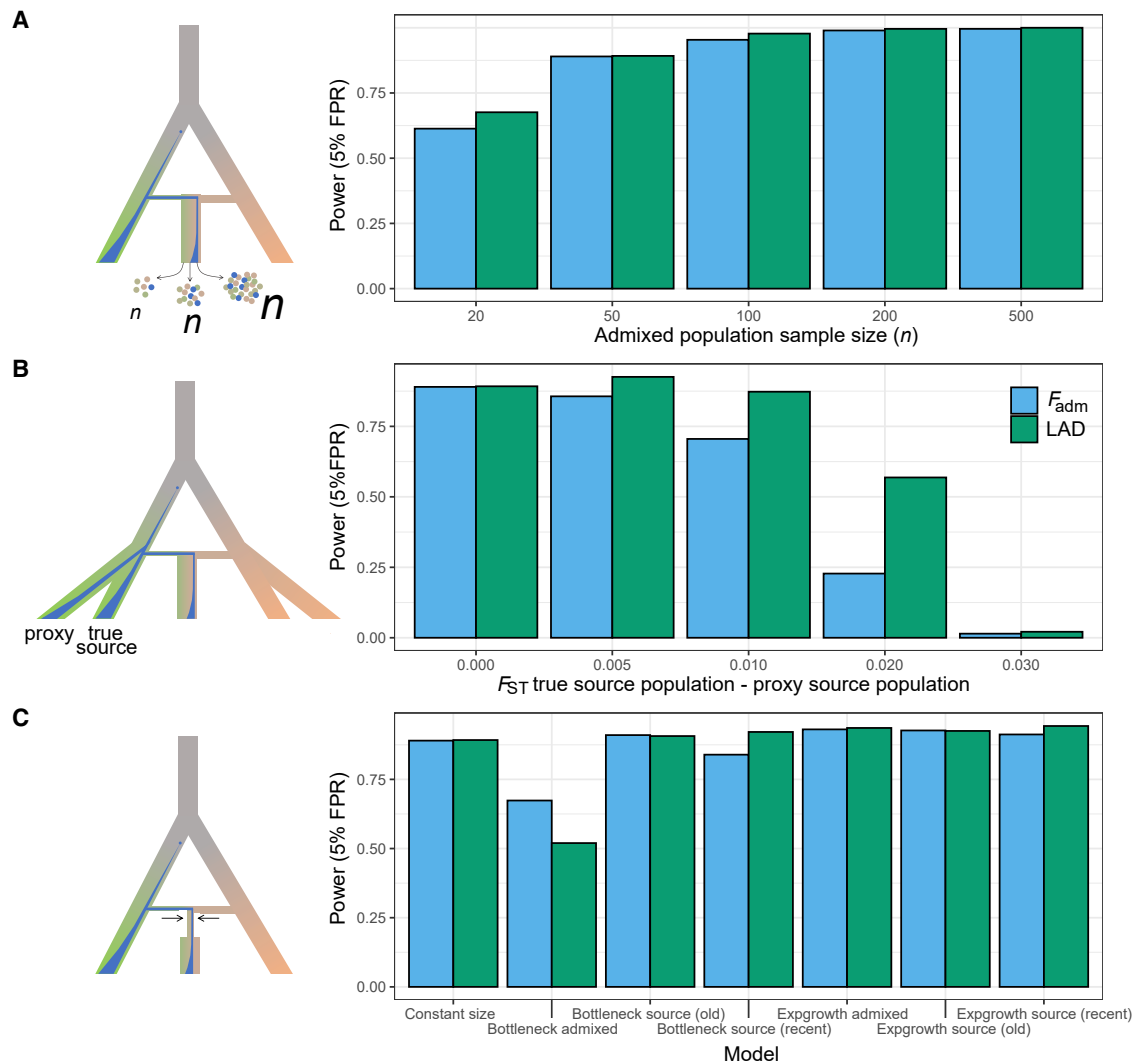
Figure 2. **Effects of study design and non-stationary demography on the power to detect adaptive admixture**
(A) Effects of the sample size of the admixed population (n) on the detection power of F_{adm} and LAD at a fixed FPR = 5%. The simulated model is shown on the left, illustrating different values of n.
(B) Effects of the use of proxy source populations on the detection power of F_{adm} and LAD at a fixed FPR = 5%. The genetic distance between the true source and its proxy was measured by F_{ST}. The simulated model is shown on the left, including the true source and its proxy.
(C) Effects of non-stationary demography on the detection power of F_{adm} and LAD at a fixed FPR = 5%. The simulated model is shown on the left, including a bottleneck in the admixed population.

were estimated as the local ancestry inferred by RFMix averaged across loci for both the simulated and observed data.

Sample size and source population choice scenarios
We explored five different values of sample sizes for the two source populations and the admixed population: $n = 20, 50, 100, 200,$ and 500 individuals (Figures 2A and S4). When exploring the values for a given population, sample size for the other two was fixed to $n = 50$ individuals. For the use of a proxy source population (Figure 2B), we simulated two additional populations that diverge 400 generations ago from each of the two source populations. We then used these proxy populations for F_{adm} and LAD calculations. To explore the effect of the genetic distance (estimated by F_{ST}) between the proxy population and the true source population on detection power, we set the population size of the proxies

to 10,000, 4,000, 1,000, and 500, resulting in F_{ST} values of 0.005, 0.01, 0.02, and 0.03, respectively.

For the scenario of selection in the proxy source population only (Figure S5), we simulated two additional populations that diverge 600 generations ago from each of the two source populations. We randomly selected a mutation that occurred in the ancestral population of the P_1 source population and its related proxy population and assigned it a selection coefficient of $s = 0.02$ in the proxy population only, 599 generations ago. Under the latter scenario, F_{adm} and LAD detect mutations that are not beneficial in the admixed population and wrongly support positive selection in the P_2 source population (Figures S5A and S5B). For comparison purposes, we thus compared F_{adm} and LAD distributions under this scenario to those obtained under a simple scenario of adaptive admixture (scenario 1, Figure 1) where the beneficial mutation is transmitted to the admixed population from the P_2 source

population (Figure S5D). In all these scenarios, the following parameters were given a fixed value: $N = 10,000$; $T_{div} = 2,000$ generations; $\alpha_1 = 0.35$; $\alpha_2 = 0.65$; $T_{adm} = 70$ generations; $s = 0.02$; and $T_{mut} = 1,400$ generations ago.

Complex admixture scenarios

We estimated detection power under two additional admixture scenarios: a double pulse model and a constant continuous model (Figures S6A and S6B). For these scenarios to be comparable to the single pulse admixture scenario, we set the sum of the admixture proportions contributed by each pulse to be equal to $\alpha_1 = 35\%$ and the average of the admixture dates to be equal to 70 generations. Namely, under the double pulse model, the admixed population originates from an admixture event that occurs 130 generations ago between two source populations, with $\alpha_1 = 17.5\%$, and receives a second admixture pulse from P_1 10 generations ago with $\alpha_1 = 17.5\%$. Under the constant continuous model, the admixed population also originates from an admixture event occurring 130 generations ago between two source populations with $\alpha_1 = 35\%/130 = 0.27\%$, when T_{adm} is not rescaled, and $a_1 = 2.7\%$, when T_{adm} is rescaled, but receives an additional pulse from P_1 of $\alpha_1 = 2.7\%$ at each generation until present. In all scenarios, the following parameters were given a fixed value: $N = 10,000$; $T_{div} = 2,000$ generations; $s = 0.02$; and $T_{mut} = 1,400$ generations ago.

Admixture parameters

Under the single pulse admixture model (Figure S1), we explored detection power as a function of different model parameters (Figures 3 and S7–S11; Table S1). In total, 32,956 compatible parameter combinations were explored. We thus reduced the number of simulations per combination from 500 to 100 to limit computational burden. For the frequency of the beneficial mutation in the source population at the time of admixture, instead of conditioning on the frequency within simulations (which would have drastically increased computations), we introduced the beneficial mutation T_{mut} generations ago in the source population on the basis of previous results.[60] For each statistic and parameter combination, we calculated the proportion of simulated sites under selection that were recovered by using a threshold of FPR = 5%. We then averaged the power across demographic parameter values to obtain a single value for each combination of T_{adm}, α_1, and s. We performed a similar procedure to obtain a single value for each combination of T_{adm}, α_1, and one of the other parameters (e.g., T_{div} and N; Figures S7–S11).

Non-stationary demography

We estimated detection power under five alternative demographic scenarios (Figures 2C and S6C), each with 500 simulations under adaptive admixture and 500 simulations with no positive selection. Demographic scenarios include: (1) a recent expansion of the source population, where the source population undergoes an expansion with a 5% growth rate since T_{adm}, from an initial $N = 10,000$; (2) a recent expansion of the admixed population, where the admixed population undergoes an expansion with a 5% growth rate since T_{adm}, from an initial $N = 10,000$; (3) an old expansion of the source population, where the source population undergoes an expansion with a 5% growth rate since T_{adm} + 500 generations, from an initial $N = 10,000$; (4) an old bottleneck in the source population, where the source population undergoes a 10-fold size reduction from T_{div} – 50 to T_{div}, from an

initial $N = 10,000$; and (5) a recent bottleneck in the admixed population, where the admixed population undergoes a 10-fold size reduction from T_{adm} – 50 to T_{adm}, from an initial $N = 10,000$. We compared these scenarios to a constant population size scenario and the size of all populations fixed to $N = 10,000$. In all scenarios, the following parameters were given a fixed value: $T_{div} = 2,000$ generations; $\alpha_1 = 0.35$; $\alpha_2 = 0.65$; $T_{adm} = 70$ generations; $s = 0.02$; and $T_{mut} = 1,400$ generations ago.

