African and Non-African Admixture Components in African Americans and An African Caribbean Population

Tanda Murray[1,*], **Terri H. Beaty**[1], **Rasika A. Mathias**[2], **Nicholas Rafaels**[2], **Audrey Virginia Grant**[3], **Mezbah U. Faruque**[4], **Harold R. Watson**[5], **Ingo Ruczinski**[6], **Georgia M. Dunston**[4], and **Kathleen C. Barnes**[1,2]

[1]Johns Hopkins School of Public Health, Epidemiology, Baltimore, Maryland [2]Johns Hopkins School of Medicine, Baltimore, Maryland [3]INSERM/Université Paris Descartes, Human Genetics of Infectious Diseases, Paris, France [4]National Human Genome Center at Howard University, Washington, District of Columbia [5]University of the West Indies, Cave Hill, Barbados [6]Johns Hopkins School of Public Health, Biostatistics, Baltimore, Maryland

Abstract

Admixture is a potential source of confounding in genetic association studies, so it becomes important to detect and estimate admixture in a sample of unrelated individuals. Populations of African descent in the US and the Caribbean share similar historical backgrounds but the distributions of African admixture may differ. We selected 416 ancestry informative markers (AIMs) to estimate and compare admixture proportions using STRUCTURE in 906 unrelated African Americans (AAs) and 294 Barbadians (ACs) from a study of asthma. This analysis showed AAs on average were 72.5% African, 19.6% European and 8% Asian, while ACs were 77.4% African, 15.9% European, and 6.7% Asian which were significantly different. A principal components analysis based on these AIMs yielded one primary eigenvector that explained 54.04% of the variation and captured a gradient from West African to European admixture. This principal component was highly correlated with African vs. European ancestry as estimated by STRUCTURE ($r^2 = 0.992$, $r^2 = 0.912$, respectively). To investigate other African contributions to African American and Barbadian admixture, we performed PCA on ~14,000 (14k) genome-wide SNPs in AAs, ACs, Yorubans, Luhya and Maasai African groups, and estimated genetic distances (F_{ST}). We found AAs and ACs were closest genetically ($F_{ST} = 0.008$), and both were closer to the Yorubans than the other East African populations. In our sample of individuals of African descent, ~400 well-defined AIMs were just as good for detecting substructure as ~14,000 random SNPs drawn from a genome-wide panel of markers.

Keywords

admixture; African Americans; African Caribbeans; African ancestry; genetic distance

INTRODUCTION

Genetic substructure (cryptic or recognized) induced by variation in geographic locale and ethno-cultural background within the sample can create confounding leading to spurious

*Correspondence to: Tanda Murray, Johns Hopkins School of Public Health, Epidemiology, 615 N Wolfe St, Suite W6517, Baltimore, Maryland. tmurray@jhsph.edu.

association between markers and the outcome in case-control studies, and can limit generalizability of study findings. If the history of each population involves one or two ancestral populations, then self-reported ethnicity may be sufficient for classification, and stratifying by race/ethnic group could prevent problems [Ziv and Burchard, 2003]. The scenario becomes complex when more than one ancestral population has contributed to the current gene pool, however.

For admixed groups such as African Americans, self-reported ethnicity is an insufficient criterion for stratification, because they are known to carry alleles from West African and European ancestral groups [Chakraborty et al., 1992; Parra et al., 2001]. The average West African and European components have been estimated to be ~80 and ~20%, respectively [Parra et al, 2004]. African Caribbean populations follow a similar pattern of genetic ancestry. A previous study by Benn-Torres et al. [2008] estimated Barbadians at near 90% West African ancestry; however, their sample comprised UK residents and included only 28 ancestry informative markers (AIMs). In analyses of substructure in more than 900 Baltimore/Washington African Americans and almost 300 Barbadians, we used ~400 AIMs, and more than 14,000 randomly selected SNPs from a genome-wide marker panel and employed both a Bayesian technique [STRUCTURE, Pritchard et al., 2002], and principal components analysis (PCA).

Although West Africans and Europeans are assumed to be the primary genetic contributors to African Americans and African Caribbeans, many studies have ignored other potential non-African components such as Amerindian and East Asian, as well as non-West African components. This is due in part to difficulty in locating proxies for the ancestral populations. Goncalves et al. [2008] showed Brazilian blacks have a considerable proportion of Southeast African ancestry (12%) based on analysis of mitochondrial markers. Although ~71% of the African ancestry among African Americans can be attributed to West African populations, other African groups account for at least 8% of the African ancestry [Tishkoff et al., 2009]. This could also be true for other populations in the African Diaspora, and failure to estimate these proportions might contribute to misclassification of individuals and a failure to appropriately correct for stratification in tests for association. In the current study, we considered other non-African components by including HapMap Han Chinese and Japanese samples [International HapMap Consortium, 2003] as a proxy ancestral population. The historical background of Africans in the Diaspora does include some degree of admixing with New World populations, i.e., Native Americans, which may be reflected in an admixture analysis. However, genotypic data on Native American populations are not easily accessible. Additionally, we also explored the genetic relationship between our African Diaspora samples (African Americans and Barbadians) with HapMap Phase II West African (Yorubans) and Phase III East African (Maasai and Luhya) populations.

Populations of African descent in the United States and the Caribbean are known to have relatively high prevalences of certain chronic diseases such as asthma and prostate cancer [Barnes, 2006; Marcella et al., 2001; Wong et al., 2009]. As research moves forward in elucidating genetic risks factors for such diseases in non-European populations, it becomes more important to accurately account for potential confounding due to genetic substructure.

