Hum Genet (2003) 112 : 387–399
DOI 10.1007/s00439-002-0896-y

ORIGINAL INVESTIGATION

**Mark D. Shriver · Esteban J. Parra · Sonia Dios
Carolina Bonilla · Heather Norton · Celina Jovel
Carrie Pfaff · Cecily Jones · Aisha Massac
Neil Cameron · Archie Baron · Tabitha Jackson
George Argyropoulos · Li Jin · Clive J. Hoggart
Paul M. McKeigue · Rick A. Kittles**

Skin pigmentation, biogeographical ancestry and admixture mapping

Received: 19 September 2002 / Accepted: 27 November 2002 / Published online: 11 February 2003

Abstract Ancestry informative markers (AIMs) are genetic loci showing alleles with large frequency differences between populations. AIMs can be used to estimate biogeographical ancestry at the level of the population, subgroup (e.g. cases and controls) and individual. Ancestry estimates at both the subgroup and individual level can be directly instructive regarding the genetics of the phenotypes that differ qualitatively or in frequency between populations. These estimates can provide a compelling foundation for the use of admixture mapping (AM) methods to identify the genes underlying these traits. We present details of a panel of 34 AIMs and demonstrate how such studies can proceed, by using skin pigmentation as a model phenotype. We have genotyped these markers in two population samples with primarily African ancestry, viz. African Americans from Washington D.C. and an African Caribbean sample from Britain, and in a sample of European Americans from Pennsylvania. In the two African population samples, we observed significant correlations between estimates of individual ancestry and skin pigmentation as measured by reflectometry (R^2=0.21, P<0.0001 for the African-American sample and R^2=0.16, P<0.0001 for the British African-Caribbean sample). These correlations confirm the validity of the ancestry estimates and also indicate the high level of population structure related to admixture, a level that characterizes these populations and that is detectable by using other tests to identify genetic structure. We have also applied two methods of admixture mapping to test for the effects of three candidate genes (*TYR*, *OCA2*, *MC1R*) on pigmentation. We show that *TYR* and *OCA2* have measurable effects on skin pigmentation differences between the west African and west European parental populations. This work indicates that it is possible to estimate the individual ancestry of a person based on DNA analysis with a reasonable number of well-defined genetic markers. The implications and applications of ancestry estimates in biomedical research are discussed.

Elecetronic database information: URLs for the data in this article are as follows:
dbSNP web page, http://www.ncbi.nlm.nih.gov/SNP/
Shriver Lab web page,
http://anthro.psu.edu/rsrch/biolab/index.html
McKeigue Lab web site,
http://www.lshtm.ac.uk/eph/eu/GeneticEpidemiologyGroup.htm
OMIM web site,
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM

M. D. Shriver (✉) · E. J. Parra · S. Dios · C. Bonilla · H. Norton
C. Jovel · C. Pfaff
Department of Anthropology, Penn State University,
409 Carpenter Bld., University Park, PA 16802, USA
Tel.: +1-814-8631078,
e-mail: mds17@psu.edu

C. Jones · A. Massac · R. A. Kittles
National Human Genome Center, Howard University,
Washington, DC 20060, USA

N. Cameron · A. Baron · T. Jackson
Takeway Media, London, EC1R OBD, UK

G. Argyropoulos
Pennington Center for Biomedical Research,
Baton Rouge, La., USA

L. Jin
Department of Environmental Health University of Cincinnati,
Cincinnati, Ohio, USA

C. J. Hoggart · P. M. McKeigue
Department of Epidemiology and Population Health,
London School of Hygiene and Tropical Medicine,
London, WC1E 7HT, UK

Present address:
E. J. Parra
Department of Anthropology,
University of Toronto at Mississauga,
Mississauga, Ontario, Canada

Introduction

Recently, there has been increased interest in association studies as a useful approach for mapping common disease genes (Risch and Merikangas 1996; Jorde 2000; Nordborg and Tavaré 2002). The issue of which populations would be best suited for linkage disequilibrium (LD) mapping has prompted much discussion and debate (see

Wright et al. 1999; Eaves et al. 2000; Nordborg and Tavaré 2002; Kaessmann et al. 2002). The extent of LD is a complex function of a number of genetic and evolutionary factors, such as mutation, recombination and gene conversion rates, demographic and selective events and the age of the mutation itself. Some of these factors affect the whole genome, whereas others only affect particular genome regions. Additionally, variations in mutation, recombination and gene conversion rates throughout the genome are expected to create LD differences between genomic regions (see, for example, Taillon-Miller et al. 2000). Some authors have proposed that small isolated inbred populations will be better suited than other populations for gene mapping, because of the lower heterogeneity and the larger extent of LD (Wright et al. 1999; Nordborg and Tavaré 2002; Kaessmann et al.2002 and references therein). Other populations well suited for mapping are recently admixed populations, such as Hispanics and African Americans, which offer the advantage that LD has been created recently because of the admixture process. Because this LD is recent, it can extend over large chromosomal regions. However, it is also extremely important to control for the genetic structure (inter-individual variation in admixture proportions) present in these populations in order to avoid false positives (Parra et al.1998; Lautenberger et al.2000; Pfaff et al. 2001; Nordborg and Tavaré 2002; Kittles et al. 2002). Interest in admixture mapping has increased in recent years (McKeigue et al. 2000; Smith et al. 2001; Collins-Schramm et al. 2002). We provide a general description of admixture mapping and some details about a statistical approach that we have developed for admixture mapping and its application to skin pigmentation as a model phenotype.

Materials and methods

Admixture mapping

Admixture generates allelic associations between all marker loci where allele frequencies are different between parental populations (Chakraborty and Weiss 1988). These associations will decay with time in a way that is dependent on the genetic distance between them. Thus, disease (or trait) risk alleles that are different between the parental populations can be mapped in admixed populations by using special panels of genetic markers showing high frequency differences between the parental populations. These markers, which we have termed ancestry informative markers (AIMs), are characterized by having particular alleles that are more common in one group of populations than in other populations. One measure of the informativeness of such markers is the allele frequency differential (δ), which is simply the absolute value of the difference of a particular allele between populations (Chakraborty and Weiss 1988; Dean et al. 1994).

In admixed populations, allelic associations have been generated recently and thus are more easily detected for a given sample size and extend over longer distances than in non-admixed populations (10–20 cM or more). The statistical basis of this approach was first explored by Chakraborty and Weiss (1988) and subsequently by Stephens et al. (1994) and Briscoe et al. (1994) who named it mapping by admixture LD (MALD). To eliminate associations of the trait with alleles at unlinked loci, it is necessary to control, in the analysis, for individual ancestry estimated from the marker data. To this end, we have employed an analysis of variance (ANOVA) test by using the estimate of individual admixture as a conditioning variable to control for the effect of individual ancestry.

We have developed an alternative approach to exploiting admixture that has little in common with classical LD mapping and that is more analogous to linkage analysis of an experimental cross (McKeigue 1998; McKeigue et al.2000). For this reason, the term "admixture mapping" has been proposed as more appropriate than MALD. Instead of testing for allelic associations, we model the underlying variation in ancestry on chromosomes of mixed descent to extract all the information about linkage that is generated by admixture. Although advanced statistical methods are required to apply this approach in practice, the underlying principle on which it relies to detect linkage is straightforward. Suppose, for instance, that a locus accounts for some of the variation in pigmentation between west Africans and Europeans. If we are able to classify individuals of mixed descent according to whether they have 0, 1 or 2 alleles of west African ancestry at this locus, then, in a comparison of these three groups with other factors held constant, the mean pigmentation level will vary with the proportion of alleles at the locus that are of west African ancestry. Controlling in the analysis for parental admixture eliminates association of the trait with ancestry at unlinked loci and ensures that the comparison is made with other factors held constant.

To infer the ancestry of the alleles at the locus from the marker genotype, we require the conditional probability of each allelic state given the ancestry of the allele (ancestry specific allele frequencies), which is west African or European in this example. To model the distribution of admixture in the population, the admixture of each individual's parents and the stochastic variation of ancestry on chromosomes inherited from these parents require a complex hierarchical model with many nuisance parameters. Fitting such models requires a Bayesian approach with Markov chain simulation, as previously described by McKeigue (1998) and McKeigue et al. (2000). ADMIXMAP, a program implementing this approach for Windows and Linux platforms is freely available through the web (http://www.lshtm.ac.uk/eph/eu/GeneticEpidemiologyGroup.htm).