Empirical detection of adaptive admixture

We analyzed the genomes of 15 admixed populations to search for signals of adaptive admixture. The datasets and references for all admixed and source populations can be found in Table S2, as well as the final number of SNPs used after merging the datasets for admixed and source populations. For each merged dataset, we (1) excluded sites with a proportion of missing genotypes > 5% via PLINK v.2.0 (Chang et al.[61]), (2) excluded A/T and C/G variant sites, (3) excluded first- and second-degree-related individuals (kinship coefficient > 0.08 computed with KING v2.2.2; Manichaikul et al.[62]), and (4) performed phasing by using SHAPEIT v.4.2.1 with default parameter values. Additionally, we verified the validity of the admixture model for each set of source/admixed populations (Table S2) by computing admixture f_3 statistics with admixr package v.0.7.1 (Petr et al.[63]).

We obtained admixture proportions by running ADMIXTURE v.1.23, considering the K value producing the lowest cross-validation error and a set of "independent" SNPs obtained by running the "--indep-pairwise" command with PLINK v.2.0 with the following parameters: 50-SNP window, 5-SNP step, and r^2 threshold of 0.5. We verified for each studied admixed population that the K value with the fewest cross-validation errors matches the number of source populations. Local ancestry was inferred with RFMix v.1.5.4, after excluding 2 Mb at telomeres and centromeres of each chromosome, as well as invariant sites and singletons, and with default parameter values except for the generation time "-G," which was given a value based on literature (Table S2).

We combined the SNP ranks for F_{adm} and LAD statistics by using Fisher's method,[64] defined as follows:

$$X_{2k}^2 = -2 \sum_{i=1}^{k} \ln(r_i),$$

where r_i is defined as the rank of a given SNP for the statistic i, divided by the total number of analyzed SNPs (i.e., the empirical p value), and $k = 2$ is the number of statistics.

Using simulations, we verified that this statistic followed a chi-squared distribution with $2k = 4$ degrees of freedom under no positive selection (Figure 4A), including when the admixed population experienced a 10-fold bottleneck. In these simulations, we used the same parameter values as those in Figure 2C for the "constant size" and "bottleneck in the admixed population" scenarios. Statistical significance was defined based on Bonferroni correction: we considered a p value threshold of 0.05 divided by the number of 0.2-cM RFMix windows analyzed (all SNPs within the same window had the same local ancestry value), which yielded, on average, a p value threshold of 3.5×10^{-6} (Table S3). To reduce the number of false positives due to positive selection in a proxy source population only (Figure S5), we computed iHS in the two source populations and excluded from the list of candidate genes any locus that includes SNPs with both an |iHS| > 2 in one source population and an excess of local ancestry from the other source population (Figure S5). To
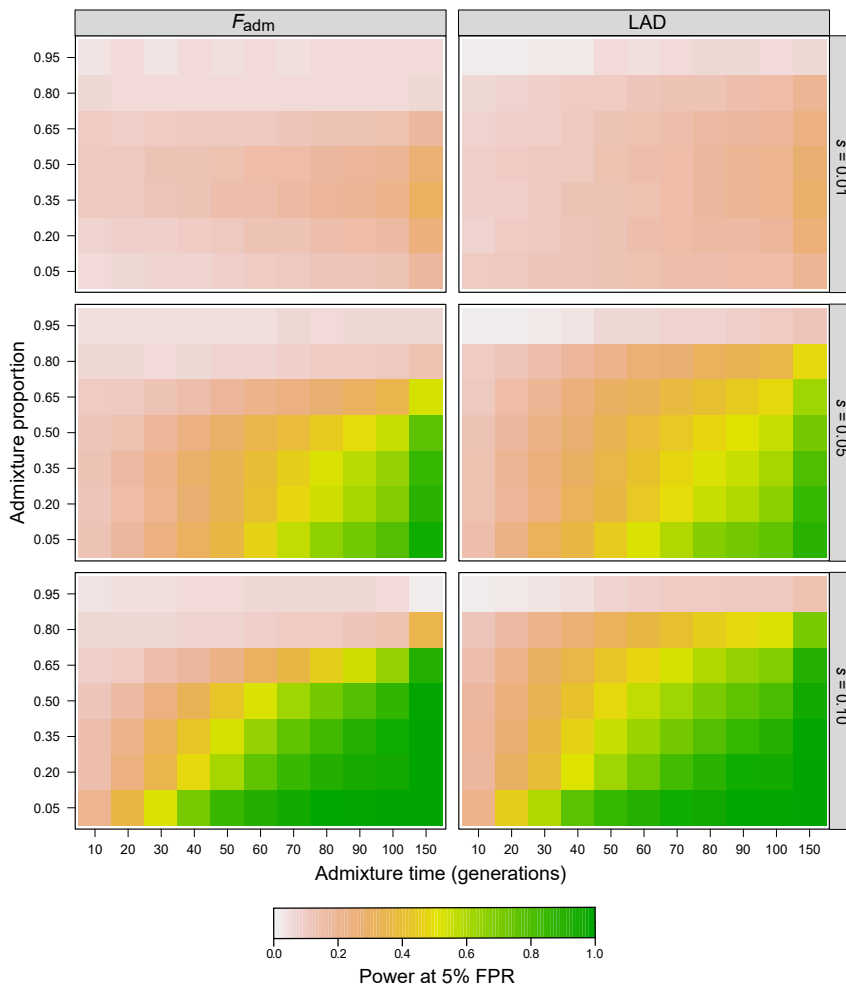
Figure 3. Effects of model parameters on the power to detect adaptive admixture
Color represents average detection power for a fixed FPR = 5% across parameter combinations. The effects of other parameters, such as population sizes and divergence time (Figure S1 and Table S1), are shown in Figures S7–S11.

annotate the different signals that passed this threshold, we chose the protein-coding gene within 250 kb of the variant with the highest V2G score.[65]

Results

Power estimation under different models of admixture with selection

To estimate the power to detect positive selection in admixed populations, we performed extensive forward-in-time simulations of a population that originates from admixture between two source populations (Figure S1). We introduced a beneficial mutation in one of the source populations, with a varying selection coefficient (material and methods). We considered three different scenarios of admixture with selection (Figures 1A and S2). Scenario 1 corresponds to adaptive admixture, where the admixed population inherits an allele that is beneficial in one of its source populations: the mutation is under positive selection in the source population, is transmitted to the admixed population, and remains beneficial—with the same selection coefficient—in the admixed population. In scenario 2, the beneficial allele is under positive selection in

the source population, is transmitted to the admixed population, and becomes neutral in the admixed population only. We simulated this scenario to verify whether some neutrality statistics wrongly support positive selection in the admixed population because of a residual signal inherited from the source population. At the same time, this scenario is also useful for evaluating the power to detect residual signals of positive selection in the admixed population, as a means to detect genes under positive selection in source populations that no longer exist in an unadmixed form.[9–15] Finally, in scenario 3, a neutral mutation in the source population becomes beneficial in the admixed population only, at the time of admixture. This case is used for determining how neutrality statistics behave when natural selection operates since admixture on standing neutral variation.

