RESEARCH ARTICLE

A Panel of Ancestry Informative Markers for Estimating Individual Biogeographical Ancestry and Admixture From Four Continents: Utility and Applications

Indrani Halder,[1]* Mark Shriver,[1] Matt Thomas,[2] Jose R Fernandez,[3] and Tony Frudakis[2]

[1]*Department of Anthropology, Pennsylvania State University, University Park, Pennsylvania;* [2]*DNAPrint Genomics, Sarasota, Florida;* [3]*Department of Nutrition Sciences, University of Alabama at Birmingham, Birmingham, Alabama*

Communicated by Pui-Yan Kwok

Autosomal ancestry informative markers (AIMs) are useful for inferring individual biogeographical ancestry (I-BGA) and admixture. Ancestry estimates obtained from Y and mtDNA are useful for reconstructing population expansions and migrations in our recent past but individual genomic admixture estimates are useful to test for association of admixture with phenotypes, as covariate in association studies to control for stratification and, in forensics, to estimate certain overt phenotypes from ancestry. We have developed a panel of 176 autosomal AIMs that can effectively distinguish I-BGA and admixture proportions from four continental ancestral populations: Europeans, West Africans, Indigenous Americans, and East Asians. We present allele frequencies for these AIMs in all four ancestral populations and use them to assess the global apportionment of I-BGA and admixture diversity among some extant populations. We observed patterns of apportionment similar to those described previously using sex and autosomal markers, such as European admixture for African Americans (14.3%) and Mexicans (43.2%), European (65.5%) and East Asian affiliation (27%) for South Asians, and low levels of African admixture (2.8–10.8%) mirroring the distribution of Y E3b haplogroups among various Eurasian populations. Using simulation studies and pedigree analysis we show that I-BGA estimates obtained using this panel and a four-population model has a high degree of precision (average root mean square error [RMSE] = 0.026). Using ancestry–phenotype associations we demonstrate that a large and informative AIM panel such as this can help reduce false-positive and false-negative associations between phenotypes and admixture proportions, which may result when using a smaller panel of less informative AIMs. Hum Mutat 0, 1–11, 2008. © 2008 Wiley-Liss, Inc.

KEY WORDS: biogeographical ancestry; admixture; ancestry informative markers; stratification

INTRODUCTION

In a recent report, the Race, Ethnicity, and Genetics Working Group of the National Human Genome Research Institute has suggested using biogeographical ancestry (BGA) estimates instead of racial, ancestral, or ethnic labels, as proxies to control for population stratification [2005]. The main advantage of measuring autosomal over uniparental ancestry lies in the ability to measure admixture within individuals contributed by all of their ancestors rather than just some of them. Individual BGA (I-BGA) estimates can therefore be treated as continuous variables in regression analyses to reconstruct aspects of our evolutionary past, infer admixture dynamics and demographic histories in populations; and provide a platform for correlating overt physical features and quantitative disease phenotypes with elements of population structure [Halder and Shriver, 2003; Shriver et al., 2003; Bonilla et al., 2004a,b; Reiner et al., 2005]. Associations between I-BGA and quantitative phenotypes help to deconstruct the sources of variation that contribute to the disease risk (i.e., genetic vs. environmental effects and gene–environment interactions) and thereby identify genetic mechanisms underlying diseases [Molokhia et al., 2003; Reiner et al., 2005]. The power of clinical trial designs can be significantly enhanced by using methods for quantifying population structure that relates more closely to the underlying biology of interest. Databases of I-BGA estimates and carefully quantified phenotypes could also benefit the forensic community and empirical methods may be used to estimate aspects of physical appearance and ultimately to map genes for normal traits (like eye, hair, and skin pigmentation) that are useful for individualizing persons.

The Supplementary Material referred to in this article can be accessed at http://www.interscience.wiley.com/jpages/1059-7794/suppmat.

Received 9 February 2007; accepted revised manuscript 7 November 2007.

*Correspondence to: Dr. Indrani Halder, 506 Old Engineering Hall, 4015 O'Hara Street, University of Pittsburgh, Pittsburgh, PA 15260. E-mail: halderi@upmc.edu

Grant sponsor: National Institute of Health (NIH); Grant number: P30AG015294.

DOI 10.1002/humu.20695

Published online in Wiley InterScience (www.interscience.wiley.com).

There are relatively few genomic regions that differ substantially among populations. Yet, based on continental origin and ethnogeographic affiliation, some phenotypes (e.g., skin color, height, facial features, and hair textures) exhibit substantial variation as a function, seemingly, of genetic ancestry. Given the substantial interindividual variability in admixture proportions within most historically intermixed populations, the relationship between overt phenotypes and genetic ancestry (or social constructs) is tenuous. For example, dark skin color imparted by eumelanin expression would not be a good indicator of West African ancestry, since many other populations such as Australian, Melanesian, and South Asians also express higher levels of eumelanin and exhibit darker skin color. In other cases, cryptic population structure contributed by recent ancestral admixture can be common for many populations, yet not always appreciable and certainly not quantifiable through self-assessment or visual cues. Hence, the practice of binning persons into single population groups can be inaccurate, and can confound genetic associations contributing to both type I and II errors.

The goal of this study was to identify a panel of ancestry informative markers (AIMs) that can distinguish between four continental populations: Europe/Eurasia, Subequatorial Africa, East Asia, and the Americas, and effectively infer I-BGA and admixture proportions with respect to these groups. We selected 176 AIMs based on high δ, F_{ST}, and locus-specific branch length (LSBL) [Shriver et al., 2004] values and implemented maximum likelihood (ML) and Bayesian methods to study the distribution of I-BGA and admixture proportions in several extant populations. We have observed that for some of the populations like those in the Americas, these markers provide very reliable ancestry estimates, while for other populations the panel serves to identify genetic similarity, rather than direct ancestry components. Our results suggest that a four-population model, may adequately describe global human genomic diversity for many applications, but not for all. Our results suggest no justification that there contemporaneously existed only four genetically defined races (or basic populations) segregated to the main continents of Africa, Eurasia, East Asia, and the Americas. Rather the markers and methods described here collapse individual human ancestry into such a model because it is convenient for use with many extant populations descended from the ancestral populations described in this way, and its elements neatly comport with various phenotypes. In many other populations these AIMs will provide a reasonable estimate of genetic structure. This report builds upon previously published AIM panels suitable for I-BGA and admixture analysis on a continental level [Shriver et al., 1997, 2003; Parra et al., 1998; Collins-Schramm et al., 2002; Bonilla et al., 2004a, b; Smith et al., 2004; Yang et al., 2005], but, is significantly larger and more accurate, and is the first autosomal AIM panel to be extensively validated with respect to both theoretical and empirical performance.

MATERIALS AND METHODS

Population Samples

Ancestral samples were selected from 100 unrelated individuals representing each of the four ancestral groups: West Africans (AF) from Nigeria, Sierra Leone, and Central African Republic [Parra et al., 2001]; Europeans (self-identified "Caucasians") (EU) from different U.S. locales; East Asian (EA) samples obtained from the Coriell cell repository (http://ccr.coriell.org) and first/second generation Asian Americans from different U.S. locales; and Indigenous Americans (IA) represented with samples obtained from Mixtec and Nahua persons from Guerrero, Mexico [Bonilla et al., 2005]. At the time of ascertainment a pedigree questionnaire was used, in which each subject described themselves, their parents, and all grandparents as belonging to either "African," "American Indian," "Asian," "Caucasian," or "Other" groups, with the option of reporting "Don't know". Since our ancestral sample selection was based on nongenetic characteristics we used the program STRUCTURE [Pritchard et al., 2000; Falush et al., 2003] with a four-population model to identify and exclude ancestral individuals who showed >15% affiliation with a second population. We identified individuals who were outliers in tests with several marker panels (30, 71, and 176 AIMs) and they were subsequently excluded from further analyses. The final ancestral groups consisted of 70 AF, 66 EU, 67 IA, and 68 EA, and allele frequencies in these groups were determined by standard gene counting methods. Since we were working with AIMs with high minor allele frequencies, calculations showed that these population sizes were generally adequate to estimate allele frequencies with a standard error of less than 3%. After excluding outliers we reclassified individuals using STRCUTURE (without prior population assignment) and confirmed each sample exhibited >85% genomic affiliation with one ancestral group.