METHODS

STUDY SAMPLE

Subjects were drawn from the "Genomic Research on Asthma in the African Diaspora" (GRAAD) consortium, an NHLBI funded project to perform a genome-wide association for asthma in populations of African descent. The African American case-control arm of this study included 447 unrelated asthma cases and 459 unrelated, unaffected controls

ascertained from eight separate studies in the Baltimore/Washington, DC area, as previously described [Mathias et al., 2010]. The African Caribbean arm of this study comprises 272 Barbadian founders from a family study of the genetics of asthma previously described by Barnes et al. [1996]. HapMap Phase II Yoruban (YRI) founders, CEPH founders (CEU) and unrelated Han Chinese and Japanese (CHB/JPT) populations were used as proxies for "continental" ancestral populations [International HapMap Consortium, 2003]. HapMap Phase III unrelated individuals from these same populations (YRI, CEU) plus two additional African populations (i.e. the Luhya of Kenya (LWK) and the Maasai (MKK) also from Kenya were also used as reference populations.

GENOTYPING

All SNPs were genotyped in a genome-wide array of 665,352 SNPs on the Illumina HumanHap650y Beadchip for the GRAAD study [Mathias et al., 2010].

MARKER SELECTION

To investigate substructure, two sets of markers were selected. First, 416 AIMs were selected from autosomal markers on the 650Y panel based on a published list of AIMs [Cheng et al., 2009] showing large differences in allele frequencies (Delta ( )>0.45) between YRI and CEU. Second, the full panel of 622,262 autosomal SNPs were mined using PLINK [v1.05, Purcell et al., 2007] to identify ~20,000 SNPs with pairwise $r^2<0.15$ in the GRAAD African American sample and having <5% missing data. From this larger set of autosomal SNPs, a total of 14,629 were found in HapMap Phase III subjects (YRI, CEU, LWK, MKK), and genotypes on these HapMap subjects were downloaded from HapMart (http://hapmart.hapmap.org/BioMart/martview), an extension of the HapMap data resource. All SNPs were tested for Hardy-Weinberg equilibrium (HWE) among African American and Barbadian GRAAD subjects. Linkage disequilibrium (LD) structures of these 416 AIMs and random SNP markers were assessed for all five populations (African Americans, Barbadians, CEU, YRI and CHB/JPT) using Haploview [Barrett et al., 2005] and PLINK.

ESTIMATION OF AFRICAN ANCESTRY

We estimated admixture for each individual using STUCTURE [v2.0, Pritchard et al., 2002] on the 416 AIMs for African American and the Barbadian GRAAD samples. HapMap CEU, YRI and the pooled Asian samples (CHB/JPT) were used as ancestral reference populations, assuming both African Americans and Barbadians could carry alleles derived from any of the three populations. The admixture model was assumed, with default priors for and , k specified as 3, and executed using 10,000 burn-ins and 50,000 iterations of the Markov Chain Monte Carlo algorithm. The Wilcoxon Rank Sum test was used to test the null hypothesis of no difference in the proportion of African or European ancestry between African Americans and Barbadians as implemented in R (v2.6.2).

PRINCIPAL COMPONENTS ANALYSIS

PCA were performed on our data using the 416 AIMs with and without HapMap CEU, YRI and CHB/JPT. Likewise, PCA was repeated on the data using the 14k random SNPs with and without the HapMap Africans (YRI, LKK and MKK). In the analysis of African Americans, Barbadians and HapMap populations using AIMs, we first estimated eigenvectors using HapMap and then projected our data unto those coefficients which is similar to a supervised STRUCTURE run employing HapMap as the ancestral populations. For all other PCA, we estimated the eigenvectors from individuals in all populations. Outliers with more than six standard deviations above the mean coefficient along the top 10 eigenvectors or principal components (PCs) were removed. Wright's fixation index (F_{ST}) between each pre-specified population was estimated for both sets of SNPs. The *smartpca*

program in the Eigensoft package [Patterson et al., 2006] was used to perform PCA on the 416 AIMs in the GRAAD African Americans and Barbadians.

COMPARISON OF SUBSTRUCTURE DETECTION: 416 AIMS VS. 14K RANDOM SNPS

Pearson's correlation coefficient was computed for the primary PC from PCA using AIMs and the corresponding PC computed using 14K random SNPs. The Q-Q plot, a common diagnostic tool, was used to check the distribution of GWAS *P*-values against expected. We repeated our GWA analysis using PC1 and PC2 from each PCA run (416 AIMs and 14K SNPs, respectively) as covariates in separate logistic regression models testing for association with asthma status among the African American cases and controls. The distribution of the resultant *P*-values from each set of analyses were checked against the null distribution to see if the PCs were able to correct any of the previous deviations which might have been due to population stratification in the data.

RESULTS

ADMIXTURE ESTIMATION AND CHARACTERIZATION USING 416 AIMS

For the GRAAD African American and Barbadian samples separately, none of the 416 AIMs showed strong pairwise LD ($r^2 > 0.35$). For the YRI, CEU and CHB/JBT HapMap subjects, these SNPs all had $r^2 < 0.2$. Although no SNPs were out of HWE for the HapMap subjects and the Barbadians, one SNP deviated from HWE ($P < 10^{-6}$) among GRAAD African Americans. A total of 14, 13 and 19 SNPs had minor allele frequencies (MAF) <0.01 in the Phase II CHB/JBT, YRI and CEU groups, respectively All SNPs within the GRAAD African American and Barbadian samples had MAF > 0.01. Allele frequency differences ( ) between CEU and YRI populations for these AIMs ranged from 0.455 to 0.925, with a mean = 0.73, reflecting the selection criteria for these AIMs.