We evaluated the performance, under each scenario, of three classic neutrality statistics, F_{ST}, ΔDAF, and iHS, as well as two statistics that are specifically designed to detect selection in the admixed population: F_{adm}, which is proportional to the squared difference between the observed and the expected allele frequency in the admixed population,[22,42,57] and LAD, the difference between the admixture proportion at the locus and its genome-wide average,[38] estimated on the basis of local ancestry inference (LAI) by RFMix (material and methods).[58] Receiver operating characteristic (ROC) curves indicate that both the classic neutrality statistics and F_{adm} and LAD are powerful to detect adaptive admixture (scenario 1) when the selection coefficient $s = 0.05$ (>70% detection power for a false positive rate [FPR] of 5%; Figures 1B and S2), in agreement with a previous study.[16] Nevertheless, the power of F_{ST}, ΔDAF, and iHS is also high when the mutation is beneficial in the source population and is no longer selected in the admixed population (scenario 2), indicating that these statistics wrongly detect selection in the source population as selection in the admixed population. In contrast, F_{adm} and LAD detect adaptive admixture specifically, as their power under scenario 2 is low or nil (Figure 1B). Of note, our simulations also imply that the power of classic statistics is substantial when using the
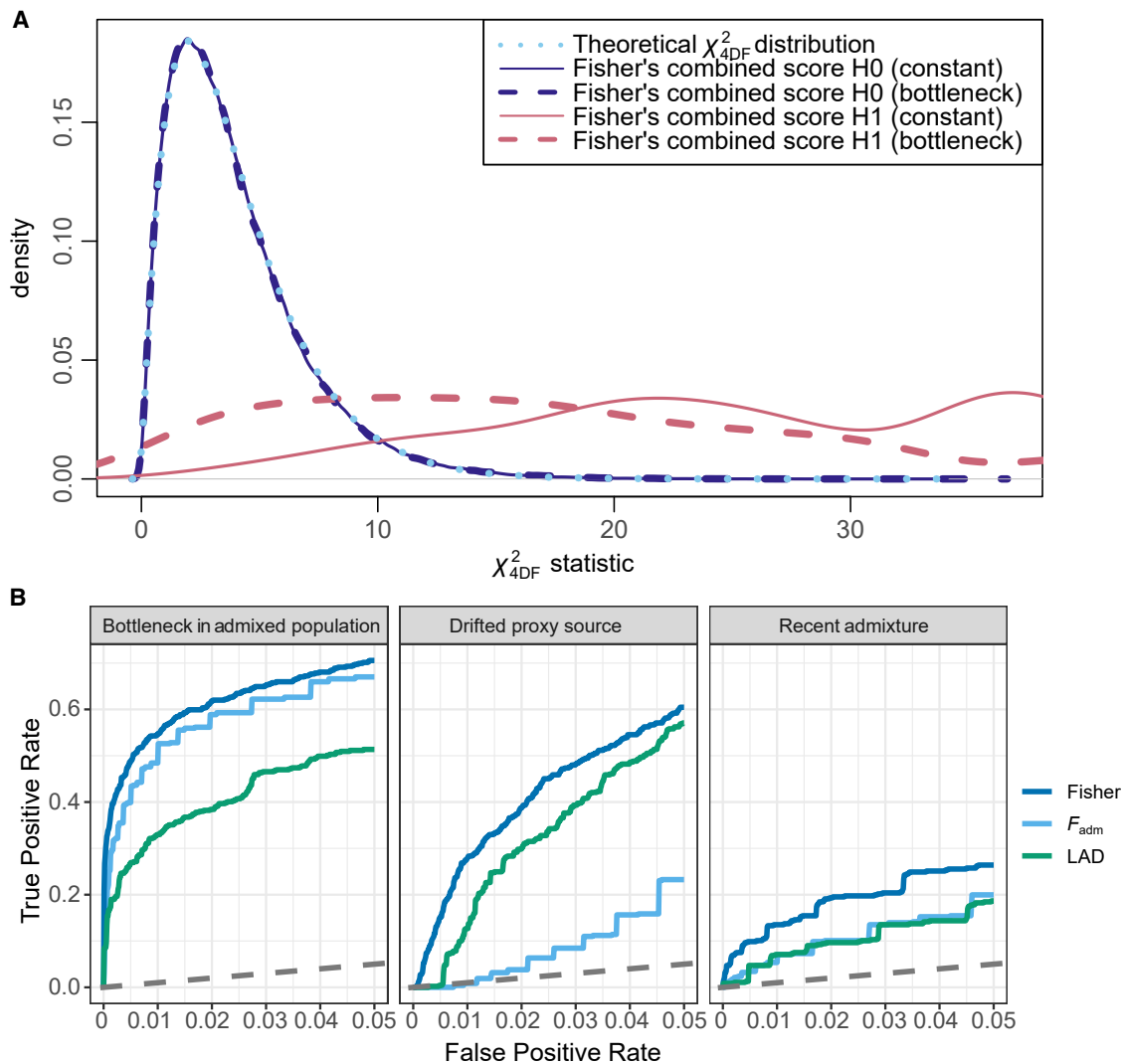
Figure 4. Performance of the Fisher's method to detect adaptive admixture

(A) Distributions of the combined Fisher's score under the null hypothesis of no positive selection (H0, blue lines) and under adaptive admixture (H1, pink lines), compared to the theoretical chi-squared distribution with 4 degrees of freedom (dotted light blue line). Solid and dashed lines indicate distributions under a constant population size and a 10-fold bottleneck in the admixed population, respectively.

(B) ROC curves for F_{adm}, LAD, and the combined Fisher's score under unfavorable scenarios for detecting adaptive admixture: a 10-fold bottleneck introduced in the admixed population, the use of a proxy source population having experienced strong drift (F_{ST} between the true source and proxy populations of 0.02), and recent admixture ($T_{adm} = 10$ generations). Only FPR < 5% are shown.

admixed population as a means to detect selection in the source populations (>65% detection power when $s = 0.05$ and FPR = 5%). Finally, LAD and iHS showed a reduced power to detect selection in the admixed population when the mutation is neutral in the source populations (scenario 3), relative to the adaptive admixture case (scenario 1). This may stem from the fact that, under scenario 3, the beneficial mutation has been selected for fewer generations than in scenario 1, resulting in a weaker signal. Furthermore, this scenario is similar to selection on standing variation, where the adaptive mutation may be present on several haplotypes, making it harder to detect.[66]

Collectively, our simulations indicate that F_{adm} and LAD are the only studied statistics that have substantial power

to specifically detect strong, ongoing selection in the admixed population and have more power to detect adaptive admixture than post-admixture selection on standing neutral variation. Because our objective is to detect the signatures of positive selection in the admixed population, and not in the source populations, we based all subsequent analyses on the F_{adm} and LAD statistics.

Effects of the study design

We investigated how sample size and the choice of source populations affect the power of F_{adm} and LAD to detect adaptive admixture signals (material and methods). We explored sample sizes ranging from $n = 20$ to $n = 500$ for both the admixed and the source populations. We found

that $n = 100$ already provides optimal power because the variance of neutrality statistics is virtually unchanged when $n \geq 100$ (Figures 2A and S4A). Conversely, we found that when $n < 50$, sampling error increases the variance of F_{adm} and LAD null distributions by as much as 5 times and ultimately decreases detection power by up to 40% (FPR = 5%). Interestingly, LAD detection power is not affected when the sample size of the source populations is low, even when $n = 20$ (Figure S4B). Consistently, RFMix accuracy was shown to be only minimally reduced when the sample size of reference panels is as small as $n = 3$, as it uses both source and admixed individuals for LAI.[58]

Because obtaining genotype data for the true source populations of an admixed population is difficult, if not impossible, population geneticists often use related, present-day populations as proxies, which may lead to false adaptive admixture signals.[16,31] We explored how detection power is affected by the genetic distance between the true source population and a related population used as a proxy for F_{adm} and LAD computations (material and methods). We observed a difference in performance between F_{adm} and LAD, the latter's being more robust to the use of a proxy (Figure 2B). LAD maintains similar detection power even if the divergence between the true and proxy populations is $F_{ST} = 0.01$, whereas power decreases by 25% for F_{adm}. Such a difference in power may result from the nature of the two statistics. In the case of F_{adm}, the expected allele frequency is directly estimated from the allele frequencies observed in the proxy, and these frequencies are decreasingly correlated with those in the true source population as their divergence increases. On the other hand, LAD is derived from LAI by RFMix, which has been shown to be robust to the use of proxy reference populations.[58]

Nonetheless, we identified a potentially problematic scenario for both F_{adm} and LAD involving population proxies: when the selection event occurs specifically in the proxy source population (i.e., the mutation is not selected in both the true source and the admixed populations; Figure S5), spurious deviations in local ancestry and in allele frequencies were observed in the admixed population. Specifically, this generates an excess of local ancestry from the other source population and expected allele frequencies higher than those observed in the admixed population (Figures S5A and S5B). We found that this scenario produces weaker LAD values (i.e., lower detection power) but larger F_{adm} values (i.e., higher detection power) relative to an adaptive admixture event (Figures S5C–S5F). To remediate this, we performed a selection scan in the proxy population by using a single-population statistic, iHS, and excluded the top 1% values. In doing so, we managed to exclude approximately 90% of the outlier values of F_{adm} and LAD generated by this scenario. More importantly, because there is no correlation between iHS in the proxy population and F_{adm} or LAD in the case of adaptive admixture, none of the outlier values generated by a true adap-

tive admixture event were excluded by this analysis step (Figures S5G and S5H).

Effects of the admixture model and non-stationary demography

Several studies have shown that admixture in humans has often involved multiple admixture pulses from two or more source populations.[8,51,67–71] We thus estimated the detection performance of F_{adm} and LAD under admixture models that are more complex than the single admixture pulse. We found that the power to detect adaptive admixture is only moderately reduced under a two-pulse admixture model or a constant, continuous admixture model: the true positive rate (TPR) decreases by <11% at a FPR = 5%, relative to the single pulse model (Figures S6A and S6B). This suggests that our power estimations are valid for a variety of admixture models.

Assuming a single-pulse admixture model, we then explored how detection power is impacted by key parameters of the adaptive admixture model, including the strength of selection s, the admixture time T_{adm}, the admixture proportion α, and the divergence time between source populations T_{div} (Figures 3 and S7–S11; Table S1). As expected, we found that detection power is high only when the selection coefficient s is strong; the TPR is up to 94% and 27% when $s = 0.05$ and 0.01, respectively (FPR = 5%; Figure 3). Power is also determined by the admixture time T_{adm}, as it affects the duration of selection; the TPR is up to 94% and 21% when $T_{adm} \geq 70$ and ≤ 20 generations, respectively. Interestingly, we observed that the higher the admixture proportion α (from the source population where the selected mutation appeared), the lower the detection power. Power decreases particularly when $\alpha > 0.65$, probably because of a threshold effect: if the beneficial allele is at high frequency and, e.g., $\alpha = 0.9$, there is little room for the observed allele frequency or local ancestry to deviate from its expectation, making it hard to detect. Finally, as the divergence time between source populations decreases, the detection power of LAD is reduced by ~15% (51% versus 67% when $T_{div} = 500$ or 2,000 generations, respectively; Figure S7), whereas that of F_{adm} is not affected (59% versus 55%, when $T_{div} = 500$ or 2,000 generations, respectively). The reduced detection power of LAD is probably due to the decreased accuracy of RFMix when T_{div} decreases.[72]

We also estimated power under scenarios where demography deviates from a constant population size model. Indeed, demographic events, such as bottlenecks, have been shown to alter the performance of several neutrality statistics.[73–79] We simulated five demographic scenarios, including 10-fold bottlenecks and 5% growth rate expansions in either the admixed or the source populations (material and methods). We found that detection power is minimally affected under all expansion models (TPR decrease of 5% at a FPR = 5%; Figure 2C). In contrast, detection power is reduced by as much as 50% under the scenario where a 10-fold bottleneck is introduced in the

admixed population, relative to the stationary model. This is mirrored by the increased variance of F_{adm} and LAD null distributions under this scenario (Figure S6C). Finally, detection power of both F_{adm} and LAD is minimally affected when the 10-fold bottleneck is introduced in the source populations, either few generations after their divergence or before the admixture pulse (TPR decrease of 5% at a FPR = 5%; Figure 2C), suggesting that both statistics are relatively robust to increased genetic drift occurring in the source populations.