Individuals used in studies of I-BGA and admixture distribution were obtained from various locations and are listed in Table 1 along with collection sites and sample sizes. Besides samples from the Coriell cell repository, all individuals self-identified as belonging to specific populations. Individual blood/DNA samples (when obtained outside of Coriell) were collected under IRB guidelines for the purposes of genetic studies of human genetic variation and written informed consent was obtained prior to ascertainment. The study was carried out in accordance with the Declaration of Helsinki [2000] of the World Medical Association.

Marker Selection

AIMs were identified in two stages by screening publicly available databases. Candidate SNPs were genotyped in ancestral population samples and of these a panel of 176 AIMs was selected. First, 400 AIMs, most with minor allele frequency >0.10, and each with $\delta > 0.4$ between any two groups were selected from previously published data on 27,000 SNPs in three populations: European American, African American, and East Asian [Akey et al., 2002]. From these, 71 AIMs that had $\delta > 0.4$ and produced consistent results on the genotyping platform (described later), were chosen such that the summed δ across all population pairs was of optimal balance (since the AIMs were high minor-allele frequency SNPs, the δ is of similar magnitude to the F_{ST}, which is the δ corrected for heterozygosity). In the second phase, 105 additional AIMs were selected to enrich the panel for markers with greater power to distinguish among EU, IA, and EA populations. LSBL values derived from F_{ST} estimates, were used to screen the previous set of 27,000 SNPs and a second set of 14,548 SNPs [Kennedy et al., 2003] typed in European American, African American, and East Asian populations. In lieu of a specific cutoff, 100 SNPs with the highest LSBL values in each ancestral group (EU, IA, and EA) (300 total) were selected. These were genotyped in the ancestral samples and the final 105 markers that provided consistent genotyping clusters and reproducible results were chosen. When selecting the second set of markers we also attempted to include only those that did not skew the balance of the total panel towards any one group (based on very high LSBL values for one population compared to the others). No a priori rule was established to set the distance between SNPs, since genomic

TABLE 1. Location and Sample Size of Populations of Known, Self-Identified Ethnicities Used for Studying I-BGA and Admixture Distribution, Using 176 AIMs*

		Mean (SD) ML estimates				Mean (SD) ADMIXMAP estimates			
Population (N)	Location	EU	AF	EA	IA	EU	AF	EA	IA
European American (207)	US[a]	**0.905** (0.1)	0.03 (0.058)	0.028 (0.049)	0.038 (0.061)	**0.966** (0.02)	0.016 (0.015)	0.006 (0.003)	0.012 (0.003)
African American (136)	US[a]	0.143 (0.133)	**0.796** (0.14)	0.028 (0.06)	0.033 (0.051)	0.182 (0.155)	**0.773** (0.161)	0.018 (0.053)	0.028 (0.035)
North African (7)	Coriell	**0.774** (0.058)	0.015 (0.073)	0.056 (0.054)	0.02 (0.034)	**0.853** (0.075)	0.131 (0.073)	0.006 (0.002)	0.01 (0.002)
North European (10)	Coriell	**0.97** (0.036)	0.01 (0.021)	0.019 (0.03)	0.04 (0.014)	**0.981** (0.006)	0.008 (0.004)	0.005 (0.002)	0.006 (0.002)
Irish (17)	Ireland[b]	**0.964** (0.043)	0.07 (0.021)	0.012 (0.027)	0.017 (0.041)	**0.978** (0.007)	0.008 (0.004)	0.005 (0.002)	0.003 (0.004)
Icelandic (12)	Coriell	**0.938** (0.055)	0.012 (0.022)	0.008 (0.014)	0.043 (0.05)	**0.966** (0.014)	0.022 (0.011)	0.004 (0.002)	0.007 (0.004)
Greek (18)	Coriell, US[a]	**0.904** (0.04)	0.048 (0.042)	0.017 (0.053)	0.047 (0.048)	**0.952** (0.038)	0.028 (0.033)	0.005 (0.002)	0.016 (0.012)
Iberians (9)	Coriell	**0.788** (0.21)	0.066 (0.071)	0.04 (0.076)	0.107 (0.167)	**0.856** (0.171)	0.055 (0.074)	0.015 (0.013)	0.074 (0.015)
Basque (10)	Coriell	**0.93** (0.052)	0.023 (0.036)	0.08 (0.025)	0.039 (0.041)	**0.961** (0.016)	0.026 (0.017)	0.005 (0.002)	0.007 (0.004)
Italian (12)	Coriell	**0.868** (0.089)	0.032 (0.048)	0.027 (0.055)	0.073 (0.059)	**0.965** (0.017)	0.013 (0.007)	0.006 (0.002)	0.017 (0.012)
Turkish (40)	Turkey[a], Us[a]	**0.853** (0.054)	0.023 (0.032)	0.073 (0.067)	0.051 (0.06)	**0.96** (0.016)	0.014 (0.01)	0.009 (0.006)	0.017 (0.018)
Ashkenazi Jews (10)	Coriell	**0.868** (0.058)	0.047 (0.039)	0.02 (0.049)	0.066 (0.036)	**0.947** (0026)	0.034 (0.022)	0.006 (0.004)	0.012 (0.009)
Middle East v1 (9)	Coriell	**0.881** (0.097)	0.028 (0.056)	0.048 (0.073)	0.042 (0.051)	**0.949** (0.056)	0.021 (0.038)	0.01 (0.012)	0.02 (0.014)
Middle East v2 (11)	Coriell	**0.822** (0.11)	0.108 (0.089)	0.045 (0.057)	0.026 (0.063)	**0.9** (0.07)	0.069 (0.063)	0.01 (0.008)	0.021 (0.013)
South Asian (South Indian) (56)	India[a]	**0.589** (0.089)	0.051 (0.047)	0.269 (0.107)	0.031 (0.088)	**0.723** (0.112)	0.045 (0.042)	0.141 (0.115)	0.091 (0.072)
South Asian (Patels, India) (8)	India[a]	**0.655** (0.077)	0.04 (0.065)	0.25 (0.103)	0.055 (0.033)	0.304 (0.033)	**0.417** (0.042)	0.265 (0.055)	0.014 (0.003)
Chinese (10)	Coriell	0.07 (0.09)	0	**0.98** (0.024)	0.013 (0.025)	0.03 (0.015)	0.006 (0.003)	**0.949** (0.018)	0.016 (0.008)
Japanese (10)	Coriell	0.011 (0.016)	0.04 (0.018)	**0.953** (0.042)	0.032 (0.04)	0.031 (0.014)	0.005 (0.002)	**0.941** (0.021)	0.023 (0.016)
Atayal (10)	Coriell	0.05 (0.016)	0	**0.976** (0.042)	0.019 (0.042)	0.027 (0.012)	0.008 (0.003)	**0.945** (0.019)	0.021 (0.019)
South East Asian (11)	Coriell	0.08 (0.111)	0.036 (0.073)	**0.822** (0.148)	0.063 (0.067)	0.142 (0.094)	0.027 (0.034)	**0.81** (0.115)	0.022 (0.017)
Pacific Islander (7)	Coriell	0.247 (0.16)	0.037 (0.045)	**0.506** (0.213)	0.21 (0.117)	**0.425** (0.022)	0.371 (0.02)	0.187 (0.026)	0.017 (0.005)
Australian Aboriginal (8)	Coriell	**0.635** (0.105)	0.01 (0.024)	0.252 (0.067)	0.103 (0.127)	0.321 0.04	0.347 0.031	0.317 0.028	0.015 0.002
American Indian1 (223)	Coriell	0.413 (0.358)	0.037 (0.124)	0.067 (0.086)	**0.476** (0.338)	0.448 (0.345)	**0.03** (0.109)	0.039 (0.063)	**0.482** (0.332)
American Indian2 (170)	Coriell	0.266 (0.276)	0.022 (0.053)	0.082 (0.032)	**0.611** (0.27)	0.338 (0.276)	0.019 (0.045)	0.048 (0.068)	**0.595** (0.276)