The proportion of African ancestry among the GRAAD African American asthmatic cases and controls was estimated to be 72.3 and 72.5%, respectively, and the corresponding estimated European ancestry was 19.7 and 19.6%, respectively [Mathias et al., 2010]. For both groups, approximately 8% of the alleles were estimated to originate from the ancestral population represented by the CHB/JBT cluster. Barbadians had 77.4% African ancestry, with an estimated 15.9% European and 6.7% Asian ancestry. As seen in Figure 1, the distribution of individual estimated African ancestry was distinct between GRAAD African Americans and the founders from Barbados, although both spanned a large range. Among Barbadians, the distribution of estimated African ancestry was tighter and shifted to the right. The Wilcoxon rank sum test was used to assess the similarity in distribution of African ancestry between these two samples (African American and Barbadian), which yielded an extremely large test statistic, corresponding to *P*-value $<10^{-16}$, which clearly rejects the null hypothesis of identical distributions in Figure 1.

PCA ON GRAAD AFRICAN AMERICANS AND BARBADIANS WITH HAPMAP PHASE II YRI, CEU AND CHB/JPT

PCA of the 416 AIMs among GRAAD African Americans Barbadians and the HapMap CEU, YRI and CHB/JPT samples yielded two significant axes of variation or PCs. The first PC explained 54.3% of the variation in these data capturing the dominant African vs. non-African admixture, while the second PC explained 5.07% of the variation and appears to capture a gradient between Asian to European populations, as seen in Figure 2. The remaining PCs each captured less than 1% of the genetic variation in the data. Inspecting a plot of PC2 vs. PC3 suggests both African Americans and Barbadian are genetically closer to Europeans than Asians, and closest to YRI. Estimated individual African ancestry from STRUCTURE described above was strongly correlated with the PC1 from this PCA using

AIMs, with $r = -0.992$ (95% CI = -0.991, -0.993), and there was a much lower, but still significant correlation between PC2 and the estimated proportion African ancestry ($r = 0.48$; 95% CI = 0.43–0.53). Estimated European ancestry was also significantly correlated with PC1 and PC2 ($r = 0.912$ and $= -0.832$, respectively). This suggests the main feature of observed genetic variance in these 416 AIMs is the difference between African and European ancestry.

PCA OF GRAAD AFRICAN AMERICANS AND BARBADIANS WITH HAPMAP PHASE III AFRICAN SAMPLES

In a PCA on 397 AIMs (subset of the 416) in the GRAAD African Americans, Barbadians and HapMap Phase III African samples from Nigeria (YRI) and Kenya (LWK and MKK), the first PC explained 6.75% of the observed genetic variation, while PC2 and PC3 captured 0.83 and 0.72%, respectively. Figure 3 plots the first two PCs (PC1 and PC2) from the analysis of AIMs (panel A), where there is very little observable clustering differentiating these subpopulations. In fact, these five samples were almost indistinguishable. GRAAD African Americans and Barbadians overlapped extensively with the YRI and the LWK, but showed greater dispersion along the PC1 axis, likely reflecting their postulated European admixture. GRAAD African Americans had more outliers who were quite distant from the African cluster (having high positive values along the PC1 axis). Table I gives estimated F_{ST} between these five populations computed with these AIMs (lower off-diagonal). Among the African populations, the MKK and YRI showed the greatest genetic distance with an F_{ST} of 0.110 (SE = 0.0037), Barbadians and African Americans were closest with $F_{ST} = 0.008$ (SE = 0.0004). Also based on these AIMs, the MKK appeared genetically closer to African Americans than either the YRI or the LWK, with $F_{ST} = 0.016$ (SE = 0.0012). However, Barbadians were genetically closer to both the YRI ($F_{ST} = 0.026$, SE = 0.0010) and the LWK ($F_{ST} = 0.011$, SE = 0.0009), than to the MKK ($F_{ST} = 0.038$, SE = 0.0021).

When 14K independent SNPs were analyzed in these same populations, quite different patterns emerged. Although the first and second PCs (PC1 and PC2) explained a much smaller proportion of total genetic variation (0.88%, 0.51%) than seen in the analysis of AIMs, both axes were highly significant in an ANOVA test of differences among these five populations ($P<10^{-10}$, $P = 2.84 \times 10^{-17}$). In a plot of PC1 vs. PC2 (Fig. 3, panel C), the African populations formed three distinct clusters along the first axis (PC1), with YRI at the rightmost extreme. All available African-derived populations (GRAAD cases and controls, Barbadians and the ASW samples from HapMap phase III) extend along this PC1 axis away from the YRI cluster. Again, the MKK are quite distinct from the West African YRI and the East Africa LWK groups, a difference which became even more apparent when PC1 and PC3 were plotted (see supplemental material). Again, African Americans and Barbadians displayed the characteristic gradient along PC1, originating near the YRI cluster. While the YRI and LWK form individually fairly tight clusters in this analysis of 14K randomly selected independent SNPs, there was slight gradient among the MKK extending toward the LWK cluster. Unlike the patterns obtained from analysis of AIMs, this PCA using 14K random SNPs revealed a second major axis (PC2) showing clear distinction between the main East to West Africa gradient, and emphasizing the uniqueness of MKK (Fig. 3, panel D). Notably, GRAAD African Americans and Barbadians formed a tighter cluster with the YRI, while the MKK and LWK clusters remained distinct. Genetic distances as represented by F_{ST} (Table I, upper off-diagonal) using these 14K independent SNPs show African Americans and Barbadians are genetically most similar ($F_{ST} = 0.001$, SE = 0.00). The YRI and MKK are genetically most distant ($F_{ST} = 0.024$; SE = 0.0004).