Evidence is accumulating that admixture mapping will be an effective means of gene identification. At least three independent groups have now reported that, in admixed populations, strong allelic association is observed between linked markers spaced at substantial distances (Parra et al.1998, 2001; McKeigue et al.2000; Lautenberger et al.2000; Smith et al.2001; Wilson and Goldstein 2000; Pfaff et al.2001). Given the very high levels of association that have been observed over long genetic distances, it is expected that phenotypes different between parental populations because of some genetic factor will also show associations with linked AIMs. A particularly well-suited phenotype to which to apply admixture mapping is skin pigmentation for the reasons detailed below.

Skin pigmentation as a model phenotype for admixture mapping

Skin pigmentation stands out as an excellent phenotype for testing admixture mapping. Pigmentation is one of the most variable phenotypes observed in human populations, possibly because of the action of natural selection. Little is known about the number of genes determining normal pigmentation variation either within or between populations, other than that many genes are probably involved. Constitutive pigmentation (pigmentation levels measured in unexposed areas of the skin) is affected by the environment to a much lesser extent than other common traits and diseases, in which a plethora of factors such as age, diet, exercise and stress play a key role. Skin pigmentation is a quantitative trait that can be readily and precisely measured by reflectometry. A number of portable instruments are currently available that accurately measure the melanin content of the skin (e.g. Shriver and Parra 2000; Wagner et al. 2002). Additionally, many of the genes that are probably involved in normal variation in pigmentation have previously been identified for their effects on hypopigmentary phenotypes and mouse coat-colour phenotypes. The best example to date of a gene affecting normal pigmentation is the melanocortin 1 receptor (*MC1R*)

gene, which determines pale skin, red hair and freckles (Valverde et al. 1995; Smith et al. 1998; Rees et al. 1999; Flanagan et al. 2000; Bastiaens et al. 2001a) and which is also involved in sun sensitivity and melanoma risk (Healy et al. 2000; Valverde et al. 1996; Kennedy et al. 2001; Bastiaens et al. 2001b). More recently, the *ASIP* gene has been shown to affect the pigmentation levels of the hair (Kanetsky et al.2002). Few studies of pigmentation genetics have involved the use of objective measurements of skin or hair pigmentation as available by reflectance spectroscopy. Two studies that stand out in this respect are Akey et al. (2001b) and Wu et al. (2001). In this report, we have also used reflectometry to investigate the relationship between normal variation in pigmentation and several candidate genes by using admixture mapping.

Populations and samples

We have examined three population samples: a sample of 232 African and African-American individuals living in Washington, D.C., a sample of 173 British African Caribbean persons, and a sample of 187 individuals of European-American ancestry living in State College, Pa. The sample from Washington, D.C. consisted of self-reported African Americans from the Howard University campus and surrounding area. Recruitment was primarily through flyers placed around the University and by word of mouth. Volunteers came from the student body, staff, faculty and the surrounding community. This sample of individuals represents a heterogeneous selection of persons with regards to personal history, national origins and cultural and biological backgrounds. In general terms, this sample is a selection of volunteers, who would declare in the US that they are Black with primarily African ancestry. The British African Caribbean sample was ascertained in Britain through flyers and advertisements regarding the study. Volunteers who phoned in were invited to participate if all four of their grandparents were of African-Caribbean ancestry. The European-American sample from State College, Pa. was collected as part of an ongoing study of the genetics of common traits in the Anthropometrics Laboratory located in the Penn State General Clinical Research Center.

For each of these surveys, volunteers contacted the study coordinators and then either came to the clinic or were then visited at their place of abode where they gave informed consent, filled out a demographic questionnaire and gave cheek cells or blood for DNA. Each person at all three sites was then measured for skin pigmentation by using a DermaSpectrometer (cyberDerm, Media, Pa.) three times on the medial aspect of each arm and these values were averaged together as previously described (Shriver and Parra 2001). Following the interview and examination, blood for DNA was taken by venipuncture, except in the British African Caribbean sample where cheek swabs were used to collect DNA samples. This study was approved by the IRB committees of both Penn State University (State College and British samples) and Howard University (Washington, D.C. samples).

Laboratory methods

Genotyping was performed by using either the melting-curve single-nucleotide polymorphism (McSNP) genotyping method (Akey et al. 2001a; Ye et al.2002) or agarose gel electrophoresis. Approximately one third of these AIMs have previously been described (Parra et al. 1998, 2001). Table 1 gives the complete list of AIMs that were analysed in this study, with information on chromosomal location and frequency differences between European, African and native American populations. Information regarding primer sequences, polymorphic sites and other relevant information has been submitted to the dbSNP NCBI database, under the submitter handle PSU-ANTH; the dbSNP submitted SNP (SS) numbers are indicated in Table 1.

Statistical methods

The admixture proportions of each study population were estimated by using a weighted least squares method (WLS; Long 1991) implemented in the ADMIX program kindly supplied by Dr. Jeffrey C. Long. Individual ancestry was estimated by two independent methods: the maximum likelihood approach described by Hanis et al. (1986) and a Bayesian method implemented in the STRUCTURE 2.0 program (Pritchard et al. 2000). Haplotype frequencies and gametic disequilibrium coefficients for pairs of loci were estimated by using the expectation maximization algorithm described by Long et al. (1995). The gametic disequilibrium coefficient is the difference between the observed haplotype frequencies, inferred by means of the expectation maximization method, and the expected frequencies, which are a function of the observed allele frequencies for a pair of markers. Hypothesis testing was performed with the likelihood ratio statistic (G), which has a $\chi 2$ distribution for large sample sizes. We used the 3LOCUS program, made available to us by Dr. Jeffrey C. Long, which also implements a permutation based test of the G statistic. Regression and ANOVA analyses were carried out by using the SPSS 10.0 statistical package. The Bayesian admixture mapping method used is described in more detail below.

Test for population structure

To test for population stratification, the STRUCTURE 2.0 program (Pritchard et al. 2000) was used. This program models population admixture, individual admixture and locus ancestry in a Bayesian approach but does not require the user to specify the number of parental populations or the allele frequencies in these populations. Instead the number of parental populations (K) and the ancestry-specific allele frequencies are estimated from the data. The program was run initially with 30,000 iterations for the burn-in period and 70,000 additional iterations to obtain parameter estimates, with a prior distribution that allowed K to take values from 1 to 3. Additional runs with longer iterations were also carried out to check the consistency of the results.

Modeling of admixture

The ADMIXMAP program was used to model the distribution of admixture and to test for linkage with the melanin index. This program is based on a Bayesian full probability model for population admixture, individual admixture and locus ancestry similar to that described previously. Admixture and marker genotypes were set up as described by McKeigue et al. (2000). In this model, the ancestry of each marker allele, the ancestry of each individual and the distribution of admixture in the population are modelled as random variables with a hierarchical structure. In contrast to the STRUCTURE program, ADMIXMAP requires the user to specify the parental populations and to supply estimates of the ancestry-specific frequencies of marker alleles and (for tightly linked loci such as the two SNPs in the DRD2 gene) haplotypes.

The distribution of individual admixture in the parental generation was assigned a non-informative prior distribution. The stochastic variation of ancestry between two states (European and African) on each set of chromosomes inherited from a parent was modelled as a Markov process with parameters α, β for the exponential distributions of segment lengths. This requires only one extra parameter, viz. $s=\alpha+\beta$, as the parental admixture specifies the ratio $\alpha/(\alpha+\beta)$. The parameter s was assigned a gamma prior. The two markers in the *DRD2* gene were modeled as a single locus, with the observed genotype being dependent on the unobserved pair of haplotypes. The posterior probability distribution of the missing data, given the observed marker data, was generated by the Markov chain Monte Carlo simulation package as described previously (McKeigue 1998; McKeigue et al. 2000).