*Mean and standard deviation of I-BGA and admixture estimates in each population is indicated. Text in bold indicates the ancestral group with highest contribution to a given population.
[a,b]Samples collected by authors for study purposes.

coverage was not the main criteria for selecting the markers. The final marker panel is listed in Supplementary Table S1 (available online at http://www.interscience.wiley.com/jpages/1059–7794/suppmat).

DNA Isolation and Genotyping

DNA was isolated from circulating lymphocytes using QIAamp 96 DNABlood kit (Qiagen, Valencia, CA) or from buccal swabs using QIAamp DNA mini kit (Qiagen). PCR primers for the chosen markers were ascertained from dbSNP (www.ncbi.nlm.nih.gov/projects/SNP). PCR was done using the single-base primer extension protocol in a tagged fluorescent assay using the 25 K SNPstream ultra-high-throughput genotyping system (Beckman Coulter, Fullerton, CA). Primers for amplification and extension are presented in Supplementary Table S2. Quality control for genotypes was achieved through the use of a modified version of the control software supplied with the Beckman system, and relied on the observation of strong genotype clusters and clear distinction between clusters corresponding to the XX, YY, and XY genotype classes. Quality assurance was achieved through repeated genotyping and statistical analysis of five control individuals (one EU, one African American, one Mexican, one Hispanic, and one EA). Departures from independence in allelic state within and between all pairwise combinations of unlinked loci were examined using a permutation-based test implemented in the MLD program [Zaykin et al., 1995]. Allele frequencies and Hardy-Weinberg equilibrium (HWE) were ascertained using the Genepop software [Raymond and Rousset, 1995].

Calculating I-BGA

An ML and a Bayesian method were used to compute I-BGA and admixture proportions. To implement the ML method we wrote a program based on the ML algorithm described by Hanis et al. [1986]. Since a four-population ancestral model was assumed for each sample, we devised a method for efficiently parsing the entire four-dimensional likelihood space. Using the multilocus genotypes of each sample, we first calculated each of the four possible three-way admixture models and chose the three-way model with the highest likelihood. For samples for which a one-way or two-way model was most appropriate (such as 100% EA), the algorithm efficiently converges on the most likely estimate within the most likely three-way model (e.g., 100% EA, 0% IA, and 0% AF). When the second-best three-way model fell within one log of the best three-way model, the confidence intervals for the genotype (2X, 5X, and 10X, which, given a bivariate Gaussian distribution for the estimates, correspond to 88%, 95%, and 98% confidence intervals, respectively) by definition extended into four-dimensions and the MLE was then calculated assuming a more computationally intensive four-population model. This grid method was used for all ML analyses of world populations, unless mentioned otherwise. For Bayesian calculations, we used the ADMIXMAP program with the "prior allele frequencies" model [McKeigue et al., 2000; Hoggart et al., 2003, 2004].

Simulations and Pedigree Analysis

Simulated data were used to explore the bias of I-BGA estimates obtained using the current AIM panel and the ML algorithm. Bias is the result of statistical uncertainty caused by the continuous nature of the allele frequencies from population to population. A program, SimSample, was written to simulate individuals (as sets of multilocus genotypes) using ancestral allele frequencies (description of the algorithm is provided in the Supplementary Appendix). We first simulated 2000 unadmixed individuals representing each ancestral population. Next, 260 individuals each with ancestry from two to four ancestral populations were simulated for whom the exact proportion of alleles inherited from each ancestral population was noted. This expected ancestry was compared to the ML I-BGA estimates computed. A total of 12 individuals from a three-generation pedigree were genotyped and their ML I-BGA was computed.

Puerto Rican Sample Data Analysis

The AIMs were genotyped in 64 Puerto Rican women from New York, NY who had previously been analyzed for ancestry-phenotype associations using fewer AIMs [Bonilla et al., 2004b]. The ML I-BGA estimates obtained using both the most likely three-population model (e.g., EU-AF-IA) and the four-population model (EU-AF-IA-EA) were compared to study the performance of the algorithm previously described. Admixture stratification was evaluated using the Individual Ancestry Correlation Test (IACT) [Shriver et al., 2005]. The program 3LOCUS (from Dr. J.C. Long et al. [1995]) was used to estimate pairwise haplotype frequencies and calculate the significance of allelic association between pairs of unlinked loci.

Statistical analyses were performed using standard tests implemented in the program SPSS v13 (SPSS Inc., Chicago, IL). I-BGA estimates obtained under different models were compared using *t*-tests and Spearman's rank correlation coefficient. Linear regression models were used to test the effect of ancestry on skin pigmentation (measured as the Melanin [M] index) and bone mineral density (BMD).

RESULTS

Characterization of AIMs and Ancestral Samples

We selected 176 AIMs from four continental populations and identified relatively unadmixed individuals (Fig. 1) to estimate ancestral allele frequencies. Allele frequencies for each population, their chromosomal location and map distance, pairwise δ and F_{ST} values and amplification primers are shown in Supplementary Tables S1 and S2. All markers were in HWE and pairwise linkage equilibrium in each ancestral group. This panel is most informative for distinguishing between African and non-African populations (sum of marker specific F_{ST}: 27.3–36.8) and least informative for distinguishing between IA and EA ancestry (sum of marker specific F_{ST}: 12.03). Using Bayesian approaches, ancestry information evaluated using total pairwise Fisher information content (*f* value: McKeigue et al. [2000]) was 76.9 (AF-EU), 97.9

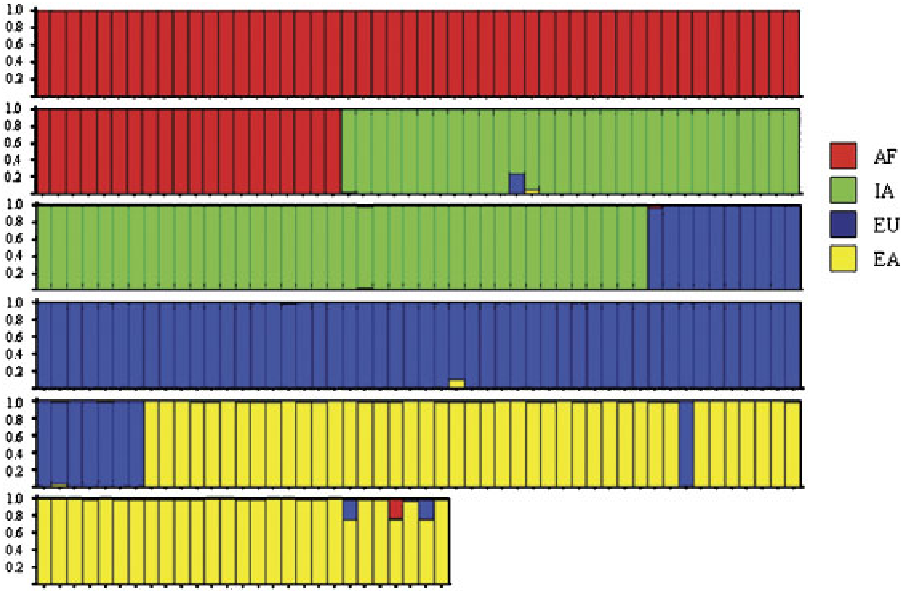

FIGURE 1. **STRUCTURE analysis of ancestral samples. All individuals with <85% affiliation with the primary group were excluded from the final analyses. Final set included 70 AF (red bars), 67 IA (green bars), 66 EU (blue bars), and 68 EA (yellow bars) individuals. Bars with more than one color depict some of the individuals who were discarded from the final analysis and are shown here for comparison purposes only.**