Empirical detection of adaptive admixture in humans

We next sought to detect candidate genes for adaptive admixture in humans by scanning, with both F_{adm} and LAD statistics, the genomes of 15 worldwide populations (Table S2) that have experienced at least one admixture event in the last 5,000 years (i.e., the upper detection limit set for accurate local ancestry inference[80]). To improve detection power and facilitate candidate prioritization, we combined the empirical p values of both statistics with Fisher's method,[64] used here as a combined test for positive selection since admixture. We confirmed with simulations that the Fisher's score follows a chi-squared distribution with 4 degrees of freedom under the null hypothesis of absence of positive selection and when assuming different demographic scenarios (Figure 4A). Consistently, we found that F_{adm} and LAD statistics are not correlated under the null hypothesis (Spearman's coefficient = 0.03), whereas they are correlated under adaptive admixture (Spearman's coefficient = 0.96). Importantly, we found that Fisher's method increases detection power under unfavorable scenarios, relative to each individual statistic (Figure 4B). In particular, Fisher's method improves power when the admixed population experienced a 10-fold bottleneck, when admixture is recent ($T_{adm} = $ 10 generations), or when using a proxy population that experienced strong drift (F_{ST} with the true source population = 0.02). Given the limited knowledge on the past population sizes of the studied populations, which could increase FPR (Figure S6C), we applied a conservative Bonferroni correction on Fisher's p values, considering the number of RFMix genomic windows as the number of tests (all SNPs within a given window have the same value for LAD). This yielded a p value threshold of approximately p = 3.5×10^{-6} (Table S3). Finally, we verified that the empirical distribution of Fisher's p values is uniform in all studied populations and found an excess of low p values for several populations (Figure S12), suggesting that adaptive admixture has occurred in these groups.

Our genome scans identified a number of previously reported signals of adaptive admixture. Among these, we found the HLA class II locus in Bantu-speaking populations from Gabon[31] (Figures 5A and 5C; top ranking SNP identified in HLA-DPA1 [MIM: 142880]; p = 7.9×10^{-8}; expected frequency of 0.33 versus observed frequency of 0.70), the HLA class I locus in Mexicans[27,35,37,81] (Figure S13; top ranked SNP identified in ABCF1; p = 2.2×10^{-6}; expected

frequency of 0.013 versus observed frequency of 0.039), the lactase persistence-associated LCT/MCM6 locus (MIM: 223100) in the Fulani nomads of Burkina Faso[82] (Figure 6A; top ranked SNP identified in CCNT2 [MIM: 603862]; p = 1.1×10^{-6}; expected frequency of 0.12 versus observed frequency of 0.47), and ACKR1 (previously referred as DARC [MIM: 613665]) in African-descent populations from Madagascar, the Sahel, and Pakistan[29,30,34] (Figures 5B and 5D and S13). For the latter locus, the top-ranking variant is rs12075 in the Malagasy (p = 3.4×10^{-9}; expected frequency 0.45 versus observed frequency of 0.93), as previously found.[30] This variant, also known as the Duffy null $FY*B^{ES}$ allele (MIM: 110700), confers resistance against Plasmodium vivax infection in sub-Saharan Africans.[39,40] Together, these results confirm that our conservative approach can recover strong, well-documented signals of adaptive admixture.

Candidate genes for adaptive admixture

We found several candidate loci for adaptive admixture (Figures 6 and S14), among which was the MYH9/APOL1 (MIM: 603743) locus in the Fulani (Figures 6A and 6C; p = 1.3×10^{-7}; top ranked SNP in IFT27 [MIM: 615870]; expected frequency of 0.15 versus observed frequency of 0.45). Common APOL1 variants confer both protection against human African trypanosomiasis (HAT, or sleep sickness) and susceptibility to common kidney diseases (MIM: 612551) in African-descent individuals.[83] Another candidate is the PKN2 (MIM: 602549) locus in East Indonesians (p = 1.1×10^{-6}; top ranked SNP in ZNF326 [MIM: 614601]; expected frequency of 0.27 versus observed frequency of 0.46), which shows a large excess of Papuan ancestry (Figures 6B and 6D). PKN2 plays a role in cellular signal transduction responses and has been reported as involved in the regulation of glucose metabolism in skeletal muscle.[84] A nearby locus, LRRC8B (MIM: 612888), has been reported as a candidate for positive selection in Solomon Islanders,[51] although it did not show signals for adaptive admixture in this population. A unique, strong signal was detected at the ARRDC4/IGF1R (MIM: 147370) locus in Solomon Islanders (p = 7.4×10^{-9}; top ranked SNP close to ARRDC4; expected frequency of 0.09 versus observed frequency of 0.58), where an excess of East Asian-related ancestry was observed (Figures S14B and S14F). This locus was previously identified as a candidate for positive selection in near and western remote Oceanians.[51] ARRDC4 is an arrestin that plays important roles in glucose metabolism and immune response to enterovirus infection,[85] whereas IGF1R, the receptor for the insulin-like growth factor, is a key determinant of body size and growth.[86,87] A last example is CXCL13 (MIM: 605149) in the Nama pastoralists from South Africa (Figures S14A and S14E; p = 2.3×10^{-6}; top ranked SNP identified in CXCL13; expected frequency of 0.51 versus observed frequency of 0.80). The CNOT6L/CXCL13 locus has previously been reported as suggestively associated with tuberculosis (TB) risk in South African populations with San ancestry.[88] However, we found that the top-ranking variants
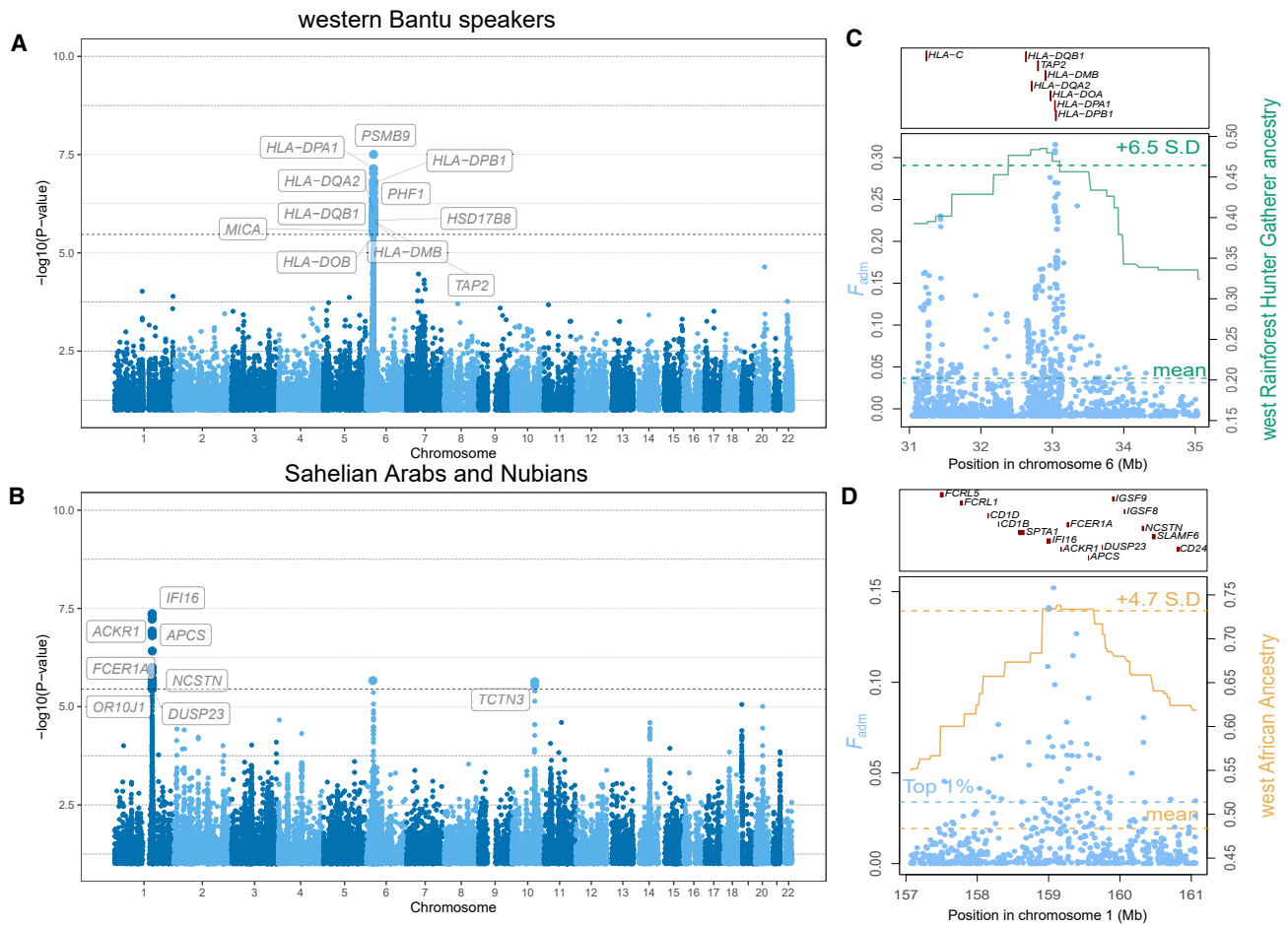
Figure 5. Iconic genomic signals of adaptive admixture

(A) Genome-wide signals of adaptive admixture in Bantu-speaking populations from Gabon.