The relationships of skin pigmentation to individual ancestry were modelled by linear regression, assigning non-informative priors to the parameters. With a large sample and non-informative priors, the posterior mode (or mean) and 95% central posterior interval obtained in this Bayesian analysis are asymptotically equivalent to a maximum likelihood estimate and 95% confidence inter-

Table 1 Ancestry informative markers. The marker name and chromosomal band, approximate location of the marker on the chromosome in megabases (Mb), the allele frequencies of the *1 alleles and difference in frequency between African and European populations (AF/EU), African and native Americans (AF/NA) and European and native Americans (EU/NA) are shown

Marker	dbSNP ss no.	Location	Mb (UCSB)	African	European	Native American	AF/EU	AF/NA	EU/NA
MID-575*1	4387042	1p34.3	36.1	0.124	0.004	0.584	0.121	0.460	0.584
MID-187*1	4387043	1p34.1	45.0	0.759	0.388	0.301	0.370	0.458	0.087
FY-NULL*1	4387025	1q23.2	160.0	0.001	0.998	1.000	0.997	0.999	0.001
*AT3**1	4387045	1q25.1	174.8	0.858	0.282	0.061	0.575	0.797	0.222
*F13B**1	4387024	1q31.3	201.1	0.704	0.063	0.018	0.641	0.687	0.045
TSC1102055*1	4387029	1q32.1	207.6	0.487	0.917	0.137	0.430	0.351	0.780
WI-11392*1	4390531	1q42.2	242.2	0.878	0.433	0.626	0.444	0.252	0.193
WI-16857*1	4387031	2p16.1	57.0	0.751	0.215	0.181	0.536	0.570	0.034
WI-11153*1	4387032	3p12.3	78.2	0.785	0.133	0.819	0.652	0.033	0.819
*GC**1F	NA	4q13.3	72.9	0.853	0.156	0.339	0.697	0.514	0.183
GC*1S	NA	4q13.3	72.9	0.069	0.607	0.542	0.538	0.473	0.065
MID-52*1	4390532	4q24	101.7	0.363	0.077	0.763	0.186	0.500	0.687
SGC30610*1	4387040	5q11.2	51.3	0.401	0.255	0.699	0.146	0.281	0.427
SGC30055*1	4387041	5q23.1	115.9	0.054	0.511	0.753	0.457	0.699	0.241
WI-17163*1	4387033	5q33.2	155.8	0.054	0.175	0.690	0.120	0.636	0.515
WI-9231*1	4387034	7	NA	0.129	0.147	0.548	0.017	0.419	0.401
*CYP3A4**1	4390533	7q22.1	101.2	0.198	0.958	0.959	0.761	0.762	0.001
WI-4019*1	4387035	7q22.1	104.5	0.430	0.306	0.618	0.124	0.168	0.311
*LPL**1	4387026	8p21.3	22.4	0.971	0.492	0.442	0.479	0.529	0.050
*CRH**1	4390534	8q13.1	64.3	0.318	0.927	0.983	0.609	0.624	0.056
WI-11909*1	4387036	9q21.31	72.4	0.805	0.881	0.181	0.075	0.587	0.663
D11S429*1	4387023	11q11	57.9	0.087	0.516	0.119	0.429	0.032	0.397
TYR-192*1	4387030	11q21	98.7	0.005	0.449	0.034	0.444	0.029	0.415
DRD2-Bcl I*1	4387021	11q23.1	118.4	0.063	0.144	0.665	0.080	0.409	0.446
DRD2-Taq I "D"*1	4387022	11q23.1	118.4	0.135	0.670	0.045	0.535	0.090	0.626
*APOA1**1	4387046	11q23.3	123.6	0.420	0.925	0.977	0.505	0.557	0.052
*GNB3**1	4387018	12p13.31	6.8	0.795	0.332	0.364	0.463	0.431	0.032
*RB1**1	4387047	13q14.2	48.3	0.926	0.315	0.175	0.611	0.733	0.122
*OCA2**1	4387028	15q13.1	24.9	0.115	0.746	0.488	0.631	0.373	0.258
WI-14319*1	4387037	15q14	30.6	0.386	0.201	0.716	0.185	0.330	0.514
*CYP19**1	4387020	15q21.2	47.9	0.332	0.296	0.741	0.037	0.409	0.446
PV92*1	4387048	16q24.1	87.7	0.225	0.152	0.792	0.073	0.568	0.640
MC1R-314*1	4387027	16q24.3	93.6	0.513	0.163	0.035	0.350	0.478	0.127
WI-14867*1	4387038	17p13.2	3.9	0.024	0.472	0.418	0.448	0.394	0.054
WI-7423*1	4387039	17p12	8.7	0.000	0.476	0.058	0.476	0.058	0.418
Sb19.3*1	4387049	19p13.11	27.2	0.415	0.903	0.645	0.488	0.230	0.258
*CKM**1	4387019	19q13.32	61.2	0.164	0.313	0.904	0.150	0.740	0.590
MID-154*1	4387044	20q11.22	32.5	0.806	0.362	0.420	0.444	0.368	0.057
MID-93*1	4390535	22q13.2	38.8	0.261	0.815	0.081	0.554	0.179	0.733

val. Because the model takes into account the uncertainty with which individual admixture can be estimated by using only a limited set of markers, this approach correctly estimates the slope of the relationship that we would observe if individual admixture were measured accurately. Tests for various null hypotheses are constructed by averaging the score (gradient of the log-likelihood) and information (curvature of the log-likelihood) over the posterior distribution as described previously (McKeigue et al. 2000). The score variance is the observed information, calculated by subtracting the missing information (variance of the score over the posterior distribution) from the complete information (expectation of the curvature of the log-likelihood over the posterior distribution). The proportion of information extracted is the ratio of the observed information to the complete information: this allows us to estimate how much more information could be extracted if we were to increase the density of markers and thus to reduce uncertainty about parental admixture and locus ancestry.

The test for linkage of the trait with a marker locus is a test for association of the trait with locus ancestry in a regression model that includes mean parental admixture as a covariate. McKeigue (1998) has shown previously that this is a specific test for linkage.

Results

Distribution of skin pigmentation in three population samples

Figure 1 shows the distribution of the melanin (M) index in three population samples. The M values range from 32.6 to 80.8 in African Americans living in Washington,

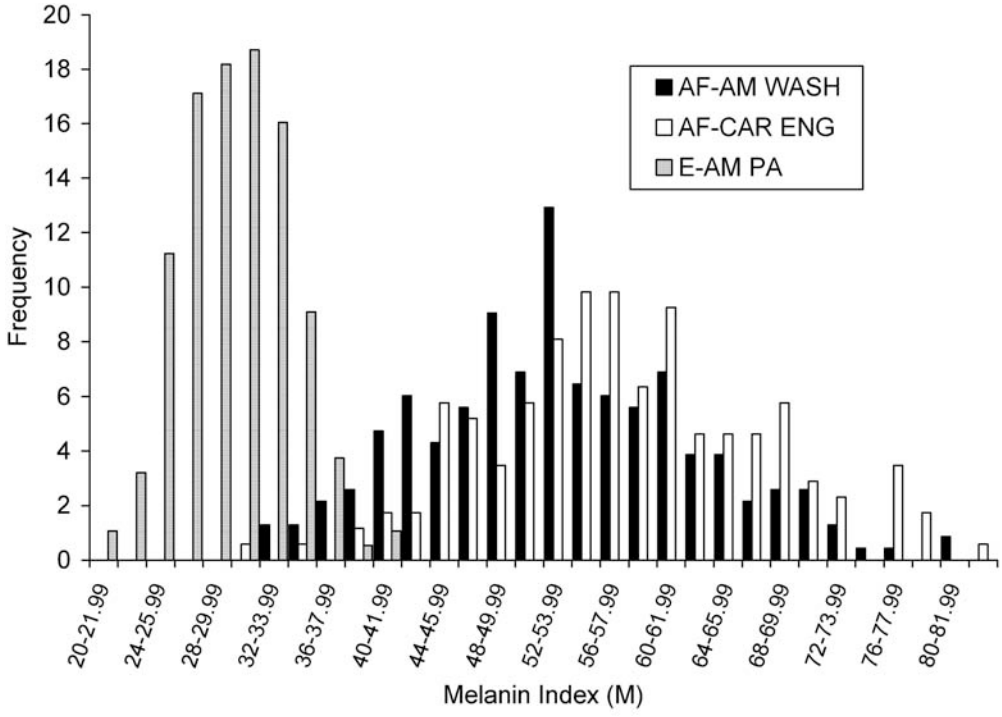

Fig. 1 Distribution of skin pigmentation in three populations. Skin pigmentation was measured by using the DermaSpectrometer and is indicated as the melanin index (*M*). The three samples are African Americans from Washington D.C. (*black bars*), African Caribbeans from Britain (*white bars*) and European Americans from State College, Pa. (*grey bars*)

D.C., 31.0 to 82.4 in African Caribbeans living in Britain, and 20.9 to 40.8 in European Americans living in State College. The M distributions are not significantly different from a normal distribution in any of the groups considered (data not shown). Average M index levels are significantly different between the European Americans (average M=29.9±0.274) and the two samples of persons of west African ancestry (Washington D.C., average M=53.4± 0.630, and African Caribbeans, average M=57.8±0.739). The difference between the two groups of west African ancestry is also significant, with African Caribbeans showing, on average, higher pigmentation levels ($P<0.0001$). This difference is also evident in Fig 1 where the distribution of skin pigmentation in African Caribbeans is shifted to the right (higher M and darker skin) with respect to African Americans from Washington D.C. We have also tested for differences in skin pigmentation between the sexes. Interestingly, in the African-Caribbean and African-American samples, males are, on average, darker than females (African Caribbeans: males M=60.1, females M= 55.9, $P=0.005$; African Americans from Washington D.C.: males M=54.4, females M=52.0, $P=0.064$). In European Americans, the opposite is true, with females being on average darker than males (males M=28.9, females M=30.4, $P=0.008$). However, further analysis of this European-American sample by using more advanced measurements of pigmentation based on diffuse reflectance spectroscopy does not show a significant difference between males and females (Wagner et al. 2002).