(AF-EA), 89.4 (AF-IA), 61.7 (EU-EA), 52.8 (EU-IA), and 32.5 (EA-IA), which correspond to average pairwise δ of 0.307 (AF-EU), 0.365 (AF-EA), 0.323 (AF-IA), 0.298 (EU-EA), 0.257 (EU-IA), and 0.178 (EA-IA). As a first test of discrimination power of the marker panel, we recalculated the I-BGA of each ancestral individual (excluding those with substantial admixture) using the allele frequencies inferred from the same samples and observed ancestral individuals to cluster closely together (Supplementary Fig. S1).

Since ML I-BGA estimates can be influenced by incorrectly specified ancestral allele frequencies (due to sampling bias), we evaluated how the ML I-BGA may change when using such altered allele frequencies. We modified allele frequencies of 60 AIMs with the highest EA-IA F_{ST} to reduce the F_{ST} of each marker by 20% and recalculated the ML I-BGA estimates for 31 individuals (with high EA and IA I-BGA). The average difference in I-BGA estimates using the original and modified sets of AIMs was only 1.7% (data not shown).

Simulation Studies

Simulation studies were used to quantify the level of bias imposed by the continuous nature of the allele frequency distributions and statistical uncertainty inherent to the MLE algorithm. For this, we assumed ancestral allele frequencies are correctly specified, and a measurement of statistical error (SE) represents an estimate of bias that is distinct from the type of error caused by deficiencies in our choice of the population model and/or ancestral representatives. SE was measured as the average proportion of outside group admixture in simulated unadmixed individuals and defined as either "population SE" (total ancestry from all noncontributing populations to an unadmixed individual) or "ancestry SE" (the total contribution from one noncontributing population to all other populations). In four samples (representing four ancestral populations), each with 2,000 simulated genomes, average I-BGA showed $>95\%$ affiliation with the expected group (Table 2) and established the average population SE at $<5\%$. Ancestry SE was also $<5\%$, as none of the populations were measured to contribute $>5\%$ across all other populations combined. Both error estimates varied across ancestral populations, with levels of individual (not population) admixture $<5\%$ being less reliable for indicating definitive admixture from another group. The lowest admixture proportions above which there is a 95% certainty of true affiliation (Table 3) ranged from $<3\%$ to 12.5% depending on the population and reflects the information content of our AIM panel as well as the genetic distances between the groups.

Next, I-BGA was estimated in 260 simulated individuals of varying EU, AF, IA, and EA admixture proportions. We found minimal departure between expected and calculated admixture proportions (Table 4), as evidenced by the strong correlations between expected and calculated I-BGA ($R^2 = 0.99$ [AF], 0.94 [EU], 0.89 [IA], and 0.77 [EA]), small root mean square of differences between expected and calculated I-BGA (average root mean square error [RMSE] = 0.026) and nonsignificant differences in mean for each ancestry axis.

TABLE 2. **Mean (ML) I-BGA Estimates in Simulated Unadmixed Individuals From Each Parental Population***

N = 2,000	AF	IA	EU	EA	Population SE
African	0.98	0.002	0.01	0.008	0.02
Indigenous American	0	0.967	0.012	0.021	0.033
European	0.004	0.017	0.964	0.015	0.036
Asian	0	0.03	0.014	0.956	0.044
Ancestry SE	0.004	0.049	0.036	0.044	

*"Population error" or the total ancestry from the noncontributing populations, and "Ancestry error" or total contribution from one noncontributing population to other populations in the analysis are both low ($P<0.05$).
AF, West African; EU, European; IA, Indigenous American; EA, East Asian; SE, statistical error.

TABLE 3. **Threshold of Affiliation Percentages for Samples of Polarized, Binary Affiliation, Above Which Results Indicate Fractional Affiliation With 95% Confidence***

	AF	IA	EU	EA
African	<0.03	<0.03	0.07	0.05
Indigenous American	<0.03	<0.03	0.075	0.115
European	0.035	0.1	<0.03	0.09
Asian	<0.03	0.125	0.08	<0.03

*$P<0.05$ for all. These results indicate that an Asian individual can show up to 12.5% IA admixture due to bias when using the current marker panel.
AF, West African; EU, European; IA, Indigenous American; EA, East Asian.

TABLE 4. **Admixture Proportions in 260 Simulated Genomes Each With Ancestry From at Least Two of the Ancestral Groups***

Ancestry axis	Expected mean ± SD	Observed mean ± SD	P
IA	0.262 ± 0.019	0.27 ± 0.019	0.199
EU	0.491 ± 0.024	0.486 ± 0.023	0.347
EA	0.124 ± 0.01	0.132 ± 0.01	0.092
AF	0.112 ± 0.018	0.121 ± 0.017	0.754

*Mean ± SD for each ancestral group. P values of *t*-test indicate nonsignificant differences in mean.

Pedigree Analyses

Using pedigree analyses we tested the effect of segregation on I-BGA estimates inferred using these AIMs. A typical three-generation pedigree is shown in Figure 2, which depicts the ratios obtained for a family of confirmed paternity (using short tandem repeat [STR] tests) with substantial EU/IA admixture. Two of the first generation individuals, Patients III and IV, were self-reported European-Americans with significant IA admixture (ascertained using current AIMs). I-BGA proportions in their son (Patient IX) and daughter (Patient X) conform to the law of independent assortment. The son's spouse (Patient VIII) self-reported as a "Hispanic" from Mexico; her BGA was 71% IA, 12% AF, and 7% EU. IA and EU admixture proportions in each of the children of Patients IX and X (Patients XI, XII, and XIII) were approximately intermediate to those of the parents, again consistent with the law of independent assortment. We noted considerable variance in estimates among siblings, Patients V, VI, VII, and VIII, notably with respect to EA and AF admixture. The father of these siblings was reported to have exhibited African pigmentation and hair texture phenotypes, and reported as of partial African ancestry. Since the variance of I-BGA estimates in siblings is a function of the age of the admixture in the parents (or how the ancestral segments are clustered along the chromosomes of the parents), and the relatively recent African contribution to the Mexican population, the variance among these siblings is most likely a function of chromosomal sampling during independent assortment. The father was not available for additional testing and we were unable to directly test effects of independent assortment in these particular individuals. Based on the simulation results, which establish $>11.5\%$ EA admixture in an IA background as significant at the 95% confidence level, Patients V and VII are most likely to have some EA I-BGA.