(B) Genome-wide signals of adaptive admixture in Sahelian Arabs and Nubians.

(A and B) Highlighted blue points indicate variants that passed the Bonferroni significance threshold (shown by a horizontal dotted line). Gene labels were attributed on the basis of the gene with the highest V2G score within 250 kb of the candidate variant. p values were obtained assuming that combined Fisher's scores follow a chi-squared distribution under no selection.

(C) Local signatures of adaptive admixture for the *HLA* region in Bantu-speaking populations from Gabon.

(D) Local signatures of adaptive admixture for the *ACKR1* region in Sahelian Arabs and Nubians.

(C and D) Light blue points indicate F_{adm} values for individual variants. The green and gold solid lines indicate average local ancestry from African rainforest hunter-gatherers and West Africans respectively.

show outlier extended haplotype homozygosity in the Ju/'hoansi San, used as source population (iHS = −3.12), while European ancestry is in excess at the locus in the Nama, suggesting a spurious signal due to positive selection in the proxy source population (Figure S5).

Lastly, we detected suggestive signals of adaptive admixture at genes shown to be strong candidates for positive selection, including the *MCM6/LCT* locus in the Bantu-speaking Bakiga of Uganda (Figure S15; p = 4.3 × 10⁻⁶; top ranked SNP in *CCNT2*; expected frequency of 0.15 versus observed frequency of 0.31) and *TNFAIP3* (MIM: 191163) in East Indonesians, who show an excess of Papuan-related ancestry at the locus (Figure 6B; p = 5.0 × 10⁻⁶; top ranked SNP in *TNFAIP3*; expected frequency of 0.27 versus observed frequency of 0.43). The *TNFAIP3* locus has not only been reported as evolving under positive selection in Papuans[51] but also as adaptively

introgressed from Denisovans.[51,89–91] TNFAIP3 plays an important role in human immune tolerance to pathogen infections.[92] Collectively, these results indicate that adaptive admixture has occurred in various admixed populations around the world and highlight the immune system and nutrient metabolism as important targets of recent genetic adaptation.

Discussion

In this study, we evaluated the power of several neutrality statistics to detect loci under positive selection in admixed populations and used these statistics to explore cases of adaptive admixture in the genomes of 15 worldwide human populations. Although F_{adm} and LAD, or closely related statistics based on the difference between observed and expected allele frequencies and admixture
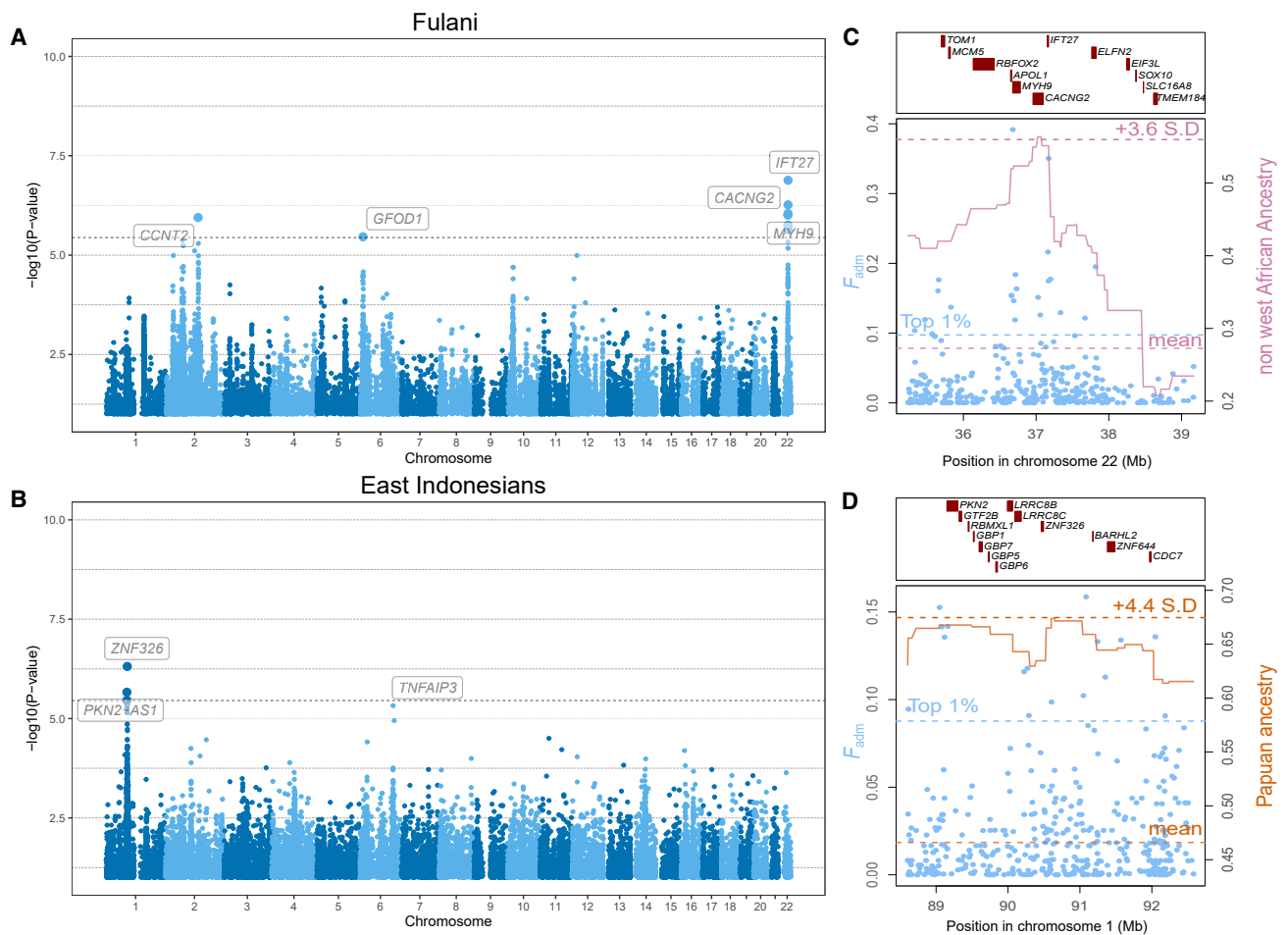
Figure 6. Genomic signals of adaptive admixture discovered in this study

(A) Genome-wide signals of adaptive admixture in the Fulani nomads of West Africa.

(B) Genome-wide signals of adaptive admixture in East Indonesians.

(A and B) Highlighted blue points indicate variants that pass the Bonferroni significance threshold (shown by a horizontal dotted line). Gene labels were attributed on the basis of the gene with the highest V2G score within 250 kb of the candidate variant. p values were obtained assuming that combined Fisher's scores follow a chi-squared distribution under no selection.

(C) Local signatures of adaptive admixture for the *IFT27/MYH9/APOL1* region in the Fulani nomads.

(D) Local signatures of adaptive admixture for the *PKN2/LRR8CB* region in East Indonesians.

(C and D) Light blue points indicate F_{adm} values for individual variants. The pink solid line indicates the local ancestry from Europeans and North Africans, and the orange solid line indicates the local ancestry from Papuans.

proportions, have been used in several empirical studies, their power has not been thoroughly evaluated. Here, we showed that these statistics are powerful to detect adaptive admixture and have little power to detect residual signals of positive selection in the source populations. Thus, F_{adm} and LAD are suited to search for loci under positive selection in admixed populations since admixture, particularly when selection is strong (i.e., $s \geq 0.05$), admixture is relatively old (i.e., $T_{adm} > 2,000$ years) and the admixture proportion is moderate-to-low (i.e., $\alpha < 0.6$). Notably, we found that power is marginally affected when admixture has been recurrent, a feature that is convenient given the difficulty to distinguish between single-pulse, double-pulse, or more complex admixture models from the genetic data.[8,51,67–71] Furthermore, F_{adm} is more powerful than LAD when selection occurs in the admixed population only and when the divergence time between source

populations is low ($T_{div} = 500$ generations), whereas LAD is more powerful than F_{adm} when source sample sizes are low (i.e., $n = 20$) and when the true and proxy source populations are distantly related (i.e., $F_{ST} \geq 0.01$; Table S4). The latter result is consistent with the known robustness of LAI to cases where the populations used as reference sources are poor proxies of the true source populations.[58] Nonetheless, caution must be taken when handling population proxies, as selection occurring only in the proxy population can produce artifactual genomic signals, for both LAD and F_{adm}, that might be misinterpreted as adaptive admixture.[16,31,51] We suggest that performing selection scans on the proxy source populations can help distinguish false from true adaptive admixture signals. We also caution that F_{adm} calculation relies on the accurate estimation of admixture proportions, which can be biased under certain scenarios.[93] Finally, we found that

combining F_{adm} and LAD statistics into a unique statistic, based on the Fisher's method, provides well-calibrated p values under different models and substantially increases power under several realistic admixture with selection scenarios, relative to individual statistics.

When applying this combined method on the empirical data, we identified several previously reported candidate variants for adaptive admixture. These include the *ACKR1* Duffy null allele detected in admixed populations from Madagascar,[30] the Sahel,[34] and Pakistan,[29] the lactase persistence c.−13910C>T *LCT* allele in the Fulani from West Africa,[82] and *HLA* alleles in Bantu-speaking populations from western Central Africa[31] and Mexicans.[27,35,37,81] These candidate loci were detected previously on the basis of LAD only or in combination with classic neutrality statistics. However, the detection of natural selection with the LAD statistic has previously been questioned because deviations in local ancestry can be explained as artifacts of long-range linkage disequilibrium (LD), which was not properly modeled by the first-generation LAI methods.[43] Our analyses reveal that these genomic regions not only show outlier LAD values but also outlier F_{adm} values. Because F_{adm} only depends on allele frequencies at the SNP of interest, these results support the view that the observed signals of adaptive admixture are true and unlikely to be explained by incorrectly modeled LD.