Population admixture and individual ancestry in three population samples

We have estimated the average African, European and native American genetic contributions in the three samples by using Long's WLS method (Long 1991). To estimate admixture proportions, we used 31–34 AIMs showing high frequency differences between African, European and native American populations (see Table 1 for additional information). The parental population frequencies were estimated by genotyping these markers in four African samples (several groups from Sierra Leone, two samples from Nigeria and a sample from Central African Republic), two European samples (Germans and Spanish) and two native American samples (Mayans from Mexico and several

Table 2 Estimates of African, European and native American ancestry in three US population samples. Estimates were obtained by using Long's WLS method and are based on 34 AIMs in European Americans from State College, 33 AIMs in African Americans from Washington, D.C. (WI-17163 was not genotyped in this sample), and 31 AIMs in African Caribbeans from Britain (WI-17163, WI-11392 and PV92 were not genotyped in this sample)

Population	African ancestry ± SE	European ancestry ± SE	Native American ancestry ± SE
African Americans Washington D.C.	78.7%±1.2%	18.6%±1.5%	2.7%±1.4%
African Caribbeans Britain	87.9%±1.1%	10.2%±1.4%	1.9%±1.3%
European Americans State College, Pa.	0.7%±0.9%	96.1%±1.6%	3.2%±1.6%

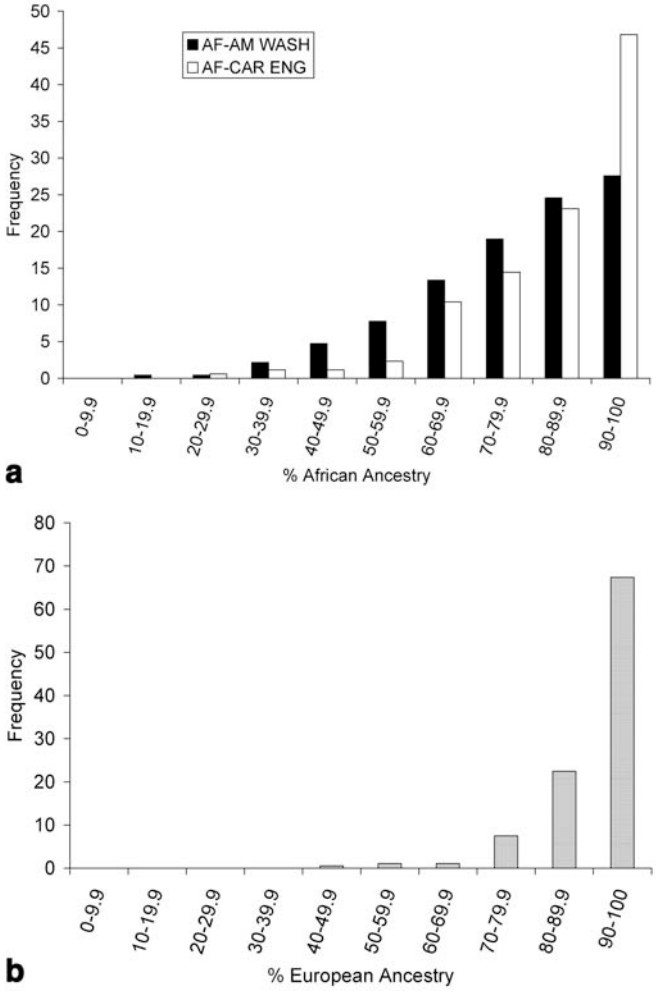

Fig. 2a, b Histograms showing the distribution of individual ancestry in three populations. **a** Distribution of percent African ancestry as estimated by using 34 AIMs in a sample of African Americans living in Washington, D.C. (*black bars*) and African Caribbeans living in Britain (*white bars*). **b** Distribution of percent European genetic contribution as estimated by using 34 AIMs in a sample of European Americans living in State College, Pa. (*grey bars*)

groups from southwest US). The average African, European and native American ancestry in each sample is reported in Table 2. Note the small standard errors associated with the estimates, which are in all cases lower than 2%. The sample of African Americans from Washington D.C. shows a significantly higher European contribution than the African Caribbean sample from Britain (18.6%± 1.5% vs 10.2%±1.4%). In both samples, the native American contribution is small, with the 95% confidence intervals overlapping 0.0 (Washington, D.C.: 2.7%; African Caribbeans: 1.9%). In European Americans from State College, the west African and native American genetic contribution are low (0.7% and 3.2%, respectively). In addition to the average admixture in each sample, we have also calculated the distribution of individual ancestry in the three groups, by using a maximum likelihood (ML) approach (Hanis et al. 1986). The distribution of west African ancestry in the African American and African Caribbean individuals is shown in Fig. 2a and the distribution of European ancestry in the sample from State College is reported in Fig. 2b. In accordance with the higher average west African genetic contribution observed in African Caribbeans, the distribution of individual west African ancestry values are shifted to the right with respect to African Americans, with more individuals showing a higher west African genetic contribution (Fig. 2a). The estimates of individual ancestry by using the ML method are very similar to the estimates produced by a totally independent method, implemented in the STRUCTURE program (Pritchard et al. 2000). Estimates of individual ancestry by ML and STRUCTURE are highly correlated (African Americans from Washington, D.C., R^2=0.983; African Caribbeans from Britain, R^2=0.950; European Americans from State College, R^2=0.762). Additionally, the average individual ancestry estimates obtained with the ML and the STRUCTURE program are in close agreement with the WLS values indicated above (data not shown).

Relationship of individual ancestry and skin pigmentation

In Fig. 3, we plot, for all the samples characterized in this study, the relationship between individual ancestry (in this case, the percentage of African genetic contribution estimated by using the ML method) and the skin pigmentation level as measured by the M index. It can be seen that the samples cover the whole range of possible west African genetic contribution (Fig. 3, X-axis). Some African Americans (circles) and African Caribbeans (asterisks) show a relatively low west African contribution and some European Americans (diamonds) show evidence of west African genetic contribution. Part of the dispersion in individual ancestry is attributable to small number of markers avail-

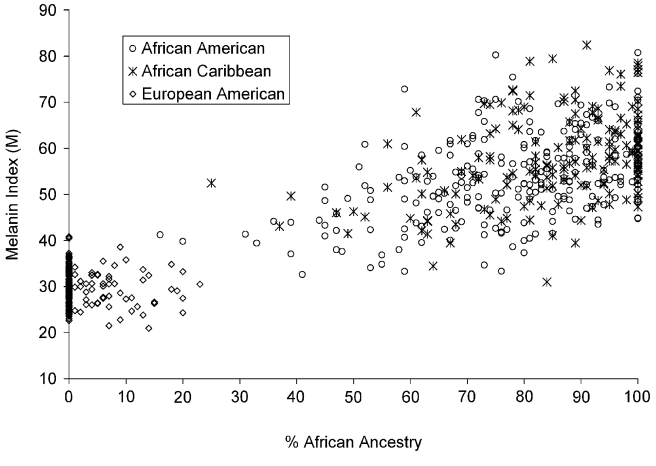

Fig. 3 The relationship between percent west African ancestry and skin pigmentation in three populations. Percent west African ancestry (obtained by using 34 AIMs and calculated by the ML method) is shown on the *y-axis* and the melanin index (*M*) on the *x-axis*. The three populations are European Americans from State College, Pa. (*diamonds*), African Americans from Washington, D.C., and State College, Pa. (*circles*) and African Caribbeans from Britain (*asterisks*)