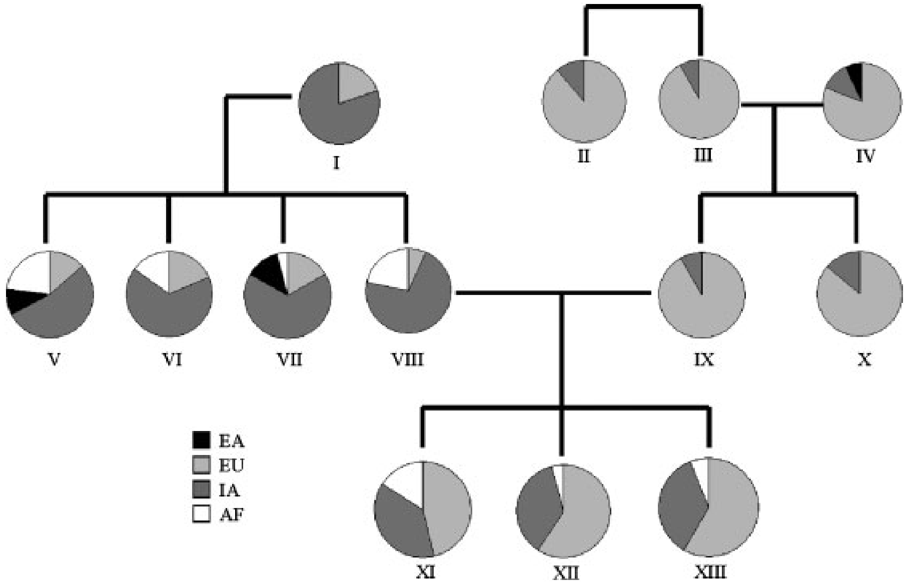

FIGURE 2. **Admixture analysis in a three-generation pedigree. Estimated I-BGA using the four-population model is shown as EA (black), EU (light grey), IA (dark grey), and AF (white) proportions. Patient I is of primarily of IA ancestry while Patients II, III, and IV are primarily of European descent.**

Survey of World Populations Using Individual BGA Estimates

We examined I-BGA proportions in several populations (Table 1) to study the global apportionment of diversity in admixture determined with this panel and population model. The apportionment of genetic diversity estimated using the ML and Bayesian methods implemented in ADMIXMAP were highly concordant indicating robustness of the AIM panel to analytical approaches (Table 1). Values of average I-BGA estimates (Table 1; Fig. 3) indicate that individuals from Europe, the Middle East, and South Asia have high EU genetic ancestry with increasing admixture moving from Northwestern Europe to South Asia (Fig. 3, ML estimates; ADMIXMAP estimates in Supplementary Figure S2). In Mediterranean and Middle Eastern populations, AF admixture increases proportionately farther south into North Africa and Southeast into Asia. This distribution appeared to mirror that for the "African" Y chromosome E3b haplogroup [Underhill et al., 2001; Jobling et al., 2003; Cruciani et al., 2004]. Populations from East and Southeast Asia affiliate primarily with EA and have increasing non-EA admixture outside of China toward Southeast Asia. African Americans revealed European admixture levels consistent with previous reports [Chakraborty and Weiss, 1986] as did Hispanics [Bonilla et al., 2005]. EA admixture increased in Eurasian populations along a Northwest to Southeast axis, and admixture levels in South Asians were comparable to previous reports from classical blood group markers [Chakraborty, 1986] (where IA and EA admixture were combined into a "Mongoloid" group) though the lower levels in continental European part of this gradient were not reproduced by the ADMIXMAP program (Table 1; compare Fig. 3 and Supplementary Fig. S2). East Asian affiliation among Eurasian populations thus deserve a more cautious interpretation, in terms of genetic distance and shared ancestry rather than in terms of direct ancestral relationships implied by our choice of ancestral group nomenclature. Interestingly, Aboriginal Australian individuals affiliated with other populations from Eurasia, exhibiting predominantly EU ancestry similar to South Asians (Table 1; Fig. 3). I-BGA in Oceanic (Pacific island) populations (Table 1; Fig. 3) were similar to those observed previously using autosomal STR AIMs and Bayesian methods [Rosenberg et al., 2002] with simpler population models, but this previous report identified the Oceanic populations as a separate cluster when a five-population model was used.

Distribution of (ML) I-BGA estimates on triangle plots (using self-identified population affiliation to group individuals) illustrate substructure, as measured by interindividual genomic variation (Fig. 4). Northern European individuals clustered as a subpopulation but Middle Eastern individuals with higher average non-EU admixture showed greater interindividual variation in I-BGA (Fig. 4A), and Greeks and Italians clustered at intermediate coordinates between the populations derived from more northern and southern regions. European Americans and continental Europeans revealed similar I-BGA profiles (Fig. 4B) but differed in that >15% EA, AF, and IA admixture was more frequently observed for the former. Non-EA admixture was exceptionally low among the Chinese, Japanese, and Atayal, but significant in Southeast Asians, South Asians, and Pacific Islanders (Fig. 4C). Ancestral AF individuals had almost no non-AF admixture, while Puerto Ricans, African Americans, and North Africans had significantly higher non-AF admixture (Fig. 4D). Among these populations too there was a notable clustering of individuals into groups with greater interindividual variation accompanying higher average admixture proportions.

On average, the group labeled as "Hispanics" was approximately 0.5 EU–0.5 IA (Fig. 4E), but some individuals were of almost 100% EU and some 100% IA. In self-identified American Indians, individuals residing on U.S. government-recognized reservations who claimed to be ">half-blood", generally showed high IA ancestry (Fig. 4F), whereas individuals from those same reservations who claimed to be "<half-blood" generally showed lower but still significant IA admixture. In contrast, individuals living outside of federally recognized reservations, in two urban locations, showed relatively low IA ancestry, despite claims of substantial American Indian ancestry by some individuals.

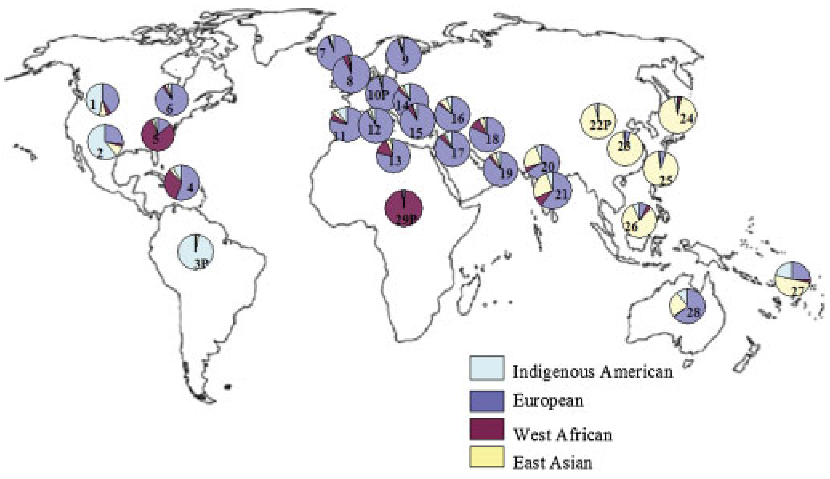

FIGURE 3. **Distribution of average ML I-BGA proportions in different global populations. P indicates ancestral population sample used for all analyses. 1) Indigenous American (includes both recognized and unrecognized tribes); 2) Indigenous American (without individuals from recognized tribes); 3) Indigenous American (ancestral IA); 4) Puerto Ricans; 5) African American; 6) European American; 7) Icelandic; 8) Irish; 9) Northern European; 10) European (ancestral); 11) Iberian; 12) Basque; 13) North African; 14) Italian; 15) Greek; 16) Turkish; 17) Ashkenazi Jews; 18) Middle Eastern (version 1); 19) Middle Eastern (version 2); 20) Indians (Patel); 21) South Asian (South Indians); 22) East Asian (ancestral); 23) Chinese; 24) Japanese; 25) Atayal; 26) South East Asian; 27) Pacific Islander; 28) Australian Aboriginal; and 29) Sub-Saharan African (ancestral).**

Admixture Analysis in Puerto Ricans and I-BGA–Phenotype Associations

To study ancestry–phenotype associations, we reanalyzed data on 64 Puerto Rican women [Bonilla et al., 2004b]. The current AIM panel has higher cumulative δ [EU-AF: 56.6, IA-AF: 59.7, and EU-IA: 47.1] compared to the previous panel of 35 AIMs used by Bonilla et al. [2004b] [EU-AF: 14.4, IA-AF: 16.6, and EU-IA: 11.7]. Using the same three-population model (EU-AF-IA) as Bonilla et al. [2004b], the mean ± standard deviation (SD) in this analysis was 0.56 ± 0.21 EU, 0.33 ± 0.24 AF, and 0.11 ± 0.09 IA, which was not significantly different for EU and AF but was significantly lower ($P < 0.0001$) for mean IA.