Our results also highlight novel signals of adaptive admixture, such as the *APOL1/MYH9* locus in the Fulani nomads of West Africa. Interestingly, an *APOL1* haplotype of non-African origin, named G3, was shown to be under positive selection in the Fulani of Cameroon,[94] in line with the excess of non-African ancestry that we detected at the locus in the Fulani from Burkina Faso. Nevertheless, the physiological effect of the G3 variants is still debated: experimental work suggests that the G3 haplotype has no lytic activity against *Trypanosoma* parasites and is not associated with increased susceptibility to common kidney diseases in African Americans.[95] Alternatively, the significant excess of non-African ancestry observed at the locus may be due to strong negative selection against HAT-resistance *APOL1* alleles (i.e., G1 and G2 haplotypes) in regions where the incidence of sleeping sickness is low, such as Burkina Faso.[96] As they do not confer a selective advantage in *Trypanosoma brucei*-free regions, the G1 and G2 haplotypes only strongly increase the risk for chronic kidney diseases[83] and thus become disadvantageous. Further epidemiological and experimental work will be needed to confirm this hypothesis.

In accordance with our simulation study, several of the putatively selected alleles detected here are known to be under strong positive selection in humans, including alleles in *ACKR1*,[97–99] *LCT*,[100,101] or *HLA*.[81] Given that we focused on admixture events occurring during the five last millennia, only alleles that confer a very strong selective advantage can leave detectable signatures in the genomes of the studied admixed individuals. In addition to their confirmatory nature, these results improve our un-

derstanding of the selective advantage conferred by these well-known beneficial alleles. First, because F_{adm} and LAD detect natural selection since admixture only, selection studies in recently admixed populations represent a valuable tool to detect recent ongoing selection. Second, admixed and source populations have often lived in different environments, so evolutionary studies of adaptive admixture can help refine correlations between signatures of natural selection and environmental pressures. An illustrative example is the Duffy null $FY*B^{ES}$ allele, which is fixed or nearly fixed in most sub-Saharan African populations.[99] It has long been proposed that natural selection has favored this allele because it protects against malaria due to *Plasmodium vivax*.[102] Indeed, cellular experiments have shown that the parasite depends on the ACKR1 protein for erythrocytic infection.[39,40] However, recent studies have casted doubt on this result because *P. vivax* has been detected in $FY*B^{ES}$ homozygous carriers,[103,104] suggesting that parasite invasion is possible when its human receptor ACKR1 is absent. We and others have found signatures of adaptive admixture for the $FY*B^{ES}$ allele in African-descent admixed populations from Madagascar,[22,30] Cabo Verde,[32] the Sahel,[34] and Pakistan[29] but not in North Americans or South Africans.[16,31] Evidence of ongoing positive selection for Duffy negativity is thus confined to regions where the current incidence of *P. vivax* malaria is estimated to be high.[105] These findings thus support the view that resistance to *vivax* malaria is the main evolutionary force driving the frequency of the $FY*B^{ES}$ allele in humans.

Overall, our study reports evidence that recent admixture has facilitated human genetic adaptation to varying environmental conditions. It has been proposed that gene flow can promote rapid evolution when the demographic structure of a species is unstable.[18] Our findings support this view, as *Homo sapiens* is a structured species that has settled a large variety of ecological niches and has undergone large-scale, massive dispersals followed by extensive gene flow.[1,7,8] We thus anticipate that more cases of adaptive admixture in humans will soon be uncovered, thanks to methodological and technological advances. Importantly, given the highly conservative nature of our approach, it is very likely that we do not recover variants that have probably been weakly to mildly selected since admixture, such as *TNFAIP3* in Indonesian populations of Papuan-related ancestry[51,89–91] or the *MCM6/LCT* locus in the Bantu-speaking Bakiga from Uganda.[31] The use of new, accurate LAI methods[80,106] and the development of novel powerful neutrality statistics, such as the integrated decay in ancestry tracts (iDAT),[32] and model-based probabilistic frameworks[107] are promising paths to improve the power to detect adaptive admixture while better accounting for the demography of admixed populations. Furthermore, many human traits are known to be highly polygenic, suggesting that polygenic adaptation is a key driver of phenotypic evolution,[108] highlighting the need for new methods to detect polygenic selection since admixture.[109] Finally, genomic studies of adaptive

admixture are expected to be more powerful when admixture is ancient, but statistical tests for admixture in modern genomes have low power when admixture time is older than 5,000 years.[8] Ancient genomics studies offer a great opportunity to circumvent this limitation by revealing how human populations interacted in the past and how beneficial alleles have spread in time and space.[41,110]

Data and code availability

Accession numbers for the genotype data used in this study are listed in Table S2. All SLiM parameter files can be found at the following: https://github.com/h-e-g/ADAD.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2022.02.011.

Web resources

1000 Genomes Phase 3 and HGDP genomic data, https://www.internationalgenome.org/data
admixr R package, https://cran.r-project.org/web/packages/admixr/index.html
ADMIXTURE software, https://dalexander.github.io/admixture/index.html
dbGAP database, https://dbgap.ncbi.nlm.nih.gov/
Estonian Biocenter public genomic data, https://evolbio.ut.ee
European Genome-Phenome archive, https://ega-archive.org/
Jakobsson Lab genomic data, http://jakobssonlab.iob.uu.se/data/
OMIM, http://www.omim.org/
PLINK software, https://www.cog-genomics.org/plink/2.0/
RFMix software, https://www.dropbox.com/s/cmq4saduh9gozi9/RFMix_v1.5.4.zip
selink software, https://github.com/h-e-g/selink
SHAPEIT software, https://odelaneau.github.io/shapeit4/
SLiM software, https://messerlab.org/slim/

References

1. Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S., and Willerslev, E. (2017). Tracing the peopling of the world through genomics. Nature 541, 302–310.

2. Novembre, J., and Di Rienzo, A. (2009). Spatial patterns of variation due to natural selection in humans. Nat. Rev. Genet. 10, 745–755.

3. Quintana-Murci, L. (2019). Human Immunology through the Lens of Evolutionary Genetics. Cell 177, 184–199.

4. Rees, J.S., Castellano, S., and Andrés, A.M. (2020). The Genomics of Human Local Adaptation. Trends Genet. 36, 415–428.

5. Fan, S., Hansen, M.E.B., Lo, Y., and Tishkoff, S.A. (2016). Going global by adapting local: A review of recent human adaptation. Science 354, 54–59.

6. Mathieson, I. (2020). Human adaptation over the past 40,000 years. Curr. Opin. Genet. Dev. 62, 97–104.

7. Pickrell, J.K., and Reich, D. (2014). Toward a new history and geography of human genes informed by ancient DNA. Trends Genet. 30, 377–389.

8. Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. (2014). A genetic atlas of human admixture history. Science 343, 747–751.

9. Johnson, N.A., Coram, M.A., Shriver, M.D., Romieu, I., Barsh, G.S., London, S.J., and Tang, H. (2011). Ancestral components of admixed genomes in a Mexican cohort. PLoS Genet. 7, e1002410.

10. Vicuña, L., Fernandez, M.I., Vial, C., Valdebenito, P., Chaparro, E., Espinoza, K., Ziegler, A., Bustamante, A., and Eyheramendy, S. (2019). Adaptation to Extreme Environments in an Admixed Human Population from the Atacama Desert. Genome Biol. Evol. 11, 2468–2479.

11. Yelmen, B., Mondal, M., Marnetto, D., Pathak, A.K., Montinaro, F., Gallego Romero, I., Kivisild, T., Metspalu, M., and Pagani, L. (2019). Ancestry-Specific Analyses Reveal Differential Demographic Histories and Opposite Selective Pressures in Modern South Asian Populations. Mol. Biol. Evol. 36, 1628–1642.

12. Lohmueller, K.E., Bustamante, C.D., and Clark, A.G. (2011). Detecting directional selection in the presence of recent admixture in African-Americans. Genetics 187, 823–835.

13. Reynolds, A.W., Mata-Míguez, J., Miró-Herrans, A., Briggs-Cloud, M., Sylestine, A., Barajas-Olmos, F., Garcia-Ortiz, H., Rzhetskaya, M., Orozco, L., Raff, J.A., et al. (2019). Comparing signals of natural selection between three Indigenous North American populations. Proc. Natl. Acad. Sci. USA 116, 9312–9317.

14. Ávila-Arcos, M.C., McManus, K.F., Sandoval, K., Rodríguez-Rodríguez, J.E., Villa-Islas, V., Martin, A.R., Luisi, P., Peñaloza-Espinosa, R.I., Eng, C., Huntsman, S., et al. (2020). Population History and Gene Divergence in Native Mexicans

Inferred from 76 Human Exomes. Mol. Biol. Evol. *37*, 994–1006.

15. Huerta-Sánchez, E., Degiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., Cardona, A., Montgomery, H.E., Cavalleri, G.L., Robbins, P.A., et al. (2013). Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. Mol. Biol. Evol. *30*, 1877–1888.

16. Bhatia, G., Tandon, A., Patterson, N., Aldrich, M.C., Ambrosone, C.B., Amos, C., Bandera, E.V., Berndt, S.I., Bernstein, L., Blot, W.J., et al. (2014). Genome-wide scan of 29,141 African Americans finds no evidence of directional selection since admixture. Am. J. Hum. Genet. *95*, 437–444.