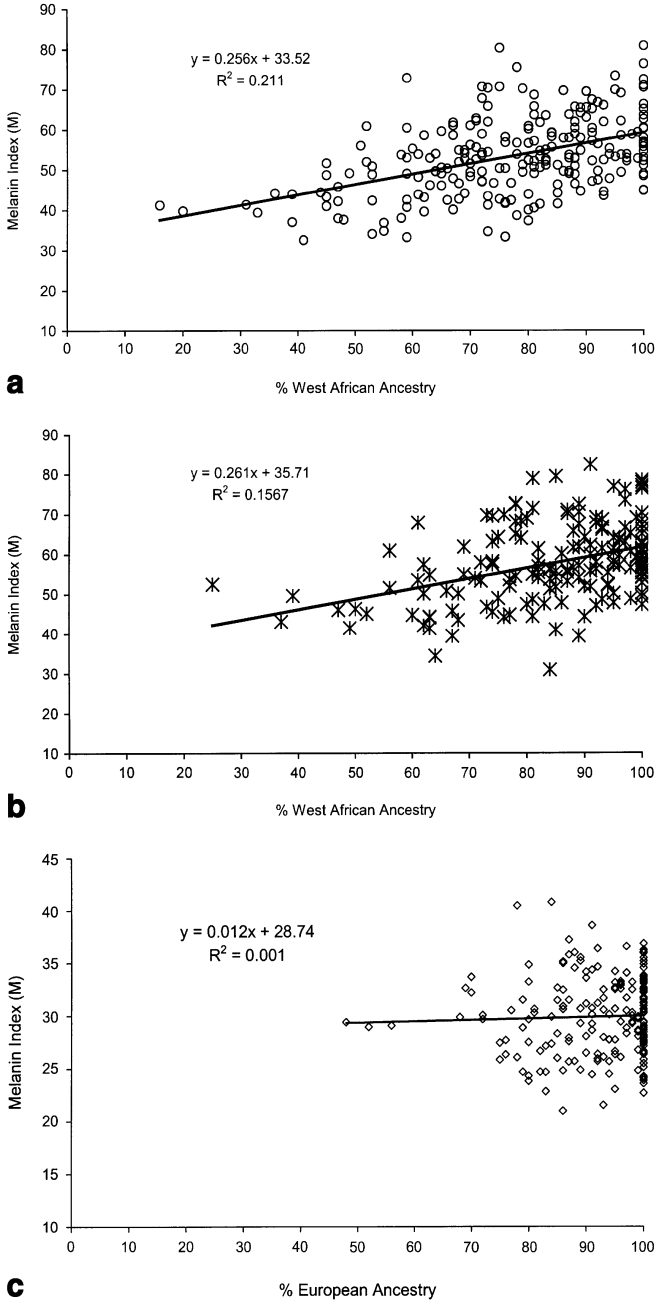

Fig. 4a–c Relationship between individual ancestry and skin pigmentation in three populations. The same data as in Fig. 3 with the samples separated into individual charts for clarity. **a** African Americans from Washington, D.C. **b** African Caribbeans from Britain. **c** European Americans from State College, Pa.

able for the current panel. Skin pigmentation also shows substantial variability (Fig. 3, y-axis), with M index values ranging between 20 and 80, but there is a clear trend towards lower pigmentation values with decreasing west African ancestry. The range in skin pigmentation is much lower among European Americans than in individuals of west African ancestry. It is important to note the wide dispersion of pigmentation values among persons classified as having 100% African ancestry by using our genetic markers (approximately from 45–82 M index units). This is probably attributable to both the error in the estimates of individual ancestry and to the natural variation in pigmentation level among the west African parental populations. Indeed, this sample includes 13 individuals born in different African countries and their pigmentation values range between 46 and 70 M index units. The range in persons showing 100% European ancestry is 20 to 40 units.

By plotting these data on independent graphs, we can investigate the data as independent samples. Figure 4a–c shows the relationship of individual ancestry and skin pigmentation in African Americans, African Caribbeans and European Americans, respectively. There is significant correlation between African individual ancestry and skin pigmentation in the two samples of primarily west African ancestry (African Americans from Washington, R^2=0.211, P<0.001; African Caribbeans from Britain, R^2=0.157, P<0.001) and the slope of the regression is similar in both cases (Fig. 4a, b). No correlation is observed between European individual ancestry and skin pigmentation in the sample with primarily European ancestry (European Americans from State College, R^2=0.001, P=0.670).

Testing the effect of candidate genes and random markers on skin pigmentation

The next question that we wanted to ask was whether any of the three AIMs that were also candidate genes for pigmentation (*TYR*, *OCA2* and *MC1R*) played a role in skin pigmentation. For the remaining 30 AIMs, there is no *a priori* reason to believe that they have an effect on skin pigmentation. However, they can be very useful for identifying and controlling for the effects of genetic structure. Additionally, given that these markers are very informative for biogeographical ancestry, they are especially powerful for detecting structure resulting from admixture. Thus, as an initial step to test for the effect of the candidate genes on pigmentation and to evaluate the extent in which the presence of genetic structure might produce false positive results, we carried out a conventional ANOVA analysis in which we tested the effect of each AIM (including the candidate genes) on skin pigmentation, with sex as a covariate. In other words, individuals were classified into the three genotypes (e.g. AA, Aa and aa) for a particular locus and the average M index for each of these groups was then compared for statistical significance. For this analysis, we included all the persons of European ancestry into one group (n=187) and the persons of African ancestry into another (n=416). In European Americans, who did not show a detectable genetic structure in our previous analysis, no significant effect was detected for any of the three candidate genes or any other ancestry informative marker, except for the Duffy gene (*FY*). For *FY*, the vast majority of European Americans were homozygotes for the allele, which is common in European populations, and only five persons were heterozygotes for the African *FY**Null allele, with no *FY**Null homozygotes. This test is, therefore, based on the comparison of only five heterozygotes with a

Table 3 Testing for an effect of single-locus genotypes on pigmentation in the combined African-American and African Caribbean sample. P-values for the test of differences in mean values for persons are classified by genotype (e.g. *1*1,*1*2, and *2*2). Markers shown in *bold* and *italics* are in or near candidate genes for pigmentation (viz. *OCA2, MC1R, TYR*)

Marker[a]	AF-EU Delta[b]	ANOVA[c]	ANOVA[d] IAE	Bayesian score test probability (% Extracted information)[e]
TYR-192	0.444	**0.000**	**0.006**	**0.0001 (16.8%)**
OCA2	0.631	**0.000**	0.249	**0.030 (19.7%)**
MC1R-314	0.350	0.552	0.731	0.666 (5.2%)
MID-575	0.121	0.491	0.219	0.683 (6.0%)
MID-187	0.370	0.086	0.958	0.743 (7.4%)
FY-NULL	0.997	**0.000**	0.979	0.389 (42.5%)
AT3	0.575	**0.030**	0.737	0.260 (13.2%)
F13B	0.641	**0.005**	0.954	0.348 (8.1%)
TSC1102055	0.430	**0.019**	0.052	0.445 (9.1%)
WI-11392	0.444	**0.038**	0.283	0.343 (5.5%)
WI-16857	0.536	0.573	0.228	0.806 (6.5%)
WI-11153	0.652	**0.000**	0.098	0.150 (11.0%)
GC	0.697	**0.046**	0.574	0.720 (16.3%)
SGC30610	0.146	**0.017**	0.202	0.284 (2.4%)
SGC30055	0.457	**0.001**	0.948	0.193 (12.4%)
WI-9231	0.017	0.647	0.157	0.459 (1.2%)
LPL	0.479	**0.000**	0.308	0.151 (11.2%)
WI-11909	0.075	0.066	0.156	0.709 (1.7%)
D11S429	0.429	**0.023**	0.171	**0.019 (11.9%)**
DRD2 *Taq*D	0.535	0.147	0.522	**0.025[f] (11.7%)**
DRD2 *Bcl*I	0.080	**0.047**	0.400	–
APOA1	0.505	**0.032**	0.673	0.060 (10.2%)
GNB3	0.463	0.223	0.542	0.858 (7.9%)
RB1	0.611	0.697	0.071	0.989 (8.8%)
WI-14319	0.185	0.322	0.136	0.086 (7.6%)
CYP19	0.037	0.547	0.565	0.310 (3.4%)
PV92	0.073	0.064	0.559	0.756 (1.9%)
WI-14867	0.448	**0.001**	0.230	0.967 (16.8%)
WI-7423	0.476	**0.003**	0.359	0.966 (30.2%)
Sb19.3	0.488	0.285	0.916	0.490 (5.8%)
CKM	0.150	0.828	0.128	0.524 (2.4%)
MID-154	0.444	0.182	0.963	0.347 (8.7%)
MID-93	0.554	0.111	0.796	0.286 (6.1%)

[a]Marker indicates the ancestry informative marker used in the test
[b]Delta is the allele frequency difference between African and European populations
[c]Analysis of variance (ANOVA) significance level where sex is the only covariate
[d]Significance level for a one-way ANOVA analysis by using individual admixture estimates (*IAE*) where the tested locus was excluded as the covariate in addition to sex
[e]Bayesian admixture mapping 1-sided probability
[f]P-value for *DRD2* TaqD-Bcl haplotype

large number of non-*FY**Null homozygotes. Contrary to expectations, these heterozygotes showed a lighter skin (P=0.019) than the average homozygotes for the common European allele.