I-BGA computed using all four ancestral populations had a mean ± SD of 0.55 ± 0.2 EU, 0.33 ± 0.25 AF, 0.09 ± 0.08 IA, and 0.04 ± 0.06AS (Fig. 4C). Estimates obtained with the three- and four-population models were highly correlated ($R^2 = 0.998$ [EU], 1.0 [AF], and 0.889 [IA]; $P < 0.0001$ for all). In individuals who had nonzero EA admixture in the four-population model (~50%), EA I-BGA estimates appeared to result from splitting of IA estimates obtained with the three-population model. Only six individuals had significant EA admixture (>12% EA).

In congruence with the previous report [Bonilla et al., 2004b], we detected admixture stratification in the sample. 9.5% of unlinked markers showed significant association ($P < 0.01$), in which only 5% were expected to associate by chance and significant correlation was observed between I-BGA estimates obtained using nonsyntenic AIM panels (Table 5).

Skin pigmentation (measured as the melanin index) was associated positively with AF ($R^2 = 0.51$) and negatively with EU ($R^2 = 0.42$) and IA ($R^2 = 0.19$; $P < 0.0001$ for all) admixture. IA I-BGA had not been associated with skin pigmentation in the previous analysis [Bonilla et al., 2004b]. With the four-population model, skin pigmentation was also associated positively with AF ($R^2 = 0.54$) and negatively with EU and IA ($R^2 = 0.47$, 0.11, respectively; $P < 0.0001$ for all), but not with EA ($P = 0.188$). Excluding six individuals with substantial EA admixture did not substantially alter the results. BMD, which had previously associated with EU admixture, did not show any association with EU (or any other ancestral group) in this study.

DISCUSSION

We present a panel of 176 AIMs for inferring I-BGA and admixture proportions from four continents that adds to and extends the list of previously described SNP AIMs. Our goal was to develop a single panel of AIMs, which can be used to infer I-BGA and genetic structure in different populations. Theoretical [Pfaff et al., 2001; Tsai et al., 2005] and empirical [Tian et al., 2006] evidence indicates that a genome-spanning panel of 100 to 160 AIMs provides reasonably robust I-BGA estimates. Results from our simulation studies and pedigree analysis show that this panel of 176 AIMs provides reliable estimates of continental admixture (or genetic structure), with low SE thresholds and relative robustness to misspecified ancestral allele frequencies. We have demonstrated the utility and applications of this AIM panel in inferring admixture related stratification and ancestry–phenotype associations by studying several extant populations and reanalyzing previous data sets. This panel is not meant for genome wide admixture mapping, which would require substantially more markers [Tian et al., 2006, Smith et al., 2004] but is a tool that will be useful in various studies of genetic structure.

The ancestry information for the current AIM panel is comparable to that of previously described panels (e.g., 199 SNP and deletion/insertion polymorphism (DIP) AIMs in Yang et al. [2005]). Some smaller AIM panels previously described have higher average allele frequency differences for some of the population pairs (e.g., the Smith et al. [2004] panels contained four partially overlapping marker sets targeting specific populations) but, in contrast, our aim was to develop a single comprehensive panel that could be used as a standard for many studies. Since a relatively balanced AIM panel is required to avoid skewing I-BGA estimates along a particular ancestry-axis (especially with models involving more than two ancestral groups), we restricted pairwise differential in certain dimensions (i.e., EU-AF)