ANALYSIS OF 416 AIMS VS. 14K SNPS IN GRAAD AFRICAN AMERICANS AND BARBADIANS

A PCA was performed on the GRAAD African Americans and Barbadians combined for both the 416 ATMs and 14,629 (14K) independent SNPs (without HapMap samples) and estimated correlations between the PC1 derived from each analysis. These correlations were –0.96 (95% CI: –0.97, –0.95) for Barbadians, –0.73 (95% CI–0.77, –0.68) for African American cases, and –0.81 (95% CI: –0.84, –0.78) for African American controls. We adjusted tests for association with asthma using the first two PCs from each PCA (based on 416 ATMs and 14K SNPs, respectively) in separate genome wide analyses of all 628,098 SNPs. The adjusted P-values showed no greater improvement in deviations from expected (under the null hypothesis), and in fact, both distributions of P-values were quite similar to the unadjusted P-values (Supplementary Fig. 1).

DISCUSSION

GENETIC ANCESTRY IN AFRICAN AMERICANS VS. BARBADIANS

Past studies have shown African Americans on average have ~80% African ancestry and ~20% European ancestry [Parra et al., 2004]. However, most publications used <100 AIMs, and Benn-Torres et al. [2008] estimated Barbadians have ~90% African ancestry based on only 28 AIMs. Here we used 416 AIMs, plus a large number of randomly selected, independent SNPs drawn from a genome-wide marker panel to explore genetic substructure in African Americans and Barbadians. In addition, our large sample size included more than 900 African Americans and 296 Barbadians. Consequently, admixture estimates from the current study should be more precise. African Americans had on average a lower proportion of African ancestry (~72%, median = 74%). The Barbadians had on average a greater proportion of African ancestry compared to the African Americans (Fig. 1), but they had a slightly lower African admixture (77.4%, median = 79%) than reported by Benn-Torres et al. [2008]. Both groups showed a wide range in the proportion of estimated individual African ancestry. Neither of these populations was genetically homogeneous compared to the proxy ancestral African and European populations (YRI and CEU, respectively), and there was a gradient of admixture extending from Africa to Europe as seen in other published studies. The proportion of apparent "Asian" ancestry among Barbadians (6.7%) was slightly lower than the estimated 8% for African Americans, which may or may not be explained by the history of admixture with indigenous populations or Asian groups in these two regions. About a quarter of the GRAAD Barbadians and 30% of the African Americans had estimated Asian ancestry > 10%. A small percentage of Chinese have been living in Barbados since the 1940s, and there has been gene flow among the other Caribbean islands and the South American mainland for centuries. Although there was no permanent Amerindian population on Barbados at the time of British settlement, there has been considerable mixing of Africans from other Caribbean islands, and on all neighboring East Caribbean Islands there were (and continues to be) with Amerindian populations. For the GRAAD African Americans from the Baltimore/Washington region, this estimated proportion of Asian ancestry could represent either East Asian, Hispanic or Native American ancestry.

PCA on both GRAAD groups and the HapMap phase II samples yielded a first PC (PC1) which accounted for 54.08% of the genetic variation in these data and summarize an admixture gradient between the ancestral populations of Africa and Europe. The second PC (representing 5.07% of variation) may represent Asian-European differences. Interestingly, the first PC in analyses of both GRAAD populations were highly correlated with the estimated proportion of African and European ancestry, suggesting PCA based on AIMs provides a reliable assessment of African-European admixture in populations of African

descent. The much smaller correlations (66 and 59%) seen between estimated Asian ancestry and PC2 suggest the presence of other non-African variation in both the ancestry of African Americans and Barbadians. Because these AIMs were originally selected to contrast African and European populations, we cannot rely on them to capture with good precision non-African components besides European ancestry.

We also selected ~14,000 (14K) random, uncorrelated SNPs across the genome and ran PCA on the Barbadians and African Americans combined. The correlation between the first PC using these random, independent markers and that based upon the 416 AIMs was stronger among the Barbadians ($r^2 = 0.96$) than among GRAAD African American ($r^2 = 0.81$ for cases and $r^2 = 0.73$ for controls). This suggests 14K random SNPs and the 416 AIMs both captured the same African-European admixture that predominates in the history of the African Diaspora in the New World. The success in detecting substructure generally increases with the number of SNPs used [Turakalov and Easteal, 2003]; however, this was not the case here. This might be explainable by the lack of any population stratification, besides clear admixture, in our data. AIMs revealed the expected African-European ancestry distribution (which did not differ between GRAAD cases and controls), and the 14K SNPs revealed a similar pattern. Adjusting for the first two PCs obtained from the AIMs or the randomly selected SNPs in a GWAS analysis of risk of asthma showed virtually identical evidence of association. Moreover, a Q-Q plot comparing P-values adjusted for PC1 and PC2 from 416 AIMs and those P-values adjusted for PC1 and PC2 from the 14K SNPs showed virtually no difference. Given both sets of SNPs yielded the same results suggest ~400 well-defined AIMs might be sufficient for analysis of population substructure where the ancestral populations are as distinct as European and African populations. A subsequent PCA on the full 622k panel of autosomal markers among GRAAD African Americans yielded a PC1 showing a slightly stronger correlation (~99%) with the PC1 obtained from the 416 AIMs. Although using thousands of SNPs to investigate population substructure has become common practice, especially when genome-wide data are available, it would be unnecessary for African Americans (and similar populations), if a smaller set of well-defined AIMs are available.