The ANOVA analyses on the combined African-American/African-Caribbean sample are markedly different. The results of these analyses are reported in Table 3. In these persons of primary west African ancestry, 17 out of 33 markers (51.5%) are significant, including two of the candidate genes (*TYR* and *OCA2*). This result is not unexpected given the strong genetic structure that is observed in African-American populations (e.g. Parra et al. 1998; Pfaff et al. 2001). There is a significant correlation between the information content of each marker (measured as the European-African delta value shown in Table 3) and the P-values of the uncorrected ANOVA analysis (R^2= 0.426, P=0.014). To control for the confounding by variation in individual ancestry, we have also carried out an ANOVA analysis including individual ancestry as a conditioning variable (4th column of Table 3). It is clear that the inclusion of individual ancestry as a covariate eliminates most of the significant results and the only marker showing a significant effect after correction is a well-known candidate gene for pigmentation, *TYR*, which remains highly significant after the correction (P=0.006).

Finally, we used the ADMIXMAP program to test for linkage with genes underlying the ethnic difference in skin pigmentation. The last column of Table 3 shows the P-values of the score test for the linkage of each marker locus with genes underlying the ethnic difference in pigmentation. The ADMIXMAP analysis shows, in agreement with the independent ANOVA analysis previously described, highly significant evidence for linkage of skin pigmentation with *TYR* (P<0.0001). There is confirmatory evidence for linkage with two loci that flank *TYR*: D11S429 (P= 0.019) and *DRD2* (P=0.025). D11S429 is approximately 25 cM centromeric to *TYR*, wherease *DRD2* is approximately 20 cM telomeric to *TYR* (GM99-GB4 map, available at http://www.ncbi.nlm.nih.gov/genemap99/). This is a multipoint analysis, so both flanking loci contribute to the information about ancestry at the *TYR* locus. Additionally, the *OCA2* locus, another pigmentation candidate gene, is also significant (P=0.03).

Discussion

Skin pigmentation and admixture mapping

Skin pigmentation is a highly variable trait both within and among human populations, with a remarkable geographical distribution. Most probably, this distribution reflects the action of natural selection through the adaptation of populations to different environments. However, the ultimate selective factors (or combination of factors) involved have yet to be clarified (see Robins 1991 for a review of major adaptive hypotheses). Knowledge of the pigmentary system has advanced dramatically over the past decades (for reviews, see Barsh 1996; Jackson 1997; Spritz 1999; Sturm et al. 2001) and much is known regarding the genes that determine both human pigmentation disorders, such as albinism and piebald trait, and the genes that determine mouse coat-colour phenotypes. However, the genes involved in normal variation in skin pigmentation remain largely unknown. Given the large differences in skin pigmentation between populations, admixture mapping can provide important insights in understanding the genetics of the geographical variation in this trait. Pigmentation is well suited for admixture mapping for two main reasons. First, the large differences observed in pigmentation between continents make admixed populations particularly useful for mapping the genes involved. Second, constitutive pigmentation (pigmentation as measured in unexposed areas) is largely independent of environmental influences. With the exception of small sex differences that can be easily controlled for, constitutive pigmentation in adults is a stable trait. These factors make the genetics of this complex polygenic trait, in principle, much easier to disentangle than other complex traits and diseases, such as obesity, diabetes, cancer and hypertension, where there are both major environmental exposures and age effects. Whereas admixture mapping will be useful for identifying the genes involved in skin pigmentation differences between major population groups, admixture mapping is not especially useful to study genetics of variability within continental populations.

In this study, we have analysed skin pigmentation in two samples from populations primarily of west African ancestry (African Americans from Washington, D.C., and African Caribbeans from Britain) and a sample of primarily European ancestry (European Americans from State College, Pa.). As shown in Fig. 1, there are major differences between the skin colour distribution observed in persons of African and European ancestry. Interestingly, even when the two samples of African ancestry show a large overlap in the distribution of M index values, African Caribbeans, on average, are significantly darker than African Americans (average M=57.8 vs 53.4, respectively, $P<0.001$). The most likely explanation for this is the significantly lower European admixture observed in African Caribbeans. Our analysis with AIMs shows the average European ancestry in African Caribbeans is 10.2% vs 18.6% in African Americans from Washington D.C. The differences in admixture are also evident in the plot of individual ancestry in these two samples (Fig. 2a). The results of this study are in accordance with previous studies showing that the average admixture of US African Americans is higher than that in the former British Colonies in the Caribbean. A notable exception is the Gullah population from South Carolina and Georgia, for which historical, linguistic and also genetic evidence indicates a very low European admixture (Parra et al. 2001). The genetic data also indicates the presence of African and native American admixture in the European-American sample from State College, a finding that has been noted in other European-American populations (Parra et al. 1998). However, the admixture in these populations is on average low, with less than 5% estimated non-European admixture in the State College sample.

We have observed pigmentation differences between males and females in the three samples. In the two samples of primarily African ancestry, males are, on average, darker than females; this difference is significant in the African Caribbean sample ($P=0.005$) and, in the Washington D.C. sample, the P value is at a suggestive level of significance ($P=0.064$). In the European-American sample, the opposite is true, with females, on average, being darker than males (the difference is significant, $P=0.008$). However, this difference between European-American males and females is not significant with two other measures of measurements of skin pigmentation (Wagner et al. 2002). Although several studies have indicated that females are, in many populations, on average lighter than males (Byard 1981; Robbins 1991), it is difficult to assess whether previous studies have controlled for potential confounding effects, such as differences in activity patterns between sexes. Notably, there are no significant differences between the sexes in average individual ancestry in any of the samples. Given this difference in pigmentation between sexes, we have included sex as a factor when carrying out the analysis to measure the effect of the genetic markers on skin pigmentation.

Evidence exists that the genetic structure observed in the samples of primarily African ancestry is attributable to European admixture. A closer look at the associations between linked and unlinked markers clearly point to this. It is important to note that, as indicated in Table 1, some of our AIMs show notable differences in frequency between Europeans and Africans, whereas others have been selected to identify native American admixture, and some of these show smaller differences in frequency between Europeans and Africans. As expected, most of the significant comparisons between unlinked markers in persons of primarily African ancestry involve markers showing large differences in frequency between Africans and Europeans. For example, there are six significant comparisons between *FY*, which is located on chromosome 1, and markers located on other chromosomes. *FY* is the most informative marker for inferring African vs non-African admixture, because there is one allele that is almost fixed in west Africa and the alternative allele is fixed in non-African populations. All the significant comparisons between *FY*

and unlinked markers involve markers also showing very high differences in frequency between Europeans and Africans (*LPL*, WI-7423, MID-187, *APOA1*, SGC30055 and WI-11153). Moreover, the alleles that are associated with FY-Null are the alleles that are more frequent in African populations. Comparisons of *FY* with markers showing small differences in frequency between Europeans and Africans are not significant (e.g. WI-9231, WI-11909, *CYP19* and PV92).

To study the effect of the pigmentation candidate genes and random markers on skin pigmentation, we carried out the analysis independently in the European-American sample and the pooled sample of individuals of primarily African ancestry (African Americans from Washington, D.C., and African Caribbeans from Britain). The main reason for pooling together the samples of African ancestry involved the similar characteristics of the samples and the nature of the analysis. Both samples have primarily west African and European ancestry. In both cases, European gene flow has created a significant genetic structure that can be detected by various tests based on the information provided by the AIMs. As mentioned above, the African-American sample shows a greater proportion of European ancestry (which is also reflected in a lower average skin pigmentation) but this is not inconvenient for the present analysis, which is powered by the observed differences in admixture between individuals. Additionally, both samples show remarkably similar trends in the observed correlation of melanin and percentage African ancestry, as expected if the decrease in the melanin index is mainly attributable to differences in European ancestry. Thus, in order to increase the power of the analysis, we have performed the analysis in the pooled sample of persons of African ancestry and report the results of this analysis in Table 3. Furthermore, for completeness we describe below the results of the independent analysis of the African-American and African-Caribbean samples.