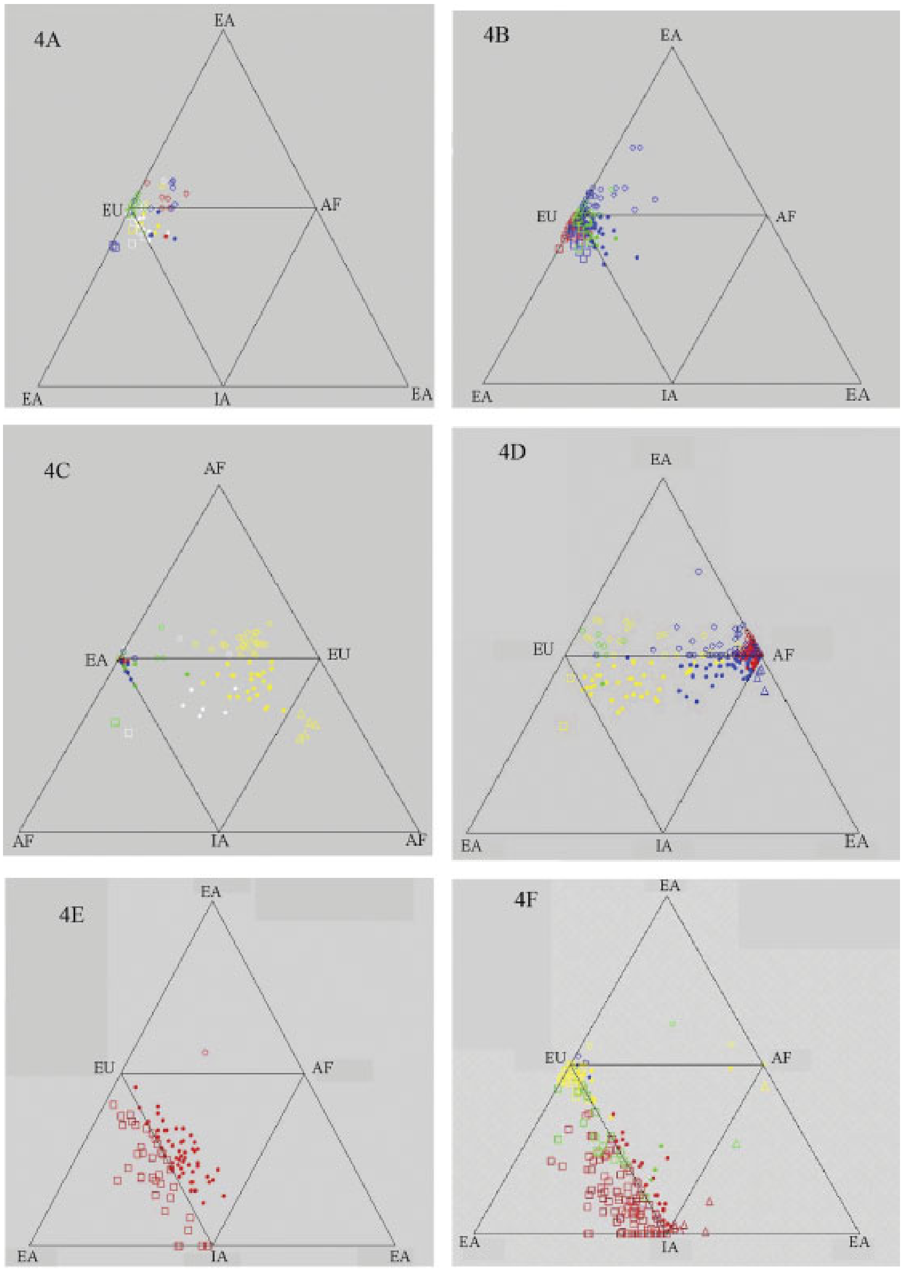

FIGURE 4. **Distribution of I-BGA in different populations: A: European (green: Northern European; red: North African; yellow: Ashkenazi Jews; and white: Greeks and Turkish), and Middle Eastern (green) populations (versions 1 and 2 from Coriell) on the tetrahedron plot. Each triangle within tetrahedron represents three-way admixture models: EU-AF-IA (filled circles): EA-IA-EU (open squares), EU-AF-EA (open circles). B: Self-identified U.S. Caucasians (blue) European ancestral individuals (red) and Continental Europeans (green) using three-way admixture models: EU-AF-IA (filled circles) EA-IA-EU (open squares) EU-AF-EA (open circles). C: Asian and Pacific Island populations: Chinese (red), Japanese (blue), South East Asian (green), South Asian (yellow), and Pacific Islander (white) under three-way admixture models: EA-IA-EU (filled circles), EA-IA-AF (open squares), EU-AF-IA (open triangles), and EU-AF-EA (open circles). D: Populations of African ancestry: West African parentals (red), African Americans (blue), North Africans (green), and Puerto Ricans (yellow) using three-way admixture models: EU-AF-IA (filled circles), EA-IA-EU (open squares), EA-IA-AF (open triangles), and EU-AF-EA (open circles). E: Hispanics under three-way admixture models: EU-AF-IA (filled circles), EA-IA-EU (open squares), and EU-AF-EA (open circles). F: Populations of self-reported Indigenous American ancestry: >0.5 blood Amerindians from recognized (R) tribes (red), >0.5 blood Amerindians from nonrecognized (NR) tribes (blue), <0.5 blood Amerindians from R tribes (green), and <0.5 blood Amerindians from NR tribes (yellow). Admixture estimates calculated under three-way admixture models: EU-AF-IA (filled circles), EA-IA-EU (open squares), EA-IA-AF (open triangles), and EU-AF-EA (open circles).**

in order to obtain comparable information across other population pairs (such as IA-EA). Some previous reports have argued that fewer markers can detect population stratification, differentiate between extant, rather than ancestral population groups, and assign group memberships to individuals [Bamshad et al., 2003; Lao et al., 2006]. Smaller marker sets are associated with lower information content and larger standard errors and as our analysis of the Puerto Rican data shows, higher information content may provide a more precise understanding of ancestry–phenotype associations by reducing false negatives (as in the IA and I-BGA-skin pigmentation association in Puerto Ricans) and false positives (as in the EU admixture-BMD association). Admixed models with

TABLE 5. Individual Admixture CorrelationTests in Puerto Ricans Under theThree-Population Model*

Ancestral axis	R^2	P
IA	0.052	<0.0001
EU	0.616	<0.0001
AF	0.743	0.005

*Pearson's correlation between estimates obtained with nonsyntenic (all odd chromosome markers vs. all even chromosome markers) marker panels for each axis of ancestry under the three-way admixture model. All axes show significant correlation indicating presence of admixture stratification.

respect to elements of extant population structure are more appropriate for human populations, as admixture underlies most of the stratification within these populations and the binning of samples to groups fails to account for such stratification. Treating ancestry in terms of individual admixture, rather than in terms of population averages (in which the grouping of populations is based on social constructs) is thus expected to help decompose elements of population structure that may be relevant for improving study designs of human genetics research. For example, we showed that the I-BGA distribution in the group labeled as "Hispanics" is continuous along the EU and IA axes, and that the term "Hispanic" shows little correspondence with genetic ancestry or even ancestral admixture (since not all Hispanics exhibit substantial Native American ancestry). The term "Hispanic" encompasses several groups including Mexicans, Cubans, and Puerto Ricans, as well as most populations from South and Central America. Individuals annotated as "Hispanics" were obtained from Coriell and included no descriptors such as origin and/or collection site, language spoken, etc. Yet I-BGA distributions were very different for self-identified Puerto Ricans (which many consider to be "Hispanics") compared to the "Hispanics" from Coriell, and even within the Coriell "Hispanics," considerable interindividual variation in EU and IA admixture was observed. Much of this admixture cannot be gleaned from self-assessments or visual cues, yet would be crucial to correct for in order to avoid the confounding influence of cryptic population structure in genetic association or epidemiological studies. Similar arguments can be made against considering "African Americans" as a single group for research purposes. However, while precise quantification of genetic ancestry serves to resolve issues related to genetic background, metapopulation/population labels will still be required as a proxy of other unmeasured sociocultural and/or environmental variables that could act as potential confounders/mediators.

It is interesting that the continental AIMs we describe show an ability to partially resolve among ethnic groups within continents, which was unexpected, since the AIMs were not selected for distinguishing intracontinental substructure. This observation has additional implications for the utility of the panel in helping resolve elements of cryptic population structure in human study samples. For instance, systematic differences in the levels of non-EU admixture (Fig. 4A) were observed among Eurasian populations as a function of geographical origin. Individuals from geographical locations intermediate to the Northern and South Eastern extremes of continental Europe, (e.g., Greeks and Italians) clustered at intermediate coordinates between these extremes. The ethnicity specific clustering within continental populations using continental AIMs thus indicates that these markers contain some ethnogeographical information, with a power to resolve among subcontinental populations with differing population histories. However, our goal was to represent as many populations as possible, at the expense of sample sizes for each, and we recognize that while our results are suggestive of an ability to discriminate elements of cryptic subcontinental population structure, they require verification with larger sample sizes.

Our choice of a four-population model was based on the previous observations that elements of modern population structure segregate to the major continents [Rosenberg et al., 2002; Underhill et al., 2001] The model is relatively simple and requires fewer markers to measure accurately, yet it accounts for most of the global diversity in uniparental haplogroups (8 of the 10 major Y haplogroup clades) and most of the anthropometric phenotype diversity. The nomenclature used for identifying ancestral populations refers to the locations from which the individuals were collected from and/or recently derived and were selected in an effort to easily communicate our results rather than to strictly delimit the geographical ranges from which they are ultimately descended. We chose the EU, AF, and EA ancestral samples by attempting to represent the ancestral diaspora, assuming that the average allele frequency among the groups more closely resembles that of the ancestral population. Using a single population to represent IA ancestors is appropriate if we assume a single wave of expansion/migration across the Bering Strait 15,000 years ago [Silva et al., 2002; Zegura et al., 2004; Mulligan et al., 2004], but would be less perfect were there two or more such expansions [Lell et al., 2002; Bortolini et al., 2003], though it is as yet unclear which of these models is appropriate. Bayesian methods such as the ADMIXMAP program [McKeigue et al., 2000; Hoggart et al., 2003, 2004] can address the uncertainty associated with allele frequencies and it is notable that the results obtained with the two programs were generally similar (Table 1).