AFRICAN ANCESTRY IN GRAAD AFRICAN AMERICANS AND BARBADIANS

AIMs were selected because they have large differences in allele frequency between current populations chosen to represent the true ancestral population. When considering African Americans, YRI (West African origin) and CEU (Northern European origin) samples are the most readily available proxy populations. It is commonly assumed HapMap YRI can adequately represent the ancestral African population for most individuals of African descent in the New World. Given the large genetic diversity among different African populations, this assumption may not be valid [Tishkoff et al., 2009]. Using random, independent SNPs makes no such assumption, and could potentially provide additional clues about components of admixture in populations of the African Diaspora. HapMap Phase in samples from East Africa (MKK, LKK) and West African (YRI) were used to explore, through PCA, the relationship between African Americans, Barbadians and other African groups. Both AIMs and 14K random, independent SNPs gave strong evidence of a primary PC reflecting differences along an "African" vs. "non-African" (presumably European) gradient. Similar results were previously demonstrated in PCA which included African and European groups in a study of world-wide populations [Tishkoff et al., 2009]. Although Europeans were excluded from this analysis, the African vs. non-African PC1 was still quite apparent, mainly because of the presence of European alleles in the GRAAD African American and Barbadian individuals. Using AIMs, the GRAAD populations overlapped with the three African groups along the axis of this primary PC.

Selection of markers is critical to interpretation of PCA. While PCA using AIMs yielded somewhat less pattern information about population structure among these five populations of African descent (three African and two African Diaspora), the use of 14K random, independent SNPs revealed greater detail (Fig. 3). There was more distinct clustering among the three African samples, especially along the secondary axis. Interestingly African Americans and Barbadians continued to cluster mainly with the Yorubans. However, the spatial clustering was not reflected in the F_{ST} estimates by the two sets of SNPs. In the analysis of AIMs, African Americans were most distant from Yorubans, followed by the Luhya, and then the Maasai and were closest to Barbadians. However, Barbadians were furthest from the Maasai, and closer to the Luhya than to the Yorubans. This pattern is reflected in the PC1 computed with AIMs. This may reflect how AIMs were selected (i.e. to maximize the difference in allele frequencies between YRI and CEU), and the East African populations may not contribute much variation at these AIMs. With 14K random, independent SNPs, both Barbadians and African Americans were genetically closest to the Yorubans and furthest from the Maasai (as revealed by the PC2—see Fig. 3 and Table I). Even though Barbadians had a greater proportion of African genes than African Americans, these two groups are genetically very close ($F_{ST}= 0.001$), relative to the continental African populations. This genetic similarity may reflect shared European admixture diversity in both the African Americans and Barbadians, which the three African populations lack [Tishkoff et al., 2009].

In summary, we have illustrated both African and non-African ancestral components in a large sample of African Americans and Barbadians, with a small but suggestive Asian contribution. Although Barbadians have a higher average African ancestry than do African Americans, there was very little genetic distance between these two groups compared to Europeans and even other Africans. Also, Barbadian average African ancestry was lower than previously reported. African Americans and Barbadians are traditional admixed populations formed primarily by two ancestral groups with highly differential SNP variation. Hence, substructure detection was as successful with a well-defined set of AIMs as with approximately 14,000 random, independent SNPs. Additionally, for our sample, hundreds of AIMs were just as good as thousands of SNPs in deriving PCs useful for accounting for potential confounding due to population stratification in an association analysis. To search for evidence of other African contributions (besides Yorubans) to African American and Barbadian ancestry, we introduced HapMap East African groups (the Luhya and Maasai) into our analysis. In terms of genetic distance, both African Americans and Barbadians were closer to the Yorubans than the East African groups, and clustered primarily with the Yorubans along the primary and secondary axes in a PCA. This is consistent with what is known about the geographical origins of slaves brought to the United States and the Caribbean. Historical references show that about 75% of African slaves were imported specifically from the West African region spanning Senegal to Nigeria [Curtins, 1969]. This alignment of historical records and genetic data supports the paradigm of West Africa being the dominant African contributor of genes to African Americans and Barbadians.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES

Barnes KC. Genetic epidemiology of health disparities in allergy and clinical immunology. J Allergy Clin Immunol. 2006; 117:243–254. [PubMed: 16461122]

Barnes KC, Neely JD, Duffy DL, Freidhoff LR, Breazeale DR, Schou C, Naidu RP, Levett PN, Renault B, Kucherlapati R, Iozzino S, Ehrlich E, Beaty TH, Marsh DG. Linkage of asthma and total serum IgE concentration to markers on chromosome 12q: evidence from Afro-Caribbean and Caucasian populations. Genomics. 1996; 37:41–50. [PubMed: 8921368]

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–265. [PubMed: 15297300]

Benn-Torres J, Bonilla C, Robbins CM, Waterman L, Moses TY, Hernandez W, Santos ER, Bennett F, Aiken W, Tullock T, Coard K, Hennis A, Wu S, Nemesure B, Leske MC, Freeman V, Carpten J, Kittles RA. Admixture and population stratification in African Caribbean populations. Ann Hum Gen. 2008; 72:90–98.