The presence of genetic structure, attributable to admixture or any other factor, is theoretically well recognized as being a potential confounder in association studies. The confounding effects of genetic structure because of admixture are clearly evident in our ANOVA analysis. In the sample of persons of African ancestry, more than 50% of the AIMs are prominent as having a significant effect on skin pigmentation at the 5% alpha level. Obviously, the number of significant associations has been tremendously inflated by the structure introduced by admixture. We can detect a significant effect for most of these markers because they are informative for ancestry and because individual ancestry and skin pigmentation are strongly correlated in this sample (Fig. 4). Clear evidence of the influence of admixture is provided by the correlation between the European-African difference in frequency and the *P*-values for the various markers in the uncorrected ANOVA model (R=0.426; *P*=0.014, see also Table 3). This result clearly emphasizes the importance of detecting, and controlling for, the genetic structure introduced by admixture. We have shown here that a simple ANOVA analysis, in which individual ancestry is used as a conditioning variable, can control for the effects of structure resulting from admixture. However, a better approach is one specifically designed for admixture mapping, such as the ADMIXMAP program described here. ADMIXMAP can take into account the uncertainty with which individual ancestry is estimated and can also combine the information about ancestry provided by linked markers in a multipoint analysis. The results of ADMIXMAP reported here are in close agreement with the simpler ANOVA analysis. It is important to note that the number of markers that we are currently using to infer ancestry (31–34), although large enough to detect the genetic structure present in admixed populations, is still insufficient to infer with precision individual ancestry and we estimate that 80–100 markers with the characteristics reported in Table 1 will greatly decrease the standard error of these estimates. The ultimate goal is to compile a whole-genome-spanning panel of 1000–2000 ancestry informative markers to be used in admixture mapping. With the increasing information available in SNP databases, this goal is clearly within reach. AIMs must show large differences in frequencies between the parental populations involved in the admixture process. It has been known for decades that the genetic variation between human populations is only a small percentage of the total genetic variation of our species (around 10%–15%; Lewontin 1972; Jorde et al. 2000). Markers showing large differences in frequency (delta >0.45) between major population groups are rare (around 5% of diallelic markers). However, it should be possible to select, from the millions of SNPs dispersed in the human genome, enough informative SNPs to compile a panel of AIMs to be used in admixture mapping. Recent research efforts have provided a substantial panel of markers for admixture mapping (Parra et al. 1998; Smith et al. 2001; Collins-Schramm et al. 2002).

With respect to skin pigmentation, the results of our analysis indicate that the tyrosinase gene (*TYR*) is involved in the differences of pigmentation between European and African populations. After controlling for individual ancestry, this marker still shows a highly significant effect on pigmentation (see Table 3). The results are also significant in an independent analysis of the African-American (ADMIXMAP, *P*=0.034) and the African Caribbean (ADMIXMAP, *P*=0.0002) samples. *TYR* is a well-known candidate gene for pigmentation, being a key enzyme involved in the synthesis of melanin. Variants of this gene produce oculocutaneous albinism type 1 (*OCA1*; see entry no. 203100 in OMIM database). However, to the best of our knowledge, the involvement of *TYR* in the normal variation of pigmentation has not been reported in the literature. It is important to note that the variant that we have genotyped, *TYR*-192 (dbSNP ss4387030; rs1042602), is a non-synonymous substitution (Ser to Tyr). Thus, this residue could be directly involved in the differences in melanin content observed between European and African populations. Alternatively, it could be in LD with one or more functional polymorphisms near this locus. Additionally, the two markers that flank *TYR*, viz. D11S429 and *DRD2*, are also significantly associated with pigmentation

level, providing confirmatory evidence that this is indeed a true signal.

The melanocortin receptor 1 gene (*MC1R*) is another locus that has been repeatedly associated with pigmentation, particularly with fair skin and red hair in individuals of European origin (Valverde et al. 1995; Smith et al. 1998; Rees et al. 1999; Flanagan et al. 2000; Bastiaens et al. 2001a). Variants of this gene have been also associated with sun sensitivity (Healy et al., 2000) and melanoma risk (Valverde et al. 1996; Kennedy et al. 2001; Bastiaens et al. 2001b). However, the *MC1R* variant that we have analysed (dbSNP ss4387027; rs2228478) does not show any significant effect on pigmentation in the ANOVA or ADMIXMAP analysis, either in the pooled sample or independently in the African-American or African-Caribbean samples. This variant has a difference in frequency of 35% between European and African populations, so we would expect that, if this polymorphism or any other linked variant within the *MC1R* gene causes differences in pigmentation between these populations, it should be reflected in a significant result in this analysis. Therefore, it seems that, even though this locus has been widely recognized as being involved in the variation in skin colour within populations of European ancestry, it probably does not play as significant role as other genes in pigmentation variation between African and European populations.

The third polymorphism located within a pigmentation candidate gene that we have characterized in these samples, which also provides information about ancestry (delta Eur-Afr =63%), is a variant in the *OCA2* gene (dbSNP: ss4387028; rs1800404). Mutations in the *OCA2* gene (or *P* gene) produce the most common form of albinism (oculocutaneous albinism type 2; OMIM entry no. 203200). It is also known that OCA2 is a key determinant of melanin synthesis, because of its role in mediating the pH environment within the melanosome (Sturm et al. 2001; Brilliant 2001). The *OCA2* marker tested shows a significant effect (P=0.03) on skin pigmentation in the combined African-American and African-Caribbean population in the ADMIXMAP analysis. In terms of the ANOVA analyses, this variant is significant when not correcting for admixture or when correcting with the individual ancestry estimate excluding the *OCA2* locus (data not shown). However, when correcting for the individual ancestry estimate using all markers, we have found that the effect is not significant (column 4, Table 3). Indeed, in an independent analysis of the two samples of African ancestry, we have observed that the *OCA2* shows a significant effect after correction for ancestry in the African-American sample from Washington, D.C. (P=0.045, with individual ancestry as a covariate) but the analysis in the African-Caribbean sample shows no significant effect for this gene. Similarly, the Bayesian analyses showed significant results for the African-American sample and not the African-Caribbean sample.

Conclusions

Admixture can be both a blessing and a curse for gene mapping efforts: a blessing, because the process of admixture between previously isolated populations creates new LD that can extend for long distances, so that the density of markers required for mapping is much lower than that number needed in other populations, and a curse, because the admixture process can also introduce genetic structure that, if not controlled for, can greatly affect case-control, measured genotype and mapping efforts by producing false positive results. The extent of structure introduced by admixture depends on several factors, including the extent of admixture, the dynamics of the admixture process and the social structure in the population that is related to ancestry (Pfaff et al. 2001). Fortunately, there are several ways in which it is possible to detect structure attributable to admixture and to control for its effects. Many persons who would describe themselves as African-American, Hispanic or European-American trace their ancestry to populations on different continents that had been previously isolated for many generations, coming back together to the New World sometime within the past five centuries. Within each of these groups, particularly African Americans and Hispanics, and to a lesser extent European Americans, there is variation in individual ancestry levels or population structure that can create important problems for gene mapping. Therefore, in these populations, it is critical to control for this structure to avoid being misled by false positive linkage results (Pfaff et al. 2002).

In this report, we present a panel of AIMs that has high information content about west African, European and native American ancestry and that can be used both to detect and to control for the population structure that results from admixture. We have also demonstrated the way in which to proceed with an admixture mapping analysis by using skin pigmentation as a model phenotype. Given the limited number of AIMs currently available, we have restricted our effort to analysing the effect of three known candidate genes for pigmentation (*TYR*, *OCA2* and *MC1R*) and present evidence that the tyrosinase gene is strongly associated with and probably plays a role in the differences in pigmentation observed between African and European populations. The development of a panel of AIMs spanning the whole genome at distances of 2–3 cM is well within reach. The application of this panel of markers to admixture mapping can and should provide important insights about the genes involved in traits and diseases showing prevalence differences between major population groups, such as obesity, type 2 diabetes, prostate cancer, dementia and hypertension.