However, in three of the populations studied we observed discrepancies between the ML and Bayesian I-BGA estimates (South Asian [Patels], Australian Aboriginals, and Pacific Islanders), which illustrate some limitations of this panel. Small sample sizes could partly account for this difference between ML and Bayesian estimates. It is also possible that a different ancestral population model (with three, five, or four other populations) would be more appropriate (for instance including "Central Asian" or "Australians" as ancestors), and not including these groups lead to greater uncertainty in assigning I-BGA in these groups. Choice of ancestral population model deserves careful consideration and hypothesis-free cluster analyses or other objective criteria (see for example the statistic proposed by Hoggart et al. [2004]) may assist one in selecting an ancestral population model most appropriate for a given phenotype, study design, or research problem. Thus, this AIM panel may be considered adequate for inferring ancestral proportions in some world populations but not in all. Nonetheless, I-BGA estimates obtained with this marker panel and ancestral population model still provides an assessment of genetic structure in populations (including the Australian aborigines, Pacific Islanders, and South Asian [Patel community] in our study) when interpreted with respect to genetic distances. These measures are ultimately useful in quantifying and controlling for genetic stratification and associating elements of population structure with phenotypes of interest.

ACKNOWLEDGMENTS

We thank Zack Gaskin, Siva Ginjupalli, Viswanathan Ponnuswamy, and Timothy Terravainen for help with data analysis, Carolina Bonilla for providing access to previously analyzed data, Paul McKeigue for access to his work on ancestry informativeness, and three anonymous reviewers for their comments. Data collection from the Puerto Rican sample was supported by NIH grant P30AG015294 (to J.F.).

REFERENCES

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002. Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12:1805–1814.

Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB. 2003. Human population genetic structure and inference of group membership. Am J Hum Genet 72:578–589.

Bonilla C, Parra EJ, Pfaff CL, Dios S, Marshall JA, Hamman RF, Ferrell RE, Hoggart CL, McKeigue PM, Shriver MD. 2004a. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. Ann Hum Genet 68:139–153.

Bonilla C, Shriver MD, Parra EJ, Jones A, Fernandez JR. 2004b. Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York city. Hum Genet 115:57–68.

Bonilla C, Gutiérrez G, Parra EJ, Kline C, Shriver MD. 2005. Admixture analysis of a rural population of the state of Guerrero, Mexico. Am J Phys Anthropol 128:861–869.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. Am J Hum Genet 73:524–539.

Chakraborty R. 1986. Gene admixture in human populations: models and predictions. Yearb Phys Anthropol 29:1.

Chakraborty R, Weiss KM. 1986. Frequencies of complex diseases in hybrid populations. Am J Phys Anthropol 70:489–503.

Collins-Schramm HE, Phillips CM, Operario DJ, Lee JS, Weber JL, Hanson RL, Knowler WC, Cooper R, Li H, Seldin MF. 2002. Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. Am J Hum Genet 70:737–750.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R. 2004. Phylogeographic analysis of haplogroup E3b E-M215 y chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet 74:1014–1022.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Halder I, Shriver MD. 2003. Measuring and using admixture to study the genetics of complex diseases. Hum Genomics 1:52–62.

Hanis CL, Chakraborty R, Ferrell RE, Schull WJ. 1986. Individual admixture estimates: disease associations and individual risk of diabetes and gallbladder disease among Mexican-Americans in Starr County, Texas. Am J Phys Anthropol 70:433–441.

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM. 2003. Control of confounding of genetic associations in stratified populations. Am J Hum Genet 72:1492–1504.

Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM. 2004. Design and analysis of admixture mapping studies. Am J Hum Genet 74:965–978.

Jobling MA, Hurles ME, Tyler-Smith C. 2003. Human evolutionary genetics: origins, peoples and disease. London: Garland Science. p 290–292.

Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J, Liu W, Yang G, Di X, Ryder T, He Z, Surti U, Phillips MS, Boyce-Jacino MT, Fodor SP, Jones KW. 2003. Large-scale genotyping of complex DNA. Nat Biotechnol 21: 1233–1237.

Lao O, van Duijn K, Kersbergen P, de Knijff P, Kayser M. 2006. Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. Am J Hum Genet 78:680–690.

Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC. 2002. The dual origin and Siberian affinities of Native American Y chromosomes. Am J Hum Genet 70:192–206.

Long JC, Williams RC, Urbanek M. 1995. An E-M algorithm and testing strategy for multiple-locus haplotypes. Am J Hum Genet 56: 799–810.

McKeigue PM, Carpenter JR, Parra EJ, Shriver MD. 2000. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. Ann Hum Genet 64:171–186.

Molokhia M, Hoggart C, Patrick AL, Shriver M, Parra E, Ye J, Silman AJ, McKeigue PM. 2003. Relation of risk of systemic lupus erythematosus to West African admixture in a Caribbean population. Hum Genet 112:310–318.

Mulligan CJ, Hunley K, Cole S, Long JC. 2004. Population genetics, history, and health patterns in Native Americans. Annu Rev Genomics Hum Genet 5:295–315.

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD. 1998. Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet 63:1839–1851.

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. 2001. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. Am J Phys Anthropol 114:18–29.

Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD. 2001. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am J Hum Genet 68:198–207.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Race, Ethnicity, and Genetics Working Group. 2005. The use of racial, ethnic, and ancestral categories in human genetics research. Am J Hum Genet 77:519–532.

Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity 86:248–249.

Reiner AP, Ziv E, Lind DL, Nievergelt CM, Schork NJ, Cummings SR, Phong A, Burchard EG, Harris TB, Psaty BM, Kwok PY. 2005. Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. Am J Hum Genet 76:463–477.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. Science 298: 2381–2385.

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrell RE. 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. Am J Hum Genet 60:957–964.

Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM, Kittles RA. 2003. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 112:387–399.

Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW. 2004. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Hum Genomics 1:274–286.

Shriver MD, Mei R, Parra EJ, Sonpar V, Halder I, Tishkoff SA, Schurr TG, Zhadanov SI, Osipova LP, Brutsaert TD, Friedlaender J, Jorde LB, Watkins WS, Bamshad MJ, Gutierrez G, Loi H, Matsuzaki H, Kittles RA, Argyropoulos G, Fernandez JR, Akey JM, Jones KW. 2005. Large-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. Hum Genomics 2:81–89.

Silva WA, Jr, Bonatto SL, Holanda AJ, Ribeiro-Dos-Santos AK, Paixao BM, Goldman GH, Abe-Sandes K, Rodriguez-Delfin L, Barbosa M, Paco-Larson ML, Petzl-Erler ML, Valente V, Santos SE, Zago MA. 2002. Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. Am J Hum Genet 71:187–192.

Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, Malasky MJ, Scafe C, Le E, De Jager PL, Mignault AA, Yi Z, De The G, Essex M, Sankale JL, Moore JH, Poku K, Phair JP, Goedert JJ, Vlahov D, Williams SM, Tishkoff SA, Winkler CA, De La Vega FM, Woodage T, Sninsky JJ, Hafler DA, Altshuler D, Gilbert DA, O'Brien SJ, Reich D. 2004. A high-density admixture map for disease gene discovery in African Americans. Am J Hum Genet 74:1001–1013.

Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF. 2006. A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. Am J Hum Genet 79:640–649.

Tsai HJ, Choudhry S, Naqvi M, Rodriguez-Cintron W, Burchard EG, Ziv E. 2005. Comparison of three methods to estimate genetic ancestry and control for stratification in genetic association studies among admixed populations. Hum Genet 118:424–433.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62.

Yang N, Li H, Criswell LA, Gregersen PK, Alarcon-Riquelme ME, Kittles R, Shigeta R, Silva G, Patel PI, Belmont JW, Seldin MF. 2005. Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. Hum Genet 118:382–392.

Zaykin D, Zhivotovsky L, Weir BS. 1995. Exact tests for association between alleles at arbitrary numbers of loci. Genetica 96:169–178.

Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol 21:164–175.