Chakraborty R, Kamboh MI, Nwankwo M, Ferrell RE. Caucasian genes in American blacks: new data. Am J Hum Genet. 1992; 1:145–155. [PubMed: 1729885]

Cheng CY, Kao WH, Patterson N, Tandon A, Haiman CA, Harris TB, Xing C, John EM, Ambrosone CB, Brancati FL, Coresh J, Press MF, Parekh RS, Klag MJ, Meoni LA, Hsueh WC, Fejerman L, Pawlikowska L, Freedman ML, Jandorf LH, Bandera EV, Ciupak GL, Nalls MA, Akylbekova EL, Orwoll ES, Leak TS, Miljkovic I, Li R, Ursin G, Bernstein L, Ardlie K, Taylor HA, Boerwinckle E, Zmuda JM, Henderson BE, Wilson JG, Reich D. Admixture mapping of 15,280 African Americans identifies obesity susceptibility loci on chromosomes 5 and X. PLoS Genet. 2009; 5:e1000490. [PubMed: 19461885]

Curtins, PD. The Atlantic Slave Trade A Census. Madison: The Univeristy of Wisconsin Press; 1969.

Goncalves VF, Carvalho CMB, Bortolini MC, Bydlowski SP, Pena SDJ. The Phylogeography of African Brazilians. Hum Hered. 2008; 65:23–32. [PubMed: 17652961]

Marcella S, Miller JE. Racial differences in colorectal cancer mortality. The importance of stage and socioeconomic status. J Clin Epidemiol. 2001; 54:359–366. [PubMed: 11297886]

Mathias RA, Grant AV, Rafaels N, Hand T, Gao L, Vergara C, Tsai YJ, Yang M, Campbell M, Foster C, Gao P, Togias A, Hansel NN, Diette G, Adkinson NF, Liu MC, Faruque M, Dunston GM, Watson HR, Bracken MB, Hoh J, Maul P, Maul T, Jedlicka AE, Murray T, Hetmanski JB, Ashworth R, Ongaco CM, Hetrick KN, Doheny KF, Pugh EW, Rotimi CN, Ford J, Eng C, Burchard EG, Sleiman PM, Hakonarson H, Forno E, Raby BA, Weiss ST, Scott AF, Kabesch M, Liang L, Abecasis G, Moffatt MF, Cookson WO, Ruczinski I, Beaty TH, Barnes KC. A genome-wide association study on African-ancestry populations identifies novel asthma genes. Allergy Clin Immunol. 2010; 125:336–346.

Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C, Sylvester N, Parrish-Gause D, Garvey WT, Jin L, McKeigue PM, Kamboh MI, Ferrell RE, Pollitzer WS, Shriver MD. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. Am J Phys Anthropol. 2001; 114:18–29. [PubMed: 11150049]

Parra EJ, Kittlles RA, Shriver MD. Implications of correlations between skin color and genetic ancestry for biomedical research. Nat Genet. 2004; 36:S55–S60.

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLOS Genet. 2006; 2:2074–2093.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotyping data. Genetics. 2002; 155:945–959. [PubMed: 10835412]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]

The International HapMap Consortium. The International HapMap Project. Nature. 2003; 426:739.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM. The genetic structure and history of Africans and African Americans. Science. 2009; 324:1035–1044. [PubMed: 19407144]

Turakalov R, Easteal S. Number of Snp Loci needed to detect population structure. Hum Hered. 2003; 55:37–45. [PubMed: 12890924]

Wong MD, Ettner SL, Boscardin WJ, Shapiro MF. The contribution of cancer incidence, stage at diagnosis and survival to racial differences in years of life expectancy. J Gen Intern Med. 2009; 24:475–481. [PubMed: 19189193]