Acknowledgments We thank the volunteers who gave their time and DNA for this research, and Mark Jobling for assistance with sample preparation. We also acknowledge the support of the Penn State University General Clinical Research Center (M01 RR10732). This work was also supported in part by grants from the NIH/NHGRI (HG002154) to M.D.S., from NIH/NIMH (MH60343) to P.M.M., a Howard University New Faculty Award to R.A.K., and by the British Broadcasting Corporation to A.B.

References

Akey JM, Sosnoski D, Parra E, Dios S, Hiester K, Su B, Bonilla C, Jin L, Shriver MD (2001a) Melting curve analysis of SNPs (McSNP): a simple gel-free low-cost approach to SNP genotyping and DNA fragment analysis. Biotechniques 30:358–367

Akey JM, Wang H, Xiong M, Wu H, Liu W, Shriver MD, Jin L (2001b) Interaction between the melanocortin-1 receptor and P genes contributes to inter-individual variation in skin pigmentation phenotypes in a Tibetan population. Hum Genet 108: 516–520

Barsh GS (1996) The genetics of pigmentation: from fancy genes to complex traits. Trends Genet 8:299–305

Bastiaens M, Huurne J ter, Gruis N, Bergman W, Westendorp R, Vermeer BJ, Bouwes Bavinck JN (2001a) The melanocortin-1-receptor gene is the major freckle gene. Hum Mol Genet 10: 1701–1708

Bastiaens MT, Huurne JA ter, Kielich C, Gruis NA, Westendorp RG, Vermeer BJ, Bavinck JN, Leiden Skin Cancer Study Team (2001b) Melanocortin-1 receptor gene variants determine the risk of nonmelanoma skin cancer independently of fair skin and red hair. Am J Hum Genet 68:884–894

Brilliant MH (2001) The mouse p gene (pink-eyed dilution) and the human P gene, oculocutaneous albinism (OCA2) and melanosomal pH. Pigment Cell Res 14:86–93

Briscoe D, Stephens JC, O'Brien SJ (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. J Hered 85:59–63

Byard PJ (1981) Quantitative genetics of human skin color. Yearb Phys Anthrop 24:123–137

Chakraborty R, Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci USA 85:9119–9123

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 linage disequilibrium. Am J Hum Genet 70:737–750

Dean M, Stephens J C, Winkler C, Lomb DA, Ramsburg M, Boaze R, Stewart C, Charbonneau L, Goldman D, Albaugh BJ, Goedert JJ, Beasley RP, Hwang L-Y, Buchbinder S, Weedon M, Johnson P, Eichelberger M, O'Brien SJ (1994) Polymorphic admixture typing in human ethnic populations. Am J Hum Genet 55: 788–808

Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA (2000) The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat Genet 25:320–323

Flanagan N, Healy E, Ray A, Philips S, Todd C, Jackson IJ, Birch-Machin MA, Rees JL (2000) Pleiotropic effects of the melanocortin 1 receptor (MC1R) gene on human pigmentation. Hum Mol Genet 9:2531–2537

Hanis C L, 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

Healy E, Flannagan N, Ray A, Todd C, Jackson IJ, Matthews JN, Birch-Machin MA, Rees JL (2000) Melanocortin-1-receptor gene and sun sensitivity in individuals without red hair. Lancet 355:1072–1073

Jackson IJ (1997) Homologous pigmentation mutations in human, mouse and other model organisms. Hum Mol Genet 6:1613–1624

Jorde LB (2000) Linkage disequilibrium and the search for complex disease genes. Genome Res 10:1435–1444

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer M (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 66:979–988

Kaessmann H, Zollner S, Gustafsson AC, Wiebe V, Laan M, Lundeberg J, Uhlén M, Paabo S (2002) Extensive linkage disequilibrium in small human populations in Eurasia. Am J Hum Genet 70:673–685

Kanetsky PA, Swoyer J, Panossian S, Holmes R, Guerry D, Rebbeck TR (2002) A polymorphism in the Agouti signaling protein gene is associated with human pigmentation. Am J Hum Genet 70:770–775

Kennedy C, Huurne J ter, Berkhout M, Gruis N, Bastiaens M, Bergman W, Willemze R, Bavinck JN (2001) Melanocortin 1 receptor (MC1R) gene variants are associated with an increased risk for cutaneous melanoma which is largely independent of skin type and hair color. J Invest Dermatol 117:294–300

Kittles RA, Chen W, Panguluri RK, Ahaghotu C, Jackson A, Adebamowo CA, Griffin R, Williams T, Ukoli F, Adams-Campbell L, Kwagyan J, Isaacs W, Freeman V, Dunston GM (2002) CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification? Hum Genetics 110:553–560

Lautenberger JA, Stephens JC, O'Brien SJ, Smith MW (2000) Significant admixture linkage disequilibrium across 30 cM around the FY locus in African Americans. Am J Hum Genet 66:969–978

Lewontin RC (1972) The apportionment of human diversity. In: Dobzhansky TH, Hecht MK, Steere WC (eds) Evolutionary biology, vol 6. Appleton-Century-Crofts, New York, pp 381–398

Long JC (1991) The genetic structure of admixed populations. Genetics 127:417–428

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 (1998) Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. Am J Hum Genet 63:241–251

McKeigue PM, Carpenter J, 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

Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18:83–90

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 Anthrop 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

Pfaff CL, Kittles RA, Shriver MD (2002) Adjusting for population structure in admixed populations. Genet Epidemiol 22:196–201

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

Rees JL, Birch-Machin M, Flanagan N, Healy E, Phillips S, Todd C (1999) Genetic studies of the human melanocortin-1 receptor. Ann N Y Acad Sci 885:134–42

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Robins AH (1991) Biological perspectives on human pigmentation. Cambridge University Press, Cambridge

Shriver MD, Parra EJ (2000) Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry. Am J Phys Anthrop 112:17–27

Smith R, Healy E, Siddiqui S, Flanagan N, Steijlen PM, Rosdahl I, Jacques JP, Rogers S, Turner R, Jackson IJ, Birch-Machin MA, Rees JL (1998) Melanocortin 1 receptor variants in an Irish population. J Invest Dermatol 111:119–122

Smith MW, Lauternberger JA, Doo Shine H, Chreetien J, Shrestha S, Gilbert DA, O'Brien SJ (2001) Markers for mapping by admixture linkage disequilibrium in African-American and Hispanic populations. Am J Hum Genet 69:1080–1094

Spritz RA (1999) Multi-organellar disorders of pigmentation: intracellular traffic jams in mammals, flies and yeast. Trends Genet 15:337–340

Stephens JC, Briscoe D, O'Brien SJ (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am J Hum Genet 55:809–824

Sturm RA, Teasdale RD, Box NF (2001) Human pigmentation genes: identification, structure and consequences of polymorphic variation. Gene 277:49–62

Taillon-Miller P, Bauer-Sardiña I, Saccone NL, Putzel J, Laitinen T, Cao A, Kere J, Pilia G, Rice J, Kwok P-Y (2000) Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. Nat Genet 25:324–328

Valverde P, Healy E, Jackson I, Rees JL, Thody AJ (1995) Variants of the meloancyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat Genet 11:328–330

Valverde P, Healy E, Sikkink S, Haldane F, Thody AJ, Carothers A, Jackson IJ, Rees JL (1996) The Asp84Glu variant of the melanocortin 1 receptor (MC1R) is associated with melanoma. Hum Mol Genet 5:1663–1666

Wagner JK, Siccard C, Norton HL, Parra EJ, Shriver MD (2002) Skin responses to ultraviolet light: effects of constitutive pigmentation, sex, and ancestry. Pigment Cell Res (in press)

Wilson JF, Goldstein DB (2000) Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. Am J Hum Genet 67:926–935

Wright AF, Carothers AD, Pirastu M (1999) Population choice in mapping genes for complex diseases. Nat Genet 23:397–404

Wu H, Wang H, Li H, Akey J, Xiao F, Ke Y, Xu H, Xiao J, Lu D, Parra E, Shriver M, Xiong M, Barton SA, Hewett-Emmett D, Liu W, Jin L (2001) Skin reflectance in the Han Chinese and Tibetan populations. Hum Biol 73:461–466

Ye J, Parra EJ, Sosnoski D, Hiester K, Underhill P, Shriver MD (2002) Melting curve SNP (McSNP) genotyping: a useful approach to SNP genotyping in forensic science. J Forens Sci 47: 593–600