17. Refoyo-Martínez, A., da Fonseca, R.R., Halldórsdóttir, K., Árnason, E., Mailund, T., and Racimo, F. (2019). Identifying loci under positive selection in complex population histories. Genome Res. *29*, 1506–1520.

18. Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. Science *236*, 787–792.

19. Racimo, F., Sankararaman, S., Nielsen, R., and Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. Nat. Rev. Genet. *16*, 359–371.

20. Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., et al. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am. J. Hum. Genet. *91*, 83–96.

21. Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., Hammer, M., Bustamante, C.D., and Ostrer, H. (2010). Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proc. Natl. Acad. Sci. USA *107* (*Suppl 2*), 8954–8961.

22. Hodgson, J.A., Pickrell, J.K., Pearson, L.N., Quillen, E.E., Prista, A., Rocha, J., Soodyall, H., Shriver, M.D., and Perry, G.H. (2014). Natural selection for the Duffy-null allele in the recently admixed people of Madagascar. Proc. Biol. Sci. *281*, 20140930.

23. Breton, G., Schlebusch, C.M., Lombard, M., Sjödin, P., Soodyall, H., and Jakobsson, M. (2014). Lactase persistence alleles reveal partial East African ancestry of southern African Khoe pastoralists. Curr. Biol. *24*, 852–858.

24. Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D.B., Pritchard, J.K., Beall, C.M., and Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. Nat. Commun. *5*, 3281.

25. Macholdt, E., Lede, V., Barbieri, C., Mpoloka, S.W., Chen, H., Slatkin, M., Pakendorf, B., and Stoneking, M. (2014). Tracing pastoralist migrations to southern Africa with lactase persistence alleles. Curr. Biol. *24*, 875–879.

26. Rishishwar, L., Conley, A.B., Wigington, C.H., Wang, L., Valderrama-Aguirre, A., and Jordan, I.K. (2015). Ancestry, admixture and fitness in Colombian genomes. Sci. Rep. *5*, 12376.

27. Zhou, Q., Zhao, L., and Guan, Y. (2016). Strong Selection at MHC in Mexicans since Admixture. PLoS Genet. *12*, e1005847.

28. Busby, G., Christ, R., Band, G., Leffler, E., Le, Q.S., Rockett, K., Kwiatkowski, D., and Spencer, C. (2017). Inferring adaptive gene-flow in recent African history. Preprint at bioRxiv. https://doi.org/10.1101/205252.

29. Laso-Jadart, R., Harmant, C., Quach, H., Zidane, N., Tyler-Smith, C., Mehdi, Q., Ayub, Q., Quintana-Murci, L., and Patin, E. (2017). The Genetic Legacy of the Indian Ocean Slave Trade: Recent Admixture and Post-admixture Selection in the Makranis of Pakistan. Am. J. Hum. Genet. *101*, 977–984.

30. Pierron, D., Heiske, M., Razafindrazaka, H., Pereda-Loth, V., Sanchez, J., Alva, O., Arachiche, A., Boland, A., Olaso, R., Deleuze, J.-F., et al. (2018). Strong selection during the last millennium for African ancestry in the admixed population of Madagascar. Nat. Commun. *9*, 932.

31. Patin, E., Lopez, M., Grollemund, R., Verdu, P., Harmant, C., Quach, H., Laval, G., Perry, G.H., Barreiro, L.B., Froment, A., et al. (2017). Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. Science *356*, 543–546.

32. Hamid, I., Korunes, K.L., Beleza, S., and Goldberg, A. (2021). Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. eLife *10*, e63177.

33. Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M.G.B., et al. (2012). Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science *338*, 374–379.

34. Triska, P., Soares, P., Patin, E., Fernandes, V., Cerny, V., and Pereira, L. (2015). Extensive Admixture and Selective Pressure Across the Sahel Belt. Genome Biol. Evol. *7*, 3484–3495.

35. Deng, L., Ruiz-Linares, A., Xu, S., and Wang, S. (2016). Ancestry variation and footprints of natural selection along the genome in Latin American populations. Sci. Rep. *6*, 21766.

36. Jin, W., Xu, S., Wang, H., Yu, Y., Shen, Y., Wu, B., and Jin, L. (2012). Genome-wide detection of natural selection in African Americans pre- and post-admixture. Genome Res. *22*, 519–527.

37. Norris, E.T., Rishishwar, L., Chande, A.T., Conley, A.B., Ye, K., Valderrama-Aguirre, A., and Jordan, I.K. (2020). Admixture-enabled selection for rapid adaptive evolution in the Americas. Genome Biol. *21*, 29.

38. Tang, H., Choudhry, S., Mei, R., Morgan, M., Rodriguez-Cintron, W., Burchard, E.G., and Risch, N.J. (2007). Recent genetic selection in the ancestral admixture of Puerto Ricans. Am. J. Hum. Genet. *81*, 626–633.

39. Miller, L.H., Mason, S.J., Clyde, D.F., and McGinniss, M.H. (1976). The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. N. Engl. J. Med. *295*, 302–304.

40. Tournamille, C., Colin, Y., Cartron, J.P., and Le Van Kim, C. (1995). Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat. Genet. *10*, 224–228.

41. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

42. Long, J.C. (1991). The genetic structure of admixed populations. Genetics *127*, 417–428.

43. Price, A.L., Weale, M.E., Patterson, N., Myers, S.R., Need, A.C., Shianna, K.V., Ge, D., Rotter, J.I., Torres, E., Taylor, K.D., et al. (2008). Long-range LD can confound genome scans in admixed populations. Am. J. Hum. Genet. *83*, 132–135, author reply 135–139.

44. Pasaniuc, B., Sankararaman, S., Torgerson, D.G., Gignoux, C., Zaitlen, N., Eng, C., Rodriguez-Cintron, W., Chapela, R., Ford, J.G., Avila, P.C., et al. (2013). Analysis of Latino populations from GALA and MEC studies reveals genomic loci

with biased local ancestry estimation. Bioinformatics *29*, 1407–1415.

45. Haller, B.C., and Messer, P.W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright-Fisher Model. Mol. Biol. Evol. *36*, 632–637.

46. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

47. Francioli, L.C., Polak, P.P., Koren, A., Menelaou, A., Chun, S., Renkens, I., van Duijn, C.M., Swertz, M., Wijmenga, C., van Ommen, G., et al. (2015). Genome-wide patterns and properties of de novo mutations in humans. Nat. Genet. *47*, 822–826.

48. Lohmueller, K.E., Albrechtsen, A., Li, Y., Kim, S.Y., Korneliussen, T., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Feder, A.F., Grarup, N., et al. (2011). Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. PLoS Genet. *7*, e1002326.

49. Boyko, A.R., Williamson, S.H., Indap, A.R., Degenhardt, J.D., Hernandez, R.D., Lohmueller, K.E., Adams, M.D., Schmidt, S., Sninsky, J.J., Sunyaev, S.R., et al. (2008). Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. *4*, e1000083.

50. Haller, B.C., and Messer, P.W. (2016). SLiM: An Evolutionary Simulation Framework. http://benhaller.com/slim/SLiM_Manual.pdf.

51. Choin, J., Mendoza-Revilla, J., Arauna, L.R., Cuadros-Espinoza, S., Cassar, O., Larena, M., Ko, A.M.-S., Harmant, C., Laurent, R., Verdu, P., et al. (2021). Genomic insights into population history and biological adaptation in Oceania. Nature *592*, 583–589.

52. Bernstein, F. (1931). Die geographische Verteilung der Blutgruppen und ihre anthropologische Bedeutung. In Comitato Weak Representations of the Data. Italiano per Lo Studio Dei Problemi Della Popolazione (Roma: Istituto Poligrafico dello Stato), pp. 227–243.

53. Charlesworth, B., and Charlesworth, D. (2010). Elements of Evolutionary Genetics (W. H. Freeman).

54. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

55. Workman, P.L., Blumberg, B.S., and Cooper, A.J. (1963). Selection, Gene Migration and Polymorphic Stability in a U. S. White and Negro Population. Am. J. Hum. Genet. *15*, 429–437.

56. Reed, T.E. (1969). Caucasian genes in American Negroes. Science *165*, 762–768.

57. Cavalli-Sforza, L.L., and Bodmer, W.F. (1971). The Genetics of Human Populations (Freeman & Co).

58. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–288.

59. Delaneau, O., Zagury, J.-F., Robinson, M.R., Marchini, J.L., and Dermitzakis, E.T. (2019). Accurate, scalable and integrative haplotype estimation. Nat. Commun. *10*, 5436.

60. Stephan, W., Wiehe, T.H.E., and Lenz, M.W. (1992). The effect of strongly selected substitutions on neutral polymorphism: Analytical results based on diffusion theory. Theor. Popul. Biol. *41*, 237–254.

61. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

62. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.

63. Petr, M., Vernot, B., and Kelso, J. (2019). admixr-R package for reproducible analyses using ADMIXTOOLS. Bioinformatics *35*, 3194–3195.

64. Fisher, R.A. (1925). Statistical Methods for Research Workers (Edinburgh: Oliver and Boyd).

65. Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M., Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A., et al. (2021). Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res. *49* (D1), D1311–D1320.

66. Peter, B.M., Huerta-Sanchez, E., and Nielsen, R. (2012). Distinguishing between selective sweeps from standing variation and from a de novo mutation. PLoS Genet. *8*, e1003011.

67. Baharian, S., Barakatt, M., Gignoux, C.R., Shringarpure, S., Errington, J., Blot, W.J., Bustamante, C.D., Kenny, E.E., Williams, S.M., Aldrich, M.C., and Gravel, S. (2016). The Great Migration and African-American Genomic Diversity. PLoS Genet. *12*, e1006059.