Ziv E, Burchard EG. Human population structure and genetic association studies. Pharmacogenomics. 2003; 4:431–441. [PubMed: 12831322]
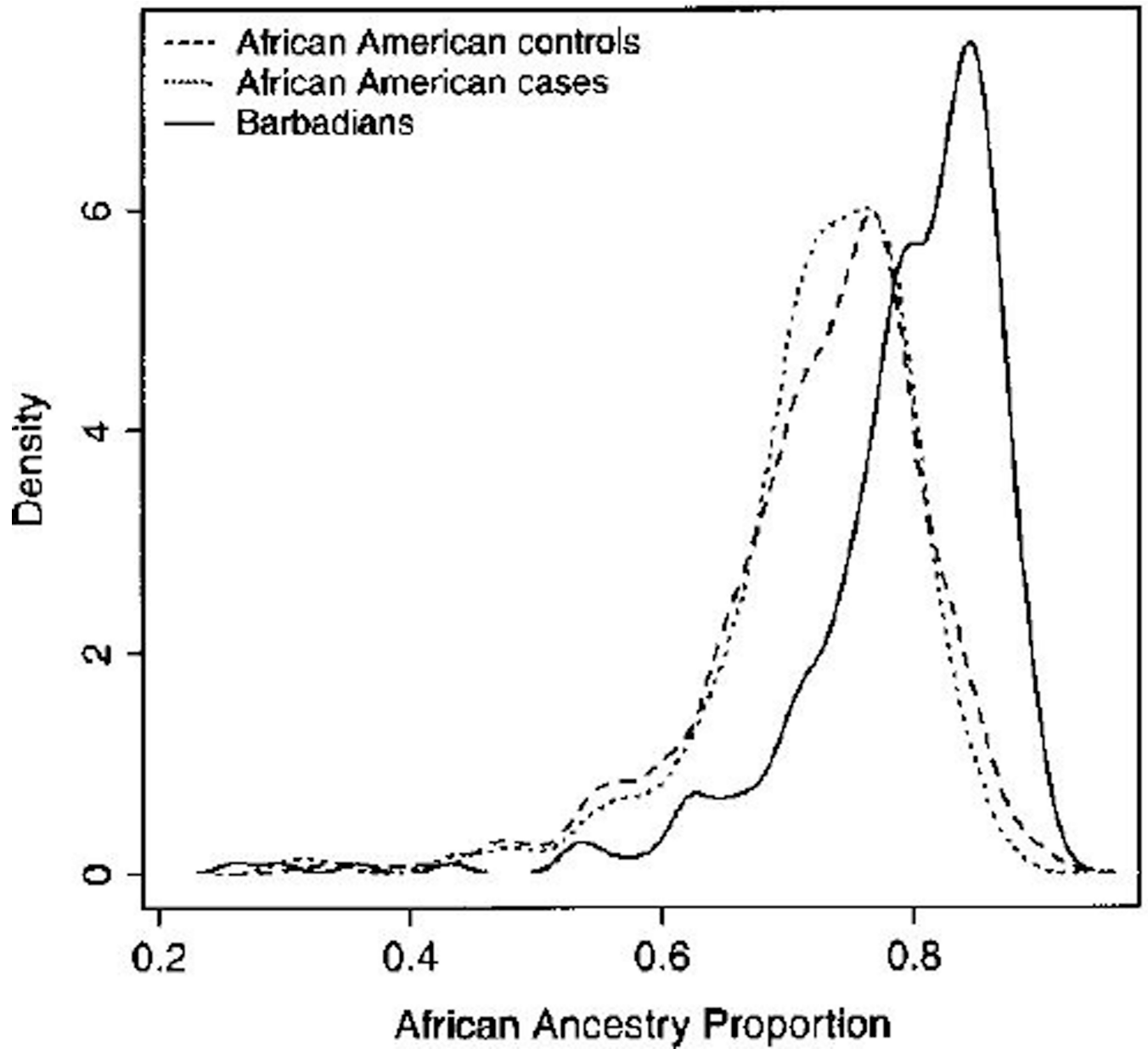
Fig. 1.
Distribution of African Admixture Proportions among GRAAD African Americans cases (*n* = 447) and controls (459) and Barbadians Founders (272).
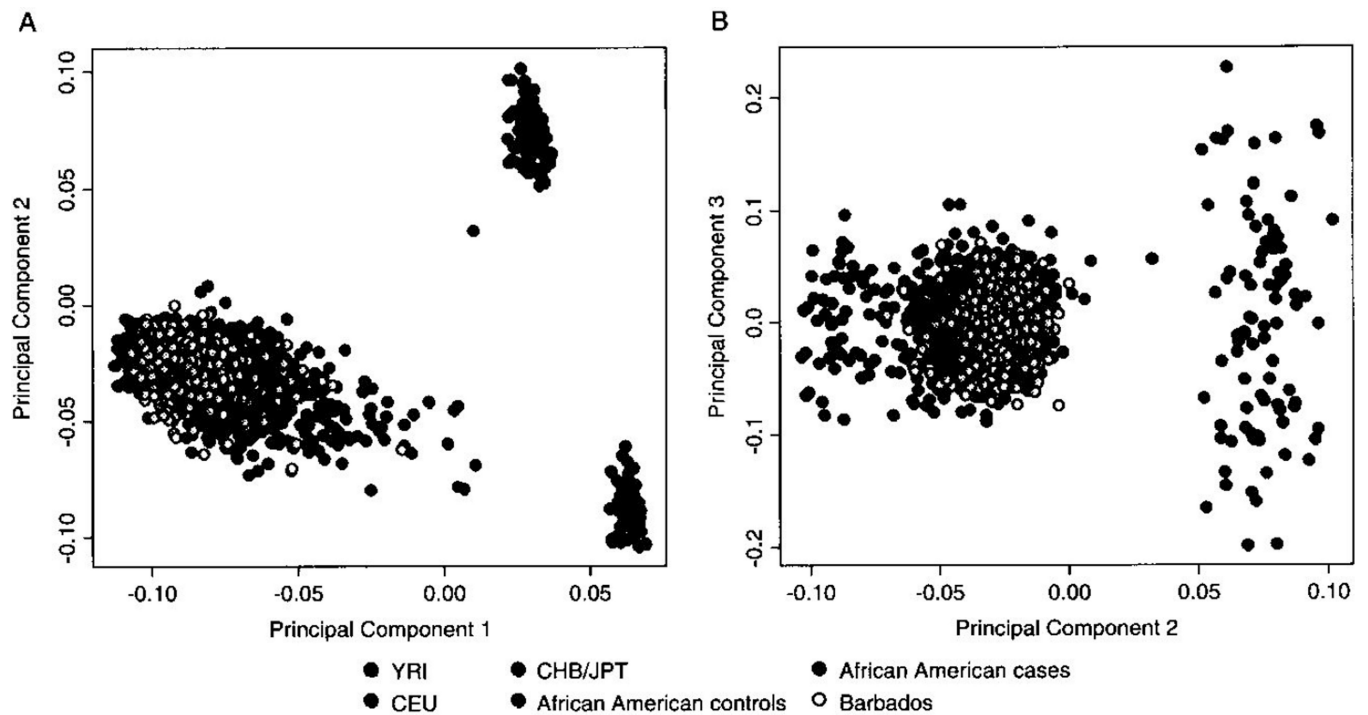
Fig. 2.
Plots of top three principal components from analysis of 416 ancestry informative markers
(AIMs) in GRAAD African American unrelated cases (447), controls (459), GRAAD
Barbadians (272), with HapMap YRI (60), CEU (60), CHB/JPT (90). (A) Plot of PC1 and
PC2 which captured 54.04% and 5.07% of the variation. (B) Plot of PC2, and PC3 which
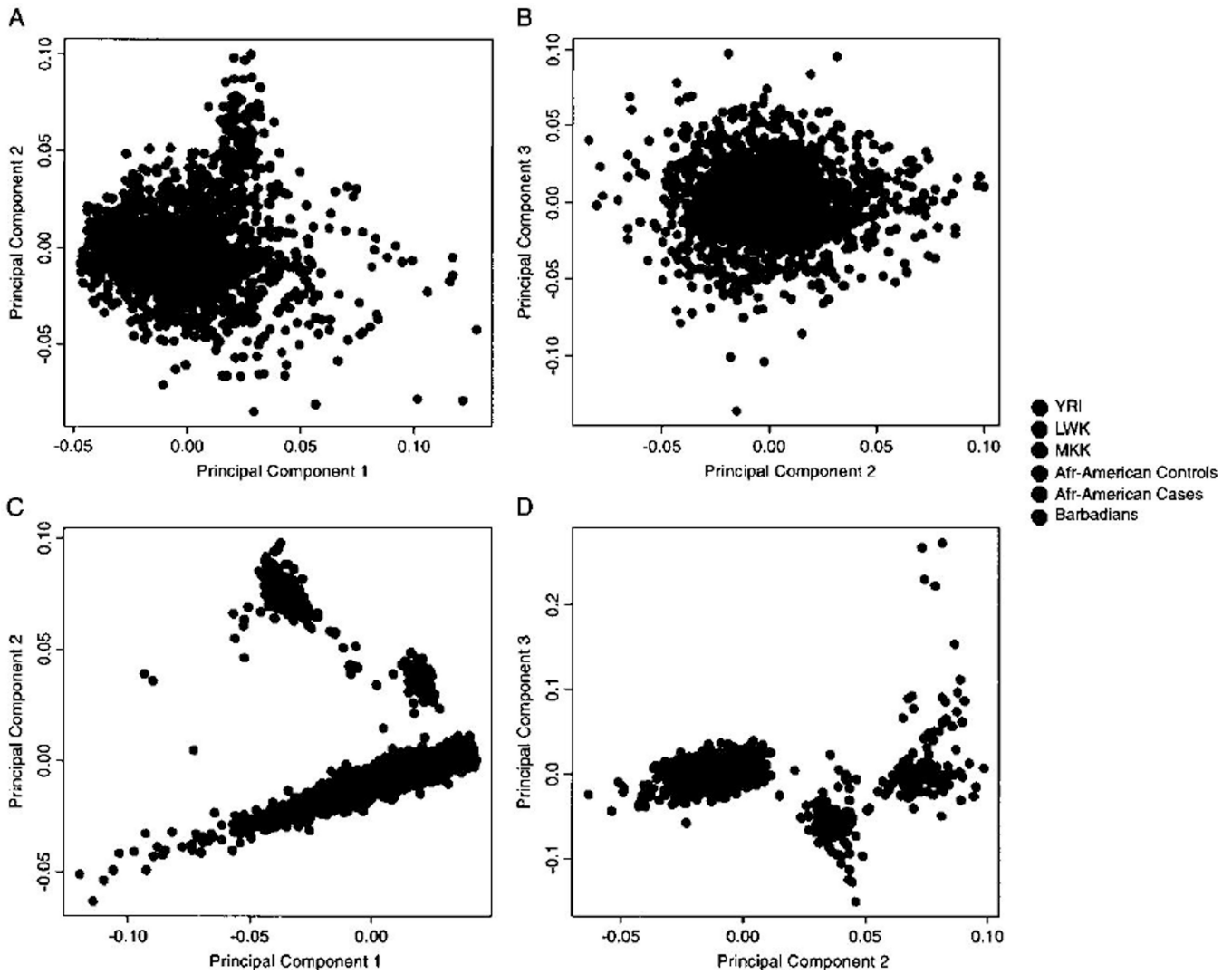captured 0.65% of the variation.

Fig. 3.
Plots of top three principal components (PC) from analysis on African American cases and controls (906), Barbadians (272), HapMap Phase III Yorubans (112), LWK (90), MKK (121). (A) Plot of PC1 vs. PC2 from PCA on 397 African-European AIMs. (B) Plot of PC2 vs.PC3 from PCA on 397 AIMS. (C) Plot of PC1 vs. PC2 from PCA on 14,629 uncorrelated SNPs. (D) Plot of PC2 vs. PC3 from PCA on 14,629 SNPs.

TABLE 1

Summarized pairwise F_{ST} estimated by 397 AIMs (lower off-diagonal) and 14,629 SNPs (upper off-diagonal)

F_{ST} (SE)	AAs controls	AA cases	Barbados	YRI	LWK	MKK
AA[a] controls		0.00 (0.00)	0.001 (0.00)	0.007 (0.0005)	0.009 (0.0005)	0.014 (0.0005)
AA cases	0.0 (0.0001)		0.001 (0.00)	0.008 (0.0005)	0.009 (0.0005)	0.014 (0.0005)
Barbados	0.008 (0.0004)	0.009 (0.0004)		0.004 (0.0005)	0.008 (0.0005)	0.018 (0.0005)
YRI[a]	0.060 (0.0015)	0.061 (0.0016)	0.026 (0.0010)		0.007 (0.0002)	0.024 (0.0004)
LWK[a]	0.032 (0.0015)	0.033 (0.0015)	0.011 (0.0009)	0.013 (0.0014)		0.015 (0.0002)
MKK[a]	0.016 (0.0012)	0.016 (0.0012)	0.038 (0.0021)	0.110 (0.0037)	0.066 (0.0027)	

[a] AA, African Americans; YRI, Yorubans; LWK, Luhya; MKK, Maasai.