68. Fortes-Lima, C.A., Laurent, R., Thouzeau, V., Toupance, B., and Verdu, P. (2021). Complex genetic admixture histories reconstructed with Approximate Bayesian Computation. Mol. Ecol. Resour. *21*, 1098–1117.

69. Malaspinas, A.-S., Westaway, M.C., Muller, C., Sousa, V.C., Lao, O., Alves, I., Bergström, A., Athanasiadis, G., Cheng, J.Y., Crawford, J.E., et al. (2016). A genomic history of Aboriginal Australia. Nature *538*, 207–214.

70. Medina, P., Thornlow, B., Nielsen, R., and Corbett-Detig, R. (2018). Estimating the Timing of Multiple Admixture Pulses During Local Ancestry Inference. Genetics *210*, 1089–1107.

71. Pickrell, J.K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. Proc. Natl. Acad. Sci. USA *111*, 2632–2637.

72. Molinaro, L., Marnetto, D., Mondal, M., Ongaro, L., Yelmen, B., Lawson, D.J., Montinaro, F., and Pagani, L. (2021). A Chromosome-Painting-Based Pipeline to Infer Local Ancestry under Limited Source Availability. Genome Biol. Evol. *13*, evab025.

73. Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.

74. Tajima, F. (1989). The effect of change in population size on DNA polymorphism. Genetics *123*, 597–601.

75. Wakeley, J., and Aliacar, N. (2001). Gene genealogies in a metapopulation. Genetics *159*, 893–905.

76. Fu, Y.X., and Li, W.H. (1993). Statistical tests of neutrality of mutations. Genetics *133*, 693–709.

77. Przeworski, M. (2002). The signature of positive selection at randomly chosen loci. Genetics *160*, 1179–1189.

78. Coop, G., Pickrell, J.K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R.M., Cavalli-Sforza, L.L., Feldman, M.W., and Pritchard, J.K. (2009). The role of geography in human adaptation. PLoS Genet. *5*, e1000500.

79. Ferrer-Admetlla, A., Liang, M., Korneliussen, T., and Nielsen, R. (2014). On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol. Biol. Evol. 31, 1275–1291.

80. Dias-Alves, T., Mairal, J., and Blum, M.G.B. (2018). Loter: A Software Package to Infer Local Ancestry for a Wide Range of Species. Mol. Biol. Evol. 35, 2318–2326.

81. Meyer, D., C Aguiar, V.R., Bitarello, B.D.C., C Brandt, D.Y., and Nunes, K. (2018). A genomic perspective on HLA evolution. Immunogenetics 70, 5–27.

82. Vicente, M., Priehodová, E., Diallo, I., Podgorná, E., Poloni, E.S., Černý, V., and Schlebusch, C.M. (2019). Population history and genetic adaptation of the Fulani nomads: inferences from genome-wide data and the lactase persistence trait. BMC Genomics 20, 915.

83. Genovese, G., Friedman, D.J., Ross, M.D., Lecordier, L., Uzureau, P., Freedman, B.I., Bowden, D.W., Langefeld, C.D., Oleksyk, T.K., Uscinski Knob, A.L., et al. (2010). Association of trypanolytic ApoL1 variants with kidney disease in African Americans. Science 329, 841–845.

84. Ruby, M.A., Riedl, I., Massart, J., Åhlin, M., and Zierath, J.R. (2017). Protein kinase N2 regulates AMP kinase signaling and insulin responsiveness of glucose metabolism in skeletal muscle. Am. J. Physiol. Endocrinol. Metab. 313, E483–E491.

85. Meng, J., Yao, Z., He, Y., Zhang, R., Zhang, Y., Yao, X., Yang, H., Chen, L., Zhang, Z., Zhang, H., et al. (2017). ARRDC4 regulates enterovirus 71-induced innate immune response by promoting K63 polyubiquitination of MDA5 through TRIM65. Cell Death Dis. 8, e2866.

86. Wit, J.M., and Walenkamp, M.J. (2013). Role of insulin-like growth factors in growth, development and feeding. World Rev. Nutr. Diet. 106, 60–65.

87. Warrington, N.M., Beaumont, R.N., Horikoshi, M., Day, F.R., Helgeland, Ø., Laurin, C., Bacelis, J., Peng, S., Hao, K., Feenstra, B., et al. (2019). Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. Nat. Genet. 51, 804–814.

88. Chimusa, E.R., Zaitlen, N., Daya, M., Möller, M., van Helden, P.D., Mulder, N.J., Price, A.L., and Hoal, E.G. (2014). Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum. Mol. Genet. 23, 796–809.

89. Vernot, B., Tucci, S., Kelso, J., Schraiber, J.G., Wolf, A.B., Gittelman, R.M., Dannemann, M., Grote, S., McCoy, R.C., Norton, H., et al. (2016). Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science 352, 235–239.

90. Gittelman, R.M., Schraiber, J.G., Vernot, B., Mikacenic, C., Wurfel, M.M., and Akey, J.M. (2016). Archaic Hominin Admixture Facilitated Adaptation to Out-of-Africa Environments. Curr. Biol. 26, 3375–3382.

91. Jacobs, G.S., Hudjashov, G., Saag, L., Kusuma, P., Darusallam, C.C., Lawson, D.J., Mondal, M., Pagani, L., Ricaut, F.-X., Stoneking, M., et al. (2019). Multiple Deeply Divergent Denisovan Ancestries in Papuans. Cell 177, 1010–1021.e32.

92. Zammit, N.W., Siggs, O.M., Gray, P.E., Horikawa, K., Langley, D.B., Walters, S.N., Daley, S.R., Loetsch, C., Warren, J., Yap, J.Y., et al. (2019). Denisovan, modern human and mouse TNFAIP3 alleles tune A20 phosphorylation and immunity. Nat. Immunol. 20, 1299–1310.

93. Toyama, K.S., Crochet, P.-A., and Leblois, R. (2020). Sampling schemes and drift can bias admixture proportions inferred by structure. Mol. Ecol. Resour. 20, 1769–1785.

94. Ko, W.-Y., Rajan, P., Gomez, F., Scheinfeldt, L., An, P., Winkler, C.A., Froment, A., Nyambo, T.B., Omar, S.A., Wambebe, C., et al. (2013). Identifying Darwinian selection acting on different human APOL1 variants among diverse African populations. Am. J. Hum. Genet. 93, 54–66.

95. Limou, S., Nelson, G.W., Lecordier, L., An, P., O'hUigin, C.S., David, V.A., Binns-Roemer, E.A., Guiblet, W.M., Oleksyk, T.K., Pays, E., et al. (2015). Sequencing rare and common APOL1 coding variants to determine kidney disease risk. Kidney Int. 88, 754–763.

96. Franco, J.R., Simarro, P.P., Diarra, A., and Jannin, J.G. (2014). Epidemiology of human African trypanosomiasis. Clin. Epidemiol. 6, 257–275.

97. Hamblin, M.T., and Di Rienzo, A. (2000). Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am. J. Hum. Genet. 66, 1669–1679.

98. Hamblin, M.T., Thompson, E.E., and Di Rienzo, A. (2002). Complex signatures of natural selection at the Duffy blood group locus. Am. J. Hum. Genet. 70, 369–383.

99. McManus, K.F., Taravella, A.M., Henn, B.M., Bustamante, C.D., Sikora, M., and Cornejo, O.E. (2017). Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. PLoS Genet. 13, e1006560.

100. Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., et al. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. Nat. Genet. 39, 31–40.

101. Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., Rhodes, M., Reich, D.E., and Hirschhorn, J.N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. Am. J. Hum. Genet. 74, 1111–1120.

102. Livingstone, F.B. (1984). The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. Hum. Biol. 56, 413–425.

103. Ménard, D., Barnadas, C., Bouchier, C., Henry-Halldin, C., Gray, L.R., Ratsimbasoa, A., Thonier, V., Carod, J.-F., Domarle, O., Colin, Y., et al. (2010). Plasmodium vivax clinical malaria is commonly observed in Duffy-negative Malagasy people. Proc. Natl. Acad. Sci. USA 107, 5967–5971.

104. Popovici, J., Roesch, C., and Rougeron, V. (2020). The enigmatic mechanisms by which Plasmodium vivax infects Duffy-negative individuals. PLoS Pathog. 16, e1008258.

105. Battle, K.E., Lucas, T.C.D., Nguyen, M., Howes, R.E., Nandi, A.K., Twohig, K.A., Pfeffer, D.A., Cameron, E., Rao, P.C., Casey, D., et al. (2019). Mapping the global endemicity and clinical burden of Plasmodium vivax, 2000-17: a spatial and temporal modelling study. Lancet 394, 332–343.

106. Guan, Y. (2014). Detecting structure of haplotypes and local ancestry. Genetics 196, 625–642.

107. Sugden, L.A., Atkinson, E.G., Fischer, A.P., Rong, S., Henn, B.M., and Ramachandran, S. (2018). Localization of adaptive variants in human genomes using averaged one-dependence estimation. Nat. Commun. 9, 703.

108. Sella, G., and Barton, N.H. (2019). Thinking About the Evolution of Complex Traits in the Era of Genome-Wide

Association Studies. Annu. Rev. Genomics Hum. Genet. *20*, 461–493.

109. Racimo, F., Berg, J.J., and Pickrell, J.K. (2018). Detecting Polygenic Adaptation in Admixture Graphs. Genetics *208*, 1565–1584.

110. Dehasque, M., Ávila-Arcos, M.C., Díez-Del-Molino, D., Fumagalli, M., Guschanski, K., Lorenzen, E.D., Malaspinas, A.-S., Marques-Bonet, T., Martin, M.D., Murray, G.G.R., et al. (2020). Inference of natural selection from ancient DNA. Evol. Lett. *4*, 94–108.